Joint Multiuser Transmit-Receive Optimization Using Linear Processing

Adam J. Tenenbaum and Raviraj S. Adve
Dept. of Elec. and Comp. Eng., University of Toronto
10 King's College Road, Toronto, Ontario, M5S 3G4, Canada
Email: {adam,rsadve}@comm.utoronto.ca

***Abstract*— In this paper we propose a novel method for joint transmit-receive linear optimization in the downlink of a multiuser MIMO communication system. This new method adapts existing joint linear optimization algorithms from the single user domain for application to the multiuser domain. The optimum transmit matrix is obtained using an iterative procedure based on a minimum mean-squared error (MMSE) criterion and a per-user power constraint; the optimum receive matrices for each user are then derived under an MMSE constraint. The proposed technique improves performance and increases data throughput in multiuser scenarios.**

I. Introduction

The desire to introduce multimedia technologies into mobile networks has motivated a great deal of research into increasing reliability and data throughput. A key development in this regard is the introduction of multiple antenna arrays at both the base station and mobile terminal [1], [2].

Within this broad field of research, one particular branch investigates optimal linear precoding at the transmitter and decoding at the receiver. The goal, usually, is to minimize the mean squared error between transmit and receive data streams, generally using spatial linear processing methods [3]–[6]. In [3], Sampath *et al.* consider a weighted minimum mean squared error (MMSE) criterion and find the optimum transmit and receive matrices through eigendecomposition of the channel. In [4], Scaglione *et al.* develop space-time precoding, again for the MMSE criterion. Palomar *et al.* generalize these results to several design criteria, classified into Schur-concave and Schur-convex objective functions [5]. The concavity or convexity of the objective function then determines the method used for finding the optimum transmit and receive matrices. The receive matrix is derived as the MMSE (Wiener) filter. The optimum transmit matrix is then determined through SVD of the whitened channel.

The aim of the work in [3]–[5] is to optimally transmit multiple data streams to a *single user*. An interesting question therefore arises as to how to transmit multiple data streams to *multiple users*. In such a multiuser environment (that is, the broadcast channel), individual users' receivers cannot cooperate with each other. In [7], Yu determines that a generalized decision feedback equalizer (GDFE) located at the transmitter is capable of achieving sum capacity. This GDFE is implemented using nonlinear Tomlinson-Harashima precoding (THP). In a similar approach, Liu and Duel-Hallen introduce THP for interference rejection [8]. In [9], Fischer and Windpassinger also use the THP concept for precoding at the transmitter.

In this paper we focus on linear processing methods, as reducing the complexity and processing requirements is desirable. As well, we focus only on processing in the spatial domain, making results directly applicable to existing time-domain based multiple access systems such as code division or time division multiple access (CDMA/TDMA). We develop linear processing for joint transmit-receive optimization in a multiuser system. In [6], Khaled and Bourdoux derive such a scheme by imposing a null-space constraint on the transmit matrix to ensure orthogonality between the effective channels of each user. This scalarization/diagonalization method reduces the global multiuser problem to a set of single user problems to which the approach of [3] can be applied. While the methods demonstrated in [6] are effective, they place a strict requirement on the number of transmit/receive antennas. Satisfying the null-space constraint requires that there be at least as many transmit antenna elements as the sum of receive antenna elements. When a realistic number of available transmit antenna elements is used, this constraint severely restricts the possibility of receiver diversity.

The algorithm developed in this paper eliminates the constraint on the number of transmit antennas imposed in [6]. In fading channels, the algorithm provides the largest possible diversity order while increasing data throughput. Available single user algorithms [3], [5] are extended to the multiuser situation. An optimal receive matrix is derived for an arbitrary precoding matrix. The eigendecomposition method of [3] is then applied iteratively, under a per-user power constraint, to obtain the individually optimum transmit matrix.

Section II describes the system model that is assumed in the derivations and results that follow. Section III derives the optimum receive and transmit matrices. Section IV presents some simulated results to illustrate the performance of the proposed algorithms. Finally, Section V presents our conclusions and discusses possibilities for future research.

