

*Les mathématiques sont un jeu
que l'on exerce selon des règles simples en manipulant
des symboles ou des concepts qui n'ont en soi,
aucune importance particulière.*

David Hilbert (1862-1943)

JOUER EFFICACEMENT AU

SUDOKU

Marc Laforest
École Polytechnique
de Montréal

Le paradoxe du Sudoku est que la simplicité de ses contraintes engendre tout de même un casse-tête d'une étonnante complexité. Selon la structure de la solution du Sudoku et des cases qui sont vides au début, on peut se retrouver soit avec un casse-tête taquinant le débutant, soit avec un casse-tête exaspérant l'expert. Bien qu'il soit souvent possible dans d'autres jeux, comme les échecs et les mots-croisés, de gagner une vue d'ensemble en mémorisant une fraction des combinaisons, en revanche, au Sudoku, l'analyse directe de toutes les combinaisons dépasse rapidement les facultés des plus doués. Toutefois, il existe un grand nombre de techniques élémentaires qui permettent de rapidement réduire le nombre de combinaisons admissibles. C'est grâce à ces astuces, et beaucoup de pratique, qu'en 2012 Jan Mrozowski a complété 10 Sudokus en 54 minutes et 4 secondes, le couronnant ainsi Champion du Monde du Sudoku.

À partir des règles du Sudoku, nous déduirons une famille de techniques qui nous permettront d'exclure certains candidats des cases vides. Ces **règles d'exclusion** sont parmi celles utilisées par les experts, mais nous ne nous attarderons qu'à un échantillon représentatif de ces techniques et à leur formulation mathématique. Toutefois, celles que nous présenteront devraient permettre au lecteur assidu de facilement compléter la majorité des Sudokus de nos hebdomadaires régionaux.

