

Bargaining and Multi-user Detection in MIMO Interference Networks

Matthew Nokleby*, A. Lee Swindlehurst†

* Electrical and Computer Engineering Department, Brigham Young University, Provo, Utah, 84602, USA
Email: nokleby@ee.byu.edu, Phone: (801) 422-5349

† Electrical and Computer Engineering Department, University of California, Irvine, CA, 92697, USA
Email: swindle@uci.edu, Phone: (949) 824-2818

Abstract—We investigate the use of multi-user detection to improve performance in MIMO interference networks. Unfortunately, while multi-user detection often allows higher data rates, it greatly complicates the problem: in addition to choosing a transmit covariance for each transmitter, we must decide which signals each receiver will detect and which data rates make such detection feasible. We discuss methods to optimize the data rates in two ways: maximizing the sum throughput of the network, and choosing rates based on the Kalai-Smorodinsky bargaining solution from cooperative game theory. Simulation results suggest that, while sum-rate maximization yields higher average throughput, the Kalai-Smorodinsky solution provides a superior solution in terms of fairness. The simulations also suggest that multi-user detection significantly improves network performance.

Index Terms—MIMO networks, Multi-user detection, Kalai-Smorodinsky solution.

I. INTRODUCTION

Mitigating the effects of mutual interference in networks composed of multiple-input multiple output (MIMO) nodes is critical to realizing their potential throughput advantages. In MIMO links, interference can be partially avoided by spatially coordinating users' signals. For example, in [1, 2], nodes spatially “steer” their transmissions to increase the total throughput of the network. However, interference-cancellation techniques based on multi-user detection can further improve the performance of wireless systems [3]. Conceptually, if a receiver can correctly decode an interfering signal, it can “subtract” that interference from the incoming signal, allowing the intended signal to be decoded more easily and at a higher data rate. However, successful multi-user detection imposes limitations on the data rates of the interfering transmitters so that their signals may be detected and subtracted out.

We study the information-theoretic advantages of incorporating multi-user detection in optimizing users' data rates in a MIMO interference network. Each transmitter must choose an input covariance matrix, which spatially characterizes the transmitted signal, and each receiver must decide which interfering signals to detect. These decisions define a set of feasible

data rates. We therefore choose covariances and detection decisions in order to maximize users' rates.

Of course, in a network of interfering links, it is impossible to simultaneously maximize each user's rate. So, we choose the optimal rates in two different ways. First, we simply maximize the sum of the users' rates, or the total network throughput. However, maximizing throughput often results in solutions where weaker links are forced to transmit at low data rates. Therefore, we also consider the Kalai-Smorodinsky bargaining solution [4] from cooperative game theory. This approach axiomatically defines an efficient solution that explicitly considers individual users' rates. Our simulation results suggest multi-user detection significantly improves the performance in both solutions. They also suggest that the Kalai-Smorodinsky solution significantly improves the fairness of the users' rates, but has lower total throughput than sum-rate maximization.

II. SYSTEM MODEL

A. Signal Model

In a MIMO interference system, we have L point-to-point links; that is, L unique transmitters send data to L unique receivers. We assume that each transmitter and each receiver has N antennas, although our results easily generalize. We assume a narrowband model where the i th transmitter sends complex baseband signal \mathbf{x}_i . The signal at the i th receiver is

$$\mathbf{y}_i = \mathbf{H}_{i,i}\mathbf{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^L \mathbf{H}_{i,j}\mathbf{x}_j + \mathbf{n}_i, \quad (1)$$

where $\mathbf{H}_{i,j}$ is the $N \times N$ channel matrix giving the gains between j th transmitter and i th receiver antennas. The $N \times 1$ vector \mathbf{n}_i is additive i.i.d. Gaussian noise normalized to have unit covariance.

We treat each transmitted signal as a zero-mean complex Gaussian random vector with covariance $E\{\mathbf{x}_i\mathbf{x}_i^H\} = \mathbf{P}_i$. The transmit covariance defines the spatial power allocation of each transmitter. We impose a power constraint on each transmitter by constraining the trace of the covariance matrix. Normalizing the channel matrices by the transmit power, the constraint becomes $\text{tr}\{\mathbf{P}_i\} = 1$.

*Corresponding author.

This work has been partially supported by the U.S. Army Research Office under Multi-University Research Initiative (MURI) grant no. W91-1NF-04-1-0224.

