KŌDAI MATH. SEM. REP.
21 (1969), 418—425

THE STRONG CONVERSE THEOREM IN THE DECODING SCHEME OF LIST SIZE L

BY SHŌICHI NISHIMURA

1. Introduction.

The coding theorem, discovered by Shannon [5], states that information can be transmitted with arbitrarily small error probability by means of words lengthening. A question will arise on what will become of an asymptotic behavior of an error probability. Denoting by M the number of input messages and by N the length of corresponding input words, the rate is defined as $R=(1/N)\log M$. If we lengthen the word with a fixed rate, the error probability approaches exponentially zero. For a broad class of channels its first exponential error-bound was given by Fano (1956). Its precise upper estimate was obtained by Gallager [4], and its precise lower estimate was obtained by Shannon, Gallager, and Berlekamp [6].

On the other hand, the weak converse theorem for Shannon's coding theorem was proved by Feinstein [3] using Fano's inequality. It states that if the rate is above the channel capacity, for a sufficiently large N the error probability is positive. Wolfowitz [7] proved the strong converse theorem; there exists a positive constant K such that there does not exist $M=e^{NC+K\sqrt{N}}$ input messages such that its error probability is below $\lambda>0$. In this paper we derive the strong converse theorem in the decoding scheme of list size L. The rate is defined as

$$R=\frac{1}{N}\log\frac{M}{L}.$$

If $R>C+\varepsilon$, then the average error probability approaches one. The decoding scheme of list size L which was mentioned in Shannon, Gallager, and Berlekamp [6], is that the decoder, rather than mapping the output words into a single message, maps it into a list of messages. If the transmitted source message is not on the list of decoded message, we say that a list decoding error has occurred.

2. Channel and list decoding.

Let input alphabets be $i=1,\cdots,I$ ($I\leqq a$) and output alphabets be $j=1,\cdots,J$ ($J\leqq a$). Let X_N be the set of all input words of length N that can be transmitted, and let Y_N be the set of all output words of length N that can be received. Let $P(y\mid x_m)$, for $x_m\in X_N$ and $y\in Y_N$, be the conditional probability of received word y, given that

Received June 17, 1969.

x_m was transmitted. We assume memoryless channel such that

$$P(y\mid x_m)=\prod_{n=1}^{N} P(j_m\mid i_{m,n}) \qquad \text{and} \qquad P(j\mid i)\neq 0 \quad \text{for all } (i,j).$$

For a given code and a list decoding scheme, let A_m be the set of output words for which message m is on the list of decoding messages. That is, we put X_y as $X_y=\{m; y\in A_m\}$, then X_y contains at most L elements. We assume that we transmit input words with equi-probability $1/M$, then the average error probability is

$$P(e)=\frac{1}{M}\sum_{m=1}^{M}\sum_{y\in A_m^C} P(y\mid x_m)$$

and the rate as

$$R=\frac{1}{N}\log\frac{M}{L}.$$

In the case of $L=1$ the decoding scheme satisfies $A_m\cap A_{m'}=\phi$ $(m\neq m')$ which is an ordinary decoding scheme and $R=(1/N)\log M$ is an ordinary rate. In the case of $L=M$ while the error probability is equal to zero $(P(e)=0)$, the rate R is also equal to zero $(R=0)$.

Shannon, Gallager, and Berlekamp [6] proved that if

$$R_{\text{crit}}<R=\frac{1}{N}\log\frac{M}{L}\leqq C-\varepsilon$$

then for the same R the average error probability has respectively the same exponential order with N independent as L and M.

There is a limit to bits per second which is transmitted through a given channel. We, for example, consider that a moon's figures is transmitted to the earth by means of television. To obtain a clear picture of the moon, a code is needed redundancy. The longer the word for a point becomes, the rougher the picture must become. Vice versa. The finer the picture becomes, the worse the accuracy for a point becomes. In such a case we could use a list decoding scheme. If the camera is not moved quickly, there is no difference between a received information at present and received informations in the immediate past. From a received word we do not decide a message, but we obtain a list and make a picture by referring to informations in the immediate past.

3. Strong converse theorem.

At first we shall begin with several probabilistic lemmas of Bernoulli trials.

