MULTICAST BEAMFORMING WITH ANTENNA SELECTION

Omar Mehanna, Nicholas D. Sidiropoulos*† *and Georgios B. Giannakis*‡

Dept. of Electrical and Computer Engineering
University of Minnesota

ABSTRACT

Multicast beamforming exploits subscriber channel state information at the base station to form multiple beams that steer radiated power towards users of interest, while minimizing leakage to other users and systems. Such functionality has been provisioned in the long-term evolution (LTE) enhanced multimedia broadcast multicast service (EMBMS). In this context, the present paper deals with joint multicast beamforming and antenna selection. Whereas this problem (and even plain multicast beamforming) is NP-hard, it is shown that using ℓ_1-norm *squared* (instead of ℓ_1-norm) as a surrogate for the ℓ_0-norm yields a natural semidefinite programming relaxation - something not obvious with the ℓ_1-norm. Simulations indicate that the proposed algorithm significantly reduces the number of antennas required to meet prescribed service levels, at a relatively small cost of excess transmission power. Furthermore, its performance is close to that attained by exhaustive search, at far lower complexity.

1. INTRODUCTION

Consider a base station (BS) transmitter using an antenna array to broadcast common information to multiple radio subscribers. Instead of broadcasting isotropically, the BS can exploit subscriber channel state information to select different weights for each antenna in order to steer power in the directions of the subscribers while controlling interference to other users. In line with mandates for green communications, the goal of multicast beamforming is to minimize the total transmission power at the BS while guaranteeing a certain minimum signal-to-noise ratio (SNR) per subscriber.

Multicast beamforming is part of the enhanced multimedia broadcast multicast service (EMBMS) of the long term evolution (LTE) standard. After market-related delays, initial roll-out of EMBMS is expected to begin in 2012. EMBMS can markedly boost spectral efficiency and reduce energy and infrastructure costs per bit when the same content must be delivered wirelessly to multiple end users.

In practice, a BS may have more antennas than (expensive) radio transmission chains, and it is desired to automatically switch the available chains to the most appropriate subset of antennas in an adaptive fashion. Each radio transmission chain requires a digital-to-analog (D/A) converter, a mixer, and a low-noise power amplifier. Antenna elements, on the other hand, are becoming smaller and cheaper; thus, antenna selection strategies are becoming increasingly desirable.

Antenna subset selection has been considered for point-to-point MIMO links [3, 10], and for receive beamforming [5, 2]. Beampattern synthesis with antenna selection was pursued in [5], using a convex optimization formulation that controls mainlobe and sidelobes while minimizing the sparsity-inducing ℓ_1-norm to produce a sparse beamforming weight vector involving fewer antennas. The setup in [5] only applies to uniform linear array (ULA) far-field scenarios, whereas the present paper's approach works for arbitrary channel (or *steering*) vectors. Another important difference is that [5] restricts the beamforming weights to be conjugate symmetric in order to turn the non-convex lower bound constraints on the beampattern into affine ones. This gives up half of the problem's degrees of freedom, thereby yielding suboptimal solutions when only the magnitude of the beampattern is important, as in transmit beamforming. This restriction is not present in this paper's formulation. In a similar vain, [2] considered using the ℓ_1-norm to obtain sparse solutions to convex beampattern synthesis problems. Whereas [2] does not restrict the weight vector to be conjugate symmetric, it does unnecessarily constrain the phase of the beampattern; without such a constraint on the phase, the problem is non-convex, and thus challenging.

In this paper, joint multicast beamforming and antenna selection is considered. Whereas this problem (and even plain multicast beamforming) is NP-hard, it is shown that using ℓ_1-norm *squared* (instead of ℓ_1-norm) as a surrogate for the ℓ_0-norm yields a natural semidefinite programming relaxation - something non-obvious with the ℓ_1-norm. In order to further enhance sparsity, an iterative ℓ_1-norm reweighting scheme similar to the one used in Lasso [9] is employed, along with related methods for compressed sensing [1]. Simulations indicate that the proposed algorithm significantly reduces the number of antennas required to meet prescribed service levels, at a small cost in excess transmission power. Furthermore, its performance is close to that attained by exhaustive search, at far lower complexity. A related antenna selection problem for multicast rate maximization (not multicast beamforming) using exhaustive search was considered in [6], where the emphasis was on asymptotic performance analysis.