B. Information-theoretic model

In ordinary (single-user) detection, each receiver treats the interfering signals as noise, and the achievable data rates are limited simply by the mutual information between each transmitter/receiver pair, or

$$r_i \leq I(\mathbf{x}_i; \mathbf{y}_i) \quad (2)$$

When we allow multi-user detection, each receiver may consider itself the destination of a multiple-access channel, where multiple transmitters communicate with a single receiver. Instead of simply detecting the desired signal and treating the interference as noise, each receiver may choose to detect interfering signals. The detected interference may then be subtracted out, allowing the intended transmitter to use a higher data rate. The multiple-access channel has been well-studied in both the scalar and vector case [5, 6]. If, in a two-link example, receiver 1 chooses to detect signals \mathbf{x}_1 and \mathbf{x}_2 , then we get the familiar pentagonal rate region:

$$\begin{aligned} r_1 &\leq I(\mathbf{x}_1; \mathbf{y}_1 | \mathbf{x}_2) \\ r_2 &\leq I(\mathbf{x}_2; \mathbf{y}_1 | \mathbf{x}_1) \\ r_1 + r_2 &\leq I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1), \end{aligned}$$

where $I(\mathbf{c}; \mathbf{d} | \mathbf{e})$ is the mutual information between \mathbf{c} and \mathbf{d} given that \mathbf{e} is known at the receiver. Since we always have $I(\mathbf{x}_1; \mathbf{y}_1 | \mathbf{x}_2) \geq I(\mathbf{x}_1; \mathbf{y}_1)$, this feasible region allows r_1 to be at least as great as under single-user detection. However, in order to successfully detect both signals and achieve this higher rate, we impose constraints that may force r_2 lower than was possible under single-user detection. Care must be taken, then, in selecting which signals are detected.

In general, let $D_i \subseteq \{1, 2, \dots, L\}$ denote the signals detected by the i th receiver.¹ That is, if $j \in D_i$, then the i th receiver detects the signal \mathbf{x}_j . Obviously we require that $i \in S_j$, since each receiver needs to detect its intended signal. Also, for any set $T \subseteq \{1, 2, \dots, L\}$, let \mathbf{x}_T denote the joint random vector given by stacking each $\mathbf{x}_i, i \in T$. Then, the achievable rates for the network are given by the *intersection* of the rates defined by the multiple-access regions for each receiver, or the rates that satisfy the inequalities

$$\sum_{k \in T} r_k \leq I(\mathbf{x}_T; \mathbf{y}_i | \mathbf{x}_{T^c}), \quad (3)$$

for all $T \subseteq D_i$, and for all i , and where T^c denotes the complement of the set T . For disjoint sets T and G , the mutual information between \mathbf{x}_T and \mathbf{y}_i given \mathbf{x}_G is

$$I(\mathbf{x}_T; \mathbf{y}_i | \mathbf{x}_G) = \log_2 \frac{\left| \sum_{j \notin G} \mathbf{H}_{i,j} \mathbf{P}_j \mathbf{H}_{i,j}^H + \mathbf{I} \right|}{\left| \sum_{k \notin T \cup G} \mathbf{H}_{i,k} \mathbf{P}_k \mathbf{H}_{i,k}^H + \mathbf{I} \right|}. \quad (4)$$

While potentially allowing improved performance, multi-user detection greatly complicates the rate optimization problem. To maximize an objective function, we must optimize

¹Note that in a MIMO system transmitters may send multiple streams of information, each of which a receiver could selectively detect. However, for simplicity, we allow the receivers only to choose whether or not to detect all of the streams of a particular transmitter's signal.

with respect to three separate sets of variables. The selection of transmit covariances \mathbf{P}_i and detection sets D_i define a region of achievable rates, from which we choose a rate vector according to our optimization criteria. Further, we must optimize jointly; we cannot, for example, choose the receivers' detection sets independent of each other or independent of the transmit covariances. In the next two sections we discuss methods for solving this problem for two distinct objectives.

III. SUM THROUGHPUT

First, we consider the maximization of the sum rate of the network. Let \mathbf{p} be the “vector” of covariance matrices \mathbf{P}_i , \mathbf{D} be the vector of detection sets D_i , and $\mathbf{r} = (r_1, r_2, \dots, r_L)^T$ be the vector of rates. Let $\mathcal{C}(\mathbf{p}, \mathbf{D})$ denote the feasible region defined by the transmit covariances and detection sets in \mathbf{p} and \mathbf{D} . Then, we want to choose \mathbf{p} and \mathbf{D} so that the feasible region $\mathcal{C}(\mathbf{p}, \mathbf{D})$ permits the largest possible sum rate:

$$\begin{aligned} \max_{\mathbf{p}, \mathbf{D}, \mathbf{r}} \quad & \sum_{i=1}^L r_i \\ \text{s.t.} \quad & \mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D}), \text{tr}\{\mathbf{P}_i\} \leq 1, \forall i. \end{aligned}$$

Because of the complicated nature of our constraints, we cannot solve easily for the arguments that globally maximize the sum rate. So, we divide up the optimization process into three “layers.” We start by assuming a fixed \mathbf{p} and \mathbf{D} , and solve for the feasible rates that optimize the sum rate. Using the solution to the inner problem, we search for the \mathbf{D} that maximizes the sum rate given a fixed \mathbf{p} . We then use the solution to the middle problem to search for the input covariances \mathbf{p} that allow the greatest sum rate.

