Z. Wahrscheinlichkeitstheorie verw. Gebiete
34, 179 – 182 (1976)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete
© by Springer-Verlag 1976

Every Bad Code Has a Good Subcode: A Local Converse to the Coding Theorem

R. Ahlswede* and G. Dueck

Mathematische Fakultät der Universität Bielefeld,
Kurt-Schumacher-Str. 6, D-4800 Bielefeld, Federal Republic of Germany

The Local Converse

The transmission probabilities P of a discrete memoryless channel (DMC) with alphabets $\mathscr{X}$ and $\mathscr{Y}$ are given by

$$P(y^n|x^n)=\prod_{t=1}^{n} w(y_t|x_t) \tag{1}$$

where $x^n=(x_1,\ldots,x_n)\in\mathscr{X}^n=\prod_1^n \mathscr{X}$, $y^n=(y_1,\ldots,y_n)\in\mathscr{Y}^n=\prod_1^n \mathscr{Y}$, and where w is a $|\mathscr{X}|\times|\mathscr{Y}|$-stochastic matrix.

An (n, N, λ)-code for the DMC is a system of pairs $\{(u_i, D_i)|i=1,\ldots,N\}$ with $u_i\in\mathscr{X}^n$ and pairwise disjoint subsets D_i of $\mathscr{Y}^n$ $(i=1,\ldots,N)$, and with

$$P(D_i|u_i)>1-\lambda \qquad \text{for } i=1,\ldots,N. \tag{2}$$

If $N(n,\lambda)$ denotes the maximal N for which an (n, N, λ)-code exists, then

$$\lim_{n\to\infty} n^{-1}\log N(n,\lambda)=C, \qquad 0<\lambda<1. \tag{3}$$

This result was stated (without proof) in [1] as Theorem 12. The inequality

$$\liminf_{n\to\infty} n^{-1}\log N(n,\lambda)\geqq C, \qquad 0<\lambda<1 \tag{4}$$

(the coding theorem) was proved in [3] and in [5]. It was shown in [2] that (weak converse)

$$\inf_{\lambda>0}\limsup_{n\to\infty} n^{-1}\log N(n,\lambda)\leqq C \tag{5}$$

and finally in [4] that the strong converse holds, i.e.

$$\limsup_{n\to\infty} n^{-1}\log N(n,\lambda)\leqq C, \qquad 0<\lambda<1. \tag{6}$$

* Research of this author was supported by the Deutsche Forschungsgemeinschaft.

(6), together with (4), establishes (3), which says that for any error probabilities λ and λ', $0<\lambda<\lambda'<1$, the asymptotic growth of the maximal codelengths $N(\mathrm{n}, \lambda)$ and $N(n, \lambda')$ is the same.

In this note we show that an even stronger statement is true:

Theorem (Local Converse). *Let P be a DMC and let ε, λ, λ' and R be real numbers such that $\varepsilon>0$, $R>0$, $0<\lambda'<\lambda<1$. Then one can find (also explicitly) an $n_0(\lambda, \lambda', \varepsilon)$ such that for all $n \geqq n_0(\lambda, \lambda', \varepsilon)$ the following is true:*

Every $(n, \exp(nR), \lambda)$-code $\{(u_i, D_i) \mid i=1, \ldots, N=\exp(nR)\}$ contains a subset of codewords $\{u_{i_k} \mid k=1, \ldots, N'=\exp(n(R-\varepsilon))\}$ with suitable decoding sets F_{i_k} $(i=1, \ldots, N')$ such that $\{(u_{i_k}, F_{i_k}) \mid k=1, \ldots, N'\}$ is an $(n, \exp(n(R-\varepsilon)), \lambda')$-code.

The result gives a new geometric insight into the coding problem. Out of a set of codewords with a certain minimal "distance" one can select a rather big subset of a prescribed larger minimal "distance". In the case of a binary symmetric channel the word "distance" as used here can be replaced by the Hamming metric, a true distance. The result may be of interest for the actual construction of small error codes.

The theorem was stated here for the DMC, the simplest and most familiar channel, even though the phenomenon "bad codes contain good codes" is of a rather general nature and occurs for much more general one-way channels as well as for multi-way channels ([6]).

Also, this theorem together with the weak converse (5) implies the strong converse (6). A new and general method to prove strong converses was presented in [7]. It applies to many multi-user coding problems for which all classical approaches fail. The idea is as follows: One enlarges the decoding sets of a given code in order to decrease the error probability. The new decoding sets are no longer disjoint, that is, one has a list code to which one applies Fano's Lemma ([2]). Surprisingly enough one can decrease the error probability significantly with a "small" increase in list size and therefore the idea works. The main estimates are contained in the lemma below. The novelty of the present method of proof for the Theorem lies in the observation that one can select at random a subcode of list size 1 out of a list code with small list size without losing to much in error probability or rate.

Proof of the Theorem

We need a result of Margulis [8] in the slightly generalized form given in [7] as Lemma 4:

Lemma. *Given a DMC with transmission probabilities $P=\prod_{1}^{n} w$ there is a constant $c=c(w)>0$ such that for any n, $B \subset \mathscr{Y}^n$ and $x^n \in \mathscr{X}^n$:*

$$P(\Gamma^k B \mid x^n) \geqq \Theta[\Theta^{-1}(P(B \mid x^n)) + n^{-\frac{1}{2}}(k-1)\, c], \tag{8}$$

where $\Gamma^k B = \{y^n=(y_1, \ldots, y_n) \mid \text{exists } y'^n=(y'_1, \ldots, y'_n) \in B \text{ with } y'_t \neq y_t \text{ for at most } k \text{ components}\}$ and $\Theta(t)=\int_{-\infty}^{t} (2\pi)^{-\frac{1}{2}} \exp(-t^2/2)\, dt$.

