EURASIP Journal on Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Improved and weighted sum rate maximization for successive zero-forcing in multiuser MIMO systems

Robert C Elliott[1,2] and Witold A Krzymień[1,2*]

Abstract

We propose an improved algorithm for optimizing the transmit covariance matrices for successive zero-forcing (SZF) precoding in multiple-input multiple-output systems. We use a conjugate gradient projection method to solve the optimization problem. Our algorithm improves upon the existing method twofold. First, the existing covariance optimization method, with higher SNRs or more simultaneously supported users, can yield a lower sum rate than when using block diagonalization (BD), when theoretically SZF should never be inferior to BD. In comparison, our method consistently provides a higher throughput than BD. Several simulations demonstrate our algorithm's enhanced performance, which exceeds the existing method's sum rate by up to 12% in the examined cases. Second, our proposed algorithm also supports the maximization of a weighted sum rate with SZF, to incorporate quality of service (QoS). To our knowledge, no other work in the literature considers a weighted sum rate and/or QoS with SZF.

1. Introduction

The topic of multiple-input multiple-output (MIMO) broadcast channels has drawn much research interest for several years. MIMO spatial multiplexing is potentially a key enabling technique to achieve increased throughput and quality of service (QoS) in commercial fourth-generation cellular and other future broadband wireless networks. The downlink of a MIMO system can be modeled as a broadcast channel (BC), where a base station transmits different data streams simultaneously to several users. It is known that the capacity of a MIMO BC can be reached through the use of dirty paper coding (DPC) [1,2]. DPC takes advantage of non-causal knowledge of each user's signal at the base station. Non-causally knowing the signal for each user, the base station can then successively encode the signal for each user k such that the effect of interference of the k - 1 previously encoded users is removed. The transmit covariance matrices for the users must then be optimized to minimize the remaining interference between users.

Unfortunately, one significant drawback of DPC is that it is highly non-linear. Due to its resulting extreme complexity, its implementation will remain impractical for the foreseeable future, even in a simplified approximate form not requiring non-causal knowledge of transmitted signals. For this reason, reduced complexity linear precoding methods are of interest to mitigate multiuser interference (MUI). Examples of such methods include zero-forcing beamforming (ZFB) [3] for systems with single-antenna users, and block diagonalization (BD) [4] for systems with multiple-antenna users. ZFB and BD are techniques that completely null the interference between users by requiring that the signal transmitted to each user falls in the null space of the channels for all other users. Thus, all MUI is removed at the base station. However, the nulling also imposes a constraint that the total number of receive antennas be no larger than the number of transmit antennas. There is, therefore, a reduction in both the system performance and the number of simultaneously supportable users compared with DPC.

In [5], the authors propose a scheme known as successive zero-forcing (SZF). This scheme nulls multiuser interference similarly to BD, but as opposed to nulling all interference between users, it instead only

* Correspondence: wak@ece.ualberta.ca
[1]Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada
Full list of author information is available at the end of the article

successively nulls the interference of a given user k on the previous k - 1 users. In a sense, it is similar in concept to DPC, but with a reverse user ordering, and with MUI removal through nulling instead of non-linear coding. It is shown in [5] that the sum-rate throughput of SZF is no worse than that of BD, and often quite better; thus, the complete removal of MUI as in BD is not necessarily always beneficial. Additionally, the less strict null space requirements mean that the transmit and receive constraints relative to BD are relaxed, and hence SZF can sometimes serve a higher number of users simultaneously than BD.

The main drawback to SZF is that its resulting sum rate is non-convex. Thus, it is difficult to find the globally optimum transmit covariance matrices that achieve the capacity. In [5], the authors propose a suboptimal covariance method based on the duality between the BC and the multiple access channel (MAC). This method first solves the less-constrained BC capacity problem by instead solving the dual MAC, which is convex, and thus for which efficient numerical methods exist (e.g., [6]). The BC covariance matrices can then be obtained from the MAC results via duality transformations [2]. Finally, SZF covariance matrices are obtained by projecting the BC matrices into the null spaces of the users' channels.

The primary impetus for this work was based on observations made during our related work on scheduling for BD [7] and SZF [8,9]. We found that, unexpectedly, the performance of our SZF scheduling algorithms was worse than for BD at high SNR. This was eventually found to be due to the SZF covariance method rather than the scheduling algorithms. Additionally, the existing covariance optimization method can only be used for sum-rate maximization. However, it is often desirable to be able to maximize a weighted sum rate. The weights on each of the user rates can account for QoS parameters to introduce more fairness into the system. A pure sum-rate maximization can result in throughput starvation for users in poorer channel conditions. Perhaps the most well-known example of weighted sum-rate maximization is the proportional fairness criterion [10,11], where the weights are the inverse of each user's average throughput.

In this paper, our contributions are twofold. First, we propose a new algorithm for optimizing the SZF covariance matrices based on a conjugate gradient projection (CGP) method. This new algorithm, while still globally suboptimal, improves significantly upon the sum-rate performance of the existing method. Second, our proposed method also enables the maximization of a weighted sum rate. This allows QoS parameters to be incorporated with the SZF precoding. To the best of our knowledge, no existing work in the literature yet considers fairness or weighted sum rates in conjunction with SZF.

The remainder of this paper is organized as follows. Section 2 describes the system model and the pertinent details of BD and SZF precoding. Section 3 describes the SZF covariance optimization problem, the existing method and the problems it faces in more detail, and outlines our proposed CGP method. Simulation results are provided in Section 4, and concluding remarks are given in Section 5.

2. System model and overview

To begin, we outline the notation used in this paper. Italic variables represent scalars, while lowercase and uppercase boldface variables denote vectors and matrices, respectively. $\lceil \cdot \rceil$ is the ceiling function. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. $\mathbf{A}^*$, $\mathbf{A}^T$, and $\mathbf{A}^H$ denote the conjugate matrix, matrix transpose, and conjugate (Hermitian) transpose of $\mathbf{A}$, respectively. $\mathrm{Tr}(\mathbf{A})$ is the trace of $\mathbf{A}$, while $\|\mathbf{A}\|_F^2$ denotes the squared Frobenius norm of $\mathbf{A}$. For a square matrix $\mathbf{A}$, $|\mathbf{A}|$ is the matrix determinant, $\mathbf{A}^{-1}$ is the matrix inverse, and $\mathbf{A}^{-1/2}$ is the matrix inverse square root. $\mathbf{A} \succcurlyeq 0$ denotes that $\mathbf{A}$ is positive semidefinite.

In this paper, we consider the downlink of a multiuser MIMO system. There exists a base station with M_T transmit antennas and a transmit power constraint of P. The base station transmits to K_0 users out of a pool of K multiple-antenna users requesting service, each with N_k receive antennas. Let $\mathbf{H}_k \in \mathbb{C}^{N_k \times M_T}$ denote the downlink channel matrix of the kth user, k = 1, 2, ..., K. We assume that at all times, the transmitter has perfect knowledge of the channel state information of all users (perfect CSIT), and that each user knows its channel perfectly. The data vector of user k, $\mathbf{s}_k \in \mathbb{C}^{N_k \times 1}$, is preprocessed at the transmitter with the precoding matrix $\mathbf{W}_k \in \mathbb{C}^{M_T \times N_k}$ to yield the transmitted signal vector $\mathbf{x}_k \in \mathbb{C}^{M_T \times 1}$. The $N_k \times 1$ received signal vector of the kth user can be expressed as

$$\mathbf{y}_k = \mathbf{H}_k \sum_{j=1}^{K_0} \mathbf{W}_j \mathbf{s}_j + \mathbf{n}_k \tag{1}$$

where $\mathbf{n}_k \in \mathbb{C}^{N_k \times 1}$ denotes zero mean additive white Gaussian noise with $E\{\mathbf{n}_k \mathbf{n}_k^H\} = \sigma_n^2 \mathbf{I}_{N_k}$. We assume herein without loss of generality that $\sigma_n^2 = 1$.

A. Block diagonalization

The following is a quick review of BD. First consider the aggregate channel matrix of K_0 users:

$$\mathbf{H} = \left[\mathbf{H}_1^T \mathbf{H}_2^T \dots \mathbf{H}_{K_0}^T\right]^T \in \mathbb{C}^{\sum_k N_k \times M_T} \tag{2}$$

Then, for each user k, remove its channel matrix $\mathbf{H}_k$ from $\mathbf{H}$ to create a set of aggregate $\left(\sum_{j=1,j\neq k}^{K_0} N_j\right) \times M_T$ channel matrices $\tilde{\mathbf{H}}_k$:

$$\tilde{\mathbf{H}}_k = \left[\mathbf{H}_1^T \dots \mathbf{H}_{k-1}^T \mathbf{H}_{k+1}^T \dots \mathbf{H}_{K_0}^T\right]^T \quad (3)$$

Block diagonalization then removes MUI by designing $\mathbf{W}_k$ to fall in the null space of $\tilde{\mathbf{H}}_k$, so that $\mathbf{H}_j\mathbf{W}_k = \mathbf{0}$ for all $k \neq j$ and $1 \leq (j, k) \leq K_0$. This gives the effective channel matrix a block diagonal structure, thereby decomposing the multiuser channel into parallel equivalent single-user channels. The received signal vector (1) for each user becomes $\mathbf{y}_k = \mathbf{H}_k\mathbf{W}_k\mathbf{s}_k + \mathbf{n}_k$. The above implies that the rank of the null space for each $\tilde{\mathbf{H}}_k$ must be greater than zero, which in turn imposes constraints on the total number of receive antennas or users that can be supported. Let $\tilde{r}_k = \text{rank}\left(\tilde{\mathbf{H}}_k\right)$. K_0 users can then be supported using BD if [4] $\max\left(\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_{K_0}\right) < M_T$. If each user's channel matrix is full rank, the above constraint is then equivalent to $\left(\sum_{k=1,k\neq j}^{K_0} N_k\right) < M_T$ for all j, $1 \leq j \leq K_0$. Furthermore, if $N_k = N$, $\forall k$, this further simplifies to $K_0 = \lceil M_T/N \rceil$.

