Capacity optimization for Rician correlated MIMO wireless channels

Mai Vu , and Arogyaswami Paulraj

Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305-9510, USA
E-mails: {mhv, apaulraj}@stanford.edu

***Abstract*– We develop an optimization program to calculate the ergodic capacity and the optimal input signal covariance of a Rician correlated MIMO wireless channel. The program employs the Newton method and the barrier interior-point method to solve convex stochastic formulations, using efficient gradient and Hessian calculations. We then use the program to study impacts of the channel mean, the transmit correlation, and the K factor on channel capacity and the optimal input signal, revealing curious optimal signal characteristics at high SNR, where mode-dropping can always occur for a channel with strong mean or strong correlation. We also compare the capacity with the mutual information obtained by the signaling solution optimal for the Jensen bound. Results reveal a close approximation for systems with equal or fewer transmit than receive antennas at all SNR, and for those with more transmit than receive antennas at low SNR. The approximation diverges for the latter system at high SNR, depending on the transmit correlation and the channel mean, or the Rician K factor.**

1. INTRODUCTION

In a MIMO wireless system, having partial channel information at the transmitter can increase the system capacity significantly. The capacity represents a fundamental limit on system transmission rate, and there now exists practical coding schemes closely approaching this fundamental performance. Therefore, optimizing the input signal to achieve the capacity, given the transmit channel side information, is of both theoretical and practical interest. We study a channel with the side information in the form of a non-zero channel mean and the transmit antenna correlation. These channel statistics cover physical Rician correlated channels; they also model a channel estimate and the estimation error covariance.

The capacity optimization problem for a Rician correlated channel involves evaluating an expectation over the non-central Wishart distribution. Although the optimal input signal is zero-mean Gaussian distributed [1], a closed-form analytical solution for its covariance given such channel statistics is still an open problem. Partial solutions exist for special cases: when the channel mean is zero (a correlated Rayleigh channel) [2, 3], or when the antennas are uncorrelated (an uncorrelated Rician channel) [4, 5]. In these cases, the eigenvectors of the input covariance matrix are known analytically, but not the eigenvalues. In terms of signal processing, these two quantities respectively function as the beamforming directions (patterns) and the power allocation on these beams.

Fortunately, the capacity optimization formulation in all cases is convex, hence allowing efficient numerical techniques to be implemented [6]. In this paper, we develop a convex optimization program using an interior point method. This program inputs the channel statistical parameters – the mean and the transmit correlation – and the SNR; it outputs the capacity-optimal signal covariance and the channel capacity. The program employs efficient techniques for calculating the gradients and the Hessians, using Monte-Carlo approximation; it has a runtime linear in the number of channel samples and quadratic-to-cubic in the number of transmit antennas.

We then use the program to study impacts of the channel mean, the transmit correlation, and the K factor on the capacity and the optimal input signal. We also compare the optimal input solution to a sub-optimal one, based on Jensen's bound on the mutual information, and establish conditions with which the sub-optimal solution is capacity-tight, hence allowing a simple approximation without performing the optimization.

The rest of this paper is organized as follows: In Section II, we introduce the channel model and formulate the capacity optimization problem. Section III outlines the implemented convex optimization methods. We analyze the program run-time complexity and provide optimization examples in Section IV. In section V, we apply the program to analyze impacts of various channel parameters on the capacity and the optimal input signal, and to assess the sub-optimal solutions. We close with some concluding remarks in Section VI.

2. CHANNEL MODEL AND PROBLEM FORMULATION

We focus on frequency flat MIMO wireless channels with Rician fading and transmit antenna correlation. Let M and N be the number of receive and transmit antennas, respectively. The channel matrix $\mathbf{H}$ of size $M \times N$, complex Gaussian distributed with mean $\mathbf{H}_m$ and transmit covariance $\mathbf{R}_t$, can then be represented as

$$\mathbf{H} \quad = \quad \mathbf{H}_m + \mathbf{H}_w \mathbf{R}_t^{\frac{1}{2}} \ , \tag{1}$$

where $\mathbf{H}_w$ has i.i.d zero-mean unit-variance Gaussian elements. The transmitter is assumed to know only the channel distribution, which includes the mean $\mathbf{H}_m$, an arbitrary complex matrix, and the covariance $\mathbf{R}_t$, a Hermitian positive semi-definite (PSD) matrix. The receiver, on the other hand, is assumed to have perfect channel knowledge at every time instance. This model also applies to a partial transmit channel knowledge form, in which the transmitter has a channel estimate with a known error covariance [7].

