Optimality of Extended Maximum Ratio Transmission

Chandra Murthy, June Chul Roh and Bhaskar D. Rao
Department of Electrical and Computer Engineering
University of California, San Diego, CA 92093-0407.
E-mail: {cmurthy,jroh,brao}@ucsd.edu

***Abstract*— In this paper, we investigate a general class of partial channel state information feedback strategies for communication over flat fading multiple antenna channels with t transmitters and r receivers. Specifically, we consider scenarios where the partial channel state information is sent back from the receiver to the transmitter in the form of n beamformer vectors with corresponding power allocations [1]. We show that, in order to maximize the channel capacity, a globally optimal strategy is to feedback the n dominant eigenvectors of $H^H H$.**

I. Introduction

Multiple antenna (MIMO) systems have received much attention in the recent years due to the higher capacity that they inherently offer compared to single antenna (SISO) systems. Studies of the information theoretic capacity of MIMO systems have typically assumed knowledge of channel state information (CSI) either at both the transmitter and the receiver (complete CSIT) [2], [4], or only at the receiver (no CSIT) [2], [3], [5]. Complete CSIT inherently provides higher capacity than no CSIT, especially at low transmit powers and when the number of transmit antennas is more than the number of receive antennas. However, achieving the higher capacity of complete CSIT entails perfect knowledge of the instantaneous channel at the transmitter. In general, this requires a fairly high feedback data rate from the receiver to the transmitter.

In order to achieve most of the gains of complete CSIT without requiring a very high feedback data rate, we consider partial channel state information feedback schemes. We consider the transmission of $n > 1$ data streams by employing n different (possibly orthonormal) beamformers, one for each data stream, and regard the n beamforming vectors (and an associated set of optimal transmit powers) as partial channel state feedback information. The reason for choosing $n > 1$ is because, optimum beamforming ($n = 1$) has a non-negligible probability of achieving the waterfilling capacity at practical signal to noise ratios only for the smallest (e.g., 2×2) MIMO systems [6]. Hence, in order to reap the benefits of the MIMO system, it is necessary to either use space-time block coding to obtain diversity gain, or transmit on at least a few beamforming directions to obtain multiplexing gain. It has been suggested [8] that the optimum beamformer is likely to be the matrix V_1 whose columns are the first n dominant eigenvectors of $H^H H$, where $H \in \mathbf{C}^{r\times t}$ is the MIMO channel matrix. (A^H denotes the conjugate-transpose of the matrix A.) This choice is intuitively satisfying, since the matrix with rank $n \le m \triangleq \text{rank}(H)$ that is closest (in the 2-norm sense) to $H = \sum_{i=1}^{m} \sigma_i u_i v_i^H$ is $H_n = \sum_{i=1}^{n} \sigma_i u_i v_i^H$, and H_n has n right singular vectors $V_1 = [v_1, v_2, \ldots v_n]$. In this paper, we prove this proposition. The result is important because it legitimizes the choice of dominant eigenvector feedback among partial channel feedback schemes. Using the dominant eigenvectors V_1 as beamforming vectors for transmission is called extended maximum ratio transmission (EMRT) [1]. In section II, we describe the system model and set down the notation. In section III, we prove the optimality of EMRT, and summarize our conclusions in section IV.

II. System Model and Preliminaries

A. Channel Model

We consider multiple antenna systems with t antennas at the transmitter, and r antennas at the receiver. We model a slow, flat-fading channel by a channel matrix $H \in \mathbf{C}^{r\times t}$. The channel input is $x \in \mathbf{C}^t$ and the channel output is $y \in \mathbf{C}^r$, and the input-output equation is:

$$y = Hx + \eta \tag{1}$$

where $\eta \in \mathbf{C}^r$ is the noise vector with *i.i.d.* circularly symmetric complex zero-mean, unit variance Gaussian random variables as its entries, i.e., $E\{\eta\eta^H\} = I_r$, where $E\{\cdot\}$ denotes the expectation operation, and I_r is the $r \times r$ identity matrix. The channel input x, the transfer matrix H, and the receiver noise η are assumed to be mutually independent.

