SIMPLE PROOFS FOR THE STRONG CONVERSE THEOREMS IN SOME CHANNELS

BY KEN-ICHI YOSHIHARA

1. Summary.

In the coding theory, one of the most important problems is to establish the "capacity" of the channel considered. But, as Wolfowitz pointed out in [5], and [6], to prove that the constant C involved is the capacity, one has to prove a coding theorem and its strong converse.

In this paper, we shall prove a general strong converse theorem which is available to various channels. In Section 2, we shall prove the theorem. As applications of the theorem, we shall consider the following two problems:

(i) Another proof of the strong converse of the coding theorem for a continuous memoryless channel with additive Gaussian noise (Section 3);

(ii) The proof of the strong converse of the time-continuous Gaussian channel with additive Gaussian noise of arbitrary spectrum (Section 4).

2. A general theorem.

At first, we shall derive a theorem which is useful to prove the strong converses of various coding theorems.

Let D' (D'') be the input space (the output space). Let $h(\cdot|\cdot)$ be a channel probability function: that is, for fixed $u \in D'$, $h(\cdot|u)$ is a generalized density with respect to a (not necessarily finite) measure μ and for fixed $v \in D''$, $h(v|\cdot)$ is a measurable function with respect to another (not necessarily finite) measure λ. Let $p(\cdot)$ be a generalized probability density with respect to the measure λ. Define

$$p(u, v)=p(u)h(v|u) \tag{1}$$

and

$$q(v)=\int_{D'} p(u)h(v|u)\lambda(du). \tag{2}$$

Next, for a positive number θ, we define $A_u(\theta)$ by

$$A_u(\theta)=\left\{v \in D'' \mid \log \frac{h(v|u)}{q(v)} \leqq \theta\right\} \qquad (u \in D'). \tag{3}$$

Received June 4, 1964.

THEOREM 1. *Let α, $0\leqq\alpha<1$, be arbitrary. Let E a set of inputs and $\{(u_1, B_1), \cdots, (u_N, B_N)\}$ $(u_i \in E, i=1, \cdots, N)$ be any (N, α) code. If we can choose a positive number θ such that*

$$\int_{\overline{A_{u_i}(\theta)}} h(v|u_i)\mu(dv) \leqq \frac{1-\alpha}{2} \qquad (i=1, \cdots, N), \tag{4}$$

then N must satisfy the relation

$$N \leqq \frac{2}{1-\alpha} 2^{\theta}. \tag{5}$$

Here, the bar over a set denotes its complement.

The proof of this theorem leans heavily on the method used by J. H. B. Kemperman in the proof of the strong converse of the coding theorem for the semi-continuous memoryless channel (cf. [5]).

Proof. For brevity, let $A_i = A_{u_i}(\theta)$. For all $u_i \in E$ $(i=1, \cdots, N)$

$$1-\alpha \leqq \int_{B_i} h(v|u_i)\mu(dv) \leqq \int_{A_i \frown B_i} h(v|u_i)\mu(dv) + \int_{\overline{A_i}} h(v|u_i)\mu(dv).$$

Therefore, using (4), we have

$$\int_{A_i \frown B_i} h(v|u_i)\mu(dv) \geqq \frac{1-\alpha}{2} \qquad (i=1, \cdots, N). \tag{6}$$

Since, for any $v \in A_i$ $(i=1, \cdots, N)$,

$$q(v) \geqq 2^{-\theta} h(v|u_i),$$

so, from (6), we obtain

$$\begin{aligned} \int_{B_i} q(v)\mu(dv) &\geqq \int_{A_i \frown B_i} q(v)\mu(dv) \\ &\geqq 2^{-\theta} \int_{A_i \frown B_i} h(v|u_i)\mu(dv) \geqq \frac{1-\alpha}{2} 2^{-\theta} \qquad (i=1, \cdots, N). \end{aligned} \tag{7}$$

Thus, summing on both side from 1 to N, and using the fact that $B_i \frown B_j = \Phi$ $(i \neq j)$ and $\cup_{i=1}^N B_i \subset D''$, we have

$$N 2^{-\theta} \frac{1-\alpha}{2} \leqq \sum_{i=1}^{N} \int_{B_i} q(v)\mu(dv) = \int_{\cup_{i=1}^N B_i} q(v)\mu(dv) \leqq 1.$$

Thus, we obtain (5) and complete the proof.

