Enumération et génération aléatoire de chemins

A. Denise *

LaBRI, Université Bordeaux I
URA CNRS 1304

Résumé

Nous donnons une nouvelle preuve bijective pour l'énumération de certains chemins dans le plan et dans l'espace à trois dimensions. On déduit de ce travail des algorithmes linéaires de génération aléatoire et uniforme des chemins étudiés.

Abstract

We give a new bijective proof for the enumeration of some paths in bidimensional and tridimensional lattices. This work provides linear-time algorithms for generating these paths uniformly at random.

1 Introduction

Nous étudions dans cet article certaines classes de chemins du plan ou de l'espace à trois dimensions. Dans le plan notamment, il s'agit de chemins composés de quatre types de pas : Nord, Sud, Est et Ouest, et particularisés par des contraintes supplémentaires.

Il existe de nombreux travaux concernant l'énumération de chemins du plan ou de l'espace, parfois en liaison avec des problèmes liés à la physique statistique (voir par exemple [4, 9]). Certains auteurs se sont intéressés à des preuves bijectives de ces résultats. Ainsi, Cori, Dulucq et Viennot [2] ont résolu le difficile problème de l'énumération des chemins fermés et contenus dans le premier quadrant du plan par des méthodes bijectives. Ces chemins, liés aux permutations de Baxter, sont comptés par le produit de deux nombres de Catalan consécutifs.

Un travail comparable a été réalisé concernant les chemins dans l'octant ($x \geq y \geq 0$) et dont l'extrémité terminale se situe sur l'axe des abscisses. Gouyou-Beauchamps [6], en les mettant en correspondance avec des couples de mots de Dyck (ou mots de parenthèses), a prouvé que le nombre de tels chemins de longueur fixée est un produit de deux nombres de Catalan.

Plus récemment, Guy, Krattenthaler et Sagan [7] ont étudié un grand nombre de types de chemins du plan, et donné des preuves bijectives pour l'énumération de la plupart d'entre eux.

Le présent article a pour but principal l'utilisation de preuves bijectives d'énumération de chemins, pour la conception d'algorithmes efficaces de génération aléatoire de ces objets. Nous étudions principalement une variété de chemins dans le plan, les *chemins du premier quadrant*, ainsi que leur généralisation dans l'espace à trois dimensions. Guy, Krattenthaler et Sagan [7] donnent une preuve bijective pour leur énumération, en tant que cas particulier d'une classe plus générale de chemins. La preuve que nous donnons ici est différente, et, contrairement à la précédente, permet de concevoir un algorithme linéaire de génération aléatoire uniforme de chemins du premier quadrant. La

*E-mail : denise@labri.u-bordeaux.fr

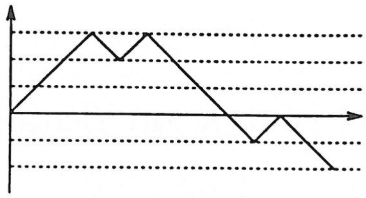

Figure 1: Un chemin de hauteur terminale -2, associé au mot $xxx\bar{x}x\bar{x}\bar{x}\bar{x}\bar{x}x\bar{x}\bar{x}$.

preuve et l'algorithme se généralisent aisément en dimension trois, pour l'énumération des chemins du premier octant en réseau cubique centré.

La section qui suit est consacrée à quelques définitions et propriétés que nous utiliserons dans la démonstration de notre résultat principal. Nous y rappelons notamment la notion de *décomposition de Catalan* d'un mot, due à Chottin et Cori [1], et définissons les chemins du premier quadrant. En section 3, nous présentons la preuve de la formule d'énumération des chemins du premier quadrant, ainsi que leur généralisation en dimension trois. La section 4 est consacrée à l'application de cette preuve à la génération aléatoire et uniforme de chemins.

2 Définitions et propriétés préliminaires

2.1 Langages et décomposition de Catalan

Soit l'alphabet $X = \{x, \bar{x}\}$. A tout mot de X^*, on peut associer un chemin du plan de la façon suivante : à une lettre x correspond un pas Nord-Est, à une lettre $\bar{x}$ correspond un pas Sud-Est (figure 1). La *hauteur terminale* d'un tel chemin (ou du mot associé) est la différence entre le nombre de pas Nord-Est et le nombre de pas Sud-Est du chemin. Si x est un mot de X^*, nous noterons $\delta(w)$ sa hauteur terminale.

