S. Ihara
Nagoya Math. J.
Vol. 58 (1975), 127-147

CODING THEORY IN GAUSSIAN CHANNEL WITH FEEDBACK II: EVALUATION OF THE FILTERING ERROR

SHUNSUKE IHARA

Introduction.

The main purpose of this paper is to give a method to evaluate the actual value of the filtering error which arises in the transmission of a signal process, using the optimal coding, over a Gaussian channel. In his earier papers ([4] and [7]), the author has shown a method to construct an optimal causal coding for which the filtering error is minimized and at the same time the mutual information is maximized.

A Gaussian channel is expressed in the form,

$$Y(t) = \Phi(t) + X(t) ,$$

where $\Phi(t)$ stands for the channel input which is a function of a Gaussian message $\{\xi(s); s \leq t\}$ and of the output $\{Y(s); s \leq t\}$, and where $X(\cdot)$ is a Gaussian noise assumed to be independent of $\xi(\cdot)$. As in [4] and [7] we assume that the input $\Phi(\cdot)$ is limited by the average power. It has been proved that the coding attaining the minimal error is given by a linear algorism. We now come to the evaluation of the error, which will be given in this paper.

In section 1, we shall prove the convergence theorem which asserts that for any Gaussian process $\xi(\cdot)$, if $\xi_n(\cdot)$, $n = 1, 2, \cdots$, form a sequence of Gaussian processes converging to $\xi(\cdot)$ in the sense of mean square for each moment t, then the error for $\xi_n(\cdot)$ tends to the one for $\xi(\cdot)$ (Theorem 1). Each $\xi_n(\cdot)$ can be taken to be a stepwise Gaussian process for which the filtering error can be obtained explicitly.

The actual value of the error for a stepwise process is given by Theorem 2, in section 2.

Received February 24, 1975.

Section 3 will be devoted to discussions on some related topics being in line with our approach to construct an optimal coding. There will be discussed an optimal but not necessarily causal coding. It will be proved that there exists a coding method with which we can send a message having the information equal to the channel capacity and the filtering error can be minimized although the way of evaluating is somewhat different from the ealier sections (Theorem 4). It is noted that an interesting difference between causal cases and non-causal ones appears in the evaluation of the filtering error.

It might be worth noting that an intrinsic meaning of the concept of multiplicity of a Gaussian process may be given from the point of view of information theory in the same spirit as the present paper (cf. [2], [3]).

The author should like to express his thanks to Professors M. Hitsuda and I. Kubo.

§ 1. A convergence theorem in Gaussian channel.

In this section, we establish a convergence theorem on the filtering errors according to messages.

Let $X(\cdot)$ be a zero mean separable Gaussian process with the canonical representation in the sense of Hida-Cramér, (cf. [1]).

$$X(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u) dB_i(u) , \qquad (N \leqq \infty) , \tag{1.1}$$

where $B_i(u)$'s are mutually independent Gaussian processes with independent increment such that $E|dB_i(u)|^2 = m_i(du)$'s are continuous measures with the property $m_i \gg m_{i+1}$. The number N is called the multiplicity of the process $X(\cdot)$. The *Gaussian channel* treated here is the following type:

$$Y(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u) A_i(u)(\xi_i(u) - f_i(u)) m_i(du) + X(t) , \qquad 0 \leq t \leq T \, (< \infty) , \tag{1.2}$$

where $\xi_i(\cdot)$'s are messages independent of $X(\cdot)$, $A_i(u)$'s are non-negative (non-random) functions and $f_i(u)$'s are $\mathscr{F}_u(Y)$*) measurable functions. Let us assume the following conditions.

(a.1) The equation (1.2) has the unique solution $Y(\cdot)$.

*) We denote by $\mathscr{F}_t(Y)$ the σ-algebra generated by $\{Y(s); 0 \leq s \leq t\}$.

(a.2) The message $\xi(\cdot) = (\xi_1(\cdot), \cdots, \xi_N(\cdot))$ is a N-dimensional Gaussian process such that $E\xi_i(t) = 0$, $0 \le t \le T$, and that $0 < \int_0^T E\xi_i^2(t)m_i(dt) < \infty$.

(a.3) The channel input satisfies the average power constraint:

$$A_i^2(t)E\,|\xi_i(t) - f_i(t)|^2 \le \rho_i(t)\ , \qquad i = 1, \cdots, N\ , \tag{1.3}$$

where ρ_i's are given positive functions such that $\sum_{i=1}^{N}\int_0^T \rho_i(u)m_i(du) < \infty$.

Here we review some results obtained in the previous papers [4] and [7], which are useful in later. Let us denote by $\boldsymbol{\Phi}$ the class of admissible codings:

$$\boldsymbol{\Phi} = \Big\{\Phi(\cdot)\,;\ \Phi(t) = \sum_{i=1}^{N}\int_0^t F_i(t, u)A_i(u)(\xi_i(u) - f_i(u))m_i(du)\ , \quad \text{satisfying (a.1) and (a.3)}\Big\}\ . \tag{1.4}$$

The capacity C_t $(0 \le t \le T)$ of the channel (1.2) under the constraint (1.3) is defined by

$$C_t = \sup I_t(\xi, Y)\ , \tag{1.5}$$

where $I_t(\xi, Y)$ is the mutual information between $\{\xi(s)\,;\, 0 \le s \le t\}$ and $\{Y(s)\,;\, 0 \le s \le t\}$, and the supremum is taken over all messages ξ satisfying (a.2) and all $\Phi \in \boldsymbol{\Phi}$. It has been shown ([4]) that

$$C_t = \frac{1}{2}\sum_{i=1}^{N}\int_0^t \rho_i(u)m_i(du)\ . \tag{1.6}$$

Let a message $\xi = (\xi_1, \cdots, \xi_N)$ be fixed, then we say that a coding $\Phi \in \boldsymbol{\Phi}$ is *optimal in information sense* if $I_t(\xi, Y) = C_t$, $0 \le t \le T$. While a coding $\Phi \in \boldsymbol{\Phi}$ is said to be *optimal in filtering sense* if the infimum of the filtering error

$$\Delta(t) = \inf \sum_{i=1}^{N} E\,|\xi_i(t) - \hat{\xi}_i(t)|^2 \tag{1.7}$$

is attained by $\Phi \in \boldsymbol{\Phi}$, where $\hat{\xi}_i(t) = E[\xi_i(t)\,|\,\mathscr{F}_t(Y)]$ and the infimum is taken over all $\Phi \in \boldsymbol{\Phi}$.

We have shown the following two lemmas.

LEMMA 1. (i) (Theorem 2 of [4] and Theorem 2 of [7]). *Let a Gaussian message* $\xi(\cdot)$ *satisfy the assumption* (a.2). *Then the coding*

$\Phi^* \in \boldsymbol{\Phi}$ defined by the following equations (1.8) *and* (1.9) *is optimal in information sense and also in filtering sense,*

$$Y(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u) A_i(u)(\xi_i(u) - \hat{\xi}_i(u)) m_i(du) + X(t) , \tag{1.8}$$

$$A_i^2(t) \Delta_i(t) = \rho_i(t) , \qquad i = 1, \cdots, N , \tag{1.9}$$

where

$$\Delta_i(t) = E\,|\xi_i(t) - \hat{\xi}_i(t)|^2 \quad \text{and} \quad \hat{\xi}_i(t) = E[\xi_i(t) \,|\, \mathscr{F}_t(Y)] . \tag{1.10}$$

(ii) (Lemma 2.4 of [4]) *The inequlity*

$$A_i^2(t) E \xi_i^2(t) \le \rho_i(t) \exp \left[\sum_{j=1}^{N} \int_0^t \rho_j(u) m_j(du) \right]$$

holds.

