When we generalize to the lattice quantizer case we have to replace $G_1 = 1/12$ by G_K and get (33).

Consider now distortion measure of the form $|x|^r$. In this case $E\{|X|^r\} \le \epsilon$ and $E\{|N|^r\} = \epsilon = \Delta^r/(r+1)2^r$. The noise entropy can be written in this case as

$$H(N) = \log \Delta = \frac{1}{r} \log \left[(r+1) 2^r \cdot \epsilon \right]$$

or $2^{2H(N)} = \Delta^2 = 4[(r+1)\epsilon]^{2/r}$.

The lower bound is easily found by substituting the source X^* that maximizes the entropy under the rth moment constraint. The entropy of such a source is given by

$$H^r(S^*) = \frac{1}{r} \log \left[e \cdot 2^r \cdot \Gamma^r\left(1 + \frac{1}{r}\right) \cdot r\epsilon \right]. \tag{53}$$

Thus, the channel capacity is lower bounded by

$$C \ge \frac{1}{2} \log \left[1 + \Gamma^2\left(1 + \frac{1}{r}\right) \cdot \left(\frac{e \cdot r}{r+1} \right)^{2/r} \right] = C_l, \tag{54}$$

which takes the values $0.755, 0.638, 0.595, \cdots, 0.5$ for $r = 1, 2, 3, \cdots, \infty$.

The upper bounding technique is slightly more complicated and the bound we get may be loose since we cannot easily get moments constraints on the output random variable Y. We can only bound the rth moment using Holder's inequality,

$$\begin{aligned} E\{|Y|^r\} &\le E\{(|S| + |N|)^r\} \\ &\le \epsilon \cdot \sum_{k=0}^{r} \binom{r}{k} \frac{(r+1)^{k/r}}{k+1} \le \epsilon \cdot 2^r. \end{aligned} \tag{55}$$

The maximum entropy of the output under this rth moment constraint is given by (53) where we substitute $2^r\epsilon$ for ϵ. Thus we get the upper bound

$$C \le \frac{1}{r} \log \left[e \cdot 2^r \cdot \Gamma^r\left(1 + \frac{1}{r}\right) \cdot \frac{r}{r+1} \right] = C_u, \tag{56}$$

where C_u takes the values $1.44, 1.254, 1.180, \cdots, 1$ for $r = 1, 2, 3, \cdots, \infty$. We see immediately that this upper bound is loose at least for $r = 2$. It can be tightened for even r, i.e., $r = 2p$. In this case, we can use the fact that the odd moments of N are zero and get

$$E\{Y^{2p}\} \le \epsilon \cdot \sum_{k=0}^{p} \binom{2p}{2k} \frac{(2p+1)^{k/p}}{2k+1} \le \epsilon \cdot 2^{2p-1}, \tag{57}$$

which improves (56) by $1/r$ and so we get

$$C \le \frac{1}{r} \log \left[e \cdot 2^{r-1} \cdot \Gamma^r\left(1 + \frac{1}{r}\right) \cdot \frac{r}{r+1} \right] = C_u, \qquad r = 2p. \tag{58}$$

Note that C_u now takes the values $0.754, 0.888, \cdots, 1$ for $r = 2, 4, \cdots, \infty$.

Similar results have been obtained in [2] for the bounds there.

References

[1] J. Ziv, "On universal quantization," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 344–347, May 1985.

[2] M. Gutman, "On universal quantization with various distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 169–171, Jan. 1987.

[3] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy constrained vector quantization," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-37, pp. 31–42, Jan. 1989.

[4] J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytops, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, Mar. 1982.

[5] H. Gish and N. J. Pierce, "Asymptotically efficient quantization," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.

[6] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[7] ——, "On the structure of vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 157–166, Mar. 1982.

[8] R. Zamir, "Randomized uniform lattice quantization for universal compression with distortion," Master's thesis, Tel-Aviv Univ., Tel Aviv, Israel, 1990

[9] T. Berger, "Minimum entropy quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 149–157, March 1982.

[10] K. Sayood, J. D. Gibson, and M. C. Rost, "An algorithm for uniform vector quantizer design," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 805–814, Nov. 1984.

