Proof of Optimality of Huffman Coding

Recall that the problem is given frequencies $f_1, \ldots, f_n$ to find the optimal prefix-free code that minimizes

$$\sum_{i}^{n} f_i \cdot (\text{length of encoding of the } i\text{-th symbol}).$$

This is the same as finding the full binary tree with n leaves, one per symbol in $1, \ldots, n$, that minimizes

$$\sum_{i=1}^{n} f_i \cdot (\text{depth of leaf of the } i\text{-th symbol})$$

Recall that we showed in class the following key claim.

Claim 1 (Huffman's Claim)**.** *There's an optimal tree where the two smallest frequency symbols mark siblings (which are at the deepest level in the tree).*

We proved this via an exchange argument. Then, we went on to prove that Huffman's coding is optimal by induction. We repeat the argument in this note.

Claim 2. *Huffman's coding gives an optimal cost prefix-tree tree.*

Proof. The proof is by induction on n, the number of symbols. The base case $n = 2$ is trivial since there's only one full binary tree with 2 leaves.

Inductive Step: We will assume the claim to be true for any sequence of $n-1$ frequencies and prove that it holds for any n frequencies. Let $f_1, \ldots, f_n$ be any n frequencies. Assume without loss of generality that $f_1 \leq f_2 \leq \ldots \leq f_n$ (by relabeling). By Claim 1, there's an optimal tree T for which the leaves marked with 1 and 2 are siblings. Let's denote the tree that Huffman strategy gives by H. Note that we are not claiming that $T = H$ but rather that T and H have the same cost.

We will now remove both leaves marked by 1 and 2 from T, making their father a new leaf with frequency $f_1 + f_2$. This gives us a new binary tree T' on $n-1$ leaves with frequencies $f_1 + f_2, f_3, f_4, \ldots, f_n$. We do the same for the Huffman tree giving us a tree H' on $n-1$ leaves with frequencies $f_1 + f_2, f_3, f_4, \ldots, f_n$. Note that H' is exactly the Huffman tree on frequencies $f_1 + f_2, f_3, f_4, \ldots, f_n$ by definition of Huffman's strategy. By the induction hypothesis,

$$cost(H') = cost(T').$$

Observe further that

$$cost(T') = cost(T) - (f_1 + f_2)$$

since to get T' from T we replaced two nodes with frequencies f_1 and f_2 at some depth d with one node with frequency $f_1 + f_2$ at depth $d - 1$. This lowers the cost by $f_1 + f_2$. Similarly,

$$cost(H') = cost(H) - (f_1 + f_2).$$

Combining the three equations together we have that

$$cost(H) = cost(H') + f_1 + f_2 = cost(T') + f_1 + f_2 = cost(T).$$

□