UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release: Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Department of Statistics; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code): University of North Carolina, Chapel Hill, North Carolina 27514	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-86-K-0039
8c. ADDRESS (City, State and ZIP Code): Statistics & Probability Program, Arlington, VA 22217	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
NR	042	269	SRC 105

11. TITLE (Include Security Classification): Information Capacity of Gaussian Channels

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 ______	October, 1986	20

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Gaussian channels; channel capacity; Shannon theory; Information Theory.

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

Information capacity of Gaussian channels is one of the basic problems of information theory. Shannon's results for white Gaussian channels and Fano's "waterfilling" analysis of stationary Gaussian channels are two of the best-known works of early information theory. Results are given here which extend to a general framework these results and others due to Gallager and to Kadota, Zakai, and Ziv. The development applies to arbitrary Gaussian channels when the channel noise has sample paths in a separable Banach space, and to a large class of Gaussian channels when the noise has sample paths in a linear topological vector space. Solutions for the capacity are given for both matched and mismatched channels.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL: C.R. Baker	22b. TELEPHONE NUMBER (Include Area Code): (919) 962-2189; 22c. OFFICE SYMBOL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Information Capacity of Gaussian Channels

Charles R. Baker

Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

LISS 18
October, 1986

To appear: Recent Advances in Communication and Control Theory, R.E. Kalman, G.I. Marchuk, A.E. Ruperti, and A.J. Viterbi, eds., New York: Optimization Software, Inc. (1987).

Research supported by ONR Contract N00014-86-K-0039

Abstract

Information capacity of Gaussian channels is one of the basic problems of information theory. Shannon's results for white Gaussian channels and Fano's "waterfilling" analysis of stationary Gaussian channels are two of the best-known works of early information theory. Results are given here which extend to a general framework these results and others due to Gallager and to Kadota, Zakai, and Ziv. The development applies to arbitrary Gaussian channels when the channel noise has sample paths in a separable Banach space, and to a large class of Gaussian channels when the noise has sample paths in a linear topological vector space. Solutions for the capacity are given for both matched and mismatched channels.

Introduction

The modern theory of information is largely based on the pioneering work of C.E. Shannon [1]. The contributions and importance of information theory to the advancement of technology are very well known, and need not be summarized here. However, new applications of a different nature seem likely to arise in the not-far-distant future. Some of these potential applications would require a much deeper development of the theory than has been needed heretofore. This is in part because of rapid advances in technology in areas such as computers and communications. Thus, one may envision computers of such high capability that their optimum use will require mathematical models using infinite-dimensional methods. Fiber optics is already leading to communication channels of extremely high bandwidth. Also to be considered is the need to develop information-theoretic models and methods for applications which do not fit into the classical mold of a communications channel with stationary Gaussian noise or a discrete memoryless channel. On the one hand, some communication channels contain nonstationary noise as a major source of interference. In another direction, information theory is viewed as a means of evaluating and designing systems in areas such as image processing, artificial intelligence, and surveillance.

Thus, the scope of information theory as presently applied may require considerable expansion in order to meet the needs of the future. In particular, mathematical models may be needed for problems of a very general nature, including channels with memory, which may be infinite-dimensional, nonstationary, and possibly nonGaussian.

The present article gives a treatment of capacity for Gaussian channels in a very general setting: when the stochastic processes of interest induce measures on a linear topological vector space. The work is an extension of previous results for induced measures on a separable Hilbert space [2], [3]. Although the latter model will be sufficiently general for most applications, it is not likely to be adequate for a treatment of nonstandard applications such as random fields, artificial intelligence, and surveillance.

In the case of stochastic processes with sample functions belonging to a separable Hilbert space, the results given in [2] and [3] represent a substantial generalization of previous work. This previous work includes Shannon's original white noise channel [1], Gallager's further work on this model [4], Kadota, Zakai, and Ziv's work on the Wiener channel [5], and the results of Fano [6] and Gallager [4] for stationary Gaussian channels. All of this prior work makes various assumptions on the channel noise.

Of course, in practical applications the coding capacity is most important. Partial results in this area for these more general models have been obtained [7], [8]. It can be expected that more complete solutions of the coding capacity problem will require the availability of general results on information capacity such as those summarized here, since proofs of coding capacity typically involve use of the information capacity.

This paper discusses the general framework in which these problems have been solved, and summarizes the solutions. Proofs will not be included; it will be seen that one can modify the proofs of the Hilbert space solutions given in [2] and [3]. This has already been done in [7] for the case of the "matched" channel analyzed in [2], and similar methods can be used for the "mismatched" channel considered in [3]. Thus, the development here will be limited to defining the framework of the problem, providing the supplementary details needed to adapt the Hilbert-space solutions and proofs of [2] and [3] to the present more general setup, and then stating the results.

