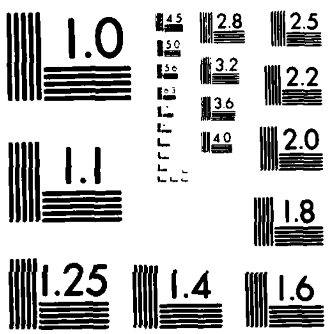
AD-A170 255 INFORMATION CAPACITY OF THE MISMATCHED GAUSSIAN CHANNEL 1/1
(U) NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF
STATISTICS C R BAKER DEC 85 N00014-81-K-0373

UNCLASSIFIED F/G 9/4 NL


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

(12)

AD-A170 255

Information Capacity of the Mismatched Gaussian Channel

Charles R. Baker
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

LISS 13
December, 1985

Research supported by ONR Contracts N00014-81-K-0373 and N00014-84-C-0212

Part of these results was presented at the 1983 IEEE Symposium on Information Theory, September 26-30, 1983, St. Jovite, Quebec, Canada.

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Office of Naval Research		N00014-81-K-0373 & N00014-84-C-0212

8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.			
Statistics & Probability Program Arlington, VA 22217	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
11. TITLE (Include Security Classification) Information Capacity of the Mismatched . . .	NR	042	269	SRO 105

12. PERSONAL AUTHOR(S)
Charles R. Baker

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM ______ TO ______	Dec., 1985	35

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	information theory; channel capacity; Gaussian channels; Shannon theory

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Information capacity is determined for the additive Gaussian channel when the constraint is given in terms of a covariance different from that of the channel noise. These results, combined with previous results on capacity when the constraint covariance is the same as the noise covariance, provide a complete and general solution for the information capacity of the Gaussian channel without feedback. They are valid for both continuous-time and discrete-time channels, and require only two assumptions: the noise energy over the observation period is finite (w.p.1), and the constraint is given in terms of a reproducing kernel Hilbert space norm. Applications include channels with ambient noise having unknown covariance, and channels subject to jamming. The results for the mismatched channel differ markedly from those for the matched channel.

TITLE CONT.: Gaussian Channel

Abstract

Information capacity is determined for the additive Gaussian channel when the constraint is given in terms of a covariance different from that of the channel noise. These results, combined with previous results on capacity when the constraint covariance is the same as the noise covariance, provide a complete and general solution for the information capacity of the Gaussian channel without feedback. They are valid for both continuous-time and discrete-time channels, and require only two assumptions: the noise energy over the observation period is finite (w.p. 1), and the constraint is given in terms of a reproducing kernel Hilbert space norm. Applications include channels with ambient noise having unknown covariance, and channels subject to jamming. The results for the mismatched channel differ markedly from those for the matched channel.

Introduction

The information capacity of the Gaussian channel without feedback, subject to a generalized energy constraint, is determined in [1]. In that work, the constraint is given in terms of the covariance of the channel noise process. However, there are many situations where one may wish to determine capacity subject to a constraint determined by a covariance that is different from that of the channel noise. Examples are jamming or countermeasures situations, or when there is insufficient knowledge of the natural environment.

Channels where the covariance of the noise is the same as that of the constraint will be called matched channels; otherwise, the channel is said to be mismatched (to the constraint). In this paper, the capacity of the mismatched Gaussian channel is determined. Results for a restricted class of mismatched channels are given elsewhere [2]. Various special cases of the mismatched channels have been treated previously [3] - [5].

The results for the mismatched channel differ significantly from those for the matched channel. A discussion of these differences follows the proof of the main result.

An example of the type of problem to which the results given here apply is the following. Suppose that one wishes to obtain the information capacity of the additive Gaussian channel with output

$$Y(t) = \int_0^t (B[X])(s)ds + N(t), \qquad t \text{ in } [0,T]$$

where $(X(t))$ is the message, B is a coding function, $(N(t))$ is zero-mean Gaussian noise independent of the message $(X(t))$, and the constraint is $E\int_0^T [B[X](t)]^2 dt \leq P$. The solution to this problem is given in Proposition 1 and Theorem 2 (if the process $(B[X](t))$ is restricted to lie in a finite-dimensional subspace) and in Theorem 3 (if there is no restriction on the dimensionality of the process $(B[X](t))$. If, for example, the signal detection problem of N vs. W is non-singular (where $(W(t))$ is the Wiener process), and if $r_N(t,s) - \min(t,s)$ is a covariance function, where r_N is the covariance of $(N(t))$, then the capacity for the unrestricted dimensionality signal will be P/2, the same as if $(N(t))$ were the Wiener process.

The relationship with the Wiener process arises because the above constraint is given in terms of the norm of the RKHS for the covariance function $\min(t,s)$: $\int_0^T \dot{y}^2(t)dt = \|y\|_W^2$ when $\dot{y}$ is in $L_2[0,T]$ and $\|\cdot\|_W$ is the norm of the reproducing kernel Hilbert space for $\min(t,s)$. When the detection problem N vs. W is singular, then the capacity can be smaller than, equal to, or larger than P/2. The expression for the capacity will depend on the covariance of $(N(t))$ and the value of P. This dependence of the expression for the capacity on the value of P does not arise when the channel is matched; that is, when the constraint is given in terms of the norm of the RKHS of the channel noise $(N(t))$. Another major difference arises in this problem when the signal process is not constrained to lie in a finite-dimensional subspace. For the matched channel, the capacity then cannot be attained; for the mismatched channel, it can be attained in some situations and not attained in others, depending again on the covariance of $(N(t))$ and the value of P. In this example, it can sometimes be attained if $r_N(t,s) - \min(t,s)$ is not a covariance function; otherwise, it can never be attained.

Definitions and Structure

The channel to be considered is the independent additive Gaussian channel without feedback. The channel output is $Y = A(X) + N$, where N is the Gaussian noise, X is the message process (independent of N), and $A(X)$ is the transmitted signal. The mathematical structure is defined below, as in [1].

The message X is represented by a probability (measure) μ_X on a measurable space $(H_1, \mathbb{B}[H_1])$, where $\mathbb{B}[H_1]$ is a σ-field of subsets of H_1. The noise N is represented by a probability μ_N on a measurable space $(H_2, \mathbb{B}[H_2])$. The transmitted signal $A(X)$ is defined by a $\mathbb{B}[H_1]/\mathbb{B}[H_2]$ measurable coding function A from H_1 into H_2. The received signal (channel output) Y is represented by the probability μ_Y on $(H_2, \mathbb{B}[H_2])$; since $Y = A(X) + N$, $\mu_Y(C) = \mu_X \otimes \mu_N\{(x,n)\colon A(x) + n \in C\}$ for C in $\mathbb{B}[H_2]$, $\mu_X \otimes \mu_N$ being the product probability. The channel probability μ_{XY} on the product measurable space $(H_1 \times H_2, \mathbb{B}[H_1 \times H_2])$ is defined by $\mu_{XY}(C) = \mu_X \otimes \mu_N\{(x,n)\colon (x, A(x)+n) \in C\}$ for C in $\mathbb{B}[H_1 \times H_2]$. The average mutual information is then $I[\mu_{XY}]$, where $I[\mu_{XY}] = \infty$ if it is false that μ_{XY} is absolutely continuous with respect to $\mu_X \otimes \mu_Y$ ($\mu_{XY} \ll \mu_X \otimes \mu_Y$), and otherwise

$$I[\mu_{XY}] = \int_{H_1 \times H_2} \log \left[\frac{d\mu_{XY}}{d\mu_X \otimes \mu_Y}\right] (x,y) d\mu_{XY}(x,y).$$

The information capacity is then $\sup_Q I[\mu_{XY}]$, where Q is a set of admissible pairs (μ_X, A).

H_1 and H_2 will be taken as real separable Hilbert spaces with $\mathbb{B}[H_i]$ the Borel σ-field of H_i. $\langle \cdot\, \cdot \rangle_1$ will be the inner product for H_1, $\langle \cdot, \cdot \rangle$ the inner product for H_2, $\|\cdot\|_1$ and $\|\cdot\|$ the corresponding norms. μ_N will be assumed to have zero mean and finite second moment: $\int_{H_2} \|x\|^2 d\mu_N(x) < \infty$. For $H_2 = L_2(0,T)$ or ℓ_2, this corresponds to an assumption of finite energy. In these cases, μ_N is induced by a path map from an underlying probability space (Ω,β,P): $\mu_N(C) = P\{\omega: N(\omega) \in C\}$ for C in $\mathbb{B}[H_2]$, where (N_t) is a measurable stochastic process with almost all paths in H_2. If $H_2 = L_2[0,T]$, and (N_t) has zero-mean and covariance function r_N, then $\int_{H_2} \|x\|^2 d\mu_N(x) = \int_0^T r_N(t,t)dt$.

A covariance operator in H_2 is (here) any bounded linear operator on H_2 which is also symmetric, non-negative, and trace-class. A probability μ on $(H_2, \mathbb{B}[H_2])$ has such a covariance operator if and only if μ has finite second moment; then, $\int_{H_2} \|x\|^2 d\mu(x) =$ Trace R = Tr R, where R is the covariance operator of μ, defined by (assuming now that μ has zero mean)

$$\langle Ru,v\rangle = \int_{H_2} \langle x,u\rangle\langle x,v\rangle d\mu(x).$$

The covariance operator of μ_N will be denoted by R_N. One can assume WLOG that $\overline{\text{range}(R_N)} = H_2$ [1], so that R_N is strictly positive and R_N^{-1} exists. For $H_2 = L_2[0,T]$, R_N can be represented by an integral operator with kernel function r_N.

A measure μ on $(H_2, \mathbb{B}[H_2])$ is Gaussian if for every v in H_2, the map $x \to \langle x,v\rangle$ defines a Gaussian distribution on $\mathbb{R}$; i.e., $F(a) \equiv \mu\{x: \langle x,v\rangle \leq a\}$ defines a Gaussian distribution. It is known that there is a 1:1 relationship between covariance operators in H_2 and zero-mean Gaussian measures on H_2.

Let R be a strictly positive covariance operator in H_2, with $R^{\frac{1}{2}}$ the positive square root of R; range$(R^{\frac{1}{2}})$ is a separable Hilbert space under the inner product $\langle u,v\rangle_R = \langle R^{-\frac{1}{2}}u, R^{-\frac{1}{2}}v\rangle$.

