On Multicast Beamforming for Minimum Outage

Vassilis Ntranos, Nicholas D. Sidiropoulos, and Leandros Tassiulas

***Abstract*—The multicast beamforming problem is considered from the viewpoint of minimizing outage probability subject to a transmit power constraint. The main difference with the point-to-point transmit beamforming problem is that in multicast beamforming the channel is naturally modeled as a Gaussian mixture, as opposed to a single Gaussian distribution. The Gaussian components in the mixture model user clusters of different means (locations) and variances (spreads). It is shown that minimizing outage probability subject to a transmit power constraint is an NP-hard problem when the number of Gaussian kernels, J, is greater than or equal to the number of transmit antennas, N. Through dimensionality reduction, it is also shown that the problem is practically tractable for $2-3$ Gaussian kernels. An approximate solution based on the Markov inequality is also proposed. This is simple to compute for any J and N, and often works well in practice.**

***Index Terms*—Multicast beamforming, outage probability, transmit power constraint.**

I. Introduction

CONSIDER a base station or wireless access point that uses an antenna array to transmit *common* information to a pool of users, each equipped with a single receive antenna. When the channel vectors of all users are known at the transmitter, it is possible to beamform in a way that directs power towards the users and limits wasteful radiation in other directions. This is a physical layer multicasting approach that has been recently investigated in a series of papers [7]–[9], [12]. The design formulations in [7]–[9], [12] target signal to (interference plus) noise ratio (SNR) guarantees: they either minimize total transmitted power subject to guaranteed SNR for each receiver, or maximize the minimum SNR subject to an overall transmitted power constraint.

Beamforming does not in general attain the multicast channel capacity - this may require a higher-rank transmit covariance [4], [12]. Beamforming, however, is a relatively simple approach that often operates close to multicast capacity - see [7], [8], [12] and the asymptotic capacity scaling results in [4]. Note that the results in [4] assume an isotropic i.i.d. Rayleigh model, which is pessimistic. Multicast beamforming is far more effective when the channel vectors are clustered around a few dominant directions [7], [8], [12].

Exact channel state information (CSI) will never be available in practice, in which case it is impossible to guarantee instantaneous SNR. An alternative is to offer average (expected) SNR guarantees. The channel correlation matrices (which vary far slower than the actual channel realizations) are then sufficient for transmit optimization, and the solutions in [7]–[9], [12] carry over almost verbatim. The drawback is that persistent deep fading can occur in this case, which is unacceptable for delay-sensitive applications. An alternative is to start with a set of nominal channel vectors, allow limited perturbation, and aim for a conservative design that guarantees a certain SNR for every allowable perturbation; see [6], [8] for related results in the context of multicasting, and references therein for earlier work on robust unicast beamforming.

A different approach is pursued here. Channel vectors are modeled as random, with a known distribution. The objective is to design the weight vector of the transmit beamformer to minimize the outage probability, i.e., the probability that the useful received signal power falls below a certain threshold. In a multicast context, this has the following interpretation: If one draws a large number of channel vectors, then the fraction of users served will be approximately one minus the outage probability. Minimizing the outage probability thus approximately maximizes the number of users served.

Several outage probability - based design problems have been considered in the wireless communications literature, e.g., [2], [3], [5], [11], [13], [15]–[18]. The power control problem for fading interference channels under outage probability specifications has been considered in [5]. Outage probability - based robust multiuser detection has been considered in [18] (see also [17]) for multicarrier CDMA systems, and [11] for space-time block-coded multiple-input multiple-output systems. Closer to our present context, unicast beamforming under an outage probability constraint has been considered in [15], [16], and minimum outage probability beamforming for point-to-point multiple-input single-output systems in [13]. Beamforming under outage probability constraints has been considered in [2] for the cellular uplink, and [3] for the cellular downlink.

The main difference of our setup is that we adopt a Gaussian mixture distribution for the channel vectors, whereas a single Gaussian is used in the aforementioned references. A Gaussian mixture model is natural for wireless multicasting, where subscribers are spatially dispersed in a non-uniform fashion (e.g., clustered in malls, squares, campuses, etc). Also, given enough kernels, it is possible to approximate almost any density by a Gaussian mixture.

II. Model

Consider a base station or access point equipped with N antennas, transmitting a common information-bearing multicast signal to K subscribers, each equipped with a single

Manuscript received June 27, 2008; revised November 10, 2008 and February 13, 2009; accepted February 13, 2009. The associate editor coordinating the review of this paper and approving it for publication was O. Simeone.

V. Ntranos and N. D. Sidiropoulos are with the Department of Electronic and Computer Engineering, Technical University of Crete, 73100 Chania - Crete, Greece (e-mail: nikos@telecom.tuc.gr; vntranos@gmail.com).

L. Tassiulas is with the Department of Computer Engineering and Telecommunications, University of Thessaly, Volos, Greece (e-mail: leandros@uth.gr).

N. D. Sidiropoulos was partially supported by EC project WIP and N-CRAVE. L. Tassiulas was partially supported by EC project N-CRAVE.

Digital Object Identifier 10.1109/TWC.2009.080850

receive antenna. While it is possible for each subscriber to estimate its $N \times 1$ channel vector $\mathbf{h}_k$ and send CSI back to the transmitter, this may not be desirable for a number of reasons. These include the need for a separate uplink channel with significant signaling overhead (especially for large K); energy considerations if battery-operated devices are involved; and privacy concerns. Ideally, subscribers should be able to join or drop listening without notifying the transmitter (logging). In this case, the transmitter must operate without knowing the instantaneous user channels $\{\mathbf{h}_k\}_{k=1}^K$ or the associated operational correlation matrices, or even the number of users currently listening (K). Still, the transmitter may exploit *prior* information about the operating scenario, in the form of the distribution of channel vectors.

In our present context, it is reasonable to assume that each channel vector, $\mathbf{h}_k$, is independently drawn from a common marginal $f(\mathbf{h})$. Independence is plausible but not crucial here - it suffices that the $\{\mathbf{h}_k\}$ process satisfies appropriate ergodic mixing conditions, as we will see. The marginal $f(\mathbf{h})$ can be fitted from samples collected in a measurement campaign or field trial.

What would be a good model for the marginal distribution $f(\mathbf{h})$? A Gaussian can be appropriate in certain cases; but it can only capture subscribers scattered around a single location, whereas there may be several 'hotspots' (squares, malls, campus) in a service area. This suggests using a Gaussian mixture

$$f(\mathbf{h}) = \sum_{j=1}^{J} p_j \mathcal{N}(\mathbf{h}; \mathbf{m}_j, \sigma_j^2 \mathbf{I}),$$

where $\mathcal{N}((\cdot); \mathbf{m}, \mathbf{C})$ denotes a multivariate Gaussian distribution of mean vector $\mathbf{m}$ and covariance matrix $\mathbf{C}$, assumed diagonal for simplicity; and $p_j \geq 0$, $\sum_{j=1}^{J} p_j = 1$ are the prior (mixture) probabilities. For high enough J and appropriate choice of parameters, the Gaussian mixture model can fit any empirical distribution.

A real-valued baseband-equivalent channel model is appropriate when one-dimensional modulation (PAM) is used, in which case there is no In-phase / Quadrature (I/Q) processing at the transmitter or the receiver. Binary PAM is sometimes used in the low SNR regime, or for hardware simplicity and cost considerations. On the other hand, most applications of wireless multicast entail high rates, and thus I/Q processing and higher-order PSK or QAM modulation. In those cases, the baseband-equivalent channel model is complex, and a complex Gaussian mixture is appropriate

$$f(\mathbf{h}) = \sum_{j=1}^{J} p_j \mathcal{CN}(\mathbf{h}; \mathbf{m}_j, \sigma_j^2 \mathbf{I}),$$

where $\mathcal{CN}$ denotes a multivariate complex Gaussian distribution.

In the real case, the beamforming weight vector $\mathbf{w}$ will be real. At the receiver, the signal will be scaled by $y := \mathbf{w}^T \mathbf{h}$, with

$$f(y; \mathbf{w}) = \sum_{j=1}^{J} p_j \mathcal{N}(y; \mathbf{w}^T \mathbf{m}_j, \sigma_j^2 ||\mathbf{w}||^2),$$

i.e., a mixture of univariate Gaussians (throughout, $||\cdot||$ denotes the Euclidean norm). In the complex case, we take the beamforming vector to be the conjugate of $\mathbf{w}$ (for convenience); at the receiver, the signal will be scaled by $z := \mathbf{w}^H \mathbf{h}$, where $(\cdot)^H$ denotes Hermitian (conjugate) transpose, and

$$f(z; \mathbf{w}) = \sum_{j=1}^{J} p_j \mathcal{CN}(z; \mathbf{w}^H \mathbf{m}_j, \sigma_j^2 ||\mathbf{w}||^2),$$

i.e., a mixture of univariate complex Gaussians.