Now the problem is to find a method evaluating the filtering error $\Delta(t) = \sum_{i=1}^N \Delta_i(t)$ determnied by (1.8)–(1.10).

LEMMA 2. (Proposition 3 of [4]). *Let $Y(\cdot)$ and A_i's be the process and the functions given by* (1.8) *and* (1.9). *Define a process $Z(\cdot) = (Z_1(\cdot), \cdots, Z_N(\cdot))$ by*

$$Z_i(t) = \int_0^t A_i(u) \xi_i(u) m_i(du) + B_i(t) , \qquad i = 1, \cdots, N . \tag{1.11}$$

Then $\mathscr{F}_t(Z) = \mathscr{F}_t(Y)$, $\hat{\xi}_i(t) = E[\xi_i(t) \,|\, \mathscr{F}_t(Y)] = E[\xi_i(t) \,|\, \mathscr{F}_t(Z)]$ and $I_t(\xi, Z) = I_t(\xi, Y)$.

It is well known that the following result can be obtained from Lemma 2.

COROLLARY. *The filtering $\hat{\xi}_i(t) = E[\xi_i(t) \,|\, \mathscr{F}_t(Y)] = E[\xi_i(t) \,|\, \mathscr{F}_t(Z)]$ is given by*

$$\hat{\xi}_i(t) = \sum_{j=1}^{N} \int_0^t H_{ij}(t, u) dZ_j(u) , \tag{1.12}$$

where $h_{ij}(t, u) = A_i(t) H_{ij}(t, u)$ is the solution of the following Wiener-Hopf equation,

$$\begin{aligned} &A_i(t) r_{ij}(t, s) A_j(s) \\ &\quad = \sum_{k=1}^{N} \int_0^t h_{ik}(t, u) A_k(u) r_{kj}(u, s) A_j(s) m_k(du) + h_{ij}(t, s) , \\ &\qquad\qquad 0 \le s \le t \le T , \ i, j = 1, \cdots, N , \end{aligned} \tag{1.13}$$

where $r_{ij}(t,s)=E[\xi_i(t)\xi_j(s)]$. *And then* $\Delta(t)=\sum_{i=1}^N E|\xi_i(t)-\hat{\xi}_i(t)|^2$ *is given by*

$$E|\xi_i(t)-\hat{\xi}_i(t)|^2=A_i^{-2}(t)h_{ii}(t,t)\,,\qquad 0\le t\le T\,. \tag{1.14}$$

The Corollary implies that if the equation (1.13) is solved then the error can be evaluated. But, unfortunately, it is difficult to get the concrete solution of (1.13) in general. Therefore we will evaluate the error $\Delta(t)$ by an approximating method. For this purpose we prepare a convergence theorem. At first we prove the continuity of the error $\Delta(t)$ with respect to the power $\rho(t)$.

LEMMA 3. *Let* $\rho_{ni}(t)$, $n=1,2,\cdots$, $i=1,\cdots,N$ *be non-negative functions such that*

$$\lim_{n\to\infty}\rho_{ni}(t)=\rho_i(t)\qquad \text{for every } 0\le t\le T\,,\ i=1,\cdots,N\,.$$

And define a process $U_n(t)$ *and functions* $D_{ni}(t)$, $i=1,\cdots,N$, *by* (1.8) *and* (1.9), *replacing* $\rho_i(t)$ *by* $\rho_{ni}(t)$, *namely,*

$$U_n(t)=\sum_{i=1}^N\int_0^t F_i(t,u)D_{ni}(u)(\xi_i(u)-\hat{\xi}_i^n(u))m_i(du)+X(t)\,,$$
$$D_{ni}^2(t)\Delta_i^n(t)=\rho_{ni}(t)\,,$$

where

$$\Delta_i^n(t)=E|\xi_i(t)-\hat{\xi}_i^n(t)|^2\quad\text{and}\quad \hat{\xi}_i^n(t)=E[\xi_i(t)\,|\,\mathscr{F}_t(U_n)]\,.$$

Then

$$\lim_{n\to\infty}\Delta_i^n(t)=\Delta_i(t)\qquad \text{for every } 0\le t\le T\,,\ i=1,\cdots,N\,. \tag{1.15}$$

Proof. (i) At first we assume that $\rho_{ni}(t)$'s are monotone decreasing as $n\nearrow\infty$, for each $0\le t\le T$ and $i=1,\cdots,N$. Then it is easily shown that $D_{ni}(t)$ is monotone decreasing (cf. Proof of Theorem 2 of [4]) and $\Delta_i^n(t)$ is monotone increasing (Lemma 2.3 of [4]) as $n\nearrow\infty$. The monotonicity enable us to define the functions $D_i(t)$ and $\Delta_i^0(t)$ by

$$D_i(t)=\lim_{n\to\infty}D_{ni}(t)\quad\text{and}\quad \Delta_i^0(t)=\lim_{n\to\infty}\Delta_i^n(t)\,.$$

In the same manner as in [6] we can show that if we define a process $U(\cdot)$ by

$$U(t)=\sum_{i=1}^{N}\int_0^t F_i(t,u)D_i(u)(\xi_i(u)-\hat{\xi}_i^0(t))m_i(du)+X(t)\ ,$$

$(\hat{\xi}_i^0(u)=E[\xi_i(u)\,|\,\mathscr{F}_u(U)])$ then $E\,|\xi_i(t)-\hat{\xi}_i^0(t)|^2=\Delta_i^0(t)$. Using the relations

$$D_i^2(t)\Delta_i^0(t)=\lim_{n\to\infty}D_{ni}^2(t)\Delta_i^n(t)=\lim_{n\to\infty}\rho_{ni}(t)=\rho_i(t)\ ,$$

and the uniqueness of the optimal coding (cf. Theorem 2 of [4]), we have

$$D_i(t)=A_i(t)\quad\text{and}\quad\Delta_i^0(t)=\Delta_i(t)$$

Thus (1.15) is proved.

(ii) In case where $\rho_{ni}(t)$'s are monotone increasing, (1.15) can be proved in the same way as above, since $A_i(t)$ is bounded ((ii) of Lemma 1).

(iii) In the general case, put

$$\rho_{ni}^+(t)=\sup_{k\geq n}(\rho_{ki}(t)\vee\rho_i(t))\quad\text{and}\quad\rho_{ni}^-(t)=\inf_{k\geq n}(\rho_{ki}(t)\wedge\rho_i(t))\ ,$$

(where $a\vee b=\max(a,b)$ and $a\wedge b=\min(a,b)$), then $\rho_{ni}^+(t)\geq\rho_{ni}(t)\geq\rho_{ni}^-(t)$ and $\rho_{ni}^+(t)$ ($\rho_{ni}^-(t)$) is monotone increasing (decreasing, respectively) to $\rho_i(t)$ as $n\nearrow\infty$. Since (i) and (ii) are applicable to ρ_n^- and ρ_n^+, respectively, if we define $D_{ni}^{\pm}$ and $\Delta_i^{\pm n}$ in the same manner as in (i) for $\rho_{ni}^{\pm}$, then we have

$$D_{ni}^+(t)\geq D_{ni}(t)\geq D_{ni}^-(t)\ ,\qquad \Delta_i^{+n}(t)\leq\Delta_i(t)\leq\Delta_i^{-n}(t)$$

and

$$D_{ni}^+(t)\searrow D_i(t)\ ,\quad D_{ni}^-(t)\nearrow D_i(t)\ ,\quad \Delta_i^{+n}(t)\nearrow\Delta_i(t)\ ,\quad \Delta_i^{-n}(t)\searrow\Delta_i(t)$$

as $n\nearrow\infty$. Thus the proof is completed.

