Acta Informatica 16, 363–370 (1981)

Acta Informatica

Best Huffman Trees

George Markowsky

Computer Sciences Department, IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598, USA

Summary. Given a sequence of positive weights, $W = w_1 \geqq \ldots \geqq w_n > 0$, there is a Huffman tree, $T\uparrow$ ("T-up") which minimizes the following functions: $\max\{d(w_i)\}$; $\Sigma d(w_i)$; $\Sigma f(d(w_i))\, w_i$ (here $d(w_i)$ represents the distance of a leaf of weight w_i to the root and f is a function defined for nonnegative integers having the property that $g(x) = f(x+1) - f(x)$ is monotone increasing) over the set of all trees for W having minimal expected length. Minimizing the first two functions was first done by Schwartz [5]. In the case of codes where W is a sequence of probabilities, this implies that the codes based on $T\uparrow$ have *all* their absolute central moments minimal. In particular, they are the least variance codes which were also described by Kou [3]. Furthermore, there exists a Huffman tree $T\downarrow$ ("T-down") which maximizes the functions considered above.

However, if $g(x)$ is monotone decreasing, $T\uparrow$ and $T\downarrow$ respectively maximize and minimize $\Sigma f(d(w_i)\, w_i$ over the set of all trees for W having minimal expected length. In addition, we derive a number of interesting results about the distribution of labels within Huffman trees. By suitable modifications of the usual Huffman tree construction, (see [1]) $T\uparrow$ and $T\downarrow$ can also be constructed in time $O(n \log n)$.

1. Introduction

The discussion in this paper will refer to trees rather than codes, since a number of arguments that we use depend on weights of internal nodes, which are much easier to visualize in the case of trees. There is a very simple relationship between trees and codes which is discussed in greater detail in [2, 3] and [4; Chap. 10]. Furthermore, only binary trees will be considered, since the results in the k-ary case can be derived in a straightforward manner from the results herein along the lines of Theorem 10.6 in [4].

At this point it might be helpful to informally recall some basic facts about Huffman trees. Essentially, we begin with a sequence of weights

$w_1 \geqq \ldots \geqq w_n > 0$ and we wish to construct a tree which has the weights at its leaves and minimizes the sum $\sum_{i=1}^{n} l_i w_i$, where l_i is the distance from the root to the leaf labeled by w_i. Huffman [2] gave a recursive procedure for constructing such a tree, T:

a) for $n=1$, it is obvious what the tree looks like;

b) for $n>1$, replace the sequence $w_1 \geqq \ldots \geqq w_n$, by the result of inserting $w_{n-1}+w_n$ into the sequence $w_1 \geqq \ldots \geqq w_{n-2}$ in the proper order and construct the tree T', for this shorter sequence;

c) take *any* leaf of T' having weight $w_{n-1}+w_n$ and create a new tree T by adding two sons to it bearing the labels w_{n-1} and w_n.

Notation. The following notation will be used throughout this paper.

1) W shall denote a sequence of positive weights $w_1 \geqq w_2 \geqq \ldots \geqq w_n > 0$.

2) $\mathscr{T}(W)$ shall denote the set of all pairs $(T, |\ |)$, where T is a (binary) tree having n leaves and $|\ |$ (called the weight function) maps the nodes of T into the positive reals in such a way that:

a) for each x, the number of leaves of T with weight equal to x is equal to the number of elements of W having value equal to x;

b) the weight of an internal node is the sum of the weight of its sons. For a node $v \in T$, we use $|v|$ to denote its weight. Note that $|\ |$ is determined by its values on the leaves. As usual, we shall just refer to $T \in \mathscr{T}(W)$ rather than $(T, |\ |)$. Note also that for the set X, $|X|$ will denote its cardinality.

3) For $T \in \mathscr{T}(W)$ and $v \in T$, $d(T, v)$ or $d(v)$ (if T is clear) will denote the distance between the root of T and v.

4) For $T \in \mathscr{T}(W)$, $L(T)$ shall denote the leaves of T.

5) Let f be a function defined for nonnegative integers and $T \in \mathscr{T}(W)$. $E(T, f)$ or $E(f)$ (if T is clear) shall denote the sum

$$\sum_{v \in L(T)} f(d(v))\,|v|.$$

$E(T, f)$ is called the expected value of f on T.

