

4. Le problème des suites de de Bruijn en dimension 2

En dimension 2, le problème des suites de de Bruijn devient celui des tores de de Bruijn. Pour $n = 2$, il existe 16 carrés 2×2 composés de 0 et de 1 (a).

La question est : peut-on remplir de 0 et de 1 un carré 4×4 représentant la surface d'un tore (les bords haut et bas sont collés, les bords droit et gauche sont collés) de telle façon qu'en faisant glisser une

petite fenêtre carrée de taille 2×2 sur les 16 positions possibles qu'elle peut prendre, on trouve les 16 carrés 2×2 énumérés ci-dessous ? Une solution est figurée en b.

Si à la place de deux symboles, on en considère quatre, il y a cette fois $4^4 = 256$ carrés possibles de taille 2×2 , et on peut donc (au mieux) espérer utiliser la surface carrée d'un tore 16×16 pour énumérer

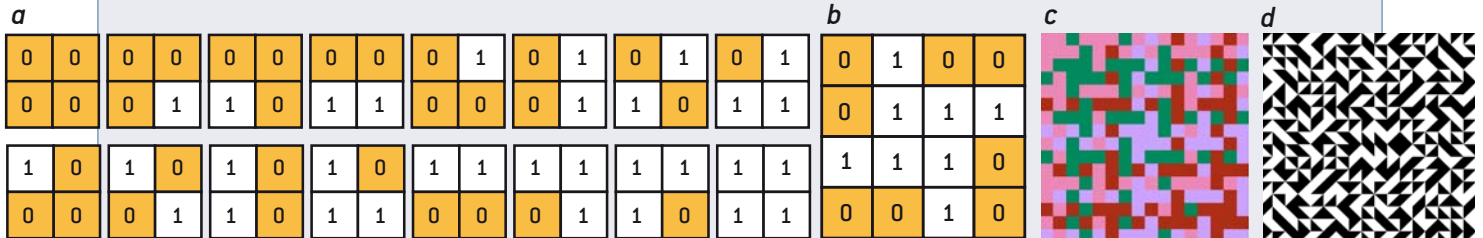
par fenêtre glissante tous les carrés possibles de taille 2×2 utilisant quatre symboles. Une solution, figurée en couleurs, est le schéma c.

Si, à la place de quatre couleurs, on utilise les quatre façons de colorier un carré en noir et blanc en le divisant par une diagonale, on obtient un intéressant tableau abstrait (d).

Un tel pavage de surface a un intérêt pratique. Un robot se

déplaçant à la surface d'une pièce où ce motif serait dessiné pourrait, en observant uniquement les quatre carrés placés sous lui et en utilisant une table convenable (comme le magicien avec son tableau caché), savoir où il se trouve précisément.

Pour d'autres informations sur cette variante du problème de de Bruijn, voir : <http://tiny.cc/9y96fx>



s'aidera d'un petit tableau (voir l'encadré 1) qui contient les réponses. Il le placera discrètement en un endroit où lui seul peut le voir, sur un petit carton qu'il a en main ou collé sur le revers de sa veste, ou sur un des objets utilisés pour son spectacle.

Un graphe pour tout comprendre

Reste une question : comment trouver les suites de Nicolaas de Bruijn ?

La suite utilisée est celle-ci :

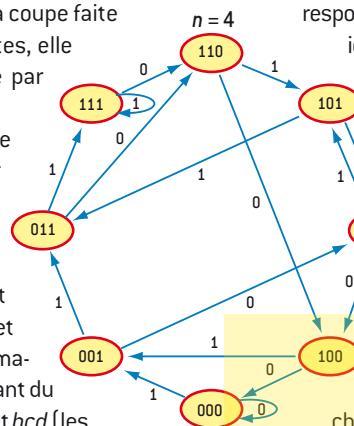
RRRRRRRRRNNRRNNNNNNRRRRNNNNRNRN.

Ramenons-nous à des bits d'information en remplaçant les N par des 1 et les R par des 0 : 00000100101100111110001101110101. Comment trouver une telle suite binaire, ou en trouver d'autres, permettant de généraliser le tour avec, par exemple, six spectateurs et un jeu de 64 cartes ?

