Enumération et génération aléatoire de cactus m-aires

Michel Bousquet[1], Cedric Chauve[2], Gilles Schaeffer[3]

1. LaCIM, Université du Québec à Montréal
2. LaBRI, Université Bordeaux I
3. LORIA, CNRS

Proceedings of Colloque LaCIM 2000 (Montréal, september 2000), P. Leroux ed., vol 27 of Publications du LaCIM, p 89-100, 2000.

Enumération et génération aléatoire de cactus m-aires

Michel Bousquet[1], Cedric Chauve[2], Gilles Schaeffer[3]

[1] LaCIM, Université du Québec à Montréal,
201 Avenue Président-Kennedy, Montréal (Québec), Canada
[2] LaBRI, Université Bordeaux I
351 Cours de la Libération, 33405 Talence, France
[3] LORIA, CNRS, Campus Scientifique
615 rue du Jardin Botanique - BP 101, 54602 Villers-lès-Nancy, France

Résumé Dans cette note nous appliquons une preuve bijective de la formule d'inversion de Lagrange multidimensionnelle pour donner une explication combinatoire de deux formules d'énumération de cactus m-aires, selon la distribution des couleurs et selon la distribution des degrés. Nous en déduisons un algorithme de génération aléatoire pour ces structures.

Abstract. In this note we apply a bijective proof of the multidimensionnal Lagrange inversion formula to give a combinatorial explanation of two formulas for the enumeration of m-ary cacti, with respect to the distribution of colors and the distribution of degrees. We deduce from this explanation a uniform random generation algorithm for these structures.

1 Introduction

Un *m-cactus* est un graphe simple connexe dans lequel chaque arête appartient à exactement un cycle élémentaire. De façon équivalente, chaque composante 2-connexe d'un cactus est un cycle élémentaire, c'est-à-dire un m-gone (un polygone à m côtés). Les premières études combinatoires de ces graphes sont dues à Harary et Uhlenbeck [14] et Harary et Palmer [13], suite à un article de Husimi [15]. Un *m-cactus planaire* est un plongement d'un m-cactus dans le plan, de sorte que chaque arête soit incidente à la région infinie. Un *cactus m-aire* est un m-cactus planaire tel que les sommets autour d'un m-gone sont coloriés successivement par les couleurs $1,2,\ldots,m$ dans le sens trigonométrique. Dans cet article, nous nous intéressons aux *cactus m-aires enracinés*, c'est-à-dire aux cactus m-aires dont un polygone (ou de manière équivalente une arête joignant un sommet de couleur 1 et un sommet de couleur 2) est distingué. La Figure 1 présente un cactus ternaire dont le triangle racine (resp. l'arête racine) est rempli (resp. hachurée).

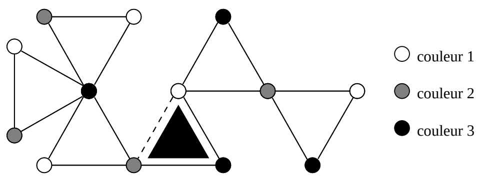


Figure 1: Un cactus ternaire enraciné.

Les cactus sont des objets combinatoires apparaissant dans divers champs des mathématiques, comme l'énumération des factorisations de permutations circulaires dans le groupe symétrique [11] ou la classification topologique des polynômes complexes [8].

On appelle *distribution des couleurs* d'un cactus C le vecteur $\mathbf{n} = (n_1, \ldots, n_m)$ tel que pour $i \in [m]$ n_i est le nombre de sommets de C de couleur i. On note $n = \sum_{i=1}^m n_i$. Par exemple, la distribution des couleurs du cactus de la Figure 1 est (5,4,4).

On vérifie aisément (cf [4, 5]) que pour une distribution des couleurs $\mathbf{n}$ donnée, il y a au moins un cactus m-aire ayant n_i sommets de couleur i, pour $i \in [m]$, si et seulement s'il existe un entier p tel que $(n-1) = p(m-1)$ (p est alors le nombre de polygones) et, pour $i \in [m]$, $1 \leq n_i \leq p$. On dit alors qu'une telle distribution $\mathbf{n}$ est *valide*.

Théorème 1 [4, 5] *Soit $\mathbf{n} = (n_1, \ldots, n_m)$ une distribution des couleurs valide. Le nombre de cactus m-aires de distribution des couleurs $\mathbf{n}$ est*

$$\frac{1}{p} \prod_{i=1}^{m} \binom{p}{n_i}, \tag{1}$$

où $n = \sum_{i=1}^m n_i$ et $p = (n-1)/(m-1)$.