2. PROBLEM FORMULATION

2.1. Basic Model

The system model is similar to [7], comprising a single BS transmitter with N antennas broadcasting a common message to M receivers, each with a single antenna. The complex vector that models the propagation loss and the frequency-flat quasi-static channel from each transmit antenna to the receive antenna of user m is denoted by $\bar{\mathbf{h}}_m$, where $m \in \{1, \ldots, M\}$. The optimization variable is the $N \times 1$ beamforming weight vector $\mathbf{w}$ applied to the N transmit-antenna elements. Assuming that the transmitted information signal is temporally white with zero-mean and unit variance, and that the noise at receiver m is also white with zero-mean and variance σ_m^2, the received SNR at the m-th user is $|\mathbf{w}^{\mathcal{H}}\bar{\mathbf{h}}_m|^2/\sigma_m^2$. If the minimum required SNR at the m-th user is γ_m, then the SNR con-

*E-mail: meha0006@umn.edu. Supported by NSF CCF grant 0747332.

†(Contact author) E-mail: nikos@umn.edu, phone: (612) 625-1242, Fax: (612) 625-4583.

‡E-mail: georgios@umn.edu.

978-1-4673-0971-4/12/$31.00 © 2012 IEEE

straint $|\mathbf{w}^{\mathcal{H}}\bar{\mathbf{h}}_m|^2/\sigma_m^2 \geq \gamma_m$ can be expressed as $|\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2 \geq 1$, where $\mathbf{h}_m := \bar{\mathbf{h}}_m/\sqrt{\sigma_m^2\gamma_m}$. It is further assumed that the $\mathbf{h}_m$'s for all users are known at the BS transmitter. Therefore, designing the beamformer that minimizes transmit-power, subject to receive-SNR constraints per user, and without performing any antenna selection, can be written as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^N} \quad & ||\mathbf{w}||_2^2 \\ \text{s.t.}: \quad & |\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2 \geq 1, \quad m = 1, \ldots, M. \end{aligned} \tag{1}$$

It is clear that the quadratic constraints are *non-convex*, which implies that (1) is a non-convex optimization problem. In fact, this problem is known to be NP-hard for general channel vectors [7].

2.2. Antenna Selection

Suppose now that only $K \leq N$ RF transmission chains are available, and thus only K antennas can be simultaneously utilized for transmission. The goal is to jointly select the *best* K out of N antennas, and find the corresponding beamforming vector $\mathbf{w}$ so that transmit-power is minimized, subject to the receive-SNR constraints per subscriber. Both objectives must be jointly considered, because the two constituent problems are tightly coupled. The resultant joint problem can be expressed as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^N} \quad & ||\mathbf{w}||_2^2 \\ \text{s.t.}: \quad & |\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2 \geq 1, \quad m = 1, \ldots, M \\ & ||\mathbf{w}||_0 \leq K \end{aligned} \tag{2}$$

where the ℓ_0-(quasi)norm is the number of nonzero entries of $\mathbf{w}$; i.e., $||\mathbf{w}||_0 := |\{n : \mathbf{w}_n \neq 0\}|$. Instead of the hard sparsity constraint, an ℓ_0 penalty can be employed to promote sparsity, leading to

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^N} \quad & ||\mathbf{w}||_2^2 + \lambda||\mathbf{w}||_0 \\ \text{s.t.}: \quad & |\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2 \geq 1, \quad m = 1, \ldots, M \end{aligned} \tag{3}$$

where λ is a positive real tuning parameter that controls the sparsity of the solution, and thus the number of antennas selected. Problem (3) strikes a balance between minimizing the transmission power and minimizing the number of selected antennas, where a larger λ implies a sparser solution.

Unfortunately, $||\mathbf{w}||_0$ is non-convex. To find the globally optimum solution, an exhaustive combinatorial search over all $\binom{N}{K}$ possible sparsity patterns of $\mathbf{w}$ is required, where the NP-hard problem (1) must be solved (or closely approximated using the algorithm in [7]) for each of these patterns. This motivates the pursuit of computationally efficient near-optimal solutions to problems (2) and (3). Whereas the ℓ_1-norm is typically used as a convex surrogate of the ℓ_0-norm in the literature, the ℓ_1-norm *squared* can be used just as well for this purpose, as detailed in the ensuing section. In fact, it will turn out that finding a convex approximation to (3) is much easier when using the latter.