II. System Model

Consider the downlink of a wireless multiuser communication system with K users in a flat fading environment. Fig. 1 illustrates the model as applied to the k^{th} user. In this system, there are N transmit antennas at the base station which are shared by all users. These antennas are used to transmit

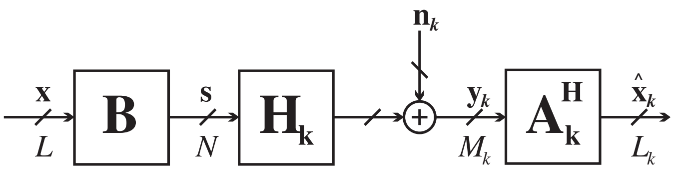

Fig. 1. Illustrating the transmit-receive scheme for the k^{th} user.

L_k symbol streams to the k^{th} user, which are received at M_k receive antennas. Each user receives a combination of all $L = \sum_{k=1}^{K} L_k$ symbol streams through its own channel $\mathbf{H}_k$. The goal of the joint transmit-receive linear processing design is to enable each user to recover its own set of L_k symbol streams. Two important constraints must be satisfied to guarantee resolvability. First, there must be at least as many transmit antennas at the base station as the total number of streams being transmitted. As well, user k must have at least as many receive antennas as it does received substreams , i.e.

$$L = \sum_{k=1}^{K} L_k \leq N; \qquad L_k \leq M_k. \tag{1}$$

The length-M_k received data vector, $\mathbf{y}_k$, and length-L_k soft-decision received data vector, $\hat{\mathbf{x}}_k$, for the k^{th} user are

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{B} \mathbf{x} + \mathbf{n}_k, & (2)\\ \hat{\mathbf{x}}_k &= \mathbf{A}_k^H \mathbf{y}_k = \mathbf{A}_k^H \mathbf{H}_k \mathbf{B} \mathbf{x} + \mathbf{A}_k^H \mathbf{n}_k, & (3) \end{aligned}$$

where the superscript ^{H} indicates the conjugate transpose operator, $\mathbf{x}$ is the $L \times 1$ vector of transmitted symbols, $\mathbf{x} = \left[\mathbf{x}_1^T \cdots \mathbf{x}_K^T\right]^T$, and $\mathbf{x}_k = [x_{k,1} x_{k,2} \cdots x_{k,L_k}]^T$. $\mathbf{B}$ is the $N \times L$ global transmit matrix for all users, $\mathbf{H}_k$ is the $M_k \times N$ channel from the base station to user k, $\mathbf{n}_k$ is the $M_k \times 1$ noise vector for user k, and $\mathbf{A}_k$ is the $L_k \times M_k$ receive matrix for user k. We assume individual data streams have unit power and are independent, i.e. $\mathbb{E}\{\mathbf{x}_k \mathbf{x}_k^H\} = \mathbf{I}_{L_k}$, where $\mathbb{E}$ represents the expectation operator and $\mathbf{I}_{L_k}$ the $L_k \times L_k$ identity matrix. We also assume that input data and noise are independent and that the receiver noise at the k^{th} terminal has correlation $\mathbf{R}_{n_k}$.

III. Optimum Receive and Transmit Matrices

For user k, the mean squared errors are the diagonal entries in the error covariance matrix, $\mathbf{E}_k$, defined as

$$\mathbf{E}_k = \mathbb{E}\left[(\hat{\mathbf{x}}_k - \mathbf{x}_k)(\hat{\mathbf{x}}_k - \mathbf{x}_k)^H\right]. \tag{4}$$

