Journal of Statistical Physics, Vol. 125, Nos. 5/6, December 2006 (© 2006)
DOI: 10.1007/s10955-006-9076-0

Capacity of Multivariate Channels with Multiplicative Noise: Random Matrix Techniques and Large-N Expansions (2)

A. Mayukh Sengupta[1] and Partha Pratim Mitra[2]

Received July 11, 2005; accepted February 21, 2006
Published Online: April 18, 2006

We study memoryless, discrete time, matrix channels with additive white Gaussian noise and input power constraints of the form $Y_i = \sum_j H_{ij} X_j + Z_i$, where Y_i, X_j and Z_i are complex, $i = 1 \ldots m$, $j = 1 \ldots n$, and H is a complex $m \times n$ matrix with some degree of randomness in its entries. The additive Gaussian noise vector is assumed to have uncorrelated entries. Let H be a full matrix (non-sparse) with pairwise correlations between matrix entries of the form $E[H_{ik} H^*_{jl}] = \frac{1}{n} C_{ij} D_{kl}$, where C, D are positive definite Hermitian matrices. Simplicities arise in the limit of large matrix sizes (the so called large-*n* limit) which allow us to obtain several exact expressions relating to the channel capacity. We study the probability distribution of the quantity $f(H) = \log\det(1 + PH^\dagger SH)$. S is non-negative definite and hermitian, with $TrS = n$ and P being the signal power per input channel. Note that the expectation $E[f(H)]$, maximised over S, gives the capacity of the above channel with an input power constraint in the case H is known at the receiver but not at the transmitter. For arbitrary C, D exact expressions are obtained for the expectation and variance of $f(H)$ in the large matrix size limit. For $C = D = I$, where I is the identity matrix, expressions are in addition obtained for the full moment generating function for arbitrary (finite) matrix size in the large signal to noise limit. Finally, we obtain the channel capacity where the channel matrix is partly known and partly unknown and of the form $\alpha I + \beta H$, α, β being known constants and entries of H i.i.d. Gaussian with variance $1/n$. Channels of the form described above are of interest for wireless transmission with multiple antennae and receivers.

KEY WORDS: Random matrix theory, multiple input, multiple output channel, capacity, large-N-matrix field theory

[1] Department of Physics and Astronomy and BioMaPS Institute, Rutgers University, Piscataway, NJ 08854.
[2] Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring Harbor, NY 11724.

1. INTRODUCTION

Shannon capacity, providing a theoretical upper bound on the amount of information that could be transmitted over a channel, is a fundamental quantity in communication theory.[1] Today, the discussion of Shannon capacity of communication channels with additive noise is a substantial part of standard textbook treatment of signal processing and information theory. Channels with multiplicative noise are in general difficult to treat and not many analytical results are known for the channel capacity and optimal input distributions. We borrow techniques from random matrix theory[3] and associated saddle point integration methods in the large matrix size limit to obtain several analytical results for the memoryless discrete-time matrix channel with additive Gaussian noise.

Apart from the intrinsic interest in multiplicative noise, these results are relevant to the study of wireless channels with multiple antennae and/or receivers.[4,5,6] It has been shown that for rich scattering environments, the Shannon capacity, goes up significantly as the number of transmitter antennas and the number of receiver antennas increase. The capacity roughly scales linearly with the smaller of the two numbers. In contrast, if the transmitter cluster and the receiver cluster are far apart with very few scattering objects around, the capacity gain is only logarithmic in the number of antennas. Communication with many antennas transmitting and receiving goes by the name of MIMO (Multiple Input Multiple Output) systems. Development of practical signal processing schemes appropriate for MIMO systems promises very high wireless data transmission rates.[7] Currently wireless networking devices based MIMO technology are available on the market. For a recent overview of the developments, see the article by Gesbert *et al.*[8]

We study communication in the multichannel setting. The channel input-output relationship is defined as

$$Y_i = \sum_{j=1}^{n} H_{ij} X_j + Z_i \tag{1}$$

where all the quantities are in general complex, and $i = 1 \ldots m, j = 1 \ldots n$. Z_i are Gaussian distributed with zero mean and a unity covariance matrix, $E[Z_i Z_j^*] = \delta_{ij}$. Note that this fixes the units for measuring signal power. For most of the paper we employ an overall power constraint

$$\sum_{j=1}^{n} E[|X_j|^2] = nP \tag{2}$$

except in one case where we are able to employ an amplitude (or peak power) constraint. The entries of the matrix H_{ij} are assumed to be chosen from a zero mean Gaussian distribution with covariance matrix

$$E[H_{ik}H_{jl}^*] = \frac{1}{n}C_{ij}D_{kl} \tag{3}$$

Here C, D are positive definite Hermitian matrices. Note that although we assume the distribution of H to be Gaussian, this assumption can be somewhat relaxed without substantially affecting some of the large n results. This kind of universality is expected from known results in random matrix theory.(3) However, for simplicity we do not enter into the related arguments.

We consider the case where C, D are arbitrary positive definite hermitian matrices, as well as the special case where C, D are identity matrices. In either case, one needs to consider the scale of H. Since H multiplies X, we absorb the scale of H into P. The formulae derived in the paper can be converted into more explicit ones exhibiting the scale of H (say h) and the noise variance σ by the simple substitution $P \to Ph^2/\sigma^2$.

A note about our choice of convention regarding scaling with n: We chose to scale the elements of the matrix H_{ij} to be order $1/\sqrt{n}$ and let each signal element X_j be order 1. In the multi-antenna wireless literature, it is common to do the scaling the other way round. In these papers,(4,5) X_j's are scaled as $1/\sqrt{n}$ but keeping H_{ij}'s are kept order 1 so that the average *total* power is P. Our choice of convention is motivated by the fact that we want to treat the systems with channel known at receiver and those with partially unknown channel within the same framework. For reasons that will become clear later, it is convenient for us to keep the scaling of the input space and the output space to be the same, i. e. to keep Y_i, X_j and Z_i all to be order 1 and to scale down H_{ij} to be order $1/\sqrt{n}$. The advantage of this is that the singular values of H happens to be order 1. For the results in the last section, it is convenient that the fluctuating part of the matrix scales this way, in order to have a meaningful result . The final answer for capacity is obviously the same in either convention. While using our results in the context of multiantenna wireless, we just have to remember that the total power, in physical units, is P, and not nP.