Parmi les langages sur l'alphabet X, certains sont bien connus. Ainsi, le *langage de Dyck* D_x est l'ensemble des mots de X^* de hauteur terminale nulle, et dont tous les facteurs gauches sont de hauteur terminale positive ou nulle :

$$D_x = \{w \in X^* \ : \ \delta(w) = 0 \text{ et } w = uv \Rightarrow \delta(u) \geq 0\}.$$

Le nombre de mots de Dyck de longueur n est le célèbre nombre de Catalan, $C_n = \frac{1}{n+1}\binom{2n}{n}$. Le langage F_x des *facteurs gauches de Dyck* est l'ensemble des mots satisfaisant la seconde des deux propriétés précédentes :

$$F_x = \{w \in X^* \ : \ w = uv \Rightarrow \delta(u) \geq 0\}.$$

On connaît bien le nombre de facteurs gauches de Dyck de longueur n, égal à $\binom{n}{\lfloor n/2 \rfloor}$.

Remarquons que les deux définitions précédentes s'interprètent particulièrement bien en termes de chemins composés de pas Nord-Est et Sud-Est : aux mots de F_x correspondent les chemins qui ne traversent jamais l'axe des abscisses. Les chemins associés aux mots de D_x présentent la particularité supplémentaire de se terminer sur l'axe des abscisses.

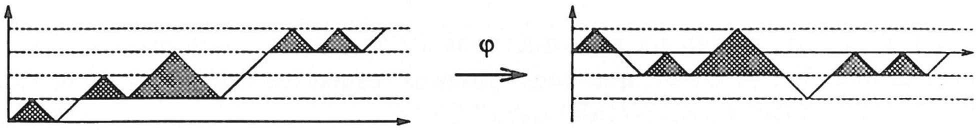

Figure 2: Décomposition de Catalan et bijection φ sur un chemin.

Nous utiliserons dans la preuve de notre résultat principal le théorème suivant, dû à Chottin et Cori [1] :

Théorème 2.1 *Il existe une bijection φ entre les facteurs gauches de Dyck de longueur n et les mots de X^* de longueur n*

- *et de hauteur terminale 0 si n est pair,*
- *et de hauteur terminale 1 si n est impair.*

La preuve de ce résultat repose sur la *décomposition de Catalan* des mots de X^*, ainsi définie par Chottin et Cori : tout mot $w \in X^*$ se décompose de façon unique en

$$w \;=\; u_0\bar{x}u_1\bar{x}u_2\ldots\bar{x}u_ixu_{i+1}\ldots xu_{j-1}xu_j$$

où u_0, u_1, $\ldots u_j$ sont des mots de Dyck (éventuellement vides). Soit $w \in F_x$ un facteur gauche de Dyck de longueur paire. Remarquons que sa décomposition de Catalan s'écrit nécessairement

$$w = u_0xu_1xu_2x\ldots xu_{2k-1}xu_{2k}$$

où k est un entier positif ou nul. On peut construit alors le mot

$$\varphi(w) = u_0\bar{x}u_1\bar{x}u_2x\ldots\bar{x}u_kxu_{k+1}x\ldots xu_{2k-1}xu_{2k}$$

qui est de hauteur terminale nulle, en transformant en $\bar{x}$ les k premières lettres x de la décomposition. On montre aisément que φ est une bijection entre les facteurs gauches de Dyck de longueur paire et les mots de X^* de même longueur et de hauteur terminale nulle, parfois appelés *mots du Grand Dyck*. On définit de même la bijection entre les facteurs gauches de Dyck de longueur impaire et les mots de X^* de même longueur et de hauteur terminale 1. Tout facteur gauche de Dyck de longueur impaire présente la décomposition de Catalan suivante :

$$w = u_0xu_1xu2x\ldots xu_{2k-1}xu_{2k}xu_{2k+1}$$

où k est un entier positif ou nul. Alors

$$\varphi(w) = u_0\bar{x}u_1\bar{x}u2x\ldots\bar{x}u_kxu_{k+1}x\ldots xu_{2k-1}xu_{2k}xu_{2k+1}.$$

La figure 2 présente la décomposition de Catalan d'un chemin facteur gauche de Dyck et son image par la bijection φ.