6) $\mathscr{O}(W)$ shall denote $\{T \in \mathscr{T}(W) \mid E(T, e) \text{ is minimal in } \mathscr{T}(W)\}$ where e is the identity function. $\mathscr{O}(W)$ is the set of *optimal merge trees*.

7) $\mathscr{H}(W)$ shall denote the set of all *Huffman trees* for the sequence W. As noted above, $\mathscr{H}(W) \subseteq \mathscr{O}(W)$. In general, not all optimal merge trees are Huffman trees.

8) For $T \in \mathscr{T}(W)$, m a non-negative integer and r a positive real number define the following functions:

a) $\mathrm{Node\#}(T, r) = |\{v \in T;\ |v| = r\}|$;

b) $\mathrm{Leaf\#}(T, r) = |\{v \in L(T);\ |v| = r\}|$;

c) $\mathrm{Minlev}(T, r) = 0$, if $\mathrm{Node\#}(T, r) = 0$
$= \min\{d(v) \mid v \in T, |v| = r\}$, else;

d) $\mathrm{Maxlev}(T, r) = 0$, if $\mathrm{Node\#}(T, r) = 0$,
$= \max\{d(v) \mid v \in L(T), |v| = r\}$, else;

e) $\mathrm{Minleaf}(T, r) = 0$, if $\mathrm{Leaf\#}(T, r) = 0$,
$= \min\{d(v) \mid v \in L(T), |v| = r\}$, else;

f) $\mathrm{Maxleaf}(T, r) = 0$, if $\mathrm{Leaf\#}(T, r) = 0$,
$= \max\{d(v) \mid v \in L(T), |v| = r\}$, else;

g) $\mathrm{Min\#}(T, r) = |\{v \in T \mid |v| = r, d(v) = \mathrm{Minlev}(T, r)\}|$;

h) $\mathrm{Max\#}(T, r) = |\{v \in T \mid |v| = r, d(v) = \mathrm{Maxlev}(T, r)\}|$;

i) $\mathrm{Nodelev}(T, r, m) = |\{v \in T \mid |v| = r, d(v) = m\}|$.

j) $\mathrm{Leaflev}(T, r, m) = |\{v \in L(T) \mid |v| = r, d(v) = m\}|$.

9) $\Delta(W)$ shall denote the sequence $w'_1 \geqq w'_2 \ldots \geqq w'_{n-1}$ which is the merge of $w_1 \geqq \ldots \geqq w_{n-2}$ and $w_{n-1} + w_n$.

Where convenient and unambiguous we will drop T as a variable in the various functions.

The goal of this paper is to show how to minimize (maximize) $E(T, f)$ for $T \in \mathscr{O}(W)$ and for various classes of f. The following lemma appears in [3] as Theorem 1 and shows that it is sufficient to just consider $T \in \mathscr{H}(W)$.

Lemma 1. *Given $T \in \mathscr{O}(W)$, there exists $T^* \in \mathscr{H}(W)$ such that for all r, m, $\mathrm{Leaflev}(T, r, m) = \mathrm{Leaflev}(T^*, r, m)$. In particular, for all f, $E(T, f) = E(T^*, f)$.*

Proof. The basic idea behind the proof is to show that we can rearrange T by switching weights around so that the two lightest weight nodes are brothers, without changing the expected value of the identity function or the distribution of weights on leaves. The proof proceeds by induction on n, the length of W. For $n=1$, the result is trivial. Assume $n \geqq 2$. Since $T \in \mathscr{O}(W)$, every internal node has degree 2. Let v_1 and v_2 be two brothers, such that $|v_2| \geqq |v_1|$, and $d(v_1)$ is maximal over all nodes of T. Let v_3 and v_4 be leaves such that $|v_3| = w_{n-1}$ and $|v_4| = w_n$. Define $(T_1, |\ |_1) \in \mathscr{T}(W)$, by $T_1 = T$ and $|v|_1 = |v|$ for $v \in L(T) - \{v_1, v_2, v_3, v_4\}$, $|v_i|_1 = |v_{5-i}|$ for $i = 1, \ldots, 4$. For $v \notin L(T)$, $|v|_1$ has the value determined by its values at the leaves, subject to the constraint that the weight of an internal node is the sum of the weights of its sons. Note that

$$(*) \qquad E(T_1, e) = E(T, e) + (|v_3| - |v_2|)\,(d(v_2) - d(v_3)) + (|v_4| - |v_1|)\,(d(v_1) - d(v_4)).$$

