Agrégation externe de mathématiques, session 2023
Épreuve de modélisation, option C : Algèbre et Calcul Formel

(*C01-public*)

Résumé : Le texte étudie plusieurs façons de calculer les premiers termes d'une suite d'entiers dénombrant certains arbres, en s'attachant à trouver des algorithmes efficaces pour cette tâche.

Mots clefs : arithmétique des polynômes, polynômes.

► *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. La présentation, bien que totalement libre, doit être organisée et le jury apprécie qu'un plan soit annoncé en préliminaire. L'exposé doit être construit en évitant la paraphrase et mettant en lumière les connaissances, à partir des éléments du texte. Il doit contenir des illustrations informatiques réalisées sur ordinateur, ou, à défaut, des propositions de telles illustrations. Des pistes de réflexion, indicatives et largement indépendantes les unes des autres, vous sont proposées en fin de texte.*

1. Arbres en chimie organique

En chimie, les *alcanes* sont des molécules formées uniquement d'atomes de carbone (C) et d'hydrogène (H), chaque atome C étant relié à quatre autres atomes (H ou C), et chaque atome H étant relié à un atome C (Figure 1). Les alcanes sont décrits de façon générique par la formule chimique brute C_nH_{2n+2}, où $n \geq 1$.

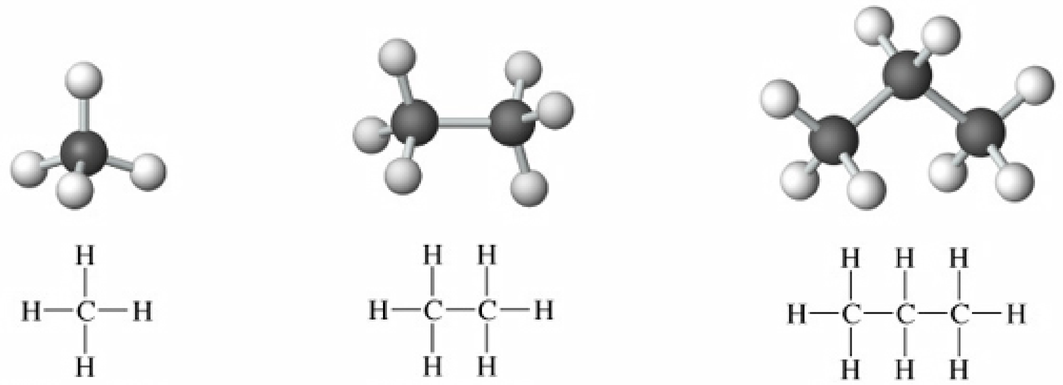

FIGURE 1. Exemples d'alcanes : CH_4 (méthane), C_2H_6 (éthane), C_3H_8 (propane).

Pour un n fixé, il existe plusieurs *isomères* différents, c'est-à-dire différentes structures des liaisons entre les atomes. Autrement dit, plusieurs molécules possèdent la même formule brute C_nH_{2n+2} mais ont des formules développées différentes. Par exemple, il y a exactement cinq isomères de l'hexane (C_6H_{14}), dont les formules semi-développées sont représentées

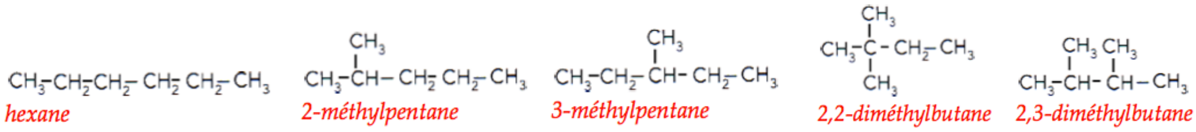

FIGURE 2. Les cinq isomères de l'hexane.

en Figure 2. Les isomères peuvent avoir des propriétés physiques, chimiques et biologiques différentes; leur énumération est un problème fondamental en chimie organique.

Mathématiquement, les composants chimiques tels les alcanes peuvent être codés par des arbres soumis à certaines contraintes. Rappelons qu'un *graphe* est défini par la donnée d'un ensemble de points nommés *sommets* et d'un ensemble de segments nommés *arêtes*, chaque arête reliant une paire de sommets, et qu'un *arbre* est un graphe connexe sans cycle. Le *degré* d'un sommet est le nombre d'arêtes le reliant aux autres sommets.

