IV. NONDYADIC p

The Shannon code $\lceil \log 1/p(x) \rceil$ is expected length optimal (a Huffman code) if p is dyadic and is within 1 of expected length optimal for arbitrary p. Similarly, $\lceil \log 1/p(x) \rceil$ is competitively optimal if p is dyadic. We now ask about the competitive performance of $\lceil \log 1/p \rceil$ in general.

Let $l(x) = \lceil \log 1/p(x) \rceil$. We now show that l competitively dominates $l' + 1$ for all uniquely decodable codes l'.

Theorem 3: If $l(x) = \lceil \log(1/p(x)) \rceil$, then

$$E \operatorname{sgn}(l(X) - (l'(X) + 1)) \le 0,$$

for all uniquely decodable assignments $l'(x)$.

Proof: From $l = \lceil \log(1/p) \rceil$ we have $2^{-l} \le p < 2 \cdot 2^{-l}$. Thus

$$\begin{aligned} E \operatorname{sgn}(l(X) - (l'(X) + 1)) &\le \sum p(x)(2^{l(x) - l'(x) - 1} - 1) \\ &= \frac{1}{2} \sum p(x) 2^{l(x) - l'(x)} - 1 \\ &< \frac{2}{2} \sum 2^{-l}(2^{l - l'}) - 1 \\ &= \sum 2^{-l'} - 1 \le 0. \end{aligned}$$

□

V. REPEATED PLAYS

The previous results easily extend to sequences of random variables. Suppose $p(x)$ is dyadic and we wish to encode blocks $(X_1, X_2, \cdots, X_n)$, where $X_1, X_2, \cdots, X_n$ are independent identically distributed according to $p(x)$. Consider the myopic encoding $l(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} l(X_i)$, where $l(x_i) = \log 1/p(x_i)$, obtained by concatenating the codewords associated with the individual symbols.

We observe that $p(x_1, \cdots, x_n)$ is also dyadic, and $l(x_1, \cdots, x_n) = \log(1/p(x_1, \cdots, x_n))$. Consequently,

$$E \operatorname{sgn}(l'(X_1, X_2, \cdots, X_n) - l(X_1, X_2, \cdots, X_n)) > 0$$

for all $l' \ne l$, for all n. Thus the short term goal of designing the competitively shortest code at time $n = 1$ is completely compatible with designing the shortest code for any time. Simply concatenate the codewords.

VI. SUMMARY

Let $l(x) = \lceil \log(1/p(x)) \rceil$. Then for any other uniquely decodable assignment $l'(x)$ we have shown that l competitively dominates $l' + 1$ and also dominates $l' + 1$ in expected value. If p is dyadic, l competitively dominates l' and also dominates l' in expected value. These results indicate that the Shannon codeword length assignment $l(x) = \lceil \log(1/p(x)) \rceil$ has optimal short run as well as optimal long run properties.

ACKNOWLEDGMENT

The author would like to thank Laura Ekroot for helpful discussions.

REFERENCES

[1] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1949.
[2] R. Blahut, *Principles and Practice of Information Theory*. New York: Addison Wesley, 1987.

A Note on D-ary Huffman Codes

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

Abstract —An upper bound on the redundancy of D-ary Huffman codes in terms of the probability p_1 of the most likely source letter is provided. For large values of p_1, the bound improves the one given by Gallager. Additionally, some results known for the binary case ($D = 2$) are extended to arbitrary D-ary Huffman codes. As a consequence, a tight lower bound that corrects a bound recently proposed by Golic and Obradovic is derived.

Index Terms —Huffman coding, entropy, codeword, source coding.

I. INTRODUCTION

Let A be a discrete source with N letters, $2 \le N < \infty$, and p_k denote the probability of letter a_k, $1 \le k \le N$. Let D, $2 \le D < \infty$, denote the size of the code alphabet. Let $\{x_1, x_2, \cdots, x_N\}$ be a set of D-ary codewords and $n_1, n_2, \cdots, n_N$ be the codeword lengths. The *Huffman* encoding algorithm provides an *optimal prefix code* C for the source A. The encoding is optimal in the sense that codeword lengths minimize the *redundancy* r, defined as the difference between the *average codeword length* E of the code and the *entropy* $H(p_1, p_2, \cdots, p_N)$ of the source:

$$r = E - H(p_1, p_2, \cdots, p_N) = \sum_{i=1}^{N} p_i n_i + \sum_{i=1}^{N} p_i \log_D p_i.$$

According to Shannon's first theorem, the redundancy of any Huffman code is always nonnegative and less than or equal to one.

In a remarkable paper [1], Gallager has proved that, knowing the probability of the most likely source letter p_1, the following upper bound holds:

$$r \le \sigma_D + p_1 D / \ln D, \tag{1}$$

where $\sigma_D = \log_D(D-1) + \log_D(\log_D e) - \log_D e + (D-1)^{-1}$. For binary codes ($D = 2$) bounds better than (1) are known [1], [3], [4], [5], and [6]. Bound (1) improves the Shannon limit, $r \le 1$, only when $p_1 < \gamma_D = (1 - \sigma_D)(\ln D)/D$. Moreover γ_D approaches 0 as D gets large. Indicatively, one has that $\gamma_3 = 0.316$, $\gamma_5 = 0.259$, $\gamma_{10} = 0.168$ and $\gamma_{20} = 0.099$. Finding upper bounds tighter than the Shannon limit for $\gamma_D < p_1 < 1$ is therefore an open problem.

A necessary and sufficient condition for the most likely letter of a discrete source to be coded by a single symbol with a binary Huffman code was first obtained by Johnsen [3]. Capocelli *et al.* [4] extended this result to the case of a two symbol codeword.