$\mathcal{L}_2[0,T]$ is the space of all Lebesque-square-integrable real-valued functions on $[0,T]$; $L_2[0,T]$ consists of the equivalence classes formed from elements of $\mathcal{L}_2[0,T]$.

Constraints

Suppose that R_W is a strictly-positive covariance operator in H_2. If H_2 is infinite dimensional, then range($R_W^{\frac{1}{2}}$) is a proper subset of H_2 and is a separable Hilbert space under the inner product

$$<u,v>_W = \sum_n <u,b_n><v,b_n>/\alpha_n = <R_W^{-\frac{1}{2}}u, R_W^{-\frac{1}{2}}v>$$

where (b_n) are c.o.n. (complete orthonormal) eigenvectors of R_W, and α_n are corresponding eigenvalues.

If $H_2 = L_2[0,T]$, then R_W has a representation as an integral operator with kernel r_W. r_W can be defined as a measurable covariance function on $[0,T]\times[0,T]$ and then defines a RKHS H_W of functions on $[0,T]$, for which r_W is the reproducing kernel, with inner product $(u,v)_W$. Let $[u]$ denote the equivalence class in $L_2[0,T]$ defined by the function u in $L_2[0,T]$. Then $[u]$ is in range($R_W^{\frac{1}{2}}$) if and only if $[u]$ is generated by an element u in H_W. Moreover, $(u,v)_W = <[u],[v]>_W$. Thus, in all that follows, one can consider $L_2[0,T]$ as a concrete example of H_2, identify H_W with range($R_W^{\frac{1}{2}}$), and consider $<\cdot,\cdot>_W$ and $\|\cdot\|_W$ as the inner product and corresponding norm for H_W.

The Wiener process is frequently used to model a white noise channel by formally considering the "integrated" channel. If W is the Wiener process on $[0,T]$, then x in H_W has RKHS norm $\|x\|_W^2 = \int_0^T [\dot{x}(t)]^2 dt$; a function x on $[0,T]$ belongs to H_W if and only if x is absolutely continuous, vanishes at the origin, and has derivative in $L_2[0,T]$. In modeling the "integrated" white noise channel by the Wiener process, the constraint $E\|x\|_W^2 \leq PT$ is then an average power constraint on the original signal x. Of course, this constraint has also been used if (W_t) is the actual (not "integrated") channel noise [6].

A constraint which is appropriate in the case of the observation time

$[0,T]$ when (W_t) is stationary with spectral density f_W is

$$E \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\hat{x}(\lambda)|^2}{f_W(\lambda)} d\lambda \leq P$$

where $\hat{x}$ is the L_2-Fourier transform of the function x. From a result of Kelly, Reed, and Root [7], for x in $L_2(-\infty,\infty)$,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\hat{x}(\lambda)|^2}{f_W(\lambda)} d\lambda = \lim_{T\to\infty} \frac{1}{T} \|x_T\|^2_{W,T}$$

where x_T is the restriction of x to $[0,T]$ and $\|\cdot\|_{W,T}$ is the RKHS norm of W restricted to $[0,T]$. If (X_t) is also stationary with spectral density f_S, then with additional assumptions one has

$$\lim_{T\to\infty} \frac{1}{T} E\|X_T\|^2_{W,T} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [f_S(\lambda)/f_W(\lambda)]d\lambda.$$

In general,

$$\lim_{T\to\infty} \frac{1}{T} E\|X_T\|^2_{W,T} \geq \frac{1}{2\pi} \int_{-\infty}^{\infty} [f_S(\lambda)/f_W(\lambda)]d\lambda.$$

An appropriate constraint is thus

$$E_{\mu_X} \|A(X)\|^2_W \leq P \tag{A-1}$$

where R_W is a covariance operator in H_2. This constraint will be used in this paper; no other assumptions will be made.

If R_W is not strictly positive, then the constraint A-1 can still be used after replacing H_2 with $\overline{\text{range}(R_W)}$. One may thus suppose WLOG that $H_2 = \overline{\text{range}(R_W)}$, so that R_W is strictly positive.

<u>Mutual Information and Channel Capacity</u>

From the results of [1], one can limit attention to cases where $\mu_{A(X)}$ is Gaussian with covariance operator

$$R_{A(X)} = \sum_n \tau_n [R_N^{1/2} u_n] \otimes [R_N^{1/2} u_n] \tag{1}$$

where $\tau_n \geq 0$ for $n \geq 1$, $\Sigma_n \tau_n < \infty$, $\{u_n, n \geq 1\}$ is a c.o.n. set and $(u \otimes v)x \equiv \langle v,x \rangle u$.

When $\mu_{A(X)}$ has (1) for covariance and is Gaussian then [1]

$$I[\mu_{XY}] = (\tfrac{1}{2}) \, \Sigma_n \log [1 + \tau_n]. \tag{2}$$

The constraint A-1 can be written as

$$E_{\mu_X} \| R_W^{-\frac{1}{2}} A(X) \|^2 = \text{Trace } R_W^{-\frac{1}{2}} R_{A(X)} R_W^{-\frac{1}{2}} \leq P. \tag{3'}$$

The supremum of (2) subject to the constraint (3') is the capacity sought and will be denoted as $C_W(P)$; the capacity for the matched channel $(R_W = R_N)$ will be denoted by $C_N(P)$.

<u>Proposition 1</u>: $C_W(P)$ is finite if and only if range$(R_W^{\frac{1}{2}}) \subset$ range$(R_N^{\frac{1}{2}})$. This is equivalent to the existence of a densely-defined self-adjoint linear operator S in H_2, as follows:

(1) $S = U(I+V)^{-1}U^* - I$

where U is unitary, V is bounded and self-adjoint, $I+V$ is strictly-positive, and $R_W^{\frac{1}{2}} = R_N^{\frac{1}{2}}(I+V)^{\frac{1}{2}}U^*$;

(2) $I+S$ is strictly positive and bounded away from zero;

(3) $R_N^{\frac{1}{2}} = U(I+S)^{\frac{1}{2}} R_W^{\frac{1}{2}}$. (A-2)

<u>Proof</u>: Range$(R_W^{\frac{1}{2}}) \subset$ range$(R_N^{\frac{1}{2}})$ if and only if there exists a bounded linear operator V in H_2 such that $R_W = R_N^{\frac{1}{2}}(I+V)R_N^{\frac{1}{2}}$[8]. This is equivalent to $R_W^{\frac{1}{2}} = R_N^{\frac{1}{2}}(I+V)^{\frac{1}{2}}U^*$ for U unitary in H_2. The constraint (A-1) is satisfied if and only if $R_{AX} = R_W^{\frac{1}{2}} C R_W^{\frac{1}{2}}$ for C trace-class with trace $C \leq P$. Then, if range$(R_W^{\frac{1}{2}}) \subset$ range$(R_N^{\frac{1}{2}})$, $R_{AX} = R_N^{\frac{1}{2}}(I+V)^{\frac{1}{2}}U^*CU(I+V)^{\frac{1}{2}}R_N^{\frac{1}{2}}$ for bounded V and unitary U. The operator $(I+V)^{\frac{1}{2}}U^*CU(I+V)^{\frac{1}{2}}$ is then trace-class, with trace bounded above by $\|I+V\|P$. From [1], $C_W(P) \leq \|I+V\|P/2$.

Conversely, suppose that $C_W(P) = K < \infty$. Then every admissible (A, μ_X) must satisfy $R_{AX} = R_N^{\frac{1}{2}} T R_N^{\frac{1}{2}}$ for trace $T \leq K$ [1]. This implies $A(X)$ is in range$(R_N^{\frac{1}{2}})$ a.e., μ_X [9]. If range$(R_W^{\frac{1}{2}})$ is not contained in range$(R_N^{\frac{1}{2}})$, then there exists z in range$(R_W^{\frac{1}{2}})$, $z = R_W^{\frac{1}{2}} u$, $\|u\|^2 = P$, with z not in range$(R_N^{\frac{1}{2}})$. Let $A = I$ and take μ_X Gaussian with covariance $R_X = z \otimes z$. Then (A, μ_X) satisfies (A-1) but $A(X)$ lies outside range$(R_N^{\frac{1}{2}})$ with probability one. Thus, $C_W(P)$ finite implies range$(R_W^{\frac{1}{2}}) \subset$ range$(R_N^{\frac{1}{2}})$.

The above proof shows that $C_W(P)$ is finite if and only if there exists a bounded self-adjoint operator V with $R_W^{\frac{1}{2}} = R_N^{\frac{1}{2}}(I + V)^{\frac{1}{2}} U^*$ with U unitary in H_2. Suppose that $R_W^{\frac{1}{2}}$ has such a representation. Since U is unitary and R_W strictly positive, $(I + V)^{-1}$ must exist. Let $S = U(I + V)^{-1} U^* - I$, so that $I + S = U(I + V)^{-1} U^*$, $(I + S)^{\frac{1}{2}} = U(I + V)^{-\frac{1}{2}} U^*$. Then, since $R_W^{\frac{1}{2}} = R_N^{\frac{1}{2}}(I + V)^{\frac{1}{2}} U^* = U(I + V)^{\frac{1}{2}} R_N^{\frac{1}{2}}$, it follows that $U(I + V)^{-\frac{1}{2}} U^* R_W^{\frac{1}{2}} = U R_N^{\frac{1}{2}} = (I + S)^{\frac{1}{2}} R_W^{\frac{1}{2}}$, so that $(I + S)^{\frac{1}{2}}$ is defined on the dense linear manifold range$(R_W^{\frac{1}{2}})$, and $\|(I + S)^{\frac{1}{2}} R_W^{\frac{1}{2}} x\|^2 = \|R_N^{\frac{1}{2}} x\|^2$ for all x in H_2. $I + S$ is obviously strictly positive on its domain $D(S)$. To prove that its smallest limit point is strictly positive, one notes that the spectrum of $(I + S)^{\frac{1}{2}}$ is bounded below by

$$\inf_{\|x\| \neq 0} \frac{\|(I+S)^{\frac{1}{2}} R_W^{\frac{1}{2}} x\|^2}{\|R_W^{\frac{1}{2}} x\|^2}, \quad \text{and} \quad \frac{\|(I + S)^{\frac{1}{2}} R_W^{\frac{1}{2}} x\|^2}{\|R_W^{\frac{1}{2}} x\|^2} = \frac{\|R_N^{\frac{1}{2}} x\|^2}{\|R_W^{\frac{1}{2}} x\|^2} = \frac{\|R_N^{\frac{1}{2}} x\|^2}{\|(I + V)^{\frac{1}{2}} R_N^{\frac{1}{2}} x\|^2} \geq \frac{1}{\|I + V\|},$$

which is strictly positive, since V is bounded. This proves (1) - (3) when $C_W(P)$ is finite; the converse is clear.