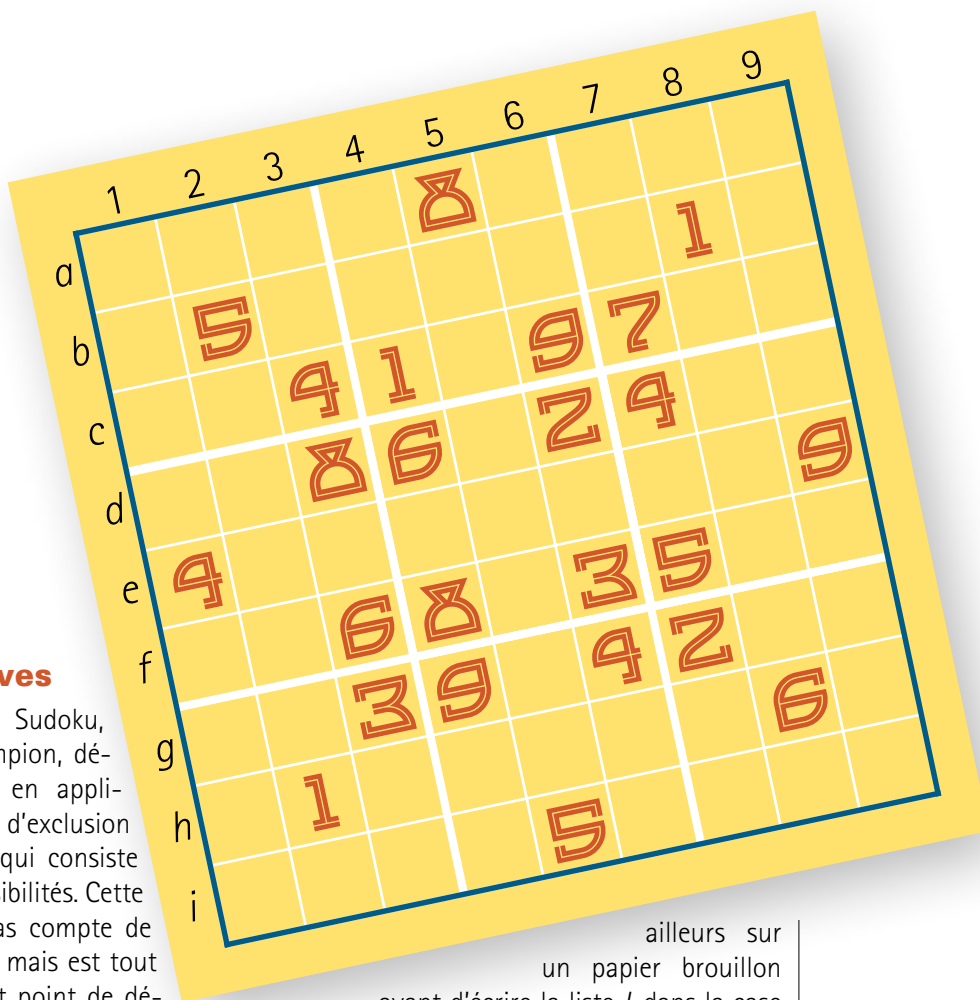
Les règles du Sudoku

Le jeu du Sudoku commence avec une figure formée de 81 cases organisées en 9 rangées et 9 colonnes. De plus, les cases sont aussi organisées en 9 blocs de 9 cases, circonscrit par les lignes plus foncées dans la figure. Il est coutume d'appeler Sudoku la figure d'un jeu de Sudoku et de dire que les blocs, les rangées, et les colonnes forment les **rangées virtuelles** du Sudoku. Une fois résolu, chaque case du Sudoku contient exactement un chiffre entre 1 et 9, mais les cases du Sudoku sont au départ majoritairement vides. L'objectif du Sudoku est de déduire les chiffres des cases vides de la **contrainte fondamentale** :

*Chaque rangée virtuelle doit
contenir les chiffres 1 à 9 une
et une seule fois.*

Malheureusement, cette prescription ne nous indique nullement comment procéder. Nous verrons comment réinterpréter cette contrainte en techniques déductives pour découvrir le contenu des cases vides.

Avant de continuer, mentionnons quelques conventions que les amateurs partagent. Les lettres de a à i sont utilisées pour énumérer les rangées du haut vers le bas. Les colonnes sont identifiées de la gauche vers la droite par un chiffre entre 1 et 9. Pour les blocs, les joueurs utilisent les trois lettres des rangées et les trois chiffres des colonnes qui se rencontrent au bloc en question.



Exclusions naïves

Tous les joueurs de Sudoku, du néophyte au champion, débent la résolution en appliquant une technique d'exclusion simple mais efficace qui consiste à exclure certaines possibilités. Cette technique ne tient pas compte de toutes les contraintes mais est tout de même un excellent point de départ pour l'application d'exclusions plus complexes.

Soit une case vide d'un Sudoku et les trois listes de chiffres connus appartenant respectivement à la rangée, la colonne, et le bloc qui se partagent la case vide. Appelons ces listes

- $R = \{\text{les chiffres connus de la rangée}\},$
- $C = \{\text{les chiffres connus de la colonne}\},$
- $B = \{\text{les chiffres connus du bloc}\}.$

À partir uniquement de cette information, la contrainte fondamentale nous permet seulement de conclure que l'ensemble L des chiffres qui sont candidats de la case vide, ou plus brièvement l'ensemble des **candidats** L de la case vide, est précisément l'ensemble des chiffres autres que ceux déjà connus, donc $L = \overline{R \cup C \cup B}$. L'objectif de l'exclusion naïve est d'exclure de $\{1, 2, \dots, 9\}$ les chiffres qui apparaissent déjà dans R, C ou B et d'écrire la courte liste L dans la case vide. Selon $L = \overline{R \cup C \cup B}$, le calcul de L doit commencer avec le calcul de l'union $R \cup C \cup B$ avant de passer à l'identification des chiffres de 1 et 9 qui n'appartiennent pas à $R \cup C \cup B$. Ceci exigerait que l'on écrive l'ensemble $R \cup C \cup B$

ailleurs sur un papier brouillon avant d'écrire la liste L dans la case vide, une procédure très peu commode.

La loi de De Morgan offre une alternative calculatoire que vous aurez peut-être déjà devinée. Appliquant cette loi à notre expression précédente pour la liste L des candidats, on déduit que

$$L = \overline{R \cup C \cup B} = \bar{R} \cap \bar{C} \cap \bar{B}.$$

Notions de la théorie des ensembles

Soit un ensemble A , et deux sous-ensembles B et C de A . On écrit $B \subset A$ et $C \subset A$. Il existe plusieurs opérations élémentaires que l'on peut effectuer sur ces ensembles.

