JOURNAL OF MULTIVARIATE ANALYSIS 5, 106–118 (1975)

Gaussian Channels and the Optimal Coding

MASUYUKI HITSUDA

Department of Mathematics, Nagoya Institute of Technology, Nagoya, Japan,

AND

SHUNSUKE IHARA

Department of Mathematics, Ehime University, Matsuyama, Japan

Communicated by T. Hida

For the Gaussian channel $Y(t) = \Phi(\xi(s), Y(s); s \leqq t) + X(t)$, the mutual information $I(\xi, Y)$ between the message $\xi(\cdot)$ and the output $Y(\cdot)$ is evaluated, where $X(\cdot)$ is a Gaussian noise. Furthermore, the optimal coding under average power constraints is constructed.

INTRODUCTION

We treat, in this paper, the Gaussian channel

$$Y(t) = \Phi(t) + X(t), \qquad 0 \leqq t \leqq T,$$

where the channel input $\Phi(t)$ is interfered with by the Gaussian noise $X(\cdot)$. The channel in this paper is with feedback, so that the channel input $\Phi(t)$ is to be a causal functional of the message $\xi(\cdot)$ and the channel output $Y(\cdot)$.

In the first section, we determine the mutual information $I_t(\xi, Y)$ of $\xi(\cdot)$ and $Y(\cdot)$, using the decomposition of the Gaussian process $X(\cdot)$, based upon the canonical representation. The formula of $I_t(\xi, Y)$ is given by a causal functional of $Y(\cdot)$ in the case where the spectral measure of $X(\cdot)$ is continuous. Such causal expressions in spacial cases were given in Kadota–Zakai–Ziv [5] and Hitsuda [3]. These results are covered in Theorem 1.

Received August 3, 1974.

AMS 1970 classification: Primary 94A15; Secondary 60G15, 60G25, 60G35, 94A10.

Key words and phrases: Gaussian channel, mutual information, canonical representation of Gaussian processes, reproducing kernel Hilbert space, optimal coding.

In the second section, we construct an optimal coding $\Phi(\cdot, \omega)$ under the average power constraint, when $\xi(\cdot)$ is a Gaussian message. Ihara [4] constructed such an optimal coding in the special case of a white Gaussian channel. The crucial difference between the present paper and the former [4] is that the noise is taken to be a general Gaussian process and that we must consider the mutual interaction among many decomposed subchannels arising form the canonical representation. Furthermore, our method enables us to evaluate the value of the channel capacity (Cor. 2).

1. Gaussian Channels with Feedback

In this section, we introduce the basic notations and evaluate the amount of mutual information between the message and the channel output.

Let the message be a stochastic process $\xi = \xi(t)$, $0 \leqq t \leqq T$, and let the noise be a separable Gaussian process $X(t)$, $0 \leqq t \leqq T$, independent of $\xi(\cdot)$. The channel to be considered is defined by

$$Y(t) = \Phi(t) + X(t), \qquad (0 \leqq t \leqq T < \infty). \tag{1.1}$$

Let us assume that the output $Y(\cdot)$ and the modulator output $\Phi(t) = \Phi(t, \omega)$ satisfies the following.

Assumption 1. $Y(t)$ is $\mathfrak{X}(t) \vee \mathfrak{\xi}(t)$-measurable for each t, where $\mathfrak{X}(t) = \sigma\{X(s); s \leqq t\}$(the σ-algebra generated by $\{X(s);\ s \leq t\}$) and $\mathfrak{\xi}(t) = \sigma\{\xi(s); s \leqq t\}$.

Assumption 2. $\Phi(t)$ is $\mathfrak{Y}(t) \vee \mathfrak{\xi}(t)$-measurable for each t, where $\mathfrak{Y}(t) = \{Y(s); s \leqq t\}$.

Assumption 3. Almost all paths of $\Phi(\cdot, \omega)$ belong to the reproducing kernel Hilbert space (*RKHS*) $\mathscr{H}(X) = \mathscr{H}_T(X)$ corresponding to the Gaussian process $X(\cdot)$.

Based upon Assumption 2, we can write formally $\Phi(t) = \Phi(t, Y_0^t, \xi_0^t)$, where Y_0^t and ξ_0^t are the paths of the respective processes up to t.

We put another assumption on the canonical representation of the noise $X(t)$.