Let $(\xi_{n1}(\cdot),\cdots,\xi_{nN}(\cdot))$ $(n=1,2,\cdots)$ be a Gaussian process satisfying the condition (a.2) and denote $r_{ij}^n(t,s)=E[\xi_{ni}(t)\xi_{nj}(s)]$. We introduce the following conditions (b.1)–(b.4):

(b.1) $\lim_{n\to\infty}r_{ii}^n(t,t)=r_{ii}(t,t),\ 0\leq t\leq T,\ i=1,\cdots,N.$

(b.2) $\displaystyle\lim_{n\to\infty}\int_0^T\int_0^T|r_{ij}^n(t,s)-r_{ij}(t,s)|^2\,m_i(dt)m_j(ds)=0\ .$

(b.3) There exist constants $0<\alpha_1<\alpha_2<\infty$ such that $\alpha_1\leq r_{ii}(t,t)\leq\alpha_2$ and $\alpha_1\leq r_{ii}^n(t,t)\leq\alpha_2$ for all $0\leq t\leq T$ and $i=1,\cdots,N$.

(b.4) There exists a constant $K>0$ such that $\rho_i(t)\leq K$ for all $0\leq t\leq T$ and $i=1,\cdots,N$.

Now we can give the statement of our theorem.

THEOREM 1. *Define processes* $Y_n(\cdot)$ *and functions* $A_{ni}(t)$ *by*

$$Y_n(t) = \sum_{i=1}^{N}\int_0^t F_i(t, u)A_{ni}(u)(\xi_{ni}(u) - \hat{\xi}_{ni}(u))m_i(du) + X(t) ,$$
$$A_{ni}^2(t)\Delta_{ni}(t) = \rho_i(t) , \qquad i = 1, \cdots, N ,$$

where

$$\Delta_{ni}(t) = E\,|\xi_{ni}(t) - \hat{\xi}_{ni}(t)|^2 \quad \textit{and} \quad \hat{\xi}_{ni}(t) = E[\xi_{ni}(t)\,|\,\mathscr{F}_t(Y_n)] .$$

Then, under the conditions (b.1)–(b.4), $\Delta_{ni}(t)$ *converges to* $\Delta_i(t)$ *for each* $0 \leq t \leq T$ *and* $i = 1, \cdots, N$.

In order to prove Theorem 1 we need a lemma. Let us introduce some notations for the lemma. We define processes $Z_n(\cdot) = (Z_{n1}(\cdot), \cdots, Z_{nN}(\cdot))$ and $Z_n^0(\cdot) = (Z_{n1}^0(\cdot), \cdots, Z_{nN}^0(\cdot))$ by

$$Z_{ni}(t) = \int_0^t A_{ni}(u)\xi_{ni}(u)m_i(du) + B_i(t) , \tag{1.16}$$

$$Z_{ni}^0(t) = \int_0^t A_{ni}(u)\xi_i(u)m_i(du) + B_i(t) \tag{1.17}$$

and define $\hat{\xi}_n^0$, Δ_n^0 and ρ_n^0 by (1.18), (1.19) and (1.20), respectively,

$$\hat{\xi}_{ni}^0(t) = E[\xi_i(t)\,|\,\mathscr{F}_t(Z_n^0)] , \tag{1.18}$$

$$\Delta_{ni}^0(t) = E|\xi_i(t) - \hat{\xi}_{ni}^0(t)|^2 , \tag{1.19}$$

$$\rho_{ni}^0(t) = A_{ni}^2(t)\Delta_{ni}^0(t) . \tag{1.20}$$

From Corollary of Lemma 1, there exist the kernels $H_{ij}^n(t, s)$ and $H_{ij}^{0n}(t, s)$ such that

$$\begin{aligned} \hat{\xi}_{ni}(t) &= \sum_{j=1}^{N}\int_0^t H_{ij}^n(t, u)dZ_{nj}(u) \\ &= \sum_{j=1}^{N}\int_0^t H_{ij}^n(t, u)A_{nj}(u)\xi_{nj}(u)m_j(du) + \sum_{j=1}^{N}\int_0^t H_{ij}^n(t, u)dB_j(u) , \end{aligned} \tag{1.21}$$

$$\begin{aligned} \hat{\xi}_{ni}^0(t) &= \sum_{j=1}^{N}\int_0^t H_{ij}^{0n}(t, u)dZ_{nj}^0(u) \\ &= \sum_{j=1}^{N}\int_0^t H_{ij}^{0n}(t, u)A_{nj}(u)\xi_j(u)m_j(du) + \sum_{j=1}^{N}\int_0^t H_{ij}^{0n}(t, u)dB_j(u) . \end{aligned} \tag{1.22}$$

If we denote

$$\tilde{\Delta}_{ni}^0(t) = E\left|\xi_i(t) - \sum_{j=1}^{N}\int_0^t H_{ij}^n(t, u)dZ_{nj}^0(u)\right|^2 \tag{1.23}$$

https://doi.org/10.1017/S0027763000016718 Published online by Cambridge University Press

and

$$\tilde{\Delta}_{ni}(t) = E\left|\xi_{ni}(t) - \sum_{j=1}^{N}\int_0^t H_{ij}^{0n}(t,u)dZ_{nj}(u)\right|^2, \tag{1.24}$$

then it is clear that

$$\Delta_{ni}(t) \leq \tilde{\Delta}_{ni}(t), \qquad \Delta_{ni}^0(t) \leq \tilde{\Delta}_{ni}^0(t). \tag{1.25}$$

LEMMA 4. *Under the conditions* (b.1)–(b.4), *for every fixed* $t \in [0, T]$, *there exist a constant* M $(= M(t)) < \infty$ *and numbers* ε_n $(= \varepsilon_n(t))$, $n = 1, 2, \cdots$, *such that* $\lim_{n\to\infty} \varepsilon_n = 0$ *and that*

$$|\Delta_{ni}(t) - \Delta_{ni}^0(t)| < M\varepsilon_n, \qquad i = 1, \cdots, N, \tag{1.26}$$

$$|\rho_i(t) - \rho_{ni}^0(t)| < M\varepsilon_n, \qquad i = 1, \cdots, N. \tag{1.27}$$

Proof. We can choose numbers ε_n $(= \varepsilon_n(t))$, $n = 1, 2, \cdots$, by the assumptions (b.1) and (b.2), such that $\lim_{n\to\infty} \varepsilon_n = 0$ and that

$$|r_{ii}(t,t) - r_{ii}^n(t,t)| \leq \varepsilon_n, \qquad i = 1, \cdots, N, \tag{1.28}$$

$$\sum_{i,j}\int_0^t\int_0^t |r_{ij}(u,v) - r_{ij}^n(u,v)|^2\, m_i(du)m_j(dv) \leq \varepsilon_n, \tag{1.29}$$

$$\sum_j \int_0^t |r_{ij}(t,u) - r_{ij}^n(t,u)|^2\, m_j(du) \leq \varepsilon_n, \qquad i = 1, \cdots, N. \tag{1.30}$$

Now we will prove the existence of a constant M_1 $(= M_1(t)) < \infty$ such that

$$|\Delta_{ni}(t) - \tilde{\Delta}_{ni}^0(t)| \leq M_1\varepsilon_n \quad \text{and} \quad |\tilde{\Delta}_{ni}(t) - \Delta_{ni}^0(t)| \leq M_1\varepsilon_n, \qquad i = 1, \cdots, N. \tag{1.31}$$