Le *complément* de B dans A , noté \bar{B} , est le sous-ensemble des éléments de A qui ne sont pas dans B .

L'*union* de B et de C , noté $B \cup C$, est l'ensemble des éléments qui appartiennent soit à B , soit à C .

L'*intersection* de B et C , noté $B \cap C$, est l'ensemble des éléments qui appartiennent à la fois à B et à C .

Ces opérations sont liées entre elles par les *lois de De Morgan*, qui sont en fait complètement évidentes

$$\overline{B \cup C} = \bar{B} \cap \bar{C} \quad \text{et} \quad \overline{B \cap C} = \bar{B} \cup \bar{C}.$$

La convention est de désigner l'ensemble vide par le symbole \emptyset .

Cette nouvelle expression pour L indique qu'il est possible de calculer L en vérifiant systématiquement, pour les chiffres de 1 à 9, qu'ils n'appartiennent ni à R , ni à C , et ni à B . Pour chaque chiffre de 1 à 9, une telle vérification peut se faire visuellement en survolant les trois rangées virtuelles, et si le chiffre appartient à L , on peut immédiatement l'écrire en petit caractères dans la case vide avant de passer au prochain chiffre de la liste. On résume donc cette technique d'exclusion naïve de la manière suivante.

Pour chaque case vide, et pour chaque chiffre entre 1 et 9, écrivez ce chiffre dans la case vide s'il n'apparaît dans aucune des trois rangées virtuelles auquel appartient la case vide.

Le résultat de cette procédure est un Sudoku semblable à celui de la figure 1. C'est une opération répétitive mais simple et que l'on gagne à pratiquer. Pour la durée du reste de cet article, nous prendrons pour acquis que tout joueur de Sudoku débute la résolution de son Sudoku avec cette technique.

	1	2	3	4	5	6	7	8	9
a	1, 2, 3, 6, 7, 9	2, 3, 6, 7, 9	1, 2, 7, 9	2, 3, 4, 6, 7	8	5, 6, 7	3, 6, 9	2, 3, 4, 5, 9	2, 3, 4, 5, 6
b	2, 3, 6, 7, 8, 9	5	2, 7, 9	2, 3, 4, 7	2, 3, 4, 6, 7	6, 7	3, 6, 8, 9	1	2, 3, 4, 6, 8
c	2, 3, 6, 8	2, 3, 6, 8	4	1	2, 3, 6	9	7	2, 3, 5, 8	2, 3, 5, 6, 8
d	1, 3, 5, 7, 9	3, 7, 9	8	6	1, 7, 9	2	4	3, 7	1, 3, 7
e	4	2, 3, 7	1, 2, 5, 7	5, 7	1, 7	1, 5, 7	1, 3, 6, 8	2, 3, 7, 8	9
f	1, 2, 7, 9	2, 7, 9	6	8	1, 4, 7, 9	3	5	2, 7	1, 2, 7
g	5, 6, 7, 8	6, 7, 8	3	9	1, 6, 7	4	2	5, 7, 8	1, 5, 7, 8
h	2, 5, 7, 8, 9	1	2, 5, 7, 9	2, 3, 7	2, 3, 7	7, 8	3, 8, 9	6	3, 4, 5, 7, 8
i	2, 6, 7, 8, 9	2, 4, 6, 7, 8, 9	2, 7, 9	2, 3, 7	5	1, 6, 7, 8	1, 3, 8, 9	3, 4, 7, 8, 9	1, 3, 4, 7, 8

figure 1 : Un Sudoku avec la liste des candidats obtenus à partir de l'exclusion naïve.

Exclusions nues et cachées

Quand vous êtes chanceux, l'exclusion naïve produit dans chaque case vide une liste ne contenant

Simplet nu
Une case avec un seul candidat s'appelle un simplet nu.

qu'un seul candidat, ce qui s'appelle un **simplet nu**. Toutefois, il existe très souvent une autre situation permettant d'identifier immédiatement le chiffre caché d'une case. Si un candidat d'une case ne se retrouve dans aucune autre case d'une des rangées virtuelles contenant la case, alors ce chiffre est la seule possibilité pour cette case, une situation nommée un **simplet caché**.