Assumption 4. The Gaussian process $X(t)$ has the canonical representation (in the sense of Hida–Cramér, cf. [2]),

$$X(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, dB_i(u), \qquad (N \leqq \infty), \tag{1.2}$$

where $F_i(t, u)$'s are Volterra kernels and $dB_i(u)$'s are mutually independent white noises with continuous measures

$$m_i(du) = E\,|\,dB_i(u)|^2, \tag{1.3}$$

such that $m_i(du) \succ m_{i+1}(du)$.

Assumptions 3 and 4 imply that the σ-algebra $\mathfrak{X}(t)$ is equal to the σ-algebra $\mathfrak{B}(t)$ generated by $\{B_i(s); i = 1,..., N, s \leqq t\}$, and that

$$\mathscr{H}(X) = \left\{ f(\cdot); f(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, a_i(u)\, m_i(du), \sum_{i=1}^{N} \int_0^T a_i(u)^2\, m_i(du) < \infty \right\}. \tag{1.4}$$

By the representation (1.4) of the *RKHS*, we obtain the following.

PROPOSITION 1. *Under Assumptions* 2, 3, *and* 4, *the modulator output* $\Phi(t)$ *in* (1.1) *has a representation in the form*

$$\begin{gathered}\Phi(t, \omega) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, \varphi_i(u, \omega)\, m_i(du), \\ \sum_{i=1}^{N} \int_0^T \varphi_i^2(u, \omega)\, m_i(du) < \infty \qquad (\text{a.e. } \omega),\end{gathered} \tag{1.5}$$

where $\varphi_i(u, \omega)$ *is* $\mathfrak{Y}(u) \vee \mathfrak{\xi}(u)$*-measurable for each* u.

Proof. The representation (1.5) follows from (1.4), if we can prove the $\mathfrak{Y}(u) \vee \mathfrak{\xi}(u)$-measurability of $\varphi_i(u, \omega)$. Let us consider the isomorphism

$$\mathscr{H}_t(X) \ni \Phi(\cdot, \omega) \leftrightarrow (\varphi_1(\cdot, \omega),..., \varphi_N(\cdot, \omega)) \in \prod_{i=1}^{N} L^2[m_i; [0, t]] \tag{1.6}$$

for any t, where $\mathscr{H}_t(X)$ is the *RKHS* of the subprocess $\{X(s); s \leqq t\}$. Since $\Phi(s, \omega)(s \leqq t)$ is $\mathfrak{Y}(t) \vee \mathfrak{\xi}(t)$-measurable, $\chi_i(s, \omega) = \int_0^s \varphi_i(u, \omega)\, m_i(du)(s \leqq t)$ is also $\mathfrak{Y}(t) \vee \mathfrak{\xi}(t)$-measurable for each i. In particular, the $\chi_i(s, \omega)$'s are $\mathfrak{Y}(s) \vee \mathfrak{\xi}(s)$-measurable. Differentiating $\chi_i(s, \omega)$, we can get the $\mathfrak{Y}(s) \vee \mathfrak{\xi}(s)$-measurability of $\varphi_i(u, \omega)$ for $u \leqq s$, by choosing a suitable version.

PROPOSITION 2. (i) *Let* $\mathbf{B}(t)$ *be the vector valued Gaussian martingale* $(B_1(t),..., B_N(t))$, *and let* $\mathbf{Z}(t) = (Z_1(t),..., Z_N(t))$ *be the vector valued process*

$$\left(B_1(t) + \int_0^t \varphi_1(u, \omega)\, m_1(du),..., B_N(t) + \int_0^t \varphi_N(u, \omega)\, m_N(du) \right). \tag{1.7}$$

Then the measure μ_Z induced by $\mathbf{Z}(\cdot)$ is absolutely continuous with respect to the measure μ_B induced by $\mathbf{B}(\cdot)$. In this case, if $E[\sum_{i=1}^{N} \int_0^t \varphi_i^2(u, \omega)\, m_i(du)] < \infty$, the density $d\mu_Z/d\mu_B$ is given by

$$\frac{d\mu_Z}{d\mu_B}(Z(\cdot)) = \exp\left\{ \sum_{i=1}^{N} \left(\int_0^T \hat{\varphi}_i(u)\, dZ_i(u) - \frac{1}{2} \int_0^T \hat{\varphi}_i^2(u)\, m_i(du) \right) \right\}, \quad (1.8)$$

where $\hat{\varphi}_i(u) = E[\varphi_i(u) \mid \mathfrak{Z}(u)]$, $\mathfrak{Z}(u) = \sigma\{\mathbf{Z}(v); v \leqq u\}$.

