PAPER *Special Section on Information Theory and Its Applications*

The Redundancy of Universal Coding with a Fidelity Criterion

Daiji ISHII† ***and*** **Hirosuke YAMAMOTO**††, ***Members***

SUMMARY The redundancy of universal lossy data compression for discrete memoryless sources is considered in terms of type and d-ball covering. It is shown that there exists a universal d-semifaithful code whose rate redundancy is upper bounded by $(A - \frac{1}{2})\ n^{-1} \ln n + o(n^{-1} \ln n)$, where A is the cardinality of source alphabet and n is the block length of the code. This new bound is tighter than known ones, and moreover, it turns out to be the attainable minimum of the universal coding proposed by Davisson.
key words: *universal d-semifaithful code, rate-distortion function, rate redundancy, n-type, d-ball covering.*

1. Introduction

The redundancy, which is defined as the difference between average code length per source symbol and its attainable limit, has been discussed in many literatures. In the case of lossless data compression, many results have already been obtained. For instance, the redundancy of rate is $O(n^{-1})$, where n is block length, when source distribution is known while it is $O(n^{-1} \ln n)$ when source distribution is unknown [1]. On the other hand, in the case of lossy data compression adequate results have not been achieved although in 1968, Pilc gave some conjecture about the redundancies of rate and distortion for discrete memoryless sources [2].

In 1995, Zhang-Yang-Wei [3], [4] evaluated tight redundancies of rate and distortion in lossy data compression. It is shown in their paper that, when source distribution is known, there exists a code such that the rate redundancy (nats/source symbol) is upper bounded by $n^{-1} \ln n + o(n^{-1} \ln n)$, and is lower bounded by $\frac{1}{2} n^{-1} \ln n + o(n^{-1} \ln n)$. Furthermore, it it shown that there exists a code such that the distortion redundancy is $-\frac{\partial}{\partial R} d(p, R) n^{-1} \ln n + o(n^{-1} \ln n)$, where $d(p, R)$ is the distortion-rate function for source distribution p and rate R (nats/source symbol). This redundancy indicates that Pilc's conjecture with respect to distortion is true. In their work, they discussed the redundancies in terms of type and d-ball covering. In particular, Lemma 3 in [4], which is shown as Lemma 1 in this paper, plays an important role in evaluating redundancies. However, the problem of redundancy in universal coding with a fidelity criterion has remained unsolved. Although Yu-Speed [5] evaluated the redundancy, their result is not tight.

In this paper*, we evaluate the redundancies of universal lossy data compression for discrete memoryless sources, based on the approach of type and d-ball covering such as [4]. Moreover, we show that our results are tighter than that of [5] and these are the attainable minimum.

Manuscript received January 29, 1997.
Manuscript revised May 12, 1997.
†The author is with C & C Media Research Laboratories, NEC Corporation, Kawasaki-shi, 216 Japan.
††The author is with the Department of Mathematical Engineering and Information Physics, Graduate School of Engineering, University of Tokyo, Tokyo, 113 Japan.

2. Preliminaries

Let $X = \{X_t\}_{t=1}^{\infty}$ be a memoryless source with a source alphabet $\mathcal{A} = \{1, \cdots, A\}$, and let $\mathcal{B} = \{1, \cdots, B\}$ be a reproduction alphabet. We denote n symbols from the source by x^n, and the set of all n symbols on $\mathcal{A}$ by $\mathcal{A}^n$.

Let $\rho : \mathcal{A} \times \mathcal{B} \rightarrow [0, \infty)$ be a single letter distortion measure. We assume ρ has an upper bound $\rho_{\max} < \infty$. For $x^n = (x_1, \cdots, x_i, \cdots, x_n)$ and $y^n = (y_1, \cdots, y_i, \cdots, y_n)$, the average single letter fidelity criterion ρ_n is defined as

$$\rho_n(x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i, y_i).$$

The rate–distortion function with a source probability distribution p and distortion level d is denoted by $R(p, d)$ while the inverse of $R(p, d)$, i.e., the distortion–rate function, is represented by $d(p, R)$.

The n-type of $x^n \in \mathcal{A}^n$ is defined as

$$t(x^n) = (t(x^n, 1), \cdots, t(x^n, A)), \tag{1}$$

where

$$t(x^n, a) = \frac{1}{n} |\{i : x_i = a\}|, \quad a \in \mathcal{A}. \tag{2}$$

$|\cdot|$ is the cardinality of a set. Let $\mathcal{T}_n(\mathcal{A})$ be the set of all n-types on $\mathcal{A}$, and we define for each $t \in \mathcal{T}_n(\mathcal{A})$

$$T_X^n(t) = \{x^n \in \mathcal{A}^n : t(x^n) = t\}, \tag{3}$$

which is the set of strings on $\mathcal{A}^n$ whose n-types are t. Similarly, let $\mathcal{T}_n(\mathcal{B})$ be the set of all n-types on $\mathcal{B}$, and we define for each $r \in \mathcal{T}_n(\mathcal{B})$

$$T_Y^n(r) = \{y^n \in \mathcal{B}^n : r(y^n) = r\}, \tag{4}$$

*This work is first presented in [8] in Japanese.

where $r(y^n)$ is the n-type of $y^n \in \mathcal{B}^n$.

