Sudoku puzzles and how to solve them

Andries E. Brouwer

2006-05-31

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

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

Figure 1: Two puzzles—the second one is difficult

1 Sudoku

A Sudoku puzzle (of 'classical type') consists of a 9-by-9 matrix partitioned into nine 3-by-3 submatrices ('boxes'). Some of the entries are given, and the puzzle is to find the remaining entries, under the condition that the nine rows, the nine columns, and the nine boxes all contain a permutation of the symbols of some given alphabet of size 9, usually the digits 1-9, or the letters A-I.

Some mathematicians will claim that since this is a finite problem, it is trivial. The time needed to solve a Sudoku puzzle is $O(1)$ - indeed, one can always try the 9^{81} possible ways of filling the grid. But one can still ask for efficient ways of finding a solution. Or, if one knows the solution already, one can ask for a sequence of logical arguments one can use to convince someone else of the fact that this really is the unique solution.

1.1 Backtrack and elegance

It is very easy to write an efficient computer solver. Straightforward backtrack search suffices, and Knuth's 'dancing links' formulation of the backtrack search for an exact covering problem takes a few microseconds per puzzle on common hardware today.

For a human solver, backtrack is the last resort. If all attempts at further progress fail, one can always select an open square, preferably with only a few possibilities, and try these possibilities one by one—maybe using pencil and eraser, or maybe copying the partially filled diagram to several auxiliary sheets of paper and trying each possibility on a separate sheet of paper.

For very difficult Sudoku puzzles, this is the fastest way to solve them, also for humans.

However, one solves puzzles not because the answer is needed, but for fun, in order to exercise one's capabilities in logical reasoning. And solving by backtrack is dull, boring, mindless, no thinking required, better left to a computer, no fun at all.

So, Backtrack, or Trial & Error, is taboo. And if it cannot be avoided one prefers some limited form. Maybe whatever can be done entirely in one's head.

1.2 Grading

Most Sudoku puzzles one meets are computer-produced, and it is necessary to have a reasonable estimate of the difficulty of these puzzles. To this end one needs computer solvers that mimic human solvers. Thus, one would also implement the solving steps described below in a Sudoku solving program, not in order to find the solution as quickly as possible, but in order to judge the difficulty of the puzzle, or in order to be able to give hints to a human player. Such AI-type Sudoku solving programs tend to be a thousand to a million times slower than straightforward backtrack.

1.3 Generating

Given the backtrack solver, generating Sudoku puzzles is easy: start with an empty grid, and each time the backtrack solver says that the solution is not unique throw in one more digit. (If now there is no solution anymore, try a different digit in the same place.) To generate a puzzle in this way requires maybe thirty calls to the backtrack solver, less than a millisecond. One can polish the puzzle a little by checking that none of the givens is superfluous.

Afterwards one feeds the puzzles that were generated to a grader. Maybe half will turn out to be very easy, and most will be rather easy ('humanly solvable'). It is very difficult to generate very difficult puzzles, puzzles that are too difficult even for very experienced humans.

2 Solving

Below we sketch a possible approach for a human solver. The goal is to be efficient. In particular, the boring and time-consuming action of writing all possibilities in every empty square is postponed as much as possible. On the other hand, some form of markup helps.

Baby steps

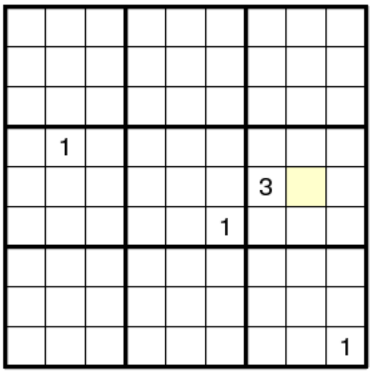
When eight digits in some row or column or box are known, one can find the last missing digit.

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

Exercise (i) Solve this puzzle using baby steps only. (ii) Show that if a puzzle can be solved using baby steps only, it has at most 21 open squares.

Singles

When there is only one place for a given digit in a given row or column or box, write it there. If there is only one digit that can go in a given square, write it there.