In this paper, we discuss two classes of problems. The first class consists of cases where H is known to the receiver but not to the transmitter. H being known to neither corresponds to problems of the second class. The case where H is known to both could be solved by a combination of random matrix techniques used in this paper and the water-filling solution.(4)

As for the first class of problems, we need to maximise the mutual information $I(X, (H, Y))$ over the probability distribution of X subject to the power constraint. Following Telatar's argument,(4) one can show that it is enough to maximise over Gaussian distributions of X, with $E(X) = 0$. Let $E(X_i^* X_j) = PS_{ij}$. $TrS = n$ so that the power constraint is satisfied. S has to be chosen so that $E(I(X, Y|H))$, i.

e. mutual information of X, Y for given H, averaged over different realisations of H, is maximum.

Most of the paper deals with the statistical properties of the quantity

$$f(H) = \log\det(1 + PH^{\dagger}SH) = \sum_{i=1}^{\mathrm{rank}(H)} \log(1 + P\mu_i) \tag{4}$$

where μ_i are the squares of the singular values of the matrix $S^{\frac{1}{2}}H$.

The conditions for optimisation over S are as follows: Let

$$E(H(1 + PH^{\dagger}SH)^{-1}H^{\dagger}) = \Lambda \tag{5}$$

Λ is a nonnegative definite matrix. Then

- S and Λ are simultaneously diagonalizable.
- In the simultaneously diagonalizing basis, let the diagonal elements $S_{ii} = s_i$ and $\Lambda_{ii} = \lambda_i$. Then for all i, such that $s_i > 0$, $\lambda_i = \lambda$.
- For i such that $s_i = 0$, $\lambda_i < \lambda$.

The derivation of these conditions are provided in Appendix A.

2. CHANNEL KNOWN AT THE RECEIVER: ARBITRARY MATRIX SIZE, UNCORRELATED ENTRIES

We start with the simplest case, in which the matrix entries are i.i.d. Gaussian, corresponding to $C = I$, $D = I$. In this case, one obtains $S = I$ for the capacity achieving distribution.[4] In this case, the joint probability density of the singular values of H is explicitly known to be given by[3]

$$P(\mu_1, \ldots, \mu_{\min(m,n)}) = \frac{1}{\mathcal{Z}} \prod_{i<j} (\mu_i - \mu_j)^2 \prod_i \mu_i^{|m-n|} e^{-n\sum_i \mu_i} \tag{6}$$

where the normalisation constant can be obtained as a consequence of the Selberg integral formula ([3], Pg. 354, Eq. 17.6.5)

$$\mathcal{Z} = \prod_{j=1}^{\min(n,m)} \Gamma(j)\Gamma(|m-n| + j) \tag{7}$$

In the following, we assume (without loss of generality) $\min(n, m) = n$.

This form has been utilised before to obtain the expectation of $f(H)$ in terms of integrals over Laguerre polynomials.[4] However, it is also fairly straightforward to obtain the full moment generating function (and hence the probability density) of $f(H)$, particularly at large P. Consider the moment generating function $F(\alpha)$

of the random variable $f(H)$, given by

$$F(\alpha) = E[\exp(\alpha f(H))] = E[\prod_i (1 + P\mu_i)^\alpha] \tag{8}$$

2.1. Large P Limit

In the limit of large P, the expectation can be simply computed as an application of the integral formula stated above. Note that the large P limit is obtained when P is much larger than the inverse of the typical smallest eigenvalue. For the case $m = n$, this would require that $P >> n$, whereas if $m/n = \beta > 1$, then we require $P >> (\sqrt{\beta} - 1)^{-1}$. Taking the large P limit, we obtain

$$F(\alpha) \approx (P)^{\alpha n} E\left[\prod_i \mu_i^\alpha\right] \tag{9}$$

$$E[\prod_i \mu_i^\alpha] = \prod_{j=1}^{n} \frac{\Gamma(\alpha + |m-n| + j)}{\Gamma(|m-n| + j)} \tag{10}$$

In this limit, it follows that the expectation of $f(H)$ is given by

$$E[f(H)] \approx n\log(P) + \sum_{j=1}^{n} \psi(m - n + j) - n\log(n) \tag{11}$$

and the variance, by

$$Var[f(H)] \approx \sum_{j=1}^{n} \psi'(|m-n| + j) \tag{12}$$

where $\psi(j) = \Gamma'(j)/\Gamma(j)$. Setting $m/n = \beta$ and for large n, we get

$$E[f(H)] \approx n\log(\beta P/e) \tag{13}$$

For $\beta > 1$ and large n,

$$Var[f(H)] \approx \log\left(\frac{m}{m-n}\right) = \log\left(\frac{\beta}{\beta - 1}\right) \tag{14}$$

For $\beta = 1$ and large $m(= n)$,

$$Var[f(H)] \approx \log(m) + 1 + \gamma \tag{15}$$

where γ is the Euler-Mascheroni constant.