A. Optimal Rates

We begin at the innermost—and simplest—layer of optimization. Given \mathbf{p} and \mathbf{D} , we solve for the feasible rate vector(s) maximize the sum rate. Fortunately, the sum rate is a simple linear function and the constraints in (3) are also linear. Therefore, we can quickly and easily solve the problem numerically using well-established linear programming techniques such as the venerable simplex algorithm [7]. However, in the outer optimization layers we require a closed-form solution.

So, we first use the simplex method to solve for a numerical solution and the active linear constraints. Using the solution and the constraints, we can easily solve for an expression for the maximizing rates as a function of the limiting mutual information terms. For brevity—and since it is quite a simple process—we omit the details.

B. Optimal Detection Sets

The next layer is to choose the best detection sets given a fixed \mathbf{p} . Since each of the L users may choose any combination of the $L - 1$ remaining users' signals to detect, there are $2^{L(L-1)}$ possible choices for \mathbf{D} . It is also difficult to construct a general expression for the maximum sum rate as a function of \mathbf{D} . So, we must evaluate the candidate solutions individually in order to find the optimum choice. When L is small,

we can feasibly perform an exhaustive search, computing the maximum rate for each possible \mathbf{D} . As L increases, however, exhaustive search quickly becomes impractical.

So, we use a simple genetic algorithm [8] to search the space of possible detection sets. A genetic algorithm is attractive because we can easily describe \mathbf{D} as a binary string of $L(L-1)$ bits, and genetic operations such as crossover and mutation are easily interpreted intuitively as changes in receivers' detection sets. After initializing the algorithm with a population of candidate solutions, each generation of the genetic algorithm comprises three main steps—selection, crossover, and mutation—each of which we briefly describe.

In the selection stage, each candidate solution in the population is evaluated in terms of fitness, which here is quantified by maximizing the sum rate as in the previous subsection. The population for the next generation is then selected by drawing randomly from the existing population, where the probability of selection is proportional to fitness.

After selection, the members of the new population are randomly paired for crossover. With a fixed probability p_c , we “cross over” the bits in each pair. After choosing a randomly selected crossover index i_c , the first member keeps its first i_c bits, but exchanges its last $i_c - L(L-1)$ bits with its pair. Conversely, the second member takes on the first i_c bits from its pair while retaining its last $i_c - L(L-1)$ bits.

Finally, we “mutate” the solutions generated by the selection and crossover stages. For each bit in each candidate solution, we invert the bit with small fixed probability p_M . This allows small, undirected changes to enter the population which, if beneficial, tend to persist in future generations.

In our trials, we use a population size of 10, which is initialized randomly. We use a crossover probability of $p_C = 0.7$ and a mutation probability of $p_M = 0.001$. We run the algorithm for a maximum of 50 generations, and keep track of the best solution generated throughout the process. While the genetic algorithm obviously cannot guarantee an optimal choice of \mathbf{D} , it is computationally feasible and provides a good solution.

C. Optimal Transmit Covariances

Finally, we use the well-known gradient projection method in order to choose the transmit covariances that make up \mathbf{p} . To do this, we must express our objective function as a differentiable function of \mathbf{p} , which we accomplish by defining the function in terms of the inner two optimization layers:

$$J_s(\mathbf{p}) = \max_{\mathbf{r}, \mathbf{D}} \sum_{k=1}^L r_k = \max_{\mathbf{D}} \max_{\mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D})} \sum_{k=1}^L r_k, \quad (5)$$

where the maximization over \mathbf{D} is carried out approximately by the genetic algorithm described in the previous subsection. Function evaluations of $J_s(\mathbf{p})$ are therefore computationally expensive since we must run the genetic algorithm to choose a good \mathbf{D} . Since, for any \mathbf{D} , the inner layer gives an expression for the maximizing \mathbf{r} as a function of the limiting mutual information terms, and since each mutual information term is differentiable with respect to \mathbf{p} , we can use gradient projection to search for transmit covariances that locally maximize $J_s(\mathbf{p})$.

