Weighted Sum Rate Maximization for the MIMO-Downlink Using a Projected Conjugate Gradient Algorithm

Ronald Böhnke and Karl-Dirk Kammeyer
Department of Communications Engineering
University of Bremen
Otto-Hahn-Allee, 28359 Bremen, Germany
e-mail: {boehnke, kammeyer}@ant.uni-bremen.de

***Abstract*— The maximization of a weighted sum of data rates is an essential point in cross-layer based resource allocation. Several algorithms have been proposed in the literature to solve this problem for the downlink of a multiple antenna system employing dirty paper precoding at the base station. However, they all suffer from a relatively slow convergence if the true number of objective function evaluations is taken into account. In this paper, an improved conjugate gradient method is presented, that takes the power constraint into account in the calculation of the search direction. Its superior convergence properties compared to existing approaches are verified by Monte-Carlo simulations for various scenarios.**

I. Introduction

In order to assure certain quality of service parameters in the downlink of a wireless communication system, the scheduler has to take information from different layers of the protocol stack into account. This task is usually performed in two stages: First, priorities are assigned to each user based on the type of application and the current queue states. Afterwards, resources are allocated by maximizing the weighted sum of achievable data rates, which requires some kind of channel knowledge at the base station. E.g., using the queue lengths as priority measures minimizes the risk of buffer overflows [1], while setting the weighting factors equal to the inverse average throughput leads to the proportional fair scheduling policy [2]. Another more sophisticated choice aims at the minimization of the average packet delay [3]. In this paper, we will focus on the second part of the described cross-layer scheduling approach, namely the optimal resource allocation for arbitrarily given weights if both the base station and the mobile terminals are equipped with multiple antennas.

In the following section, the system model is introduced and the weighted sum rate is derived. Section III contains a comprehensive overview of existing algorithms for the solution of this and related optimization problems. The discussion of their pros and cons serves as a motivation for the projected conjugate gradient method proposed in Section IV. Numerical results confirming the fast convergence of our new approach for arbitrary system parameters are provided in Section V, and concluding remarks can be found in Section VI.

II. Problem Statement

Consider a multiple-input multiple-output (MIMO) system with N_U users, where the central base station is equipped with N_B antennas and each mobile terminal has N_M antennas. It was demonstrated in [4] that in the downlink all feasible rate tuples can be achieved by so-called dirty paper precoding. Furthermore, the capacity region is identical to that of a dual multiple access channel with the same total transmit power [5]. We will focus on this equivalent uplink problem, as it is in general much easier to handle, so the system model is given by

$$\boldsymbol{y} = \sum_{u=1}^{N_\mathrm{U}} \boldsymbol{H}_u \boldsymbol{x}_u + \boldsymbol{n} = \sum_{u=1}^{N_\mathrm{U}} \boldsymbol{H}_u \boldsymbol{T}_u \boldsymbol{d}_u + \boldsymbol{n} \ , \qquad (1)$$

where $\boldsymbol{y} \in \mathbb{C}^{N_\mathrm{B}}$ is the vector of receive signals at the base station, $\boldsymbol{H}_u \in \mathbb{C}^{N_\mathrm{B} \times N_\mathrm{M}}$ and $\boldsymbol{x}_u \in \mathbb{C}^{N_\mathrm{M}}$ are the dual uplink channel matrix and transmit vector of user u, respectively, and $\boldsymbol{n} \in \mathbb{C}^{N_\mathrm{B}}$ represents additive white Gaussian noise with covariance matrix $\mathrm{E}\{\boldsymbol{n}\boldsymbol{n}^\mathrm{H}\} = \boldsymbol{I}$; furthermore, the transmit signal $\boldsymbol{x}_u$ follows from multiplying the uncorrelated unit variance Gaussian symbols in vector $\boldsymbol{d}_u$ with the filter matrix $\boldsymbol{T}_u \in \mathbb{C}^{N_\mathrm{M} \times N_\mathrm{M}}$. Denoting the transmit covariance matrices as $\boldsymbol{Q}_u = \mathrm{E}\{\boldsymbol{x}_u \boldsymbol{x}_u^\mathrm{H}\} = \boldsymbol{T}_u \boldsymbol{T}_u^\mathrm{H}$, the power constraint can be expressed as