Laplace transforming the moment generating function, one obtains the probability density of $\mathcal{C} = f(H)$. In the large P limit, the probability density is therefore

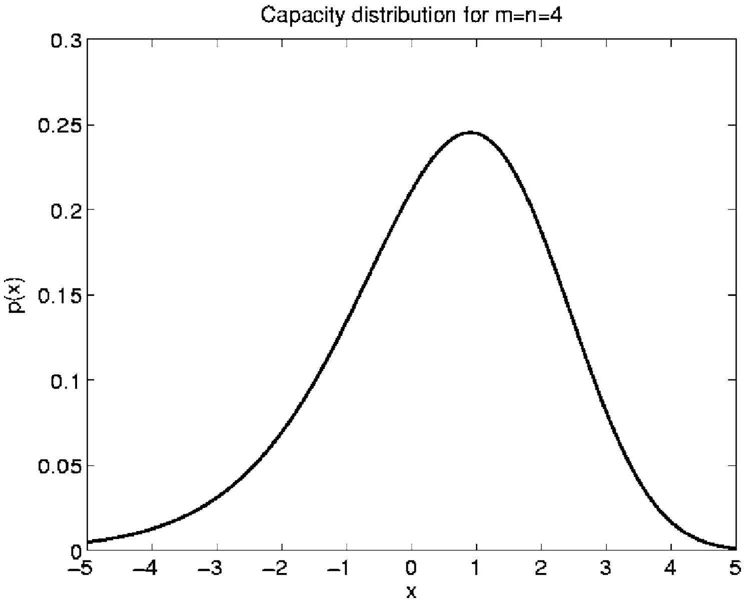

Fig. 1. The probability density function of $f(H)$ is given for $m = n = 4$ in the limit of large P. The origin is shifted to the value $4\log(P/e)$.

given by $p(\mathcal{C} - n\log(P/e))$ where $p(x)$ is given by

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\alpha e^{-i\alpha n(\log(n)-1)-ix\alpha} \prod_{j=1}^{n} \frac{\Gamma(i\alpha + |m-n| + j)}{\Gamma(|m-n| + j)} \tag{16}$$

An example of $p(x)$ is presented in Fig. 1 for $m = n = 4$.

2.2. Arbitrary P

For arbitrary P, $F(\alpha)$ does not simplify as above, but can nevertheless be written in terms of an $n \times n$ determinant as follows:

$$F(\alpha) = \frac{\det M(\alpha)}{\det M(0)} \tag{17}$$

where the entries of the complex matrix M are given by $(i, j = 1 \ldots n)$

$$M_{ij}(\alpha) = \int_0^{\infty} d\mu (1 + P\mu)^{\alpha} \mu^{i+j+|m-n|-2} e^{-n\mu} \tag{18}$$

To obtain this expression for $F(\alpha)$, one has to simply express the quantity $\prod_{i\neq j}(\mu_i - \mu_j)$ as a Vandermonde determinant and perform the integrals in the resultant sum. The integral can be expressed in terms of a Whittaker function (related to degenerate Hypergeometric functions), and can be evaluated rapidly, so that for small values of m, n this provides a reasonable procedure for numerical evaluation of the probability distribution of $f(H)$.

3. CHANNEL KNOWN AT THE RECEIVER: LARGE MATRIX SIZE, CORRELATED ENTRIES

For the more general case of correlations between matrix entries as in Eq. 3, the matrix ensemble is no longer invariant under rotations of H, so that the eigenvalue distribution used in the earlier section is no longer valid. However, by using saddle point integration,(9) it is still possible to compute the expectation and variance of $f(H)$ in the limit of large matrix sizes. In this section, we simply state the results for the expectation and variance, and explore the consequences of the formulae obtained. The saddle point method used to obtain these results was used in an earlier paper to obtain the singular value density of random matrices(9) and is described in Appendix B .

The expectation and variance of $f(H)$ are given in terms of the following equations:

$$E[f(H)] = \sum_{i=1}^{m} \log(w + \xi_i r) + \sum_{j=1}^{n} \log(w + \eta_j q) - nqr - (m+n)\log(w) \tag{19}$$

$$Var[f(H)] = -2\log|1 - g(r,q)| \tag{20}$$

where

$$w^2 = \frac{1}{P} \tag{21}$$

$$g(r,q) = \left[\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\eta_j}{w + \eta_j q}\right)^2\right]\left[\frac{1}{n}\sum_{j=1}^{m}\left(\frac{\xi_j}{w + \xi_j r}\right)^2\right] \tag{22}$$

In the above equations, ξ, η denote the eigenvalues of the matrices $\tilde{C} = S^{\frac{1}{2}} C S^{\frac{1}{2}}$, D respectively. The numbers r, q are determined by the equations

$$r = \frac{1}{n}\sum_{j=1}^{n}\frac{\eta_j}{w + \eta_j q} \tag{23}$$

$$q = \frac{1}{n}\sum_{j=1}^{m}\frac{\xi_j}{w + \xi_j r} \tag{24}$$

These equations are expected to be valid in the limit of large m, n assuming that a sufficient number of the eigenvalues ξ, η remain nonzero. These equations could be used to design optimal multi-antenna systems.[10]

4. CALCULATING CAPACITY

In this section we provide the step by step procedure for calculating capacity using the results from the previous sections. One can show that the optimal covariance matrix S and the matrix C could be diagonalized together (Appendix C). Let us work in the diagonalizing basis. Define $\tilde{C}$ as before. This is a diagonal matrix in this basis, with diagonal elements $\xi_i = c_i s_i$, where c_i, s_i are the diagonal elements of C, S respectively. We assume that c_i's are sorted in decreasing order. That is, $c_1 > c_2 > \cdots > c_m$. The optimality condition, Equation 5, becomes:

$$\frac{c_i r}{w + c_i s_i r} = \lambda, \text{ for } i = 1, \ldots, p. \tag{25}$$

p is the number for nonzero s_i's. One way to see this is as follows: Take the expression in Eq. 19, replace ξ by $c_i s_i$ and take its derivative with respect to non-zero s_i's. Note that q, r changes as ξ_i changes. However, this expression is evaluated at a point which is stationary with respect to variation in q and r. Hence, to first order, changes of q, r due to changes in ξ do not have a contribution. We just change ξ keeping q, r fixed. Since $\partial\xi_i/\partial s_i = c_i$, we got the expression in Eq. 25.

