Channel Capacity and Beamforming for Multiple Transmit and Receive Antennas with Covariance Feedback

Syed Ali Jafar, Sriram Vishwanath, Andrea Goldsmith
Dept. of Electrical Engineering
Stanford University, USA
E-mail: {syed,sriram,andrea}@systems.stanford.edu

Abstract—**We consider the capacity of a narrowband point to point communication system employing multiple-element antenna arrays at both the transmitter and the receiver with covariance feedback. Under covariance feedback the receiver is assumed to have perfect Channel State Information (CSI) while at the transmitter the channel matrix is modeled as consisting of zero mean complex jointly Gaussian random variables with known covariances. Specifically we assume a channel matrix with i.i.d. rows and correlated columns, a common model for downlink transmission. We determine the optimal transmit precoding strategy to maximize the Shannon capacity of such a system. We also derive closed form necessary and sufficient conditions on the spatial covariance for when the maximum capacity is achieved by beamforming. The conditions for optimality of beamforming agree with the notion of waterfilling over multiple degrees of freedom.**

I. INTRODUCTION

The Shannon capacity of systems with MEA (Multiple Element Antenna) arrays at the transmitter and/or the receiver has attracted much research activity recently. With perfect channel state information at both the transmitter and the receiver the (n_T, n_R) MEA array (n_T transmit and n_R receive antennas) was shown in [1] to consist of $n = \min(n_T, n_R)$ Single Input Single Output (SISO) non-interfering channels or eigenmodes. The optimal transmit strategy in this case takes the form of a waterfill over these modes. When the transmitter does not know the instantaneous channel state perfectly, it is not possible to transform the Multiple Input Multiple Output (MIMO) channel into parallel non-interfering SISO channels and the required vector coding across the antenna array significantly increases the complexity of the system.

Recent work in [2], [3] and [4] on multiple antenna channel capacity with partial Channel State Information at the Transmitter (CSIT) has produced some interesting results. It was found that unlike single antenna systems where exploiting CSIT does not significantly enhance the Shannon capacity, for multiple antenna systems the capacity improvement through even partial CSIT can be substantial. Moreover, the numerical results in [2] and [3] show that in some scenarios a beamforming transmission strategy achieves close to channel capacity. This is interesting since beamforming corresponds to scalar coding with linear preprocessing at the transmit antenna array and the complexity involved is only a fraction of the vector coding complexity for typical array sizes. For *mean feedback*, Narula and Trott [2] point out that there are cases where the capacity is actually achieved via beamforming. While they do not obtain fully general necessary and sufficient conditions for when beamforming is a capacity achieving strategy, they develop partial answers to the problem for two transmit antennas. The capacity optimization problem itself was recently solved by Visotsky and Madhow in [3] for the cases of mean feedback and covariance feedback. Their numerical results indicate that beamforming is close to the optimal strategy when the quality of feedback improves (mean feedback) or a stronger path is present (covariance feedback). The general necessary and sufficient conditions for optimality of beamforming with mean or covariance feedback are given in [4].

Note that the solution to the capacity optimization problem and the necessary and sufficient conditions for optimality of beamforming in [2], [3] and [4] all assume a single receive antenna with multiple transmit antennas. The solutions for the case of multiple antennas at both the transmitter and the receiver have not been obtained so far and form the main thrust of this paper. Specifically, in this work we extend the capacity optimization problem model in [3] to multiple receive antennas for the case of covariance feedback and develop necessary and sufficient conditions for optimality of beamforming.

The organization of this paper is as follows. The next section describes the system model and introduces our notation. The problem statement is presented in Section III. Section IV contains the solution to the capacity optimization problem. Necessary and sufficient conditions for optimality of beamforming are derived in Section V. We conclude with a summary and a waterfilling interpretation of the results in Section VI.

II. SYSTEM MODEL

We use the following notation: $x \sim \tilde{N}(\mu, \sigma^2)$ implies that x is a complex circularly symmetric Gaussian with mean μ and variance σ^2. $Z_{i\cdot}$ and $Z_{\cdot j}$ represent the i^{th} row and the j^{th} column of a matrix Z respectively. E$[x]$ denotes the expectation of random variable x. Lastly, *s.t.* is short for *subject to*.

We focus on a point-to-point communication system using n_T transmit and n_R receive antennas over a narrowband flat fading channel. The channel matrix is represented as $H = [h_{ij}]_{n_R \times n_T}$ where $h_{ij} \sim \tilde{N}(0, \sigma_{ij}^2)$ represents the channel gain from transmit antenna j to receive antenna i (Rayleigh Fading). It is assumed that the channel coefficients are available to the receiver while the transmitter knows only the correlations $\rho_{ijkl} = \text{E}[h_{ij} h_{kl}^*]$.