Mutual Information and Channel Capacity

Let (X,β) and $(Y,\mathcal{F})$ be two measurable spaces, with μ_{XY} a probability on $(X\times Y,\ \beta\times\mathcal{F})$. For the sake of clarity, μ_{XY} is called a *joint* measure. Denote by μ_X and μ_Y the projections of μ_{XY} on (X,β) and $(Y,\mathcal{F})$, $\mu_X\otimes\mu_Y$ the product measure on $(X\times Y,\ \beta\times\mathcal{F})$. The (average) mutual information of μ_{XY} is defined to be

$$I(\mu_{XY}) = \sup_{N;C_1,\ldots,C_N} \sum_{n=1}^{N} \mu_{XY}(C_n)\ \log \frac{\mu_{XY}(C_n)}{\mu_X\otimes\mu_Y(C_n)}, \tag{1}$$

where the supremum is over all $N \geq 1$ and all measurable partitions $C_1,\ldots,C_N$ of $(X\times Y,\ \beta\times\mathcal{F})$. It follows immediately that $I(\mu_{XY}) = \infty$ when it is false that μ_{XY} is absolutely continuous with respect to $\mu_X\otimes\mu_Y$ ($\mu_{XY} \ll \mu_X\otimes\mu_Y$). When $\mu_{XY} \ll \mu_X\otimes\mu_Y$, then [9]

$$I(\mu_{XY}) = \int_{X\times Y} \left[\log \frac{d\mu_{XY}}{d\mu_X\otimes\mu_Y}(x,y)\right] d\mu_{XY}(x,y) \tag{2}$$

Channel information capacity is defined as the supremum of $I(\mu_{XY})$ over all μ_{XY} in a suitable set. In the framework of most communication channels, and to be used here, the channel model is defined as follows. A measure μ_{XN} on $(X\times Y,\ \beta\times\mathcal{F})$ describes the statistical relationship between the message X and the channel noise N; usually, as we shall assume, $\mu_{XN} = \mu_X\otimes\mu_N$. The channel output Y is described by a measure $\mu_Y = \mu_X\otimes\mu_N \circ g^{-1}$, where g is $(X\times Y,\ \beta\times\mathcal{F})/(Y,\mathcal{F})$-

measurable. The joint measure μ_{XY} is then $\mu_X \otimes \mu_N \circ f^{-1}$, where $f(x,y) = (x, g(x,y))$. The most typical situation in engineering applications is for $g(x,n) = A(x)+n$, where A is an $(X,\mathcal{F})/(Y,\beta)$-measurable coding function. In general, the capacity is then defined as $\sup_Q I(\mu_{XY})$, where Q is a set of constraints on all admissible pairs (μ_X, A) of message measures μ_X and coding functions A. However, if A is 1:1 and bimeasurable, then no information is lost due to A. That is, let $\mu_{SY} = [\mu_X \circ A^{-1} \otimes \mu_N] \circ h^{-1}$, where $h(x,y) = (x, x+y)$. If A is 1:1 and $\overline{(X,\beta)}^{\mu_X} / \overline{(Y,\mathcal{F})}^{\mu_S}$ bimeasurable, then (1) shows that $I(\mu_{SY}) = I(\mu_{XY})$. If X and Y are Polish (complete, separable, metrizable), then by Kuratowski's Borel mapping theorem [10] any 1:1 Borel-measurable map $A: X \to Y$ is Borel-bimeasurable.

We shall assume here that $X = Y$, $\beta = \mathcal{F}$, $A = I$ (identity), so $g(x,y) = x + y$. The extension to the more general case can be obtained by either restricting attention to coding functions A which are 1:1 and bimeasurable, or else by computing the information lost due to a coding function which does not have these properties.

Mathematical Structure

The following assumptions will be made henceforth. E is a locally convex Hausdorff linear topological vector space over the real numbers, with topological dual E'. It will also be assumed that E is quasi-complete: every closed and bounded subset is complete. E must then be sequentially complete. $\sigma(E')$ will denote the cylindrical σ-field, generated by the elements of E', $\overline{\sigma(E')}^{\mu}$ the completion under the measure μ. For x in E and y in E', the value of y at the point x will be denoted by $\langle y,x \rangle$.

The noise measure μ_N will be defined on (E, σ(E')). μ_N will be assumed to be Gaussian and zero-mean: $\mu_N \circ \ell^{-1}$ is a zero-mean Gaussian distribution on $\mathbb{R}$ for each ℓ in E'. μ_N will be assumed to have a covariance operator R_N: E' → E. R_N is linear, self-adjoint and nonnegative: $\langle x, R_N y\rangle = \langle y, R_N x\rangle$ and $\langle x, R_N x\rangle \geq 0$ for all x,y in E'. μ_N has characteristic function given by $\hat{\mu}_N(x) = \int_E e^{i\langle x,y\rangle} d\mu_N(y) = e^{-\frac{1}{2}\langle x, R_N x\rangle}$, and $\langle x, R_N y\rangle = \int_E \langle x,u\rangle\langle y,u\rangle d\mu_N(u)$.