The channel ergodic capacity, under an average sum transmit power constraint, is achieved by Gaussian inputs with zero mean and a covariance matrix $\mathbf{Q}$ that is the optimizer of the following problem:

$$\begin{array}{ll} \min_{\mathbf{Q}} & f = -E_{\mathbf{H}}[\log\det(\mathbf{I} + \gamma \mathbf{H}\mathbf{Q}\mathbf{H}^*)] \\ \text{s.t.} & \text{tr}(\mathbf{Q}) = 1 \ , \ \ \mathbf{Q} \succcurlyeq 0 \ , \end{array} \tag{2}$$

where γ is the SNR, tr(.) is the trace of a matrix, and $\succcurlyeq$ denotes PSD. This problem is convex, hence solving for the unknown $\mathbf{Q}$ can be performed numerically up to a desired precision. Since $\mathbf{Q}$ is a Hermitian complex matrix of size $N \times N$, it contains N^2 real scalar variables (including N real entries on the diagonal and $\frac{1}{2}N(N-1)$ complex entries in the upper triangle, which contribute two real variables each). Thus, the optimization variable size is N^2.

There are special cases in which the eigenvectors of the optimal $\mathbf{Q}$ are known analytically. These are a zero-mean correlated channel ($\mathbf{H}_m = 0$ and $\mathbf{R}_t$ is arbitrary) and a non-zero mean uncorrelated channel ($\mathbf{H}_m$ is arbitrary and $\mathbf{R}_t = \mathbf{I}$). Let the eigenvalue decomposition of $\mathbf{Q}$ be $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$. The optimal eigenvectors $\mathbf{U}$ in each case are then given by the eigenvectors of $\mathbf{R}_t$ [2, 3] and of $\mathbf{H}_m^*\mathbf{H}_m$ [4, 5], respectively. These cases only require finding the optimal eigenvalues $\mathbf{\Lambda}$ (i.e. the input power distribution) via numerical optimization, reducing the number of unknowns to N real variables. The optimization problem (2) then becomes

$$\begin{aligned} \min \quad & g = -E_{\mathbf{S}}[\log\det(\mathbf{I} + \mathbf{S}\mathbf{\Lambda})] \\ \text{s.t.} \quad & \text{tr}(\mathbf{\Lambda}) = 1 \ , \ \ \mathbf{\Lambda} \succcurlyeq 0 \ , \end{aligned} \tag{3}$$

where $\mathbf{S} = \gamma\mathbf{U}^*\mathbf{H}^*\mathbf{H}\mathbf{U}$, which has a known distribution.

3. THE OPTIMIZATION METHOD

We solve both problems (2) and (3) using the Newton method. In problem (2), when the positive semidefinite constraint $\mathbf{Q} \succcurlyeq 0$ is active (i.e., some eigenvalues of the optimal $\mathbf{Q}$ are zero), we implement a barrier interior-point method, using the Newton method for the inner iterations [6]. Such a barrier implementation is not necessary for problem (3). In both problems, the optimization is complicated by the stochastic nature of the objective function, making it difficult to compute the exact function value, the gradient, and the Hessian needed at each optimization step. These values are also approximated using sets of discretized channel samples; as the number of transmit antennas N grows, the number of samples needs to be increased. The gap between the current function value and the optimal value, or the function gap-to-optimal value, used in the stopping criterion is also approximated by Monte-Carlo simulations. Therefore, the numerical precision (the tolerance) is dictated by the variance of the sample means.

3.1. Solving for the eigenvalues $\mathbf{\Lambda}$ in problem (3)

First consider the simpler optimization problem (3) with N real variables. Denote $\lambda = \text{diag}(\mathbf{\Lambda})$, a real column-vector consisting of the N unknown eigenvalues. The problem in terms of λ becomes an optimization with equality constraint $\mathbf{1}^T\lambda = 1$. The Newton method can be implemented very efficiently. The condition $\lambda \succcurlyeq 0$ is handled as follows: when ever a Newton step produces a negative λ_i value, this eigenvalue is set to zero, and the optimization continues with the other variables, re-adjusted for the unit sum. This step is equivalent to dropping a mode in a water-filling process.

The Newton method requires calculating the gradient and the Hessian. To do so, compute the following matrices:

$$\mathbf{P} = \mathbf{I} + \mathbf{S}\mathbf{\Lambda} \ , \qquad \mathbf{Y} = \mathbf{P}^{-1} \ , \qquad \mathbf{Z} = \mathbf{Y}\mathbf{S} \ . \tag{4}$$

