JOURNAL OF MULTIVARIATE ANALYSIS 10, 319–331 (1980)

On the Capacity of the Continuous Time Gaussian Channel with Feedback

SHUNSUKE IHARA

Department of Mathematics, Ehime University, Matsuyama, 790, Japan

Communicated by T. Hida

We discuss the capacity of the Gaussian channel with feedback. In general it is not easy to give an explicit formula for the capacity of a Gaussian channel, unless the channel is without feedback or a white Gaussian channel. We consider the case where a constraint, given in terms of the covariance functions of the input processes, is imposed on the input processes. It is shown that the capacity of the Gaussian channel can be achieved by transmitting a Gaussian message and using additive linear feedback.

1. INTRODUCTION

The following model for a Gaussian channel with feedback is considered:

$$Y(t) = \Phi(t) + X(t), \qquad 0 \leqslant t \leqslant T \ (<\infty), \tag{1}$$

where $X(\cdot)$ is a Gaussian process expressing a noise, and $\Phi(t)$ and $Y(t)$ are an input signal and the output signal, respectively, at time t. $\Phi(t)$ is a causal function of a message Θ to be transmitted and the output process $Y(\cdot)$. In general, some conditions are imposed on the messages and the input processes. We denote by $\mathbb{A}$ the class of all pairs (Θ, Φ) satisfying the given conditions. Denote by $I(\Theta, Y)$ the mutual information between a message Θ and the output process $Y = \{Y(t); 0 \leqslant t \leqslant T\}$. Then the capacity $C(\mathbb{A})$ of the Gaussian channel under the constraint specified by $\mathbb{A}$ is defined by

$$C(\mathbb{A}) = \sup\{I(\Theta, Y); (\Theta, \Phi) \in \mathbb{A}\}.$$

In this paper, we consider the case where the constraint $\mathbb{A}$ is prescribed by the covariance functions of the input processes. A typical example is the so-called average power constraint. The main purpose is to show that the

Received September 1, 1978; revised April 23, 1979.

AMS subject classification: Primary 94A15; Secondary 60G15, 60G35, 94A40.

Key words and phrases: Gaussian channel with feedback, capacity of channel, optimal coding, canonical representation of Gaussian process.

0047-259X/80/030319–13$02.00/0

capacity $C(\mathbb{A})$ is achieved by transmitting a Gaussian message with additive feedback. Using additive feedback, an input process $\Phi(\cdot)$ is expressible as

$$\Phi(t) = \Theta(t) + \Psi(t), \qquad 0 \leqslant t \leqslant T, \tag{2}$$

where $\Psi(\cdot)$ is a causal function of the output process $Y(\cdot)$. Define a Gaussian class $\mathbb{A}g$ by

$$\mathbb{A}g = \{(\Theta, \Phi) \in \mathbb{A};\ \Phi \text{ is given by (2) and } (\Theta(\cdot), \Psi(\cdot), Y(\cdot)) \text{ forms a Gaussian system}\}.$$

Then it is shown that the capacity $C(\mathbb{A})$ is attained in the class $\mathbb{A}g$:

$$C(\mathbb{A}) = C(\mathbb{A}g) \tag{3}$$

(Theorem 2). Furthermore, when the capacity is finite, we can analyze more detailed structure of the Gaussian message and the additive feedback by which the capacity is achieved (Theorem 1, Theorem 2).

Channel (1) is called a white Gaussian channel when the noise $X(\cdot)$ is a Brownian motion. Result (3) for the white Gaussian channel with feedback was known previouly [8, 10]. Ebert [3] gave the result (3) in the case where the noise $X(\cdot)$ is equivalent (or mutually absolutely continuous) to a Brownian motion. In the case of the Gaussian channel without feedback, (3) has also been known [1, 7].

We can get the analogous result for the discrete time Gaussian channel with feedback [9].

We will give the precise description of the results in Section 2. In Section 3, the proofs will be given.