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

figure 2 : Le 2 est un simplet nu en a5 et le 5 est un simplet caché en a7.

La notion de simplet caché s'explique bien à l'aide d'un exemple. Dans la rangée de la figure 2, les candidats des cases vides sont

$$C_1 = \{2, 9\}, C_3 = \{2, 8, 9\}, C_4 = \{1, 2, 8\}, \\ C_5 = \{2\}, C_7 = \{1, 2, 5, 8, 9\}.$$

Si l'on regarde la septième case, on remarque que le chiffre 5 n'est candidat dans aucune autre case. Mathématiquement, ceci se résume en écrivant que 5 n'appartient pas à l'ensemble $C_1 \cup C_3 \cup C_4 \cup C_5$ mais qu'il appartient quand même à C_7 ,

$$\{5\} = \overline{(C_1 \cup C_3 \cup C_4 \cup C_5)} \cap C_7.$$

Selon ce raisonnement, tous les simplets cachés du Sudoku peuvent être identifiés en visitant chaque case vide et, pour chaque candidat de la case, en vérifiant que le candidat ne se retrouve pas ailleurs dans chaque rangée virtuelle de la case.

Voici une deuxième manière beaucoup plus rapide d'identifier tous les simplets cachés, que nous justifierons plus loin,

Pour chaque rangée virtuelle et pour chaque chiffre entre 1 et 9, identifiez s'il est candidat d'une seule case. Si c'est le cas, il doit être le chiffre caché de la case.

En examinant l'application de cette méthode à l'exemple précédent, on peut comprendre son équivalence avec celle suggérée par l'expression initiale. Pour la rangée de la figure 2, supposons que l'on cherche à savoir si 5 est un simplet caché quelque part dans la rangée. En un coup d'oeil, on note les cases où 5 est candidat,

$$\begin{aligned} \{5\} \cap C_1 &= \emptyset \Leftrightarrow 5 \in \overline{C_1} \\ \{5\} \cap C_3 &= \emptyset \Leftrightarrow 5 \in \overline{C_3} \\ \{5\} \cap C_4 &= \emptyset \Leftrightarrow 5 \in \overline{C_4} \\ \{5\} \cap C_5 &= \emptyset \Leftrightarrow 5 \in \overline{C_5} \\ \{5\} \cap C_7 &= \{5\} \Leftrightarrow 5 \in C_7 \end{aligned}$$

On déduit alors que le chiffre 5 appartient aux intersections

$$\overline{C_1} \cap C_7, \overline{C_3} \cap C_7, \overline{C_4} \cap C_7, \overline{C_5} \cap C_7.$$

À l'aide de la loi de De Morgan, on conclut que 5 est un simplet caché de la septième case, car

$$\begin{aligned} \{5\} &= (\overline{C_1} \cap C_7) \cap (\overline{C_3} \cap C_7) \cap (\overline{C_4} \cap C_7) \cap (\overline{C_5} \cap C_7) \\ &= \overline{(C_1 \cup C_3 \cup C_4 \cup C_5)} \cap C_7. \end{aligned}$$

En conclusion, les deux approches diffèrent dans l'ordre dans lequel la recherche se fait et dans les questions qu'on doit se poser, mais le résultat est le même.

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

figure 3 : Les chiffres {1, 4, 8} forment un triplet caché en a_1, a_5, a_6 .

Les notions de simplet nu et de simplet caché ont des notions cousines de paires, triplets et plus généralement de k -tuplet, nus ou cachés.

Un k -tuplet nu d'une rangée virtuelle est un ensemble T de k chiffres pour lequel il existe k cases vides dont les candidats satisfont

$$C_1 \subset T, C_2 \subset T, \dots, C_k \subset T.$$

Étant donné que les chiffres du k -tuplet nu sont les seuls occupants possibles de k cases, ils ne peuvent pas apparaître dans les autres cases et on peut **exclude** les membres de T de candidats de toutes les autres cases.