LEMMA 1. *If Y be any nonnegative (discrete) random variable, and d be any positive real number, then*

$$P\{Y>d\}<\frac{1}{d}E(Y).$$

Proof.

$$E(Y)=\sum_{y} yP(y)=\sum_{0\leqq y\leqq d} yP(y)+\sum_{y>d} yP(y)$$
$$\geqq\sum_{y>d} yP(y)>d\sum_{y>d} P(y)=dP\{Y>d\}.$$

LEMMA 2. *If Y be a (discrete) random variable, b be any real number and r be any positive real number, then*

$$P\{X>b\}<e^{-rb}E[e^{rX}] \qquad \text{and} \qquad P\{X<b\}<e^{rb}E[e^{-rX}].$$

Proof. Now taking $Y=e^{rX}$ and $d=e^{rb}$, we obtain

$$P\{X>b\}<e^{-rb}E[e^{rX}].$$

In the same way we obtain

$$P\{X<b\}=P\{-X>-b\}<e^{rb}E[e^{rX}].$$

LEMMA 3. *If $Z_1, \cdots, Z_N$ be an independent identically distributed Bernoulli trials ($E[Z_n]=p$, $n=1, \cdots, N$, $1\geqq{}^{\forall}p\geqq\lambda>0$) and if S_N stand for the number of successes in N Bernoulli trials ($S_N=\sum_{n=1}^{N} Z_n$). Then there exists a positive constant c' such that*

$$P\{|S_N-Np|>Np\delta\}\leqq 2e^{-Nc'}.$$

Proof. At first we only estimate:

i) In the case of $1\leqq p(1+\delta)$

$$P\{S_N>Np(1+\delta)\}\leqq P\{S_N>N\}=0,$$

ii) In the case of $1>p(1+\delta)$.

From lemma 2 we have

$$P\{S_N>Np(1+\delta)\}\leqq e^{-rNp(1+\delta)}E[e^{rS_N}].$$

Since $Z_1, \cdots, Z_N$ are mutually independent, we have

$$P\{S_N>Np(1+\delta)\}\leqq e^{-rNp(1+\delta)}E[e^{rZ_1}]^N.$$

We put $\varphi(r)=E[e^{rZ_1}]=1-p+pe^r$ and $\mu(r)=\log\varphi(r)$, then

$$P\{S_N>Np(1+\delta)\}\leqq e^{N[\mu(r)-rp(1+\delta)]}.$$

To obtain the tightest bound we differentiate $\mu(r)-rp(1+\delta)$ by r.

$$\mu'(r)=p(1+\delta), \qquad r=\log\frac{1-p+\delta-\delta p}{1-p-\delta p}.$$

Consequently

$$\mu(r)-rp(1+\delta)=\mu(r)-r\mu'(r)$$
$$=-(1-p-\delta p)\log\left(1-\frac{\delta p}{1-p}\right)-p(1+\delta)\log(1+\delta).$$

We differentiate this by p.

$$\frac{d}{dp}(\mu(r)-r\mu'(r))=(1+\delta)\log\frac{1-p-p\delta}{(1-p)(1+\delta)}+\frac{\delta}{1-p}$$
$$\leqq-(1+\delta)\frac{\delta}{(1-p)(1+\delta)}+\frac{\delta}{1-p}=0.$$

$\mu(r)-r\mu'(r)$ is monotone decreasing with respect to p. Since $p\geqq\lambda$, we have

$$\mu(r)-r\mu'(r)\leqq-(1-\lambda-\lambda\delta)\log\left(1-\frac{\lambda\delta}{1-\lambda}\right)-\lambda(1+\delta)\log(1+\delta).$$

$$P\{S_N>Np(1+\delta)\}\leqq\exp\left[-N\left\{(1-\lambda-\lambda\delta)\log\left(1-\frac{\lambda\delta}{1-\lambda}\right)+\lambda(1+\delta)\log(1+\delta)\right\}\right].$$

In the same way as above, we have

$$P\{S_N<Np(1-\delta)\}\leqq\exp\left[-N\left\{(1-\lambda+\lambda\delta)\log\left(1+\frac{\lambda\delta}{1-\lambda}\right)+\lambda(1-\delta)\log(1-\delta)\right\}\right].$$

Since

$$p\log\frac{p}{q}+(1-p)\log\frac{1-p}{1-q}>0 \qquad (p\neq q,\ p\neq 0,1),$$

those which are two terms in the brackets are positive. Let we put the smaller of these as c', then we completed the proof.

