Analysis of Shidoku

Francisco Sanchez

April 28, 2020

Abstract

This paper is an analysis of a combinatorial number-placement puzzle named sudoku. Correctly, the variation of sudoku called shidoku is analyzed using the group's variables. The group classifies and enumerates minimal structures of the 4x4 grid puzzles. The minimal shidokus is a smaller subset of the total enumerations that gives the group a manageable set of data. Using the solutions of these puzzles, we can use a created difficulty system to assess the data based on the start of given clues. The goal of doing so is to prove that as the amount of starting clues given decreases, the difficulty of the minimal shidoku will increase on average.

1 Introduction

1.1 What is Sudoku?

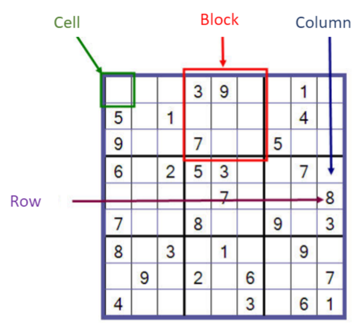


Figure 1: Illustration of constraints on Sudoku grid

Sudoku is a combinatorial number-placement puzzle played all around the world. Its origin stems from magic squares, which is also a puzzle based on placement constraints. Euler transformed these squares into Latin squares. These Latin squares have similar restrictions and ultimately lead to sudoku's creation. Traditionally sudoku is represented as a 9x9 grid with nine 3x3 sub-grids. Nine cells compose each of the nine blocks that have constraints. The position of the cell about its row, column, and block defines the restrictions. The number of cells in a block is the range of numbers placed and must be unique in its three relative positions.

1.2 Shidoku

These limitations on any nxn grid can yield similar results. There are 81 cells in traditional sudoku, which would mean 6.67x10^21 distinct solutions [2]. The number of different ways to fill all unaffected cells after filling in one. However, in the 4x4 variation, shidoku, there are a total of 288 enumerations. This number shows the relationship between one of the 4! ways to place the first and the sub sequential steps. By using an exhaustive enumeration process, it is possible to see a tree of possibilities. At each stage of filling cells, some of the options vanish. By doing this tree of possibilities, it happens that there also 12 distinct possible outcomes [1]. This relationship is 4! ways of filling the first cell by its 12 distinct solutions give 288 enumerations.

1.3 Minimal puzzles

In the set of 288, there is a group of 36 minimal puzzles. Minimal puzzles are puzzles that only have the necessary clues to solve the given problem. The number of filled cells must be 4 in shidoku and goes up to 6 clues. A graph theory proof establishes this [1]. So, they can be grouped by the number of given cells or the puzzles starting position. There is a different amount per group: 13 4-clue, 22 5-clue, and only one 6-clue. Constructing this data set of 36 puzzles is reasonably comprehendible. When doing so, the goal is to affect as many cells as possible.

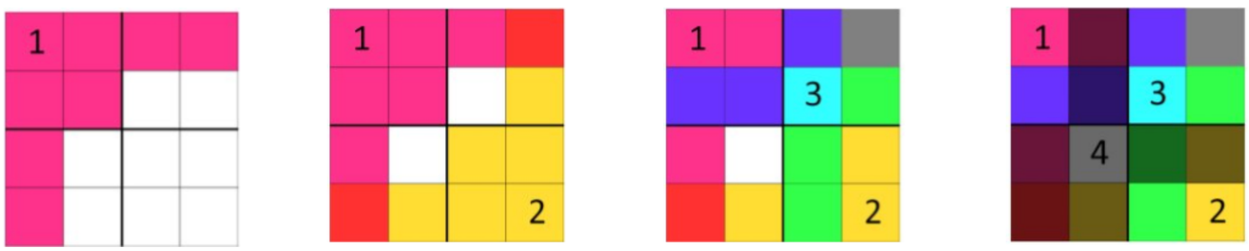


Figure 2: Creating a Minimal Shidoku

Even though the figure shows the first two being placed in opposing corners, it is not a requirement [4]. However, it does create a unique solution. After compiling all variations of minimal puzzles, the group can extract data from work done to complete said puzzles. The number of steps and the processes of finding each step in the work composes the variables the group focused on.

2. Methods

2.1 Backtracking

The group's data centers around manipulating the constraints of the number-placement in the puzzles. If manipulation doesn't occur then, the group would have to cross-check every random number placement with the restrictions. Meaning, back-tracking would occur. This method, "starts with an empty variable assignment and tries to find a solution by assigning values to variables one by one" [3].