3. Another proof of the strong converse of the coding theorem for a continuous memoryless channel with additive Gaussian noise.

As an application of Theorem 1, in this section, we shall show another proof of the strong converse of the coding theorem for a continuous memoryless channel with additive Gaussian noise.

Let $P>0$ be the maximum permissible input average power per coordinate. A set E of possible channel inputs is all sequences $u=(x_1, \cdots, x_n)$ of n real numbers such that

$$\frac{1}{n}\sum_{j=1}^{n} x_j^2 \leqq P. \tag{8}$$

Any sequence $v=(y_1, \cdots, y_n)$ is a possible output. Let $\gamma>0$ be the average noise power per coordinate. We assume that for the input $u=(x_1, \cdots, x_n)$ given, the output $v=(y_1, \cdots, y_n)$ is obtained by adding an independent Gaussian random variable with mean zero and variance γ, to each of the coordinates of u; that is,

$$h(v|u)=\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi\gamma}}e^{-(y_j-x_j)^2/2\gamma}. \tag{9}$$

For this channel, Shannon first proved that the capacity is $C=(1/2)\log_2[1+P/\gamma]$, and Wolfowitz showed another proof of the result, (cf. [3], [4] and [5]).

Now, we show a new method to prove the strong converse of the coding theorem for the channel.

Strong converse of the coding theorem:

Let α, $0\leqq\alpha<1$, and $\varepsilon>0$ be arbitrary. Let $\{(u_1, B_1), \cdots, (u_N, B_N)\}$ $(u_i\in E, i=1, \cdots, N)$ be a code (n, N, α) for the above defined channel. If n is sufficiently large, then

$$N<\frac{2}{1-\alpha}2^{n(C+\varepsilon)}.$$

Proof. Let ε_1 $(0<\varepsilon_1<\varepsilon)$ be arbitrary and let Q be a positive number such that

$$0<\log_e\left(1+\frac{Q}{\gamma}\right)-\log_e\left(1+\frac{P}{\gamma}\right)<\varepsilon_1.$$

For $p(u)$, we use

$$p(u)=\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi Q}}e^{-x_j^2/2Q} \tag{10}$$

and define

$$q(v)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(u)h(v|u)dx_1\cdots dx_n=\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi(Q+\gamma)}}e^{-y_j^2/2(Q+\gamma)}. \tag{11}$$

Let E be the set of all u which satisfies (8) and let

$$A_u=\left\{v=(y_1, \cdots, y_n) \mid \log_2 \frac{h(v|u)}{q(v)} \geqq n(C+\varepsilon)\right\}$$

$$=\left\{v=(y_1, \cdots, y_n) | \log_e \frac{h(v|u)}{q(v)} \geqq \frac{n}{2}\left\{\log_e\left(1+\frac{P}{\gamma}\right)+2\varepsilon \log_e 2\right\}\right\}.$$

Since, for any $u \in E$,

$$E\left[\log_e \frac{h(V|u)}{q(V)} \mid u\right]$$

$$=-\frac{nQ}{2(Q+\gamma)}+\frac{\sum_{j=1}^{n} x_j^2}{2(Q+\gamma)}+\frac{n}{2}\log_e\left(1+\frac{Q}{\gamma}\right)$$