Remark 1: Suppose that $H_2 = L_2[0,T]$ and that the set of admissible (A, μ_X) consists of all such that $A(X)$ is absolutely continuous with $L_2[0,T]$ derivative a.e. μ_X, and $E \int_0^T [\dot{A}(X)]^2(t)dt \leq P$. The information capacity $C_W(P)$ will then be finite if and only if μ_N is such that range$(R_N^{\frac{1}{2}})$ contains all equivalence classes in $L_2[0,T]$ that are generated by absolutely continuous functions with $L_2[0,T]$ derivative.

It can often be assumed that the operator S is bounded, from physical considerations. That is, S will be bounded if and only if range$(R_N^{\frac{1}{2}}) \subset$ range$(R_W^{\frac{1}{2}})$. In jamming applications, N may have the form $N = J + W$, where W is the original channel noise and J is a jamming noise independent of the ambient noise W. Since W will typically include wide-band receiver noise, it is not plausible that the sample functions of the jamming noise J should be more irregular than those of the ambient noise W. The path properties of N and W are determined by the properties of the RKHS of N and W (see, e.g., [10]). Thus, if the paths of $W+J$ are not to be more irregular than those of W, then it is necessary that range$(R_N^{\frac{1}{2}}) \subset$ range$(R_W^{\frac{1}{2}})$. These statements, which can be rigorously justified, imply that one can often assume S to be bounded. However, it is desirable to state the results here in maximum generality, so S will not be assumed to be bounded.

When H_2 is infinite-dimensional, θ will denote the smallest limit point of the spectrum of S, the operator defined by (A-2). The limit points of the spectrum of S consist of all eigenvalues of infinite multiplicity, limit points of distinct eigenvalues, or points of the continuous spectrum [11]. A key consequence of θ being a limit point is that there is a sequence of o.n. elements (f_n) such that $\|(S - \theta I)f_n\| \to 0$ [11, p. 364]. From Proposition 1, $1 + \theta > 0$. Moreover, a real number C with $0 < C < 1 + \theta$ can be in the spectrum of $I + S$ if and only if C is an eigenvalue of finite multiplicity for $I + S$. Thus, θ is the only possible limit point of the eigenvalues of S strictly less than θ. $\{\lambda_n, n\geq 1\}$ will denote the eigenvectors of S that are strictly less than θ; of course, this set can be empty. Similarly, $\{e_n, n\geq 1\}$ will always denote an o.n. set of H_2 eigenvalues of S corresponding to the eigenvalues $\{\lambda_n, n\geq 1\}$: $Se_n = \lambda_n e_n$, $n\geq 1$.

The case $\theta = \infty$ requires special treatment. It is simplified by the following result.

<u>Proposition 2</u>: Suppose that $\theta = \infty$. Then $I + V$ must be compact, and $\{\lambda_n, n\geq 1\}$ is an infinite set. Moreover, $P + \sum_1^K \lambda_i < K\lambda_K$ for some finite K, any fixed $P > 0$.

<u>Proof</u>: $\theta = \infty$ implies that zero is the only limit point of the spectrum of $I + V$, so that $I + V$ is compact. Since $I + V$ is self-adjoint, this operator has a c.o.n. set of eigenvectors. $I + V$ is strictly positive, so that its eigenvalues are $\{(1+\lambda_n)^{-1}, n\geq 1\}$, with $\lambda_1 > -1$.

To see that $\sum_1^K \lambda_i + P < K\lambda_K$ for some finite K, suppose not. Then $P \geq \sum_1^K(\lambda_K - \lambda_i)$ for all $K\geq 1$. This cannot hold, since $\lambda_K - \lambda_1 \to \infty$. $\square$

One can now formulate the capacity problem in terms of the operator S, as follows.

$C_W(P)$ is the supremum of (2) subject to the constraint (3'). Rewriting (3') in terms of S, and using (1), one obtains the equivalent constraint

$$\Sigma_n \tau_n \|(I+S)^{\frac{1}{2}} U^* u_n\|^2 \leq P. \tag{3}$$

Setting $X_n^2 = \tau_n \|(I+S)^{\frac{1}{2}} U^* u_n\|^2$,

$$C_W(P) = \sup(\tfrac{1}{2})\ \Sigma_n \log\ [1 + X_n^2(1+\gamma_n)^{-1}] \tag{4}$$

where the supremum is over all sequences (X_n^2) and c.o.n. sets $\{v_n, n\geq 1\}$ in the domain $D(S)$ of S such that $\Sigma_n X_n^2 \leq P$, where $\gamma \equiv \langle Sv_n, v_n\rangle$, $n\geq 1$.

When S is bounded and compact, the results given here were presented at the 1983 IEEE Symposium on Information Theory (St. Jovite, Quebec, Canada) and are partially contained in [2]. An upper bound for the capacity when S is bounded but not compact has been given by Yanagi [12].

Capacity for Finite-Dimensional Signal Space

Proofs of the following two lemmas are given in the Appendix.

Lemma 1: Let (γ_n), $n \leq M$, be any non-decreasing sequence of strictly positive real numbers. Let (X_n) be any sequence of M real numbers. Fix $P > 0$ and define

$$g(M,P,\underset{\sim}{\gamma}) = \sup_{\{\underset{\sim}{X}:\ \sum_1^M X_n^2 \leq P\}} \prod_1^M (\gamma_n + X_n^2)/\gamma_n .$$

Then

$$g(M,P,\underset{\sim}{\gamma}) = \prod_{n=1}^K \left(\sum_1^K \gamma_i + P\right)/(K\gamma_n)$$

where $K \leq M$ is the largest integer such that $\sum_1^K \gamma_i + P \geq K\gamma_K$. $g(M,P,\underset{\sim}{\gamma})$ is uniquely attained by (X_n^2) such that

$$\begin{cases} X_n^2 = \left(\sum_1^K \gamma_i + P\right)/K - \gamma_n & n \leq K \\ \quad\ = 0 & n > K. \end{cases}$$

Lemma 2: Let (λ_i), $1 \leq i \leq K$, be a non-decreasing sequence of strictly positive real numbers and fix $P > 0$. Define a sequence (γ_n) to be admissible if it is non-decreasing, $\sum_1^J \gamma_i \geq \sum_1^J \lambda_i$ for all $J \leq K$, and $\sum_1^K \gamma_i + P \geq K\gamma_K$. Define

$f_K(\underset{\sim}{\gamma}) = \prod_{n=1}^{K} (P + \sum_1^K \gamma_i)/(K\gamma_n)$. Then, for any admissible sequence (γ_n), $f_K(\underset{\sim}{\gamma}) \leq f_K(\underset{\sim}{\lambda})$ with equality if and only if $\gamma_i = \lambda_i$ for all $i \leq K$.

Corollary 1: Let (ν_i) and (γ_i), $i = 1,\ldots,M$ be two non-decreasing sequences of strictly positive real numbers. Fix $P > 0$, and let K be the largest integer $\leq M$ such that $\sum_1^K \gamma_i + P \geq K\gamma_K$. Let (X_n^2), $n = 1,\ldots,M$, be any sequence such that $\sum_1^M X_n^2 \leq P$. If $\sum_{n=1}^{J} \gamma_i \geq \sum_{i=1}^{J} \nu_i$ for all $J \leq K$, then

$$\sum_{n=1}^{M} \log \left[1 + X_n^2 \gamma_n^{-1}\right] \leq \sum_{n=1}^{K} \log \left[(P + \sum_{i=1}^{K} \nu_i)/K\nu_n\right]$$

with equality if and only if $\gamma_n = \nu_n$ for $n \leq K$ and

$$\begin{aligned} X_n^2 &= (\sum_1^K \nu_i + P)/K - \nu_n \qquad & n \geq K \\ &= 0 & n > K. \end{aligned}$$

Proof: If $\sum_{n=1}^{M} \log (1 + X_n^2 \gamma_n^{-1}) > \sum_{n=1}^{K} \log \left[(P + \sum_1^K \nu_i)/K\nu_n\right]$, then it must also strictly exceed (by Lemma 2) $\sum_{n=1}^{K} \log \left[(P + \sum_1^K \gamma_i)/(K\gamma_n)\right]$. This contradicts Lemma 1. The conditions for equality follow from lemmas 1 and 2.

□

Remark 2: If S is a bounded linear operator with a complete set of eigenvectors and non-decreasing eigenvalues $\beta_1 \leq \beta_2 \leq \ldots$, then $\sum_1^K \beta_n \leq \sum_1^K \langle Sv_n, v_n \rangle$ for any o.n. set $v_1,\ldots,v_K$ and any $K \geq 1$ [13].

Theorem 1:

Suppose that H_2 has dimension $M < \infty$. The capacity is then

$$C_W(P) = (\tfrac{1}{2}) \sum_{n=1}^{K} \log \left[\frac{\sum_1^K \beta_i + P + K}{K(1+\beta_n)} \right]$$

where $\beta_1 \leq \beta_2 \leq \dots \leq \beta_M$ are the eigenvalues of S, and K is the largest integer $\leq M$ such that $\sum_1^K \beta_i + P \geq K\beta_K$. The capacity is attained by a Gaussian $\mu_{A(X)}$ with covariance operator (1), where $u_n = Ug_n$ and $\tau_n = \left[\sum_1^K \beta_i + P - K\beta_n\right](1+\beta_n)^{-1}K^{-1}$ for $n \leq K$, $\tau_n = 0$ for $n > K$, and $\{g_n, n \geq 1\}$ are o.n. eigenvectors of S corresponding to the eigenvalues (β_n). No other Gaussian $\mu_{A(X)}$ can attain capacity. The same result is obtained if H_2 has dimension $L < \infty$ and $\mu_{A(X)}$ is constrained to have support of dimension $M < L$.