2.2 Chemins NSEO, chemins du premier quadrant et mélanges de mots

Un chemin NSEO est un chemin du plan débutant à l'origine et comportant quatre types de pas unitaires : Nord, Sud, Est et Ouest, que nous noterons respectivement **N**, **S**, **E** et **O**. Nous serons amenés à utiliser le résultat suivant, établi dans [7] :

Théorème 2.2 *Il existe une bijection entre les chemins NSEO de longueur n allant de l'origine au point (c, d), et les couples de mots de X^* de longueur n et de hauteurs terminales respectives $c + d$ et $c - d$.*

La preuve est basée sur la correspondance suivante :

$$\begin{array}{lcl} \mathbf{N} & \to & (x, x) \\ \mathbf{S} & \to & (\bar{x}, \bar{x}) \\ \mathbf{E} & \to & (x, \bar{x}) \\ \mathbf{O} & \to & (\bar{x}, x). \end{array}$$

Un *chemin du premier quadrant* est un chemin NSEO entièrement situé dans la premier quadrant du plan. En d'autres termes, un tel chemin ne traverse ni l'axe des abscisses ni l'axe des ordonnées.

Les chemins NSEO sont naturellement codés par des mots *mélanges* ("shuffles") de mots de X^* et de $Y^* = \{y, \bar{y}\}$. Rappelons à ce propos les définitions du mélange de deux mots ou de de langages, telle que données dans [8]. Soit A un alphabet, u et v deux mots de A^*. Le *mélange* de u et v est le langage défini par

$$\begin{aligned} u \sqcup\!\sqcup v \; = \; \{ w \; : \; & w = u_1 v_1 u_2 v_2 \dots u_n v_n, \\ & n \geq 0, \; u_i, v_i \in A^*, \; u = u_1 u_2 \dots u_n, \; v = v_1 v_2 \dots v_n \}. \end{aligned}$$

Si L_1 et L_2 sont deux langages, alors leur mélange, noté $L_1 \sqcup\!\sqcup L_2$, est l'union des mélanges de chacun des mots de L_1 avec chacun des mots de L_2.

Le codage d'un chemin NSEO est ainsi défini : aux pas **E**, **O**, **N**, **S** on associe respectivement les lettres x, $\bar{x}$, y, $\bar{y}$. Les chemins du premier quadrant satisfont à la propriété suivante, qui se déduit immédiatement de la définition du codage.

Propriété 2.3 *Les chemins du premier quadrant de longueur n sont en bijection avec les mots de longueur n du langage $F_x \sqcup\!\sqcup F_y$.*

Dans cette propriété, F_y représente le langage des facteurs gauches de Dyck sur l'alphabet Y, défini de façon similaire à F_x (section 2.1).

3 Enumération de chemins

3.1 Chemins du premier quadrant du plan

Théorème 3.1 *Il existe une bijection entre les chemins du premier quadrant de longueur n, et les couples de facteurs gauches de Dyck de longueurs respectives n et $n + 1$.*

Preuve. Considérons un chemin du premier quadrant de longueur n, et w le mot de $F_x \sqcup\!\sqcup F_y$ qui lui est associé (propriété 2.3). Soient u et v les uniques mots, respectivement de F_x et F_y, tels que $w \in u \sqcup\!\sqcup v$. Soient $u' \in X^*$ et $v' \in Y^*$ les mots associés respectivement à u et v par la bijection φ

du théorème 2.1 (on étend naturellement la notion de hauteur terminale et la bijection φ aux mots de Y^*). Considérons maintenant l'unique mot $w' \in u' \sqcup\!\sqcup v'$ tel que les lettres de u' apparaissent dans w' aux mêmes occurences que les lettres de u dans w. Nous examinerons plusieurs cas, selon que l'entier n est pair ou non.

Si n est pair, supposons tout d'abord que u et v sont de longueur paire. Alors les mots u' et v' sont tous deux de hauteur nulle (théorème 2.1). Le mot w' code alors un chemin NSEO finissant à l'origine. D'après le théorème 2.2, on peut associer à w' un unique couple (f, g) de mots de X^*, tous deux de longueur n et de hauteur terminale nulle. Soit g' le mot obtenu en ajoutant une lettre x au mot g. Alors (f, g') est un couple de mots associé bijectivement à w : f est de longueur n et de hauteur terminale nulle, et g' est de longueur $n + 1$, de hauteur terminale 1, et finit par une lettre x.

Supposons maintenant que u et v sont de longueur impaire. Alors la hauteur terminale de u' et v' est égale à 1 d'après le théorème 2.1. Au mot w' correspond alors un chemin NSEO dont l'extrémité terminale se trouve au point de coordonnées $(1, 1)$. Le théorème 2.2 nous permet alors d'associer à w' un couple (f, g) de mots de X^* de longueur n et de hauteurs terminales respectives 0 et 2. Soit g' le mot obtenu en ajoutant une lettre $\bar{x}$ au mot g. Alors (f, g') est un couple de mots associé bijectivement à w : f est de longueur n et de hauteur terminale nulle, et g' est de longueur $n + 1$, de hauteur terminale 1, et finit par une lettre $\bar{x}$.

