

produit maintenant de plus en plus souvent, il s'agit d'une démonstration avec un ordinateur (voir l'encadré page 78 pour d'autres exemples). Le détail des étapes du calcul n'a pas été publié, car cela donnerait un document d'une longueur colossale. L'article rendant compte de la preuve n'est donc que la description de la méthode utilisée, comportant, entre autres, les énoncés et démonstrations des propositions purement mathématiques utiles, mais pas assez d'éléments pour que l'on puisse vérifier le théorème sans tout reprogrammer et recalculer.

Des versions préliminaires de l'article étaient disponibles depuis un moment, mais

il n'a été publié dans sa forme définitive que le 12 juin 2014, dans la revue *Experimental Mathematics*. Malgré cette publication officielle, essentielle pour qu'un résultat soit considéré comme sérieux, et une vérification du calcul menée par une autre équipe (en septembre 2013), certains chercheurs, dont Joshua Cooper de l'Université de Caroline du Sud, doutent du statut de ce théorème dont la « démonstration », d'un type particulier, exclut difficilement le risque d'erreur. Nous y reviendrons.

Dans un premier temps, présentons la méthode de preuve utilisée par G. McGuire et les autres équipes qui ont abordé le

problème. Nous insisterons sur l'utilité d'une analyse mathématique préliminaire, la collaboration d'autres équipes (ayant mené d'autres calculs volumineux et transmis des programmes à G. McGuire et son équipe), la rigueur méthodologique indispensable, le délicat savoir-faire nécessaire à la programmation et à la réalisation des calculs et, enfin, l'incroyable quantité de calculs effectués.

Existe-t-il une grille complète G dont on peut extraire 16 données qui, quand on les complète (en respectant les règles du sudoku), donnent nécessairement la grille G ?

Des grilles intéressantes

La grille a comporte 17 données et n'a qu'une seule solution, b. Ce type de grille a posé la question : peut-on faire plus économique et ne fixer que 16 données dans une grille tout en préservant l'unicité de la solution ?

Le calcul de Gary McGuire et de son équipe montre que pour avoir une solution unique, il faut toujours spécifier au moins 17 cases. Mais parfois, 17 données ne suffisent pas. La grille c fournit un exemple où 18 données sont nécessaires (aucune sous-grille à 17 données n'est correcte). Cette grille permet aussi d'illustrer la puissance de la méthode des sous-ensembles inévitables au cœur des programmes de G. McGuire.

Considérons les deux colonnes verticales de trois chiffres en jaune en haut à gauche (1, 4, 7) et au centre (4, 7, 1). Si on les échange, on obtient une nouvelle grille complète de sudoku. Cela prouve que tout problème conduisant à cette grille doit comporter une donnée parmi les six cases concernées. Il en va de même pour les deux colonnes jaunes en haut au centre (4, 7, 1) et à droite (7, 1, 4), ainsi que pour

les colonnes (1, 4, 7), et (7, 1, 4) en jaune en haut. Il en résulte que parmi les neuf cases jaunes en haut, deux doivent être données dans tout problème ayant cette grille comme solution unique.

C'est aussi vrai pour les autres paquets de neuf cases (en bleu, en rouge) de la zone du haut, de la zone du milieu et de la zone du bas. Cela fait en tout neuf paquets de neuf cases, et pour chacun deux données au moins à retenir. Tout problème conduisant à c doit donc comporter au moins 18 données. Il n'y aura donc, dans ce cas particulier, aucun énoncé correct à 17 données.

Pour toute grille, un raisonnement montre qu'un problème

correct comporte au moins huit données. S'il ne comporte que sept données ou moins et doit conduire à une grille G fixée, ces données ne mentionnent pas deux chiffres, par exemple le 8 et le 9, et donc les sept données ne peuvent pas distinguer la grille complète visée G et la même grille G' où l'on a interverti les 8 et les 9. Les sept données ne constituent donc pas un problème correct.

De tels raisonnements mathématiques ne permettent pas aujourd'hui d'aller plus loin : on ne sait même pas prouver sans machine que huit données ne suffisent jamais à composer un problème correct.

a

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

b

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

c

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

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