

A Comment on "A Rate of Convergence Result for a Universal D -Semifaithful Code"

Neri Merhav, *Senior Member, IEEE*

Abstract—In the above paper,¹ Yu and Speed propose a universal pointwise D -semifaithful code whose expected compression ratio, for discrete memoryless sources, approaches the rate-distortion function at a rate $O(n^{-1} \log n)$. They also conjecture that this is the fastest achievable convergence rate for pointwise D -semifaithful codes. In this correspondence, we use a simple extension of Kraft's inequality and prove that this conjecture is true, at least for the Hamming distortion measure.

Index Terms—Rate-distortion theory, universal coding, D -semifaithful codes, memoryless sources.

I. INTRODUCTION

In the above paper, Yu and Speed have considered the following setting of rate-distortion coding: Let $\mathcal{A} = \{1, 2, \dots, \mathcal{J}\}$ and $\mathcal{B} = \{1, 2, \dots, \mathcal{K}\}$ denote the source alphabet and the reproduction alphabet, respectively, and let $P = \{P(1), P(2), \dots, P(\mathcal{J})\}$ denote the vector of letter probabilities that characterizes the source. A single-letter fidelity criterion is used to measure the distortion between a source string $x^n = (x_1, \dots, x_n) \in \mathcal{A}^n$ and its reproduction $y^n = (y_1, \dots, y_n) \in \mathcal{B}^n$, that is

$$d_n(x^n, y^n) = n^{-1} \sum_{i=1}^n d(x_i, y_i)$$

where $d(x, y)$ is assumed finite for every $(x, y) \in \mathcal{A} \times \mathcal{B}$. We consider a class of variable-length rate-distortion encoder-decoders defined as follows. At the encoder, every source string x^n is mapped to a variable-length binary codeword of $L(x^n)$ bits where the length function $L(\cdot)$ is designed such that any stream of codewords is uniquely decipherable (UD). The decoder in turn, maps the binary codeword into $y^n \in \mathcal{B}^n$. The overall encoder-decoder mapping is denoted by $y^n = M(x^n)$. A code is called *pointwise D -semifaithful* if for every $x^n \in \mathcal{A}^n$, $d_n(x^n, M(x^n)) \leq D$. The goal of designing a pointwise D -semifaithful code is, of course, to minimize the compression ratio $E\{L(x^n)\}/n$ for a given distortion level D , namely, to achieve a compression ratio as close as possible to the rate-distortion function of the source $R(D)$. A universal sequence of D -semifaithful codes is independent of the unknown source P and yet asymptotically achieves $R(D)$ as $n \rightarrow \infty$.

In the above paper, Yu and Speed have proposed a universal pointwise D -semifaithful coding scheme by partitioning \mathcal{A}^n into sets of typical sequences and covering each typical set by a sphere whose normalized radius (with respect to the given distortion measure) is D . The codeword consists of two parts: The first part is the binary representation of index of the type and the second is the index of the first sequence in the D -sphere whose normalized distance from x^n is less than or equal to D . It is easy to see that this code falls in the class of variable-rate, pointwise D -semifaithful codes described above. It has been shown in Theorem 2 of the above paper, by a random coding

Manuscript received October 16, 1994; revised February 13, 1995.

The author is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel.
IEEE Log Number 9412186.

¹B. Yu and T. P. Speed, *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 813–820, May 1993.

argument, that the expected compression ratio is less than

$$R(D) + (KJ + J + 4)n^{-1} \log n + O(n^{-1}).$$

In Section IV of the above paper, Yu and Speed conjecture that $n^{-1} \log n$ is indeed the best attainable convergence rate and this belief stems from a known lower bound on the rate-distortion performance (as a function of n) due to Pilc [2]. However, Pilc's lower bound is stated for the dual problem, i.e., fixed coding rate and minimum expected distortion. Specifically, Pilc has shown that even if the source P is known, then for a given fixed rate R , the expected distortion is essentially never less than

$$D(R) + \log n / (2n |S(R)|)$$

where $D(R)$ is the distortion-rate function and $s(R)$ is the slope of the function $R(D)$ at the point where the rate is R . Thus by applying the rate-distortion function on this expression it is seen that this fixed rate R has to be at least as large as

$$R(D) + \frac{1}{2} n^{-1} \log n$$

so that the expected distortion would be less than D (Corollary 2).