The largest computational cost is in calculating $\mathbf{Y}$ due to the matrix inversion. Then the gradient of the objective function (3) with respect to λ is a $N \times 1$ row-vector with elements given by [8]

$$(\nabla g)_i = -E\big[\text{tr}\Big(\mathbf{P}^{-1}\frac{\partial \mathbf{P}}{\partial \lambda_i}\Big)\big] = -E\big[\tilde{\mathbf{y}}_i^T\mathbf{s}_i\big] = -E[Z_{ii}] \ , \qquad i = 1 \ldots N \tag{5}$$

where $\tilde{\mathbf{y}}_i^T$ is row i of $\mathbf{Y}$, $\mathbf{s}_i$ is column i of $\mathbf{S}$, and Z_{ii} is the i^{th} diagonal element of $\mathbf{Z}$. The Hessian can also be computed efficiently as a $N \times N$ matrix with elements given by

$$\begin{aligned} (\nabla^2 g)_{ij} &= E\big[\text{tr}\Big(\mathbf{P}^{-1}\frac{\partial \mathbf{P}}{\partial \lambda_i}\mathbf{P}^{-1}\frac{\partial \mathbf{P}}{\partial \lambda_j}\Big)\big] = E\big[\tilde{\mathbf{y}}_j^T\mathbf{s}_i\tilde{\mathbf{y}}_i^T\mathbf{s}_j\big] \\ &= E[Z_{ji}Z_{ij}] \ , \quad i,j = 1 \ldots N \ . \end{aligned} \tag{6}$$

The expectation operator is approximated by taking the sample mean, using a set of sample channels. An independent sample channel set is generated for each Newton step.

3.2. Solving for the covariance $\mathbf{Q}$ in problem (2)

Similarly, problem (2) is solved using the Newton method. However, due to the matrix form of the unknown variable $\mathbf{Q}$, the PSD constraint $\mathbf{Q} \succcurlyeq 0$ needs handling in a different way. Specifically, a barrier interior-point method is implemented if this constraint is active, meaning the optimal $\mathbf{Q}$ has at least one zero eigenvalue. This condition is heuristically established when the initial Newton step without involving the PSD constraint produces a non-positive semidefinite $\mathbf{Q}$. Otherwise, the program proceeds ignoring this PSD constraint. By using a good starting point (such as the Jensen solution, see Section 5.1), this heuristic check leads to the optimal solution, while reducing the optimization time.

We now establish the formula for computing the gradient and the Hessian in each Newton step. Since $\mathbf{Q}$ is a Hermitian matrix, form a vector of unknown variables from the real and imaginary parts of $\mathbf{Q}$ as

$$\mathbf{q} = \Big[\frac{q_{11}}{2} \ldots \frac{q_{NN}}{2} \ q_{R21} \ q_{R32} \ldots q_{RN1} \ q_{I21} \ q_{I32} \ldots q_{IN1}\Big]^T \ , \tag{7}$$

where $\mathbf{Q}_R = \mathcal{R}e(\mathbf{Q})$, $\mathbf{Q}_I = \mathcal{I}m(\mathbf{Q})$, and the lower-case letters refer to the entries in the corresponding matrix. The factor $\frac{1}{2}$ is introduced for uniformity in the gradient and the Hessian (with respect to $\mathbf{q}$) formula. The length of $\mathbf{q}$ is N^2. Compute $\mathbf{G} = \gamma\mathbf{H}^*\mathbf{H}$ and

$$\mathbf{R} = \mathbf{I} + \mathbf{G}\mathbf{Q} \ , \qquad \mathbf{W} = \mathbf{R}^{-1} \ , \qquad \mathbf{X} = \mathbf{W}\mathbf{G} \ . \tag{8}$$

Again, the most intensive step is to compute $\mathbf{W}$, involving matrix inversion. Once $\mathbf{X}$ is computed, the gradient and the Hessian follow directly. Noting that $\mathbf{X}$ is Hermitian, the gradient of the objective function in (2) has elements calculated as

$$\begin{aligned} \frac{\partial f}{\partial q_{Rij}} &= -E\big[\text{tr}\Big(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial q_{Rij}}\Big)\big] \\ &= -E\big[\tilde{\mathbf{w}}_j^T\mathbf{g}_i + \tilde{\mathbf{w}}_i^T\mathbf{g}_j\big] = -2E\big[\mathcal{R}e(\mathbf{X})_{ij}\big] \\ \frac{\partial f}{\partial q_{Iij}} &= -E\big[\text{tr}\Big(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial q_{Iij}}\Big)\big] \\ &= -jE\big[\tilde{\mathbf{w}}_j^T\mathbf{g}_i - \tilde{\mathbf{w}}_i^T\mathbf{g}_j\big] = -2E\big[\mathcal{I}m(\mathbf{X})_{ij}\big] \end{aligned} \tag{9}$$

where $\tilde{\mathbf{w}}_i^T$ is row i of $\mathbf{W}$, $\mathbf{g}_i$ is column i of $\mathbf{G}$, and $\mathcal{R}](.)$ and $\mathcal{I}\Updownarrow(.)$ represent the real and imaginary parts. The Hessian has elements calculated as

