Enumerating Small Sudoku Puzzles in a First Abstract Algebra Course

Crystal Lorch
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306
clorch@bsu.edu

John Lorch
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306
jlorch@bsu.edu

Introduction

Purpose We give two methods for counting small 'essentially different' sudoku puzzles using elementary group theory, and we indicate how this topic can be incorporated into an introductory abstract algebra course. Students benefit by exploring an interesting application of algebra to counting, as well as by revisiting key concepts in group theory.

Background A *sudoku puzzle*[1] is (usually) a 9×9 array in which some slots contain one of the numbers $\{1, 2, \ldots, 9\}$ and other slots are empty (left panel of Figure 1). The goal is to complete all empty slots so that each row, each column, and each of the nine 3×3 'blocks' contains all of the numbers 1 through 9 (right panel of Figure 1). In other words, a completed sudoku puzzle is an order-9 Latin square with an additional condition on the 3×3

[1]'Sudoku' literally means 'single number' in Japanese. The puzzles supposedly originated in the United States in the late 1970's (see [5]).

blocks. A valid sudoku puzzle must have a *unique* completion[2].

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

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

Figure 1: A 9×9 sudoku puzzle and its completion

While solving sudoku puzzles requires no mathematics per se, there are natural mathematical problems associated with sudoku, including those involving the enumeration of completed puzzles. These enumeration problems are harder than one might think. For instance, the exact number of order-9 Latin squares (approximately 5.52×10^{27}) wasn't known until 1975 (see [1]); the somewhat smaller exact number of distinct completed sudoku puzzles (6670903752021072936960) was determined by Felgenhauer and Jarvis in 2005 in a computer-aided argument (see [3]).

Finding the number of distinct completed puzzles isn't the end of the story. There are many completed puzzles that are the same up to relabeling, or up to an 'allowable' symmetry of the underlying array (e.g., rotation). Puzzles that exhibit this type of similarity are *essentially the same*; those that do not are *essentially different.* Since the collection of relabelings and array symmetries can be thought of as a group (the *sudoku group*) acting on the set of completed puzzles, the problem of counting essentially different completed puzzles can be phrased in terms of counting orbits of a group action. From this perspective, Jarvis and Russell [4] used Burnside's formula to determine

[2]We do not address strategy for completing sudoku puzzles. Those interested in strategy should refer to [5] or [6].

that there are 5472730538 essentially different 9×9 completed sudoku puzzles, employing the group theory software GAP to aid the computation.

Purpose revisited: Junior sudoku puzzles The sudoku enumeration problems described above become far more tractable if we consider 4×4 *junior*[3] sudoku puzzles in place of the standard 9×9 variety (see Figure 2). We can quickly count the distinct complete junior puzzles by methodically

	1		4
4		1	2
1			
	2	4	1

2	1	3	4
4	3	1	2
1	4	2	3
3	2	4	1

Figure 2: A junior sudoku puzzle and its completion

writing down all twelve puzzles with upper left block of the form $\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}$ (see Figure 3), and then observing that any other completed junior puzzle must be among the $4! = 24$ relabelings of one of these twelve puzzles. So there are $4! \times 12 = 288$ puzzles total.

Later, we will use two different methods to show that *these* 288 *puzzles boil down to only* 2 *essentially different junior puzzles.* One method relies on a property of junior puzzles that is invariant under the action of the sudoku group, while the other illustrates the use of Burnside's formula as in [4]. Both methods only require background usually acquired in a first abstract algebra course, with the former method requiring less than the latter. Along the way questions are provided to help instructors incorporate this material into their algebra classes.

The structure of the paper is as follows: The first section contains material about the sudoku group and its action on completed puzzles that is necessary for both counting methods. The next two sections illustrate the two methods of counting essentially different complete junior puzzles. The final section

[3]This terminology is suggested in the puzzles column of [2].