We will consider the problem of choosing $\mathbf{w}$ to minimize the outage probability. In the real case,

$$\min_{||\mathbf{w}||^2 = P} Pr[|y| < \gamma] \iff$$

$$\min_{||\mathbf{w}||^2 = P} \sum_{j=1}^{J} p_j \int_{-\gamma}^{\gamma} \mathcal{N}(y; \mathbf{w}^T \mathbf{m}_j, \sigma_j^2 ||\mathbf{w}||^2) \iff$$

$$\min_{||\mathbf{w}||^2 = P} \sum_{j=1}^{J} p_j \int_{-\gamma}^{\gamma} \mathcal{N}(y; \mathbf{w}^T \mathbf{m}_j, \sigma_j^2 P),$$

and likewise in the complex case.

The motivation for this formulation is two-fold. First, suppose that a new potential subscriber ('customer') wishes to join the particular multicast. Then the above formulation maximizes the probability that the new customer will be served. Second, if one draws a large number of customer channel vectors, then $1 - Pr[|y| < \gamma]$ is an estimate of the fraction of them that will be served. Thus picking $\mathbf{w}$ to minimize $Pr[|y| < \gamma]$ approximately maximizes the number of customers served in the large sample regime: the fraction converges to $1 - Pr[|y| < \gamma]$ under quite general ergodic mixing conditions, notably when the channel vectors are drawn independently from $f(\mathbf{h})$.

Remark 1: Consider the case of a point-to-point multiple-input single-output link as in [13], but this time without assuming any CSI feedback from the receiver to the transmitter. Suppose there are J possible channel states, each associated with a different Gaussian distribution of the channel vector, and let p_j denote the probability of the j-th channel state. Under ergodic mixing conditions on the channel process, minimizing outage approximately maximizes the fraction of time that the link meets a minimum SNR requirement. In this context, averaging is with respect to the temporal channel variation, instead of the number of multicast customers.

III. RESULTS

Most of our results in the sequel are applicable in both the real and the complex case. For brevity and clarity of exposition, however, we first state and prove results for the real case, then discuss extensions to the complex case.

A. Special case: $J = 1$

When there is only one Gaussian kernel ($J = 1$), minimizing the outage probability under $||\mathbf{w}||^2 = P$ reduces to maximizing $|\mathbf{w}^T \mathbf{m}_1|$ in the real case, or $|\mathbf{w}^H \mathbf{m}_1|$ in the complex case, under the same constraint. From the Cauchy-Schwartz inequality, the optimum $\mathbf{w}$ is simply $\mathbf{m}_1$ scaled to power P (note there is freedom to choose the sign in the real case, or phase in the complex case; if $\mathbf{m}_1 = \mathbf{0}$, then any $\mathbf{w}$

on the sphere of radius $\sqrt{P}$ is equally good). The solution is trivial - but also interesting in the following way: for a large number of customers, it approximately maximizes the number of customers served. This is interesting, because even if the channel vectors were exactly known at the transmitter, exactly maximizing the number of customers served is NP-hard, and even approximate solutions are non-trivial, see [10]. Thus, when the number of customers is large, we can approximately maximize the number of customers served by matching the weight vector to the mean vector; but exactly maximizing the number of customers served is prohibitive, even when all channel vectors are known exactly at the transmitter.

Unfortunately, there's no escape from NP-hardness for large-enough J, as the next result shows.

B. NP-hardness

We have the following result, whose proof can be found in the Appendix. The intuition behind it is as follows. Consider the case where the components (kernels) of the mixture are Dirac atoms, $\delta(\mathbf{h} - \mathbf{m}_j)$. Then, a beamforming vector will either miss or hit a certain component; each p_j will either be counted as outage or not *as a whole*, and a combinatorial counting problem emerges. Letting all p_j's be equal, minimizing outage is equivalent to maximizing the number of hits, which has been shown to be NP-hard in this context. The technical difficulty in the proof is to show that the same holds for Gaussian kernels of sufficiently narrow but non-singular support - which inevitably overlap.

Claim 1: Computing $\min_{||\mathbf{w}||^2=P} Pr[|y| < \gamma]$ is NP-hard for $J \geq N$.

Remark 2: For brevity, the proof given in the Appendix is for the real-valued case. The proof relies on the Cauchy-Schwartz inequality, continuity of the kernel densities, and the NP-hardness proof in [12] - all of which are valid in both the real and the complex case. The proof is therefore applicable in the complex case as well. Furthermore, there is nothing unique to the Gaussian distribution in so far as this proof is concerned - other continuous kernel densities could be employed just as well.

C. Real Case: $J = 2$, general N

We have seen that minimizing outage is trivial for $J = 1$, and NP-hard for $J \geq N$. Small values of $J < N$ are of practical interest, and in this section we present a practical algorithm for $J = 2$ real Gaussian kernels. In this case, the problem becomes

$$\min_{||\mathbf{w}||^2=P} \sum_{j=1}^{2} p_j \int_{-\gamma}^{\gamma} \mathcal{N}(y; \mathbf{w}^T\mathbf{m}_j, \sigma_j^2 P). \iff$$

$$\min_{||\mathbf{w}||^2=P} \sum_{j=1}^{2} p_j \left[\mathcal{Q}\left(\frac{-\gamma - \mathbf{w}^T\mathbf{m}_j}{\sigma_j} \right) - \mathcal{Q}\left(\frac{\gamma - \mathbf{w}^T\mathbf{m}_j}{\sigma_j} \right) \right].$$

Let

$$\mathcal{M}(\mathbf{w}) = \sum_{j=1}^{2} p_j \left[\mathcal{Q}\left(\frac{-\gamma - \mathbf{w}^T\mathbf{m}_j}{\sigma_j} \right) - \mathcal{Q}\left(\frac{\gamma - \mathbf{w}^T\mathbf{m}_j}{\sigma_j} \right) \right]$$

Then, $\mathbf{w}_o = arg \min_{||\mathbf{w}||^2=P} \mathcal{M}(\mathbf{w})$ will lie on the subspace $\mathcal{V}$, spanned by the mean vectors $\mathbf{m}_1, \mathbf{m}_2$ (otherwise, power allocated in a direction out of $\mathcal{V}$ would be wasted). Then $\mathbf{w}^T\mathbf{m}_j$ can be parameterized as $||\mathbf{w}||||\mathbf{m}_j|| \cos(\angle w - \angle m_j)$, where $\angle x := \arccos(\mathbf{x}^T\mathbf{v}_r/||\mathbf{x}||||\mathbf{v}_r||)$ is the angle between a vector $\mathbf{x}$ and a reference vector $\mathbf{v}_r$ in the two dimensional space $\mathcal{V}$. For simplicity we may take $\mathbf{v}_r = \mathbf{m}_1/||\mathbf{m}_1||$, and the objective function becomes

$$\mathcal{M}(\angle w) = \sum_{j=1}^{2} p_j \left[\mathcal{Q}\left(\frac{-\gamma - \sqrt{P}||\mathbf{m}_j|| \cos(\angle w - \angle m_j)}{\sigma_j} \right) - \mathcal{Q}\left(\frac{\gamma - \sqrt{P}||\mathbf{m}_j|| \cos(\angle w - \angle m_j)}{\sigma_j} \right) \right],$$

and $\angle w_o = arg \min_{||\mathbf{w}||^2=P} \mathcal{M}(\angle w)$ determines $\mathbf{w}_o = arg \min_{||\mathbf{w}||^2=P} \mathcal{M}(\mathbf{w})$ via

$$\mathbf{w}_o = \sqrt{P}\mathbf{V}\mathbf{Q}\mathbf{V}^T\mathbf{v}_r,$$

where

$$\mathbf{Q} := \begin{bmatrix} \cos(\angle w) & -\sin(\angle w) \\ \sin(\angle w) & \cos(\angle w) \end{bmatrix}$$

is a rotation matrix, $\mathbf{V} := [\mathbf{v}_1, \mathbf{v}_2]$ is the orthonormal basis of $\mathcal{V}$ and $\mathbf{v}_r$ is the reference vector used above. We can therefore find the optimal $\angle w_o$ and $\mathbf{w}_o$ (up to desired accuracy) by one-dimensional line search over $\angle w_o$.

