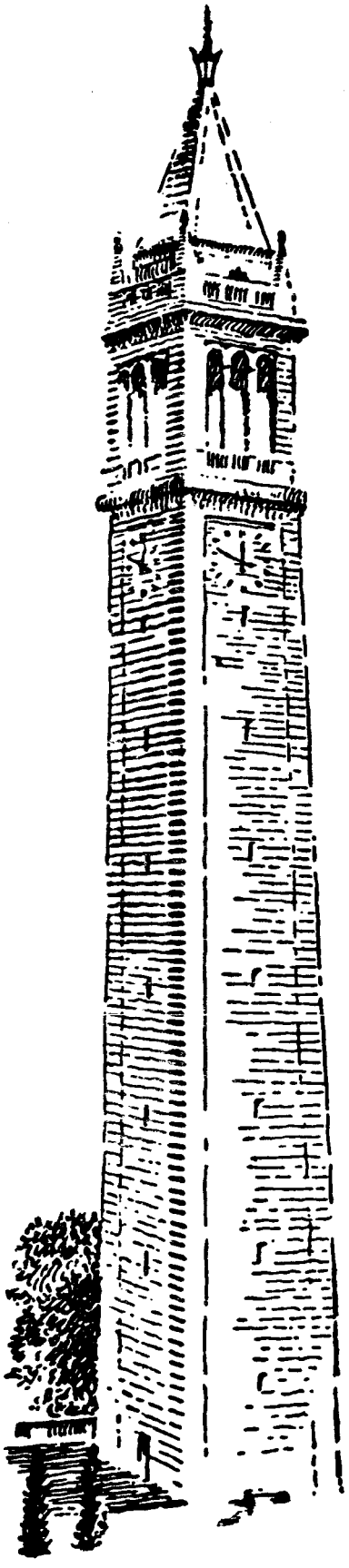

CAPACITY OF CLASSES OF GAUSSIAN CHANNELS

PART I: DISCRETE-TIME

by

W. L. Root

P. P. Varaiya

Memorandum No. ERL-M208

9 March 1967

ELECTRONICS RESEARCH LABORATORY

College of Engineering

University of California, Berkeley

CAPACITY OF CLASSES OF GAUSSIAN CHANNELS
PART I: DISCRETE-TIME

by

W. L. Root

P. P. Varaiya

Memorandum No. ERL-M208

9 March 1967

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Capacity of Classes of Gaussian Channels
Part I: Discrete-Time

by

W. L. Root
University of Michigan
Ann Arbor, Michigan

P. P. Varaiya
University of California
Berkeley, California

ABSTRACT

The usual definition of the capacity of a discrete-time, memoryless Gaussian channel is generalized to the case of a collection of such channels. Each member of the collection is specified by a pair (A, Ω) where A represents the deterministic transmission matrix, possibly infinite-dimensional, and Ω is the covariance matrix of the additive Gaussian noise. The definition is justified by showing that the capacity is the supremum of the attainable rates.

The research reported herein was supported in part by the National Aeronautics and Space Administration under Grant NsG-2-59 to the University of Michigan and by the National Science Foundation under Grant GK-716 to the University of California, Berkeley.

1. Introduction

Suppose a communications system transforms a vector-valued input signal x into a vector-valued output signal y according to an equation of the form

$$y = Ax + z$$

where A is a linear transformation and z is a Gaussian noise vector. Suppose further that neither the transformation A nor the covariance matrix Ω of z are precisely known, but are known only to belong to a certain specified class. Then each possible pair (A, Ω) defines a certain Gaussian channel, and the collection of all pairs (A, Ω) determines a class of channels. For such a class, we define a channel capacity and then prove that the supremum of attainable rates is equal to the capacity.

Section 2 contains the proof of the direct coding theorem when x, y and z are vectors of fixed finite dimension. Section 3 contains the proof of the converse, and Section 4 extends the results of Sections 2 and 3 to infinitely many dimensions under the conditions that the operators A are Hilbert-Schmidt and the noise is white noise.

It will be noted that the A's may be integral operators (of finite rank to meet the conditions of Section 2, or Hilbert-Schmidt to meet the conditions of Section 4) carrying L_2 fundtions defined on a finite interval into L_2 functions on a finite interval (not necessarily the same

interval). The conditions imposed here require that the channel be reset to "zero state" after each transmission before being reused. In the special case of convolution type operators it is more natural to let the time interval for transmitting and receiving grow continuously without limit (Ref.6). The different but related problem of defining a capacity and proving a coding theorem for classes of channels in that case is the subject of a paper with the same title, Part II.

The general outline of the proof used here is taken from Blackwell, Breiman and Thomasian.[1]

2. Preliminaries

We consider communication channels and classes of channels that can be described as follows. A transmitted signal x and a received signal y are (column) vectors of dimension p, with real-valued components, and are related by

$$y = Ax + z$$

where A is a $p \times p$ matrix and z, the noise, is a Gaussian random (column) vector of dimension p. We assume $Ez = 0$ and denote the covariance matrix of z, Ezz', by Ω. A channel is a pair (A, Ω). Specification of a channel determines the statistics of the random vector y once the vector x is given. We are concerned with classes $\mathcal{C}$ of channels satisfying the following condition: there are numbers

a, α_0, α_1, $0 < a$, $0 < \alpha_0 < \alpha_1$ such that for each $(A,\Omega)_\gamma \in \mathcal{b}$

(1) $||A|| \leq a$, where $||A||$ is the operator norm of the matrix A (i.e., $||Ax|| \leq a||x||$ for all p-vectors x, where $||x||$ is the usual "distance norm").

(2) $\alpha_0 \leq \dfrac{x'\Omega x}{||x||^2} \leq \alpha_1$ for all p-vectors x.

Henceforth when we speak of a class $\mathcal{b}$ of channels these conditions will be assumed. We shall sometimes refer to the set of indices γ itself as $\mathcal{b}$.

The n-extension of a channel (A,Ω) is denoted by $(A,\Omega)^n$; it carries n-sequences of p-dimensional input vectors into n-sequences of p-dimensional output vectors according to

$$y_i = Ax_i + z_i, \quad i = 1, 2, \cdots, n$$

where the z_i are mutually independent random Gaussian-vectors, each with mean zero and covariance matrix Ω. It is convenient to denote sequences of x's by $u = (x_1, x_2, \cdots, x_n)$ and sequences of y's by $v = (y_1, y_2, \cdots, y_n)$, and to let $U_n(V_n)$ respectively, be the set of all $u(v)$. The ith component of $x(y)$ is written $x^i(y^i)$; thus x_k^i is the ith component of the input vector x_k, and similarly y_k^i is the ith component of y_k.

The average power constraint on the input implies that we are allowed only to send signals u via the n-extension channel which satisfy $||u||^2 = \Sigma_{j=1}^{n} \Sigma_{i=1}^{p} (x_j^i)^2 \leq nM$, where M is a fixed constant. M is the maximum allowable average signal energy per use of channel.