We denote the singular value decomposition of H by $H = U\Sigma V^H$, and the rank of H by m. $V \in \mathbf{C}^{t\times t}$ and $U \in \mathbf{C}^{r\times r}$ are unitary matrices spanning the input space $\mathbf{C}^t$ and the output space $\mathbf{C}^r$, respectively. $\Sigma \in \mathbf{R}^{r\times t}$ contains the singular values σ_i, $1 \le i \le m$, with $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_m > 0$. The total power constraint is imposed as $E\{x^H x\} \le P_T$. For notational convenience, we also define $V_1 = [v_1, \ldots, v_n]$, where v_i is the i-th column vector of V.

B. Ergodic Channel Capacity

In the sequel, our performance measure will be ergodic channel capacity [2]. The ergodic channel capacity of a random MIMO channel with transmit power constraint P_T is given by:

$$C(P_T) = E_H\{C(P_T; H)\}$$

where $E_H\{\cdot\}$ denotes the expectation over channel realizations; and $C(P_T;H)$ is the conditional capacity for a given channel realization H with a power constraint P_T. That is,

$$C(P_T;H) = \max_{p(x):\ E\{x^H x\}\le P_T} I(x;y)$$

The mutual information $I(x;y)$ satisfies the following inequality:

$$I(x;y) \le \log\ \det(I_r + H\Phi_x H^H) \qquad (2)$$

where $\log(\cdot)$ is the base-2 logarithm, $\det(A)$ denotes the determinant of A, and $\Phi_x = E\{xx^H\}$. The equality in (2) holds if and only if x is a circularly symmetric Gaussian random vector. Thus, if we maximize the mutual information for every channel realization H, we would then maximize the ergodic channel capacity. The channel capacity conditioned on H is given by:

$$C(P_T;H) = \max_{\mathrm{tr}\,\Phi_x\le P_T}\ \log\ \det(I_r + H\Phi_x H^H) \qquad (3)$$

We assume that perfect CSI is known at the receiver. In a system where *single* transmit data stream is transmitted over t transmit antennas after passing through a beamformer $w \in \mathbf{C}^t$, the optimum choice of the beamformer w is the first eigenvector of $H^H H$, or, v_1. This transmission scheme is called maximum-ratio transmission (MRT). The choice of the beamformer as $w = v_1$ is optimal in terms of maximizing the received signal-to-noise ratio [7]. In addition, it can be shown that this choice of beamformer is also optimal in the sense of maximizing the mutual information.

C. Preliminaries

We now make a few key observations and summarize a few results that motivate and facilitate the development of the main proof.

1) MIMO channels with complete CSIT: We denote by C_{HH} the conditional capacity of MIMO channels with CSI H fully known to both the transmitter and the receiver. With complete CSIT, it is optimum to use the m columns of V, the eigenvectors of $H^H H$, as beamforming vectors, with corresponding optimal power allocation determined by water-filling over m independent spatial channels [2]. The capacity is given by:

$$\begin{aligned} C_{HH}(P_T;H) &= \max_{\substack{P_1\ge 0,\dots,P_m\ge 0\\ P_1+\dots+P_m\le P_T}} \sum_{i=1}^{m} \log(1+P_i\lambda_i) \\ &= \sum_{i=1}^{m} [\log(\nu\lambda_i)]^+ \end{aligned} \qquad (4)$$

where $\lambda_i = \sigma_i^2$ is the i-th largest eigenvalue of $H^H H$, and P_i is the transmit power allocated on the i-th data stream (i.e., $x = Vs$, with $\Phi_s = \mathrm{diag}(P_1,\dots P_t)$). $[a]^+$ is defined as $\max\{a,0\}$ and ν is the water-filling level satisfying the power constraint:

$$\sum_{i=1}^{m}\left[\nu - \frac{1}{\lambda_i}\right]^+ = P_T$$

The optimum power allocation P_i for $1 \le i \le n$ can be shown to be given by:

$$P_i = \left[1 + \sum_{l=1}^{n}(P_T\lambda_l)^{-1} - n(P_T\lambda_i)^{-1}\right]^+ P_T/n \qquad (5)$$