Minimizing a function of a real scalar variable may appear benign, but can be very hard. In our particular context, however, we aim to minimize a continuous (albeit multi-modal) function over a bounded interval, and the function is easy to evaluate. We therefore use a two-step grid search: we run a grid search once, select the best bin, then run another grid search that zooms-in on the chosen bin. We call this two-step process *fine grid search*. In our experiments, we chose grids that were fine enough to yield accurate results. Alternatively, the same accuracy can be obtained using a relatively coarser grid search, followed by gradient descent or Newton iterations. The objective function

$$\mathcal{M}(\mathbf{w}) = \sum_{j=1}^{J} p_j \left[\mathcal{Q}\left(\frac{-\gamma - \mathbf{w}^T\mathbf{m}_j}{\sigma_j} \right) - \mathcal{Q}\left(\frac{\gamma - \mathbf{w}^T\mathbf{m}_j}{\sigma_j} \right) \right],$$

is differentiable with

$$\nabla\mathcal{M}(\mathbf{w}) = \sum_{j=1}^{J} \frac{p_j \mathbf{m}_j}{\sigma_j^2 \sqrt{2\pi}} \left[e^{\frac{(-\gamma - \mathbf{w}^T\mathbf{m}_j)^2}{2\sigma_j^2}} - e^{\frac{(\gamma - \mathbf{w}^T\mathbf{m}_j)^2}{2\sigma_j^2}} \right].$$

This combination of grid search and gradient descent can be computationally more efficient than fine grid search, depending on the required accuracy. We used fine grid search throughout our simulations.

D. Real Case: $J = 3$, general N

The complexity of grid search is exponential in the number of dimensions. This is one reason why dimensionality reduction using properties of the objective function is a very important step, prior to any numerical minimization. We show that the case of three real Gaussian kernels can be reduced to a two-dimensional minimization problem, which is the practical limit of what we can handle using grid search. Consider

$$\mathcal{M}(\mathbf{w}) = \sum_{j=1}^{3} p_j \left[\mathcal{Q}\left(\frac{-\gamma - \mathbf{w}^T \mathbf{m}_j}{\sigma_j} \right) - \mathcal{Q}\left(\frac{\gamma - \mathbf{w}^T \mathbf{m}_j}{\sigma_j} \right) \right].$$

Again, $\mathbf{w}_o = arg \min_{||\mathbf{w}||^2=P} \mathcal{M}(\mathbf{w})$ will lie on the subspace $\mathcal{V}$, spanned by the mean vectors $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$. Let $\mathbf{V} := [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ be an orthonormal basis of subspace $\mathcal{V}$. We can use spherical coordinates (r, θ, ϕ) to write the weight and mean vectors as:

$$\mathbf{w} = \sqrt{P}\mathbf{V}\hat{\mathbf{r}}_{\theta_w,\phi_w}, \ \mathbf{m}_j = ||\mathbf{m}_j||\mathbf{V}\hat{\mathbf{r}}_{\theta_j,\phi_j},$$

where

$$\hat{\mathbf{r}}_{\theta_j,\phi_j} = \begin{bmatrix} \cos(\theta_j)\sin(\phi_j) \\ \sin(\theta_j)\sin(\phi_j) \\ \cos(\phi_j) \end{bmatrix}$$

is the spherical position vector, θ_j is the angle between the projection of the j-th mean vector on the $\mathbf{v}_1$, $\mathbf{v}_2$ plane, and ϕ_j is the angle between the j^{th} mean vector and $\mathbf{v}_3$.

We may now write $\mathcal{M}(\mathbf{w})$ as a function of θ_w, ϕ_w, and perform a two-dimensional fine grid search of $\mathcal{M}(\theta_w, \phi_w)$ to compute $\{\theta_{w_o}, \phi_{w_o}\} = arg \min_{||\mathbf{w}||^2=P} \mathcal{M}(\theta_w, \phi_w)$, up to desired accuracy. The associated beamforming vector is then given by $\mathbf{w}_o = \sqrt{P}\mathbf{V}\hat{\mathbf{r}}_{\theta_{w_o},\phi_{w_o}}$.

E. Complex Case: $J = 2$, general N

We now generalize to the complex case. Here, we have

$$f(\mathbf{h}) = \sum_{j=1}^{2} p_j \mathcal{CN}(\mathbf{h}; \mathbf{m}_j, \sigma_j^2 \mathbf{I}),$$

and with $z := \mathbf{w}^H \mathbf{h}$,

$$f(z; \mathbf{w}) = \sum_{j=1}^{2} p_j \mathcal{CN}(z; \mathbf{w}^H \mathbf{m}_j, \sigma_j^2 ||\mathbf{w}||^2).$$

The optimal beamforming vector can be found by

$$\min_{||\mathbf{w}||^2=P} Pr[|z| < \gamma] \Longleftrightarrow$$

$$\min_{||\mathbf{w}||^2=P} \sum_{j=1}^{2} p_j \iint_{\mathbf{A}} \mathcal{CN}(z; \mathbf{w}^H \mathbf{m}_j, \sigma_j^2 ||\mathbf{w}||^2),$$

where $\mathbf{A}$ is a disc of radius γ in the complex plane. The above integral is given by [1]

$$\iint_{\mathbf{A}} \mathcal{CN}(z; \mathbf{w}^H \mathbf{m}_j, \sigma_j^2 ||\mathbf{w}||^2) = \mathcal{P}\left[\left(\frac{\gamma}{\sigma_j ||\mathbf{w}||} \right)^2 \Big|_2, \left(\frac{|\mathbf{w}^H \mathbf{m}_j|}{\sigma_j ||\mathbf{w}||} \right)^2 \right]$$

where $\mathcal{P}[\chi^2|_2, \lambda]$ is the cdf of the non-central χ^2 distribution with two degrees of freedom and non-centrality parameter λ. Let

$$\mathcal{C}(\mathbf{w}) := \sum_{j=1}^{2} p_j \mathcal{P}\left[\left(\frac{\gamma}{\sigma_j ||\mathbf{w}||} \right)^2 \Big|_2, \left(\frac{|\mathbf{w}^H \mathbf{m}_j|}{\sigma_j ||\mathbf{w}||} \right)^2 \right].$$

Again, $\mathbf{w}_o = arg \min_{||\mathbf{w}||^2=P} \mathcal{C}(\mathbf{w})$ will lie on the subspace spanned by the complex mean vectors $\mathbf{m}_1, \mathbf{m}_2$. Thus, all candidate beamforming vectors can be written as $\mathbf{w} = c_1\mathbf{m}_1 + c_2\mathbf{m}_2$, with c_1, c_2 complex numbers such that $||\mathbf{w}||^2 = P$. We can use this constraint to find a relationship between c_1 and c_2:

$$||\mathbf{w}||^2 = P \Longleftrightarrow$$

$$(c_1^* \mathbf{m}_1^H + c_2^* \mathbf{m}_2^H)(c_1 \mathbf{m}_1 + c_2 \mathbf{m}_2) = P \Longleftrightarrow$$

$$|c_1|^2 ||\mathbf{m}_1||^2 + |c_2|^2 ||\mathbf{m}_2||^2 + 2\Re\{c_1^* c_2 \mathbf{m}_1^H \mathbf{m}_2)\} = P.$$

Using the rotational invariance of the outage probability in the complex plane

$$Pr[|\mathbf{w}^H \mathbf{h}| < \gamma] = Pr[|e^{j\omega} \mathbf{w}^H \mathbf{h}| < \gamma],$$

we can take c_2 to be real without loss of generality. Thus:

$$|c_1|^2 ||\mathbf{m}_1||^2 + c_2^2 ||\mathbf{m}_2||^2 + 2c_2 \Re\{c_1^* \mathbf{m}_1^H \mathbf{m}_2)\} = P.$$

We can now compute the optimal beamforming vector by performing a two-dimensional grid search in $\mathcal{C}(\mathbf{w})$ - one dimension for $\angle c_1 \in [0, 2\pi)$ and one for $|c_1| \in (0, \sqrt{P}/||\mathbf{m}_1||)$. For every c_1, we can compute c_2 through the constraint equation

$$c_{21,2} = \frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha},$$

where

$$\alpha = ||\mathbf{m}_2||^2,$$

$$\beta = 2\Re\{c_1^* \mathbf{m}_1^H \mathbf{m}_2\},$$

$$\gamma = |c_1|^2 ||\mathbf{m}_1||^2 - P.$$

Note that we need to check only one root for c_2: for every $c_1, c_1' = -c_1 \in \mathbb{C}$ it is easy to see that $c_{21} = -c_{22}' \Rightarrow c_1\mathbf{m}_1 + c_{21}\mathbf{m}_2 = -(c_1'\mathbf{m}_1 + c_{22}'\mathbf{m}_2) \Rightarrow \mathbf{w} = -\mathbf{w}'$. These two beamforming vectors are equivalent in terms of minimizing the outage probability because $\mathcal{C}(\mathbf{w}) = \mathcal{C}(-\mathbf{w}) = \mathcal{C}(\mathbf{w}')$.