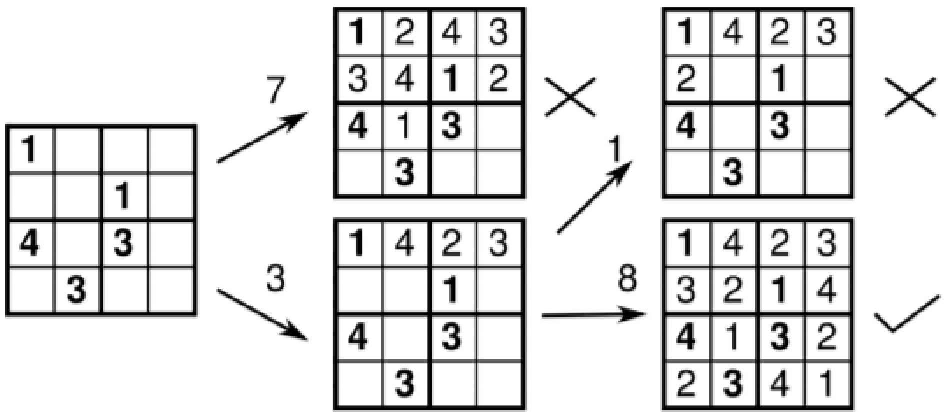


Figure 3: An example of Backtracking

As we can these steps in Figure 3, the steps look tedious, and they are. Having to do this on every minimal puzzle would not provide meaningful data because, at random, the number of steps will differ for the same puzzle. So, the group must choose a different method of solving each puzzle. Two techniques allow for checking the limitations before placing first numbers randomly.

2.2 Naked Singles and Hidden singles

These techniques are naked singles and hidden singles. At the beginning of the puzzle, one can assess the known possibilities of the empty cells using the constraints.

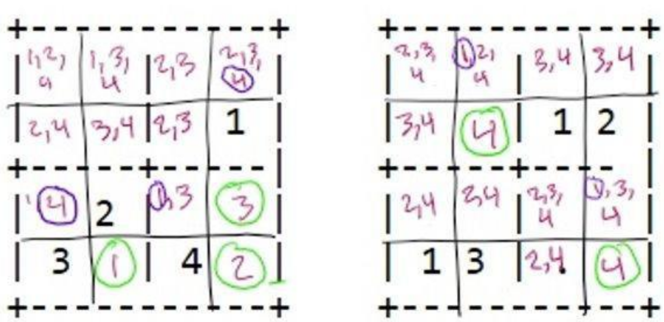


Figure 4: Distinguishing the singles

After penciling in all the options cells with only one possibility penciled in is a naked single (green circle)[3]. Though there might be more than one possibility in a cell, some might be clear after cross-examining all constraints, and these are hidden singles (purple circle) [3]. The group chose to separate the two techniques. Keeping them separate was a way to tell if there was any difference in outcome based on which method filled more cells. Also, count how many cells filled by each method at the beginning of each step. By knowing this information, the group intended for it to represent several ways. Firstly, and most obvious, as there are more known cells, there are fewer unknown cells. It is possible to conclude the number of times needed to complete each puzzle, which the group saw as a direct relation to the overall difficulty of the puzzle. The group can represent n-clue subsets amongst the data.

2.3 Data Collection

The method of collecting the data at first seemed an insurmountable amount of work. However, once the group saw the 36 minimal puzzle 4x4 grids, it became the focus. Mainly, it was possible to solve these 36 puzzles by hand, and that was best for our method of collecting data. Because of being written it was easy to see the work done, each group member was able to concur about the results. Each cell considered a naked single, and a hidden single was distinguished, and each 'step' had a uniform definition between each puzzle.