Manuscript received April 4, 1988; revised August 7, 1989. This work was supported in part by the Italian Ministry of Education and by the Italian National Council for Research. This work was presented in part at IEEE 1988 International Symposium on Information Theory, Kobe, Japan, June 1988.

R. M. Capocelli is with the Dipartimento di Matematica, Università di Roma, 00185 Roma, Italy.

A. Desantis was at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, while on leave from the Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (Salerno), Italy.

IEEE Log Number 9039284.

Finally, Montgomery and Kumar [8] provided a solution for arbitrary codeword lengths.

Katona and Nemetz [7] considered the connections between the probability p_a of a generic source letter a and its codeword length n_a in a binary Huffman code. They also derived an upper bound on n_a in terms of p_a.

A lower bound on the redundancy of binary Huffman codes in terms of p_1 was first obtained by Johnsen [3], for the case $p_1 \geq 0.4$. Capocelli and De Santis [6] and Montgomery and Abrahams [9] extended Johnsen's result to arbitrary $0 < p_1 \leq 1$.

Golic and Obradovic [10] have recently extended Johnsen's results to the case where the code alphabet is of arbitrary size D. However, the lower bound they derived, is not correct.

The main result of this correspondence is an upper bound that improves (1) for large values of p_1 and any D. Our upper bound holds for $\mu_D < p_1 < 1$, where μ_D is proven to be less than γ_D. Since it is always less than 1, whenever $p_1 < 1$, it provides an improvement on the Shannon limit $r \leq 1$. Therefore, our bound used in conjunction with Gallager's bound provides an upper bound that is always better than the Shannon limit.

In the paper we also obtain a tight lower bound on r for D-ary codes in terms of p_1. This bound extends the bound previously known for the binary case [6], [9] and corrects the one given in [10]. Furthermore, a general necessary and sufficient condition for the most likely symbol to be coded with a codeword of length l in a D-ary Huffman code is established, extending the results of [8] and [10]. Finally, generalizing the results of [7] to arbitrary D-ary Huffman code, an upper bound on the codeword length n_a associated to the source letter a in terms of p_a is provided.

It is also shown that as D gets large the redundancy of a D-ary Huffman code can be bounded by $p_1 - \epsilon \leq r < 1$, where ϵ is an arbitrary positive constant.

II. Preliminaries

As is well known, a D-ary prefix code can be represented as a conveniently labeled rooted D-ary code tree in which each source letter corresponds to a leaf on the tree and where the associated codeword is the sequence of labels on the path from root to leaf. The *level* of a node is the number of links from the root to it. A node a is called the *parent* of node b, which is the *child* of node a, if there is a link between them and the level of a increased by 1 is equal to the level of b. Two nodes with a common parent are called *siblings*.

Each leaf of the code tree has a probability assigned to it, namely, the probability of the corresponding source letter. We also assign a probability to each intermediate node, defined as the sum of the probabilities of all leaves for which the path from root to leaf passes through the given node.

We allow $D-1$ of the source messages to have zero probability. Thus we can assume without loss of generality that $N = c(D-1)+1$ for some integer c.

Property 1: In any D-ary Huffman code tree, for each $l \geq 1$, the probability of each node at level l is less than or equal to the probability of each node at level $l-1$.

Property 2: Let a and b be two nodes on the same level in a D-ary Huffman code tree and p_a and p_b their probabilities. If b is not a leaf then $p_b \leq Dp_a$.

Definition: [1] A D-ary code tree has the *sibling property* if the nodes (except the root) can be listed in order of nonincreasing probability such that for each i, $1 \leq i \leq c$, nodes $iD, iD-1, \cdots, iD-D+1$ are all siblings of each other.

Gallager [1] has proved the following

Theorem 1: A D-ary prefix code is a Huffman code iff the code tree has the sibling property.

The following Lemma is an extension of a lemma by Montgomery and Kumar [8] to arbitrary D-ary codes.

Lemma 1: If a D-ary Huffman code for the source A has a minimum codeword length of l, $l \geq 1$, then the probability p_1 of the most likely letter of A satisfies:

$$\frac{1}{D^2 - D + 1} \leq p_1 \leq 1, \qquad \text{if } l = 1, \tag{2}$$

$$\frac{1}{D^{l+1} - D + 1} \leq p_1 \leq \frac{D}{D^l + D - 1}, \qquad \text{if } l > 1. \tag{3}$$

Proof: Let $C = \{x_1, x_2, \cdots, x_N\}$ be a D-ary Huffman code for A and let x_1 be the codeword with minimum length l encoding the source symbol with probability p_1. Let $q_1 \geq q_2 \geq \cdots \geq q_L$, $L = D^l$, be the probabilities of nodes at level l in the Huffman code tree. Let g be the node with smallest probability, namely, $q_g = q_{L-D+1} + \cdots + q_L$, at level $l-1$. Then all the codewords that correspond to the leaves of the subtree rooted at g have the common prefix σ of length $l-1$. We may assume, without loss of generality, that g is not the father of the node corresponding to x_1.

From Property 2 we get $q_i \leq Dp_1$, $1 \leq i \leq L$, thus obtaining $(1-p_1)/(D^l - 1) \leq \max q_i \leq Dp_1$, where the maximization is over all $i \in [1, L]$, $q_i \neq p_1$. This proves (2) and the left inequality of (3).

