INFORMATION AND CONTROL 51, 153–173 (1981)

On the Capacity of Channels with Gaussian and Non-Gaussian Noise*

IAN W. MCKEAGUE

Department of Statistics, Florida State University, Tallahassee, Florida 32306

We evaluate the information capacity of channels for which the noise process is a Gaussian measure on a quasi-complete locally convex space. The coding capacity is calculated in this setting and for time-continuous Gaussian channels using the information capacity result. The coding capacity of channels with non-Gaussian noise having finite entropy with respect to Gaussian noise of the same covariance is shown not to exceed the coding capacity of the Gaussian channel. The sensitivity of the information capacity to deviations from normality in the noise process is also investigated.

1. INTRODUCTION

In this article we extend some of the classical results of information theory to the setting of topological spaces. The capacity of a channel is currently defined in two closely connected ways. In the information capacity sense it is the supremum of the mutual information between an allowable input and the output of the channel. In the coding capacity sense it is the highest rate at which coded messages can be sent with arbitrarily small probability of error. We shall evaluate the information capacity for the additive Gaussian channel, where the Gaussian noise is defined on any quasi-complete locally convex space. The coding capacity is calculated in this setting and also for time-continuous channels. For the case of non-Gaussian noise, an upper bound on the coding capacity is given. The question of robustness of the information capacity is also investigated.

Shannon (1948) determined the information capacity of the white noise Gaussian channel with bandlimited input signals. Kadota *et al.* (1971) rigorously treated the case with causal feedback and the Wiener process as noise. Hitsuda and Ihara (1975) extended these results to a large class of Gaussian channels with causal feedback by using the Cramér–Hida representation of the noise. Baker (1978a) obtained the information capacity of the general Gaussian channel without feedback, assuming that signal and noise

* This research was supported by the Office of Naval Research under Grant N00014–74–C–0491.

0019-9958/81/110153–21$02.00/0

belong to separable Banach spaces. Baker proved his results for separable Hilbert spaces and extended them to separable Banach spaces using the Banach–Mazur Theorem. Here we consider the problem of extending these results to some class of topological spaces beyond Banach spaces and to a case that is important for applications, the product space $\mathfrak{R}^T$, where T is an arbitrary index set.

The evaluation of the coding capacity of additive Gaussian channels has been carried out by Shannon (1948) and Wyner (1966) for bandlimited channels with stationary Gaussian noise having a flat spectral density, by Shannon (1959) for the time-discrete Gaussian channel, and by Fortet (1961), Bethoux (1962), and Ash (1965) for a channel with stationary Gaussian noise and signals satisfying an average energy constraint. We shall evaluate the coding capacity for a generalization of the time-discrete Gaussian channel and use this to find the capacity of the time-continuous channel with arbitrary Gaussian noise. The coding capacity of channels with non-Gaussian noise satisfying an entropy condition is shown not to exceed the coding capacity of the Gaussian channel.

2. Preliminaries

Throughout this work E will denote a locally convex, Hausdorff topological vector space (we contract this to locally convex space) with topological dual E'. The Borel σ-algebra on E will be denoted $\mathscr{B}(E)$. A probability measure on $\mathscr{B}(E)$ is called a Borel measure.

Let F be a finite dimensional subspace of E'. A subset C of E is a cylinder set based at F if it is of the form

$$C = \{x \in E: (\langle f_1, x\rangle, \ldots, \langle f_n, x\rangle) \in D\},$$

where $n \geqslant 1, f_1, \ldots, f_n \in F$, and $D \in \mathscr{B}(\mathfrak{R}^n)$. Let $\mathscr{C}_F$ denote the σ-algebra of all cylinder sets based on F, and $\mathscr{C} = \bigcup \mathscr{C}_F$, where F runs over all finite dimensional subspaces of E'. $\mathscr{C}$ is an algebra. The σ-algebra generated by $\mathscr{C}$ is called the cylindrical σ-algebra, denoted $\sigma(E')$, and it is the smallest σ-algebra of subsets of E such that each linear functional $f \in E'$ is measurable. Clearly $\sigma(E') \subseteq \mathscr{B}(E)$. If E is separable and metrizable then $\sigma(E')$ and $\mathscr{B}(E)$ coincide, but not in general (see Vakhania and Tarieladze, 1978).

A set function $\mu: \mathscr{C} \to [0, 1]$ is called a cylinder set measure (csm) or a cylindrical measure if, for each finite dimensional subspace F of E', $\mu|\mathscr{C}_F$ is a probability measure. By Kolmogorov's Theorem, a csm may be extended to be countably additive on E'^a (the algebraic dual of E'). Let μ be either a csm on E or a probability measure on $\sigma(E')$. We note the following definitions. The characteristic functional $\hat{\mu}$ is given by $\hat{\mu}(f) =$

$\int_E \exp\{i\langle f, x\rangle\}\, d\mu(x)$ for $f \in E'$. μ is Gaussian if each $f \in E'$ is a Gaussian random variable under μ. If H is a Hilbert space then the csm with characteristic functional $\exp(-\|\cdot\|^2/2)$ is called the canonical Gaussian csm and is denoted γ_H. μ is of weak second order if, for each $f \in E'$, $\int_E \langle f, x\rangle^2\, d\mu(x) < \infty$. The mean $m \in E'^a$ and covariance operator $R \in \mathscr{L}(E', E'^a)$ of a weak second-order csm μ are defined by

$$\langle m, f\rangle = \int_E \langle f, x\rangle\, d\mu(x),$$

$$\langle f, g\rangle = \int_E \langle f, x\rangle\langle g, x\rangle\, d\mu(x) - \langle m, f\rangle\langle m, g\rangle,$$

where f, $g \in E'$.

The covariance operator R is non-negative and symmetric: $\langle Rf, g\rangle = \langle f, Rg\rangle$, $\langle Rf, f\rangle \geqslant 0$, for all f, g in E'. If E is separable and quasi-complete then the covariance operator of a weak second-order probability measure on $(E, \sigma(E'))$ belongs to $\mathscr{L}(E', E)$, and conversely every symmetric non-negative linear operator $R: E' \to E$ is the covariance of some weak second-order probability measure on E. If E is a normed space then a Borel measure μ is said to be of strong second order if $\int_E \|x\|^2\, d\mu(x) < \infty$. All Gaussian Borel measures on a separable Banach space are of strong second order by Fernique (1970). The notion of a covariance operator was introduced for strong second-order measures on separable Hilbert spaces by Prohorov (1956) and for weak second-order measures on Banach spaces Vakhania (1968). We refer to Vakhania and Tarieladze (1978) for further material on covariance operators.

A Borel measure μ on E is said to be Radon if, for each $B \in \mathscr{B}(E)$, $\mu(B) = \sup\{\mu(K): K \subset B,\ K \text{ compact}\}$. If E is Polish (i.e., separable, complete, metrizable) then every Borel measure on E is Radon (see Schwartz (1973, p. 122)).

By Prohorov's Theorem, a cylinder set measure μ on E can be extended to a Radon measure on E if and only if for every $\varepsilon > 0$ there is a compact set $K \subset E$ such that $\mu^*(K) \geqslant 1 - \varepsilon$, where μ^* is the outer measure derived from μ (see Schwartz, 1973).

If μ is a Gaussian Radon measure on E, then (E, μ) can be considered as an abstract Wiener space; i.e., there exists a separable Hilbert space H and an injection $j: H \to E$ such that $\mu = j(\gamma_H)$, where γ_H is the canonical Gaussian csm on H (see Borell, 1976). H is called the reproducing kernel Hilbert space (RKHS) of μ. The mean of μ belongs to E and its covariance operator belongs to $\mathscr{L}(E', E)$.

Schwartz (1964) showed that if E is quasi-complete then each linear operator, $R: E' \to E$, which is symmetric and non-negative, has a unique

Hilbert space H, which is a vector subspace of E, such that the injection j of H into E is continuous and $R = jj^*$. The Hilbert space H is called the RKHS of R. The methods employed in this article depend on the existence of a RKHS for the covariance operators. For this reason, it is assumed throughout that the locally convex spaces E are quasi-complete and that the covariance operators map E' into E. If the RKHS of a covariance operator $R: E' \to E$ is separable with a CONS $\{e_n, n \geqslant 1\}$ then the covariance operator admits a series representation $R = \sum_n e_n \otimes e_n$, where $(e_n \otimes e_n)(f) = \langle f, e_n \rangle e_n$, for $f \in E'$, and the series converges to R in the strong operator topology: $\sum_1^N \langle f, e_n \rangle e_n \to Rf$ in E for all $f \in E'$.

Let μ_X be a zero-mean Gaussian measure on $\sigma(E')$ with covariance operator $R_X: E' \to E$ and RKHS H_X. The linear space of μ_X, denoted $\mathscr{L}_X$, is the closed subspace of $L^2(E, \mu_X)$ spanned by E'. The space $\mathscr{L}_X$ is a Hilbert space under the inner product induced from $L^2\,(E, \mu_X)$. It is well known that there exists a unitary operator $U_X: \mathscr{L}_X \to H_X$ such that $U_X f = j_X^* f$, for all $f \in E'$.