Proof: Since H_2 is finite-dimensional, the self-adjoint operator S is bounded and has a complete set of eigenvectors. From (4),

$C_W(P) = \sup \left[\frac{1}{2} \sum_1^M \log \left[1 + X_n^2 \gamma_n^{-1}\right] \right]$, where $\gamma_n = 1 + \langle Sv_n, v_n \rangle$, $\{v_n, n \leq M\}$ is a c.o.n. set, and the supremum is over all such c.o.n. sets and all (X_n^2) such that $\sum_1^M X_n^2 \leq P$. Since $\beta_1 \leq \beta_2 \leq \dots \leq \beta_M$ are the non-decreasing eigenvalues of S, by Remark 2, $\sum_1^J [1 + \langle Sv_n, v_n \rangle] \geq \sum_1^J [1 + \beta_n]$ for all $J \leq M$ and any fixed c.o.n. set $\{v_n, n \leq M\}$. The expression for $C_W(P)$ and the unique covariance of the maximizing Gaussian $\mu_{A(X)}$ both now follow from Corollary 1. $\square$

Remark 3: The result holds if $\dim(H_2) = L < \infty$ and $\dim[\mathrm{supp}(\mu_{A(X)})] \leq M \leq L$, since in this case S again has M smallest eigenvalues.

Theorem 2:

Suppose that $\theta<\infty$, H_2 is infinite-dimensional, and support $(\mu_{A(X)})$ is restricted to have dimension $\leq M < \infty$.

(a) If $\{\lambda_n, n\geq 1\}$ is empty, then $C_W(P) = (M/2) \log [1 + PM^{-1}(1+\theta)^{-1}]$. Capacity can be attained if and only if S has θ as an eigenvalue of multiplicity $\geq M$. In this case $C_W(P)$ is attained by a Gaussian $\mu_{A(X)}$ with covariance (1), where $u_i = Ug_i$ and $\tau_i = PM^{-1}(1+\theta)^{-1}$ for $i \leq M$ with $\{g_1,\ldots,g_M\}$ any o.n. set in the null space of $S-\theta I$.

(b) If $K\lambda_K \leq \sum_1^K \lambda_i + P < K\lambda_{K+1}$ for some $K \leq M$, then the capacity is as in Theorem 1, with $\lambda_i = \lambda_i$, $i = 1,\ldots,K$, and can be similarly attained.

(c) Let $K = \min(L,M)$, where $L \geq 1$ is the number of eigenvalues (λ_n) of S whose value is strictly less than θ, and suppose that $P + \sum_1^K \lambda_i \geq K\lambda_K$. The capacity is then

$$C_W(P) = \tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \left(\frac{M}{2}\right) \log \left[1 + \frac{P + \sum_{i=1}^{K} (\lambda_i - \theta)}{M(1+\theta)}\right].$$

The capacity can be attained if and only if θ is an eigenvalue of S with multiplicity $\geq M-K$. The capacity is then achieved by a Gaussian μ_{AX} with covariance (1), where $u_n = Ug_n$ and $\tau_n = (\sum_1^K \lambda_i + P - M\lambda_n + (M-K)\theta)(1+\lambda_n)^{-1}M^{-1}$ for $n \leq K$, with $Sg_n = \lambda_n g_n$ and $\{g_1,\ldots,g_K\}$ an o.n. set; and with $u_n = Uv_n$ and $\tau_n = (P + \sum_1^K \lambda_i - K\theta)M^{-1}(1+\theta)^{-1}$ for $K+1 \leq n \leq M$, where $Sv_n = \theta v_n$ and $v_{K+1},\ldots,v_M$ is an o.n. set. The sets $\{u_1,\ldots,u_K\}$ and $\{\tau_1,\ldots,\tau_K\}$ are uniquely defined for any maximizing Gaussian $\mu_{A(X)}$.

Proof: If $S > \theta I$, then $S - \theta I$ does not have zero as an eigenvalue. However, there exist [11] o.n. elements (f_n) in $D(S)$ such that $\|(S-\theta I)f_n\| \to 0$, so that $\langle(S-\theta I)f_n, f_n\rangle \to 0$, $\langle Sf_n, f_n\rangle \to \theta$, $\langle Sf_n, f_n\rangle > \theta$ for every n. Thus, for any $\varepsilon > 0$ there are o.n. elements $f_1^\varepsilon, \ldots, f_M^\varepsilon$ such that setting $\gamma_i^\varepsilon = \langle Sf_i^\varepsilon, f_i^\varepsilon\rangle$, $\theta < \gamma_i^\varepsilon < \theta + \varepsilon$ for $i \leq M$. Using this sequence in (2) and (4), one obtains

$$I[\mu_{XY}] = (\tfrac{1}{2}) \sum_1^M \log\,[1 + \tau_n]$$

$$= (\tfrac{1}{2}) \sum_1^M \log\,[1 + X_n^2\,(1 + \gamma_n^\varepsilon)^{-1}] \geq (\tfrac{1}{2}) \sum_1^M \log\,[1 + X_n^2\,(1 + \theta + \varepsilon)^{-1}].$$

The expression on the right of the inequality is maximized, over all (X_n^2) such that $\sum_1^M X_n^2 \leq P$, by defining $X_n^2 = P/M$, $n \leq M$. Thus, $C_W(P) \geq (\tfrac{1}{2})\sum_1^M \log\,[1 + (1 + \theta + \varepsilon)^{-1}P/M]$ for all $\varepsilon > 0$, and so $C_W(P) \geq (\tfrac{1}{2}) \sum_1^M \log\,[1 + PM^{-1}(1 + \theta)^{-1}]$. For the reverse inequality, one notes that under the constraint $E_{\mu_X} \|R_N^{-\frac{1}{2}}A(X)\|^2 \leq P$, it is shown in [1] that $C_N(P) = (M/2)\log\,(1 + P/M)$. For $S \geq \theta I$, $\|R_N^{-\frac{1}{2}}A(X)\|^2 \leq (1 + \theta)^{-1}\,\|R_W^{-\frac{1}{2}}A(X)\|^2$. Thus, $E_{\mu_X}\|R_W^{-\frac{1}{2}}A(X)\|^2 \leq P$ implies $E_{\mu_X}\|R_N^{-\frac{1}{2}}A(X)\|^2 \leq (1 + \theta)^{-1}P$, giving $C_W(P) \leq \left(\frac{M}{2}\right) \log\,[1 + \frac{P}{M(1 + \theta)}]$, so that $C_W(P) = (M/2)\log\,[1 + PM^{-1}(1 + \theta)^{-1}]$.

If $S \geq \theta I$, with θ an eigenvalue of multiplicity K, the above argument is modified in an obvious way ($\gamma_1^\varepsilon = \theta$ for $i = 1, \ldots, \min(K,M)$) to again obtain $C_W(P) = (M/2)\log\,[1 + PM^{-1}(1 + \theta)^{-1}]$.

To prove (b), the proof of Theorem 1 is repeated after substituting λ_i for β_i, $i \geq K+1$.

For (c), suppose that S has $K \leq M$ eigenvalues $\lambda_1 \leq \ldots \leq \lambda_K$ strictly less than θ and that $\sum_1^K \lambda_i + P \geq K\lambda_K$. $C_W(P) = \sup_{(P_1,\underline{v})} C_W(P_1,\underline{v})$ where

$$C_W(P_1,\underline{v}) = \sup \left[\tfrac{1}{2} \sum_{n=1}^{M} \log \left[1 + X_n^2 \left(1 + \langle Sv_n, v_n \rangle_2\right)^{-1}\right]\right],$$

$\underline{v} = \{v_n, n \le M\}$ is any o.n. set, $0 \le P_1 \le P$, and the supremum is over all (X_n^2) such that $\sum_1^K X_n^2 \le P_1$, $\sum_1^M X_n^2 \le P$. From the proofs of Theorem 1 and part (a),

$$C_W(P_1,\underline{v}) = (\tfrac{1}{2}) \sum_{n=1}^{J} \log \left[\frac{\sum_1^J \lambda_i + P_1 + J}{J(1+\lambda_n)}\right] + (\tfrac{1}{2})(M-K) \log \left[1 + \frac{P - P_1}{(M-K)(1+\theta)}\right]$$

where $J \le K$ is the largest integer such that $\sum_1^J \lambda_i + P_1 \ge J\lambda_J$. Since this result holds for any o.n. set $\{v_n, n \le M\}$ in $D(S)$, it remains only to determine the value of P_1 that maximizes $C_W(P_1,\underline{v})$ (a differentiable function of P_1 in $[0,P]$). Differentiating, one sees that $C_W(P_1,\underline{v})$ is increasing with P_1 so long as $P_1 < [JP + (M-K)(J\theta - \sum_1^J \lambda_i)](M-K+J)^{-1}$. Since $P_1 < J\lambda_{J+1} - \sum_1^J \lambda_i$, the preceding inequality is satisfied as long as $(M-K+J)\lambda_{J+1} - \sum_1^J \lambda_i < P + (M-K)\theta$ and this is satisfied because $P + \sum_1^J \lambda_i \ge J\lambda_{J+1}$, $\lambda_{J+1} < \theta$. It follows that $C_W(P_1,\underline{v})$ is an increasing function of P_1 for $P_1 < -\sum_1^K \lambda_i + K\lambda_K$. Assuming that $P_1 \ge -\sum_1^K \lambda_i + K\lambda_K$, the maximum of $C_W(P_1,\underline{v})$ is attained uniquely by $P_1 = M^{-1}[KP - (M-K)\sum_1^K \lambda_i + (M-K)K\theta]$. Using this value of P_1 in the expression for $C_W(P_1,\underline{v})$, one obtains $C_W(P)$ as in (c). The value of $C_W(P)$ when $L = M$ in (c) follows as in the proof of Theorem 1. The statement on attaining capacity follows from the results of (a) and (b).

<u>Corollary 2</u>: If $\theta = \infty$, then $C_W(P)$ has the value given in Theorem 2(b), and can be similarly attained.

<u>Proof</u>: Follows from Proposition 2 and the proof of Theorem 1.

Capacity for Infinite-Dimensional Signal Space

Theorems 1 and 2 give the solution to the capacity problem when the dimension of the signal space is finite. We now proceed to the case of an infinite-dimensional signal space.

Lemma 3: Suppose that $\theta < \infty$, $\{\lambda_n, n \geq 1\}$ is an infinite set, and $P > 0$. Then $P + \sum_1^K \lambda_n \geq K\lambda_K$ for all $K \geq 1$ if and only if $P \geq \sum_{n \geq 1} (\theta - \lambda_n)$.