$$\sum_{u=1}^{N_\mathrm{U}} \mathrm{tr}(\boldsymbol{Q}_u) = \sum_{u=1}^{N_\mathrm{U}} \|\boldsymbol{T}_u\|_\mathrm{F}^2 \leq P \ . \qquad (2)$$

Dirty paper coding at the base station corresponds to successive interference cancelation in the dual uplink, where a user is not interfered by previous ones. Assume that the users are decoded in the order N_U to 1. Then, the receive covariance matrix for the k-th user becomes

$$\boldsymbol{\Phi}_k = \boldsymbol{I} + \sum_{u=1}^{k} \boldsymbol{H}_u \boldsymbol{Q}_u \boldsymbol{H}_u^\mathrm{H} = \boldsymbol{\Phi}_{k-1} + \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^\mathrm{H} \ , \qquad (3)$$

where $\boldsymbol{\Phi}_{k-1}$ with $\boldsymbol{\Phi}_0 = \boldsymbol{I}$ represents the effective noise including interference from subsequent users. The corresponding data rate is given by

$$R_k = \log_2 \det(\boldsymbol{\Phi}_k) - \log_2 \det(\boldsymbol{\Phi}_{k-1}) \ . \qquad (4)$$

Note that the term containing $\boldsymbol{\Phi}_k$ appears in both R_k and R_{k+1} with different sign, so using (4) the weighted sum rate may be written as

$$\sum_{k=1}^{N_\mathrm{U}} w_k R_k = \sum_{k=1}^{N_\mathrm{U}} w_k \left(\log_2 \det(\boldsymbol{\Phi}_k) - \log_2 \det(\boldsymbol{\Phi}_{k-1})\right) \quad (5)$$

$$= \sum_{k=1}^{N_\mathrm{U}} (w_k - w_{k+1}) \log_2 \det(\boldsymbol{\Phi}_k) \quad (6)$$

with $w_{N_\mathrm{U}+1} = 0$. For notational simplicity, we will use the abbreviation $\Delta_k = w_k - w_{k+1}$ in the following.

In order to maximize (6) the users must be sorted according to their priorities such that $w_1 \geq \ldots \geq w_{N_\mathrm{U}}$ and consequently $\Delta_k \geq 0 \; \forall k$, which will be assumed in the following without loss of generality [6]. For the optimal decoding order, the weighted sum rate is a concave function with respect to the covariances $\boldsymbol{Q} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_{N_\mathrm{U}}]$, and the feasible set of positive semidefinite matrices fulfilling the total power constraint (2) is convex. Hence, the optimization problem could be solved by standard interior point methods [7], but the possibly large number of parameters calls for more efficient algorithms.

III. Existing Algorithms

For the special case of equal priorities, (6) degenerates to the pure sum rate. In [8], an iterative waterfilling approach based on the eigenvalue decomposition of effective channel matrices was presented for the uplink with individual power constraints per user. Unfortunately, updating all users simultaneously may fail to converge if only the total transmit power is fixed in each iteration. In [9], two variants of a globally convergent cyclic coordinate ascent algorithm are proposed, where transmit covariance matrices are averaged over the current and previous iterations. The speed of convergence was significantly increased by a simple approximate line search in [10]. Instead of that, the covariance matrices of randomly selected user pairs are optimized in [11], while in [12] users are successively updated for fixed waterfilling level, which is adjusted in an outer loop until the desired total transmit power is reached.