Eq. 25, along with Eq. 23 and Eq. 24, provide $p + 2$ equations for $p + 3$ unknowns, namely r, q and s_i, $i = 1, \ldots, p$. The additional condition comes from total power constraint $\sum_i s_i = P$. Once we find such a solution, we could check whether the conditions $s_i > 0$ and $\lambda_i = c_i r/w < \lambda$ is satisfied for all $i > p$. If any of them is not satisfied, we need to change p, the number of non-zero eigenvalues of S. After getting a consistent set of solutions we use Eq. 19 to calculate capacity.

Schematically, the algorithm is as follows:

1. Diagonalize C and arrange eigenvalues in the decreasing order along the diagonal.
2. Start with $p = 1$.
3. Solve equations 25,23,24 along with the power constraint.
4. Check whether $s_i > 0$ for $i = 1, .., p$, and, $c_{p+1} r/w < \lambda$.
5. If any of the previous conditions are not satisfied, go back to step 3 with p incremented by 1. Otherwise, proceed to next step.
6. Calculate capacity using Eq. 19.

5. CHANNEL KNOWN AT THE RECEIVER: LARGE MATRIX SIZE, UNCORRELATED ENTRIES

The results of the previous section simplify if we assume that the matrix entries are uncorrelated with unit variance. In this case, the equations become

$$E[f(H)] = m\log(w+r) + n\log(w+q) - nqr - (m+n)\log(w) \tag{26}$$

$$Var[f(H)] = -2\log\left|1 - \frac{1}{(w+q)^2}\frac{\beta}{(w+r)^2}\right| \tag{27}$$

$$r = \frac{1}{w+q} \tag{28}$$

$$q = \frac{\beta}{w+r} \tag{29}$$

First, consider the special case where $m = n$. In this case, we obtain

$$E[f(H)] = n\left[\log\left(\frac{P}{e}\right) + \log\left(1+\frac{1}{x}\right) + \frac{x}{P}\right] \tag{30}$$

$$Var[f(H)] = 2\log\left(\frac{(1+x)^2}{(2x+1)}\right) \tag{31}$$

where $x^2 + x = P$ (x positive). For large P, the expectation and variance tend to $n\log(P/e)$ and $\log(P)$ respectively. Note that the variance grows logarithmically with power, but does not depend on the number of channels.

For m, n not equal, one obtains expressions which are analogous by solving the simultaneous equations above for q and r (which lead to quadratic equations for either q or r by elimination of the other variable):

$$r(w) = \frac{-(w^2+m-n)+\Delta}{2w} \tag{32}$$

$$q(w) = \frac{-(w^2-m+n)+\Delta}{2w} \tag{33}$$

$$\Delta = \sqrt{(w^2+m+n)^2 - 4mn} \tag{34}$$

Substituting these formulae in Eq. 26 and Eq. 27 gives the desired expressions for the expectation and variance of the capacity $f(H)$.

6. H UNKNOWN AT BOTH RECEIVER, TRANSMITTER: LARGE MATRIX SIZE, UNCORRELATED ENTRIES

The case where H is unknown both to the transmitter and receiver is in general hard.[6] For example, analytical formulae for the capacity are not available even in the scalar case. However, in the case that the matrix entries are uncorrelated, the problem reduces to an effective scalar problem which exhibits simple behaviour at large m. To proceed, one first obtains the conditional distribution $p(\vec{Y}|\vec{X})$. This can be done by noting that for fixed $\vec{X}$, $\vec{Y}$ is a linear superposition of zero mean Gaussian variables and is itself Gaussian with zero mean and variance given by

$$E[Y_i Y_j^*] = \left(1 + \frac{1}{n}\sum_k |X_k|^2\right)\delta_{ij} \tag{35}$$

Note that only the magnitude of the vector $\vec{X}$ enters into the equation, and the distribution of $\vec{Y}$ is isotropic. Effectively, since the transfer matrix is unknown both at the transmitter and receiver, only magnitude information and no angular information can be transmitted. Since we are free to choose the input distribution of $x = |\vec{X}|/\sqrt{n}$, we can henceforth regard x as a positive scalar variable. As for $y = |\vec{Y}|/\sqrt{m}$ ($\sqrt{m}$ is just to arrange the right scaling),we still have to keep track of the phase space factor y^{2m-1} which comes from transforming to $2m$ dimensional polar coordinates. Note that we need $2m$ dimensions since $\vec{Y}$ is a complex vector. Thus, the problem can be treated as if it were a scalar channel, keeping track only of the magnitudes y and x, except that the measure for integration over y should be $d\mu(y) = \Omega_{2m} y^{2m-1} dy$ where Ω_{2m} is from the angular integral. The conditional probability $p(y|x)$ is given by

$$p(y|x) = \left[\frac{m}{\pi(1+x^2)}\right]^m \exp\left(-\frac{my^2}{2(1+x^2)}\right) \tag{36}$$

The conditional entropy of y given x is easy to compute from the original obervation that the conditional distribution is Gaussian, and is given by

$$H(y|x) = mE_x\left[\log\left(\frac{\pi e}{m}(1+x^2)\right)\right] \tag{37}$$

The entropy of the output is

$$H(y) = -E_x \int d\mu(y) p(y|x) \log(E_{x'} p(y|x')) \tag{38}$$

Thus, the mutual information between input and output is given by subtracting the two expressions above and rearranging terms:

$$I = -E_x \int d\mu(y) p(y|x) \log\left(E_{x'}\left[\left(\frac{1+x^2}{1+x'^2}\right)^m \exp\left(-\frac{my^2}{(1+x'^2)} + m\right)\right]\right) \tag{39}$$