F. General covariance matrix case

Up to this point we have made the assumption that channel vectors are drawn from a Gaussian mixture distribution with diagonal component covariance matrices. In general, one may have to drop that assumption to better fit a given channel (e.g., an ‘urban canyon’ ellipsoidal scenario). In this case, $f(\mathbf{h}) = \sum_{j=1}^{J} p_j \mathcal{CN}(\mathbf{h}; \mathbf{m}_j, \mathbf{C}_j)$ and the minimization problem becomes (recall that $\mathbf{A}$ is the disc of radius γ in the complex plane)

$$\min_{||\mathbf{w}||^2=P} \sum_{j=1}^{J} p_j \iint_{\mathbf{A}} \mathcal{CN}(z; \mathbf{w}^H \mathbf{m}_j, \mathbf{w}^H \mathbf{C}_j \mathbf{w}).$$

Note that the optimal beamforming vector no longer lies on the subspace spanned by the J mean vectors since $\mathbf{w}$ affects not only the mean of each univariate Gaussian in the projected mixture, but its variance as well. We will only consider a special but important case in the sequel.

*a) **Special case: J=1**:* When $J = 1$, the outage probability is given by

$$\mathcal{C}(\mathbf{w}) = \iint_{\mathbf{A}} \mathcal{CN}(z; \mathbf{w}^H\mathbf{m},\ \mathbf{w}^H\mathbf{C}\mathbf{w}).$$

Consider problem

$$Q: \min_{||\mathbf{w}||^2=P} \mathcal{C}(\mathbf{w}).$$

We have the following result, whose proof can be found in the Appendix.

Claim 2: Let $\mathbf{p}$ be the unit-norm principal component of $\left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right)\mathbf{C}\left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right)$, where $\mathbf{I}$ is the identity matrix. The optimal beamforming vector, $\mathbf{w}_o = arg\min_{||\mathbf{w}||^2=P}\mathcal{C}(\mathbf{w})$, lies on the subspace spanned by $\mathbf{m}$ and $\mathbf{p}$, and it can be found via the one-dimensional line search

$$\min_{0\leq c\leq\ P||\mathbf{m}||^2} \mathcal{C}\left(\sqrt{c}\frac{\mathbf{m}}{||\mathbf{m}||^2} + \sqrt{P - c/||\mathbf{m}||^2}\mathbf{p}\right).$$

G. Markov Approximation

From Markov's inequality we have that $Pr[x \geq t] \leq t^{-1}E[x]$, for any non-negative random variable. We can thus consider approximating

$$\min_{||\mathbf{w}||^2=P} Pr[|y| < \gamma] \Longleftrightarrow \max_{||\mathbf{w}||^2=P} Pr[|y| \geq \gamma] \Longleftrightarrow$$

$$\max_{||\mathbf{w}||^2=P} Pr[|y|^2 \geq \gamma^2]$$

by

$$\max_{||\mathbf{w}||^2=P} E[|y|^2],$$

thus maximizing an upper bound on the actual objective function (when put in maximization form). Now,

$$E[|y|^2] = \sum_{j=1}^{J} p_j\left(|\mathbf{w}^T\mathbf{m}_j|^2 + \sigma_j^2 P\right),$$

thus we may

$$\max_{||\mathbf{w}||^2=P} \sum_{j=1}^{J} p_j|\mathbf{m}_j^H\mathbf{w}|^2.$$

Solution of the latter problem is easy. Let

$$\mathbf{D} := diag([\sqrt{p_1}, \cdots, \sqrt{p_J}\ ]),\ \mathbf{M} := [\mathbf{m}_1, \cdots, \mathbf{m}_J]^H,$$

then $\mathbf{w}_{app} = arg\max_{||\mathbf{w}||^2=P} E[|y|^2]$ is given by the principal right singular vector of the matrix $\mathbf{DM}$ scaled to power P.

Of course, $\mathbf{w}_o$ does not in general solve the original problem of minimizing outage (maximizing service) probability; but it is interesting to note that in the special case of $J = 1$ (single Gaussian kernel) it does. Also note that $\mathbf{w}_o$ *is not* $\sum_{j=1}^{J} p_j\mathbf{m}_j$ normalized to power P, as quick intuition would perhaps suggest. To appreciate this, consider for example what happens when $J = 2$, $p_1 = p_2 = 1/2$, and $\mathbf{m}_2 = -\mathbf{m}_1$.

We note that it would have been preferable to maximize an achievable lower bound on the objective function, as opposed to an upper bound. Finding a suitable lower bound appears non-trivial here - this is an NP-hard problem which, unlike its perfect-CSI counterpart [12], does not appear amenable to semidefinite relaxation. Our original motivation for the Markov approximation was that it is optimal in the case of $J = 1$, *and* very simple to compute for any J. Surprisingly, at least in some cases, it turns out to perform remarkably well in terms of outage probability. We explore this and other issues next.

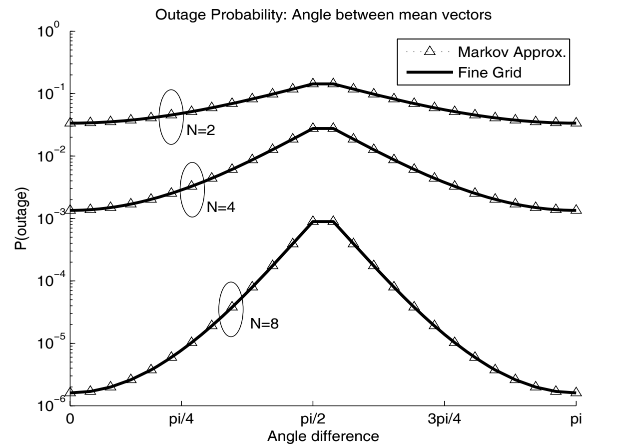

Fig. 1. Angle between mean vectors($J = 2$): Outage Probability as a function of the angle: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus $\hat{\phi}$.

IV. Numerical Results

A. Real case

When $J = 2$ or 3, we can effectively compute the optimal beamforming vector via low-dimensional fine grid search. In this case, we can evaluate how far the solution based on Markov approximation is from the optimal one. In four different scenarios for each case ($J = 2, J = 3$), we computed $\mathbf{w}_{opt} = arg\min_{||\mathbf{w}||^2=P}\mathcal{M}(\mathbf{w})$ and $\mathbf{w}_{app} = arg\max_{||\mathbf{w}||^2=P} E[|y|^2]$, through fine grid search and the Markov approximation respectively. The parameters for the different scenarios are given below. The results are summarized in Figs. 1 - 8, where curves are parameterized by the number of transmit antennas, N.

*1) **Angle between mean vectors (J=2)**:* Fig. 1 plots outage probability results for $p_1 = p_2 = 1/2$, $\sigma_1^2 = \sigma_2^2 = 1$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$, $||\mathbf{w}||^2 = P = 4$, as $\hat{\phi} := \angle m_2 - \angle m_1$ varies in $[0, \pi)$. The worst-case outage occurs when the two mean vectors are orthogonal, whereas the best situation is when the mean vectors are aligned, as expected. The Markov approximation is indistinguishable from the fine grid solution in this perfectly balanced scenario.

*2) **Magnitude of mean vectors (J=2)**:* Fig. 2 plots outage probability results for $p_1 = p_2 = 1/2$, $\sigma_1^2 = \sigma_2^2 = 1$, $\hat{\phi} = \pi/3$, $||\mathbf{m}_1||^2 = N$, $||\mathbf{w}||^2 = P = 4$, as $||\mathbf{m}_2||$ varies in $(0, 10]$. Notice that the Markov approximation breaks down when $||\mathbf{m}_2||^2$ is larger than about twice the value of $||\mathbf{m}_1||^2$ - a near-far situation. The gap between the Markov approximation and fine grid search becomes quite pronounced as the number of transmit antennas (N) increases.

*3) **Variance of the Gaussian Kernels (J=2)**:* Fig. 3 plots outage probability results for $p_1 = p_2 = 1/2$, $\hat{\phi} = \pi/3$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$, $\sigma_1^2 = 1$, $||\mathbf{w}||^2 = P = 4$, as σ_2^2 varies in $(0, 10]$. Notice that the outage probability increases

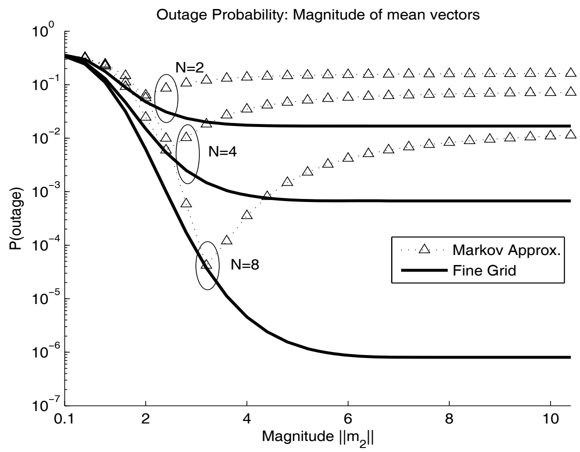

Fig. 2. Magnitude of mean vectors($J = 2$): Outage Probability as a function of the magnitude: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus $||\mathbf{m}_2||$.

