and that there exist exactly two nonequivalent $[18, 6, 8]$ codes $G_{18}(1)$ and $G_{18}(2)$. All codes G_n, $G_{18}(1)$, and $G_{18}(2)$ are shortened of the extended binary $[24, 12, 8]$ Golay code. Their weight distributions are listed as follows:

	A_0	A_8	A_{12}	A_{16}
$G_{18}(1)$	1	46	16	1
$G_{18}(2)$	1	45	18	
G_{19}	1	78	48	1
G_{20}	1	130	120	5
G_{21}	1	210	280	21
G_{22}	1	330	616	77
G_{23}	1	506	1288	253

A straightforward application of Theorem 2 shows that all these codes are proper.

ACKNOWLEDGMENT

The authors wish to thank the referees for their detailed comments on the first version of the correspondence which led to significant improvements in the presentation. The work has been done during a visit of S. M. Dodunekov to the Department of Information Theory, Chalmers University of Technology. He would like to thank A. Svensson for his hospitality and Mrs. E. Axelsson and L. Kollberg for their assistance.

REFERENCES

[1] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[2] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.
[3] J. K. Wolf, A. M. Michelson, and A. H. Levesque, "On the probability of undetected error for linear block codes," *IEEE Trans. Commun.*, vol. COM-30, pp. 317–324, Feb. 1982.
[4] V. I. Korzhik, "Bounds on undetected error probability and optimum group codes in a channel with feedback," *Radiotech.*, vol. 20, no. 1, pp. 27–33, 1965. (English translation: *Telecommun. Radio Eng.*, vol. 20, no. 1, pp. 87–92, Jan. 1965.)
[5] J. Massey, "Coding techniques for digital data networks," in *Proc. Int. Conf. on Information Theory and Systems* (NTG-Fachberichte, Berlin, Germany, Sept. 18–20, 1978), vol. 65.
[6] S. K. Leung-Yan-Cheong, E. R. Barnes, and D. U. Friedman, "Some properties of undetected error probability of linear codes," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 110–112, Jan. 1979.
[7] T. Kasami and S. Lin, "On the probability of undetected error for the maximum distance separable codes," *IEEE Trans. Commun.*, vol. COM-32, pp. 998–1006, Sept. 1984.
[8] T. Kløve, "Near-MDS codes for error detection," in *Proc. Int. Workshop on Optimal Codes and Related Topics* (Sozopol, Bulgaria, May 26–June 1, 1995), pp. 103–107.
[9] R. Dodunekova and S. M. Dodunekov, "On the probability of undetected error for near-MDS codes," Preprint 1995-25/ISSN 0347-2809, Chalmers Univ. of Technol. and Göteborg University, Göteborg, Sweden, 1995.
[10] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.
[11] W. W. Peterson and E. J. Weldon Jr., *Error-Correcting Codes*, 2nd ed. Cambridge, MA: MIT Press, 1972.
[12] T. Kløve and V. Korzhik, *Error Detecting Codes*. Boston, MA: Kluver, 1995.
[13] S. Ross, *A First Course in Probability*, 4th ed. New York: Macmillan, 1994.
[14] S. M. Dodunekov and S. B. Encheva, "Uniqueness of some subcodes of the extended binary Golay code," *Probl. Inform. Transm.*, vol. 29, no. 1, pp. 38–43, 1993.

Existence of Optimal Prefix Codes for Infinite Source Alphabets

Tamás Linder, *Member, IEEE*, Vahid Tarokh, *Member, IEEE*, and Kenneth Zeger, *Senior Member, IEEE*

***Abstract*—It is proven that for every random variable with a countably infinite set of outcomes and finite entropy there exists an optimal prefix code which can be constructed from Huffman codes for truncated versions of the random variable, and that the average lengths of any sequence of Huffman codes for the truncated versions converge to that of the optimal code. Also, it is shown that every optimal infinite code achieves Kraft's inequality with equality.**

***Index Terms*—Huffman, lossless coding, prefix codes.**

I. INTRODUCTION

An *alphabet* $\mathcal{A}$ is a finite set and $\mathcal{A}^*$ is the set of all finite-length words formed from the elements of $\mathcal{A}$. For each word $w \in \mathcal{A}^*$, let $l(w)$ denote the word length of w. A D-ary *prefix code* C over an alphabet $\mathcal{A}$ (with $|\mathcal{A}| = D$) is a subset of $\mathcal{A}^*$ with the property that no word in C is the prefix of another word in C. Let $\mathcal{Z}^+$ denote the positive integers.