It is known that two Gaussian measures μ_X, μ_Y on $\sigma(E')$ are either mutually absolutely continuous ($\mu_X \sim \mu_Y$) or orthogonal ($\mu_X \perp \mu_Y$) in cases that E is a product space $\mathfrak{R}^T$ (see Feldman, 1958; Pan, 1966, for instance), a separable Hilbert space (Rao and Varadarajan, 1963) and a separable Banach space (Kuelbs, 1970; Baker, 1978b). The following theorem extends these results to quasi-complete locally convex spaces.

THEOREM 2.1. *Let μ_X, μ_Y be two Gaussian measures on $\sigma(E')$ with means $m_X = 0$ and $m_Y = m \in E$, and covariance operators $R_X: E' \to E$, $R_Y: E' \to E$, respectively. Then $\mu_X \sim \mu_Y$ if and only if the following conditions are satisfied.*

(i) *$H_X = H_Y$ as subsets of E;*

(ii) *$m \in H_X$;*

(iii) *$I_Y - JJ^*$ is Hilbert–Schmidt, where $J: H_X \to H_Y$ is the injection of H_X into H_Y and I_Y is the identity on H_Y.*

Moreover, if conditions (i), (ii) *and* (iii) *hold, and if $\{\lambda_n\}$ denotes the eigenvalues of JJ^* which are different from* 1, *while $\{v_n\}$ denotes the corresponding sequence of normalized eigenvectors, then*

$$\frac{d\mu_Y}{d\mu_X}(x) = \exp\left\{U_X^{-1}[(JJ^*)^{-1/2}\, m](x) - \frac{1}{2}\langle m, (JJ^*)^{-1}\, m\rangle_{H_X}\right. \\ \left. - \frac{1}{2}\sum_1^\infty \left[(U_X^{-1} v_n)^2\,(x)\left(\frac{1}{\lambda_n} - 1\right) + \log \lambda_n\right]\right\}.$$

Furthermore, if one or more of the conditions (i), (ii) *or* (iii) *does not hold, then $\mu_X \perp \mu_Y$.*

Proof. Let E be identified with its image under the injection of E into E'^a defined by $\rho(x)(f) = f(x)$, for $x \in E, f \in E'$. $\sigma(E')$ is identified with the σ-algebra on $\rho(E)$ induced by the cylindrical σ-algebra on E'^a. Under the $\sigma(E'^a, E')$ topology E'^a is isomorphic to a product space $\mathfrak{R}^A$, where A is an algebraic basis of E' (Robertson and Robertson (1964, p. 96)). In this way, the measures μ_X, μ_Y are identified with Gaussian measures on a product space $\mathfrak{R}^A$. The proof of the theorem now follows from the product space result (Pan, 1966). ∎

The next result was given by Baker (1973) for the case that B and E are separable Banach spaces with B reflexive.

THEOREM 2.2. *Let E be a locally convex space, μ a probability measure on $(E, \sigma(E'))$. Suppose B is a separable or reflexive Banach space and $j: B \to E$ is a continuous linear injection. Then the following are equivalent:*

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

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

Moreover, if (i) *or* (ii) *holds then μ is Gaussian if and only ν is Gaussian. If B is both separable and reflexive then $j(B) \in \overline{\sigma(E')^\mu}$ so that condition* (i) *may be written $\bar{\mu}(j(B)) = 1$.*

Proof. Write $M = j(B)$. Then it is easy to show that $M \cap \sigma(E') = j[\sigma(j^*(E'))]$, where $\sigma(j^*(E'))$ is the σ-algebra generated by $j^*(E')$. Also, $j^*(E')$ is total in B' from the assumptions that E is locally convex and j is injective. By Perlman (1972, Theorem 8) it follows that $\sigma(j^*(E')) = \sigma(B')$, since it is assumed that B is separable or reflexive. Thus $M \cap \sigma(E') = j[\sigma(B')]$. Suppose that (i) holds. By Theorem 1.1 of Doob (1937) we may define a probability measure $\tilde{\mu}$ on $M \cap \sigma(E')$ by $\tilde{\mu}(M \cap A) = \mu(A)$, for $A \in \sigma(E')$. Define $\nu(A) = \tilde{\mu}(j(A))$, for $A \in \sigma(B')$. Then ν is a probability measure on $\sigma(B')$ and $\nu \circ j^{-1}(A) = \tilde{\mu}(A \cap M) = \mu(A)$, for $A \in \sigma(E')$. Each $A \in \sigma(B')$ is of the form $j^{-1}(A)$ for some $A \in \sigma(E')$, so that ν is unique.

Conversely, suppose (ii) holds. Let $A \in \sigma(E')$ and $A \supset M$. Then $\mu(A) = \nu \circ j^{-1}(A) = \nu(B) = 1$. Thus $\mu^*(j(B)) = 1$ and (i) holds. For $f \in E'$, $\nu\{x: \langle j^*f, x\rangle \leqslant k\} = \nu\{x: \langle f, jx\rangle \leqslant k\} = \nu \circ j^{-1}\{y: \langle f, y\rangle \leqslant k\} = \mu\{y: \langle f, y\rangle \leqslant k\}$, and it follows by Perlman (1972) that μ is Gaussian if and only if ν is Gaussian.

Finally, suppose that B is both separable and reflexive. The separability implies that ν is a Radon measure so that $\mu = \nu \circ j^{-1}$ is a Radon measure. Now $B = \bigcup_{n=1}^\infty nU$, where U is the unit ball of B. Thus, since B is reflexive, U is weakly compact, and since j is weakly continuous (see Schaefer, 1966,

p. 158) we have that $j(B)$ is a Borel subset of E. Hence $j(B) \in \overline{\sigma(E')}^{\mu}$, since by a result of Badrikian and Chevet (1974, p. 347), $\overline{\mathscr{B}(E)}^{\mu} = \overline{\sigma(E')}^{\mu}$ for Radon measures μ. ∎

Let $(\Omega, \mathscr{B})$, $(\Lambda, \mathscr{F})$ be measurable spaces, μ_{XY} a probability measure on the product space $(\Omega \times \Lambda, \mathscr{B} \times \mathscr{F})$ with marginals μ_X and μ_Y. The average mutual information $I(\mu_{XY})$ of the measure μ_{XY} is defined by $I(\mu_{XY}) = \sup \sum_{i,j} \mu_{XY}(A_i \times B_j) \log(\mu_{XY}(A_i \times B_j)/\mu_X(A_i)\,\mu_Y(B_j))$, where the upper bound is taken over all possible finite measurable partitions $\{A_i\}$, $\{B_j\}$ of Ω and Λ, respectively, such that $\mu_X(A_i) > 0$, $\mu_Y(B_j) > 0$.

Let E_X, E_Y be quasi-complete locally convex spaces, and let μ_{XY} be a zero-mean Gaussian measure on $(E_X \times E_Y,\ \sigma(E_X') \times \sigma(E_Y'))$ with covariance operator $R: E_X' \times E_Y' \to E_X \times E_Y$. The marginal and cross-covariance operators of R are denoted R_X, R_Y, and R_{XY}. It is known (Baker, 1973; McKeague, 1980) that $R_{XY} = j_X V_{XY} j_Y^*$, where $V_{XY}: H_Y \to H_X$ is a unique bounded linear operator. The following result was proved by Baker (1978a) for E_X, E_Y Hilbert spaces and a similar proof works for any locally convex E_X, E_Y.

PROPOSITION 2.3. (a) *$\mu_{XY} \sim \mu_X \otimes \mu_Y$ if and only if V_{XY} is Hilbert–Schmidt and $\| V_{XY} \| < 1$.*

(b) *If $\mu_{XY} \sim \mu_X \otimes \mu_Y$ then $I(\mu_{XY}) < \infty$ and $I(\mu_{XY}) = -\frac{1}{2} \sum_n \log(1 - \gamma_n)$, where $\{\gamma_n, n \geqslant 1\}$ are the eigenvalues of $V_{XY}^* V_{XY}$.*

3. THE INFORMATION CAPACITY OF GAUSSIAN CHANNELS

Suppose that messages are selected according to a probability distribution μ_X on $(E_X, \sigma(E_X'))$ and encoded through a measurable map $A: E_X \to E_Y$. A noise process μ_N on $(E_Y, \sigma(E_Y')$ is assumed to be known. The joint distribution of the transmitted message and received signal is given by $\mu_{XY}(D) = \mu_X \otimes \mu_N\{(x, y): (x, Ax + y) \in D\}$, $D \in \sigma(E_X') \times \sigma(E_Y')$. The information capacity, subject to some constraints Q on μ_X and A, is defined to be $\sup_Q I(\mu_{XY})$. For the remainder of this section we shall assume that μ_N is Gaussian with zero mean and covariance operator $R_N: E_Y' \to E_Y$. The RKHS of R_N will be denoted H_N and the injection into E_Y by j_N. Let the measure $\mu_X \circ A^{-1}$ be denoted μ_{AX}. The constraints Q that will be put on μ_X and A are the minimal ones to ensure that the information capacity is finite. One such constraint is $\mu_{AX}^*(H_N) = 1$ (see Baker, 1979a). By Theorem 2.2, if H_N is separable then this constraint may be written $\bar{\mu}_{AX}(H_N) = 1$. The following result was obtained by Baker (1978a) for the Hilbert space case and the proof for general E_X, E_Y is along similar lines.

