

FAST POWER MINIMIZATION WITH QOS CONSTRAINTS IN MULTI-USER MIMO DOWNLINKS

Quentin H. Spencer, A. Lee Swindlehurst *
Dept. of Electrical & Computer Engineering
Brigham Young University
459 CB, Provo, UT 84602

Martin Haardt †
Communications Research Laboratory
Ilmenau University of Technology
P.O. Box 100565, D-98684 Ilmenau, Germany

ABSTRACT

In the downlink of a multi-user MIMO (Multiple Input Multiple Output) communication system where each user has an arbitrary QoS requirement, intelligent algorithms are needed to choose transmit vectors. Here we present a new method of choosing transmit vectors that minimizes total transmitted power. The approach is based on previous iterative interference balancing algorithms, but it is initialized by applying a “block-diagonalization” algorithm that helps improve convergence speed. When the channel supports multiple data streams per user, power is distributed among the data streams by bit-loading using the channel gains derived from the block-diagonalization step. The result is a solution which is not guaranteed to converge to the global optimum, but will reach a solution that is either optimal or near-optimal with high probability and at minimal computational cost.

1. INTRODUCTION

There has been considerable recent attention to the problem of multi-user MIMO (Multiple Input Multiple Output) communications systems, where a base station array is used to communicate with mobile stations that also have arrays. In the downlink of such systems where advance channel information is available, it is possible to achieve spatial multiplexing of the users by using intelligent algorithms to choose the base station transmit vectors. This is particularly important since many current and future consumer services have asymmetric bandwidth requirements, where the downlink must be able to provide more bandwidth than the uplink. Two closely linked optimization problems can be considered in this scenario. The first is maximizing the sum capacity of the entire system subject to a power constraint. The second problem, which is the focus of this paper, is to mini-

mize transmit power subject to meeting a minimum Quality of Service (QoS) for each user.

In an effort to simplify the receivers as much as possible, we make the common assumption that no multi-user detection (MUD) is used at the receivers, and all interference due to signals intended for other users is therefore treated as noise. So far, there are two general ways in which this problem has been addressed. First, iterative interference balancing strategies have been used to generate a set of transmit vectors that satisfy the given constraints [1, 2, 3]. These approaches require a certain amount of computation, but result in robust solutions. The second option is non-iterative “block-diagonalization” [4]. This approach is computationally inexpensive, near-optimal at high SNR, and can be easily adapted to various problems. However, in some situations it is sensitive to channel estimation errors. In this paper, we combine the two approaches by using the block-diagonalization solution to find a good initialization point for the iterative algorithms. We show that using this initialization point results in faster convergence than other initializations, and the resulting solutions are at or near the global optimum with high probability.

2. PROBLEM DEFINITION

We assume that the channel for user j is narrowband and quasi-static. The base station has n_T antennas, and user j has n_{R_j} antennas, so that the channel from the base to the particular user can be represented by a $n_{R_j} \times n_T$ matrix \mathbf{H}_j . Let \mathbf{M}_j (of dimension $n_T \times m_j$) represent the m_j transmit vectors for user j , and the vector \mathbf{d}_j represent the m_j data symbols transmitted, so that the desired component of the signal received by user j can be represented by $\mathbf{H}_j \mathbf{M}_j \mathbf{d}_j$. Taking into account the signals simultaneously transmitted to all other users, the total received signal \mathbf{x}_j for user j is:

$$\mathbf{x}_j = \mathbf{H}_j \mathbf{M}_j \mathbf{d}_j + \sum_{i \neq j} \mathbf{H}_i \mathbf{M}_i \mathbf{d}_i + \mathbf{n}_j, \quad (1)$$

where the vector \mathbf{n}_j is assumed to be spatially white with power σ_n^2 . At the receiver, an estimate of the data vector $\hat{\mathbf{d}}$

*Work supported by the National Science Foundation under Wireless Initiative Grant CCR-9979452 and Information Technology Research Grant CCR-0081476.