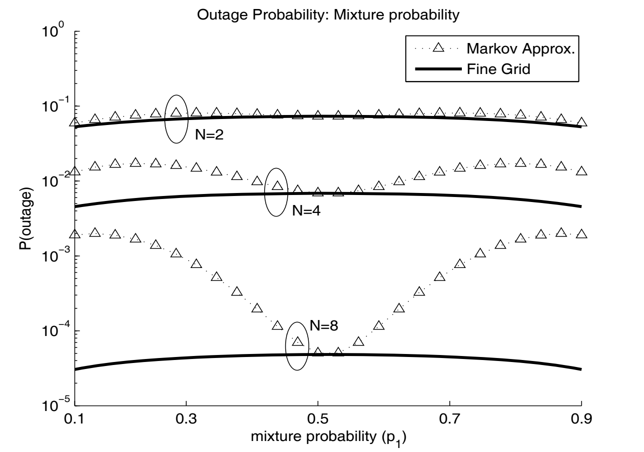

Fig. 4. Mixture Probability($J = 2$): Outage Probability as a function of the mixture probability: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus p_1.

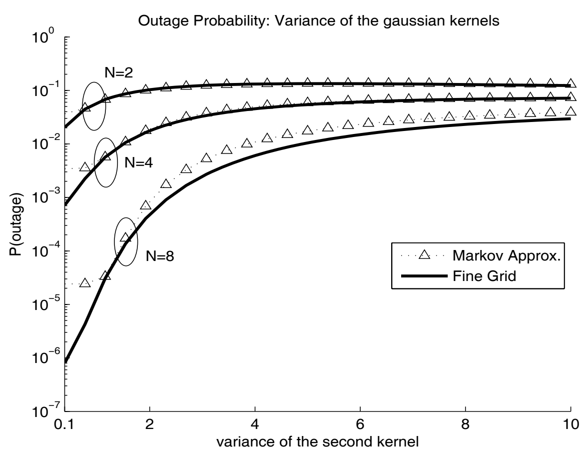

Fig. 3. Variance of the Gaussian Kernels($J = 2$): Outage Probability as a function of the variance: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus σ_2^2.

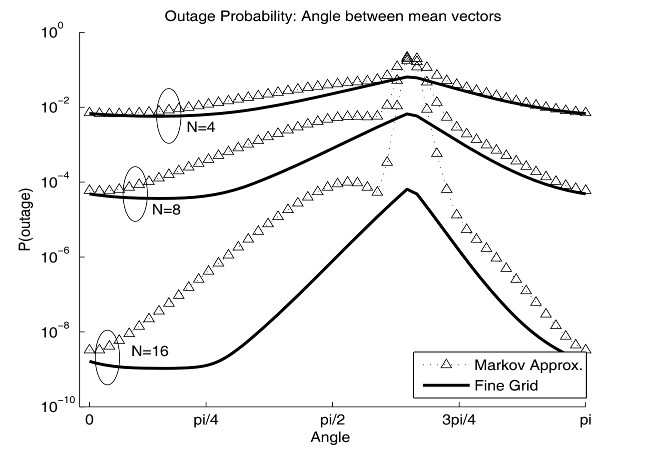

Fig. 5. Angle between mean vectors($J = 3$): Outage Probability as a function of the angle: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus ϕ_3.

with kernel spread, as expected (the scenario becomes less directional and the beamforming gain diminishes). Also note that the Markov approximation remains fairly close to fine grid search, despite the imbalance in kernel spreads.

*4) **Mixture probability (J=2)**:* Fig. 4 plots outage probability results for $\sigma_1^2 = \sigma_2^2 = 1$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$ $\hat{\phi} = \pi/3$, $||\mathbf{w}||^2 = P = 4$, as p_1 varies in $[0.1, 0.9]$. Whereas fine grid search is robust to mixture probability imbalance, this is not the case for the Markov approximation - an effect that becomes more pronounced as the number of transmit antennas increases.

*5) **Angle between mean vectors (J=3)**:* For $J = 3$, the numerical results are qualitatively similar to the $J = 2$ case, with one notable difference: the Markov approximation is considerably worse relative to fine grid search. Fine grid search provides clearly superior performance, especially in difficult scenarios with near-far or mixture probability imbalance effects.

Fig. 5 plots outage probability results for $p_j = 1/3, \sigma_j^2 = 1$, $||\mathbf{m}_j||^2 = N$, $\forall j$, $(\theta_1, \phi_1) = (\frac{\pi}{4}, \frac{\pi}{4})$, $(\theta_2, \phi_2) = (\frac{3\pi}{4}, \frac{\pi}{4})$, $\theta_3 = \frac{\pi}{4}$, $||\mathbf{w}||^2 = P = 4$, as ϕ_3 varies in $[0, \pi)$.

*6) **Magnitude of mean vectors (J=3)**:* Fig. 6 plots outage probability results for $p_j = 1/3$, $\sigma_j^2 = 1$, $\forall j$, $(\theta_1, \phi_1) = (\frac{\pi}{4}, \frac{\pi}{4})$, $(\theta_2, \phi_2) = (\frac{3\pi}{4}, \frac{\pi}{4})$, $(\theta_3, \phi_3) = (\frac{\pi}{2}, \frac{\pi}{2})$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$, $||\mathbf{w}||^2 = P = 4$, as $||\mathbf{m}_3||$ varies in $(0, 5]$.

*7) **Variance of the Gaussian Kernels (J=3)**:* Fig. 7 plots outage probability results for $p_j = 1/3$, $||\mathbf{m}_j||^2 = N$, $\forall j$, $(\theta_1, \phi_1) = (\frac{\pi}{4}, \frac{\pi}{4})$, $(\theta_2, \phi_2) = (\frac{3\pi}{4}, \frac{\pi}{4})$, $(\theta_3, \phi_3) = (\frac{\pi}{2}, \frac{\pi}{2})$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $||\mathbf{w}||^2 = P = 4$, as σ_3^2 varies in $(0, 5]$.

*8) **Mixture probability (J=3)**:* Fig. 8 plots outage probability results for $\sigma_j^2 = 1$, $||\mathbf{m}_j||^2 = N$, $\forall j$, $(\theta_1, \phi_1) = (\frac{\pi}{4}, \frac{\pi}{4})$, $(\theta_2, \phi_2) = (\frac{3\pi}{4}, \frac{\pi}{4})$, $(\theta_3, \phi_3) = (\frac{\pi}{2}, \frac{\pi}{2})$, $p_1 = p_2 = (1 - p_3)/2$,$||\mathbf{w}||^2 = P = 4$, as p_3 varies in $[0.1, 0.9]$.

B. Complex case

We now turn to the complex case. Only $J = 2$ kernels are considered. The results are summarized in Figs. 9 - 12. The key difference with respect to the $J = 2$ real case is that the Markov approximation is much closer to fine grid search in the complex case. There is no noticeable breakdown due to near-far or mixture probability imbalance effects.

Analytical explanation of this behavior is not easy - we are dealing with an NP-hard problem. Worst-case analysis of

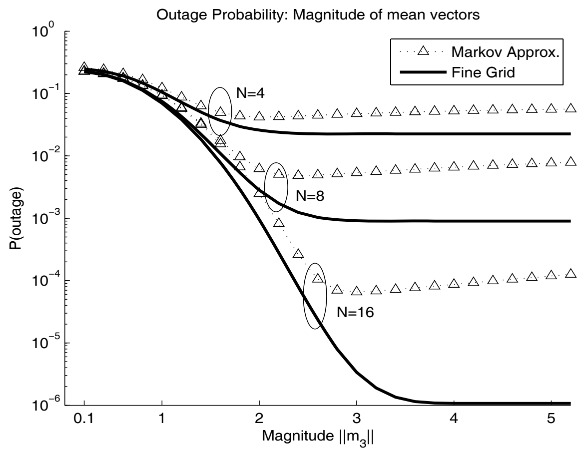

Fig. 6. Magnitude of mean vectors($J = 3$): Outage Probability as a function of the magnitude: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus $||\mathbf{m}_3||$.