2. Statement of the Results

Let $(\Omega, \mathscr{F}, P)$ be a basic probability space. The noise $X = \{X(t);\ 0 \leqslant t \leqslant T\}$ is a zero mean separable Gaussian process defined on $(\Omega, \mathscr{F}, P)$ such that $\int_0^T E[X^2(t)]\, dt < \infty$. A message Θ is a random variable, defined on $(\Omega, \mathscr{F}, P)$, taking values in some measurable space, in general. However, we may regard messages $\Theta = \{\Theta(t);\ 0 \leqslant t \leqslant T\}$ as stochastic processes. A model for a Gaussian channel with feedback is given by (1), where $\Phi = \{\Phi(t);\ 0 \leqslant t \leqslant T\}$ and $Y = \{Y(t);\ 0 \leqslant t \leqslant T\}$ are an input process and the corresponding output process defined on $(\Omega, \mathscr{F}, P)$, respectively. It is reasonable to assume the following conditions (A.1)–(A.3):

(A.1) Θ is independent of X;

(A.2) For each t, $\Phi(t)$ is $\mathscr{F}(\Theta)^{\vee} \mathscr{F}_t(Y)$-measurable, where $\mathscr{F}_t(Y)$ (resp. $\mathscr{F}(\Theta)$) is a σ-algebra generated by $Y(u)$, $0 \leqslant u \leqslant t$ (resp. $\Theta(u)$, $0 \leqslant u \leqslant T$);

(A.3) Stochastic equation (1) has a unique solution $Y(\cdot)$.

The condition (A.2) means that the channel is with feedback. The feedback is additive type if input processes are given by (2), where $\Psi(t)$ is $\mathscr{F}_t(Y)$-measurable.

We consider the case where the constraint for the input is given in terms of the covariance. That is, whether a process Φ can be an input to the channel or not is determined only by the covariance function of Φ. In mathematics, the constraint can be formulated as follows. Let $\mathbb{R}$ be a family of symmetric nonnegative definite functions (i.e., covariance functions) $R(s, t)$ defined on $[0, T]^2$ such that $\int_0^T R(t, t)\, dt < \infty$. We define a class $\mathbb{A}(\mathbb{R})$ of all admissible pairs (Θ, Φ) of a message Θ and an input process Φ, and the corresponding Gaussian class $\mathbb{A}g(\mathbb{R})$ by

$$\mathbb{A}(\mathbb{R}) = \{(\Theta, \Phi);\ (\Theta, \Phi) \text{ satisfies (A.1)–(A.3) and the covariance function of } \Phi \text{ belongs to } \mathbb{R}\},$$

$$\mathbb{A}g(\mathbb{R}) = \{(\Theta, \Phi) \in \mathbb{A}(\mathbb{R});\ \Phi \text{ is given by (2) and } (\Theta(\cdot), \Psi(\cdot), Y(\cdot)) \text{ forms a Gaussian system}\}.$$

Every separable Gaussian process has a canonical representation in the sense of Hida-Cramér [4]. We will be concerned with the case where the canonical representation of the Gaussian process X has no discrete spectrum, namely, we assume that X has a canonical representation of the form

$$X(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, dB_i(u), \qquad 0 \leqslant t \leqslant T, \tag{4}$$

where $F_i(t, u)$, $i = 1,..., N$, are canonical kernels and $dB_i(\cdot)$, $i = 1,..., N$, are mutually independent white Gaussian noises with continuous spectral measures,

$$dm_i(t) = E[|dB_i(t)|^2], \qquad i = 1,..., N,$$

such that m_{i+1} is absolutely continuous with respect to m_i $(i = 1,..., N-1)$. We denote by $\mathbb{R}^*$ the class of all covariance functions of the form,

$$R(s, t) = \sum_{i,j=1}^{N} \int_0^t \int_0^s F_i(s, u)\, F_j(t, v)\, r_{ij}(u, v)\, dm_i(u)\, dm_j(v), \tag{5}$$

where $r(u, v) \equiv (r_{ij}(u, v))_{i,j=1,...,N}$ is a symmetric nonnegative definite function defined on $[0, T]^2$, satisfying

$$\sum_{i=1}^{N} \int_0^T r_{ii}(u, u)\, dm_i(u) < \infty. \tag{6}$$

Hereafter, for brevity, we denote by $\mathbb{L}$ the family of all $r(u, v) \equiv (r_{ij}(u, v))_{i,j=1,\dots,N}$ such that

$$\sum_{i,j=1}^{N} \int_0^T \int_0^T r_{ij}^2(u, v)\, dm_i(u)\, dm_j(v) < \infty. \tag{7}$$