†Work supported by the Thuringian Ministry of Science, Research, and Art (TMWFK) under Grant B 607-01024.

is generated using a linear receiver matrix \mathbf{W}_j of dimension $n_{Rj} \times m_j$:

$$\hat{\mathbf{d}} = \mathbf{W}_j^* \mathbf{x}_j.$$

One way of minimizing inter-user interference is to eliminate it altogether by forcing the constraint $\mathbf{H}_i \mathbf{M}_j = \mathbf{0}$ for $i \neq j$, resulting in the product

$$[\mathbf{H}_1^T \quad \mathbf{H}_2^T \quad \dots \quad \mathbf{H}_K^T]^T [\mathbf{M}_1 \quad \mathbf{M}_2 \quad \dots \quad \mathbf{M}_K]$$

having a block-diagonal structure. A “block-diagonalization” algorithm for optimally choosing $\mathbf{M}_1 \dots \mathbf{M}_K$ to achieve this constraint is found in [4].

An alternative method of managing inter-user interference is to design the transmit vectors so that interference is allowed, while meeting all individual QoS requirements [1, 2, 3]. For the special case when all users have $n_{Rj} = 1$, the channel matrices are row vectors (which we denote as \mathbf{h}_j^T). When the channels are all known to the transmitter, the SINR at the j^{th} receiver (γ_j) is a function of the channels and the transmit vectors \mathbf{m}_j :

$$\gamma_j = \frac{|\mathbf{h}_j^T \mathbf{m}_j|^2}{\sum_{i \neq j} |\mathbf{h}_j^T \mathbf{m}_i|^2 + \sigma_j^2}. \quad (2)$$

The QoS requirement for each user can often be directly mapped to a SINR requirement (γ_j) given the available signal and code designs, and an optimal set of transmit vectors \mathbf{m}_j can be designed to satisfy all requirements with minimum power using algorithms such as those in [1, 2, 5]. To be more specific, define $\mathbf{R}_j = \mathbf{h}_j^* \mathbf{h}_j^T$ and let $\mathbf{m}_j = \sqrt{\lambda_j} \mathbf{u}_j$, the product of a real scalar and a unit-length vector. We want to minimize the total transmitted power ($\sum \lambda_j$) such that

$$\lambda_j \mathbf{u}_j^* \mathbf{R}_j \mathbf{u}_j - \sum_{j \neq k} \gamma_j \lambda_k \mathbf{u}_k^* \mathbf{R}_j \mathbf{u}_k \geq \gamma_j \sigma_j^2. \quad (3)$$

This optimization problem has multiple solutions in the literature [5].

The above problem can be generalized for cases where $n_{Rj} > 1$ by assuming that only one data stream is transmitted to each user ($m_j = 1$) and linear processing is used at the receivers [3]. The receiver matrix \mathbf{W}_j becomes a vector \mathbf{w}_j , and the SINR takes the form:

$$\gamma_j = \frac{|\mathbf{w}_j^* \mathbf{H}_j \mathbf{m}_j|^2}{\sum_{i \neq j} |\mathbf{w}_j^* \mathbf{H}_j \mathbf{m}_i|^2 + \sigma_j^2}. \quad (4)$$

In this case, the interference balancing methods would work if the receiver weight vectors \mathbf{w}_j were known. On the other hand, since the transmitter knows what the received signal at the receiver should be, it can predict \mathbf{w}_j for a predetermined linear receiver design based on, for example, the Minimum Mean Squared Error (MMSE) or Zero Forcing (ZF) criteria. If the transmitter first guesses at an initial set of \mathbf{w}_j vectors,

we can define $\mathbf{R}_j = \mathbf{H}_j^* \mathbf{w}_j \mathbf{w}_j^* \mathbf{H}_j$ and find the corresponding optimal transmit vectors \mathbf{m}_j . Repeated alternating recalculation of the \mathbf{w}_j and \mathbf{m}_j vectors will reduce the required transmitted power until it converges to a global optimum [3]. Care must be taken in choosing an initial set of \mathbf{w}_j vectors, because it is possible for the initial point to not have a feasible solution even when such a solution exists. Furthermore, the computational cost of this method will increase with the distance of the initialization point from the final solution.