Insofar as it is reasonable to do so, we shall use the notation and definitions of Blackwell, Breiman and Thomasian[1] in what follows. A (G, ϵ, n) code with constraint set E_n for a class $\mathcal{b}$ of channels for $G \geq 1$, $\epsilon > 0$ and n a positive integer, is a set of $[G]$ ($[G]$ denotes "greatest integer contained in G") distinct sequences $u_k = (x_{k1}, x_{k2}, \cdots x_{kn})$, $k = 1, 2, \cdots, [G]$ lying in $E_n \subset U_n$, and a set of $[G]$ disjoint subsets $B_1, \cdots, B_{[G]}$ of the collection V_n of all sequences $v = (y_1, \cdots, y_n)$, such that

$$P_\gamma(B_i^c | u_i) \leq \epsilon \quad \text{for} \quad i = 1, 2, \cdots, [G] \text{ and all } \gamma \in \mathcal{b},$$

where $P_\gamma(B|u_i)$, $B \subset V_n$, $u_i \in U_n$ is the probability of the set B of output sequences, given the input sequence u_i and the channel $(A, \Omega)_\gamma$. The P_γ probabilities are always given by pn-variate Gaussian density functions, and $P_\gamma(B_i^c | u_i)$ will be defined for all γ if B_i (and hence its complement B_i^c) is any Lebesgue measurable subset of the pn-dimensional Euclidean space V_n. We call B_i the decoding set for the code word u_i. Suppose a (G, ϵ, n) code is given, then when a code word u_i from this code is transmitted over the channel an output word v is received. If v falls in some B_k it is decoded as u_k. Thus v is

correctly decoded if and only if it falls in B_i, and by the definition of a (G, ϵ, n) code the probability of error is uniformly dominated by ϵ.

A number $R \geq 0$ is an <u>attainable rate</u> for a class $\mathcal{b}$ of channels if there exists a sequence of codes (e^{Rn}, ϵ_n, n) for $\mathcal{b}$ where $\epsilon_n \to 0$ as $n \to \infty$. The quantity $\hat{C}(\mathcal{b})$ for the class $\mathcal{b}$ is defined to be the supremum of the attainable rates for $\mathcal{b}$.

We choose to define the capacity C of a class of channels $\mathcal{b}$ artifically in terms of the mutual information, and then prove that $C = \hat{C}$, the supremum of the attainable rates. Let $q(u) = q(x_1^1, \cdots, x_1^p; x_2^1, \cdots, x_2^p; \cdots; x_n^1, \cdots, x_n^p)$ be an np-variate probability density function, to be regarded as providing a probability distribution for the input vectors u, which are to be statistically independent of the noise z. Let $p_\gamma(v|u)$ be the np-variate Gaussian density function determined by $P_\gamma(B|u)$, $B \subset V_n$. Then

$$p(v) = \int p_\gamma(v|u)\, q(u)\, du$$

(where the integral is actually an np-fold integral over R^{pn}) defines a probability density function for output vectors v. The <u>mutual information</u> for the input density $q(u)$ and the channel $(A,\Omega)_\gamma$ is defined to be

$$J_\gamma(u, v) = \log \frac{p_\gamma(v/u)}{p(v)} ,$$

where the dependence on $q(u)$ is not made explicit. It turns out that we

need consider only Gaussian densities for the input density p(u), as might be expected since the additive noise is Gaussian. Let $\mathcal{S}$ be the class of p-variate Gaussian density functions with mean zero which have the property that their covariance matrices each has trace less than or equal to M. It will be convenient sometimes to let $S \in \mathcal{S}$ denote both a Gaussian density p(u) belonging to $\mathcal{S}$ and its $p \times p$ covariance matrix, and this should cause no confusion.

We now define the channel capacity for the class $\mathcal{C}$ of channels subject to an average allowable signal power M to be

$$C(\mathcal{C}) = \sup_{p(u) \in \mathcal{S}} \inf_{\gamma \in \mathcal{C}} E_\gamma J_\gamma$$

where $E_\gamma J_\gamma$ is the expected value of J_γ according to the distributions given by p(u) and $p_\gamma(v/u)$ with u and z statistically independent. Let $\Gamma(S, A, \Omega)$ be the matrix $ASA' + \Omega$. Then it is essentially well known, but will be verified below, that

$$E_\gamma J_\gamma = \log \frac{|\Gamma(S, A_\gamma, \Omega_p)|^{1/2}}{|\Omega_\gamma|^{1/2}}$$

where $|\Gamma|$, $|\Omega|$ denote the determinants of the matrices Γ and Ω. Hence,

$$C(\mathcal{C}) = \sup_{S \in \mathcal{S}} \inf_{\gamma \in \mathcal{C}} \log \frac{|\Gamma(S, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \tag{1}$$

3. Proof of the Coding Theorem

The coding theorem asserts that for any class $\mathcal{C}$ of channels as defined in the previous section, the supremum of the attainable rates is at least as great as the capacity of the class of channels as defined by Eq. (1), i.e., $\hat{C}(\mathcal{C}) \geq C(\mathcal{C})$.

Our proof is a modification of the proof given by Blackwell, Breiman and Thomasian in (1) for a class of finite-state channels and is based on their fundamental lemma with adaptations to take care of the particular structure of the channels we are considering. We also use results and methods of Thomasian.[2] The proof proceeds by a sequence of lemmas, the first two of which are taken directly from the above mentioned references.

Lemma 1 to follow applies to a larger class of channels than were defined in the previous section; in particular, it applies to any channel in which each input vector x determines a probability density for the output vector y, $p(y|x)$. This generality is needed temporarily in Lemma 2, as will be seen. We shall denote such a channel in the customary way, by $(U_1, V_1, p(y|x))$. Each (A,Ω) channel is, of course, a $(U_1, V_1, p(y|x)$ channel, but not vice versa.

Lemma 1.

For any channel $(U_1, V_1, p(y|x))$ with any fixed input density function $q(u)$, and for any integer $G \geq 1$, $\alpha > 0$, and measurable subset

E of U_1, there exists a $(G, \epsilon, 1)$ code with constraint set E where ϵ satisfies

$$\epsilon = G e^{-\alpha} + P\{J(x,y) \leq \alpha\} + P\{E^c\},$$

where

$$P\{E^c\} = \int_{E^c} q(x)\, dx$$

and

$$P\{J_\alpha \leq \alpha\} = \int_{J \leq \alpha} p(y|x)\, q(x)\, dx\, dy .$$

Proof. See Thomasian,[2] Theorem 2.

Lemma 2.

Let $(A,\Omega)_\gamma$, $\gamma \in \mathcal{G}' = \{1, 2, \cdots, L\}$, be a finite class of channels, and let $q(x)$ be an input probability density function determining $p_\gamma(x,y)$ and $J_\gamma(x,y)$.

(a) Define a channel $(U_1, V_1, p(y|x))$ by $p(y|x) = \frac{1}{L} \Sigma_{\gamma=1}^{L} p_\gamma(y|x)$ and let $q(x)$ determine $p(x,y)$, $J(x,y)$. Then for all $\alpha > 0$, $\delta > 0$

$$P\{J \leq \alpha\} \leq \frac{1}{L} \Sigma_{\gamma=1}^{L} P_\gamma\{J_\gamma \leq \alpha + \delta\} + L e^{-\delta}$$

(b) Let $E \subset U_1$ be a fixed constraint set. Then for any $\alpha > 0$, $\delta > 0$, $G \geq 1$, there exists a $(G, \epsilon, 1)$ code for $\mathcal{G}'$ with code words in E such

$$\epsilon = LGe^{-\alpha} + L^2 e^{-\delta} + LP\{E^c\} + \Sigma_1^L P_\gamma (J_\gamma \leq \alpha + \delta).$$

Proof. The proof is a trivial modification of that of Lemma 3 (1). We have stated more than we need, for only part (b) of the lemma is used.

The following lemma gives the known result for the expected value of the mutual information for a Gaussian channel.

Lemma 3. For any channel (A,Ω)

$$EJ(x,y) = \log \frac{|\Gamma|^{1/2}}{|\Omega|^{1/2}}$$

where Γ is the matrix $ASA' + \Omega$.