À chaque alcane on associe un arbre, son "squelette de carbone", obtenu en enlevant tous les atomes d'hydrogène du graphe associé à sa formule semi-développée. Déterminer le nombre b_n d'isomères de C_nH_{2n+2} se ramène ainsi à dénombrer les arbres à n sommets, dont chaque sommet a un degré au plus 4. Pour tout $N \geq 0$, notons $B_N(x)$ le N-ième polynôme générateur de la suite $(b_n)_{n\geq 0}$, c'est-à-dire le polynôme $B_N(x) = \sum_{n=0}^{N-1} b_n x^n$, où par convention $b_0 = 1$. Une énumération exhaustive montre que

$$B_8(x) = 1 + x + x^2 + x^3 + 2x^4 + 3x^5 + 5x^6 + 9x^7.$$

Un problème lié, mais plus simple, est l'énumération des isomères des *alcools*. Par définition, ceux-ci sont obtenus à partir d'alcanes en remplaçant un atome d'hydrogène par un radical OH, comportant un atome d'oxygène et un atome d'hydrogène liés. Les n-alcools sont en bijection avec les *arbres enracinés* (c'est-à-dire ayant un sommet distingué, appelé *racine*) à n sommets, dont la racine est de degré au plus 3 et dont chaque sommet est de degré au plus 4.

On ne connaît aucune formule explicite pour le nombre a_n de n-alcools. En revanche, il est connu que le polynôme générateur $A_N(x) = \sum_{n=0}^{N-1} a_n x^n$ vérifie la formule (qu'on admettra)

$$A_N(x) = 1 + x\left(\frac{1}{3}A_N(x^3) + \frac{1}{2}A_N(x)A_N(x^2) + \frac{1}{6}A_N(x)^3\right) \bmod x^N, \tag{1}$$

où, pour deux polynômes A et B, on écrit $A = B \bmod x^N$ lorsque x^N divise $A - B$.

Ceci permet le calcul de a_n pour tout n. Par exemple,

$$A_{11}(x) = 1 + x + x^2 + 2x^3 + 4x^4 + 8x^5 + 17x^6 + 39x^7 + 89x^8 + 211x^9 + 507x^{10}.$$

Une étude de l'équation (1) dépassant le cadre de cette épreuve permet de prouver que a_n croît asymptotiquement (lorsque $n \to \infty$) en $\kappa\rho^n/n^{3/2}$, pour $\rho \approx 2,815$ et $\kappa \approx 0,518$.

2. Énumération d'arbres 2-3-4

Plus généralement, il est naturel de s'intéresser à des familles d'arbres dont les degrés sont soumis à certaines contraintes. Dans la suite du texte, nous considérerons une famille

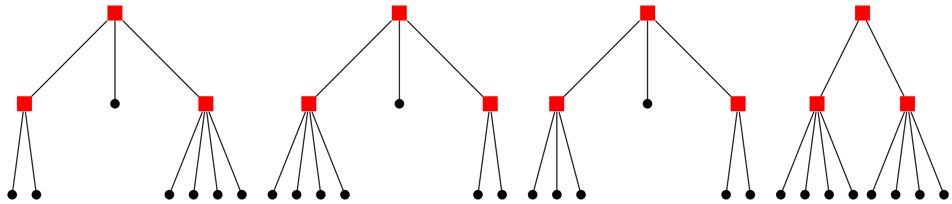

FIGURE 3. Quatre parmi les 333 arbres 2-3-4 à trois sommets internes.

d'arbres enracinés plus simples à étudier que ceux des alcanes et des alcools. Il s'agit des *arbres 2-3-4*, définis de façon récursive comme suit. Un tel arbre est soit constitué d'un unique *sommet feuille* (représenté par le symbole •), soit composé d'une racine reliée à deux, trois ou quatre sous-arbres de même type. Les *sommets internes* d'un arbre 2-3-4 sont ceux de degré au moins égal à 2 (représentés par des carrés ■), c'est-à-dire ceux qui ne sont pas des sommets feuilles. La taille d'un arbre est égale au nombre de ses sommets internes.