Furthermore, the following relations hold: $|v_2| \geqq |v_1| \geqq w_n = |v_4|$; $|v_2| \geqq w_{n-1} = |v_3|$; $d(v_1) = d(v_2) \geqq d(v_3), d(v_4)$. The relations and (*) imply that $E(T_1, e) \leqq E(T, e)$. However, since $E(T, e)$ is minimal, we must have

$$(|v_3| - |v_2|)\,(d(v_2) - d(v_3)) = (|v_4| - |v_1|)\,(d(v_1) - d(v_4)) = 0.$$

Thus in the switch we made, either we switched identical weights or weights at identical distances from the root. Thus for all r, m, $\mathrm{Leaflev}(T, r, m) = \mathrm{Leaflev}(T_1, r, m)$.

Let $(T_2, |\ |_2)$ be defined as follows. T_2 is T_1 with v_1 and v_2 dropped and $|\ |_2$ is the restriction of $|\ |_1$ to T_2. Clearly, $T_2 \in \mathscr{T}(\Delta(w))$. We claim that $T_2 \in \mathscr{O}(\Delta(W))$. Note that $E(T_2, e) = E(T_1, e) - (w_{n-1} + w_n)$. If $T_2 \notin \mathscr{O}(\Delta(W))$, there would exist $T_3 \in \mathscr{O}(W)$ with $E(T_3, e) < E(T_1, e)$. By adding nodes with weights w_{n-1} and w_n as sons of a node of T_3 with weight $w_{n-1} + w_n$ we could get a tree T_4 with $E(T_4, e) = E(T_3, e) + w_{n-1} + w_n < E(T_2, e) + w_{n-1} + w_n = E(T_1, e)$, which is impossible since

$T_1 \in \mathscr{C}(W)$. Thus $T_2 \in \mathscr{C}(\Delta(w))$ and from induction it follows that there exists $T^{**} \in \mathscr{H}(\Delta(w))$ such that $\mathrm{Leaflev}(T_2, r, m) = \mathrm{Leaflev}(T^{**}, r, m)$ for all r, m. Let $T^* \in \mathscr{H}(W)$ be constructed from T^{**} by adding sons to a leaf of T^{**} having weight $w_{n-1}+w_n$ and the same value of d as the leaf used to derive T_1 from T_2. One can easily check that $\mathrm{Leaflev}(T^*, r, m) = \mathrm{Leaflev}(T', r, m) = \mathrm{Leaflev}(T, r, m)$ for all r, m. □

2. Main Results

The following lemma is essentially Lemma 5 in [3].

Lemma 2. *Let $T, T' \in \mathscr{H}(W)$. Then:*

a) $\mathrm{Node}\#(T, r) = \mathrm{Node}\#(T', r)$ *for all* r;

b) *if $v, v^* \in T$ are such that $|v| > |v^*|$, then $d(v) \leqq d(v^*)$;*

c) *if $v_1, v_2 \in T$ are such that $|v_1| = |v_2|$, then $|d(v_1) - d(v_2)| \leqq 1$.*

Proof. a) This follows immediately from induction on n since all Huffman trees are constructed by picking the two smallest weights and replacing them by their sum.

b) We proceed by induction on n, the case $n=1$ being trivial. As noted earlier, T can be constructed from some $T^* \in \mathscr{H}(\Delta(W))$ in the usual way. Let v_1 and v_2 be the two nodes with weights w_{n-1} and w_n added to T^* in order to obtain T. By induction, b) holds for T^*. Thus, the only way b) can fail for T is for there to exist $v \in T^*$ with $|v| > |v_2|$ and $d(T^*, v) = d(T, v) > d(T, v_2) = d(T, v_1)$. Since v is not the root it has a brother v' such that $|v'| \geqq w_{n-1}$, since $v' \in T^*$ and thus was added after the two *smallest* labels (w_{n-1} and w_n) were. However, the father, v^*, of v and v' has weight $|v| + |v'| > |v_1| + |v_2| = |v^+|$ where v^+ is the father of v_1 and v_2. But $v^+ \in T^*$, whence $d(v^*) \leqq d(v^+)$ and $d(v) = d(v^*) + 1 \leqq d(v^+) + 1 = d(v_2)$, yielding a contradiction.

