

On Capacity of Ergodic Multiple-Input Multiple-Output Channels

Leif Hanlen and Alex Grant

Abstract—Capacity results for Gaussian matrix channels are investigated where the receiver has knowledge of the channel realization and the transmitter has knowledge only of the channel statistics. We extend beamforming results and examine arbitrary ergodic random vector channels, under the condition that the transmit covariance is independent of the channel.

Index Terms—Capacity, Gaussian Channels, Multiple-Input Multiple-Output Channels

I. INTRODUCTION

Multiple-input multiple-output (MIMO) channels have become a significant topic of theoretical research. In particular the capacity results obtained in [1, 2] have motivated many theoretical generalizations and have also spawned incredible research and development activity in practical methods (i.e. space-time coding strategies) of attaining these capacities.

From an information theoretic point of view, the main problem is to find the maximum possible rate of transmission over additive white Gaussian noise channels of the form

$$y[i] = \sqrt{P}H[i]x[i] + n[i] \quad (1)$$

where $y[i] \in \mathbb{C}^{r \times 1}$ is a complex column vector of matched filter outputs at symbol time i , $H[i] \in \mathbb{C}^{r \times t}$ is the corresponding matrix of complex channel coefficients. The vector $x[i] \in \mathbb{C}^{t \times 1}$ is the vector of input signals, and $n[i] \in \mathbb{C}^{r \times t}$ is a complex, circularly symmetric Gaussian vector with $E\{n[i]n[i]^\dagger\} = I_r$. Let $n = \max(t, r)$ and $m = \min(t, r)$. The channel has an average power constraint

$$E\{\text{tr}(xx^\dagger)\} = \text{tr}(E\{xx^\dagger\}) \leq 1 \quad (2)$$

and accordingly, the signal-to-noise ratio is defined as P . Throughout this paper we will use $E_a\{f(a)\}$ to denote expectation of a function of variable a , with respect to a . Where the independent variable is clear, we drop the subscript, we assume familiarity with expectation with respect to random matrices.

Henceforth, let $Q = E\{xx^\dagger\}$ denote the input signal covariance [3]. The power constraint (2) assumes that the power received from the collection of transmit signals at any point in space (eg. at some imaginary point close to the transmitter) is given by the summation of the individual signal powers, ie. zero mutual coupling.

L. Hanlen is with National ICT Australia. National ICT Australia is funded through the Australian Government's *Backing Australia's Ability initiative*, in part through the Australian Research Council.

A. Grant is with the Institute for Telecommunications Research, University of South Australia. A part of this work was carried out while A. Grant visited National ICT Australia.

There are several categories of channels (1) that have been investigated in the literature:

- 1) Ergodic channels in which the $H[i]$, $i = 1, 2, \dots$ are random matrices, selected independently of each other and independently of the $x[i]$, according to some matrix probability density function $p(H)$, which is known at the transmitter. The specific channel realizations are unknown at the transmitter, but are known at the receiver.
- 2) Non-ergodic channels in which $H[i] = H$, $i = 1, 2, \dots$ and H is selected once for all time according to a density $p(H)$. The transmitter knows only $p(H)$ and the receiver knows the realization H .
- 3) Deterministic channels, in which $H[i] = H$, $i = 1, 2, \dots$ with H known at both the transmitter and receiver.

In this paper, we focus on case 1). For this ergodic (over i) channel, whose realizations are known only at the receiver, capacity is the solution to the following optimization problem [1]

$$C = \max_{\substack{\text{tr}(Q) \leq 1 \\ Q=Q^\dagger}} E_H \{\log \det (I_r + PHQH^\dagger)\}, \quad (3)$$

where Q is a $t \times t$ hermitian matrix and the expectation is over the random matrix H . In the i.i.d. Gaussian case with $H \sim N_{t,r}(0, I)$ and H independent of x , Telatar showed that the optimizing $Q = I_t/t$,

$$C = E \left\{ \log \det \left(I_r + \frac{P}{t} HH^\dagger \right) \right\}, \quad (4)$$

and gave an expression for computation of (4).