(ii) *The measure μ_Y induced by $Y(\cdot)$ is absolutely continuous with respect to the measure μ_X induced by $X(\cdot)$.*

Proof. In the case of $N = 1$, (i) is proved in Kailath [6] and in Liptzer–Shiryaev [7, 8]. Even in the case of $N \leqq \infty$, the analogous argument to that in [6] is available, because

$$\sum_{i=1}^{N} \int_0^T \varphi_i^2(u)\, m_i(du) < \infty \qquad \text{(a.e. } \omega).$$

In order to prove (ii), we note that

$$Y(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, dZ_i(u). \quad (1.9)$$

By (i), there exists a density $M(\omega)$, such that $\mathbf{B}(\cdot)$ under the measure $\tilde{P}(d\omega) = M(\omega)\, P(d\omega)$ has the same distribution as $\mathbf{Z}(\cdot)$. Hence

$$X(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, dB_i(u)$$

has the same distribution as $Y(\cdot)$ under the measure $\tilde{P}(d\omega)$. Therefore, μ_Z is absolutely continuous with respect to μ_X.

PROPOSITION 3. (i) *The σ-algebras $\mathfrak{Y}(t)$ and $\mathfrak{Z}(t)$ are equivalent.*

(ii) *The mutual information $I_t(\xi, Y)$ between the message $\{\xi(s); s \leqq t\}$ and the channel output $\{Y(s); s \leqq t\}$ is equal to the mutual information $I_t(\xi, \mathbf{Z})$ between $\{\xi(s); s \leqq t\}$ and $\{\mathbf{Z}(s); s \leqq t\}$.*

Proof. (i) Since $\mathfrak{X}(t)$ and $\mathfrak{B}(t)$ are equivalent, the results follow immediately by the measure transformation stated in Proposition 2.

(ii) This statement is trivially derived from (i).

Based upon these Propositions, we can evaluate the mutual information as in the following theorem.

THEOREM 1. *Under Assumptions* 1–4, *if*

$$E\left(\sum_{i=1}^{N} \int_0^T \varphi_i^2(u, \omega)\, m_i(du) \right) < \infty, \tag{1.10}$$

then the mutual information $I_t(\xi, Y)$ *is given by the formula*

$$I_t(\xi, Y) = I_t(\xi, \mathbf{Z}) = \frac{1}{2} \sum_{i=1}^{N} \int_0^t E \mid \varphi_i(u, \omega) - \hat{\varphi}_i(u, \omega)|^2\, m_i(du), \qquad 0 \leqq t \leqq T, \tag{1.11}$$

where $\hat{\varphi}_i(u, \omega) = E[\varphi_i(u, \omega) | \mathfrak{Y}(u)]$.

Proof. By Gelfand–Yaglom [1], we know that

$$I(Y, \xi) = I(\mathbf{Z}, \xi) = E\{\log(d\mu_{(Z,\xi)}/d(\mu_Z \times \mu_\xi))(Z, \xi)\}. \tag{1.12}$$

In order to show that $\mu_{(Z,\xi)}$ is absolutely continuous with respect to $\mu_Z \times \mu_\xi$, and to evaluate the right hand side of (1.12), we use the densities of the measures μ_Z and $\mu_{Z|\xi}$ related to the Wiener measure μ_B, where $\mu_{Z|\xi}$ means the induced measure of $\mathbf{Z}(\cdot)$ for a fixed message $\xi = \xi(\cdot)$. Note that

$$d\mu_{Z|\xi}/d\mu_B(Z) = d\mu_{(Z,\xi)}/d\mu_{(B,\xi)}(Z, \xi) \quad \text{(a.e.)}, \tag{1.13}$$

and

$$d\mu_Z/d\mu_B(Z) = d(\mu_Z \times \mu_\xi)/d\mu_{(B,\xi)}(Z, \xi) \quad \text{(a.e.)}, \tag{1.14}$$

because $\mu_{(B,\xi)} = \mu_B \times \mu_\xi$. Let us consider the ratio of (1.13) and (1.14). Then we obtain

$$\frac{d\mu_{Z|\xi}/d\mu_B(Z)}{d\mu_Z/d\mu_B(Z)} = \frac{d\mu_{(Z,\xi)}}{d(\mu_Z \times \mu_\xi)}(Z, \xi) \qquad \text{(a.e.)}$$