c) If $v_1 = v_2$, we are done. If $v_1 \neq v_2$, then neither is the root. Let v_3 be the father of v_1 and v_4 the father of v_2. Since $|v_3| > |v_1| = |v_2|$ and $|v_4| > |v_2| = |v_1|$, we can use b) to conclude that $d(v_2) \geqq d(v_3) = d(v_1) - 1$ and $d(v_1) \geqq d(v_4) = d(v_2) - 1$, which implies that $|d(v_1) - d(v_2)| \leqq 1$. □

We are now ready to define the Huffman trees of greatest interest to us.

Definition. 1. To construct $T\uparrow(W)$ or $T\uparrow$, which we call *T-up*, we follow the recipe given in the beginning of the paper, except that we modify step c) by adding the sons to a leaf, v, of $T\uparrow(\Delta(W))$ having $|v| = w_{n-1} + w_n$ and $d(v) = \mathrm{Minleaf}(T\uparrow(\Delta(W)), |v|)$. The definition of $T\downarrow(W)$, *T-down*, is similar except that we require $d(v) = \mathrm{Maxleaf}(T\downarrow(\Delta(W)), |v|)$.

We can now describe some of the key properties of $T\uparrow$ and $T\downarrow$. Note that $T\uparrow$ and $T\downarrow$ are not unique in general.

Theorem 1. *For all $T \in \mathscr{H}(W)$ and all r, $\mathrm{Minlev}(T\uparrow, r) \leqq \mathrm{Minlev}(T, r)$, $\mathrm{Maxlev}(T\downarrow, r) \geqq \mathrm{Maxlev}(T, r)$. Furthermore, in case of equality, $\mathrm{Min}\#(T\uparrow, r) \geqq \mathrm{Min}\#(T, r)$, $\mathrm{Max}\#(T\uparrow, r) \geqq \mathrm{Max}\#(T, r)$.*

Proof. The proof is by induction on n. Clearly, the only values of r we must consider are w_{n-1} and w_n. There are two cases to consider.

Case 1. $w_{n-1} > w_n$.

Here we must have that $\mathrm{Node}\#(T, w_{n-1}+w_n) = \mathrm{Leaf}\#(T, w_{n-1}+w_n)$ for all $T \in \mathscr{H}(\Delta(W))$, since the sum of any two weights in $\Delta(W)$ exceeds $w_{n-1}+w_n$. Note also that for all $T \in \mathscr{H}(W)$,

$$\mathrm{Minlev}(T, w_n) = 1 + \mathrm{Minlev}(T, w_{n-1}+w_n) = 1 + \mathrm{Minlev}(T', w_{n-1}+w_n) \geqq$$
$$1 + \mathrm{Minlev}(T\uparrow(\Delta W), w_{n-1}+w_n) = \mathrm{Minlev}(T\uparrow(W), w_n)$$

where $T' \in \mathscr{H}(\Delta(W))$ is the tree from which T was constructed using Huffman's procedure. Finally, since $\mathrm{Node}\#(T, w_n) = 1$ for all $T \in \mathscr{H}(W)$, we are done in the case $r = w_n$.

If $w_{n-2} > w_{n-1}$, the same argument as for w_n holds in this case. Thus it only remains to consider the situation with $w_{n-2} = w_{n-1}$. Let $v \in T \in \mathscr{H}(\Delta(W))$ be such that $|v| = w_{n-2}$. Since the father v' of v has $|v'| \geqq 2w_{n-2} > w_{n-1} + w_n$, by Lemma 2 b) $d(v') \leqq \mathrm{Minlev}(T, w_{n-1}+w_n)$. Thus, if $T' \in \mathscr{H}(W)$ is derived from T in the usual way $\mathrm{Minlev}(T, w_{n-1}) = \mathrm{Minlev}(T', w_{n-1})$. Hence, for all $T \in \mathscr{H}(W)$, $\mathrm{Minlev}(T, w_{n-1}) \geqq \mathrm{Minlev}(T\uparrow(W), w_{n-1})$.