Under these assumptions, it is known [11] that there exists a unique Hilbert space H_N contained in E, such that the natural (canonical) injection j_N: H_N → E is continuous, $R_N = j_N j_N^*$, and H_N is the closure of range(R) under the inner product $\langle Ru, Rv\rangle_N = \langle u, Rv\rangle$. Here, H_N' is always identified with H_N. H_N is termed the reproducing kernel Hilbert space (RKHS) of R_N (or μ_N); it is actually the RKHS for the covariance function R_0: E'×E' → $\mathbb{R}$, $R_0(u,v) = \langle u, Rv\rangle$. It will be further assumed that H_N is separable; instances where this assumption is not necessary will be noted. If μ_N is Radon, then H_N is necessarily separable [12].

The message measure μ_X is a probability on (E, σ(E')). The constraints to be imposed will ensure that μ_X has a covariance operator R_X: E' → E; it can be assumed (WLOG) that μ_X has zero mean. As in the previous section, the measure of interest is μ_{XY}, defined by $\mu_{XY} = \mu_X \otimes \mu_N \circ f^{-1}$, where f(x,y) = (x, x+y).

A basic result in the Shannon theory is that if the supports of μ_N and μ_X are restricted to be of finite dimension and the covariance of μ_X is fixed, then $I(\mu_{XY})$ is maximized when μ_X is Gaussian. From this one obtains the result that the channel capacity problem can be solved by assuming μ_X to be Gaussian (see [2, Lemma 6]). This assumption will be made henceforth.

The *observation* measure $\mu_Y = \mu_X \otimes \mu_N \circ g^{-1}$, where $g(x,y) = x + y$, is thus Gaussian, with covariance operator $R_Y: E' \to E$, $R_Y = R_X + R_N$. Of course, R_Y has a RKHS H_Y contained in E and $R_Y = j_Y j_Y^*$, where $j_Y: H_Y \to E$ is the natural injection and is continuous.

The joint Gaussian measure μ_{XY} has a joint covariance operator $\mathcal{R}_{XY}: E' \times E' \to E \times E$ [13], [7]. This operator and its properties are characterized by the following result. It does not require that H_N be separable. Moreover, the result holds for any joint Gaussian measure on $(E \times E, \sigma(E') \times \sigma(E'))$ having a covariance operator $\mathcal{R}_{XY}: E' \times E' \to E \times E$.

<u>Lemma 1</u> [13], [7]:

(1) $\mathcal{R}_{XY} = \mathcal{J}(\mathcal{I}+\mathcal{V})\mathcal{J}^*$, where $\mathcal{J}: H_X \times H_Y \to E \times E$ is the natural injection, $\mathcal{I}$ is the identity in $E \times E$, and $\mathcal{V}$ is a self-adjoint bounded linear operator in $H_X \times H_Y$ with $\|\mathcal{V}\| \leq 1$.

(2) $\mathcal{V}(x,y) = (V_{XY}y, V_{XY}^*x)$, where $V_{XY}: H_Y \to H_X$ is a bounded linear operator with $\|V_{XY}\| \leq 1$. The operator V_{XY} is uniquely defined by $\int_E \langle u,x\rangle\langle v,y\rangle d\mu_{XY}(x,y) = \langle u, j_X V_{XY} j_Y^* v\rangle$ for all u,v in E'.

(3) $I(\mu_{XY}) < \infty$ if and only if V_{XY} is Hilbert-Schmidt with $\|V_{XY}\| < 1$.

(4) When V_{XY} is Hilbert-Schmidt with $\|V_{XY}\| < 1$, then $I(\mu_{XY}) = -\frac{1}{2}\Sigma_n \log(1-\gamma_n)$ where (γ_n) are the eigenvalues of $V_{XY}V_{XY}^*$.

Lemma 1 is fundamental to the solution of the channel capacity problem. It enables one to calculate the mutual information, yielding the following result.

Lemma 2 [2], [7]: Suppose that μ_X is Gaussian. Then:

(1) $I(\mu_{XY}) < \infty$ if and only if $\bar{\mu}_X[\text{range}(j_N)] = 1$, where $\bar{\mu}_X$ is the extension of μ_X to $\overline{\sigma(E')}^{\mu_X}$;

(2) $I(\mu_{XY}) < \infty$ if and only if $R_X = j_N T j_N^*$, where $T: H_N \to H_N$ is trace-class. When this is satisfied, $I(\mu_{XY}) = \frac{1}{2} \Sigma_n \log (1+\tau_n)$, where (τ_n) are the eigenvalues of T.