(cf. [8, p. 336]). It is well known that for each $Z(\cdot)$, the numerator $d\mu_{Z|\xi}/d\mu_B(Z)$ is given by

$$\exp\left\{ \sum_{i=1}^{N} \left(\int_0^T \varphi_i(u)\, dZ_i(u) - \frac{1}{2} \int_0^T \varphi_i^2(u)\, m_i(du) \right) \right\}$$

(cf. [3] in Gaussian case) and the denominator is already given by Proposition 2. So the right-hand side of (1.12) is calculated as follows.

$$E\left\{ \sum_{i=1}^{N} \left(\int_0^T \varphi_i(u)\, dZ_i(u) - \frac{1}{2} \int_0^T \varphi_i^2(u)\, m_i(du) \right) \right.$$
$$\left. - \sum_{i=1}^{N} \left(\int_0^T \hat{\varphi}_i(u)\, dZ_i(u) - \frac{1}{2} \int_0^T \hat{\varphi}_i^2(u)\, m_i(du) \right) \right\}$$

$$= E\left\{ \sum_{i=1}^{N} \left(\int_0^T (\varphi_i(u) - \hat{\varphi}_i(u))\, dB_i(u) + \frac{1}{2} \int_0^T (\varphi_i^2(u) - \hat{\varphi}_i^2(u))\, m_i(du) \right) \right\}$$

$$= \frac{1}{2} \sum_{i=1}^{N} \int_0^T E(\varphi_i(u) - \hat{\varphi}_i^2(u))\, m_i(du). \qquad \text{(by (1.7))}$$

The last equality is justified by the fact that

$$E\left(\sum_{i=1}^{N} \int_0^T \varphi_i(u)\, dB_i(u) \right) = E\left(\sum_{i=1}^{N} \int_0^T \hat{\varphi}_i(u)\, dB_i(u) \right) = 0,$$

which follows from the assumption (1.10). Thus, Theorem 1 is proved.

2. Capacity of the Channel

In this section, we impose an average power constraint on the channel input Φ for Gaussian channels which have been formulated by (1.1) with Assumptions 1–4.

Let $P_i(t) \geqq 0$ $(i = 1,..., N)$ be a monotone nondecreasing function. The constraint is formulated by

$$E\left\{ \int_0^t \varphi_i^2(u)\, m_i(du) \right\} \leqq P_i(t), \qquad 0 \leqq t \leqq T, \quad i = 1,..., N. \tag{2.1}$$

If we denote the norm in the *RKHS* $\mathscr{H}_t(X)$ by $\|\cdot\|_t$, we then have

$$E[\| \Phi \|_t^2] \leqq \sum_{i=1}^{N} P_i(t) \equiv P_0(t). \tag{2.2}$$

It is easy to obtain the following proposition by the use of Theorem 1.

Proposition 4. *If the input signal Φ satisfies the constraint* (2.1), *the mutual information $I_t(\xi, Y)$ is evaluated by*

$$I_t(\xi, Y) \leqq (1/2)\, P_0(t). \tag{2.3}$$

It follows from this proposition that if we could find the coding

$$\Phi(t) \equiv \Phi(t, Y_0^t, \xi_0^t)$$

by which the equality holds in (2.3) then the coding would give the optimal one in the sense of information transmission. In what follows it is shown that the optimal coding does exist and indeed it is found in the class of Gaussian processes. From now on, the construction of such an optimal coding will be discussed in the case of Gaussian message $\xi(\cdot)$.

Let the message $\xi = \{\xi(t);\ 0 \leqq t \leqq T\}$ be a mean-zero Gaussian process satisfying the two conditions (a) and (b):

(a) $\xi(\cdot)$ belongs to $\mathscr{H}(X)$ for almost all ω, i.e.,

$$\xi(t, \omega) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, \xi_i(u, \omega)\, m_i(du) \qquad \text{a.e.} \quad \omega, \tag{2.4}$$

where $\sum_{i=1}^{N} \int_0^T \xi_i^2(u)\, m_i(du) < \infty$.

(b) $E[\xi_i^2(t)] \neq 0, \qquad 0 \leqq t \leqq T, \qquad i = 1,..., N.$

Now we get the following theorem.

