Lecture 7 (Huffman Encoding cotd)

1 Theorem

Let $F = (f_1, f_2 \cdots, f_n)$ represent frequencies of symbols $(a_1, a_2, \cdots, a_n)$ and i, j be such that $\exists$ an optimal tree in which a_i, a_j are siblings. Then $F^* = (F \setminus \{f_i, f_j\}) \cup \{f^*\}$ where $f^* = f_i + f_j$ satisfies

$$opt(F) = f_i + f_j + opt(F^*)$$

2 Proof

2.1 $opt(F) \leq opt(F^*) + f_i + f_j$

This is obvious from construction (too lazy to write it formally here)

3 $opt(F^*) \leq opt(F) - (f_i + f_j)$

Consider the optimal tree for F, i, j are siblings and the total length of string of F^* will be $f_i + f_j$ lesser than the length of F since we have *atleast* one level lesser for f^*. Thus, we have an upper bound for $opt(F^*)$.

4 BFS

Discussed in the later half of the lecture