To conclude Case 1 it is only necessary to analyze the situation where $\mathrm{Minlev}(T, w_{n-1}) = \mathrm{Minlev}(T\uparrow(W), w_{n-1})$ for some $T \in \mathscr{H}(W)$. As usual, let $T' \in \mathscr{H}(\Delta(W))$ be the predecessor for T. We need to show that $\mathrm{Min}\#(T\uparrow, w_{n-1}) \geqq \mathrm{Min}\#(T, w_{n-1})$. Let i be such that $w_{n-j} = w_{n-1}$ for all $1 \leqq j \leqq i \geqq 2$, but $w_{n-i-1} > w_{n-1}$.

If $i = 2k+1$, k nodes of weight $2w_{n-1}$ have two sons of weight w_{n-1} and one node of weight $w_{n-1}+w_n$ has one son of weight w_{n-1} for any $T \in \mathscr{H}(W)$. From the definition of $T\uparrow$, there will be $2 \times \min\{\mathrm{Min}\#(T\uparrow, 2w_{n-1}), k\}$ nodes of weight w_{n-1} having distance $= \mathrm{Minlev}(T\uparrow, 2w_{n-1}) + 1$ and one node having distance $\mathrm{Minlev}(T\uparrow, w_{n-1}+w_n)$. Since, nodes of weight $> w_{n-1}$ exist in trees in $\mathscr{H}(\Delta W)$, $\mathrm{Min}\#(T\uparrow, 2w_{n-1}) \geqq \mathrm{Min}\#(T, 2w_{n-1})$ for all $T \in \mathscr{H}(W)$ with $\mathrm{Minlev}(T, w_{n-1}) = \mathrm{Minlev}(T\uparrow, w_{n-1})$ and $\mathrm{Minlev}(T\uparrow, w_{n-1}+w_n) \leqq \mathrm{Minlev}(T, w_{n-1}+w_n)$. This shows that $\mathrm{Min}\#(T\uparrow, w_{n-1}) \geqq \mathrm{Min}\#(T, w_{n-1})$.

If $i = 2k$, essentially the same argument holds. In this case there will be a node of weight $w^* = \min\{w_{n-1-i} + w_{n-1}, 2w_{n-1} + w_n\}$ having one son of weight w_{n-1}. It becomes necessary to compare $\mathrm{Minlev}(T\uparrow, w^*)$ and $\mathrm{Minlev}(T\uparrow, 2w_{n-1})$, but otherwise the argument is the same.

Case 2. $w_{n-1} = w_n$.

This argument is very similar to the one in Case 1. First, one must consider the situation where $w_{n-2} > w_{n-1}$. Here $\mathrm{Min}\#(T\uparrow, w_n) = 2$ and $\mathrm{Minlev}(T\uparrow, w_n) = \mathrm{Minlev}(T\uparrow, 2w_n) + 1$.

If $w_{n-2} = w_{n-1}$, then let i be such that $w_{n-j} = w_n$ for all $1 \leqq j \leqq i \geqq 2$, but $w_{n-i-1} > w_{n-1}$. As in Case 1, we consider the consequences of i being odd or even in almost the identical way to conclude our proof.

For $T\downarrow$, the arguments are very similar and we leave them to the reader. □

Before proceeding to results dealing with $E(T\uparrow, f)$, we show that $T\uparrow$ has the properties considered by E.S. Schwartz in [5].

Corollary 1. *For all* $T \in \mathscr{H}(W)$, $\max\{d(v) \mid v \in L(T\uparrow)\} \leqq \max\{d(v) \mid v \in L(T)\}$ *and* $\sum_{v \in L(T\uparrow)} d(v) \leqq \sum_{v \in L(T)} d(v)$.