$$\begin{aligned}
\frac{\partial^2 f}{\partial q_{Rij}\partial q_{Rkl}} &= 2E\left[\mathcal{R}e(X_{lj}X_{ik} + X_{kj}X_{il})\right] \\
\frac{\partial^2 f}{\partial q_{Iij}\partial q_{Ikl}} &= 2E\left[\mathcal{R}e(X_{lj}X_{ik} - X_{kj}X_{il})\right] \qquad (10) \\
\frac{\partial^2 f}{\partial q_{Rij}\partial q_{Ikl}} &= 2E\left[\mathcal{I}m(X_{lj}X_{ik} + X_{kj}X_{il})\right] \\
\frac{\partial^2 f}{\partial q_{Iij}\partial q_{Rkl}} &= 2E\left[\mathcal{I}m(-X_{lj}X_{ik} + X_{kj}X_{il})\right] \ .
\end{aligned}$$

Again sample means are used to approximate the expected values; an independent sample set is generated for each Newton step.

When the PSD constraint on $\mathbf{Q}$ is tight, equivalently the optimal input signal drops modes, we use the barrier method [6] and iteratively solve the following problem

$$\begin{aligned}
\min \quad & -E_{\mathbf{G}}[\log\det(\mathbf{I} + \mathbf{G}\mathbf{Q})] - \log\det(\mathbf{Q})/t \qquad (11) \\
\text{s.t.} \quad & \mathrm{tr}(\mathbf{Q}) = 1 \ .
\end{aligned}$$

The second term in the objective function, $\log\det(\mathbf{Q})/t$, is to ensure that $\mathbf{Q}$ stays PSD. This term is deterministic and independent of the channel, thus its gradient and Hessian, established similarly to (9) and (10) respectively, can be calculated exactly with little overhead. The Newton method is used in the inner iterations to solve the above problem. Then at each outer iteration, the barrier value t is increased by multiplying with μ, an optimization parameter, until the desired tolerance is satisfied. The barrier implementation, therefore, takes longer to execute due to this double loop.

4. OPTIMIZATION COMPLEXITY AND EXAMPLES

4.1. Complexity assessment

The optimization programs were written using Matlab. Figure 1 shows the computational complexity in terms of program runtime per Newton iteration versus the number of channel samples, and versus the number of transmit antennas N. The runtime scales linearly with the number of channel samples, but at different rates for each problem (2) and (3). For the number of transmit antennas N in the range of interest ($N \leq 10$), the runtime for problem (3) scales as N^2, faster than the order of N^3 predicted by theory [6]. For problem (2) without the barrier implementation, the runtime scales by an order of N^3, not N^6 as theoretically predicted.

4.2. Numerical examples

Figure 2 presents an example of the optimization process, showing the mutual information value and its gap to the optimal value versus the number of iterations. The optimization parameters are given in the Appendix. We choose the starting point of the optimization as the Jensen covariance solution (see Section 5.1), resulting in a fast convergence. The fluctuation in the function value and its gap-to-optimal value over different iterations is due to Monte Carlo simulation (the use of sample mean to approximate the expectation). Thus, the function error floor is dictated by the number of channel samples, but not the number of iterations as the case in deterministic function optimization.

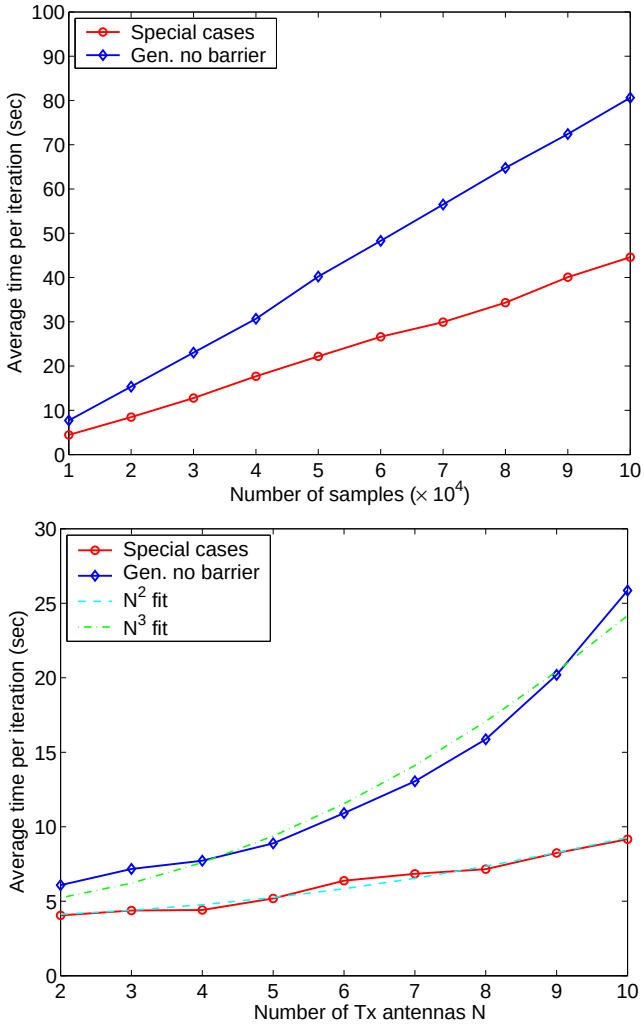

Fig. 1. Runtime complexity versus the number of channel samples (above) and the number of transmit antennas (below).