By the definitions, $\tilde{\Delta}_{ni}^0(t)$ and $\Delta_{ni}(t)$ can be written in the form,

$$\begin{aligned}\tilde{\Delta}_{ni}^0(t) = r_{ii}(t,t) &- 2\sum_j \int_0^t H_{ij}^n(t,u)A_{nj}(u)r_{ij}(t,u)m_j(du)\\ &+ \sum_{j,k}\int_0^t\int_0^t H_{ij}^n(t,u)H_{ik}^n(t,v)A_{nj}(u)r_{jk}(u,v)A_{nk}(v)m_j(du)m_k(dv)\\ &+ \sum_j \int_0^t [H_{ij}^n(t,u)]^2 m_j(du)\end{aligned}$$

and

https://doi.org/10.1017/S0027763000016718 Published online by Cambridge University Press

$$\Delta_{ni}(t) = r^n_{ii}(t,t) - 2\sum_j \int_0^t H^n_{ij}(t,u)A_{nj}(u)r^n_{ij}(t,u)m_j(du)$$
$$+ \sum_{j,k}\int_0^t\int_0^t H^n_{ij}(t,u)H^n_{ik}(t,v)A_{nj}(u)r^n_{jk}(u,v)A_{nk}(v)m_j(du)m_k(dv)$$
$$+ \sum_j \int_0^t [H^n_{ij}(t,u)]^2 m_j(du)\ .$$

So we get the inequality:

$$|\tilde{\Delta}^0_{ni}(t) - \Delta_{ni}(t)| \le I_1 + 2I_2 + I_3\ ,$$

where

$$I_1 = |r_{ii}(t,t) - r^n_{ii}(t,t)|\ ,$$
$$I_2 = \sum_j \left|\int_0^t H^n_{ij}(t,u)A_{nj}(u)[r_{ij}(t,u) - r^n_{ij}(t,u)]m_j(du)\right|$$

and

$$I_3 = \sum_{j,k}\left|\int_0^t\int_0^t H^n_{ij}(t,u)H^n_{ik}(t,v)A_{nj}(u) \times [r_{jk}(u,v) - r^n_{jk}(t,u)]A_{nk}(v)m_j(du)m_k(dv)\right|.$$

We have $I_1 \le \varepsilon_n$ from (1.28). By the use of (ii) of Lemma 1 and the assumptions (b.3) and (b.4) we get

$$A^2_{ni}(t) \le [r_{ii}(t,t)]^{-1}\rho_i(t)\exp\left[\sum_i \int_0^t \rho_i(u)m_i(du)\right] \le \alpha_1^{-1}KL\ ,$$

and we get

$$\sum_j \int_0^t [H^n_{ij}(t,u)]^2 m_j(du) \le E\,|\hat{\xi}_{ni}(t)|^2 \le r_{ii}(t,t) \le \alpha_2\ ,$$

from (1.21). Therefore it holds that

$$I_2^2 = \left\{\sum_j \left|\int_0^t H^n_{ij}(t,u)A_{nj}(u)[r_{ij}(t,u) - r^n_{ij}(t,u)]\,m_j(du)\right|\right\}^2$$
$$\le \sum_j \int_0^t [H^n_{ij}(t,u)]^2 A^2_{nj}(u)m_j(du)\sum_j\int_0^t [r_{ij}(t,u) - r^n_{ij}(t,u)]^2 m_j(du)$$
$$\le \frac{\alpha_2}{\alpha_1}KL\varepsilon_n\ , \qquad \left(L = \exp\left[\sum_j \int_0^t \rho_i(u)\,m_i(du)\right]\right),$$

and in the same way, it holds that $I_3^2 \le ((\alpha_2/\alpha_1)KL)^2\varepsilon_n$. We obtain the inequality $|\tilde{\Delta}^0_{ni}(t) - \Delta_{ni}(t)| \le M_1\varepsilon_n$, putting $M_1 = 1 + 2\sqrt{(\alpha_2/\alpha_1)KL} + (\alpha_2/\alpha_1)KL$. The other inequality of (1.31) can be obtained in the same manner.

The desired inequality (1.26) follows from (1.25) and (1.31). Finally,

https://doi.org/10.1017/S0027763000016718 Published online by Cambridge University Press

the inequality (1.27) is derived from the inequality,

$$|\rho_i(t) - \rho_{ni}^0(t)| = |A_{ni}^2(t)\Delta_{ni}(t) - A_{ni}^2(t)\Delta_{ni}^0(t)| \leq \alpha_1^{-1}KLM_1\varepsilon_n \ .$$

Now we proceed to

Proof of Theorem 1. By (1.26), it holds that

$$\lim_{n\to\infty} |\Delta_{ni}(t) - \Delta_{ni}^0(t)| = 0 \ . \tag{1.32}$$

On the other hand, using (1.27) we have the following relation by Lemma 3,

$$\lim_{n\to\infty} \Delta_{ni}^0(t) = \Delta_i(t) \ . \tag{1.33}$$

Thus we have, from (1.32) and (1.33), the result:

$$\lim_{n\to\infty} \Delta_{ni}(t) = \Delta_i(t) \ .$$

§ 2. Evaluation of the filtering error.

We can take stepwise processes as the approximating processes ξ_n's in Theorem 1. In this section, we give a method to evaluate the minimal filtering error $\Delta(t)$ for such a stepwise process in a special case where the noise $X(\cdot) = B(\cdot)$ is a standard Wiener process (Theorem 2). Consequently, we can evaluate the minimal filtering error for any Gaussian process, using Theorem 1 and Theorem 2. The method presented in this section is applicable to the case of multiplicity one.

Let us assume, throughout this section, that the noise $X(\cdot) = B(\cdot)$ be a standard Wiener process. Then the message $\xi(\cdot)$ of (a.2) is a Gaussian process such that $0 < \int_0^T E\xi^2(t)dt < \infty$, and the optimal coding for $\xi(\cdot)$ can be presented by

$$Y(t) = \int_0^t A(u)(\xi(u) - \hat{\xi}(u))du + B(t) \ , \tag{2.1}$$

$$A^2(t)\Delta(t) = \rho(t) \ , \tag{2.2}$$

where

$$\Delta(t) = E\,|\xi(t) - \hat{\xi}(t)|^2 \ , \qquad \hat{\xi}(t) = E[\xi(t)\,|\,\mathscr{F}_t(Y)] \ , \tag{2.3}$$

and $\rho(t)$ is a given positive function.

Remark. In the transmission of the Gaussian message

$$\xi(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\xi_i(u)m_i(du)$$

over the channel (1.8), if we assume that $\xi_1(\cdot), \cdots, \xi_N(\cdot)$ are mutually independent, then

$$Y_i(t) = \int_0^t F_i(t, u)A_i(u)(\xi_i(u) - \hat{\xi}_i(u))m_i(du) + B_i(t) \,, \qquad (i = 1, \cdots, N)$$

are mutually independent and

$$\hat{\xi}_i(t) = E[\xi_i(t) \,|\, \mathscr{F}_t(Y)] = E[\xi_i(t) \,|\, \mathscr{F}_t(Y_i)] \,.$$

Therefore, for the evaluation of

$$\Delta(t) = \sum_{i=1}^{N} \Delta_i(t) = \sum_{i=1}^{N} E\,|\xi_i(t) - \hat{\xi}_i(t)|^2 \,,$$

it is enough to consider the case of multiplicity one.

As for the minimal filtering error $\Delta(t)$ of (2.3), the following examples are known.