THEOREM 2. *Let the Gaussian message* $\xi = \{\xi(t); 0 \leqq t \leqq T\}$ *satisfy the conditions* (a) *and* (b). *If the constraint functions* $P_i(t)$ *are absolutely continuous with respect to the spectral measures* $m_i(du) = E\mid dB_i(u)|^2 (i = 1,..., N)$, *then there exists a unique output* $Y(\cdot)$ *and positive functions* $A_i(t) (i = 1,..., N)$ *satisfying*

$$Y(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, A_i(u)(\xi_i(u) - \hat{\xi}_i(u))\, m_i(du) + X(t) \tag{2.5}$$

and

$$A_i^2(t)\, E \mid \xi_i(t) - \hat{\xi}_i(t)|^2 = \rho_i(t) \qquad (i = 1,..., N), \tag{2.6}$$

where the $\rho_i(t)$ *are the densities of the* $P_i(t)$ *with respect to the* $m_i(dt)$ *and where* $\hat{\xi}_i(t) = E[\xi_i(t) \mid \mathfrak{Y}(t)]$.

Remark. The channel given by (2.5) satisfies the Assumptions 1–4, and we obtain

$$I_t(\xi, Y) = \frac{1}{2} \sum_{i=1}^{N} \int_0^t \rho_i(u)\, m_i(du) = \frac{1}{2} P_0(t), \tag{2.7}$$

by the use of Theorem 1. Therefore, the coding is the optimal one. It is interesting that the optimal coding is attained by a linear functional of $\xi(\cdot)$ and $X(\cdot)$.

For the proof of Theorem 2, we prepare four lemmas.

LEMMA 2.1. *If there exist a unique vector valued process* $\mathbf{Z}(\cdot) = (Z_1(\cdot),..., Z_N(\cdot))$ *and the functions* $A_i(t) (i = 1,..., N)$ *satisfy*

$$Z_i(t) = \int_0^t A_i(u)(\xi_i(u) - \hat{\xi}_i(u))\, m_i(du) + B_i(t) \qquad (i = 1,..., N), \tag{2.8}$$

and

$$A_i^2(t)\, E[\xi_i(t) - \hat{\xi}_i(t)]^2 = \rho_i(t), \qquad \hat{\xi}_i(u) = E[\xi_i(u) \mid \mathfrak{Z}(u)], \tag{2.9}$$

then the statement of Theorem 2 *holds.*

Proof. Define

$$\begin{aligned} Y(t) &= \sum_{i=1}^{N} \int_0^t F_i(t, u)\, dZ_i(u) \\ &= \sum_{i=1}^{N} \int_0^t F_i(t, u)\, A_i(u)(\xi_i(u) - \hat{\xi}_i(u))\, m_i(du) + X(t). \end{aligned} \tag{2.10}$$

Then the process $Y(t)$ and the functions $A_i(t)$ are desired ones, since $\mathfrak{Z}(t) = \mathfrak{Y}(t)$ for each t and we can apply Theorem 1.

The following three lemmas, except the inequalities (2.15) and (2.16), are proved in the same way as in Ihara [4], where the corresponding lemmas are proved in the case of $N = 1$.

LEMMA 2.2. *Let* $\xi(\cdot)$ *satisfy the condition* (a), *and let* $\mathbf{f}(s, t) = (f_{ij}(s, t))$, *where* $f_{ij}(s, t)$ *is a Volterra kernel satisfying*

$$\sum_{i,j} \int_0^T \int_0^s f_{ij}^2(s, t)\, m_i(ds)\, m_j(dt) < \infty.$$

Then the stochastic equation

$$\begin{aligned} Z_i(t) &= \int_0^t (\xi_i(u) - f_i(u))\, m_i(du) + B_i(t) \qquad (i = 1,\dots, N), \\ f_i(u) &= \int_0^u \sum_{j=1}^{N} f_{ij}(u, s)\, dZ_j(s), \end{aligned} \tag{2.11}$$

has a unique solution $\mathbf{Z}(\cdot) = \mathbf{Z}_{\mathbf{f}}(\cdot)$ *and the* σ*-algebras* $\mathfrak{Z}(t)$, $t \in [0, T]$, *are invariant under* $\mathbf{f}$ *viewed as a Volterra operator. Moreover, the mutual information* $I_t(\xi, \mathbf{Z})$ *is also invariant under* $\mathbf{f}$.