Proof. y is a Gaussian random vector with mean zero and covariance matrix,

$$Eyy' = ASA' + \Omega = \Gamma.$$

Γ is nonsingular, since Ω is nonsingular. Hence

$$p(y) = \frac{1}{(2\pi)^{p/2}|\Gamma|^{1/2}} \exp[-\frac{1}{2} y'\Gamma^{-1} y].$$

The Gaussian random vector $z = y - Ax$ has covariance matrix Ω and mean zero, hence

$$p(y|x) = \frac{1}{(2\pi)^{p/2}|\Omega|^{1/2}} \exp[-\frac{1}{2} (y-Ax)^1 \Omega^{-1} (y - Ax)]$$

Thus $J(x,y)$ is given by

$$J(x,y) = \log \frac{|\Gamma|^{1/2}}{|\Omega|^{1/2}} + \{\frac{1}{2} y' \Gamma^{-1} y - \frac{1}{2} (y - Ax)' \Omega^{-1} (y - Ax)\} .$$

The first term is a constant; the expectation of the second term is

$$\frac{1}{2} E [y' \Gamma^{-1} y - z' \Omega^{-1} z] = \frac{1}{2} (p - p) = 0 ,$$

hence the lemma is proved.

We now obtain an estimate as to how rapidly the distribution of $J(u,v)$ peaks around its mean value. The calculation is an extension of one given in (2).

Lemma 4.

Let (A,Ω) be a fixed channel and consider its nth extension. Let $q(x)$ be a Gaussian distribution for x with covariance matrix S, and let $q_n(x_1, \cdots, x_n) = \prod_n q(x_i)$. Let A, Ω and $q(x)$ determine $J(u,v) = \sum_n J(x_i, y_i) = \sum_n \log \frac{p(y_i|x_i)}{p(y_i)}$. Then, for any $\delta > 0$,

$$P\{J(u,v) \leq E J(u,v) - n\delta\} \leq \exp \left\{ \frac{-np}{4} \left[\left(1 + \frac{4\delta^2}{p^2} \right)^{1/2} - 1 \right] \right\}$$

Proof. We have already observed in the preceding lemma that

$$J(x_i, y_i) = \log \frac{|\Gamma|^{1/2}}{|\Omega|^{1/2}} + \xi_i$$

where

$$\xi_i = -\frac{1}{2}(y_i - Ax_i)'\Omega^{-1}(y_i - Ax^i) + \frac{1}{2}\, y_i'\Gamma^{-1}y_i \tag{2}$$

Now $P\{J(u,v) \leq E - J(u,v) - n\delta\}$

$$= P\{\Sigma^n \xi_i \leq -n\delta\} = P\{-(n\delta + \Sigma^n \xi_i) \geq 0\}$$

$$\leq E\, e^{-t(n\delta + \Sigma^n \xi_i)}$$

$$= e^{-tn\delta}\, E\, e^{-nt\xi}, \quad \text{for any} \quad t \geq 0,$$

since the ξ_i are statistically independent and identically distributed, and where ξ is a random variable with the same distribution as each ξ_i. We now put $h(t) = E\, e^{-t\xi}$, so that

$$P\{J(u,v) \leq EJ(u,v) - n\delta\} \leq (e^{-t\delta} h(t))^n. \tag{3}$$

In order to compute $h(t)$ we introduce the Gaussian random vector $w = \text{column}\,[x,y]$ of dimension $2p$. Since x and y have mean zero, w has mean zero, and its covariance matrix can be written

$$E\, w w' = \begin{bmatrix} S & SA' \\ AS & \Gamma \end{bmatrix}$$

where the matrix is a partitioned matrix with $p \times p$ blocks. ξ is given by Eq. (2) to be $\frac{1}{2}[y'\Gamma^{-1}y - (y - Ax)'\Omega^{-1}(y - Ax)]$. If we define partitioned matrices

$$Y = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \Gamma^{-1} \end{array}\right]$$

and

$$Z = \left[\begin{array}{c|c} A'\Omega^{-1}A & -A'\Omega^{-1} \\ \hline -\Omega^{-1}A & \Omega^{-1} \end{array}\right] ,$$

we can write

$$2\xi = w'Yw - w'Zw .$$

Then,

$$h(t) = E e^{-t\xi} = \frac{1}{(2\pi)^p |W|^{1/2}} \int \exp\left(-\frac{1}{2} w'W^{-1}w\right)$$

$$\times \exp\left[-\frac{t}{2}(w'Yw - w'Zw)\right] dw$$

where the integral is a 2p-fold integral over R^{2p}. It follows that

$$h(t) = \frac{1}{(2\pi)^p |W|^{1/2}} \int \exp\left(-\frac{1}{2} w'Q(t)w\right) dw$$

$$= \left(|Q(t)W|\right)^{-1/2} \qquad \text{for} \quad t < t_0 ,$$

where the $2p \times 2p$ matrix $Q(t)$, which is given by

$$Q(t) = W^{-1} + t(Y - Z) ,$$

is nonsingular for t less than some $t_0 > 0$. Substitution for Y, Z and W, and use of the fact that $\Gamma = ASA' + \Omega$ gives

$$Q(t)W = \begin{bmatrix} I & tA' \\ t\Gamma^{-1}AS & I \end{bmatrix}$$

By a standard result for partitioned matrices,

$$|Q(t)W| = |I - t^2 \Gamma^{-1} ASA'|$$

$$= |\Gamma^{-1}| \, |\Gamma - t^2 ASA'| = |\Gamma^{-1}| \, |\Omega + (1 - t^2) ASA'|$$

$$= |\Gamma^{-1}| \, |\Omega^{1/2} (I + (1 - t^2)\Omega^{-1/2} ASA'\Omega^{-1/2})\Omega^{1/2}|$$

$$= |\Gamma^{-1}| \, |\Omega| \, |I + (1 - t^2)\Omega^{-1/2} ASA'\Omega^{-1/2}| .$$

Let $\lambda_1, \cdots, \lambda_p$ be the eigenvalues of the positive semi-definite matrix

$\Omega^{-1/2}ASA'\Omega^{1/2}$. Then, $|I+(1-t^2)\Omega^{-1/2}ASA'\Omega^{-1/2}| = \prod^{p}(1+(1-t^2)\lambda_i)$.

Since $\Gamma = ASA' + \Omega = \Omega^{1/2}[\Omega^{-1/2}ASA'\Omega^{-1/2} + I]\Omega^{1/2}$,

$$|\Gamma| = |\Omega| \prod^{p} (1+\lambda_i)$$

and hence,

$$|Q(t)W| = \prod^{p} \frac{1+(1-t^2)\lambda_i}{1+\lambda_i} = \prod^{p}\left(1-t^2\ \frac{\lambda_i}{1+\lambda_i}\right)$$

so that $h(t) = \prod^{p}\left(1-t^2\ \frac{\lambda_i}{1+\lambda_i}\right)^{-1/2}$, $\quad 0 \leq t \leq 1$.

Then $h(t) \leq \frac{1}{(1-t^2)^{p/2}}$,

and $(e^{-\delta t}h(t))^{2/p} \leq \frac{e^{-2\frac{\delta}{p}t}}{1-t^2}$, $\quad 0 \leq t \leq 1$.

If we put

$$t = \frac{p}{2\delta}\left[-1+\left(1+\frac{4\delta^2}{p^2}\right)\right]^{1/2}$$

then $0 < t < 1$, and $(1-t^2)^{-1}\ e^{-2\frac{\delta}{p}t}$ is equal to

$$\left(1+\frac{1}{2}\left[-1+\left(1+\frac{4\delta^2}{p^2}\right)^{1/2}\right]\right)\exp\left(-\left[-1+\left(1+\frac{4\delta^2}{p^2}\right)^{1/2}\right]\right).$$

Since $(1+\frac{1}{2}x)\,e^{-x} \leq e^{-x/2}$ for $x \geq 0$, we have

$$e^{-\delta t}h(t) \leq \exp\left(-\frac{p}{4}\left[\left(1+\frac{4\delta^2}{p^2}\right)^{1/2}-1\right]\right)$$

which, combined with the inequality (3), proves the lemma.

The same sort of argument is now used to obtain an exponential bound on $P\{E^c\}$.

Lemma 5.

