

Les démonstrations par ordinateur

Le recours à l'ordinateur pour démontrer des résultats mathématiques est de plus en plus fréquent. Souvent, l'ordinateur n'intervient qu'associé à des résultats purement mathématiques démontrés à la main et qui posent les bases des algorithmes confiés à la machine. C'est le cas pour le théorème des 17 données du sudoku.

Les démonstrations de théorèmes dépendant de calculs informatiques ont souvent été jugées insatisfaisantes, car d'une part, aucun humain seul ne peut les contrôler, d'autre part, les programmes et leurs calculs sont susceptibles d'erreurs de toutes sortes. Pour limiter ce risque d'erreurs, on a développé une méthode de « preuve formelle » par ordinateur, qui diminue considérablement ce risque et qui permet même une meilleure garantie de justesse que les preuves humaines dans certains cas. (voir cette rubrique de juin 2011, *Du rêve à la réalité des preuves*). On met en œuvre des « assis-

tants de preuves », des logiciels spécialisés que les mathématiciens utilisent pour écrire pas à pas toutes les étapes d'une démonstration, sans en oublier aucune (oubli qui est possible quand la preuve est menée par un humain en utilisant de nombreux raccourcis).

La preuve complète mise au point est vérifiée par une procédure indépendante et très poin-



a

tilleuse ; éventuellement, cette vérification est confiée à plusieurs autres logiciels de contrôle. Les logiciels de contrôle peuvent eux-mêmes avoir été soumis à des contrôles très stricts « prouvant » leur bon fonctionnement. On parvient ainsi à une certitude quasi parfaite qu'une preuve est correcte.

Parmi les théorèmes soumis à cette vérification par la méthode des assistants de preuves, deux ont retenu particulièrement l'attention. L'un est le théorème des quatre couleurs, qui indique que « toute carte peut être coloriée avec quatre couleurs sans que deux pays voisins soient colorés de la même façon » (a) ; il a été démontré en 1976 et vérifié formellement en 2005 par une équipe de l'Inria.

L'autre est le théorème de Thomas Hales qui résout la conjecture de Kepler sur l'empilement le plus dense des

b



sphères (b), démontré en 1998 et formellement vérifié par T. Hales et son équipe en 2014.

Malheureusement, dans le cas du résultat de Gary McGuire sur le sudoku, la méthode des assistants de preuve ne pourra sans doute pas s'appliquer avant longtemps : le calcul effectué est énorme et, à moins d'avancées mathématiques nouvelles, une preuve formelle ne pourra que refaire le calcul dans un cadre plus rigoureux pour éviter les erreurs, ce qui demande encore plus de calculs, chose impossible à envisager aujourd'hui.

La méthode de résolution se décompose en trois étapes : (A) énumérer toutes les grilles complètes de sudoku ; (B) pour chacune, considérer toutes les façons d'en extraire 16 données ; et (C) pour chaque jeu de 16 données ainsi extrait, faire fonctionner une méthode de résolution de manière à trouver au moins deux solutions différentes.

Si toutes les grilles complètes passent ce test, alors on aura démontré que 16 données ne sont jamais suffisantes et que 17 est donc la bonne réponse. Le faire brutalement comme je viens de le décrire demanderait des siècles de calcul.

Trois remarques économisent du temps de calcul et, bien employées, donnent une durée de calcul raisonnable.

Point (A) : limiter l'énumération des grilles complètes qu'on fait dérouler.

Point (B) : trouver des méthodes évitant

d'énumérer tous les sous-ensembles de 16 éléments d'un ensemble de 81 éléments, car il y en a trop !

Point (C) : disposer d'un algorithme de test le plus efficace possible pour tester si une grille partielle est un problème correct.

Un catalogue intelligent et complet des grilles

Remarquons qu'on aurait pu procéder autrement : (A') énumérer toutes les grilles partielles possibles de 16 données, puis (B') montrer que chacune a plusieurs solutions. Toutefois, la théorie des « ensembles inévitables », qui permet de réduire considérablement les calculs de (B) (expliquée plus loin), rend la méthode en trois étapes (A-B-C) plus efficace que la méthode en deux étapes (A'-B').

Le nombre de grilles complètes différentes semble avoir été connu par Mark Brader dès 2003, mais c'est en 2006 qu'un document de Bertram Felgenhauer et Frazer Jarvis l'a justifié. Ce nombre est énorme et égal à $9! \times 2^{13} \times 3^4 \times 27\,704\,267\,971 = 6\,670\,903\,752\,021\,072\,936\,960 = 6,6... \times 10^{21}$.

C'est, en ordre de grandeur, la mémoire cumulée numérique de tous les dispositifs informatiques (téléphones compris), et ces dispositifs opèrent tous ensemble environ 10^{21} instructions par seconde. Ces chiffres sont imprécis, en particulier parce qu'on ne peut pas savoir quelles machines ont cessé de fonctionner, mais à un facteur 10 près, ils indiquent où nous en sommes.