Let the singular value decomposition (SVD) of $\tilde{\mathbf{H}}_k$ be denoted as $\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \left(\tilde{\mathbf{\Sigma}}_k \mathbf{0}\right) \left(\tilde{\mathbf{V}}_k^1 \tilde{\mathbf{V}}_k^0\right)^H$. Then the $\tilde{r}_k \times \tilde{r}_k$ diagonal matrix $\tilde{\mathbf{\Sigma}}_k$ contains the $\tilde{r}_k$ non-zero singular values of $\tilde{\mathbf{H}}_k$. $\tilde{\mathbf{V}}_k^0$ holds the $M_T - \tilde{r}_k$ right singular vectors, which form a basis for the null space of $\tilde{\mathbf{H}}_k$. Constructing the precoding matrix $\mathbf{W}_k$ with the columns of $\tilde{\mathbf{V}}_k^0$ will satisfy the zero MUI condition. The decoupled, non-interfering, equivalent single-user MIMO channels can be expressed as

$$\mathbf{H}_{k,e} = \mathbf{H}_k\tilde{\mathbf{V}}_k^0 \quad (4)$$

With the transmit power constraint P, the achievable throughput of BD is

$$R_{BD} = \max_{\mathbf{Q}_k:\mathbf{Q}_k \succcurlyeq \mathbf{0}} \sum_{k=1}^{K_0} \log_2 \left|\mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{H}_{k,e}\mathbf{Q}_k\mathbf{H}_{k,e}^H\right| \quad (5)$$

such that $\left(\sum_{k=1}^{K_0} \text{Tr}\left(\mathbf{Q}_k\right)\right) \leq P$. $\mathbf{Q}_k$ is the covariance matrix for the equivalent channel of user k. The solution of (5) is obtained by performing the water-filling power allocation over the singular values of the block-diagonal matrix $\mathbf{H}_e = \text{diag}\left(\mathbf{H}_{1,e}, \mathbf{H}_{2,e}, \dots, \mathbf{H}_{K_0,e}\right)$ for the sum-power constraint of P [4], where $\mathbf{H}_{k,e}$, $k = 1, 2, \dots, K_0$ is defined by (4).

B. Successive zero-forcing

Unlike BD, SZF does not completely pre-eliminate the multiuser interference. As the name implies, precoding in SZF is successive, and hence a user precoding order must be defined. For a given set of users with an order π, for each user $k \in \{1, \dots, K_0\}$ the received signal can be expressed as [5]

$$\mathbf{y}_{\pi(k)} = \mathbf{H}_{\pi(k)}\left(\mathbf{W}_{\pi(k)}\mathbf{s}_{\pi(k)} + \sum_{j<k}\mathbf{W}_{\pi(j)}\mathbf{s}_{\pi(j)} + \sum_{j>k}\mathbf{W}_{\pi(j)}\mathbf{s}_{\pi(j)}\right) + \mathbf{n}_{\pi(k)} \quad (6)$$

In SZF, the precoding matrix $\mathbf{W}_{\pi(k)}$ is designed such that it lies in the null space of the aggregate channel $\bar{\mathbf{H}}_k$ of the k - 1 previously precoded users' channels (compared with all other users with BD):

$$\bar{\mathbf{H}}_k = \left[\mathbf{H}_{\pi(1)}^T \mathbf{H}_{\pi(2)}^T \dots \mathbf{H}_{\pi(k-1)}^T\right]^T \quad (7)$$

With this null space constraint, the third term in (6) is cancelled. Then, (6) reduces to

$$\mathbf{y}_{\pi(k)} = \mathbf{H}_{\pi(k)}\left(\mathbf{W}_{\pi(k)}\mathbf{s}_{\pi(k)} + \sum_{j<k}\mathbf{W}_{\pi(j)}\mathbf{s}_{\pi(j)}\right) + \mathbf{n}_{\pi(k)} \quad (8)$$

SZF of K_0 users' channels is possible if $M_T > \text{rank}\left(\bar{\mathbf{H}}_{K_0-1}\right)$.

Let us denote the SVD of (7) as

$$\bar{\mathbf{H}}_k = \bar{\mathbf{U}}_k\bar{\mathbf{\Sigma}}_k\bar{\mathbf{V}}_k^H = \bar{\mathbf{U}}_k\bar{\mathbf{\Sigma}}_k\left[\bar{\mathbf{V}}_k^1\bar{\mathbf{V}}_k^0\right]^H \quad (9)$$

where $\bar{\mathbf{V}}_k \in \mathbb{C}^{M_T \times M_T}$. $\bar{\mathbf{V}}_k^0 \in \mathbb{C}^{M_T \times \bar{\nu}_k}$ holds the $\bar{\nu}_k = M_T - \text{rank}\left(\bar{\mathbf{H}}_k\right)$ right column vectors, which define a basis for the null space of $\bar{\mathbf{H}}_k$. $\bar{\mathbf{V}}_1^0$ is defined as $\mathbf{I}_{M_T}$. The precoding matrix $\mathbf{W}_{\pi(k)}$ is constructed from the columns of $\bar{\mathbf{V}}_k^0$.

Assuming the transmitted signal vectors to be Gaussian distributed [4,5], for a given set of K_0 users and a specific ordering $\pi = \pi_i$ of those users, the maximum achievable rate of each user is given by

$$R_{\pi_i(k)} = \log_2 \frac{\left|\mathbf{I} + \mathbf{H}_{\pi_i(k)}\left(\sum_{j=1}^{k} \bar{\mathbf{V}}_j^0\mathbf{B}_{\pi_i(j)}\left(\bar{\mathbf{V}}_j^0\right)^H\right)\mathbf{H}_{\pi_i(k)}^H\right|}{\left|\mathbf{I} + \mathbf{H}_{\pi_i(k)}\left(\sum_{j=1}^{k-1} \bar{\mathbf{V}}_j^0\mathbf{B}_{\pi_i(j)}\left(\bar{\mathbf{V}}_j^0\right)^H\right)\mathbf{H}_{\pi_i(k)}^H\right|} \quad (10)$$

where the precoder input covariance matrices $\mathbf{B}_{\pi_i(k)}$ and the channel input covariance matrices $\mathbf{Q}_{\pi_i(k)}$ of the users are defined such that $\mathbf{Q}_{\pi_i(k)} = \mathbf{W}_{\pi_i(k)}\mathbf{W}_{\pi_i(k)}^H = \bar{\mathbf{V}}_k^0\mathbf{B}_{\pi_i(k)}\left(\bar{\mathbf{V}}_k^0\right)^H$.

The achievable sum rate of SZF precoding for a given user order π_i is

$$R_{\text{SZF}}^{\pi_i} = \max_{\{\mathbf{Q}_{\pi_i(k)}\}_{k\in\{1,\dots,K_0\}}:\mathbf{Q}_{\pi_i(k)}\succcurlyeq\mathbf{0},\sum_k \text{Tr}(\mathbf{Q}_{\pi_i(k)})\leq P} \sum_{k=1}^{K_0} R_{\pi_i(k)} \quad (11)$$

The maximum achievable sum rate R_{SZF} of SZF precoding is then obtained by maximizing (11) over all $K_0!$ possible user orders:

$$R_{\mathrm{SZF}} = \max_{\pi_i, i=1,2,\ldots,K_0!} R_{\mathrm{SZF}}^{\pi_i} \tag{12}$$

C. Further comments on SZF

We note that for both BD and SZF, it is possible to support additional users beyond what is described above by not transmitting the maximum number of streams N_k to certain users. However, due to the zero-forcing constraints, it is insufficient to simply reduce the number of data streams alone. For example, say a user is to receive data in a null space of rank two, but is only sent one data stream. It will, in general, be optimal to send that sole data stream using a linear combination of both null space basis vectors rather than just using one of the vectors. Thus, in such a case, the number of supportable users would still be unchanged. Instead, the transmitter requires some knowledge about the receive processing at the users; this will increase the rank of the effective null space for other users through consideration of the effective $d_k \times M_T$ channel matrices $\mathbf{H}_{eff,k} = \mathbf{R}_k\mathbf{H}_k$, where $\mathbf{R}_k \in \mathbb{C}^{d_k \times N_k}$ is the receive-processing matrix of user k, and d_k is the number of data streams sent to user k. In other words, the effective rank of a user's channel would be reduced, thus allowing further users to be scheduled. For example, the system could use coordinated beamforming with joint design of the transmit precoding and receive-processing matrices [4,12]. Research has indicated that such a joint consideration is in fact necessary to truly maximize the (weighted) sum rate of the MIMO BC under linear precoding (see for example [13]). However, this also requires a great deal of additional complexity. As mentioned, the transmitter must know how each receiver is processing its data, or at least make some assumption on its part regarding that processing. For example, a receiver with linear processing may possibly use receive antenna selection [13,14], SVD-based processing [3,14] ([15] also uses this to reduce the rank of effective channels), or minimum mean-squared error processing [12]. Alternately, the transmitter could calculate processing matrices/filters for the receivers [16,17]. Either way, there would generally be a significant amount of additional overhead to calculate and/or signal this information between the transmitter and receivers, above and beyond that required to obtain channel state information. With a large pool of active users, this also would increase the complexity of scheduling, as the base station would now also have to optimize the allocation of data streams as well as the selection of users and their ordering. Nevertheless, the SZF optimization method we propose later would also work in this case if the users' channel matrices were replaced with their effective channel matrices, which incorporate the receive-processing matrices.

In theory, a better overall system performance could possibly be obtained by a general linear precoding scheme without the zero-forcing constraints of SZF. Those constraints do result in a restriction on the degrees of freedom that the transmitter has to transmit data to users. However, from a practical standpoint, we wish to note that often when linear precoding is implemented in practice, the base station may be restricted in its choice of precoding vectors and/or covariance matrices. For instance, the base station may have to perform precoding based on selections from a codebook [18,19]. ZFB and BD are then approximated at the base station by selecting beamforming vectors for a given user that are aligned with that user's channel, but are not aligned with (i.e., near-orthogonal to) the channels of the other scheduled users. SZF can thus also be approximated in much the same way, with about the same complexity. The difference is that instead of selecting vectors unaligned with all other users, the base station would instead select ones that are unaligned with only the previously considered users.