The n_T dimensional vector symbol $\overline{x}(\tau) = [x_1(\tau), x_2(\tau), \cdots, x_{n_T}(\tau)]^T$ is transmitted at time instant τ to yield the n_R dimensional received vector $\overline{r}(\tau)$ as

$$\overline{r}(\tau) = H(\tau)\overline{x}(\tau) + \overline{n}(\tau). \quad (1)$$

The n_R components of AWGN $\overline{n}(\tau)$ are assumed to be i.i.d. $\tilde{N}(0, \sigma^2)$ and uncorrelated with the signals. For convenience the time index τ will be dropped in this paper.

Recall that in general for partial information $\mathcal{U}$ at the transmitter, codes need to be defined over an extended alphabet of

0-7803-7097-1/01/$10.00 ©2001 IEEE

functions $\mathcal{U} \rightarrow \mathcal{X}$ where $\mathcal{X}$ is the input alphabet. However when the CSIT is a deterministic function of the CSIR optimal codes can be constructed directly over the input alphabet $\mathcal{X}$ [6]. For our case since the receiver knows the channel perfectly, the spatial correlations available to the transmitter are also available to the receiver. Thus the CSIT is a deterministic function of the CSIR and the capacity is easily shown to be

$$C = \max_{Q\,:\,\text{trace}(Q)=P} C(Q), \tag{2}$$

where

$$C(Q) \triangleq \mathrm{E}\left[\log\left|I_{n_R} + \frac{HQH^\dagger}{\sigma^2}\right|\right] \tag{3}$$

is the capacity with the input covariance matrix $\mathrm{E}[\overline{x}\overline{x}^\dagger] = Q$. The capacity $C(Q)$ is achieved by transmitting independent complex circular Gaussian symbols along the eigenvectors of Q. The powers allocated to each eigenvector are given by the eigenvalues of Q. Thus the capacity optimization problem involves finding the optimum Q to maximize $C(Q)$ subject to the transmit power constraint trace(Q)= P. Consistent with [3] and [7], we define beamforming as a transmission strategy where the input covariance matrix Q has rank one. Beamforming capacity therefore refers to the maximum capacity $C(Q)$ subject to $\text{trace}(Q) = P$ and $\text{rank}(Q) = 1$.

A key assumption for this paper is that the entries in a given row of H are correlated while those belonging to different rows of H are uncorrelated. More specifically we assume that the rows of H are i.i.d. while the columns are correlated. This is typical of the channel correlations obtained using the 'one-ring' model employed by Shiu *et. al.* in [5]. The model is a ray tracing model appropriate for a scenario where the base station (BS) is unobstructed and the subscriber unit (SU) is surrounded by local scatterers. A detailed description of the model is presented in [5]. For our purpose it suffices to point out the following two features of the correlations obtained using the one ring model:

1. $\mathbf{E}[h_{ik}h_{il}^*] = \mathbf{E}[h_{jk}h_{jl}^*]$ for $1 \le i,j \le n_R, 1 \le k,l \le n_T$. So the rows of H are identically distributed.
2. The scatterers surrounding the SU impose random phase shifts onto the waves incident upon them, decorrelating the fades associated with any two distinct antennas at the SU. Thus, for a wavelength λ and an angle spread Δ (typical values range from 0.6 degrees to 15 degrees), the minimum decorrelating antenna spacing at the receiver is just 0.38λ, while at the transmitter it is $0.38\lambda\Delta^{-1}$ for a broadside transmitting antenna array and $1.53\lambda\Delta^{-2}$ for an inline transmitting antenna array. For GHz frequency operation, these correspond to antenna spacings on the order of centimeters for receive antenna array, meters for broadside transmit antennas and hundreds of meters for inline transmit antennas. So practical systems demonstrate strong correlation between fades associated with different transmit antennas on the downlink while the fades associated with different receive antennas are fairly uncorrelated. This justifies our assumption that the rows of H are uncorrelated. Since they are jointly Gaussian, they are also independent.

III. Problem Definition

We address the following problems:

Problem 1 : Characterize the optimal input covariance matrix Q that maximizes the Shannon capacity of our system given by Equation (2) where the elements of H are all zero mean complex Gaussians and H has i.i.d. rows and correlated columns. The distribution of the i^{th} row of H is given by $H_{i.} \sim \tilde{N}(\mathbf{0}, \Sigma)$ $\forall i \in \{1, 2, \cdots, n_R\}$.

Problem 2 : Find necessary and sufficient conditions on Σ for beamforming to be the optimal transmission strategy, i.e. for the optimal input covariance matrix Q to have rank one.

IV. Solution to Problem 1

With just one receive antenna, i.e. when H is just a row vector distributed as $H \sim \tilde{N}(\mathbf{0}, \Sigma)$, the optimal solution was shown in [3] to consist of independent Gaussian inputs along the eigenvectors of Σ. We show that the same is true for our case even with multiple receive antennas.