This above method for $n_{Rj} > 1$ can be generalized one step further to accommodate the transmission of multiple data streams per user (when the channel dimensions and rank allow it). This idea has been proposed for single-user channels [6]. If the SINR for each data stream is specified (let $\gamma_{j,k}$ represent the SINR of the k^{th} data stream and $\mathbf{w}_{j,k}$ the corresponding column of \mathbf{W}_j for user j), then

$$\gamma_{j,k} = \frac{|\mathbf{w}_{j,k}^* \mathbf{H}_j \mathbf{m}_{j,k}|^2}{\sum_{j \neq l, k \neq m} |\mathbf{w}_{j,k}^* \mathbf{H}_j \mathbf{m}_{l,m}|^2 + \sigma_j^2}. \quad (5)$$

The problem with this approach is determining the optimal SINR requirements. The simplest solution is to use equal power for all channels. However, this could easily result in a situation where sub-channels with low gain have unusually high amounts of power forced into them. In iterative interference balancing algorithms, this results in more interference for other channels to deal with, and ultimately results in solutions that require higher total power and more iterations for convergence (assuming that convergence is possible).

Finding an optimal SINR distribution among multiple data streams is solvable in closed form (the classic water-filling solution) when all sub-channels are orthogonal (such as with block-diagonalization). In the presence of inter-user interference, this is a much more challenging problem. The gains of each sub-channel are not known until the iterative solution is completed, and the iterative solution requires the SINRs to be fixed in advance. It may be possible to add an additional outer iteration loop where an initial power distribution is guessed, and after the iterative solution is computed, the channel gains are used to update the power distribution, and then the iterative solution is recomputed. However, this would substantially increase the computational cost of an already expensive algorithm.

At this point it is important to consider the fact we are really dealing with a discrete optimization problem. Water-filling solutions are useful in illustrating theoretical limits, but they require infinite granularity in the size of the signal constellation and assume that Shannon’s capacity bound is met perfectly. In practice, however, only a finite set of signal constellations and codes are available, each of which has a known set of SNR requirements for the error rate of interest. Since we are already interested in guessing at a good initialization point for the iterative algorithm, if the channel gains at the initialization point are close to the channel

gains at the iterative solution, the discrete nature of the bit-loading problem makes it possible that solving for the optimal bit distribution using an intelligent preliminary guess at the channel gains will find the globally optimal solution. Even if the initial guess is not the globally optimal solution, it will converge to a point that is better than if equal power distribution were used, and may likely be near-optimal.

The bit-loading problem was originally solved with multi-carrier modulation schemes in mind, but its application to this problem is straightforward. Specifically, we are interested in solutions to the “Margin Maximization Problem” in [7]. The most basic algorithm for solving this consists of incrementally adding 1 bit to the sub-channel with the lowest incremental cost until the total transmission rate requirement is met.

3. HYBRID ALGORITHM

For channels where $\sum n_{R_j} \leq n_T$, we propose using the block-diagonalization algorithm in [4] to decompose the channel, using the resulting receiver vectors for \mathbf{W}_j , and using the channel gains Σ_j for the bit-loading algorithm. The block-diagonalization algorithm optimally balances the \mathbf{M}_j matrices among all users, although it is subject to the sub-optimal constraint that $\mathbf{H}_i \mathbf{M}_j = \mathbf{0}$ for $i \neq j$. The inter-user balancing property means that the resulting solution is likely close to the global optimal solution. Given the finite set of signal designs, there is now a reasonable probability that the resulting output of the bit-loading algorithm will result in the globally optimal solution, particularly when operating at high SNR, where the optimal solution would inherently require minimal inter-user interference. The fact that the block-diagonalization solution is close to the globally optimal solution also means that using the associated receiver vectors to initialize \mathbf{W}_j for iterative interference balancing results in a reduced overall number of required iterations. The following is a summary of the hybrid algorithm that combines block-diagonalization and iterative interference balancing:

1. For $j = 1, \dots, K$:

- (i) Compute $\tilde{\mathbf{V}}_j^{(0)}$, the right null space of $\tilde{\mathbf{H}}_j$.
- (ii) Compute the SVD

$$\mathbf{H}_j \tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j \Sigma_j \mathbf{V}_j^*.$$

- (iii) Using the sub-channel gains $\sigma_{j,1} \dots \sigma_{j,L_j}$, find the sub-channel SINRs $\gamma_{j,1} \dots \gamma_{j,L_j}$, such that the total rate constraint R_j is met and total power is minimized [7].
- (iv) Let $\mathbf{w}_{j,k}$ be the k^{th} column of \mathbf{U}_j .