Denote by $\mathscr{H}(X)$ the reproducing kernel Hilbert space (RKHS) corresponding to the Gaussian process X. It is known [4] that if

$$P\{\Phi(\cdot, \omega) \in \mathscr{H}(X)\} = 1, \tag{8}$$

then the process Φ is represented as

$$\Phi(t, \omega) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, \varphi_i(u, \omega)\, dm_i(u), \qquad \text{a.e. } \omega, \tag{9}$$

with φ_i, $i = 1,\dots, N$, such that

$$\sum_{i=1}^{N} \int_0^T \varphi_i^2(t, \omega)\, dm_i(t) < \infty, \qquad \text{a.e. } \omega. \tag{10}$$

When the process Φ in (9) is Gaussian, it is also known that (10) is equivalent to the following condition

$$\sum_{i=1}^{N} \int_0^T E[\varphi_i^2(t)]\, dm_i(t) < \infty. \tag{11}$$

Thus, when Φ is a Gaussian process, condition (8) is equivalent to that the covariance function of Φ belongs to $\mathbb{R}^*$.

First of all, we sate the following:

THEOREM 1. *Let $R \in \mathbb{R}^*$ and $(\Theta, \Phi) \in \mathbb{A}(\{R\})$, and assume that Φ satisfies* (8). *Then there exists a Gaussian pair $(\Theta_0, \Phi_0) \in \mathbb{A}g(\{R\})$ having the following properties*:

(B.1) *The processes $(\varphi_1,\dots, \varphi_N, B_1,\dots, B_N)$ and $(\varphi_{01},\dots, \varphi_{0N}, B_1,\dots, B_N)$ have same covariance, where $(\varphi_1,\dots, \varphi_N)$ and $(\varphi_{01},\dots, \varphi_{0N})$ are related to the processes Φ and Φ_0, respectively, by* (9);

(B.2) *Almost all sample paths $\Theta_0(\cdot, \omega)$ of the process Θ_0 belong to $\mathscr{H}(X)$*;

(B.3) *$\Psi_0(t) \equiv \Phi_0(t) - \Theta_0(t)$ is of the form,*

$$\Psi_0(t) = -\sum_{i,j=1}^{N} \int_0^t \int_0^u F_i(t, u)\, l_{ij}(u, v)\, dZ_{0j}(v)\, dm_i(u), \tag{12}$$

where $Z_{0j}(\cdot)$, $j = 1,\dots, N$, are given by

$$Y_0(t) = \Phi_0(t) + X(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\, dZ_{0i}(u), \tag{13}$$

and $l(u, v) \equiv (l_{ij}(u, v)) \in \mathbb{L}$ is a Volterra kernel (i.e., $l_{ij}(u, v) = 0$ for $u < v$).

Then main purpose is to show that the capacity of the Gaussian channel is achieved within the Gaussian system. We will show this dividing into two cases, according as $\mathbb{R} \subset \mathbb{R}^*$ or $\mathbb{R} \not\subset \mathbb{R}^*$.

THEOREM 2. *If* (i) $\mathbb{R} \subset \mathbb{R}^*$ *and the condition* (8) *is satisfied by every* $(\Theta, \Phi) \in \mathbb{A}(\mathbb{R})$, *or if* (ii) $\mathbb{R} \not\subset \mathbb{R}^*$, *then*

$$C(\mathbb{A}(\mathbb{R})) = C(\mathbb{A}g(\mathbb{R})). \tag{14}$$

In the case (i), *the capacity $C(\mathbb{A}(\mathbb{R}))$ is achieved by transmitting a Gaussian message with additive feedback given by* (B3). *In the case* (ii), *the capacity $C(\mathbb{A}(\mathbb{R}))$ is infinite and is attained by transmitting a Gaussian message without feedback.*

Theorem 2 is a generalization of the result due to Ebert [3]. He considered the case where the noise X is equivalent to a Brownian motion (namely, the measure induced on the function space by the process X and the Wiener measuree are mutually absolutely continuous), and the input processes are limited in the form

$$\Phi(t) = \int_0^t \varphi(u)\, du \qquad \text{with} \quad \int_0^T E[\varphi^2(t)]\, dt \leqslant \rho T \tag{15}$$