For arbitrary weight factors, the problem gets much more involved. The basic concepts of [9] and [10] are adopted in [13] for mobile terminals with a single antenna, but now a function has to be inverted numerically to obtain the transmit powers for a given water level. Similarly, the principle of [12] is used in [14] for a MISO-OFDM system. The whole algorithm consists of three nested loops, where each has to be repeated until a certain accuracy is reached. Furthermore, if the power of user u is modified, then all inverse receive covariance matrices $\boldsymbol{\Phi}_k^{-1}$ for $k \geq u$ must be recalculated, so from this point of view it is better to update all users at the same time. Nevertheless, these waterfilling-based procedures are supposed to be faster than the steepest descent method of [15], which can be used for general MIMO systems, though. Here, the gradient of the weighted sum rate with respect to the transmit covariance matrices is determined, and a line search is performed in the direction of the eigenvector corresponding to the largest eigenvalue among all users. Recently, a projected conjugate gradient algorithm was proposed in [6], that was originally developed for pure sum rate maximization in [16]. The Fletcher-Reeves [17] search direction is used without taking the power constraint into account. It will be demonstrated in Section IV that this is a major drawback.

Algorithm 1 Conjugate Gradient with Projection

1: Initialize $\boldsymbol{T}_u = \sqrt{\frac{P}{N_\mathrm{U} N_\mathrm{M}}} \boldsymbol{I} \; \forall u$, $\boldsymbol{S} = \boldsymbol{0}$, $\rho = 1$, $\alpha = 1$
2: **repeat**
3:   Store $\boldsymbol{T}_\mathrm{old} = \boldsymbol{T}$, $\boldsymbol{S}_\mathrm{old} = \boldsymbol{S}$, $\rho_\mathrm{old} = \rho$
4:   Calculate gradient $\boldsymbol{G} = [\boldsymbol{G}_1, \ldots, \boldsymbol{G}_{N_\mathrm{U}}]$
5:   Normalize gradient $\bar{\boldsymbol{G}} = \sqrt{\frac{P}{\|\boldsymbol{G}\|_\mathrm{F}^2}} \boldsymbol{G}$
6:   Project gradient $\check{\boldsymbol{G}} = \bar{\boldsymbol{G}} - \frac{\mathrm{tr}(\boldsymbol{T}^H \check{\boldsymbol{G}})}{\mathrm{tr}(\boldsymbol{T}^H \boldsymbol{T})} \boldsymbol{T}$
7:   Calculate Frobenius norm $\rho = \left\|\check{\boldsymbol{G}}\right\|_\mathrm{F}^2$
8:   Update search direction $\boldsymbol{S} = \check{\boldsymbol{G}} + \frac{\rho}{\rho_\mathrm{old}} \boldsymbol{S}_\mathrm{old}$
9:   Step in search direction $\tilde{\boldsymbol{T}} = \boldsymbol{T}_\mathrm{old} + \alpha \boldsymbol{S}$
10:   Normalize transmit filters $\boldsymbol{T} = \sqrt{\frac{P}{\|\tilde{\boldsymbol{T}}\|_\mathrm{F}^2}} \tilde{\boldsymbol{T}}$
11:   **if** no (sufficient) improvement **then**
12:     Decrease step size α (and set $\boldsymbol{S} = \check{\boldsymbol{G}}$)
13:     Go to step 9
14:   **end if**
15: **until** desired accuracy reached

The orthogonal projection of the transmit covariance matrices onto the feasible set requires an eigenvalue decomposition for each user. A very simple projected gradient method based on the transmit filters $\boldsymbol{T} = [\boldsymbol{T}_1, \ldots, \boldsymbol{T}_{N_\mathrm{U}}]$ was used in [18] for the related weighted sum mean square error minimization. The projection onto the power constraint now simplifies to a mere scaling, and, in contrast to the transmit covariance matrices, $\boldsymbol{T}_u$ does not have to be positive semidefinite. However, after this change of variables, the objective function (6) is not concave anymore, but it was argued in [19] that every stationary point corresponds to a global optimum. This is also exploited in the alternating optimization approach [20], whose complexity is roughly cubic in the number of users, though.