If the RKHS H_N is not separable, then part (1) of Lemma 2 holds with the condition $\bar{\mu}_X[\text{range}(j_N)] = 1$ replaced by $\mu_X^*[\text{range}(j_N)] = 1$, where μ_X^* is the outer measure obtained from μ_X [7]. The following result is then useful.

Lemma 3 [7]: Suppose that E is a locally convex l.t.v.s., μ a probability measure on $(E,\sigma(E'))$. Suppose that B is a separable or reflexive Banach space and that $j: B \to E$ is a continuous linear injection. Then, the following are equivalent:

(1) $\mu^*[j(B)] = 1$;

(2) $\mu = \nu \circ j^{-1}$, where ν is a unique probability measure on $(B,\sigma(B'))$.

If (1) or (2) holds, then μ is Gaussian if and only if ν is Gaussian. If B is both separable and reflexive, then $j[B] \in \overline{\sigma(E')}^{\mu}$ so that (1) is equivalent to $\bar{\mu}[j(B)] = 1$.

In the mismatched channel to be considered subsequently, the constraints are given in terms of the norm for another Hilbert subspace of E. The following result is then useful. It does not require that H_N be separable.

<u>Proposition 1:</u> Suppose that H_W is a Hilbert subspace of E. Let $j_W: H_W \to E$ be the natural injection map, and suppose that $\mu_X^*[\text{range}(j_W)] = 1$. Then $I(\mu_{XY}) < \infty$ if and only if H_W is a vector subspace of H_N. If $H_W \subset H_N$, then j_W is continuous, the natural injection $J: H_W \to H_N$ is continuous, and H_W is the RKHS for the covariance operator $j_W j_W^*$; if H_N is separable, then H_W is also separable.

<u>Proof</u>: Suppose that $H_W \subset H_N$. Since H_W is a Hilbert space contained in the RKHS H_N, H_W is also a RKHS of functions on E' and $\|Jx\|_N^2 \leq k\|x\|_W^2$ for all x in H_W, some $k < \infty$ [12], so that the natural injection $J: H_W \to H_N$ is continuous. Since $j_W = j_N J$, j_W must also be continuous, so that $j_W j_W^*$ is a covariance operator mapping E' → E. By definition, H_W is the (unique) RKHS for $j_W j_W^*$. To see that H_W is separable (assuming that H_N is separable), one notes that the linear map $L: H_N \to H_W$, $L j_N^* u = j_W^* u$, is continuous and has dense range in H_W, so that L^* has only {0} in its null space. Thus, if $\{u_n, n\geq 1\}$ is such that $\{j_N^* u_n, n\geq 1\}$ is dense in H_N, then $\{L j_N^* u_n, n\geq 1\}$ must be dense in H_W. $I(\mu_{XY}) < \infty$, by Lemma 2, since $\bar{\mu}_X[\text{range}(j_N)] = 1$.

If H_W is not contained in H_N, then there exists z in range(j_W), $z \notin \text{range}(j_N)$. The Gaussian measure μ_X with covariance $z \otimes z$ has $\mu_X^*[\text{range}(j_N)] = 0$; by Lemma 2, $I(\mu_{XY}) = \infty$. □

<u>Constraints</u>

The constraints that will be used to define the admissible set Q of message measures μ_X are the following:

(A-1) $\mu_X^*[\text{range}(j_W)] = 1$.

(A-2) $\int_{H_W} \|x\|_W^2 d\upsilon_X(x) \le P,$

where H_W is a Hilbert space contained in H_N, with norm $\|\cdot\|_W$, $j_W\colon H_W \to E$ is the natural injection, and υ_X is the Borel measure on H_W satisfying $\mu_X = \upsilon_X \circ j_W^{-1}$.

Since we wish to have the constraint (A-2) apply a.e. $d\mu_X$, it is first necessary to require (A-1). The existence of the measure υ_X such that $\mu_X = \upsilon_X \circ j_W^{-1}$ follows from Lemma 3; H_W is separable, from Proposition 1. Also by Proposition 1, the capacity will be infinite if H_W is not a vector subspace of H_N.

The constraint (A-2) is motivated by the typical application when E is $L_2[0,T]$. In the case of formal white noise, the constraint is usually $E \int_0^T X_t^2(\omega)dt \le P$. This can be viewed as a constraint on $E\|X\|_W^2$, where W is the RKHS of the identity operator; this is the covariance of formal white noise. When white noise is viewed as the formal derivative of the Wiener process, then the "integrated" channel is analyzed [5]. In that case, the transmitted signal X is defined by $X_t = \int_0^t u(s)ds$, u in $L_2[0,T]$, and the constraint is typically $E\|U\|_{L_2}^2 \le PT$. $\|\dot{x}\|_{L_2}^2$ is the norm of x in the RKHS of Wiener measure. Finally, one may note that in his treatment of stationary power-and-frequency-limited Gaussian channels when the noise has integrable spectral density [4], Gallager first assumes a constraint on the message of the form $E\|X\|_{L_2[T]}^2 \le PT$. However, the transmitted signal is obtained by passing the message through a linear filter whose transfer function G satisfies $\int_{-\infty}^{\infty} \frac{|G(\lambda)|^2}{\phi_N(\lambda)} d\lambda < \infty$, where ϕ_N is the noise spectral density. Such a transmitted signal satisfies both (A-1) and a constraint of the type A-2, with (assuming that $|G|^2/\phi_N$