On appelle *degré* d'un sommet dans un cactus le nombre de polygones auxquels ce sommet est incident. La *distribution des degrés* d'un cactus est une matrice $D = (d_{i,j})$, de taille $m \times \infty$, où $d_{i,j}$ est le nombre de sommets de couleur i et de degré j. On considère chaque ligne $(d_{i,1}, d_{i,2}, \ldots)$ comme un vecteur $\mathbf{d_i}$. On note $n_i = \sum_j d_{i,j}$ le nombre de sommets de couleur i, $n = \sum_{i=1}^m n_i$ le nombre de sommets et on dit qu'une matrice D est *valide* si et seulement s'il existe un entier p tel que $(n-1) = p(m-1)$ et, pour $i \in [m]$, $1 \leq n_i \leq p$. Par exemple, la distribution des degrés du cactus de la Figure 1 est

$$D = \begin{bmatrix} 4\,1\,0\,0 \ldots \\ 2\,2\,0\,0 \ldots \\ 3\,0\,1\,0 \ldots \end{bmatrix}.$$

Théorème 2 [11, 4] *Soit $D = (d_{i,j})_{m \times \infty}$ une distribution des degrés valide. Le nombre de cactus m-aires de distribution des degrés D est*

$$p^{m-1} \prod_{i=1}^{m} \frac{1}{n_i} \binom{n_i}{\mathbf{d_i}}, \tag{2}$$

où $n_i = \sum_{j \geq 1} d_{i,j}$, $n = \sum_{i=1}^m n_i$ et $p = (n-1)/(m-1)$.

Les preuves des Théorèmes 1 et 2 données par Goulden et Jackson [11] ou Bóna, Bousquet, Labelle et Leroux [4] utilisent la forme explicite de la formule d'inversion de séries formelles multivariées de Good-Lagrange (voir les ouvrages [10, 3] pour une description détaillée de cette formule et de ses différentes variantes).

Théorème 3 [9] *Soient $\mathbf{R}(\mathbf{x}) = (R_1(\mathbf{x}), \ldots, R_m(\mathbf{x}))$ un vecteur de séries formelles en les variables $\mathbf{x} = (x_1, \ldots, x_m)$, telles que $R_i(\mathbf{0}) \neq 0$, pour $i \in [m]$, et $\mathbf{A}(\mathbf{x})$ un vecteur de séries formelles satisfaisant*

$$A_i(x_1, \ldots, x_m) = x_i R_i(A_1, \ldots, A_m), \quad i = 1, \ldots, m. \tag{3}$$

Alors, pour toute série formelle $F(\mathbf{x})$ et tout vecteur $\mathbf{n}$ de m entiers on a

$$[\mathbf{x^n}]F(\mathbf{A}(\mathbf{x})) = [\mathbf{x^n}] \left\{ \det \left(\delta_{i,j} - \frac{x_i}{R_j(\mathbf{x})} \frac{\partial R_j(\mathbf{x})}{\partial x_i} \right)_{m \times m} F(\mathbf{x}) \prod_{i=1}^{m} R_i^{n_i}(\mathbf{x}) \right\} \tag{4}$$

où $[\mathbf{x^n}]G(\mathbf{x})$ est le coefficient de $x_1^{n_1} \cdots x_m^{n_m}$ dans $G(\mathbf{x})$ et $\delta_{i,j} = 1$ si $i = j$ et 0 sinon.

Il n'existe pas d'interprétation bijective directe connue de cette formule permettant d'expliquer combinatoirement les résultats découlant de son application. Cela est notamment dû à la présence de termes négatifs dans le déterminant du membre droit de (4). Récemment, Bender et Richmond, motivés par la recherche de formules asymptotiques pour les structures lagrangiennes [1, 2], et Goulden et Kulkarni [12] ont proposé une reformulation du Théorème 3, que l'on appelle *forme arborescente*, basée sur la notion de dérivée d'un vecteur de séries formelles selon une arborescence orientée.

Définition 1 *Soient $\mathcal{G}$ un graphe orienté étiqueté ayant ensemble de sommets $S = \{1,\ldots,d\}$ et ensemble d'arcs A, $\mathbf{x} = (x_1,\ldots,x_d)$ un vecteur de variables formelles et $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}),\ldots,f_d(\mathbf{x}))$ un vecteur de séries formelles en $\mathbf{x}$. On définit la dérivée de $\mathbf{f}(\mathbf{x})$ par rapport à $\mathcal{G}$ par*

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathcal{G}} = \prod_{j\in S}\left\{\left(\prod_{(i,j)\in A}\frac{\partial}{\partial x_i}\right) f_j(\mathbf{x})\right\}.$$

Théorème 4 [12, 1] *Soient $\mathbf{R}(\mathbf{x}) = (R_1(\mathbf{x}),\ldots,R_m(\mathbf{x}))$ un vecteur de séries formelles en les variables $\mathbf{x}$, telles que $R_i(\mathbf{0}) \neq 0$, pour $i \in [m]$, et $\mathbf{A}(\mathbf{x})$ un vecteur de séries formelles satisfaisant le système d'équations fonctionnelles (3). Alors, pour toute série formelle $F(\mathbf{x})$ et tout vecteur $\mathbf{n}$ de m entiers on a*

$$[\mathbf{x}^{\mathbf{n}}]F(\mathbf{A}(\mathbf{x})) = \left(\prod_{i=1}^{m}\frac{1}{n_i}\right)[\mathbf{x}^{\mathbf{n-1}}]\left\{\sum_{\mathcal{T}}\frac{\partial(F(\mathbf{x}),R_1^{n_1}(\mathbf{x}),\ldots,R_m^{n_m}(\mathbf{x}))}{\partial \mathcal{T}}\right\}, \qquad (5)$$

où $\mathbf{n}-\mathbf{1} = (n_1-1,\ldots,n_m-1)$, la somme étant sur toutes les arborescences $\mathcal{T}$ ayant $\{0,1,\ldots,m\}$ pour ensemble de sommets, de racine 0, et dont les arcs sont orientés vers 0.