5. APPLICATIONS IN CAPACITY ANALYSIS

5.1. The Jensen solution

Of interest is a fast way to approximate the capacity and the optimal input covariance. Consider optimizing Jensen's upper-bound on the mutual information, for which a closed-form analytical solution is available. The Jensen bound on the average mutual information is

$$E[\log\det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^*)] \quad \leq \quad \log\det(\mathbf{I} + E[\mathbf{H}^*\mathbf{H}]\mathbf{Q}) \ .$$

Perform the eigenvalue decomposition $E[\mathbf{H}^*\mathbf{H}] = \mathbf{U}_s\mathbf{D}\mathbf{U}_s{}^*$; the covariance $\mathbf{Q}_J$ that maximizes the Jensen bound has the eigenvectors given by $\mathbf{U}_s$, and the eigenvalues obtained by the standard water-filling on $\mathbf{D}$ as

$$\lambda_i = \left(\mu - \frac{1}{\gamma d_i}\right)_+ , \quad i = 1 \ldots N \ , \qquad (12)$$

where d_i are the diagonal elements of $\mathbf{D}$, and μ is chosen to satisfy $\sum_{i=1}^{N} \lambda_i = 1$. Using the Jensen covariance $\mathbf{Q}_J$ as the input covariance results in a mutual information value, termed the *Jensen mutual information*, smaller than the channel capacity.

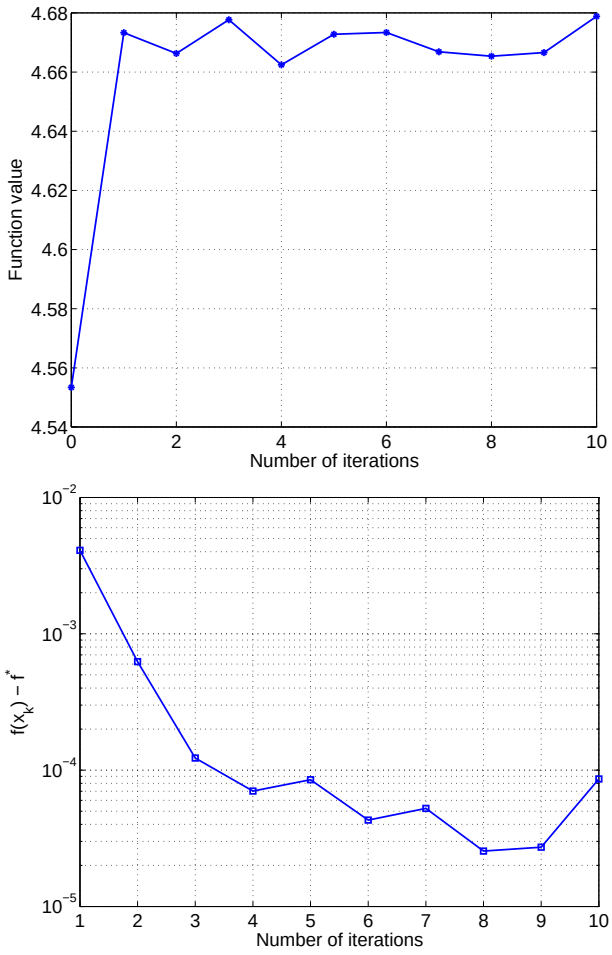

Fig. 2. An optimization example of a 4×2 channel at SNR = 10dB and $K = 0.1$, using 20000 independently generated channel samples in each iteration. Above: mutual information value (nats/Hz); below: its gap to optimal.