Furthermore, for $(x^n, y^n) \in \mathcal{A}^n \times \mathcal{B}^n$, the joint n-type is defined as

$$s(x^n, y^n) = (s(x^n, y^n : 1, 1), \cdots, s(x^n, y^n : A, 1), \cdots, s(x^n, y^n : a, b), \cdots, s(x^n, y^n : A, B)), \quad (5)$$

where

$$s(x^n, y^n : a, b) = \frac{1}{n}|\{(i,j) : x_i = a, y_j = b\}|. \quad (6)$$

Letting $\mathcal{T}_n(\mathcal{A} \times \mathcal{B})$ be the set of all joint n-types on $\mathcal{A} \times \mathcal{B}$, the set of strings with joint n-type $s \in \mathcal{T}_n(\mathcal{A} \times \mathcal{B})$ is defined as

$$T^n_{X,Y}(s) = \{(x^n, y^n) \in \mathcal{A}^n \times \mathcal{B}^n : s(x^n, y^n) = s\}. (7)$$

We next describe the d-ball covering. In general, the d-ball centered at $x^n \in \mathcal{A}^n$ is defined as

$$B(x^n, d) = \{y^n \in \mathcal{B}^n : \rho_n(x^n, y^n) \leqq d\}. \quad (8)$$

According to [4], the restricted d-ball for $r \in \mathcal{T}_n(\mathcal{B})$ is defined as

$$B(x^n, r, d) = B(x^n, d) \cap T^n_Y(r). \quad (9)$$

Since the cardinality of $B(x^n, r, d)$ depends on x^n only through the n-type $t(x^n) = t$, we denote it by $F_n(t, r, d) = |B(x^n, r, d)|$.

Then, the following lemma holds.

Lemma 1 (Zhang-Yang-Wei [4, Lemma 3]): For sufficiently large n,

$$\ln F_n(t, r, d) \leqq nH_u(t, r, d) - nH(t) - \frac{B}{2}\ln n + c_1 \ln\ln n, \quad (10)$$

$$\ln F_n(t, r, d) \geqq nH_u(t, r, d) - nH(t) - \frac{B}{2}\ln n - c_2, \quad (11)$$

where $c_1, c_2 > 0$.

$H_u(t, r, d)$ in Lemma 1 is defined by

$$H_u(t, r, d) \equiv \sup_{s \in \mathcal{S}(t,r,d)} H(s), \quad (12)$$

where

$$\mathcal{S}(t, r, d) = \{s : t \text{ and } r \text{ are the marginals of } s, \text{ and } E_s\rho(X, Y) \leqq d\}, \quad (13)$$

and $H(s)$ is the joint entropy of $s \in \mathcal{T}_n(\mathcal{A} \times \mathcal{B})$. $H_u(t, r, d)$ is called the upper joint entropy in [4]. Moreover,

$$I_l(t, r, d) \equiv H(t) + H(r) - H_u(t, r, d) \quad (14)$$

is defined as the lower mutual information. Note that

$$R(p, d) = I_l(p, Q^*, d), \quad (15)$$

where Q^* is the optimal distribution on $\mathcal{B}$ associated with $R(p, d)$.

In this paper, two kinds of redundancy, i.e., rate redundancy and distortion redundancy, are considered. For the rate redundancy of fixed distortion coding, we introduce d-semifaithful code C_n as a fixed distortion code that satisfies, for block length n,

$$\rho_n(x^n, C_n) \leqq d, \quad \text{for any } x^n \in \mathcal{A}^n, \quad (16)$$

where

$$\rho_n(x^n, C_n) = \min_{y^n \in C_n} \rho_n(x^n, y^n). \quad (17)$$

Letting $r_n(x^n, C_n)$ (nats/source symbol) be the codeword length per source symbol when $x^n \in \mathcal{A}^n$ is encoded by C_n, and letting $r_n(C_n)$ (nats/source symbol) be the coding rate, i.e., the average codeword length per source symbol, then the rate redundancy of d-semifaithful code C_n is defined as

$$R_n(C_n) = r_n(C_n) - R(p, d). \quad (18)$$

Similarly the distortion redundancy of fixed rate coding is defined. Letting $\rho_n(C_n)$ be the average distortion with respect to block code C_n whose rate is R (nats/source symbol), the distortion redundancy of block code C_n is defined as

$$D_n(C_n) = \rho_n(C_n) - d(p, R). \quad (19)$$

3. Main Results

In this paper, we assume that a source is known to belong to a class of memoryless sources. However, the probability distribution p of the source is unknown. We consider the universal codes proposed by Davisson [5], [6], in which a proper code is prepared for each n-type t and a source sequence x^n is encoded by its code if $t(x^n) = t$. Encoded data and the information about the n-type of the sequence are sent from the encoder to the decoder. We call this kind of codes *Davisson-like* universal codes.