EXAMPLE 1. ([5], [10]). Let $\xi(t) = \theta$, $0 < t \leq T$, be a deterministic process, where θ is a Gaussian random variable with distribution $N(0, \gamma)$. Then

$$\Delta(t) = \gamma \exp\left[-\int_0^t \rho(u)du\right] \,, \qquad 0 \leq t \leq T \,. \tag{2.4}$$

EXAMPLE 2. (Liptzer-Shiryaev [10], [11]). Let $\xi(\cdot)$ be a Gaussian Markov process presented by

$$d\xi(t) = a(t)\xi(t)dt + b(t)dW(t) \quad \text{and} \quad E\xi(0) = 0 \,, \; E\xi^2(0) = \gamma > 0 \,,$$

where $W(\cdot)$ is a standard Wiener process independent of $B(\cdot)$. Then $\Delta(t)$ is given by

$$\Delta(t) = \gamma \exp\left[2\int_0^t a(s)ds\right] \exp\left[-\int_0^t \rho(s)ds\right] + \int_0^t b^2(s) \exp\left[2\int_s^t a(u)du\right] \exp\left[-\int_s^t \rho(u)du\right] ds \,.$$

We give a lemma which is used in later.

LEMMA 5. (i) *The process* $Y(\cdot)$ *given by* (2.1) *is a standard Wiener process.*

(ii) *The equation* (2.1) *can be rewritten in the form,*

https://doi.org/10.1017/S0027763000016718 Published online by Cambridge University Press

(2.5) $$\tilde{Y}(t) = \int_{t_0}^{t} A(u)(\tilde{\xi}(u) - \hat{\tilde{\xi}}(u))du + \tilde{B}(t) , \qquad t \geq t_0 \,(> 0) ,$$

where $\tilde{Y}(t) = Y(t) - Y(t_0)$, $\tilde{B}(t) = B(t) - B(t_0)$, $\tilde{\xi}(u) = \xi(u) - E[\xi(u) | \mathscr{F}_{t_0}(Y)]$ *and* $\hat{\tilde{\xi}}(u) = E[\tilde{\xi}(u) | \mathscr{F}_u(\tilde{Y})]$.

Proof. (i) It is easily shown that $Y(\cdot)$ is the innovation process for the process $Y_0(t) = \int_0^t A(u)\xi(u)du + B(t)$ (cf. Kailath [8]).

(ii) It is enough to show that $\tilde{\xi}(u) - \hat{\tilde{\xi}}(u) = \xi(u) - \hat{\xi}(u)$ $(u \geq t_0)$. Since $Y(s)$, $s \leq t_0$, is independent of $\tilde{Y}(t) = Y(t) - Y(t_0)$, $t \geq t_0$, $\mathscr{F}_{t_0}(Y)$ is independentof $\mathscr{F}_u(\tilde{Y})$ and then $\mathscr{F}_u(Y) = \mathscr{F}_{t_0}(Y) \vee \mathscr{F}_u(\tilde{Y})$ $(u \geq t_0)$. Therefore, $\hat{\xi}(u) = E[\xi(u) | \mathscr{F}_u(Y)] = E[\xi(u) | \mathscr{F}_{t_0}(Y)] + E[\xi(u) | \mathscr{F}_u(\tilde{Y})]$ and $\xi(u) - \hat{\xi}(u) = \tilde{\xi}(u) - E[\xi(u) | \mathscr{F}_u(\tilde{Y})] = \tilde{\xi}(u) - \hat{\tilde{\xi}}(u)$.

Remark. (i) The property (i) of Lemma 5 was first pointed out by Liptzer-Shiryaev [10].

(ii) The property (ii) of Lemma 5 makes it possible to treat t_0 as the starting point of the channel, by replacing the message ξ by $\tilde{\xi}$.

Now we consider a stepwise process. Let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of the interval $[0, T]$. And let $\xi(\cdot)$ be a Gaussian process such that

(2.6) $$\xi(t) = \theta_i , \qquad t_{i-1} \leq t < t_i ,$$

where $(\theta_1, \cdots, \theta_n)$ is a system of Gaussian random variables with

$$E\theta_i = 0 \quad \text{and} \quad E\theta_i\theta_j = r_{ij} , \quad i, j = 1, \cdots, N .$$

Then we can give a formula to calculate the error $\Delta(t)$ for the process $\xi(\cdot)$.

THEOREM 2. *Let* $\xi(\cdot)$ *be the Gaussian process given by* (2.6). *Then,*

(2.7) $$\Delta(t) = \gamma_{kk-1} \exp\left[-\int_{t_{k-1}}^{t} \rho(u)du\right] , \qquad t_{k-1} \leq t < t_k, \ k = 1, \cdots, n ,$$

where γ_{kk-1} *is determined by the following equations*:

(2.8) $$\gamma_{kk-1} = \begin{cases} r_{kk} - \sum_{i=1}^{k-1} \lambda_{ki} , & k = 2, 3, \cdots, n-1 , \\ r_{11} , & k = 1 , \end{cases}$$

(2.9) $$\lambda_{k\ell} = a_{k\ell}^2 \gamma_{\ell\ell-1}\left[1 - \exp\left[-\int_{t_{\ell-1}}^{t_\ell} \rho(u)du\right]\right] , \qquad \ell = 1, \cdots, k ,$$

$$(2.10)\qquad a_{k\ell} = \begin{cases} \gamma_{\ell\ell-1}^{-1}\left(r_{k\ell} - \sum_{i=1}^{\ell-1} a_{ki}a_{\ell i}\lambda_{ii}\right), & \ell = 1, \cdots, k-1\,, \\ 1\,, & \ell = k\,. \end{cases}$$

Remark. γ_{kk-1} is uniquely determined by (2.8)–(2.10), inductively.

Proof. For the process $Y(\cdot)$ given by (2.1) and (2.2), we define a process $Y_k(t)$ by (2.5), replacing t_0 by t_{k-1}:

$$(2.11)\qquad Y_k(t) = \int_{t_{k-1}}^{t} A(u)(\xi_k(u) - \hat{\xi}_k(u))du + B_k(t)\,, \qquad t_{k-1} \le t < t_k\,,$$

where $Y_k(t) = Y(t) - Y(t_{k-1})$, $B_k(t) = B(t) - B(t_{k-1})$, $\xi_k(u) = \xi(u) - E[\xi(u)\,|\,\mathscr{F}_{t_{k-1}}(Y)]$ and $\hat{\xi}_k(u) = E[\xi_k(u)\,|\,\mathscr{F}_u(Y_k)]$ $(u \ge t_{k-1})$. Define random variables $\hat{\theta}_{k\ell}$ and $\theta_{k\ell}$ by

$$(2.12)\qquad \hat{\theta}_{k\ell} = E[\theta_{k\ell-1}\,|\,\mathscr{F}(Y_\ell)]\,, \qquad \ell = 1, \cdots, k\,,$$

$$(2.13)\qquad \theta_{k\ell} = \begin{cases} \theta_{k\ell-1} - \hat{\theta}_{k\ell}\,, & \ell = 1, \cdots, k\,, \\ \theta_k\,, & \ell = 0\,, \end{cases}$$

where $\mathscr{F}(Y_\ell) = \mathscr{F}_{t_\ell}(Y_\ell)$, and put

$$(2.14)\qquad \gamma_{k\ell} = E\,|\theta_{k\ell}|^2\,, \quad \lambda_{k\ell} = E\,|\hat{\theta}_{k\ell}|^2\,, \quad \ell = 1, \cdots, k\,.$$

Then it follows from Lemma 5 that $\hat{\theta}_{k\ell}$, $\ell = 1, \cdots, k$, are mutually independent and that

$$(2.15)\qquad E[\theta_k\,|\,\mathscr{F}_{t_\ell}(Y)] = \sum_{i=1}^{\ell} \hat{\theta}_{ki}\,, \qquad \ell = 1, \cdots, k\,.$$

Since

$$\theta_{kk-1} = \theta_{kk-2} - \hat{\theta}_{kk-1} = \cdots = \theta_k - \sum_{i=1}^{k-1} \hat{\theta}_{ki} = \theta_k - E[\theta_k\,|\,\mathscr{F}_{t_{k-1}}(Y)]\,,$$

(2.11) can be rewritten as follows:

$$(2.16)\qquad Y_k(t) = \int_{t_{k-1}}^{t} A(u)(\theta_{kk-1} - \hat{\theta}_{kk-1}(u))du + B_k(t)\,, \qquad t_{k-1} \le t < t_k\,,$$

where $\hat{\theta}_{kk-1}(u) = E[\theta_{kk-1}\,|\,\mathscr{F}_u(Y_k)]$.