PROPOSITION 3.1. *Suppose that* μ_X *is Gaussian with zero mean and covariance operator* $R_X: E'_X \to E_X$. *Let* $A: E_X \to E_Y$ *be a continuous linear map and assume that* H_N *is separable and* $\bar{\mu}_{AX}(H_N) = 1$. *Then there exists a unique trace-class covariance operator* T *on* H_N *such that* $AR_XA^* = j_N T j_N^*$ *and* $I(\mu_{XY}) = \frac{1}{2} \sum_n \log(1 + \tau_n)$, *where* $\{\tau_n, n \geqslant 1\}$ *are the eigenvalues of* T.

Proof. Since A is linear and μ_X is Gaussian, μ_{AX} is Gaussian with covariance operator AR_XA^*. Since $\bar{\mu}_{AX}(H_N) = 1$ and H_N is separable, by Theorem 2.2 there exists a unique Gaussian measure ν on H_N such that $\mu_{AX} = \nu \circ j_N^{-1}$. Let T be the covariance operator of ν. T is trace-class and the covariance operator of $\nu \circ j_N^{-1}$ is $j_N T j_N^*$. Thus $AR_XA^* = j_N T j_N^*$.

In order to evaluate $I(\mu_{XY})$ we use Proposition 2.3 and the representation $R_{XY} = j_X V j_Y^*$, where $V = V_{XY}: H_Y \to H_X$ is a unique bounded linear operator. Let $\{\tau_n, n \geqslant 1\}$ denote the eigenvalues of T. It suffices to show that the eigenvalues of V^*V are $\{\tau_n(1+\tau_n)^{-1},\ n \geqslant 1\}$. We have $R_Y = ARA^* + R_N = j_N T j_N^* + j_N j_N^*$. Thus $j_Y j_Y^* = j_N(I + T) j_N^*$, where I is the identity on H_N. This implies that H_Y coincides with the RKHS of $I + T$. Let ρ denote the injection of H_Y into H_N, so that $j_Y = j_N \rho$. Since $R_{XY} = R_X A^*$ (see Baker, 1978a, p. 83), it follows that $j_X V j_Y^* = R_X A^* = j_X j_X^* A^*$, so that $V j_Y^* = j_X^* A^*$ and $A j_X = j_Y V^*$. Hence $AR_XA^* = A j_X j_X^* A^* = j_Y V^* V j_Y^* = j_N \rho V^* V \rho^* j_N^*$. But we also have $AR_XA^* = j_N T j_N^*$. Therefore $T = \rho V^* V \rho^*$. This last equation uniquely determines V^*V. Let $\{e_n, n \geqslant 1\}$ be a CONS in H_N consisting of eigenvectors of T so that $T = \sum_n \tau_n e_n \otimes_{H_N} e_n$. It is now easily checked that $V^*V = \sum_n \tau_n e_n \otimes_{H_Y} e_n$. Moreover, $u_n = (1+\tau_n)^{1/2} e_n$, $n \geqslant 1$, is a CONS for H_Y and $V^*V = \sum_n \tau_n(1+\tau_n)^{-1} u_n \otimes_{H_Y} u_n$. Therefore, $\{\tau_n(1+\tau_n)^{-1},\ n \geqslant 1\}$ is the point spectrum of V^*V, as required. ∎

Remark. The hypotheses in Proposition 3.1 may be weakened slightly. Instead of requiring that H_N be separable and $\bar{\mu}_{AX}(H_N) = 1$, it suffices to have a separable subspace H_N^1 of H_N such that $\bar{\mu}_{AX}(H_N^1) = 1$.

The next two results extend Theorems 1 and 2 of Baker (1978a) to quasi-complete locally convex spaces under the restriction that A be one-to-one. This restriction allows us to use a result of Dobrushin (1959) in order to reduce to the case of Gaussian messages and avoid the martingale arguments of Baker (1979b). The approach in Baker (1978, Lemma 6) fails in our situation because the cylindrical σ-algebra on E_Y is not necessarily separable and the limiting arguments used there no longer hold.

THEOREM 3.2. *Suppose that* $\dim(H_N) \geqslant M$ $(M < \infty$ *and fixed*$)$, $\dim(E_X) \geqslant M$, *and the following constraints* Q *are imposed on* μ_X *and* A:

(a) A *is* $\sigma(E'_X)/\sigma(E'_Y)$ *measurable;*

(b) $\mu_X \circ A^{-1}$ *is concentrated on an* M*-dimensional subspace of* H_N;

(c) *A is 1–1 and* $\overline{\sigma(E_X')}^{\mu_X}/\overline{\sigma(E_Y')}^{\mu_{AX}}$ *bimeasurable on a* $\overline{\sigma(E_X')}^{\mu_X}$ *measurable subset of* E_X *which supports* $\bar{\mu}_X$*;*

(d) $\int_{E_X} \|Ax\|_{H_N}^2 \, d\bar{\mu}_X(x) \leqslant P_0$, *where* $P_0 < \infty$ *is fixed.*

Then

(1) $\sup_Q I(\mu_{XY}) = (M/2)\log(1 + P_0/M)$;

(2) *the supremum is attained when* $\mu_X \circ A^{-1}$ *is Gaussian with zero mean and covariance operator* $(P_0/M)\sum_{n=1}^M e_n \otimes e_n$, *where* $\{e_n, n = 1,..., M\}$ *is an orthonormal subset of* H_N. *The maximizing* $\mu_X \circ A^{-1}$ *can be obtained with a Gaussian* μ_X *and a continuous linear A.*

Proof. First we show that it suffices to consider the case that $E_X = E_Y = E$ and A is the identity on E. Let Q denote the set of pairs (A, μ_X) satisfying constraints (a)–(d), Q' the set of probability measures μ_W on $\sigma(E_Y')$ such that μ_W is concentrated on an M-dimensional subspace L_W of H_N and $\int_{E_Y} \|x\|_{H_N}^2 \, d\bar{\mu}_W(x) \leqslant P_0$. Let $\mu_W \in Q'$ and let L_X denote an M-dimensional subspace of E_X, A_0 an isomorphism from L_X onto L_W. By a consequence of the Hahn–Banach Theorem (Robertson and Robertson, 1964, p. 29), A_0 can be extended to a continuous linear map $A: E_X \to L_W$. A is $\sigma(E_X')/\sigma(E_Y')$ measurable. Since μ_W is concentrated on a finite dimensional subspace on which it is Radon it can be extended to a Radon measure on the whole of E_Y. Thus, by Badrikian and Chevet (1974, p. 374), $\mathscr{B}(E_Y) \subset \overline{\sigma(E_Y')}^{\mu_W}$ and since μ_W is concentrated on L_W we may define a Radon measure μ_X on E_X by $\mu_X(D) = \bar{\mu}_W[A(D \cap L_X)]$ for $D \in \mathscr{B}(E_Y)$. Note that $L_X \in \mathscr{B}(E_X) \subset \overline{\sigma(E_X')}^{\mu_X}$ and μ_X is concentrated on L_X. Let $C \in \sigma(E_Y')$. Then $\mu_W(C) = \bar{\mu}_W(C \cap L_W) = \bar{\mu}_W[AA^{-1}(C \cap L_W)] = \bar{\mu}_W[A(A^{-1}(C) \cap L_X)] = \mu_X[A^{-1}(C)]$, so that $\mu_W = \mu_X \circ A^{-1}$. Therefore $\bar{\mu}_W = \bar{\mu}_X \circ A^{-1}$ so μ_X and A satisfy constraints (b) and (d). A is 1–1 and $\overline{\sigma(E_X')}^{\mu_X}/\overline{\sigma(E_Y')}^{\mu_W}$ bimeasurable on L_X so that μ_X and A satisfy constraint (c). Hence (A, μ) belongs to Q, and as in Baker (1978a, Lemma 4) we have $I(\mu_{X,AX+N}) = I(\mu_{AX,AX+N})$. Conversely, given $(A, \mu_X) \in Q$ it is clear that $\mu_W \equiv \mu_X \circ A^{-1}$ belongs to Q', and as before $I(\mu_{X,AX+N}) = I(\mu_{W,W+N})$. Therefore $\sup_{A,\mu_X \in Q} I(\mu_{X,AX+N}) = \sup_{\mu_W \in Q'} I(\mu_{W,W+N})$.