is bounded) an upper bound of $E_X\|X\|^2_{N,T} \le \frac{PT}{2\pi} \sup_\lambda \frac{|G(\lambda)|^2}{\phi_N(\lambda)} d\lambda$ for any $T > 0$, where now X refers to the filtered message and $\|\cdot\|_{N,T}$ is the RKHS of the noise covariance for the interval [0,T]. Of course, the constraint (A-2) is not placed explicitly on the transmitted signal in [4]; instead, it appears in the solution for the capacity. Gallager's analysis is for the water-filling model treated by Fano [6]. Fano's treatment does not yield finite capacity, precisely because the constraints (A-1) and (A-2) are not imposed.

In addition to its use in previous more specialized analyses, the use of a Hilbert space norm is plausible from two other considerations. First, as can be seen from Lemma 2, the capacity will be infinite unless the constraint used implies $E_X\|X\|^2_N \le P'$ for some $P' < \infty$. Proposition 1 shows that H_W must be a RKHS of functions on E' if the capacity is to be finite. Second, a RKHS norm actually places a dual constraint on the signal; this corresponds to limitations on the amount and frequency distribution of the signal energy in typical applications.

The capacity subject to the constraints (A-1) and (A-2) will be denoted by $\mathcal{C}_W(P)$. If $H_W = H_N$ (consisting of the same elements and the identical inner product), then the capacity will be denoted by $\mathcal{C}_N(P)$ and the channel is said to be *matched* (to the constraint). If $H_W \neq H_N$ as Hilbert spaces, then the channel is said to be *mismatched*. It will be seen that in the matched case, the results can be directly related to results obtained by Shannon [1] and Gallager [4] for the white noise case and by Kadota, Zakai, and Ziv [5] when the noise is the Wiener process (without a dimensionality constraint). Thus, these results extend the aforementioned results to an arbitrary Gaussian noise, rather than for formal white noise or for the Wiener process.

In the case of the mismatched channel, a completely new set of results is obtained. These results differ from those of the matched channel not only in the value of the capacity, but also by the properties of the solution. These differences will be discussed after the main results are presented.

Information Capacity of the Matched Channel

The solution for $\mathscr{C}_N(P)$ is given by the following theorem.

Theorem 1 [2], [7]:

(a) Suppose that H_N is of dimension $\geq M$, and that (in addition to (A-1) and (A-2)), μ_X is required to satisfy dim $[\text{supp}(\mu_X)] \leq M$. Then $\mathscr{C}_N(P) = (M/2) \log (1 + P/M)$. The supremum is attained, and only attained, when μ_X is Gaussian with zero mean and covariance operator $R_X = j_N T j_N^*$, where $T: H_N \to H_N$ is any self-adjoint linear operator with M-dimensional range space and with a single non-zero eigenvalue of value P/M.

(b) Suppose that H_N is infinite-dimensional. Subject only to the constraints (A-1) and (A-2), $\mathscr{C}_N(P) = P/2$. The capacity cannot be attained.

Shannon's original work [1] considered capacity for the white Gaussian channel with noise of spectral density $N_0/2$ and the signal S constrained in time to T seconds, constrained in bandwidth to W hertz, and constrained in average power by $E\int_0^T S_t^2 dt \leq PT$. His result, one of the best-known results of early information theory, was that the capacity is WT log $(1 + P/(WN_0))$. Since the formal RKHS norm for such noise is $\|x\|_N^2 = \int_0^T x_t^2 dt/(N_0/2)$, Theorem 1(a) gives Shannon's result by setting the dimensionality $M = 2WT$. Gallager [4] has obtained this result for the white noise channel by considering the signal as a point in a space of 2WT dimensions, and applying the same energy constraint as that used by Shannon.

Part (b) of Theorem 1 is also a generalization of known results for the white noise case, again generalizing results of Shannon and Gallager. For white noise of spectral density $N_0/2$ and signal S such that $E\int_0^T S_t^2 dt \leq PT$, Shannon showed that the capacity without a band-width constraint is at least P/N_0. Gallager proved that the capacity is P/N_0 if there is no dimensionality constraint on the signal.

The result of Theorem 1(b) has also been obtained by Kadota, Zakai, and Ziv [5] for the case where the channel noise is the Wiener process (the "integrated white noise" channel).