Goulden et Kulkarni ont montré que la formule (5), qui ne fait apparaître dans son membre droit qu'une somme de termes positifs, est équivalente à (4), et ont donné une interprétation combinatoire des deux membres de (5), en termes d'endofonctions coloriées et d'arborescences coloriées. Deux preuves bijectives de cette forme arborescente de la formule de Good-Lagrange ont été proposées [12, 6], toutes deux basées sur une bijection entre arborescences m-coloriées et endofonctions m-coloriées.

Le but de cette note est d'illustrer, en prenant l'exemple des cactus m-aires, deux aspects intéressants de ces preuves. Premièrement, elles permettent d'expliquer combinatoirement des formules obtenues par application de la formule de Lagrange multidimensionnelle. Deuxièmement, de telles explications combinatoires peuvent induire des algorithmes de génération aléatoire de structures lagrangiennes (structures dont la série génératrice vérifie une équation de la forme (3)). Dans la section suivante, nous présentons rapidement les structures combinatoires utilisées dans la preuve bijective du Théorème 4, puis nous décrivons une bijection entre les cactus m-aires et une famille d'arborescences, que nous appelons *arborescences cactacées*, qui met en évidence le caractère lagrangien des cactus m-aires enracinés. Finalement, nous déduisons de ces deux constructions une bijection entre les arborescences cactacées et une famille d'endofonctions que nous appelons *endofonctions cactacées*. La Section 3 contient les preuves combinatoires des Théorèmes 1 et 2, basées sur cette bijection. Finalement la Section 4 propose un algorithme de génération aléatoire et uniforme pour les cactus m-aires de distribution de couleurs ou de degrés fixée, à n sommets, en temps et espace $O(n)$, basé sur les preuves de la Section 3.

Remarque 1. Les preuves bijectives du Théorème 4 de [12, 6] utilisant des structures coloriées étiquetées, nous considérons des cactus m-aires enracinés étiquetés : pour chaque couleur $i \in [m]$, on étiquette les n_i sommets de couleur i par les entiers de 1 à n_i (on note alors j_i le $j^{ème}$ élément de couleur i). Les cactus enracinés étant des structures asymétriques (du fait de l'enracinement), il y a $\mathbf{n}! = \prod_{i=1}^{m} n_i!$ fois plus de cactus m-aires enracinés étiquetés ayant n_i sommets de couleur i que de cactus m-aires enracinés non étiquetés ayant n_i sommets de couleur i.

2 Des cactus aux endofonctions

Dans cette section, nous nous plaçons dans le cadre de la théorie des espèces de structures développée par l'école de combinatoire montréalaise et particulièrement adaptée à l'énumération de structures arborescentes (voir [3, 16]). Le premier paragraphe est un rappel de [12, 6] (voir aussi [7]).

2.1 Arborescences, endofonctions et formule de Lagrange multidimensionnelle

Soient $\mathbf{R} = (R_1, \ldots, R_m)$ un vecteur de m espèces de structures m-coloriées et F une espèce m-coloriée, de séries génératrices exponentielles multivariées respectives $R_i(\mathbf{x})$, pour $i \in [m]$, et $F(\mathbf{x})$ ($\mathbf{x} = (x_1, \ldots, x_m)$). Les preuves bijectives du Théorème 4 font intervenir deux types de structures combinatoires : les arborescences $(F,\mathbf{R})$-enrichies et les endofonctions restreintes $(F,\mathbf{R})$-enrichies.

On note $A_{\mathbf{R},i}$ l'espèce des arborescences m-coloriées (chaque sommet est colorié par une couleur choisie parmi les couleurs $1, \ldots, m$) $\mathbf{R}$-enrichies (pour $j \in [m]$, la fibre - l'ensemble des fils - d'un sommet de couleur j est munie d'une structure d'espèce R_j) dont la racine est de couleur i. L'espèce des arborescences $(F,\mathbf{R})$-enrichies est l'espèce des F-assemblées de $(\cup_{i=1}^{m} A_{\mathbf{R},i})$-structures. L'ensemble des arborescences $(F,\mathbf{R})$-enrichies ayant n_i éléments de couleur i, pour $i \in [m]$, est noté $F(\mathbf{A_R})[\mathbf{n}]$. Dans la suite de cet article, on considère une arborescence $(F,\mathbf{R})$-enrichie sur $[\mathbf{n}]$ comme une arborescence sur $\{0\} \cup [n_1] \cup \cdots \cup [n_m] = [0,\mathbf{n}]$, de racine 0, telle que la fibre de 0 est munie d'une F-structure, les fibres des sommets de couleur i étant toujours munies d'une R_i-structure. Ces arborescences sont reliées au Théorème 4 par le fait que, si on note $A_{\mathbf{R},i}(\mathbf{x})$ la série génératrice exponentielle multivariée des $A_{\mathbf{R},i}$, on a l'interprétation combinatoire suivante du membre gauche de (5) [12, 6] :

$$|F(\mathbf{A_R})[\mathbf{n}]| = \mathbf{n}![\mathbf{x^n}]F(A_{\mathbf{R},1}(\mathbf{x}), \ldots, A_{\mathbf{R},m}(\mathbf{x})). \tag{6}$$

