On the Convex Hull of Huffman Trees

Thanh Hai Nguyen[1], Jean - François Maurras[1], Viet Hung Nguyen[2]

[1] Laboratoire d'Informatique Fondamentale, Université d'Aix-Marseille, Faculté des Sciences de Luminy, F-13288 Marseille Cedex 9, France.
{jean-francois.maurras, thanh.hai.nguyen}[at]lif.univ-mrs.fr

[2] Laboratoire d'Informatique de Paris 6, Université de Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cedex 05, France.
hung.nguyen[at]lip6.fr

Mots-Clés : *Polyhedral combinatorics, Polytope, Huffman code, Huffman tree, Fibonacci, Hyperplan, Facet.*

1 Introduction

In computer science and information theory, Huffman coding is an entropy encoding algorithm used for lossless data compression. It was developed by David A. Huffman [1] in 1952 and produces prefix-free, variable-length code words based on the frequency of each character. Huffman gives a greedy algorithm constructing an optimal tree in O(n log n) time (in term of number of letters). Initially, all nodes are leaf nodes, which contain the symbol itself, the weight (frequency of appearance) of the symbol. Two smallest weight items are combined into a single node with weight equal to the sum of weights of these small items. The procedure is repeated until only one node remains. The tree obtained is a labeled fully binary tree that is called *Huffman tree*. The construction of Huffman tree is described more precisely in [2].

We propose the following way to associate each Huffman tree of n leaves with a point in $\mathbb{Q}^n$:

> Considering an alphabet of n letters, $\Lambda = \{c_1, c_2, c_3, \ldots, c_n\}$. To each Huffman tree T_0 obtained from these n letters, we associate the point $x^{T_0} = (x_1^{T_0}, x_2^{T_0}, \ldots, x_n^{T_0})^t \in \mathbb{Q}^n$ where $x_i^{T_0}$ is the number of edges of the path between the root and the leaf corresponding to the letter c_i. The labels of the coordinates of x^{T_0} (read left-to-right) should be in the same order as in the original input sequence. This point associated to T^0 is called a *Huffman point.* Let PH_n called *Huffmanhedron* denote the convex hull of all the Huffman points in $\mathbb{Q}^n$.

2 Results

2.1 A partial description of PH_n

In the following theorem, we introduce a family of facet-defining inequalities for PH_n, called *Fibonacci inequalities*, whose coefficients almost form a Fibonacci sequence (the first coefficient is equal to 1 instead of 0). More precisely, we prove the following theorem.

Theorem 1. FIBONACCI INEQUALITY *Let $\alpha \in \mathbb{Z}^n$ whose components $\alpha_1 = \alpha_2 = \alpha_3 = 1$ and $\alpha_i = \alpha_{i-1} + \alpha_{i-2}$, $\forall i \geq 4$. The inequality*

$$\text{FIBO}_n \equiv \sum_{i=1}^{n} \alpha_i x_i \geq F_{n+4} - 3. \; \forall n \geq 3. \tag{1}$$

with F_{n+4} is the $(n+4)^{th}$ Fibonacci number, defines a facet of PH_n.

Given 2 binary trees of respectively $n-k$ and $k+1$ leaves where $n > k \geq 1$ integers, we can see that if we identify the root of one of them with one leaf of the other, we obtain a binary tree of n leaves. We show that under a simple condition C on coefficients (which is not detailed in this abstract), we can do the same for facet-defining inequalities, i.e. given a facet-defining inequality for PH_{n-k} and an other for PH_{k+1} satisfying condition C, we can compose them to obtain a facet-defining inequality of PH_n.

Proposition 1. *The Fibonacci inequalities satisfy condition C.*

A classical lifting method such as zero-lifting can be also applied to Fibonacci inequalities

Theorem 2. Zero-lifting Fibonacci
Let

$$\text{FIBO}_n \equiv \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_{n-1} x_{n-1} + \alpha_n x_n \geq F_{n+4} - 3,$$

be a Fibonacci facet of PH_n. The inequality:

$$\text{LIFT}_k \equiv \begin{bmatrix} \underbrace{0 \; 0 \; \cdots \; 0}_{k} & \alpha_1 = 1 & \alpha_2 = 1 & \alpha_3 = 1 & \alpha_4 & \cdots & \alpha_{n-1} & \alpha_n \end{bmatrix} \mathbf{x} \geq F_{n+4} - 2 \tag{2}$$

defines a facet of PH_{n+k} with $k \neq 2$.

2.2 A characterization of the deepest Huffman trees

Let L_n be the partial description of PH_n containing the Fibonacci inequalities and all facet-defining inequalities for PH_n obtained by the composition and zero-lifting operations described above, applied to the inequalities in L_i for all $i = 3, \ldots, n-1$.

Definition 1. *A Huffman tree A of n leaves is called a deepest tree if A has two leaves at depth $n-1$, and the corresponding Huffman vertex has the form $(n-1, n-1, n-2, n-3, \ldots, 3, 2, 1)$.*

Definition 2. *A cost function f is said to be super-fibonacci if the Huffman algorithm outputs a deepest tree as an optimal solution minimizing f.*

By means of a primal-dual type algorithm, we prove that

Theorem 3. *Optimizing a f super-fibonacci and non-negative over L_n will also give a deepest tree as an optimal solution.*

References

[1] David Albert Huffman. A method for the construction of minimum–redundancy codes *Procedings of The I.R.E*, Vol. 40 :1098 - 1101, 1951.

[2] Jon Kleinberg and Éva Tardos. *Algorithm Design*, Pearson Addison Wesley, :161 - 176, 2006.