Proof: It suffices to show that $\sum_1^K (\lambda_n - \theta) + P \geq K(\lambda_K - \theta)$ for all $K \geq 1$ implies $\sum[\theta - \lambda_n] \leq P$. Suppose not. Then there exists $K > 1$ and $\Delta > 0$ such that $P + \sum_1^K (\lambda_n - \theta) = -\Delta$. Thus $[\theta - \lambda_{K+1}] \geq [-\sum_1^K (\lambda_n - \theta) - P]/K = \Delta/K$. Suppose $[\theta - \lambda_{K+p}] \geq \Delta/K$ for $1 \leq p \leq N$. Then $[\theta - \lambda_{K+N+1}] \geq [-\sum_1^{K+N} (\lambda_n - \theta) - P]/(K+N)$ $= [\Delta - \sum_{K+1}^{K+N} (\lambda_n - \theta)]/(K+N) \geq [\Delta + N\Delta/K]/(K+N) = \Delta/K$. Thus, the induction hypothesis would yield that $[\theta - \lambda_{K+N}] \geq \Delta/K$ for all $N \geq 1$. However, θ is the smallest limit point of the spectrum of S, and since (λ_n) is a bounded infinite sequence, (λ_n) must contain a limit point. Thus, $\theta - \lambda_{K+N} \geq \Delta/K$ for all $N \geq 1$ would mean that (λ_n) has a limit point strictly less than θ. This contradiction implies that $P + \sum_1^K (\lambda_n - \theta) = -\Delta < 0$ must be false.

$\square$

Lemma 4: Suppose that $\theta < \infty$ and $S - \theta I$ is negative-definite with an infinite set of strictly negative eigenvalues

(a) If $P \geq \sum_n (\theta - \lambda_n)$, then $C_W(P) = \frac{1}{2} \sum_n \log \frac{1+\theta}{1+\lambda_n} + \frac{1}{2} \left[\frac{P + \sum(\lambda_n - \theta)}{1 + \theta}\right]$.

The capacity can be attained if and only if $P = \sum_n(\theta-\lambda_n)$. It is then attained by a Gaussian μ_{AX} with covariance operator (1), where $u_n = Ue_n$ and $\tau_n = (\theta-\lambda_n)(1+\lambda_n)^{-1}$ for all $n \geq 1$.

(b) If $P < \sum_n(\theta-\lambda_n)$, then $C_W(P) = \frac{1}{2}\sum_{n=1}^{K} \log \left[\frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)}\right]$ where $K < \infty$ is the largest integer such that $P + \sum_1^K \lambda_n \geq K\lambda_K$. The capacity can be attained by a unique Gaussian μ_{AX} with covariance operator (1), where $u_n = Ue_n$ and

$$\tau_n = \frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)} - 1 \quad \text{for } n \leq K$$

$$\tau_n = 0 \qquad\qquad \text{for } n > K.$$

Proof: (a). The fact that

$$2C_W(P) \geq \sum_{n\geq 1} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{P + \sum_m(\lambda_m-\theta)}{1+\theta}$$

follows from (c) of Theorem 2, letting $M\to\infty$ in that result. To prove the reverse inequality, suppose that $C_W(P)$ is strictly greater than its value as given in (a). Then for ε in $(0, C_W(P))$ there exists a Gaussian μ^{ε}_{AX} with covariance $R^{\varepsilon}_{AX} = \sum_n \tau^{\varepsilon}_n [R_N^{\frac{1}{2}} u^{\varepsilon}_n] \otimes [R_N^{\frac{1}{2}} u^{\varepsilon}_n]$ where $\{u^{\varepsilon}_n, n \geq 1\}$ is a c.o.n. set, all $\tau^{\varepsilon}_n \geq 0$, $\sum_n \tau^{\varepsilon}_n \|(I+S)U^* u^{\varepsilon}_n\| \leq P$, and

$C_W(P) = \frac{1}{2}\sum_n \log(1+\tau_n^\varepsilon) + \varepsilon$. Since $\sum_1^M \log(1+\tau_n^\varepsilon)$ is non-decreasing with M, for some $\varepsilon > 0$ there must exist $M^\varepsilon < \infty$ such that

$$\tfrac{1}{2}\sum_1^{M^\varepsilon} \log(1+\tau_n^\varepsilon) > \tfrac{1}{2}\sum_{n\geq 1} \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{P+\Sigma_m(\lambda_m-\theta)}{1+\theta}. \tag{5}$$

The LHS of (5) is the value of the information when $\mu_{A(X)}$ has covariance operator

$$R_{A(X)} = \sum_{n=1}^{M^\varepsilon} \tau_n^\varepsilon R_N^{\frac{1}{2}} u_n \otimes R_N^{\frac{1}{2}} u_n .$$

From Theorem 2(c), the LHS (5) can be no greater than

$$\tfrac{1}{2}\sum_1^{M^\varepsilon} \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{M^\varepsilon}{2}\log\left[\frac{1+P+\sum_{i=1}^{M^\varepsilon}(\lambda_i-\theta)}{M(1+\theta)}\right].$$

As $M^\varepsilon \to \infty$, this last expression converges upward to RHS (5). Thus, the inequality (5) cannot hold.

To see that the capacity is attained as stated in (a), one notes that from (2) the Gaussian measure with covariance (1) will achieve capacity if and only if

$$\Sigma \log(1+\tau_n) = \sum_n \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{P+\Sigma_m(\lambda_m-\theta)}{1+\theta} \tag{6}$$

and

$$\Sigma\tau_n(1+\lambda_n) \leq P, \tag{7}$$

the latter requirement following from the definition of (X_n^2), Corollary 1, and Remark 1. Both (6) and (7) are satisfied if $\tau_n = (\theta-\lambda_n)(1+\lambda_n)^{-1}$ for all $n \geq 1$ and $\Sigma(\theta-\lambda_n) = P$. Conversely, if (τ_n) satisfies (6), then

$$\frac{P+\Sigma_m(\lambda_m-\theta)}{1+\theta} \leq \sum_n (\tau_n+\lambda_n+\tau_n\lambda_n-\theta)/(1+\theta), \text{ or } P \leq \sum_n \tau_n(1+\lambda_n),$$

with equality if and only if $\tau_n+\lambda_n+\tau_n\lambda_n = \theta$ for all $n \geq 1$. If (τ_n) also satisfies (7), then necessarily $P = \Sigma\tau_n(1+\lambda_n)$, and so $\tau_n+\lambda_n+\tau_n\lambda_n = \theta$ for all $n \geq 1$. Thus, if (τ_n) satisfies both (6) and (7), $\tau_n = (\theta-\lambda_n)(1+\lambda_n)^{-1}$ for all $n \geq 1$, and $P = \sum_n(\theta-\lambda_n)$.

$$\text{(b).}\quad C_W(P) \geq \tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{\sum_{i=1}^{K} \lambda_i + P + K}{K(1+\lambda_n)} \right] \tag{8}$$

follows from (b) of Theorem 2. Suppose that $C_W(P) >$ RHS (8). Then by (4) there exists a c.o.n. set $\{v_n,\ n \geq 1\}$ and a sequence (X_n^2) with infinite number of non-zero terms (using (b) of Theorem 2) such that

$$\text{RHS (8)} \quad < \tfrac{1}{2} \sum_n \log \left[1 + X_n^2 (1 + \langle Sv_n, v_n \rangle)^{-1}\right] \tag{9}$$

with $\sum_n X_n^2 \leq P$. Since RHS(9) is finite and the sum of non-negative terms, there must exist $M < \infty$ such that

RHS (8) $< \frac{1}{2} \sum_{n=1}^{M} \log \left[1 + X_n^2 (1 + \langle Sv_n, v_n \rangle)^{-1}\right]$. This contradicts (b) of Theorem 2.

$\square$

<u>Lemma 5</u>: If $\theta < \infty$ and $S \geq \theta I$, then $C_W(P) = \dfrac{P}{2(1+\theta)}$.

<u>Proof</u>: $C_W(P) \geq P(1+\theta)^{-1} 2^{-1}$ follows from part (a) of Theorem 2 by letting $M \to \infty$. To prove the reverse inequality, one notes that for the constraint $E_{\mu_X} \|R_N^{-\frac{1}{2}} A(X)\|^2 \leq (1+\theta)^{-1} P$, the capacity $C_N([1+\theta]^{-1} P)$ is $P(1+\theta)^{-1} 2^{-1}$ [1, Theorem 2]. Since $E_{\mu_X} \|R_W^{-\frac{1}{2}} A(X)\|^2 \leq P$ implies $E_{\mu_X} \|R_N^{-\frac{1}{2}} A(X)\|^2 \leq (1+\theta)^{-1} P$, optimization w.r.t. the former constraint is over a smaller set than w.r.t. the latter constraint; thus $C_W(P) \leq C_N[(1+\theta)^{-1} P]$.

$\square$

The capacity for channels with an infinite-dimensional signal space can now be given.

<u>Theorem 3</u>: Suppose that $\theta < \infty$, H_2 is infinite-dimensional, and $\dim[\text{supp}(\mu_{AX})]$ is not constrained.

(a) If $\{\lambda_n, n \geq 1\}$ is not empty, and $\sum_n (\theta - \lambda_n) \leq P$, then

$$C_W(P) = \tfrac{1}{2} \sum_n \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \tfrac{1}{2} \, \frac{P + \sum_m (\lambda_m - \theta)}{1 + \theta} \quad .$$

(b) If $\{\lambda_n, n\geq 1\}$ is not empty, and $P < \Sigma_n(\theta-\lambda_n)$, then there exists a largest integer K such that $\sum_1^K \lambda_i + P \geq K\lambda_K$, and $C_W(P) = \frac{1}{2}\sum_{n=1}^K \log\left[\frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)}\right]$.

(c) If $\{\lambda_n, n\geq 1\}$ is empty, then $C_W(P) = \frac{P}{2(1+\theta)}$.

(d) In (a), the capacity can be attained if and only if $\Sigma_n(\theta-\lambda_n) = P$. It is then attained by a Gaussian μ_{AX} with covariance operator as in (1), where $u_n = Ue_n$ and $\tau_n = (\theta-\lambda_n)(1+\lambda_n)^{-1}$ for all $n\geq 1$. In (b), the capacity can be attained by a unique Gaussian μ_{AX} with covariance operator (1), where $u_n = Ue_n$ and $\tau_n = \frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)} - 1$ for $n \leq K$; $\tau_n = 0$ for $n > K$. In (c), the capacity cannot be attained.

<u>Proof</u>: From the preceding, one must find sup $\Sigma_n \log[1+\tau_n]$, where (τ_n) is a non-negative summable sequence, subject to the constraint $\Sigma_n\tau_n[1 + \langle USU^*u_n,u_n\rangle] \leq P$, where $\{u_n, n\geq 1\}$ is any c.o.n. set in H_2, and $R_N^{\frac{1}{2}} = R_W^{\frac{1}{2}}(I+S)^{\frac{1}{2}}U^*$, with U unitary.