Now assume $l > 1$. Consider the encoding of A by means of the code C' with codewords $x_1' = \sigma$, $x_i' = x_1\tau$ if $x_i = \sigma\tau$ and $x_i' = x_i$ otherwise. In words C' is obtained from C by interchanging the subtree rooted at g with the node corresponding to x_1. C' is a prefix code. The difference between the expected codeword lengths of C and C' is $E(C) - E(C') = p_1 - q_g \geq p_1 - (1-p_1)D/(D^l - 1)$. From the optimality of C we get $E(C) - E(C') \leq 0$, which in turn implies the right inequality of (3). □

A different but equivalent statement of the relation between the probability p_1 of the source A and the minimum codeword length in the associated D-ary Huffman code is stated in the following lemma, which extends the nice result of [8] and is in accordance with Theorem 1 of Golic and Obradovic [10].

Lemma 2: If

$$\frac{D}{D^{l+1} + D - 1} < p_1 < \frac{1}{D^l - D + 1}, \tag{4}$$

for some $l \geq 1$, then the minimum codeword length of any D-ary Huffman code for a source A, where the most likely letter has probability p_1, is l. Furthermore, if

$$\frac{1}{D^{l+1} - D + 1} \leq p_1 \leq \frac{D}{D^{l+1} + D - 1}, \tag{5}$$

for some $l \geq 1$, then any D-ary Huffman code for A has a minimum codeword length of either l or $l+1$.

Proof: The lemma can be easily proved by distinguishing the two cases: (4) and (5), and by making use of Lemma 1. □

Consider the source B with alphabet $\{a_2, a_3, \cdots, a_N\}$ and entropy

$$H_B = H\left(\frac{p_2}{1-p_1}, \cdots, \frac{p_N}{1-p_1}\right).$$

The entropy $H = H(p_1, p_2, \cdots, p_N)$ of A is related to H_B by

$$H = \mathscr{H}_D(p_1) + (1-p_1)H_B, \tag{6}$$

where $\mathscr{H}_D(x)$ is the Shannon function $-x\log_D x - (1-x)\log_D(1-x)$. The following lemma provides a useful link between the expected codeword length L and L_B of the D-ary Huffman codes C and C_B for sources A and B, respectively. This lemma will be instrumental in deriving our claimed upper bound on r.

Lemma 3: The expected codeword lengths, L and L_B, of the D-ary Huffman codes C and C_B corresponding to source $A=\{a_1,a_2,\cdots,a_N\}$ and subset $B=\{a_2,a_3,\cdots,a_N\}$, respectively, satisfy

$$L \le 1+(1-p_1)(L_B-1+2/D). \tag{7}$$

Moreover, for every $p_1 \ge 1/(D+1)$, a source A exists that satisfies (7) with equality.

Proof: If $N \le D$, then $L=L_B=1$ and the lemma is true. Now suppose $N>D$ and let $x_{B,i}$, $i=1,\cdots,N$, be the codewords for the D-ary Huffman code of the source B. Its expected codeword length is

$$L_B=\sum_{i=2}^{N}\frac{p_i}{1-p_1}|x_{B,i}|.$$

Let $q_1 \ge q_2 \ge \cdots \ge q_D$ be the probabilities of nodes $g_1,g_2,\cdots,g_D$ at level 1 in the Huffman code tree of the source B.

Let $C[i]$, $1\le i\le D$, be the set of codewords that correspond to the leaves of the subtree rooted at g_i. All the codewords in $C[i]$ have as prefix the same letter σ_i.

Consider the code R for the source A with codewords $x_1'=\sigma_D$, $x_i'=\sigma_{D-1}x_{B,i}$ when the first letter of $x_{B,i}$ is either σ_{D-1} or σ_D and $x_i'=x_{B,i}$ otherwise. R is a prefix code and its expected codeword length is

$$\begin{aligned}L_R&=\sum_{i=1}^{N}p_i|x_i'|\\&=p_1+\sum_{x_{B,i}\notin C[D]\cup C[D-1]}p_i|x_{B,i}|\\&\quad+\sum_{x_{B,i}\in C[D]\cup C[D-1]}p_i(|x_{B,i}|+1)\\&=p_1+(1-p_1)L_B+(1-p_1)(q_{D-1}+q_D).\end{aligned} \tag{8}$$

From $\sum_{i=1}^{D}q_i=1$ and $q_1\ge q_2\ge\cdots\ge q_D$ follows $q_D+q_{D-1}\le 2/D$, and thus

$$L_R \le 1+(1-p_1)(L_B-1+2/D).$$

Since the expected codeword length of the Huffman code is minimum, one has $L \le L_R$, and (7) follows.

Bound (7) is tight for every $p_1\ge 1/(D+1)$. Indeed, the source consisting of $N=D+1$ letters having probabilities $p_1=p$ and $p_i=(1-p)/D$, $i=2,3,\cdots,D+1$, satisfies (7) with equality. □

The following lemma will be useful in deriving the lower bound.

Lemma 4: If a D-ary Huffman code has a minimum codeword length of l, $l\ge 1$, then its redundancy satisfies

$$r \ge l-\mathscr{H}_D(p_1)-(1-p_1)\log_D(D^l-1).$$

Proof: By the hypothesis, $n_1=l$. Let $q_1,q_2,\cdots,q_L$, $L=D^l$, be the probabilities of nodes $g_1,g_2,\cdots,g_L$ at level l in the Huffman code tree T. We may assume, without loss of generality, that $p_1=q_1$. Let T_i, $2\le i\le L$, be the subtree of the tree T rooted at g_i. Let P_i be the set of source letters that correspond to leaves to T_i. Subtrees $T_2,T_3,\cdots,T_L$ determine a partition $P_2,P_3,\cdots,P_L$ of the source letters $\{a_2,a_3,\cdots,a_L\}$. Let H_i denote the entropy of P_i, and E_i the average codeword length of the Huffman code for subsource P_i.