The y integral contains the factor

$$y^{2m-1} \exp\left(-\frac{my^2}{(1+x^2)}\right) \tag{40}$$

which is sharply peaked around $y^2 = (1 + x^2)$ for m large. Thus, the y integral can be evaluated using Laplace's method to obtain (for m large)

$$I \approx -E_x \log E_{x'} \left[\left(\frac{1+x^2}{1+x'^2}\right)^m \exp\left(-m\frac{(1+x^2)}{(1+x'^2)} + m\right)\right] \tag{41}$$

Applying Laplace's method again to perform the integral inside the logarithm, assuming that the distribution over x is given by a continuous function $p(x)$, we finally obtain

$$I = \frac{1}{2}\log\left(\frac{2m}{\pi}\right) + \int dxp(x)\log\left[\frac{x}{1+x^2}\frac{1}{p(x)}\right] \tag{42}$$

The capacity and optimal input distribution is straightforwardly obtained by maximising the above. It is easier to treat the case where a peak power constraint is used, namely $x \le \sqrt{P}$. In this case, the optimal input distribution is ($x \in [0, \sqrt{P}]$)

$$p(x) = \frac{1}{\log(1+P)}\frac{2x}{1+x^2} \tag{43}$$

and the channel capacity is

$$\mathcal{C} = \frac{1}{2}\log\left(\frac{m}{2\pi}\right) + \log(\log(1+P)) \tag{44}$$

Notice that the capacity still grows with m, which is somewhat surprising, but this growth is only logarithmic. Secondly, the dependence on the peak power is through a double logarithm.

With an average power constraint $\int x^2 dxp(x) = P$ the optimal input distribution is given by

$$p(x) = a\frac{2x}{1+x^2}e^{-\frac{x^2}{a(1+P)}} \tag{45}$$

where a is a constraint determined by the normalisation condition, which yields the equation

$$a = \int_0^\infty \frac{dy}{1+y}e^{-\frac{y}{a(1+P)}} \tag{46}$$

The capacity is given by

$$\mathcal{C} = \frac{1}{2}\log\left(\frac{m}{2\pi}\right) + \log(a) + \frac{P}{1+P}\frac{1}{a} \tag{47}$$

For large P, $a \approx \log(1+P)$, thus recovering the double logarithm behaviour.

7. INFORMATION LOSS DUE TO MULTIPLICATIVE NOISE

We could generalize the calculation in the previous section to a problem which interpolates smoothly between usual additive noise channel and the case considered above. This is a problem with same number of transmitters and receivers ($m = n$) and is defined by

$$Y_i = \sum_{j=1}^{n} (\alpha\delta_{ij} + \beta H_{ij}) X_j + Z_i \tag{48}$$

$\beta = 0$ is the usual channel with additive gaussian noise. $\alpha = 0$ corresponds the problem we have just discussed. In the first case, capacity increases logarithmically with input power, whereas in the second case it has a much slower (double logarithmic) dependence on input power. Apart from the theoretical interest in studying the crossover between these two kinds of behavior, this problem has much practical importance.[11]

The easy thing to calculate is $c = \lim_{n\to\infty} \mathcal{C}/n$. Notice that this quantity is zero in the limit $\alpha \to 0$, capacity being logarithmic in n in that limit. For simplicity, we choose the input power constraint $\sum_i |X_i|^2 \leq nP$. We relegate the details of the saddle point calculation to Appendix D. The result is

$$c = \log\left[1 + \frac{\alpha^2 P}{1 + \beta^2 P}\right] \tag{49}$$

The result tells us that, in the large N limit, the effect of multiplicative noise is similar to that if an additive noise whose strength increases with the input power.

It is of particular interest to note that there exists a lower bound to the channel capacity, which is given by the capacity of a fictitious additive gaussian channel with the same covariance matrix for $(\vec{X}, \vec{Y})$ as the channel in question.[11] Remarkably, this bound coincides with the saddle point answer.

8. DISCUSSIONS

This papers discusses how saddle point methods could be useful in dealing with many multichannel communication scenarios with different degrees of knowledge of the transfer matrix. Our original manuscript was first uploaded to arXiv$_{(2)}$ on October 2000. In the intervening years, this has been a fairly active field of research. A number of papers have appeared on the subject, several of which cite the arxiv version of the preprint. Rather than refer to all these relevant papers, it might serve our readers better to provide some idea of different research directions in which our tools could be used and have been used.

The most important application has been to analyze the role of correlation transfer matrix in MIMO systems. It turns out that the geometry of the antenna cluster determines the nature of correlations and hence our insights could be used to optimize MIMO antenna design.[10,12] Other applications involve how knowledge of the channel could be used to optimize the transmitting system.[13]

Another area of research benefitting from this work is the study of fluctuating, partially unknown, channel. As we showed, one could use our methods to model the crossover between the fixed random channel to the completely unspecified channel (with only some power constraints). This model has been used as an approximation to understand the limits of data transfer rates in nonlinear optical fibers.[11,14] The channels in this case correspond to the different frequencies and the nontrivial channel mixing is due to non-linear interactions.

Finally, one possible set of applications of this method could be to nonstationary MIMO channels that are correlated in time. Some preliminary results in this direction come from studies in which the channel keeps changing to something totally uncorrelated after a fixed interval of time (6, 15). See also the work of Abdi and Kaveh[16] on more realistic models of space-time correlation in mobile fading channels. A systematic study of this problem would teach us about the crossover between long correlation time and extremely short correlation times: another way to go from frozen channels to unknown ones.

APPENDIX A

The condition of optimality with respect to S is

$$E[Tr\{(1 + PH^{\dagger}SH)^{-1}H^{\dagger}\delta SH\}] = Tr(\Lambda\delta S) \leq 0 \tag{50}$$

for all allowed small δS. δS has to satisfy two conditions: that $S + \delta S$ is non-negative definite and that $Tr(\delta S) = 0$. The matrix Λ has been defined in the first section. It is a non-negative definite hermitian matrix.