Gradient projection is used to optimize a scalar function $f(\mathbf{x})$, where \mathbf{x} is constrained to be an element of some convex set. Since the matrices that form \mathbf{p} are positive semi-definite and subject to a trace constraint, the convexity of the feasible set of covariances is immediate. Properly implemented, the procedure is guaranteed to converge to a local optimum. The inner optimization layers give $J_s(\mathbf{p})$ as a linear combination of M mutual information terms:

$$J_s(\mathbf{p}) = \sum_{k=1}^M c_k I(\mathbf{x}_{T_k}; \mathbf{y}_k | \mathbf{x}_{G_k}). \quad (6)$$

So, the gradient with respect to a particular transmit covariance \mathbf{P}_i is

$$\nabla_{\mathbf{P}_i(t)} J_s(\mathbf{p}) = \sum_{k=1}^M c_k \nabla_{\mathbf{P}_i(t)} I(\mathbf{x}_{T_k}; \mathbf{y}_k | \mathbf{x}_{G_k}). \quad (7)$$

For a complex number $z = x + jy$, we define the gradient as $\nabla_z f(z) = \partial f(z)/\partial x + j\partial f(z)/\partial y$. Using this definition, it can be shown that the gradient of the mutual information $I(\mathbf{x}_{T_k}; \mathbf{y}_k | \mathbf{x}_{G_k})$ with respect to \mathbf{P}_i is given as follows. Defining

$$\mathbf{S} = \sum_{j \in T_k} \mathbf{H}_{k,j} \mathbf{P}_j \mathbf{H}_{k,j}^H \quad (8)$$

and

$$\mathbf{R} = \sum_{j \notin T_k \cup G_k} \mathbf{H}_{k,j} \mathbf{P}_j \mathbf{H}_{k,j}^H + \mathbf{I}, \quad (9)$$

we can express the gradient as

$$\nabla_{\mathbf{P}_i} I = \begin{cases} \frac{2}{\ln 2} \mathbf{H}_{k,i}^H (\mathbf{S} + \mathbf{R})^{-1} \mathbf{H}_{k,i}, & \text{for } i \in T_k \\ \mathbf{0}, & \text{for } i \in G_k \\ \frac{2}{\ln 2} \mathbf{H}_{k,i}^H ((\mathbf{S} + \mathbf{R})^{-1} - (\mathbf{R})^{-1}) \mathbf{H}_{k,i}, & \text{otherwise} \end{cases}. \quad (10)$$

At iteration t , we start with the covariances contained in \mathbf{p}^t and step in the direction of the gradient for each transmit covariance, forming a new vector $\tilde{\mathbf{p}}^t$:

$$\tilde{\mathbf{P}}_i^t = \mathbf{P}_i^t + s \nabla_{\mathbf{P}_i^t} J_s(\mathbf{p}^t), \quad (11)$$

where s is a fixed step size. Of course, simply following the gradient may lead to covariances that do not obey the trace constraint. Therefore, we project each matrix in $\tilde{\mathbf{p}}^t$ onto the set of feasible matrices. Using the usual matrix inner product $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}\{\mathbf{A}^H \mathbf{B}\}$, the induced norm is $\|\mathbf{A}\|_F = \sqrt{\text{tr}\{\mathbf{A}^H \mathbf{A}\}}$, the Frobenius norm. We project by choosing the feasible vector of matrices $\hat{\mathbf{p}}^t$ that minimizes the sum of the squared Frobenius norm of the matrices in $\hat{\mathbf{p}}^t - \tilde{\mathbf{p}}^t$.

We omit the details, but it is straightforward to show that the projected matrices have the form

$$\hat{\mathbf{P}}_i^t = \mathbf{X}_i^t [\mathbf{D}_i^t - \nu_i^t \mathbf{I}]^+ (\mathbf{X}_i^t)^H, \quad (12)$$

$\mathbf{X}_i^t \mathbf{D}_i^t (\mathbf{X}_i^t)^H = \tilde{\mathbf{P}}_i^t$ is the eigen-decomposition of $\tilde{\mathbf{P}}_i^t$, $[\mathbf{A}]^+$ zeros out any negative entries in the matrix \mathbf{A} , and ν_i^t is chosen to satisfy the the trace constraint.