Une endofonction partielle $(F,\mathbf{R})$-enrichie sur $[\mathbf{n}]$ est une fonction f de $[\mathbf{n}]$ dans $[0,\mathbf{n}]$ dont chaque fibre $f^{-1}(u)$ est munie d'une R_i-structure si u est de couleur $i \in [m]$ et d'une F-structure si $u = 0$. Soit f une endofonction partielle $(F,\mathbf{R})$-enrichie sur $[\mathbf{n}]$: on appelle *graphe des couleurs* de f, noté $G(f)$, le graphe orienté (non ordonné : les voisins d'un sommet sont structurés en un ensemble non ordonné) étiqueté ayant $[0,m]$ pour ensemble de sommets et ayant, pour $i \in [m]$, un arc de i vers j ($j \in [0,m]$) si et seulement si 1_i est dans la fibre d'un élément de couleur j ($1_i \in f^{-1}(k_j)$, avec $k \in [n_j]$), et un arc de i vers 0 si 1_i est dans la fibre de 0 ($1_i \in f^{-1}(0)$). On représente habituellement une endofonction partielle f par le produit des fibres de ses éléments. Par exemple, si $m = 3$, $n_1 = 4$, $n_2 = 4$ et $n_3 = 3$, la figure suivante représente une endofonction partielle f de $[\mathbf{n}]$ (un arc de i_j vers k_l indique que $f(i_j) = k_l$) et son graphe des couleurs $G(f)$.

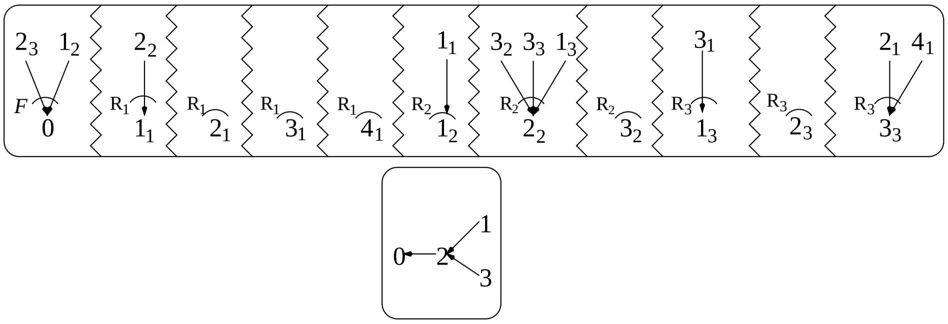


Figure 2: Une endofonction partielle restreinte.

Une endofonction partielle $(F,\mathbf{R})$-enrichie f est dite *restreinte* si et seulement si $G(f)$ est une arborescence de racine 0 dont tous les arcs sont orientés vers la racine. L'endofonction partielle précédente est donc une endofonction partielle restreinte. Ces structures sont reliées au Théorème 4 par la relation suivante : si on note $End^{\mathcal{R}}_{F,\mathbf{R}}[\mathbf{n}]$ l'ensemble des endofonctions restreintes $(F,\mathbf{R})$-enrichies sur $[\mathbf{n}]$, on a l'interprétation combinatoire suivante du membre droit de (5) [12, 6] :

$$\left|End^{\mathcal{R}}_{F,\mathbf{R}}[\mathbf{n}]\right| = (\mathbf{n}-\mathbf{1})![\mathbf{x}^{\mathbf{n}-\mathbf{1}}]\sum_{\mathcal{T}} \frac{\partial(F(\mathbf{x}),R_1^{n_1}(\mathbf{x}),\ldots,R_m^{n_m}(\mathbf{x}))}{\partial \mathcal{T}}. \tag{7}$$

On a alors le résultat suivant, reliant arborescences et endofonctions (et qui, combiné à (6) et (7), prouve bijectivement le Théorème 4).

Proposition 1. [12, 6] *Il existe une bijection Φ entre les arborescences $(F,\mathbf{R})$-enrichies sur $[\mathbf{n}]$ et les endofonctions restreintes $(F,\mathbf{R})$-enrichies sur $[\mathbf{n}]$.*

Cette bijection Φ vérifie de plus la propriété suivante (dite *de conservation de la structure des fibres*), qui s'avèrera nécessaire à la preuve du Théorème 2.

Propriété 1 *Soit A une arborescence $(F,\mathbf{R})$-enrichie sur $[\mathbf{n}]$ et u un sommet quelconque de A : la fibre de u dans A contient j sommets de couleur $k \in [m]$ si et seulement si la fibre de u dans l'endofonction restreinte $(F,\mathbf{R})$-enrichie sur $[\mathbf{n}]$ $\Phi(A)$ contient elle aussi j sommets de couleur k.*

2.2 Cactus m-aires et endofonctions

On note L l'espèce des ordres linéaires, X_i l'espèce des singletons de couleur i et, pour un vecteur $(V_1,\ldots,V_m)$ d'espèces, $\hat{V_i} = \prod_{j\neq i} V_j$ l'espèce du produit de tous les V_j excepté V_i.

On appelle *arborescences cactacées* l'espèce des arborescences $(F,\mathbf{R})$-enrichies telles que $F = X_1$ et $R_i = L(\hat{X_i})$: la fibre de 0 est réduite à un seul sommet, de couleur 1, et pour tout sommet j_i, sa fibre est structurée en une liste ordonnée de *i-blocs*, un i-bloc étant un ensemble de $(m-1)$ sommets de couleurs différentes de i et deux à deux différentes.