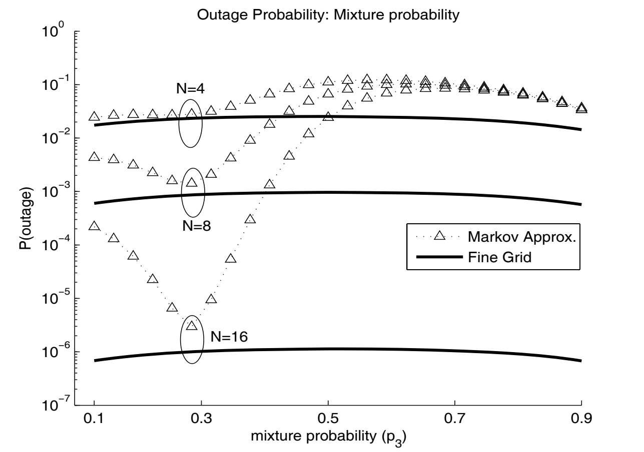

Fig. 8. Mixture Probability($J = 3$): Outage Probability as a function of the mixture probability: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus p_3.

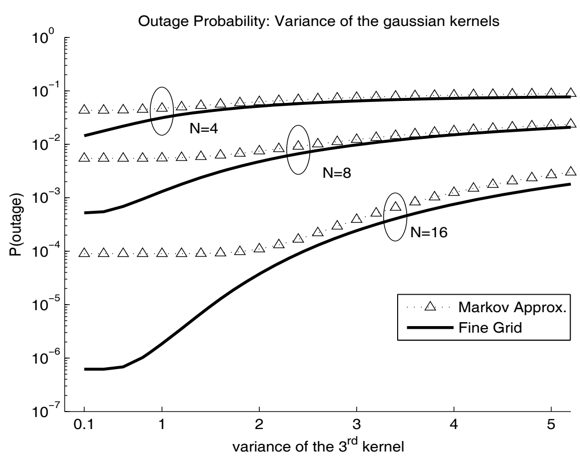

Fig. 7. Variance of the Gaussian Kernels($J = 3$): Outage Probability as a function of the variance: $\mathcal{M}(\mathbf{w}_{opt})$ and $\mathcal{M}(\mathbf{w}_{apr})$ versus σ_3^2.

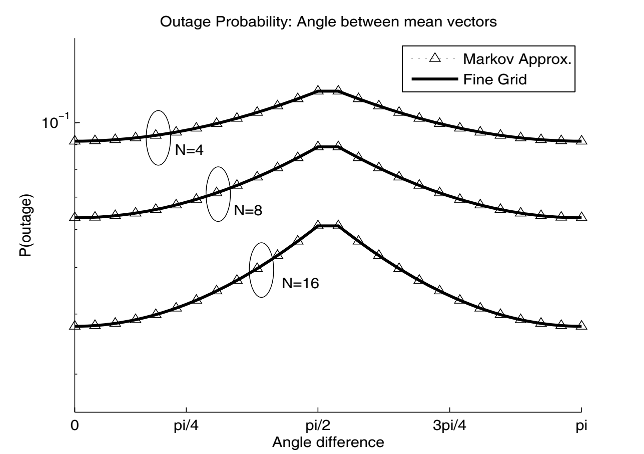

Fig. 9. Angle between mean vectors: Outage Probability as a function of the angle: $\mathcal{C}(\mathbf{w}_{opt})$ and $\mathcal{C}(\mathbf{w}_{apr})$ versus θ.

related problems has revealed that, in the case of perfect CSI, the approximation problem is easier in the complex case [9]. While the worst-case analysis in [9] is useful and sheds some light on this puzzle, it is not clear how to adapt it in the present context - the type of approximation employed in [9] is very different. Furthermore, it was also shown in [9] that worst-case analysis can be very pessimistic: average approximation performance can be orders of magnitude better than what the worst-case analysis predicts. This limits the engineering value of worst-case analysis. Still, it is interesting that our numerical findings corroborate those in [9].

*1) **Angle between mean vectors**:* Fig. 9 plots outage probability results for $p_1 = p_2 = 1/2$, $\sigma_1^2 = \sigma_2^2 = 1$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$, $||\mathbf{w}||^2 = P = 1$, as the angle between the mean vectors, θ, varies in $[0, \pi)$.

*2) **Magnitude of mean vectors**:* Fig. 10 plots outage probability results for $p_1 = p_2 = 1/2$, $\sigma_1^2 = \sigma_2^2 = 1$, $\theta = \pi/3$, $||\mathbf{m}_1||^2 = N$, $||\mathbf{w}||^2 = P = 1$, as $||\mathbf{m}_2||$ varies in $(0, 5]$.

*3) **Variance of the Gaussian Kernels**:* Fig. 11 plots outage probability results for $p_1 = p_2 = 1/2$, $\theta = \pi/3$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$, $\sigma_1^2 = 1$, $||\mathbf{w}||^2 = P = 1$, as σ_2^2 varies in $(0, 5]$.

*4) **Mixture Probability**:* Fig. 12 plots outage probability results for $\sigma_1^2 = \sigma_2^2 = 1$, $||\mathbf{m}_1||^2 = ||\mathbf{m}_2||^2 = N$ $\theta = \pi/3$, $||\mathbf{w}||^2 = P = 1$, as p_1 varies in $[0.1, 0.9]$.

V. CONCLUSIONS

The multicast beamforming problem was considered from the viewpoint of minimizing outage probability subject to a transmit power constraint. In a multicast context, the channel is naturally modeled as a Gaussian mixture, as opposed to a single Gaussian distribution. The different Gaussian kernels model user clusters of different means (locations) and variances (spreads). It was shown that minimizing outage probability subject to a transmit power constraint is an NP-hard problem when the number of Gaussian kernels, J, is greater than or equal to the number of transmit antennas, N. Through dimensionality reduction, it was also shown that the problem is practically tractable for $2-3$ Gaussian kernels. An approximate solution based on the Markov inequality was also proposed.

In the real case, the Markov approximation can be very accurate, but appears sensitive to near-far and mixture probability imbalance effects. For a large number of transmit

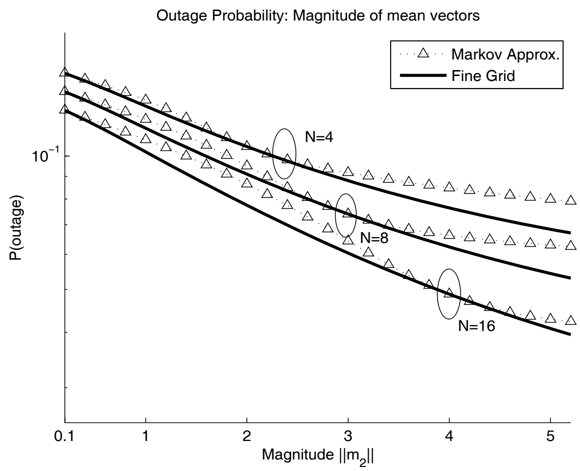

Fig. 10. Magnitude of mean vectors: Outage Probability as a function of the magnitude: $\mathcal{C}(\mathbf{w}_{opt})$ and $\mathcal{C}(\mathbf{w}_{apr})$ versus $||\mathbf{m}_2||$.

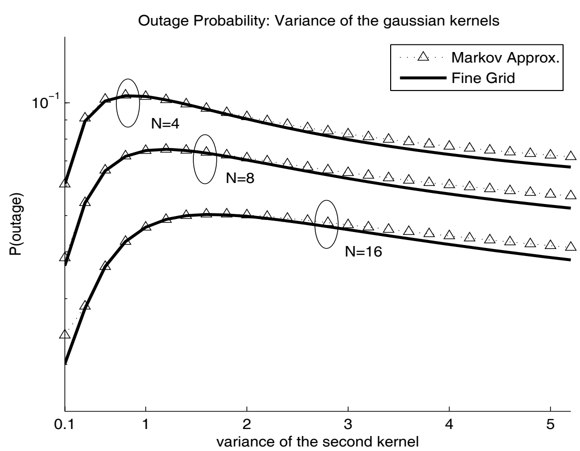

Fig. 11. Variance of the Gaussian Kernels: Outage Probability as a function of the variance: $\mathcal{C}(\mathbf{w}_{opt})$ and $\mathcal{C}(\mathbf{w}_{apr})$ versus σ_2^2.

antennas, N, the Markov approximation brakes down in the presence of such imbalances - the gap from the optimal solution is significant. The reason is that the principal right singular vector of the matrix $\mathbf{DM}$ then tends to align with the dominant component(s), effectively ignoring weaker ones.

Interestingly, the Markov approximation seems to be far more accurate in the complex case. This corroborates findings in [9], which showed that related approximation problems are easier in the complex case.

Acknowledgment: The authors would like to express their gratitude to an anonymous reviewer, whose insightful comments and refreshing sportsmanship not only helped improve the quality of this paper, but also made the review process interesting and enjoyable.