A Note on the Competitive Optimality of the Huffman Code

Meir Feder, *Member, IEEE*

Abstract—**It is known that a bound on the probability that the length of any source code will be shorter than the self information by γ bits can be obtained using a Chebychev-type argument. From this bound, one can establish the competitive optimality of the self information and of the Shannon-Fano code (up to one bit). In general, however, the Huffman code cannot be examined using this technique. Nevertheless, in this correspondence the competitive optimality (up to one bit) of the Huffman code for general sources is also established using a different technique.**

Index Terms—**Competitive optimality, Huffman code, self information, Chebychev inequality.**

I. Introduction

Given the probability of a source one can design a uniquely decodable (UD) source code that minimizes the expected codelength. This expected code length must be, of course, greater than the entropy of the source. The optimal code in this sense would assign to each outcome x a codeword of length $-\log p(x)$, the self information, and its expected length would exactly be the entropy. (Throughout the correspondence $\log x = \log_2 x$.) However, the self information may not be an integer. Incorporating the Diophantine constraints, it is well known that the Huffman code minimizes the expected codelength.

Manuscript received August 7, 1990; revised May 24, 1991. This work was supported by The Andrew W. Mellon Foundation, Woods Hole Oceanographic Institution.

The author is with the Department of Electrical Engineering—Systems, Faculty of Engineering, Tel-Aviv University, Tel-Aviv, 69978, Israel.

IEEE Log Number 9103963.

In some applications, one may be interested in the *competitive optimality* of the code (a term originally expressed and investigated in [1] and [2]), i.e., in comparing two source codes by the probability that the codelength of one scheme will be shorter than the codelength of the other scheme. In [1] and [2], it was shown that the self information is also optimal in this respect for dyadic sources in which the values of the self information are integers for all source's outcome. For nondyadic sources, the Shannon–Fano code, whose length for each outcome is bounded by the self information plus unity, was shown to be competitively optimal within one bit. The Huffman codelength, however, cannot be bounded in terms of the self information for every outcome and so no claim about its competitive optimality could have been made using the techniques of [2]. In this correspondence, it is shown, by a different technique, that the Huffman code is also competitively optimal within one bit.

The properties of the self-information and Shannon–Fano code are rederived in the next section. The Huffman code and its competitive optimality are discussed in Section III.

II. The Competitive Optimality of the Self Information

Let $x = x_1 \cdots x_n$ be a sample outcome of length n of a general discrete source. Let Ω be the sample space ($\Omega = \{0, 1\}^n$ for binary source) and let the probability $p(x)$ be given for all $x \in \Omega$. Denote the self information $L_n^*(x) = -\log p(x)$ and denote by $L_n(x)$ the codelength associated by any UD code to the outcome x. The following theorem, which appeared, for example in [3], is similar to the result in [1] Section VI and provides an upper bound for the probability that the codelength $L_n(x)$ of an arbitrary UD code is shorter by γ bits than the self-information. It is derived via an argument used in Chebychev-type inequalities.

Theorem 1 (Baron, Cover, and others):

$$\Pr\{L_n(x) \le L_n^*(x) - \gamma\} \le 2^{-\gamma}. \tag{1}$$

Proof: The lengths of any code that satisfy Kraft's inequality can be presented as $L_n(x) = -\log q(x)$, where $\sum_{x \in \Omega} q(x) = \sum_{x \in \Omega} 2^{-L_n(x)} \le 1$. Thus,

$$\begin{aligned}\Pr\{L_n(x) \le L_n^*(x) - \gamma\} &= \Pr\{\log q(x) \ge \log(2^\gamma p(x))\}\\ &= \Pr\{q(x) > 2^\gamma p(x)\}.\end{aligned}$$

Denote the event $A_\gamma = \{x \mid q(x) \ge 2^\gamma p(x)\}$. For $x \in A_\gamma$, $p(x) \le 2^{-\gamma} q(x)$; thus

$$\begin{aligned}\Pr\{q(x) \ge 2^\gamma p(x)\} &= \sum_{A_\gamma} p(x)\\ &\le \sum_{A_\gamma} 2^{-\gamma} q(x) \le 2^{-\gamma} \sum_{\Omega} q(x) \le 2^{-\gamma}.\end{aligned}$$

□

Theorem 1 establishes another sense in which the self information is the optimal codelength. To emphasize this optimality even further, consider the following corollary.