It is lastly worth noting that SZF can also be combined with DPC to remove MUI. The precoding of DPC partially removes MUI, while the zero-forcing constraints of SZF remove the remaining MUI. This is referred to as SZFDPC in [5], and is a generalization of zero-forcing DPC [20] to users with multiple receive antennas. SZFDPC is known to asymptotically achieve the sum rate of DPC in the low and high SNR regimes, at low SNR with optimal ordering, and at high SNR for any ordering. We do not consider SZFDPC in this paper, though; we restrict our focus to just the linear precoding of SZF.

3. SZF Covariance Optimization

Attempting to solve (10)-(12) to determine optimal covariance matrices for the SZF sum rate is quite complex. The optimization in (11) is non-convex, unlike that of BD, where the complete decoupling of the users' effective channels creates a convex problem. Thus, finding a global optimum can be difficult. In the less-constrained case of DPC, the issue of non-convexity for the broadcast channel can be avoided by operating on the dual MAC instead. The capacity region of the MAC equals that of the BC, and the covariance matrices for one can be found from those of the other [2]. Moreover, the MAC capacity is convex, so the globally optimum solution for it (and thus indirectly for the BC) can be found relatively easily. Unfortunately, the transformation does not support the additional null space constraints in SZF. There do exist alternative, more general MAC-BC

dualities and transformations, such as mean-squared-error duality [21], signal-to-interference-plus-noise ratio (SINR) duality [12,22], and rate duality [23], which also account for linear precoding. However, the results for these dualities indicate that even if the null space constraints can be accounted for in the transformations (which is not necessarily guaranteed), the problem on the dual MAC would still be a non-convex problem. Thus, regardless of operating on the MAC or the BC, finding the global optimum would remain difficult. Finding a local optimum solution is somewhat easier, although how far that solution is from the global optimum may be uncertain.

A. Existing method

For a given user order π, the authors in [5] have proposed a suboptimal DPC-based numerical technique to solve (11). This technique makes use of the duality between the MIMO BC and MAC. The proposed technique uses the following steps [5]:

1. Using some optimization method for the MAC, such as the one in [6], find optimal MAC covariance matrices $\mathbf{P}_{\pi(k)}$ for the transmit power constraint P.
2. For the given order π, convert the MAC covariance matrices $\mathbf{P}_{\pi(k)}$ to BC DPC covariance matrices $\mathbf{D}_{\pi(k)}$ using the transformation method described in [2] (assuming user 1 is encoded last in DPC).
3. For users $\pi(1)$, $\pi(2)$, ..., $\pi(K_0 - 1)$, where user $\pi(1)$ is ordered first in SZF, project the DPC matrices $\mathbf{D}_{\pi(k)}$ to the SZF null space constraints by

$$\mathbf{Q}_{\pi(k)} = \bar{\mathbf{V}}_k^0(\bar{\mathbf{V}}_k^0)^H \mathbf{D}_{\pi(k)} \bar{\mathbf{V}}_k^0(\bar{\mathbf{V}}_k^0)^H \tag{13}$$

4. For user $\pi(K_0)$,
   (a) Find a new $\mathbf{D}_{\pi(K_0)}$ by waterfilling over the effective channel matrix

$$\mathbf{H}_{\text{eff}} = \left[\mathbf{I} + \mathbf{H}_{\pi(K_0)} \left(\sum_{i=1}^{K_0-1} \mathbf{Q}_{\pi(i)}\right) \mathbf{H}_{\pi(K_0)}^H\right]^{-1/2} \times \mathbf{H}_{\pi(K_0)} \bar{\mathbf{V}}_{K_0}^0 (\bar{\mathbf{V}}_{K_0}^0)^H \tag{14}$$

   with the power constraint $P_{\pi(K_0)} = P - \sum_{k=1}^{K_0-1} \text{Tr}(\mathbf{Q}_{\pi(k)})$.
   (b) Obtain $\mathbf{Q}_{\pi(K_0)}$ from (13) with the new $\mathbf{D}_{\pi(K_0)}$.

We note that the sum-rate expression for DPC is essentially the same as that for SZF, except without the null space constraints, and for a reversed ordering. Thus, the DPC optimization is basically a relaxation of the SZF optimization. As noted in [5], if the projections in (13) were unitary, the optimal SZF waterfilling solution would be obtained. Unfortunately, $\bar{\mathbf{V}}_k^0(\bar{\mathbf{V}}_k^0)^H$ is generally not unitary.

B. Problems with existing method

The above suboptimal method performs reasonably well. The sum rate of SZF exceeds that of BD in the simulation results provided in [5]. However, we have found two main deficiencies with the existing method. The first was discovered during our related work on scheduling for BD in [7] and SZF in [8,9]. We found in one case at high SNR, our scheduling algorithms for SZF performed significantly worse than the equivalent ones for BD at the same SNR. To isolate the problem, we ran an optimal exhaustive search scheduling algorithm. It was found that even with optimal scheduling, the throughput for SZF was worse than that for BD. This does not make sense; the BD optimization problem is a more constrained version of the SZF optimization problem. Any solution that satisfies the BD constraints also satisfies the SZF constraints. Therefore, the performance of SZF must always be no worse that of BD. With the problem lying not in the scheduling algorithm, we were able to determine the deficiency to be in the SZF covariance optimization method. Subsequent work, which we shall describe later in this paper, showed that the existing method becomes increasingly suboptimal as the number of supportable users K_0 and/or the SNR increase. The authors in [5] acknowledge that their method is suboptimal, and that better methods can likely be found, but to date, we are unaware of any results in the literature examining exactly how suboptimal the existing method is.

The second deficiency is that the existing covariance method only accounts for maximization of a pure (unweighted) sum rate. However, the method cannot be directly applied to a weighted sum-rate (WSR) maximization; i.e., to maximize $\sum_k w_{\pi(k)} R_{\pi(k)}$, where $w_{\pi(k)}$ is a weight for user $\pi(k)$ and $R_{\pi(k)}$ is as defined in (10). It may be possible to extend the method of [5] to a WSR by first solving a WSR maximization for the MAC, then proceeding as normal. However, a WSR maximization for the MAC (and thus for the BC due to duality) is found for one specific ordering of users. Namely, it is known that users should be decoded on the MAC in the reverse order of the size of the weights of the users, such that the user with the largest weight is decoded last [24,25]. Equivalently, the user with the largest weight should be encoded first on the BC. Because of this, the existing method's transformations and projections may not be the best if a different user ordering is to be considered for the SZF WSR. Neither is it necessarily the case that the same ordering that is optimal to

maximize the WSR for SZF is the same ordering required for the MAC/BC.

In the following section, we propose a new method for SZF covariance optimization that addresses both of these issues.

C. Proposed conjugate gradient projection method

We propose a CGP algorithm to optimize the covariance matrices for SZF. CGP methods are particularly useful in MIMO systems, as the solutions can be found using gradients and functions of complex-valued matrix variables. Some other optimization methods are only well defined for functions of real-valued vectors, so in those circumstances the covariance matrices and functions would have to be decoupled and expressed in terms of those vectors. CGP algorithms or gradient projection algorithms have been used for covariance optimization in other similar circumstances. For example, CGP is used in a weighted MAC sum-rate maximization in [25] and [26], and a gradient projection method is used for MIMO interference systems in [27] and for the MIMO MAC in [28]. We model our CGP algorithm after the one in [26], which operates on transmit filter matrices $\mathbf{T}_u$ instead of on the covariance matrices $\mathbf{Q}_u$ directly. This method has the advantage of guaranteeing a positive semidefinite covariance matrix $\mathbf{Q}_u = \mathbf{T}_u\mathbf{T}_u^H$ (this is a Cholesky decomposition [29]). Operating on $\mathbf{Q}_u$ directly would require a projection during each iteration to ensure the solution is in the set of positive semidefinite matrices (cf. [25]).

Let us rewrite and combine (10) and (11) to account for a weighted sum rate. Without loss of generality, we assume $\pi(k) = k$ for brevity of notation:

$$R_{\mathrm{WSZF}} = \max_{\mathbf{B}_k \succcurlyeq 0, \sum_k \mathrm{Tr}(\mathbf{B}_k) \le P} \sum_{k=1}^{K_0} w_k \log_2 \frac{\left|\mathbf{I} + \mathbf{H}_k \left(\sum_{i=1}^{k} \bar{\mathbf{V}}_i^0 \mathbf{B}_i (\bar{\mathbf{V}}_i^0)^H\right) \mathbf{H}_k^H\right|}{\left|\mathbf{I} + \mathbf{H}_k \left(\sum_{i=1}^{k-1} \bar{\mathbf{V}}_i^0 \mathbf{B}_i (\bar{\mathbf{V}}_i^0)^H\right) \mathbf{H}_k^H\right|} \quad (15)$$

Note in the above that $\mathrm{Tr}(\mathbf{Q}_k) = \mathrm{Tr}\left[\bar{\mathbf{V}}_k^0 \mathbf{B}_k (\bar{\mathbf{V}}_k^0)^H\right] = \mathrm{Tr}\left[\mathbf{B}_k (\bar{\mathbf{V}}_k^0)^H \bar{\mathbf{V}}_k^0\right] = \mathrm{Tr}(\mathbf{B}_k)$, as the columns of $\bar{\mathbf{V}}_k^0$ are orthonormal, so $(\bar{\mathbf{V}}_k^0)^H \bar{\mathbf{V}}_k^0 = \mathbf{I}$. Thus, there is the same transmit power constraint on $\mathbf{B}_k$ as on $\mathbf{Q}_k$.

Let us further define $\mathbf{B}_k = \mathbf{T}_k\mathbf{T}_k^H$, where $\mathbf{T}_k$ is a $\bar{v}_k \times \min(\bar{v}_k, N_k)$ matrix. Thus, $\mathbf{W}_k = \bar{\mathbf{V}}_k^0 \mathbf{T}_k$. Defining $\mathbf{T}_k$ in such a manner helps reduce the complexity of the optimization by reducing the number of optimization variables [5]. The power constraint can also be re-expressed as $\sum_k \|\mathbf{T}_k\|_F^2 \le P$, since $\|\mathbf{T}_k\|_F^2 = \mathrm{Tr}(\mathbf{B}_k)$. The CGP algorithm that operates on $\mathbf{T}_k$ is described as Algorithm 1 in the following.