Using (3) and expanding the result yields

$$\begin{aligned} \mathbf{E}_k &= \mathbf{A}_k^H \mathbb{E}\left[\mathbf{y}_k \mathbf{y}_k^H\right] \mathbf{A}_k - \mathbf{A}_k^H \mathbb{E}\left[\mathbf{y}_k \mathbf{x}_k^H\right] \\ &\qquad - \mathbb{E}\left[\mathbf{x}_k \mathbf{y}_k^H\right] \mathbf{A}_k + \mathbf{I}_{L_k} & (5)\\ &= \mathbf{A}_k^H \mathbf{R}_{yk} \mathbf{A}_k - \mathbf{A}_k^H \mathbf{H}_k \mathbf{B}_k - \mathbf{B}_k^H \mathbf{H}_k^H \mathbf{A}_k + \mathbf{I}_{L_k}, & (6) \end{aligned}$$

with

$$\begin{aligned} \mathbf{R}_{yk} &= \mathbb{E}\left[\mathbf{y}_k \mathbf{y}_k^H\right] \\ &= \mathbf{H}_k \mathbf{B} \mathbf{B}^H \mathbf{H}_k^H + \mathbf{R}_{n_k}. & (7) \end{aligned}$$

The goal then, using a per user power constraint, is to find the optimal matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ (for each user k) that minimize the sum MSE, i.e. trace of $\mathbf{E}_k$,

$$\left(\mathbf{A}_k^{\text{opt}}, \mathbf{B}_k^{\text{opt}}\right) = \arg \min_{\mathbf{A}_k, \mathbf{B}_k, \text{tr}\left[\mathbf{B}_k \mathbf{B}_k^H\right] \leq P/K} \text{tr}\left[\mathbf{E}_k\right]. \tag{8}$$

Note that the power constraint restricts the transmit matrix only.

A. Optimum Receive Matrix

The derivation for the receive matrix is presented first, as the optimum solution can be found as a general solution for any transmit matrix $\mathbf{B}$. A similar derivation for the single user case is available in [5].

Given a transmit matrix $\mathbf{B}$, the problem of finding the optimum receive matrix $\mathbf{A}_k$ (that minimizes the trace of $\mathbf{E}_k$) is equivalent to:

$$\mathbf{A}_k^{\text{opt}} = \arg \min_{\mathbf{A}_k} \mathbf{c}^H \mathbf{E}_k \mathbf{c} \quad \forall \mathbf{c}. \tag{9}$$

Since for arbitrary matrices $\mathbf{C}$ and $\mathbf{D}$, with consistent dimensions, $\text{tr}\left[\mathbf{CD}\right] = \text{tr}\left[\mathbf{DC}\right]$, and the fact that $\mathbf{c}^H \mathbf{E}_k \mathbf{c}$ is a scalar, we can write $\mathbf{c}^H \mathbf{E}_k \mathbf{c} = \text{tr}\left[\mathbf{c}^H \mathbf{E}_k \mathbf{c}\right]$, and (9) becomes

$$\mathbf{A}_k^{\text{opt}} = \arg \min_{\mathbf{A}_k^H} \text{tr}\left[\mathbf{E}_k \mathbf{c} \mathbf{c}^H\right] \ \forall \mathbf{c}. \tag{10}$$

Taking the gradient of the new objective function and equating it to zero,

$$\nabla_{\mathbf{A}_k^H} \text{tr}\left[\mathbf{E}_k \mathbf{c} \mathbf{c}^H\right] = \mathbf{R}_{yk} \mathbf{A}_k \mathbf{c} \mathbf{c}^H - \mathbf{H}_k \mathbf{B}_k \mathbf{c} \mathbf{c}^H = 0. \tag{11}$$

Thus, we find that the optimum solution is the linear MMSE solution (Wiener filter):

$$\begin{aligned} \mathbf{A}_k^{\text{opt}} &= \mathbf{R}_{y_k}^{-1} \mathbf{H}_k \mathbf{B}_k \\ &= (\mathbf{H}_k \mathbf{B} \mathbf{B}^H \mathbf{H}_k^H + \mathbf{R}_{n_k})^{-1} \mathbf{H}_k \mathbf{B}_k. & (12) \end{aligned}$$

This result illustrates that the optimum receive matrix for user k, $\mathbf{A}_k^{\text{opt}}$, can be obtained through linear processing when knowledge of the transmitting filter coefficients and the channel coefficients from the base station to the user are available. The receiver *does not* require knowledge of the channels to other users, i.e. it does not require cooperation with other users. On the other hand, the receiver does require knowledge of the transmit linear precoding matrix $\mathbf{B}$ to all users. One can imagine this information being transmitted to the user or the Wiener filter being estimated in a training phase.