Let x_i, $i=1,\cdots,n$, be independent identically distributed p-dimensional Gaussian random vectors with mean zero and covariance matrix S. Let the trace of S be equal to M. Then, for any $\delta > 0$,

$$P\{\Sigma^n ||x_i||^2 \geq n(M+\delta)\} \leq \left[(1+\frac{\delta}{M})\,e^{-\frac{\delta}{M}}\right]^{n/2}$$

where $||x_i||^2 = \Sigma_{j=1}^{p}(x_i^j)^2$.

Proof. Tr(S), the trace of S, is equal to

$\Sigma_{j=1}^{p} E(x_i^j)^2 = E||x_i||^2 = M$ for all i. Let x be a random vector with the same distribution as the x_i, then

$$P\{\Sigma^n ||x_i||^2 \geq n(M+\delta)\} = P\{\Sigma^n ||x_i||^2 - n(M+\delta) \geq 0\}$$

$$\leq E\, e^{t(\Sigma^n ||x_i||^2 - n(M+\delta))}$$

$$= \left[e^{-(M+\delta)t}\, E\, e^{t||x||^2} \right]^n \tag{4}$$

since the x_i are mutually independent. By a standard calculation,

$$E\, e^{t||x||^2} = E\, e^{tx'x} = \prod^{p} (1 - 2t\mu_i)^{1/2}, \quad t < t_0$$

where $\mu_1, \cdots, \mu_p$ are the eigenvalues of S, for t_0 small enough so that all the factors are positive, whether S is non-singular or singular (i.e., the equation holds even if some of the $\mu_i = 0$). Since

$$\prod^{p} (1 - 2t\mu_i) \geq 1 - 2t(\mu_1 + \cdots + \mu_p) = 1 - 2tM,$$

$$e^{-(M+\delta)t}\, E\, e^{t||x||^2} \leq \frac{e^{-(M+\delta)t}}{(1 - 2tM)^{1/2}}, \quad 0 < t < \frac{1}{2M}.$$

Putting $t = \frac{1}{2} \frac{\delta}{M(\delta + M)}$ yields.

$$e^{-(M+\delta)t}\, E\, e^{t||x||^2} \leq (1 + \frac{\delta}{M})^{1/2} e^{-\delta/2M}$$

which combined with the inequality (4) proves the lemma.

Only the case where $\mathcal{C}$ is finite in the following lemma on approximation is needed for Theorem 1, but the stronger result is needed for Theorem 2 and it is convenient to put it all together.

Lemma 6.

Let $\epsilon > 0$ be fixed arbitrarily. Then, there is an $S \in \mathcal{S}$ such that (1) $\mathrm{Tr}\, S < M$, (2) $\log \dfrac{|\Gamma(S, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \geq C(\mathcal{C}) - \epsilon$ for $\gamma \in \mathcal{C}$.

Proof. First consider a finite set of γ's $\mathcal{C}'$, $\mathcal{C}' = \{1, 2, \cdots, L\}$. By the definition of $C(\mathcal{C})$ one can find an $S_0 \in \mathcal{S}$ such that

$$\log \frac{|\Gamma(S_0, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \geq C(\mathcal{C}') - \frac{\epsilon}{2} \quad \text{for all } \gamma \in \mathcal{C}'.$$

If $\mathrm{Tr}\, S_0 < M$, there is nothing to prove. So suppose $\mathrm{Tr}\, S_0 = M$. For fixed γ, $\log \dfrac{|\Gamma|^{1/2}}{|\Omega|^{1/2}}$ is a continuous function of S (where the topology on $\mathcal{S}$ is given, say, by regarding each S as an element of R^{p^2} with the usual Euclidean norm). One can therefore change S_0 slightly to give $S_1 \in \mathcal{S}$ so that $\mathrm{Tr}\, S_1 < M$ while

$$\log \frac{|\Gamma(S_1, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \geq C - \epsilon, \quad \gamma = 1, 2, \cdots, L.$$

This can be accomplished, for example, by writing $S_0 = O'DO$ where O is an orthogonal matrix and D is diagonal and positive semi-definite, and decreasing one of the positive diagonal elements of D.

Now consider an arbitrary class $\mathcal{b}$ meeting the conditions stipulated in the previous Section. A triple (S, A, Ω), $S \in \mathcal{S}$, $(A, \Omega) \in \mathcal{b}$, can be regarded as a point in $3p^2$ - dimensional Euclidean space where each of the matrices is a point in R^{p^2}. The product set $\mathcal{Q}$ of all $S \in \mathcal{S}$ with all $(A, \Omega) \in \mathcal{b}$ is a conditionally compact set, because the conditions on the matrices A, on the positive semi-definite matrices S, and on the positive definite matrices Ω guarantee that this product set is bounded in R^{3p^2}. Now the function $\zeta = \log \frac{|\Gamma(S, A, \Omega)|^{1/2}}{|\Omega|^{1/2}}$ is continuous on the closure $\bar{\mathcal{Q}}$ of $\mathcal{Q}$ and hence is uniformly continuous on $\bar{\mathcal{Q}}$. As before, one can find $S_0 \in \mathcal{S}$ such that

$$\log \frac{|\Gamma(S_0, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \geq C(\mathcal{b}) - \frac{\epsilon}{2} \quad \text{for all} \quad \gamma \in \mathcal{b}$$

If $\operatorname{Tr} S_0 < M$, there is nothing to prove. If $\operatorname{Tr} S_0 = M$, we can find, by the uniform continuity of ζ on $\bar{P}$, a number $\delta > 0$ such that $|\zeta(S, \gamma) - \zeta(S_0, \gamma)| \leq \frac{\epsilon}{2}$ for all γ if $||S - S_0|| \leq \delta$. We can then, as before, change S_0 into S_1 in such a way that $||S_1 - S_0|| \leq \delta$ and $\operatorname{Tr} S_1 < \operatorname{Tr} S_0 = M$. S_1 satisfies the conditions of the lemma.

We now prove the direct half of the coding theorem for a finite class of channels.

Theorem 1

If $\mathcal{C}' = \{(A,\Omega)_\gamma\}$, $\gamma = 1, 2, \cdots, L$, then $\hat{C}(\mathcal{C}') \geq C(\mathcal{C}')$.

Proof. Let R be any positive number less than $C(\mathcal{C}')$ and put $\theta = C(\mathcal{C}') - R$. By Lemma 6 one can find $S_1 \in \mathcal{S}$ such that $\mathrm{Tr}\, S_1 = M - \beta$, $\beta > 0$, and

$$E J_\gamma = \log \frac{|\Gamma(S_1, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \geq R + \frac{\theta}{2}. \tag{5}$$

for all $\gamma \in \mathcal{C}'$. For the n-extension channel let E_n be the set of all input sequences u such that $||u||^2 = \Sigma^n ||x_i||^2 \leq nM$. Then, by Lemma 5, $P\{||u||^2 \geq n(\hat{M} + \beta)\} \leq e^{-n\frac{\beta}{2}}$ where $\hat{M} = M - \beta$ and $\hat{\beta} = \frac{\beta}{M} - \log\left(1 + \frac{\beta}{M}\right) > 0$. Now define $G = e^{nR}$, $\alpha = n(R + \frac{\theta}{8})$, $\delta = n\frac{\theta}{8}$.