If S has only positive eigenvalues then adding a small enough hermitian δS to it does not make any of the eigenvalues zero or negative. Then only way the optimisation condition can be satisfied is by choosing Λ to be proportional to the Identity matrix. This can be seen as follows: for $\Lambda = \lambda I$, $Tr\Lambda\delta S = \lambda Tr\delta S = 0$. If $\Lambda \neq \lambda I$, then, in general, $Tr\Lambda\delta S \neq 0$ even though $\delta S = 0$, and can therefore be chosen to be positive.

What if S has few zero eigenvalues? Let us choose a basis so that S is diagonal. The eigenvalue of S s_i are ordered so that $s_1, \ldots, s_k$ are positive and $s_i = 0$ for $i > k$. We could choose δS_{ij} to be non zero only for $1 \leq i, j \leq k$ and repeating the argument of the last paragraph, $\Lambda_{ij} = \lambda\delta_{ij}$, for $1 \leq i, j \leq k$. In fact, even if we choose δS_{ij} to be nonzero for $i \leq k < j$, and $j \leq k < i$ we do not violate, to first order in δS, non negativity of eigenvalues of $S + \delta S$. This would give us $\Lambda_{ij} = 0$ for $i \leq k < j$ and $j \leq k < i$. Hence Λ is of block-diagonal form. The $k \times k$ block is already constrained to be proportional to Identity matrix. We would now constrain the other block of Λ which is of size $(n - k) \times (n - k)$.

Since the last $n-k$ eigenvectors of S correspond to zero eigenvalues, we are free to rotate them among each other. Using this freedom, we diagonalise the lower $(n-k)\times(n-k)$ block of Λ. Choosing diagonal δS_{ij} with with negative values for $i=j\le k$ but positive values $i=j>k$, and satisfying $Tr(\delta S)=0$, we can show that the last $n-k$ eigenvalues of Λ are smaller than or equal to λ.

APPENDIX B

In this section, it is assumed without loss of generality that $m\ge n$. We consider first the case $S=I$, but derive the results for arbitrary C, D. It is easy to recover the results for general S by making the transformation $H\to S^{\frac{1}{2}}H$ and $C\to S^{\frac{1}{2}}CS^{\frac{1}{2}}$.

We start from the identity

$$\det([w\; iH\;;\; -iH^\dagger\; w])^{-\alpha} = \int d\mu(X)d\mu(Y) \times \exp\left(-\frac{1}{2}\sum_{a=1}^{\alpha}[w(Y_a^\dagger Y_a + X_a^\dagger X_a) + i(Y_a^\dagger H X_a - X_a^\dagger H^\dagger Y_a)]\right) \tag{51}$$

where

$$d\mu(X) = \prod_{i=1}^{n}\prod_{a=1}^{\alpha}\frac{dX_{ia}^R dX_{ia}^I}{2\pi} \tag{52}$$

with R, I denoting real and imaginary parts respectively. $d\mu(Y)$ is defined analogously. The introduction of multiple copies of the Gaussian integration is the well known 'replica trick'.[9] This allows us to compute $f(H)$, since it is easily verified that

$$\det([w\; iH\;;\; -iH^\dagger\; w])^{-\alpha} = w^{-(m+n)\alpha}e^{-\alpha f(H)} \tag{53}$$

where we have set $w^2=n/P$. The moment generating function of $f(H)$ can be obtained by studying the expectation of the determinant above with respect to the probability distribution of H. We therefore obtain for the moment generating function

$$F(-\alpha) = w^{(m+n)\alpha}\int d\mu(X)d\mu(Y)\exp\left(-\frac{1}{2}\left[w\sum_{a=1}^{\alpha}(Y_a^\dagger Y_a + X_a^\dagger X_a) + \frac{1}{2n}\sum_{a,b=1}^{\alpha}(Y_a^\dagger C Y_b X_b^\dagger D X_a)\right]\right) \tag{54}$$

The last term in the exponent can be decoupled by introducing the $\alpha \times \alpha$ complex matrices P, Q with contour integrals over the matrix entries in the complex plane to obtain

$$F(-\alpha) = w^{(m+n)\alpha} \int d\mu(X) d\mu(Y) d\mu(R) d\mu(Q) \exp\left(-\frac{1}{2}S\right) \tag{55}$$

where

$$S = w \sum_{a=1}^{\alpha} (Y_a^\dagger Y + X_a^\dagger X) + \sum_{a,b=1}^{\alpha} (Y_a^\dagger C Y_b R_{ab} + Q_{ab} X_a^\dagger D X_b - n R_{ab} Q_{ba}) \tag{56}$$

$$d\mu(R) d\mu(Q) = \prod_{a,b=1}^{\alpha} \frac{d R_{ab} d Q_{ab}}{2\pi} \tag{57}$$

The R, Q integrals, in contrast with the X, Y integrals, are complex integrals along appropriate contours in the complex plain. For example, if the Q_{ij} integrals are along the imaginary axis, so that the Q integrals give rise to delta functions which can then be integrated over R to obtain the above equation. The integrals over X, Y can now be performed to obtain

$$\begin{aligned} F(-\alpha) = w^{(m+n)\alpha} \int d\mu(R) d\mu(Q) \exp(&-\log(\det(w + CR)) \\ &- \log(\det(w + DQ)) + n Tr(RQ)) \end{aligned} \tag{58}$$

where CR and DQ are understood to be outer products of the matrices. Introducing the eigenvalues ξ, η of C, D the exponent may be written as

$$\sum_{i=1}^{m} \log(\det(w + \xi_i R)) + \sum_{j=1}^{n} \log(\det(w + \eta_j Q)) - n Tr(RQ) \tag{59}$$