The following theorems about the redundancy of universal codes can be proved by using *Davisson-like* universal coding.

Theorem 1: For a class of memoryless sources and sufficiently large n, there exists a universal d-semifaithful code C_n^U such that

$$r_n(C_n^U) \leqq R(p, d) + \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + o\left(\frac{\ln n}{n}\right). \quad (20)$$

Theorem 2: For a class of memoryless sources and sufficiently large n, there exists a universal block code C_n^U with rate R such that

$$\rho_n(C_n^U) \leqq d(p, R) - \left(A - \frac{1}{2}\right)\frac{\partial}{\partial R}d(p, R)\frac{\ln n}{n} + o\left(\frac{\ln n}{n}\right). \quad (21)$$

In the next section, we will prove Theorem 1. Theorem 2 can be proved similarly, and hence only the outline of the proof is given in Appendix A.

4. Proof of Theorem 1

For each n-type t, prepare a data base Y_t such as

$$Y_t=\{Y^n_{t1},Y^n_{t2},\cdots,Y^n_{tj},\cdots Y^n_{t2^{\lambda(t)}}\ :\ Y^n_{tj}\in T^n_Y(Q_t)\}, \tag{22}$$

where $||Q_t - Q^*_t|| \leqq \frac{\beta}{n}, \beta \geqq 0$, Q^*_t is the optimal distribution on $\mathcal{B}$ associated with $R(t,d)$, and $||\cdot||$ is the Euclidean distance. We assume that $\lambda(t)$ determining the size of Y_t depends on t via a parameter $\tilde{R}_t$, which is given later, as follows.

$$\lambda(t) = \left\lceil \frac{n\tilde{R}_t}{\ln 2}\right\rceil < \frac{n\tilde{R}_t}{\ln 2} + 1. \tag{23}$$

For a given data base Y_t and a source output X^n, the recurrence time $N_n(Y_t, X^n, d)$ is defined as the smallest positive integer j such that for d-ball $B(X^n, d)$,

$$Y^n_{tj} \in B(X^n, d). \tag{24}$$

If any Y^n_{tj} of Y_t does not satisfy (24), then we set $N_n(Y_t, X^n, d) = \infty$.

We use the set of Y_t, say $\mathcal{C}^U_n$, as a universal code book of block length n, i.e.,

$$\mathcal{C}^U_n = \{Y_t\ :\ t \in \mathcal{T}_n(\mathcal{A})\}. \tag{25}$$

If $t(X^n) = t$ and $N_n(Y_t, X^n, d) = i \leq 2^{\lambda(t)}$, then we encode X^n as (t, i). Note that the distortion tolerance d can be achieved since the decoder can reproduce Y^n_{ti} from (t, i) and $\mathcal{C}^U_n$. In case of $N_n(Y_t, X^n, d) = \infty$, we send $Y^n_* \in \mathcal{B}^n$ that satisfies $d(X^n, Y^n_*) = \min_{Y^n\in\mathcal{B}^n} d(X^n, Y^n)$.

In order to combine these two cases, we use the following binary codeword,

$$C(X^n) = \begin{cases} 0[t]_{\lambda_0}[i]_{\lambda(t)}, \\ \quad \text{if } N_n(Y_t, X^n, d) = i \leqq 2^{\lambda(t)} \\ \qquad \text{(Case A)}, \\ 1\langle Y^n_*\rangle_{\lambda_1}, \\ \quad \text{otherwise (Case B)}, \end{cases} \tag{26}$$

where $[i]_\lambda$ is the ordinary binary representation of i with λ bits and $\langle Y^n_*\rangle_\lambda$ is λ bits representation of Y^n_*. In (26), the first bit of $C(X^n)$ is used as a flag bit.