It follows from Lemma 2 applied to the n-extension channels that there is a (G, ϵ_n, n) code for $\mathcal{C}'$ with

$$\epsilon_n \leq L e^{nR} \cdot e^{-n(R+\frac{\theta}{8})} + L^2 e^{-\frac{n\theta}{\theta}} + L e^{-\frac{\beta n}{2}} + \sum_{\gamma=1}^{L} P\{J_\gamma(u,v) \leq n(R+\tfrac{\theta}{4})\}. \tag{6}$$

Since $E\, J_\gamma(u,v) = n\, E\, J_\gamma(x,y)$, it follows from Eq. (5) and Lemmas 3 and 4 that

$$P\{J_\gamma(u,v) \leq n(R+\frac{\theta}{4})\} = P\{J_\gamma(u,v) = n(R+\frac{\theta}{2}) - n\frac{\theta}{4}\}$$

$$\leq P\{J_\gamma(u,v) \leq E\,J_\gamma(u,v) - n\frac{\theta}{4}\}$$

$$\leq \exp\left\{-\frac{np}{4}\left[1+\frac{\theta^2}{4^2_p}^{1/2} - 1\right]\right\}.$$

Then, from (6),

$$\epsilon_n \leq (L+L^2)e^{-\frac{n\theta}{8}} + Le^{-\frac{n\beta}{2}}$$

$$+ L\exp\left\{-\frac{np}{4}\left[1+\frac{\theta^2}{4^2_p}^{1/2} - 1\right]\right\} \qquad (7)$$

which approaches zero as $n \to \infty$. Since R is any number less than $C(\mathcal{b}')$ the theorem is proved.

To extend the theorem to arbitrary classes $\mathcal{b}$ we need to establish an approximation inequality, and a probabilistic bound on output power to make the approximation inequality applicable.

Lemma 7.

Let (A,Ω), $(\hat{A},\hat{\Omega})$ be two channels and let u be an input n-sequence of vectors x_i. Let $p_{A,\Omega}\{v/u\}$ be the np-variate probability density for the output signal sequence v, given u, for the n-extension of the (A,Ω) channel, and $p_{\hat{A},\hat{\Omega}}\{v/u\}$ be the corresponding density for the $(\hat{A},\hat{\Omega})$

channel. Then, for those v satisfying $||v||^2 \leq nT$

$$\frac{p_{A,\Omega}\{v\ u\}}{p_{\hat{A},\hat{\Omega}}\{v\ u\}} \leq \frac{|\hat{\Omega}|^{n/2}}{|\Omega|^{n/2}} \exp\left\{ \frac{n}{2\alpha_0^2}\left[T + a^2 M + a\sqrt{MT}\right] ||\Omega - \Omega|| \right.$$

$$\left. + \frac{n}{\alpha_0}\left[\sqrt{MT} + aM\right] ||A - A|| \right\}$$

where the norm signs on the matrices denote operator norms, and where α_0, a, M are numbers such that: $\alpha_0 \leq \frac{z'\Omega z}{||z||^2}$ and $\frac{z'\Omega z}{||z||^2}$ $\left(\text{i.e., } ||\Omega^{-1}||, ||\hat{\Omega}^{-1}|| \leq \frac{1}{\alpha_0}\right)$, $||A||$ and $||A|| \leq a$, $||u||^2 = \Sigma^n ||x_i||^2 \leq nM$.

Proof. $$\frac{p_{A,\Omega}(y|x)}{p_{\hat{A},\hat{\Omega}}(y|x)} = \frac{|\hat{\Omega}|^{1/2}}{|\Omega|^{1/2}} \exp\left\{ -\frac{1}{2}(y - Ax)'\Omega^{-1}(y - Ax) + \frac{1}{2}(y - \hat{A}x)'\hat{\Omega}^{-1}(y - \hat{A}x) \right\} \tag{8}$$

Now, $$|(y - Ax)'\Omega^{-1}(y - Ax) - (y - \hat{A}x)'\hat{\Omega}^{-1}(y - \hat{A}x)|$$

$$\leq |(y - Ax)'\Omega^{-1}(y - Ax) - (y - Ax)'\hat{\Omega}^{-1}(y - Ax)|$$

$$+ |(y - Ax)'\hat{\Omega}^{-1}(y - Ax) - (y - \hat{A}x)'\hat{\Omega}^{-1}(y - \hat{A}x)| \tag{9}$$

The first term enclosed within absolute value signs on the right side of (9) is dominated by

$$||\Omega^{-1} - \hat{\Omega}^{-1}|| \left[||y||^2 + ||A||^2 ||x||^2 + 2||A||\ ||x||\ ||y|| \right].$$

Since $\Omega^{-1} - \hat{\Omega}^{-1} = \Omega^{-1}(\hat{\Omega} - \Omega)\hat{\Omega}^{-1}$, this in turn is dominated by

$\frac{1}{\alpha_0^2} ||\hat{\Omega} - \Omega|| \left[||y||^2 + a^2 ||x||^2 + 2a||x|| \, ||y|| \right]$. The second term enclosed within absolute value signs on the right side of (9) is dominated by

$$2||\hat{A} - A|| \, ||x|| \, ||\hat{\Omega}^{-1}|| \, ||y|| + ||\hat{\Omega}^{-1/2} Ax||^2 - ||\hat{\Omega}^{-1/2} Ax||^2$$

$$= 2||\hat{A} - A|| \, ||x|| \, ||\hat{\Omega}^{-1}|| \, ||y|| + (||\hat{\Omega}^{-1/2} \hat{A}x|| - ||\hat{\Omega}^{-1/2} Ax||)$$

$$\cdot \, (||\hat{\Omega}^{-1/2} \hat{A}x|| + ||\hat{\Omega}^{-1/2} Ax||)$$

$$\leq \frac{2}{\alpha_0} ||\hat{A} - A|| \, ||x|| \, ||y|| + \frac{2}{\sqrt{\alpha_0}} a||x|| \, ||\Omega^{-1/2}(\hat{A} - A) x||$$

$$\leq \frac{2}{\alpha_0} ||\hat{A} - A|| \, ||x|| \, ||y|| + \frac{2a}{\alpha_0} ||\hat{A} - A|| \, ||x||^2$$

Now, using these inequalities to dominate the absolute value of the argument of the exponential in (8), using $||x||^2 \leq M$ and requiring $||y||^2 \leq T$ gives the lemma for $n=1$. Since $p_{A,\Omega}\{v|u\} = \prod_{i=1}^{n} p_{A,\Omega}(y_i|x_i)$ and similarly for $p_{\hat{A},\hat{\Omega}}\{v|u\}$, the lemma follows immediately.

Lemma 8.

Let (A, Ω) be any channel satisfying the conditions (1) and (2) of Section 2. Let u be any input n-sequence satisfying $||u||^2 \leq nM$. Then the output sequence satisfies

$$\Delta = P_{A,\Omega}\{||v||^2 \geq n(2a^2 M + 2p\alpha_1 + 2)|u\} \leq \left[\left(1 + \frac{1}{p\alpha_0}\right) e^{-\frac{1}{p\alpha_0}} \right]^{n/2}$$

Proof. $||v||^2 = \Sigma^n ||y_i||^2 = \Sigma^n ||z_i + Ax_i||^2 \leq 2\ \Sigma^n (||z_i||^2 + ||Ax_i||^2)$

$$\leq 2\ \Sigma^n ||z_i||^2 + 2a^2 nM$$

Hence, $\Delta \leq P\{2\ \Sigma^n ||z_i||^2 + 2a^2 n\, M \geq n(2a^2 M + 2p\alpha_1 + 2)\}$

$$= P\{\Sigma^n ||z_i||^2 \geq n(p\alpha_1 + 1)\} \leq P\{\Sigma^n ||z_i||^2 \geq n(\mathrm{Tr}\,\Omega + 1)\}$$

$$\leq \left[\left(1 + \frac{1}{p\alpha_1}\right) e^{-\frac{1}{p\alpha_1}} \right]^{n/2}$$

by Lemma 5 and the fact that $p\alpha_0 \leq \mathrm{Tr}\,\Omega \leq p\alpha_1$.

The direct half of the coding theorem for an arbitrary number of matrix channels follows.

Theorem 2.