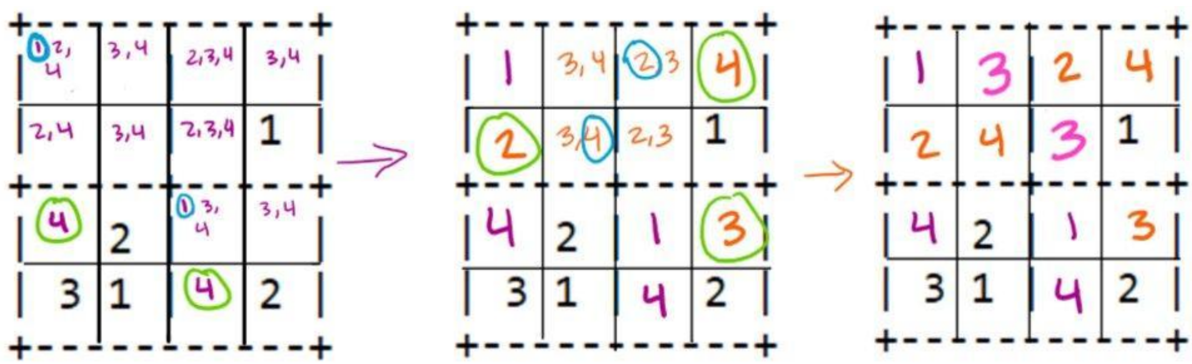


Figure 5: Working out a Shidoku

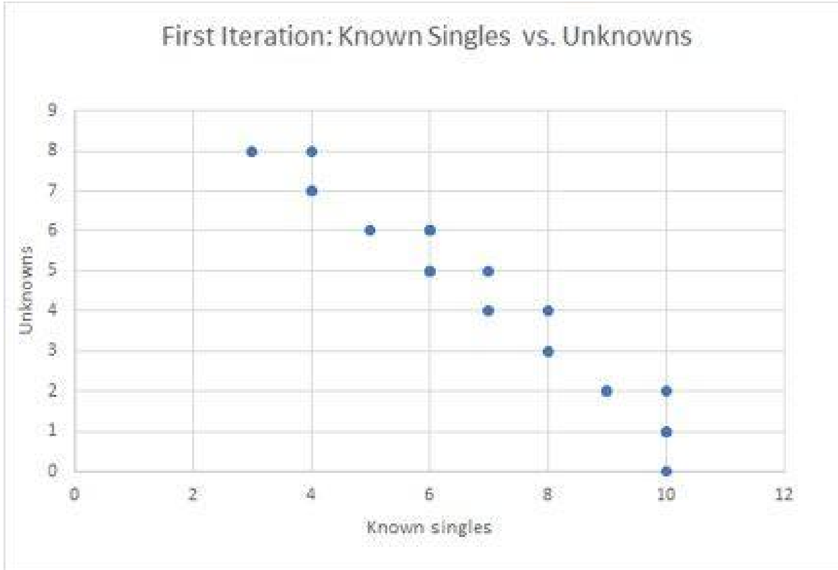
The group ordered the puzzles 1-36 and recorded the number of naked singles and hidden them by looking at the initial clues. By knowing the number of known singles, the group can see the number of unknown cells and can record them. By the hypothesis, there should be a lower ratio of known singles and unknowns in minimal puzzles of 4-clues. Also, across the n-clues there should be fewer steps as the ratio of gets higher.

3 Results

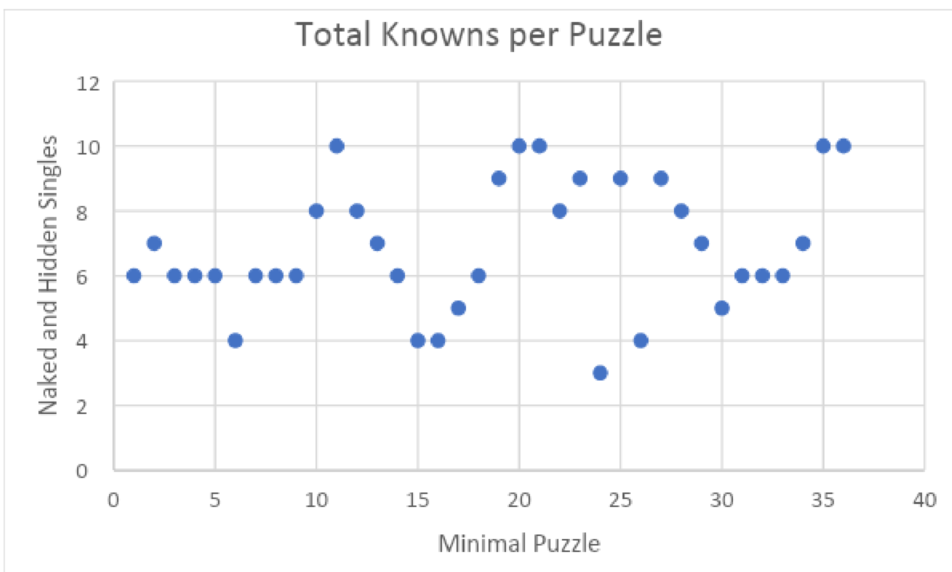
3.1 Difficulty of Techniques

Once the group began collecting data, suspected trends occurred for some puzzles and not for others. The initial goal was to create a rating system to see what exactly makes certain puzzles harder. When first researching sudoku, there are many ways to define the difficulty in sudoku. However, because of the group's focus on shidoku, it was hard to see what exactly would make them hard. This problem is that shidoku only requires the two techniques described above. Also, in previous papers, time can be used as a metric, but all the shidoku were able to be solved in two or three steps. Besides these problems, there were a few relationships established.