Algorithm 1 CGP Algorithm for SZF covariance optimization

Initialize: $\mathbf{T}_k$; $\mathbf{S}_k = \mathbf{0}$, $\forall k$; $\rho = 1$; $\alpha = 1$.
Calculate *WSR* from (15).
repeat
Store $\mathbf{T}_{k\_old} = \mathbf{T}_k$, $\forall k$; $\mathbf{S}_{k\_old} = \mathbf{S}_k$, $\forall k$; $\rho_{old} = \rho$; $WSR_{old} = WSR$.
Calculate gradients: $\mathbf{G}_k$, $\forall k$ from (16)
Normalize gradients: $\bar{\mathbf{G}}_k = \sqrt{\frac{P}{\sum_k \| \mathbf{G}_k \|_F^2}}\mathbf{G}_k, \forall k$
Project gradients: $\hat{\mathbf{G}}_k = \bar{\mathbf{G}}_k - \frac{\sum_k \mathrm{Tr}(\mathbf{T}_k^H \bar{\mathbf{G}}_k)}{\sum_k \mathrm{Tr}(\mathbf{T}_k^H \mathbf{T}_k)}\mathbf{T}_k, \forall k$
Calculate Frobenius norm: $\rho = \sum_k ||\hat{\mathbf{G}}_k||_F^2$
Determine search directions: $\mathbf{S}_k = \hat{\mathbf{G}}_k + \frac{\rho}{\rho_{old}}\mathbf{S}_{k\_old}, \forall k$
Step in search directions: $\hat{\mathbf{T}}_k = \mathbf{T}_{k\_old} + \alpha\mathbf{S}_k, \forall k$
Normalize transmit filter sum-power: $\mathbf{T}_k = \sqrt{\frac{P}{\sum_k ||\hat{\mathbf{T}}_k||_F^2}}\hat{\mathbf{T}}_k, \forall k$
Calculate *WSR* from (15).
Set *LoopCounter* = 0.
while $WSR < WSR_{old}$ **do**
Decrease step size α.
Set $\mathbf{S}_k = \hat{\mathbf{G}}_k, \forall k$.
LoopCounter = *LoopCounter* + 1
if *LoopCounter* = *LoopThresh* **then**
Set $WSR_{old} = WSR$.
Reset α to 1.
end if
Recalculate $\hat{\mathbf{T}}_k$, $\mathbf{T}_k$, and *WSR*.
end while
until desired accuracy reached

Because the SZF WSR maximization problem is not convex, the CGP algorithm may not necessarily find the global optimum. Furthermore, the optimal $\mathbf{T}_k$ is not necessarily unique, since $\mathbf{B}_k$ is positive semidefinite. (For example, multiply $\mathbf{T}_k$ by any unitary matrix, and the new $\mathbf{T}_k$ will yield the same $\mathbf{B}_k$, and thus the same WSR.) The local optimum that the algorithm finds is also to some degree dependent on the initial values for $\mathbf{T}_k$. Often, when optimizing covariance matrices, an initial choice of a scaled identity matrix is used, but this in general cannot be done here, as generally $\mathbf{T}_k$ is not a square matrix. Furthermore, even if the algorithm was operating on $\mathbf{B}_k$ instead of $\mathbf{T}_k$, a scaled identity would still not be an appropriate starting point, as the rank would likely be too large; the rank of $\mathbf{B}_k$ would be $\bar{v}_k$ instead of $\min(\bar{v}_k, N_k)$. Instead, we initialize $\mathbf{T}_k$ by distributing values of $\sqrt{P/K_0/\bar{v}_k}$ to the columns of $\mathbf{T}_k$ in a round-robin fashion. This is equivalent to creating a

$\min(\bar{v}_k, N_k) \times \min(\bar{v}_k, N_k)$ identity matrix, vertically concatenating copies of the rows of that identity matrix until there are $\bar{v}_k$ rows, then finally multiplying by $\sqrt{P/K_0/\bar{v}_k}$. For example, if $\mathbf{T}_k$ was 3 × 2, entries (1,1), (2,2), and (3,1) of $\mathbf{T}_k$ would be initialized to $\sqrt{P/K_0/3}$, while the remaining entries would be 0.

The gradient can be calculated using matrix calculus from the partial differential of (15) with respect to $\mathbf{T}_k^H$ [30]. Specifically, $\nabla_k = 2\frac{\partial R_{\mathrm{WSZF}}}{\partial \mathbf{T}_k^*} = 2\left[\frac{\partial R_{\mathrm{WSZF}}}{\partial \mathbf{T}_k^H}\right]^T$. Since the gradients will be normalized, leading constants can be left off. It can be shown that the gradient for user k is proportional to

$$\begin{aligned}\mathbf{G}_k = &\left(\bar{\mathbf{V}}_k^0\right)^H \left(\sum_{i=k}^{K_0} w_i \mathbf{H}_i^H \left[\mathbf{I} + \mathbf{H}_i \left(\sum_{j=1}^{i} \bar{\mathbf{V}}_j^0 \mathbf{T}_j \mathbf{T}_j^H \left(\bar{\mathbf{V}}_j^0\right)^H\right) \mathbf{H}_i^H\right]^{-1} \mathbf{H}_i \right. \\ &\left. - \sum_{i=k+1}^{K_0} w_i \mathbf{H}_i^H \left[\mathbf{I} + \mathbf{H}_i \left(\sum_{j=1}^{i-1} \bar{\mathbf{V}}_j^0 \mathbf{T}_j \mathbf{T}_j^H (\bar{\mathbf{V}}_j^0)^H\right) \mathbf{H}_i^H\right]^{-1} \mathbf{H}_i \right) \bar{\mathbf{V}}_k^0 \mathbf{T}_k\end{aligned} \tag{16}$$

The above gradients seem quite complex at first glance. However, some computational savings can be obtained by a successive calculation of part of the gradients. To begin, the sums $\mathbf{\Phi}_i = \sum_{j=1}^{i} \bar{\mathbf{V}}_j^0 \mathbf{T}_j \mathbf{T}_j^H (\bar{\mathbf{V}}_j^0)^H$ can first be calculated and stored for each $i = 1, \ldots, K_0$ to avoid calculating these sums multiple times. Next, we can define $\mathbf{Z}_k$ as

$$\begin{aligned}\mathbf{Z}_k = &w_k \mathbf{H}_k^H [\mathbf{I} + \mathbf{H}_k \mathbf{\Phi}_k H_k^H]^{-1} \mathbf{H}_k \\ &- w_{k+1} \mathbf{H}_{k+1}^H [\mathbf{I} + \mathbf{H}_{k+1} \mathbf{\Phi}_k \mathbf{H}_{k+1}^H]^{-1} \mathbf{H}_{k+1}\end{aligned} \tag{17}$$

Then, each gradient $\mathbf{G}_k$ can be calculated starting from $k = K_0$ downwards, using a running sum for $\mathbf{Z}_k$. For example, if $K_0 = 4$, $\mathbf{G}_4 = \left(\bar{\mathbf{V}}_4^0\right)^H (\mathbf{Z}_4) \bar{\mathbf{V}}_4^0 \mathbf{T}_4$, $\mathbf{G}_3 = \left(\bar{\mathbf{V}}_3^0\right)^H (\mathbf{Z}_3 + \mathbf{Z}_4) \bar{\mathbf{V}}_3^0 \mathbf{T}_3$ and so on.

In [26], the authors define aggregate matrices **G**, **S**, and **T**, which are the horizontal concatenation of the matrices $\mathbf{G}_u$, $\mathbf{S}_u$, and $\mathbf{T}_u$, respectively. This primarily allows them to avoid the summation of squared F-norms and traces in the notation for their algorithm. For example, in the gradient normalization step, $\sum_u ||\mathbf{G}_u||_F^2$ can be represented more compactly as $||\mathbf{G}||_F^2$. This notation, strictly speaking, is not possible with our adaptation for SZF, as the gradients $\mathbf{G}_k$ and matrices $\mathbf{T}_k$ are generally of different dimensions for each k. An equivalent notation could still be used by instead defining aggregate matrices as a block-diagonal formation of the component matrices instead of a horizontal concatenation. However, this could potentially require additional memory and computational complexity unless the algorithm can account for the sparseness of the aggregate matrices (i.e., the many matrix entries after block-diagonalization that equal zero), and is not strictly necessary in the first place.

In the "step in search directions" portion of the algorithm, it is possible to find an approximately best step size, for example via an inexact line search like Armijo's Rule [31]. However, we find just as in [26] that it is generally sufficient to simply reduce the step size by a factor if there is no increase in the WSR. For example, we had good results when using equal weights of $w_k = 1$, $\forall k$, by simply multiplying α by about 0.8. We did, however, notice on rare occasions when the algorithm did not converge properly[a]. This is likely due to the non-convexity of the problem; the algorithm is likely stalling near a saddle point in these cases. Repeated decreases in α did not result in an increase in the WSR, and often led to a small decrease in the WSR. This may also be due to the fact that when a non-linear function is being optimized, an inexact line search (or lack of one, in our case) can lead to the search not being in the correct direction [31]. For example, if a function is being maximized, although the search should be in a direction of ascent, the search direction may actually be in one of descent. Thus, we implement the addition of a loop counter to compensate for these rare cases. If the loop counter reaches a certain threshold (we use a threshold of 100), the previous best WSR is set to the currently found value for the WSR, and α is reset[b] to 1. Since this updated value is often smaller than the previous value, there is a guaranteed larger value that the algorithm can head towards. This slight decrease in WSR and resetting of α is generally enough for the algorithm to get sufficiently far enough away from wherever it has stalled to continue finding a better solution (i.e., even better than where it stalled). If the algorithm's WSR still does not increase notably at this point, then it means the algorithm has found a local solution to the problem, as the change in WSR should be less than the desired accuracy. Thus, the algorithm can stop and return the current solution.

While we found that a decrease in the step size α by a factor of 0.8 worked well for our simulations, this value can likely be tuned depending on the specific system parameters and/or channel experienced by the users, to improve the convergence of the algorithm. The loop threshold, however, is best made dependent on the step size factor and the numerical precision of the system, and also optionally on the desired accuracy of the WSR. For instance, with our values of 0.8 and 100, note that $0.8^{100} \approx 2 \times 10^{-10} \approx 2^{-32}$. Thus, when the loop threshold is reached, the gradients would be changing around the 10th decimal place, or the 32nd bit of a floating point representation. Further decreases in the step size would result in changes in the gradients (and thus the WSR)

that are quite insignificant and likely below whatever accuracy of the solution that would be required in practice. Thus, our loop threshold of 100 is reasonably logical in conjunction with the step size factor of 0.8.