Let Q be the projection operator onto the (closed linear) subspace spanned by $\{Ue_n, n\geq 1\}$, where $\{e_n, n\geq 1\}$ are eigenvectors of $S-\theta I$ corresponding to strictly negative eigenvalues. Let $Q^\perp$ be the projection onto the orthogonal complement of range(Q) Range(Q) is obviously an invariant subspace for $U(S-\theta I)U^*$ and thus for USU^*; since USU^* is self-adjoint, range($Q^\perp$) is also invariant for USU^*.

The set of covariance operators $T = \Sigma\tau_n u_n \otimes u_n$ satisfying the above constraint consists of those T such that Trace $T^{\frac{1}{2}}[Q + QUSU^*Q]T^{\frac{1}{2}} = P_1$ and Trace $T^{\frac{1}{2}}[Q^{\perp} + Q^{\perp}USU^*Q^{\perp}]T^{\frac{1}{2}} \leq P - P_1$, where $P_1 = P_1(T)$ is contained in $[0,P]$.

Thus, the capacity problem is to determine $\sup_{P_1 \text{ in } [0,P]} \sup_{A_1 \cap A_2} \text{Trace} \log (I + T)$, where

$$A_1 = A_1(P_1) = \{\text{covariance operators } T\text{: Tr } T^{\frac{1}{2}}Q(I + USU^*)QT^{\frac{1}{2}} = P_1\}$$

and

$$A_2 = A_2(P_1) = \{\text{cov. operators } T\text{: Tr } T^{\frac{1}{2}}Q^{\perp}(I + USU^*)Q^{\perp}T^{\frac{1}{2}} \leq P - P_1\}.$$

Now, $\sup_{P_1 \text{ in } [0,P]} \sup_{A_1 \cap A_2} \text{Tr} \log (I + T)$

$$\leq \sup_{P_1 \text{ in } [0,P]} \left\{ \sup_{A_1} \text{Tr} \log (I + T_1) + \sup_{A_2} \text{Tr} \log (I + T_2) \right\}.$$

It is shown below that this inequality is actually an equality.

Proof of (a). First suppose that $\{\lambda_n, n \geq 1\}$ is an infinite set, and fix P_1 in $[0,P]$. Suppose that $\Sigma(\theta - \lambda_n) > P_1$. By Lemma 3, there exists a largest integer K such that $\sum_1^K \lambda_n + P \geq K\lambda_K$. $\sup_{A_1(P_1)} \frac{1}{2}$ Trace $\log (I + T)$ is then (from part (b) of Lemma 4) $\frac{1}{2} \sum_{n=1}^{K} \log \left[\frac{\sum_{i=1}^{K} \lambda_i + P_1 + K}{K(1 + \lambda_n)} \right]$. From Lemma 5, $\sup_{A_2(P_1)} \frac{1}{2}$ Trace $\log (I + T) = \frac{1}{2} (P - P_1)(1+\theta)^{-1}$. Thus, for this value of P_1,

$$\sup_{A_1 \cap A_2} \tfrac{1}{2} \text{Trace} \log (I + T) \leq C_+(P_1, K) \equiv \tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{\sum_{i=1}^{K} \lambda_i + P_1 + K}{K(1 + \lambda_n)} \right] + \tfrac{1}{2}(P - P_1)(1+\theta)^{-1}.$$

$$\frac{d}{dP_1} C_+(P_1,K) = \frac{K}{2}\,\frac{1}{\sum\limits_1^K \lambda_i + P_1 + K} - \frac{1}{2(1+\theta)}$$ whose sign is determined by that

of $K\theta - P_1 - \sum\limits_1^K \lambda_i$, since $\sum\limits_1^K \lambda_i + P_1 + K > 0$ (because $\sum\limits_1^K \lambda_i + P_1 \geq K\lambda_K$ and $\lambda_K > -1$).

$P_1 + \sum\limits_1^K \lambda_i < K\lambda_{K+1} < K\theta$, so C_+ is increasing for P_1 increasing when

$K\lambda_K \leq P_1 + \sum\limits_1^K \lambda_i < K\lambda_{K+1}$. Define $P_1^K = -\sum\limits_{i=1}^{K+1} \lambda_i + (K+1)\lambda_{K+1}$. Then

$$2C_+(P_1^K,K) = \sum_{n=1}^{K} \log\left[\frac{1+\lambda_{K+1}}{1+\lambda_n}\right] + \frac{P + \sum\limits_1^K \lambda_i - K\lambda_{K+1}}{1+\theta}$$. Using the inequalities

$\theta \geq \lambda_{K+2} \geq \lambda_{K+1}$ and $1 + \lambda_{K+2} > 0$, $C_+(P^K,K)$ is seen to be a strictly increasing function of K. Since $\sum(\theta - \lambda_n) \leq P$, $K(\theta - \lambda_K) \to 0$, and thus

$$\lim_K P_1^K = \lim_K \left[-\sum_{i=1}^{K+1} (\lambda_i - \theta) + (K+1)(\lambda_{K+1} - \theta)\right] = \sum_1^{\infty}(\theta - \lambda_i).$$

This gives as an upper bound for the capacity, for all P_1 such that $P_1 < \sum\limits_n(\theta - \lambda_n)$,

$$\lim_K C_+(P^K,K) = \tfrac{1}{2} \sum_n \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \tfrac{1}{2}\left[\frac{P - \sum(\theta - \lambda_n)}{1+\theta}\right].$$

Since $P > \sum(\theta - \lambda_n)$, there exist P_1 values satisfying the constraint with $P_1 \geq \sum(\theta - \lambda_n)$. In this case, from part (a) of Lemma 4,

$$C(P_1) = \tfrac{1}{2} \sum_n \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \tfrac{1}{2}\left[\frac{P_1 + \sum(\lambda_n - \theta)}{1+\theta}\right] + \tfrac{1}{2}\,\frac{P - P_1}{1+\theta}$$

$$= \tfrac{1}{2} \sum_n \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \tfrac{1}{2}\,\frac{[P + \sum_n(\lambda_n - \theta)]}{1+\theta}.$$

The last expression is thus an upper bound on $C_W(P)$ for all P_1 in $[0,P]$.

From part (a) of Lemma 4, this is the value of the capacity when the operator $S - \theta I$ is negative definite with an infinite set of strictly negative eigenvalues. Thus, this is the capacity when $P_1 = P$, and so (a) is proved if $\{\lambda_n, n\geq 1\}$ is an infinite set.

Suppose next that the set $\{\lambda_n, n\geq 1\}$ is finite, $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_K < \theta$. Proceeding as above, for P_1 such that $P_1 + \sum_1^K \lambda_n < K\lambda_K$, the upper bound $C(P_1,J)$ on the capacity is increased by increasing P_1 up to the value $P_1 = K\lambda_K - \sum_1^K \lambda_n$. If $P_1 + \sum_1^K \lambda_n > K\lambda_K$, then from above,

$$C(P_1,K) = \tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{\sum_{i=1}^{K} \lambda_i + P_1 + K}{K(1+\lambda_n)}\right] + \tfrac{1}{2}\, \frac{(P - P_1)}{1+\theta} \quad .$$

Differentiating w.r.t. P_1, one sees that the derivative is positive for $\sum_1^K \lambda_i + P_1 - K\theta < 0$, negative for $\sum_1^K \lambda_i + P_1 - K\theta > 0$, and so the unique maximum occurs for $P_1 = \sum_1^K (\theta - \lambda_i)$. This gives an upper bound on the capacity of

$$\tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \tfrac{1}{2}\, \frac{P + \sum_1^K (\lambda_i - \theta)}{1+\theta} \quad .$$

To show that this upper bound is actually the capacity, a sequence (T_M^{ε}) of covariance operators will be exhibited, each satisfying the constraint, and such that $\sup_{\varepsilon>0, M>K} \tfrac{1}{2} \operatorname{trace} \log (I + T_M^{\varepsilon})$ is equal to the upper bound. Thus, fix $\varepsilon > 0$ such that $\varepsilon < 1+\theta$. For $M > K$, define T_M^{ε} by $T_M^{\varepsilon} = \sum_{n=1}^{M} \tau_n^{\varepsilon,M} u_n^{\varepsilon} \otimes u_n^{\varepsilon}$ where

$$\begin{aligned}
\tau_n^{\varepsilon,M} &= (\theta - \lambda_n)(1+\lambda_n)^{-1} && \text{for} \quad 1 \leq n \leq K \\
&= \frac{P + \sum_1^K (\lambda_i - \theta)}{(M-K)(1+\theta+\varepsilon)} && \text{for} \quad K < n \leq M \\
&= 0 && n > M\ .
\end{aligned}$$

Set $u_n^\varepsilon = Ue_n$, $n = 1,\ldots,K$. Choose the o.n. elements u_n^ε for $n = K+1,\ldots,M$ such that $\langle USU^* u_n^\varepsilon, u_n^\varepsilon\rangle < \theta + \varepsilon$; this is possible for any M, because θ is the smallest limit point of the spectrum of S. One now obtains

$$\sum_{n=1}^{M} \tau_n^{\varepsilon,M}[1 + \langle USU^* u_n^\varepsilon, u_n^\varepsilon\rangle] \le \sum_{n=1}^{K} (\theta - \lambda_n) + \sum_{n=K+1}^{M} \tau_n^{\varepsilon,M}[1 + \theta + \varepsilon] = P.$$

Moreover

$$\tfrac{1}{2} \text{ Trace } \log (I + T_M^\varepsilon) = \tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{M-K}{2} \log \left[1 + \frac{P + \sum_{i=1}^{K} (\lambda_i - \theta)}{(M-K)(1+\theta+\varepsilon)}\right].$$

For fixed ε, the limit as $M\to\infty$ of thisexpression is

$$\tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{P + \sum_1^K(\lambda_i - \theta)}{2(1+\theta+\varepsilon)} .$$

Since $\varepsilon > 0$ is arbitrary, one sees that the supremum over all $\varepsilon > 0$, $M > K$ is equal to the upper bound previously obtained. That upper bound is thus the capacity $C_W(P)$, completing the proof of (a).