Proposition 2. *Il existe une bijection Ψ entre les arborescences cactacées sur $[\mathbf{n}]$ et les cactus m-aires sur $[\mathbf{n}]$.*

Il s'agit d'un résultat classique concernant les cactus m-aires enracinés, basé sur un parcours en profondeur des polygones d'un cactus, en partant du polygone racine, au cours duquel on remplace chaque polygone rencontré par un i-bloc si on est entré dans ce polygone par un sommet de couleur i. On associe donc un bloc à chaque polygone (excepté le polygone racine). La figure suivante, dans laquelle on a représenté les i-blocs par des rectangles marqués i et on a "fusionné" les arêtes joignant un bloc à son père en une seule arête, illustre cette construction.

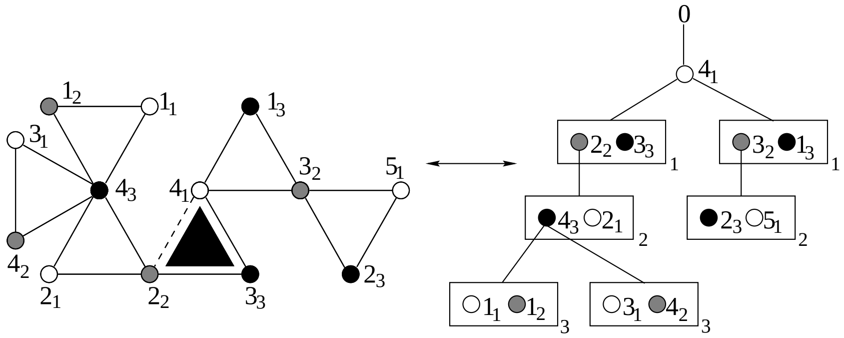


Figure 3: Un cactus m-aire et l'arborescence cactacée correspondante.

Si on appelle degré d'un élément d'une arborescence cactacée le nombre de blocs présents dans sa fibre, on vérifie alors immédiatement la propriété suivante.

Propriété 2 *Soit C un cactus m-aire. Si un sommet j_i de C différent du (resp. égal au) sommet de couleur 1 du polygone racine est incident à k polygones dans C ($k > 0$), alors j_i a degré $(k-1)$ (resp. k) dans l'arborescence cactacée $\Psi(C)$.*

Par la Proposition 1, on peut alors affirmer que les arborescences cactacées sont en bijection avec la famille des endofonctions restreintes $(F,\mathbf{R})$-enrichies telles que $F = X_1$ et $R_i = L(\hat{X}_i)$, que nous appelons *endofonctions cactacées.*

Propriété 3 *Une endofonction cactacée f sur $[\mathbf{n}]$ satisfait notamment les contraintes suivantes :*

1. *la fibre $f^{-1}(0)$ de 0 est constituée du seul élément 1_1 (par la contrainte sur le graphe des couleurs d'une endofonction restreinte),*
2. *le graphe des couleurs est une arborescence de racine 0 ayant pour seul fils le sommet 1_1 (conséquence de la contrainte précédente) et dont tous les arcs sont orientés vers 0.*

3 Enumération de cactus m-aires

Dans cette section, nous proposons une explication combinatoire des Théorèmes 1 et 2 basée sur la bijection entre cactus m-aires et endofonctions cactacées. Dans un premier temps, nous faisons un rappel sur une construction classique reliant arborescences et mots, le codage de Prüfer.

3.1 Codage de Prüfer

Il s'agit d'une bijection entre les arborescences sur $[n]$ et les mots de longueur $(n-1)$ sur l'alphabet $[n]$ [17]. Soit m un mot de $(n-1)$ lettres sur l'alphabet $[n]$. On construit une arborescence A_m à partir de m de la manière suivante :

– au départ A_m est constituée des seuls sommets $1,\ldots,m$, sans aucune arête;

– soit F l'ensemble des lettres de $[n]$ non présentes dans m;
– pour i allant de 1 à $(n-1)$, soit $u = m(i)$ la $i^{ème}$ lettre de m et v le plus grand élément de F : ajouter un arc de u vers v à A_m, supprimer v de F et si u n'apparaît plus dans m, ajouter u à F.

On peut alors faire les remarques suivantes.

– Tous les arcs sont orientés vers la racine, la dernière lettre de m étant la racine de A_m.
– Cet algorithme induit un étiquetage sur $[n-1]$ des arêtes de A_m : si l'arête allant de u vers v est créée lors de la $i^{ème}$ étape, on l'étiquette par i.

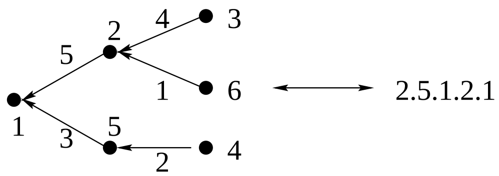


Figure 4: Arborescence et mot associés suivant le codage de Prüfer.