We note that λ_1 must satisfy†,

$$\lambda_1 = \lceil n \log_2 B\rceil < n \log_2 B + 1 \tag{27}$$

because of $Y^n_* \in \mathcal{B}^n$. On the other hand, λ_0 can be bounded by

$$\lambda_0 = \lceil \log_2 |\mathcal{T}_n(\mathcal{A})|\rceil < (A - 1) \log_2(n + 1) + 1 \tag{28}$$

because we have

$$|\mathcal{T}_n(\mathcal{A})| \leqq (n + 1)^{A-1} \tag{29}$$

from the Type Counting Lemma [7]. Hence, for a given x^n, the codeword length per source symbol, $r_n(x^n, \mathcal{C}^U_n)$ (nats/source symbol), can be bounded as follows.

$$r_n(x^n,\mathcal{C}^U_n)\leqq \begin{cases} \tilde{R}_t+\dfrac{(A-1)\ln(n+1)+3\ln 2}{n}, \\ \quad \text{(Case A)}, \\ \ln B+\dfrac{2\ln 2}{n}, \\ \quad \text{(Case B)}. \end{cases} \tag{30}$$

In order to optimize $\tilde{R}_t$ and the coding rate, we introduce a random code ensemble where $Y^n_{tj} \in Y_t$ is i. i. d. and uniformly distributed on $T^n_Y(Q_t)$. In this code ensemble, we have that for $x^n \in T^n_X(t)$,

$$\begin{aligned} &\Pr\left[\frac{\ln N_n(Y_t, X^n, d)}{n} > R_t | X^n = x^n\right] \\ &\leqq \left(1 - \frac{F_n(t, Q_t, d)}{|T^n_Y(Q_t)|}\right)^{e^{nR_t}} \\ &\leqq e^{-\frac{F_n(t,Q_t,d)}{|T^n_Y(Q_t)|}e^{nR_t}}. \end{aligned} \tag{31}$$

Since we can easily show that

$$|T^n_Y(Q_t)| = e^{nH(Q_t)-\frac{B-1}{2}\ln n+\mathrm{O}(1)}, \tag{32}$$

(31) is further upper bounded as follows:

$$\begin{aligned} e^{-\frac{F_n(t,Q_t,d)}{|T^n_Y(Q_t)|}e^{nR_t}} &\overset{1)}{\leqq} e^{-e^{nR_t-nI_l(t,Q_t,d)-\frac{1}{2}\ln n+\mathrm{O}(1)}} \\ &\overset{2)}{=} e^{-e^{nR_t-nI_l(t,Q^*_t,d)-\frac{1}{2}\ln n+\mathrm{O}(1)}} \\ &= e^{-e^{nR_t-nR(t,d)-\frac{1}{2}\ln n+\mathrm{O}(1)}}, \end{aligned} \tag{33}$$

where 1) is derived from (11) and (14), and 2) is due to the following fact: Since $I_l(\cdot,\cdot,\cdot)$ is lower semicontinuous in its domain [4], it satisfies that for $||Q_t - Q^*_t|| \leqq \frac{\beta}{n}$,

$$\begin{aligned} I_l(p, Q_t, d) &= I_l(p, Q^*_t, d) + \mathrm{O}\left(||Q_t - Q^*_t||\right) \\ &= I_l(p, Q^*_t, d) + \mathrm{O}\left(\frac{1}{n}\right). \end{aligned} \tag{34}$$

We note that if we use

$$\begin{aligned} R_t = \tilde{R}_t \equiv R(t, d) + \frac{1}{2}\frac{\ln n}{n} + \frac{\ln\ln n^\sigma}{n} \\ + \mathrm{O}\left(\frac{1}{n}\right), \quad \sigma > 0, \end{aligned} \tag{35}$$

then $\tilde{R}_t$ satisfies from (31) and (33) that for a proper σ,

$$\Pr\left[\frac{\ln N_n(Y_t, X^n, d)}{n} > \tilde{R}_t | X^n = x^n\right] \leqq \frac{1}{n}. \tag{36}$$

This means that in the random code ensemble the probability of "Case B" is bounded by

† $\lceil k\rceil$ is the smallest integer that is not less than k.

$$\Pr[\text{Case B}] \leqq \frac{1}{n}. \tag{37}$$

Now, we consider the ensemble average of coding rate $r_n(\mathcal{C}_n^U)$, which can be represented for some $\alpha > 0$ as follows:

$$\begin{aligned} &E_{\mathcal{C}_n^U} r_n(\mathcal{C}_n^U) \\ &= \sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} \sum_{x^n \in T_X^n(t)} p^n(x^n) \\ &\qquad \times E_{\mathcal{C}_n^U} r_n(x^n, \mathcal{C}_n^U) \\ &+ \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} \sum_{x^n \in T_X^n(t)} p^n(x^n) \\ &\qquad \times E_{\mathcal{C}_n^U} r_n(x^n, \mathcal{C}_n^U). \end{aligned} \tag{38}$$

In case of $x^n \in T_X^n(t)$, we have from (30), (35), and (37) that

$$\begin{aligned} &E_{\mathcal{C}_n^U} r_n(x^n, \mathcal{C}_n^U) \\ &\leqq \tilde{R}_t + \frac{(A-1)\ln(n+1)+3\ln 2}{n} \\ &\quad + \Pr[\text{Case B}]\left\{\ln B + \frac{2\ln 2}{n}\right\} \\ &\leqq \tilde{R}_t + \frac{(A-1)\ln(n+1)+3\ln 2}{n} \\ &\quad + \frac{1}{n}\left\{\ln B + \frac{2\ln 2}{n}\right\} \\ &= R(t,d) + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right). \end{aligned} \tag{39}$$