IV. Proposed Algorithm

In this section, we propose a projected conjugate gradient algorithm for the optimization of transmit filter matrices. The whole procedure is outlined in Algorithm 1, where the iteration index has been omitted in order to simplify notation. First, the gradient $\boldsymbol{G}$ is calculated. Using elementary matrix calculus, it can be shown that the partial derivative of the weighted sum rate (6) with respect to $\boldsymbol{T}_u^\mathrm{H}$ is proportional to

$$\boldsymbol{G}_u = \boldsymbol{H}_u^\mathrm{H} \left(\sum_{k=u}^{N_\mathrm{U}} \Delta_k \boldsymbol{\Phi}_k^{-1} \right) \boldsymbol{H}_u \boldsymbol{T}_u = \boldsymbol{H}_u^\mathrm{H} \boldsymbol{\Psi}_u \boldsymbol{H}_u \boldsymbol{T}_u \; , \quad (7)$$

where the cumulative sum $\boldsymbol{\Psi}_u = \boldsymbol{\Psi}_{u+1} + \Delta_u \boldsymbol{\Phi}_u^{-1}$ can be computed in a recursive manner. Likewise, the matrix inversion lemma yields the update equation

$$\boldsymbol{\Phi}_k^{-1} = \boldsymbol{\Phi}_{k-1}^{-1} - \boldsymbol{B}_k \left(\boldsymbol{I} + \boldsymbol{A}_k^\mathrm{H} \boldsymbol{B}_k\right)^{-1} \boldsymbol{B}_k^\mathrm{H} \quad (8)$$

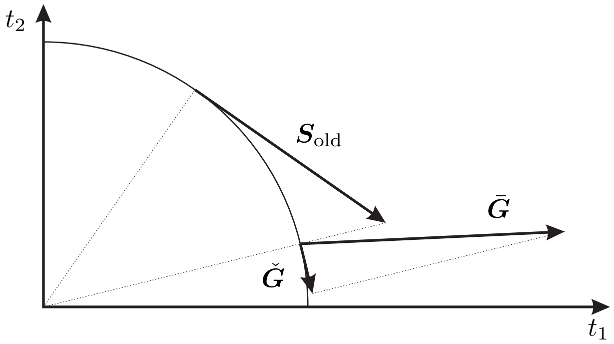

Fig. 1. Illustration of the gradient projection

with $\boldsymbol{A}_k = \boldsymbol{H}_k \boldsymbol{T}_k$ and $\boldsymbol{B}_k = \boldsymbol{\Phi}_{k-1}^{-1} \boldsymbol{A}_k$. For the important case where the base station has more antennas than the mobiles, (8) is much more efficient than directly inverting $\boldsymbol{\Phi}_k$.

The normalization of the gradient ensures that the speed of convergence becomes more or less independent of P, but the key feature is the projection of the gradient before the deflection. This is illustrated in Figure 1 for a simple example with two single-antenna terminals. In [6], the conjugate gradient $\boldsymbol{S}$ is calculated exactly as in the unconstrained case based on $\boldsymbol{G}$. However, for the constrained optimization problem considered here, the gradient is usually not zero in the optimum, but orthogonal to the boundary of the feasible set. As a consequence, the old search direction $\boldsymbol{S}_{\text{old}}$ always significantly contributes to the new one, and the algorithm may even try to step out of the optimum again. This undesirable behavior is prevented by taking the tangential component of the gradient $\check{\boldsymbol{G}}$ instead. If the projected gradients point into similar directions in consecutive iterations, they add up and the effective step size is increased. On the other hand, the oscillations around the optimum often encountered in pure gradient ascent algorithms are largely avoided.