(12)

$$\leqq -\frac{nQ}{2(Q+\gamma)}+\frac{nP}{2(Q+\gamma)}+\frac{n}{2}\log_e\left(1+\frac{Q}{\gamma}\right)$$

$$\leqq \frac{n}{2}\left\{\log_e\left(1+\frac{P}{\gamma}\right)+\varepsilon_1\right]$$

and

$$\operatorname{var}\left\{\log_e \frac{h(V|u)}{q(V)} \mid u\right\}$$

$$=\sum_{j=1}^{n} E\left[\left\{\frac{1}{2(Q+\gamma)}\left(-\frac{Q}{\gamma}(Y_j-x_j)^2+2x_j(Y_j-x_j)+Q\right)\right\}^2 | x_j\right]$$

(13)

$$=\frac{1}{4(Q+\gamma)^2}\sum_{j=1}^{n} E\left[\frac{Q^2}{\gamma^2}(Y_j-x_j)^4-\frac{4x_jQ}{\gamma}(Y_j-x_j)^3\right.$$

$$\left.+\left(4x_j^2-\frac{2Q^2}{\gamma}\right)(Y_j-x_j)^2+4x_jQ(Y_j-x_j)+Q^2 | x_j\right]$$

$$=\frac{nQ^2+2\gamma\sum_{j=1}^{n} x_j^2}{2(Q+\gamma)^2} \leqq \frac{n(Q^2+2\gamma P)}{2(Q+\gamma)^2},$$

so, by Tchebychev's inequality,

$$P\{V \in A_u | u\} \leqq \frac{2(Q^2+2\gamma P)}{n(\varepsilon-\varepsilon_1)^2(Q+\gamma)^2}.$$

Consequently, if we put $\theta=n(C+\varepsilon)$, then the condition (4) in Theorem 1 is satisfied for sufficiently large n and by the theorem, we have the desired result.

4. The proof of the strong converse of the time-continuous Gaussian channel with additive Gaussian noise of arbitrary spectrum.

(a) Definitions.

Recently, in [1], Ash proved a coding theorem and its weak converse for a time-continuous channel with additive Gaussian noise of arbitrary spectrum. In this section, we shall prove the strong converse of Ash's result as another application of Theorem 1.

Let $n(t)$ be a stationary Gaussian stochastic process with zero mean, continuous covariance function $R(\tau)$, and spectral density $N(\omega)$ with

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\{N(\omega)\}^2 d\omega<\infty. \tag{14}$$

Here, to avoid degeneracy, we assume that the eigenfunctions of the integral equation

$$\int_{-T}^{T} R(t-\tau)g(\tau)d\tau=\rho g(t) \tag{15}$$

span the entire L_2 space of square integrable functions over $[-T, T]$. Consider the class of real functions of integrable square over $(-\infty, \infty)$ whose Fourier transforms are zero whenever $N(\omega)$ is zero. If $S(\omega)$ is the Fourier transform of a function $s(t)$ in this class, let

$$F(\omega)=\frac{S(\omega)}{\sqrt{2\pi N(\omega)}} \tag{16}$$

and let $f(t)$ be the inverse Fourier transform of $F(\omega)$. (Since $F(\omega)$ is of integrable square, $F(\omega)$ has an inverse Fourier transform, at least in the sense of a limit in the mean.) For any positive real number T, define $s_T(t)$ and $n_T(t)$ as follows:

$$s_T(t)=\begin{cases} s(t) & \text{if } -T\leqq t\leqq T,\\ 0 & \text{otherwise;}\end{cases} \tag{17}$$

and

$$n_T(t)=\begin{cases} n(t) & \text{if } -T\leqq t\leqq T\\ 0 & \text{otherwise.}\end{cases} \tag{18}$$

DEFINITION 1. A function $s(t)$ (and its corresponding $s_T(t)$) will be called *allowable* if

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|S(\omega)|^2}{N(\omega)}d\omega\leqq KT \tag{19}$$

where K is a positive constant. The integrand is defined to be zero whenever $N(\omega)=0$.

