8-2022

Verifying Sudoku Puzzles

Chelsea Schweer

.

Verifying Sudoku Puzzles

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Chelsea Schweer

August 2022

Verifying Sudoku Puzzles

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Chelsea Schweer

August 2022

Approved by:

Dr. Jeremy Aikin, Committee Chair

Dr. Lynn Scow, Committee Member

Dr. Youngsu Kim, Committee Member

Dr. Madeleine Jetter, Chair, Department of Mathematics

Dr. Corey Dunn, Graduate Coordinator

Abstract

Sudoku puzzles, created by Meki Kaji around 1983, consist of a square 9 by 9 grid made up of 9 rows, 9 columns, and nine 3 by 3 square sub-grids called blocks. The goal of the puzzle is to be able to place the numbers 1 through 9 in every row, column, and block where no number is repeated in each row, column, and block. Imagine being given a completed Sudoku puzzle and having to check that it was solved correctly. You could just check all the rows columns and blocks (27 items), but is there a smaller number of checks that would suffice? In fact, what is the least amount of rows, columns, and blocks you would need to check? This question has been investigated by Emil Jeřábek, Tony Huynh, and others on various internet posts, where an answer has been posted, along with involved arguments that use a variety of mathematical tools. In this thesis, we investigate this question in the context of matroid theory and present a matroid theoretic solution by using the concept of circuits of a matroid.

Acknowledgements

First I would like to thank my family and close friends who have supported me throughout my journey. Without their help I would not have been able to accomplish my goals and finish the program. Next, I would like to thank Dr.Aikin for all of his help with this thesis and his support in all my endeavors and goals both during this program and in my future goals as well. Thank you to all the Professors and staff in the math department at CSUSB, without each of you I would not be where I am today.

Table of Contents

List of Figures

.

Chapter 1

Introduction

A puzzle and game meant to be enjoyed by all ages, Sudoku puzzles, created by Meki Kaji around 1983, gained popularity world wide in 2007 [Mak21]. A Sudoku puzzle consists of a square 9 by 9 grid made up of 9 rows, 9 columns, and nine 3 by 3 square sub-grids called blocks. The goal of the puzzle is to place the numbers 1 through 9 in every row, column, and block where no number is repeated in each row, column, and block. One might assume, rightfully so, that there are many different ways that a 9 by 9 Sudoku puzzle can be filled in using the numbers 1 through 9. So, it is customary for a Sudoku puzzle to have certain numbers already positioned in the puzzle at the start so there is a unique way to complete the puzzle. Perhaps you are thinking that completing a Sudoku puzzle is little more than a calming way to spend an hour or two. However, if the reader is anything like the writer, then you may finish one of these puzzles only to notice that one of your rows, columns, or blocks has a repeated number making your puzzle incorrect. All of that time trying to position numbers in each row, column, and block only to lead to failure. You may begin to wonder if other puzzles that you have "solved" are actually right or if there a chance that while you believed yourself to be a Sudoku savant, in reality you are just a Sudoku simpleton. Is there a way you can go back and check some of your other solved puzzles? Do you have to spend the rest of your days searching through every single row, column, and block that you have ever completed to see if all your puzzles are correct? Or is there a minimum amount of rows, columns, and blocks that one can check to determine if they have a correctly solved a Sudoku puzzle? That is the question we are going to answer in this thesis.

This question seemed to have originated from a MathOverflow post where it was claimed that in order to verify whether or not a Sudoku puzzle has been completed correctly, one must check at least 21 of the rows, columns, and 3 by 3 sub-blocks [Je7]. Emil Jeřábek posted a rather complicated proof that showed this fact. At the heart of his proof seems to be a combinatorial structure known as a matriod. This question was further explored by Tony Hyuen [Hyu14]. Our goal is to prove just that.

In Chapter 1, we will introduce the concept of a Sudoku puzzle provide important definitions. Then, in Chapter 2 we will classify specific sets of rows, columns, and blocks in a Sudoku puzzle. Chapter 3 we will focus on matroid theory. The theorems we introduce in Chapter 3 will be used in Chapter 4 to analyze Sudoku puzzles through the lens of matroid theory in order to answer questions about verifying Sudoku puzzles. Finally, in Chapter 5 we will discuss some consequences and extensions of our analysis in Chapter 4.

Throughout this thesis, we will introduce new terminology that will be useful to describe both the components of a Sudoku puzzle and the actions we perform on a puzzle. For example, instead of saying that we have solved a Sudoku puzzle correctly, we will say the completed puzzle is consistent. More specifically, a solved Sudoku puzzle is said to be *consistent* if every row, column, and block contain the numbers 1 through 9. We denote the rows by r_i, where $i \in \{1, 2, ..., 9\}$, so that r_1 is the top row and the remaining rows are labeled sequentially from top to bottom. Similarly we denote the columns by c_j, where $j \in \{1, 2, ..., 9\}$, so that c_1 is the furthest left column and the remaining columns are labeled sequentially from left to right. We denote the blocks by b_k, for $k \in \{1, 2, ..., 9\}$, where b_1 is the top left block and the remaining blocks are labeled sequentially in the order we would read a page in a book. The collection of rows, columns, and blocks is called set K of *elements* of a Sudoku puzzle. The set K is also called the *ground set*. Evidently, the cardinality of K, denoted $|K|$, is 27. Each element in K is made up of 9 unit squares, called *cells*, into which we place the numbers 1 through 9. Each of the 81 cells in a Sudoku puzzle can be unambiguously identified by an ordered triple (i, j, k) where $i, j, k \in \{1, 2, .., 9\}$ and i, j and k represent the row, column, and block, respectively, the cell is in. For example, $(4, 6, 5)$ would be the cell that is contained in with row 4, column 5, and block 6. See Figure 1.1.

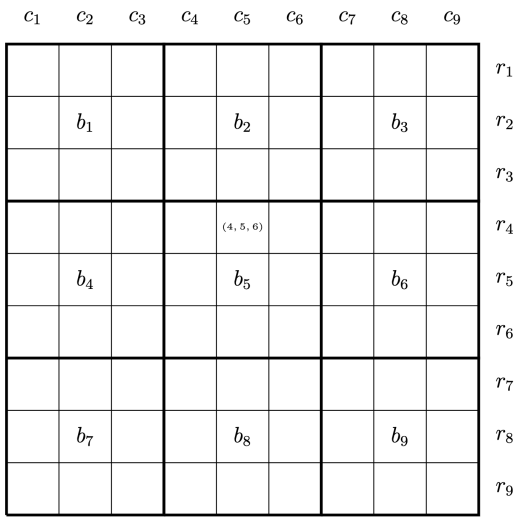

Figure 1.1: The elements and cells of a Sudoku puzzle.

Throughout this thesis, we will denote our ground set by K. Now that we know what the elements of a Sudoku puzzle are we are ready to introduce some important terms. If $A \subset K$ and $e \in (K - A)$, we say that A *verifies* e or A is a verification set for e, denoted $A \to e$, when it can be deduced that the element e is *consistent* from assuming the elements in A are consistent. If $A \subseteq K$ and $B \subseteq (K - A)$, we say that A *verifies* the set B, denoted $A \to B$, if under the assumption that every element if A is consistent, it can be deduced that every element in B is consistent. In this event, we also say that A is a *verification set* for B. When $A \to (K - A)$, we call A a *full verification set* for the Sudoku puzzle. We note that when $A \to B$, it is usually the case that the elements of B can be ordered sequentially as $B = \{b_1, b_2, ..., b_t\}$ such that $A \to b_1, (A \cup b_1) \to b_2, ...,$ and $(A \cup b_1 \cup b_2 \cup ... \cup b_{t-1} \to b_t$. If the consistency of an element e cannot be deduced by the set A, we write $A \not\to e$. However, this does not necessarily mean that e is inconsistent. It just means that the elements assumed to be consistent in A cannot verify e. We call an element e of the ground set K *unverified* if we have not yet determined it to be consistent or inconsistent. For example, it can be easily argued that if we knew 26 elements in K were consistent, we would be able to deduce that the remaining element is consistent, which would lead to having a consistent

puzzle. But the question becomes, how do we go about verifying an unverified element? In order for us to answer this question, we first need to introduce a few more terms. A *row block* is a set of three rows and three block such that each of the rows has cells contained in each of the blocks. There are three such row blocks in a Sudoku puzzle, and we denote them by RB_1, RB_2 and RB_3. Specifically, let $RB_1 = \{r_1, r_2, r_3, b_1, b_2, b_3\}$, $RB_2 = \{r_4, r_5, r_6, b_4, b_5, b_6\}$, and $RB_3 = \{r_7, r_8, r_9, b_7, b_8, b_9\}$. Similarly, we introduce the analogous notion of *Column Block*. We denote the three the three column blocks by CB_1, CB_2, and CB_3 where, $CB_1 = \{c_1, c_2, c_3, b_1, b_4, b_7\}$, $CB_2 = \{c_4, c_5, c_6, b_2, b_5, b_8\}$, and $CB_3 = \{c_7, c_8, c_9, b_3, b_6, b_9\}$. See Figure 1.2.

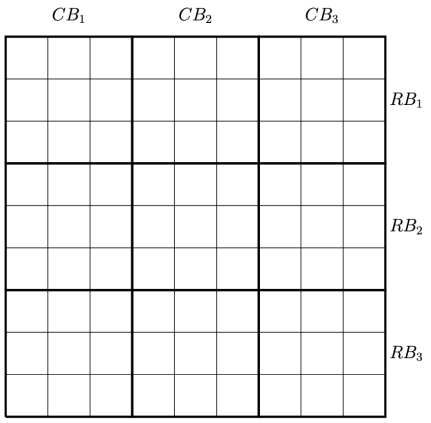

Figure 1.2: Row blocks and column blocks.

