Polyèdre des Arbres d'Huffman

Thanh Hai NGUYEN[1], Jean-François MAURRAS et Viet Hung NGUYEN[2]

1. Laboratoire d'Informatique Fondamentale,
163 av Luminy 13009 Marseille
{Thanh.Hai.Nguyen, Jean-Francois.Maurras}@lif.univ-mrs.fr

2. LIP6 – Université Pierre et Marie Curie – Paris 6,
4 place Jussieu, 75252 Paris Cedex 05.
Hung.Nguyen@lip6.fr

Mots-clefs : Polyèdre combinatoire, enveloppe convexe, arbre d'Huffman.

1 Introduction

Dans cette présentation, nous étudions l'enveloppe convexe des arbres binaires à racine sur n feuilles. Ce sont des arbres d'Huffman ayant n feuilles. Rappelons tout d'abord qu'un arbre d'Huffman est utilisé principalement pour représenter l'alphabet d'une langue écrite. On attribue aux caractères un certain poids qui est souvent sa fréquence dans la langue en question. La construction de l'arbre d'Huffman consiste tout d'abord à considérer les caractères comme les arbres d'un seul nœud. On remplace ensuite 2 arbres du plus petit poids en créant un nouvel arbre acceptant ces 2 arbres comme fils. Le poids du nouvel arbre est la somme des poids de ses deux fils. On répète ce processus jusqu'à ce qu'il reste un seul arbre. Chaque caractère donc peut-être codé par une séquence, composée de 0s et de 1s, qui représente le chemin unique de la racine vers ce caractère (0si on tourne à gauche, 1 si on tourne à droite). Le codage de Huffman est optimal dans le sens où il minimise la fonction linéaire $\sum_{i=1}^{n} p_i \times l_i$, où p_i est le poids du caractère i et l_i est la longueur en $0s$ et en $1s$ de son codage (ici l_i est variable).
On choisit la façon suivante pour associer un point de l'espace $\mathbf{Z}^n$ à un arbre d'Huffman A donné. Les variables de l'espace dans lequel nous allons décrire ce polyèdre sont indicées par les caractères. À l'arbre A donné, nous associons le point de l'espace dont la coordonnée associée au caractère c est la longueur de son codage déterminé par A. Pour les quatre caractères {a, b, c, d}, quelques soient les poids attribués, on a 13 arbres Huffman possibles dont les points correspondants sont les suivants :
$(3,3,2,1), (3,3,1,2), (3,2,3,1), (3,2,1,3), (3,1,3,2), (3,1,2,3), (2,3,3,1), (2,3,1,3), (2,1,3,3),$
$(1,3,3,2), (1,3,2,3), (1,2,3,3), (2,2,2,2)$.
Nous notons $\mathbb{EH}_n$ l'ensemble de ces points représentant tous les arbres Huffman possibles pour n caractères.
Nous nous intéressons dans cet exposé à la description de l'enveloppe convexe des points de $\mathbb{EH}_n$. Nous appelons PAH^n ce polyèdre. Nous rappelons l'algorithme de construction progressive d'un arbre Huffman en dimension $n+1$ à partir d'un arbre en dimension n, ce qui nous permet d'établir les liens entre les facettes en dimension n et celles en dimension $n+1$. Nous donnons alors quelques familles d'inégalités du polyèdre d'Huffman. Parmi celles-ci, il y a une famille de facettes dont les coefficients forment une suite de Fibonacci. Les facettes que nous avons obtenues ne constituent qu'une description partielle du polyèdre d'Huffman en dimension n. La contribution principale de ce travail repose essentiellement sur les liens que nous avons établis entre la construction des arbres et celle des facettes, en dimension $n+1$, à partir des arbres et des facettes en dimension n.

2 Polyèdre PAH^n

Théorème 2.0.1. *Les points dans* $\mathbb{EH}_n$ *sont tous les* ***sommets extrêmes*** *de* PAH^n.

2.1 Facettes de Fibonacci

Théorème 2.1.1. *Soit* α_i *tels que* $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\alpha_i = \alpha_{i-1} + \alpha_{i-2}$, $\forall i \geq 4$, *l'inégalité suivante définit une facette de* PAH^n *:*

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i \geq \alpha_{n+4} - 3. \ \forall n \geq 3$$

Par permutation des coefficients croissants, on obtient une famille de $\mathbf{\frac{n!}{6}}$ facettes de Fibonacci.

2.2 Facettes de Fibonacci 1-saut

Théorème 2.2.1. *Étant donnée la facette Fibonacci 0-saut de* PAH^n, $\forall n \geqslant 4$ *:*

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i \geq b_n = \alpha_{n+4} - 3.$$

où $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\alpha_i = \alpha_{i-1} + \alpha_{i-2}$, $\forall i \geq 4$.