The result of (b) can be obtained from the proof of (a). Since now $\sum_n(\theta - \lambda_n) > P$, there exists a largest integer K such that $P + \sum_1^K \lambda_n \ge K\lambda_K$, by Lemma 3. Choose $P_1 \le P$ and proceed as in the proof of (a) to obtain an upper bound on the capacity of

$$C(P_1,M) = \tfrac{1}{2} \sum_{n=1}^{M} \log \left[\frac{\sum_{i=1}^{M} \lambda_i + M + P_1}{M(1+\lambda_n)}\right] + \tfrac{1}{2}\, \frac{(P - P_1)}{1 + \theta}$$

where M is the largest integer (note $M \le K$) such that $P_1 + \sum_1^M \lambda_n \ge M\lambda_M$. Defining P_1^M as in the proof of (a), the sequence $(C[P_1^M,M])$ is non-decreasing as M increases, and since K is the largest integer such that $P + \sum_1^K \lambda_n \ge K\lambda_K$, one has that $C[P,K]$ is an upper bound on the capacity. This is the value of the capacity, from (b) of Lemma 4, when $S - \theta I$ is negative definite (i.e., $P = P_1$).

(c) follows from Lemma 5.

The statement of (d) that the capacity cannot be attained when $S - \theta I$ is non-negative definite (part (c)) can be proved by noting that $R_N^{\frac{1}{2}} = R_W^{\frac{1}{2}}(S - \theta I + (1+\theta)I)^{\frac{1}{2}}U^*$. Thus $||x||_N^2 = ||R_N^{-\frac{1}{2}}x||^2$ $\leq (||S - \theta I + (1+\theta)I||^{-1})||R_W^{-\frac{1}{2}}x||^2 \leq (1+\theta)^{-1}||x||_W^2$. A solution attaining the capacity $P/[2(1+\theta)]$, subject to the constraint $E_{\mu_X}||A(x)||_W^2 \leq P$, would thus satisfy the constraint $E_{\mu_X}||A(x)||_N^2 \leq P/(1+\theta)$; this is impossible, by [1, Theorem 2].

The statements in (d) on attaining the capacity in parts (a) and (b) follow directly from corresponding statements in Lemma 4, as follows. The fact that the capacity in (a) cannot be attained when $\{\lambda_n, n\geq 1\}$ is an infinite set follows from the fact that $P_1 = P$ (in the proof of (a)) uniquely gives the capacity, and this gives the same result as when $S - \theta I$ is negative definite. The fact that the capacity cannot then be attained is contained in part (a) of Lemma 4. If $\{\lambda_n, n\geq 1\}$ is a finite (nonempty) set, then the capacity is uniquely obtained by setting $P - P_1 = P + \sum_1^K(\lambda_n - \theta)$, corresponding to the constraint Trace $T^{\frac{1}{2}}[Q^{\perp} + Q^{\perp}USU^*Q^{\perp}]T^{\frac{1}{2}}$. Since $Q^{\perp} + Q^{\perp}USU^*Q^{\perp}$ is non-negative definite, application of the result for part (c) shows that the capacity cannot be attained. Finally, the statements on attaining capacity in case (b) follow directly from part (b) of Lemma 4, since the capacity in (b) is uniquely obtained by setting $P_1 = P$, equivalent to $S - \theta I$ being negative definite.

□

Corollary 3: If $\theta = \infty$, then $C_W(P)$ has the value given in Theorem 3(b), and can be similarly attained.

Proof: Apply Proposition 2 and the proof of Theorem 3(b).

Comparison of $C_W(P)$ and $C_N(P)$

For the finite-dimensional channel, the capacity $C_W(P)$ given in Theorem 1 is strictly greater than $C_N(P)$ $(= \frac{1}{2} \log [1 + P/M])$ if $\sum_1^M \beta_i \leq 0$, or if $P + \sum_1^K \beta_i \leq 0$. $C_W(P) \leq C_N(P)$ if $0 \leq \beta_1 < \beta_M$. The verification is omitted.

For the infinite-dimensional channel, a general statement can be made if $\{\lambda_n, n \geq 1\}$ is empty. Then, $C_W(P) > C_N(P)$ if $\theta < 0$, $C_W(P) < C_N(P)$ if $\theta > 0$, $C_W(P) = C_N(P)$ if $\theta = 0$; see Theorem 2 (a) and Theorem 3 (c). Note that $C_N(P) = P/2$ for the unconstrained channel [1, Theorem 2].

If $\{\lambda_n, n \geq 1\}$ is not empty, then for the unconstrained channel the value of $C_W(P)$ given in Theorem 3(a) is greater than $\frac{P}{2(1+\theta)}$, using $\log x^{-1} \geq 1-x$. This inequality can also be shown for the value given in Theorem 3(b), proceeding as in the proof of part (b) of the Theorem in [2]. Thus, for the unconstrained channel, $C_W(P) > C_N(P)$ if $\theta \leq 0$ and $\{\lambda_n, n \geq 1\}$ is not empty. A similar result can be obtained for the constrained channel.

Discussion

The mismatched channel differs from the matched channel in several ways. First, the value of the capacity can be very different, as already seen. Secondly, the problem of attaining capacity is much more significant. Even in the finite-dimensional channel the vectors $u_1, \ldots, u_M$ must be a specific set of vectors, not just any o.n. set. If H_2 is infinite-dimensional with $\dim[\mathrm{supp}(\mu_{A(X)})] \leq M$, the situation is even worse in (c) of Theorem 2. That is, capacity can then be attained only if S has zero as an eigenvalue of multiplicity $\geq M$ when $S \leq \theta I$, or of multiplicity $\geq M-K$ when S has $K < M$ eigenvalues $\lambda_1 \leq \ldots \leq \lambda_K < \theta$ and $P + \sum_1^K \lambda_i \geq K\lambda_K$. Otherwise, in order to approach capacity, one will need to put part of the available "energy" P in elements (Ue_n) where (e_n) are eigenvectors of S corresponding to successively smaller eigenvalues. In practical applications, this typically corresponds to eigenfunctions at higher and higher frequencies.

For the infinite-dimensional channel without a constraint on $\dim[\mathrm{supp}(\mu_{AX})]$, again there can be significant differences between $C_W(P)$ and $C_N(P)$, depending on $\{\theta; \lambda_n, n \geq 1\}$. However, in this case one sees a rather different situation in the problem of attaining capacity. $C_N(P)$ can never be attained; $C_W(P)$ can be attained if and only if $\{\lambda_n, n\geq 1\}$ is not empty and $P \leq \Sigma_n(\theta - \lambda_n)$.

It may be noted that the results given in Theorem 1 and Theorem 2(b) are similar to those obtained in [4, p. 170], although the developments are quite different. However, these previous results are given in terms of a constraint on $E\|A(X)\|^2$, and assume that the noise variance components can be arranged in ascending order. This can only be done if the channel is finite-dimensional. In that case, one can take $R_W = I$, the identity, and thereby use a true power constraint. (A-2) then becomes $R_N = I + S$, and the capacity is as given in Theorem 1; this agrees with the referenced results in [4].

Applications and Extensions

The results given here provide a complete and general solution to the information capacity problem for the Gaussian channel without feedback, so long as the constraint can be given in terms of any covariance (or RKHS norm). Moreover, the formulation of the problem as developed here, and the availability of these results, are already leading to a number of related results, and additional applications and extensions seem likely.

For example, it is well-known that feedback does not increase information capacity of a large class of matched Gaussian channels, including the "white noise" (Wiener process as noise) channel [6], [15]. It can be shown, using the results given here, that capacity is increased by feedback for a large class of mismatched channels, thus validating a long-held conjecture. In another direction, as discussed above, these results enable one to analyze jamming channels when information capacity is used as the criterion. Other related results can also be obtained, based directly or indirectly on the formulation and results given here. Examples include capacity-per-unit time for mismatched channels with and without feedback, coding capacity for various types of channels, and new relations between optimum filtering and optimum transmission in the Shannon sense.

The framework used here requires that the signal and noise sample functions lie in a real separable Hilbert space. This is easily extended to separable Banach spaces, such as $C[0,1]$ (see [1, p.88]).

Appendix

Proof of Lemma 1: Define f_M: $\mathbf{R}^M \to \mathbf{R}$ by $f_M(\underline{y}) = \sum_{n=1}^{M} \log\,[1 + y_n\,\gamma_n^{-1}]$. f_M is to be maximized subject to the constraints

$$g(\underline{y}) = \sum_{1}^{M} y_n - P \leq 0$$

$$h_i(\underline{y}) = -y_i \leq 0, \quad i = 1,\ldots,M.$$

This is a constrained optimization problem with objective function f_M which is strictly concave over the convex set $\{Z \text{ in } \mathbf{R}^M;\ Z_i \geq 0,\ i = 1,\ldots,M\}$. Moreover, each constraint function is linear. Thus, a solution to this problem will define a unique global maximum for f_M[14]. In order that $\underset{\sim}{y^*}$ be this unique solution, it is necessary and sufficient that the following set of equations be satisfied [14]:

$$\frac{1}{y_i^* + \gamma_i} + \beta - \alpha_i = 0 \qquad i = 1,\ldots,M \tag{a-1}$$

$$\sum_{1}^{M} y_n^* - P \leq 0, \quad \beta\Big[\sum_{1}^{M} y_n^* - P\Big] = 0 \tag{a-2}$$

$$-y_i^* \leq 0, \quad \alpha_i y_i^* = 0,\ i = 1,\ldots,M \tag{a-3}$$

for some set of non-positive real numbers $\{\beta,\alpha_1,\ldots,\alpha_M\}$.

