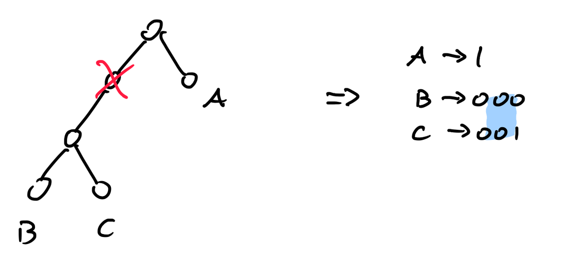
WE NOW PROVE SEVERAL PROPERTIES OF OPTIMAL SOLUTIONS, IN ORDER TO SIMPLIFY OUR SEARCH FOR AN ALGORITHM (I.E., IN ORDER TO SHRINK THE SEARCH SPACE).

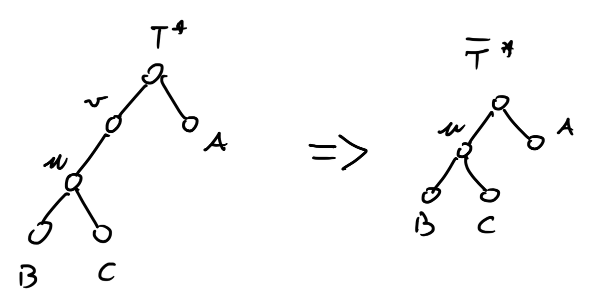
RECALL THAT $f_x > 0 \quad \forall x \in S$

L1: SUPPOSE THAT T^* IS AN OPTIMAL TREE. THEN, EACH INTERNAL NODE OF T^* HAS EXACTLY TWO CHILDREN.



P: BY CONTRADICTION, SUPPOSE THAT v IS AN INTERNAL NODE OF T^* WITH EXACTLY ONE CHILD u.

THEN, THE TREE $\bar{T}^*$ THAT HAS THE SUBTREE OF T^* ROOTED IN u IN PLACE OF THE SUBTREE OF T^* ROOTED IN v,



GIVES TO EACH LETTER A DEPTH NOT LARGER THAN THE ONE IT HAD IN T^*.

THAT IS, FOR EACH $x \in S$, $\mathrm{DEPTH}_{\bar{T}^*}(x) \le \mathrm{DEPTH}_{T^*}(x)$.

MOREOVER, IF $x \in S$ IS ONE OF THE LABELS OF THE LEAVES IN THE TREE ROOTED IN v, THEN $\mathrm{DEPTH}_{\bar{T}^*}(x) = \mathrm{DEPTH}_{T^*}(x) - 1 < \mathrm{DEPTH}_{T^*}(x)$.

GIVEN THAT THE TREE ROOTED IN v HAS AT LEAST ONE LEAF, IT MUST HOLD THAT $\mathrm{ABL}(\bar{T}^*) < \mathrm{ABL}(T^*)$. ◻

THEN, NO OPTIMAL TREE CONTAINS NODES WITH EXACTLY ONE CHILD.

FURTHER,

L2: IF T^* IS AN OPTIMAL SOLUTION (AN OPTIMAL LABELLED BINARY TREE) AND u, v ARE TWO LEAVES OF T^* SUCH THAT

(i) $\mathrm{DEPTH}_{T^*}(u) < \mathrm{DEPTH}_{T^*}(v)$, AND

(ii) THE LABEL OF u IS $y \in S$, AND THE LABEL OF v IS $x \in S$

THEN $f_y \ge f_x$.

P: RECALL THAT THE OBJECTIVE FUNCTION IS

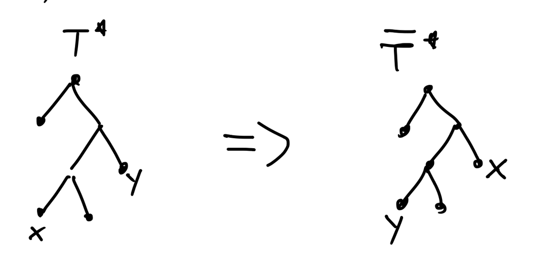
$$\mathrm{ABL}(T^*) = \sum_{z \in S} \left(f_z \, \mathrm{DEPTH}_{T^*}(z) \right).$$

THEN,

$$\mathrm{ABL}(T^*) = \sum_{z \in S - \{x,y\}} \left(f_z \, \mathrm{DEPTH}_{T^*}(z) \right) + f_x \, \mathrm{DEPTH}_{T^*}(x) + f_y \, \mathrm{DEPTH}_{T^*}(y)$$

BY CONTRADICTION, ASSUME THAT $f_y < f_x$.

LET $\bar{T}^*$ BE THE TREE T^* WITH THE LABELS x AND y SWAPPED.