The following example illustrates the usefulness of column and row blocks and the role they play in verifying elements.

Example 1.
Let $A = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, b_1, b_2, b_3, b_4,$
$b_5, b_6, b_7, b_8\}$. We wish to show that $A \to b_9$. We begin by assuming all elements in A are consistent. Let us look specifically at the column block CB_3, which contains block b_9. Suppose, to the contrary, that b_9 is inconsistent. Without loss in generality assume that there is a repeated 1 in b_9. Since all the rows and columns are consistent in A, we know these repeated 1's occur in different columns and in different rows in b_9. We lose no generality in assuming that cells $(9, 7, 9)$ and $(8, 9, 9)$ have a 1 in them. Since $b_6 \in A$, it

must be that b_6 is consistent, so the number 1 must appear somewhere in block b_6. Since columns c_7 and c_9 already have 1's in them, a 1 must be positioned in a cell contained in c_8 and b_6. We may assume that a 1 occurs in the cell $(6, 8, 6)$. Thus, column block CB_3 must have the general arrangement of 1's found in Figure 1.3.

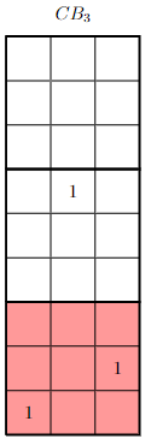

Figure 1.3: Column block CB_3, where b_9, as seen in red, has a repeated 1.

Now, we have positioned a 1 in every consistent column in CB_3, leaving b_3 without a 1. It follows that b_3 is inconsistent. However, $b_3 \in A$, leading to a contradiction. So, b_9 can not contain any repeated 1's and is therefore verified by A. It follows that $A \to b_9$. Consequently, A is a full verification set since $K - A = \{b_9\}$.

In future chapters, we will see more involved verification arguments. We note that our justification for $A \to b_9$ in Example 1 relied primarily on the Pigeonhole Principle. In general, this will be a useful tool for justifying the consistency of elements in a Sudoku puzzle. Before diving deeper into Sudoku puzzle verification, we introduce other important subsets of elements of a Sudoku puzzle, similar to row and column blocks, that will be of great importance in our work in Chapter 4.

Chapter 2

Subsets of Sudoku Puzzles

Our goal for this section is to find subsets A of the ground set K of a Sudoku puzzle that have the property $(A - e) \to e$, for all $e \in A$. It would first be be helpful to distinguish these subsets of K from those that do not have this property. To do this, we will focus on the cells that compose each element in the subsets A if K. We need to distinguish between a cell that is in an already consistent row, column, or block from one that is not. Furthermore, we must distinguish between cells that are in multiple consistent rows, columns, and blocks. We say a cell is *covered* when it is contained in at least one consistent element in a subset A. To distinguish further, a cell is called *triple covered* when covered by three consistent elements in A, *double covered* when covered by two consistent elements in A, *single covered* when covered by one consistent element in A, or *uncovered* when it is not contained in any consistent elements in A. It would stand to reason that having a cell be uncovered by a set A would mean that cell would be left unchecked by every assumed consistent element in that set A. Therefore, one would not be able to deduce that the element that cell is contained in is consistent by the elements in that subset A.

Lemma 2.1. *Let $A \subset K$ and $e \in K - A$. If e contains cells that are uncovered by A, then $A \not\to e$*

Proof. Assume there exists a cell (i, j, k) contained in $e \in K - A$. Then there is no elements in A that contains (i, j, k). So, no matter what number is positioned in the cell (i, j, k) the consistency of the elements in A will not be contradicted. So, (i, j, k) could contain a duplicate in e. Therefore it cannot be deduced by the elements in A that e is

consistent or inconsistent. Thus, $A \not\rightarrow e$. □

Lemma 2.2. *If $A \subset K$ that contains single covered cells covered by A, then there exists an element $e \in A$ such that $(A - e) \not\rightarrow e$.*

Proof. Let $A \subset K$ that contains single covered cells covered by A. Let $e \in X$ such that e covers some of the single covered cells. Then $(A - e)$ would contain cells that are uncovered. By Lemma 2.1 $(A - e) \not\rightarrow e$. □

Lemma 2.2 will be useful in determining whether or not a subset A of K has the desired property that $(A - e) \rightarrow e$, for all $e \in A$. Indeed, Lemma 2.2, if A contains any cells that are single covered in A, then A does not have this property. Our next step is to construct subsets that both have this property of verification and are minimal with respect to this property.

2.1 Configuration Sets

Configuration sets are nonempty subsets C of K that are minimal with respect to the property that for every $x \in C$ then $C - x \rightarrow x$. In order for us to properly describe these configurations sets, we first have to describe a specific collection of elements similar to a row block and a column block. Let a *row set* be the collection of three rows whose cells compose the same three blocks, specifically $RS_1 = \{r_1, r_2, r_3\}$, $RS_1 = \{r_4, r_5, r_6\}$, $RS_2 = \{r_7, r_8, r_9\}$. Let a *column set* be the collection of three columns whose cells compose the same three blocks, specifically $CS_1 = \{c_1, c_2, c_3\}$, $CS_1 = \{c_4, c_5, c_6\}$, $CS_2 = \{c_7, c_8, c_9\}$. With these subsets we are able to better define these configuration sets and the types that they fall into. We will start with one that should seem familiar to the reader.

Configuration Type 1. C is the collection of one row set or column set and the blocks that intersect those rows or columns respectively.

Proof. Without loss of generality, let C contain a row block with all consistent elements, specifically RB_1. We will show that for any element, e in C, then $(C - e) \rightarrow e$. Without loss of generality, we will first look at $(C - r_1)$. Since $r_1 \notin (C - r_1)$, it follows that r_1 is unverified. So, we can position a repeated 1 somewhere in r_1. However, since $b_1, b_2, b_3 \subset C - r_1$ that implies that those blocks are consistent and each contain a single

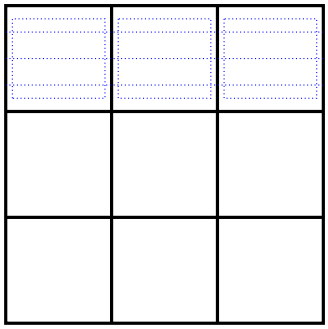

Figure 2.1: One Example of configuration Type 1 where all consistent elements are in blue.

1 in them. So, without loss of generality we can position a 1 in $(1, 2, 1)$ and $(1, 5, 2)$. Lastly, since $r_2 \in (C - r_1)$, r_2 must be consistent so it follows that there is a 1 positioned somewhere in r_2. Since b_3 needs to have a 1 positioned in it, without losing generality we can put a 1 in $(2, 8, 3)$. So, our row block would like Figure 2.2

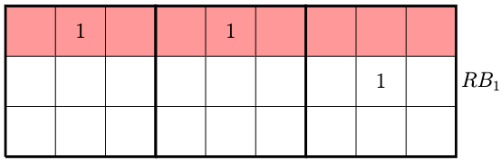

Figure 2.2: Row block RB_1 where $r_1 \notin C - r_1$ is seen in red.

Now, since all blocks have a 1 in them, there is no where to position a 1 in r_3, it follows that r_3 is inconsistent. However, $r_3 \in (C - r_1)$, which would mean r_3 is consistent, hence a contradiction. So, r_1 must not contain a repeated 1, meaning r_1 is consistent. So, $(C - r_1) \rightarrow r_1$. We can follow the same reasoning for column blocks missing a column. Now, we will look at the case trying to verify a block. Without loss of generality lets look at $(C - b_3)$. Since $b_3 \notin (C - b_3)$ it follows that b_3 is unverified so we can position two 1's in b_3. Since $r_2, r_3 \subset (C - b_3)$ it follows that those elements are consistent so we will position a 1 in $(2, 8, 3)$ and $(3, 9, 3)$. Since $r_1 \in (C - b_3)$, it must be consistent. Since $b_1 \in (C - b_3)$ it must also be consistent. So, we will position a 1 in $(1, 2, 1)$. So, our row block would look like:

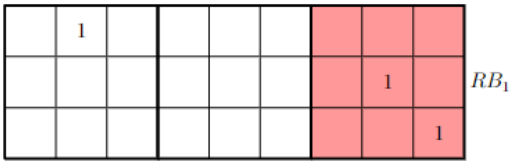

Figure 2.3: Row block RB_1 where $b_3 \notin (C - b_3)$ is seen in red.

Now, since all rows have a 1 position in them, it would follow that b_2 is inconsistent. However, $b_2 \in (C - b_3)$ so it would follow that b_2 is consistent, leading to a contradiction. Thus, b_3 cannot contain a repeated 1. So, $(C - b_3) \rightarrow b_3$. The same reasoning can be applied to column blocks.
Thus, C is a configuration since $(C - e) \rightarrow e$, for all $e \in C$. By Lemma 2.2 removing any of these elements from C would lead to cells being single covered by $(C - e)$. The remaining elements would then be unable to be verified by $(C - e)$. Thus, C is minimal. □

Configuration Type 1 is a special case because there are some subsets that could contain multiples of this type. Subsets that contain 2 rows sets and the blocks that intersect those rows or contains 2 columns sets with the blocks that intersect those columns, would not be considered configurations, see Figure 2.4. This is because of our definition of configuration sets require that these subsets be minimal. The same would be true if a subset contained only 3 row sets and the blocks that intersected those rows or 3 column sets with the blocks that intersect those columns, see Figure 2.5. Instead we would only say that those subsets contain configuration Type 1.