3. RELAXATION

3.1. Convex Sparsity-Inducing Norms

It is well known that the ℓ_1-norm (defined as $||\mathbf{w}||_1 := \sum_{n=1}^N |w_n|$) is the closest convex approximation to the ℓ_0-norm, albeit a weaker and indirect measure of sparsity [1]. In this subsection, it is argued that the ℓ_1-norm squared, namely $||\mathbf{w}||_1^2$, also induces sparsity. Replacing the ℓ_0-norm in (3) with the ℓ_1-norm, the problem becomes

$$\min_{\mathbf{w}\in\mathcal{F}} ||\mathbf{w}||_2^2 + \lambda||\mathbf{w}||_1 \tag{4}$$

where $\mathcal{F} := \{\mathbf{w} \in \mathbb{C}^N : |\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2 \geq 1, \forall m \in \{1, \ldots, M\}\}$. Problem (4) is equivalent to

$$\min_{\mathbf{w}\in\mathcal{F}} ||\mathbf{w}||_2^2 \qquad \text{s.t.} \quad ||\mathbf{w}||_1 \leq \tau \tag{5}$$

since for any λ, one can find a τ such that the both problems yield the same optimum sparse solution. By squaring both sides of the constraint, problem (5) can be written as

$$\min_{\mathbf{w}\in\mathcal{F}} ||\mathbf{w}||_2^2 \qquad \text{s.t.} \quad ||\mathbf{w}||_1^2 \leq \bar{\tau} \tag{6}$$

where $\bar{\tau} := \tau^2$. If the Pareto boundary is convex, then there exists a $\bar{\lambda}$ such that problem

$$\min_{\mathbf{w}\in\mathcal{F}} ||\mathbf{w}||_2^2 + \bar{\lambda}||\mathbf{w}||_1^2 \tag{7}$$

is equivalent to (6), i.e., (7) is just a re-parametrization of (4). This is always true for convex problems, e.g., the Lasso[1] [9], suggesting that the ℓ_1-norm squared can be used as a sparsity-inducing regularization. In our case $\mathcal{F}$ is non-convex, hence convexity of the Pareto boundary is not guaranteed. Still, the above discussion motivates using the ℓ_1-norm squared as a sparsity-inducing penalty in place of the ℓ_0 penalty in (3). The reason why the ℓ_1-norm squared is preferred over the ℓ_1-norm will become clear in the next subsection.

One limitation of using the ℓ_1-norm (or the ℓ_1-norm squared) as a convex-approximation to the ℓ_0-norm in problem (2) or (3) is that the resulting solution is no longer necessarily the minimum power solution. This is because the ℓ_0-norm counts the number of nonzero entries of $\mathbf{w}$ without regard to their size, whereas the ℓ_1-norm is size-sensitive. The solution of the relaxed problem thus strikes a balance between minimizing the ℓ_1- and ℓ_2-norms. This implies that after obtaining an approximate sparse solution to (2) or (3), one should consider a reduced-size ℓ_2-norm minimization problem of type (1) as a last step, omitting the antennas corresponding to the zero entries of the sparse solution.

3.2. Semidefinite Program Formulation

After replacing the ℓ_0-norm in (3) by the ℓ_1-norm squared, the resulting problem is still NP-hard (it contains [7]), but it will be possible to relax it to a semidefinite program (SDP) as follows: Define $\mathbf{Q}_m := \mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}$ and $\mathbf{X} := \mathbf{w}\mathbf{w}^{\mathcal{H}}$, and use the transformations

$$||\mathbf{w}||_2^2 = \text{tr}(\mathbf{w}\mathbf{w}^{\mathcal{H}}) = \text{tr}(\mathbf{X})$$

[1] It is also easy to check that the soft thresholding (shrinkage) property of the Lasso holds when the ℓ_1-norm squared is used instead of the ℓ_1-norm to induce sparsity, albeit with a different scaling for the threshold.