- 2. Let $\mathbf{R}_{j,k} = \mathbf{H}_j^* \mathbf{w}_{j,k} \mathbf{w}_{j,k}^* \mathbf{H}_j$. Find the unit vectors $\mathbf{u}_{j,k}$ and power coefficients $\sqrt{\lambda_{j,k}}$ such that $\sum \lambda_{j,k}$ is minimized and

$$\lambda_{j,k} \mathbf{u}_{j,k}^* \mathbf{R}_{j,k} \mathbf{u}_{j,k} - \sum_{j \neq l, k \neq m} \gamma_{j,k} \lambda_{l,m} \mathbf{u}_{l,m}^* \mathbf{R}_{j,k} \mathbf{u}_{l,m} \geq \gamma_{j,k} \sigma_{j,k}^2,$$

using one of the available algorithms [1, 2, 5].

- 3. Repeat until convergence (optional):

- (i) Recalculate predicted MMSE receiver weights,

$$\mathbf{W}_j = (\mathbf{H}_j \mathbf{M}_S \Lambda \mathbf{M}_S^* \mathbf{H}_j^* + \sigma_n^2 \mathbf{I})^{-1} \mathbf{H}_j \mathbf{M}_j,$$

and normalize so that $\mathbf{w}_{j,k}^* \mathbf{w}_{j,k} = 1$.

- (ii) Recalculate \mathbf{M}_S and Λ (repeat step 2).

The reason that step 3 is optional is that after step 2 is completed, a feasible solution exists. Step 3 generally accounts for most of the computational cost of this algorithm, and as will be seen in the next section, provides only a small gain in performance.

Another advantage of interference balancing strategies over zero-forcing methods such as block-diagonalization is that they can easily be generalized to accommodate noisy estimates of \mathbf{H}_j . To do so, the procedure will be the same, but $\mathbf{R}_{j,k}$ will be redefined to incorporate noise statistics.

4. SIMULATION RESULTS

We assume in the simulations that the required rate for each user is an integer number of bits/use. The available set of signals are the QAM constellations from 1-8 bits/symbol (including BPSK and QPSK as special cases), and the power requirement is based on the corresponding upper bound for the symbol error rate from [8], equation (5-2-80):

$$P_M \leq 4Q \left(\sqrt{\frac{3kE_b}{(M-1)N_0}} \right), \quad (6)$$

where k is the number of bits and $M = 2^k$. We assume that no additional coding is used. The symbol error rate has been fixed at 10^{-5} .

We compare the performance of two algorithms: (i) the hybrid block-diagonalization and iterative interference balancing algorithm, referred to in the plots as “BD-IIB,” and (ii) iterative interference balancing (i.e. step 3 of the algorithm outlined in the last section), but instead of using block-diagonalization to initialize the algorithm, we perform an SVD on each \mathbf{H}_j , use the singular values as the channel gains for bit-loading, and the left singular vectors as the initial \mathbf{W}_j matrices. This is referred to in the plots as “SVD-IIB.” This approach would be a good one for a single-user

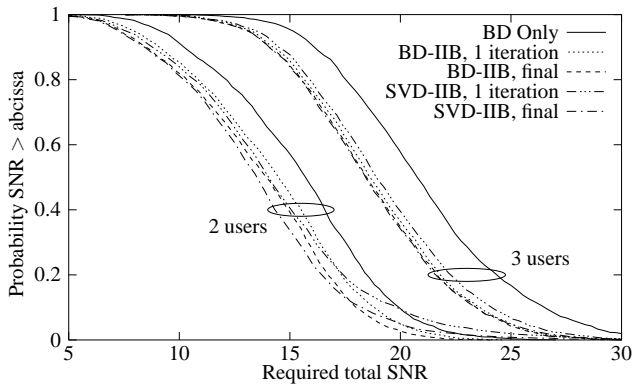


Fig. 1. A comparison of the power minimization capability of the algorithms in different channel conditions.