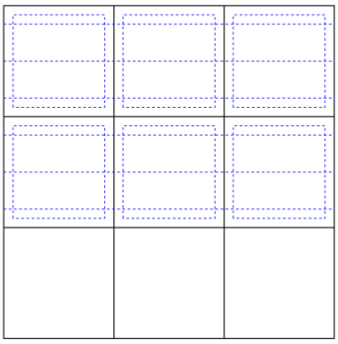

Figure 2.4: A subset that contains configuration Type 1 but is not a configuration set.

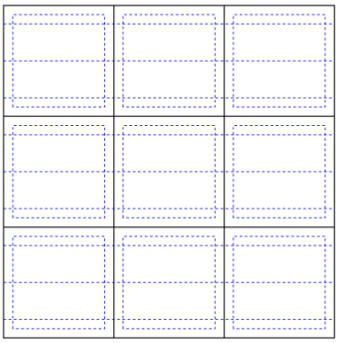

Figure 2.5: A subset that contain configuration Type 1 that is not a configuration set.

Now, we have to ask ourselves what would other subsets that would fit our definition of configuration sets look like? It would stand to reason that we can keep including a combination of row sets and column sets. Then by choosing to not include the block that intersect a row set and column set we can start to build our other configuration types. First, we will see and prove that an intersecting row set and column set with the blocks that only intersect one of those respectively will be a configuration set.

Configuration Type 2. Let C be the collection of one row set and one column set with the inclusion of the blocks that intersection with only one of those respective row sets and column sets.

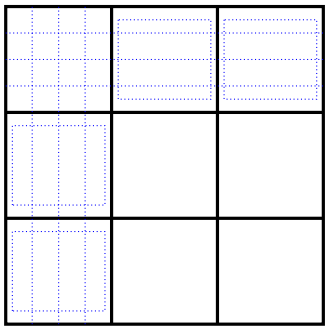

Figure 2.6: One example of configuration Type 2 where all consistent elements are seen in blue.

Proof. With out loss of generality let C contain the consistent elements from RS_1 and CS_1 with blocks b_2,b_3,b_4 and b_7. We can show that for all elements, e in C, $(C - e) \rightarrow e$. First, without loss of generality, we will look at $(C - r_1)$. Since $r_1 \notin (C - r_1)$, r_1 is

unverified. We can position two 1's somewhere in r_1. Since $\{c_1, c_2, c_3\} \subset (C - r_1)$ there has to be a 1 positioned somewhere in each of these elements. If both 1's are positioned in the intersections of r_1 and two of the consistent columns, without loss of generality let those columns be $c_1$and c_2, then a 1 still has to be position somewhere in blocks b_4 and b_7 since $\{b_4, b_7\} \subset (C - r_1)$. We can position a 1 in one of the cells that are in the intersection of c_3 and b_4. It would follow that b_7 is inconsistent, since there is no where to position a 1 since all three consistent columns that intersect that block already has a 1. However, $b_7 \in (C - r_1)$ leading to a contradiction.

If a 1 is position in the intersection r_1, one consistent column, and a consistent block that

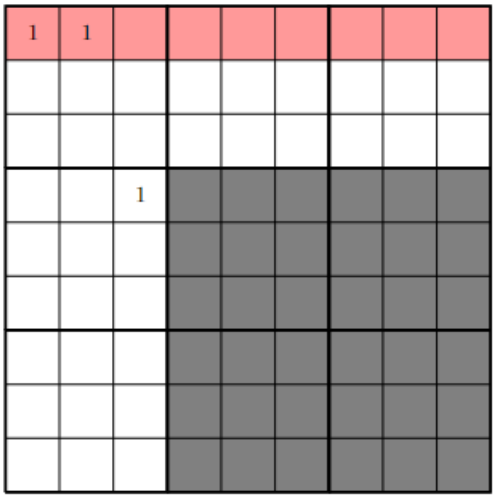

Figure 2.7: The entire Sudoku puzzle where $r_1 \notin (C - r_1)$ is seen in red and the elements not in C are seen in gray.

does not intersect with that consistent column, then we can follow the same reasoning as above that leads to the same contradiction.

If both 1's are positioned in the intersection of r_1 and the two consistent blocks, b_2 and b_3, then we would be able to position a 1 in r_2, specifically in the cells that are not in b_2 and b_3. It would follow that r_3 is inconsistent. However, $r_3 \in (C - r_1)$, leading to a contradiction. Thus, Thus $(C - r_1) \rightarrow r_1$.

Next, let us try to verify one of the blocks in C. Without loss of generality, let $C = C - b_2$. Since $b_2 \notin (C - b_2)$, therefore we can deduce that b_2 is unverified. We can position two 1s somewhere in b_2. Since $\{r_1, r_2, r_3\} \subset (C - b_2)$ we have to position the 1's in different rows.

Without loss of generality we can position a 1 in one of the cells in the intersection of r_1 and b_2 as well as in one of the cells in the intersection of r_2 and b_2. Since $b_3 \in (C - b_2)$, it must be assumed that b_3 is consistent. So, we can position a 1 in one of the cells in the intersection of r_3 and b_3. Next, we have to position a 1 somewhere in the intersections of c_1,c_2,c_3, b_4, and b_7. Without loss of generality we can position a 1 in one of the cells in the intersection of c_3 and b_4 and in one of the cells in the intersection of c_2 and b_7. It would follow that c_1 is inconsistent. However, $c_1 \in (C - b_2)$. Thus, $(C - b_2) \rightarrow b_2$.

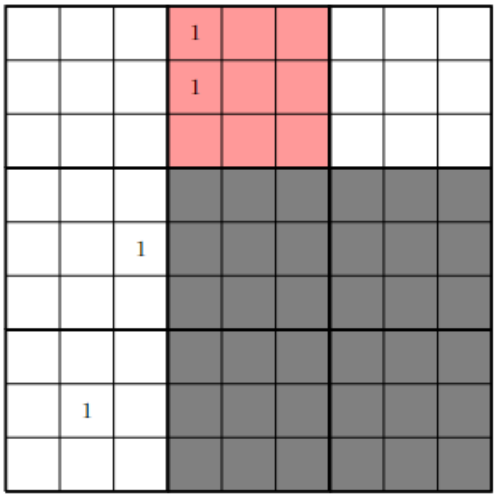

Figure 2.8: The entire Sudoku puzzle where $b_2 \notin (C - b_2)$ is seen in red and the elements not in C are seen in gray.

Since for all elements, $e \in C$, $(C - e) \rightarrow e$, C is a configuration.
By Lemma 2.2 removing any of these elements from C would lead to cells being single covered by $(C - e)$. The remaining elements would then be unable to be verified by $(C - e)$. Thus, C is minimal. □

We can continue this construction of these configuration types by intersecting various row sets and column sets and including the blocks that specifically intersect with only one of those sets.

Configuration Type 3. Let C be the collection of either two row sets and one column set or one row set and two columns with the inclusion of the blocks that intersect with only one of those respective row sets and column sets.

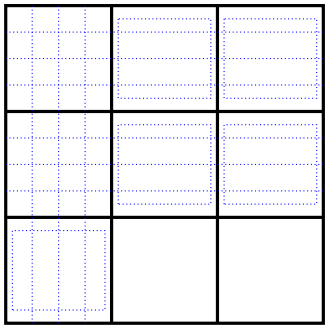

Figure 2.9: One example of configuration Type 3 where all consistent elements are seen in blue.

Proof. Without loss of generality let $\{r_1, r_2, r_3, r_4, r_5, r_6, b_2, b_3, b_5, b_6, b_7\} \in C$. Without loss of generality, we will show that $(C - r_4) \rightarrow r_4$. Since $r_4 \notin (C - r_4)$, implies that r_4 is unverified. So we can position two 1's somewhere in r_4. Following the same logic as above, we can position the two 1's in the intersections of the consistent columns, blocks, or one in each. Placing both 1's in just the intersection of the consistent columns and r_4 would lead to the same argument seen above, without needing to use all consistent elements. However, positioning a 1 in at least one of the consistent blocks would lead to use needing all other elements in $(C - r_4)$. We can position our two 1's in the intersection of c_1, r_4 and a 1 in one of the cells in the intersection of b_5, r_4. Then there still has to be a 1 positioned somewhere in r_5, r_6, and b_6. We can position a 1 in one of the cells in the intersection of r_5 and b_6. Then since $\{c_2, c_3\} \subset (C - r_4)$, means we have to position a 1 somewhere in these columns. First, position a 1 in the intersection of c_3 and r_6. Then since $b_7 \in (C - r_4)$, we can position a 1 in one of the cells in the intersection of c_2, b_7. Since $\{r_1, r_2, b_2, b_3\} \subset (C - r_4)$ each of these elements must have a 1 positioned somewhere in them. Without loss of generality we can position a 1 in one of the cells that make up the intersection of r_1, b_2 and in one of the cells that make up the intersection of r_2, b_3. It would follow that r_3 is inconsistent, since we cannot position a 1 anywhere in r_3 since our consistent elements that intersect with r_3 already have a 1 positioned in them. However, $r_3 \in (C - r_4)$. Leading to a contradictions. Thus, $(C - r_4) \rightarrow r_4$.