Corollary: For any fixed $\delta > 0$, the codelength, per symbol, of any UD code satisfies

$$\Pr\left\{\frac{1}{n} L_n(x) \le -\frac{1}{n}\log p(x) - \delta\right\} \le 2^{-n\delta}. \tag{2}$$

The corollary is proved by choosing $\gamma = n\delta$ in Theorem 1. Note that for any stationary and ergodic process, with probability 1, $\lim_{n\to\infty} -n^{-1}\log p(x) = H$, the entropy of the source. We note the exponential decay of the probability in (2) for any $\delta > 0$, and point out that it is proved for every n and every source using a simple Chebychev-type argument. Note that similar optimality results in the context of gambling and investment are also known, see e.g., [1, Section 17], [4], [5], and more.

As previously noted, a result similar to Theorem 1 was provided in [1], Section VI. There, the codelength of any code was compared to the codelength of the Shannon–Fano code, which is $\lceil -\log p(x) \rceil$. Indeed, when we consider any "approximately optimal" code whose codelengths, $L_n^a(x)$, are integers satisfying

$$L_n^a(x) < L_n^*(x) + 1 = -\log p(x) + 1, \qquad \text{for all } x, \tag{3}$$

e.g., the Shannon–Fano code, we get, using (3) and (1),

$$\begin{aligned}&\Pr\{L_n^a(x) \ge L_n(x) + \gamma + 1\}\\ &\qquad \le \Pr\{L_n^*(x) \ge L_n(x) + \gamma\} \le 2^{-\gamma},\end{aligned}$$

where $L_n(x)$ is the length associated with an arbitrary code. Now, since both $L_n^a(x)$ and $L_n(x)$ are integers, we get for any integer γ and for all codes satisfying (3),

$$\begin{aligned}&\Pr\{L_n^a(x) > L_n(x) + \gamma\}\\ &\qquad = \Pr\{L_n^a(x) \ge L_n(x) + \gamma + 1\} \le 2^{-\gamma},\end{aligned} \tag{4}$$

which is the result in [1]. Note that for $\gamma = 1$,

$$\Pr\{L_n^a(x) > L_n(x) + 1\} \le \tfrac{1}{2} \le \Pr\{L_n^a(x) \le L_n(x) + 1\}, \tag{5}$$

that is, the Shannon–Fano code most of the time provides a code that is shorter (up to one bit per block, or $1/n$ bits per symbol) than any other competing code.

Another result presented in [2] and [1] considers the case where the source is dyadic, i.e., $L_n^*(x) = -\log p(x)$ is an integer for all x. In this case, the Huffman codelength, as well as the Shannon–Fano codelength, is the (integer) self-information $L_n^*(x)$. Using Theorem 1 and incorporating the fact that all lengths are integers, we get

$$\begin{aligned}&\Pr\{L_n^*(x) \ge L_n(x) + \gamma\}\\ &\qquad = \Pr\{L_n^*(x) > L_n(x) + \gamma - 1\} \le 2^{-\gamma},\end{aligned} \tag{6}$$

for any integer γ. For $\gamma = 1$,

$$\Pr\{L_n^*(x) > L(x)\} \le \tfrac{1}{2} \le \Pr\{L_n^*(x) \le L(x)\}. \tag{7}$$

For $\gamma = 1$ a stronger result may be stated. Denote the probability of the event $\{L_n^*(x) = L_n(x)\}$ by Q. For dyadic sources, we get

$$\begin{aligned}\Pr\{L_n^*(x) > L(x)\} &= \Pr\{L_n^*(x) \ge L(x) - 1\}\\ &= \Pr\left\{p(x) \le \frac{q(x)}{2}\right\},\end{aligned}$$

where $L(x) = -\log q(x)$. Using these arguments,

$$\Pr\{L_n^*(x) > L(x)\} \le \frac{1}{2} \sum_{\{x \mid p(x) \le q(x)/2\}} q(x) \le \frac{1-Q}{2}, \tag{8}$$

where the last inequality follows since

$$\sum_{\{x \mid p(x) \le q(x)/2\}} q(x) \le 1 - \sum_{\{x \mid p(x) = q(x)\}} q(x) = 1 - \sum_{\{x \mid p(x) = q(x)\}} p(x) = 1 - Q.$$