MIMO channel, but in the multi-user case the initialization doesn't take into account the effects of the interference from other users.

In the first scenario, a 4-antenna base station communicates with two 2-antenna users with equal mean channel gains. In the second scenario, we consider a more challenging case where an 8-antenna base station communicates with 3 users with 1, 2, and 3 antennas, respectively, and random attenuations varying uniformly from 0-10 dB. All channel gains are independent Gaussian. Figure 1 shows the total SNR required to transmit at the requested QoS for both scenarios. We compare 5 different transmit vector computation methods: block-diagonalization (labeled BD), BD-IIB with 1 iteration and after convergence, and SVD-IIB with 1 iteration and after convergence.

Figure 2 illustrates the number of iterations required for the iterative algorithm to converge for both initializations, and for both of the two channel scenarios considered. In all cases, the computation time is considerable, although it is clear that the cost is generally cut in half by using the block-diagonalization approach. However, as seen in Figure 1, the excellent results obtained after only a single iteration make that perhaps the most attractive option.

5. CONCLUSION

We have proposed a new method of calculating transmit vectors for multi-user MIMO downlinks with arbitrary QoS constraints for each user. This involves using previous zero-forcing or "block-diagonalization" methods to arrive at a feasible, but sub-optimal solution. We use this solution to initialize an iterative interference balancing solution, including using bit-loading on the channel gains to guess at the globally optimum bit-distribution. Simulation results show that on average, less than 1 dB of savings in system power is available by letting the algorithm converge to the global optimum, so it may be practical to only use one iteration of

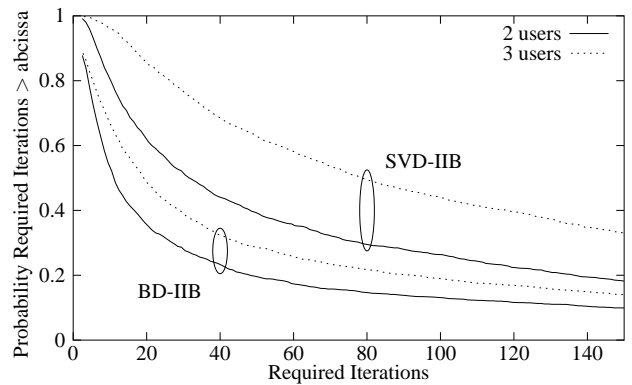


Fig. 2. A comparison of the computational cost of the two iterative algorithms

the algorithm when computational cost is a priority.

6. REFERENCES

- [1] H. Boche and M. Schubert, "Multi-Antenna Downlink Transmission with Individual SINR Receiver Constraints for Cellular Wireless Systems," in *Proc. 4th International ITG Conf. on Source and Channel Coding*, pp. 159–166, VDE Verlag GmbH, January 2002.
- [2] M. Bengtsson and B. Ottersten, *Handbook of Antennas in Wireless Communications*, ch. Optimal and Suboptimal Beamforming. CRC Press, August 2001.
- [3] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint Transmitter Receiver Diversity for Efficient Space Division Multiaccess," *IEEE Trans. on Wireless Communications*, vol. 1, pp. 16–27, January 2002.
- [4] Q. H. Spencer and M. Haardt, "Capacity and Downlink Transmission Algorithms for a Multi-user MIMO Channel," in *36th Asilomar Conference on Signals, Systems and Computers*, November 2002.
- [5] M. Bengtsson, "A Pragmatic Approach to Multi-User Spatial Multiplexing," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, August 2002.
- [6] D. P. Palomar, M. A. Lagunas, and J. Cioffi, "Optimum Linear Joint Transmit/Receive Processing for MIMO Channels with QoS Constraints," submitted to *IEEE Trans. on Signal Processing*, May 2002.
- [7] J. Campello, "Practical Bit Loading for DMT," in *Proc. IEEE International Conf. on Communications*, pp. 801–805, IEEE, 1999.
- [8] J. G. Proakis, *Digital Communications*. McGraw Hill, third ed., 1989.