Next, without loss of generality, we will show that $(C - b_3) \rightarrow b_3$. Since $b_3 \notin (C - b_3)$, it can be deduced that b_3 is unverified. So, we can position two 1's in the cells that make up b_3. Since $\{r_1, r_2\} \subset (C - b_3)$, by assumption that means that r_1, r_2 are consistent.

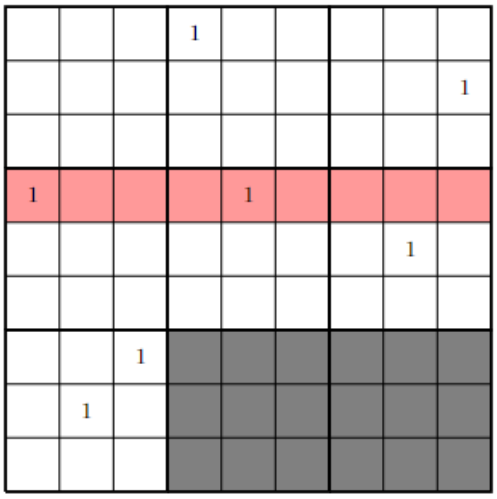

Figure 2.10: The entire Sudoku puzzle where $r_4 \notin (C - r_4)$ is seen in red and the elements not in $(C - r_4)$ are seen in gray.

Without loss of generality we can position a 1 in one of the cells in the intersection of r_1, b_3 and r_2, b_3. Since $\{r_3, b_2\} \subset (C - b_3)$, we can position a 1 in one of the cells in the intersection of r_3, b_2. Now, since $\{r_4, b_5, c_3, b_7\} \subset (C - b_3)$, we can position 1's in the intersections of these elements. Without loss of generality, we can position a 1 in one of the cells in the intersections of r_4, b_5, r_5, b_6, and c_3, b_7. Then since $\{r_6, c_2\} \subset (C - b_3)$, we can position a 1 in the intersection of r_6 and c_2. It follows that c_1 is inconsistent, since we are unable to position a 1 anywhere in c_1. However, $c_1 \in (C - b_3)$. Leading to a contradiction. Thus, $(C - b_3) \to b_3$.

Thus, for every element, e, in C, $(C - e) \to e$. Thus, C is a configuration.
By Lemma 2.2 removing any of these elements from C would lead to cells being single covered by $(C - e)$. The remaining elements would then be unable to be verified by $(C - e)$. Thus, C is minimal. □

Configuration Type 4. Let C be the collection of two row sets and two column sets with the inclusion of the blocks that intersect with only one of those respective row sets and column sets.

Proof. Without loss of generality let $C = \{r_1, r_2, r_3, r_4, r_5, r_6, c_1, c_2, c_3, c_4, c_5, c_6, b_3, b_6, b_7, b_8\}$. Without loss of generality, we will show that $(C - c_4) \to c_4$. Since $c_4 \notin (C - c_4)$, it can

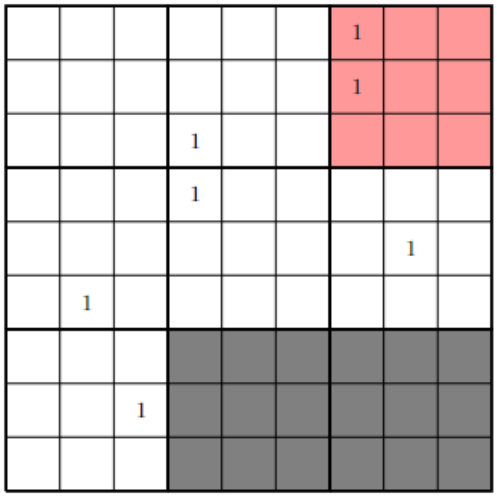

Figure 2.11: The entire Sudoku puzzle where $b_3 \notin (C - b_3)$ is seen in red and elements not in $C - b_3$ are seen in gray.

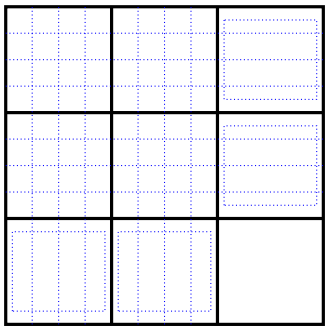

Figure 2.12: One example of configuration Type 4 where all consistent elements are seen in blue.

be deduced that c_4 is unverified. So, we can position two 1's somewhere in c_4. Since $\{b_8, r_1\} \subset (C - c_4)$, we can position a 1 in the intersection of c_4 and r_1 as well as one of the cells in the intersection of c_4 and b_8. Now, $\{b_3, b_6, b_7\} \subset (C - c_4)$, so each has to have a 1 positioned somewhere in the blocks. Notice that each of these blocks are composed of either consistent rows or columns, so when we position these 1's, we also position them in a consistent row or column respectively. Now, we only have consistent rows and columns left to position a 1 into. Notice that each consistent row intersects with a consistent column, so when we position a 1 in consistent row then we also position a 1 in that intersecting column. So, there are a remainder of 3 consistent rows that intersect with 4

consistent columns. We can position a 1 in the cells that intersect these consistent rows and columns. Leading to a column, without loss of generality c_6, being inconsistent, since we cannot position a 1 anywhere in c_6. However, $c_6 \in (C - c_4)$ leading to a contradiction. Thus, $(C - c_4) \to c_4$.

Next, we will show that $(C - b_8) \to b_8$. Since $b_8 \notin (C - b_8)$, it can be assumed that b_8 is

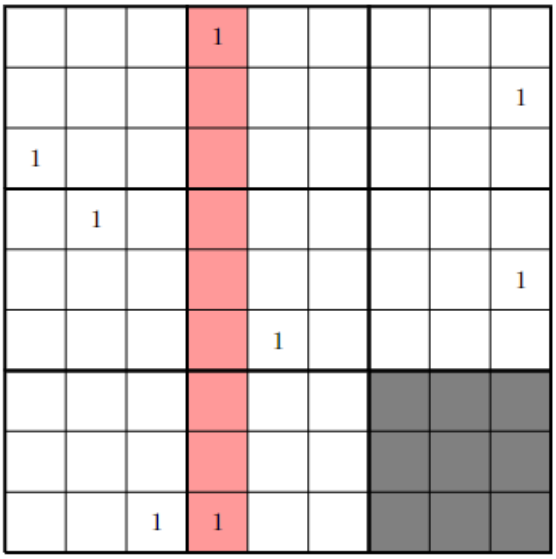

Figure 2.13: The entire Sudoku puzzle where $c_4 \notin (C - c_4)$ is seen in red and the elements not in C are seen in gray.

unverified. So, we can position two 1's in b_8. Since $\{r_1, r_4\} \subset (C - b_8)$, we can position a 1 in the cells in the intersection of those respective rows and consistent blocks. Next, since $b_7 \in (C - b_8)$ we can position a 1 in each block. Just as above, since there are 5 consistent columns but 4 consistent rows remaining after position a 1 in each of the consistent blocks, without loss of generality, we are unable to position a 1 in c_1, leading to c_1 being inconsistent. However, $c_1 \in (C - b_8)$, which is a contradiction. Thus $(C - b_8) \to b_8$.

Thus, for every element,e, in C, $(C - e) \to e$. Thus, C is a configuration. By Lemma 2.2 removing any of these elements from C would lead to cells being single covered by $(C - e)$. The remaining elements would then be unable to be verified by $(C - e)$. Thus, C is minimal. □

Configuration Type 5. Let C be the collection of either three row sets and one column set or three column sets and one row set with the inclusion of the blocks that

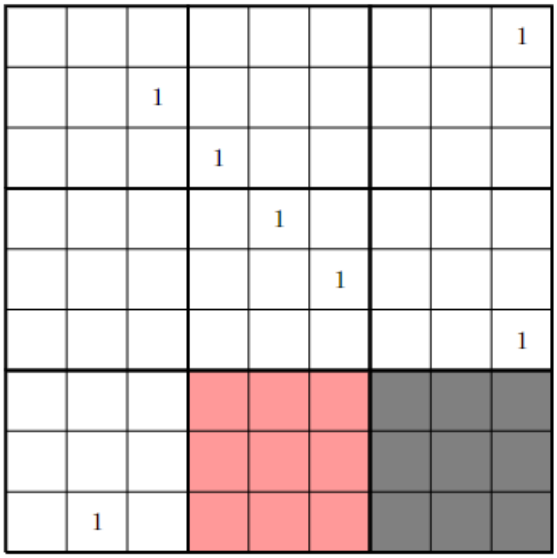

Figure 2.14: The entire Sudoku puzzle where $b_8 \notin (C - b_8)$ is seen in red and the elements not in $(C - b_8)$ as seen in gray.

intersect with only one of those respective row sets and column sets.

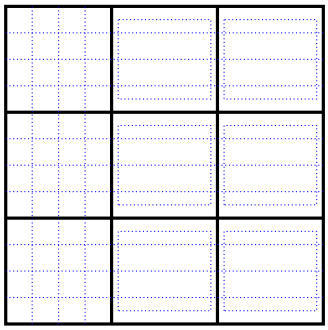

Figure 2.15: One example of configuration Type 5 where all consistent elements are seen in blue.