– *En dimension* $n+1$ *l'inégalité :*
$$f_{f1s}(\mathbf{x}) = \sum_{i=1}^{n-1} \alpha_i \mathbf{x}_i + \alpha_{n+1}\mathbf{x}_n + \alpha_{n+1}\mathbf{x}_{n+1} \geq b_{n+1} = \alpha_{n+5} + 2\alpha_{n-1} - 3.$$
définit une facette de PAH^{n+1}.

2.3 Facettes de Fibonacci 1-saut Fibonacci

Théorème 2.3.1. *Étant donnée la facette Fibonacci 1-saut de* PAH^n, $\forall n \geqslant 5$ *:*

$$^n f(\mathbf{x}) = \sum_{i=1}^{n-2} \alpha_i \mathbf{x}_i + \alpha_n \mathbf{x}_{n-1} + \alpha_n \mathbf{x}_n \geq \alpha_{n+4} + 2\alpha_{n-2} - 3.$$

où $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\alpha_i = \alpha_{i-1} + \alpha_{i-2}$, $\forall i \geq 4$.

– *En dimension* $n+k$, $k \geq 1$, *l'inégalité :*
$$^{n+k} f(\mathbf{x}) = \sum_{i=1}^{n-2} \alpha_i \mathbf{x}_i + \alpha_n \mathbf{x}_{n-1} + \alpha_n \mathbf{x}_n + \alpha_n \left(\alpha_4 \mathbf{x}_{n+1} + \ldots + \alpha_{k+3} \mathbf{x}_{n+k} \right) \geq \alpha_{n+k+4} + (\alpha_{k+5} - 1)\alpha_{n-2} - 3.$$
définit une facette de PAH^{n+k}.

3 Polyèdre Dominant des arbres d'Huffman

3.1 Facettes 0-lifting de Fibonacci

Théorème 3.1.1. *Étant donnée la facette Fibonacci de* PAH^n *:*

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i \geq b_n = \alpha_{n+4} - 3. \ \forall n \geq 3$$

où $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\alpha_i = \alpha_{i-1} + \alpha_{i-2}$, $\forall i \geq 4$.

1 *La famille de facettes Fibonacci de* PAH^n *est celle de* $DPAH^n$

2 *La facette Fibonacci de $DPAH^n$ est aussi celle de $DPAH^{n+k}$ avec $b_{n+k} = b_n + 1 = \alpha_{n+4} - 2$ et $\alpha_{n+k} = \alpha_{n+k-1} = \ldots = \alpha_{n+1} = 0$.*

Ces facettes de la famille Fibonacci nous permettent de **caractériser les sommets** ($\langle\ n-1,\ n-1,\ n-2,\ n-3,\ \ldots,\ 3,\ 2,\ 1\ \rangle$). Ce sont les sommets générés par l'arbre le plus profond.

4 Conclusion et Perspectives

Nous avons obtenu certains types de facettes qui permettent décrire partiellement le polyèdre PAH^n, et complètement le polyèdre $PAHt$, PAH^5, PAH^6, PAH^7. Nous avons aussi trouver la génération de facettes en dimension n + 1 à partir de la dimension n. Nous espérons l'exploiter pour trouver un moyen de décrire complètement PAH^n vu que optimiser sur ce polyèdre peut se faire en temps polynômial.

Références

[1] T. Christof, *http ://www.zib.de/Optimization/Software/Porta/*, Logiciel Porta.

[2] K. Fukuda, *http ://www.ifor.math.ethz.ch/fukuda/cdd_home/*, Logiciel cdd, cdd+.

[3] R. G. Gallager, *Information theory and reliable communications,* J. Wiley, 1968, pp. 45.

[4] D. A. Huffman, *A method for the construction of minimum redundancy codes,* Proc. IRE, 40 (1951), pp. 1098 – 1101.

[5] M. M. Kapranov, *The permutoassociahedron, Mac Lane's coherence theorem and asymtotic zones for the KZ equation,* (J. of Pure and Applied Algebre 85 (1993)). pp. 119 – 142.

[6] Carl W. Lee, *The Associahedron and Triangulations of the n–gon,* (Europ. J. Combinatorics, 10 (1989)). pp. 551 – 560.

[7] J. F. Maurras, *Programmation linéaire, complexité,* (Springer, 2002).

[8] D. S. Parker et P. Ram, *The construction of Huffman codes is a submodular ('convex') optimization problem over a lattice of binary trees,* SIAM. J. Comput. 28(5), 1999, p. 2 – 5.

[9] M. Pouzet, K. Reuter, I. Rival and N. Zaguia. *A generalized permutahedron,* Algebra Universalis, 34 (1995), pp. 496 – 509.

[10] N. J. A Sloane, S. Plouffe, *The encyclopedia of integer sequences,* Academic Press, 1995.