1	2	3	4
3	4	1	2
2	1	4	3
4	3	2	1

x_1

1	2	4	3
3	4	1	2
2	1	3	4
4	3	2	1

x_2

1	2	3	4
3	4	2	1
2	1	4	3
4	3	1	2

x_3

1	2	4	3
3	4	2	1
2	1	3	4
4	3	1	2

x_4

1	2	3	4
3	4	1	2
4	1	2	3
2	3	4	1

x_5

1	2	4	3
3	4	2	1
4	1	3	2
2	3	1	4

x_6

1	2	3	4
3	4	1	2
2	3	4	1
4	1	2	3

x_7

1	2	4	3
3	4	2	1
2	3	1	4
4	1	3	2

x_8

1	2	3	4
3	4	1	2
4	3	2	1
2	1	4	3

x_9

1	2	4	3
3	4	1	2
4	3	2	1
2	1	3	4

x_{10}

1	2	3	4
3	4	2	1
4	3	1	2
2	1	4	3

x_{11}

1	2	4	3
3	4	2	1
4	3	1	2
2	1	3	4

x_{12}

Figure 3: Twelve puzzles with a fixed upper left block

introduces an application to the 'minimum hint' problem. Throughout, the following vocabulary will hold sway:

- **grid** = the underlying 4×4 array of square slots in a junior sudoku puzzle.
- **blocks** = the 2×2 arrays occurring at the corners of a grid. The top blocks are labeled I and II from left to right, the bottom blocks are similarly labeled III and IV.
- **completed puzzle** = a grid such that each slot contains one of the numbers $\{1, 2, 3, 4\}$, and each of the numbers 1 through 4 appears in every row, every column, and each of the four 2×2 blocks. (*We often abuse the terminology and use the word* ***puzzle*** *to mean completed puzzle.*)
- **rows** = the rows of a puzzle, labeled 1 to 4 from top to bottom
- **columns** = the columns of a puzzle, labeled 1 to 4 from left to right.

The sudoku group and its action on puzzles

Consider the following (inefficient) collection of elementary manipulations that carry one completed junior puzzle to another:

- relabeling entries
- swapping rows 1&2, or rows 3&4
- swapping columns 1&2, or columns 3&4
- swapping the top two blocks with the bottom two blocks, or the right two blocks with the left two blocks
- 90° clockwise rotation
- reflection across the diagonal (i.e., matrix transpose)

Viewing these manipulations as functions on the set X of completed puzzles, we may form the *sudoku group* G (under function composition) generated by these manipulations. The sudoku group acts naturally on X by

$$g.x = g(x), \qquad g \in G,\ x \in X.$$

We say that two puzzles are *essentially the same* if they lie in the same orbit under this action (that is, they are the same up to some elementary puzzle manipulation), and are *essentially different* otherwise. Counting the number of essentially different puzzles is tantamount to counting orbits in X under G.

Discussion

1. Is swapping rows 1&3 of a puzzle an acceptable manipulation of completed puzzles? Why or why not?
2. What properties must a set S of functions possess in common in order for S to be a group under function composition?
3. Is G a finite group? Why or why not?
4. The generating set for G given above is redundant. Can you produce a smaller set of generators that still yields G?
5. Why is the action of G on X truly a group action?

Counting puzzles using an invariant property

Here we present the first of two methods for counting essentially different completed puzzles.

Let X_{fix} denote the set of puzzles $\{x_1, \dots, x_{12}\}$ with fixed upper left block as given in Figure 3, and let K be the subgroup of G preserving X_{fix}. Since K must fix the upper left block of every puzzle, it follows that $K = \{e, r, c, rc\}$ where r swaps rows 3&4 and c swaps columns 3&4. Luckily both X_{fix} and K are small, so we can quickly determine that

$$\{x_1, x_4, x_9, x_{12}\}, \quad \{x_2, x_3, x_{10}, x_{11}\}, \quad \text{and} \quad \{x_5, x_6, x_7, x_8\} \tag{1}$$

are the the orbits of X_{fix} under K.

Since G contains all possible relabelings, each G-orbit in X must intersect X_{fix}. This together with the fact that K is a subgroup of G implies that the number of G-orbits in X is bounded above by the number of K-orbits in X_{fix} (see Discussion questions). Therefore, we have:

PROPOSITION 1. There are at most three orbits in X under G.

The distinct G-orbits in X are potentially $G.x_1$, $G.x_2$, and $G.x_5$ (with one base-point picked from each of the K-orbits in X_{fix}). After some quick trial and error, we find that a rotation followed by a relabeling will send x_2 to $x_7 \in G.x_5$, and hence $G.x_2 = G.x_5$.