Now consider the case in which $E_X = E_Y = E$ and A is the identity on E. By Dobrushin (1959) we have

$$I(\mu_{XY}) = \sup_F \sup_{\{C_i\}\{D_j\}} \sum_{i,j} \log \left[\frac{\mu_{XY}(C_i \times D_j)}{\mu_X(C_i)\,\mu_Y(D_j)}\right] \mu_{XY}(C_i \times D_j),$$

where the outside supremum runs over all finite dimensional subspaces F of E' and the inside supremum over all possible finite $\mathscr{C}_F$-measurable partitions $\{C_i\}$, $\{D_j\}$ of E, such that $\mu_X(C_i) > 0$, $\mu_Y(D_j) > 0$. Since C_i, D_j belong to $\mathscr{C}$, there exist A_i, $B_j \in \mathscr{B}(\Re^n)$ such that $C_i = \pi_n^{-1}(A_i)$, $D_j = \pi_n^{-1}(B_j)$, where

$\pi_n: E \to \mathfrak{R}^n$ is defined by $\pi_n(x) = (\langle f_1, x\rangle, ..., \langle f_n, x\rangle)$, $x \in E$, for some $f_1, ..., f_n \in E'$. Thus,

$$\begin{aligned}\mu_{XY}(C_i \times D_j) &= \mu_X \otimes \mu_N\{(x, y): (x, x+y) \in \pi_n^{-1}(A_i) \times \pi_n^{-1}(B_j)\}\\ &= \mu_X \otimes \mu_N\{(x, y): (\pi_n x, \pi_n x + \pi_n y) \in A_i \times B_j\}\\ &= \mu_X^n \otimes \mu_N^n\{(u, v) \in \mathfrak{R}^n: (u, u+v) \in A_i \times B_j\},\end{aligned}$$

where $\mu_X^n = \mu_X \circ \pi_n^{-1}$ and $\mu_N^n = \mu_N \circ \pi_n^{-1}$. Similarly $\mu_Y(D_j) = \mu_X^n \otimes \mu_N^n\{(u, v) \in \mathfrak{R}^n: u + v \in B_j\}$. Therefore, the inside supremum in the expression for $I(\mu_{XY})$ is equal to the mutual information for a channel with message μ_X^n and noise μ_N^n on $\mathfrak{R}^n$. For such a channel it is known (see Baker, 1978a, Lemma 6) that, when the covariance of μ_X^n is fixed and its distribution allowed to vary, the mutual information is maximized when μ_X^n is Gaussian. We conclude that in order to evaluate $\sup_Q I(\mu_{XY})$ it suffices to consider Gaussian μ_X belonging to Q. We already know an expression for $I(\mu_{XY})$ in this case. By Proposition 3.1 and the remark following its proof, $I(\mu_{XY}) = \frac{1}{2}\sum_{n=1}^{M} \log(1 + \tau_n)$, where $\tau_n \geqslant 0$, $R_X = \sum_{n=1}^{M} \tau_n e_n \otimes e_n$ and $\{e_n,\ n = 1, ..., M\}$ are orthonormal in H_N. Constraint (d) implies that $\sum_{n=1}^{M} \tau_n \leqslant P_0$. It follows (see Baker, 1978, p. 84) that $I(\mu_{XY})$ is maximized when $\tau_n = P_0/M$ for $n = 1, ..., M$ and we conclude that $\sup_Q I(\mu_{XY}) = (M/2) \log(1 + P_0/M)$. ∎

THEOREM 3.3. *Suppose that E_X and H_N are infinite dimensional and H_N is separable. Impose the following constraints Q on μ_X and A:* (a), (c), (d) *of Theorem* 3.2 *and* (b), $\bar{\mu}_{AX}(H_N) = 1$. *Then* $\sup_Q I(\mu_{XY}) = P_0/2$, *and the supremum cannot be attained.*

Proof. First constrain $\mu_X \circ A^{-1}$ to have M-dimensional support. By Theorem 3.2 there exists a pair (μ_X^M, A^M) satisfying Q and such that $I(\mu_{XY}^M) = (M/2) \log(1 + P_0/M)$, where μ_{XY}^M is the corresponding joint distribution of input and output. Thus, $\lim_{M\to\infty} I(\mu_{XY}^M) = P_0/2$, showing that $\sup_Q I(\mu_{XY}) \geqslant P_0/2$.

To show the converse inequality, reduce to the case where $E_X = E_Y = E$, $A =$ identify on E, and μ_X is Gaussian, as in the proof of Theorem 3.2. By Proposition 3.1 and (c) it follows that $\sup_Q I(\mu_{XY}) \leqslant \frac{1}{2}\sum_n \tau_n \leqslant P_0/2$.

If the supremum is attained it must be attained by a Gaussian μ_X on E. Then, using the notation of Proposition 3.1, we have $\sum_n \log(1 + \tau_n) = P_0$. But from Theorem 2.2 and the constraints $\bar{\mu}_X(H_N) = 1$, $\int_E \|x\|_{H_N}^2 \, d\mu_X(x) \leqslant P_0$, it follows that $\sum_n \tau_n \leqslant P_0$. These two equations can hold simultaneously only if $P_0 = 0$. Hence $\sup_Q I(\mu_{XY})$ cannot be attained. ∎

4. Robustness of the Information Capacity

In this section we investigate the robustness and sensitivity of the information capacity to small deviations from normality in the noise process μ_N. Gualtierotti (1979, 1980) introduced a class of contaminated Gaussian laws, called QN-laws, and studied the information capacity problem for channels having QN-laws as noise. Here we make use of an inequality of Ihara (1978) to put bounds on the information capacity of these contaminated Gaussian channels and give conditions under which their information capacity tends to the information capacity of the corresponding Gaussian channels.

Let P_1 and P_2 be two probability measures defined on the same measurable space $(\Omega, \mathscr{F})$. The entropy $H_{P_2}(P_1)$ of P_1 with respect to P_2 is defined by $H_{P_2}(P_1) = \sup \sum_i P_1(C_i) \log(P_1(C_i)/P_2(C_i))$, where the supremum is taken over all finite measurable partitions $\{C_i\}$ of Ω. Let E be a locally convex space and let μ_N be a noise distribution on $(E, \sigma(E'))$ which is not necessarily Gaussian. Assume, however, that μ_N is of weak second order. In the notation of Section 3 we shall here only be considering channels with $E_X = E_Y = E$ and $A =$ the identity on E. Let $\mathscr{K}$ be a class of Gaussian covariance operators on E. Let $\mathscr{X}$ be the class of weak second-order measures μ_X on $(E, \sigma(E'))$ having a covariance operator belonging to $\mathscr{K}$. $\mathscr{X}$ represents a class of allowable input distributions determined by their covariance operators. For example, the constraints in Theorem 3.3 (with $E_X = E_Y = E$, $A =$ identity) have $\mathscr{K}$ equal to the class of covariance operators R_X which have a representation $R_X = j_N T j_N^*$, where T is a trace-class covariance operator on H_N such that $\mathrm{Trace}(T) \leqslant P_0$. Let $C(\mu_N; \mathscr{X})$ denote the information capacity $\sup\{I(\mu_{XY}) : \mu_X \in \mathscr{X}\}$. The following result is an extension of Ihara's inequality to locally convex spaces.

Theorem 4.1. *Suppose that there exists a zero-mean Gaussian measure μ_N^0 on $(E, \sigma(E'))$ having the same covariance operator as μ_N. Then*

$$C(\mu_N^0; \mathscr{X}) \leqslant C(\mu_N; \mathscr{X}) \leqslant C(\mu_N^0; \mathscr{X}) + H_{\mu_N^0}(\mu_N).$$

Remark. Ihara stated this result for the case $E = L^2[0, T]$. His proof is based on a result of Huang and Johnson (1962) which holds only for a small class of signal and noise covariances (see Baker, 1969). However, Baker (1978a, Lemma 6) extended Huang and Johnson's result to arbitrary Gaussian covariance operators for the signal and noise so that Ihara's result is valid for $E = L^2[0, T]$. For locally convex spaces E the same reasoning as in the proof of Theorem 3.2 works to reduce to the Gaussian case from which the result follows.

Let H denote a separable Hilbert space, W a covariance operator on H and let k be a real number. Let P be a zero-mean Gaussian probability measure on H with covariance operator R. Let $c_Q^{-1} = k^2 + \operatorname{Tr} WR$ and $q(x) = c_Q(k^2 + \| W^{1/2}x\|^2)$. Then $q(x) \geqslant 0$, $\int_H q\, dP = 1$ so that the relation $dQ = q\, dP$ defines a probability measure Q on H. Q is said to be a QN-law with parameters k, W, and R, and we write $Q \div QN(k, W, R)$. $\|W\|$ and $\operatorname{Tr} WR$ are rough measures of the "degree of deviation of Q from normality." Q has mean zero and covariance operator $R^{1/2}(I + 2c_Q R^{1/2} WR^{1/2}) R^{1/2}$ (see Gualtierotti, 1979). Let $\{Q_n, n \geqslant 1\}$ be a sequence of QN-laws, $Q_n \div QN(k_n, W_n, R)$, with R and P fixed. We pose the following question: If Q_n converges to P (in some sense) then does the information capacity for the channel with noise Q_n converge to that with noise P?

In order to give a satisfactory answer to this question we use the following notation. If ν is a probability measure on H with RKHS H_ν then for $S < \infty$ fixed let $\mathscr{X}_\nu$ denote the set of input distributions given by

$$\mathscr{X}_\nu = \left\{ \mu_X : \mu_X(H_\nu) = 1 \text{ and } \int_H \|x\|^2_{H_\nu}\, d\mu_X(x) \leqslant S \right\}.$$

Note that $\mathscr{X}_\nu$ is a constraint on the covariance operator of μ_X and so falls under the framework for Ihara's Theorem. The following result gives bounds on the information capacity in terms of the degree of contamination of the Gaussian noise.

THEOREM 4.2. *Suppose that* $Q \div QN(k, W, R)$, *where* $k > 0$, *and* $P \div N(0, R)$. *Then*

$$C(P; \mathscr{X}_P) \leqslant C(Q; \mathscr{X}_Q) \leqslant C(P; \mathscr{X}_P) + \frac{3 \operatorname{Tr} WR}{k^2}.$$

COROLLARY 4.3. *Suppose that* $Q_n \div QN(k_n, W_n, R)$ *for* $n \geqslant 1$ *and* $\operatorname{Tr}(W_n R) \to 0$, $\liminf_{n\to\infty} k_n > 0$. *Then* $C(Q_n; \mathscr{X}_{Q_n}) \to C(P; \mathscr{X}_P)$ *as* $n \to \infty$.