Traiter une grille de sudoku pour savoir si elle possède des sous-grilles correctes à 16 données est un gros travail. Procéder

naïvement, même en mettant tous les dispositifs de calcul terrestres à la tâche, n'aurait même pas permis de résoudre le problème. Réduire le nombre de grilles à étudier est le premier point où les mathématiques ont aidé l'équipe de G. McGuire.

Si, dans une grille complète de sudoku, on change tous les 1 en 9 et tous les 9 en 1, on obtient une autre grille complète de sudoku. Si la première admet des problèmes corrects à 16 données, c'est aussi le cas de la seconde : partant d'un problème correct pour la première, on change les 1 en 9 et les 9 en 1, et on obtient un problème correct pour la seconde. Il suffit donc d'étudier l'une des deux. L'échange des 1 et des 9 n'est pas la seule transformation qui conduit à trouver des grilles équivalentes. En fait, toute permutation des neuf chiffres 1, 2, ..., 9 permet, à partir d'une grille complète, d'en construire une autre équivalente pour notre problème. Pour chaque grille, ce procédé donne $9! = 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 = 362\,880$ grilles équivalentes. Voilà déjà une belle réduction du calcul à entreprendre.

Cependant, la permutation des chiffres n'est pas la seule façon d'obtenir des grilles équivalentes. Voici la liste de quelques autres opérations qui, appliquées à une grille complète, donnent des grilles équivalentes.

- Choisir deux bandes horizontales et les échanger. Une bande horizontale est une série de trois carrés 3×3 placés les uns à côtés des autres (il y en a trois : une en haut, une au milieu, une en bas).
- Échanger deux bandes verticales. Une bande verticale est une série de trois carrés 3×3 placés les uns au-dessus des autres (il y en a trois : une à droite, une au centre, une à gauche).
- Échanger deux colonnes d'une même bande verticale.
- Échanger deux lignes d'une même bande horizontale.
- Transposer la grille (la ligne 1 devient la colonne 1, la ligne 2 devient la colonne 2, etc.).

Finalement, une grille est équivalente à un très grand nombre d'autres. Les grilles se regroupent en classes et il suffit de traiter une grille de chacune des classes.

9	3	7	8	5	6	2	4	1
5	6	2	1	9	4	3	8	7
4	8	1	2	7	3	5	6	9
8	2	3	6	4	7	9	1	5
6	1	5	9	3	2	4	7	8
7	4	9	5	8	1	6	2	3
3	7	8	4	6	9	1	5	2
1	9	6	7	2	5	8	3	4
2	5	4	3	1	8	7	9	6

DEUX « ENSEMBLES INÉVITABLES » sont indiqués (en bleu et en orange) dans cette grille. Chacun nécessite une donnée au moins pour que la solution au sudoku soit unique.

L'utilisation d'outils de la théorie des groupes, dont le lemme de Burnside (voir *Le miraculeux « lemme de Burnside » dans la rubrique de décembre 2006*), a permis de connaître le nombre exact de grilles complètes non équivalentes : il y en a $5\,472\,730\,538 \approx 5,5 \times 10^9$.

Glenn Fowler, chercheur aux Laboratoires AT&T, a constitué un catalogue complet des grilles en ne retenant qu'une seule grille par classe d'équivalence. Un algorithme de compression spécialement conçu pour les grilles de sudoku a permis de stocker tout le catalogue de G. Fowler sur six gigaoctets, ce qui est assez remarquable puisque cet algorithme permet de n'utiliser qu'un octet (8 bits) environ par grille.

Chercher les « ensembles inévitables »

Si l'on s'y prend sans finesse, pour chacune des grilles retenues, il faut envisager toutes les façons d'extraire 16 données. Or il y a $33\,594\,090\,947\,249\,085 \approx 3,4 \times 10^{16}$ façons de choisir 16 cases parmi 81. Cela conduirait à résoudre un grand nombre de problèmes de sudoku pour découvrir à chaque fois qu'il y a deux solutions ou plus. Ce nombre serait énorme, car égal à : $5\,472\,730\,538 \times 33\,594\,090\,947\,249\,085 = 183\,851\,407\,423\,359\,414\,572\,057\,730 \approx 1,8 \times 10^{26}$.

Nous savons que ce nombre est bien au-delà de ce qui est envisageable, même en utilisant toute la puissance de calcul disponible sur la planète. Les mathématiques à nouveau ont dû proposer une idée intéressante qui s'applique à d'autres problèmes, la « méthode des ensembles inévitables ». Expliquons cette idée sur l'exemple de la grille complète G ci-contre.

Pour construire un problème correct ayant pour solution cette grille G, il faut que l'une des données du problème soit l'une des quatre cases en bleu, en haut de la grille. En effet, il existe une autre grille complète G' très proche de G obtenue en échangeant seulement, dans les quatre cases en bleu, les 9 et les 5. Si les données de la grille partielle n'indiquent rien concernant ces quatre cases, il sera impossible