$$
\begin{aligned}
||\mathbf{w}||_1^2 &= \left(\sum_{n=1}^{N} |w_n|\right)^2 = \mathbf{1}_{N\times 1}^T |\mathbf{X}| \mathbf{1}_{N\times 1} \\
&= \mathrm{tr}(\mathbf{1}_{N\times N} |\mathbf{X}|) \\
|\mathbf{w}^{\mathcal{H}} \mathbf{h}_m|^2 &= \mathbf{h}_m^{\mathcal{H}} \mathbf{w}\mathbf{w}^{\mathcal{H}} \mathbf{h}_m = \mathrm{tr}(\mathbf{h}_m^{\mathcal{H}} \mathbf{w}\mathbf{w}^{\mathcal{H}} \mathbf{h}_m) \\
&= \mathrm{tr}(\mathbf{h}_m \mathbf{h}_m^{\mathcal{H}} \mathbf{w}\mathbf{w}^{\mathcal{H}}) = \mathrm{tr}(\mathbf{Q}_m \mathbf{X})
\end{aligned}
$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator, $|\mathbf{X}|$ is the element-wise absolute value of matrix $\mathbf{X}$, $\mathbf{1}_{N\times 1}$ is an $N \times 1$ all one vector, and $\mathbf{1}_{N\times N}$ is an $N \times N$ matrix with all one entries. Note that $\mathbf{X} = \mathbf{w}\mathbf{w}^{\mathcal{H}}$ if and only if $\mathbf{X} \succeq 0$ and $\mathrm{rank}(\mathbf{X}) = 1$. By dropping the non-convex rank constraint $\mathrm{rank}(\mathbf{X}) = 1$, problem (1) can be relaxed to the SDP

$$
\begin{aligned}
\min_{\mathbf{X}\in\mathbb{C}^{N\times N}} \quad & \mathrm{tr}(\mathbf{X}) \\
\text{s.t.}: \quad & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m) \geq 1,\ m = 1, \ldots, M, \quad \mathbf{X} \succeq 0.
\end{aligned} \tag{8}
$$

Similarly, problem (3) can be relaxed to the SDP

$$
\begin{aligned}
\min_{\mathbf{X}\in\mathbb{C}^{N\times N}} \quad & \mathrm{tr}(\mathbf{X} + \lambda \mathbf{1}_{N\times N} |\mathbf{X}|) \\
\text{s.t.}: \quad & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m) \geq 1,\ m = 1, \ldots, M, \quad \mathbf{X} \succeq 0.
\end{aligned} \tag{9}
$$

Both SDP problems can be efficiently solved (in polynomial time) using interior point methods, and a MATLAB implementation of modern interior point solvers is SeDuMi [8]. Matrix $|\mathbf{X}|$ can be readily brought to the standard form for SDP by introducing an auxiliary matrix $\mathbf{Y} \in \mathbb{R}^{N\times N}$, replacing $|\mathbf{X}|$ by $\mathbf{Y}$ in the given SDP problems, and upper bounding the absolute value of complex entries of $\mathbf{X}$ by the corresponding real entries of $\mathbf{Y}$ using the LMI constraints: $\begin{bmatrix} Y(i,j) - \Re\{X(i,j)\} & \Im\{X(i,j)\} \\ \Im\{X(i,j)\} & Y(i,j) + \Re\{X(i,j)\} \end{bmatrix} \succeq 0$, for each $i, j = 1, \ldots, N$.

Let $\mathbf{X}^{(s)}$ denote the sparse solution of (9). Its zero diagonal entries correspond to the antennas that should be left out, whereas the nonzero ones correspond to the selected antennas. Suppose that the number of nonzero diagonal entries of $\mathbf{X}^{(s)}$ is $\hat{N} \leq N$, and let $S \subseteq \{1, \ldots, N\}$ denote the corresponding subset of antennas that should be utilized, where the cardinality of S is $\hat{N}$. Due to the influence of the ℓ_1-norm squared minimization, the minimum power beamforming vector cannot be directly extracted from $\mathbf{X}^{(s)}$. Thus, to find the minimum power solution, problem (8) is solved for the reduced size problem, namely $\mathbf{X} \in \mathbb{C}^{\hat{N}\times\hat{N}}$, where $\mathbf{Q}_m$ is an $\hat{N} \times \hat{N}$ matrix obtained by omitting the channel entries corresponding to the left-out antennas. Due to the rank relaxation, the solution to problem (8), denoted by $\mathbf{X}^{(o)}$, might not be a rank-one matrix in general; hence, the optimum beamforming vector cannot be directly extracted from the obtained $\mathbf{X}^{(o)}$. However, it is possible to adopt the approach of [7], where an approximate solution to the original problem (1) can be found using a randomization technique by generating candidate beamforming vectors from $\mathbf{X}^{(o)}$, and choosing the vector that when scaled to satisfy the SNR constraints, yields the minimum total power. When $\mathbf{X}^{(o)}$ is rank-one, its principal component will be the optimal beamforming vector solution to problem (1).