Note that given a channel matrix H, there is a critical total transmit power $P_T^{(c)}(n)$ such that for $P_T \le P_T^{(c)}(n)$, the water-filling procedure would allocate non-zero power only for the first n spatial channels. For this range of P_T, the problem of finding optimum beamforming vectors is trivial: we simply use the optimum solution with complete CSIT and note that the transmitter only needs to know n beamforming vectors, which are the first n eigenvectors of $H^H H$. $P_T^{(c)}(n)$ is given by:

$$P_T^{(c)}(n) = n\lambda_{n+1}^{-1} - \sum_{i=1}^{n}\lambda_i^{-1} \qquad (6)$$

The goal of this paper is to prove that the set of beamforming vectors which are optimum for $P_T \le P_T^{(c)}(n)$ are also optimum when $P_T > P_T^{(c)}(n)$.

2) Interlacing property of eigenvalues: Here, we state a result from matrix algebra that is useful for our analysis. It relates the eigenvalues of a Hermitian matrix to that of its principal submatrices. Let $H = [H_1|H_2]$, where H_1 is the first n columns of H and H_2 is the remaining $t-n$ columns. Let $\lambda_1,\dots,\lambda_m$ be the m non-zero eigenvalues of $H^H H$ arranged in decreasing order, and $\lambda_{s1},\dots,\lambda_{sn}$ be that of $H_1^H H_1$. Then, since $H^H H \in \mathbf{C}^{t\times t}$ is Hermitian and since $H_1^H H_1$ is a principal submatrix of $H^H H$, for every k such that $1 \le k \le n$, the eigenvalues λ_{sk} of $H_1^H H_1$ satisfy (see thm. 4.3.15 in [10]):

$$\lambda_{t-n+k} \le \lambda_{sk} \le \lambda_k \qquad (7)$$

We will use this property in proving the optimality of EMRT.

III. Extended Maximum-Ratio Transmission

When n $(1 < n < m)$ spatial channels are known at the transmitter, a generalization of the MRT for a system with n transmit data streams is to employ n different beamformers, one for each data stream. That is, the transmitter uses the n given spatial beamforming vectors to transmit n data streams along n spatial directions.

In this scheme, the transmitted signal $x \in \mathbf{C}^t$ can be modelled as

$$x = w_1 s_1 + \ldots + w_n s_n = Ws \qquad (8)$$

where $W = [w_1,\dots,w_n] \in \mathbf{C}^{t\times n}$ is referred to as the beamformer matrix, and it is assumed that the columns w_i of W are linearly independent. The vector $s = [s_1,\dots,s_n]^T \in \mathbf{C}^t$ contains the information being transmitted on the n beams. In this section, we will show that the n dominant eigenvectors of $H^H H$, i.e., V_1, which are considered as partial spatial information of the channel, are the optimal choice of the beamformer matrix in the sense of maximizing the channel capacity for any given H.

Under this formulation, the transmit covariance matrix $\Phi_x = E\{xx^H\}$ is given by $\Phi_x = W\Phi_s W^H$ where $\Phi_s = E\{ss^H\}$. If Φ_s is not constrained to be diagonal, then one can constrain matrix W to have orthonormal columns without loss of generality. To elaborate, if $\tilde{W}$ is an orthogonal basis for the span of W, then $W = \tilde{W}Q$, where Q is an invertible $n \times n$ matrix. Then $\Phi_x = W\Phi_s W^H = \tilde{W}\tilde{\Phi}_s\tilde{W}^H$, where $\tilde{\Phi}_s = Q\Phi_s Q^H$. This implies that (W, Φ_s) and $(\tilde{W}, \tilde{\Phi}_s)$ are equivalent parameterizations. Furthermore with W constrained to have orthonormal columns we have[1] $\operatorname{tr}\Phi_x = \operatorname{tr}\Phi_s$, allowing the power constraint to be satisfied by controlling Φ_s. Thus, the constraints we impose on W and Φ_s are as follows: $W \in \mathcal{S}$ and $\operatorname{tr}\Phi_s \le P_T$, where $\mathcal{S}$ is the set of $t \times n$ complex matrices defined as follows