On déduit des deux cas qui précèdent qu'à tout chemin du premier quadrant de longueur n paire est associé bijectivement un couple de mots de X^* de longueurs respectives n et $n + 1$ et de hauteurs terminales respectives 0 et 1. Le théorème 2.1 nous permet alors d'affirmer qu'il existe une bijection entre les chemins du premier quadrant de longueur n paire et les couples de mots facteurs gauches de Dyck de longueurs respectives n et $n + 1$.

Si n est impair, alors les longueurs de u et v sont de parités différentes. Considérons le cas où u est de longueur paire, tandis que v est de longueur impaire. Alors la hauteur terminale de u' est nulle, tandis que celle de v' est égale à 1. En conséquence, le chemin correspondant au mot w' se termine au point de coordonnées $(0, 1)$. Donc le couple de mots (f, g) de mots associé à w' (théorème 2.2) est tel que la hauteur terminale de f est égale à 1, et celle de g est -1. En ajoutant une lettre x à g, on obtient un mot g' de longueur $n + 1$ et de hauteur terminale nulle.

Le cas où u est de longueur impaire et v de longueur paire se traite de façon similaire. On montre alors qu'à w est associé un couple (f, g'), de façon que f est un mot de longueur n et de hauteur terminale 1, tandis que g' est de longueur $n + 1$, de hauteur terminale nulle et se termine par une lettre $\bar{x}$.

Comme précédemment, on déduit de ces deux derniers cas et du théorème 2.1 qu'il existe une bijection entre les chemins du premier quadrant de longueur n impaire et les couples de mots facteurs gauches de Dyck de longueurs respectives n et $n + 1$. □

On déduit bien évidemment du théorème précédent que le nombre de chemins du premier quadrant de longueur n est $\binom{n}{\lfloor n/2 \rfloor}\binom{n+1}{\lfloor (n+1)/2 \rfloor}$.

3.2 Chemins du premier octant de l'espace

Le résultat précédent se généralise facilement dans l'espace à trois dimensions, en réseau cubique centré. Ce réseau autorise huit types de pas, comme indiqué en figure 3. Leur projection sur un plan horizontal comporte les quatre type de pas **N**, **S**, **E**, **O**. Dans l'espace, à chacun de ces pas est associée une direction verticale, montante ou descendante. Les chemins sur ce réseau sont codés par des mots sur l'alphabet $Z = \{x_d, x_m, \bar{x}_d, \bar{x}_m, y_d, y_m, \bar{y}_d, \bar{y}_m\}$. Les lettres x, $\bar{x}$, y et $\bar{y}$ codent respectivement les pas **E**, **O**, **N** et **S**. Chacune de ces lettres est indicée par d ou m, selon que le

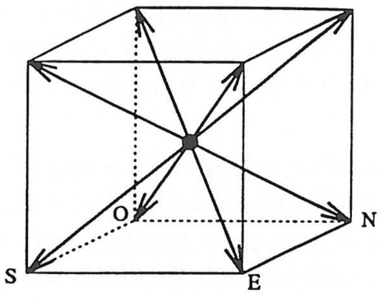

Figure 3: Les huit pas autorisés en réseau cubique centré.

pas est descendant ou montant. Un chemin du premier octant est un chemin dont tous les pas sont contenus dans l'octant $(x \geq 0, y \geq 0, z \geq 0)$. On montre alors aisément la proposition suivante.

Proposition 3.2 *Les chemins du premier octant de longueur n sont en bijection avec les triplés de mots facteurs gauches de Dyck, de longueurs respectives n, n et $n+1$.*

Il en découle que le nombre de chemins du premier octant de longueur n est $\binom{n}{\lfloor n/2 \rfloor}^2 \binom{n+1}{\lfloor (n+1)/2 \rfloor}$.