We lastly wish to note that the proposed algorithm, like the existing method from [5], is meant to find covariance matrices for a given selection of users and their order. The problem of user scheduling and the selection of an order is a complicated issue in and of itself, and for the most part outside the scope of this paper. Where scheduling is involved herein, we generally consider an optimal exhaustive search. The goal of this paper is rather to provide an improved algorithm that applies to whatever selection and order that the scheduler may wish to examine. The capability of examining alternative choices might be desired by the scheduler to, for example, meet certain QoS guarantees for the users, such as a minimum throughput.

D. Discussion of complexity

In [8,9], we calculate the complexity of the method for calculating covariance matrices from [5] in terms of the number of flops (floating point operations) required. It was found that the existing method has complexity order $\mathcal{O}(K_0M_T^3)$, assuming all users have the same number of receive antennas N. That is also the order of finding the null space basis vectors for SZF. Since our CGP algorithm also requires these vectors, it too must have a complexity order of at least $\mathcal{O}(K_0M_T^3)$. This in fact is exactly the order of complexity of the CGP algorithm, as the other steps have no greater complexity. M_T is the largest matrix dimension encountered, but it is never necessary to multiply two $M_T \times M_T$ matrices together (with complexity $\mathcal{O}(M_T^3)$ [32]) for all K_0 users. Each gradient requires a matrix inversion, but this is of an $N_k \times N_k$ matrix, which would have a complexity order $\mathcal{O}(N_k^3)$ [32]. The only other comparable order term is the multiplication of $(\bar{\mathbf{V}}_k^0)^H\mathbf{Z}_k$ for each user with complexity $\mathcal{O}(\bar{\nu}_k M_T^2)$, which when summed over all K_0 users also works out to $\mathcal{O}(K_0M_T^3)$.

In fact, any precoding method that requires calculating an SVD, a QR decomposition, or a pseudoinverse of an $M \times N$ or $N \times M$ matrix for each of K_0 users (for example, to find beamforming vectors for the precoder) will have complexity order $\mathcal{O}(K_0M^2N)$, where M is the larger matrix dimension [32]. As a comparison, again assuming all users have N receive antennas, the regularized BD method of [15] performs an SVD of a $(K_0\text{-}1)N \times M_T$ matrix for each user. As generally a system will be looking to schedule as many users and/or send as many data streams as possible, the product K_0N is of the same order as M_T, so the method of [15] is also approximately of order $\mathcal{O}(K_0M_T^3)$. The efficient WSR method of [16] is also technically a scheduling algorithm, so there would be a factor of K in its complexity. However, to compare it on equal terms, we will ignore the complexity of determining the best user to allocate a data stream to, and just assume that decision has been made. At each step i of the method, when allocating the ith stream, the method calculates the pseudoinverse of an $i \times M_T$ composite channel matrix to perform waterfilling. In the case where the base station transmits the maximum of M_T data streams, the total order of complexity would theoretically then be $\mathcal{O}\left(\sum_{i=1}^{M_T} M_T^2 i\right) = \mathcal{O}\left(M_T^4\right)$. However, some complexity savings might be possible if the pseudoinverse can be recursively calculated as each row is added to the composite channel matrix[c]. We are uncertain what the exact order of complexity would be in such a case, however.

In fairness, we note that while the precoding methods compared above scale with around the same order of complexity as our proposed CGP algorithm, the power allocation for those methods is of closed form instead of the iterative power allocation of our algorithm. Thus, there would be a smaller constant on the highest order term for those algorithms compared with the CGP algorithm.

4. Simulation results

In this section, we begin by presenting simulation results comparing the performance of our proposed SZF CGP covariance optimization method with the existing method. For comparison, we also present the performance when using BD, and for DPC to provide an upper bound on the achievable performance of the MIMO BC. We consider the case for which the transmitter has no knowledge of the receive-processing matrices of the users, and thus the maximum number of supported users for both BD and SZF is $K_0 = \lceil M_T/N \rceil$. This is the same as that considered in [5].

Later, we also compare the performance of our CGP algorithm to other existing precoding methods, namely the regularized BD (RBD) method and iterative RBD (IRBD) method from [15], and the efficient WSR method from [16]. We note, though, that RBD and IRBD are only meant for maximizing an unweighted sum rate, not a WSR. All three methods are capable of serving additional users by reducing the user null spaces with effective channel matrices when allocating data streams to the users. Thus, for a fair comparison, we also implement receive antenna selection (RAS) in conjunction with our CGP algorithm. RAS also reduces the effective rank of a user's channel, and requires a very small amount of extra overhead for the transmitter to

tell certain users to switch off given antennas; only the indices of the antennas would need to be sent. Given the notably increased complexity of having to search over all possible selections of antennas, we limit the comparison to the very simple scenarios also considered in [15] and [16]. (This increased scheduling complexity is also partially why we do not consider receive processing for the scenarios described in the previous paragraph, in addition to the extra signalling overhead.)

For all the simulations, we assume a spatially uncorrelated flat Rayleigh fading channel model, i.e., the elements of $\mathbf{H}_k$ are independent and identically distributed complex Gaussian random variables with a variance of 0.5 per dimension. Quasi-static block fading is assumed, such that the channel remains fixed for a given transmission interval, and changes independently between intervals. All users are assumed to have the same number of receive antennas. The SNR is defined as P/σ_n^2.

A. Comparison with existing method

We begin by comparing two simple cases also examined in [5]. The average sum rate is determined through a Monte Carlo simulation. Figure 1 shows the unweighted sum rate (i.e., all users have a weight of 1) for BD and the existing and proposed SZF covariance optimization methods. In this first case, $M_T = 4$, $K = K_0 = 2$, and $N_1 = N_2 = 2$. Figure 2 shows a second unweighted case with $M_T = 6$, $K = K_0 = 3$, and $N_1 = N_2 = N_3 = 2$. In both of these cases, strictly speaking scheduling is not necessary, as the number of available users K equals the number of simultaneously supportable users K_0. However, we do still consider all possible subsets of those users, and all possible orders of those users for SZF, to find the ordered selection that gives the maximum sum rate.

It can be seen that at low SNR, there is essentially no difference between our proposed CGP algorithm and the existing SZF covariance optimization method from [5]. However, as the SNR increases, there is an increasing gain in the throughput of our proposed algorithm relative to the existing algorithm. In Figure 1, the gains are rather modest; the sum rate is about 0.35 bit/s/Hz larger at 10 dB, about 0.65 bit/s/Hz larger at 14 dB, and about 0.6 bit/s/Hz larger at 20 dB. These represent percentage gains of about 3.5, 5, and 3%, respectively. However, the gains are much more significant in Figure 2. The throughput increase is about 0.75 bit/s/Hz at 10 dB, and about 2.45 bit/s/Hz at 20 dB. This is a percentage gain of over 5 and 9%, respectively. More importantly, we note that the performance of the original method is worse than that of BD above an SNR of 16 dB. This result was not visible in [5], as the graph for $M_T = 6$, $K = 3$ in that paper only went up to 16 dB. In comparison, our proposed CGP algorithm performance is consistently above that of BD. We can thus see that the performance gains increase both with the number of supported users and with the SNR.

In Figure 3, we present a somewhat more complicated scenario more related to our scheduling work in [8,9]. In this case, we consider a larger user pool size of $K = 16$ with $M_T = 8$. Each user in the pool has $N_k = 2$ receive antennas, so at most $K_0 = 4$ users can be served simultaneously. We use an exhaustive search for

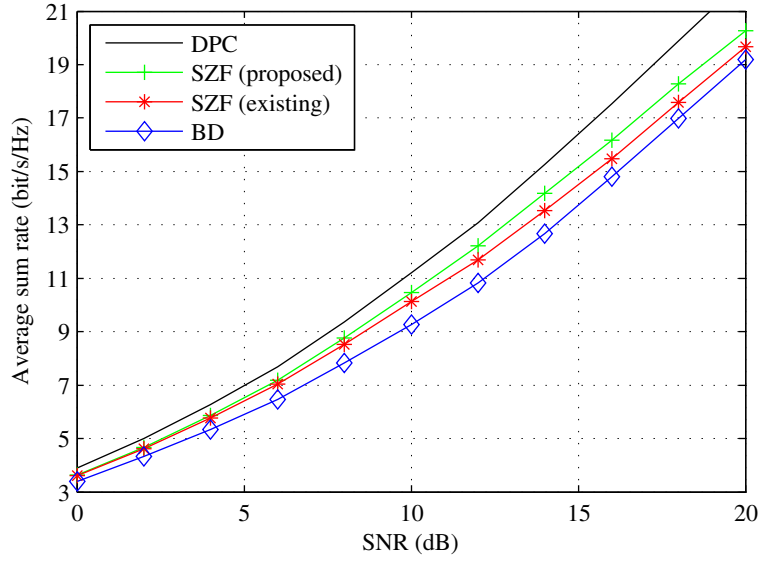

Figure 1 Average sum rate versus SNR with proposed and existing SZF covariance optimization methods and BD; $M_T = 4$, $K = K_0 = 2$, $N_1 = N_2 = 2$.