Theorem 1 thus extends some of the well-known results of information theory, previously obtained only for the formal white noise channel (or, for part (b), the Wiener process channel) to a general Gaussian channel without feedback. The results of Theorem 1 hold for any Gaussian noise measure whose covariance operator maps into E and which has a separable RKHS. In particular, they hold whenever the space E is a separable Banach space.

Capacity of the Mismatched Gaussian Channel

The capacity problem for the case where $H_W \neq H_N$ (as Hilbert spaces) provides a degree of flexibility which is lacking in the matched channel. By Lemma 2, necessary and sufficient conditions for finite capacity are that $\bar{\mu}_X[\text{range}(j_N)] = 1$ and $E_X\|X\|_N^2 \leq P'$ for some

$P' < \infty$. However, one may wish to use more selectivity in the choice of constraint. As noted above, the RKHS norm $\|x\|_W^2$ can be viewed as a constraint on both the amount and the frequency distribution of the signal energy.

In the matched case, the constraint $E_X\|X\|_N^2 \leq P$ constrains the frequency content of the energy in a manner determined solely by the channel noise. It is more desirable to constrain the frequency distribution in a manner which not only satisfies constraints imposed by the RKHS of the channel noise, but also satisfies additional constraints. Such additional constraints are needed in order to analyze channels with partially unknown noise, including jamming channels.

In this section, the results corresponding to Theorem 1 will be given, now assuming a mismatched channel. For more details, reference is made to [3], where the corresponding results are obtained for the case where E is a separable Hilbert space. Those results can be extended to E a separable complete metric space by using the Banach-Mazur theorem and Kuratowski's Borel mapping theorem. However, in applications one may deal with a linear space that is not metrizable, or not complete, or not separable. The extension of those results to the present framework can be carried out by using Lemmas 1-3 and Proposition 1 to adapt the proofs given in [3].

The constraints (A-1) and (A-2) involve the unique Gaussian Borel measure ν_X on H_W satisfying $\mu_X = \nu_X \circ j_W^{-1}$. Of course, ν_X has a covariance operator $R_\nu: H_W \to H_W$, and $R_X = j_W R_\nu j_W^*$. At the same time, $R_X = j_N T j_N^*$ where T is the covariance operator (in H_N) of the Gaussian measure ν_T satisfying $\nu_T \circ j_N^{-1} = \mu_X$. $T = \Sigma\, \tau_n u_n \otimes u_n$ for a CONS $\{u_n,\ n \geq 1\}$

in H_N, and $I(\mu_{XY}) = \frac{1}{2} \Sigma_n \log(1+\tau_n)$. Moreover, $\nu_T = \nu_X \circ J^{-1}$, $J: H_W \to H_N$ the imbedding map, and so $R_X = j_N J R_\nu J^* j_N^*$.

From this it follows that one can assume that $\overline{\text{range}(J)}^{H_N} = H_N$ (otherwise, restrict attention to $\overline{\text{range}(J)}^{H_N}$). One then has $R_\nu = J^{-1}TJ^{-1*}$, and the constraint (A-2) becomes $\Sigma_n \tau_n \|J^{-1}u_n\|_W^2 \leq P$. Define

$$I_N + S = (J^{-1})^* J^{-1} = (JJ^*)^{-1} \tag{3}$$

where I_N is the identify in H_N. The operator $S: H_N \to H_N$ is densely defined. The limit points of the spectrum of S consist of all eigenvalues of infinite multiplicity, all limit points of distinct eigenvalues, and all points of the continuous spectrum [13]. Let θ be the smallest limit point of the spectrum of S, and let $\{\lambda_n, n \geq 1\}$ be the set of all eigenvalues of S which are strictly less than θ, with corresponding o.n. eigenvectors $\{e_n, n \geq 1\}$. Of course, $\{\lambda_n, n \geq 1\}$ can be empty, finite, or countably infinite.

The capacity problem now becomes that of determining

$$\mathscr{C}_W(P) = \sup \left(\tfrac{1}{2}\right) \sum_n \log\left(1 + X_n^2[1+\gamma_n]^{-1}\right) \tag{4}$$

subject to the constraint

$$\sum_n X_n^2 \leq P, \tag{5}$$

where

$$X_n^2 = \tau_n(1 + \gamma_n), \tag{6}$$

$$1 + \gamma_n = \|J^{-1}u_n\|_W^2 , \tag{7}$$

$$R_X = \sum \tau_n j_N (u_n \otimes u_n) j_N^* , \tag{8}$$

where (τ_n) is any nonnegative summable sequence of real numbers, $\{u_n, n \geq 1\}$ any CONS in H_N which belongs to range(J).

The results for the mismatched channel which correspond to part (1) of Theorem 1 can now be stated.