Let $\mathcal{C}$ be a class of channels $(A, \Omega)_\gamma$ satisfying the conditions (1) and (2) of the Section 2. Then

$$\hat{C}(\mathcal{C}) \geq C(\mathcal{C}) = \sup_{S \in \mathcal{S}} \inf_{\gamma \in \mathcal{C}} \log \frac{|\Gamma(S, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}}\ .$$

Proof. Let R be any positive number less than $C(\mathcal{C})$ and put $2\theta = C(\mathcal{C}) - R$. By Lemma 6 one can find $S_1 \in \mathcal{S}$ such that $\mathrm{Tr}\, S_1 = M - \beta$, $\beta > 0$, and

$$E\, J_\gamma(x,y) = \log \frac{|\Gamma(S_1, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \geq R + \theta \tag{10}$$

for all $\gamma \in \mathcal{b}$.

We now pick a finite subset $\mathcal{b}'$ of $\mathcal{b}$ such that for every $(A,\Omega) \in \mathcal{b}$ there is an $(\hat{A},\hat{\Omega}) \in \mathcal{b}'$ with the property that $||A - \hat{A}|| \leq \eta$, $||\Omega - \hat{\Omega}|| \leq \eta$. This can be done because $\mathcal{b}$ is a bounded subset of a finite-dimensional Euclidean space and hence is totally bounded. By the inequality (10), and since $C(\mathcal{b}') \geq C(\mathcal{b})$, $C(\mathcal{b}') \geq R + \theta$. Hence, by the calculations of Theorem 1 there is an (e^{Rn}, ϵ_n', n) code for $\mathcal{b}'$ such that:

(a) The code words $u = (x_1, \cdots, x_n)$ are constrained to lie in E_n, i.e., $||u||^2 \leq nM$.

(b) The probability of error is uniformly dominated by

$$\epsilon_n' \leq (L_\eta + L_\eta^2)\, e^{-\frac{n\theta}{8}} + L_\eta\, e^{-n\frac{\tilde{\beta}}{2}}$$

$$+ L_\eta \exp\left\{-\frac{np}{4}\left[\left(1 + \frac{\theta^2}{4p^2}\right)^{1/2} - 1\right]\right\} \tag{11}$$

where β is independent of n, and L_η is the number of elements in $\mathcal{b}'$. Note that the θ appearing in (11) is $C(\mathcal{b}') - R$, whereas we can take the θ appearing in (11) to be $\frac{C(\mathcal{b}) - R}{2} \leq C(\mathcal{b}') - R$, which actually

weakens the inequality and hence is permissible. θ does not then, however, depend on the approximating class $\mathcal{b}'$.

We now consider the use of the code words and decoding sets belonging to the (e^{Rn}, ϵ_n', n) code for $\mathcal{b}'$ with the larger class of channels $\mathcal{b}$. Let $(A,\Omega) \in \mathcal{b}$ and $(\hat{A},\hat{\Omega}) \in \mathcal{b}'$ and such that $||A - \hat{A}|| \leq \eta$, $||\Omega - \hat{\Omega}|| \leq \eta$. Let u be a code word for $\mathcal{b}'$ and B the corresponding decoding set. Let $F = \{v \mid ||v||^2 \leq nT\}$ where $T = 2a^2M + 2p\alpha_1 + 2$. Then

$$\epsilon_n = P_{A,\Omega}\{B^c|u\} = \underset{A,\Omega}{P}\{(B^c \cap F) \cup (B^c \cap F^c)|u\}$$

$$\leq \underset{A,\Omega}{P}\{B^c \cap F|u\} + \underset{A,\Omega}{P}\{F^c|u\}. \tag{12}$$

By Lemma 8, $P_{A,\Omega}\{F^c|u\} \leq \left[\left(1 + \frac{1}{p\alpha_0}\right) e^{-\frac{1}{p\alpha_0}}\right]^{n/2}$.

By Lemma 7,

$$P_{A,\Omega}\{B^c \cap F|u\} \leq \frac{|\hat{\Omega}|^{n/2}}{|\Omega|^{n/2}} \exp\left\{\frac{n}{2\alpha_0^2}\left[T + a^2M + a\sqrt{MT}\right]||\Omega - \Omega||\right\}$$

$$\times \exp\left\{\frac{n}{\alpha_0}\left[\sqrt{MT} + aM\right]||A - A||\right\} P_{\hat{A},\hat{\Omega}}\{B^c \cap F|u\} \tag{13}$$

Now $P_{\hat{A},\hat{\Omega}}\{B^c \cap F|u\} \leq P_{\hat{A},\hat{\Omega}}\{B^c|u\} \leq \epsilon_n'$.

Hence, using the fact that

$$\alpha_0^p \leq |\Omega_\gamma| \leq \alpha_1^p$$

for all $\gamma \in \mathcal{C}$, we have by substituting for ϵ_n',

$$\epsilon_n \leq \left[\left(1 + \frac{1}{p\alpha_0}\right) e^{-\frac{1}{p\alpha_0}}\right]^{n/2}$$

$$+ \left(\frac{\alpha_1}{\alpha_0}\right)^p \exp\left\{n\left[\frac{T + a^2 M + a\sqrt{MT}}{2\alpha_0^2} + \frac{\sqrt{MT} + aM}{\alpha_0}\right]\eta\right\}$$

$$\times \left[(L_\eta + L_\eta^2) e^{-n\frac{\theta}{8}} + L_\eta e^{-n\frac{\tilde{\beta}}{2}}\right.$$

$$\left. + L_\eta \exp\left\{-n\frac{p}{4}\left[\left(1 + \frac{\theta^2}{4p^2}\right)^{1/2} - 1\right]\right\}\right] \qquad (14)$$

It is evident that the exponentials in (14) are all of the form: $e^{-n(K_1 - K_2\eta)}$ where K_1 and K_2 do not depend on η, where K_1 is positive and K_2 is non-negative.

Consequently for η sufficiently small the (e^{Rn}, ϵ_n', n) code for $\mathcal{C}'$ is an (e^{RN}, ϵ_n, n) code for $\mathcal{C}$ with $\epsilon_n \to 0$ as $n \to \infty$.

Thus $\hat{C}(\mathcal{b}) \geq R$ for every $R < C(\mathcal{b})$, which proves the theorem.

4. Converse of the Coding Theorem

In this section we prove the weak converse of Theorem 2. The proof is a trivial modification of the one given by Ash (3) for the case where the class $\mathcal{b}$ consists of a single element. We shall need the following lemma which is a straightforward extension of a result due to Fano (4, p. 144).

Lemma 9. Let q be the distribution function of a p-dimensional random variable y, with covariance matrix Γ. Then

$$H_q(y) \leq \frac{1}{2}(p + \log(2\pi)^p |\Gamma|)$$

where H is the entropy function and $|\Gamma|$ is the determinant of Γ. The equality is achieved if q is Gaussian.

Corollary. Let $(A,\Omega)_\gamma \in \mathcal{b}$. Let q be a distribution on the p-dimensional input vectors x. Let y be the output vector. Then $I_\gamma(q) = E_{\gamma,q} J(x,y)$ $\leq \log \frac{|\Gamma(S, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}}$ where S is the covariance matrix of x due to the distribution q. Again, the equality is acheived if q is Gaussian.

Proof. We have

$$I_\gamma(q) = H_{\gamma,q}(y) - H_{\gamma,q}(y|x)$$

Now the random variable $y = Ax + z$ has covariance matrix $\Gamma = \Gamma(S, A_\gamma, \Omega_\gamma)$ so that by Lemma 9,

$$H_{\gamma, q}(y) \leq \frac{1}{2} (p + \log(2\pi)^p |\Gamma|) \tag{15}$$

Also, $H_{\gamma, q}(y|x) = H(z)$ and z is a Gaussian random variable with covariance matrix Ω so that by Lemma 9

$$H(z) = \frac{1}{2} (p + \log (2\pi)^p |\Omega|).$$

Combining the above equality with (15) gives the result.