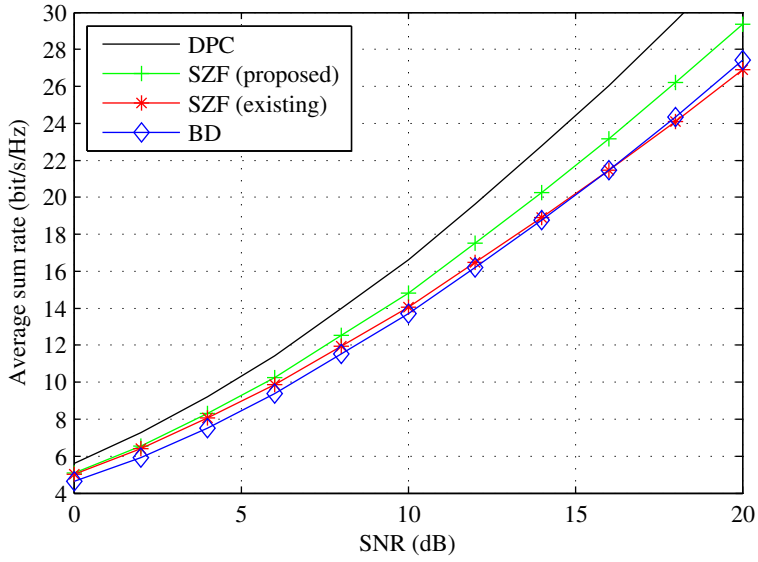

Figure 2 Average sum rate versus SNR with proposed and existing SZF covariance optimization methods and BD; $M_T = 6$, $K = K_0 = 3$, $N_1 = N_2 = N_3 = 2$.

scheduling that considers all possible subsets of users and all possible user orders for those subsets. This decouples the effect of the specific scheduling algorithm and allows us to focus on the performance of the SZF covariance optimization methods.

For the existing numerical covariance optimization method [5], the average sum rate for SZF quickly becomes less than that of BD, at just above 10 dB. However, the average sum rate for SZF using our CGP algorithm remains higher than that of BD at least up to an SNR of 20 dB. The improvement in performance over the existing algorithm is about 0.68 bit/s/Hz (about 5%) at 5 dB, about 1.5 bit/s/Hz (about 7%) at 10 dB, and about 4.85 bit/s/Hz (about 12%) at 20 dB.

We note, though, that the gain in throughput relative to BD starts to decrease at higher SNR. The throughput may likely become less than that of BD at an SNR somewhere larger than 20 dB. This would serve to

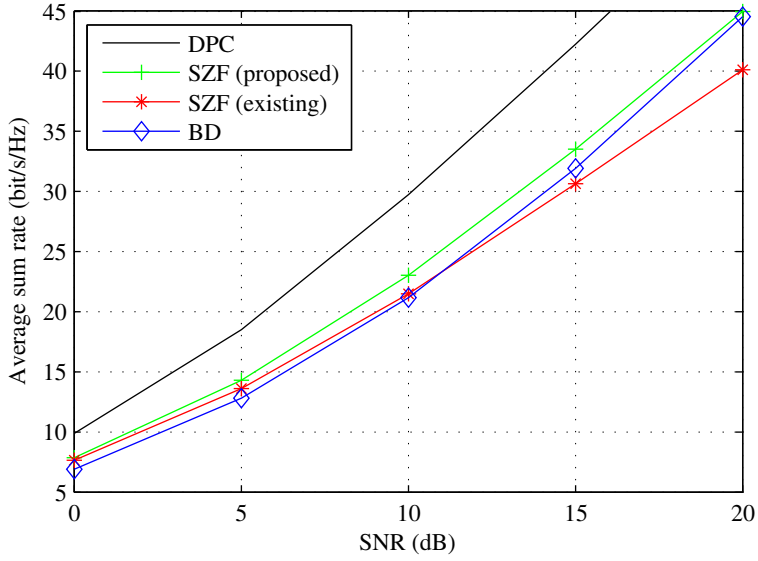

Figure 3 Average sum rate versus SNR with proposed and existing SZF covariance optimization methods and BD, using exhaustive search scheduling; $M_T = 8$, $K = 16$, $N_k = 2$ for each user, $K_0 = 4$.

demonstrate that our CGP algorithm, though improved, is still globally suboptimal. However, a worse performance than BD can be avoided with our algorithm. Rather than the round-robin initialization described in Section 3-C, instead the matrices $\mathbf{T}_k$ can be initialized based on the BD-optimal covariance matrices. For example, at 20 dB, the difference in the average sum rate between the round-robin initialization and the BD-optimal initialization is about 0.01 bit/s/Hz, which is negligible. Obtaining the BD-optimal matrices will require some additional complexity, due to an additional set of null space basis vector calculations and waterfilling. The added complexity may be partially offset, though, as the CGP algorithm might have to run for fewer iterations. Nevertheless, this for the most part should be unnecessary, as that extremely high of an SNR (or SINR, in the case of interference-limited cellular systems) is unlikely to be seen in practice.

We have also noticed that, while the existing SZF covariance method is worse than our proposed CGP method, the covariance matrices $\mathbf{Q}_{k,o}$ provided by the existing method sometimes provide a better starting point for our CGP than the round-robin initialization. This is particularly the case at high SNR. For example, in Figure 2, the average throughput for our CGP at 20 dB using $\mathbf{Q}_{k,o}$ for initialization increases from about 29.3 bit/s/Hz to about 29.7 bit/s/Hz. This extra throughput represents on average about an additional 1.4% increase. However, in most cases, this small additional throughput is likely not worth the added computational complexity. To get that extra percent, in effect, two optimizations must be run. The first is on the MAC (followed by transformations and projections) to find $\mathbf{Q}_{k,o}$, then a second with our CGP algorithm using $\mathbf{Q}_{k,o}$ for initialization.

Furthermore, as K increases, this effect seems to essentially disappear. If we consider now the scenario from Figure 3, there is virtually no difference in the average sum rate between the two initialization methods. Our simulations only showed an improvement of about 0.03 bit/s/Hz at 20 dB, which is certainly negligible and within the error margin of the simulation. It appears that the larger user pool and scheduling have the effect of mostly removing any initialization-based gains. In part, this is because the larger user pool means that the scheduled users' channels are closer to orthogonal. The larger pool also means that the scheduling algorithm has more options to choose a different set of users or encoding order that may negate any effect from the different initialization point.

B. Comparison of weighted sum rate

We now consider a simple scenario for a weighted sum rate. We examine the case where $M_T = 8$, $K = K_0 = 4$, and $N_k = 2$, $\forall k$. We set the weight for each user proportional to that user's index, i.e., $w_k = k/\sum_k w_k$. (The sum in the denominator is just for normalization and does not affect the rates the scheduled users receive.) Such a scenario might arise in practice if each user belongs to a different class of service, such as if they are carrying different types of traffic, or they have paid for higher average data rates. Figure 4 shows the WSR performance of our proposed CGP method relative to a WSR using BD. All possible user subsets and orderings are considered. Recall that there exists no prior method for weighted SZF covariance optimization, so we cannot compare our performance with any such algorithm. We observe that the WSR of SZF is larger than that when using BD. The SZF algorithm performs better than BD in this scenario by about 0.5 dB in SNR.

Our simulations for this scenario also indicate only a minor correlation between the best user ordering and the relative sizes of the weights. Figure 5 shows a histogram of how often each user index is ordered in a given position for the best obtained WSR. A user index of 0 indicates that no user has been encoded in that position (i.e., transmitting to less than the maximum supportable number of users maximizes the WSR). It can be seen that there is somewhat of a tendency to order the users in the decreasing order of their weights. This trend is strongest at lower SNRs. However, as the SNR increases, this trend diminishes. At 20 dB, for example, it is approximately equally likely that users 3 and 4 (with weights 0.3 and 0.4, respectively) will be ordered first. User 2 is ordered first about half as often as 3 or 4, but also ordered second about half as often as 3 or 4. Thus, there is no hard rule to determine the optimal user ordering for a WSR for SZF. This is in stark contrast to the MAC or when using DPC on the BC; e.g., on the MAC users should always be decoded in the increasing order of their weights.

We also note that even at high SNR, it is often best in terms of maximizing the WSR to not transmit to the maximum possible number of users. In this scenario, with the limited user pool to choose from, it is better to transmit to less than the maximum number about 43-57% of the time. We made a similar observation in [8,9] for unweighted sum rates when scheduling to small user pools. The likelihood of scheduling the maximum possible number of users increases with the size of the user pool K, due to multiuser diversity and the increased chance of finding users with orthogonal channels. This fact is unlikely to change using our proposed CGP optimization algorithm here.

C. Comparison with other precoding methods

We now compare the performance of our SZF CGP algorithm to some other precoding methods. We start with a case considered in [15], with $K = 3$ users, each

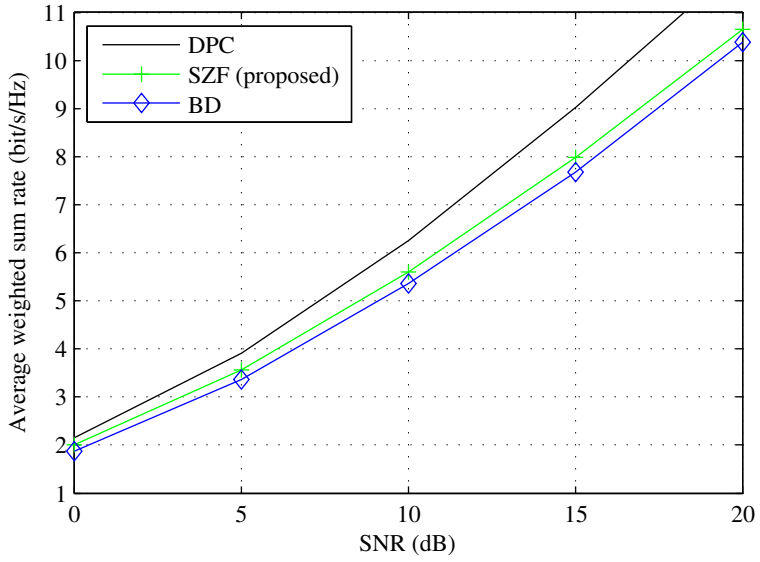

Figure 4 Average weighted sum rate versus SNR with proposed SZF covariance optimization method and with BD; $M_T = 8$, $K = K_0 = 4$, $N_k = 2$, $\forall k$, $w_k = k/10$.