Proof. Without loss of generality let C contain $r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, c_1, c_2, c_3, b_2, b_3, b_5, b_6, b_8, b_9$.
Without loss of generality, first we will show that $(C - r_1) \rightarrow r_1$. Since $r_1 \notin (C - r_1)$, it can be assumed that r_1 is unverified. So, we can position two 1's in r_1. Positioning both 1's in the unverified block, would lead to us not needing all of the other elements

in $(C - r_1)$. So, since $b_3 \in (C - r_1)$, we will assume that one of the repeated 1's be in the intersection of r_1 and b_3. Without loss of generality, we will position a 1 in one of the cells in the intersection of r_1 and b_3. Then since $c_1 \in (C - r_1)$, c_1 is consistent, so we can position a 1 in the intersection of r_1, c_1. Next, since $\{b_2, b_5, b_6, b_8, b_9\} \subset (C - r_1)$ we can position a 1's in each of the other consistent blocks. Notice that each of these blocks intersect with either consistent row sets or consistent column sets. So, when we position 1's in each of these blocks, there are a remainder of 3 consistent rows and 2 consistent columns. Each of these consistent rows intersect with a consistent column. Without loss of generality, r_9 would be inconsistent, since we cannot place a 1 in r_9. However, $r_9 \in (C - r_1)$, leading to a contradiction. Thus, $(C - r_1) \rightarrow r_1$.

Next, we will show that $(C - b_3) \rightarrow b_3$. Since $b_3 \notin (C - b_3)$, b_3 is inconsistent. So,

1								1
			1					
	1							
			1					
						1		
		1						
			1					
						1		

Figure 2.16: The entire Sudoku puzzle where r_1 is seen in red.

lets position two 1's in b_3. Notice that b_3 intersect with the consistent rows in RS_1. So, without loss of generality lets position a 1 in one of the cells in the intersection of r_1, b_3 and in one of the cells in the intersection of r_2, b_3. Next, since $\{b_2, b_5, b_6, b_8, b_9\} \subset (C - b_3)$, those blocks are consistent. So, we can position a 1 in each of these blocks. Notice that each of these blocks also intersect with consistent row sets or column sets. Lastly, we have to position a 1 in the remaining 3 consistent columns and 2 consistent rows, however, notice that these columns intersect with these consistent rows. Without loss of generality,

c_1 is inconsistent since we cannot position a 1 anywhere in c_1. However, $c_1 \in (C - b_3)$, leading to a contradiction. Thus, $(C - b_3) \to b_3$.

Thus, for every element, e, in C, $(C - e) \to e$. Thus, C is a configuration.

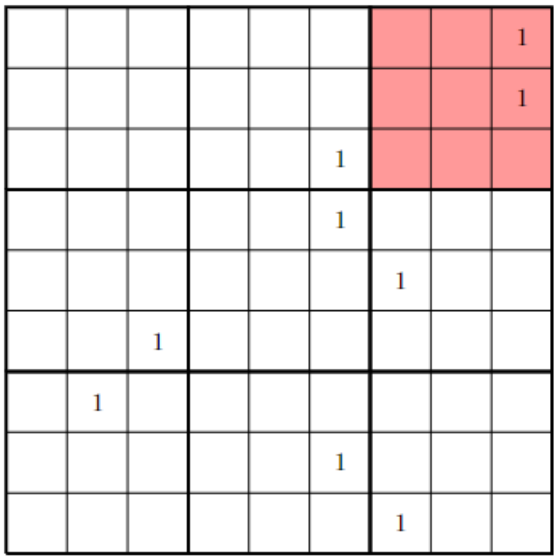

Figure 2.17: The entire Sudoku puzzle where b_3 is seen in red.

By Lemma 2.2 removing any of these elements from C would lead to cells being single covered by $(C - e)$. The remaining elements would then be unable to be verified by $(C - e)$. Thus, C is minimal. □

Configuration Type 6. Let C be the collection of either three row sets and two column sets or two row sets and three column sets with the inclusion of the blocks that intersect with only one of those respective row sets and column sets.

Proof. Without loss of generality, let $C = r_1, r_2, r_3, r_4, r_5, r_6, c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, b_7, b_8, b_9$. Without loss of generality, first we will show that $(C - r_1) \to r_1$. Since $r_1 \notin (C - r_1)$, r_1 is unverified. So, we can position two 1's somewhere in r_1. Without loss of generality, since $\{c_1, c_2\} \subset (C - r_1)$ we can position a in the intersecting cells of r_1, c_1 and r_1, c_2. Next, since $\{b_7, b_8, b_9\} \subset (C - r_1)$, those blocks are consistent. Lets position a 1 in each of these blocks, and notice that each block is composed of consistent column sets. This means that when we position a 1 in each block, we also have to place a 1 in a consistent column. This means that we have 5 consistent rows and 4 consistent columns that we have to

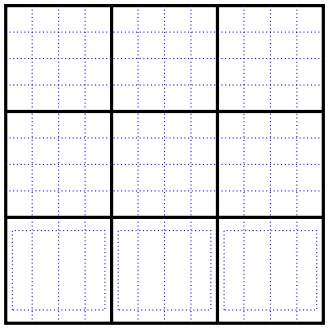

Figure 2.18: One example of configuration Type 6 where all consistent elements are seen in blue.

position 1's into. Each 1 must be in the intersection of a consistent row and column. Without loss of generality, it would follow that r_6 is inconsistent. However, $r_6 \in (C - r_1)$, leading to a contradiction. Thus, $(C - r_1) \rightarrow r_1$.

Next, we will show that $(C - b_7) \rightarrow b_7$. Since $b_7 \notin C - b_7$, b_7 is unverified. So, we

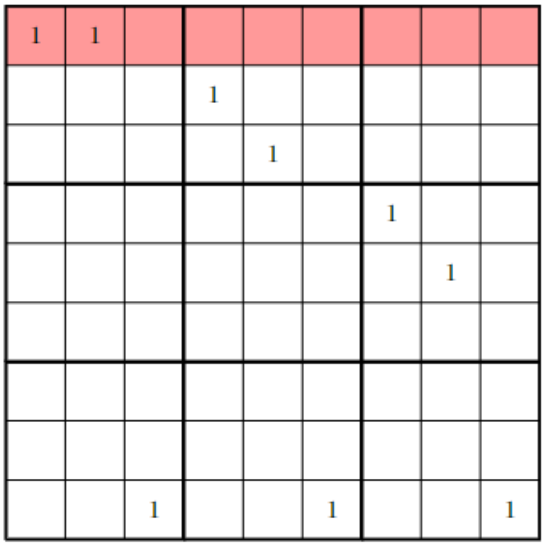

Figure 2.19: The entire Sudoku puzzle where r_1 is seen in red.

can position two 1's somewhere in b_7. Notice that b_7 intersects with a consistent column set, CS_1. Without loss of generality, since $\{c_1, c_2\} \subset (C - b_7)$, c_1, c_2 are consistent. So, lets position a 1 in one of the cells in the intersection of c_1, b_7 and in c_2, b_7. Next, since $\{b_8, b_9\} \subset (C - b_7)$, both are consistent and we can place a 1 in each of these blocks.

Notice that these blocks intersect with consistent column sets, so position a 1 in a block also positions a 1 in a consistent column. Now, we have a remaining 6 consistent rows and 5 consistent columns that we can position 1's in. Notice that each of the consistent rows intersects with the consistent columns, so when we position a 1 in a consistent row, we also position a 1 in a consistent column. Leading to c_6 being inconsistent. However, $c_6 \in (C - b_7)$, leading to a contradiction. Thus, $(C - b_7) \to b_7$.

Thus, for every element,e, in C, $(C - e) \to e$. Thus, C is a configuration.

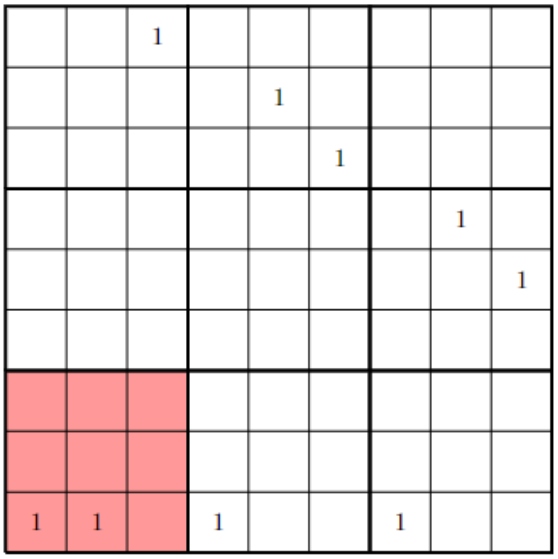

Figure 2.20: The entire Sudoku puzzle where $\{b_7\}$ is seen in red.