Development of the theory beyond [1, 2] has, by and large, focused on extending these results to increasingly general assumptions regarding the channel statistics $p(H)$. These well-known and oft-cited results have however suffered widespread mis-application in recent literature.

One common mis-quotation is to refer to (3) as the capacity when the transmitter has no channel knowledge. However from (3), it is clear that the optimal transmit covariance Q is a (possibly trivial) function of the statistics $p(H)$ of a random channel. Thus the transmitter is required by (3) to have knowledge of the statistics of the channel. If the transmitter truly knew nothing at all about the channel, the underlying information theoretic problem is completely different and (3) does not apply. In fact, even properly setting up such a problem has difficulties beyond the MIMO nature of the channel.

The most frequent abuse of the literature however, is the application of (4) to various channels in which the elements of H are correlated, are non-central or have other statistical

properties resulting in the sub-optimality of (4). The result of (4) arises from [1, Theorem 1] and holds *only* for independent, identically distributed, circularly symmetric Gaussian channel matrix H , independent of transmit symbols (as stated in that paper). In general, $Q = I_t/t$ is not the optimal input distribution, and thus provides only a lower bound to capacity. In general, the mutual information on the right hand side of (4) is an achievable rate rather than capacity. To stress this point, the following lemma expresses mutual information as a function of the covariance matrix.

Lemma 1. *Consider the channel (1), with the matrices $H[i]$ chosen independently of each other and independently of the transmitted data at each time i according to the matrix probability density function $p(H)$. The transmitter sends signals with input covariance $\mathbb{E}\{xx^\dagger\} = Q$. The mutual information $I(x; y|H) = I(Q)$ for this channel is given by:*

$$I(Q) = \int \log \det (I + PHQH^\dagger) p(H) dH \quad (5)$$

The purpose of this paper we investigate the optimization problem (3) for general $p(H)$. This is approached from several different perspectives. First, in Sections II and III we consider the low- and high-SNR regimes and review the asymptotic forms for the optimal Q in these cases (which are already reasonably well known). In the low SNR regime, we also find the second-order optimal Q . Next, in Section IV we give a fixed-point equation for the optimal Q at arbitrary SNR, which leads directly to an iterative method for numerically solving (3). Proofs of all results are omitted.

II. LOW SIGNAL-TO-NOISE RATIO

This section reviews some known results for low SNR ($P \rightarrow 0$). Consider the matrix channel (1) and define $S = HQH^\dagger$. By Taylor series expansion, (5) may be approximated near $P = 0$ by

$$I(Q) \approx \sum_{n=1} (-1)^{n-1} \frac{P^n}{n} \mathbb{E}\{\text{tr}(S^n)\}. \quad (6)$$

Of particular interest are the first and second order approximations,

$$I(Q) \approx P \text{tr} (QE\{H^\dagger H\}) \quad (7)$$

$$\approx P \text{tr} (QE\{H^\dagger H\}) - \frac{P^2}{2} \text{tr} (Q^2 \mathbb{E}\{(H^\dagger H)^2\}) \quad (8)$$

Theorem 1 (Low SNR First Order Approximation). *Consider a matrix channel (1), with $\mathbb{E}\{HH^\dagger\} = U\Lambda U^\dagger$, with U unitary and $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_t\}$, $\lambda_1 > \lambda_2 > \dots > \lambda_t > 0$. For low SNR, $P\lambda_1 \ll 1$ the capacity achieving distribution is $Q = u_1 u_1^\dagger$ where u_1 is the first column of U . The resulting capacity is*

$$C(P) \approx \max_{\text{tr}(Q)=1} P \text{tr} (\mathbb{E}_H\{HQH^\dagger\}) \quad (9)$$

$$= P\lambda_1 \quad (10)$$

Corollary 1 (Low SNR First Order, Equal Eigenvalues). *Consider a matrix channel (1), with $\mathbb{E}\{HH^\dagger\} = U\Lambda U^\dagger$, with U unitary and Λ diagonal with $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_t\}$ and*

