

LOGIQUE & CALCUL

Le problème du sudoku

La donnée de 16 chiffres dans une grille de sudoku est insuffisante pour que la solution du problème soit unique. Pour le prouver, il a fallu énumérer toutes les grilles, en usant d'astuces afin de raccourcir les calculs.

Jean-Paul DELAHAYE

C'est une illusion de croire que la puissance des machines nous dispense de réfléchir. À propos du problème des données minimales au sudoku, nous allons voir que la limite du faisable est vite atteinte, et que pour la repousser, il faut de l'ingéniosité, des mathématiques et quand même beaucoup de patience.

Rappelons les règles du sudoku et fixons quelques termes. Une « grille de sudoku complète » est un tableau carré de $9 \times 9 = 81$ cases contenant chacune un chiffre de 1 à 9 et tel que :

(a) chaque ligne et chaque colonne contiennent chaque chiffre une fois exactement ;
(b) chaque sous-tableau 3×3 (résultat du découpage de la grille en neuf carrés de neuf cases) contient chaque chiffre une fois exactement.

Une grille partielle est un « problème de sudoku correct » s'il existe une façon unique de compléter la grille en une grille de sudoku complète, ce qui est le but du jeu.

Précisons qu'au sudoku, l'ordinateur est gagnant sur l'humain : il existe un grand nombre de programmes permettant à l'ordinateur de battre de vitesse le meilleur humain. Les bons programmes trouvent la solution d'un problème, aussi difficile soit-il, en moins d'un millième de seconde. On joue donc pour le plaisir de l'exercice.

Quand on cherche à formuler un énoncé, si l'on spécifie trop peu de cases, la grille

partielle ne constitue pas un problème correct (c'est évident si on ne retient qu'une case), car plusieurs solutions seront possibles. Dans les journaux, les grilles proposées comportent environ 25 cases remplies. On sait cependant qu'il existe des grilles correctes de sudoku ne comportant que 17 données (voir l'encadré ci-contre).

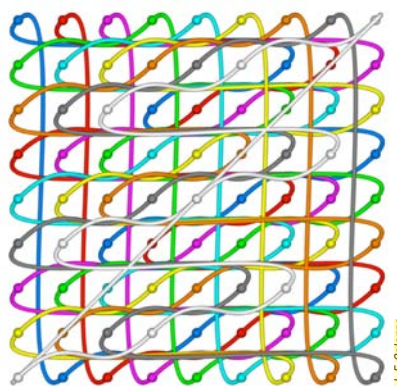
Le problème des données minimales est : *Quel est le plus petit nombre de données d'un problème de sudoku correct ?*

L'exemple à 17 données montre que ce minimum est 17 ou moins. On a longtemps cherché, sans succès, des problèmes corrects à 16 données ; on a donc conjecturé que 17 est la réponse.

Au moins 16 ou 17 cases ?

Le problème est resté ouvert jusqu'à ce que, en décembre 2011, la réponse soit fournie par Gary McGuire, de l'Université de Dublin, et son équipe. À la suite d'un calcul étalé sur une année et correspondant à 800 années de calcul d'un processeur unique, la conclusion fut qu'il n'existe aucune grille correcte de sudoku à 16 données et donc que le nombre minimal de données d'un problème correct de sudoku est 17.

Il s'agit d'un théorème mathématique, et puisque c'est la question mathématique concernant le sudoku qui a demandé le plus d'effort, on l'appellera « Le théorème du sudoku ». Le travail qui a conduit à l'énoncer est une démonstration. Comme cela se



VISUALISATION D'UN SUDOKU
par Jean-François Colonna. Chaque couleur est associée à un chiffre.

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