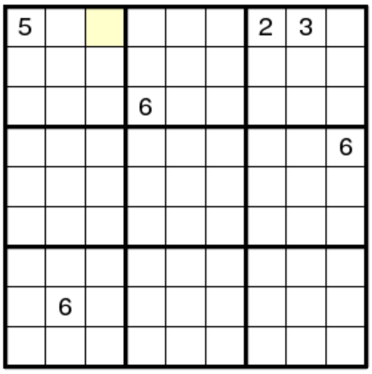


Figure 2: The 1 in the middle right box must be in the yellow square.

Figure 3: The 6 in the top row must be in the yellow square.

Baby steps are particularly easy cases of singles. Checking for singles requires 324 steps. Knuth's 'dancing links' backtracker will take 324 steps if and only if the puzzle can be solved by singles only. It is unknown how many open squares a puzzle can have and be solvable by singles only. There are examples with 17 givens. It is unknown whether any Sudoku puzzles exist with 16 givens and a unique solution.

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

5		2				4		
			7	1				3
					4	6		
	7		2					
	1							
6					2			
				3			1	
4								

Figure 4: Solve using singles only

Figure 5: 16 clues and 2 solutions

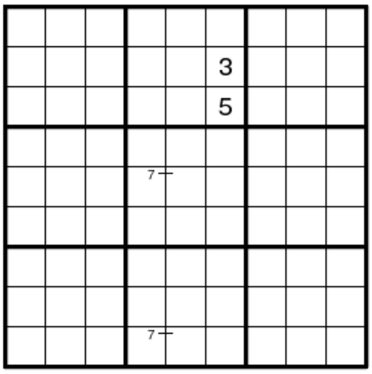
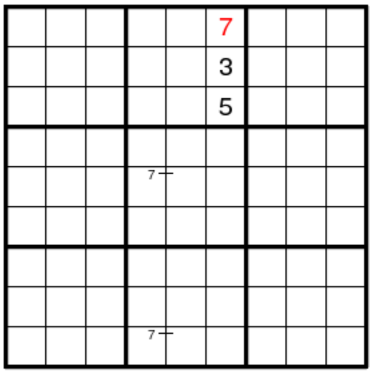
Pair markup

If one checks where a given digit can go in some row or column or box and finds that there are precisely two possibilities, then it helps to note this down. That is efficient, one does not do the same argument over and over again, and helps in further reasoning.

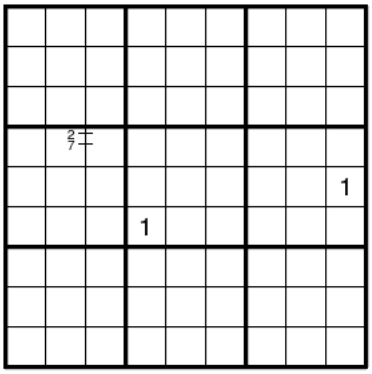
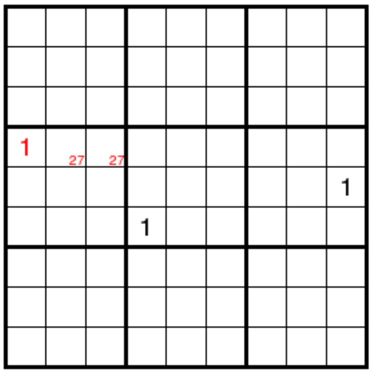
1	2							
								7
3	4							
5	6					3		
8	9					4	5	

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

Given two pairs for the same digit straddling the same two rows or columns, the digit involved cannot occur elsewhere in those rows or columns.

Given two pairs between the same two squares, one concludes that these two squares only contain the two digits involved and no other digit.