$\lambda_1 = \dots = \lambda_k > \lambda_{k+1} \dots > \lambda_t > 0$. For low SNR, $P\lambda_1 \ll 1$ the capacity achieving distribution is $Q = U\hat{Q}U^\dagger$ where \hat{Q} is diagonal and

$$\hat{Q} = \text{diag} \left\{ \underbrace{\frac{1}{k}, \dots, \frac{1}{k}}_{k \text{ terms}}, 0, \dots, 0 \right\}$$

and $C = Pk\lambda_1$.

Theorem 1 states that to first order, beamforming in the direction of the largest eigenvector of $\mathbb{E}\{HH^\dagger\}$ is optimal. This is intuitively satisfying and aligns with well known results [4, 5]. This result must be taken with care: the approximation is for $P\lambda_1 \ll 1$ so that large channel gains will necessitate a correspondingly smaller value of P before the expansion of (9) is valid. From a waterfilling perspective, Theorem 1 corresponds to placing *all* water on the best eigenmode. Although this result is well known, we use it as a generalization of several MIMO results, and extend it to higher order terms shortly.

Example 1 (Line-of-Sight Correlated Channel). *Consider the channel (1) with $H \sim N_{t,r}(M, R \otimes T)$ (in the notation of [6]) which may be written as*

$$H = M + R^{1/2} X T^{1/2}$$

where $X \sim N_{t,r}(0, I)$.

This model corresponds to correlated Rayleigh fading with a line of sight (LOS) component M . R and T are respectively the receive-side and transmit-side channel covariance matrices. The matrices M , R and T are known to the transmitter, while the particular realizations of H is known only to the receiver. None of M , R or T are assumed to be diagonal, or jointly diagonalisable. From [6, pp. 251], $S = HH^\dagger$ is a quadratic normal form and

$$\mathbb{E}\{HH^\dagger\} = T \text{tr}(R) + M^\dagger M.$$

From Theorem 1

$$C(P)|_{P \rightarrow 0} = P\lambda_1 \quad (11)$$

where λ_1 is the largest eigenvalue of $T \text{tr}(R) + M^\dagger M$. This makes it clear that the most fortuitous arrangement of T and M is when they share a common largest eigenvector.

There are several special cases that result in simpler forms for λ_1 .

- 1) In the case of identity transmit covariance $T = I_t$, $\lambda_1 = \text{tr}(R) + \lambda_1(M^\dagger M)$.
- 2) $M = \alpha I$. Then $\lambda_1 = \alpha^2 + \text{tr}(R)\lambda_1(T)$.
- 3) Weak LOS component, $T \text{tr}(R) \gg M^\dagger M$. Then $\lambda_1 = \text{tr}(R)\lambda_1(T) + \epsilon$, where $|\epsilon| \leq \lambda_1(M^\dagger M)$. Obviously if $M = 0$, $\epsilon = 0$.
- 4) Strong LOS component, $M^\dagger M \gg T \text{tr} R$. Then $\lambda_1 = \lambda_1(M^\dagger M) + \delta$, where $|\delta| \leq \text{tr}(R)\lambda_1(T)$.
- 5) For $r = t = 2$ it is easy to obtain a closed form solution for λ_1 .

Considering now the second-order approximation (8), the optimal Q is in general no longer diagonal, since $\mathbb{E}\{H^\dagger H\}$

and $E\{(H^\dagger H)^2\}$ may not be simultaneously diagonalizable. However if $H \sim N_{t,r}(0, R \otimes T)$, then

$$E\{H^\dagger H\} = \text{tr}(R)T \quad \text{and} \\ E\{(H^\dagger H)^2\} = \text{tr}(R^2) \text{tr}(T)T + \text{tr}^2(R)T^2 + \text{tr}(R^2)T^2$$

which are simultaneously diagonalizable. Using Lagrange multipliers and the Kuhn-Tucker condition, the second-order optimal covariance is given as follows.