We complete iteration t by stepping in the feasible direction defined by $\hat{\mathbf{p}}^k$:

$$\mathbf{p}^{t+1} = \mathbf{p}^k + a_t(\hat{\mathbf{p}}^t - \mathbf{p}^t), \quad (13)$$

where $a_t \in [0, 1]$ is a variable step size. Since (13) defines a convex combination of two elements in the feasible set, \mathbf{p}^{t+1} is also feasible. We choose a_t using Armijo's rule along the feasible direction $\hat{\mathbf{p}}^t - \mathbf{p}^t$, which specifies that $a_t = \gamma^{m_t}$, where $\gamma \in [0, 1]$ and m_t is the smallest nonnegative integer such that

$$J_s(\mathbf{p}^{t+1}) - J_s(\mathbf{p}^t) \geq \sigma \gamma^{m_t} \langle \nabla J_s(\mathbf{p}^t), \hat{\mathbf{p}}^t - \mathbf{p}^t \rangle \quad (14)$$

$$= \sigma \gamma^{m_t} \sum_{i=1}^L \text{tr}((\nabla_{\mathbf{P}_i} J_s(\mathbf{p}^t))^H (\hat{\mathbf{P}}_i - \mathbf{P}_i)) \quad (15)$$

for a small constant σ . In order to find the constant a_t , we must repeatedly evaluate J_s , which is expensive computationally. Therefore, we greatly reduce the computational complexity by holding \mathbf{D} fixed—and using only the innermost layer of optimization—as we find a_t . After solving for the new matrices, we evaluate $J_s(\mathbf{p}^{t+1})$ using both inner optimization layers in preparation for computing the gradient in the next iteration.

After each iteration, we check to see whether or not the convergence criterion is met, which is

$$\max |\mathbf{p}^{t+1} - \mathbf{p}^t| < \epsilon \quad (16)$$

for a small constant ϵ . If (16) is met, iterations stop and \mathbf{p}^{t+1} , along with the associated \mathbf{D} and \mathbf{r} , are chosen.

While maximizing the sum rate is ideal in terms of network throughput, it may give poor results in terms of individual rates. In maximizing the sum rate, it is often advantageous to force a weak user to a very low rate—or to shut off entirely—so that stronger users may increase their rates. To address this issue, we consider a game-theoretic bargaining solution which aims to define a fair, but still efficient, solution.

IV. KALAI-SMORODINSKY SOLUTION

A. Theory

The Kalai-Smorodinsky (K-S) solution selects a unique operating point according to a set of axioms chosen to ensure fairness and efficiency. Generally, a bargaining problem is defined by a set K composed of L players, a *disagreement point* δ , and a set of feasible payoffs $S \subset \mathbb{R}^L$. The disagreement point represents the “status quo” prior to bargaining, or the payoff that each player receives if bargaining should fail. In our problem, $K = \{1, 2, \dots, L\}$, where each player represents a transmitter/receiver pair. The set of feasible payoffs is simply the set of rate vectors achievable, or $S = \{\mathbf{r} : \mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D}), \text{tr}\{\mathbf{P}_i\} \leq 1, \forall i\}$.

In an arbitrary bargaining problem, define the *negotiation set* N as the set of feasible payoffs for which $r_i \geq \delta_i, \forall i \in K$. The Kalai-Smorodinsky solution to the bargaining problem (K, δ, S) selects a unique point $\mu(K, \delta, S) = \mathbf{r}^*$ characterized by the following axioms:

- 1) **Pareto efficiency:** If $\mathbf{r} \in N$ is a vector such that $r_i \geq r_i^*$, then $\mathbf{r} = \mathbf{r}^*$. That is, if there is an i for which $r_i > r_i^*$, there must be at least one j for which $r_j < r_j^*$. Pareto efficiency ensures that we do not overlook any payoff vectors which improve a player's payoff without cost to other players.
- 2) **Invariance to positive affine transformations:** If we scale and shift the set of feasible payoffs by positive affine transformations, the solution \mathbf{r}^* is shifted by the same transformations.
- 3) **Symmetry:** Let T be a permutation of the players in K . Then, $\mu(T(K), T(\delta), T(S)) = T(\mathbf{r}^*)$.
- 4) **Monotonicity:** Define $b_i(N) = \sup\{r_i : \mathbf{r} \in N\}$, the best-case payoff for player i . Next, let $\mathbf{r}_{-i} = (r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_L)^T$ denote an $L-1$ -dimensional vector of payoffs. Then, define $g_i(\mathbf{r}_{-i}, N) = \sup\{r_i : (r_i, \mathbf{r}_{-i}) \in N\}$, the best-case payoff for player i assuming that the other $L-1$ players receive \mathbf{r}_{-i} . If (K, δ, S_1) and (K, δ, S_2) are bargaining problems such that $b_i(N_1) = b_i(N_2)$ and $g_i(\mathbf{r}_{-i}, N_1) \leq g_i(\mathbf{r}_{-i}, N_2)$ for all feasible \mathbf{r}_{-i} , then $\mu_i(K, \delta, S_1) \leq \mu_i(K, \delta, S_2)$.