Let the eigendecomposition of Σ be given as $\Sigma = U_\Sigma \Lambda_\Sigma U_\Sigma^\dagger$ where U_Σ is a unitary matrix and Λ_Σ is a diagonal matrix containing the eigenvalues of Σ arranged in decreasing order. We assume that Σ has full rank so that $\lambda_{11}^\Sigma \ge \lambda_{22}^\Sigma \ge \cdots \ge \lambda_{n_T n_T}^\Sigma > 0$. Our goal is to show that the optimal input covariance matrix has a spectral decomposition $Q = U_\Sigma \Lambda_Q U_\Sigma^\dagger$. Equivalently we wish to show that for the optimal Q the matrix $U_\Sigma^\dagger Q U_\Sigma$ is diagonal. Note that the optimum Q may not have full rank, as is typical of a water filling solution.

Define the matrix

$$Z \triangleq H U_\Sigma \Lambda_\Sigma^{-\frac{1}{2}}. \tag{4}$$

So the i^{th} row of Z is given as $Z_{i.} = H_{i.} U_\Sigma \Lambda_\Sigma^{-\frac{1}{2}}$. Since the rows of H are zero mean and i.i.d., so are the rows of Z. Also, since $U_\Sigma \Lambda_\Sigma^{-\frac{1}{2}}$ is the whitening filter for the random vector $H_{i.}$, the covariance matrix of $Z_{i.}$ is given by the identity matrix. Thus Z consists of i.i.d. zero mean and unit variance complex Gaussian elements. Substituting back $H = Z\Lambda_\Sigma^{\frac{1}{2}} U_\Sigma^\dagger$ in the capacity expression (2) we get

$$C = \max_Q \mathrm{E}\left[\log\left|I_{n_R} + \frac{Z\Lambda_\Sigma^{\frac{1}{2}} U_\Sigma^\dagger Q U_\Sigma \Lambda_\Sigma^{\frac{1}{2}} Z^\dagger}{\sigma^2}\right|\right] \quad \text{s. t. trace}(Q) = P. \tag{5}$$

We define $\hat{Q} \triangleq \Lambda_\Sigma^{\frac{1}{2}} U_\Sigma^\dagger Q U_\Sigma \Lambda_\Sigma^{\frac{1}{2}}$. Note that since U_Σ is unitary the non-negative definite matrices Q and $U_\Sigma^\dagger Q U_\Sigma = \Lambda_\Sigma^{-\frac{1}{2}} \hat{Q} \Lambda_\Sigma^{-\frac{1}{2}}$ have the same set of eigenvalues. Therefore we have

$$\text{trace}(Q) = \text{trace}(U_\Sigma^\dagger Q U_\Sigma) = \text{trace}(\Lambda_\Sigma^{-\frac{1}{2}} \hat{Q} \Lambda_\Sigma^{-\frac{1}{2}}), \tag{6}$$

and the capacity expression (5) can be rewritten as

$$C = \max_{\hat{Q}\,:\,\text{trace}(\Lambda_\Sigma^{-\frac{1}{2}} \hat{Q} \Lambda_\Sigma^{-\frac{1}{2}})=P} \mathrm{E}\left[\log\left|I_{n_R} + \frac{Z\hat{Q}Z^\dagger}{\sigma^2}\right|\right], \tag{7}$$

where the rows of Z are independent and identically distributed as $Z_{i.} \sim \tilde{N}(\mathbf{0}, I)$. As stated earlier, our aim is to show that

0-7803-7097-1/01/$10.00 ©2001 IEEE

this optimal $U_\Sigma^\dagger Q U_\Sigma$ is a diagonal matrix. Now, since $\Lambda_\Sigma^{-\frac{1}{2}}$ is a diagonal matrix, this is equivalent to showing that the optimal $\hat{Q} = \Lambda_\Sigma^{\frac{1}{2}} U_\Sigma^\dagger Q U_\Sigma \Lambda_\Sigma^{\frac{1}{2}}$ is a diagonal matrix. Next we prove that the optimal $\hat{Q}$ that maximizes the capacity in (7) is indeed diagonal.

Let the optimal $\hat{Q}$ have a spectral decomposition $\hat{Q} = \hat{U}\hat{\Lambda}\hat{U}^\dagger$ where as before $\hat{U}$ is a unitary matrix and $\hat{\Lambda}$ is a diagonal matrix of eigenvalues arranged in decreasing order. Since each element of Z is i.i.d. and zero mean and $\hat{U}$ is unitary it is easy to see that $Z\hat{U} \sim Z$, i.e. $Z\hat{U}$ and Z are identically distributed. This implies that

$$\mathrm{E}\left[\log\left|I_{n_R} + \frac{Z\hat{Q}Z^\dagger}{\sigma^2}\right|\right] = \mathrm{E}\left[\log\left|I_{n_R} + \frac{Z\hat{\Lambda}Z^\dagger}{\sigma^2}\right|\right]$$