4 Application à la génération aléatoire

Etant donné un ensemble E d'objets combinatoires et un entier n, le problème de la génération aléatoire et uniforme est le suivant : il s'agit de tirer un objet de taille n de façon équiprobable parmi tous les objets de taille n dans E. Il existe principalement deux méthodes générales pour résoudre ce type de problème. La méthode récursive [5, 10] peut être utilisée lorsque l'on connaît une interprétation constructive d'une relation de récurrence permettant de compter les objets considérés. Dans d'autres cas, on utilise parfois une méthode de rejet, consistant à tirer des objets dans un surensemble de l'ensemble considéré, jusqu'à obtention d'un objet convenable. Le point faible de ce dernier type de méthode réside dans le fait que sa complexité n'est pas bornée; cependant, on obtient parfois une très bonne complexité *en moyenne*. La méthode récursive, par contre, donne des algorithmes déterministes, donc de complexité bornée.

Dans le cas des chemins du premier quadrant, il existe une méthode de rejet pour engendrer des chemins aléatoires en temps moyen linéaire [3]. Cependant, on ne connaît pas de méthode récursive efficace. La preuve du théorème 3.1 fournit un algorithme déterministe de complexité linéaire pour résoudre le problème. Supposons par exemple que n est pair. Il s'agit tout d'abord d'engendrer un couple de mots (f, g') de X^*, l'un de longueur n et hauteur terminale 0, l'autre de longueur $n+1$ et hauteur terminale 1. Chacun de ces mots peut être engendré à l'aide d'un algorithme très classique de tirage aléatoire et uniforme d'une partie à $\lfloor n/2 \rfloor$ éléments dans l'ensemble $\{1, 2, \ldots n\}$ [10]; les entiers tirés correspondent aux occurrences des lettres $\bar{x}$ dans le mot. Soit g le mot obtenu en ôtant la dernière lettre de g'. Par la correspondance du théorème 2.2, on associe au couple (f, g) un chemin NSEO, codé naturellement par un mot $w' \in X^* \sqcup\!\sqcup Y^*$. Soient $u' \in X^*$ et $v' \in Y^*$ les deux mots composant w', et u et v les mots de F_x et F_y qui leur sont respectivement associés par la bijection φ^{-1} (théorème 2.1). Alors l'unique mot $w \in u \sqcup\!\sqcup v$ tel que les lettres de u apparaissent dans w aux mêmes occurences que celles de u' dans w' codent un chemin du premier quadrant. Celui-ci est bien engendré de façon équiprobable, en raison de l'uniformité du tirage du couple de mots (f, g), et de la correspondance bijective entre ce couple et le chemin obtenu. Toutes les opérations précédentes peuvent bien sûr être effectuées en temps et place mémoire linéaires. La procédure de génération d'un chemin de longueur impaire se déduit tout aussi aisément de la preuve du théorème 3.1. L'algorithme se généralise naturellement aux chemins du premier octant en réseau cubique centré, à l'aide de la proposition 3.2.

Bibliographie

[1] Chottin (L.) et Cori (R.). – Une preuve combinatoire de la rationnalité d'une série génératrice associée aux arbres. *R.A.I.R.O*, vol. 16, n° 2, 1982, pp. 113–128.

[2] Cori (R.), Dulucq (S.) et Viennot (X. G.). – Shuffles of parenthesis systems and Baxter permutations. *Journal of Combinatorial Theory A*, vol. 43, n° 1, 1986, pp. 1–22.

[3] Denise (A.). – *Méthodes de génération aléatoire d'objets combinatoires de grande taille et problèmes d'énumération.* – Thèse, Université Bordeaux I, Janvier 1994.

[4] Domb (C.). – On the theory of cooperative phenomena in crystals. *Advances in Physics*, vol. 9, n° 34, 1960, pp. 150–361.

[5] Flajolet (Ph.), Zimmermann (P.) et Van Cutsem (B.). – *A Calculus for the Random Generation of Combinatorial Structures.* – Rapport technique n° 1830, INRIA, janvier 1993. A paraître dans Theoretical Computer Science.

[6] Gouyou-Beauchamps (D.). – *Codages par des mots et des chemins : problèmes combinatoires et algorithmiques.* – Mai 1985. Thèse d'Etat.

[7] Guy (R. K.), Krattenthaler (C.) et Sagan (B. E.). – Lattice paths, reflections & dimension-changing bijections. *Ars Combinatoria*, vol. 34, 1992, pp. 3–15.

[8] Lothaire (M.). – *Combinatorics on Words.* – Addison-Wesley, 1983. Encyclopedia of Mathematics and its applications.

[9] Percus (J.K.). – *Combinatorial methods.* – Springer-Verlag, 1971.

[10] Wilf (H. S.). – A unified setting for sequencing, ranking, and selection algorithms for combinatorial objects. *Advances in Mathematics*, vol. 24, 1977, pp. 281–291.