A sequence of D-ary prefix codes $C_1, C_2, C_3, \cdots$, *converges* to an infinite prefix code C if for every $i \geq 1$, the ith codeword of C_n is eventually constant (as n grows) and equals the ith codeword of C. D-ary prefix codes are known to satisfy Kraft's inequality $\Sigma_{w \in C} D^{-l(w)} \leq 1$. Conversely, any collection of positive integers that satisfies Kraft's inequality corresponds to the codeword lengths of a prefix code [1].

Let X be a source random variable whose countably infinite range is (without loss of generality) $\mathcal{Z}^+$, with respective probabilities $p_1 \geq p_2 \geq p_3 \geq \cdots$, where $p_i > 0$ for all i. The average length of a prefix code $C = \{w_1, w_2, \cdots\}$ to encode X is $\Sigma_{i=1}^{\infty} p_i l(w_i)$. A prefix code C is called *optimal* for a source X if no other prefix code has a smaller average length. The entropy of the random variable X is defined as

$$H(X) = -\sum_{i=1}^{\infty} p_i \log p_i.$$

It is known that the average length of an optimal prefix code is no smaller than $H(X)$ and is smaller than $H(X) + 1$ [1].

The well-known Huffman algorithm [2] gives a method for constructing optimal prefix codes for sources with finite ranges. For each $n \geq 1$, let X_n be a random variable with a finite range and with outcome probabilities $p_i^{(n)} = p_i / S_n$ for $1 \leq i \leq n$ and $p_i^{(n)} = 0$ for $i > n$, where $S_n = \Sigma_{j=1}^{n} p_j$. A *$D$-ary truncated Huffman code* of size n for X is defined to be a D-ary Huffman code for X_n.

Manuscript received June 5, 1996. This work was supported in part by the National Science Foundation, the Hungarian National Foundation for Scientific Research, and the Foundation for Hungarian Higher Education and Research.

T. Linder is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA, on leave from the Technical University of Budapest, Budapest, Hungary.
V. Tarokh is with AT&T Laboratories, Florham Park, NJ 07932 USA.
K. Zeger is with the Department of Electrical and Computer Engineering, the University of California at San Diego, La Jolla, CA 92093-0407 USA.

Publisher Item Identifier S 0018-9448(97)06817-X.

For sources with infinite ranges, several approaches have been taken to construct optimal codes [3]–[7], but in each case some condition on the tail of the probability mass function of the source random variable was assumed. To the best of our knowledge there is no known proof in the literature that optimal codes always exist for sources with infinite ranges.

In this correspondence we present such a proof for sources with finite entropy. In particular, we show that a subsequence of Huffman codes designed for truncated versions of the source random variable X leads to an optimal infinite code for X. We provide an existence proof and cannot, however, specify which Huffman code subsequence is needed. Still, this theorem does suggest that recursive Huffman code construction algorithms might exist for any source, regardless of how fast the tail of its probability mass function decays. We also show that any sequence of truncated Huffman codes indeed converges in the average length sense, whereas only a subsequence is guaranteed to converge in the code sense.

If a source random variable has a finite range then an optimal binary code satisfies Kraft's condition with equality, but not necessarily for D-ary codes when $D \geq 3$. In contrast, our theorem also establishes that for all $D \geq 2$ an optimal D-ary code for a source with an infinite range must satisfy the Kraft inequality with equality.

In [6] it was noted that an optimal code for a source with an infinite range must have a full encoding tree. However, a full encoding tree does not guarantee that Kraft's inequality is satisfied with equality.

A simple counterexample to demonstrate this fact for $D = 2$ is given next. For any $A, B \subset \{0,1\}^*$ let

$$AB = \{ab \in \{0,1\}^* \colon a \in A, b \in B\}.$$

For $n \geq 0$, let $T_n = \{0,1\}^n \setminus \{0^n\}$be the set of all n-bit binary words excluding the all-zeros word, and let Π denote binary word concatenation. Define the prefix code

$$\begin{aligned} C &= \left(\bigcup_{k=2}^{\infty}\left(\prod_{n=2}^{k} T_n\right)0^{k+1}\right) \cup \{00\} \\ &= \{00, 01000, 10000, 11000, \cdots\} \end{aligned}$$

and note that the Kraft sum for C is

$$\begin{aligned} \sum_{w\in C} 2^{-l(w)} &= \tfrac{1}{4} + \sum_{k=2}^{\infty}\left|\left(\prod_{n=2}^{k} T_n\right)0^{k+1}\right| 2^{-\Sigma_{i=2}^{k+1} i} \\ &= \tfrac{1}{4} + \sum_{k=2}^{\infty}(2^{-\Sigma_{i=2}^{k+1} i}) \prod_{n=2}^{k}(2^n - 1) \\ &< \tfrac{1}{4} + \sum_{k=2}^{\infty} 2^{\Sigma_{i=2}^{k} i} 2^{-\Sigma_{i=2}^{k+1} i} \\ &= \sum_{k=1}^{\infty} 2^{-(k+1)} \\ &= 1/2. \end{aligned}$$

Thus the Kraft inequality is strict in this case and it is easy to see that the encoding tree of the code C is full.