The result in (12) was derived in general terms; as such, it is valid for any transmit filter $\mathbf{B}$. By rewriting the $\mathbf{B}\mathbf{B}^H$ term as $\sum_{j=1}^{K} \mathbf{B}_j \mathbf{B}_j^H$, we can rewrite (12) as

$$\mathbf{A}_k^{\text{opt}} = (\mathbf{H}_k \mathbf{B}_k \mathbf{B}_k^H \mathbf{H}_k^H + \mathbf{R}_{(n+I)_k})^{-1} \mathbf{H}_k \mathbf{B}_k, \tag{13}$$

where

$$\mathbf{R}_{(n+I)_k} = \mathbf{R}_{n_k} + \sum_{j=1, j \neq k}^{K} \mathbf{H}_k \mathbf{B}_j \mathbf{B}_j^H \mathbf{H}_k^H. \tag{14}$$

The result in (13) is directly analogous to the result in [5] for the single user case. The noise covariance matrix in [5] is replaced by a noise-plus-interference covariance matrix.

B. Optimum Transmit Matrix

Given the optimal receive matrix in (13) and the original expression for MSE in (6), the resulting MSE between the transmitted and received streams is given by

$$\begin{aligned} \mathbf{E}_k &= \mathbf{I} - \mathbf{B}_k^H \mathbf{H}_k^H (\mathbf{H}_k \mathbf{B}_k \mathbf{B}_k^H \mathbf{H}_k^H + \mathbf{R}_{(n+I)_k})^{-1} \mathbf{H}_k \mathbf{B}_k \\ &= \left(\mathbf{I} + \mathbf{B}_k^H \mathbf{R}_{H_k} \mathbf{B}_k\right)^{-1}, \end{aligned} \tag{15}$$

where $\mathbf{R}_{H_k} \doteq \mathbf{H}_k^H \mathbf{R}_{(n+I)_k}^{-1} \mathbf{H}_k$ and (15) follows from the matrix inversion lemma.

A convenient solution for the optimal transmit matrix of user k, $\mathbf{B}_k$, would also have been analogous to that for the single user case in [5]. However, when a similar derivation is attempted in the multiuser case (to find a global solution to the constrained MMSE problem), the noise covariance matrix is replaced by the noise-plus-interference covariance matrix. The result of this substitution is that, unlike in the single user case, each user's MSE matrix becomes a complicated function of all of the other users' transmit matrices, and not just that of the user under consideration (user k).

In order to overcome this difficulty, we propose an iterative solution. In our algorithm, the i^{th} step of the iteration optimizes the k^{th} user's transmit matrix using the transmit matrices calculated for users $1, \ldots, k-1$ in the current iteration, i, and those determined for users $k+1, \ldots, K$ in the previous iteration, $i-1$. Using this iterative method, we solve a set of individually concave optimizations, rather than trying to find an optimum solution based on the complicated global MSE matrix. In particular, we assign equal power to individual users, which may be sub-optimal.