($\rho > 0$ is a given constant). In this case, we can analyze more detailed structure of the pair (Θ, Φ) by which the capacity is achieved. Due to Hitsuda [5], the Gaussian process X, equivalent to a Brownian motion, can be canonically represented as

$$X(t) = B(t) + \int_0^t \int_0^s f(s, u)\, dB(u)\, ds, \tag{16}$$

where $B(\cdot)$ is a Brownian motion and $f(s, u) \in L^2([0, T]^2)$ is a Volterra kernel. Note that the canonical kernel $F(t, u)$ of X is

$$F(t, u) = 1 + \int_u^t f(s, u)\, ds, \qquad t \geqslant u, \tag{17}$$

and the spectral measure is Lebesgue and that $\mathscr{H}(X)=\mathscr{H}(B)$. Define a class $\mathbb{R}_0$ of covariance functions by

$$\mathbb{R}_0=\left\{R;\, R(s,t)=\int_0^t\int_0^s r(u,v)\,du\,dv \text{ with } \int_0^T r(u,u)\,du\leqslant \rho T\right\}.$$

Then we can easily show that $\mathbb{R}_0$ satisfies the conditions in (i) of Theorem 2. Thus, applying Theorem 2, the capacity C of the channel with the noise X in (16) and the constraint (15) is given by $C=C(\mathbb{A}(\mathbb{R}_0))=C(\mathbb{A}g(\mathbb{R}_0))$. Let $\theta=\{\theta(t);\ 0\leqslant t\leqslant T\}$ be a Gaussian process, independent of X, such that $\int_0^T E[\theta^2(t)]\,dt<\infty$. Then we can show that the stochastic equation,

$$Y(t)=\int_0^t [\theta(u)-E[\theta(u)\,|\,\mathscr{F}_u(Y)]]\,du+X(t),\qquad 0\leqslant t\leqslant T,\tag{18}$$

has a unique solution $Y(\cdot)$, where $E[\theta(u)\,|\,\mathscr{F}_u(Y)]$ is the conditional expectation (see Lemma 5 of Section 3). Giving a message Θ and an input Φ expressed in the form

$$\Theta(t)=\int_0^t \theta(u)\,du,\qquad 0\leqslant t\leqslant T,\tag{19}$$

$$\Phi(t)=\int_0^t [\theta(u)-E[\theta(u)\,|\,\mathscr{F}_u(Y)]]\,du,\qquad 0\leqslant t\leqslant T,\tag{20}$$

we define a subclass $\hat{\mathbb{A}}(\mathbb{R}_0)$ of $\mathbb{A}g(\mathbb{R}_0)$ by

$$\hat{\mathbb{A}}(\mathbb{R}_0)=\{(\Theta,\Phi)\in\mathbb{A}g(\mathbb{R}_0);\ \Theta \text{ and } \Phi \text{ are given by (19) and (20)}\}.$$

Then we can prove that the optimal pair is found in $\hat{\mathbb{A}}(\mathbb{R}_0)$.

THEOREM 3. *Under the average power constraint* (15), *the capacity C of the Gaussian channel with feedback and with the noise X equivalent to a Brownian motion is given by*

$$C=C(\mathbb{A}(\mathbb{R}_0))=C(\hat{\mathbb{A}}(\mathbb{R}_0)).\tag{21}$$

Throughout the paper, we assume that the channel is with instantaneous feedback. However, we should note that our results are still valid for the channel with time-lag feedback.

3. Proof of the Results

In this section we will give the proofs of the theorems. Without loss of generality, we may assume that the expectations of processes are zero.

First of all, we give a formula for the mutual information $I(\Theta, Y)$ in the Gaussian channel (1). Let us assume that (4) and conditions (A1)–(A3) are satisfied. The following result has been known.

Lemma 1. (i) (*Hitsuda and Ihara* [6]) *Assume that an input Φ is given by* (9) *and* (11). *Then $I(\Theta, Y)$ is given by*

$$I(\Theta, Y) = \frac{1}{2} \sum_{i=1}^{N} \int_0^T E[|\varphi_i(t) - \hat{\varphi}_i(t)|^2]\, dm_i(t), \tag{22}$$

where $\hat{\varphi}_i(t) = E[\varphi_i(t) \mid \mathscr{F}_t(Y)]$.

(ii) (*Pitcher* [12]) *Assume that the channel* (1) *is without feedback and that Φ is a Gaussian process. Then $I(\Phi, Y)$ is finite if and only if Φ satisfies* (8).

For the proof of Theorem 1, we prepare two lemmas.