Using (ii) of Lemma 5 and (2.4), it it concluded that the error $\Delta(t)$ is given by (2.7).

From (2.14), (2.15) and (2.16), γ_{kk-1} is given by (2.8), since $\hat{\theta}_{ki}$, $i = 1, \cdots, k$, are mutually independent.

To prove (2.9) and (2.10) completes the proof of Theorem. Define constants $a_{k\ell}$'s by

(2.17) $$\theta_{k\ell-1} = a_{k\ell}\theta_{\ell\ell-1} + \varphi_{k\ell}\,, \qquad \ell = 1, \cdots, k\,,$$

where $\varphi_{k\ell}$ is independent of $\theta_{\ell\ell-1}$. Since $\varphi_{k\ell}$ is independent of $Y_\ell(\cdot)$ and

$$E\,|\theta_{\ell\ell-1} - \hat{\theta}_{\ell\ell}|^2 = \gamma_{\ell\ell-1}\exp\left[-\int_{t_{\ell-1}}^{t_\ell}\rho(u)du\right]$$

(by (2.4)), we have

$$\lambda_{k\ell} = E\,|\hat{\theta}_{k\ell}|^2 = a_{k\ell}^2 E\,|\hat{\theta}_{\ell\ell}|^2 = a_{k\ell}^2\gamma_{\ell\ell-1}\left[1 - \exp\left(-\int_{t_{\ell-1}}^{t_\ell}\rho(u)du\right)\right].$$

Thus (2.9) is proved. In order to prove (2.10), we put

$$\alpha_{k\ell}(m) = E[\theta_{km}\theta_{\ell m}]\,, \qquad 0 \leq m \leq \ell \leq k\,.$$

Then, by (2.12), (2.17) and (2.14),

(2.18)
$$\begin{aligned}\alpha_{k\ell}(m) &= E[(\theta_{km-1} - \hat{\theta}_{km})(\theta_{\ell m-1} - \hat{\theta}_{\ell m})]\\ &= \alpha_{k\ell}(m-1) - a_{km}a_{\ell m}E\,|\hat{\theta}_{mm}|^2 = \cdots\\ &= \alpha_{k\ell}(0) - \sum_{j=1}^{m} a_{kj}a_{\ell j}\lambda_{jj} = r_{k\ell} - \sum_{j=1}^{m} a_{kj}a_{\ell j}\lambda_{jj}\,.\end{aligned}$$

On the other hand, from (2.17), it follows that

(2.19) $$\alpha_{k\ell}(\ell-1) = E[\theta_{k\ell-1}\theta_{\ell\ell-1}] = a_{k\ell}E\,|\theta_{\ell\ell-1}|^2 = a_{k\ell}\gamma_{\ell\ell-1}\,.$$

The relation (2.10) follows from (2.18) and (2.19).

Theorem 1 and Theorem 2 give us a method to evaluate the minimal filtering error $\Delta(t)$.

THEOREM 3. *Let* $\xi(t), 0 \leq t \leq T$, *be a zero mean Gaussian process such that* $r(t,s) = E\xi(t)\xi(s)$ *is continuous in* (t,s) *and that* $r(t,t) \neq 0$ *for all* t. *Let* $0 = t_{n0} < t_{n1} < \cdots < t_{nk_n} = T\ (n = 1, 2, \cdots)$ *be a partition of* $[0, T]$ *such that* $\max_{1\leqslant k\leqslant k_n}|t_{nk} - t_{nk-1}| \to 0$ *as* $n \to \infty$, *and define a Gaussian process* $\xi_n(\cdot)$ *by*

$$\xi_n(t) = \xi(t_{nk}) \quad \textit{if} \quad t_{nk} \leq t < t_{nk+1}\,.$$

If we denote by $\Delta_n(t)$ *the minimal filtering error for* $\xi_n(\cdot)$, *then, under the assumption* (b.4), $\Delta_n(t)$ *is given by* (2.7)–(2.10) (*replacing* r_{ij} *by* $r_{ij}^n = E[\xi_n(t_{ni})\xi_n(t_{nj})]$). *And the minimal filtering error* $\Delta(t)$ *for* $\xi(\cdot)$ *is given by*

(2.20) $$\Delta(t) = \lim_{n\to\infty}\Delta_n(t)\,.$$

Proof. We can easily show that the conditions (b.1)–(b.3) are satisfied, because $r(t,s)$ is continuous and $r(t,t) \neq 0$. Thus we can apply

Theorem 2 and we know that $\Delta_n(t)$ is given by (2.7)–(2.10). And we get (2.20) from Theorem 1.

§ 3. Filtering error under non-causal coding.

We take into account non-causal codings in this section, while we treated only causal codings in the preceding sections. We will show that we can construct a non-causal coding which is optimal in information sense and, at the same time, in filtering sense, and that the minimal filtering error is determined by the ε-entropy of the message and the channel capacity (Theorem 4).

Atfirst we will interpret the messages and the codings to be considered in this section. Let the message $\xi(t) = (\xi_1(t), \cdots, \xi_M(t))$, $0 \leq t \leq S\ (< \infty)$ be a M-dimensional Gaussian process such that

$$E\xi_i(t) = 0\ , \qquad E[\xi_i(s)\xi_j(t)] = r_{ij}(s, t)$$

and that

$$\sum_{i=1}^{M} \int_0^S E\xi_i^2(t)\nu_i(dt) < \infty\ ,$$

where ν_i's are measures on $[0, S]$. The ε-entropy $H_\varepsilon(\xi)$ of $\xi(\cdot)$ is defined (cf. [9], [13]) by

$$H_\varepsilon(\xi) = \inf I(\xi, \eta)\ , \qquad (\varepsilon > 0)\ , \tag{3.1}$$

where the infimum is taken over all processes $\eta(t) = (\eta_1(t), \cdots, \eta_M(t))$, $0 \leq t \leq S$, such that

$$\sum_{i=1}^{M} \int_0^S E\,|\xi_i(t) - \eta_i(t)|^2\,\nu_i(dt) \leq \varepsilon^2\ . \tag{3.2}$$

On the other hand, the channel and the codings are as follows:

$$Y(t) = \Phi(t) + X(t)\ , \qquad 0 \leq t \leq T\ , \tag{3.3}$$

where the noise $X(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)dB_i(u)$ is of (1.1) and the channel input $\Phi(t)$ is a functional of $\{\xi(s)\,; 0 \leq s \leq S\}$ and $\{Y(u)\,; 0 \leq u \leq t\}$. Note that $\Phi(\cdot)$ can be non-causal in the message $\xi(\cdot)$. Let us assume that $\Phi(\cdot)$ can be written in the form,

$$\Phi(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)\Phi_i(u)m_i(du)\ , \quad \text{with probability one}\ , \tag{3.4}$$

where $\int_0^T \Phi_i^2(t)m_i(dt) < \infty$ (with probability one), and assume that the average power constraint is imposed on $\Phi(\cdot)$:

$$E\Phi_i^2(t) \leq \rho_i(t)\,, \qquad 0 \leq t \leq T\,, \quad i = 1, \cdots, N\,, \tag{3.5}$$

where ρ_i's are given positive functions with $\sum_{i=1}^{N}\int_0^T \rho_i(u)m_i(du) < \infty$. Then the capacity C_t of the channel (3.3)–(3.5) is given by (1.6).