To finish, it remains to determine whether $G.x_1 = G.x_2$. We accomplish this by considering a property possessed by some of the puzzles in X. We say a puzzle possesses property P if *each row in blocks II, III, and IV is a permutation of a row in block I, and each column in blocks II, III, and IV is a permutation of a column in block I.* (Note: A row or column 'in a block' will consist of only two entries.) We can see that if $x \in X$ has P and g is one of the generators of G mentioned in the previous section (see bulleted list), then $g.x$ also has P. Therefore P is *invariant* under G (i.e., if $x \in X$ has P then so does $g.x$ for any $g \in G$). Then, since x_1 has P and x_2 does not, it follows that the orbits $G.x_1$ and $G.x_2$ cannot coincide. Therefore:

THEOREM 2. There are exactly two orbits in X under G (namely $G.x_1$ and $G.x_2$), and hence exactly two essentially different completed junior sudoku puzzles.

Discussion

1. Suppose $x, y \in X$ and $g_1, g_2 \in G$ are such that $g_1.x$ and $g_2.y$ both lie in X_{fix}. If $g_1.x$ and $g_2.y$ lie in the same K-orbit in X_{fix}, explain why x and y lie in the same G-orbit in X.
2. Use the result of the previous problem to argue that the number of G-orbits in X is bounded above by the number of K-orbits in X_{fix}.
3. Find the combination of rotation and relabeling that sends x_2 to x_7.
4. If a property of certain puzzles in X is preserved by the generators of G, why is it preserved by the whole of G?
5. Why doesn't $G.x_1$ coincide with $G.x_2$? Give a careful argument.
6. How many elements are in $G.x_1$? In $G.x_2$?

Counting puzzles using Burnside's formula

Here we count essentially different puzzles via the approach [4] using Burnside's formula. Burnside's formula says

$$N \cdot |G| = \sum_{g \in G} |X^g|, \tag{2}$$

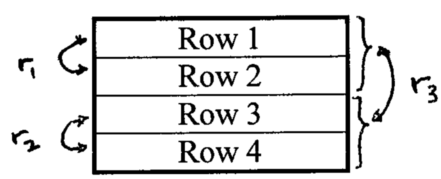
where N is the number of orbits in X under G and $X^g = \{x \in X \mid g.x = x\}$. Though this sum may become unwieldy if both G and X are large, we will see that (2) admits considerable reduction depending on the structure of G and the group action.

More on the sudoku group To apply Burnside's formula efficiently, we need more information about the sudoku group G. Since the relabelings in G commute with the underlying grid symmetries (e.g., rotation), we have

$$G = S_4 \times H \tag{3}$$

where H is the subgroup of G consisting of underlying grid symmetries. By labeling the squares in a grid 1 through 16, elements of H may be viewed as elements of S_{16}.

We attend to the structure of H. All we need to generate H are the row swappings, the column swappings, and reflection across the diagonal (see bulleted list of generators for G). Let r_1, r_2, r_3 denote the basic row swappings that interchange the first two rows, the last two rows, and the two rows of blocks, respectively (see Figure 4). These elements generate a



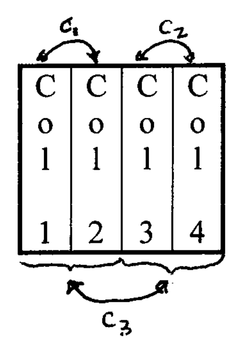


Figure 4: Row and column permutations

subgroup of H isomorphic to D_4. If c_1, c_2, c_3 denote the corresponding column permutations, then the corresponding subgroup of G is also isomorphic to D_4, and any interchange of columns will commute with any interchange of rows. Finally, matrix transpose (denoted τ) generates a group isomorphic to $\mathbb{Z}_2$, and the row and column permutation subgroups of H are related by the fact that $\tau r_j = c_j \tau$ for $j \in \{1, 2, 3\}$. These observations imply that any element of H may be uniquely written in the form

$$\text{(row interchange)} \cdot \text{(column interchange)} \cdot \text{(transpose or identity)}$$

More specifically:

THEOREM 3. The sudoku group G is isomorphic to $S_4 \times H$, where $H \cong (D_4 \times D_4) \rtimes \mathbb{Z}_2$. Therefore the order of G is $24 \cdot 8 \cdot 8 \cdot 2 = 3072$.