The final axiom specifies a criterion of fairness in defining how the solution varies with changes in S . If, for every fixed \mathbf{r}_{-i} , player i can obtain a higher feasible payoff, then the payoff assigned to player i by μ is increased. Assuming that S is compact and convex, these four axioms characterize a unique solution with a simple geometric representation. Define $L(\delta, \mathbf{b}(N))$ as the line which passes through the disagreement point and the vector of best-case payoffs for each player. Then, the K-S solution is defined as the greatest point in N that lies on $L(\delta, \mathbf{b}(N))$:

$$\mu(K, \delta, S) = \max_{\mathbf{r} \in N \cap L(\delta, \mathbf{b}(N))} \|\mathbf{r}\|, \quad (17)$$

where we use the standard Euclidean norm. The axioms, as well as the form of the objective function defined by the solution, give reason to anticipate that the K-S solution will provide an effective solution. Pareto efficiency guarantees an efficient solution, while symmetry and monotonicity guarantee at least a notion of fairness.

While there are several reasonable choices for the disagreement point δ , we choose $\delta = \mathbf{0}$ for simplicity. Also, we note that, for our problem, the feasible set $S = \{\mathbf{r} : \mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D}), \text{tr}\{\mathbf{P}_i\} \leq 1, \forall i\}$ is compact, but not generally convex. Fortunately, it is shown in [9] that, by slightly weakening the first and last axioms, the unique solution defined in (17) generalizes to non-convex sets.

B. Implementation

We employ similar optimization techniques to search for the Kalai-Smorodinsky solution as we did in maximizing the rates in Section III. First, however, we must solve for the best-case rates that define the vector \mathbf{b} . With $\delta = \mathbf{0}$, this is straightforward and does not involve the receivers' detection sets. To find b_i , we assume single-user detection and set $\mathbf{P}_j = \mathbf{0}, i \neq j$. Then, the i th transmitter chooses its covariance

optimally using the well-established water-filling technique [10]. Then we set $b_i = I(\mathbf{x}_i; \mathbf{y}_i)$, the maximum rate possible when all other users do not transmit. Now, we search for the rates that maximize the objective function defined by the K-S solution:

$$\begin{aligned} & \max_{\mathbf{p}, \mathbf{D}, \mathbf{r}} \quad \|\mathbf{r}\| \\ & \text{s.t.} \quad \mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D}) \cap L(\mathbf{0}, \mathbf{b}), \text{tr}\{\mathbf{P}_i\} \leq 1, \forall i. \end{aligned}$$

As with the sum rate, we accomplish this maximization through three optimization layers.

1) *Optimal Rates:* For a fixed \mathbf{p} and \mathbf{D} , we wish to find the rate vector $\mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D}) \cap L(\mathbf{0}, \mathbf{b})$ that maximizes $\|\mathbf{r}\|$. Fortunately, this is a straightforward problem with a simple solution. The region $\mathcal{C}(\mathbf{p}, \mathbf{D})$ is defined by a set of constraints of the form $\sum_{i \in T} r_i \leq I(\mathbf{x}_T; \mathbf{y}_k | \mathbf{x}_G)$. To find the optimal rate vector, we simply solve for the intersection of $L(\mathbf{0}, \mathbf{b})$ with the hyperplane defined by each linear constraint, and keep the intersecting point with the smallest norm. Since it has the smallest norm, is a member of $\mathcal{C}(\mathbf{p}, \mathbf{D})$, but it still lies on the boundary.

Neglecting the measure-zero case where $L(\mathbf{0}, \mathbf{b})$ simultaneously intersects multiple hyperplanes on the boundary of $\mathcal{C}(\mathbf{p}, \mathbf{D})$, the norm of the optimal rate vector is of the form

$$\|\mathbf{r}\| = cI(\mathbf{x}_T; \mathbf{y}_k | \mathbf{x}_G), \quad (18)$$

for some T, k, G , and constant c . That is, the norm is a single mutual information term multiplied by a constant, which is easily computed from the numerical solution for \mathbf{r} .

2) *Optimal Detection Sets:* Next, we choose detection sets \mathbf{D} for a fixed \mathbf{p} . The discussion in Section III-B also applies to the K-S solution. There are still $2^{L(L-1)}$ possibilities for \mathbf{D} and no clear way to optimize analytically. So, as before, we employ a simple genetic algorithm to search for a good \mathbf{D} . At each generation, we select a new population according to the old population members' fitness (which here is quantified by the norm of the corresponding rate vector), crossover the new population's bit strings, and mutate the resulting solutions. We again use a population size of 10, a crossover probability of $p_C = 0.7$, and a mutation probability of $p_M = 0.001$, and run the algorithm for a maximum of 50 generations.