LEMMA 2.3. (i) *Let*

$$A_i^{(1)}(t) \geqslant A_i^{(2)}(t), \qquad 0 \leqslant t \leqslant T, \quad i = 1,\dots, N.$$

And let $\mathbf{Z}^{(j)}(t) = (Z_1^{(j)}(t),\dots, Z_N^{(j)}(t))$, *where*

$$Z_i^{(j)}(t) = \int_0^t A_i^{(j)}(u)\, \xi_i(u)\, m_i(du) + B_i(t) \qquad (j = 1, 2). \tag{2.12}$$

Then the inequalities

$$E \mid \xi_i(t) - \hat{\xi}_i^{(1)}(t)|^2 \leqslant E \mid \xi_i(t) - \hat{\xi}_i^{(2)}(t)|^2, \tag{2.13}$$

and

$$I_t(\xi, \mathbf{Z}^{(1)}) \geqq I_t(\xi, \mathbf{Z}^{(2)}) \tag{2.14}$$

hold, where $\hat{\xi}_i^{(j)}(t) = E[\xi_i(t) | \mathfrak{Z}^{(j)}(t)]$ *and* $\mathfrak{Z}^{(j)}(t) = \sigma\{\mathbf{Z}^{(j)}(s); s \leqq t\}$.

(ii) *In addition, let us assume that there exists a measurable set* Γ *such that* $m_i(\Gamma) > 0$ *for some* i *and that*

$$A_i^{(1)}(t) > A_i^{(2)}(t), \qquad t \in \Gamma.$$

Then the following inequality holds,

$$I_T(\xi, \mathbf{Z}^{(1)}) > I_T(\xi, \mathbf{Z}^{(2)}). \tag{2.15}$$

Since $B_i(\cdot)$, $i = 1,..., N$, are mutually independent processes with independent increments, we can prove the inequalities (2.14) and (2.15), using some basic properties on the mutual information.

LEMMA 2.4. *If for each* t

$$A_i^2(t)\, E \mid \xi_i(t) - \hat{\xi}_i(t)|^2 \leqq \rho_i(t), \qquad i = 1,..., N, \tag{2.16}$$

then inequalities

$$E \mid \xi_i(t) - \hat{\xi}_i(t)|^2 \geqslant E[\xi_i^2(t)]\, e^{-P_0(t)}, \tag{2.17}$$

and

$$A_i^2(t)\, E[\xi_i^2(t)] \leqslant P_i(t)\, e^{P_0(t)}, \qquad i = 1,..., N, \tag{2.18}$$

hold.

Remark. If we give a system of positive functions $A_i(t)$, $i = 1,..., N$, as a multiplier $Y(\cdot)$ can be determined by (2.5). So the conditional expectation $\hat{\xi}_i(t)$ has a definite meaning.

Now, we turn back to the proof of Theorem 2.

Proof of Theorem 2. At first, we wish to construct an approximating sequence $A_{k,i}(t)$ to $A_i(t)$. Define $A_{k,i}(\cdot)$, $Z_i^{(k)}(\cdot)$, $\hat{\xi}_{k,i}(\cdot)$ and $\sigma_{k,i}^2(t)$ by induction:

$$\sigma_{0,i}^2(t) = E \mid \xi_i(t)|^2,$$

$$A_{k,i}^2(t)\, \sigma_{k-1,i}^2(t) = P_i(t),$$

$$Z_i^{(k)}(t) = \int_0^t A_{k,i}(u)\, \xi_i(u)\, m_i(du) + B_i(t),$$

$$\hat{\xi}_{k,i}(t) = E[\xi_i(t) \mid \mathfrak{Z}^{(k)}(t)],$$

and

$$\sigma_{k,i}^2(t) = E \mid \xi_i(t) - \hat{\xi}_{k,i}(t)|^2,$$

where $\mathfrak{Z}^{(k)}(t) = \sigma\{(Z_1^{(k)}(s),..., Z_N^{(k)}(s));\ s \leqq t\}$. Then, by the use of (2.13) in Lemma 2.3, we get the monotonicity of $A_{k,i}(t)$ and $\sigma_{k,i}^2(t)$:

$$A_{k,i}(t) \leqslant A_{k+1,i}(t) \qquad \text{and} \qquad \sigma_{k,i}^2(t) \geqslant \sigma_{k+1,i}^2(t).$$