DEFINITION 2. A *code* (T, M, α) for a time-continuous Gaussian channel is a set

$$\{(s_1(t), A_1), (s_2(t), A_2), \cdots, (s_M(t), A_M)\} \tag{20}$$

where each $s_j(t)$ is an allowable function $s_T(t)$ and the A_j are disjoint Borel sets in function space such that

$$P\{s_j(t)+n_T(t)\in A_j\}\geqq 1-\alpha \quad (j=1, 2, \cdots, M). \tag{21}$$

DEFINITION 3. A number R is called a *permissible rate of transmission* if for each T, there is a code $(T, [2^{RT}], \beta(T))$ such that $\beta(T)\to 0$ as $T\to\infty$. The *channel capacity* C is the supremum of all permissible transmission rate.

(b) The strong converse of the coding theorem for a time-continuous Gaussian channel.

In [1], Ash proved the coding theorem, that is,

$$C\geqq \frac{(\log_2 e)K}{2}$$

and its weak converse

$$C\leqq \frac{1}{1-2\alpha}\,\frac{(\log_2 e)K}{2}.$$

We shall now show the following strong converse of the coding theorem, which is new.

THEOREM 2. *Let α, $0\leqq\alpha<1/2$, and $\varepsilon>0$ be arbitrary. Any code (T, M, α) for the time-continuous Gaussian channel must satisfy*

$$M\leqq \frac{1}{1-2\alpha}\,2^{T((\log_2 e)K/2+\varepsilon)}. \tag{22}$$

Proof. As Ash has done, we shall prove the theorem by approximating a given coding (T, M, α) by a code for a discrete memoryless channel. Since the eigenfunctions of the integral equation

$$\int_{-T}^{T} R(t-\tau)g(\tau)d\tau=\rho g(t), \quad -T\leqq t\leqq T, \tag{23}$$

span the Hilbert space $L_2[-T, T]$, there is a one to one correspondence between square integrable functions over $[-T, T]$ and square summable sequences. The sequence $(x_1, x_2, \cdots)$ corresponding to a function $x(t)$ with respect to the "basis functions" $g_n(t)$. Thus, for each decoding set A_i in function space, there corresponds a Borel set A_i^* in sequence space such that if

$$s_i(t)=\sum_{n=1}^{\infty} s_{in}g_n(t), \quad -T\leqq t\leqq T; \qquad i=1, 2, \cdots, M, \tag{24}$$

where

$$s_{in}=\int_{-T}^{T} s_i(t)g_n(t)dt$$

and if the series (24) converges in the mean, then

(25) $$P\{(s_{i1}, s_{i2}, \cdots)+(z_1, z_2, \cdots)\in A_i^*\}\geqq 1-\alpha \qquad (i=1, \cdots, M)$$

where $A_i^*\frown A_j^*=\Phi$ $(i\neq j)$ and

$$z_n=\int_{-T}^{T} n_T(t)g_n(t)dt$$

(the z_n are independent, normally distributed random variables with zero mean and variance ρ_n, ρ_n being the eigenvalue corresponding to $g_n(t)$).

For each A_i^*, there is a measurable cylinder B_i with the following properties: (i) for each i $(i=1, \cdots, M)$

(26) $$P\{(s_{i1}, s_{i2}, \cdots)+(z_1, z_2, \cdots)\in B_i\}\geqq 1-2\alpha,$$

and

(ii) if $i\neq j$, then $B_i\frown B_j=\Phi$.

To prove this fact, corresponding to each A_i^*, we choose a measurable cylinder B_i^* in sequence space such that

(27) $$P\{B_i^*-A_i^*\frown B_i^*\}\leqq\frac{\alpha}{4M}$$

and

(28) $$P\{A_i^*-A_i^*\frown B_i^*\}\leqq\frac{\alpha}{4M} \qquad (i=1, \cdots, M).$$

Since $B_i^*=(A_i^*\frown B_i^*)\frown(B_i^*-(A_i^*\frown B_i^*))$ for each i, and $A_i^*\frown A_j^*=\Phi$ $(i\neq j)$, so, by (27), for any i and j $(i\neq j)$,