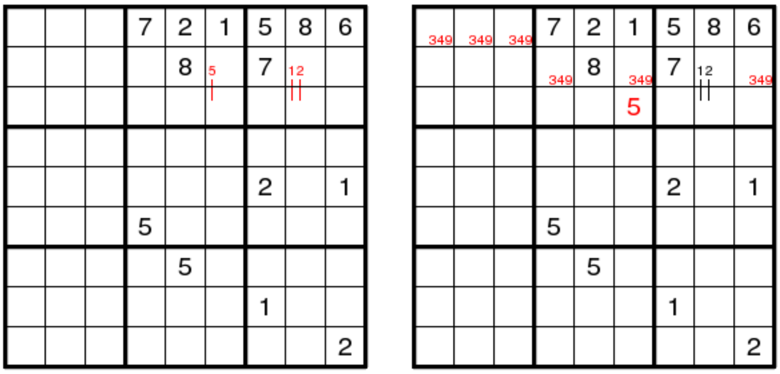
 

Matchings

A sudoku defines 36 matchings (1-1 correspondences) of size 9: between positions and digits given a single row or column or box, and between rows and columns given a single digit. For each of these 1-1 correspondences between sets X and Y of size 9, if we have identified subsets $A \subset X$ and $B \subset Y$ of the same size n, $1 \leq n \leq 9$, such that we know that the partners of every element of B must be in A, then A and B are matched, and nothing else has a partner in A. More explicitly:

(A) If for some set of n positions in a single row, column, or box there are n digits that can be only at these positions, then these positions do contain these digits (and no other digits).

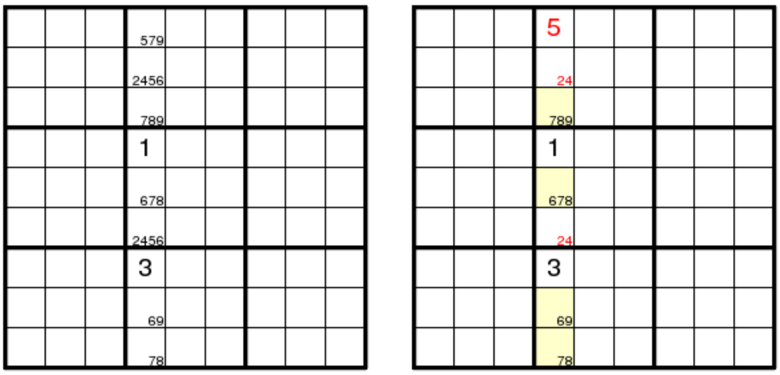
For example,



The three digits 3, 4, 9 in the second row must be in columns 4, 6, 9, so the digit 5 cannot be there.

(B) If for some set of n digits there are n positions in a single row, column, or box, that cannot contain any digits other than these, then these digits must be at those positions (and not elsewhere in the same row, column, or box).

For example,

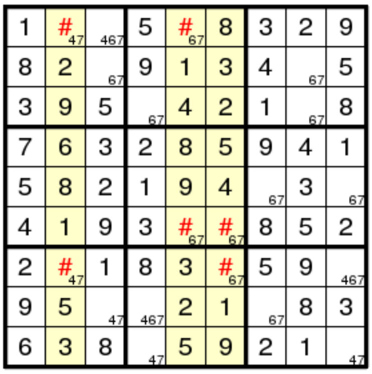
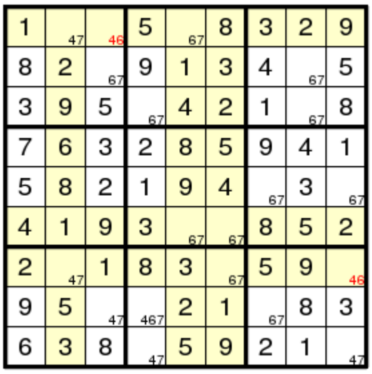


Here all possibilities for the fields in column 4 are given. Note the four yellow fields: together, they only have the four possibilities 6, 7, 8, 9. So, these yellow fields contain 6, 7, 8, 9 in some order, and we can remove 6, 7, 8, 9 from the possibilities of the other fields in that column.

(C) Pick a digit d. If for some set of n rows R there is a set of n columns C such that all occurrences of d in these rows must be in one of the columns in C, then the digit d does not occur in a column in C in a row not in R (and the same with rows and columns interchanged).