Proof of Theorem 4.2. Let $Q \div QN(k, W, R)$ and let Q^0 denote the zero-mean measuye on H with the same covariance operator as Q, namely $R^{1/2}(I + 2c_Q R^{1/2} WR^{1/2}) R^{1/2}$. The operator $T = 2c_Q R^{1/2} WR^{1/2}$ is Hilbert–Schmidt, and since $c_Q > 0$ and W is non-negative, T does not have -1 as an eigenvalue. Hence, by Theorem 5.1 of Rao and Varadarajan (1963), we have that P and Q^0 are mutually absolutely continuous. Therefore $Q \ll Q^0$ and

$$H_{Q^0}(Q) = \int_H \left[\log \frac{dQ}{dQ^0}\right] dQ = \int \left[\log \frac{dQ}{dP}\right] \frac{dQ}{dP}\, dP + \int \left[\log \frac{dP}{dQ^0}\right] dQ. \qquad (*)$$

The first term on the rhs of $(*)$ is

$$\int [\log c_Q(k^2 + \| W^{1/2}x\|^2)]\, c_Q(k^2 + \| W^{1/2}x\|^2)\, dP$$

$$= \int \left[\log c_Q k^2 + \log\left(1 + \frac{1}{k^2}\| W^{1/2}x\|^2\right)\right] c_Q(k^2 + \| W^{1/2}x\|^2)\, dP$$

$$\leqslant c_Q \int \| W^{1/2}x\|^2\, dP + \frac{c_Q}{k^2}\int \| W^{1/2}x\|^4\, dP \qquad (\text{since } c_Q k^2 \leqslant 1)$$

$$= c_Q \operatorname{Tr}(WR) + \frac{c_Q}{k^2}\{2 \operatorname{Tr}(WR)^2 + \operatorname{Tr}(WR)^2\}$$

$$= \frac{k^2 \operatorname{Tr}(WR) + 2 \operatorname{Tr}(WR)^2 + (\operatorname{Tr} WR)^2}{k^2(k^2 + \operatorname{Tr} WR)}.$$

The second term on the rhs of $(*)$ equals

$$\int \left\{\tfrac{1}{2} \sum_{i=1}^{\infty} [\log(1 + \tau_i) - (\eta_i(x))^2\, \tau_i(1 + \tau_i)^{-1}]\right\} dQ(x),$$

where $\eta_i = \sum_{j=1}^{\infty} \lambda_j^{-1/2}\langle e_j, v_i\rangle\langle e_j, x\rangle$; $\{\lambda_j, e_j, j \geqslant 1\}$ are the eigenvalues, orthonormal eigenvectors of R, and $\{\tau_i, v_i, i \geqslant 1\}$ are the eigenvalues, orthonormal eigenvectors of $2c_Q R^{1/2} W R^{1/2}$ (see Rao and Varadarajan, 1963, p. 318). Therefore the second term is less than

$$\frac{1}{2}\sum_{i=1}^{\infty} \log(1 + \tau_i) \leqslant c_Q \operatorname{Tr}(R^{1/2} W R^{1/2}) = \frac{\operatorname{Tr}(R^{1/2} W R^{1/2})}{k^2 + \operatorname{Tr} WR}.$$

Note that $\dim(H_P) = \dim(H_Q)$ since $P \sim Q^0$. Therefore, by Theorem 3.2 if $\dim(H_P) < \infty$ (or by Theorem 3.3 if $\dim(H_P) = \infty$), we have $C(Q^0; \mathscr{X}_Q) = C(P; \mathscr{X}_P)$. However, by Ihara's inequality

$$C(Q^0; \mathscr{X}_Q) \leqslant C(Q; \mathscr{X}_Q) \leqslant C(Q^0; \mathscr{X}_Q) + H_{Q^0}(Q),$$

with which the previous statements complete the proof of the theorem. ∎

5. Calculation of the Coding Capacity

A basic tool for the construction of codes is a lemma of Feinstein (see Ash, 1965, p. 232) and its generalization by Kadota (1970) to channels for which the noise process does not possess a probability density. A version of these results, to be used in this section, is stated below.

Suppose that we are given measurable spaces $(\Omega, \mathscr{B})$ of transmitted signals and $(\Lambda, \mathscr{F})$ of received signals. Let $f: \Omega \times \Lambda \to \Lambda$ be a $\mathscr{B} \times \mathscr{F} / \mathscr{F}$ measurable function and μ_N a probability measure on $(\Lambda, \mathscr{F})$. Here, μ_N represents a noise process. For the additive channels considered in this article, the spaces Ω and Λ are locally convex spaces and f is the addition operation. If a message $x \in \Omega$ is transmitted then the received signal is a random element of Λ having distribution $\mu_N \circ f_x^{-1}$, where $f_x: \Lambda \to \Lambda$ is defined by $f_x(y) = f(x, y)$. Note that f_x is $\mathscr{F}/\mathscr{F}$ measurable (see Halmos, 1950, p. 142). A code (k, F, ε) is a set of k code words $x_1, \ldots, x_k$ belonging to $F \subset \Omega$, and a measurable partition of Λ into k decoding sets $C_1, \ldots, C_k$ such that $\mu_N \circ f_{x_i}^{-1}(C_i) > 1 - \varepsilon$, for $i = 1, \ldots, k$. Such a code is used as follows: If a received signal belongs to C_i then the receiver concludes that x_i has been transmitted. The probability of error is less than ε, regardless of which code word is transmitted.

Now let μ_X be an arbitrary probability measure on $(\Omega, \mathscr{B})$ and define μ_{XY} on $\Omega \times \Lambda$ by $\mu_{XY}(D) = \mu_X \otimes \mu_N\{(x, y): (x, f(x, y)) \in D\}$. Let μ_Y be the projection of μ_{XY} onto Λ.

LEMMA 5.1 (Baker, 1979b). *Suppose $\mu_N \circ f_x^{-1} \sim \mu_N$ a.e. $d\mu_X(x)$. Then $\mu_{XY} \sim \mu_X \otimes \mu_Y$ and $\mu_Y \sim \mu_N$. Moreover, if $[d\mu_N \circ f_x^{-1} \mid d\mu_Y](y)$ is $\mathscr{B} \times \mathscr{F}$ measurable then*

$$[d\mu_{XY} | d\mu_X \otimes \mu_Y](x, y) = [d\mu_N \circ f_x^{-1} | d\mu_Y](y) \qquad \text{a.e.} \quad d\mu_X \otimes \mu_Y(x, y). \quad \blacksquare$$

LEMMA 5.2 (Kadota, 1970). *Suppose that $\mu_N \circ f_x^{-1} \sim \mu_N$ a.e. $d\mu_X(x)$ and $[d\mu_N \circ f_x^{-1} | d\mu_Y](y)$ is $\mathscr{B} \times \mathscr{F}$ measurable. For any real α let $A = \{(x, y) \in \Omega \times \Lambda: \log[d\mu_{XY} | d\mu_X \otimes \mu_Y](x, y) > \alpha\}$. Then for each integer k and $F \in \mathscr{B}$ there exists a code (k, F, ε) such that*

$$\varepsilon \leqslant k e^{-\alpha} + \mu_{XY}(A^c) + \mu_X(F^c). \quad \blacksquare$$

For the remainder of this section we take $\Omega = \Lambda = E$, where E is a quasi-complete locally convex space. The function $f: E \times E \to E$ is taken to be the addition operation $(f(x, y) = x + y, \text{ for } x, \ y \in E)$ which is $\sigma(E') \times \sigma(E') | \sigma(E')$ measurable. Let μ_N be a zero-mean weak second-order probability measure on $\sigma(E')$, with covariance operator $R_N: E' \to E$ which has RKHS denoted H_N. This set-up is known as the channel with additive noise. We do not assume that μ_N is Gaussian unless explicitly mentioned.

A code (k, n, ε) for this channel will be as before but with F replaced by n, a positive integer, representing the constraints on code words $x_1, \ldots, x_k$ given by

(a) $x_i \in H_N$, for $i = 1, \ldots, k$;

(b) $\|x_i\|_{H_N}^2 \leqslant n P_0$, for $i = 1, \ldots, k$;

(c) $\dim sp\{x_i, i = 1, \ldots, k\} \leqslant n$.

These constraints form the natural generalization of the constraints for the time-discrete Gaussian channel with average power limitation P_0, as in Shannon (1959), in which case $x_1,..., x_k \in \Re^n$ and $\sum_{j=1}^n x_{ij}^2 \leqslant nP_0$, for $i = 1,..., k$. Let $[a]$ denote the integer part of $a \in \Re$. A real number $R \geqslant 0$ is said to be a permissible rate of transmission if there exist codes $([e^{nR}], n, \varepsilon_n)$ with $\varepsilon_n \to 0$ as $n \to \infty$. The coding capacity, denoted C_0, is defined as the supremum of all permissible transmission rates. Let μ_n^0 denote the Gaussian cylindrical measure on E having zero-mean and the same covariance operator as μ_N. If μ_N^0 is countably additive then the entropy $H_{\mu_N^0}(\mu_N)$ of μ_N with respect to μ_N^0 may be defined as in Section 4. Otherwise we define $H_{\mu_N^0}(\mu_N)$ by regarding μ_N and μ_N^0 as measures on E'^a on which μ_N^0 is countably additive by Kolmogorov's Theorem.