(29) $$P\{B_i^*\frown B_j^*\}\leqq P\{B_i^*-A_i^*\frown B_i^*\}+2P\{B_j^*-A_j^*\frown B_j^*\}\leqq\frac{3\alpha}{4M}.$$

Now, we define B_i $(i=1, \cdots, M)$ inductively, as follows:

$$B_1\equiv B_1^*$$

and

$$B_j=B_j^*-(B_1^*\smile\cdots\smile B_{j-1}^*)\frown B_j^* \qquad \text{for any } j\ (2\leqq j\leqq M).$$

Then, by (25), (28) and (29), for any j $(2\leqq j\leqq M)$.

$$P\{B_j\}=P\{B_j^*\}-P\{B_1^*\smile\cdots\smile B_{j-1}^*)\frown B_j^*\}$$

$$\geqq P\{B_j^*\}-M\max_{i,j} P\{B_i^*\frown B_j^*\}$$

$$=P\{B_j^*\}-\frac{3\alpha}{4}\geqq P\{A_j^*\frown B_j^*\}-\frac{3\alpha}{4}$$

$$=P\{A_j^*\}-P\{A_j^*-(A_j^*\frown B_j^*)\}-\frac{3\alpha}{4}$$

$$\geqq(1-\alpha)-\frac{\alpha}{4M}-\frac{3\alpha}{4}\geqq 1-2\alpha.$$

Thus we establish the existence of such B_i $(i=1, \cdots, M)$.

Now, membership in a measurable cylinder in a sequence space is determined by a finite number of coordinates. Since there is only a finite number of code words, there is an integer n such that the base of each B_i is n dimensional. Consequently,

(30) $$P\{(s_{i1}, s_{i2}, \cdots, s_{in})+(z_1, z_2, \cdots, z_n)\in B_{in}\}\geqq 1-2\alpha \qquad (i=1, \cdots, M)$$

where B_{in} is the base of B_i. (30) is equivalent to

(31) $$P\left\{\left(\frac{s_{i1}}{\sqrt{\rho_1}}, \cdots, \frac{s_{in}}{\sqrt{\rho_n}}\right)+\left(\frac{z_1}{\sqrt{\rho_1}}, \cdots, \frac{z_n}{\sqrt{\rho_n}}\right)\in B_{in}'\right\}\geqq 1-2\alpha$$
$$(i=1, \cdots, M)$$

where B_{in}' is formed B_{in} by dividing the j-th component of each vector in B_{in} by $\sqrt{\rho_j}$ $(j=1, \cdots, n)$. The sets B_{in}' are of course disjoint. Thus, the vectors $(s_{i1}/\sqrt{\rho_1}, \cdots, s_{in}/\sqrt{\rho_n})$, $i=1, \cdots, M$, may be considered as code words of a code $(n, M, 2\alpha)$ for a time-discrete memoryless channel with noise variance unity. Since each code word is allowable, so we have

(32) $$\sum_{j=1}^{n}\frac{s_{ij}^2}{\rho_j}\leqq KT \qquad (i=1, \cdots, M).$$

Therefore, the coordinate $(x_1, \cdots, x_n)$ of any code word satisfy

(33) $$\frac{1}{n}\sum_{j=1}^{n}x_j^2\leqq\frac{KT}{n}.$$

We shall denote this set "E".

Now, the method used in the proof of the strong converse of the coding theorem for a time-discontinuous Gaussian channel is completely carried over to this case. Let

$$h(y_1, \cdots, y_n|x_1, \cdots, x_n)=\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-(y_j-x_j)^2/2}$$

Let ε_1, $0<\varepsilon_1<\varepsilon$, be arbitrary and let δ be any positive number such that

$$0<\log_e\left(1+\frac{KT+\delta}{n}\right)-\log_e\left(1+\frac{KT}{n}\right)<\frac{T\varepsilon_1}{n}.$$

Furthermore, let

$$p(x_1, \cdots, x_n)=\prod_{j=1}^{n} \frac{\sqrt{n}}{\sqrt{2\pi(KT+\delta)}} e^{-n x_j^2/2(KT+\delta)}.$$