(This argument is called *X-wing* for $n = 2$, *Swordfish* for $n = 3$, *Jellyfish* for $n = 4$.)

For example,

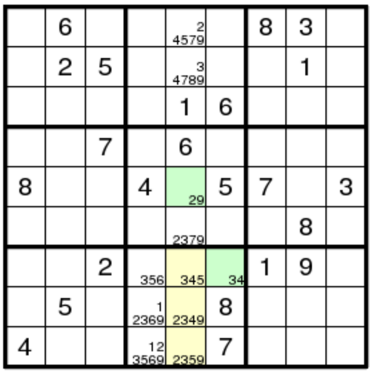
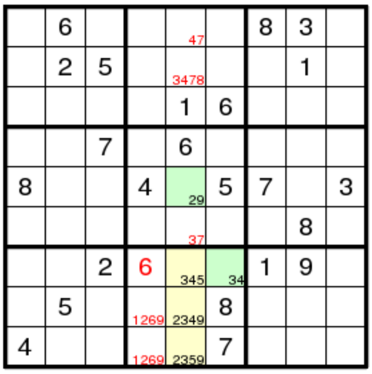
For digit 7, the only possibilities in columns 2, 5, 6 do occur in rows 1, 6, 7. Therefore, digit 7 cannot occur outside columns 2, 5, 6 in these rows.

Exercise Complete this Sudoku.

The subset principle

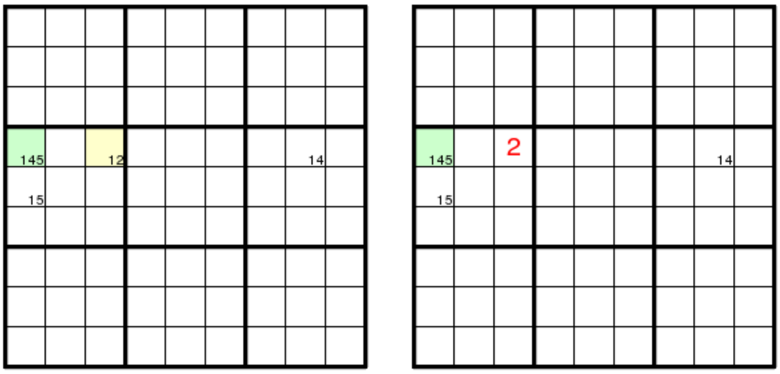
Let S be a subset of the set of cells of a partially filled Sudoku diagram, and let for each digit d the number of occurrences of d in S be at most n_d. If $\sum n_d = |S|$, then the situation is tight: each digit d must occur precisely n_d times in S. In this case we can eliminate a digit d from the candidates of any cell C such that the presence of a d in C would force the number of d's in S to be less than n_d.

For example,

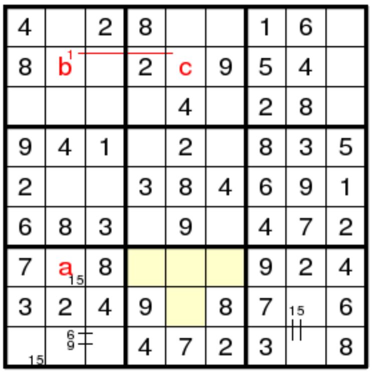
In the five colored squares, the five digits 2,3,4,5,9 each occur at most once (since all occurrences of 3,4,5 are in a single box and all occurrences of 2,5,9 in a single column). Since the situation is tight, digits 3,4,5 do not occur elsewhere in this box, and digits 2,5,9 do not occur elsewhere in this column.

More generally, one may remove a candidate for a cell outside S if its presence would force $\sum n_d < |S|$.



Hinge

The previous subsection used that each cell contains at least one digit. Conversely, each digit is in at least one cell in any given row, column or box. For example,



Consider the above diagram. If cell **a** has a 1, then cell **b** does not have a 1, and then the 1 in row 2 must be in cell **c**. But then the yellow area cannot contain a 1, impossible. (So, cell **a** has a 5.)