Proof. From Lemma 2 it follows that for all $T \in \mathscr{H}(W)$, $\max\{d(v) \mid v \in L(T)\} = \mathrm{Maxlev}(T, w_n)$. Theorem 1 implies that $\mathrm{Minlev}(T\uparrow, w_n) \leqq \mathrm{Minlev}(T, w_n)$ for all $T \in \mathscr{H}(W)$. If the inequality is strict, we are done since $\mathrm{Maxlev}(T, w_n) \leqq \mathrm{Minlev}(T, w_n)+1$. On the other hand if $\mathrm{Minlev}(T\uparrow, w_n) = \mathrm{Minlev}(T, w_n)$, then $\mathrm{Min}\#(T\uparrow, w_n) \geqq \mathrm{Min}\#(T, w_n)$. Thus if $\mathrm{Maxlev}(T\uparrow, w_n) \neq \mathrm{Minlev}(T\uparrow, w_n)$, $\mathrm{Maxlev}(T, w_n) \neq \mathrm{Minlev}(T, w_n)$. This implies that $\mathrm{Maxlev}(T\uparrow, w_n) = \mathrm{Minlev}(T\uparrow, w_n)+1 = \mathrm{Minlev}(T, w_n)+1 = \mathrm{Maxlev}(T, w_n)$.

The proof that the second property holds proceeds by induction. If when we go from $T\uparrow(\Delta W)$ to $T\uparrow(W)$, we attach nodes of weight w_{n-1} and w_n to a node v with $|v| = w_{n-1}+w_n$ and $d(v) = \mathrm{Minlev}(T\uparrow(W), |v|)$, and the minimality of the sum for $T\uparrow(W)$ follows immediately from the minimality of the sum for $T\uparrow(\Delta W)$. However, if $d(v) = \mathrm{Minlev}(T\uparrow(W), |v|)+1$, $w_{n-1} = w_n$ and all nodes of weight $|v|$ at distance $\mathrm{Minlev}(T\uparrow(W), |v|)$ from the root must already have two sons of weight $w_{n-1} = w_n$. From Theorem 1, this is the largest possible number of nodes of weight w_n as close to the root as possible, which clearly minimizes $\sum_{v \in T} d(v)$. □

Theorem 2. *Let f be a function defined for all nonnegative integers and $g(x) = f(x+1)-f(x)$ for all nonnegative integers. If g is monotone increasing, then $E(T\uparrow, f) \leqq E(T, f)$ and $E(T\downarrow, f) \geqq E(T, f)$ for all $T \in \mathcal{O}(W)$. Similarly, if g is monotone decreasing, then $E(T\downarrow, f) \leqq E(T, f)$ and $E(T\uparrow, f) \geqq E(T, f)$ for all $T \in \mathcal{O}(W)$.*

Proof. For all $T \in \mathscr{H}(W)$, we have $E(T, f) = E(T', f) + (w_{n-1}+w_n)(f(A+1) - f(A))$ where $T' \in \mathscr{H}(\Delta W)$ is a predecessor of T and A is the level of the node to which we add the nodes of weight w_{n-1} and w_n. If for $T\uparrow(W)$, $A \underset{\mathrm{def}}{=} \alpha = \mathrm{Minlev}(T\uparrow(W), w_{n-1}+w_n)$, we are done since $E(T\uparrow(\Delta W), f) \leqq E(T', f)$ and by Theorem 1 and the fact that g is monotone increasing $(w_{n-1}+w_n)(f(A+1) - f(A))$ is the smallest possible increment. In fact, the only troublesome case occurs when $A = \alpha$ for some T' and $T \in \mathscr{H}(W)$, while $A = \alpha+1$ for $T\uparrow(\Delta W)$ and $T\uparrow(W)$. This can only happen when $w_{n-1} = w_n$. In this case, consider T^* and $T\uparrow(W^*)$ where $T^* \in \mathscr{H}(W^*)$ is the largest ancestor of T having no brothers of weight w_n and $T\uparrow(W^*)$ is the corresponding ancestor of $T\uparrow(W)$ where W^* is the corresponding sequence. In particular, if $W = w_1 \geqq w_2 \geqq \ldots > w_k = w_{k+1} = \ldots = w_{n-1} = w_n$ and $n-k$ is odd, W^* is the merge of $w_1 \geqq \ldots \geqq w_{k-1}$ and a sequence of $(n-k+1)/2$ terms all equal to $2w_n$. If $n-k$ is even, W^* depends on whether $w_{k-1} < 2w_n$ or not. If $w_{k-1} < 2w_n$, W^* is the merge of $w_1 \geqq \ldots \geqq w_{k-2}$ and a sequence beginning with $w_{k-1}+w_n$ and followed by $(n-k)/2$ terms all equal to $2w_n$. If $w_{k-1} \geqq 2w_n$, W^* is the merge of $w_1 \geqq \ldots \geqq w_{k-1}$ and a sequence beginning with $3w_n$ and followed by $(n-k-2)/2$ terms all equal to $2w_n$.