Let us assign a probability vector $\boldsymbol{P}=(p_1,\cdots,p_I)$ on input alphabets such shat p_i is a multiple of $1/N$, then we have a probability vector $\boldsymbol{Q}=(q_1,\cdots,q_J)$ on output alphabets such that $q_j=\sum_i P(j\,|\,i)p_i$. Even though we assigned input probability as has been described, discussion from now on should not merely relied upon the usage of these words. Attention should also be given to the fact that probabilistic approach only will be applied concering the relation between input and output, that is, the number of generated sequence, etc.

We define the function H as follows:

$$H(\boldsymbol{Q})=-\sum_j q_j\log q_j,$$
$$H(\boldsymbol{Q}\,|\,\boldsymbol{P})=-\sum_{ij}p_iP(j\,|\,i)\log P(j\,|\,i).$$

Let $N(i\,|\,x)$ be the number of alphabets i in x, $N(j\,|\,y)$ be the number of alphabets j in y and $N(ij\,|\,xy)$ be the number of pairs $(i_n,j_n)=(i,j)$ in (x,y). For convenience' sake, we introduce following sets dependent on $\delta, \eta>0$.

DEFINITION 1. $x \in X_N$ is called to be ***P**-seq.* if $N(i|x)=Np_i$ for all i.
Now we put

$$V_{ij}(x)^*=\begin{cases}\{y;\ N(ij\,|\,xy)>N(i\,|\,x)P(j\,|\,i)(1+\eta)\} & \text{for } p_i\geqq\delta,\\ \phi & \text{for } p_i<\delta,\end{cases}$$

$$V_{ij}(x)_*=\begin{cases}\{y;\ N(ij\,|\,xy)<N(i\,|\,x)P(j\,|\,i)(1-\eta)\} & \text{for } p_i\geqq\delta,\\ \phi & \text{for } p_i<\delta\end{cases}$$

and $V(x)=\cup_{ij}(V_{ij}(x)^*\cup V_{ij}(x)_*)$.

DEFINITION 2. $y\in Y_N$ is called to *be generated by **P**-seq.* if $y\in V(x)^c$.

LEMMA 4. $$P\{V(x)|x\}\leqq 2a^2e^{-Nc''}.$$

Proof. For a given channel, the conditional probability $P(j\,|\,i)$ is fixed and the number of pairs (i,j) is finite, then there exists a positive constant λ such that $\lambda=\min_{(i,j)}P(j\,|\,i)>0$. If $V_{ij}(x)^*$ or $V_{ij}(x)_*$ is not empty, $N(i\,|\,x)$ must be larger than $N\delta$. Using lemma 3 there exists c' such that

$$P\{V_{ij}(x)^*\,|\,x\}\leqq e^{-N\delta c'} \quad\text{and}\quad P\{V_{ij}(x)_*\,|\,x\}\leqq e^{-N\delta c'} \qquad \text{for all} \quad (i,j).$$

If we put $c''=\delta c'$, we completed the proof.

Let $B(\boldsymbol{P})$ be the number of output words which is generated by $\boldsymbol{P}$-seq. x.

LEMMA 5. *We have*

$$B(\boldsymbol{P})\leqq\exp N\{H(\boldsymbol{Q})-a(\eta+\delta)\log\lambda\}.$$

Proof. By

$$N(ij\,|\,xy)\leqq N(i\,|\,x)P(j\,|\,i)(1+\eta)=Np_iP(j\,|\,i)(1+\eta) \qquad \text{for} \quad p_i\geqq\delta$$

and

$$N(ij\,|\,xy)\leqq N(p_iP(j\,|\,i)+\delta) \qquad \text{for} \quad p_i<\delta,$$