So the diagonal matrix $\hat{\Lambda}$ achieves the same capacity as the optimal $\hat{Q}$. However note that $\hat{Q}$ needs to satisfy the additional constraint given by

$$\mathrm{trace}(\Lambda_\Sigma^{-\frac{1}{2}}\hat{Q}\Lambda_\Sigma^{-\frac{1}{2}}) = P. \tag{8}$$

To complete the proof we need to show that if $\hat{Q}$ satisfies the trace constraint, then so does $\hat{\Lambda}$. Now

$$\mathrm{trace}(\Lambda_\Sigma^{-\frac{1}{2}}\hat{Q}\Lambda_\Sigma^{-\frac{1}{2}}) = \sum_{i=1}^{n_T} \frac{\hat{q}_{ii}}{\lambda_{ii}^\Sigma}, \tag{9}$$

$$\text{and } \mathrm{trace}(\Lambda_\Sigma^{-\frac{1}{2}}\hat{\Lambda}\Lambda_\Sigma^{-\frac{1}{2}}) = \sum_{i=1}^{n_T} \frac{\hat{\lambda}_{ii}}{\lambda_{ii}^\Sigma}, \tag{10}$$

where $\hat{q}_{ii}$, $\hat{\lambda}_{ii}$ and λ_{ii}^Σ are the diagonal elements of $\hat{Q}$, $\hat{\Lambda}$, and Λ^Σ respectively.

We need the following theorems in order statistics:

Theorem 1: For a Hermitian matrix A the vector of diagonal entries $\{a_{ii}\}$ *majorizes* the vector of eigenvalues $\{\lambda_{ii}^A\}$.

Proof: This is Theorem 4.3.26 in [8]. ∎

Recall that a real vector $\alpha = [\alpha_i] \in \mathbb{R}^n$ majorizes another real vector $\beta = [\beta_i] \in \mathbb{R}^n$ if and only if the sum of the k smallest entries of α is greater than or equal to the sum of the k smallest entries of β for $k = 1, 2, \cdots, n-1$ and the sums of the entries of α and β are equal. This is a mathematical way to capture the vague notion that the components of a vector α are "less spread out" or "more nearly equal" than are the components of a vector β. Majorization is the precise relationship between the eigenvalues and the diagonal entries of a Hermitian matrix. That is, for any real vector α that majorizes another real vector β there exists a Hermitian matrix with its main diagonal given by α and its eigenvalues given by β.

Theorem 2: For any two given positive real vectors $\alpha = [\alpha_i], \beta = [\beta_i] \in \mathbb{R}_+^n$ the permutation π^* that minimizes the sum $\sum_{i=1}^n \frac{\alpha_{\pi(i)}}{\beta_i}$ is such that $\alpha_{\pi^*(i)}$ and β_i are in the same order. That is, $\forall i, j \in \{1, 2, \cdots, n\}$, if $\alpha_{\pi^*(i)} < \alpha_{\pi^*(j)}$, then $\beta_i < \beta_j$.

Proof: Let the minimizing permutation be such that for some $i, j \in \{1, 2, \cdots, n\}$, $\alpha_{\pi^*(i)} < \alpha_{\pi^*(j)}$ and $\beta_i > \beta_j$. We show that this leads to a contradiction. Specifically, let

$$\alpha_{\pi^*(i)}\left(\frac{1}{\beta_j} - \frac{1}{\beta_i}\right) < \alpha_{\pi^*(j)}\left(\frac{1}{\beta_j} - \frac{1}{\beta_i}\right)$$

$$\Rightarrow \quad \frac{\alpha_{\pi^*(i)}}{\beta_j} + \frac{\alpha_{\pi^*(j)}}{\beta_i} < \frac{\alpha_{\pi^*(i)}}{\beta_i} + \frac{\alpha_{\pi^*(j)}}{\beta_j} \tag{11}$$

Thus the sum $\sum_{i=1}^n \frac{\alpha_{\pi(i)}}{\beta_i}$ can be reduced further by switching the values of $\pi^*(i)$ and $\pi^*(j)$. Thus π^* can not be the minimizing permutation and we have a contradiction. ∎

Theorem 3: If $\alpha = [\alpha_i], \beta = [\beta_i]$ and $\gamma = [\gamma_i] \in \mathbb{R}_+^n$ are three vectors with components arranged in descending order, i.e. if $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$, $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$ and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n$, and if α majorizes β then the following is true:

$$\sum_{i=1}^n \frac{\alpha_i}{\gamma_i} \geq \sum_{i=1}^n \frac{\beta_i}{\gamma_i} \tag{12}$$

Proof: We outline the first two steps here. The rest of the proof follows in an iterative fashion.