Thus, (38) is upper bounded by

$$\begin{aligned} E_{\mathcal{C}_n^U} r_n(\mathcal{C}_n^U) \leqq &\sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))\left\{R(t,d)\right. \\ &\left. + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right\} \\ &+ \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))\left\{R(t,d)\right. \\ &\left. + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right\}. \end{aligned} \tag{40}$$

We first consider the case of $||t-p|| > \alpha\sqrt{\frac{2\ln n}{n}}$ in (40). In this case, the relative entropy $D(t||p)$ satisfies

$$D(t||p) > \alpha^2\frac{\ln n}{n} + \mathrm{O}\left(\left(\frac{\ln n}{n}\right)^{3/2}\right) \tag{41}$$

from the Appendix B. Hence for $\alpha^2 > A+1$ and sufficiently large n, we have

$$\begin{aligned} &\sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t)) \\ &\overset{1)}{\leqq} \sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} e^{-nD(t||p)+\mathrm{O}(1)} \\ &\overset{2)}{<} (n+1)^{A-1} e^{-\alpha^2\ln n - \mathrm{O}\left(n^{-1/2}(\ln n)^{3/2}\right)} \\ &\leqq \frac{1}{n^2} + \mathrm{o}\left(\frac{1}{n^2}\right), \end{aligned} \tag{42}$$

where 1) holds because for $x^n \in T_X^n(t)$,

$$\begin{aligned} p^n(T_X^n(t)) &= |T_X^n(t)|\, p^n(x^n) \\ &= e^{-nD(t||p) - \frac{A-1}{2}\ln n + \mathrm{O}(1)} \\ &\leqq e^{-nD(t||p)+\mathrm{O}(1)}, \end{aligned} \tag{43}$$

and 2) is derived from (29) and (41). Since $R(t,d) \leqq \ln B$, we obtain that for $\alpha^2 > A+1$,

$$\begin{aligned} &\sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))\left\{R(t,d)\right. \\ &\left. + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right\} \\ &< \left(\frac{1}{n^2} + \mathrm{o}\left(\frac{1}{n^2}\right)\right) \\ &\quad \times \left\{\ln B + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right\} \\ &= \mathrm{O}\left(\frac{1}{n^2}\right). \end{aligned} \tag{44}$$

Substituting (44) into (40), we have

$$\begin{aligned} E_{\mathcal{C}_n^U} r_n(\mathcal{C}_n^U) \leqq &\sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))R(t,d) \\ &+ \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right). \end{aligned} \tag{45}$$

The Taylor expansion of $R(t,d)$ around $t = p$ is given by

$$\begin{aligned} R(t,d) = &R(p,d) + \left\langle \frac{\partial R(p,d)}{\partial t}, t-p \right\rangle \\ &+ (t-p)\left(\frac{1}{2}\frac{\partial^2 R(p,d)}{\partial t^2}\right)(t-p)' + \mathrm{o}(||t-p||^2), \end{aligned} \tag{46}$$

where $< \cdot, \cdot >$ is inner product, and t' represents the transpose of vector t. Substituting (46) into (45),

$$\begin{aligned} &E_{\mathcal{C}_n^U} r_n(\mathcal{C}_n^U) \\ &\leqq \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))R(p,d) \\ &+ \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))\left\langle \frac{\partial R(p,d)}{\partial t}, t-p \right\rangle \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{t:\|t-p\|\leq\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))(t-p) \\
&\cdot \left(\frac{1}{2}\frac{\partial^2 R(p,d)}{\partial t^2}\right)(t-p)' + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} \\
&+\mathrm{o}\left(\frac{\ln n}{n}\right) \\
&\overset{1)}{=} \sum_{t\in\mathcal{T}_n(\mathcal{A})} p^n(T_X^n(t))R(p,d) \\
&+ \sum_{t\in\mathcal{T}_n(\mathcal{A})} p^n(T_X^n(t))\left\langle\frac{\partial R(p,d)}{\partial t}, t-p\right\rangle \\
&+ \sum_{t\in\mathcal{T}_n(\mathcal{A})} p^n(T_X^n(t))(t-p)\left(\frac{1}{2}\frac{\partial^2 R(p,d)}{\partial t^2}\right) \\
&\cdot(t-p)' + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right), \quad (47)
\end{aligned}
$$

where 1) holds because $R(p,d)$, $\left\langle\frac{\partial R(p,d)}{\partial t}, t-p\right\rangle$, and $(t-p)\left(\frac{1}{2}\frac{\partial^2 R(p,d)}{\partial t^2}\right)(t-p)'$ are at most $\mathrm{O}(1)$, respectively, and (42) means

$$
\sum_{t:\|t-p\|>\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t))\mathrm{O}(1) < \mathrm{O}\left(\frac{1}{n^2}\right).
$$

