SAT Solving Sudoku

Aresh Pourkavoos

March 17, 2022

Sudoku is a puzzle game which consists of a 9×9 square grid, subdivided into 3×3 blocks and partially filled with the digits 1-9. The goal is to fill in the rest of the digits such that every row, column and block contains each digit exactly once. Sudoku is a well-studied problem, and there are many algorithms out there to solve it, from pencil-and-paper tricks used mostly by human players to trial-and-error methods used by computers, such as backtracking. However, I will focus on one particular approach: SAT solving.

SAT is short for "satisfiability," and a SAT solver is a program that checks whether a given statement is satisfiable. In this case, a statement is a logical proposition such as "P and Q", and a statement is satisfiable if there is some way to assign each variable (P and Q, in this case) to either true or false such that the entire statement is true. "P and Q," for example, is satisfiable: it has exactly one solution, where P and Q are both true. "P and not P," on the other hand, is not satisfiable, since for both possible values of P, the statement is false. A brute-force SAT solver would simply check every possible combination, of which there are 2^n, where n is the number of variables. However, there are much more optimized algorithms in use, making SAT solvers practical for huge formulas with thousands of variables or more.

The following table shows a way to formulate the statement that at most one of 8 variables $a_0, \ldots, a_7$ is true, using a binary divide-and-conquer method.

a_0, a_1 free$\vert$	a_2, a_3 free$\vert$	a_4, a_5 free	a_6, a_7 free$\vert$
$\underline{\neg a_0 \vee \neg a_1\vert}$	$\neg a_2 \vee \neg a_3\vert$	$\neg a_4 \vee \neg a_5$	$\neg a_6 \vee \neg a_7\vert$
$a_{0-1} := a_0 \vee a_1$	$a_{2-3} := a_2 \vee a_3\vert$	$a_{4-5} := a_4 \vee a_5$	$a_{6-7} := a_6 \vee a_7\vert$
$\neg a_{0-1} \vee a_0 \vee a_1$	$\neg a_{2-3} \vee a_2 \vee a_3\vert$	$\neg a_{4-5} \vee a_4 \vee a_5$	$\neg a_{6-7} \vee a_6 \vee a_7\vert$
$\neg a_0 \vee a_{0-1}$	$\neg a_2 \vee a_{2-3}\vert$	$\neg a_4 \vee a_{4-5}$	$\neg a_6 \vee a_{6-7}\vert$
$\neg a_1 \vee a_{0-1}$	$\neg a_3 \vee a_{2-3}\vert$	$\neg a_5 \vee a_{4-5}$	$\neg a_7 \vee a_{6-7}\vert$
	$\underline{\neg a_{0-1} \vee \neg a_{2-3}\vert}$		$\neg a_{4-5} \vee \neg a_{6-7}\vert$
	$a_{0-3} := a_{0-1} \vee a_{2-3}$		$a_{4-7} := a_{4-5} \vee a_{6-7}\vert$
	$\neg a_{0-3} \vee a_{0-1} \vee a_{2-3}$		$\neg a_{4-7} \vee a_{4-5} \vee a_{6-7}\vert$
	$\neg a_{0-1} \vee a_{0-3}$		$\neg a_{4-5} \vee a_{4-7}\vert$
	$\neg a_{2-3} \vee a_{0-3}$		$\neg a_{6-7} \vee a_{4-7}\vert$
			$\underline{\neg a_{0-3} \vee \neg a_{4-7}\vert}$

i	Free vars	Aux vars	Formulas
1	2	0	1
2	4	2	9
3	8	6	25

At level i of the tree, there are $n = 2^i$ free variables, $n-2$ auxiliary variables (for a total of $2n-2$ variables),

and $4n-7$ formulas. In contrast, the naive method with no auxiliary variables uses $1+(n^2-n)/2$ formulas.

$$
\begin{array}{r|r|r}
a_0, a_1, a_2 \text{ free} & a_3, a_4, a_5 \text{ free} & a_6, a_7, a_8 \text{ free} \\
\neg a_0 \vee \neg a_1 & \neg a_3 \vee \neg a_4 & \neg a_6 \vee \neg a_7 \\
\neg a_0 \vee \neg a_2 & \neg a_3 \vee \neg a_5 & \neg a_6 \vee \neg a_8 \\
\underline{\neg a_1 \vee \neg a_2} & \underline{\neg a_4 \vee \neg a_5} & \underline{\neg a_7 \vee \neg a_8} \\
a_{0-2} := a_0 \vee a_1 \vee a_2 & a_{3-5} := a_3 \vee a_4 \vee a_5 & a_{6-8} := a_6 \vee a_7 \vee a_8 \\
\neg a_{0-2} \vee a_0 \vee a_1 \vee a_2 & \neg a_{3-5} \vee a_3 \vee a_4 \vee a_5 & \neg a_{6-8} \vee a_6 \vee a_7 \vee a_8 \\
\neg a_0 \vee a_{0-2} & \neg a_3 \vee a_{3-5} & \neg a_6 \vee a_{6-8} \\
\neg a_1 \vee a_{0-2} & \neg a_4 \vee a_{3-5} & \neg a_7 \vee a_{6-8} \\
\neg a_2 \vee a_{0-2} & \neg a_5 \vee a_{3-5} & \neg a_8 \vee a_{6-8} \\
 & & \neg a_{0-2} \vee \neg a_{3-5} \\
 & & \neg a_{0-2} \vee \neg a_{6-8} \\
 & & \underline{\neg a_{3-5} \vee \neg a_{6-8}}
\end{array}
$$

i	Free vars	Aux vars	Formulas
1	3	0	3
2	9	3	24

At level i of the tree, there are $n = 3^i$ free variables, $(n-3)/2$ auxiliary variables (for a total of $(3n-3)/2$ variables), and $(7n-15)/2$ formulas.