Thus the desired nonnegative function $A_i(t)$ $(i = 1,..., N)$ can be defined by

$$A_i(t) = \lim_{k\to\infty} A_{k,i}(t).$$

By Lemma 2.4, $A_{k,i}^2(t)\, E[\xi_i^2(t)]$ is bounded uniformly in k, hence $A_i^2(t)\, E[\xi_i^2(t)]$ is also bounded. So we can define

$$Z_i^{(0)}(t) = \int_0^t A_i(u)\, \xi_i(u)\, m_i(du) + B_i(t),$$

$$\hat{\xi}_i^{(0)}(t) = E[\xi_i(t) \mid \mathfrak{Z}^{(0)}(t)],$$

$$\sigma_i^2(t) = E \mid \xi_i(t) - \hat{\xi}_i^{(0)}(t)|^2,$$

where $\mathfrak{Z}^{(0)}(t) = \sigma\{(Z_1^{(0)}(s),..., Z_N^{(0)}(s));\ s \leqq t\}$. Then,

$$\sigma_i^2(t) = \lim_{k\to\infty} \sigma_{k,i}^2(t)$$

is easily derived. Therefore,

$$A_i^2(t)\, \sigma_i^2(t) = \lim_{k\to\infty} A_{k,i}^2(t)\, \sigma_{k,i}^2(t) = P_i(t)$$

hold. Let us put

$$Z_i(t) = \int_0^t A_i(u)(\xi_i(u) - \hat{\xi}_i(u))\, m_i(du) + B_i(t).$$

By Lemma 2.2, $\mathfrak{Z}(t) = \sigma\{\mathbf{Z}(s) = (Z_1(s),..., Z_N(s));\ s \leqq t\}$ is equal to $\mathfrak{Z}^{(0)}(t)$ for any t, so we get

$$\hat{\xi}_i(t) = E[\xi_i(t) \mid \mathfrak{Z}(t)] = E[\xi_i(t) \mid \mathfrak{Z}^{(0)}(t)] = \xi_i^{(0)}(t).$$

Therefore, we can derive (2.9).

Now we prove the uniqueness. We assume that there exist two pairs $(A_i^{(j)}(t),$ $i = 1,..., N; \mathbf{Z}^{(j)}(\cdot))$ $(j = 1, 2)$ of the solution of (2.8) and (2.9). Define the functions $A_i^*(t)$, $i = 1,\dots, N$, and the vector valued process $\mathbf{Z}^*(\cdot) = (Z_1^*(\cdot),\dots, Z_N^*(\cdot))$ by

$$A_i^*(t) = \max(A_i^{(1)}(t), A_i^{(2)}(t)),$$

and

$$Z_i^*(t) = \int_0^t A_i^*(u)\, \xi_i(u)\, m_i(du) + B_i(t).$$

Then, from Lemma 2.2 and 2.3, we have

$$E \mid \xi_i(t) - \hat{\xi}_i^*(t)|^2 \leqslant E \mid \xi_i(t) - \hat{\xi}_i^{(j)}(t)|^2, \qquad j = 1, 2, \tag{2.19}$$

where $\hat{\xi}_i^*(t) = E[\xi_i(t) | \mathfrak{Z}^*(t)]$ and $\hat{\xi}_i^{(j)}(t) = E[\xi_i(t) | \mathfrak{Z}^{(j)}(t)]$. Using (2.19) and the formula (1.11), we get

$$\begin{aligned} I_t(\xi, \mathbf{Z}^*) &= \frac{1}{2} \sum_{i=1}^N \int_0^t (A_i^*(u))^2\, E \mid \xi_i(u) - \hat{\xi}_i^*(u)|^2\, m_i(du) \\ &\leqq \frac{1}{2} \sum_{i=1}^N \int_0^t (A_i^{(j)}(u))^2\, E \mid \xi_i(u) - \hat{\xi}_i^{(j)}(u)|^2\, m_i(du) \\ &= I_t(\xi, \mathbf{Z}^{(j)}), \qquad 0 \leqq t \leqq T, \quad (j = 1, 2). \end{aligned}$$

On the other hand, if we assume that $A_i^{(1)}(t) \neq A_i^{(2)}(t)$, for $t \in \Gamma$, where Γ is a measurable set such that $m_i(\Gamma) > 0$, then from Lemma 2.2 and 2.3 we have

$$I_T(\xi, \mathbf{Z}^*) > I_T(\xi, \mathbf{Z}^{(j)}). \tag{2.21}$$

The inequality (2.21) contradicts to the inequality (2.20).