3.2 Enumération de cactus selon la distribution des couleurs

Compte-tenu de la bijection entre cactus m-aires et arborescences cactacées et de la Proposition 1, il suffit d'énumérer les endofonctions m-cactacées ayant n_i eléments de couleur $i \in [m]$. Nous réalisons l'énumération de ces endofonctions en trois étapes : répartition des blocs, placement des éléments minimaux (éléments 1_i) et placement des éléments non minimaux. On rappelle qu'une endofonction cactacée ayant n_i sommets de couleur i comporte exactement

$$p = \frac{(\sum_{i=1}^{m} n_i) - 1}{m-1}$$

blocs, et que la fibre de l'élément 0 est constituée du seul élément 1_1. Un i-bloc étant constitué de $m-1$ éléments de couleurs deux à deux distinctes et autres que c_i, on en déduit la propriété suivante.

$$\text{Le nombre de 1-blocs est } p - n_1 + 1 \text{ et le nombre de } i\text{-blocs est } p - n_i, \text{ pour } i > 1. \tag{8}$$

1. Répartition des blocs. Le nombre de partitions d'un ensemble de k éléments indistinguables en l parts (ordonnées entre elles) éventuellement vides est

$$\binom{k+l-1}{l-1}. \tag{9}$$

On appelle *partition des blocs* d'une endofonction cactacée le vecteur $(k_{1,1}, \ldots, k_{1,n_1}, \ldots, k_{m,1}, \ldots, k_{m,n_m})$ de n entiers positifs ou nuls tel que $k_{i,j}$ est le nombre de i-blocs présents dans la fibre de l'élément j_i. On déduit de (8) et (9) que le nombre de partitions des blocs possibles pour les endofonctions cactacées sur $[\mathbf{n}]$ est

$$\binom{p}{n_1 - 1} \prod_{i=2}^{m} \binom{p-1}{n_i - 1} = \frac{1}{(p - n_1 + 1)p^{m-1}} \prod_{i=1}^{m} n_i \binom{p}{n_i}. \tag{10}$$

Maintenant, pour une partition des blocs fixée, nous regardons le nombre de répartitions des éléments dans les blocs satisfaisant à la contrainte 2 de la Propriété 3 sur le graphe des couleurs d'une endofonction cactacée.

2. Placement des éléments minimaux. Rappelons que l'élément 1_1 étant l'unique élément présent dans la fibre de 0, il n'appartient à aucun bloc. On considère alors les éléments minimaux 1_i, $i = 2, \ldots, m$, qui sont les seuls à intervenir dans la contrainte sur le graphe des couleurs.

$$\text{Il y a } p^{m-2}(p - n_1 + 1) \text{ façons de placer les éléments } 1_i,\ i = 2, \ldots, m. \tag{11}$$

En effet, supposons les p blocs numérotés arbitrairement de 1 à p et considérons un mot w de longueur $(m-1)$ sur l'alphabet $[p]$: à chaque lettre de w correspond un des p blocs de l'endofonction. A partir de w, on construit le mot w' en remplaçant chaque numéro de bloc par son type : si le bloc a dans w est un i-bloc, on remplace a par i dans w'. Ce mot w' est alors un mot de longueur $(m-1)$ sur l'alphabet $[m]$. En appliquant le codage de Prüfer au mot w', on obtient une arborescence orientée sur $[m]$ dont les arcs sont étiquetés sur $[m-1]$ et orientés vers la racine. On peut alors associer à cette arborescence le placement suivant des éléments minimaux : si on a un arc de i vers j, étiqueté k, on place 1_i dans le bloc associé à la $k^{ième}$ lettre de w (il est clair que ce bloc ne peut être un i-bloc). Par exemple, pour $m = 4$ et $\mathbf{n} = (3{,}3{,}3{,}4)$, on considère la partition des blocs donnée à la Figure 5 (les 4 blocs sont numérotés de 1 à 4 de gauche à droite). Si $w = 2.4.1$, on a alors $w' = 1.3.1$ et le codage de Prüfer appliqué à w' donne l'arborescence apparaissant au bas de la Figure 5.

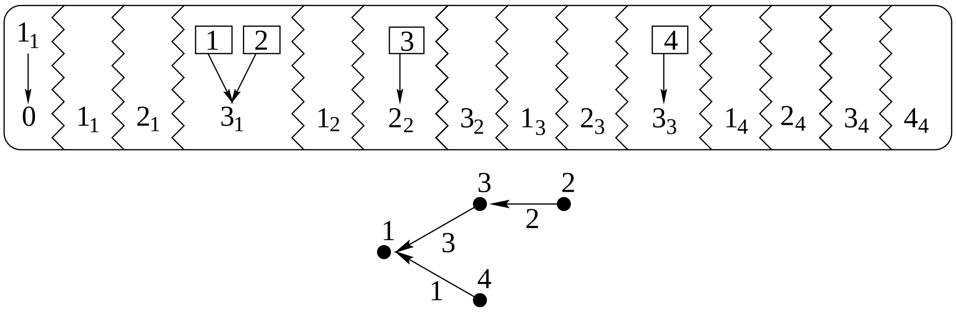


Figure 5: Arborescence correspondant au mot $w = 2.4.1$

On place donc 1_2 dans le bloc 4, 1_3 dans le bloc 1 et 1_4 dans le bloc 2. Pour que la contrainte sur le graphe des couleurs soit vérifiée par ce placement des éléments minimaux, il faut et il suffit que sa racine soit le sommet 1, ce qui est équivalent à imposer que la dernière lettre de w' est la lettre 1 (on obtient alors une arborescence de racine 1 que l'on relie à 0 par un arc de joignant 1 et 0) et donc que le bloc associé à la dernière lettre de w est un 1-bloc. On a donc p choix pour chacune des $(m-2)$ premières lettres de w et $(p - n_1 + 1)$ pour sa dernière lettre, ce qui prouve (11).