Theorem 2 [3]: Suppose that H_N has dimension $M < \infty$. The capacity is then

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

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 $\Sigma_1^K \beta_i + P \geq K\beta_K$. The capacity is attained by a Gaussian μ_X with covariance operator (8), where $\tau_n = \left[\Sigma_1^K \beta_i + P - K\beta_n\right](1+\beta_n)^{-1}K^{-1}$ for $n \leq K$, $\tau_n = 0$ for $n > K$, and $\{u_n, n \geq 1\}$ are o.n. eigenvectors of S corresponding to the eigenvalues (β_n). No other Gaussian μ_X can attain capacity. The same result is obtained if H_N has dimension $L < \infty$ and μ_X is constrained to have support of dimension $M < L$.

The above result assumes H_N to be finite-dimensional. The following theorem extends to the case where H_N is infinite-dimensional, but the support of μ_X is restricted to be of finite dimension.

Theorem 3 [3]: Suppose that H_2 is infinite-dimensional, and that support (μ_X) is restricted to have dimension $\leq M < \infty$.

(1) Suppose that $\theta < \infty$.

(a) If $\{\lambda_n, n \geq 1\}$ is empty, then $\mathscr{C}_W(P) = (M/2) \log [1 + PM^{-1}(1+\theta)^{-1}]$. Capacity can be attained if and only if S has θ as an eigenvalue of multiplicity $\geq M$. In this case $\mathscr{C}_W(P)$ is attained only by a Gaussian μ_X with covariance (8), where $\tau_i = PM^{-1}(1+\theta)^{-1}$ for $i \leq M$ with $\{u_1, \dots, u_M\}$ any o.n. set in the null space of $S - \theta I$.

(b) If $K\lambda_K \le \Sigma_1^K\lambda_i + P < K\lambda_{K+1}$ for some $K \le M$, then the capacity is as in Theorem 2, with $\beta_i = \lambda_i$, $i=1,\ldots,K$, and can be similarly attained.

(c) Let $K = \min(L,M)$, where $L \ge 1$ is the number of eigenvalues (λ_n) of S whose value is strictly less than θ, and suppose that $P + \Sigma_1^K\lambda_i \ge K\theta$. The capacity is then

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

The capacity can be attained if and only if θ is an eigenvalue of S with multiplicity $\ge M-K$. The capacity is then achieved only by a Gaussian μ_X with covariance (8), where $\tau_n = (\Sigma_1^K\lambda_i + P - M\lambda_n + (M-K)\theta)(1+\lambda_n)^{-1}M^{-1}$ for $n \le K$, with $Su_n = \lambda_n u_n$ and $\{u_1,\ldots,u_K\}$ an o.n. set; and with $u_n = v_n$ and $\tau_n = (P + \Sigma_1^K\lambda_i - K\theta)M^{-1}(1+\theta)^{-1}$ for $K+1 \le n \le M$, where $Sv_n = \theta v_n$ and $v_{K+1},\ldots,v_M$ is an o.n. set.

(2) If $\theta = \infty$, then $\mathscr{C}_W(P)$ has the value given in part 1(b), and can be similarly attained.

Theorems 2 and 3 together are parallel to part 1 of Theorem 1. The solution for the mismatched channel is seen to be considerably more complex than that for the matched channel. The final generalization is to permit μ_X to have infinite-dimensional support. The solution will again be much more complex than the solution to the corresponding problem for the matched channel (part 2 of Theorem 2).

Theorem 4 [3]:

(1) Suppose that $\theta < \infty$, H_2 is infinite-dimensional, and $\dim[\mathrm{supp}(\mu_X)]$ is not constrained.

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

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

(b) If $\{\lambda_n, n \geq 1\}$ is not empty, (λ_n) is an infinite sequence, and $P < \Sigma_n(\theta - \lambda_n)$, then there exists a largest integer K such that $\Sigma_1^K \lambda_i + P \geq K\lambda_K$, and

$$\mathscr{C}_W(P) = \tfrac{1}{2} \sum_{n=1}^{K} \log \left[\frac{\Sigma_1^K \lambda_i + P + K}{K(1+\lambda_n)}\right].$$

(c) If $\{\lambda_n, n \geq 1\}$ is empty, then $\mathscr{C}_W(P) = \dfrac{P}{2(1+\theta)}$.

(d) In (a), the capacity can be attained if and only if $\Sigma_n(\theta - \lambda_n) = P$. It is then attained, and only attained, by a Gaussian μ_X with covariance operator as in (8), where $u_n = e_n$ 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_X with covariance operator (8), where $u_n = e_n$ and $\tau_n = \dfrac{\Sigma_1^K \lambda_i + P + K}{K(1+\lambda_n)} - 1$ for $n \leq K$; $\tau_n = 0$ for $n > K$. In (c), the capacity cannot be attained.

(2) If $\theta = \infty$, then $\mathscr{C}_W(P)$ has the value given in part 1(b), and can be similarly attained.