From the definitions, we know that if

$$H_\varepsilon(\xi) > C \ (= C_T)\,,$$

then their is no method of coding which transmit $\xi(\cdot)$ with reproduction accuracy ε. The Shannon's fundamental problem ([14]) is that if

$$H_\varepsilon(\xi) \leq C$$

then is it possible to construct a coding which transmit $\xi(\cdot)$ with reproduction accuracy ε. The first result of this section is to construct such a coding.

THEOREM 4. *Assume that*

$$H_\varepsilon(\xi) = C_T\,, \tag{3.6}$$

then there exists a coding method $\Phi^*(t) = \sum_{i=1}^{N}\int_0^t F_i(t, u)\Phi_i^*(u)m_i(du)$, *satisfying* (3.5), *such that*

$$\sum_{i=1}^{M}\int_0^S E\,|\xi_i(t) - \eta_i^*(t)|^2\,\nu_i(dt) = \varepsilon^2\,, \tag{3.7}$$

where

$$\eta_i^*(t) = E[\xi_i(t)\,|\,\mathscr{F}_T(Y^*)] \quad \textit{and} \quad Y^*(t) = \Phi^*(t) + X(t)\,.$$

Remark. In case where $M = N = 1$ and $X(\cdot)$ is a standard Wiener process, the result has been obtained by Ovseevich [12].

In order to prove Theorem 4 we prepare two lemmas. The one is concerning the ε-entropy and the other is concerning the optimal coding for a deterministic process.

A formula to calculate the ε-entropy can be obtained in the same manner as in 1-dimensional case (cf. [9], [13]). Let $\mathscr{L}$ be a Hilbert space

$$\mathscr{L} = \{\varphi(t) = (\varphi_1(t), \cdots, \varphi_M(t))\,;\, \varphi_i \in L^2([0, S], \nu_i)\}$$

with an inner product

$$(\varphi, \psi) = \sum_{i=1}^{M} \int_0^S \varphi_i(t)\psi_i(t)\nu_i(dt) .$$

Then the covariance operator R on $\mathscr{L}$, given by

$$(R\varphi)_i(t) = \sum_{j=1}^{M} \int_0^S r_{ij}(t, s)\varphi_j(s)\nu_j(ds) , \qquad \varphi \in \mathscr{L} ,$$

is symmetric, positive definite and Hilbert-Schmidt type. Denote by $\{\lambda_k\}$ the eigenvalues of R, then we have

LEMMA 6. *The ε-entropy $H_\varepsilon(\xi)$ is given by*

(3.8) $$H_\varepsilon(\xi) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(\frac{\lambda_k}{a} \vee 1\right) ,$$

where a is a constant determined by

$$\sum_{k=1}^{\infty} (\lambda_k \wedge a) = \varepsilon^2 .$$

Let θ be a Gaussian random variable with $E\theta = 0$ and $E\theta^2 = \gamma > 0$. Then it is known ([9], [13]) that

(3.9) $$H_\varepsilon(\theta) = \inf \{I(\theta, \tilde{\theta})\,;\, E\,|\theta - \tilde{\theta}|^2 \leq \varepsilon^2\} = \frac{1}{2} \log \left(\frac{\gamma}{\varepsilon^2} \vee 1\right) .$$

By Lemma 1, the optimal coding (in Φ of (1.4)) for θ is given by

(3.10) $$Y(t) = \sum_{i=1}^{N} \int_0^t F_i(t, u)A_i(u)(\theta - \hat{\theta}(u))m_i(du) + X(t) ,$$

(3.11) $$A_i^2(t)E\,|\theta - \hat{\theta}(t)|^2 = \rho_i(t) ,$$

where $\hat{\theta}(t) = E[\theta\,|\,\mathscr{F}_t(Y)]$.

LEMMA 7. *Under the optimal coding* (3.10) *and* (3.11), *if we put*

$$\varepsilon^2(t) = E\,|\theta - \hat{\theta}(t)|^2 ,$$

then the following relation holds:

(3.12) $$H_{\varepsilon(t)}(\theta) = C_t , \qquad 0 \leq t \leq T .$$

And then the error $\varepsilon^2(t)$ is

(3.13) $$\varepsilon^2(t) = \gamma \exp \left[-\sum_{i=1}^{N} \int_0^t \rho_i(u)m_i(du)\right] .$$

Proof. By Corollary of Lemma 2, it holds that

(3.14) $$h_{ii}(t,t) = A_i^2(t)E\,|\theta - \hat{\theta}(t)|^2 = A_i^2(t)\varepsilon^2(t) = \rho_i(t)\ ,$$

where $h_{ij}(t,s)$ is the solution of the equation

(3.15) $$A_i(t)\gamma A_j(s) = \sum_{k=1}^{N}\int_0^t h_{ik}(t,u)A_k(u)\gamma A_j(s)m_k(du) + h_{ij}(t,s)\ , \qquad s \le t\ .$$

Noting that h_{ij} must be in the form $h_{ij}\ (t,s) = \gamma H_i(t)A_j(s)$, we have, from (3.14) and (3.15), that

(3.16) $$\varepsilon^2(t) = \gamma\left[\gamma\sum_{k=1}^{N}\int_0^t A_k^2(u)m_k(du) + 1\right]\ .$$

According to (3.9) we have

(3.17) $$H_{\varepsilon(t)}(\theta) = \tfrac{1}{2}\log\left[\gamma\sum_{k=1}^{N}\int_0^t A_k^2(u)m_k(du) + 1\right]\ .$$

On the other hand, it is easily shown, from (1.6), (3.14) and (3.16), that

(3.18) $$C_t = \tfrac{1}{2}\sum_{i=1}^{N}\int_0^t \gamma A_i^2(u)\left[\gamma\sum_{k=1}^{N}\int_0^u A_k^2(v)m_k(dv) + 1\right]^{-1} m_i(du)\ .$$

The desired equation (3.12) is obtained by differentiating the right hand sides of (3.17) and (3.18). And finally (3.13) is derived from (1.6), (3.9) and (3.12).

By the use of Lemma 6 and Lemma 7, Theorem 4 can be proved in the same manner as in [12]. So we show only the outline of the proof.

Outline of the proof of Theorem 4. The desired coding method

$$\Phi^*(t) = \sum_{i=1}^{N}\int_0^t F_i(t,u)\Phi_i^*(u)m_i(du)$$

is given by the following progresses (i)–(iv):

(i) Define constants a and L, for ε in (3.6), by the equation

$$\varepsilon^2 = \sum_{k=1}^{\infty}(\lambda_k \wedge a) = La + \sum_{k>L}\lambda_k$$

(we may assume that λ_k's are arranged in decreasing order).

(ii) Define $0 < T_1 < \cdots < T_L$ by the equations

$$\sum_{i=1}^{N}\int_{T_{k-1}}^{T_k}\rho_i(u)m_i(du) = \log\frac{\lambda_k}{a}\ , \qquad k = 1,\cdots,L\ ,$$

then it can be shown that $T_L = T$.