VI. Appendix: Proofs

A. Claim 1

Proof: Consider the special case where $\sigma_j = \sigma$, $p_j = 1/J$, $\forall j$. Using the Cauchy-Schwartz inequality, it can be shown (cf. [14]) that

$$|\mathbf{w}^T\mathbf{m}_j| - \epsilon||\mathbf{w}|| \leq |\mathbf{w}^T\mathbf{h}_j| \leq$$

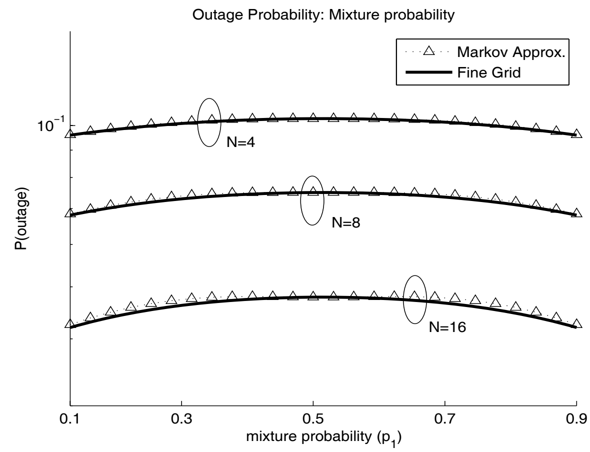

Fig. 12. Mixture Probability: Outage Probability as a function of the mixture probability: $\mathcal{C}(\mathbf{w}_{opt})$ and $\mathcal{C}(\mathbf{w}_{apr})$ versus p_1.

$$|\mathbf{w}^T\mathbf{m}_j| + \epsilon||\mathbf{w}||, \forall \mathbf{h}_j \in B_\epsilon(\mathbf{m}_j),$$

where $B_\epsilon(\mathbf{m}_j)$ denotes a ball of radius ϵ centered at $\mathbf{m}_j$. Let $\mathbf{h}_j$ be drawn from the j-th component pdf $\mathcal{N}(\mathbf{h}; \mathbf{m}_j, \sigma^2\mathbf{I})$. Given ϵ and $\delta > 0$, we can pick $\sigma = \sigma(\epsilon, \delta)$ such that $Pr[\mathbf{h}_j \in B_\epsilon(\mathbf{m}_j)] \geq 1 - \delta$. Let

$$p_{out}(\mathbf{w}) := \frac{1}{J}\sum_{j=1}^{J} p_{out|j}(\mathbf{w}),\ p_{out}^* := \min_{||\mathbf{w}||^2=P} p_{out}(\mathbf{w}),$$

where $p_{out|j}(\mathbf{w}) := Pr[|\mathbf{w}^T\mathbf{h}_j| < \gamma]$. With $1(\text{condition})$ denoting the indicator function, and $E_{\mathbf{h}_j}[\cdot]$ the expectation conditioned on the j-th component,

$$p_{out|j}(\mathbf{w}) = E_{\mathbf{h}_j}[1(|\mathbf{w}^T\mathbf{h}_j| < \gamma)].$$

For $\mathbf{h}_j \in B_\epsilon(\mathbf{m}_j)$, it holds

$$|\mathbf{w}^T\mathbf{m}_j| + \epsilon||\mathbf{w}|| < \gamma \Rightarrow |\mathbf{w}^T\mathbf{h}_j| < \gamma,$$

and therefore

$$1(|\mathbf{w}^T\mathbf{m}_j| < \gamma - \epsilon||\mathbf{w}||) \leq 1(|\mathbf{w}^T\mathbf{h}_j| < \gamma).$$

It follows that, for all $\mathbf{h}_j$,

$$1(|\mathbf{w}^T\mathbf{m}_j| < \gamma - \epsilon||\mathbf{w}||)1(\mathbf{h}_j \in B_\epsilon(\mathbf{m}_j)) \leq 1(|\mathbf{w}^T\mathbf{h}_j| < \gamma).$$

Taking $E_{\mathbf{h}_j}[\cdot]$ we obtain

$$(1-\delta)1(|\mathbf{w}^T\mathbf{m}_j| < \gamma - \epsilon||\mathbf{w}||) \leq p_{out|j}(\mathbf{w}).$$

In a similar way, for $\mathbf{h}_j \in B_\epsilon(\mathbf{m}_j)$, it holds

$$|\mathbf{w}^T\mathbf{h}_j| < \gamma \Rightarrow |\mathbf{w}^T\mathbf{m}_j| - \epsilon||\mathbf{w}|| < \gamma,$$

and therefore

$$1(|\mathbf{w}^T\mathbf{h}_j| < \gamma) \leq 1(|\mathbf{w}^T\mathbf{m}_j| < \gamma + \epsilon||\mathbf{w}||).$$

It follows that, for all $\mathbf{h}_j$,

$$1(|\mathbf{w}^T\mathbf{h}_j| < \gamma) \leq 1(|\mathbf{w}^T\mathbf{m}_j| < \gamma + \epsilon||\mathbf{w}||)1(\mathbf{h}_j \in B_\epsilon(\mathbf{m}_j)) + 1 - 1(\mathbf{h}_j \in B_\epsilon(\mathbf{m}_j)),$$

where the last term is a trivial upper bound that applies to the complement of $B_\epsilon(\mathbf{m}_j)$. Again taking $E_{\mathbf{h}_j}[\cdot]$ we obtain

$$p_{out|j}(\mathbf{w}) \leq (1-\delta)1(|\mathbf{w}^T\mathbf{m}_j| < \gamma + \epsilon||\mathbf{w}||) + \delta.$$

Combining the two inequalities, we have

$$(1-\delta)1(|\mathbf{w}^T\mathbf{m}_j| < \gamma - \epsilon||\mathbf{w}||) \leq p_{out|j}(\mathbf{w}) \leq$$
$$(1-\delta)1(|\mathbf{w}^T\mathbf{m}_j| < \gamma + \epsilon||\mathbf{w}||) + \delta,$$

Averaging out over j and taking the minimum over $\mathbf{w}$ with $||\mathbf{w}|| = \sqrt{P}$ yields

$$\frac{1-\delta}{J} \min_{||\mathbf{w}||=\sqrt{P}} \sum_{j=1}^{J} 1(|\mathbf{w}^T\mathbf{m}_j| < \gamma - \epsilon\sqrt{P}) \leq p^*_{out}(\gamma) \leq$$
$$\frac{1-\delta}{J} \min_{||\mathbf{w}||=\sqrt{P}} \sum_{j=1}^{J} 1(|\mathbf{w}^T\mathbf{m}_j| < \gamma + \epsilon\sqrt{P}) + \delta,$$

where we have also made explicit that p^*_{out} depends on γ. It follows that

$$p^*_{out}(t-\epsilon\sqrt{P}) - \delta \leq \frac{1-\delta}{J} \min_{||\mathbf{w}||=\sqrt{P}} \sum_{j=1}^{J} 1(|\mathbf{w}^T\mathbf{m}_j| < t) \leq$$
$$p^*_{out}(t+\epsilon\sqrt{P}),$$

for all $t \in (\epsilon\sqrt{P}, 1-\epsilon\sqrt{P})$. Notice now that $p^*_{out}(\cdot)$ is a continuous function, whereas

$$\frac{1-\delta}{J} \min_{||\mathbf{w}||=\sqrt{P}} \sum_{j=1}^{J} 1(|\mathbf{w}^T\mathbf{m}_j| < t)$$

only takes discrete values, separated by $\frac{1-\delta}{J}$. Recall that $\epsilon > 0$, $\delta > 0$, but otherwise up to our control. Pick $0 < \delta < \frac{1}{J+1}$ (which implies $\delta < \frac{1-\delta}{J}$) and ϵ sufficiently small to sandwich

$$\frac{1-\delta}{J} \min_{||\mathbf{w}||=\sqrt{P}} \sum_{j=1}^{J} 1(|\mathbf{w}^T\mathbf{m}_j| < t)$$

within an interval strictly less than $\frac{1-\delta}{J}$. This leaves no ambiguity - computing $p^*_{out}(t \pm \epsilon\sqrt{P})$ pin-points the exact value of $\min_{||\mathbf{w}||=\sqrt{P}} \sum_{j=1}^{J} 1(|\mathbf{w}^T\mathbf{m}_j| < t)$. In particular, this answers the question of whether or not it is possible to find a $\mathbf{w}$ of norm $\sqrt{P}$ such that $|\mathbf{w}^T\mathbf{m}_j| \geq t$, $\forall j \in \{1, \cdots, J\}$. The latter is the decidability version of a problem shown to be NP-hard in [12] for $J \geq N$. ■

B. Claim 2

Proof: Problem Q can be equivalently written in two stages. By conditioning on the mean $\mathbf{w}^H\mathbf{m}$, minimizing outage is equivalent to maximizing variance. Due to circular symmetry, only $|\mathbf{w}^H\mathbf{m}|$ matters, and problem Q is equivalent to

$$Q': \quad \min_{0 \leq c \leq P||\mathbf{m}||^2} \mathcal{C}\begin{pmatrix} arg\max \mathbf{w}^H\mathbf{C}\mathbf{w} \\ s.t.: ||\mathbf{w}||^2 = P \\ |\mathbf{w}^H\mathbf{m}|^2 = c \end{pmatrix}.$$