3.2 Unknown Cell over Known Singles

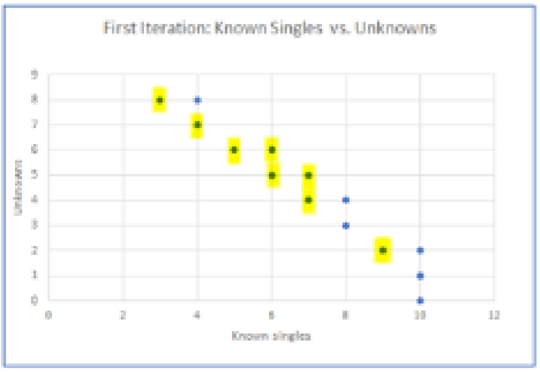


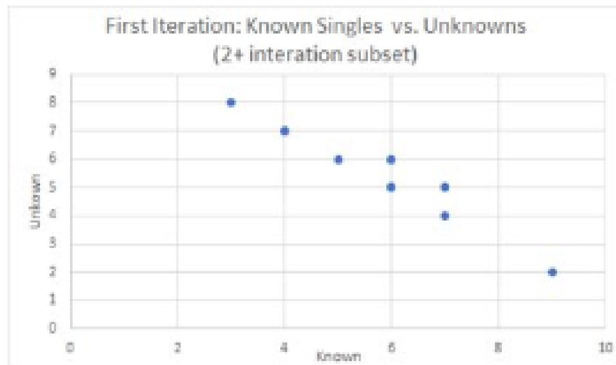
The group chose to define a step as an iteration of the puzzle when the known cells filled in. The first graph shows the fundamental relationship necessary to support the hypothesis. This graph shows the change between the known cells filled versus the cells that require another step to be complete, unknowns. It is clear and quite evident that as unknowns decrease. However, this relationship doesn't coincide with the group's assumption of how the starting n-clues affect the number of iterations.



As this graph shows, the relationship between n-clues and their knowns versus unknowns ratios is hard to distinguish in the first iteration. The orange indicates the 4-clue, blue the 5-clue and the red dot is the one 6-clue. The assumption was that as the ratio of unknowns over knowns increases, meaning the unknowns outweigh the knowns, that this would translate to the difficulty. The difficulty is the number of steps or iterations needed to solve a minimal shidoku.

However, two things occurred that the group didn't expect. One was that the ratios seemed to be almost at random. Second, these ratios didn't seem to affect whether or not the group solved the puzzle after iteration one. Also interesting was that as n, in n-clue minimal shidokus, increased there was no evidence of a decrease in the defined difficulty. In fact, there was a higher percentage of 5-clue among themselves with high unknown ratios than in the 4-clue. So, it seems the hypothesis was not accurate. The only thing the team accomplished through data was the relations between unknowns and knowns. The two graphs side by side best expresses the relationship.





In the second iteration, it shows the highlighted points from the first iteration. This correlation indicates that in general, the higher ratio of unknowns to knowns will result in a necessary third step. Though this is mostly true, there seems to not be a correlation to the n-clue. There is a higher percentage of 5- clue needing a third step, which does coincide with 5-clue puzzles having an overall higher unknown ratio. However, this does not support the group's hypothesis of having a higher degree of difficulty with a lower number of starting clues. It seems independent of one another as it seems to be centered around the position of said puzzles.

4 Discussion

4.1 Conclusion

By the end, the project had grown then shrunk. The team had made many relations about what creates a combinatorial based puzzle. After deciding the focus was sudoku, the project had not finished changing. There exists no definitive proof for every type of sudoku, only the number of them are. Also, seeing that there was also no absolute difficulty rating system for sudoku, the group was inspired. However, due to the time parameters of the discussion, the group saw its limits. Shidoku seemed like the answer, yet it was not enough. The manageable data set had its pros and cons, but ultimately the group was unable to achieve their goal. The group did have enough data to reject the proposed hypothesis, though. The starting number of n-clues does not affect the difficulty of the solution for shidoku. There are also not enough techniques to create a difficulty rating around. The two discussed were too similar, and the data shows this. Ultimately, the group's ambition didn't match the results of the project.

4.2 Application

The application of this data is smaller than initially intended. The goal was to have a set of functions that describe any nxn grid of the same parameters. However, because of a lack of data, this cannot be seen all the way through. The group did prove that the number of starting clues is not a good representation of difficulty for any combinatorial puzzle. But the number of steps needed taken to finish a minimal is a good representation. So, moving forward, the two simple techniques can serve as a metric for bigger grids. For instance, if traditional sudokus have more cells filled with the two, than it could require fewer steps to solve versus other techniques.

References

1) Sudopedia. (2010).http://sudopedia.enjoysudoku.com/Shi_Doku.html

2) Oddson, K. (2016) Math and Sudoku. Portland State University. http://web.pdx.edu/~caughman/Oddson-Math%20and%20Sudoku.pdf

3) Pelanek, R. Difficulty Rating of Sudoku Puzzles: An Overview and Evaluation. Masaryk University. https://www.fi.muni.cz/~xpelanek/publications/sudoku-arxiv.pdf

4) Puskar, L. (2017) An Exploration of the Minimum Clue Sudoku Problem. Sacred Heart University.https://digitalcommons.sacredheart.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1089&context=acadfest