On the other hand,

$$\Pr\{L_n^*(x) < L(x)\} = 1 - \Pr\{L_n^*(x) > L(x)\} - \Pr\{L_n^*(x) = L(x)\} \ge \frac{1-Q}{2}. \tag{9}$$

Thus, combining (8) and (9),

$$\Pr\{L_n^*(x) > L(x)\} \le \Pr\{L_n^*(x) < L(x)\}. \tag{10}$$

This result is stronger than (7) since the event in the right-hand side does not include the case $L_n^*(x) = L_n(x)$. It also appeared originally, with a different proof, in [2].

Unfortunately, we cannot prove (4) and (5) for the Huffman code in the nondyadic case following Theorem 1. Nevertheless, we show in the next section through a different technique that the Huffman code is also competitively optimal within one bit.

III. The Competitive Optimality of the Huffman Code

The Huffman code minimizes the expected codelength for a given source under the constraint that the lengths are integers. It is well known that $E\{L^H(x)\} < E\{-\log p(x)\} + 1 = H + 1$. This does not imply, however, the bound (3) on the length of each outcome. An example of a source whose Huffman codelengths do not satisfy (3) (due to [6]) is the source with an infinite (countable) number of outcomes, $x_0, x_1, \cdots$ whose probabilities are given by,

$$\begin{aligned} P(x_0) &= 1 - \alpha \\ &\vdots \\ P(x_1) &= (1-\alpha)\alpha \\ &\vdots \\ P(x_i) &= (1-\alpha)\alpha^i. \end{aligned}$$

For $\frac{1}{2} \le \alpha < \frac{\sqrt{5}-1}{2} = 0.618$ (the golden number) the Huffman codebook, for the previous source, is $\{1, 01, 001, 0001, \cdots\}$. Clearly, in this codebook the codelength assigned to the symbol x_i, denoted $L^H(x_i)$, is $i + 1$. On the other hand $-\log p(x_i) = -\log(1-\alpha) + i \cdot -\log\alpha$. Since, for any $\frac{1}{2} < \alpha < 0.618$, $-\log\alpha < 1$, we can find j such that for any $i \ge j$, $L^H(x_i) > -\log p(x_i) + 1$. Additional discussion on the Huffman code for a geometric distribution on an infinite alphabet can be found in [7].

As a side remark we note that an upper bound on the lengths of the Huffman code, which is weaker than (3), can be derived. This bound states that

$$L^H(x) \le n, \tag{11}$$

where n is such that $\frac{1}{F_n} > p(x) \ge \frac{1}{F_{n+1}}$ and $F_n = 1, 2, 3, 5, 8, \cdots$ for $n = 1, 2, \cdots$ is the Fibbonaci sequence. This bound follows from the sibling property of the Huffman code, [8]. This bound appears e.g., in [9]–[11], and for completeness it is also derived in the Appendix.

Now, despite the fact that for some input symbols the Huffman codelength can be much longer than the self information, it is competitively optimal within one bit, as shown in the following theorem.

Theorem 2: Let $L(x)$ be the length of the code associated with the symbol x by an arbitrary UD code. Let $L^H(x)$ be the Huffman codelength of the symbol x. Then,

$$\Pr\{L(x) < L^H(x) - 1\} < \Pr\{L(x) > L^H(x) - 1\}. \tag{12}$$

Proof: Let A be the set of source symbols $\{x \mid L(x) < L^H(x) - 1\}$, B the set $\{x \mid L(x) = L^H(x) - 1\}$, and let C be the set $\{x \mid L(x) > L^H(x) - 1\}$. Note that the competing code is shorter than the Huffman code, i.e., $L(x) < L^H(x)$ for symbols in A and B, while $L(x) \ge L^H(x)$ for symbols in C.