Once the search direction is found, the transmit filters are updated and normalized such that the power constraint is fulfilled. An approximate line search could be performed at this point, e.g. according to Armijo's rule as in [6], but it usually suffices to reduce the step size α by a factor if the weighted sum rate does not increase. Additionally, it may be advantageous to reset the search direction to the projected gradient in this case.

A. Optimization of Covariance Matrices

With some modifications, the described algorithm can also be used for the optimization of the transmit covariance matrices $\boldsymbol{Q}$. Then, the gradient simplifies to $\boldsymbol{G}_u = \boldsymbol{H}_u^{\text{H}} \boldsymbol{\Psi}_u \boldsymbol{H}_u$, and the normalization should be such that the sum of $\operatorname{tr}\big(\bar{\boldsymbol{G}}_u\big)$ over all active users equals P; note that with (7), switched off users automatically do not contribute to the gradient when considering transmit filters. The projection of the gradient now corresponds to $\check{\boldsymbol{G}}_u = \bar{\boldsymbol{G}}_u - \mu \boldsymbol{I}$, where the constant μ is chosen such that the diagonal elements of all $\check{\boldsymbol{G}}_u$ (again only for active users) sum up to zero. Hence, the total transmit power does not change when adding $\check{\boldsymbol{G}}$ to the covariance matrices $\boldsymbol{Q}$. The proper choice of the deflection coefficient ρ is not so obvious. The best results were obtained by summing up the absolute values of the diagonal elements of $\check{\boldsymbol{G}}_u$, which is motivated by the relation $\|\boldsymbol{T}_u\|_{\text{F}}^2 = \operatorname{tr}(\boldsymbol{Q}_u)$. Unfortunately, the projection of $\tilde{\boldsymbol{Q}}$ onto the feasible set is much more involved than before, as the transmit covariance matrices additionally have to be positive semidefinite. According to [6], the eigenvalue decomposition $\tilde{\boldsymbol{Q}}_u = \boldsymbol{V}_u \boldsymbol{\Lambda}_u \boldsymbol{V}_u^{\text{H}}$ has to be calculated to this end, which leads to

$$\boldsymbol{Q}_u = \boldsymbol{V}_u \max(\boldsymbol{\Lambda}_u - \nu \boldsymbol{I}, 0)\, \boldsymbol{V}_u^{\text{H}}\ , \tag{9}$$

where the maximum is taken element-wise, and ν has to be determined iteratively such that the power constraint is fulfilled.

V. Numerical Results

The performance of the proposed conjugate gradient algorithm with projection (CGP) was analyzed for various system parameters. The gradient projection method (GP) from [18] using the search direction $\boldsymbol{S} = \bar{\boldsymbol{G}}$ serves as a benchmark. As a measure for the complexity, we use the number of transmit filter or covariance matrix updates per user until the relative error between the current and the true weighted sum rate is smaller than 10^{-4}, because this always necessitates a recalculation of the inverse receive covariance matrices $\boldsymbol{\Phi}_k^{-1}$. The channel matrices are assumed to contain uncorrelated complex Gaussian entries with unit variance, and the priorities are chosen as uniformly distributed integers between 1 and 10. Some results are shown in Figure 2. It can be observed that the proposed CGP method features excellent convergence properties irrespective of the number of users, mobile antennas, and total transmit power, in particular when applied to the transmit filters $\boldsymbol{T}$. Especially at small signal to noise ratios, the implicit adaptation of the step size resulting from the conjugate search directions leads to a significant speed-up. The optimization of the transmit covariance matrices is mostly attractive for single-antenna mobiles due to the required eigenvalue decomposition for $N_{\text{M}} > 1$, but the performance of both variants is quite similar in this case. Interestingly, the simple gradient algorithm is slightly better than CGP in Figure 2c. This happens because for $N_{\text{U}} N_{\text{M}} \leq N_{\text{B}}$ the equal power initialization is already close to optimal, and using the projected gradient the CGP method may step over the optimum in the first iteration. As mentioned before, this could be avoided by an approximate line search, but the additional complexity does not pay off in general.