3. Placement des éléments non minimaux. Une fois que les éléments minimaux sont placés, la répartition des éléments restants dans les blocs ne pose aucun problème. Par (10), (11), la Remarque 1, et le fait qu'il y a $(\mathbf{n} - \mathbf{1})!$ façons de répartir les éléments non minimaux, et après simplification, on obtient la formule (1).

3.3 Enumération de cactus selon la distribution des degrés

La preuve du Théorème 2 est basée sur le même principe que la preuve précédente. Si on appelle degré d'un élément d'une endofonction cactacée le nombre de blocs présents dans sa fibre, par la bijection entre cactus m-aires et arborescences cactacées, la Proposition 1, la Propriété 1 de conservation de la structure des fibres et la Propriété 2, l'ensemble des cactus m-aires de partition des degrés D et de racine de degré l (le sommet de couleur 1 du polygone pointé est incident à l polygones) est en bijection avec l'ensemble des endofonctions cactacées ayant

(a) $d_{i,j}$ éléments de couleur i et de degré $j-1$ si $i>1$ ou ($i=1$, $j \neq l$ et $j \neq (l+1)$),

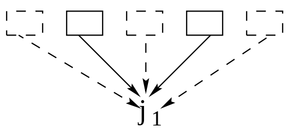
(b) $d_{1,l}-1$ éléments de couleur 1 et de degré $l-1$ et

(c) $d_{1,l+1}+1$ éléments de couleur 1 et de degré l.

1.a. Répartition des i-blocs pour $i>1$. Pour chaque couleur $i>1$, on a donc $d_{i,j}$ éléments de degré $(j-1)$ et le nombre de façons de répartir ces i-blocs dans les fibres des n_i sommets de couleur i est

$$\binom{n_i}{d_{i,1}, d_{i,2}, \ldots}. \tag{12}$$

1.b. Répartition et pointage des 1-blocs. En ce qui concerne la couleur 1, après avoir réparti $d_{1,j}$ 1-blocs, pour tous les $j>0$, il découle de (b) et (c) qu'il reste encore un 1-bloc à placer dans la fibre d'un élément de couleur 1. Comme le montre la Figure 6 (dans laquelle les blocs en pointillés représentent les positions possibles d'insertion d'un bloc supplémentaire), on a exactement $k+1$ façons de placer ce 1-bloc dans la fibre d'un élément de degré k en le distinguant des autres blocs déjà placés,



et on donc a exactement

$$n_1 + \sum_{j \geq 1} (j-1) d_{1,j} - 1 = n_1 + (p - n_1 + 1) - 1 = p \tag{13}$$

façons de placer un bloc distingué dans la fibre d'un élément de couleur 1. On en déduit qu'on a

$$p \prod_{i=1}^{m} \binom{n_i}{d_{i,1}, d_{i,2}, \ldots}. \tag{14}$$

façons de placer les p blocs tout en distinguant un 1-bloc.

2. Placement des éléments minimaux. De la même manière que pour la preuve précédente, on remarque que le nombre de placements des éléments minimaux 1_i ($i>1$) respectant la contrainte sur le graphe des couleurs est donné, pour toute numérotation arbitraire des blocs sur $[p]$, par le nombre de mots de longueur $(m-1)$ sur l'alphabet $[p]$ tels que la dernière lettre est le numéro du 1-bloc distinguée. Il y en a exactement

$$p^{m-2}. \tag{15}$$

3. Placement des éléments non minimaux. Finalement, comme précédemment, le placement des éléments non minimaux ne pose aucun problème et on en a $(\mathbf{n}-\mathbf{1})!$, ce qui, combiné à (14), (15) et à la Remarque 1, et après simplification, donne la formule (2).

4 Génération aléatoire de cactus m-aires

Les preuves présentées dans la section précédente, outre le fait qu'elles donnent la première explication combinatoire, à notre connaissance, des formules d'énumération de cactus m-aires, permettent aussi la mise au point d'algorithmes efficaces de génération aléatoire uniforme de cactus m-aires, selon la distribution des couleurs ou des degrés.

Pour énumérer les endofonctions cactacées selon la partition des couleurs ou des degrés, nous avons suivi un cheminement en trois étapes qui sous-tendent les grandes lignes d'un algorithme de génération aléatoire de cactus m-aires. Pour engendrer aléatoirement et uniformément une telle structure selon la distribution des couleurs (resp. des degrés), on peut alors appliquer l'Algorithme 1 suivant.

Algorithme 1: Génération aléatoire de cactus m-aire

Données : distribution $\mathbf{n}$ (resp. D) des couleurs (resp. des degrés)
Résultat : cactus m-aire
début
  génération d'une endofonction cactacée ayant distribution des couleurs $\mathbf{n}$ (resp. des degrés D);
    1. partition des blocs;
    2. placement des éléments minimaux;
    3. placement des éléments non minimaux;
  application de la bijection Φ entre endofonctions cactacées et arborescences cactacées;
  application de la bijection Ψ entre arborescences cactacées et cactus m-aires;
fin

Lemme 1. *L'algorithme 1 permet d'engendrer aléatoirement et uniformément un cactus m-aire enraciné à n sommets selon la distribution des couleurs ou des degrés, en temps et espace $O(n)$.*

Preuve. D'après les preuves des Théorèmes 1 et 2, il est clair que cet algorithme engendre un cactus m-aire. Il reste à vérifier que cette génération est uniforme et que les complexités en temps et en espace sont linéaires.