Un k -tuplet caché d'une rangée virtuelle avec $k + j$ cases vides est un ensemble T de k chiffres pour lequel les $k + j$ listes de candidats $C_1, \dots, C_k, C_{k+1}, \dots, C_{k+j}$ peuvent s'énumérer de manière à ce que

$$T \subset C_1 \cup \dots \cup C_k.$$

$$T \cap (C_{k+1} \cup \dots \cup C_{k+j}) = \emptyset.$$

Étant donné que les k chiffres du k -tuplet caché doivent occuper les k premières cases vides, on peut exclure tous les autres chiffres $C_{k+1} \cup \dots \cup C_{k+j}$ des listes C_1, \dots, C_k . Contrairement aux notions de simplet nu ou caché, l'objectif n'est pas de découvrir la position de certains chiffres mais plutôt de dire où ils ne peuvent pas apparaître. De plus chaque exclusion peut avoir un effet boule de neige et vous mener à découvrir de nouveaux k -tuplets nus ou cachés. Voici la méthodologie à suivre si vous désirez identifier ces exclusions.

1. Procédez une rangée virtuelle à la fois.
2. Éliminez les chiffres des simplets nus et cachés des autres cases des rangées virtuelles se partageant les cases des simplets identifiés.
3. Déterminez visuellement s'il existe deux cases vides avec des listes identiques de deux candidats. Éliminez ces doublets nus de toutes les autres listes de candidats.
4. Parmi les listes contenant 3 candidats ou plus, cherchez deux chiffres qui n'apparaissent que dans deux cases. Éliminez de vos deux cases les chiffres autres que ceux du doublet caché.
5. Continuez successivement avec les k -tuplets nus et cachés, $k \geq 3$.

Exclusions de Jedi

Dans cette configuration du Sudoku, la force est la logique.

Soit quatre cases vides à l'intersection de deux rangées et de deux colonnes. On dit qu'ils forment une aile-X¹ s'il existe un chiffre n qui n'est candidat dans les deux rangées qu'à l'intersection des deux colonnes.

Bien que la définition d'une aile-X exclue déjà la présence de n dans les autres cases des deux rangées, comme nous expliquerons plus tard, cette configuration exclut aussi n des autres cases des deux colonnes. On aurait aussi pu définir une aile-X en inversant le rôle des rangées et des colonnes et l'exclusion aurait eu lieu dans les autres cases des deux rangées.

L'exclusion de Jedi s'explique aisément dans un exemple concret comme celui du Sudoku de la figure 4. À l'intersection des rangées c , h et des colonnes 2, 7, on observe que le chiffre

1. L'expression aile-X est une traduction de X-Wing, le célèbre vaisseau de Luke Skywalker dans la série Star Wars.

	1	2	3	4	5	6	7	8	9
a	1, 2, 6	4	8	7	9	3, 6	1, 2, 3, 5	1, 2, 3, 5	2, 3,
b	1, 6, 9	5	1, 3	8	2	3, 6	7	1, 3, 4	3, 4, 9
c	2, 9	2, 3, 9	7	5	4	1	2, 3, 9	6	8
d	3	8	5	2	1	9	4	7	6
e	7	6	2	3	5	4	8	9	1
f	4	1	9	6	7	8	2, 3	2, 3	5
g	8	7	6	4	3	5	1, 2, 9	1, 2	2, 9
h	1, 5, 9	3, 9	4	1, 9	6	2	1, 3, 5, 9	8	7
i	1, 2, 5, 9	2, 3, 9	1, 3	1, 9	8	7	6	3, 4, 5	3, 4

figure 4 : Une exclusion de Jedi

3 n'est candidat que dans deux cases des rangées c et h . Soit que 3 appartienne à $c2$, soit qu'il appartienne à $c7$. Dans le premier cas, le 3 est alors exclu de $c7$ et de toutes les autres cases de la colonne 2. Dans la rangée h , le 3 ne peut donc plus appartenir à $h2$ et il doit donc appartenir à $h7$. La décision de placer le 3 en $c2$, permet de l'exclure de toutes les cases des colonnes 2 et 7 sauf $c2$ et $h7$. Vice versa, si le chiffre 3 était d'abord placé dans la case $c7$, alors une analyse semblable démontrerait que 3 ne peut appartenir à aucune des cases des colonnes 2 et 7 sauf $c7$ et $h2$. En conclusion, la présence d'une aile-X exclut le 3 des listes de candidats aux cases $i2$, $a7$ et $f7$. La valeur d'une exclusion dépend en grande partie de la facilité avec laquelle elle peut être utilisée. Pour une exclusion de ce type, la technique est identique à sa formulation.