The sparsest solution, meaning the one with the minimum number of antennas, that can be obtained using this approach corresponds to using $\lambda \to \infty$ in (9), or equivalently

$$
\begin{aligned}
\min_{\mathbf{X}\in\mathbb{C}^{N\times N}} \quad & \mathrm{tr}(\mathbf{1}_{N\times N} |\mathbf{X}|) \\
\text{s.t.}: \quad & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m) \geq 1,\ m = 1, \ldots, M, \quad \mathbf{X} \succeq 0.
\end{aligned} \tag{10}
$$

Note that the SNR constraints can be satisfied even with only one antenna transmitting at sufficiently high power if

$$
\max_{m\in\{1,\cdots,M\}} \min_{n\in\{1,\cdots,N\}} |h_{mn}| =: \tilde{h} > 0,
$$

where h_{mn} is the n-th component of $\mathbf{h}_m$, at the cost of perhaps very high transmission power ($= \tilde{h}^{-1}$). Use of the *size-sensitive* ℓ_1-norm (or ℓ_1-norm squared), however, often precludes very sparse solutions, simply because they cost too much in terms of ℓ_1 cost. This motivates adapting the sparsity-enhancing iteratively re-weighted ℓ_1-norm idea, originally proposed in the context of (linear) compressed sensing problems [1], to the present context.

3.3. Enhancing Sparsity: Iterative Algorithm

To further increase the sparsity of $\mathbf{w}$, the iteratively re-weighted ℓ_1-norm penalty in [1] is adapted to suit our problem. Consider the weight vector $\mathbf{u}$, where $u_1, u_2, \ldots, u_N$ are positive weights, and $\mathbf{X} := \mathbf{w}\mathbf{w}^{\mathcal{H}}$ as before. The weighted ℓ_1-norm squared of $\mathbf{w}$ can be expressed as

$$
\begin{aligned}
\left(\sum_{n=1}^{N} u_n |w_n|\right)^2 &= \mathbf{u}^T |\mathbf{X}| \mathbf{u} = \mathrm{tr}(\mathbf{U}|\mathbf{X}|) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} U_{i,j} |X_{i,j}|
\end{aligned}
$$

where $\mathbf{U} = \mathbf{u}\mathbf{u}^T$ denotes the weight matrix.

The iterative algorithm is as follows:

1. *Initialize* the iteration count to $r = 0$, and the weight matrix to $\mathbf{U}^{(0)} = \mathbf{1}_{N\times N}$.
2. *Solve* the weighted ℓ_1-norm squared minimization SDP problem

$$
\begin{aligned}
\min_{\mathbf{X}^{(r)}\in\mathbb{C}^{N\times N}} \quad & \mathrm{tr}(\mathbf{U}^{(r)} |\mathbf{X}^{(r)}|) \\
\text{s.t.}: \quad & \mathrm{tr}(\mathbf{X}^{(r)}\mathbf{Q}_m) \geq 1,\ m = 1, \ldots, M, \\
& \mathbf{X}^{(r)} \succeq 0
\end{aligned} \tag{11}
$$

   to obtain the optimum $\mathbf{X}^{(r)}$ at the r-th iteration.
3. *Update* the weight matrix to be used in the next iteration as $\mathbf{U}_{i,j}^{(r+1)} = 1/(|\mathbf{X}_{i,j}^{(r)}| + \epsilon)$ for each $i, j = 1, \ldots, N$.
4. Terminate on convergence, or, when a certain maximum number of iterations for r is reached. Otherwise, increment r, and go to step 2.

The weight matrix updates force small entries of $|\mathbf{X}|$ to zero and avoid unduly restraining large entries. The small parameter ϵ provides stability, and ensures that a zero-valued entry of $\mathbf{X}^r$ does not strictly prohibit a nonzero estimate at the next step. In the initial step of the iterative algorithm, problem (10) is solved for initialization. Convergence of this algorithm is very fast ($\sim 5 - 10$ iterations), as observed in the simulations.