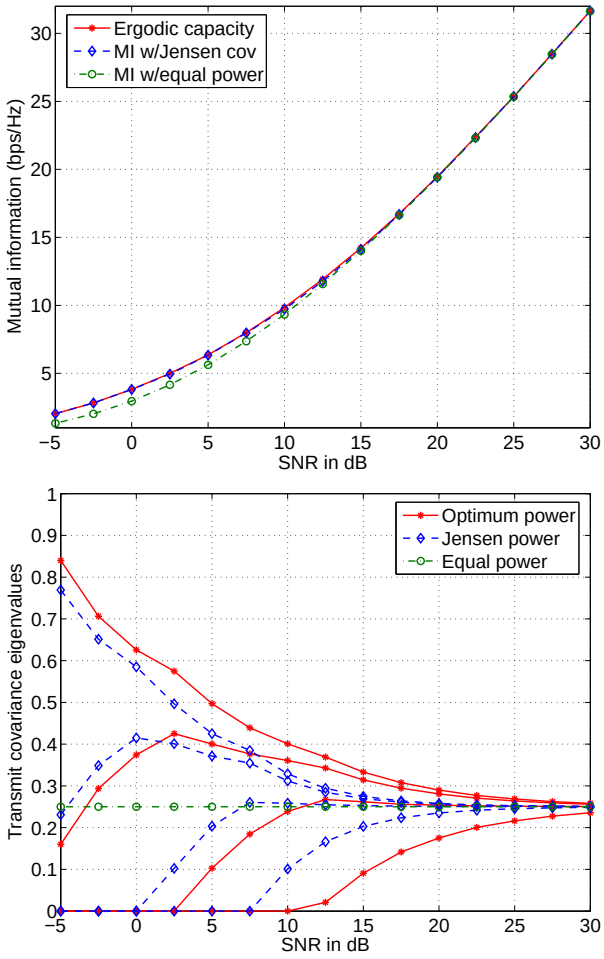

Fig. 3. Capacity and mutual information of a 4×4 system (above) and the corresponding power allocations (below).

5.2. Channel capacity and the Jensen mutual information

Figure 3 shows an example of the channel capacity and the Jensen mutual information with the associated eigenvalues of $\mathbf{Q}$ for a system with equal numbers of transmit and receive antennas, 4×4. The channel has a non-zero mean and a transmit correlation, given in the Appendix. The mutual information with equal power allocation is also included for comparison. The results show that the Jensen mutual information closely approximates the channel capacity at all SNR. Any minor penalty due to the Jensen input covariance occurs only at mid-range SNR. At high SNR, the capacity optimal power allocation approaches equi-power, as does the Jensen solution. Similar observations apply to systems with equal or fewer transmit than receive antennas in general.

For systems with more transmit than receive antenna, the comparison can be different. Figure 4 illustrates the results for a 4×2 channel with the mean and the transmit correlation matrices given in the Appendix. The Jensen mutual information also closely approximates the channel capacity at low SNR. However, at high SNR, the Jensen mutual information exhibits a gap to the capacity. This gap occurs depending on the channel mean and the transmit correlation – a more correlated channel (measured by, for example, a higher condition number of the correlation matrix) will result in a bigger gap. The main reason for the gap at high SNR is the difference in the power allocation. In contrast to the equi-power allocation of the Jensen solution at high SNR, the capacity-optimal input converges to a non-equi-power solution for this channel. These convergence values are still unknown analytically. Asymptotic analyses on the conditions for mode-dropping (resulting in at least one zero-power mode) at high SNR can be found in [7].

5.3. Effects of the K factor

The channel Rician K factor affects the capacity and the gap between the capacity and the Jensen mutual information. Figure 5 shows the capacity versus K at two different SNR values for 4×2 channels with the mean and covariance given in the Appendix. Notice that at a low SNR (e.g. -2dB), the channel capacity is a non-monotonous function of the K factor, and a minimum exists. This effect is partly due to the transmit correlation impact on the capacity: at low K, the correlation impact becomes more dominant, and it is well-known that correlation helps increase the capacity at low SNR. At a higher SNR (e.g. 12 dB), the correlation impact diminishes for channels with a full-rank correlation matrix; provided that the channel mean is also full-rank, the capacity monotonically increases with the K factor.

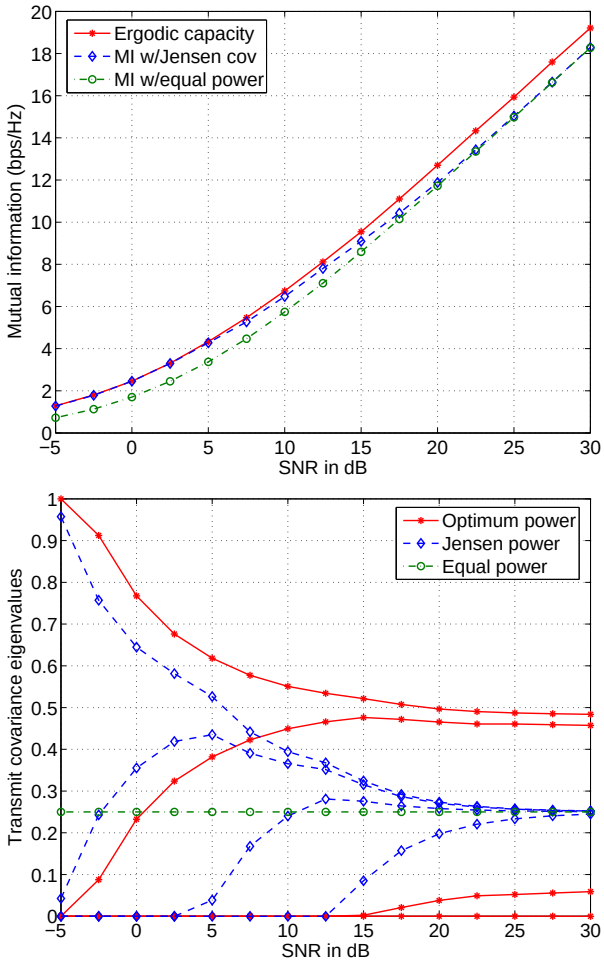

Fig. 4. Capacity and mutual information of a 4×2 system (above) and the corresponding power allocations (below).