Discussion

The results summarized in Theorems 1-4 provide a general solution to the capacity problem for the Gaussian channel without feedback, requiring a minimal set of assumptions. The solution for the mismatched channel is markedly different from that for the matched channel. The value of the capacity can be very different, as already seen; it can be less or more than the capacity of the matched channel, depending on θ and $\{\lambda_n, n\geq 1\}$. The expression for the capacity varies as a function of P, H_N and H_W. Moreover, 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_N is infinite-dimensional with $\dim[\mathrm{supp}(\mu_X)] \leq M$, the situation is even worse in (a) and (c) of Theorem 3. That is, capacity can then be attained only if S has θ as an eigenvalue of multiplicity $\geq M$ when $S \geq \theta I$, or of multiplicity $\geq M-K$ when S has $K < M$ eigenvalues $\lambda_1 \leq \ldots \leq \lambda_K < \theta$ and $P + \Sigma_1^K \lambda_i \geq K\theta$.

For the infinite-dimensional channel without a constraint on $\dim[\mathrm{supp}(\mu_X)]$, there can again be significant differences between $\mathscr{C}_W(P)$ and $\mathscr{C}_N(P)$, depending on $\{\theta; \lambda_n, n\geq 1\}$. Moreover, there is again a rather different situation in the problem of attaining capacity. $\mathscr{C}_N(P)$ can never be attained; $\mathscr{C}_W(P)$ can be attained if and only if $\{\lambda_n, n\geq 1\}$ is not empty and $P \leq \Sigma_n(\theta-\lambda_n)$.

A comparison of the value of the capacity $\mathscr{C}_W(P)$ for the mismatched channel with the capacity $\mathscr{C}_N(P)$ of the matched channel can be made from the preceding results. For the finite-dimensional channel, $\mathscr{C}_W(P)$ is strictly greater than $\mathscr{C}_N(P)$ if $\Sigma_1^M \beta_i \leq 0$, or if $P + \Sigma_1^K \beta_i \leq 0$. $\mathscr{C}_W(P) \leq \mathscr{C}_N(P)$ if $0 \leq \beta_1 < \beta_M$. For the infinite-dimensional channel, suppose that $\{\lambda_n, n\geq 1\}$ is empty. Then, $\mathscr{C}_W(P) > \mathscr{C}_N(P)$ if $\theta < 0$, $\mathscr{C}_W(P) < \mathscr{C}_N(P)$ if $\theta > 0$, $\mathscr{C}_W(P) = \mathscr{C}_N(P)$ if $\theta = 0$. If $\{\lambda_n, n\geq 1\}$ is not empty, then for the unconstrained channel $\mathscr{C}_W(P)$ is greater than

$P/[2(1+\theta)]$. Thus, $\mathcal{C}_W(P) > \mathcal{C}_N(P)$ if $\theta \leq 0$ and $\{\lambda_n, n \geq 1\}$ is not empty. A similar result can be obtained for the constrained channel.

Applications

The results on the mismatched channel can be used to analyze channels with partially unknown noise, including jamming channels. The results can also be applied to compare the capacity of channels with and without feedback. It has been possible to show that the capacity of a large class of mismatched Gaussian channels is increased by adding linear feedback [16]; for example, this class includes the time-discrete correlated noise channel with a pure power constraint. This illustrates another difference between matched and mismatched channels; it is not possible to increase capacity of matched Gaussian channels by adding feedback.

References

1. C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* 27 (1948), 379-423, 623-656.

2. C.R. Baker, Capacity of the Gaussian channel without feedback, *Inform. Contr.* 37 (1978), 70-89.

3. C.R. Baker, Information capacity of the mismatched Gaussian channel, to appear, *IEEE Trans. on Inform. Theory*.

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

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

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

7. I.W. McKeague, On the capacity of channels with Gaussian or non-Gaussian noise, *Inform. Contr.* 51 (1981), 153-173.

8. I.W. McKeague and C.R. Baker, The coding capacity of mismatched Gaussian channels, *IEEE Trans. on Inform. Theory*, IT-32 (1986), 431-436.

9. R.L. Dobrushin, General formulation of Shannon's basic theorems of the theory of information, *Uspehi Mat. Nauk* 14 (1959), 3-104; translated in *Amer. Math. Soc. Transl.* (2) 33 (1963), 323-438.

10. K.R. Parthasarathy, "Probability Measures on Metric Spaces," Academic Press, New York (1967).

11. L. Schwartz, Sous-espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associes, *J. Analyse Math.* 13 (1964), 115-256.

12. C. Borell, Gaussian Radon measures on locally convex spaces, *Math. Scand.* 38 (1976), 265-284.

13. C.R. Baker, Joint measures and cross-covariance operators, *Trans. Amer. Math. Soc.* 186 (1973), 272-289.

14. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* 68 (1950), 337-404.

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

16. C.R. Baker, Information capacity of the Gaussian channel with feedback, *Abstracts, IEEE International Symp. on Inform. Theory* (1986).