Lemma 10. If there exists a (G, ϵ, n) code for $\mathcal{b}$ with G an integer, and with average power constraint M, then

$$\log G < \frac{nC(\mathcal{b}) + \log 2}{1 - \epsilon}$$

Proof. Again we use the letter q, with various affixes to denote distribution functions on the input space R^p, and the expected mutual information corresponding to a distribution q and a channel $\gamma \in \mathcal{b}$ will be denoted by

$$I_\gamma(q) = H_{\gamma, q}(x) - H_{\gamma, q}(x|y) = H_{\gamma, q}(y) - H_{\gamma, q}(y|x) \tag{16}$$

Let the codewords be given by $u_1 = (x_{11}, \cdots, x_{1n}), \cdots, u_G = (x_{G1}, \cdots, x_{Gn})$. We will say that x_{ij} is a component of the codewords. Also if

$z_i = (z_i^1, \cdots, z_i^p)$ for $i=1, 2$ are any two vectors in R^p, we say $z_1 < z_2$ if and only if $z_1^i \leq z_2^i$ for every i. Now for each p-vector x let

$$\bar{q}(x) = \frac{1}{Gn} \text{ (number of components } x_{ij},\ i=1, \cdots, G,\ j=1, \cdots, n \text{ which are } \leq x.)$$

and for $j = 1, \cdots, n$ let

$$q_j(x) = \frac{1}{G} \text{ (number of components } x_{ij},\ i=1, \cdots, G \text{ which are } \leq x.)$$

Then $\bar{q}(x) = \frac{1}{n} \Sigma_{j=1}^{n} q_j(x)$ so that by the concavity of I (see (5), p. 131), we have for each $\gamma \in \mathcal{b}$

$$I_\gamma(\bar{q}) \geq \frac{1}{n} \Sigma_{j=1}^{n} I_\gamma(q_j) \tag{17}$$

Also if S is the covariance matrix on the input vectors induced by $\bar{q}$ then the trace of $S \leq M$ and by the corollary for each $(A,\Omega)_\gamma \in \mathcal{b}$

$$I_\gamma(\bar{q}) \leq \log \frac{|\Gamma(S, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} \tag{18}$$

Now let $q(u)$ be the pn-dimensional distribution which assigns probability $\frac{1}{G}$ to each of the codewords u_i, $i=1, \cdots, G$. Then (see 5, p. 125) we have for each $\gamma \in \mathcal{b}$

$$I_\gamma(q) \leq \Sigma_{j=1}^{n} I_\gamma(q_j) \tag{19}$$

Finally if we express $I_\gamma(q)$ as

$$I_\gamma(q) = H_{\gamma,q}(u) - H_{\gamma,q}(u|v) = \log G - H_{\gamma,q}(u|v)$$

then by Fano (5, p. 187) we have

$$H_{\gamma,q}(u|v) \leq \log 2 + \epsilon \log G$$

so that for each $\gamma \in \mathcal{C}$,

$$I_\gamma(q) \geq \log G - \log 2 - \epsilon \log G \tag{20}$$

The chair of inequalities (17) - (20) yields

$$\log G \leq n \log \frac{\dfrac{|\Gamma(S, A_\gamma, \Omega_\gamma)|^{1/2}}{|\Omega_\gamma|^{1/2}} + \log 2}{1 - \epsilon} \tag{21}$$

for each $\gamma \in \mathcal{C}$. Taking the infininum over $\gamma \in \mathcal{C}$ first and then the supremum over $S \in \mathcal{S}$ on both sides of (21) gives the result.

Theorem 3. $\hat{C}(\mathcal{C}) \leq C(\mathcal{C})$.

Proof. Let R be an attainable rate for $\mathcal{C}$, so that there is a sequence

of (e^{nR}, ϵ_n, n) codes for $\mathcal{C}$ with $\epsilon_n \to 0$ as $n \to \infty$. By Lemma 10

$$[nR] \leq \frac{n C(\mathcal{C}) + \log 2}{1 - \epsilon_n}$$

Dividing both sides by n and taking the limit as $n \to \infty$ we see that $R \leq C(\mathcal{C})$.

5. The Case of ∞-Dimensional Channels

In this section we extend the results of the previous sections to to the case where the matrix A, and hence the input vector x and the output y are ∞-dimensional. For simplicity, we assume that the additive noise is white. Thus if x is the input vector to the channel A, then the output vector y is given by

$$y = Ax + z$$

where $z = (z^1, z^2, \cdots)$ is an ∞-dimensional random vector with independent components each of which is a Gaussian random variable with zero mean and variance σ^2. The n-extension of A is defined as before so that it carries an n-sequence of input vectors $u = (x_1, \cdots, x_n)$ into an n-sequence of output vector $v = (y_1, \cdots, y_n)$ with

$$y_i = Ax_i + z_i, \quad i = 1, \cdots, n$$

where the z_i are mutually independent.

As before $\mathcal{b}$ will represent a class of channels, i.e., a class of matrices A. We assume the following:

(1) Each matrix $A \in \mathcal{b}$, is Hilbert-Schmidt, i.e., if $A = \{a_{ij}\}$ then $\Sigma_{i,j} a_{ij}^2 < \infty$.

(2) For any two Hilbert-Schmidt matrices $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$, define $||A - B||^2 = \Sigma_{i,j} |a_{ij} - b_{ij}|^2$. Then $||\ \ ||$ defines a metric (in fact a norm). We assume that $\mathcal{b}$ is a totally bounded subset of the metric space of all Hilbert-Schmidt matrices.

As before we impose the average input power constraint M. We define in a similar manner, a (G, ϵ, n) code for $\mathcal{b}$, and an attainable rate for $\mathcal{b}$. Again let $\hat{C}(\mathcal{b})$ be the supremum of all attainable rates. Now let $\mathcal{s}$ be the set of all ∞-dimensional covariance matrices S whose trace is less than or equal to M. For each $A_\gamma \in \mathcal{b}$ and $S \in \mathcal{s}$, the matrix $A_\gamma S A_\gamma'$ is symmetric and positive semi-definate. Let its eigenvalues be $\lambda_1^\gamma \geq \lambda_2^\gamma \geq \cdots$. We define the capacity of $\mathcal{b}$ to be

$$C(\mathcal{b}) = \sup_{S \in \mathcal{s}} \inf_{\gamma \in \mathcal{b}} \frac{1}{2} \Sigma_{i=1}^{\infty} \log \left(1 + \frac{\lambda_i^\gamma}{\sigma^2}\right)$$

We now proceed to show that $\hat{C}(\mathcal{b}) = C(\mathcal{b})$.

For any matrix $B = \{b_{ij}\}$ and positive integer k, let $B^k = \{b_{ij}^k\}$ be the matrix given by $b_{ij}^k = b_{ij}$ if $i \leq k$, $j \leq k$ and $b_{ij}^k = 0$ otherwise. For $S \in \mathcal{s}$ and $A_\gamma \in \mathcal{b}$ we denote the eigenvalues of $A_\gamma^k S^k A_\gamma^{k'}$ by

$\lambda_1^{\gamma,k} \geq \lambda_2^{\gamma,k} \geq \cdots$. Note that $S^k \in \mathcal{S}$.