$$\alpha_n \geq \beta_n \Rightarrow \frac{\alpha_n}{\gamma_n} \geq \frac{\beta_n}{\gamma_n} \tag{13}$$

$$\alpha_n + \alpha_{n-1} \geq \beta_n + \beta_{n-1} \tag{14}$$

$$\Rightarrow \quad \frac{1}{\gamma_{n-1}}(\alpha_n + \alpha_{n-1}) + \alpha_n\left(\frac{1}{\gamma_n} - \frac{1}{\gamma_{n-1}}\right) \geq \frac{1}{\gamma_{n-1}}(\beta_n + \beta_{n-1}) + \beta_n\left(\frac{1}{\gamma_n} - \frac{1}{\gamma_{n-1}}\right) \tag{15}$$

$$\Rightarrow \quad \frac{\alpha_n}{\gamma_n} + \frac{\alpha_{n-1}}{\gamma_{n-1}} \geq \frac{\beta_n}{\gamma_n} + \frac{\beta_{n-1}}{\gamma_{n-1}} \tag{16}$$

Proceeding in this fashion we arrive at the desired result. Note that (13) and (14) make use of the fact that α majorizes β. ∎

Now we use these theorems to solve our original problem. Recall that we need to show that the trace of $\Lambda_\Sigma^{-\frac{1}{2}}\hat{\Lambda}\Lambda_\Sigma^{-\frac{1}{2}}$ is never greater than the trace of $\Lambda_\Sigma^{-\frac{1}{2}}\hat{Q}\Lambda_\Sigma^{-\frac{1}{2}}$. Specifically we need to prove the following:

$$\sum_{i=1}^{n_T} \frac{\hat{q}_{ii}}{\lambda_{ii}^\Sigma} \geq \sum_{i=1}^{n_T} \frac{\hat{\lambda}_{ii}}{\lambda_{ii}^\Sigma} \tag{17}$$

where λ_{ii}^Σ and $\hat{\lambda}_{ii}$ are arranged in descending order. Let π be a permutation on the indices so that $\hat{q}_{\pi(1)} \geq \hat{q}_{\pi(2)} \geq \cdots \geq \hat{q}_{\pi(n)}$. So then it follows from Theorem 2 that

$$\sum_{i=1}^{n_T} \frac{\hat{q}_{ii}}{\lambda_{ii}^\Sigma} \geq \sum_{i=1}^{n_T} \frac{\hat{q}_{\pi(i)}}{\lambda_{ii}^\Sigma}. \tag{18}$$

Using Theorem 1 we have that the vector $\{\hat{q}_{\pi(i)}\}$ majorizes the vector $\hat{\lambda}_{ii}$. Combining this with the result from Theorem 3 we get

$$\sum_{i=1}^{n_T} \frac{\hat{q}_{ii}}{\lambda_{ii}^\Sigma} \geq \sum_{i=1}^{n_T} \frac{\hat{q}_{\pi(i)}}{\lambda_{ii}^\Sigma} \geq \sum_{i=1}^{n_T} \frac{\hat{\lambda}_{ii}}{\lambda_{ii}^\Sigma}. \tag{19}$$

This completes the proof and we conclude that the optimal input covariance matrix that maximizes the capacity in (2) has the same eigenvectors as the covariance matrix Σ of each row of the channel matrix. Therefore the optimal transmit strategy is to transmit independent complex Gaussians along the eigenvectors of Σ. The powers allocated to each of the eigenvectors are given by the diagonal matrix Λ_Q and need to be determined through numerical optimization. The solution resembles waterfilling in the sense that Λ_Q and Λ_Σ are both arranged in descending order, i.e., stronger channel modes get allocated more power than weaker modes.

V. Solution to Problem 2

Now we turn our attention to deriving the necessary and sufficient conditions for optimality of beamforming with multiple transmit and receive antennas. Assume $n_R \geq 2$. Starting with Z as defined in (4) and making the substitution $Q = U_\Sigma \Lambda_Q U_\Sigma^\dagger$ in (5) we obtain

$$\begin{aligned} C &= \max_{\Lambda_Q : \text{trace}(\Lambda_Q)=P} \mathrm{E}\left[\log\left|I_{n_R} + \frac{Z\Lambda_\Sigma^{\frac{1}{2}}\Lambda_Q\Lambda_\Sigma^{\frac{1}{2}}Z^\dagger}{\sigma^2}\right|\right] \\ &= \max_{\Lambda_Q : \text{trace}(\Lambda_Q)=P} \mathrm{E}\left[\log\left|I_{n_R} + \sum_{i=1}^{n_T}\frac{Z_{\cdot i}Z_{\cdot i}^\dagger\lambda_i^\Sigma\lambda_i^Q}{\sigma^2}\right|\right] \end{aligned}$$