First, attempt to obtain a solution by setting $\alpha_1 = \alpha_2 = \ldots = \alpha_M = 0$. This requires $\beta(\gamma_i + y_i^*) = -1$ for $i = 1,\ldots,M$; thus,

$$\sum_{1}^{M} y_i^* + \sum_{1}^{M} \gamma_i = -M\beta^{-1}, \text{ and so } \ y_n^* = \Big(\sum_{1}^{M} y_i^* + \sum_{1}^{M} \gamma_i\Big)/M - \gamma_n$$

for $n = 1,2,\ldots,M$. This definition of $\underset{\sim}{y^*}$ and the constraints (a-3) require that

$$\sum_{1}^{M} y_i^* + \sum_{1}^{M} \gamma_i \geq M\gamma_n$$

for $n \leq M$; this inequality is satisfied for all $n \leq M$ if and only if it is satisfied for $n = M$. Also, $\beta^{-1} = -(y_i^* + \gamma_i)$ for $i \leq M$ implies $\beta < 0$, so that $\sum_{1}^{M} y_i^* = P$ by constraints (a-2). Thus, if $P + \sum_1^M \gamma_i \cdot M\gamma_M$, an optimum solution is given by

$$y_i^* = (P + \sum_1^M \gamma_n - M\gamma_i)/M, \qquad i \leq M.$$

If there exists $K < M$ such that $K\gamma_K \leq P + \sum_1^K \gamma_i < (K+1)\gamma_{K+1}$, then constraints (a-1)-(a-3) are satisfied by choosing $\beta = -K[P + \sum_1^K \gamma_i]^{-1}$,

$$\alpha_1 = \alpha_2 = \ldots = \alpha_K = 0,$$

$$\sum_1^K y_i^* = P$$

$$y_i^* = 0, \quad i > K$$

$$y_i^* = K^{-1}[P + \sum_1^K \gamma_n - K\gamma_i], \quad i \leq K$$

$$\alpha_i = -K[P + \sum_1^K \gamma_n]^{-1} + \gamma_i^{-1}, \quad i > K.$$

Thus,

$$\sup_{\{\underset{\sim}{X}:\ \sum_1^M X_n^2 \leq P\}} \prod_{n=1}^M (\gamma_n + X_n^2)/\gamma_n = \prod_{n=1}^K (\Sigma\gamma_i + P)/(K\gamma_n)$$

where $K \leq M$ is the largest integer such that $\sum_1^K \gamma_i + P \geq K\gamma_K$. The supremum is attained by $\underset{\sim}{y}^*$ as defined above, or for

$$X_n^2 = [P + \sum_1^K \gamma_i]/K - \gamma_n \qquad n \leq K$$

$$= 0 \qquad n > K.$$

□

<u>Proof of Lemma 2</u>: Note that $f_K(\underset{\sim}{\gamma})$ increases with decreasing γ_n for $1 \leq n \leq K$, using $\sum_1^K \gamma_i + P \geq K\gamma_K$. Also, one can assume that $\sum_{n=1}^K \gamma_n = \sum_{n=1}^K \gamma_n$. To see this,

suppose that $\sum_{n=1}^{K} \gamma_n > \sum_{n=1}^{K} \lambda_n$. Define admissible sequences $(\gamma_n^{(j)})$, $j \geq 1$, as follows. $\gamma_n^{(1)} = \gamma_n$, $1 \leq n \leq K$. Given $(\gamma_n^{(j)})$, let p_j be the largest integer $i \leq K$ such that $\sum_{n=1}^{p_j} (\gamma_n^{(j)} - \lambda_n) = 0$. If no such integer exists for $j = 1$, set $p_1 = 0$. Define $(\gamma_n^{(j+1)})$ by

$$
\begin{aligned}
\gamma_n^{(j+1)} &= \gamma_n^{(j)} & n \leq p_j \\
&= \gamma_n^{(j)} - \varepsilon_j & p_j < n \leq K
\end{aligned}
$$

where

$$
\varepsilon_j = \min\left\{\frac{1}{J - p_j} \sum_{n=p_j+1}^{J} (\gamma_n^{(j)} - \lambda_n) : \; p_j < J \leq K\right\}
$$

To see that $(\gamma_n^{(j+1)})$ is non-decreasing, it suffices to check that $\gamma_{p_j+1}^{(j+1)} \geq \gamma_{p_j}^{(j+1)}$. This holds because $\varepsilon_j \leq \gamma_{p_j+1}^{(j)} - \gamma_{p_j}^{(j)}$, noting that $\lambda_{p_j+1} \geq \lambda_{p_j} \geq \gamma_{p_j}^{(j)}$ and $\varepsilon_j \leq \gamma_{p_j+1}^{(j)} - \lambda_{p_j+1}$. Also, $\sum_{n=1}^{J} (\gamma_n^{(j+1)} - \lambda_n) \geq 0$ for $1 \leq J \leq K$: by the similar property for $(\gamma_n^{(j)})$ for $J \leq p_j$; by the definition of ε_j if $J > p_j$. Finally, $\sum_1^K \gamma_n^{(j+1)} + P = \sum_1^K (\gamma_j^{(j)}) + P - (K - p_j)\varepsilon_j \geq K\gamma_K - (K-p_j)\varepsilon_j \geq K\gamma_K^{(j)} - K\varepsilon_j = K\gamma_K^{(j+1)}$. The sequence $(\gamma_n^{(j+1)})$ is thus admissible.

Since $P_{j+1} > p_j$, the above procedure must terminate in at most $K - p_1$ steps. Moreover, $\gamma_n^{(j+1)} \leq \gamma_n^{(j)}$ for $1 \leq n \leq K$, so that $f_K[(\gamma_n^{(j+1)})] \geq f_K[(\gamma_n^{(j)})]$.

Assuming now that $\sum_1^K \gamma_n = \sum_1^K \lambda_n$, it is sufficient to show that $\prod_1^K \gamma_n \geq \prod_1^K \lambda_n$ with equality if and only if $\gamma_n = \lambda_n$, $1 \leq n \leq K$.

Define $(\beta_n^{(1)}) = (\gamma_n)$; given $(\beta_n^{(j)})$ such that $\sum_{n=1}^{K} \beta_n^{(j)} = \sum_{n=1}^{K} \lambda_n$ and $(\beta_n^{(j)}) \neq (\lambda_n)$, define $(\beta_n^{(j+1)})$ as follows. Let b_j be the largest integer i such that $\beta_i^{(j)} \neq \lambda_i$; necessarily $\lambda_{b_j} > \beta_{b_j}^{(j)}$. Let a_j be the largest integer i such that $\lambda_i < \beta_i^{(j)}$. Let $\Delta_j = \min(\lambda_{b_j} - \beta_{b_j}^{(j)}, \beta_{a_j}^{(j)} - \lambda_{a_j})$. Define the sequence $(\beta_n^{(j+1)})$ by

$$\beta_n^{(j+1)} = \beta_n^{(j)} \qquad n \notin \{a_j, b_j\}$$

$$\beta_{b_j}^{(j+1)} = \beta_{b_j}^{(j)} + \Delta_j$$

$$\beta_{a_j}^{(j+1)} = \beta_{a_j}^{(j)} - \Delta_j.$$

Clearly $\sum_{n=1}^{K} \beta_n^{(j+1)} = \sum_{n=1}^{K} \beta_n^{(j)} = \sum_{n=1}^{K} \lambda_n$. Since $\beta_{b_j}^{(j+1)} \geq \beta_{b_j}^{(j)} \geq \gamma_{b_j}$ and $\beta_{a_j}^{(j+1)} \leq \beta_{a_j}^{(j)} \leq \gamma_{a_j}$, and (γ_n) is non-decreasing, $\beta_{b_j}^{(j+1)} \geq \beta_{a_j}^{(j+1)}$. Thus,

$$\prod_{n=1}^{K} \beta_n^{(j+1)} = \prod_{n \neq a_j, b_j} \beta_n^{(j)} [\beta_{b_j}^{(j)} + \varepsilon][\beta_{a_j}^{(j)} - \varepsilon] = \prod_{n=1}^{K} \beta_n^{(j)} + \prod_{n \neq a_j, b_j} \beta_n^{(j)} [-\varepsilon(\beta_{b_j}^{(j)} - \beta_{a_j}^{(j)}) - \varepsilon^2]$$

$< \prod_{n=1}^{K} \beta_n^{(j)}$. $\beta_n^{(j)} \geq \lambda_n$ if $\gamma_n > \lambda_n$; $\lambda_n \geq \beta_n^{(j)} \geq \gamma_n$ if $\gamma_n < \lambda_n$; $\beta_n^{(j)} = \lambda_n$ if $\gamma_n = \lambda_n$, these relations holding for all $j \geq 1$. There are at most K elements (γ_n) such that $\gamma_n \neq \lambda_n$, and the number of such elements is reduced by at least one whenever the sequence $(\beta_n^{(j+1)})$ is formed from $(\beta_n^{(j)})$. The procedure must terminate in at most K steps, and will terminate when and only when a sequence

$(\beta_n^{(j)})$ is formed with $(\beta_n^{(j)}) = (\lambda_n)$. Since $\sum_{n=1}^{K} \beta_n^{(j)} = \sum_{1}^{K} \lambda_n$ and

$\prod_{n=1}^{K} \beta_n^{(j+1)} < \prod_{n=1}^{K} \beta_n^{(j)}$ for all $j \geq 1$, one has that $f_K(\underline{\lambda}) > f_K(\underline{\gamma})$.

$\square$

References

1. C.R. Baker, Capacity of the Gaussian channel without feedback, Information and Control, 37, 70-89 (1978).

2. C.R. Baker, Channel models and their capacity, in "Essays in Statistics: Contributions in Honor of Norman L. Johnson," pp. 1-16, P.K. Sen, ed., North-Holland, New York (1983).

3. P.M. Ebert, The capacity of the Gaussian channel with feedback, Bell System Tech. J., 49, 1705-1712 (1970).

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

5. R.G. Gallager, "Information Theory and Reliable Communication," Wiley, New York (1968).

6. T.T. Kadota, M. Zakai, and J. Ziv, Mutual information of the white Gaussian channel with and without feedback, IEEE Trans. on Information Theory, 17, 368-371 (1971).

7. E.J. Kelly, I.S. Reed, and W.L. Root, The detection of radar echoes in Noise, Part I, Siam J. Applied Mathematics, 8, 309-341 (1960).

8. R.G. Douglas, On majorization, factorization, and range inclusion of operators in Hilbert space, Proceedings Amer. Math. Soc., 17, 413-415 (1966).

9. C.R. Baker, On equivalence of probability measures, Annals of Probability, 1, 690-698 (1973).

10. K. Itô, The topological support of Gauss measure on Hilbert space, Nagoya Math. J., 38, 181-183 (1970).

11. F. Riesz and B. Sz.-Nagy, "Functional Analysis," Ungar, New York (1955).

12. K. Yanagi, On the capacity of a class of mismatched Gaussian channels, IEEE Trans. Information Theory, 30, 840-842 (1984).

13. K. Fan, On a theorem of Weyl concerning eigenvalues of a linear transformation. I, Proc. National Academy of Sciences, 35, 652-655 (1949).

14. G.R. Walsh, "Methods of Optimization," Wiley, New York (1975).

15. M. Hitsuda and S. Ihara, Gaussian channels and the optimum coding, J. Multivariate Analysis, 5, 106-118 (1975).

END

DTIC

8—86