Proposition: The solution to the Schur-concave optimization problem[1]

$$\min_{\mathbf{B}_k,\, \mathrm{tr}\left[\mathbf{B}_k^H \mathbf{B}_k \leq P/K\right]} \mathrm{tr}\left[\mathbf{E}_k\right], \tag{16}$$

is given by

$$\mathbf{B}_k = \mathbf{U}_k \mathbf{\Sigma}_k, \tag{17}$$

where $\mathbf{U}_k$ are the $\tilde{L}_k$ eigenvectors corresponding to the $\tilde{L}_k$ largest eigenvalues of $\mathbf{R}_{H_k}$, $\tilde{L}_k = \min\{L_k, \mathrm{rank}\left[\mathbf{H}_k\right]\}$. $\mathbf{\Sigma}_k = [\mathbf{0}; \mathrm{diag}\{\sigma_{k,l}\}]$ is a $\tilde{L}_k \times L_k$ matrix with zero elements except along the rightmost main diagonal. Furthermore,

$$\sigma_{k,l} = \left[\mu^{-1/2} \lambda_{k,l}^{-1/2} - \lambda_{k,l}^{-1}\right]^+, \tag{18}$$

where $[x]^+ = \{x, x \geq 0; 0, x < 0\}$. $\lambda_{k,l}$ are the eigenvalues of $\mathbf{R}_{H_k}$ and μ is chosen to satisfy the transmit power constraint per user, $\mathrm{tr}\left[\mathbf{B}_k \mathbf{B}_k^H\right] = P/K$.

Proof: The detailed proof is omitted due to space constraints. However, the proposition is valid as the optimization equation in (16) is directly analogous to Eqn. (8) in [5], with the noise covariance term being replaced by the noise-plus-interference covariance matrix. Since the derivation in [5] is valid for an arbitrary noise covariance matrix, (17) holds. For the same reasons, (18) holds. As such, we follow methods similar to [3] and [5] in decomposing the resulting channel and allocating power for each user over its best available eigenchannels.

[1] Proof that this problem is Schur-concave is available in [5].

C. Proposed Algorithm

The steps in the proposed iterative algorithm are:

1) *Initialization:* Before beginning the iterative process, we initialize the transmit matrix for each user. One possible starting point for the iterative process is with a zero transmit matrix ($\mathbf{B}(0) = \mathbf{0}$). The selection of zero as an initialization point allows the first user to form its first transmit matrix as if it were in a single user environment. Simulations demonstrate that while solutions using different initialization points converge to the same sum MSE, selecting the zero initialization point provides faster convergence. An example in Section IV illustrates this phenomenon.
2) *Iteration:* For each user $k = 1, \ldots, K$,

$$\begin{aligned} \mathbf{R}_{(n+I)_k} &= \mathbf{R}_{n_k} + \sum_{j=1}^{k-1} \mathbf{H}_k \mathbf{B}_j(i) \mathbf{B}_j^H(i) \mathbf{H}_k^H \qquad (19) \\ &\quad + \sum_{j=k+1}^{K} \mathbf{H}_k \mathbf{B}_j(i-1) \mathbf{B}_j^H(i-1) \mathbf{H}_k^H \\ \mathbf{R}_{H_k} &= \mathbf{H}_k^H \mathbf{R}_{(n+I)_k}^{-1} \mathbf{H}_k. \qquad (20) \end{aligned}$$

   Using these matrices with Eqns. (17) and (18), determine the transmit matrix $\mathbf{B}_k(i)$ for user k.
3) *Termination:* The process of iteration is terminated when the transmit matrix $\mathbf{B}(i)$ has converged to a solution. Convergence may be defined in some convenient fashion, such as in terms of the average MSE.

IV. SIMULATION RESULTS AND ANALYSIS

In this section we provide Monte Carlo simulation results to illustrate certain features of the joint optimization algorithm. All cases assume a flat Rayleigh fading environment and zero mean Gaussian noise with variance σ^2 ($\mathbf{R}_{n_k} = \sigma^2 \mathbf{I}$). Channel coefficients are generated for each realization of the channel matrix $\mathbf{H}$ as independent and identically distributed (*i.i.d.*) samples of a complex Gaussian process with zero mean and unit variance. Full knowledge of the channel matrix $\mathbf{H}$ is assumed at the transmitter, and each user k is assumed to have knowledge of its own channel matrix $\mathbf{H}_k$. As well, each receiver is assumed to have knowledge of the transmit matrix $\mathbf{B}$. The number of iterations in transmit optimization is limited to 60; however, it is rare that more than 30 iterations are required for the algorithm to converge.