Similar to the approach in [4], let us allocate power $P-p$ to the dominant eigenvector and p to the next strongest eigenvector. This gives us a capacity of

$$C(p) = \mathrm{E}\left[\log\left|I_{n_R} + (P-p)\tfrac{Z_{\cdot 1}Z_{\cdot 1}^\dagger\lambda_1^\Sigma}{\sigma^2} + p\tfrac{Z_{\cdot 2}Z_{\cdot 2}^\dagger\lambda_2^\Sigma}{\sigma^2}\right|\right] \quad (20)$$

In [4], for a single receive antenna the necessary condition for optimality of beamforming was obtained from the inequality $\left.\frac{\partial C(p)}{\partial p}\right|_{p=0} < 0$ and the condition was found to be sufficient since the second derivative $\frac{\partial^2 C(p)}{\partial p^2}$ was negative for all p. However, for multiple receive antennas the form of the capacity expression (20) does not lend itself easily to differentiation. So instead, we use bounds to derive necessary and sufficient conditions. We start with the necessary condition.

A. Necessary Condition

We use the following theorems.

Theorem 4: Let A, B be Hermitian matrices, and let $\lambda^A = [\lambda_i^A]$, $\lambda^B = [\lambda_i^B]$, and $\lambda^{A+B} = [\lambda_i^{A+B}]$ denote the column vectors in $\mathbb{R}^n$ whose components are the eigenvalues of A, B, and $A+B$, respectively, arranged in increasing order. The vector λ^{A+B} majorizes the vector $\lambda^A + \lambda^B$.

Proof: This is Theorem 4.3.27 in [8]. ∎

Theorem 5: If g:$\mathbb{R}^+ \to \mathbb{R}$ is strictly concave then the symmetric concave function $\phi(x) = \sum_{i=1}^n g(x_i)$ is Schur-concave. Recall that a real function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be Schur concave if for all $x, y \in \mathbb{R}^n$ such that y majorizes x we have $\phi(y) \geq \phi(x)$.

Proof: This is Theorem 3.C.1 in [9]. ∎

Returning to our problem let us define the Hermitian matrices

$$A \triangleq I_{n_R} + (P-p)\frac{Z_{\cdot 1}Z_{\cdot 1}^\dagger\lambda_1^\Sigma}{\sigma^2}, \quad (21)$$

$$B \triangleq p\frac{Z_{\cdot 2}Z_{\cdot 2}^\dagger\lambda_2^\Sigma}{\sigma^2}. \quad (22)$$

The n_R eigenvalues of these matrices, arranged in decreasing order, are as follows:

$$\lambda^A = \{(P-p)\frac{Z_{\cdot 1}^\dagger Z_{\cdot 1}\lambda_1^\Sigma}{\sigma^2} + 1, 1, 1, \cdots, 1\}, \quad (23)$$

$$\lambda^B = \{p\frac{Z_{\cdot 2}^\dagger Z_{\cdot 2}\lambda_2^\Sigma}{\sigma^2}, 0, 0, \cdots, 0\}. \quad (24)$$

The capacity expression (20) now becomes

$$\begin{aligned} C(p) &= \mathrm{E}\left[\sum_{i=1}^{n_R}\log(\lambda_i^{A+B})\right] && (25) \\ &\geq \mathrm{E}\left[\sum_{i=1}^{n_R}\log(\lambda_i^A + \lambda_i^B)\right] && (26) \\ &= \mathrm{E}\log(1 + (P-p)\frac{Z_{\cdot 1}^\dagger Z_{\cdot 1}\lambda_1^\Sigma}{\sigma^2} + p\frac{Z_{\cdot 2}^\dagger Z_{\cdot 2}\lambda_2^\Sigma}{\sigma^2}) \\ &\triangleq C_l(p). && (27) \end{aligned}$$

Thus $C_l(p)$ is a lower bound on the capacity $C(p)$. We used Theorems 4 and 5 to obtain (26) from (25). Note that $C(0) = C_l(0)$. This is crucial for the derivation that follows.

A necessary condition for optimality of beamforming is that $\left.\frac{\partial C(p)}{\partial p}\right|_{p=0} \leq 0$. While it is difficult to obtain this derivative directly we can use the lower bound to obtain a necessary condition. Since $C(0) = C_l(0)$ and $C(p) \geq C_l(p)$ for $p > 0$, if $\left.\frac{\partial C(p)}{\partial p}\right|_{p=0} \leq 0$ then we must have $\left.\frac{\partial C_l(p)}{\partial p}\right|_{p=0} \leq 0$. Then a necessary condition for optimality of beamforming is $\left.\frac{\partial C_l(p)}{\partial p}\right|_{p=0} \leq 0$.