3) *Optimal Transmit Covariances:* Finally, we search for the optimal transmit covariances described by \mathbf{p} . As before, we express the objective function as a differentiable function of \mathbf{p} :

$$J_{KS}(\mathbf{p}) = \max_{\mathbf{D}} \max_{\mathbf{r} \in \mathcal{C}(\mathbf{p}, \mathbf{D}) \cap L(\mathbf{0}, \mathbf{b})} \|\mathbf{r}\|. \quad (19)$$

Using the inner two layers, the objective function has the form

$$J_{KS}(\mathbf{p}) = cI(\mathbf{x}_T; \mathbf{y}_k | \mathbf{x}_G), \quad (20)$$

where c, T, k , and G may change with \mathbf{p} .

As before, we use the gradient projection method to find a local optimum for $J_{KS}(\mathbf{p})$. The process is identical to that in Section III-C, except that instead of being a linear combination of mutual information terms, the objective function is a (multiple of a) single mutual information term. Essentially, the

gradient projection algorithm works to maximize the limiting mutual information term until a different mutual information becomes the limiting term. Of course, it is easy to maximize a single mutual information since it is either convex or concave in each covariance matrix. Rather than converging when a single mutual information term is maximized, the algorithm converges when small changes in \mathbf{p} cause alternations between several limiting mutual information terms which cannot be further increased simultaneously.

V. SIMULATIONS

To examine the effectiveness of our methods, we perform simulations on randomly-generated networks. We assume Rayleigh-fading channels with channel matrices

$$\mathbf{H}_{i,j} = \sqrt{\frac{\rho_{i,j}}{N}} \mathbf{H}'_{i,j}, \quad (21)$$

where the entries of $\mathbf{H}'_{i,j}$ are independently drawn from the standard complex normal distribution, and $\rho_{i,j}$ is the expected signal-to-noise ratio between the j th transmitter and the i th receiver when $\text{tr}\{\mathbf{P}_j\} = 1$. We determine $\rho_{i,j}$ by placing transmitters and receivers on a plane, giving us

$$\rho_{i,j} = \frac{M}{d(i,j)^\alpha}, \quad (22)$$

where $d(i,j)$ is the Euclidean distance between the j th transmitter and the i th receiver, M is an arbitrary constant, and α is the path loss exponent.

For our simulations, we place transmitters and receivers randomly and uniformly on the unit square. We set $\alpha = 4$ and choose $M = 5/8$, which forces $\rho_{i,j} = 10\text{dB}$ when $d(i,j) = 1/2$. In Figure 1 we plot the average per-user rate for a variety of values of L and N for both the sum rate and K-S solutions. When $2 \leq L \leq 4$, we choose $N = 2$, and $N = 3$ for $L = 5$. As a baseline, we also compute results for single-user detection. Each data point represents the average of 100 independent trials. Naturally, maximizing the sum rate gives the best average rate, while the K-S solution performs somewhat worse. Indeed, in terms of sum rate, it is better to maximize using single-user detection than to use the K-S solution under multi-user detection. Next, in Figure 2 we plot the average of the *worst* user's rate as a partial measure of the fairness of each solution. In terms of fairness, the K-S solution clearly performs better. Although the average rate is lower, users enjoy improved worst-case performance over sum rate maximization.

For a more complete picture, we plot the empirical cumulative distribution function (CDF) for the 100 trials where $L = 5$ and $N = 3$ in Figure 3. The CDF emphasizes the fact that, while the best users perform better under sum rate maximization, the worst users perform quite poorly. On the other hand, under the K-S solution, the effects of interference are distributed fairly while still maintaining an efficient solution.

Finally, we point out the significant benefits gained by employing multi-user detection, particularly in the K-S solution. In terms of average rate, multi-user detection modestly

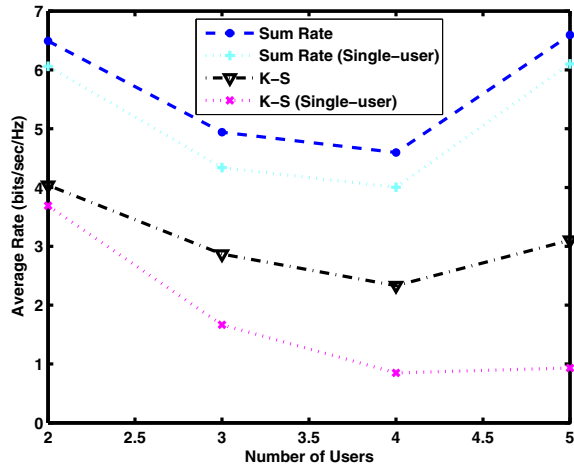


Fig. 1. Average rate per user for sum rate and K-S solutions.