L'idée repose sur une représentation en graphe du problème. Elle va nous conduire à une solution générale qui prouvera que les suites de de Bruijn existent pour tout ordre n . Il est plus simple d'expliquer la méthode avec

un jeu de seulement 16 cartes (on ne retient par exemple que les cœurs et les piques) et quatre spectateurs : les couleurs des cartes des quatre spectateurs dont prend connaissance le magicien fournissent maintenant 4 bits d'information au magicien ; cela lui permet de retrouver la coupe faite dans le jeu de 16 cartes, elle aussi est déterminée par 4 bits d'information.

Le graphe ci-contre permet de trouver toutes les suites de de Bruijn d'ordre 4. Il est composé de huit sommets qui sont les huit triplets possibles de 0 et de 1. On a tracé systématiquement un arc partant du triplet abc vers le triplet bcd (les deux derniers symboles de abc sont identiques aux deux premiers symboles de bcd). On a écrit d sur l'arc qui joint abc à bcd . Cette façon de dessiner le graphe assure que de chaque sommet partent exactement deux arcs, et que sur chaque sommet arrivent exactement deux arcs. Il y a en tout 16 arcs.



Si l'on considère un chemin suivant une série de flèches liées et passant d'un sommet à un autre, on construit une suite finie de 0 et de 1. Par exemple, en partant du sommet 110 en haut, on passe à 100, puis à 001, puis à 011, puis à 110. À un tel chemin, on fait correspondre une notation compactée, ici 1100110, obtenue en notant

les uns derrière les autres les noms des sommets, sans répéter les deux symboles communs à deux sommets consécutifs (en bleu) : 110-100-001-011-110 → 1100110.

Trouver une suite de de Bruijn d'ordre 4 consiste maintenant à trouver un chemin revenant à son point de départ (un « circuit ») qui passe par tous les arcs une fois et une fois seulement. Un tel circuit passe deux fois par chaque sommet du graphe, puisque de chaque sommet partent deux arcs. Un tel circuit donne une suite de de Bruijn, car si on écrit la notation compactée du circuit, on y trouvera tous les

5. Les graphes de de Bruijn pour l'assemblage des génomes

Les séquenceurs d'ADN produisent de nombreuses petites séquences extraites de la longue séquence de quatre lettres A, G, C, T à laquelle on s'intéresse (un gène, un chromosome, etc.). Les petites séquences ont des parties communes qui déterminent leur assemblage correct. Une petite séquence

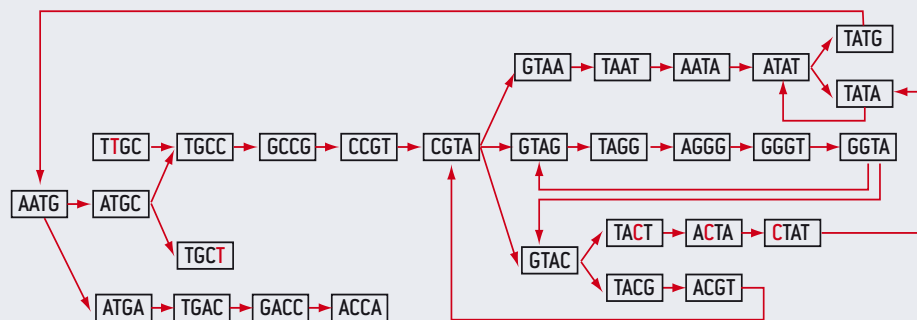
peut parfois s'assembler avec plusieurs, et on est donc face à un énorme puzzle : toutes les petites séquences produites par les machines doivent être utilisées, afin de ne déterminer qu'une seule grande séquence.

Depuis les travaux de Pavel Pevzner, Haixu Tang et Michael Waterman, les méthodes utili-

sées pour traiter ces énormes problèmes combinatoires sont fondées sur la recherche de chemins eulériens dans des graphes analogues à ceux de l'encadré 3. Les sommets de ces graphes (dont un exemple est représenté ci-dessous) sont les séquences données par la machine et les arcs indiquent les

paires de séquences ayant une partie commune (car l'une commence par ce qui est la fin de l'autre). Mener l'assemblage se ramène à trouver un chemin eulérien dans le graphe.