4. PROPOSED ALGORITHM

The proposed algorithm that jointly selects $K \leq N$ antennas and finds the corresponding beamforming vector that minimized transmit-power, subject to receive-SNR constraints for each user can be summarized as follows:

1. Apply the weighted ℓ_1-norm iterative algorithm, and find the sparsest solution $\mathbf{X}^*$ and corresponding weights $\mathbf{U}^*$ when the algorithm terminates. Let $\hat{N}^*$ denote the number of nonzero diagonal entries in $\mathbf{X}^*$. If $\hat{N}^* > K$, then the algorithm fails to provide a sparse enough solution. If $\hat{N}^* = K$, then pick S to be the antennas corresponding to the nonzero diagonal entries of $\mathbf{X}^*$ and skip to step 3. Otherwise, go to the next step.

2. Solve the SDP problem

$$\begin{array}{ll} \min\limits_{\mathbf{X}\in\mathbb{C}^{N\times N}} & \mathrm{tr}(\mathbf{X} + \lambda\mathbf{U}^*|\mathbf{X}|) \\ \text{s.t.}: & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m) \geq 1,\ m = 1,\ldots,M, \quad \mathbf{X} \succeq 0. \end{array} \tag{12}$$

   using the obtained weights $\mathbf{U}^*$, which is problem (9) with $\mathbf{1}_{N\times N}$ replaced by $\mathbf{U}^*$, and use binary search to find λ that gives the required number of antennas K. The binary search procedure works as follows. For a given upper bound λ_{UB} and lower bound λ_{LB}, set $\lambda = (\lambda_{UB} - \lambda_{LB})/2 + \lambda_{LB}$ and solve the SDP problem. Let $\mathbf{X}^{(s)}$ denote the solution of (12) having $\hat{N}$ nonzero diagonal entries. If $\hat{N} = K$, then find the subset of selected antennas S corresponding to the nonzero diagonal entries of $\mathbf{X}^{(s)}$, and move to the next step. Otherwise, if $\hat{N} > K$ then set $\lambda_{LB} = \lambda$ while if $\hat{N} < K$ then set $\lambda_{UB} = \lambda$, and repeat until $\hat{N} = K$.

3. Now that K antennas have been selected, (8) is solved for the reduced-size problem, namely $\mathbf{X} \in \mathbb{C}^{K\times K}$, to find the minimum power beamforming vector. If the solution, denoted as $\mathbf{X}^{(o)}$, is rank-one, then its principal component is the minimum power beamforming vector. Otherwise, use the randomization technique of [7] to generate candidate beamforming vectors from $\mathbf{X}^{(o)}$, and select the one that when scaled to satisfy the SNR constraints, gives the minimum power.

Note that early termination of the binary search when a solution with fewer than the desired K antennas has been obtained will result in higher transmission power, but maybe used to speed up the process. Suitable λ_{LB} and λ_{UB} can be obtained empirically, using prior information on N, M, and the channel type.

Although the binary search over λ may require solving (12) more than once for different values of λ until the appropriate one is found, an important advantage over the exhaustive search method is that the number of iterations is independent of N and K, unlike exhaustive search, which requires solving $\binom{N}{K}$ problems of type (8). The solution obtained using the novel algorithm occasionally coincides with that obtained using exhaustive search, while the transmission power increase for the other cases is insignificant, as demonstrated in the next section.

5. NUMERICAL TESTS

To test the proposed SDP-based algorithms, YALMIP was used since it is implemented as a free toolbox for MATLAB [4]. YALMIP uses SeDuMi, a MATLAB implementation of second-order interior-point methods, for the actual computations [8].

K	Selected antennas S	Power (dBm)	Power inc. (dB)	λ ($\times 10^{-6}$)	Total itrs.
8	{1,...,8}	28.14 (30.9)	0 (0)	0	1
7	{1,2,3,4,5,7,8}	28.20 (31.16)	0.06 (0.26)	6.25	13
6	{1,2,3,4,5,7}	28.90 (31.30)	0.76 (0.35)	7.813	15
5	{2,3,4, 5,7}	29.37 (31.44)	1.23 (0.54)	12.5	12
4	{2,4,5,7}	29.72 (32.26)	1.58 (1.36)	25	11
3	{2,5,7}	30.49 (32.42)	2.34 (1.52)	50	10
2	{5,7}	32.87 (33.86)	4.73 (2.96)	75	11
1	{5}	35.78 (35.78)	7.64 (4.88)	∞	8

Table 1. Performance of the proposed algorithm for a particular channel realization for different antenna selection requirements.