Theorem 2 (Low SNR, Second Order). *Consider the channel (1) with $H \sim N_{t,r}(0, R \otimes T)$, where w.l.o.g., $T = \text{diag}(\tau_1, \dots, \tau_t)$ and $Q = \text{diag}(q_1, \dots, q_t)$, and optimize $I(Q)$ over the q_i (i.e. q_i is the power transmitted in the direction of eigenvector i of T). The second-order optimal Q satisfies*

$$\tau_k - q_k \tau_k \left(\frac{\text{tr}(R^2) \text{tr}(T)}{\text{tr}(R)} + \tau_k \frac{\text{tr}^2(R) + \text{tr}(R^2)}{\text{tr}(R)} \right) = \mu$$

if $q_k > 0$ (LHS $\leq \mu$ if $q_k = 0$).

In the case $H \sim N_{t,r}(M, R \otimes T)$, terms involving M prevent simultaneous diagonalization (except in the case that T and M are simultaneously diagonalizable).

Example 2 (Line-of-Sight Correlated Channel). *Let $H \sim N_{t,r}(0, R \otimes T)$ where $R = I_2$ and $T = \text{diag}(\tau, 2 - \tau)$. For the second order approximation, the optimal input covariance is*

$$q_1 = \frac{1}{2} + \frac{3 - 3\tau}{8 + 3(-2 + \tau)\tau} \quad (12)$$

and $q_2 = 1 - q_1$.

III. HIGH SIGNAL-TO-NOISE RATIO

For large z , $\log(1 + z) \rightarrow \log(z)$, and hence at high SNR,

$$I(Q) \rightarrow t \log P + \log \det Q + \log \det(H^\dagger H). \quad (13)$$

Care must be taken in the definition of ‘‘high’’ SNR. The approximation (13) is only valid when $PQ_{ii} \gg \lambda_{\min}$, i.e. the high SNR, is based on high *received* SNR over all modes, not necessarily high transmit power.

Theorem 3 (High SNR capacity). *Consider a matrix channel (1) with H a random variable, independent of Q . Then the capacity achieving distribution is $Q = I_t/t$ and the resulting capacity is*

$$C \rightarrow t \log \left(\frac{P}{t} \right) + E\{\log \det(HH^\dagger)\} \quad (14)$$

Theorem 3 holds for any probability density function $p(H)$, provided that H is independent of Q . Regardless of the characteristics of the channel, the optimal transmit strategy at high SNR is equal power, independent white signals. This is not surprising when it is seen that for large received power, the variation in channel strength is meaningless. From a waterfilling perspective, we have a very deep pool, with tiny pebbles on the bottom: allocation of power is irrelevant.

Note also that at high SNR, $t \log(P/t)$ is asymptotic to the capacity resulting from transmitting independent data across t non-interfering AWGN channels (each channel getting P/t of the available power). The remaining term is either a capacity

loss or gain over this parallel channel scenario, depending on the statistics of the channel.

In the case of Wishart matrices, $H \sim N_{t,r}(0, R \otimes I)$ (14) has a known closed-form solution [7]. For numerical purposes, $E\{\log \det(HH^\dagger)\}$ may be obtained by Monte-Carlo methods.

IV. ARBITRARY SIGNAL-TO-NOISE RATIO

In this section we consider conditions for optimality of a covariance matrix in the maximization (3). In general, this is a semidefinite program, since the maximization is over the cone of positive semidefinite hermitian matrices $Q \geq 0$. In certain cases however, the problem simplifies, and we can obtain convenient conditions for optimality from the Kuhn-Tucker conditions.