$$\mathcal{S} \triangleq \left\{B : B \in \mathbf{C}^{t\times n}; B^H B = I_n; \dim(\operatorname{span}(B)) = n\right\} \quad (9)$$

We can now restate the optimization problem (3) as follows:

$$C(P_T; H) = \max_{\substack{W\in\mathcal{S};\\ \operatorname{tr}\Phi_s \le P_T}} \log \det(I_r + HW\Phi_s W^H H^H) \quad (10)$$

Thus, we wish to find the arguments W and Φ_s that solve (10). The method we use is as follows. Given some choice for the beamforming matrix W, we maximize the capacity for that W over all possible choices for Φ_s. Then, we find the supremum of this capacity over all possible choices of rank-n beamforming matrices W.

Denote the capacity (10) given a particular choice for W by $C(P_T, W)$. Now, the problem of maximizing the channel capacity for a given choice of W is equivalent to the complete-CSIT problem with a new (virtual) channel matrix $\tilde{H} \triangleq HW$. Therefore, the solution given by (4) can be directly applied, and the channel capacity is upper-bounded by:

$$\begin{aligned} C(P_T, W; H) &\le C_{\tilde{H}\tilde{H}}(P_T; \tilde{H}) \\ &= \max_{\substack{P_1\ge 0,\ldots,P_n\ge 0\\ P_1+\ldots+P_n\le P_T}} \sum_{i=1}^{n} \log(1 + P_i\tilde{\lambda}_i) \\ &= \sum_{i=1}^{n} \log(1 + \tilde{P}_i\tilde{\lambda}_i) \end{aligned} \quad (11)$$

where we have used tildes to highlight the dependence of the parameters on $\tilde{H}$ rather than on H itself (for example, $\tilde{\lambda}_i,\ 1 \le i \le n$ are the eigenvalues of $\tilde{H}^H\tilde{H}$). The upper bound is achieved by choosing $\Phi_s = \tilde{V}\operatorname{diag}(\tilde{P}_1, \ldots, \tilde{P}_n)\tilde{V}^H$ thereby decoupling the channel $\tilde{H}$ into n parallel independent channels, where $\tilde{V}$ is a matrix whose columns are the eigenvectors of $\tilde{H}^H\tilde{H}$, and $\tilde{P}_1, \ldots, \tilde{P}_n$ is the water-filling power allocation.

Next, we state the following lemma that helps us maximize $C(P_T, W)$ over all possible choices of W.

Lemma 1: With $\tilde{P}_i$ and $\tilde{\lambda}_i,\ 1 \le i \le n$ as defined by (11),

$$\sum_{i=1}^{n} \log(1 + \tilde{P}_i\tilde{\lambda}_i) \le \sum_{i=1}^{n} \log(1 + P_i\lambda_i) \quad (12)$$

where $\lambda_i = \sigma_i^2$ are the n largest eigenvalues of $H^H H$, and P_i are the optimal (water-filling) power allocation across n parallel independent channels with gains $\sigma_i,\ 1 \le i \le n$.