By Lemma 2.2 removing any of these elements from C would lead to cells being single covered by $(C - e)$. The remaining elements would then be unable to be verified by $(C - e)$. Thus, C is minimal. □

Configuration Type 7. *C is the collection of all three row sets and three column sets.*

Proof. Let C contain all the rows and columns. Without loss of generality, we will look at $(C - c_9)$. Since $c_9 \notin (C - c_9)$, it follows that c_9 is unverified. So, we can position a repeated 1 somewhere in c_9. First, we will look at the case of a repeated 1 in the same block. Without loss of generality we will position a 1 in the cell $(1, 9, 3)$ and the cell $(2, 9, 3)$. Since $\{r_3, c_8\} \subset (C - c_9)$ it follows that those elements are consistent so we can

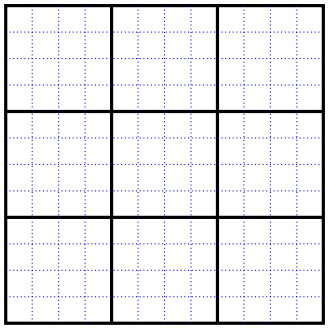

Figure 2.21: One example of configuration Type 7 where all consistent elements are seen in blue.

place a 1 in their intersection. We can follow the same reasoning for positioning a 1 in r_4, c_7. In fact we can follow this pattern of placing a 1 in the intersection the bottom row and left column of the previously placed 1. So, our rows and columns will look like:

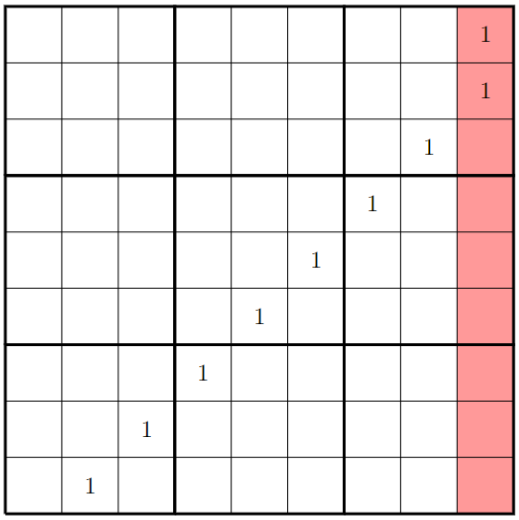

Figure 2.22: The entire Sudoku puzzle where c_9 is seen in red.

Since we can have 8 columns and 7 rows to position these 1's in, it would follow that one column would be inconsistent, here we have c_1 as inconsistent. However, $c_1 \in (C - c_9)$ leading to a contradiction. So, c_9 cannot contain a repeated number. It follows that $(C - c_9) \rightarrow c_9$. This same logic can be applied if C did not contain a row.
Thus, for every element,e, in C, $(C - e) \rightarrow e$. Thus, C is a configuration.

Notice that removing any element from C would lead to cells being single covered by $(C-e)$. By Lemma 2.2the remaining elements would then be unable to be verified by by $(C-e)$. Thus C is minimal. □

This is where our pattern ends. There are no more combinations of row sets or column sets left to consider. So, now we can ask ourselves how do we know these are all the configuration sets? Since there are only three row sets and three column sets in a Sudoku puzzle, that means that there are only 9 total possible subsets that are a combination of row sets and column sets. However, since our definition of configuration sets do not include those subsets that contain either just 2 row sets or column sets or 3 row sets or column sets, there would only be 7 total configuration types possible. Now that we have defined what our configuration sets look like we can describe some of the properties that these subsets have.

2.2 Properties of Configuration Sets

We now have a way to categorize these configuration sets into types, but we want to be able to distinguish if a subset is one of these configuration types. The question is how do we do that. Would we have to remove every element from that subset and try to verify it in order to say that this set is a configuration set? First, we need to explore what characteristics subsets must have in order to be categorized as one of our configuration types.

Lemma 2.3. *Let $X \subset K$ where all the cells in X are double covered by X. Then X must contain at least one row set or column set. Moreover, either*

(i) X has at least one row set and at least one column set. In this case X must be one of the seven configuration types; or

(ii) X has zero column sets or zero row sets. In this case X musts contain configuration Type 1

Proof. Let $X \subset K$ where all cells in X are double covered by X. In order to have a cell be double covered by X, there are three possibilities: a cell is in the intersection of a consistent row and column, it is in the intersection of a consistent row and block, or

it is in the intersection of a consistent column and block. If a cell (i, j, k) composes an element in X then it has to be in the intersection of one of the three previous choices. If the cell (i, j, k) is one of the cells that compose row r_i and block b_k then that means that $\{b_k, r_i\} \in X$, furthermore there are three cells in r_i that are double covered by X since three cells compose the both of r_i, b_k. Furthermore, the cells of b_k also compose two rows in the same row set as r_i. Since all cells in X must be double covered by X, the other two rows in the row set of r_i must be in X, since adding any columns would make the three cells composing r_i and b_k to be triple covered by X. Thus, X contains at least one row set when the cells in X are double covered by X. Notice that if the cell (i, j, k) is one of the cells that compose a column c_j and a block b_k we could us the same reasoning, so X must contain at least one columns set when the cells in X are double covered by X.

Lastly, if the cell (i, j, k) is the cell in the that make up the row r_i and the column c_j there exists at least 4 cells that are single covered by X, the cells in the column c_j that intersect with the other rows in the row set that r_i belongs in and the cells in r_i that intersect with the column set that c_j belongs to. Since all cells in X have to be double covered by X, and we cannot add b_k, we have to add the other rows in the same row set as r_i and the columns in the same column set as c_j. Thus, X must contain at least one row set or column set when all the cells in X are double covered by X.

Now, we have a total of three row sets and three column sets that can be in X. That means we have a total of 9 possible combinations of row sets and column sets. By the rotational symmetry of a square, we can look at our possible combinations of row sets and column sets.

If X consists of three column sets and three row sets then all cells would be double covered by X since each cell is in the intersection of a consistent row and column, furthermore we would not be able to have any consistent blocks in X since that would lead to triple covered by X cells. This gives us configuration Type 7. Since we are choosing all row blocks and all column blocks, there is only one way for Type 7 to look.

If X consists of three row sets and two column sets, then in the cells in the intersection of the three row sets and two column sets would be double covered by X. However, the cells that compose the row sets that do not intersect with the consistent column set would be single covered by X. Since X only has double covered cells, X must have the blocks that these single covered cells compose. Leading to configuration Type

6.

If X consists of three row sets and one column set then we can follow the same reasoning as above. Now, X must contain the blocks composed of the cells that compose the rows that do not intersect with the consistent columns. Leading to configuration Type 5.

If X consists of two row sets and two column sets then there will be cells that are single covered by X. Specifically the cells that are not in the intersection of the consistent row sets and column sets. So, X has to include the blocks that these single covered cells compose. Leading to configuration Type 4.

If X consists of two row sets and one column set then there are some cells that are single covered by X, specifically the ones not in the intersection of the row set and column set. So, the blocks that are composed of the single covered cells must be in X. Leading to configuration Type 3.

If X consists of one row set and one column set, then there are some cells that are single covered by X, specifically the ones not in the intersection of the row set and column set. X must contain the blocks that are composed of the single covered cells. Leading to configuration Type 2.

If X consists of one row set and no column sets, then all the cells that compose the rows in the row set are single covered by X. Since the no columns can be added, X must contain the three blocks that are composed by these single covered cells. Leading to configuration Type 1.

Now, assume that X contains 2 rows sets and no column sets. However, notice that then X would have to include all 6 blocks that intersect with those row sets. However, this would mean that this subset contain two disjoint configuration type 1's. If X contains 3 row sets then X would have to include all the blocks. Meaning X also contain three disjoint configuration Type 1s as a subset. Since there are no other possible combinations of 3 row sets and 3 column sets, X either has to be one of the seven configuration sets or contains configuration set Type 1. □

Now we know that having double covered cells that are covered by the elements of a subset means that this subset must be one of our configuration types or contains configuration type one but we can take this a step further. Based on how we defined these configuration types we can make the argument that these subsets that contain only

double covered cells must also have the same property as configuration sets. We will formally define that next.

Lemma 2.4. *If $X \subset K$ where every cell is exactly double covered by X then for all $e \in X$, $(X - e) \rightarrow e$.*

Proof. From Lemma 2.3 we know that if X contains all double covered cells then X must look like one of the seven configuration types or X contains configuration Type 1. By definition of configuration sets, for any $e \in X$, $(X - e) \rightarrow e$. □

Now, we can tie these two previous lemmas together to clearly state and show that when a subset contains only double covered cells covered by a set then that subset is a configuration set or contains configuration Type 1.

Lemma 2.5. *If $X \subset K$ in which every cell is exactly double covered by X then X is a configuration or contains configuration Type 1.*

Proof. If all the cells in X are double covered by X then from Lemma 2.3 we know that X must contain at least one row set and one column set and that X will look like one of the configuration types or contain configuration Type 1. Furthermore we know from Lemma 2.4 that each of the elements in X, when removed, can be verified by the other elements. By definition that means X is a configuration. □

We already know, by Lemma 2.2, that if there are cells that are singularly covered by X, then that subset cannot be a verification set, let alone a full verification set. Now, we have seen what happens if a subset only contains double covered cells. So, the next question is what happens if there are triple covered cells? Is this subset a configuration set or does the addition of more elements rule out this subset from being one?

Lemma 2.6. *If $X \subset K$ that has no cells that are single covered by X, but does have cells triple covered by X, then X contains a configuration.*

Proof. Let X have no single covered cells. That means that X must contain double covered and triple covered cells. Let the cell (i, j, k) be a triple covered cell in X. This means that the cell (i, j, k) would be in the intersection of a r_i, c_j and b_k, that is in X. By Lemma 2.3 since X has cells that are double covered X must contain all the rows in that

row set and columns in that column set. Thus, all the cells that compose b_k are triple covered. So, $X - \{b_k\}$ would have double covered cells. This process can be repeated for all the blocks that are composed of triple covered cells in X, leading to a subset of X with only double covered cells. By Lemma 2.5, this subset would be a configuration. Thus, X contains a configuration. □

Now we have a way to determine if a subset is a configuration type by looking at the cell coverage. But, what about configuration sets themselves? Does this property about double covered cells still hold for our configuration sets? We will prove that in fact a configuration set will have that same property.