Quelques arbres 2-3-4 de taille 3 sont représentés en Figure 3. Contrairement au cas des alcanes et des alcools, l'ordre des sous-arbres est pris en compte ; par exemple les deux premiers arbres en Figure 3 sont considérés distincts. Soit f_n le nombre d'arbres 2-3-4 de taille n. Par énumération exhaustive, on constate que les premières valeurs de la suite $(f_n)_{n\geq 0}$ sont :

$$f_0 = 1, \quad f_1 = 3, \quad f_2 = 27.$$

La suite du texte est dédiée à la conception de plusieurs méthodes pour le calcul d'un ou de plusieurs termes de la suite $(f_n)_{n\geq 0}$.

2.1. *Une récurrence non linéaire*

Une première remarque est que la suite $(f_n)_{n\geq 0}$ vérifie une récurrence non linéaire qui permet de calculer ses termes de proche en proche.

Proposition 1. *Pour tout $n \geq 1$, on a :*

$$f_n = \sum_{i+j=n-1} f_i f_j + \sum_{i+j+k=n-1} f_i f_j f_k + \sum_{i+j+k+\ell=n-1} f_i f_j f_k f_\ell. \tag{2}$$

Démonstration. La formule (2) découle de la définition récursive des arbres 2-3-4. □

L'utilisation de la Proposition (1) entraîne la conséquence algorithmique suivante.

Corollaire 2. *Pour tout $N \geq 1$, on peut calculer $f_0, \ldots, f_N$ en $O(N^4)$ opérations dans $\mathbb{Q}$.*

2.2. *Un calcul par itération*

Pour tout $N \geq 0$, nous introduisons le N-ième polynôme générateur de la suite $(f_n)_{n\geq 0}$

$$F_N(x) = \sum_{n=0}^{N-1} f_n x^n \quad \in \mathbb{Q}[x].$$

En utilisant la méthode de la section 2.1, on détermine

$$F_7(x) = 1 + 3x + 27x^2 + 333x^3 + 4752x^4 + 73764x^5 + 1209492x^6.$$

Proposition 3. *F_N est l'unique polynôme de degré $\le N-1$ dans $\mathbb{Q}[x]$ qui vérifie*

$$F_N = 1 + xF_N^2 + xF_N^3 + xF_N^4 \bmod x^N. \tag{3}$$

Démonstration. Pour $\ell \ge 2$ et $n < N$, le coefficient de x^n du polynôme $xF_N^\ell \bmod x^N$ vaut

$$\sum_{i_1+\cdots+i_\ell=n-1} f_{i_1}\cdots f_{i_\ell}.$$

L'égalité (3) est donc équivalente à la récurrence (2) avec la condition initiale $f_0 = 1$. □

Introduisons la suite de polynômes $(H_k(x))_{k\ge0}$ de $\mathbb{Q}[x]$ définie par

$$H_0 = 1, \qquad H_{k+1} = 1 + xH_k^2 + xH_k^3 + xH_k^4 \quad \text{pour} \quad k \ge 0. \tag{4}$$

Rappelons que la valuation d'un polynôme non identiquement nul $P = \sum_{n=0}^{\deg P} p_n x^n \in \mathbb{Q}[x]$, notée $\mathrm{val}(P)$, est par définition égale au plus petit entier n tel que $p_n \ne 0$. On pose $\mathrm{val}(0) = \infty$.

Proposition 4. *La valuation de $F_N - H_N$ est au moins égale à N.*

Démonstration. Pour $k = 0, 1, \ldots, N-1$, les égalités (3) et (4) entraînent que le polynôme $F_N - H_{k+1}$ s'écrit $x(F_N - H_k)Q_k + x^N R_k(x)$ pour deux polynômes Q_k, R_k de $\mathbb{Q}[x]$. Il vient l'inégalité $\mathrm{val}(F_N - H_{k+1}) \ge \min(N, \mathrm{val}(F_N - H_k) + 1)$. Le résultat découle par récurrence sur k. □

En utilisant la Proposition 4, on peut modifier l'itération (4) en menant les calculs modulo x^{N+1}, ce qui entraîne la conséquence suivante.

Corollaire 5. *Pour tout $N \ge 1$, on peut calculer $f_0, \ldots, f_N$ en $O(N^3)$ opérations dans $\mathbb{Q}$.*

2.3. *Une itération plus efficace*

On introduit le polynôme $P(x, y) = y - 1 - x(y^2 + y^3 + y^4)$ et on note P_y sa dérivée par rapport à la variable y. On définit la suite de polynômes $(G_k(x))_{k\ge0}$ par

$$G_0 = 1, \qquad G_{k+1} = G_k - \frac{P(x, G_k)}{P_y(x, G_k)} \bmod x^{2^{k+1}} \ \text{ et } \ \deg(G_{k+1}) < 2^{k+1} \quad \text{pour} \quad k \ge 0. \tag{5}$$

Remarquons que $P_y(x, G_k)$ est premier avec $x^{2^{k+1}}$, donc l'expression précédente est correctement définie. Nous allons prouver le résultat ci-dessous.