1. Génération uniforme. Le fait que cet algorithme engendre uniformément les cactus m-aires, c'est-à-dire que tous les cactus de distribution des couleurs $\mathbf{n}$ (resp. des degrés D) ont la même probabilité d'être engendré, découle du fait que lors de l'étape de génération aléatoire des endofonctions cactacées :

- toutes les partitions des blocs ont la même probabilité d'être engendrées,
- le nombre de façons de placer les éléments minimaux ne dépend pas de la partition des blocs (et tous les placements sont donc équiprobables),
- le nombre de façons de placer les éléments non minimaux ne dépend ni de la partition des blocs, ni du placement des éléments minimaux (et tous les placements sont donc équiprobables).

2. Complexité linéaire en temps et en espace. Dans les deux cas, la phase de répartition des blocs consiste à partitionner l'ensemble des i-blocs en n_i parts, ce qui se fait en temps et espace linéaires. Le placement des éléments minimaux consiste à choisir un mot aléatoirement sur un alphabet fixé, puis à appliquer le codage de Prüfer, ce qui là aussi ce fait en temps et espace linéaires. Le placement des éléments non minimaux consiste à choisir aléatoirement une permutation sur les éléments non minimaux, ce qui là aussi se fait en temps et espace linéaires. Finalement, les bijections Φ et Ψ pouvant être implantées en temps et espace $O(n)$, on obtient bien une complexité linéaire. □

Remarque 2. Récemment, Dimitri Zvonkine [18] a proposé une autre preuve combinatoire du Théorème 2.

Remerciements. Nous remercions Pierre Leroux pour de très utiles conversations.

Références

1. Bender (E. A.) et Richmond (L. B.). – A multivariate Lagrange inversion formula for asymptotic calculations. *Electron. J. Combin.*, vol. 5, n° 1, 1998, pp. Research Paper 33, 4 pp.
2. Bender (E. A.) et Richmond (L. B.). – Multivariate asymptotics for products of large powers with applications to Lagrange inversion. *Electron. J. Combin.*, vol. 6, n° 1, 1999, pp. Research Paper 8, 21 pp.
3. Bergeron (F.), Labelle (G.) et Leroux (P.). – *Combinatorial species and tree-like structures.* – Cambridge, Cambridge University Press, 1998.
4. Bóna (M.), Bousquet (M.), Labelle (G.) et Leroux (P.). – Enumeration of m-ary cacti. *Adv. in Appl. Math.*, vol. 24, 2000, pp. 22–56.
5. Bousquet (M.). – *Théorie des espèces et applications au dénombrement de cartes et de cactus planaires.* – Thèse de Doctorat, LaCIM, Université du Québec à Montréal, 1999.
6. Bousquet (M.), Chauve (C.), Labelle (G.) et Leroux (P.). – A bijective proof of the arborescent form of the multivariable Lagrange's inversion formula. *In : Colloquium on Mathematics and Computer Science*, Trends in Mathematics. – Birkhauser, 2000. (à paraître).
7. Chauve (C.). – *Structures arborescentes : problèmes algorithmiques et combinatoires.* – Thèse de Doctorat, LaBRI, Université Bordeaux I, 2000. (en préparation).
8. El Marraki (M.), Hanusse (N.), Zipperer (J.) et Zvonkin (A.). – Cacti, braids and complex polynomials. *Sém. Lothar. Combin.*, vol. 37, 1996, pp. Art. B37b, 36 pp.
9. Good (I. J.). – Generalizations to several variables of Lagrange's expansion, with applications to stochastic processes. *Proc. Cambridge Philos. Soc.*, vol. 56, 1960, pp. 367–380.
10. Goulden (I. P.) et Jackson (D. M.). – *Combinatorial enumeration.* – New York, John Wiley & Sons Inc., 1983.
11. Goulden (I. P.) et Jackson (D. M.). – The combinatorial relationship between trees, cacti and certain connection coefficients for the symmetric group. *European J. Combin.*, vol. 13, n° 5, 1992, pp. 357–365.
12. Goulden (I. P.) et Kulkarni (D. M.). – Multivariable Lagrange inversion, Gessel-Viennot cancellation, and the matrix tree theorem. *J. Combin. Theory Ser. A*, vol. 80, n° 2, 1997, pp. 295–308.
13. Harary (F.) et Palmer (E. M.). – *Graphical enumeration.* – New York, Academic Press, 1973.
14. Harary (F.) et Uhlenbeck (G. E.). – On the number of Husimi trees. *Proc. Nat. Acad. Sci. U.S.A.*, vol. 39, 1953, pp. 315–322.
15. Husimi (K.). – Note on Mayer's theory of cluster integrals. *J. Chem. Phys.*, vol. 18, 1950, pp. 682–684.
16. Labelle (G.). – Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange. *Adv. in Math.*, vol. 42, n° 3, 1981, pp. 217–247.
17. Moon (J. W.). – *Counting labelled trees.* – Canadian Mathematical Congress, Montréal, Québec, 1970. Canadian Mathematical Monographs, No. 1.
18. Zvonkine (D.). Communication personnelle.