Let us define

$$w_1 \triangleq Z_{\cdot 1}^\dagger Z_{\cdot 1} \quad (28)$$

$$w_2 \triangleq Z_{\cdot 2}^\dagger Z_{\cdot 2} \quad (29)$$

where E[w_1]=E[w_2]=n_R and w_1, w_2 are i.i.d. with distribution given by $p(w) = \frac{w^{(n_R-1)}}{(n_R-1)!}e^{-w}$. Differentiating (27), rearranging terms and integrating, we obtain the necessary condition as

$$\left.\frac{\partial C_l}{\partial p}\right|_{p=0} = \mathrm{E}\left[\frac{\frac{\lambda_2^\Sigma}{\sigma^2}w_2 - \frac{\lambda_1^\Sigma}{\sigma^2}w_1}{1 + \frac{P\lambda_1^\Sigma}{\sigma^2}w_1}\right] \leq 0,$$

which can be simplified as

$$\frac{P\lambda_2^\Sigma}{\sigma^2} \leq \frac{1}{n_R}\frac{1}{\left(\frac{\sigma^2}{P\lambda_1^\Sigma}\right)^{n_R} e^{\frac{\sigma^2}{P\lambda_1^\Sigma}}}\Gamma(1-n_R, \frac{\sigma^2}{P\lambda_1^\Sigma}) - \frac{1}{n_R}. \quad (30)$$

Here $\Gamma(\alpha, \beta) = \int_\beta^\infty x^{\alpha-1}e^{-x}dx$ is the incomplete Gamma function.

B. Sufficient Condition

To derive a sufficient condition for optimality of beamforming we need an upperbound on the capacity expression (20). We need the following theorem:

Theorem 6: If $A, B \in \mathbb{C}^n$ are n-dimensional column vectors then

$$|I + AA^\dagger + BB^\dagger| \leq |I + AA^\dagger||I + BB^\dagger| \qquad (31)$$

Proof: Note that $AA^\dagger$ and $BB^\dagger$ are rank one matrices with the nonzero eigenvalue given by $A^\dagger A$ and $B^\dagger B$ respectively. So the right hand side of the inequality equals $(1 + A^\dagger A)(1 + B^\dagger B)$. The matrix $AA^\dagger + BB^\dagger$ has rank at most 2. Let the nonzero eigenvalues be λ_1 and λ_2 where $\lambda_1 \geq \lambda_2$. So the LHS equals $(1 + \lambda_1)(1 + \lambda_2)$. Equating traces we must have $\lambda_1 + \lambda_2 = A^\dagger A + B^\dagger B$. Also, using the interlacing property of eigenvalues of Hermitian matrices (Theorem 4.3.4 in [8]) we must have $\lambda_1 \geq A^\dagger A, B^\dagger B \geq \lambda_2$. So the vector $\{1 + \lambda_1, 1 + \lambda_2\}$ is majorized by the vector $\{1 + A^\dagger A, 1 + B^\dagger B\}$ and the result follows (see Problem 11 on page 199 of [8]). ■

With A and B defined as $\frac{\sqrt{(P-p)\lambda_1^\Sigma}}{\sigma} Z_{.1}$ and $\frac{\sqrt{p\lambda_2^\Sigma}}{\sigma} Z_{.2}$, applying Theorem 6 to the capacity expression in (20) gives us the following upperbound:

$$\begin{aligned} C(p) &\leq \mathrm{E}\log(1 + (P-p)\frac{w_1\lambda_1^\Sigma}{\sigma^2}) + \mathrm{E}\log(1 + p\frac{w_2\lambda_2^\Sigma}{\sigma^2}) \\ &\triangleq C_u(p), \end{aligned} \qquad (32)$$

where w_1, w_2 are as defined in (28) and (29). Again note that the upperbound $C_u(p)$ is equal to the actual capacity $C(p)$ for $p = 0$ and $C_u(p) \geq C(p)$ for $p > 0$. This implies that if $\left.\frac{\partial C_u(p)}{\partial p}\right|_{p=0} \leq 0$ then $\left.\frac{\partial C(p)}{\partial p}\right|_{p=0} \leq 0$. Also it is easy to see that the second derivative of $C_u(p)$ is negative for $p \in (0, P)$. Thus a sufficient condition for optimality of beamforming is

$$\left.\frac{\partial C_u}{\partial p}\right|_{p=0} = -\mathrm{E}\left[\frac{\frac{\lambda_1^\Sigma}{\sigma^2} w_1}{1 + \frac{P\lambda_1^\Sigma}{\sigma^2} w_1}\right] + \mathrm{E}[\frac{w_2\lambda_2^\Sigma}{\sigma^2}] \leq 0 \qquad (33)$$

$$\Rightarrow \quad \mathrm{E}\left[\frac{1}{1 + \frac{P\lambda_1^\Sigma}{\sigma^2} w_1}\right] \leq 1 - \frac{n_R P\lambda_2^\Sigma}{\sigma^2} \qquad (34)$$

$$\Rightarrow \quad \frac{P\lambda_2^\Sigma}{\sigma^2} \leq \frac{1}{n_R} - \frac{1}{n_R}\left(\frac{\sigma^2}{P\lambda_1^\Sigma}\right)^{n_R} e^{\frac{\sigma^2}{P\lambda_1^\Sigma}}\Gamma(1 - n_R, \frac{\sigma^2}{P\lambda_1^\Sigma})$$

Figure 1 shows how the necessary and sufficient conditions divide the $(\frac{P\lambda_1^\Sigma}{\sigma^2}, \frac{P\lambda_2^\Sigma}{\sigma^2})$[1] space into regions where beamforming is the capacity achieving strategy and where beamforming cannot achieve capacity for two receive antennas. Also shown is the necessary and sufficient condition for a single receive antenna found in [4].