The simulation setup assumes a BS having $N = 8$ transmit antennas broadcasting a common message to $M = 16$ receivers. Independent identically distributed (i.i.d.) Rayleigh fading channel vectors $\{\mathbf{h}_m\}_{m=1}^{M}$ were generated, each with entries i.i.d. circularly symmetric zero-mean complex Gaussian random variables of variance 1. The noise variance was set to $\sigma^2 = 1$, and the minimum required SNR to $\gamma = 1$ at all users.

To gain insights, detailed results are provided first for a single "typical" channel realization, which allows comparing the selected antenna subsets with the baseline exhaustive search solution. Running the first step of the proposed algorithm in Section 4 results in the sparsest solution of $\hat{N}^* = 1$ antenna, which corresponds to selecting antenna number 5. This result is obtained when the weighted ℓ_1-norm iterative algorithm converges after 8 iterations. It is worth noting that after the initial step of the iterative weighted ℓ_1-norm algorithm (which is equivalent to solving problem (10)), the resulting sparse solution has $\hat{N} = 6$ antennas, many more than the single antenna solution obtained after the iterative weighted ℓ_1-norm algorithm terminates. It was empirically found that for $N = 8$ and Rayleigh fading channels, setting $\lambda_{UB} = 10^{-4}$ and $\lambda_{LB} = 0$ is sufficient to cover the required range of λ for the binary search step in the proposed algorithm.

Table 1 summarizes the results obtained using the novel algorithm for this representative channel realization. The required number of antennas to be selected (or, the available number of RF chains) K is given in column 1. The subset of selected antennas is given in column 2. The minimum transmit-power obtained for each K is listed in column 3 (in dBm units), where the lower bound on the transmission power $\mathrm{tr}(\mathbf{X}^{(o)})$ and the actual power of the beamforming vector extracted from $\mathbf{X}^{(o)}$ are both reported. The increase of transmission power (compared to the case of using all $N = 8$ antennas) due to antenna selection is given in column 4 (in dB units). Finally, the value of λ used in problem (12) to select the required K antennas is given in column 5, whereas the total number of SDP problems solved in order to obtain the required solution is shown in the last column. Similarly, Table 2 summarizes the results obtained when exhaustive search is used for antenna selection.

The results in Table 1 demonstrate that as λ increases, the num-

K	Selected antennas S	Power (dBm)	Power inc. (dB)	Total itrs.
8	{1,...,8}	28.14 (30.9)	0 (0)	1
7	{1,2,3,4,5,7,8}	28.20 (31.16)	0.06 (0.26)	8
6	{1,2,4,5,7,8}	28.53 (31.59)	0.39 (0.85)	28
5	{2,4,5,7,8}	28.96 (31.88)	0.82 (0.98)	56
4	{2,4,5,7}	29.72 (32.26)	1.58 (1.36)	70
3	{2,5,7}	30.49 (32.42)	2.34 (1.52)	56
2	{5,7}	32.87 (33.86)	4.73 (2.96)	28
1	{5}	35.78 (35.78)	7.64 (4.88)	8

Table 2. Performance of the exhaustive search algorithm for a particular channel realization for different antenna selection requirements.

K	Proposed Algorithm				Exhaustive Search		
	Power Inc. (dB)		SDP Itrs.		Power Inc. (dB)		SDP
	Avg.	Max.	Avg.	Max.	Avg.	Max.	Itrs.
6	0.6 (0.1)	1.2 (1.9)	15.7	28	0.37 (0.01)	0.7 (1.12)	28
4	1.77 (1)	2.9 (2.3)	17.3	41	1.36 (0.5)	1.8 (2.3)	70
2	4.2 (2.8)	6.6 (6.7)	11.2	13	3.9 (2)	6.6 (6.7)	28

Table 3. Performance comparison between the proposed algorithm and the exhaustive search.