Proof: We have $\tilde{H} = HW$. Denote by $W^\perp$, a matrix with $t - n$ orthonormal columns that are also orthogonal to the space spanned by W, i.e.,

$$(W^\perp)^H W^\perp = I_{t-n},\ W^H W^\perp = 0$$

For example, $W^\perp$ can be chosen as the eigenvectors corresponding to the nonzero eigenvalues of $I_t - WW^H$. Next, let $\hat{H} \triangleq H[W|W^\perp] = [\tilde{H}|HW^\perp]$. Then, $\hat{H}^H\hat{H}$ has the same eigenvalues λ_i as $H^H H$ since H and $\hat{H}$ are related by multiplication by an orthonormal matrix $[W|W^\perp]$. Also, since $\tilde{H}^H\tilde{H}$ is a principal submatrix of $\hat{H}^H\hat{H}$, by virtue of the interlacing property (7), we have $\tilde{\lambda}_i \le \lambda_i$, whence $\sum_{i=1}^n \log(1 + \tilde{P}_i\tilde{\lambda}_i) \le \sum_{i=1}^n \log(1 + \tilde{P}_i\lambda_i)$. Now, we can use the water-filling procedure as in (4) and (5) to re-distribute the total available power P_T among the n channels with gains σ_i, and the resultant optimal power allocation $P_i,\ 1 \le i \le n$ would satisfy $\sum_{i=1}^n \log(1 + \tilde{P}_i\lambda_i) \le \sum_{i=1}^n \log(1 + P_i\lambda_i)$, which completes the proof. Note that we have slightly abused the notation here: in (4), P_i was the optimal power allocation assuming complete CSIT (that is, P_i was the optimal power allocation across $m = \operatorname{rank}(H)$ decoupled channels), whereas here, P_i is the optimal power allocation across $n < m$ decoupled channels with gains $\sigma_i,\ 1 \le i \le n$. ■

Note that the upper bound of (12) can be achieved by choosing $W = V_1$, and using the water-filling power allocation across the decoupled channels with gains $\sigma_i, 1 \le i \le n$. Thus, we state the following result:

Theorem 1: In the beamforming scheme of (8), the conditional capacity for any channel instantiation H has a global maximum given by

$$C(P_T; H) = \sum_{i=1}^{n} \log(1 + \lambda_i P_i) \quad (13)$$

where λ_i is the i-th largest eigenvector of $H^H H$, and $P_i,\ 1 \le i \le n$ is given by (5), that is

$$P_i = \left[1 + \sum_{l=1}^{n} (P_T\lambda_l)^{-1} - n(P_T\lambda_i)^{-1}\right]^+ P_T/n$$

This maximum can be achieved when (a) $W = V_1$ and (b) the powers P_i are chosen by water-filling across n decoupled channels with gains $\sigma_i,\ 1 \le i \le n$.

In other words, feedback of the dominant eigenvectors of $H^H H$ along with the waterfilling power allocation P_i on each (decoupled) channel maximizes mutual information of an EMRT MIMO system.

Proof: From (11) and (12), we have

$$C(P_T, W; H) \le \sum_{i=1}^{n} \log(1 + P_i\lambda_i) \quad (14)$$

[1]tr denotes the trace operator.

Since the right hand side of the above equation is independent of W, we have

$$\sup_{W\in\mathcal{S}} C(P_T, W; H) \leq \sum_{i=1}^{n} \log(1 + P_i\lambda_i) \qquad (15)$$

where $\mathcal{S}$ is as in (9). However, with $W = V_1$, we have

$$HW = U\Sigma V^H V_1 = U\Sigma_1$$

where $\Sigma_1 \in \mathbf{R}^{r\times n}$ is a diagonal matrix with $\sigma_1, \ldots, \sigma_n$, the n largest singular values of H, as its diagonal elements. Substituting in (10) we have

$$\begin{aligned} C(P_T, V_1; H) &= \max_{\text{tr}\,\Phi_s\leq P_T} \log\det(I_r + U\Sigma_1\Phi_s\Sigma_1^H U^H) \\ &= \max_{\text{tr}\,\Phi_s\leq P_T} \log\det(I_r + \Sigma_1\Phi_s\Sigma_1^H) \\ &= \max_{\text{tr}\,\Phi_s\leq P_T} \log\det(I_n + \Phi_s\text{diag}(\lambda_1, \ldots, \lambda_n)) \\ &= \sum_{i=1}^{n} \log(1 + \lambda_i P_i) \qquad (16) \end{aligned}$$