II. Main Result

Theorem 1: Let X be a random variable with a countably infinite set of possible outcomes and with finite entropy. Then for every $D > 1$, the following hold:

(I) There exists a sequence of D-ary truncated Huffman codes for X which converges to an optimal code for X.

(II) The average codeword lengths in any sequence of D-ary truncated Huffman codes converge to the minimum possible average codeword length for X.

(III) Any optimal D-ary prefix code for X must satisfy the Kraft inequality with equality.

Proof: For each $n \geq 1$, let C_n be a D-ary truncated Huffman code of size n for X, and denote the sequence of n codeword lengths of C_n (followed by zeros) by

$$l^{(n)} = \{l_1^{(n)}, l_2^{(n)}, \cdots, l_n^{(n)}, 0, 0, 0, \cdots\}.$$

Let $\mathcal{F}$ denote the set of all sequences of positive integers. For each n, the average length $\Sigma_{i=1}^{\infty} l_i^{(n)} p_i^{(n)}$ of Huffman code C_n is not larger than $H(X_n) + 1$, where the entropy of X_n is

$$\begin{aligned} H(X_n) &= -\sum_{i=1}^{n} p_i^{(n)} \log p_i^{(n)} \\ &= -\frac{1}{S_n}\sum_{i=1}^{n} p_i \log p_i - \log \frac{1}{S_n} \\ &\to H(X), \qquad \text{as } n \to \infty \end{aligned}$$

since $S_n = \sum_{i=1}^{n} p_i \to 1$ as $n \to \infty$. Hence

$$H(X_n) + 1 \leq H(X) + 2$$

for n sufficiently large. For each positive integer n, we have

$$\sum_{i=1}^{\infty} p_i l_i^{(n)} \leq (H(X_n) + 1) S_n$$

and, therefore,

$$p_i l_i^{(n)} \leq (H(X_n) + 1) S_n \leq H(X_n) + 1$$

for all i. This implies that

$$l_i^{(n)} \leq (H(X) + 2)/p_i$$

for n sufficiently large.

Thus for each i, the sequence of codeword lengths $\{l_i^{(1)}, l_i^{(2)}, l_i^{(3)}, \cdots\}$ is bounded and therefore the corresponding sequence of codewords can only take on a finite set of possible values. Hence, for each i, there is a convergent subsequence of codewords. In fact, every infinite indexed subset of this sequence of codewords has a convergent subsequence of codewords. We conclude (using a minor modification of [8, Theorem 7.23]) that there exists a subsequence of codes $C_{n_1}, C_{n_2}, C_{n_3}, \cdots$, that converges to an infinite code $\hat{C}$. Clearly, $\hat{C}$ is a prefix code since it is a limit of finite Huffman codes. Furthermore, the subsequence $\{l^{(n_k)}\}$, of elements of $\mathcal{F}$, converges to a sequence $\hat{l} = \{\hat{l}_1, \hat{l}_2, \cdots\} \in \mathcal{F}$, in the sense that for each $i \in \mathcal{Z}^+$, the sequence $l_i^{(n_k)}$ converges to $\hat{l}_i$.

To show the optimality of $\hat{C}$, let $\lambda_1, \lambda_2, \lambda_3, \cdots$, be the codeword lengths of an arbitrary prefix code. For every k, there exists a $j \geq k$ such that $\hat{l}_i = l_i^{(n_m)}$ for every $i \leq k$ provided that $m \geq j$. Thus for

all $m \geq j$, the optimality of Huffman codes implies

$$\sum_{i=1}^{k} p_i^{(k)} \hat{l}_i = \sum_{i=1}^{k} p_i^{(k)} l_i^{(n_m)} = \sum_{i=1}^{n_m} p_i^{(k)} l_i^{(n_m)} \leq \frac{S_{n_m}}{S_k} \sum_{i=1}^{n_m} p_i^{(n_m)} l_i^{(n_m)} \leq \frac{S_{n_m}}{S_k} \sum_{i=1}^{n_m} p_i^{(n_m)} \lambda_i.$$

Therefore,

$$\sum_{i=1}^{k} p_i \hat{l}_i \leq \sum_{i=1}^{n_m} p_i l_i^{(n_m)} \leq \sum_{i=1}^{n_m} p_i \lambda_i \leq \sum_{i=1}^{\infty} p_i \lambda_i \tag{1}$$

and thus

$$\sum_{i=1}^{\infty} p_i \hat{l}_i \leq \sum_{i=1}^{\infty} p_i \lambda_i.$$

This implies that the infinite code $\hat{C}$ is optimal.