More on Burnside's formula To complete our count of the orbits in X under G, we simplify the righthand side of Burnside's formula (2). For $h \in H$, let $\widetilde{X^h} = \{x \in X \mid h.x = \sigma.x \text{ for some } \sigma \in S_4\}$ and $\widetilde{X^h_{\text{fix}}} = \{x \in X_{\text{fix}} \mid h.x = \sigma.x \text{ for some } \sigma \in S_4\}$. Since each element of X is a relabeling of a unique element of X_{fix}, it follows that $|\widetilde{X^h}| = 24 \cdot |\widetilde{X^h_{\text{fix}}}|$. Further, due to (3), we may manipulate the righthand side of Burnside's formula as follows:

$$\sum_{g \in G} |X^g| = \sum_{h \in H} \sum_{\sigma \in S_4} |X^{\sigma h}| = \sum_{h \in H} \left| \bigcup_{\sigma \in S_4} X^{\sigma h} \right| = \sum_{h \in H} |\widetilde{X^h}| = 24 \sum_{h \in H} |\widetilde{X^h_{\text{fix}}}|. \quad (4)$$

So, in our specific circumstances, Burnside's formula becomes

$$N \cdot |G| = 24 \sum_{h \in H} |\widetilde{X^h_{\text{fix}}}|. \quad (5)$$

Using a software package such as *Mathematica*, we can show that the righthand side of (5) is 6144. (A *Mathematica* program for this computation may be found at `http://www.cs.bsu.edu/homepages/jdlorch/preps.htm`.) Putting this together with the fact that $|G| = 3072$ (Theorem 3), we conclude:

COROLLARY 5. There are exactly two orbits in X under G, and hence exactly two essentially different completed junior sudoku puzzles.

Discussion

1. Show how a 90° clockwise rotation can be obtained in terms of row permutations, column permutations, and reflection across the diagonal.
2. In [4], Jarvis and Russell use the following version of Burnside's formula:

$$N \cdot |G| = \sum_{h \in H}' c(h) |\widetilde{X^h}|,$$

where $c(h)$ is the number of elements in the conjugacy class containing h, and the primed sum picks up precisely one representative h from each conjugacy class. Show that for our G, H, X, etc., this version of Burnside's formula follows from that given in (2).

3. Write r_1 and c_2 as elements of S_{16}. Then, show that $r_1 c_2 = c_2 r_1$ by direct computation in S_{16}. Can you verify that any of the row swaps commute with any of the column swaps?
4. List the elements of the subgroup of H generated by r_1, r_2, r_3. Show that this subgroup is isomorphic to D_4.
5. The symbol $\rtimes$ denotes a semi-direct product. Why does this symbol appear in Theorem 3?
6. Verify each equality given in (4).
7. Write a program in *Mathematica* or some other package that will evaluate the righthand side of (5).

Application: The minimal number of entries determining a unique puzzle

In the 9×9 case, there are complete puzzles that are uniquely determined by 17 entries (see Figure 5), but it is an open problem as to whether any

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

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

Figure 5: A puzzle with 17 entries and its unique completion

completed puzzles are uniquely determined by 16 entries. For junior puzzles, this 'minimum hint' problem is made far simpler by the fact there are only 288 completed puzzles overall, and most of these are essentially the same. While each of the essentially different puzzles x_1 and x_2 are uniquely determined by four entries, no junior puzzle is uniquely determined by only three entries.

Discussion

1. For each of x_1 and x_2, find four entries that uniquely determine the puzzle. What does this result say about *all* completed junior puzzles?
2. Show that no junior puzzle is uniquely determined by three entries.
3. Check that the puzzle in Figure 5 has a unique completion.

References

[1] S. Bammel and J. Rothstein, *The number of* 9×9 *Latin squares*, Discrete Mathematics **11** (1975), 93-95.

[2] E. Berlekamp and J. Buhler, Puzzles column, MSRI Emissary, Fall 2005.

[3] B. Felgenhauer and F. Jarvis, *Mathematics of Sudoku I*, Math. Spectrum, to appear. (Preprint available at www.afjarvis.staff.shef.ac.uk/maths/felgenhauer_jarvis_spec1.pdf)

[4] F. Jarvis and E. Russell, *Mathematics of Sudoku II*, Math. Spectrum, to appear. (Preprint available at www.afjarvis.staff.shef.ac.uk/maths/russell_jarvis_spec2.pdf)

[5] R. Wilson, *The Sudoku epidemic*, MAA Focus **26** no. 1 (2006), 5-7.

[6] R. Wilson, *How to Solve Sudoku: A Step by Step Guide*, Infinite Ideas, 2005.