Consider the channel (1) with $H[i] \sim N_{m,m}(M, R \otimes T)$, i.e. the correlated Rayleigh LOS channel introduced earlier. It is known that in the case $M = 0$ that the optimal Q has the form

$$Q = U\hat{Q}U^\dagger \quad (15)$$

$$\hat{Q} = \text{diag}(q_1, q_2, \dots, q_t) \quad (16)$$

where U diagonalizes T . This means that for zero-mean channels, the optimization problem reduces to finding the best allocation of power to each eigenvector of T .

In this case, the condition $Q > 0 \implies \hat{Q} > 0$, together with $\text{tr}(Q) = \text{tr}(\hat{Q}) = 1$ allow the application of the Kuhn-Tucker condition for maximization of a convex function over the space of probability vectors [3, p. 87] to yield the following lemma.

Lemma 2. *Consider the channel (1) with $H[i] \sim N_{m,m}(M, R \otimes T)$. The optimizing Q from $\max_{\text{tr}(Q)=1} I(Q)$ has the form (15) and satisfies the Kuhn-Tucker conditions [3, p. 87]*

$$\frac{\partial I(Q)}{\partial q_i} = \mu \quad q_i > 0$$

$$\frac{\partial I(Q)}{\partial q_i} \leq \mu \quad q_i = 0$$

where μ is a constant independent of q_i .

We emphasize that in general, the capacity achieving input covariance Q from Lemma 2 is not the identity, is not even diagonal and may have no particular structure, other than being hermitian. The following theorem is a result of Lemma 2, obtained via differentiation of $I(Q)$.

Theorem 4 (Optimal Covariance). *Consider the channel, with (1) with $H[i] \sim N_{m,m}(M, R \otimes T)$ where both R and T are known at the transmitter, and power limit P . A necessary and sufficient condition for the optimality of \hat{Q} in (15) is*

$$E_S \left\{ \left((I + S\hat{Q})^{-1} S \right)_{kk} \right\} = \mu \quad q_k > 0 \quad (17)$$

$$E_S \left\{ \left((I + S\hat{Q})^{-1} S \right)_{kk} \right\} < \mu \quad q_k = 0 \quad (18)$$

for $k = 1, 2, \dots, t$ and some constant μ . The expectation is with respect to the random matrix $S = U^\dagger H^\dagger H U$, $H \sim N_{r,t}(0, R \otimes T)$

In the case $Q > 0$, the condition (17) may be re-written as a fixed-point equation

$$\hat{Q} = \nu E_S \left\{ \left(\hat{Q}^{-1} + S \right)^{-1} S \right\}, \quad (19)$$

which suggests the following iterative procedure for numerically finding the optimal \hat{Q} . Starting from an initial diagonal $\hat{Q}^{(0)} > 0$, compute

$$q_k^{(i+1)} = \nu^{(i+1)} \left[E_S \left\{ \left((\hat{Q}^{(i)})^{-1} + S \right)^{-1} S \right\} \right]_{kk},$$

selecting $\nu^{(i)}$ at each step to keep $\text{tr} \hat{Q}^{(i)} = P$. Although there is no existing closed form solution for $E_S \left\{ \left(\hat{Q}^{-1} + S \right)^{-1} S \right\}$, it may be accurately estimated using monte-carlo techniques.

The conditions (17), (18) may be compared with the corresponding condition for parallel Gaussian channels. Suppose $y = Sx + n$ where S is a deterministic diagonal matrix known to both the transmitter and receiver, then the condition for optimality of the input covariance is

$$\begin{aligned} \left((I + SQ)^{-1} S \right)_{kk} &= \mu & q_k > 0 \\ \left((I + SQ)^{-1} S \right)_{kk} &< \mu & q_k = 0. \end{aligned}$$

Thus Theorem 4 can be recognised as a direct generalization of the classical water-pouring result for parallel channels.