We can easily show that

$$
\sum_{t\in\mathcal{T}_n(\mathcal{A})} p^n(T_X^n(t))\left\langle\frac{\partial R(p,d)}{\partial t}, t-p\right\rangle = 0. \quad (48)
$$

Moreover, noting that for $t=(t_1,t_2,\cdots,t_a,\cdots,t_A)$ and $p_a=\Pr[X=a], a\in\mathcal{A}$,

$$
\begin{aligned}
&\sum_{t\in\mathcal{T}_n(\mathcal{A})} p^n(T_X^n(t))(t_a-p_a)(t_{\tilde a}-p_{\tilde a}) \\
&= \begin{cases} \frac{1}{n}p_a(1-p_a), & \text{if } a=\tilde a, \\ \frac{1}{n}p_a p_{\tilde a}, & \text{if } a\neq\tilde a, \end{cases} \quad (49)
\end{aligned}
$$

we have

$$
\begin{aligned}
&\sum_{t\in\mathcal{T}_n(\mathcal{A})} p^n(T_X^n(t))(t-p)\left(\frac{1}{2}\frac{\partial^2 R(p,d)}{\partial t^2}\right)(t-p)' \\
&= \mathrm{O}\left(\frac{1}{n}\right). \quad (50)
\end{aligned}
$$

Hence, from (47), (48) and (50), we finally obtain that

$$
\begin{aligned}
E_{\mathcal{C}_n^U} r_n(\mathcal{C}_n^U) &\leq R(p,d) + \left(A-\frac{1}{2}\right)\frac{\ln n}{n} \\
&\quad + \mathrm{o}\left(\frac{\ln n}{n}\right). \quad (51)
\end{aligned}
$$

Since, in the code ensemble, there exists at least one code $\mathcal{C}_n^U$ such that $r_n(\mathcal{C}_n^U) \leq E_{\mathcal{C}_n^U} r_n(\mathcal{C}_n^U)$, we obtain (20). □

5. Concluding Remarks

In this paper, we obtained the new upper bounds of the redundancy in *Davisson-like* universal codes.

In [5], the upper bound of the rate redundancy in *Davisson-like* universal d-semifaithful code is obtained as

$$
(AB+A+4)\frac{\ln n}{n} + \mathrm{O}\left(\frac{1}{n}\right). \quad (52)
$$

Comparing this bound with (20) in Theorem 1, we note that our bound is much tighter than (52). Actually, we can show that our upper bound is the attainable minimum in any *Davisson-like* universal d-semifaithful codes as follows.

$|\mathcal{T}_n(\mathcal{A})|$ is given by

$$
\begin{aligned}
&|\mathcal{T}_n(\mathcal{A})| \\
&= \binom{n+A-1}{A-1} = \frac{(n+A-1)!}{n!(A-1)!} \\
&\overset{1)}{=} \frac{\sqrt{2\pi(n+A-1)}}{(A-1)!\sqrt{2\pi n}}\frac{(n+A-1)^{n+A-1}}{n^n}e^{-A+1+\varepsilon} \\
&= \frac{\left(1+\frac{A-1}{n}\right)^{n+\frac{1}{2}}}{(A-1)!}(n+A-1)^{A-1}e^{-A+1+\varepsilon}, \quad (53)
\end{aligned}
$$

where $|\varepsilon| \leq \mathrm{O}(1/n)$, and 1) follows from the well-known Stirling formula $n! = \sqrt{2\pi n}n^n e^{-n+\mathrm{O}(\frac{1}{n})}$. Noting

$$
\left(1+\frac{A-1}{n}\right)^{n+\frac{1}{2}} \geq 1, \quad (54)
$$

$|\mathcal{T}_n(\mathcal{A})|$ is lower bounded by $\mathrm{O}(n^{A-1})$. On the other hand, it is known that when source distribution is given, the lower bound of rate redundancy is $\frac{1}{2}\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)$ [4]. Hence, combining these two bounds, the rate redundancy in *Davisson-like* universal d-semifaithful code must be lower bounded by $(A-\frac{1}{2})\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)$. This equals the upper bound obtained by Theorem 1, which means that Theorem 1 is tight in a *Davisson-like* universal d-semifaithful code.

Similarly, we can easily show that the upper bound of the distortion redundancy in Theorem 2 is actually the attainable minimum in any *Davisson-like* universal fixed rate codes.