H_i and E_i are related to the entropy $H(p_1,p_2,\cdots,p_N)$ of the source and to the average codeword length E of the Huffman code by the following relations:

$$H(p_1,p_2,\cdots,p_N)=H(q_1,q_2,\cdots,q_L)+\sum_{i=2}^{L}q_iH_i$$

and

$$E=l+\sum_{i=2}^{L}q_iE_i.$$

Finally, using $E_i\ge H_i$, we get

$$\begin{aligned}r&=E-H(p_1,p_2,\cdots,p_N)\\&=l-H(q_1,q_2,\cdots,q_L)+\sum_{i=2}^{L}q_i(E_i-H_i)\\&\ge l-H(q_1,q_2,\cdots,q_L)\\&=l-\mathscr{H}_D(p_1)-(1-p_1)H\left(\frac{q_2}{1-p_1},\frac{q_3}{1-p_1},\cdots,\frac{q_L}{1-p_1}\right)\\&\ge l-\mathscr{H}_D(p_1)-(1-p_1)\log_D(D^l-1).\end{aligned}$$ □

III. The Upper Bound

Gallager [1] has proved that, knowing the probability of the most likely source letter p_1, the following upper bound holds:

$$r \le \sigma_D+p_1D/\ln D, \tag{9}$$

where

$$\sigma_D=\log_D(D-1)+\log_D(\log_D e)-\log_D e+(D-1)^{-1}. \tag{10}$$

Equation (9) is a generalization of Theorem 2 in [1]. As a numerical example for (10), $\sigma_3=0.135$, $\sigma_5=0.194$, $\sigma_{10}=0.269$, $\sigma_{20}=0.335$ have been calculated in [1]. Unfortunately, bound (9) improves the Shannon limit, $r\le 1$, only when $p_1<\gamma_D=(1-\sigma_D)(\ln D)/D$. It is thus an open problem finding sharper upper bounds on the redundancy r for larger values of p_1.

In the sequel we will provide an upper bound that improves (9) for large values of p_1, whatever is $D\ge 2$.

Recalling (6), from Lemma 3 we get

$$r \le 1-\mathscr{H}_D(p_1)+(1-p_1)(L_B-H_B-1+2/D) \tag{11}$$

and, since $L_B\le H_B+1$, we obtain

$$r\le 1-\mathscr{H}_D(p_1)+(1-p_1)2/D. \tag{12}$$

Unfortunately, as D gets large, bound (12) approaches 1. The same is true for Gallager's bound (9) since, as stressed in [1], $\sigma_i\to 1$ as D increases. In both cases the convergence is not rapid. This is not just the case of (9) and (12); it is rather inherent in the definition of the redundancy, as the following theorem shows.

Theorem 2: Let $\phi^+(D,p_1)\le 1$ be an upper bound on the redundancy of any D-ary Huffman code for the source A in terms of the probability p_1 of the most likely source letter. Then for any p_1, $1>p_1>0$, $\lim_{D\to\infty}\phi^+(D,p_1)=1$.

Proof: Let A be a source consisting of N letters. For every $D>N$, the average codeword length of the corresponding D-ary Huffman code is 1. For a fixed N, the entropy of the source A approaches 0 as D gets large, as it is computed with base D logarithms. Therefore any upper bound on r must approach 1 as D increases. □

Lemma 5: For any $D\ge 2$, the function

$$f_D(x)=1-\mathscr{H}_D(x)+(1-x)2/D$$

is a continuous and convex ∪ function of x that reaches its minimum at point $x_{D,\min}=D^{2/D}/(1+D^{2/D})$, and $0.5<x_{D,\min}<1$. Moreover

$$f_D(x)<1, \qquad \text{iff } \delta_D<x<1 \tag{13}$$

where δ_D is the unique solution of the equation $f_D(x)=1$ in the interval $]0,1[$. The sequence $\{\delta_D\}_{D\geq 2}$ satisfies the following properties, for $D\geq 2$:

1) $\delta_{D+1}<\delta_D$,
2) $\delta_D\leq(\log_2 D)/(D+\log_2 D)$.

Proof: The convexity property as well as the computation of the minimum can be easily seen by using standard analysis arguments. To prove (13) note that, since $f_D(0)=1+2/D>1$ and $f_D(1)=1$, from the convexity there exists a unique solution of the equation $f_D(x)=1$ in the interval $]0,1[$.

The equation $f_D(x)=1$ can also be written as $\mathscr{H}_2(x)=(1-x)2(\log_2 D)/D$. From $(\log_2(D+1))/(D+1)<(\log_2 D)/D$ results

$$\begin{aligned}\mathscr{H}_2(\delta_{D+1})&=(1-\delta_{D+1})2(\log_2(D+1))/(D+1)\\&<(1-\delta_{D+1})2(\log_2 D)/D,\end{aligned}$$

hence $f_D(\delta_{D+1})>1$, which, by using (13), leads to $\delta_{D+1}<\delta_D$.

A direct computation shows that $\delta_2=0.227$. For $0\leq x\leq 0.5$, we have $\mathscr{H}_2(x)\geq 2x$. Using this inequality we get

$$\mathscr{H}_2(\delta_D)=\frac{2(1-\delta_D)\log_2 D}{D}\geq 2\delta_D,$$

which leads to $\delta_D\leq(\log_2 D)/(D+\log_2 D)$. □

We have now two bounds, both improving the Shannon bound, namely, bound (9) for $0<p_1<\gamma_D$ and bound (12) for $\delta_D<p_1<1$. Their combined use gives a bound that is definitively better than the Shannon limit, for any value of $0<p_1<1$. Notice that the width of the interval in which Gallager's bound (9) holds approaches 0 as D gets large, whereas that of (12) approaches 1.