Then

$$q(y_1, \cdots, y_n)=\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi(1+(KT+\delta)/n)}} e^{-y_j^2/2(1+(KT+\delta)/n)}$$

For any $u\in E$, we put

$$\begin{aligned} A_u&=\left\{v=(y_1,\cdots, y_n)|\log_2 \frac{h(v|u)}{q(v)} \geqq \frac{n}{2}\left(\log_2\left(1+\frac{KT}{n}\right)+\frac{2T\varepsilon}{n}\right)\right\} \\ &=\left\{v=(y_1, \cdots y_n)|\log_e \frac{h(v|u)}{q(v)} \geqq \frac{n}{2}\left(\log_e\left(1+\frac{KT}{n}\right)+\frac{2T\varepsilon\log_e 2}{n}\right)\right\}. \end{aligned}$$

Since

$$E\left[\log_e \frac{h(V|u)}{q(V)}\bigg|u\right]\leqq \frac{n}{2}\left(\log_e\left(1+\frac{KT}{n}\right)+\frac{T\varepsilon_1}{n}\right)$$

and

$$\operatorname{var}\left[\log_e \frac{h(V|u)}{q(V)}\bigg|u\right]\leqq \frac{n\left\{\left(\frac{KT+\delta}{n}\right)^2+\frac{KT}{n}\right\}}{2\left(1+\frac{KT+\delta}{n}\right)^2},$$

so, by Tshebychev's inequality

$$P\{V\in A_u|u\}\leqq \frac{2\left\{\left(\frac{KT+\delta}{n}\right)^2+\frac{KT}{n}\right\}}{n\left(\frac{T\varepsilon}{n}-\frac{T\varepsilon_1}{n}\right)^2\left(1+\frac{KT+\delta}{n}\right)^2}=\frac{2\left\{\frac{KT}{n}\left(1+\frac{\delta}{KT}\right)^2+K\right\}}{T(\varepsilon-\varepsilon_1)^2\left(1+\frac{KT+\delta}{n}\right)^2}.$$

Thus the condition (4) of Theorem 1 is satisfied for suitably chosen T and for all sufficiently large n. Accordingly, we can conclude from the theorem that any code $(n, M, 2\alpha)$ whose code words meet the constraint (33) must satisfy

$$M\leqq \frac{2}{1-2\alpha}\exp_2\left[n\left\{\frac{1}{2}\log_2\left(1+\frac{KT}{n}\right)+\frac{T\varepsilon}{n}\right\}\right]$$

for all sufficiently large n. Thus, we have

$$M\leqq \frac{2}{1-2\alpha}\exp_2\left[T\left\{\frac{(\log_2 e)K}{2}+\varepsilon\right\}\right]$$

and the proof is completed.

Therefore, from the coding theorem (proved by Ash) and Theorem 2, we can conclude that $(\log_2 e)K/2$ is the capacity for the time-continuous Gaussian channel.

Achnowledgement. The author wishes to thank Prof. M. Udagawa, Tokyo University of Education, for many helpful suggestions and kind guidance.

References

[1] Ash, R. B., Capacity and error bounds for a time-continuous Gaussian channel. Information and Control **6** (1963), 14–27.

[2] Fortet, R., Hypothesis testing and estimation for Laplacian functions. Fourth Berkeley Symp. on Math. Stat. and Prob. **1** (1961), 289–305. Univ. of California Press.

[3] Shannon, C. E., Probability of error for optimal codes in a Gaussian channel. Bell System Tech. J. **38** (1959), 611–656.

[4] Thomasian, A. J., Error bounds for continuous channels. Fourth London Symp. on Information Theory (1961), 46–60. Butterworth Scientific Publications.

[5] Wolfowitz, J., Coding theorems of information theory. Springer-Verlag (1961).

[6] Wolfowitz, J., On channels without a capacity. Information and Control **6** (1963), 49–54.

Faculty of Art Education, Yokohama National University.