(iii) Define random variables ξ_k, $k = 1, 2, \cdots$ by

https://doi.org/10.1017/S0027763000016718 Published online by Cambridge University Press

$$\xi_k = \sum_{i=1}^{M} \int_0^S \xi_i(t)\varphi_{ki}(t)\nu_i(dt) \ ,$$

where $\varphi_k(t) = (\varphi_{k1}(t), \cdots, \varphi_{kM}(t)) \in \mathscr{L}$ is an eigenfunction of R corresponding to λ_k such that $\{\varphi_k\}$ forms a complete orthonormal system of $\mathscr{L}$. And define processes $\tilde{\xi}_i(\cdot), i = 1, \cdots, N$ by

$$\tilde{\xi}_i(t) = \xi_k \quad \text{if} \quad T_{k-1} < t \leq T_k, \quad i = 1, \cdots, N \ .$$

(iv) The coding $\Phi_i^*, i = 1, \cdots, N$ is given by

$$\Phi_i^*(t) = A_i(u)(\tilde{\xi}_i(t) - \hat{\tilde{\xi}}_i(t)) \ ,$$

where $A_i(t)$ and $\hat{\tilde{\xi}}_i$ are determined by

$$E\,|\Phi_i^*(t)|^2 = A_i^2(t)E\,|\tilde{\xi}_i(t) - \hat{\tilde{\xi}}_i(t)|^2 = \rho_i(t)$$
$$\hat{\tilde{\xi}}_i(t) = E[\tilde{\xi}_i(t)\,|\,\mathscr{F}_u(Y^*)] \ , \qquad Y^*(t) = \Phi^*(t) + X(t) \ .$$

Then we can show (3.7), applying Lemma 7 to ξ_k on each time interval $(T_{k-1}, T_k]$.

From Theorem 4 we can get the following fact in the same manner as in [5].

COROLLARY. *Let $\xi(\cdot) = (\xi_1(\cdot), \cdots, \xi_M(\cdot))$ be a M-dimensional Gaussian message. Then in all non-causal codings of* (3.4) *and* (3.5), *the coding Φ^* in Theorem* 4 *minimizes the error*

$$\sum_{i=1}^{M} \int_0^S E\,|\xi_i(t) - \hat{\hat{\xi}}_i(t)|^2\, \nu_i(dt) \ , \tag{3.19}$$

where $\hat{\hat{\xi}}_i(t) = E[\xi_i(t)\,|\,\mathscr{F}_T(Y)]$. And then the minimal filtering error ε^2 is given by (3.6).

Proof. Suppose that there exists another coding Ψ such that $\delta^2 = \sum_{i=1}^{M} \int_0^S E\,|\xi_i(t) - \zeta_i(t)|^2\, \nu_i(dt) \lneqq \varepsilon^2$, where $\zeta_i(t) = E[\xi_i(t)\,|\,\mathscr{F}_T(Z)]$ and $Z(t) = \Psi(t) + X(t)$. Then by the definitions we have

$$H_\varepsilon(\xi) \lneqq H_\delta(\xi) \leq I(\xi, \zeta) \leq C_T \tag{3.20}$$

$\Big($where we use the fact that the ε-entropy $H_\varepsilon(\xi)$ is strictly decreasing in $\varepsilon^2\Big(< \sum_{i=1}^{M} \int_0^S E\xi_i^2(t)\nu_i(dt)\Big)\Big)$. The inequality contradicts the equality (3.6).

Corollary asserts that, if we take into account non-causal codings the minimal filtering error is determined only by the sum $P(T) =$

https://doi.org/10.1017/S0027763000016718 Published online by Cambridge University Press

$\sum_{i=1}^{N}\int_0^T \rho_i(t)m_i(dt)$, since the capacity C_T depends only upon $P(T)$. But if we treat only causal codings $\Phi \in \boldsymbol{\Phi}$ (of (1.4)), such a fact is not expected.

To illustrate such a situation we give an example.

EXAMPLE. Let $\xi(\cdot) = (\xi_1(\cdot), \cdots, \xi_N(\cdot))$ be a N-dimensional process such that $\xi_i(t) = \theta_i$, $0 \le t \le T$, where $\theta_1, \cdots, \theta_N$ are mutually independent and the distribution of θ_i is $N(0, \gamma_i)$ (suppose that $\gamma_{i_1} \ge \gamma_{i_2} \ge \cdots \ge \gamma_{i_N}$). Then it is easily shown that the minimal filtering error $\Delta(t) = \sum_{i=1}^N \Delta_i(t)$ is not determined only by the sum $P(t) = \sum_{i=1}^{N}\int_0^t \rho_i(u)m_i(du)$. Here we assume that the sum $P(t)$, $0 \le t \le T$, is given. Then the method, to choose $\rho_i(t)$ such that $\Delta(t)$ is minimized, is as follows:

$$\rho_{i_k}(t) = \begin{cases} \dfrac{1}{\ell}\tilde{\rho}(t)\,, & k = 1, \cdots, \ell\,, \\ 0\,, & k = \ell + 1, \cdots, N\,, \end{cases} \qquad \text{if } t_{\ell-1} < t \le t_\ell$$

where $\tilde{\rho}_i(t) = \rho_i(t)(m_i(dt)/m_1(dt))$, $\tilde{\rho}(t) = \sum_{i=1}^N \tilde{\rho}_i(t)$ and t_ℓ's are determined by

$$\gamma_{i_{\ell+1}} = \gamma_{i_\ell} \exp\left[-\frac{1}{\ell}\int_{t_{\ell-1}}^{t_\ell} \tilde{\rho}(u)m_1(du)\right].$$

The proof follows from Lemma 7 and the proof of Theorem 4.

REFERENCES

[1] T. Hida, Canonical representation of Gaussian processes and their application, Mem. Coll. Sci. Univ. Kyoto, **33** (1960), 109–155.

[2] M. Hitsuda, Multiplicity of some classes of Gaussian processes, Nagoya Math. J., **52** (1973), 39–46.

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

[4] M. Hitsuda, S. Ihara, Gaussian channels and the optimal coding, J. Mult. Anal., **5** (1975), 106–118.

[5] S. Ihara, Optimal coding in white Gaussian channel with feedback, Proc. Second Japan-USSR Symp. on Prob. Theory, (Lecture Notes in Math. 330), pp. 120–123, Springer-Verlag, 1973.

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

[7] S. Ihara, Coding theory in Gaussian channel with feedback, Trans. 7-th Prague Conference on Inform. theory, Stat. Dec. Functions, and Random Processes, to appear.

[8] T. Kailath, The innovations approach to detection and estimation theory, Proc. IEEE, **58** (1970), 680–695.

[9] A. N. Kolmogolov, Theory of transmission of information. A. M. S. Translations Ser. 2, **33** (1963), 291–321.

[10] Р. Ш. Липцер, А. Н. Ширяев, Статистика случайных процессов (нелинейная фильтрация и смежные вопросы), М., Наука, 1974.
[11] Р. Ш. Липчер, Оптимальное кодирование и декодирование при передачи Гауссовского Марковского сигнала по каналу с бесшумной обратной связью, Проблемы передачи информации, **10-4** (1974), 3-15.
[12] I. A. Ovseevich, Optimum transmission of a Gaussian message over a channel with Gaussian noise and feedback, Problems of Information Transmission, **6-3** (1970), 191-199.
[13] С. М. Пинскер, Гауссовские источники. Проблемы передачи информации, **14** (1963), 59-100.
[14] C. E. Shannon, A mathematical theory of communication, Bell Syst. Thch. J., **27** (1948), 379-423, 623-656.

Department of Mathematics
Ehime University