References

[1] J. Rissanen, "Universal coding, information, prediction and estimation," IEEE Trans. Inf. Theory, vol.IT-30, no.4, pp.629–636, July 1984.
[2] R.J. Pilc, "The transmission distortion of a source as a function of the encoding block length," Bell Syst. Tech. J., vol.47, pp.827–885, 1968.
[3] Z. Zhang, E. Yang, and V. Wei, "On the redundancy of lossy source coding," Proc. 1995 IEEE ISIT, p.190, Sept. 1995.
[4] Z. Zhang, E. Yang, and V. Wei, "The redundancy of source

coding with a fidelity criterion—Part one: Known statistics," IEEE Trans. Inf. Theory, vol.43, no.1, pp.71–91, Jan. 1997.
[5] B. Yu and T.P. Speed, "A rate of convergence result for a universal D-semifaithful code," IEEE Trans. Inf. Theory, vol.39, no.3, pp.813–820, May 1993.
[6] L.D. Davisson, "Universal noiseless coding," IEEE Trans. Inf. Theory, vol.IT-19, no.6, pp.783–795, Nov. 1973.
[7] I. Csiszár and J. Körner, "Information Theory: Coding Theorems for Discrete Memoryless Systems," Academic Press, New York, 1981.
[8] D. Ishii, "The asymptotic property of source coding with a fidelity criterion," Master thesis, Department of Mathematical Engineering and Information Physics, Univ. of Tokyo, Tokyo, March 1996.

Appendix A: Rough Proof of Theorem 2

For each n-type t, prepare a fixed rate code $\mathcal{C}_t$ such as

$$\mathcal{C}_t = \{Y_{t1}^n, Y_{t2}^n, \cdots, Y_{tM}^n : Y_{tj}^n \in T_Y^n(Q_t) \text{ and } M = \lfloor e^{nR-\lambda_0 \ln 2} \rfloor\}, \qquad \text{(A·1)}$$

and a universal code $\mathcal{C}_n^U = \{\mathcal{C}_t : t \in \mathcal{T}_n(\mathcal{A})\}$. When $t(x^n) = t$, we select Y_{tj}^n from $\mathcal{C}_t$ that minimizes $\rho_n(X^n, Y_{tj}^n)$. Then, we encode X^n as (t, j) and the decoder reproduces $Y_{tj}^n \in \mathcal{C}_t$. Letting $\lambda_0 = \lceil |\log_2 |\mathcal{T}_n(\mathcal{A})| \rceil$ bits, (t, j) can be represented by a fixed length code with rate R.

In order to evaluate the average distortion of this code, we introduce a random code ensemble where $Y_{tj}^n \in \mathcal{C}_t$ is i. i. d. and uniformly distributed on $T_Y^n(Q_t)$.

For $\tilde{d}_t$ that satisfies

$$nR - \lambda_0 \ln 2 - nR(t, \tilde{d}_t) - \frac{\ln n}{2} + \mathrm{O}(1) = \ln \ln n^\sigma, \ \sigma > 0, \qquad \text{(A·2)}$$

we can show in the same way as (36) that for a proper σ,

$$\Pr[\rho_n(X^n, \mathcal{C}_t) > \tilde{d}_t | X^n = x^n] \leqq \frac{1}{n}. \qquad \text{(A·3)}$$

From (A·2) and (28), we obtain that

$$\begin{aligned}\tilde{d}_t &= d\left(t, R - \frac{\lambda_0 \ln 2}{n} - \frac{\ln n}{2n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right) \\ &\leqq d\left(t, R - \frac{(A-1)\ln(n+1) + \ln 2}{n} - \frac{\ln n}{2n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right) \\ &= d\left(t, R - \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right). \qquad \text{(A·4)}\end{aligned}$$

We now evaluate the ensemble average of $\rho_n(\mathcal{C}_n^U)$. For some $\alpha > 0$,

$$\begin{aligned}&E_{\mathcal{C}_n^U} \rho_n(\mathcal{C}_n^U) \\ &= \sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} \sum_{x^n \in T_X^n(t)} p^n(x^n) E_{\mathcal{C}_n^U} \rho_n(x^n, \mathcal{C}_t) \\ &\quad + \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} \sum_{x^n \in T_X^n(t)} p^n(x^n) E_{\mathcal{C}_n^U} \rho_n(x^n, \mathcal{C}_t) \\ &\leqq \sum_{t:||t-p||>\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t)) \rho_{\max} \\ &\quad + \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} \sum_{x^n \in T_X^n(t)} p^n(x^n) E_{\mathcal{C}_n^U} \rho_n(x^n, \mathcal{C}_t).\end{aligned} \qquad \text{(A·5)}$$

Using (A·3) and (A·4), we obtain that for $x^n \in T_X^n(t)$,

$$\begin{aligned}&E_{\mathcal{C}_n^U} \rho_n(x^n, \mathcal{C}_t) \\ &\leqq \tilde{d}_t + \Pr\{\rho_n(X^n, \mathcal{C}_t) > \tilde{d}_t | X^n = x^n\} \rho_{\max} \\ &\leqq d\left(t, R - \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right) + \mathrm{O}\left(\frac{1}{n}\right).\end{aligned} \qquad \text{(A·6)}$$