In the above, we have used the determinant identity $\det(I + FG) = \det(I + GF)$. The last equality in (16) follows because the water-filling solution $P_i,\ 1 \leq i \leq n$ maximizes the conditional capacity as in (4). Thus, the upper bound of (14) is achieved by choosing $W = V_1$ and the water-filling power allocation. ∎

It is worth noting that the optimum receive filter R is U_1^H, where $U_1^H \in \mathbf{C}^{n\times r}$ is a matrix consisting of the first n rows of U^H. That is, the received vector y is pre-multiplied by U_1^H to obtain $U_1^H y = \Sigma_{1,1}s + U_1^H\eta$, where $\Sigma_{1,1} = \text{diag}(\sigma_1, \ldots, \sigma_n)$. The MIMO channel is thus decoupled into the n strongest parallel spatial sub-channels. Furthermore, due to the water-filling power allocation, more power is allocated to the stronger modes. It is also interesting to compare this result with [9], where the authors jointly design the transmit filter W and the receive filter R to minimize the mean-squared error (MMSE) subject to a total average transmit power P_T constraint. Further, a fixed data-rate is assumed to be used for all the n data-streams. The optimization problem is:

$$\min_{\text{tr}\,\Phi_s\leq P_T} E_{s,\eta}\left\{\|s - (RHWs + R\eta)\|_2^2\right\} \qquad (17)$$

where $\|\cdot\|$ is the l_2 (Euclidean) norm on $\mathbf{C}^n$ [10]. The authors solve the optimization problem stated in (17) using the method of Lagrange multipliers to obtain:

$$\begin{aligned} W &= V_1 \\ R &= U_1^H \\ \Phi_s &= \left[\frac{1}{\nu}\Sigma_{1,1}^{-1} - \Sigma_{1,1}^{-2}\right]^+ \qquad (18) \end{aligned}$$

where ν in the last equation is chosen such that the total transmit power $\text{tr}\,\Phi_s \leq P_T$ is satisfied. The optimum transmit and receive filters under the capacity maximization criterion and the MMSE criterion are identical, but the optimum power allocation is completely different: water-filling maximizes channel capacity, whereas the so-called *inverse* water-filling minimizes the mean-square error. In order to maximize the capacity we transmit higher data-rate on better channels, while to minimize the mean-square error under a fixed data-rate assumption, we apply more power to weaker modes of the channel.

IV. Conclusion

We have shown that the EMRT using the n dominant eigenvectors of $H^H H$ as the spatial beamforming vectors for n independent data streams is optimal in the sense of maximizing the channel capacity. Equivalently, the first n eigenvectors, V_1, are optimal choice of spatial information of the channel when n is limited to be less than m, the rank of H.

Acknowledgement

This work was supported in part by CoRe grant nos. com 02-10105, com 02-10109, and by Intersil and Ericsson.

References

[1] J. C. Roh and B. D. Rao, "Multiple antenna channels with partial channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, to appear.

[2] I. .E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Labs Tech. Memo.*, 1995.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311-335, Mar. 1998.

[4] E. Biglieri, G. Caire and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273-1289, May 2001.

[5] A. Narula, M. D. Trott and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418-2433, Mar. 1999.

[6] M. Wennstrom, M. Helin, A. Rydberg and T. Oberg, "On the optimality and performance of transmit and receive space diversity in MIMO channels," *IEE Tech. Seminar on MIMO Commun. Systems: From Concept to Implementation*, London, Dec 2001.

[7] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, pp. 1458–1461, Oct. 1999.

[8] H. Sampath, P. Stoica and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, pp. 2198-2206, Dec. 2001.

[9] H. Sampath and A. Paulraj, "Joint transmit and receive optimization for high data rate wireless communication using multiple antennas" *IEEE Asilomar Conference on Signals, Systems and Computers* vol. 1, pp. 215-219, Oct. 1999

[10] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.