THEOREM 5.3. (1) *If* $H_{\mu_N^0}(\mu_N) < \infty$ *then* $C_0 \leqslant \frac{1}{2}\log(1 + P_0)$.

(2) *If* $H_{\mu_N^0}(\mu_N) < \infty$ *and* $\dim(H_N) < \infty$ *then* $C_0 = 0$.

(3) *If* μ_N *is Gaussian and* $\dim(H_N) = \infty$ *then* $C_0 = \frac{1}{2}\log(1 + P_0)$.

Remark 5.4. An intuitive explanation of (2) is that a finite dimensional space H_N is "too small" to allow a code to attain any positive rate of transmission.

Proof of Theorem 5.3. First let us suppose that $\dim(H_N) = \infty$ and μ_N is Gaussian. Let $\{e_n, n \geqslant 1\}$ be an orthonormal sequence in H_N. Choose $Q < P_0$ and let μ_X be the distribution of the random element $X = \sqrt{Q}\sum_{i=1}^n \gamma_i e_i$, where $\{\gamma_i, i = 1,..., n\}$ are i.i.d. $N(0, 1)$ random variables. Then μ_X has mean zero and covariance operator $Q\sum_{i=1}^n e_i \otimes e_i$. By Theorem 2.1, $\mu_N \circ f_x^{-1} \sim \mu_N$ a.e. $d\mu_X(x)$, since μ_X is concentrated on H_N. Using Theorem 2.1 it is possible to evaluate $[d\mu_N \circ f_x^{-1}/d\mu_Y](y)$ and check that it is $\sigma(E') \times \sigma(E')$ measurable. For full details see McKeague (1980). Thus, Lemma 5.2 is applicable for μ_X.

Let $F_n = \{x \in sp(e_1,..., e_n): \|x\|_{H_N}^2 \leqslant nP_0\}$ and note that $\mu_X(F_n^c) = P\{n^{-1}\sum_{i=1}^n \gamma_i^2 > P_0/Q\} \to 0$ as $n \to \infty$, by the law of large numbers. Baker (1978a, p. 78) has evaluated $d\mu_{XY}/d\mu_X \otimes \mu_Y$ in the Hilbert space case and it is clear that the same method works here. Let V denote the unique bounded linear operator in the representation $R_{XY} = j_X V j_Y^*$. Baker (1978a, p. 78) shows that

$$\log \frac{d\mu_{XY}}{d\mu_X \otimes \mu_Y}(x, y) = \frac{1}{2}\sum_{i=1}^{\infty} \{a_i^2(x, y) - b_i^2(x, y) - \log(1 + \delta_i) - \log(1 - \delta_i)\},$$

where the series converges a.e. $d\mu_{XY}$, $\{\delta_i^2, i \geqslant 1\}$ are the eigenvalues of V^*V, both a_i and b_i, $i \geqslant 1$, are zero-mean Gaussian random variables with respect to μ_{XY} and $\{a_i, i \geqslant 1; b_j, j \geqslant 1\}$ is a set of independent random variables with respect to μ_{XY}. From the proof of Proposition 3.1 we have $\delta_i = (Q/(1+Q))^{1/2}$, $i = 1,..., n$, and $\delta_i = 0$, $i > n$, so that

$$\log \frac{d\mu_{XY}}{d\mu_X \otimes \mu_Y}(x, y) = \frac{1}{2} \sum_{i=1}^{n} \{a_i^2(x, y) - b_i^2(x, y)\} + \frac{1}{2} n \log(1 + Q),$$

where $\{a_1,..., a_n, b_1,..., b_n\}$ are i.i.d., zero-mean Gaussian random variables with variance $(Q/(1+Q))^{1/2}$ with respect to μ_{XY}. Let $\delta > 0$, $\alpha_n = \frac{1}{2} n \log(1+Q) - n\delta$ and $A_n = \{(x, y): \log(d\mu_{XY}/d\mu_X \otimes \mu_Y)(x, y) > \alpha_n\}$, so that $A_n^c = \{(x, y): \frac{1}{2} \sum_{i=1}^{n} (a_i^2 - b_i^2) \leqslant n\delta\}$ and $\mu_{XY}(A_n^c) \to 0$ as $n \to \infty$, by the law of large numbers. Let $R < \frac{1}{2} \log(1 + P_0)$ and $k_n = [e^{nR}]$. Note that $k_n e^{-\alpha_n} \leqslant e^{nR-(1/2)n \log(1+Q)+n\delta} \to 0$ as $n \to \infty$, provided Q is chosen sufficiently close to P_0 and δ is sufficiently small. Thus, by Lemma 5.2 there exist codes $([e^{nR}], F_n, \varepsilon_n)$ such that $\varepsilon_n \to 0$ as $n \to \infty$. Therefore $C_0 \geqslant \frac{1}{2} \log(1 + P_0)$.

For the converse inequality suppose that $H_{\mu_N^0}(\mu_N) < \infty$ and do not assume that μ_N is Gaussian. Let $0 < \varepsilon < \frac{1}{4}$ and suppose that we have a code (k, n, ε) with code words $x_1,..., x_k$ and corresponding decoding sets $C_1,..., C_k$. Write $\mu_i \equiv \mu_N \circ f_{x_i}^{-1}$ so that $\mu_i(C_i) \geqslant 1 - \varepsilon$, for $i = 1,..., k$. Using a standard technique (see Ash, 1965, p. 253) it is possible to approximate the sets $C_1,..., C_k$ by disjoint cylinder sets $D_1,..., D_k$ and produce a new code $(k, n, 2\varepsilon)$ with code words $x_1,..., x_k$ and decoding sets $D_1,..., D_k$. Let F be a finite dimensional subspace of E' with basis $\{f_1,..., f_r\}$ such that all the cylinder sets D_i, $i = 1,..., k$, are based on F. Let π_r be the map $\pi_r: E \to \mathfrak{R}^r$ defined by $\pi_r(x) = (\langle f_1, x\rangle,..., \langle f_r, x\rangle)$ so that $D_i = \pi_r^{-1}(A_i)$, for some $A_i \in \mathscr{B}(\mathfrak{R}^r)$, $i = 1,..., k$.

Since the D_i are disjoint so are the A_i. Define $y_i = \pi_r(x_i)$, $\mu_i^r = \mu_i \circ \pi_r^{-1}$, for $i = 1,..., k$. Then

$$\mu_i^r(A_i) = \mu_i \circ \pi_r^{-1}(A_i) = \mu_i(D_i) \geqslant 1 - 2\varepsilon$$

so that the (y_i, A_i) form a code for the channel on $\mathfrak{R}^r$ with noise $\mu_N \circ \pi_r^{-1}$. Notice that the y_i, $i = 1,..., k$, are distinct because we have assumed that $\varepsilon < \frac{1}{4}$; if two coincide then we would have $\mu_i^r(A_i) > \frac{1}{2}$, $\mu_i^r(A_j) > \frac{1}{2}$ for some $i \neq j$, which would be a contradiction.

Let μ_X be the measure on E given by $\mu_X = (1/k) \sum_{i=1}^{k} \delta_{x_i}$. Define $\rho: E \times E \to \mathfrak{R}^r \times \mathfrak{R}^r$ by $\rho(x, y) = (\pi_r x, \pi_r y)$ and put $\mu_{XY}^r = \mu_{XY} \circ \rho^{-1}$. Then $I(\mu_{XY}^r) \leqslant I(\mu_{XY})$ (see Baker, 1978a, Lemma 4). However, since $x_1,..., x_k$ satisfy the constraints (a), (b), (c) above, μ_X satisfies the following: μ_X is concentrated on an n dimensional subspace of H_N and $\int \|x\|_{H_N}^2 \, d\mu_X(x) \leqslant nP_0$.

Therefore, by Theorem 3.2 and Ihara's inequality, $I(\mu^r_{XY}) \leqslant (n/2) \log(1+P_0) + H_{\mu^0_N}(\mu_N)$. Since $H_{\mu^0_N}(\mu_N) < \infty$, if follows (see Pinsker, 1960) that $\mu_N \ll \mu^0_N$. Thus, each $\mu^r_{Y|x_i}$ has a density on $\mathfrak{R}^r$ and it is possible to define the conditional entropy $H^r(X|Y)$ of X given Y on $\mathfrak{R}^r$, as in Ash (1965, p. 241). Since $I(\mu^r_{XY}) = H^r(X) - H^r(X|Y)$, where $H^r(X)$ is the entropy of μ^r_X which equals $\log k$, we have, using Fano's inequality (see Ash, 1965, p. 244),

$$\begin{aligned}\log k &= I(\mu^r_{XY}) + H^r(X|Y)\\ &\leqslant \frac{n}{2}\log(1+P_0) + H_{\mu^0_N}(\mu_N) + \log 2 + 2\varepsilon \log k.\end{aligned}$$