Thus (A·5) is upper bounded from (42) and (A·6) as follows: For $\alpha^2 > A + 1$,

$$\begin{aligned}E_{\mathcal{C}_n^U} \rho_n(\mathcal{C}_n^U) \leqq& \sum_{t:||t-p||\leqq\alpha\sqrt{\frac{2\ln n}{n}}} p^n(T_X^n(t)) \\ &\times d\left(t, R - \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right) + \mathrm{O}\left(\frac{1}{n}\right).\end{aligned} \qquad \text{(A·7)}$$

Expanding $d\left(t, R - \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right)$ around $t = p$ in the same way as the proof of Theorem 1, we finally obtain that

$$\begin{aligned}&E_{\mathcal{C}_n^U} \rho_n(\mathcal{C}_n^U) \\ &\leqq d\left(p, R - \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right)\right) + \mathrm{o}\left(\frac{\ln n}{n}\right) \\ &= d(p, R) - \frac{\partial}{\partial R} d(p, R) \left(A - \frac{1}{2}\right)\frac{\ln n}{n} + \mathrm{o}\left(\frac{\ln n}{n}\right).\end{aligned} \qquad \text{(A·8)}$$

□

Appendix B: Derivation of (41)

For $t = (t_1, t_2, \cdots, t_A)$ and $p = (p_1, p_2, \cdots, p_A)$, relative entropy $D(t||p)$ is defined as

$$D(t||p) = \sum_{a \in \mathcal{A}} t_a \ln \frac{t_a}{p_a}. \qquad \text{(A·9)}$$

The Taylor expansion of $D(t||p)$ at $t = p$ is given by

$$D(t||p) = D(p||p) + \left\langle \frac{\partial D(p||p)}{\partial t}, t-p \right\rangle + (t-p)\left(\frac{1}{2}\frac{\partial^2 D(p||p)}{\partial t^2}\right)(t-p)' + O(||t-p||^3), \quad \text{(A· 10)}$$

where $\langle \cdot, \cdot \rangle$ is inner product, and t' represents the transpose of vector t.

It is well known that

$$D(p||p) = 0. \quad \text{(A· 11)}$$

Since

$$\frac{\partial D(p||p)}{\partial t_a} = 1 \quad \text{(A· 12)}$$

holds for every t_a, we have

$$\left\langle \frac{\partial D(p||p)}{\partial t}, t-p \right\rangle = \sum_{a \in \mathcal{A}} (t_a - p_a) = 0. \quad \text{(A· 13)}$$

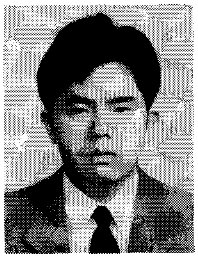
Furthermore, since we can easily show that

$$\frac{\partial^2 D(p||p)}{\partial t_a \partial t_{\tilde{a}}} = \begin{cases} 1/p_a, & \text{if } a = \tilde{a}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{(A· 14)}$$

we obtain that for $||t-p|| > \alpha\sqrt{\frac{2\ln n}{n}}$,

$$(t-p)\left(\frac{1}{2}\frac{\partial^2 D(p||p)}{\partial t^2}\right)(t-p)' \geq \frac{1}{2}||t-p||^2 > \alpha^2 \frac{\ln n}{n}. \quad \text{(A· 15)}$$

Thus, substituting (A· 11), (A· 13) and (A· 15) into (A· 10), (41) is obtained.



Daiji Ishii was born in Kanagawa, Japan, on November 22, 1970. He received the B.E. and M.E. degrees in mathematical engineering from University of Tokyo, Tokyo, Japan, in 1994 and 1996, respectively. From 1996, he joined NEC Corp. His research interest is in information theory, in particular, rate distortion theory.

Hirosuke Yamamoto was born in Wakayama, Japan, on November 15, 1952. He received the B.E. degree from Shizuoka University, Shizuoka, Japan, in 1975 and the M.E. and Dr.E. degrees from the University of Tokyo, Tokyo, Japan, in 1977 and 1980, respectively, all in electrical engineering. In 1980 he joined Tokushima University, Tokushima, Japan. He was an Associate Professor at Tokushima University from 1983 to 1987, and at University of Electro-Communications, Tokyo, Japan, from 1987 to 1993. Since 1993 he has been an Associate Professor in the Department of Mathematical Engineering and Information Physics, Faculty of Engineering, University of Tokyo, Tokyo, Japan. In 1989-90, he was a Visiting Scholar at the Information Systems Laboratory, Stanford University. His research interests are in Shannon theory, coding theory, cryptography, and communication theory. Dr. Yamamoto is a member of the IEEE.