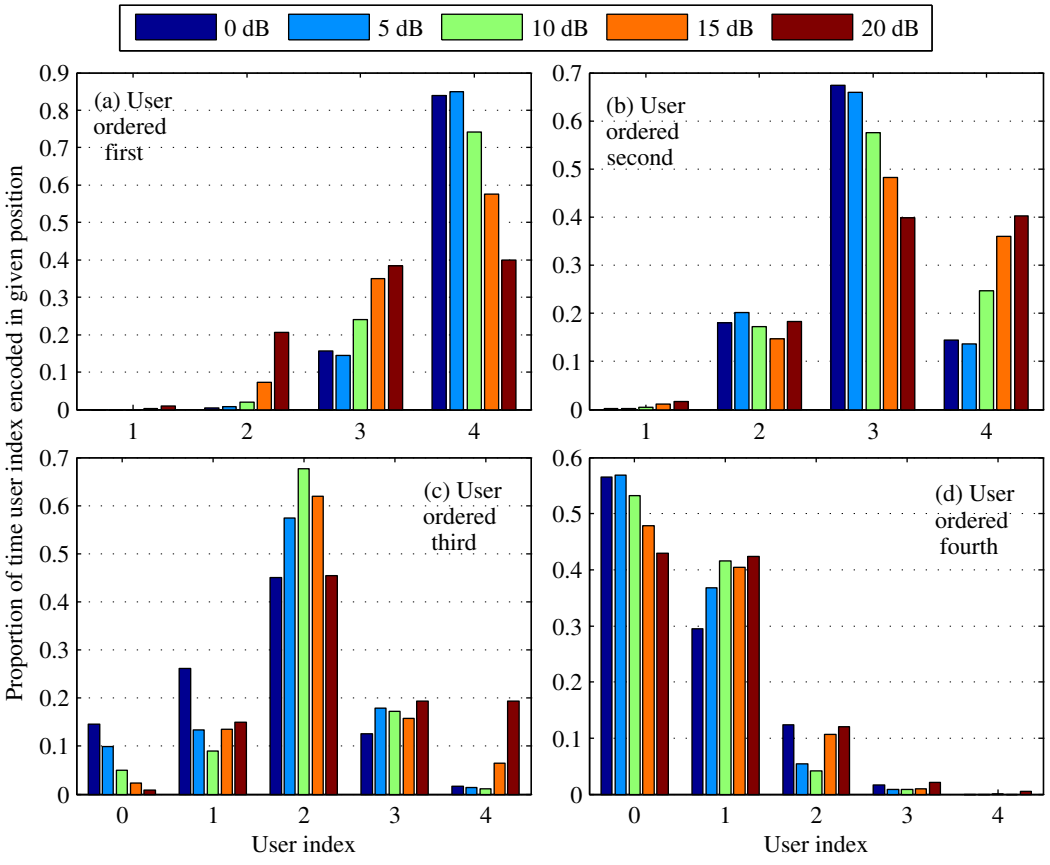

Figure 5 Proportion of time that user index is ordered in a given position to maximize SZF weighted sum rate, where user weights equal 1/10 of user indices ($w_k = k/10$); $M_T = 8$, $K = K_0 = 4$, $N_k = 2$ for each user. Index "0" indicates no user encoded in that position. (a) User index is ordered first. (b) User index ordered second. (c) User index ordered third. (d) User index ordered fourth.

with $N_k = 4$ receive antennas, served by a base station with $M_T = 4$ transmit antennas. Without considering receive processing, the system could only support one user at a time with BD or SZF. However, with receive processing, all three users could potentially be supported at once by transmitting a single stream to each of them (with one single user potentially receiving two streams). We consider RBD and IRBD from [15], the WSR method of [16] (with equal user weights of 1), and our SZF CGP algorithm with RAS and an exhaustive search over all user subsets, orders, and antennas. We also include the performance with DPC as an upper bound. The simulation results are shown in Figure 6.

It can be seen that both IRBD and the efficient WSR methods provide very similar performance, both yielding a throughput very close to that of DPC. IRBD is slightly better at higher SNR by about 0.3 bit/s/Hz. The performance of our CGP algorithm for SZF with RAS provides a lower throughput by about 1.2 bit/s/Hz, or with a loss of just under 1 dB in SNR. However, we note that RAS, though simple, is also a suboptimal receive-processing method. Were we to consider a better method of generating an effective channel matrix for the users, the performance would almost certainly improve. RBD is the worst of the methods examined by a significant margin, as unlike IRBD, it does not iteratively optimize the transmit filters for the users to account for their unused channel subspaces.

Furthermore, we compare the WSR performance in a scenario used in [16], with $K = 4$ users, each with $N_k = 3$ receive antennas, served by a base station with $M_T = 4$ transmit antennas. Users 1 and 2 have weights of $w_1 = w_2 = 2$, while users 3 and 4 have weights of $w_3 = w_4 = 1$. For this scenario, we cannot compare the performance of RBD or IRBD, since those methods do not support the maximization of a weighted sum rate. The simulation results are shown in Figure 7.

It can be seen that there is virtually no difference in the WSR achieved by the method of [16] and our SZF CGP algorithm with RAS in this scenario. This is despite the suboptimality of RAS. Both precoding methods achieve a weighted sum rate reasonably close to that of DPC, giving about 92-94% of the WSR of DPC, or with a loss of at most about 1.3 dB in SNR.

It is lastly interesting to note that the covariance matrices generated by the method of [16] satisfy null space constraints for their effective channels. Thus, those same matrices, and the selection of users and their effective channel matrices, can also be used as an initialization point for our SZF CGP algorithm. It is then only necessary to find the best ordering of the selected users for SZF. The CGP algorithm would operate on the effective channel matrices, but would not change the incorporated receive filter matrices, only the transmit covariance matrices. However, such an initialization only results in a negligible increase in the (weighted) sum rate beyond that already provided by the efficient WSR method. For example, in the scenario of Figure 6, the sum rate increases by less than 0.1 bit/s/Hz. For the scenario of Figure 7, the increase in WSR is less than 0.2 bit/s/Hz. We have used an exhaustive search to find the best user order in both cases.

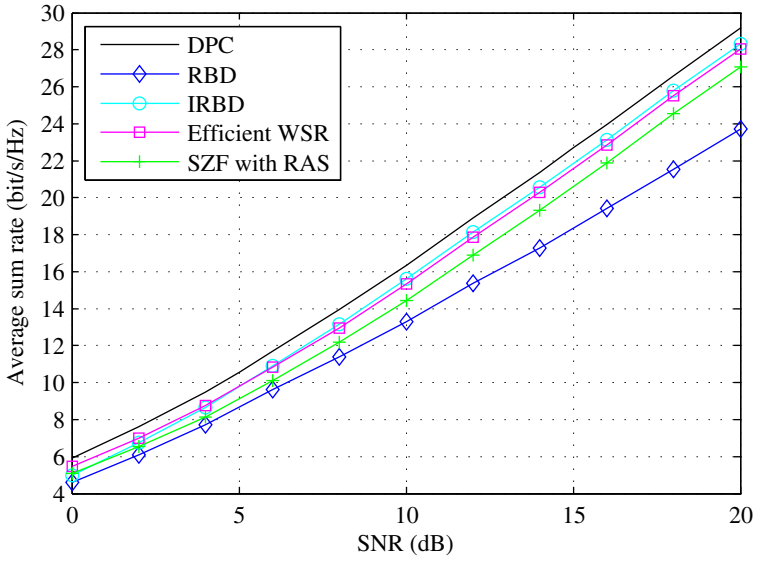

Figure 6 Average sum rate versus SNR of various precoding methods; $M_T = 4$, $K = 3$, $N_k = 4$ for each user.

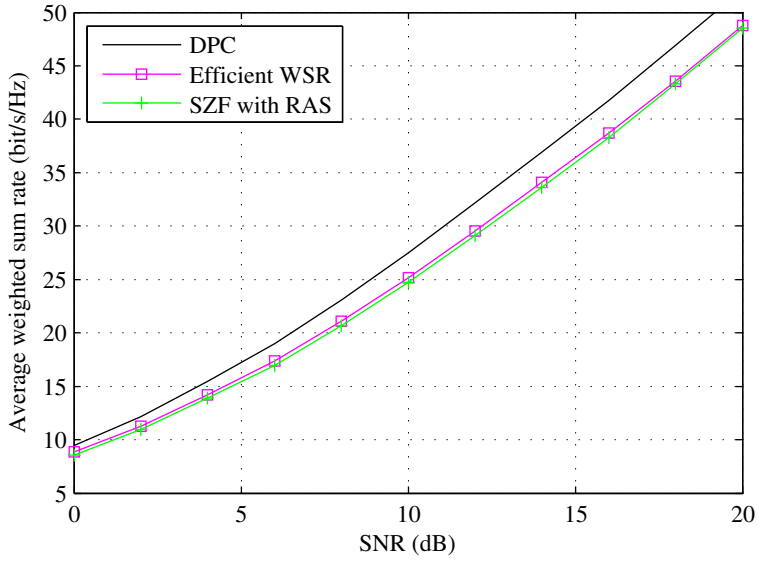

Figure 7 Average weighted sum rate versus SNR of various precoding methods; M_T = 4, K = 4, N_k = 3 for each user. User weights are $w_1 = w_2 = 2$, $w_3 = w_4 = 1$.

5. Conclusions

We have proposed and analyzed an improved method based on CGP for optimizing the covariance matrices in SZF precoding. This proposed method outperforms the existing method from [5] by up to an additional 12% in sum rate for the cases analyzed. It was also seen that there is an increasing gain in the performance of our method over the prior method both with increasing SNR and with higher numbers of simultaneously supportable users K_0. Our proposed method also consistently ensured a throughput larger than that when using BD; the throughput of the existing SZF covariance optimization scheme was seen to drop below that of BD at higher SNR and K_0.

Our CGP method also supports the maximization of a WSR using SZF. Such a weighted sum rate is important in various applications. To our knowledge, there is no prior method for WSR maximization using SZF in the literature. We demonstrated with a simple case that even when considering a WSR, our proposed method still provided a higher weighted throughput than when using BD.

We further compared SZF employing our CGP covariance optimization method with other precoding methods in the literature that can allocate data to users on a per-data-stream basis. For this comparison, we also incorporated RAS in the simulation of our method to adjust the size of the SZF null spaces available for users to receive data in. It was seen that our CGP algorithm provided comparable, though slightly worse, performance to those existing methods. The inferior performance was in part due to the suboptimality of RAS compared with the receiver filters and/or effective channel formations used in the other methods. We have also found that our CGP algorithm scales with about the same order of complexity as the other methods, though since those methods employ a closed-form power allocation compared with our iterative algorithm, the constant on the highest order complexity term for the other methods is smaller than for our CGP method.