Example 1: Comparison to Existing Techniques: In this first example we compare the performance of our iterative algorithm to the null-space criterion method described in [6]. The system uses two antennas to transmit to two users ($N = 2$, $K = 2$). A single QPSK modulated data stream is transmitted to each user. To satisfy the "number of transmit antennas" constraint of [6], the implementation of their method uses one receive antenna element per user ($M_k = 1$). To illustrate the fact that the method proposed in this paper does not require

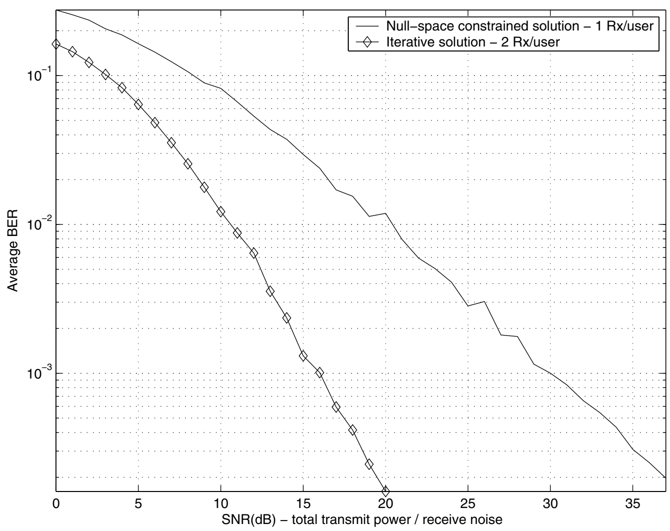

Fig. 2. Performance improvement using iterative method

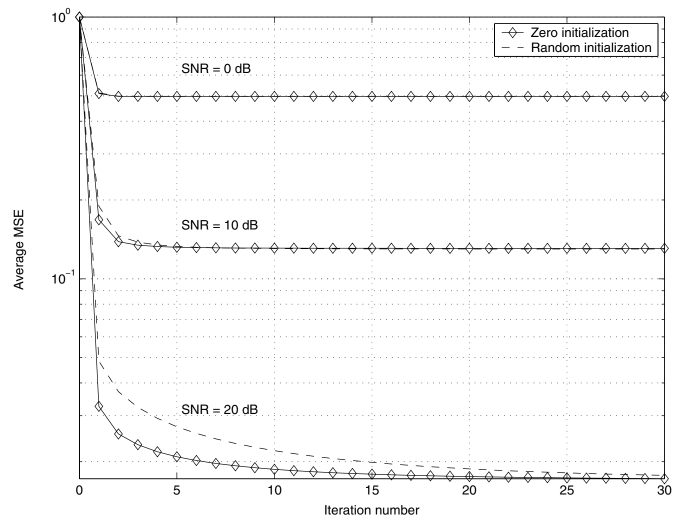

Fig. 3. Illustrating the effects of initialization on convergence, $M_k = 2$

that constraint, the algorithm is implemented with two receive antenna elements per user ($M_k = 2$). 200 QPSK symbols are transmitted for each channel realization, and bit error rates (BER) are averaged over at least 1000 channel realizations. A minimum of 10^4 bit errors are generated for each data point.

Fig. 2 plots the average BER as a function of SNR for the two cases, $M_k = 1$ (using the null-space constrained solution proposed in [6]) and $M_k = 2$. Using two receive elements per user, a gain of approximately 14 dB is achieved at a BER of 10^{-3}. The fact that the gain arises is not surprising. However, what is interesting is that the scheme proposed in Section III-C allows us to achieve these gains, even though the null-space constraint of [6] is violated.