Dans chaque rangée, identifiez les chiffres qui ne sont candidats que dans deux cases. Pour chaque chiffre, vérifiez s'il est candidat à l'intersection des deux colonnes avec une autre rangée. Si ce chiffre n'est pas candidat ailleurs dans cette nouvelle rangée, alors vous pouvez l'exclure des autres cases des deux colonnes.

Les joueurs aguerris du Sudoku cherchent aussi pour la présence d'un chiffre n qui soit candidat à l'intersection de k rangées et de k colonnes, mais pas ailleurs sur ces k rangées. Comme auparavant, cette configuration permet d'exclure le chiffre n des candidats dans toutes les cases des k colonnes à l'extérieur des k rangées. En anglais, la configuration avec $k = 3$ se nomme *Swordfish* et avec $k = 4$, elle se nomme *Jellyfish*. Toutefois, une configuration avec $k \geq 5$ est sans intérêt car il est possible de montrer que les autres $9 - k \leq 4$ colonnes et rangées formeront une exclusion de Jedi nécessairement plus simple.

La structure de Tuleja

Plutôt que de présenter une dernière technique d'exclusion, cet article se terminera avec la présentation d'un résultat général de Gregory Tuleja, directeur du Williston Northhampton School, concernant la structure de la solution d'un Sudoku.

Théorème de Tuleja

Supposons qu'un joueur ait réussi à remplir toutes les cases vides du casse-tête initial, tout en satisfaisant la contrainte fondamentale, et qu'il ou elle admire joyeusement son œuvre.

Le contenu d'un triplet horizontal de blocs possède l'une des deux propriétés suivantes.

i) Il existe une subdivision de

$$\{1, 2, \dots, 9\}$$

en trois ensembles C_1, C_2, C_3 de trois chiffres, telle que chaque triplet de cases horizontales d'un bloc est l'un de ces trois ensembles.

ii) Il existe une subdivision de

$$\{1, 2, \dots, 9\}$$

en quatre ensembles, les trois premiers C_1, C_2, C_3 contiennent chacun deux chiffres et le dernier contient les trois derniers chiffres.

$$L = \overline{C_1 \cup C_2 \cup C_3}.$$

Chaque triplet de cases horizontales d'un bloc est l'union d'un des ensembles C_i et d'un des chiffres de L .

Grâce à la symétrie des Sudokus, il est évident que ce résultat s'applique autant aux triplets horizontaux de blocs qu'aux triplets verticaux de blocs. Dans la figure 5, le premier cas s'applique aux blocs des rangées abc . Plus précisément, dans le bloc abc , on a

$$C_1 = \{1, 2, 3\}, C_2 = \{4, 5, 6\}, C_3 = \{7, 8, 9\}$$

et le triplet C_1 apparaît en $a123$, $c456$ et $b789$. Le deuxième cas s'applique aux triplet de blocs des rangées def .

Ce très joli résultat peut aussi être utilisé dans la résolution de Sudoku bien qu'il ne se prête pas à une démarche algorithmique comme les techniques précédentes.

	1	2	3	4	5	6	7	8	9
a	1	2	3	4	5	6	7	8	9
b	6	4	5	7	9	8	3	2	1
c	9	8	7	3	1	2	5	4	6
d	8	5	9	6	7	3	4	1	2
e	4	1	6	2	8	5	9	3	7
f	7	3	2	9	4	1	8	6	5
g	5	6	4	8	2	9	1	7	3
h	3	7	1	5	6	4	2	9	8
i	2	9	8	1	3	7	6	5	4

figure 5 – Une illustration du théorème de Tuleja