Lemma 2. *Let $\mathbb{R} \subset \mathbb{R}^*$, $(\Theta, \Phi) \in \mathbb{A}(\mathbb{R})$ and Φ be given by* (9). *Then for every $s \in [0, T]$ the function $E[\varphi_i(s) B_j(\cdot)]$ is absolutely continuous with respect to the measure m_j, where $dB_j(\cdot)$ is the white noise in* (4). *Moreover if we denote by $h_{ij}(s, \cdot)$ the Radon–Nikodym derivative:*

$$h_{ij}(s, \cdot) = \frac{d}{dm_j} E[\varphi_i(s) B_j(\cdot)], \tag{23}$$

then $h(s, t) \equiv (h_{ij}(s, t))_{i,j=1,\dots,N}$ is a Volterra kernel in $\mathbb{L}$.

Proof. Denote by $\bar{\varphi}_{ij}(s, t)$ the linear least mean square estimate of $\varphi_i(s)$ based on the observation of $\{B_j(u);\ 0 \leqslant u \leqslant t\}$. Then it is easy to show that there is a function $h(u, v) = (h_{ij}(u, v)) \in \mathbb{L}$ such that

$$\bar{\varphi}_{ij}(s, t) = \int_0^t h_{ij}(s, v)\, dB_j(v), \qquad i, j = 1,\dots, N. \tag{24}$$

Since $E[(\varphi_i(s) - \bar{\varphi}_{ij}(s, t))\, B_j(t)] = 0$, it follows from (24) that

$$E[\varphi_i(s) B_j(t)] = \int_0^t h_{ij}(s, v)\, dm_j(v). \tag{25}$$

Equation (25) means that the function $E[\varphi_i(s) B_j(\cdot)]$ is absolutely continuous with respect to m_j and the Radon–Nikodym derivative is $h_{ij}(s, \cdot)$. If $t > s$ and

$|\delta| < t-s$, then $\varphi_i(s)$ is independent of $B_j(t+\delta) - B_j(t)$. Hence $h_{ij}(s,t) = 0$ for $t > s$ and $h(s,t)$ is a Volterra kernel.

The following property, concerning the Volterra kernel, is known.

LEMMA 3. *Let $h(s,t) = (h_{ij}(s,t))_{i,j=1,\dots,N} \in \mathbb{L}$ be a Volterra kernel. Then there exists uniquely a resolvent kernel $l(s,t) = (l_{ij}(s,t))_{i,j=1,\dots,N} \in \mathbb{L}$ such that*

$$l_{ij}(s,t) + h_{ij}(s,t) + \sum_{k=1}^{N} \int_t^s l_{ik}(s,u)\, h_{kj}(u,t)\, dm_k(u) = 0.$$

We now proceed to the proof of Theorem 1.

Proof of Theorem 1. Let $(\varphi_{01},\dots,\varphi_{0N})$ be an N-dimensional Gaussian process such that $(\varphi_{01},\dots,\varphi_{0N}, B_1,\dots,B_N)$ is a Gaussian process with the same covariance as that of $(\varphi_1,\dots,\varphi_N, B_1,\dots,B_N)$, and define a process Φ_0 by $\Phi_0(t) = \sum_{i=0}^N \int_0^t F_i(t,u)\, \varphi_{0i}(u)\, dm_i(u)$. We define a Gaussian process $Y_0(t) = \sum_{i=1}^N \int_0^t F_i(t,u)\, dZ_{0i}(u)$ by

$$Z_{0i}(t) = \int_0^t \varphi_{0i}(s)\, dm_i(s) + B_i(t), \qquad 0 \leqslant t \leqslant T, \tag{26}$$

to see that $Y_0(t) = \Phi_0(t) + X(t)$. And we define a Gaussian process $\Theta_0(t) = \sum_{i=1}^N \int_0^t F_i(t,u)\, \theta_{0i}(u)\, dm_i(u)$ by

$$\theta_{0i}(t) = \varphi_{0i}(t) + \sum_{j=1}^{N} \int_0^t l_{ij}(t,u)\, dZ_{0j}(u), \qquad 0 \leqslant t \leqslant T, \tag{27}$$

where $l(s,t) = (l_{ij}(s,t)) \in \mathbb{L}$ is the resolvent kernel of the Volterra kernel $h(s,t) = (h_{ij}(s,t))$ given by (23). What we have to show is that (Θ_0, Φ_0) satisfies the conditions (A1)–(A3) and (B1)–(B3). The verifications of (A2), (B1) and (B2) are straightforward. By (26) and (27), we see that $(Z_{01},\dots,Z_{0N})$ satisfies the stochastic equations