Therefore,

$$\log k \leqslant \left[\frac{n}{2}\log(1+P_0) + K\right] \Big/ (1-2\varepsilon), \tag{$\dagger$}$$

where K is a constant which is independent of n. Let $R > \frac{1}{2}\log(1+P_0)$ and suppose $k \geqslant e^{nR}$. Then from ($\dagger$) it follows that

$$\varepsilon \geqslant \frac{n[R - \frac{1}{2}\log(1+P_0)] - K}{2nR} \to \frac{R - \frac{1}{2}\log(1+P_0)}{2R} > 0,$$

as $n \to \infty$. Thus, if $R > \frac{1}{2}\log(1+P_0)$, no sequence of codes $([e^{nR}], n, \varepsilon_n)$ can exist with $\varepsilon_n \to 0$ as $n \to \infty$, so that $C_0 \leqslant \frac{1}{2}\log(1+P_0)$. Finally, assuming $\dim(H_N) = M < \infty$, we have instead of ($\dagger$),

$$\log k \leqslant \left[\frac{M}{2}\log\left(1 + \frac{nP_0}{M}\right) + K\right] \Big/ (1-2\varepsilon).$$

Let $R > 0$ and suppose $k \geqslant e^{nR}$. Then,

$$\varepsilon \geqslant \left[nR - \frac{M}{2}\log\left(1+\frac{nP_0}{M}\right) - K\right] \Big/ 2nR \to \frac{1}{2},$$

as $n \to \infty$. Thus, if $R > 0$, no sequence of codes $([e^{nR}], n, \varepsilon_n)$ can exist with $\varepsilon_n \to 0$ as $n \to \infty$, so that $C_0 = 0$. ∎

6. The Coding Capacity of Time-Continuous Channels

Let $N = \{N_t, -\infty < t < \infty\}$ be a real-valued second-order stochastic process on a probability space $(\Omega, \mathscr{F}, P)$. N represents a noise process and is not assumed Gaussian unless explicitly stated. The RKHS of N is denoted H_N and is assumed to be separable. Code words are given by real-valued functions $S: (-\infty, \infty) \to \mathfrak{R}$ which vanish outside the time interval $[0, T]$. A decision is made as to the identity of the code word transmitted during the interval $[0, T]$ after observing the output during that interval. Let H_N^T denote the RKHS of $\{N_t, 0 \leqslant t \leqslant T\}$. The following constraints are to be imposed on code words $S = \{S_t, -\infty < t < \infty\}$:

(a) S vanishes outside $[0, T]$;

(b) S restricted to $[0, T]$ belongs so H_N^T;

(c) $\|S\|_{H_N^T}^2 \leqslant P_0 T$.

A code (k, T, ε) for this channel is a set of code words $S^{(1)}, ..., S^{(k)}$ satisfying the constraints (a), (b), (c) together with disjoint decoding sets $C_1, ..., C_k$, belonging to $\mathscr{B}_T$, the cylindrical σ-algebra on $\mathfrak{R}^{[0,T]}$, such that

$$P\{(S_t^{(i)} + N_t)_0^T \in C_i\} \geqslant 1 - \varepsilon, \qquad \text{for } i = 1, ..., k.$$

A real number $R \geqslant 0$ is said to be a permissible rate of transmission if there exist codes $([e^{RT}], T, \varepsilon_T)$ with $\varepsilon_T \to 0$ as $T \to \infty$. The coding capacity, denoted C_0, is defined as the supremum of all permissible transmission rates.

Let N^0 denote a zero-mean Gaussian stochastic process with the same covariance as N and let μ_N, μ_{N^0} be the measures on $\mathfrak{R}^{(-\infty,\infty)}$ induced by N and N^0, respectively. Denote by $H_{N^0}(N)$ the entropy of μ_N with respect to μ_{N^0} as defined in Section 4.

THEOREM 6.1. *For the time-continuous channel with additive noise N:*

(1) *If* $H_{N^0}(N) < \infty$ *then* $C_0 \leqslant P_0/2$.

(2) *If* $H_{N^0}(N) < \infty$ *and* $\dim(H_N) < \infty$ *then* $C_0 = 0$.

(3) *If* N *is Gaussian and* $\dim(H_N^T) = \infty$ *for some* $T > 0$ *then* $C_0 = P_0/2$.

Proof. First let us suppose that N is Gaussian and $\dim(H_N^{T_0}) = \infty$, where $T_0 > 0$. Let m be a fixed positive integer and let $n = mT$ vary as $T \geqslant T_0$ varies over positive integers. Consider the general channel on $\mathfrak{R}^{[0,T_0]}$ with noise $\{N_t, 0 \leqslant t \leqslant T_0\}$ and the following constraints on code words $x_1, ..., x_k \in \mathfrak{R}^{[0,T_0]}$:

(a) $x_i \in H_N^{T_0}$, for $i = 1, ..., k$;

(b) $\|x_i\|_{H_N^{T_0}}^2 \leqslant n(P_0/m)$, for $i = 1, ..., k$;

(c) $\dim sp\{x_i, i = 1, ..., k\} \leqslant n$.

Since N is Gaussian and $\dim(H_N^{T_0}) = \infty$ we have, by Theorem 5.3, that the coding capacity for this channel is $\frac{1}{2}\log(1 + P_0/m)$. Let $\varepsilon > 0$ and $R' < \frac{1}{2}\log(1 + P_0/m)$. Then there exists a code $([e^{nR'}], n, \varepsilon)$ with code words

$S^{(i)}$, $i=1,\ldots,[e^{nR'}]$, and corresponding decoding sets $C_i^* \in \mathscr{B}_{T_0}$. Let $C_i = C_i^* \times \Re^{(T_0,T]} \in \mathscr{B}_T$, where T is determined by the choice of n. The C_i will act as decoding sets since they are disjoint. Since there is a natural norm preserving injection of $H_N^{T_0}$ into H_N^T we shall identify each $S^{(i)} \in H_N^{T_0}$ with its corresponding element in H_N^T. Define $S^{(i)}$ to be zero outside $[0, T]$. In this way we have a code $([e^{nR'}], T, \varepsilon)$ for the time-continuous channel and it follows that if $R = mR' < (m/2)\log(1+P_0/m)$ there exists a code $([e^{RT}], T, \varepsilon)$ for the time-continuous channel. Therefore $C_0 \geqslant (m/2)\log(1+P_0/m)$ and letting $m \to \infty$, we conclude that $C_0 \geqslant P_0/2$.

For the converse inequality suppose that $H_{N^0}(N) < \infty$ and do not assume that N is Gaussian. Let $0 < \varepsilon < \frac{1}{3}$ and suppose that we have a code (k, T, ε) with code words $S^{(1)},\ldots, S^{(k)}$ and corresponding decoding sets $C_1,\ldots, C_k$ for the time-continuous channel. Write $P_i(A) \equiv P\{(S_t^{(i)} + N_t)_0^T \in A\}$, for $A \in \mathscr{B}_T$ so that $P_i(C_i) \geqslant 1-\varepsilon$ for $i=1,\ldots,k$. There exist cylinder sets D_i^*, $i=1,\ldots,k$, whose base sets are finite unions of semiclosed intervals of the form $\{(x_1,\ldots,x_n) \in \Re^n : x_i \in [a_i, b_i),\ i=1,\ldots,n\}$, such that $P_j(C_i \Delta D_i^*) \leqslant \varepsilon/3k$, for $i,j=1,\ldots,k$ (see Halmos, 1950, pp. 21, 56). Now define inductively

$$D_1 \equiv D_1^*,$$

$$D_j \equiv D_j^* - \bigcup_{i=1}^{j-1} D_i^*, \qquad j=2,\ldots,k.$$

Then, by a standard argument (see Ash, 1965, p. 253) it can be shown that $P_j(D_j) \geqslant 1-2\varepsilon$, for $j=1,\ldots,k$.

Let $\{e_i, i \geqslant 1\}$ be a CONS (possibly a finite sequence) in H_N^T, which exists since H_N is assumed to be separable. Let π_n denote the projection of H_N^T onto span $\{e_1,\ldots,e_n\}$ and define $S_n^{(i)} \equiv \pi_n S^{(i)}$. We will show that for n sufficiently large $P\{(S_{n,t}^{(i)} + N_t)_0^T \in D_i\} \geqslant 1-3\varepsilon$, for $i=1,\ldots,k$. Let $\tau = \{t_1,\ldots,t_r\} \subset [0,T]$ be such that all the cylinder sets D_i, $i=1,\ldots,k$, are based on $\Re^\tau$. In other words there exist $A_i \subset \Re^\tau$, which are finite unions of semiclosed intervals, such that $D_i = A_i \times \Re^{[0,T]-\tau}$, $i=1,\ldots,k$. Since $H_{N^0}(N) < \infty$ we have $\mu_N \ll \mu_{N^0}$, where μ_N and μ_{N^0} are the measures induced on $\Re^{(-\infty,\infty)}$ by N and N^0, respectively. Therefore $\mu_N^\tau \ll \mu_{N^0}^\tau$ and we may write

$$\begin{aligned} P\{(S_{n,t}^{(i)} + N_t)_0^T \in D_i\} &= P\{(S_{n,t}^{(i)} + N_t)_{t\in\tau} \in A_i\} \\ &= \int_{\Re^\tau} \chi_{A_i}(x + S_n^{(i)}) \frac{d\mu_N^\tau}{d\mu_{N^0}^\tau}(x)\, d\mu_{N^0}^\tau(x). \end{aligned}$$