Theorem 3: Let p_1 be the probability of the most likely source letter of a discrete source A. The redundancy of the corresponding D-ary Huffman code is upper bounded by

$$r\leq\sigma_D+p_1D/\ln D,\qquad \text{if } 0<p_1\leq\theta_D,\quad(14)$$

$$r\leq 1-\mathscr{H}_D(p_1)+(1-p_1)2/D,\qquad \text{if } \theta_D<p_1<1,\quad(15)$$

where θ_D is the unique solution of the equation $\sigma_D+xD/\ln D=1-\mathscr{H}_D(x)+(1-x)2/D$ in the interval $x\in]0,1[$. Moreover the sequence $\{\theta_D\}_{D\geq 2}$ satisfies the following inequalities, for $D\geq 3$;

1) $\delta_D<\theta_D<\gamma_D$,
2) $\theta_D<(1+\ln\ln D)/D$.

Proof: First observe that

$$1+\ln\ln D>D\gamma_D>0.82+\ln\ln D.\qquad(16)$$

Indeed,

$$\begin{aligned}D\gamma_D&=(1-\sigma_D)\ln D\\&=\frac{D-2}{D-1}\ln D+1+\ln\ln D-\ln(D-1)\\&<\ln D+1+\ln\ln D-\ln(D-1)\\&<1+\ln\ln D\end{aligned}$$

proves the first inequality of (16). $(x-2)(\ln x)/(x-1)+1-\ln(x-1)$ is a convex $\cup$ function of x. A direct computation shows that it reaches the minimum at $x=4.9215$, where its value is 0.8207. Hence

$$\begin{aligned}D\gamma_D&=\frac{D-2}{D-1}\ln D+1+\ln\ln D-\ln(D-1)\\&>0.8207+\ln\ln D,\end{aligned}$$

which proves the second inequality of (16).

Now we prove that $\delta_D<\gamma_D$. To this aim notice that, because of (13) and (16), it is sufficient to prove $f_D(\beta_D/D)<1$, where $\beta_D=0.82+\ln(\ln D)$. This latter inequality can be written as

$$(D-2+2\beta_D/D)\ln D-\beta_D\ln\beta_D-(D-\beta_D)\ln(D-\beta_D)>0.\qquad(17)$$

A standard but tedious calculus argument shows that the function on the left side of the inequality in (17), that we denote by $g(D)$, is increasing for $D\geq 3$. Hence, from $g(D)>g(3)=0.3165$ we have (17), which in turn proves $\delta_D<\gamma_D$.

The function $\sigma_D+xD/\ln D$ is less than 1, for $0<x<\gamma_D$. From Lemma 5 and from $\delta_D<\gamma_D$ one obtains that the functions $f_D(x)$ and $\sigma_D+xD/\ln D$ intersect in the interval $x\in]\delta_D,\gamma_D[$ at one point θ_D.

From $\theta_D<\gamma_D$ and (16), it follows $\theta_D<(1+\ln\ln D)/D$. This concludes the proof. □

A direct computation shows that $\theta_3=0.2895$, $\theta_5=0.2451$, $\theta_{10}=0.1625$, and $\theta_{20}=0.0971$.

The bound provided by Theorem 3 can be further improved employing a refinement of Johnsen's technique [3].

Since the minimum of $f_D(x)$ satisfies $0.5<x_{D,\min}<1$ and $f_D(1)=1$, $f_D(\theta_D)=\sigma_D+\theta_DD/\ln D<1$ for $\theta_D<0.5$ and $f_D(x)$ is convex, there exists a unique solution of the equation $f_D(x)=f_D(\theta_D)=\sigma_D+\theta_DD/\ln D$ in the interval $]0.5,1[$. Denote this solution by η_D.

Suppose that only an upper bound on p_1, say p, is known and its exact value $0<p_1\leq p<1$, is unknown. Any upper bound on the redundancy r of the Huffman code for the source A, as a function of p, must be a nondecreasing function. Making use of Theorem 3 we obtain the following upper bound on r:

$$r\leq r_{\max}(p),\qquad(18)$$

where

$$r_{\max}(p)=\begin{cases}\sigma_D+pD/\ln D, & \text{if } 0<p\leq\theta_D\\ \sigma_D+\theta_DD/\ln D, & \text{if } \theta_D<p\leq\eta_D\\ 1-\mathscr{H}_D(p)+(1-p)2/D, & \text{if } \eta_D<p<1.\end{cases}$$

Letting $r_B=L_B-H_B$ the redundancy of the subsource $B=\{a_2,\cdots,a_N\}$, inequality (11) can be rewritten as

$$r\leq 1-\mathscr{H}_D(p_1)+(1-p_1)(2/D+r_B-1).\qquad(19)$$

The probability p_1' of the most likely source letter in B satisfies $p_1'\leq(\max_i p_i)/(1-p_1)=p_1/(1-p_1)$, and provides an upper bound on p_1' better than the trivial $p_1'\leq 1$, when $p_1<0.5$. From (18) and (19) we then obtain for $p_1<0.5$ the bound

$$r\leq 1-\mathscr{H}_D(p_1)+(1-p_1)(2/D+r_{\max}(p_1/(1-p_1))-1).\qquad(20)$$

The previous bound improves the bound stated in Theorem 3 for $p_1\in[\theta_D,0.5[$, since for $p<1$, $r_{\max}(p)<1$. Recalling the definition of $r_{\max}(p)$, and using the identity $\mathscr{H}_D(x)+(1-x)\mathscr{H}_D(x/(1-x))=\mathscr{H}_D(2x)+2x\log_D 2$, inequality (20) can be written as

$$r\leq\begin{cases}1-\mathscr{H}_D(p_1)+(1-p_1)(2/D+\sigma_D+\theta_DD/\ln D-1),\\ \qquad\text{if } 0<p_1\leq\eta_D/(1+\eta_D)\\ 1+4/D-2p_1(3/D+\log_D 2)-\mathscr{H}_D(2p_1),\\ \qquad\text{if } \eta_D/(1+\eta_D)<p_1<0.5.\end{cases}$$

We will now combine Gallager's bound, (9), and the two new derived bounds, (15) and (20), all together.