ber of antennas selected for transmission decreases (as the solution becomes more sparse), and the corresponding minimum transmission power increases, due to the decrease in degrees of freedom, as expected. Interestingly, the simulations suggest that the number of transmit antennas can be significantly reduced at only a small cost in terms of excess transmission power. Halving the number of antennas from 8 to 4, for example, entails about 1.5 dB extra power. Comparing with the exhaustive search results in Table 2, one can verify that exhaustive search slightly outperforms the proposed algorithm only for the cases of $K = 5$ and $K = 6$ antennas, by utilizing antenna number 8 instead of antenna number 3. However, the number of SDP problems that must be solved for the exhaustive search is significantly larger. The maximum number of iterations required for the binary search process, namely step 2 in the proposed algorithm, is 6 - these are needed to reach the value of $\lambda = 7.8125 \times 10^{-6}$ to select $K = 6$ antennas, where 8 SDP problems are solved for step 1, 6 for step 2, and 1 for the final step, yielding a worst-case total of 15 SDP problems. On the other hand, the exhaustive search algorithm requires solving $\binom{8}{4} = 70$ SDP problems to select $K = 4$ antennas. If for example it is required to select $K = 8$ antennas from an $N = 16$ antenna transmitter, exhaustive search requires solving $\binom{16}{8} = 12870$ SDP problems, which is clearly intractable. Finally, Table 3 reports the average and maximum increase in transmission power (in dB), along with the average and maximum number of SDP problems solved, for the proposed algorithm and for exhaustive search. The results are obtained for 100 different Rayleigh channel matrix realizations. The two main conclusions from the table are:

- The number of transmit antennas can be considerably reduced at a relatively small cost in terms of excess transmission power. If we halve the number of antennas, the transmission power increases by less than 3 dB (less than 2dB on average) to satisfy the SNR constraints.
- When compared to the exhaustive search, the proposed algorithm incurs much lower complexity solution at a very small additional power cost. The difference in power is less than 1 dB, on average.

6. CONCLUSIONS

We studied the problem of multicast beamforming with antenna selection, where the objective is to select a sparse beamforming vector of minimum power that meets the SNR constraints of all subscribers. Instead of using the ℓ_1 norm to promote sparsity, we argued that the ℓ_1-norm squared offers a more prudent sparsity-inducing regularization for our purposes. The reason is that it naturally (and elegantly) yields a semidefinite relaxation that is similar in spirit to the corresponding one for the baseline multicasting problem without antenna selection, considered in [7]. One interesting result is that the number of transmit antennas can be considerably reduced with only minimal increase in the transmission power. We also showed that our proposed algorithm performs joint antenna selection and weight optimization at significantly lower complexity than the exhaustive search algorithm, and at negligible excess power.

The novel algorithm can be modified to handle antenna selection for the related problem of maximizing the minimum received SNR over all subscribers, subject to a bound on the transmitted power, and the problem of maximizing the common mutual information of the (non-degraded) Gaussian broadcast channel in which the transmitter has N antennas, and each of the (non-cooperative) receivers has a single antenna. We will present these and other extensions in a forthcoming journal version.

7. REFERENCES

[1] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[2] J. de Andrade Jr., M. Campos, and J. Apolinario, "Sparse solutions for antenna arrays," in *Proc. XXIX Simposio Brasileiro de Telecomunicacoes*, Curitiba, Oct. 2-5, 2011.

[3] A. Dua, K. Medepalli, and A. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Communications*, vol. 5, no.9, pp. 2353–2357 Sept. 2006.

[4] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD*, Taipei, Sep. 4, 2004.

[5] S. Nai, W. Ser, Z. Yu, and H. Chen, "Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization," *IEEE Trans. Antennas and Propagation*, vol. 58, no. 12, Dec. 2010.

[6] S. Park and D. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2524–2534, June 2008.

[7] N. Sidiropoulos, T. Davidson, and Z. Luo, "Transmit beamforming for physical layer multicasting," *IEEE Trans. Signal Processing*, vol. 54, no. 6, part 1, pp. 2239–2251, June 2006.

[8] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1, pp. 625–653, 1999.

[9] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal. Statist. Soc B.*, vol. 58, no. 1, pp. 267–288, 1996.

[10] M. Yukawa and I. Yamada, "Minimal antenna-subset selection under capacity constraint for power-efficient MIMO systems: A relaxed ℓ_1 minimization approach," in *Proc. ICASSP*, March 14-19, 2010, Dallas, Texas.