Forcing Chains

Consider propositions $(i, j)d$ 'cell (i, j) has value d' and $(i, j)!d$ 'cell (i, j) has a value different from d'. By observing the grid one finds implications among such propositions.

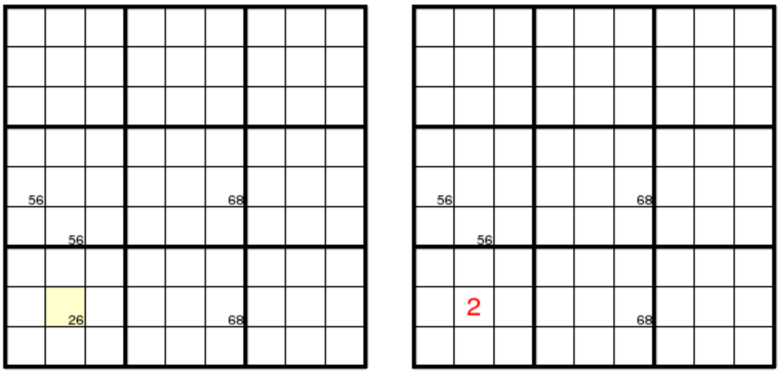
There are at least three obvious types of such implications. Let us say that two cells are 'adjacent' (or, 'see each other') when they lie in the same row, column or box, so that they must contain different digits. This gives the first type: If (i, j) is adjacent to (k, l) then $(i, j)d > (k, l)!d$ where $>$ denotes implication.

In case (i, j) is adjacent to (k, l) and (k, l) only has the two possibilities d and e, then $(i, j)d > (k, l)e$. This is the second type of implication.

Finally, if some row, column or box has only two possible positions (i,j) and (k,l) for some digit d, then $(i,j)!d > (k,l)d$. This is the third type.

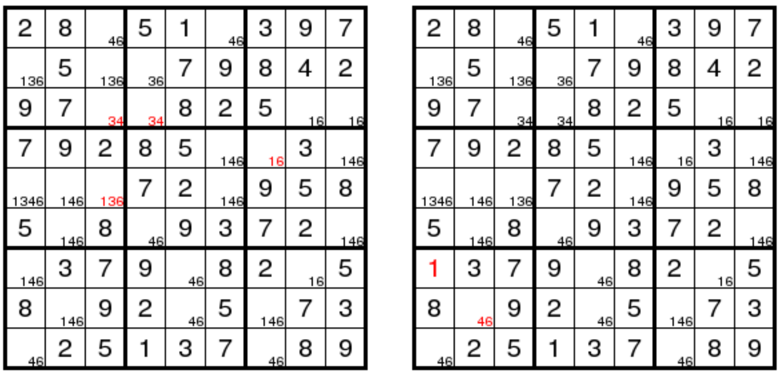
Consider chains of implications. If $(i,j)d > \ldots > (i,j)!d$ then $(i,j)!d$.

For example



Here $(8,2)6 > (8,6)8 > (5,6)6 > (5,1)5 > (6,2)6 > (8,2)2$ was used to conclude $(8,2)2$.

For example



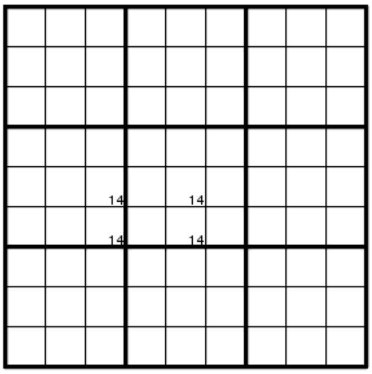
(Check the given possibilities in red!) Here $(8,2)1 > (6,2)!1 > (6,9)1 > (4,7)!1 > (8,7)1 > (8,2)!1$ was used to conclude $(8,2)!1$. For such chains that involve a single digit only, and where the implication types alternate between I and III, one often uses a simplified notation like $1:\ (8,2)-(6,2)=(6,9)-(4,7)=(8,7)-(8,2)$ where $-$ denotes that at most one is true and $=$ that at least one is true.

Finding useful chains may be nontrivial, and there are various techniques such as ‘colouring’ that help.