we have

$$N(j\,|\,y)\leqq N(q_j+a(\eta+\delta)) \qquad \text{for all} \quad j.$$

Since $q_j=\sum_i p_iP(j\,|\,i)$, we have

$$q_j\geqq\lambda,$$

$$P(y)=\prod_{j=1}^{J}(q_j)^{N(j|y)}\geqq\prod_{j=1}^{J}(q_j)^{N(q_j+a(\delta+\eta))}$$

$$\geqq\exp[-N\{H(\boldsymbol{Q})-a(\eta+\delta)\log\lambda\}].$$

The probability of set y which is generated by $\boldsymbol{P}$-seq. is less than or equal to one. We have

$$B(\boldsymbol{P}) \leqq \exp N\{H(\boldsymbol{Q}) - a(\eta+\delta) \log \lambda\}.$$

LEMMA 6. *If x be $\boldsymbol{P}$-seq. and y be generated by x, then*

$$P\{y \mid x\} \leqq \exp[-N\{H(\boldsymbol{Q} \mid \boldsymbol{P}) + a(\eta+\delta) \log \lambda\}].$$

Proof.

$$P\{y \mid x\} = \prod_j P(j \mid i)^{N(ij \mid xy)} \leqq \prod_j P(j \mid i)^{N\{p_i P(j|i) - \eta - \delta\}}$$

$$\leqq \exp[-N\{H(\boldsymbol{Q} \mid \boldsymbol{P}) + a(\eta+\delta) \log \lambda]\}.$$

Now we can prove the strong converse theorem. At first we attach conditions to x_m and A_m, and step by step detach them.

LEMMA 7. *Let $x_m (m=1, \cdots, M)$ be $\boldsymbol{P}$-seq. and A_m be the subset of output words y which is generated by x_m. Let $(x_1, A_1), \cdots, (x_M, A_M)$ be a decoding scheme of list size L. If $R \geqq C + \varepsilon/2$ where R is the rate of list size L and C is the channel capacity, then for a sufficiently small δ and η there exists a positive constant c_1' such that*

$$P(e) \geqq 1 - e^{-Nc_1'}$$

Proof. From lemma 5 and definition of list decoding, we have

$$\sum_{m=1}^{M} (\text{the number of } y \text{ which is contained in } A_m)$$

$$\leqq L B(\boldsymbol{P}) \leqq L \exp[N\{H(\boldsymbol{Q}) - a(\eta+\delta) \log \lambda\}].$$

From lemma 6, we have

$$P(y \mid x) \leqq \exp[-N\{H(\boldsymbol{Q} \mid \boldsymbol{P}) + a(\eta+\delta) \log \lambda\}].$$

Then we obtain

$$1 - P(e) = \frac{1}{M} \sum_{m=1}^{M} P(A_m \mid x_m)$$

$$\leqq \frac{1}{M} L \exp[N\{H(\boldsymbol{Q}) - a(\eta+\delta) \log \lambda\}] \exp[-N\{H(\boldsymbol{Q} \mid \boldsymbol{P}) + a(\eta+\delta) \log \lambda\}]$$

$$\leqq \exp[-N\{R - C + 2a(\eta+\delta) \log \lambda\}]$$

$$\leqq \exp\left[-N\left\{\frac{\varepsilon}{2} + 2a(\eta+\delta) \log \lambda\right\}\right].$$

For sufficiently small δ and η, we have

$$c_1' = \frac{\varepsilon}{2} + 2a(\eta+\delta)\log\lambda > 0,$$

where c_1' is dependent on ε, δ, η and λ, but independent of N. Hence

$$P(e) \geqq 1 - e^{-Nc_1'}.$$

Next we detach the condition to A_m.

LEMMA 8. *Let $(x_1, A_1), \cdots, (x_M, A_M)$ be a decoding scheme such that x_m is **P**-seq. and A_m is any list size L decoding set. If $R > C + \varepsilon/2$, then there exists positive constants c_1'' and c_2'' such that*

$$P(e) \geqq 1 - c_2'' e^{Nc_1''}.$$

Proof. Let A_m be the intersection of A_m and $V(x_m)^c$, A_m' be the intersection of A_m and $V(x_m)$.

$$1 - P(e) = \frac{1}{M}\sum_{m=1}^{M} P(A_m \mid x_m) = \frac{1}{M}\sum_{m=1}^{M} P(A_m' \mid x_m) + \frac{1}{M}\sum_{m=1}^{M} P(A_m'' \mid x_m).$$

From lemma 7 the first term can not exceed $e^{-Nc_1'}$. From lemma 4 the second term can not exceed $2a^2 e^{-Nc''}$. Then for fixed sufficiently small δ and η, there exists positive constant c_1'' and c_2'' such that

$$P(e) \geqq 1 - c_2'' e^{-Nc_1''}.$$

Thus we completed the proof.