First observe that $\eta_D/(1+\eta_D)>1/3$ since $\eta_D>0.5$. Hence we get $\eta_D/(1+\eta_D)>\theta_D$. Set $h_D(x)=1-\mathscr{H}_D(x)+(1-x)(2/D+\sigma_D+\theta_DD/\ln D-1)$. Function $h_D(x)$ is a continuous, convex $\cup$ function of x, which, evaluated at $x=0$, gives $h_D(0)>\sigma_D$. Since $\sigma_D+\theta_DD/\ln D<1$, $h_D(x)<f_D(x)$, for $x<0.5$, then at the point $x=\theta_D$, we find $h_D(\theta_D)<f_D(\theta_D)=\sigma_D+$

$\theta_D D/\ln D$. Therefore the equation $h_D(x)=\sigma_D + xD/\ln D$, as a function of x, has a unique solution μ_D in the interval $]0,\gamma_D[$.

Thus bound (20) improves (9) for $\mu_D < p_1 \le \theta_D$, and improves (15) for $\theta_D \le p_1 < 0.5$. Combining (9), (15), and (20) we can finally summarize our results in the following theorem.

Theorem 4: Let p_1 be the probability of the most likely source letter of a discrete source A. The redundancy of the corresponding D-ary Huffman code is upper bounded by:

$$r \le \sigma_D + p_1 D/\ln D, \qquad \text{if } 0 < p_1 \le \mu_D, \tag{21}$$

$$r \le 1 - \mathscr{H}_D(p_1) + (1-p_1)(2/D + \sigma_D + \theta_D D/\ln D - 1), \qquad \text{if } \mu_D < p_1 \le \eta_D/(1+\eta_D), \tag{22}$$

$$r \le 1 + 4/D - 2p_1(3/D + \log_D 2) - \mathscr{H}_D(2p_1), \qquad \text{if } \eta_D/(1+\eta_D) < p_1 \le 0.5, \tag{23}$$

$$r \le 1 - \mathscr{H}_D(p_1) + (1-p_1)2/D, \qquad \text{if } 0.5 < p_1 < 1. \tag{24}$$

As a numerical example consider the case $D=3$, $\eta_3 = 0.9805$. Bounds (21)–(24) can be written as:

$$r \le 0.135 + 2.7307p_1, \qquad \text{if } 0 < p_1 \le 0.2769,$$

$$r \le 1.8925 - 0.8925p_1 - \mathscr{H}_3(p_1), \qquad \text{if } 0.2769 < p_1 \le 0.49507,$$

$$r \le 7/3 - 3.2618p_1 - \mathscr{H}_3(2p_1), \qquad \text{if } 0.49507 < p_1 \le 0.5,$$

$$r \le 5/3 - 2p_1/3 - \mathscr{H}_3(p_1), \qquad \text{if } 0.5 < p_1 < 1.$$

Again, this bound can be further improved by employing the same procedure. But the improvement achieved is not worth the effort. Therefore we do not proceed further in the recursive use of this technique.

IV. Additional Results

Now we derive a tight lower bound. Let

$$\alpha_{D,0} = 1,$$

$$\alpha_{D,m} = 1 - \left(\log_D \frac{D^{m+1}-1}{D^m - 1}\right)^{-1}, \qquad \text{if } m \ge 1.$$

For $m \ge 0$, it is $\alpha_{D,m+1} < \alpha_{D,m}$; moreover $\lim_{m\to\infty} \alpha_{D,m} = 0$.

Theorem 5: If for some $m \ge 1$,

$$\alpha_{D,m} < p_1 \le \alpha_{D,m-1},$$

the redundancy of any D-ary Huffman code satisfies

$$r \ge m - \mathscr{H}_D(p_1) - (1-p_1)\log_D(D^m - 1). \tag{25}$$

The bound is tight.

Proof: Assume that $\alpha_{D,m} < p_1 \le \alpha_{D,m-1}$. Then, since $1/(D^{m+1} - D + 1) \le \alpha_{D,m} < \alpha_{D,m-1} \le D/(D^m + D - 1)$ for $m > 1$, Lemma 2 implies that the minimum codeword length of the D-ary Huffman code can be $m-1$, m or $m+1$, if $m>1$, and 1 or 2 if $m=1$. If it is $m-1$, $m>1$, then from Lemma 4 we get

$$r \ge m - 1 - \mathscr{H}_D(p_1) - (1-p_1)\log_D(D^{m-1} - 1)$$
$$\ge m - \mathscr{H}_D(p_1) - (1-p_1)\log_D(D^m - 1)$$

since $p_1 \le \alpha_{D,m-1}$. If it is m, $m \ge 1$, then the lemma follows from Lemma 4. If it is $m+1$, $m \ge 1$, then from Lemma 4 we get

$$r \ge m + 1 - \mathscr{H}_D(p_1) - (1-p_1)\log_D(D^{m+1} - 1)$$
$$\ge m - \mathscr{H}_D(p_1) - (1-p_1)\log_D(D^m - 1)$$

since $\alpha_{D,m} < p_1$.