Lemma 2.7. *If C is a configuration, then every cell in X is double covered by X.*

Proof. Let C be a configuration. Assume C contains any cells single covered by X then by Lemma 2.2, C could not verify all elements in C, so C would not be a configuration, which is a contradiction. Next, assume that C contains cells that are triple covered by X, then C would contain a configuration by Lemma 2.6. By definition of configurations, C would not be a configuration, which is a contradiction. Thus, C can only contain cells that are double covered by X. □

This means that we are able to both determine if a subset is a configuration set and describe configuration sets by only examining the coverage of the cells that compose its elements. We can put together our previous lemma's for our first theorem.

Theorem 2.8. *A set $C \subset K$ is a configuration, or contains configuration type 1, if and only if every cell in C is exactly double covered by X.*

Proof. Let C be a configuration. Then by Lemma 2.7 every cell in C is double covered by X.
If every cell in C is double covered by X then by Lemma 2.4 then C is a configuration or contains configuration type 1.
Thus, C is a configuration or contains configuration type 1 if and only if every cell in C is double covered by X. □

Now, one may be wondering how will this help us with out goal of finding a full verification set? These configuration sets seem to only really help us with verifying removed elements from that subset. The answer lies within matroid theory.

Chapter 3

Introduction to Matroids

A major goal of this thesis is to view Sudoku puzzles through the lens of matorid theory. In this chapter, we will introduce the concept of a matroid. One of the advantages of studying matroids is the various equivalent perspectives from which we can develop the theory. We will explore several of these perspectives in this chapter and provide connection between these perspectives. In chapter 4, we will see how Sudoku puzzles and matroid theory are connected and see how the matroid theory tools we develop in this chapter will aid in our investigation of Sudoku puzzle verification.

The beginnings of matroids came in 1935 when Hassler Whitney worked to abstract the several mathematical concepts of dependence that arise in subjects such as graph theory and linear algebra. In fact, one may think of matroid theory as the collection of ideas from a wide variety of other mathematical fields including graph theory, linear algebra, combinatorics and geometry. This gives matroid theory the unique characteristic of being able to connect different concepts together through various equivalency statements. First, we will define a matroid based of independence.

3.1 Independent Sets of a Matroid

A matroid starts with a collection of elements that we call the ground set. Studying special subsets of the ground set distinguishes one matroid from another. One such special collection of subsets are called independent sets. In order for a collection of sets to be the collection of independent sets of a matroid, that collection must adhere to the following axioms:

Definition 3.1 ([GM12], definition 2.1)**.** *Let E be a finite set and let $\mathcal{I}$ be a family of subsets of K. Then the family $\mathcal{I}$ forms the independent sets of a matroid M if $\mathcal{I}$ satisfies:*

(I1) $\mathcal{I} \neq \emptyset$

(I2) if $J \in \mathcal{I}$ *and* $I \subset J$*, then* $I \in \mathcal{I}$

(I3) if $I, J, \in \mathcal{I}$*, with* $|I| < |J|$*, then there is some element* $x \in J - I$ *with* $I \cup \{x\} \in \mathcal{I}$*.*

A subset that is not independent is then called *dependent*. Matroid theory has a very special property where we are able to convert from one system to another. This ability to use equivalent definitions to define a matroid is called *cryptomorphism* and this is how we will be able to not only show we have a matroid but be able to describe, to an extent, some of the components of this matroid. We will first start of by formally proving the equivalency between subsets called circuits and independent sets.

3.2 Matroids Through the Lens of Circuits

A minimal dependent set of a matroid is called a *circuit*. Another way to describe circuits is by verifying that every proper subset of that dependent set is independent. By minimally dependent, we mean that the set is dependent, but every proper subset is independent. First formally define what a collection of circuits are and the properties that they have.

Theorem 3.2 ([GM12])**.** *Let K be a finite set and let $\mathcal{C}$ be a family of subsets of K. The family $\mathcal{C}$ forms the circuits of a matroid M, with independent sets $\mathcal{I}$, if:*

(C1) $\emptyset \notin \mathcal{C}$

(C2) if $C_1, C_2 \in \mathcal{C}$ *and* $C_1 \subset C_2$*, then* $C_1 = C_2$

(C3) if $C_1, C_2 \in \mathcal{C}$ *with* $C_1 \neq C_2$*, and* $x \in C_1 \cap C_2$*,*
then $C_3 \subset (C_1 \cup C_2) - x$ *for some* $C_3 \in \mathcal{C}$*.*

Proof. Assume that $\mathcal{C}$ is a family of circuits. First, we will prove $(C1)$. Notice that by our definition of independent sets, specifically Axiom $(I1)$, $\mathcal{I}$ is non trivial. So there exists $A \subset K$ such that $A \in \mathcal{I}$. By Axiom $(I2)$ that means that every subset of A is also in $\mathcal{I}$. Since $\emptyset \subset A$, that implies that $\emptyset \in \mathcal{I}$. Thus $\emptyset$ is an independent set and cannot be a circuit. Therefore, we can deduce that $\emptyset \notin \mathcal{C}$.

Secondly, we will prove Axiom $(C2)$ by assuming that $C_1, C_2 \in \mathcal{C}$ and $C_1 \subset C_2$, but $C_1 \neq C_2$. Then, by definition of a circuit, all proper subsets of C_2 are independent. Then C_1 must be independent, so $C_1 \notin \mathcal{C}$, leading to a contradiction. Thus, $C_1 = C_2$.
Lastly, we will prove Axiom $(C3)$. Assume that there exist some element x such that $x \in C_1 \cap C_2$. From the definition of circuits, that means that $C_1 - x$ and $C_2 - x$ are independent sets. Assume that $(C_1 \cup C_2) - x$ does not contain any circuits and is therefore independent. From Axiom $(C2)$ we know that $C_1 \not\subset C_2$, thus there exists $a \in C_2 - C_1$ where $C_2 - a$ is independent by the definition of a circuit. Assume that there exists a maximal independent set $B \subset C_1 \cup C_2$ where $C_2 - a \in B$. Thus, we can deduce that $a \notin B$. Then there exists some $y \in C_1 - B$. Notice that a and y are distinct since $a \in C_2 - C_1$. Therefore,

$$|B| \leq |(C_1 \cup C_2) - \{a, y\}| = |(C_1 \cup C_2) - 2| < |C_1 \cup C_2| - e$$

Now, we can apply axiom $(I3)$ to the sets B and $C_1 \cup C_2 - e$ to find some element $z \in (C_1 \cup C_2 - e) - B$ so that $B \cup z \in \mathcal{I}$. This contradicts the assumption that B is a maximal independent set. Thus, Axiom $(C3)$ holds. □

One may be wondering how do we connect circuits back to independent sets. The answer lies in a special property that matroids have, cryptomorphism. Using various cryptomorphisms we are able to convert from one system of defining a matroid to another. Specifically we can state the connection between circuits and independent sets.

Theorem 3.3 ([GM12]. Theorem 2.10)**.** *Let E be a finite set and let $\mathcal{C}$ be a family of subsets of E satisfying (C1),(C2),(C3). Then $(E, \mathcal{C})$ is cryptomorphic to the matroid $M = (E, \mathcal{I})$ and $\mathcal{C}$ is the collection of circuits of M*

The proof to this cryptomorphism requires two things. The first being that a collection of independent sets of a matroid, $\mathcal{I}$, can lead to a collection of subsets satisfying the axioms $(C1) - (C3)$. We have already done that with the proof of Theorem 4.6. Lastly, we would have to show that the collection of circuits of a matroid, $\mathcal{C}$, can lead to a collection of subsets satisfying the axioms $(I1) - (I3)$. The proof of which can be found in reference [Oxl92], specifically his proof for Theorem 1.1.4. With these two things we are able to prove this cryptomorphism.
There is a strong connection between independent sets and circuits. As mentioned previously, circuits are the minimal subsets of a ground set that are not independent. From

this, we can also think of independent sets as being subsets of our ground set that do not contain any circuits. This means that if we are able to find what the circuits of our Sudoku puzzle look like, we can describe what the independent sets will look like as well. So far we have discussed what these independent sets and circuits do in terms of a Sudoku puzzle, however, neither have fully described the subsets that we are trying to find, full verification sets. Knowing what these specific subsets look like would help us answer our main question, how many elements does one need to claim a Sudoku puzzle is correct. That is where the bases of a matroid come into play.

3.3 Independent Sets and Bases

If M is a matroid with independent sets $\mathcal{I}$,then B is a basis of the matroid M if B is a maximal independent set. This means that if any element, $e \in K - B$ is added to B, then $B \cup e$ would be dependent. Furthermore, we can show that a collection of subsets form the bases of our matroid if those subsets meet three specific criteria.

Theorem 3.4 ([GM12]). *Let K be a finite set and let $\mathcal{B}$ be a family of subsets of K. The family $\mathcal{B}$ forms the bases of a matroid M, with independent sets $\mathcal{I}$, if:*

(B1) $\mathcal{B} \neq \emptyset$

(B2') If $B_1, B_2 \in \mathcal{B}$, then $|B_1| = |B_2|$.

(B3) If $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 - B_2$, then there is an element $y \in B_2 - B_1$ so that $(B_1 - x) \cup \{y\} \in \mathcal{B}$

Proof. Assume that $\mathcal{I}$ is a family of of independent subsets. Then assume that $\mathcal{B}$ is a family of maximal subsets of $\mathcal{I}$. First, we will show that axiom $(B1)$ holds. By axiom $(I1)$ we know that $\emptyset \in \mathcal{I}$. Since K is finite, we are able to find some subset $B \subset K$ such that $B \in \mathcal{I}$ but where B is maximal. Thus, $B \in \mathcal{B}$.