VI. Conclusions

We solved the capacity optimization problem for multiple transmit and receive antennas when the rows of the channel matrix are independent and identically distributed while the columns are correlated. The optimum transmit strategy involves transmitting independent complex Gaussian symbols along the eigenvectors of the channel covariance matrix. We also determined necessary and sufficient conditions on the spatial correlations for beamforming to be a capacity achieving strategy. As in the single receive antenna case the optimality of beamforming depends upon the total transmit power, noise variance, and the two largest eigenvalues of the channel covariance matrix.

It is difficult to use the single antenna approach [4] directly to solve the beamforming optimality problem for multiple receive antennas since the form of the capacity expression does not lend itself to simple differentiation. Instead, we derived lower and upper bounds on the capacity, and then used these bounds to derive necessary and sufficient conditions for optimality of beamforming. We find that even with multiple receive antennas and imperfect feedback beamforming becomes the optimal strategy as the eigenvalues of the channel covariance matrix become more disparate, as the variance of noise increases, or as the transmit power decreases. This is consistent with the notion of waterfilling over multiple degrees of freedom. Multiple transmit antennas add degrees of freedom. Beamforming corresponds to using only one of these. If the returns associated with using each degree of freedom are too disparate (this corresponds to strong spatial correlation between the columns of the channel matrix), or if the transmit power is too small, the waterfill covers only the deepest (strongest) mode and all others are left dry - leading to a beamforming solution.

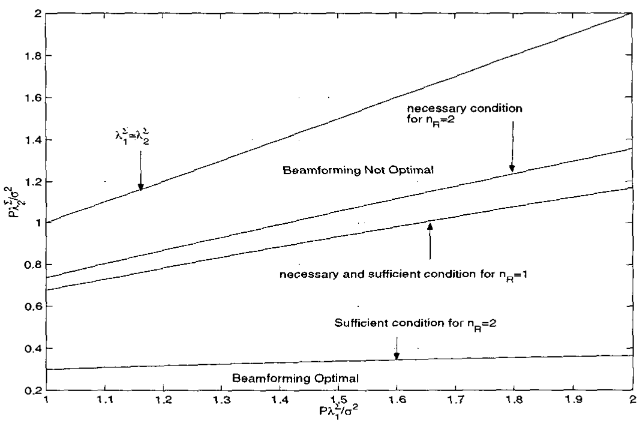

Fig. 1. Necessary and Sufficient Conditions for Optimality of Beamforming

[1] Note that λ_1^Σ is the dominant eigenvalue, so we are only interested in the region $\lambda_1^\Sigma > \lambda_2^\Sigma$.

References

[1] E. Telatar, "Capacity of Multi-antenna Gaussian Channels", AT&T-Bell Labs Internal Tech. Memo., June 1995.

[2] A. Narula, M. J. Lopez, M. D. Trott, G. W. Wornell, "Efficient Use of Side Information in Multiple-Antenna Data Transmission over Fading Channels", IEEE JSAC, Vol. 16, No. 8, October 1998.

[3] E. Visotsky, U. Madhow, "Space-Time Transmit Precoding with Imperfect Feedback", Proceedings of ISIT 2000.

[4] A. J. Syed, A. J. Goldsmith, "On Optimality of Beamforming for Multiple Antenna Systems with Imperfect Feedback", Submitted.

[5] D. Shiu, G.J. Foschini, M.J.Gans, J.M. Kahn, "Fading Correlation and Its Effect on the Capacity of Multi-Element Antenna Systems", Pages: 502-513, IEEE Transactions on Communications, March 2000.

[6] G. Claire, S. Shamai, "On the Capacity of Some Channels with Channel State Information", IEEE Trans. on Info. Theory, Vol. 45, No. 6, Sep. 1997.

[7] A Narula, M. D. Trott, G. W. Wornell, "Performance Limits of Coded Diversity Methods for Transmitter Antenna Arrays", IEEE Trans. on Info. Theory, Vol. 45, No. 7, Nov. 1999.

[8] R. Horn, C. R. Johnson, Matrix Analysis, Cambridge University Press.

[9] A. W. Marshall, I. Olkin, Inequalities: Theory of Majorization and its applications, Academic Press, 1979.