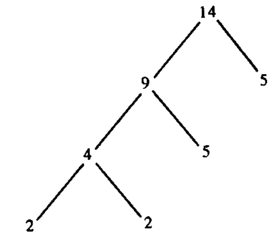

Fig. 1

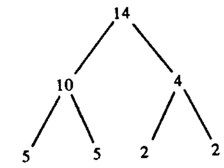

Fig. 2

We know that $E(T\uparrow(W^*), f) \leqq E(T^*, f)$ and in computing $E(T\uparrow(W), f)$ we add as many $(f(\alpha+1)-f(\alpha))$'s and as few $(f(\alpha+2)-f(\alpha+1))$'s as possible, establishing the minimality of $E(T\uparrow(W), f)$. The other arguments are similar. □

Note that f need not be monotone increasing itself nor need it be nonnegative. As a result of this, we have the following Corollary.

Theorem 3. *Let n be a nonnegative integer, c any real number and $f(x) = |x-c|^n$, then $E(T\uparrow, f) \leqq E(T, f)$ for all $T \in \mathscr{H}(W)$, i.e., all absolute moments (central, at a point, etc.) for W are simultaneously minimized by $T\uparrow$. In particular, $T\uparrow$ has the smallest variance of any $T \in \mathcal{O}(W)$.*

Proof. To apply Theorem 2, we need only show that $g(x+1) \geqq g(x)$, i.e., that $|x+2-c|^n + |x-c|^n \geqq 2|x+1-c|^n$. There are several cases to consider.

Case 1. $x \geqq c$ $(x+2 \leqq c)$.

Let $a = x-c$ $(a = c-x-2)$, then we need only show $(a+2)^n + (a)^n \geqq 2(a+1)^n$, for $a \geqq 0$. This however follows quickly from the binominal theorem.

Case 2. $x < c \leqq x+1$ $(x+1 < c \leqq x+2)$.

Let $a = x+1-c$ $(a = c-x-1)$, then we need only show that $2a^n \leqq (a+1)^n + (1-a)^n$ for $0 \leqq a < 1$, but this is immediate since $2a^n \leqq a^n + 1 \leqq (a+1)^n$. □

If we are dealing with codes so that the w_i's are probabilities with sum 1, Theorem 3 shows that the distribution of lengths of the code given by $T\uparrow$ is clustered more sharply around the mean (or any other point for that matter) than the distributions for any other $T \in \mathcal{O}(W)$. This result, together with the others in this section show why we feel $T\uparrow$ deserves to be called the "best" Huffman tree.
Note that Theorem 2 is false in general if we drop the requirement that $T \in \mathcal{O}(W)$ as can be seen in Example 1.

Example. 1. Let $W = 5, 5, 2, 2$. Then $T\uparrow$ is given in Fig. 1. Let T be the tree given in Fig. 2. Note that $E(T\uparrow, e) = \cancel{21}$ 27 but $E(T, e) = \cancel{24}$ 28, whence $T \notin \mathcal{O}(W)$. Let $f(x) = x^2$. Clearly, in this case $g(x)$ is strictly increasing. However, $E(T\uparrow, f) = 61$ whereas $E(T, f) = 56$.

Acknowledgement. The author would like to thank Larry Kou for a number of very stimulating discussions concerning Huffman trees and codes, and Larry Carter for reading a preliminary draft of this paper and making many helpful suggestions for improving its readability.

References

1. Even, S.: Algorithmic combinatorics. New York: Macmillan 1973
2. Huffman, D.A.: A method for the construction of minimum-redundancy codes. Proc. IRE **40**, 1098–1101 (1952)
3. Kou, L.T.: Minimal variance Huffman codes. IBM Technical Report RC8333, T.J. Yorktown Heights, NY: Watson Research Center 1980
4. McEliece, R.J.: The theory of information and coding. Reading, Mass.: Addison-Wesley 1977
5. Schwartz, E.S.: An optimal encoding with minimum longest code and total number of digits. Information and control, **7**, 37–44 (1964)

Received January 12, 1981