Let

$$\mathbf{w}_o(c) := arg\max \mathbf{w}^H\mathbf{C}\mathbf{w} \quad s.t.: ||\mathbf{w}||^2 = P, \quad |\mathbf{w}^H\mathbf{m}|^2 = c$$

Due to phase invariance, it suffices to consider the following solution parametrization

$$\mathbf{w} = \sqrt{c}\frac{\mathbf{m}}{||\mathbf{m}||^2} + \mathbf{v},$$

with $\mathbf{v}^H\mathbf{m} = 0$ and $||\mathbf{v}||^2 = P - c/||\mathbf{m}||^2$. The maximization problem can now be reformulated in terms of $\mathbf{v}$

$$\mathbf{v}_o(c) := arg\max \mathbf{v}^H\mathbf{C}\mathbf{v} \quad s.t.: ||\mathbf{v}||^2 = P - c/||\mathbf{m}||^2, \quad \mathbf{v}^H\mathbf{m} = 0.$$

Introduce the eigen-decomposition $\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^H$, and decompose $\mathbf{U} = \mathbf{U}_m + \mathbf{U}_\perp$, where $\mathbf{U}_m := \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\mathbf{U}$ and $\mathbf{U}_\perp := \mathbf{U} - \mathbf{U}_m = \left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right)\mathbf{U}$ are parallel and perpendicular to $\mathbf{m}$, respectively. Under $\mathbf{v}^H\mathbf{m} = 0$, $\mathbf{v}^H\mathbf{C}\mathbf{v} = \mathbf{v}^H\mathbf{U}_\perp\mathbf{D}\mathbf{U}_\perp^H\mathbf{v}$; but $\mathbf{U}_\perp^H$ annihilates components in the direction of $\mathbf{m}$, and thus

$$\mathbf{v}_o(c) := arg\max \mathbf{v}^H\mathbf{U}_\perp\mathbf{D}\mathbf{U}_\perp^H\mathbf{v} \quad s.t.: ||\mathbf{v}||^2 = P - c/||\mathbf{m}||^2.$$

It follows that $\mathbf{v}_o(c)$ is the principal component of $\mathbf{U}_\perp\mathbf{D}\mathbf{U}_\perp^H = \left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right)\mathbf{U}\mathbf{D}\mathbf{U}^H\left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right) = \left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right)\mathbf{C}\left(\mathbf{I} - \frac{\mathbf{m}\mathbf{m}^H}{||\mathbf{m}||^2}\right)$, scaled to squared norm $P - c/||\mathbf{m}||^2$. Let $\mathbf{p}$ denote the said principal component, scaled to unit norm. The optimal beamforming vector can be found by one-dimensional line search over c:

$$\min_{0 \leq c \leq P||\mathbf{m}||^2} \mathcal{C}\left(\sqrt{c}\frac{\mathbf{m}}{||\mathbf{m}||^2} + \sqrt{P - c/||\mathbf{m}||^2}\mathbf{p}\right).$$

■

REFERENCES

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions and Formulas, Graphs, and Mathematical Tables*, chapter 26.3.24. Government Printing Office, Washington, D.C., 1972.

[2] B. K. Chalise and A. Czylwik, "Robust uplink beamforming based upon minimum outage probability criterion," in *Proc. Global Telecommun. Conf. (GLOBECOM)* Nov. 2004, vol. 6, pp. 3974-3978.

[3] B. K. Chalise, S. Shahbazpanahi, A. Czylwik, and A. B. Gershman, "Robust downlink beamforming based on outage probability specifications," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3498-3503, Oct. 2007.

[4] N. Jindal and Z.-Q. Luo, "Capacity limits of multiple antenna multicast," in *Proc. IEEE International Symp. Inform. Theory* (ISIT 2006), pp. 1841-1845, July 2006.

[5] S. Kandikuri and S. Boyd, "Optimal power control in interference limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46-55, Jan. 2002.

[6] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Convex transmit beamforming for downlink multicasting to multiple co-channel groups," in *Proc. IEEE ICASSP 2006*, May 2006, Toulouse, France.

[7] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min-fair transmit beamforming to multiple co-channel multicast groups," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1268-1279, Mar. 2008.

[8] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Processing*, vol. 55, no. 10, pp. 4916-4927, Oct. 2007.

[9] Z.-Q. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang, "Approximation bounds for quadratic optimization with homogeneous quadratic constraints," *SIAM J. Optimization*, vol. 18, no. 1, pp. 1-28, Feb. 2007.

[10] E. Matskani, N. D. Sidiropoulos, and L. Tassiulas, "On multicast beamforming and admission control for UMTS-LTE," in *Proc. ICASSP 2008*, Mar. 2008, Las Vegas, NV.

[11] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receivers for multiaccess space-time block-coded MIMO systems: a probabilistically constrained approach," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1560-1570, Aug. 2006.

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

[13] Y. Xie, C. N. Georghiades, and A. Arapostathis, "Minimum outage probability transmission with imperfect feedback for MISO fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1084-1091, May 2005.

[14] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313-324, Feb. 2003.

[15] S. A. Vorobyov, Y. Rong, and A. B. Gershman, "Robust adaptive beamforming using probability-constrained optimization," in *Proc. IEEE Statistical Signal Processing Workshop*, July 2005, pp. 934-939.

[16] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformmers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Processing*, vol. 56, pp. 5719-5724, Nov. 2008.

[17] S. A. Vorobyov, "Robust CDMA multiuser detectors: probability-constrained versus the worst-case based design," *IEEE Signal Processing Lett.*, vol. 15, pp. 273-276, 2008.

[18] R. Wang, H. Li, and T. Li, "Robust multiuser detection for multicarrier CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 673-683, Mar. 2006.

Vassilis Ntranos received the Diploma in Electronic and Computer Engineering (ECE) from the Technical University of Crete (TUC), Greece, in 2008. His work in this paper was part of his undergraduate Diploma thesis. He is currently a graduate student in the Department of ECE at TUC. His research interests are in convex approximation theory and algorithms, with current emphasis on fascinating applications in social network analysis.

Nicholas D. Sidiropoulos (F'09) received the Diploma degree from the Aristotelian University of Thessaloniki, Greece, and M.Sc. and Ph.D. degrees from the University of Maryland at College Park (UMCP), in 1988, 1990, and 1992, respectively, all in Electrical Engineering. He has been a Postdoctoral Fellow (1994-1995) and Research Scientist (1996-1997) at the Institute for Systems Research, UMCP, and has held positions as Assistant Professor, Department of Electrical Engineering, University of Virginia-Charlottesville (1997-1999), and Associate Professor, Department of Electrical and Computer Engineering, University of Minnesota - Minneapolis (2000-2002). Since 2002, he is a Professor in the Department of Electronic and Computer Engineering at the Technical University of Crete, Chania-Crete, Greece, and Adjunct Professor at the University of Minnesota. His current research interests are primarily in signal processing for communications, convex approximation of NP-hard problems, cross-layer resource allocation for wireless networks, and multiway analysis. Prof. Sidiropoulos received the U.S. NSF CAREER Award (1998), the IEEE Signal Processing Society Best Paper Award (in 2001 and in 2007), and is currently a Distinguished Lecturer of the IEEE SP Society (2008-2009). He has served as member (2000-2004), vice-chair (2005-2006), and chair (2007-2008) of the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC), and as member of the Sensor Array and Multichannel Technical Committee (SAM-TC) of the IEEE SP Society (2004-2009). He has served as Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING (2000-2006) and IEEE SIGNAL PROCESSING LETTERS (2000-2002), and currently serves on the editorial board of IEEE SIGNAL PROCESSING MAGAZINE.

Leandros Tassiulas (F'07) was born in 1965 in Katerini, Greece. He received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1989 and 1991, respectively. He is a Professor in the Department of Computer and Telecommunications Engineering, University of Thessaly, Greece, and a Research Professor in the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, since 2001. He has held positions as Assistant Professor at Polytechnic University New York (1991-1995), Assistant and Associate Professor at University of Maryland, College Park (1995-2001), and Professor at University of Ioannina, Greece (1999-2001). His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, sensor networks, high-speed internet and satellite communications. Dr. Tassiulas received a National Science Foundation (NSF) Research Initiation Award in 1992, an NSF CAREER Award in 1995, an Office of Naval Research (ONR) Young Investigator Award in 1997, a Bodosaki Foundation Award in 1999, the IEEE INFOCOM '94 Best Paper Award, and the inaugural INFOCOM Achievement award (2007) for Fundamental contributions to resource allocation in communication networks.