The lower bound is tight. Indeed, it is reached by a source consisting of $D^{m+1} - D + 1$ letters having probabilities $p_1 = p$ and $p_i = (1-p)/(D^{m+1} - D)$ for $2 \le i \le D^{m+1} - D + 1$. Notice that a Huffman code for this source has minimum codeword length $n_1 = m$, whereas remaining codeword lengths are $n_i = m + 1$. □

Notice that, putting $m=1$ in Theorem 5, we obtain a bound that corrects Golic and Obradovic's Theorem 2 [10]. Indeed, being greater than our tight lower bound, their bound is erroneous. The difference between the two bounds is $(1-p_1)\log_D(D-1)$, which is positive for any $p_1 < 1$, $D \ge 2$.[1]

Let $\phi^-(D, p_1)$ be the lower bound on r as stated in Theorem 2. Then it is easily seen that for any p_1, $1 > p_1 > 0$, it results $\lim_{D\to\infty}\phi^-(D,p_1) = p_1$. On the other hand, Theorem 2 states that for large D any upper bound approaches 1. Informally speaking, for large values of D the redundancy of D-ary Huffman codes can be bounded by $p_1 - \epsilon \le r \le 1$, $\epsilon > 0$, both inequalities being the best possible. Formally, using Theorem 2 and of the aforementioned limit, $\lim_{D\to\infty}\phi^-(D,p_1) = p_1$, the following is easily obtained.

Theorem 6: For every $\epsilon > 0$, there exists $D_0 \ge 2$ such that for $D \ge D_0$ the redundancy of D-ary Huffman codes can be bounded by $p_1 - \epsilon \le r \le 1$. Moreover, for every $\epsilon_1 > 0$, $\epsilon_2 > 0$ and $1 > p_1 > 0$ there exist two sources A_1 and A_2; where p_1 is the probability of the most likely letter such that redundancies r_1 and r_2 of the corresponding D-ary Huffman codes for A_1 and A_2 satisfy $p_1 - \epsilon < r_1 < p_1$ and $1 - \epsilon_2 < r_2 < 1$.

In the following theorem an upper bound on the codeword length n_a in a D-ary Huffman code in terms of the probability p_a of the source letter a is given. Notice that this bound differs from the one provided by Lemma 2. Now the letter a is not assumed to be the most likely letter of the source. The theorem is a generalization of Katona and Nemetz's Theorem 1 [7] to arbitrary D-ary Huffman codes.

Theorem 7: If for some $s \ge 2$ the probability p_a of a source letter a satisfies

$$\frac{1}{f_{D,s+1}} \le p_a < \frac{1}{f_{D,s}}, \tag{26}$$

where $f_{D,n} = f_{D,n-1} + (D-1)f_{D,n-2}$ for $n \ge 3$, with initial conditions $f_{D,1} = 1$ and $f_{D,2} = 1$, then its codeword length n_a in any D-ary Huffman code satisfies

$$n_a \le s - 1,$$

and this is the best possible bound.

Proof: Let $q_{i_1} = p_a, q_{i_2}, \cdots, q_{i_t} = 1$ be the probabilities of nodes on the path from the leaf corresponding to source letter a to the roof in its D-ary Huffman code tree. Then q_{i_j}, $1 \le j \le t$, is the probability corresponding to the node on level $t-j$ on such a path, and $t-1$ is the length of the codeword associated to a. We claim that for $j = 2, \cdots, t-1$,

$$q_{i_j} \ge f_{D,j} p_a. \tag{27}$$

Inequality (27) can be proved by induction. Indeed, for $j=2$, $q_{i_2} \ge q_{i_1} = p_a$ provides a basis for the induction. Now suppose (27) is true for $j \le k$ and consider the case $j = k+1$. Then $q_{i_{k+1}}$ is equal to the sum of D terms: one of them is q_{i_k} and each of the other $D-1$ terms is not smaller than $q_{i_{k-1}}$ by Property 1. Thus $q_{i_{k+1}} \ge q_{i_k} + (D-1)q_{i_{k-1}}$. And the claim follows recalling the definition of $f_{D,i}$.

Substituting $q_{i_t} = 1$ in (27) we obtain $1/f_{D,t} \ge p_a$, which, compared with (26), gives $t \le s+1$ and thus the theorem follows.

[1]One of the referees has communicated to us the existence of the manuscript "Corrections to "A lower bound on the redundancy of D-ary Huffman codes"" by J. Dj. Golic and M. M. Obradovic that gives a bound that still is not tight.

This is the best possible bound; i.e., for each p_a satisfying (26) a source A' exists that has $n_a = s-1$ in its D-ary Huffman code. Indeed, consider the source consisting of $2+(D-1)(s-2)$ letters where two letters have probabilities p and $1/f_{D,s}-p$ and, for each $j=1,2,\cdots,s-2$, there are $D-1$ letters with probability $f_{D,j}/f_{D,s}$. This source can be encoded by a D-ary Huffman code that has 2 codewords of length s and $D-1$ codewords of length j, for $s-1 \geq j \geq 1$. □

From

$$f_{D,n}=\frac{1}{\sqrt{4D-3}}\left[\left(\frac{1+\sqrt{4D-3}}{2}\right)^n-\left(\frac{1-\sqrt{4D-3}}{2}\right)^n\right]$$

we obtain the following.

Corollary 1: Let n_a be the codeword length for the source letter a in a D-ary Huffman code. Then

$$\limsup_{p_a\to 0}\frac{n_a}{I_D(a)}=\left[\log_D\frac{1+\sqrt{4D-3}}{2}\right]^{-1}, \tag{28}$$

where $I_D(a)=-\log_D p_a$ is the self-information of letter a.