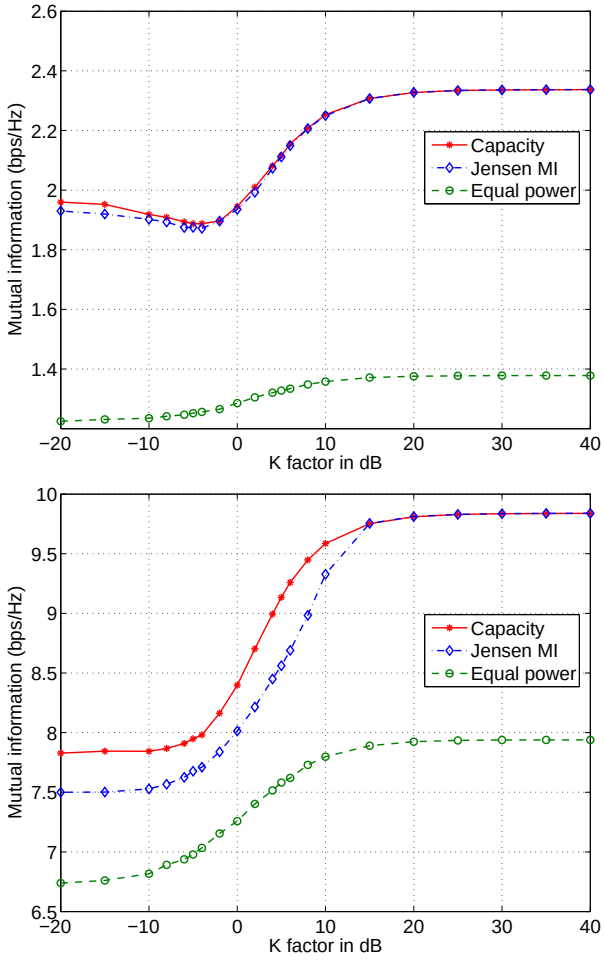

Fig. 5. Channel capacity and mutual information versus the K factor at SNR = -2dB (above) and SNR = 12dB (below).

A higher K factor also causes the SNR point, at which the Jensen mutual information starts diverging from the channel capacity, to increase, implying that the Jensen mutual information closely approximates the capacity for a larger range of SNR. Figure 6 presents this K factor threshold versus the SNR for these 4×2 channels. When K is above this threshold, the Jensen mutual information can be used to accurately approximate the capacity; the difference is less than 0.03 bps/Hz, which is within the numerical precision for optimizing the capacity.

6. CONCLUDING REMARKS

We have implemented convex optimization routines to maximize the mutual information of MIMO wireless channels with a non-zero mean and transmit antenna correlation, using the Newton method and the barrier interior-point method. The gradient, the Hessian, and the objective function values are evaluated using Monte-Carlo simulations. The computational cost grows linearly with the number of channel samples, and quadratically-to-cubically with the number of transmit antennas. Comparisons with the Jensen bound optimizer are performed, illustrating that the Jensen mutual information often approximates the capacity well for systems with equal or fewer transmit than receive antennas. For systems with more transmit than receive antennas, the approximation is accurate at low SNR but diverges at high SNR. This gap at high SNR is due to the difference in power allocation and depends on the transmit correlation and the channel mean, or the K factor.

REFERENCES

[1] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley & Sons, Inc., 1991.

[2] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Info. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[3] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. on Wireless Comm.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[4] S. Venkatesan, S. Simon, and R. Valenzuela, "Capacity of a gaussian MIMO channel with nonzero mean," *Proc. IEEE Vehicular Tech. Conf.*, vol. 3, pp. 1767–1771, Oct. 2003.

[5] D. Hösli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," *Proc. 5th Int'l ITG Conf. on Source and Channel Coding*, Jan. 2004.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2003. [Online]. Available: http://www.stanford.edu/∼boyd/cvxbook.html

[7] M. Vu and A. Paulraj, "Characterizing the capacity for MIMO wireless channels with non-zero mean and transmit covariance," *Proc. Forty-Third Allerton Conf. on Comm., Control, and Comp.*, Sep. 2005.