Le problème est rendu encore plus difficile par certaines erreurs présentes dans les résultats fournis par les machines à séquencer. Les lettres en rouge indiquent les endroits où il y a des erreurs de séquençage. La mise au point de ces algorithmes d'assemblages à base de graphes de de Bruijn fut une étape importante dans l'histoire du séquençage et aujourd'hui encore de nombreux travaux sont consacrés au perfectionnement de ces méthodes.



quadruplets possibles de 0 et de 1. En effet, le circuit passe deux fois par tout sommet (un triplet) ; une première fois, il le quitte en choisissant un 0, la seconde fois en choisissant un 1. Chaque quadruplet est obtenu une seule fois, sinon le circuit passerait deux fois par le même arc.

Donnons un exemple. Soit le circuit passant par les sommets 000-000-001-011-110-100-001-010-101-011-111-110-101-010-100. Sa notation compacte est la suite de de Bruijn d'ordre 4 notée 0000110010111101.

On vérifie qu'en prenant quatre symboles consécutifs, on obtient les 16 suites possibles de 4 symboles 0 ou 1 dans l'ordre suivant : 0000-0001-0011-0110-1100-1001-0010-0101-1011-0111-1111-1110-1101-1010-0100-1000.

La démonstration complète que, par cette méthode, on trouvera des suites de de Bruijn d'ordre n pour tout n est presque terminée. Il suffit d'évoquer un célèbre résultat de théorie des graphes dû à Leonhard Euler : **Théorème d'Euler** : Un graphe orienté possède un circuit empruntant tous les arcs une fois exactement (dénomé circuit eulérien) si et seulement si :

- il est connexe (on peut toujours joindre deux sommets en empruntant les arcs du graphe) ;
- à chaque sommet, le nombre d'arcs entrants est le même que le nombre d'arcs sortants.

Circuits eulériens

Pour les graphes considérés ici, la seconde condition du théorème est vérifiée, car de chaque sommet partent deux arcs, et en arrivent deux. Pour la première condition, il faut réfléchir un peu : partant de abc (cas $n = 3$), peut-on arriver à efg ? Oui, il suffit de suivre les arcs abc , bce , cef , efg . Pour n quelconque, l'idée s'adapte.

Les graphes de de Bruijn vérifient les conditions du théorème d'Euler. Ils ont donc tous des circuits eulériens, et donc pour tout entier n , il existe des suites de de Bruijn. Conclusion : quel que soit n , on peut faire le tour de cartes avec n spectateurs et un jeu de 2^n cartes. La méthode des circuits eulériens permet de trouver les suites de de Bruijn d'ordre n , pour tout n , et elle les donne toutes. Cela permet de les dénombrer. Il y en a exactement $2^{2^n - 1 - n}$.

Une multitude de résultats ont été démontrés concernant les suites et les

graphes de Nicolaas de Bruijn et leurs variantes. Ces résultats sont utiles par exemple en robotique (voir l'encadré 4) ou dans le domaine du traitement des séquences génétiques pour assembler les courtes séquences que les machines déchiffrent en grand nombre et qu'il faut recoller pour connaître, par exemple, la totalité d'un chromosome (voir l'encadré 5).

Le merveilleux livre *Magical Mathematics* de Persi Diaconis et Ron Graham (voir la bibliographie) consacre trois chapitres aux tours de magie fondés sur les suites de Nicolaas de Bruijn et décrit plusieurs autres applications de ces suites. On y trouve aussi des informations sur l'histoire du tour présenté au début d'article. Elle est intéressante, car elle montre qu'avant que les mathématiciens ne s'en mêlent, l'idée du tour était déjà comprise, mais n'était qu'imparfaitement mise en œuvre. La première version du tour fut exposée en 1919 par Charles Jordan, un fermier californien, éleveur de poulets, qui inventait et vendait des tours de magie. Cette première version n'était pas la meilleure possible : le magicien, qui utilisait pourtant un jeu de 32 cartes, interrogeait six spectateurs au lieu de cinq. Le rangement du jeu ne