Reconstruct a new code, whose length function is denoted by $L'(x)$, as follows. Use the competing code whenever it is shorter, i.e., in A and B, use the Huffman code otherwise, i.e., in C, and add a one bit prefix to indicate which code is used. Clearly this code is uniquely decodable. For symbols $x \in A$ the length of this code satisfies $L'(x) < L^H(x) - 1 + 1 = L^H(x)$, or since the lengths are integers $L'(x) \le L^H(x) - 1$. For symbols $x \in B$ we have $L'(x) = L^H(x) - 1 + 1 = L^H(x)$. For symbols $x \in C$ the code is identical to the Huffman code plus a one bit prefix and thus $L'(x) = L^H(x) + 1$. Thus, the expected length of this code satisfies

$$\begin{aligned} E\{L'\} &\le E\{L^H\} - \sum_{x \in A} p(x) + \sum_{x \in C} p(x) \\ &= E\{L^H\} + \Pr(C) - \Pr(A). \end{aligned}$$

By reconstruction, the binary tree representing the codebook associated with $L'(\cdot)$ is incomplete. Thus, there is a length function $L''(\cdot)$ of a uniquely decodable code such that $E\{L''\} < E\{L'\}$, and so $E\{L''\} < E\{L^H\} + \Pr(C) - \Pr(A)$, or,

$$\Pr(C) > \Pr(A) + E\{L''\} - E\{L^H\}.$$

Since the Huffman code minimizes the expected code length, $E\{L''\} - E\{L^H\} \ge 0$, and so $\Pr(C) > \Pr(A)$, i.e.,

$$\Pr\{L(x) > L^H(x) - 1\} > \Pr\{L(x) < L^H(x) - 1\}. \quad \square$$

Note that as compared to what can be stated for the Shannon–Fano code, this result (12) on Huffman code, proved in Theorem 2, is stronger than (5) but less general than (4).

Acknowledgment

The author acknowledges Prof. T. M. Cover for bringing [1] and [2] to his attention and for pointing out an error in the early version of the correspondence.

Appendix
Proof of (11)

Let x be a source symbol whose probability is $p(x)$ and its Huffman code length is $L^H(x)$, i.e., it is represented by a leaf at level $L^H(x)$ at the Huffman code tree. Thus, at level $L^H(x) - 1$ of the tree the node that this leaf descends from has a probability greater than $p(x)$, and by the sibling property its sibling has a probability of at least $p(x)$. Thus, the probability of the node where the two siblings descend from, at level $L^H(x) - 2$, is greater than $2p(x)$, while its sibling has a probability of at least $p(x)$. At the previous level ($L^H(x) - 3$) the probability of the node where this subtree descends from is greater than $3p(x)$ while its sibling has a probability of at least $2p(x)$. We immediately see that the probability of the node at the $(L^H(x) - n)$th level, where this source leaf descends from, is greater than $F_n \cdot p(x)$, where $F_n = F_{n-1} + F_{n-2}$ is the Fibbonaci sequence. The root, which will be at level $0 = L^H(x) - L^H(x)$ must have a probability greater than $F_{L^H} \cdot p(x)$. Since the root probability is 1, we must have $1 > F_{L^H} \cdot p(x)$ which, together with the requirement that the codelength is an integer, leads to (11). □

References

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

[2] T. M. Cover, "On the competitive optimality of Huffman code," *IEEE Trans. Inform. Theory*, vol. 37, pp. 172–174, Jan. 1991.

[3] A. R. Barron, "Logically Smooth Density Estimation," Ph.D. thesis, Stanford Univ., Stanford, CA, 1985.

[4] P. H. Algoet and T. M. Cover, "Asymptotic optimality and asymptotic equipartition properties of log-optimal investment," *Annals Probab.*, vol. 16, pp. 876–898, 1988.

[5] R. Bell and T. M. Cover, "Competitive optimality of logarithmic investment," *Math. Oper. Res.*, vol. 5, no. 2, pp. 161–166, 1980.

[6] T. M. Cover, private communication.

[7] R. G. Gallager and D. C. Van Voorhis, "Optimal source codes for generally distributed integer alphabets," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 228–230, 1975.

[8] R. G. Gallager, "Variation on a theme by Huffman," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 668–674, Nov. 1978.

[9] G. O. H. Katona and T. O. H. Nemetz, "Huffman codes and self information," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 337–340, 1976.

[10] K. C. Chu and J. Gill, "Upper bounds on Huffman codewords lengths," in *Proc. IEEE Int. Symp. Inform. Theory*, Budapest, Hungary, June 24–28, 1991.