If m, n become large and the number of non-zero ξ_i, η_i grow linearly with m, n, then we can perform the R, Q integrals using saddle point methods. If we assume that at the saddle point the matrices R, Q do not break the replica symmetry , i.e $R = rI$, $Q = qI$ where I is the identity matrix, then the saddle point equations are $\partial\mathcal{C}/\partial r = \partial\mathcal{C}/\partial q = 0$, where $\mathcal{C}$ is defined below, leading to

$$r = \frac{1}{n} \sum_{j=1}^{n} \frac{\eta_j}{w + \eta_j q} \tag{60}$$

$$q = \frac{1}{n} \sum_{j=1}^{m} \frac{\xi_j}{w + \xi_j r} \tag{61}$$

Expanding the exponent upto quadratic order around the saddle point and performing the resulting Gaussian integral, we obtain

$$F(\alpha) = \exp\left(\alpha\mathcal{C}(r,q) + \frac{\alpha^2}{2}\mathcal{V}(r,q)\right) \tag{62}$$

$$\mathcal{C}(r,q) = \sum_{i=1}^{m}\log(w+\xi_i r) + \sum_{j=1}^{n}\log(w+\eta_j q) - nqr - (m+n)\log(w) \tag{63}$$

$$\mathcal{V}(r,q) = -2\log|1-g(r,q)| \tag{64}$$

$$g(r,q) = \left[\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\eta_j}{w+\eta_j q}\right)^2\right]\left[\frac{1}{n}\sum_{j=1}^{m}\left(\frac{\xi_j}{w+\xi_j r}\right)^2\right] \tag{65}$$

Since $F(\alpha)$ is the moment generating function for $f(H)$, the expressions for $\mathcal{C}$, V give the expressions for the expectation and variance of $f(H)$, as presented in Section (3).

Note that these results could be obtained by many other means. We use the replica symmetric ansatz as a quick way to obtain the equations. An alternative way of arriving at these equations is by re-summing planar diagrams.[9] The replica trick here is a just a way of organizing the perturbation theory in power P. As long as $\mathcal{V}$ remains real and positive, one does not run into any inconsistencies. Most of the papers dealing with more rigorous methods tend to focus on the case with uncorrelated entries. For a treatment of the case with correlated entry, directly dealing with the matrix integrals, see the recent paper by Simon and Moustakas.[18]

APPENDIX C

We showed that S and Λ are simultaneously diagonalizable, and, therefore, commute with each other. Doing the transformation $\tilde{H} = C^{-\frac{1}{2}}H$ and $\tilde{S} = C^{\frac{1}{2}}SC^{\frac{1}{2}}$, and following the same trend of argument, one shows that $\tilde{S} = C^{\frac{1}{2}}SC^{\frac{1}{2}}$ and $\tilde{\Lambda} = C^{-\frac{1}{2}}\Lambda C^{-\frac{1}{2}}$ are simultaneously diagonalizable as well. Then

$$C^{\frac{1}{2}}S\Lambda C^{-\frac{1}{2}} = \tilde{S}\tilde{\Lambda} = \tilde{\Lambda}\tilde{S} = C^{-\frac{1}{2}}\Lambda SC^{\frac{1}{2}} \tag{66}$$

or

$$CS\Lambda = S\Lambda C \tag{67}$$

Now use the basis where S and Λ are simultaneously diagonalized. Eqn. 67 leads to $C_{ij}(s_i\lambda_i - s_j\lambda_j) = 0$. Remebering the condition for optimality of S and Λ, $(s_i\lambda_i - s_j\lambda_j) = 0$ only if $s_i = s_j$. So $C_{ij} \neq 0$ only if $s_i = s_j$. If there is subspace V, with dimension more than 1, spanned by all eigenvectors of S with the eigenvalue

s, one could choose a new basis in the subspace V to make C diagonal in that subspace/block, while keeping S diagonal. Continuing this process, one can find a basis where both S and C are simultaneously diagonal.

APPENDIX D

In this case,

$$P(\vec{Y}|\vec{X}) = \frac{1}{[\pi(1+\beta^2|X|^2)]^n} e^{-\frac{|\vec{Y}-\alpha\vec{X}|^2}{(1+\beta^2|X|^2/n)}} \tag{68}$$

Let us redefine $\vec{x} = \vec{X}$ and $\vec{y} = \vec{Y}/\sqrt{n}$. The optimal probability distribution of $\vec{x}$ depends only on its norm $x = |\vec{x}|/\sqrt{n}$. Let $q(x)$ to be the probability distribution of x.

Once more,

$$H(\vec{y}|\vec{x}) = E_{\vec{x}}[n\log(\pi e(1+\beta^2x^2)/n)] = n\int dx q(x)\log\left[\frac{\pi e}{n}(1+\beta^2x^2)\right] \tag{69}$$

However,

$$p(\vec{y}) = E_{\vec{x}}\,[p(\vec{y}|\vec{x})] \approx \int dx q(x)\frac{n^n}{[\pi(1+\beta^2x^2)]^n} e^{-\frac{n(y^2+\alpha^2x^2)}{(1+\beta^2x^2)}+2n\phi\left(\frac{\alpha xy}{1+\beta^2x^2}\right)} \tag{70}$$

where

$$\phi(a) = \lim_{d\to\infty}\frac{1}{d}\log\left[\frac{\int_0^\pi d\theta\,\sin^{d-2}(\theta)e^{da\cos(\theta)}}{\int_0^\pi d\theta\,\sin^{d-2}(\theta)}\right] \tag{71}$$

Saddle point evaluation of $\phi(a)$ (which is equivalent to doing an expansion of the Bessel functions $I_\nu(z)$ with large order ν and large argument z, but the ratio z/ν held fixed) gives

$$\phi(a) = a\cos\theta(a) + \log\sin\theta(a) \tag{72}$$

$$\cos\theta(a) = a\sin^2\theta(a) \tag{73}$$

In fact we would need $d\phi(a)/da$.