[8] H. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation and Modulation Theory*. Wiley & Sons, Inc., 2002.

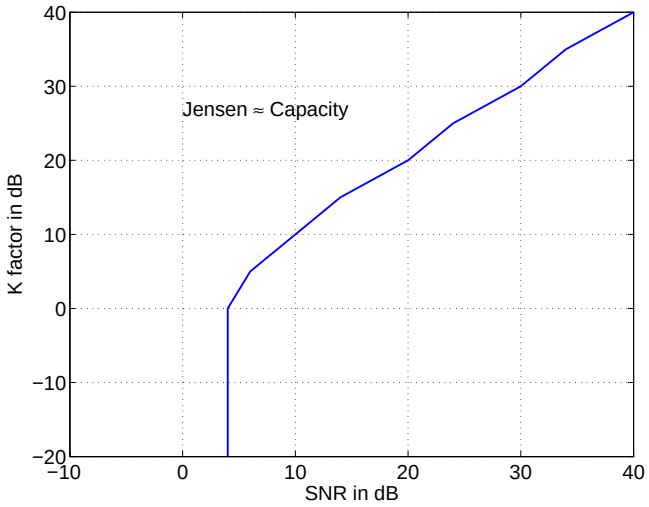

Fig. 6. K factor threshold for tight Jensen mutual information approximation of the capacity of 4×2 channels (difference < 0.03 bps/Hz).

ACKNOWLEDGMENTS

This work was supported in part by the Rambus Stanford Graduate Fellowship and the Intel Foundation PhD Fellowship, and also in part by NSF Contract DMS-0354674-001 and ONR Contract N00014-02-0088.

APPENDIX

All channels used in the simulations are normalized for a constant average power gain of MN (the product of the number of receive and transmit antennas, respectively) as

$$\text{tr}(E[\mathbf{H}^*\mathbf{H}]) = \mathbf{H}_m^*\mathbf{H}_m + M\mathbf{R}_t = MN \ .$$

For the channel model (1), the K factor is defined as

$$K = \frac{||\mathbf{H}_m||_F^2}{M\text{tr}(\mathbf{R}_t)} \ , \tag{13}$$

where $||.||_F$ is the Frobenius norm. Except for the results in Figures 5 and 6 with a varying K factor, all other simulations use $K = 0.1$. The non-normalized transmit correlation matrix is

$$\mathbf{R}_t = \begin{bmatrix} 0.8758 & -0.0993 - 0.0877i & -0.6648 - 0.0087i & 0.5256 - 0.4355i \\ -0.0993 + 0.0877i & 0.9318 & 0.0926 + 0.3776i & -0.5061 - 0.3478i \\ -0.6648 + 0.0087i & 0.0926 - 0.3776i & 1.0544 & -0.6219 + 0.5966i \\ 0.5256 + 0.4355i & -0.5061 + 0.3478i & -0.6219 - 0.5966i & 1.1379 \end{bmatrix} \tag{14}$$

This matrix has the eigenvalues $[2.717\ ,\ 0.997\ ,\ 0.237\ ,\ 0.049]$ and a condition number of 55.5, representing a strong correlation. The non-normalized mean for the 4×2 channel is

$$\mathbf{H}_m = \begin{bmatrix} 0.0749 - 0.1438i & 0.0208 + 0.3040i & -0.3356 + 0.0489i & 0.2573 - 0.0792i \\ 0.0173 - 0.2796i & -0.2336 - 0.2586i & 0.3157 + 0.4079i & 0.1183 + 0.1158i \end{bmatrix} \tag{15}$$

The non-normalized mean for the 4×4 channel is

$$\mathbf{H}_m = \begin{bmatrix} 0.2976 + 0.1177i & 0.1423 + 0.4518i & -0.0190 + 0.1650i & -0.0029 + 0.0634i \\ -0.1688 - 0.0012i & -0.0609 - 0.1267i & 0.2156 - 0.5733i & 0.2214 + 0.2942i \\ 0.0018 - 0.0670i & 0.1164 + 0.0251i & 0.5599 + 0.2400i & 0.0136 - 0.0666i \\ -0.1898 + 0.3095i & 0.1620 - 0.1958i & 0.1272 + 0.0531i & -0.2684 - 0.0323i \end{bmatrix} \tag{16}$$

The parameters used in the optimization programs are

Maximum number of iterations	MAXITER	= 10
Maximum number of line searches each step	MAXLINES	= 50
Barrier method update factor	μ	= 100
Initial barrier multiplying factor	barr_t	= 100
Tolerance ($\Delta x_{\text{nt}}^T \nabla^2 f(x) \Delta x_{\text{nt}} \leq \epsilon$)	ϵ	= 10^{-5}
Number of channel samples in Monte-Carlo simulations	NSAMPLE	= 20000