How may this result be extended to non-zero mean channels? In the general case, we cannot restrict attention to optimization over a probability vector q_1, \dots, q_t and the cone of positive semidefinite matrices must be considered.

One approach is to relax the positive semidefinite condition and to apply the fixed-point equation (19) anyway, since a covariance matrix that satisfies this equation will result in a stationary point. Note also that starting from a p.s.d. hermitian $Q^{(0)}$ the iteration

$$Q^{(i+1)} = \nu^{(i+1)} E_H \left\{ \left(\left(Q^{(i)} \right)^{-1} + H^\dagger H \right)^{-1} H^\dagger H \right\}, \quad (20)$$

will remain in the p.s.d. cone. Hence we may use (20) for optimization of channels with arbitrary $p(H)$.

Another approach is to approximate a non-zero mean random variable with a correlated zero-mean random variable.

Theorem 5 (Wishart Approximation). *Consider a channel (1), where the density of H is known at both transmitter and receiver, and is independent of the transmit signal, with T and M arbitrary and $R = I$. Then $S = HQH^\dagger$ may be approximated by a central Wishart matrix [6, p. 125]*

$$S \sim W_n(0, \Sigma) \quad (21)$$

$$\Sigma = T^{1/2} Q T^{1/2} + \frac{1}{n} M^\dagger M \quad (22)$$

Then the optimal input distribution is given by Theorem 4, with S given by (21).

The relation between correlation and line-of-sight (non-zero mean) has been heuristically established in MIMO channel measurement literature [8–10].

Example 3 (Ricean Example). *Consider a MIMO system with $t = r = n$ transmit and receive elements in Rician fading.*

Let

$$H = \sqrt{\frac{\kappa}{\kappa + 1}} M + \sqrt{\frac{1}{\kappa + 1}} R^{1/2} X.$$

From (22), create an approximate correlated Rayleigh model, for an approximation to capacity:

$$H \approx \sqrt{\frac{1}{\kappa + 1}} \left(R^{1/2} + \sqrt{\frac{\kappa}{n^2}} I \right) X$$

The optimal input covariance may be found from Theorem 4, with appropriate substitutions.

V. CONCLUSION

We have given methods for determining the optimal input covariance for a variety of ergodic multiple-input multiple-output channels. In particular we considered low and high signal-to-noise ratio limits for arbitrary channel statistics. For zero-mean correlated Rayleigh channels we gave a fixed point equation that characterizes the optimal transmit covariance. This particular characterization reveals a close link between the optimality condition for deterministic channels (water filling) and that for ergodic channels. The fixed point equation also yields an iterative method for performing this optimization that is also applicable to channels with non-central, non-Gaussian statistics. We also described a method of approximating line of sight channels to zero-mean correlated channels, which aids in computation of capacity.

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
- [2] 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, 1998.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*. New York, USA: John Wiley & Sons, 1968.
- [4] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 406–417, Apr. 2003.
- [5] S. A. Jafar and A. Goldsmith, "Multi-antenna capacity in correlated Rayleigh fading with channel covariance information," in *Proceedings IEEE International Symposium on Information Theory, ISIT*, Yokohama, Japan, June 29–July 4 2003, p. 470.
- [6] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, ser. Monographs and Surveys in Pure and Applied Mathematics. Boca Raton: Chapman & Hall/CRC, 2000, vol. 104.
- [7] Y.-H. Kim and A. Lapidoth, "On the log determinant of non-central Wishart matrices," in *International Symposium on Information Theory ISIT2003*, Yokohama, Japan, June 29–July 4 2003, p. 54.
- [8] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. Wolniansky, "Correlation analysis based on MIMO channel measurements in an indoor environment," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 713–720, June 2003.
- [9] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 1002–1012, May 2001.
- [10] K. Yu and B. Ottersten, "Models for MIMO propagation channels: A review," *Wireless Communications and Mobile Computing*, vol. 2, pp. 653–666, Nov. 2002.