Suppose now we are given the (n, N, λ)-code $\{(u_i, D_i) | i = 1, \dots, N\}$. Define

$$E_i = \Gamma^{k_n} D_i \quad \text{for } i = 1, \dots, N \tag{9}$$

with $k_n = n^{\frac{1}{2}} \log n$ (actually every k_n with $n^{\frac{1}{2}} k_n^{-1} = o(1)$ and $n^{-1} k_n = o(1)$ would work). By the lemma

$$P(E_i | u_i) \geqq \Theta[\Theta^{-1}(1-\lambda) + n^{-\frac{1}{2}}(k_n - 1)\, c]. \tag{10}$$

Since $k_n n^{-\frac{1}{2}} \to \infty$ for $n \to \infty$ and since $\Theta(t) \to 1$ for $t \to \infty$, the right side converges to 1, and therefore certainly exceeds $1 - \lambda'/4$ for $n \geqq n_0(\lambda, \lambda')$, suitable. On the other hand, the decoding list $\mathscr{J}(y^n) = \{u_i | 1 \leqq i \leqq N,\ y^n \in E_i\}$ satisfies:

$$|\mathscr{J}(y^n)| \leqq |\Gamma^{k_n}(y^n)| \leqq \binom{n}{k_n} |\mathscr{Y}|^{k_n} \leqq (n |\mathscr{Y}|)^{k_n}$$

and hence

$$|\mathscr{J}(y^n)| \leqq \exp(n \delta_n) \quad \text{for all } y^n \in \mathscr{Y}^n, \tag{11}$$

where $\delta_n = |\mathscr{Y}|\, n^{-\frac{1}{2}} \log^2 n \to 0$ as $n \to \infty$.

We complete now the proof by a random coding argument ([5]). Let U_i $(i = 1, \dots, M)$ be independent, identically distributed random variables with distribution

$$\Pr(U_i = u_k) = 1/N \quad \text{for } k = 1, \dots, N.$$

With every outcome $(u_{i_1}, \dots, u_{i_M})$ of $(U_1, \dots, U_M)$ we associate decoding sets $(F_{i_1}, \dots, F_{i_M})$, where

$$F_{i_k} = \{y^n \in E_{i_k} \,|\, |\mathscr{J}(y^n) \cap \{u_{i_1}, \dots, u_{i_M}\}| = 1\}. \tag{12}$$

Equivalently: $F_{i_k} = E_{i_k} - \bigcup_{j \neq k} E_{i_j}$.

For reasons of symmetry the expected average error probability for this decoding rule

$$E\lambda(U_1, \dots, U_M) = M^{-1} \sum_{k=1}^{M} E \sum_{y^n \in F_{i_k}^c} P(y^n | U_k)$$

equals $E \sum_{y^n \in F_{i_1}^c} P(y^n | U_1)$, and this expression is upper bounded by

$$E \sum_{y^n \in E_{i_1}^c} P(y^n | U_1) + E \sum_{y^n \in E_{i_1} \cap F_{i_1}^c} P(y^n | U_1).$$

The first sum is smaller than $\lambda'/4$. Assume therefore that $y^n \in E_{i_1}$ and also that $U_1 = u_{i_1}$. Since $|\mathscr{J}(y^n)| \leqq \exp(n\delta_n)$, the probability for

$$\{U_2, \dots, U_M\} \cap \mathscr{J}(y^n) \neq \emptyset$$

is smaller than

$$1 - (1 - \exp(n\delta_n)/N)^{M-1} \leqq M \exp(n\delta_n)/N$$

and hence

$$E \sum_{y^n \in E_{i_1} \cap F_{i_1}^c} P(y^n | U_1) \leqq E \sum_{y^n \in E_{i_1}} M \exp(n\delta_n)\, P(y^n | U_1)/N \leqq M \exp(n\delta_n)/N.$$

With $M=2\exp(n(R-\varepsilon))$ we obtain for n large enough

$$E\sum_{y_n\in F_{i_1}^c} P(y^n\,|\,U_1)\leqq\lambda'/4+M\exp(n\delta_n)/N\leqq\lambda'/2.$$

From a subcode of length M and with average error $\lambda'/2$ we can pass to a further subcode of length $N'=\exp(n(R-\varepsilon))$ and *maximal* error λ'.

References

1. Shannon, C.E.: The mathematical theory of communication. Bell System Techn. J. **27**, 379-423, 623-656 (1948)
2. Fano, R.M.: Statistical theory of communication. Notes on a course given at MIT, 1952, 1954
3. Feinstein, A.: A new basic theorem of information theory. Trans. IRE, PGIT, Sept. 1954, 2-22
4. Wolfowitz, J.: The coding of messages subject to chance errors. Illinois J. Math. 1, 591-606 (1957)
5. Shannon, C.E.: Certain results in coding theory for noisy channels. Information and Control **1**, 6-25 (1957)
6. Ahlswede, R.: Multi-way communication channels. 2nd Internat. Sympos. on Inform. Theory 1971. Publishing House of the Hungarian Academy of Sciences, pp. 23-52
7. Ahlswede, R., Gàcs, P., Körner, J.: Bounds on conditional probabilities with applications in multi-user communication. Z. Wahrscheinlichkeitstheorie verw. Gebiete **34**, 157-177 (1976)
8. Margulis, G.A.: Veroyatnostniye characteristiki grafov s bolshoy svyaznostyu [In Russian]. Problemy Peredači. Informačii **X**, 101-108 (1974)

Received April 25, 1975