$$\frac{d\phi(a)}{da} = \cos\theta(a) = \frac{\sqrt{1+4a^2}-1}{2a} \tag{74}$$

Variation of $H(\vec{y}) = \int d\vec{y}p(\vec{y})\log\frac{1}{p(\vec{y})}$ with respect to $q(x)$ produces

$$\frac{\delta H(\vec{y})}{\delta q(x)} = -\int d\vec{y}p(\vec{y}|x)(1+\log p(\vec{y})) \tag{75}$$

where

$$p(\vec{y}|x) = \left[\frac{n}{\pi(1+\beta^2x^2)}\right]^n \exp(-nf(x,y)) = p(y|x) \tag{76}$$

and

$$f(y,x) = \frac{y^2+\alpha^2x^2}{(1+\beta^2x^2)} - 2\phi(\frac{\alpha xy}{1+\beta^2x^2}) \tag{77}$$

Now we can do the $\vec{y}$ integral in Eq. 75 by the saddle point method. After going over to polar coordinates and doing some straightforward calculations, we find that the integral peaks at $y = y(x)$ given by

$$y(x)^2 = (1+(\alpha^2+\beta^2)x^2) \tag{78}$$

This is expected, as variance of $\vec{y}$ given a uniform angular distribution of $\vec{x}$ with a fixed norm x is the right hand side of (78). On the other hand, the variance is $y(x)^2$ in the saddle point approximation.

Thus finally, we have the condition for the stationarity of the mutual information,

$$-\mathcal{C} = \log\int dx' q(x') p(y(x)|x') + n\log\left[\frac{\pi e}{n}(1+\beta^2x^2)\right] \tag{79}$$

where $\mathcal{C}$ is a constant, which turns out to be the channel capacity. The constant is fixed by the condition that $q(x)$ is a normalised probability distribution. This condition, along with the fact $\int d\vec{y}\, p(y|x) = \Omega_{2n}\int dy y^{2n-1} p(y|x) = 1$, $\Omega_{2n} = 2\pi^n/\Gamma(n)$, can be used to determine C.

$$1 = \Omega_{2n}\int dx y'(x) y(x)^{2n-1}\int dx' q(x') p(y(x)|x') \tag{80}$$

$$= e^{-\mathcal{C}}\Omega_{2n}\int_0^{\sqrt{P}} dx \left[\frac{n}{\pi e(1+\beta^2x^2)}\right]^n \frac{y'(x)}{y(x)} y(x)^{2n} \tag{81}$$

$$\approx e^{-\mathcal{C}}\sqrt{\frac{2n}{\pi}}\int_0^{\sqrt{P}} dx \frac{y'(x)}{y(x)}\left[\frac{y(x)^2}{(1+\beta^2x^2)}\right]^n \tag{82}$$

For any $\alpha > 0$,

$$f(x) = \log\left[\frac{y(x)^2}{(1+\beta^2x^2)}\right] = \log\left[\frac{1+(\alpha^2+\beta^2)x^2}{1+\beta^2x^2}\right] \tag{83}$$

is a monotonically increasing function of x, for positive x. Hence the last integral is dominated by the contribution from the region near the upper limit. For a monotonically increasing function $f(x)$,

$$\int_0^z g(x)\exp(nf(x)) \approx \frac{g(z)\exp(nf(z))}{nf'(z)}. \tag{84}$$

Using this, we get

$$c = \lim_{n\to\infty} \mathcal{C}/n = \log\left[\frac{1+(\alpha^2+\beta^2)P}{1+\beta^2 P}\right] \tag{85}$$

ACKNOWLEDGMENTS

The authors would like to thank Jerry Foschini and Emre Telatar for many useful and inspiring discussions. One of the authors (AMS) thanks Harold Baranger, Aris Moustakas and Steve Simon for collaboration and discussions on further applications of ideas elaborated in the paper.

REFERENCES

1. T. M. Cover and J. Thomas,*Elements of Information Theory* (Wiley, New York 1989).
2. A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: Random matrix techniques and large-N expansions for full transfer matrices," arXiv:physics/0010081 v1 31 Oct 2000.
3. M. L. Mehta, *Random Matrices* (Academic Press, New York 1991).
4. I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.* 10(6): 585–595 (1999).
5. G. J. Foschini "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.* 1(2): 41–59 (1996).
6. T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in a Raleigh flat-fading environment," *IEEE Trans. Info. Theory.* 45(1): 139–157 (1999).
7. P. W. Wolniansky, G. J. Foschini, G. D. Golden and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over rich-scattering wireless channel," *URSI Int. Symp. Signals, Syst. Electron.* 295–300 (1998).
8. D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.* 21(3): 281–302 (2003).
9. A. M. Sengupta and P. P. Mitra, "Distribution of Singular Values of some random matrices," *Phys. Rev. E* 60(3): 3389–3392 (1999).
10. A. L. Moustakas et al., "Communication through a diffusive medium: Coherence and capacity," *Science* 287(5451): 287–290 (2000).
11. P. P. Mitra and J. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature* 411(6841): 1027–1030 (2001).
12. D. Chizhik, F. Rashid-Farrokhi, J. Ling and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Commun. Lett.* 4(11): 337–339 (2000).
13. S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.* 21(3): 406–417 (2003).
14. J. Stark, P. P. Mitra and A. M. Sengupta, "Information capacity of nonlinear wavelength division multiplexing fiber optic transmission line," *Opt. Fiber Technol.* 7(4): 275–288 (2001).
15. L. Zheng and D. N. C. Tse, ÀCommunicating on the Grassmann Manifold: A Geometric Approach to the Non-coherent Multiple Antenna Channelö, *IEEE Transactions on Information Theory* 48(2): 359–383 (2002).

16. A. Abdi and M. Kaveh, ÀA space-time correlation model for multielement antenna systems in mobile fading channelsö, *IEEE J. Select. Areas Commun.* 20(3): 550–560 (2002).
17. K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, New York 1991).
18. S. H. Simon and A. L. Moustakas, "Eigenvalue density of correlated complex random Wishart matrices," *Phys. Rev. E* 69(6) pt 2, 065101 (2004).