The improvements on the SZF WSR we have seen herein with our proposed method have been for a relatively simple channel model of uncorrelated quasi-static Rayleigh fading and with perfect channel knowledge. Future work should also consider the effects of a more realistic scenario, including imperfect channel knowledge and temporal and/or spatial correlation.

Although our proposed method improves on the performance of the existing method, our method is still not globally optimal. Since the SZF optimization problem is non-convex, finding the global optimum is very difficult. It is thus hard to say how far away our scheme is from the global optimum for SZF. There are a few global optimization techniques which could find the best overall solution. For example, a branch-and-bound with reformulation linearization technique such as that described in [33,34] may assist in finding the global optimum. However, such techniques may be extremely complex and not meant for real-time implementation in practical systems. Nonetheless, the global optimization problem remains as possible future work.

Endnotes

[a]During our simulations, these rare cases seemed to primarily occur at low SNR.

[b]This is similar and related in concept to the notion of "restarting" the CGP search during non-linear optimizations, as discussed in [31].

[c]A similar reduction in complexity may also be possible for SZF by recursively calculating the null space basis vectors. However, we have not investigated this further, since the vectors only need to be calculated once when starting the CGP algorithm, and then used repeatedly during the iterations. Thus, any complexity savings would be minor.

Acknowledgements

The authors would like to thank Dr. Shreeram Sigdel for his comments on the paper. Our work made use of the infrastructure and computational resources of Academic Information and Communication Technologies (AICT) at the University of Alberta, and of the Western Canada Research Grid (WestGrid). The authors also gratefully acknowledge funding for this research provided by TRLabs, the Rohit Sharma Professorship, and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Author details

[1]Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada [2]TRLabs, Edmonton, AB T5K 2M5, Canada

Competing interests

The authors declare that they have no competing interests.

Received: 1 November 2010 Accepted: 14 October 2011
Published: 14 October 2011

References

1. MHM Costa, Writing on dirty paper. IEEE Trans Inform Theory **29**(3), 439–441 (1983). doi:10.1109/TIT.1983.1056659
2. S Vishwanath, N Jindal, A Goldsmith, Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. IEEE Trans Inform Theory **49**(10), 2658–2668 (2003). doi:10.1109/TIT.2003.817421
3. T Yoo, A Goldsmith, On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming. IEEE J Select Areas Commun. **24**(3), 528–541 (2006). doi:10.1109/JSAC.2005.862421
4. QH Spencer, AL Swindlehurst, M Haardt, Zero forcing methods for Downlink spatial multiplexing in multiuser MIMO channels. IEEE Trans Signal Processing **52**(2), 461–471 (2004). doi:10.1109/TSP.2003.821107
5. AD Dabbagh, DJ Love, Precoding for multiple antenna Gaussian Broadcast channels with successive zero-forcing. IEEE Trans Signal Processing **55**(7), 3837–3850 (2007). doi:10.1109/TSP.2007.894285
6. N Jindal, W Rhee, S Vishwanath, SA Jafar, A Goldsmith, Sum power iterative water-filling for multi-antenna Gaussian broadcast channels. IEEE Trans Inform Theory **51**(4), 1570–1580 (2005). doi:10.1109/TIT.2005.844082
7. S Sigdel, RC Elliott, WA Krzymień, M Al-Shalash, Greedy and genetic user scheduling algorithms for multiuser MIMO systems with block diagonalization, in *Proc IEEE Veh Technol Conf (VTC'09-Fall)*, 1–6 (2009). doi:10.1109/VETECF.2009.5378984
8. RC Elliott, S Sigdel, WA Krzymień, M Al-Shalash, ACK Soong, Genetic and greedy user scheduling for multiuser MIMO systems with successive zero-forcing, in *Proc 5th IEEE Broadband Wireless Access Workshop (2009 IEEE GLOBECOM Workshops)*, 1–6 (2009). doi:10.1109/GLOCOMW.2009.5360758
9. RC Elliott, S Sigdel, WA Krzymień, Low complexity greedy, genetic, and hybrid user scheduling algorithms for multiuser MIMO systems with successive zero-forcing. [Submitted in March 2011 for publication in *Eur. Trans. Telecommun.* under review]
10. FP Kelly, AK Maulloo, DKH Tan, Rate control for communication networks: shadow prices, proportional fairness, and stability. J Oper Res Soc. **49**(3), 237–252 (1998). doi:10.1057/palgrave.jors.2600523
11. A Jalali, R Padovani, R Pankaj, Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system, in *Proc IEEE Veh Technol Conf (VTC'00-Spring)*. **3**, 1854–1858 (2000). doi:10.1109/VETECS.2000.851593
12. M Codreanu, A Tölli, M Juntti, M Latva-aho, Joint design of Tx-Rx beamformers in MIMO downlink channel. IEEE Trans Signal Processing. **55**(9), 4639–4655 (2007). doi:10.1109/ICC.2007.825
13. BC Lim, WA Krzymień, C Schlegel, Efficient sum rate maximization and resource allocation in block-diagonalized space-division multiplexing. IEEE Trans Veh Technol. **58**, 478–484 (2009). doi:10.1109/TVT.2008.924973
14. F Boccardi, H Huang, A near-optimum technique using linear precoding for the MIMO broadcast channel, in *Proc IEEE Int Conf Acoustics, Speech and Signal Process. (ICASSP'07)*. **3**, III-17–III-20 (2007). doi:10.1109/ICASSP.2007.366461
15. V Stankovic, M Haardt, Generalized design of multi-user MIMO precoding matrices. IEEE Trans Wireless Commun. **7**(3), 953–961 (2008). doi:10.1109/LCOMM.2008.060709
16. C Guthy, W Utschick, R Hunger, M Joham, Efficient weighted sum rate maximization with linear precoding. IEEE Trans Signal Processing **58**(4), 2284–2297 (2010). doi:10.1109/TSP.2009.2040016
17. C-B Chae, D Mazzarese, T Inoue, RW Heath Jr, Coordinated beamforming for the multiuser MIMO broadcast channel with limited feedforward. IEEE Trans Signal Processing **56**(12), 6044–6056 (2008). doi:10.1109/TSP.2008.929869
18. M Trivellato, F Boccardi, H Huang, On transceiver design and channel quantization for downlink multiuser MIMO systems with limited feedback. IEEE J Select Areas Commun. **26**(8), 1494–1504 (2008). doi:10.1109/JSAC.2008.081015
19. IH Kim, DJ Love, On the capacity and design of limited feedback multiuser MIMO uplinks. IEEE Trans Inform Theory **54**(10), 4712–4724 (2008). doi:10.1109/TIT.2008.928997
20. G Caire, S Shamai (Shitz), On the achievable throughput of a multiantenna Gaussian broadcast channel. IEEE Trans Inform Theory **49**(7), 1691–1706 (2003). doi:10.1109/TIT.2003.813523
21. R Hunger, M Joham, W Utschick, On the MSE-duality of the broadcast channel and the multiple access channel. IEEE Trans Signal Processing **57**(2), 698–713 (2009). doi:10.1109/TSP.2008.2008253
22. M Codreanu, A Tölli, M Juntti, M Latva-aho, Uplink-downlink SINR duality via Lagrange duality, in *Proc IEEE Wireless Commun Netw Conf (WCNC'08)*, 1160–1165 (2008). doi:10.1109/WCNC.2008.209
23. R Hunger, M Joham, A general rate duality of the MIMO multiple access channel and the MIMO broadcast channel, in *Proc IEEE Global Telecommun Conf (GLOBECOM'08)*, 1–5 (2008). doi:10.1109/GLOCOM.2008.ECP.178
24. H Viswanathan, S Venkatesan, H Huang, Downlink Capacity Evaluation of Cellular Networks With Known-Interference Cancellation. IEEE J Select Areas Commun. **21**(5), 802–811 (2003). doi:10.1109/JSAC.2003.810346
25. J Liu, YT Hou, HD Sherali, On the Maximum Weighted Sum-Rate of MIMO Gaussian Broadcast Channels, in *Proc IEEE Int Commun Conf (ICC'08)*, 3664–3668 (2008). doi:10.1109/ICC.2008.689
26. R Böhnke, K-D Kammeyer, Weighted Sum Rate Maximization for the MIMO-Downlink Using a Projected Conjugate Gradient Algorithm, in *Proc Int Workshop on Cross Layer Design (IWCLD'07)*, 82–85 (2007). doi:10.1109/IWCLD.2007.4379043
27. S Ye, RS Blum, Optimized Signalling for MIMO Interference Systems with Feedback. IEEE Trans Signal Processing **51**(11), 2839–2848 (2003). doi:10.1109/TSP.2003.818339
28. R Hunger, DA Schmidt, M Joham, W Utschick, A general covariance-based optimization framework using orthogonal projections, in *Proc IEEE 9th Workshop on Signal Process. Advances in Wireless Commun. (SPAWC'08)*, 76–80 (2008). doi:10.1109/SPAWC.2008.4641573
29. RA Horn, CR Johnson, *Matrix Analysis* (New York, NY: Cambridge University Press, 1985)
30. KB Petersen, MS Pedersen, The Matrix Cookbook, (Technical University of Denmark, 2008) http://www2.imm.dtu.dk/pubdb/p.php?3274. [Version 20081114]
31. J Nocedal, SJ Wright, *Numerical Optimization*, 2nd edn. (New York, NY: Springer, 2006)

32. GH Golub, CF Van Loan, *Matrix Computations*, 3rd edn. (Baltimore, MD: The John Hopkins Univ Press, 1996)
33. J Liu, YT Hou, Y Shi, HD Sherali, S Kompella, On the capacity of multiuser MIMO networks with interference. IEEE Trans Wireless Commun. **7**(2), 488–494 (2008). doi:10.1109/TWC.2008.060732
34. J Liu, YT Hou, HD Sherali, Optimal power allocation for achieving perfect secrecy capacity in MIMO wire-tap channels, in *Proc 43rd Conf on Inform Sciences and Systems 2009 (CISS'09)*, 606–611 (2009). doi:10.1109/CISS.2009.5054790

doi:10.1186/1687-1499-2011-133
Cite this article as: Elliott and Krzymień: **Improved and weighted sum rate maximization for successive zero-forcing in multiuser MIMO systems.** *EURASIP Journal on Wireless Communications and Networking* 2011 **2011**:133.