$$Z_{0i}(t) = \int_0^t \left[\theta_{0i}(s) - \sum_{j=1}^{N} \int_0^s l_{ij}(s,u)\, dZ_{0j}(u)\right] dm_i(s) + B_i(t), \qquad i = 1,\dots,N, \tag{28}$$

from which (B3) follows. Using (23), (26), (27) and Lemma 3, we see that for every i and j

$$E[\theta_{0i}(s)\, B_j(t)] = \int_0^{s\wedge t} \left[h_{ij}(s,u) + l_{ij}(s,u) + \sum_{k=1}^{N} \int_0^s l_{ik}(s,v)\, h_{kj}(v,u)\, dm_k(v)\right] dm_j(u) = 0,$$

where $s \wedge t = \min(s, t)$. Hence $(\theta_{01}, ..., \theta_{0N})$ is independent of $(B_1, ..., B_N)$. Thus (A1) is satisfied. Finally (A3) follows from the uniqueness of the solution of stochastic equations (28).

We now can prove Theorem 2 in the case (i).

Proof of Theorem 2 (*case* (i)). In order to prove (14) it is enough to show, for each fixed $R \in \mathbb{R}$, that

$$C(\mathbb{A}(\{R\})) = C(\mathbb{A}g(\{R\})). \tag{29}$$

Let $R \in \mathbb{R} \subset \mathbb{R}^*$ be given by (5) and let $(\Theta, \Phi) \in \mathbb{A}(\{R\})$. From condition (8), Φ is represented as (9). And it follows that

$$E[\varphi_i(s)\, \varphi_j(t)] = r_{ij}(s, t).$$

By Theorem 1, there exists a Gaussian pair $(\Theta_0, \Phi_0) \in \mathbb{A}g(\{R\}))$ satisfying (B1)–(B3). Let Y and Y_0 be the outputs corresponding to the inputs Φ and Φ_0, respectively. Noting that condition (11) follows from the assumption $\mathbb{R} \subset \mathbb{R}^*$ and applying (i) of Lemma 1, we have

$$I(\Theta, Y) = \frac{1}{2} \sum_{i=1}^{N} \int_0^T E[|\varphi_i(t) - \hat{\varphi}_i(t)|^2]\, dm_i(t), \tag{30}$$

$$I(\Theta_0, Y_0) = \frac{1}{2} \sum_{i=1}^{N} \int_0^T E[|\varphi_{0i}(t) - \hat{\varphi}_{0i}(t)|^2]\, dm_i(t), \tag{31}$$

where $\hat{\varphi}_i(t) = E[\varphi_i(t) \mid \mathscr{F}_t(Y)]$ and $\hat{\varphi}_{0i}(t) = E[\varphi_{0i}(t) \mid \mathscr{F}_t(Y_0)]$. Denote by $\bar{\varphi}_i(t)$ (resp. $\bar{\varphi}_{0i}(t)$) the linear least mean square estimate of $\varphi_i(t)$ (resp. $\varphi_{0i}(t)$) based on the observation of $\{Y(u); 0 \leqslant u \leqslant t\}$ (resp. $\{Y_0(u); 0 \leqslant u \leqslant t\}$). From (B1) we have

$$E[|\varphi_i(t) - \bar{\varphi}_i(t)|^2] = E[|\varphi_{0i}(t) = \bar{\varphi}_{0i}(t)|^2].$$

Since $(\varphi_{01}, ..., \varphi_{0N}, Y_0)$ is Gaussian, $\hat{\varphi}_{0i}(t) = \bar{\varphi}_{0i}(t)$. Hence we have

$$E[|\varphi_i(t) - \hat{\varphi}_i(t)|^2] \leqslant E[|\varphi_i(t) - \bar{\varphi}_i(t)|^2] = E[|\varphi_{0i}(t) - \hat{\varphi}_{0i}(t)|^2]. \tag{32}$$

Combining (30)–(32), we get the inequality,

$$I(\Theta, Y) \leqslant I(\Theta_0, Y_0).$$

This inequality implies that

$$C(\mathbb{A}(\{R\})) \leqslant C(\mathbb{A}g(\{R\})).$$

The converse inequality is clear from the definition, and we have the desired equality (29). Thus the proof is completed.