THEN,

$$\mathrm{ABL}(\bar{T}^*) = \sum_{z \in S - \{x,y\}} \left(\mathrm{DEPTH}_{\bar{T}^*}(z) \, f_z \right) + f_x \, \mathrm{DEPTH}_{\bar{T}^*}(x) + f_y \, \mathrm{DEPTH}_{\bar{T}^*}(y)$$

$$= \sum_{z \in S - \{x,y\}} \left(\mathrm{DEPTH}_{T^*}(z) \cdot f_z \right) + f_x \, \mathrm{DEPTH}_{T^*}(y) + f_y \, \mathrm{DEPTH}_{T^*}(x)$$

THUS,

$$\mathrm{ABL}(T^*) - \mathrm{ABL}(\bar{T}^*) = f_x \, \mathrm{DEPTH}_{T^*}(x) + f_y \, \mathrm{DEPTH}_{T^*}(y) - f_x \, \mathrm{DEPTH}_{T^*}(y) - f_y \, \mathrm{DEPTH}_{T^*}(x)$$

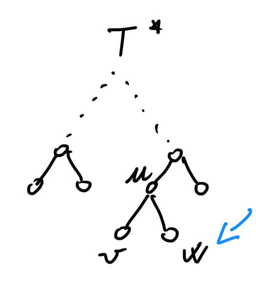
$$= \mathrm{DEPTH}_{T^*}(x)\left(f_x - f_y\right) - \mathrm{DEPTH}_{T^*}(y)\left(f_x - f_y\right)$$

$$= \underbrace{\left(\mathrm{DEPTH}_{T^*}(x) - \mathrm{DEPTH}_{T^*}(y)\right)}_{>0} \underbrace{\left(f_x - f_y\right)}_{>0} > 0$$

GIVEN THAT $\mathrm{ABL}(T^*) > \mathrm{ABL}(\bar{T}^*)$, THE TREE T^* IS NOT OPTIMAL. CONTRADICTION. ◻

THEN, GIVEN THE STRUCTURE OF THE OPTIMAL TREE, WE SHOULD ASSIGN LETTERS "GREEDILY": LESS FREQUENT LETTERS SHOULD BE ASSIGNED TO DEEPER LEAVES.

NOW, LET v BE A LEAF OF AN OPTIMAL T^* HAVING MAXIMUM DEPTH, LET u BE THE PARENT OF v, AND LET w BE THE OTHER CHILD OF u.

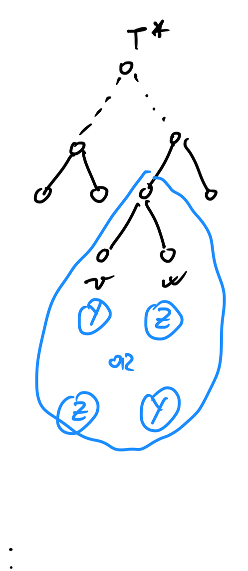


w EXISTS BECAUSE OF L1.

L3: w IS A LEAF OF T^*.

P: IF w WAS NOT A LEAF, THEN ITS SUBTREE CONTAINS A LEAF w' THAT IS DEEPER THAN w. BUT $\mathrm{DEPTH}_{T^*}(w) = \mathrm{DEPTH}_{T^*}(v)$, SINCE v AND w ARE SIBLINGS. GIVEN THAT v IS A LEAF OF MAXIMUM DEPTH, w' CANNOT EXIST. CONTR. ◻

THEN v AND w ARE TWO SIBLING LEAVES OF MAXIMUM DEPTH. THEN THERE EXISTS AN OPTIMAL LABELLING OF T^* THAT ASSIGNS TWO LETTERS OF SMALLEST FREQ. TO v AND w.

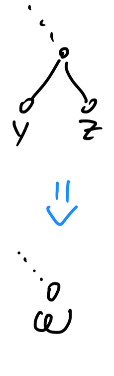


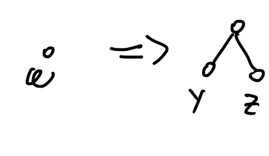
$f_A \ge f_B \ge \dots \ge f_Y \ge f_Z$

WE THEN PROVED:

T4: $\exists$ OPTIMAL PREFIX CODE, WITH TREE T^*, THAT ASSIGNS TWO LETTERS OF MINIMUM FREQUENCY TO TWO SIBLING LEAVES OF T^*.

SINCE THE TWO LETTERS OF SMALLEST FREQUENCY (SAY, Y AND Z) WILL BE CHILDREN OF THE SAME PARENT, AN ALGORITHM COULD REMOVE THE SUBTREE ROOTED IN THIS PARENT AND SUBSTITUTE IT WITH A NEW META-LETTER ω.

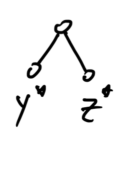




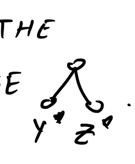
WE CAN THEN GREEDILY ITERATE.

HUFFMAN(S, f): $\quad (|S| \ge 2)$

IF $|S| == 2$:

- LET $S = \{y^*, z^*\}$
- ENCODE y^* WITH 0 AND z^* WITH 1 (OR, VICEVERSA).
- SET $T =$ 

ELSE:

- LET y^* AND z^* BE TWO LETTERS OF SMALLEST FREQUENCIES
- LET $S' = S - \{y^*, z^*\} \cup \{\omega_{y^*,z^*}\}$
- LET f' BE SUCH THAT
  (i) $f'_{\omega_{y^*,z^*}} = f_{y^*} + f_{z^*}$, AND
  (ii) $f'_x = f_x \quad \forall x \in S - \{y^*, z^*\}$
- RECURSIVELY BUILD AN OPTIMAL PREFIX-CODE FOR S', f'. LET T' BE TREE ASSOCIATED TO THIS CODE.
- CREATE A TREE T BY SUBSTITUTING THE LEAF LABELED ω IN T', WITH THE TREE 

RETURN T.