To prove part (II) of the theorem, notice that by the optimality of Huffman codes

$$\sum_{i=1}^{n} p_i l_i^{(n)} = S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} \leq S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n+1)} = \sum_{i=1}^{n} p_i l_i^{(n+1)} \leq \sum_{i=1}^{n+1} p_i l_i^{(n+1)}.$$

The sequence $\Sigma_{i=1}^{n} p_i l_i^{(n)}$ is thus an increasing sequence which is bounded above by $H(X)+2$ and has a limit. It follows from (1) that

$$\sum_{i=1}^{\infty} p_i \hat{l}_i \leq \lim_{m \to \infty} \sum_{i=1}^{n_m} p_i l_i^{(n_m)} = \lim_{n \to \infty} \sum_{i=1}^{n} p_i l_i^{(n)}.$$

Next by the optimality of Huffman codes

$$\lim_{n \to \infty} \sum_{i=1}^{n} p_i l_i^{(n)} = \lim_{n \to \infty} S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} \leq \lim_{n \to \infty} S_n \sum_{i=1}^{n} p_i^{(n)} \hat{l}_i = \sum_{i=1}^{\infty} p_i \hat{l}_i.$$

Thus

$$\lim_{n \to \infty} \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} = \lim_{n \to \infty} \frac{1}{S_n} \sum_{i=1}^{n} p_i l_i^{(n)} = \sum_{i=1}^{\infty} p_i \hat{l}_i.$$

This proves the second part of the theorem.

Next we prove part (III) of the theorem. Let the codeword lengths of an optimal code be denoted $l_1 \leq l_2 \leq l_3 \leq \cdots$, and assume to the contrary that the Kraft inequality is strict, i.e., $\Sigma_i D^{-l_i} < 1$. Let $\delta = 1 - \Sigma_i D^{-l_i} > 0$. Then there exists a positive integer k such that $D^{-l_i} < \delta$ for all $i \geq k$. Let j be an integer such that $l_j > l_k$. Define a collection of integers $\hat{l}_1, \hat{l}_2, \cdots$ such that $\hat{l}_i = l_i$ for all $i \neq j$ and such that $\hat{l}_j = l_k$. Then

$$\sum_{i=1}^{\infty} D^{-\hat{l}_i} = \sum_{i=1}^{\infty} D^{-l_i} - D^{-l_j} + D^{-l_k} < \sum_{i=1}^{\infty} D^{-l_i} + \delta = 1.$$

Thus the integers $\hat{l}_1, \hat{l}_2, \cdots$, satisfy Kraft's inequality, so that there exists a prefix code having them as codeword lengths. Since $\hat{l}_j < l_j$, such a prefix code will have a strictly smaller average codeword length for X than the optimal code whose codeword lengths are $l_1, l_2, \cdots$. This is a contradiction. □

References

[1] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, pp. 1098–1101, Oct. 1952.

[3] J. Abrahams, "Huffman-type codes for infinite source distributions," *J. Franklin Inst.*, vol. 331B, no. 3, pp. 265–271, 1994.

[4] R. A. Gallager and D. C. Van Voorhis, "Optimal source coding for geometrically distributed integer alphabets," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 228–230, Mar. 1975.

[5] P. A. Humblet, "Optimal source coding for a class of integer alphabets," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 110–112, Jan. 1978.

[6] B. Montgomery and J. Abrahams, "On the redundancy of optimal binary prefix-condition codes for finite and infinite sources," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 156–160, Jan. 1987.

[7] A. Kato, T. S. Han, and H. Nagaoka, "Huffman coding with infinite alphabet," *IEEE Trans. Inform. Theory*, vol. 42, pp. 977–984, May 1996.

[8] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

A New Bound for the Data Expansion of Huffman Codes

Roberto De Prisco and Alfredo De Santis

***Abstract*— In this correspondence, we prove that the maximum data expansion δ of Huffman codes is upper-bounded by $\delta < 1.39$. This bound improves on the previous best known upper bound $\delta < 2$. We also provide some characterizations of the maximum data expansion of optimal codes.**

***Index Terms*—Data expansion of optimal codes, Huffman codes, redundancy of optimal codes, source coding.**

I. Introduction

Huffman encoding is one of the most widely used compression techniques. Let F be a data file of size $|F|$ over an N-ary source alphabet $(a_1, a_2, \cdots, a_N)$. We assume that the original uncompressed file F is encoded using $\lceil \log N \rceil$ bits per source letter. Huffman's

Manuscript received July 19, 1996; revised February 17, 1997.

R. De Prisco is with MIT Laboratory for Computer Science, Cambridge, MA 02139 USA.

A. De Santis is with the Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (SA), Italy.

Publisher Item Identifier S 0018-9448(97)06704-7.