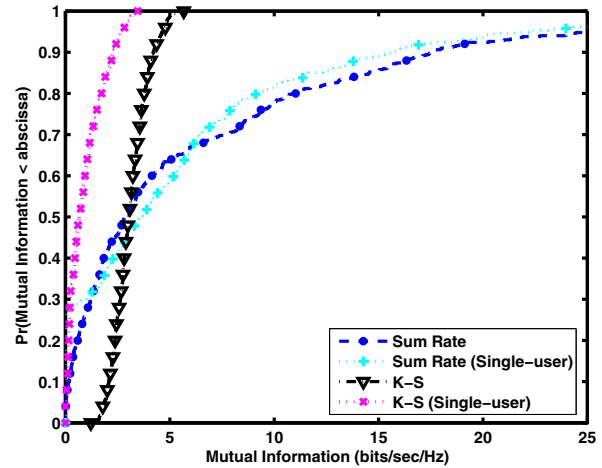


Fig. 3. Empirical CDF of users' rates. $L = 5$ and $N = 3$.

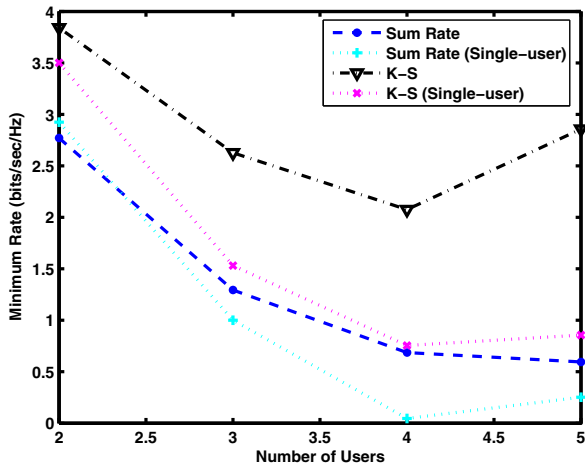


Fig. 2. Worst user's rate for sum rate and K-S solutions.

improves the sum rate solution while greatly improving the K-S solution. In terms of worst-case performance, the K-S solution again is greatly improved by the use of multi-user detection. Note, however, that the fairness of the sum rate solution increases as well. In addition to increasing the achievable sum rate, multi-user detection also allows the worst users to perform at least slightly better on average.

VI. CONCLUSION

We have proposed methods for optimizing a MIMO interference network according to two objective functions: network sum rate and the Kalai-Smorodinsky solution. Using linear programming, genetic algorithms, and gradient projection, we search for sub-optimal solutions to these challenging maximization problems.

Our empirical results suggest that multi-user detection im-

proves performance in randomly generated networks. The average rate is improved in sum-rate maximization, and in the Kalai-Smorodinsky approach—where we attempt to find a fair solution—multi-user detection drastically improves network performance. While the K-S approach does lower the overall sum rate, it ensures that individual rates are kept high.

Finally, further work is needed to improve the optimization processes described in Sections III and IV. The genetic algorithm used to find the best detection sets is fairly simple, and we probably could obtain better performance with a more sophisticated algorithm. Also, since the objective functions J_s and J_{KS} are piecewise continuous in the transmit covariances, the gradient projection method is particularly prone to getting “stuck” at local optima. Methods to overcome these shortcomings may improve performance substantially.

REFERENCES

- [1] S. Ye and R. S. Blum, “Optimized signaling for MIMO interference systems with feedback,” *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2839–2848, November 2003.
- [2] Y. Rong and Y. Hua, “Optimal power schedule for distributed MIMO links,” in *Proc. 25th Army Science Conference*, November 2006.
- [3] S. Verdú, *Multuser Detection*. Cambridge University Press, 1998.
- [4] E. Kalai and M. Smorodinsky, “Other solutions to Nash’s bargaining problem,” *Econometrica*, vol. 43, no. 3, pp. 513–518, May 1975.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley and Sons, 2006.
- [6] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, “Iterative water-filling for Gaussian vector multiple-access channels,” *IEEE Trans. Info. Theory*, vol. 50, no. 1, pp. 145–152, January 2004.
- [7] D. P. Bertsekas, *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific, 1997.
- [8] D. Whitley, “A genetic algorithm tutorial,” *Statistics and Computing*, vol. 4, no. 2, pp. 65–85, 1994.
- [9] J. L. Hougaard and M. Tvede, “The Kalai-Smorodinsky solution: A generalization to nonconvex n-person bargaining,” *Discussion Papers, Univ. Copenhagen Dept. Economics*, no. 98-21, 1998.
- [10] E. Telatar, “Capacity of multi-antenna Gaussian channels,” AT&T Bell Labs, Tech. Rep., 1995.