In this correspondence, we show that the above conjecture of Yu and Speed is true at least for the case where $\mathcal{A} = \mathcal{B}$ and $d(\cdot, \cdot)$ is the Hamming distortion measure. In other words, we show that even if the source is *known*, the lower bound on the expected rate is essentially

$$R(D) + \frac{1}{2} n^{-1} \log n$$

under the setting of variable-rate, pointwise D -semifaithful coding as studied by Yu and Speed. If, in addition, the letter probabilities of the source are unknown, then the coefficient associated with the $n^{-1} \log n$ term must be larger.

It should be noted that Zhang *et al.* [3] have proved independently the above conjecture and its dual distortion rate form for general additive distortion measures by using much heavier technical tools that are applicable to finite-alphabet sources. The message of the present correspondence remains, however, that at least for the widely used Hamming distortion measure, the same lower bound can be obtained in a relatively simple manner, where the main underlying idea is in generalizing Kraft's inequality for rate-distortion coding. This generalization also enables an analogous derivation for continuous-valued sources and difference distortion measures, e.g., the Gaussian source and the squared error distortion measure.

II. THE LOWER BOUND

We start from an extended version of Kraft's inequality. Suppose that $\mathcal{A} = \mathcal{B}$ and define the subtraction operation $x - y$, $x, y \in \mathcal{A}$ as subtraction modulo J . Let $d(\cdot, \cdot)$ be a difference distortion measure, that is, $d(x, y) = \rho(x - y)$ where ρ is a function from \mathcal{A} to the nonnegative reals. Let $S_n(D)$ denote the set of all sequences $z^n = (z_1, \dots, z_n) \in \mathcal{A}^n$ with the property

$$n^{-1} \sum_{i=1}^n \rho(z_i) \leq D$$

and let $|S_n(D)|$ denote the cardinality of $S_n(D)$. The following is an extension of Kraft's inequality for variable-rate pointwise D -semifaithful codes for n -sequences.

Lemma 1: For any pointwise D -semifaithful code for n -sequences

$$\sum_{x^n \in \mathcal{A}^n} 2^{-L(x^n)} \leq |S_n(D)|. \quad (1)$$

Proof: Let us denote by C_n the collection of all reproduction vectors $\{M(x^n), x^n \in \mathcal{A}^n\}$. Also, since $L(x^n)$ depends only on the codeword of x^n , which in turn, is related to y^n by a one-to-one correspondence (with no loss in optimality), we shall also denote the length function by $L(y^n)$ with a slight abuse of notation. We then have

$$\begin{aligned} \sum_{x^n \in \mathcal{A}^n} 2^{-L(x^n)} &= \sum_{y^n \in C_n} \sum_{x^n: M(x^n)=y^n} 2^{-L(x^n)} \\ &\leq |S_n(D)| \sum_{y^n \in C_n} 2^{-L(y^n)} \end{aligned} \quad (2)$$

where the inequality follows from the obvious facts that all sequences that are mapped to the same y^n have the same codeword length and that for a pointwise D -semifaithful code no more than $|S_n(D)|$ sequences can be mapped to the same codeword, for if this were not the case then at least one source sequence would have been mapped to a reproduction vector y^n with distortion strictly larger than D . Finally, since y^n is assumed uniquely decipherable from the stream of codewords, then the ordinary version of Kraft's inequality must be satisfied with respect to the set of all y^n in C_n , that is

$$\sum_{y^n \in C_n} 2^{-L(y^n)} \leq 1. \quad (3)$$

This completes the proof of the Lemma.

The above Lemma tells us that the best one can expect from a length function of a pointwise D -semifaithful code is that $2^{-L(x^n)}/|S_n(D)|$ would be a probability measure, namely

$$L(x^n) \geq -\log Q(x^n) - \log |S_n(D)| \quad (4)$$

for some probability measure $Q(\cdot)$. The first term can be optimized by an appropriate design of a code, while the second depends only on the distortion measure ρ and the distortion level D . Specifically, if P is known then $Q = P$ minimizes the expectation of the right-hand side and we get

$$E\{L(x^n)\} \geq nH - \log |S_n(D)| \quad (5)$$

where H is the entropy of the source defined as

$$H = -\sum_{j=1}^J P(j) \log P(j). \quad (6)$$

It remains, therefore, to provide a tight upper bound on $\log |S_n(D)|$. We shall henceforth assume the Hamming distortion measure, i.e., $\rho(z) = 0$ if $z = 0$ and $\rho(z) = 1$ otherwise. The exact expression for $|S_n(D)|$ (assuming that nD is integer) is the following:

$$|S_n(D)| = \sum_{l=0}^{nD} \frac{n!}{l!(n-l)!} \cdot (J-1)^l \quad (7)$$

where the l th term in this summation is the number of sequences z^n for which $\rho(z^n) = l$: For each value of l there are $n!/l!(n-l)!$ combinations of error positions, and in each one of them, every one

of the $J-1$ erroneous symbols can replace the correct source symbol. Our next step is to show that this sum is upper-bounded, within a constant that depends only on J and D , by the last term provided that $D < 1/2$; namely, the number of sequences in the interior of $S_n(D)$ is within a constant factor, the same as the number of sequences on the surface. Finally, we shall estimate the last summand by Stirling's formula.

The first goal can be achieved by bounding the above summation by a geometric series. Let us denote the l th term in this summation by α_l . Then

$$\begin{aligned} \frac{\alpha_{l-1}}{\alpha_l} &= \frac{l}{n-l+1} \cdot \frac{1}{J-1} \\ &\leq \frac{nD}{n-nD} \cdot \frac{1}{J-1} \\ &= \frac{D}{(1-D)(J-1)} \triangleq \beta < 1 \end{aligned} \quad (8)$$

where the first inequality is obtained by omitting the $+1$ term in the denominator and by using the fact that $l/(n-l)$ is monotonically increasing with l . Since we have shown that $\alpha_{l-1} \leq \beta\alpha_l$, and hence by iterating, $\alpha_{l-j} \leq \beta^j\alpha_l$, we can now bound $|S_n(D)|$ in the following manner:

$$\begin{aligned} |S_n(D)| &= \sum_{l=0}^{nD} \alpha_{nD-l} \\ &\leq \alpha_{nD} \sum_{l=0}^{nD} \beta^l \\ &\leq \alpha_{nD} \sum_{l=0}^{\infty} \beta^l \\ &= \frac{\alpha_{nD}}{1-\beta} \\ &= \frac{1}{1-\beta} \cdot (J-1)^{nD} \frac{n!}{(nD)!(n-nD)!}. \end{aligned} \quad (9)$$

Assuming that D is strictly positive, we now apply Stirling's formula, $n! \approx (n/e)^n \sqrt{2\pi n}$, and get

$$\log |S_n(D)| \leq nD \log(J-1) + nh(D) - \frac{1}{2} \log n + O(1) \quad (10)$$

where

$$h(D) = -D \log D - (1-D) \log(1-D)$$

and the $O(1)$ term absorbs the constant term associated with β and the one associated with the Stirling approximation. By substituting the right-hand side of (10) into (5) and normalizing by n , we get

$$\frac{E\{L(x^n)\}}{n} \geq H - D \log(J-1) - h(D) + \frac{1}{2} \cdot \frac{\log n}{n} + O\left(\frac{1}{n}\right) \quad (11)$$

where the first three terms form the rate-distortion function of the source with respect to the Hamming distortion [1].

Remark: The above technique can be generalized to other difference distortion measures. In the general case, the leading term of the bound will be

$$H - \max_{Z: E\rho(Z) \leq D} H(Z)$$

which is the Shannon lower bound [1] of the rate-distortion function, and the convergence rate term stems from the difference between the

log-cardinality of $S_n(D)$ and the quantity

$$\max_{Z: E\rho(Z) \leq D} H(Z).$$

Thus this technique is useful whenever the Shannon lower bound is tight. As another example, consider a Gaussian memoryless source and the squared error distortion measure $\rho(z) = z^2$. In this case, the extended Kraft inequality of Lemma 1 is as follows:

$$\int_{\mathbb{R}^n} dx^n \cdot 2^{-L(x^n)} \leq \text{Vol}\{S_n(D)\} \quad (12)$$

where

$$\text{Vol}\{S_n(D)\} = 2(\pi n D)^{n/2} / [n\Gamma(n/2)]$$

is the volume of the n -dimensional Euclidean sphere with radius \sqrt{nD} . It is easy to check that the difference between the normalized log-volume of this sphere and the maximum entropy of Z subject to the second moment constraint $EZ^2 \leq D$ (which is given by $\frac{1}{2} \log(2\pi e D)$) behaves like $1.5n^{-1} \log n$.

ACKNOWLEDGMENT

Useful discussions with M. Feder are acknowledged.

REFERENCES

- [1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [2] R. J. Pile, "The transition 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. K. Wei, "The redundancy of source coding with a fidelity criterion," preprint, 1994.