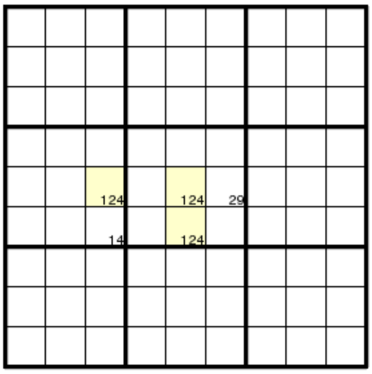
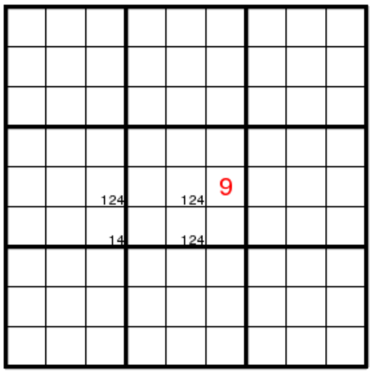
Uniqueness

A properly formulated Sudoku puzzle has a unique solution. One can assume that a given puzzle actually is properly formulated, and use that in the reasoning, to exclude branches that would not lead to a unique solution.

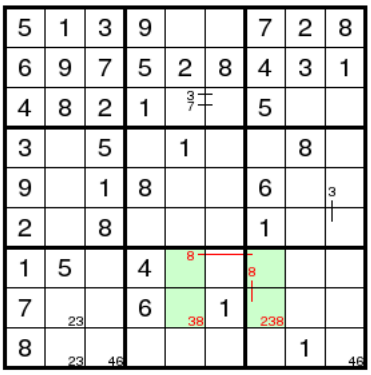
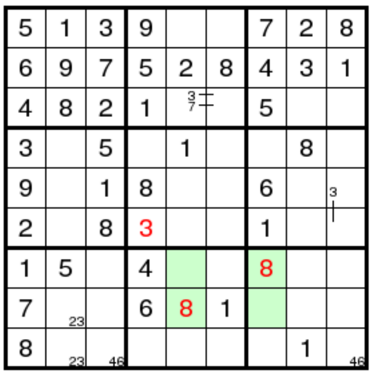
For example, a grid



can be completed in at least two ways, violating the uniqueness assumption. That means that this must be avoided, so that

At least one of the corners of the rectangle is 2.
For example,

Look at the green rectangle. If $(7,7)3$, then $(7,5)8$ and $(8,7)8$ so that $(8,5)3$ and we have a forbidden rectangle with pattern 83-38. So, $(7,7)!3$, which means that we have an X-wing: digit 3 in columns 2,7 can only be in rows 8,9 and does not occur elsewhere in these rows. In particular $(9,4)!3$ so that $(6,4)3$, and $(8,5)!3$ so that $(8,5)8$.

More generally:

Theorem *Suppose one writes some (more than 0) candidate numbers in* *some of the initially open cells of a given Sudoku diagram, 0 or 2 in each cell, such that each value occurs 0 or 2 times in any row, column, or box. Then this Sudoku diagram has an even number of completions that agree with at least one of the candidates in each cell where candidates were given. In particular, if the Sudoku diagram has a unique solution, then that unique solution differs from both candidates in at least one cell.*

Digit patterns and jigsaw puzzles

A more global approach was described by `Myth Jellies`. Solve a puzzle until no further progress is made. Then, for each of the nine digits, write down all possible solution patterns for that digit. One hopes to find not more than a few dozen patterns in all. Now the actual solution has one pattern for each digit, where these 9 patterns partition the grid. Regard each digit pattern ('jigsaw piece') as a boolean formula ('this pattern occurs in the solution'). Write down the formulas that express that for each of the nine digits exactly one pattern occurs, and that overlapping pieces cannot both be true. Solve the resulting system of propositional formulas.

This approach allows one to solve some otherwise unapproachable puzzles.

References

[1] There is an enormous amount of literature on Sudoku on the web. This note is a condensed version of `http://homepages.cwi.nl/~aeb/games/sudoku/` .