Note that $S_{n,t}^{(i)} \to S_t^{(i)}$ as $n \to \infty$, for each $t \in [0,T]$ and $i=1,\ldots,k$, since $S_n^{(i)} \to S^{(i)}$ in H_N^T as $n \to \infty$. If $\mu_{N^0}^t$ is degenerate at 0 for some $t \in [0,T]$ then $S_{n,t}^{(i)} = S_t^{(i)} = 0$, for all n, $i=1,\ldots,k$. Therefore, since the A_i are finite unions of intervals in $\mathfrak{R}^\tau$, it can be seen that $\chi_{A_i}(x + S_n^{(i)}) \to \chi_{A_i}(x + S^{(i)})$ as $n \to \infty$, a.e. $d\mu_{N^0}^\tau(x)$, for $i = 1,..., k$. Thus, by the dominated convergence theorem

$$P\{(S_{n,t}^{(i)} + N_t)_0^T \in D_i\} \to P\{(S_t^{(i)} + N_t)_0^T \in D_i\} \qquad \text{as} \quad n \to \infty,$$

for $i = 1,..., k$. But we have already shown that $P\{(S_t^{(i)} + N_t)_0^T \in D_i\} \geqslant 1 - 2\varepsilon$. Therefore, for n sufficiently large

$$P\{(S_{n,t}^{(i)} + N_t)_0^T \in D_i\} \geqslant 1 - 3\varepsilon, \qquad \text{for } i = 1,..., k.$$

Also

$$\| S_n^{(i)} \|_{H_N^T}^2 \leqslant n(P_0 T/n), \qquad i = 1,..., k,$$

and

$$\dim sp\{S_n^{(1)},..., S_n^{(k)}\} \leqslant n.$$

Therefore, by the converse for the general channel in Theorem 5.3 we have

$$\log k \leqslant \left[\frac{n}{2} \log \left(1 + \frac{P_0 T}{n}\right) + H_{\{N_t^0\}_0^T}(\{N_t\}_0^T) + \log 2\right] \Big/ (1 - 3\varepsilon). \qquad (*)$$

Note that $H_{\{N_T^0\}_0^T}(\{N_t\}_0^T) \leqslant H_{N^0}(N) < \infty$ and $(n/2) \log(1 + P_0 T/n) \leqslant P_0 T/2$ and it follows that $C_0 \leqslant P_0/2$. Finally, assuming $\dim(H_N) = M < \infty$, we have instead of $(*)$, $\log k \leqslant [(M/2) \log(1 + P_0 T/M) + K]/(1 - 3\varepsilon)$, where K is independent of T and it follows that $C_0 = 0$. ∎

Remark. If in Theorem 6.1(1) we have instead of $H_{N^0}(N) < \infty$,

$$\lim_{T \to \infty} \frac{H_{\{N_t^0\}_0^T}(\{N_t\}_0^T)}{T} = \bar{H}_{N^0}(N),$$

where $\bar{H}_{N^0}(N)$ is called the entropy rate of N with respect to N^0 (see Pinsker, 1960, p. 77), the most we can conclude is that $C_0 \leqslant P_0/2 + \bar{H}_{N^0}(N)$.

Acknowledgments

This work is a part of the author's doctoral disseration in Statistics at the University of North Carolina at Chapel Hill and he gratefully acknowledges the help and support of his advisor, professor Charles R. Baker. The author also thanks the referee for his comments.

RECEIVED: March 6, 1981; REVISED: April 30, 1982

REFERENCES

ASH, R. B. (1965), "Information Theory," Interscience, New York.

BADRIKIAN, A., AND CHEVET, S. (1974), "Measures Cylindriques, Espaces de Wiener et Fonctions Aléatoires Gaussiennes," Lecture Notes in Mathematics No. 379, Springer-Verlag, Berlin/New York.

BAKER, C. R. (1969), Complete simultaneous reduction of covariance operators, *SIAM J. Appl. Math.* **17**, 972–983.

BAKER, C. R. (1973), Zero-one laws for Gaussian measures on Banach space, *Trans. Amer. Math. Soc.* **186**, 291–308.

BAKER, C R. (1978a), Capacity of the Gaussian channel without feedback, *Inform. and Control* **37**, 70–89.

BAKER, C. R. (1978b), Absolute continuity for a class of probability measures, *in* "Vector Space Measures and Applications, I, Proceedings, Dublin, 1977," pp. 44–50, Lecture Notes in Mathematics No. 644, Springer-Verlag, New York/Berlin.

BAKER, C. R. (1979a), Mutual information, strong equivalence, and signal sample path properties for Gaussian processes, *Inform. and Control* **41**, 156–164.

BAKER, C. R. (1979b), Calculation of the Shannon information, *J. Math. Anal. Appl.* **69**, 115–123.

BETHOUX, P. (1962), Tests et estimations concernant certaines fonctions aléatoires, en particulier Laplaciennes, *Ann. Inst. H. Poincaré* **17**, 255–337.

BORELL, C. (1976), Gaussian Radon measures on locally convex vector spaces, *Math. Scand.* **38**, 265–284.

DOBRUSHIN, R. L. (1959), General formulation of Shannon's basic theorems of the theory of information, *Uspehi Mat. Nauk.* **14**, 3–104; *Amer. Math. Soc. Transl.* (*2*) **33** (1963), 323–438. [English Transl.]

DOOB, J. L. (1937), Stochastic processes depending on a continuous parameter, *Trans. Amer. Math. Soc.* **42**, 107–140.

FELDMAN, J. (1958), Equivalence and perpendicularity of Gaussian processes, *Pacific J. Math.* **8**, 699–708. Correction, **9** (1959), 1295–1296.

FERNIQUE, X. (1970), Integrabilite des vecteurs Gaussiens, *C. R. Acad. Sci. Paris Sér. A* **270**, 1698–1699.

FORTET, R. (1961), Hypothesis testing and estimation for Laplacian functions, *in* "Fourth Berkeley Symp. on Math. Stat. and Probability," Vol. 1, pp. 289–305, Univ. of California Press, Berkeley, California.

GUALTIEROTTI, A. F. (1979), Sur la détection des signaux, l'information de Shannon, et la capacité de retransmission, quand les lois du modèle ne sont pas exactement normales, *C. R. Acad. Sci. Paris Sér. A* **288**, 69–71.

GUALTIEROTTI, A. F. (1980), On average mutual information and capacity for a channel without feedback and contaminated Gaussian noise, *Inform. and Control* **46**, 46–70.

HALMOS, R. P. (1950), "Measure Theory," Van Nostrand, New York.

HITSIDA, M., AND IHARA, S. (1975), Gaussian channels and the optimal coding, *J. Multivariate Anal.* **5**, 106–118.

HUANG, R. Y., AND JOHNSON, R. A. (1962), Information capacity of time-continuous channels, *IEEE Trans. Inform. Theory* **IT–8**, S191–S198.

IHARA, S. (1978), On the capacity of channels with additive non-Gaussian noise, *Inform. and Control* **37**, 34–39.

KADOTA, T. T. (1970), Generalization of Feinstein's fundamental lemma, *IEEE Trans. Inform. Theory* **IT–16**, 791–792.

KADOTA, T. T., ZAKAI, M., AND ZIV, J. (1971), Mutual information of the white Gaussian channel with and without feedback, *IEEE Trans. Inform. Theory* **IT–17**, 368–371.

KUELBS, J. (1970), Gaussian measures on a Banach space, *J. Funct. Anal.* **5**, 354–367.
McKEAGUE, I. W. (1980), "Covariance Operators and Their Applications in Probability and Information Theory," Ph.D. dissertation, University of North Carolina at Chapel Hill.
PAN, Y.–M. (1966), Simple proofs of equivalence conditions for measures induced by Gaussian processes, *Shuxue Jinzhan* **9**, 85–90; *Selected Transl. Math. Statist. Prob.* **12** (1973), 109–118. [English Transl.]
PERLMAN, M. D. (1972), Characterizing measurability, distribution and weak convergence of random variables in Banach space by total subsets of linear functionals, *J. Multivariate Anal.* **2**, 174–188.
PINSKER, M. S. (1960), "Information and Information Stability of Random Variables and Processes," Izd. Akad. Nauk. SSSR, Moscow; Holden–Day, San Francisco, 1964. [English Transl.]
PROHOROV, YU. V. (1956), Convergence of random processes and limit theorems in probability theory, *Theory Probab. Appl.* **1**, 157–124.
RAO, C. R., AND VARADARAJAN, V. S. (1963), Discrimination of Gaussian processes, *Sankhyā Ser. A* **25**, 303–330.
ROBERTSON, A. P., AND ROBERTSON, W. (1964), "Topological Vector Spaces," Cambridge Univ. Press, Cambridge.
SCHAEFER, H. H. (1966), "Topological Vector Spaces," Macmillan, New York.
SCHWARTZ, L. (1964), Sous-espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés, *J. Analyse Math.* **13**, 115–256.
SCHWARTZ, L. (1973), "Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures," Oxford Univ. Press, Bombay.
SHANNON, C. E. (1948), A mathematical theory of communication, *Bell System Tech. J.* **27**, 379–423, 623–656.
SHANNON, C. E. (1959), Probability of error for optimal codes in a Gaussian channel, *Bell System Tech. J.* **38**, 6–25.
VAKHANIA, N. N. (1968), A covariance operator of a probability distribution in a Banach space, *Inf. Georgian SSR Academy of Sciences* **51**, 35–40. [In Russian]
VAKHANIA, N. N., AND TARIELADZE, V. I. (1978), Covariance operators of probability measures in locally convex spaces, *Theory Probab. Appl.* **23**, 1–21.
WYNER, A. D. (1966), The capacity of the band-limited Gaussian channel, *Bell System Tech. J.* **45**, 359–395.