Proposition 6. *Pour tout entier $k \ge 0$, la valuation du polynôme*

$$G_k - (1 + xG_k^2 + xG_k^3 + xG_k^4) \tag{6}$$

est au moins égale à 2^k. En particulier, on a $F_{2^k} = G_k$.

Démonstration. Grâce à l'égalité (5), il suffit de montrer que la valuation de $G_{k+1} - G_k$ est au moins égale à 2^k. Nous prouvons ceci par récurrence sur k. L'assertion est clairement vraie pour $k = 0$. On vérifie par calcul que le polynôme $(y_1 - y_2)^2$ divise

$$P(x, y_2) - P(x, y_1) - (y_2 - y_1)P_y(x, y_1) \tag{7}$$

dans l'anneau $\mathbb{Q}[y_1, y_2, x]$. En substituant y_1 par $G_k(x)$ et y_2 par $G_{k+1}(x)$, cela entraîne que

$$P(x, G_{k+1}) = (G_{k+1} - G_k)^2 S_k \bmod x^{2^{k+1}}, \tag{8}$$

pour un certain polynôme S_k. Comme par hypothèse de récurrence, le polynôme $G_{k+1} - G_k$ est de valuation au moins 2^k, l'égalité précédente implique que $P(x, G_{k+1})$ est de valuation au moins 2^{k+1}. L'unicité prouvée en proposition 3 entraîne $F_{2^k} = G_k$. □

Une conséquence importante de la Proposition 6 est un algorithme plus efficace pour calculer les premiers termes de la suite $(f_n)_{n\geq 0}$.

Corollaire 7. *Pour tout $N \geq 1$, on peut calculer $f_0, \dots, f_N$ en $O(N^2)$ opérations dans $\mathbb{Q}$.*

2.4. *Une récurrence linéaire*

Nous allons utiliser une stratégie différente. Le point de départ est de remarquer que la série entière $F(x) = \sum_{n\geq 0} f_n x^n$ (dont on admettra que le rayon de convergence est strictement positif), qui est solution de $P(x, F(x)) = 0$, vérifie l'équation différentielle

$$F'(x) = -\frac{P_x}{P_y}(x, F(x)), \tag{9}$$

où P_x et P_y désignent les polynômes obtenus par dérivation de P par rapport à x et à y.

Les polynômes P et P_y, vus dans l'anneau $\mathbb{Q}(x)[y]$, sont premiers entre eux. On peut donc construire deux éléments u et v de $\mathbb{Q}(x)[y]$ tels que : $uP + vP_y = 1$. L'égalité (9) entraîne alors :

$$F'(x) = Q_1(x, F(x)), \tag{10}$$

où $Q_1 \in \mathbb{Q}(x)[y]$ est le reste de la division euclidienne de $-vP_x$ par P.

En dérivant de façon répétée les deux membres de l'égalité (10), nous concluons qu'il existe des polynômes Q_j dans $\mathbb{Q}(x)[y]$, de degré au plus 3, tels que

$$F^{(j)}(x) = Q_j(x, F(x)), \quad \text{pour tout } j \geq 0. \tag{11}$$

Autrement dit, nous venons d'exprimer les dérivées successives $F', F'', F''', \dots$ de F comme combinaisons linéaires sur $\mathbb{Q}(x)$ des quatre séries $1, F, F^2, F^3$. Il s'ensuit que les séries $1, F, F'$, F'' et F''' vivent dans un $\mathbb{Q}(x)$-espace vectoriel de dimension au plus 4, d'où nous déduisons :

Proposition 8. *La série entière $F(x)$ vérifie une égalité de la forme :*

$$c_3(x)F'''(x) + c_2(x)F''(x) + c_1(x)F'(x) + c_0(x)F(x) = b(x), \tag{12}$$

où $b(x)$ et les $c_i(x)$ sont des polynômes dans $\mathbb{Q}[x]$.