Note that (28) approaches 2 as D gets large. Finally, following the reasoning of Katona and Nemetz's Theorem 2 [7], it is possible to prove the following theorem, which is an extension of their Theorem 2 to D-ary Huffman codes.

Theorem 8: If for some $s \geq 2$ the probabilities p_a and p_b of the source letters a and b satisfy

$$\frac{1}{f_{D,s+1}} \leq \frac{p_a}{p_b} < \frac{1}{f_{D,s}},$$

then their codeword lengths n_a and n_b in any D-ary Huffman code satisfy

$$n_a - n_b \leq s,$$

and this is the best possible bound.

V. Conclusion

We have investigated the redundancy of D-ary Huffman codes. In particular, we derived upper and lower bounds when the probability p_1 of the most likely source letter is provided. The upper bound improves a bound given by Gallager, while the lower bound corrects a bound recently proposed by Golic and Obradovic. We also showed that as D gets large the redundancy of a D-ary Huffman code can be bounded by $p_1-\epsilon \leq r < 1$, where ϵ is an arbitrary positive constant. As a complement, some results known for the binary case have been extended to arbitrary D-ary Huffman codes.

References

[1] R. G. Gallager, "Variation on a theme by Huffman," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 6, pp. 668–674, Nov. 1978.

[2] D. A. Huffman, "A method for the construction of minimum redundancy codes," *Proc. IRE*, vol. 40, no. 2, pp. 1098–1101, 1952.

[3] O. Johnsen, "On the redundancy of Huffman codes," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 2, pp. 220–222, Mar. 1980.

[4] R. M. Capocelli, R. Giancarlo, and I. J. Taneja, "Bounds on the redundancy of Huffman codes," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 6, pp. 854–857, Nov. 1986.

[5] R. M. Capocelli and A. De Santis, "Tight upper bounds on the redundancy of Huffman codes," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1084–1091, Sept. 1989.

[6] ____, "Tight bounds on the redundancy of Huffman codes," IBM Research Report RC-14154.

[7] G. O. H. Katona and T. O. H. Nemetz, "Huffman codes and self-information," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 3, pp. 337–340, May 1976.

[8] B. L. Montgomery and B. V. K. Vijaya Kumar, "On the average codeword length of optimal binary codes for extended sources," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 2, pp. 293–296, Mar. 1987.

[9] B. L. 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, no. 1, pp. 156–160, Jan. 1987.

[10] J. Dj. Golic and M. M. Obradovic, "A lower bound on the redundancy of D-ary Huffman codes," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 6, pp. 910–911, Nov. 1987.

A New Binary Code of Length 10 and Covering Radius 1

Patric R. J. Östergård, *Student Member, IEEE*

***Abstract*—A mixed code $C \subseteq F_4^1F_2^7$ of covering radius 1 that has 60 codewords is constructed. This code is then used to show that $K(10,1) \leq 120$.**

***Index Terms*—Binary codes, covering radius, simulated annealing.**

Let $K(n,R)$ denote the minimal number of codewords in any binary code of length n and covering radius R. Tables of bounds for $K(n,R)$ in the range $n \leq 33$, $R \leq 10$ appear in [1]. From [1] we know that $103 \leq K(10,1) \leq 128$. This work improves the upper bound to 120. The bound is also mentioned in [2], the construction of the result, however, is not presented in that paper.

We consider the space $F_4^aF_2^b$, where $F_4=\{0,1,2,3\}$ and $F_2=\{0,1\}$. The Hamming distance $d(x,y)$ between two words $x,y \in F_4^aF_2^b$ is the number of coordinates in which they differ. The covering radius of a code $C \subseteq F_4^aF_2^b$ is R if each word in $F_4^aF_2^b$ is within Hamming distance R from at least one codeword of C, and R is the smallest integer with this property.

Theorem 1: Let $C \subseteq F_4^1F_2^b$ be a code of covering radius R that has n codewords. Then there is a binary code $C' \subseteq F_2^{b+3}$ of covering radius R that has $2n$ codewords. Thus $K(b+3,R) \leq 2n$.

Proof: We prove that the code C' obtained by encoding the first coordinate of C using the rules

$$0 \to 000 \text{ or } 111$$
$$1 \to 001 \text{ or } 110$$
$$2 \to 010 \text{ or } 101$$
$$3 \to 011 \text{ or } 100$$

in all possible ways, fulfills the conditions. Apparently, $C' \subseteq F_2^{b+3}$ and the number of codewords of C' is $2n$. It then remains to show that C' has a covering radius of R. Consider an arbitrary $x=(x_1,x_2) \in F_2^{b+3}$, where $x_1 \in F_2^3$, $x_2 \in F_2^b$. Let $y \in F_4^1$ be the word that is encoded into x_1 according to the previous scheme. We now know that there is a codeword $c=(c_1,c_2) \in C \subseteq F_4^1F_2^b$ ($c_1 \in F_4^1$, $c_2 \in F_2^b$) such that $d(x',c) \leq R$, where $x'=(y,x_2) \in F_4^1F_2^b$. Let c' and c'' be the codewords of C' obtained from c as c_1 is encoded into c_1' and c_1''. If $c_1=y$, then $c_1'=x_1$ or $c_1''=x_1$, and so $d(x,\{c',c''\})=d(x_2,c_2)=d(x',c) \leq R$. If $c_1 \neq y$, then

Manuscript received April 17, 1990.

The author is with the Digital Systems Laboratory, Helsinki University of Technology, 02150 Espoo, Finland.

IEEE Log Number 9039287.