Lemma 11. Let $S \in \mathcal{S}$ be a fixed, diagonal matrix, i.e., if $S > \{s_{ij}\}$ then $s_{ij} = 0$ for $i \neq j$. Then for each $\epsilon > 0$ there exists $k_0 = k_0(\epsilon) < \infty$ such that for all $k \geq k_0$ and for all $\gamma \in \mathcal{b}$

$$\Delta(\gamma,k)\left| \sum_{i=1}^{\infty} \log\left(1+\frac{\lambda_i^{\gamma}}{\sigma^2}\right) - \sum_{i=1}^{\infty} \log\left(1+\frac{\lambda_i^{\gamma}}{\sigma^2}\right) \right| \leq \epsilon . \tag{22}$$

Proof: Since S is adiagonal, the operator represented by $A_\gamma S A'_\gamma$ dominates the operator represented by $A_\gamma S^k A'_\gamma$. Hence the eigenvalues of $A_\gamma S A'_\gamma$ dominated the eigenvalue of $A_\gamma S^k A'_\gamma$. Now $A_\gamma^k S^k A_\gamma^{k'} = P^k A_\gamma S^k A'_\gamma P^k$ for some projection operator P^k. Hence the eigenvalues of $A_\gamma S^k A'_\gamma$ dominate the eignenvalues of $A_\gamma^k S^k A_\gamma^{k'}$. Therefore $\lambda_i^{\gamma} \geq \lambda_i^{\gamma,k}$ for each i,k and γ. Hence

$$\Delta(\gamma,k) = \sum_{i=1}^{\infty} \left\{ \log\left(1+\frac{\lambda_i^{\gamma}}{\sigma^2}\right) - \log\left(1+\frac{\lambda_i^{\gamma}}{\sigma^2}\right)\right\}$$

$$\leq \sum_{i=1}^{\infty} \left(\frac{\lambda_i^{\gamma}}{\sigma^2} - \frac{\lambda_i^{\gamma,k}}{\sigma^2}\right) \quad \text{since} \quad \lambda_i^{\gamma} \geq \lambda_i^{\gamma,k}$$

$$= \frac{1}{\sigma^2}\left(\text{Trace } A_\gamma S A'_\gamma - \text{Trace } A_\gamma^k S^k A_\gamma^{k'}\right)$$

Let $A_\gamma = \{a_{ij}\}$. Then using the fact that S is diagonal we obtain

$$\text{Trace}\,(A_\gamma S A_\gamma' - A_\gamma^k S^k A_\gamma^{k'}) = \sum_{j=1}^{\infty} s_{jj} \sum_{i=1}^{\infty} a_{ij}^2 - \sum_{j=1}^{k} s_{jj} \sum_{i=1}^{k} a_{ij}^2$$

$$\leq \sum_{j=k+1}^{\infty} s_{jj} \sum_{i=1}^{\infty} a_{ij}^2 + \sum_{j=1}^{\infty} s_{jj} \sum_{i=k+1}^{\infty} a_{ij}^2 .$$

Since the matrices $A_\gamma \in \mathcal{b}$ are uniformly bounded, there is a number $N < \infty$ such that $\Sigma_{i=1}^{\infty} a_{ij}^2 < N$ for every $A_\gamma \in \mathcal{b}$. Now $S \in \mathcal{S}$ so that $\Sigma_{j=1}^{\infty} s_{jj} \leq M$. Hence there is a $k_1 < \infty$ such that $\Sigma_{j=k_1+1}^{\infty} s_{jj} \leq \frac{\epsilon}{2N}$. Also since the set is totally bounded there is $k_2 < \infty$ such that $\Sigma_{i=k_2+1}^{\infty} a_{ij}^2 \leq \frac{\epsilon}{2M}$ for every $\gamma \in \mathcal{b}$. Then $k_0 = \max(k_1, k_2)$ satisfies (22).

Theorem 4. $\hat{C}(\mathcal{b}) \geq C(\mathcal{b})$.

Proof. Let $R < C(\mathcal{b})$ be fixed. We want to show that R is an attainable rate. By definition of $C(\mathcal{b})$ there exists $S \in \mathcal{S}$ such that

$$\frac{1}{2} \sum_{i=1}^{\infty} \log \left(1 + \frac{\lambda_i^\gamma}{\sigma^2}\right) \geq R + \theta$$

for some $\theta > 0$ for every $\gamma \in \mathcal{b}$. By choosing an appropriate basis for the input and output space we can assume that S is diagonal. We

note that the matrix representation of a channel relative to this new basis may be different, but this does not change $C(\mathcal{b})$ which is defined in terms of the eigenvalues and these are invariant under change of basis. Since S can be assumed diagonal, it follows from Lemma 11, that there is a finite k such that

$$\frac{1}{2} \sum_{i=1}^{\infty} \log \left(1 + \frac{\lambda_i^{\gamma,k}}{\sigma^2}\right) \geq R + \theta/2 \tag{23}$$

for every $\gamma \in \mathcal{b}$. But the $\lambda_i^{\gamma,k}$ are the eigenvalues of the matrix $A_\gamma^k S^k A_\gamma^{k'}$ which is effectively a k-dimensional matrix channel so that from (23) and Theorem 2 we conclude that $\hat{C}(\mathcal{b}) \geq R + \theta/2$ and the theorem is proved.

The detailed proof of the weak converse of Theorem 4 is laborious but straightforward and hence only a sketch is given.

Theorem 5. $\hat{C}(\mathcal{b}) \leq C(\mathcal{b})$.

Proof: Let there be a (G, ϵ, n) code for $\mathcal{b}$ with G an integer. Let the codewords be $u_1 = (x_{11}, \cdots, x_{1n}), \cdots, u_G = (x_{G1}, \cdots, x_{Gn})$ and let the disjoint decoding sets be $B_1, \cdots, B_G$. Then $P_\gamma(B_i^c | u_i) \leq \epsilon$ for each i and γ. Clearly the same codewords and decoding sets define a (G, ϵ, n) code for every finite subset $\mathcal{b}_f$ of $\mathcal{b}$. Now we can find a $k_1 = k_1(\epsilon) < \infty$, and disjoint sets $B_1, \cdots, B_G$ (see 5) such that $P_\gamma(B_i^c | u_i) \leq 2\epsilon$ for all $\gamma \in \mathcal{b}_f$ and such that the B_i are cylinder sets determined by the first k_1

components of the output. Thus we obtain a $(G, 2\epsilon, n)$ code for $\mathcal{b}_f$ with the same codewords and with decoding sets given by the first k_1 outputs. Next we find a $k_2 = k_2(\epsilon) < \infty$ such that the codewords $u_1, \cdots, u_G$ obtained from $u_1, \cdots, u_G$ by setting all but the first k_2 components of the x_{ij} to zero satisfy, $P_\gamma(B_i^c | u_i) < 3\epsilon$ for all $i = 1, \cdots, G$ and all $\gamma \in \mathcal{b}_f$. Thus the codeword $u_1, \cdots, u_G$ and deconding sets $B_1, \cdots, B_G$ determine a $(G, 3\epsilon, n)$ code for $\mathcal{b}_f$ which effectively uses only the first $k = \max(k_1, k_2)$ inputs and output components. By Lemma 10 therefore,

$$\log G < \frac{n C(\mathcal{b}_f) + \log 2}{1 - 3\epsilon} .$$

Taking the infimum of both sides over all finite subsets $\mathcal{b}_f$ of $\mathcal{b}$ yields

$$\log G < \frac{n C(\mathcal{b}) + \log 2}{1 - 3\epsilon} .$$

A reproduction of the proof of Theorem 3 now gives the result.

REFERENCES

1. Blackwell, D., Breiman, L., and Thomasian, A. J. (1959), The capacity of a class of channels, Ann. Math. Statist., 30, pp. 1229-1241.

2. Thomasian, A. J. (1960), Error bound for continuous channels," Fourth London Symposium on Information Theory," pp. 46-60, Butterworth Scientific Publications, London.

3. Ash, R. B. (1963), Capacity and error bound for a time-continuous Gaussian channel. Information and Control 6, pp. 14-27.

4. Fano, R. M. (1961), "Transmission of Information," M. I. T. Press and Wiley, New York.

5. Ash, R. B. (1964), "Further Discussion of a Time-Continuous Gaussian Channel," Info and Control 7 pp. 78-83.

6. Gallager, R. G., "Continuous Time Channels," Draft of a chapter of a book in preparation.