The number of probabilistic vector $\boldsymbol{P}$ is at most $(N+1)^a$. For each $\boldsymbol{P}$ lemma 8 was prove. We number the class of $\boldsymbol{P}$-seq. into which we classify x_m $(m=1, \cdots, M)$.

$$1 - P(e) = \frac{1}{M}\sum_{m=1}^{M} P(A_m \mid x_m) = \sum_{k=1}^{K} \frac{M_k}{M}\frac{1}{M_k}\sum P(A_{m_k} \mid x_{m_k}),$$

where M_k is the number of k-th class of $\boldsymbol{P}$-seq. x_m and $K \leqq (N+1)^a$.

We assume that $R \geqq C + \varepsilon$.

i) If $M_k > Le^{N(C+\varepsilon/2)}$, then from lemma 8 we have

$$\frac{M_k}{M}\frac{1}{M_k}\sum P(A_{m_k} \mid x_{m_k}) \leqq \frac{M_k}{M} c_2'' e^{-Nc_1''} \leqq c_2'' e^{-Nc_1''}.$$

ii) If $M_k \leqq Le^{N(C+\varepsilon/2)}$, then we have

$$\frac{M_k}{M}\frac{1}{M_k}\sum P(A_{m_k} \mid x_{m_k}) \leqq \frac{M_k}{M} \leqq e^{-N\varepsilon/2}.$$

Hence we obtain

$$1-P(e)\leqq(N+1)^a \max(c_2''e^{-Nc_1''}, e^{-N\varepsilon/2}).$$

Then for a sufficiently large N there exist positive constants c_1 and c_2 such that $P(e)\geqq 1-c_2e^{-Nc_1}$.

The above will be summarized as follows:

THEOREM. (*Strong converse theorem*) *Let a given channel be a discrete memoryless channel such that* $P(j\,|\,i)\neq 0$ *for all* (i,j). *Let input messages* $m=(1,\cdots,M)$ *be transmitted with equi-probability* $1/M$. *If* $R=(1/N)\log(M/L)\geqq C+\varepsilon$, *where* C *is the channel capacity, then for any code* $(x_1,A_1),\cdots,(x_M,A_M)$ *which is list size* L, *there exist positive constants* c_1, *and* c_2 *such that for a sufficiently large* N,

$$P(e)=\frac{1}{M}\sum_{m=1}^{M}P(A_m^c\,|\,x_m)\geqq 1-c_2e^{-Nc_1}.$$

ACKNOWLEDGEMENT. The author would like to express his most sincere appreciation to Prof. K. Kunisawa for his valuable suggestion and encouragement.

REFERENCES

[1] ASH, R. B., Information theory. Wiley, New York (1965).
[2] FANO, R. G., Transmission of information. M.I.T. Press, Cambridge (1961).
[3] FEINSTEIN, A., Foundations of information theory. McGraw-Hill, New York (1958).
[4] GALLAGER, R. G., A simple derivation of the coding theorem and some applications. I.E.E.E. Trans. IT-**11**, (1965), 3-18.
[5] SHANNON, C. E., A mathematical theory of communication. B.S.T.J. **27** (1948), 379-423.
[6] SHANNON, C. E., R. G. GALLAGER, AND E. R. BERLEKAMP, Lower bound to error probability for coding on discrete memoryless channel (1); (2). Information and Control **10** (1967), 15-103; 522-552.
[7] WOLFOWITZ, J., Coding theorems of information theory. Printice-Hall Englewood Cliffs N.J. (1961).

DEPARTMENT OF MATHEMATICS,
TOKYO INSTITUTE OF TECHNOLOGY.