Example 2: Initialization Conditions: In this example, we examine the convergence and effects of initialization of the transmit matrix on the iterative algorithm. The data illustrated in this example are the result of averaging over Monte Carlo simulations of 5000 channel realizations for each SNR value. The numerical results represent the average MSE over all symbol streams for all users. Here, the two conditions tested are zero initialization and random initialization. In the random case, the entries of $\mathbf{B}(0)$ are drawn as *i.i.d.* samples of a complex Gaussian random process with zero mean and variance $\frac{1}{NL}$, so that $\mathbf{B}(0)$ meets the specified power constraint.

Individual simulations suggest that the algorithm converges to the same point regardless of the initialization point. However, upon examining the average performance over many simulations, it appears that when using zero initialization, convergence occurs in fewer iterations. Fig. 3 confirms this behaviour. Here, two symbol streams (one per user, $L_k = 1$) are transmitted over two transmit antenna elements ($N = 2$) to two users ($K = 2$), each of whom has two receive antenna elements ($M_k = 2$). This case demonstrates an interesting result: while the two initialization points yield equally rapid convergence at low SNR, using a zero initialization provides much quicker convergence at higher SNR. This behaviour is due to the fact that the first step, following the zero initialization, is unaffected by the other users at low SNR (where noise dominates the noise+interference covariance matrix). However, at high SNR, where multiuser interference is dominant, the zero initialization point leads to a better 'first estimation' (i.e. the single-user estimation). In contrast, the random initialization point causes less desirable interference (when compared to that induced by the iterative algorithm). More iterations must thus be dedicated to cancelling out this initial interference in order to reach the convergence point.

Example 3: Diversity Gain with Additional Antennas: This example serves to illustrate the increase in diversity order that is exhibited when multiple antennas are added at the transmitter or receiver. With the exception of the antenna parameters N and M_k, simulations for this example are configured identically to those in Example 1. In Fig. 4 we see an increase in diversity order when the number of transmit antennas is fixed ($N = 2$) and the number of receive antennas per user is varied ($M_k = 2, 3, 4$); this increase is exhibited by the changing slope of the BER curve. Similar observations can be made when the number of receive antennas is fixed ($M_k = 2$) but the number of transmit antennas increases ($N = 2, 3, 4$). An equivalence can be seen between adding antennas at the transmitter and receiver, as the ($N = 2, M_k = 3$) and ($N = 2, M_k = 4$) configurations offer nearly identical performance to the ($N = 3, M_k = 2$) and ($N = 4, M_k = 2$) configurations respectively.

Example 4: Performance Comparison to Other Numerical Methods: In this final example, we examine the performance of the proposed algorithm in the context of other numerical solutions. In particular, we apply the Sequential Quadratic Programming method [10] to the MMSE problem. SQP uses numerical approximations to decompose a problem into quadratic programming subproblems that are solved using well known methods. As SQP is a more computationally intensive algorithm, it is not comparable to our proposed algorithm