Une équation différentielle de type (12) peut être déterminée par un calcul d'algèbre linéaire sur $\mathbb{Q}(x)$. On peut en trouver une dont les coefficients ont degré au plus 8, par exemple

$$c_3 = x^2(12x^3 + 48x^2 + 572x - 27)(36x^3 + 90x^2 + 405x - 25), \quad c_0 = 144(21x^3 + 36x^2 - 80x - 10).$$

Corollaire 9. *Pour tout $N \geq 1$, on peut calculer $f_0, \dots, f_N$ en $O(N)$ dans $\mathbb{Q}$.*

Démonstration. On extrait le coefficient de x^k dans les deux membres de (12) et on utilise le fait que le coefficient de x^k dans $x^i F^{(j)}(x)$ vaut $(j+k-i)(j+k-i-1)\cdots(k-i+1) f_{j+k-i}$. Cela fournit une récurrence linéaire vérifiée par la suite $(f_n)_{n\geq 0}$, qu'il suffit de dérouler. □

2.5. *Une méthode pour calculer la parité de f_N*

Une variation de la méthode précédente permet de calculer la parité du coefficient f_N. Le point de départ est l'observation qu'il est possible de déterminer un multiple du polynôme $P(x,y) = y-1-x(y^2+y^3+y^4)$ contenant uniquement des monômes de la forme $x^i y^{2^j}$. Un tel polynôme peut être déterminé de la façon suivante. On commence par effectuer la division euclidienne de $y, y^2, y^4, y^8, \dots$ par P. Ceci permet de trouver des polynômes R_k dans $\mathbb{Q}(x)[y]$, de degrés au plus 3, tels que $F(x)^{2^k} = R_k(x, F(x))$, pour tout $k \geq 0$. Par un calcul d'algèbre linéaire, suivi d'une réduction modulo 2, nous déduisons l'existence de la relation suivante

$$x + F(x) + (x^2+x+1)F(x^2) + xF(x^4) + x^2F(x^8) = 0 \quad \text{modulo } 2.$$

Par extraction de coefficients, nous tirons l'égalité :

$$f_n \bmod 2 = \begin{cases} f_{(n-1)/2} \bmod 2, & \text{si } n \equiv 3 \bmod 4. \\ f_{(n-1)/2} + f_{(n-1)/4} \bmod 2, & \text{si } n \equiv 1 \bmod 4, \\ f_{n/2} + f_{n/2-1} + f_{(n-2)/8} \bmod 2, & \text{si } n \equiv 2 \bmod 8, \\ f_{n/2} + f_{n/2-1} \bmod 2, & \text{sinon.} \end{cases} \qquad \text{pour tout } n \geq 2.$$

Corollaire 10. *Pour tout $N \geq 1$, il est possible de calculer la parité de f_N en $O(\log N)$ opérations.*

Suggestions et pistes de réflexion

► *Les pistes de réflexion suivantes ne sont qu'indicatives et il n'est pas obligatoire de les suivre. Vous pouvez choisir d'étudier, ou non, certains des points proposés, de façon plus ou moins approfondie, mais aussi toute autre question à votre initiative. Vos investigations comporteront une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats. À défaut, si vos illustrations informatiques n'ont pas abouti, il est conseillé d'expliquer ce que vous auriez souhaité mettre en œuvre.*

— Tout au long du texte, plusieurs faits ou résultats sont énoncés sans démonstration, ou avec des démonstrations incomplètes. On pourra en justifier certains.
— En s'inspirant de la méthode de la Section 2.2, calculer le nombre a_N de N-alcools puis vérifier expérimentalement que $\rho \approx 2,815$ et $\kappa \approx 0,518$.
— Revisiter les analyses de complexité du texte, en supposant qu'on dispose d'un algorithme de multiplication polynomiale en degré N de complexité $O(N^\alpha)$, pour $\alpha \in]1,2[$.
— Utiliser un logiciel de calcul formel pour déterminer la parité des entiers f_{9^9} et $f_{19^{19}}$.
— Montrer que f_n est divisible par 3 pour tout $n \geq 1$ et par 9 pour tout $n \geq 2$.
— En admettant que f_n croît en $\gamma n^\alpha \beta^n$, estimer numériquement les valeurs α, β, γ.
— Discuter le cas des arbres 2-3-$\cdots$-M, pour un entier $M \geq 5$.