[11] K. C. Chu, "Lossless data compression and its VLSI implementation," Ph.D. thesis, Stanford Univ., Stanford, CA, 1990.

On the Redundancy of Optimal Codes with Limited Word Length

Renato M. Capocelli, *Senior Member, IEEE,* and
Alfredo De Santis, *Member, IEEE*

***Abstract*—Some limitations are given on the redundancy of *D*-ary codes with maximal codeword length *L*. First we give an upper bound that improves on a previous result of Gilbert. Then, it is shown that the redundancy of these constrained codes is very close to that of the unconstrained Huffman codes when the number of codewords *N* is such that ND^{1-L} becomes negligible. Further, a tight bound is given on the redundancy when only the most likely probabilities are known. Finally, in the binary case, a tight lower bound is given on the redundancy when only the least likely probability is known.**

***Index Terms*—Limited codeword length, redundancy, bounds on the redundancy, optimal codes.**

Manuscript received January 3, 1989; revised July 11, 1991. This work was supported in part by the Italian Ministry of Education, Project Progetto ed Analisi di Algoritmi, and in part by the Italian National Council for Research under Contract No. 90.01552.12. This work was presented in part at the 23rd Annual Conference on Information Sciences and Systems, Baltimore, MD, March 1989. This work was done while A. De Santis was at IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598.

R. M. Capocelli is with the School of Computer Science, Dipartimento di Matematica, Università di Roma "La Sapienza," Via Salaria 113, 00198 Roma, Italy.

A. De Santis is with the Dipartimento di Informatica ed Applicazioni, Universitá di Salerno, 84081 Baronissi, Salerno, Italy.

IEEE Log Number 9103968.

I. Introduction

In coding theory, it is usually assumed that all the statistics of a message source are known with perfect accuracy. This assumption seems reasonable as we can measure all statistical parameters accurately by examining sufficiently long sample messages. If we want, thus, to design an efficient code for a message source S whose letter probabilities are not known, we can use a two-stage procedure. First we determine the relative frequencies $f_1, f_2, \cdots, f_N$ of the letters of S, using a fairly long sample message. Then, we design an optimal code using such frequencies as the actual probabilities $p_1, p_2, \cdots, p_N$.

There are situations in which the measurements required to design increasingly efficient codes for a given source become increasingly difficult. If the probabilities needed to design an elaborate code are not estimated precisely enough, the elaborate code can be less efficient than simpler codes. The knowledge of upper bounds on the redundancy of optimal codes in terms of most likely probabilities may help to avoid the design of elaborate codes that are only a little more efficient than simpler codes. This is particularly true if only the probability of the most likely source letter is precisely known [1], [4], [5], and [6].

The underestimation of the probabilities can lead to very long codewords which is non desirable (limited buffer, hardware configuration [12], etc.). Thus it may be desirable to design a code with limited maximum length.

Therefore, instead of constructing an efficient code using the estimated probabilities $p_1, p_2, \cdots, p_N$, a more conservative approach is recommended, namely, construct an optimal code, based on the estimated probabilities, with the constraint that each codeword length is less than or equal to a given maximal length L [8].

Given a source S consisting of N letters with probabilities $p_1, p_2, \cdots, p_N$ and given an integer $L > 0$, an *optimal D-ary code with limited (word) length L* for S is defined as a D-ary code whose N codeword lengths satisfy $n_i \le L$ and minimize the redundancy

$$r = \sum_{i=1}^{N} p_i n_i - H_D(p_1, p_2, \cdots, p_N), \tag{1}$$

where $H_D(p_1, p_2, \cdots, p_N) = -\sum_{i=1}^{N} p_i \log_D p_i$ is the entropy of S. Since the entropy is independent of the code, minimizing (1) is equivalent to minimizing the average codeword length $\sum_{i=1}^{N} p_i n_i$.

Algorithms for constructing optimal codes with limited word length have been investigated in [9]–[17].

Notice that a D-ary code with N codewords and maximum length L exists iff $N \le D^L$. Also notice that the redundancy of an optimal code with maximum length L may be larger than the redundancy of an optimal (Huffman) code with no constraint. The redundancy will be the same if L is less than or equal to the

0018-9448/92$03.00