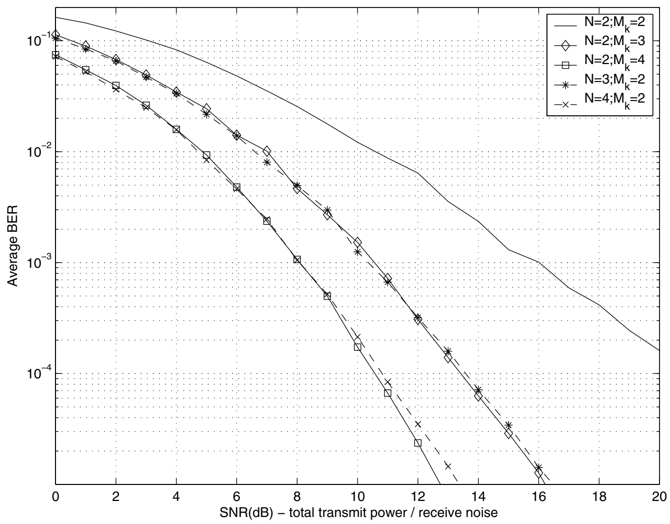

Fig. 4. Increasing diversity by adding transmit / receive antennas

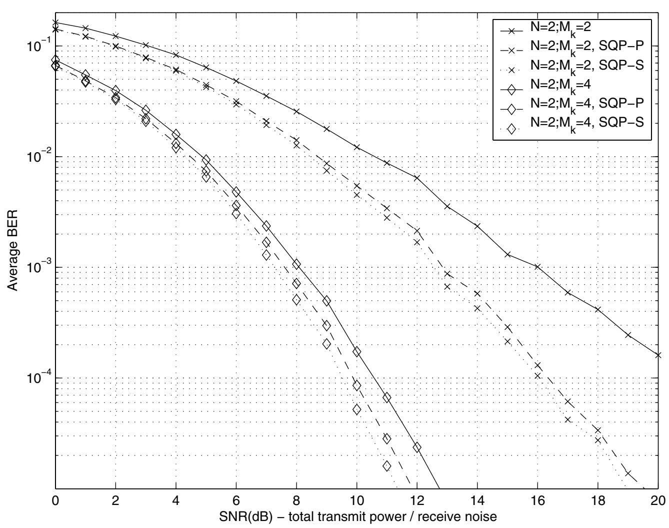

Fig. 5. Comparison of proposed algorithm to SQP

in terms of efficiency; however, it does provide a point of comparison in terms of possible BER performance.

In Fig. 5, we compare the BER performance of our proposed algorithm with the SQP solution under both a per-user and sum power constraint. Simulation parameters are identical to selected cases in Example 3. Here, we see that while SQP based methods can achieve higher diversity order (and thus better performance) than our iterative approach, the change in diversity order when adding antennas is similar. The improvement in performance when applying SQP to the problem suggests that the algorithm proposed in Section III converges to a false minimum; this is most likely caused by the presence of the k^{th} user's transmit matrix in the MMSE subproblems of the other $K-1$ users. The results in Fig. 5 also confirm that the per-user power constraint (SQP-P) is indeed sub-optimal when compared to a sum power constraint (SQP-S).

V. Conclusions

This paper has presented a technique for joint transmit-receive optimization in *multiuser* situations. A general solution was derived for the optimum (MMSE) receive matrix, and an iterative procedure was developed to find an individually optimum transmit matrix for each user under a per-user power constraint. Simulation results were presented to illustrate the possible increases in performance made possible by the linear processing optimization method and to examine the impact of initialization conditions on convergence.

A significant drawback of the proposed iterative approach is that it appears to converge to a false, local, minimum. More computationally intensive approaches, such as SQP, on the other hand, provide better performance. The theory to overcome this performance loss would be a significant extension of this work. On the other hand, the results in Section IV show the proposed iterative algorithm to provide excellent results and make full use of the available transmit-receive spatial resources. Finally, the techniques developed within this paper operate only in the spatial domain, and, as such, can be applied without modification to current multiple access schemes, such as CDMA or TDMA.

References

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Dec. 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[4] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[5] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint tx-rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[6] A. Bourdoux and N. Khaled, "Joint TX-RX optimisation for MIMO-SDMA based on a null-space constraint," in *Proc. of IEEE Vehicular Technology Conference (VTC-02 Fall)*, Vancouver, Canada, Sept. 2002, pp. 171–174.

[7] W. Yu, "Competition and cooperation in multiuser communication environments," Ph.D. dissertation, Stanford University, 2002, available at `http://www.comm.utoronto.ca/˜weiyu/publications.html`.

[8] J. Liu and A. Duel-Hallen, "Tomlinson-Harashima transmitter precoding for synchronous multiuser communications," in *Proc. of the 37th Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2003.

[9] R. F. H. Fischer and C. A. Windpassinger, "Improved MIMO precoding for decentralized receivers resembling concepts from lattice reduction," in *Proc. of IEEE Globecom*, vol. 4, San Francisco, CA, Dec. 2003, pp. 1852–1856.

[10] P. T. Boggs and J. W. Tolle, "Sequential quadratic programming," in *Acta Numerica*, 1995, pp. 1–51.