VI. Conclusion

A new algorithm for the maximization of the weighted sum rate in the MIMO-downlink was presented. In contrast to a recently published conjugate gradient approach, a projection is performed before the deflection of the search direction. The method was applied to the optimization of both transmit covariance and filter matrices, where the latter converged faster in most cases and has the advantage that no eigenvalue decomposition is required.

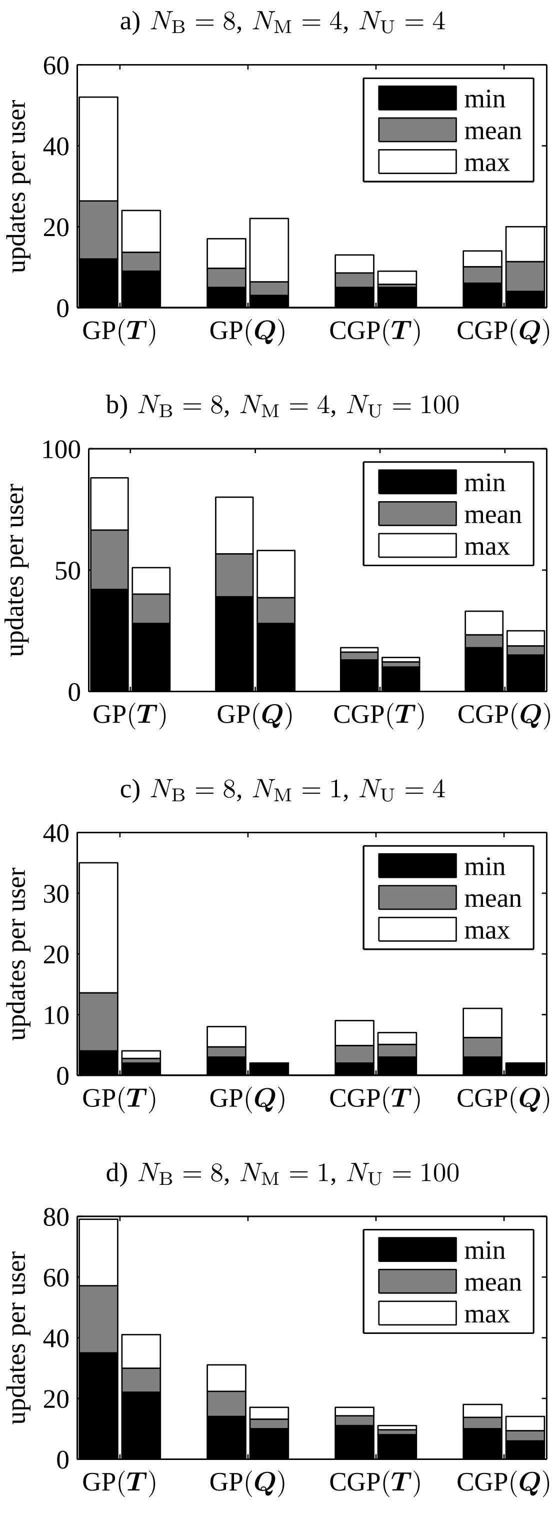

Fig. 2. Convergence behavior of the optimization algorithms for different system parameters. The left bars correspond to $P = 1$ and the right ones to $P = 100$, respectively.

REFERENCES

[1] 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.

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

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

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

[8] W. Yu, W. Rhee, S. Boyd, and J. 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. 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. Kammeyer, “Fast Sum Rate Maximization for the Downlink of MIMO-OFDM Systems,” in *Proc. Canadian Workshop on Information Theory*, Montreal, Canada, June 2005.

[11] 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.

[12] 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.

[13] 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.

[14] ——, “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.

[15] 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.

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

[17] 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.

[18] 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.

[19] R. Hunger, D. 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.

[20] 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.