In case (ii), for the proof of Theorem 2, we need a zero-one law for a Gaussian process which is a slight generalization of the result due to Kallianpur [11] and Driscoll [2].

LEMMA 4. *Let $K(s,t)$ be a symmetric nonnegative definition function defined on $[0,T]^2$ satisfying $\int_0^T\int_0^T K^2(s,t)\,ds\,dt<\infty$, and let $\Phi=\{\Phi(t); 0\leqslant t\leqslant T\}$ be a Gaussian process with covariance $R(s,t)$ such that $\int_0^T R(t,t)\,dt<\infty$. Then it holds that*

$$P\{\Phi(\cdot,\omega)\in\mathscr{H}(K)\}=1 \qquad or \qquad P\{\Phi(\cdot,\omega)\in\mathscr{H}(K)\}=0,$$

where $\mathscr{H}(K)$ is the RKHS with the reproducing kernel K.

Proof of Theorem 2 (*case* (ii)). Let $\Phi=\{\Phi(t); 0\leqslant t\leqslant T\}$ be a Gaussian process having a function $R\in\mathbb{R}\backslash\mathbb{R}^*$ as the covariance and let $Y(t)=\Phi(t)+X(t)$. We will show that

$$P\{\Phi(\cdot,\omega)\notin\mathscr{H}(X)\}=1. \tag{33}$$

On the contrary, suppose that (33) were not true. Then it follows from Lemma 4 that

$$P\{\Phi(\cdot,\omega)\in\mathscr{H}(X)\}=1.$$

Hence Φ can be represented as (9) with (11). Putting $r_{ij}(s,t)=E[\varphi_i(s)\,\varphi_j(t)]$, $i,j=1,\ldots,N$, $R(s,t)$ is written as (5). Hence, from (11), we have that $R\in\mathbb{R}^*$, a contradiction. Therefore (33) is true, and we know that

$$I(\Phi,Y)=\infty, \tag{34}$$

by (ii) of Lemma 1. Giving a message $\Theta=\{\Theta(t); 0\leqslant t\leqslant T\}$ by $\Theta(t)=\Phi(t)$, it is clear that $(\Theta,\Phi)\in\mathbb{A}g(\mathbb{R})$ and we have the desired equations $C(\mathbb{A}(\mathbb{R}))=C(\mathbb{A}g(\mathbb{R}))=\infty$, from (34).

We now turn to the proof of Theorem 3. Let the noise X be equivalent to a Brownian motion and be given by (16). Let $\theta=\{\theta(t); 0\leqslant t\leqslant T\}$ be a Gaussian process, independent of X,such that $\int_0^T E[\theta^2(t)]\,dt<\infty$ and define a process $Y_1=\{Y_1(t); 0\leqslant t\leqslant T\}$ by

$$Y_1(t)=\int_0^t\theta(s)\,ds+X(t).$$

Then we can prove the following lemma, in the similar manner to the case where $X(\cdot)$ is a Brownian motion.

LEMMA 5. *Given a Volterra kernel $k(u, v) \in L^2([0, T]^2)$, the stochastic equation*

$$Y(t) = \int_0^t \theta(s)\, ds - \int_0^t \int_0^s k(s, u)\, dY(u)\, ds + X(t), \qquad 0 \leqslant t \leqslant T,$$

has the unique solution,

$$Y_2(t) = Y_1(t) + \int_0^t \int_0^s l(s, u)\, dY_1(u)\, ds,$$

where $l(u, v) \in L^2([0, T]^2)$ is the resolvent kernel of $k(u, v)$. The stochastic equation

$$Y(t) = \int_0^t [\theta(s) - E[\theta(s) \mid \mathscr{F}_s(Y)]]\, ds + X(t), \qquad 0 \leqslant t \leqslant T,$$

has a unique solution $Y_3(\cdot)$, and $\hat{\theta}(s) = E[\theta(s) \mid \mathscr{F}_s(Y)]$ is presented as

$$\hat{\theta}(s) = \int_0^s h(u, v)\, dY(u),$$

where $h(u, v) \in L^2([0, T]^2)$ is a Volterra kernel. Moreover it holds that

$$\mathscr{F}_t(Y_1) = \mathscr{F}_t(Y_2) = \mathscr{F}_t(Y_3), \qquad 0 \leqslant t \leqslant T.$$