Thus the proof is completed.

Now we give two corollaries which are easily seen from Theorem 2.

COROLLARY 1. *Let $\xi(\cdot)$ satisfy the conditions* (a) *and* (b). *Then there exists an optimal coding $\Phi(t, \omega)$ satisfying*

$$I_t(\xi, Y) = 1/2\, P_0(t), \qquad t \geqq 0, \tag{2.22}$$

if and only if there exist nonnegative densities $\rho_i(t)$, $1 \leq i \leq n$, and $P_0(t)$ is represented as

$$P_0(t) = \sum_{i=1}^N P_i(t), \qquad P_i(t) = \int_0^t \rho_i(u)\, m_i(du).$$

Proof. The "if" part is none other than the statement of Theorem 2. The "only if" part is a direct result from the equality for the optimal coding $\Phi(\cdot, \omega)$,

$$I_t(\xi, Y) = \frac{1}{2} \sum_{i=1}^{N} \int_0^t E \mid \varphi_i(u) - \hat{\varphi}_i(u)|^2 \, m_i(du) = \frac{1}{2} P_0(t),$$

(cf. Theor. 1).

The following Corollary 2 discusses the case in which some of $P_i(t)$'s are not absolutely continuous with respect to $m_i(du)$. Let $C_t(0 \leqq t \leqq T)$ be the capacity of the channel (1.1) satisfying Assumptions 1–4 and (2.1):

$$C_t = \sup_{(\xi,\Phi) \in \mathcal{Ol}} I_t(\xi, Y),$$

$\mathcal{Ol} = \{(\xi, \Phi); X(\cdot), \Phi(\cdot)$ and $Y(\cdot) = \Phi(\cdot) + X(\cdot)$ satisfy Assumptions 1–4 and (2.1)$\}$.

COROLLARY 2. *The channel capacity C_t is evaluated by*

$$C_t = 1/2 \, \tilde{P}_0(t), \qquad 0 \leqq t \leqq T,$$

where

$$\tilde{P}_0(t) = \sum_{i=1}^{N} \tilde{P}_i(t), \qquad \tilde{P}_i(t) = \sup_{Q_i \in \mathcal{Q}_i} Q_i(t)$$

and

$$\mathcal{Q}_i = \left\{ Q_i(\cdot); Q_i(s) = \int_0^s \rho(u) \, m_i(du) \text{ and } Q_i(s) \leqq P_i(s) \text{ for any } s \right\}.$$

Proof of Corollary 2 can be easily shown from the proof of Theorem 2.

ACKNOWLEDGMENT

The authors thank Professor Dawson for his advice.

REFERENCES

[1] GELFAND, I. M. AND YAGLOM, A. M. (1957). Calculation of the amount of information about a random function contained in another such function. *Uspekhi Math. Nauk* **12** 3–52 (in Russian); translation in *Amer. Math. Soc. Transl. Ser. 2* **12** 199–247.

[2] HIDA, T. (1960). Canonical representation of Gaussian processes and their applications. *Mem. Coll. Sci. Univ. Kyoto Ser. A* **33** 109–155.

[3] HITSUDA, M. (1974). Mutual information in Gaussian channels. *J. Mult. Anal.* **4** 66–73.

[4] IHARA, S. (1974). Coding theory in white Gaussian channel with feedback. *J. Mult. Anal.* **4** 74–87.

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

[6] KAILATH, T. (1971). The structure of Radon–Nikodym derivatives with respect to Wiener and related measures. *Ann. Math. Stat.* **42** 1054–1067.

[7] LIPTZER, R. SH. AND SHIRYAEV, A. N. (1972). Absolutely continuity with respect to Wiener measure of measures corresponding to process of diffusion type. *Izv. Akad. Nauk SSSR Ser. Math.* **36** 847–889 (in Russian); translation in *Math. USSR-Izv.* **6** 839–882.

[8] LIPTZER, R. SH. AND SHIRYAEV, A. N. (1974). *Statistics of Stochastic Processes* (in Russian). Nauka, Moscow, U.S.S.R.