Proof of Theorem 3. In order to prove Theorem 3, it is enough to show that for each pair $(\Theta, \Phi) \in \mathbb{A}g(\mathbb{R}_0)$ there exists a pair $(\Theta_0, \Phi_0) \in \hat{\mathbb{A}}(\mathbb{R}_0)$ such that

$$I(\Theta, Y) = I(\Theta_0, Y_0), \tag{35}$$

where Y and Y_0 are the outputs corresponding to the inputs Φ and Φ_0, respectively. From Theorem 1, $Y(\cdot)$ is expressed in the form

$$Y(t) = \int_0^t F(t, u) \left[\theta(u) - \int_0^u l(u, s)\, dZ(s)\right] du + X(t),$$

where $\Theta(t) = \int_0^t F(t, u)\, \theta(u)\, du$, $Y(t) = \int_0^t F(t, u)\, dZ(u)$ and $l(u, v) \in L^2([0, T]^2)$ is a Volterra kernel. Denoting by $g(u, v) \in L^2([0, T]^2)$ the resolvent kernel of $f(u, v)$ in (16), we define a Gaussian process $\Theta_0 = \int_0^t \theta_0(u)\, du$ and a Volterra kernel $k(u, v) \in L_2([0, T]^2)$ by

$$\theta_0(t) = \theta(t) + \int_0^t f(t, u)\, \theta(u)\, du, \tag{36}$$

$$k(u, v) = l(u, v) + \int_0^u f(u, s)\, l(s, v)\, ds + \int_0^u l(u, s)\, g(s, v)\, ds$$
$$+ \int_0^u \int_0^s f(u, s)\, l(s, w)\, g(w, v)\, dw\, ds.$$

Then we get the following relations

$$\theta(t) = \theta_0(t) + \int_0^t g(t, u)\, \theta_0(u)\, du, \tag{37}$$

$$Y(t) = \int_0^t \left[\theta_0(s) - \int_0^s k(s, u)\, dY(u) \right] ds + X(t). \tag{38}$$

From (38) and the definition of the class $\mathbb{R}_0$, we know that

$$\int_0^T E\left[\left| \theta_0(s) - \int_0^s k(s, u)\, dY(u) \right|^2 \right] ds \leqslant \rho T. \tag{39}$$

We now define a coding scheme by

$$Y_0(t) = \int_0^t [\theta_0(s) - \hat{\theta}_0(s)]\, ds + X(t) \equiv \Phi_0(t) + X(t),$$

where $\hat{\theta}_0(s) = E[\theta_0(s) \mid \mathscr{F}_s(Y_0)]$. It follows from Lemma 5 that

$$\mathscr{F}_t(Y_0) = \mathscr{F}_t(Y), \qquad 0 \leqslant t \leqslant T. \tag{40}$$

Therefore, from (39), we have the inequality

$$\int_0^T E[|\theta_0(t) - \hat{\theta}_0(t)|^2]\, dt \leqslant \rho T,$$

which means $(\Theta_0, \Phi_0) \in \hat{\mathbb{A}}(\mathbb{R}_0)$. And the desired equality (35) follows from (36), (37), and (40).

References

[1] Baker, C. R. (1978). Capacity of the Gaussian channel without feedback. *Inform. Contr.* **37** 70–89.

[2] Driscoll, M. F. (1973). The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **26** 309–316.

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

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

[5] HITSUDA, M. (1968). Representation of Gaussian processes equivalent to Wiener process. *Osaka J. Math.* **5** 299–312.

[6] HITSUDA, M. AND IHARA, S. (1975). Gaussian channels and the optimal coding. *J. Multivar. Anal.* **5** 106–118.

[7] HUANG, R. Y. AND JOHNSON, R. A. (1962). Information capacity of time-continuous channels. *IEEE Trans. Information Theory* **IT-8** s191–s198.

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

[9] IHARA, S. (1979). On the capacity of the discrete time Gaussian channel with feedback. In *Transactions, 8th Prague Conference on Information Theory, Statistical Decisions, Random Processes, Prague 1978*, Vol. C, pp. 175–186.

[10] 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.

[11] KALLIANPUR, G. (1970). Zero-one laws for Gaussian processes. *Trans. Amer. Math. Soc.* **149** 199–211.

[12] PITCHER, T. S. (1963). On the sample functions of processes which can be added to a Gaussian process. *Ann. Math. Statist.* **34** 329–333.