Second we will show axiom $(B2')$ holds. Assume that $B_1, B_2 \in \mathcal{B}$, then $|B_1| \neq |B_2|$. Without loss of generality assume that $|B_1| < |B_2|$. By definition of a basis we know that $B_1, B_2 \in \mathcal{I}$. By axiom $(I3)$ that means there exists some element $e \in B_2 - B_1$ where $B_1 \cup e \in \mathcal{I}$. However, $B_1 \subset B_1 \cup e$ meaning B_1 is not maximal. Therefore, $B_1 \notin \mathcal{B}$ leading to a contradiction. Thus, $|B_1| = |B_2|$.

Finally, we will show that axiom $(B3)$ holds. Assume that $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 - B_2$.

From axiom $(B2')$ we know that $|B_1 - x| < |B_2|$. We know that $B_1 - x$ is an independent set. Applying axiom $(I3)$, we know that there must exist some element $y \in B_2 - (B_1 - x)$ such that $(B_1 - x) \cup y \in \mathcal{I}$. Further, we know that $|(B_1 - x) \cup y| = |B_1|$. Now, we have to show that $(B_1 - x) \cup y$ is maximal. Assume that $(B_1 - x) \cup y$ is not maximal. Then there must exists some $I \in \mathcal{I}$ where $|(B_1 - x) \cup y| < |\mathcal{I}|$. Then by axiom $(I3)$ there must exist some element $e \in I - (B_1 - x) \cup y$ where $(B_1 - x) \cup \{y, e\} \in \mathcal{I}$. But, $|(B_1 - x) \cup \{y, e\}| > |B_1|$ which would mean that B_1 is not maximal and is not a basis, leading to a contradiction. Thus, $(B_1 - x) \cup y$ must be maximal. Therefore, $(B_1 - x) \cup y \in \mathcal{B}$.
Thus, all three axioms hold. □

These theorems and definitions are what we will be using in the following chapter to show that the elements of our ground set form a matroid. Moreover, we will use these theorems to answer the main question of this thesis, how many elements does one have to check in a completed Sudoku puzzle in order to know that the entire puzzle is correct?

3.4 Example of a Matroid

Now that we have discussed some important definitions and theorem of a matroid, we should look at an example. More specifically we will first examine a matroid represented by a graph and then a matroid represented by a matrix.

Example 1.

First, we will look at a matroid as a graph, see Figure 3.1. Now, we can describe the dependent sets as being the cycles of the graph. More specifically we can describe these subsets by the edges of the graph. The dependent sets would include $\{e_1, e_2, e_4\}$, $\{e_2, e_3, e_5\}$ and $\{e_1, e_3, e_6\}$. We can then describe some of the independent sets as those sets that do not contain these cycles. For example $\{e_1, e_5\}$ and $\{e_2, e_3, e_6\}$ would both be independent sets. Further, we can find the subsets that are basis by looking for the spanning trees of this graph. For example, $\{e_1, e_2, e_3\}$ would be a basis.

This matroid is a special case because we can express this matroid both as a graph and as a binary matrix. To create this binary matrix the columns of that matrix would be represented by the edges of the graph. While the rows of the matrix will be represented by the vertices of the graph. Now, we will position a 1 in the intersection of a row and column when an edge connects one vertex to another. For example, e_4 connected vertices

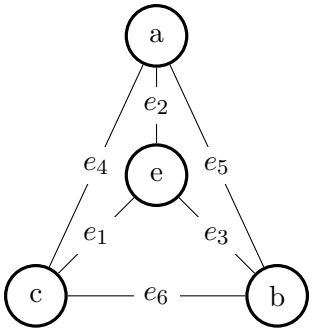

Figure 3.1: A graph that represents a matorid.

a and c. So, we would position a 1 in the the intersection of the column represented by e_4 and the rows that represent vertices a and c. See Figure 3.2. From the matrix represented

$$\mathbf{A} = \begin{array}{c} \\ a \\ b \\ c \\ e \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Figure 3.2: A binary matrix A that represents the graph from Figure 3.1.

by Figure 3.2 we can row reduce it to get the matrix represented by Figure 3.3.

From here we can find the independent sets, bases, and circuits of the matrix represented

$$\mathbf{A} = \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Figure 3.3: The row reduced matrix represented by Figure 3.2.

by Figure 3.3 to be the same as those we found from Figure 3.1.

Chapter 4

Sudoku Puzzles and Matroid Theory

Now that we have described these configuration types we have to ask ourselves how does this connect to one of the properties in matroid theory. In this chapter we are going to show that the these configuration sets that we defined in chapter 2 are the circuits of a matroid where the ground set of that matroid are all the elements of a Sudoku puzzle. In order for us to prove this, we are going to need to prove more properties about these configuration sets. Specifically we will see what happens if we union two configuration sets.

4.1 Symmetric Difference

The symmetric difference of two configuration sets, denoted $C_1 \Delta C_2$, means that we are looking at the union of C_1 and C_2 with the removal of all elements in $C_1 \cap C_2$. For example, if I wanted to take the symmetric difference of a configuration set of Type 2 with one that is Type 3, then the symmetric difference would only include the elements in the union of those subsets but not the elements in the intersection, see Figure 4.1. Just like we did with configuration types, there is some significance to the coverage of cells. We will first prove that the symmetric differences of configuration sets cannot have cells that are triple covered by the symmetric difference of two configuration sets.

Lemma 4.1. *If $C = C_1 \Delta C_2$ where C_1 and C_2 are configurations, then C has no triple*

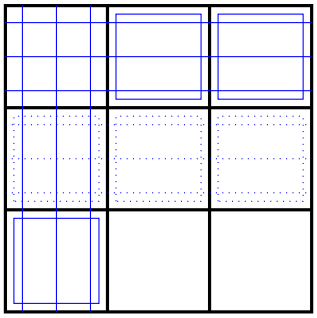

Figure 4.1: The symmetric difference of a Type 2 configuration set and a Type 3 configuration set where the solid lines are elements in their intersection that have been removed.

covered cells.

Proof. Let $C = C_1 \cap C_2$. Assume that C contains cells that are triple covered by C. By definition of triple covered cells that means $C_1 \Delta C_2$ must contain at least one block, row, and column that intersect with a cell. By definition of $C_1 \Delta C_2$, either C_1 of C_2 contains the row set, column set, or block that triple covers these cells. Without loss of generality assume that C_1 contains cells that are triple covered by C. By Lemma 2.3 that means C_1 contains a block and the row set and column set that intersect with that block. However, by Lemma 2.6 that means that C_1 would contain a configuration leading to a contradiction since C_1 is a configuration. Thus, the st $C_1 \Delta C_2$ cannot contain any cells that are triple covered by C. □

Not only can we not have any triple covered cells in the symmetric difference but these subsets cannot have any cells that are single covered by the symmetric difference of any two configurations.

Lemma 4.2. *If $C = C_1 \Delta C_2$ where C_1 and C_2 are configurations then C contains no cells that are single covered by C.*

Proof. First notice that if a row, r, is in element in a configuration, then the entire row set that contains r also belongs to the configuration. Same goes for a column and its respective column set. Because of this we can view the elements of configuration sets as row sets, column sets, and blocks.

Suppose the contrary, that C_1 and C_2 are configurations where $C_1 \Delta C_2$ contains cells that are single covered by $C_1 \Delta C_2$. It follows from the reasoning above that the set

of single covered cells in $C_1 \Delta C_2$ can be represented as a union of some collection of row sets, column sets, and blocks in $C_1 \Delta C_2$. We know that $C_1 \Delta C_2$ either contains a block composed of single covered cells or a row set or column set composed of single covered cells.

If $C_1 \Delta C_2$ contains a block, call it b, that is composed of single covered cells then either $b \in C_1$ or $b \in C_2$ but not in both. Since C_1 andC_2 are configurations, by the double configuration property each cell must be double covered. However, since the row set, call it R, and column set, call it C, that intersect with b are not elements in $C_1 \Delta C_2$, that means that $R \notin C_1 \Delta C_2$ and $C \notin C_1 \Delta C_2$. It would follow that $R \in C_1 \cap C_2$ and $C \in C_1 \cap C_2$. However, b is either an element in C_1 or in C_2 leading to cells in either C_1 or C_2 being triple covered. By the double covered property C_1 and C_2 should only have cells that are double covered leading to a contradiction. Thus, $C_1 \Delta C_2$ cannot contain any blocks composed of single covered cells.

Next, assume that $C_1 \Delta C_2$ contains a row set or column set that is composed of single covered cells. Without loss of generality, assume that $C_1 \Delta C_2$ contains RS_1 that contains single covered cells. Then RS_1 is either in C_1 or C_2 but cannot be in both. Assume that $RS_1 \in C_1 - C_2$. Since $RS_1 \in C_1$, the cells in RS_1 must be double covered hence C_1 either contains b_1, b_2, and b_3 or CS_1, CS_2 and CS_3. If C_1 contains b_1, b_2 and b_3, then $\{b_1, b_2, b_3\} \notin C_1 \Delta C_2$ which means that $\{b_1, b_2, b_3\} \in C_2$. But, since $RS_1 \notin C_2$ and cells in C_2 are double covered, then $CS_1, CS_2, CS_3 \in C_2$. See Figure 4.2 for what C_1 looks like and Figure 4.3 for what C_2 looks like.

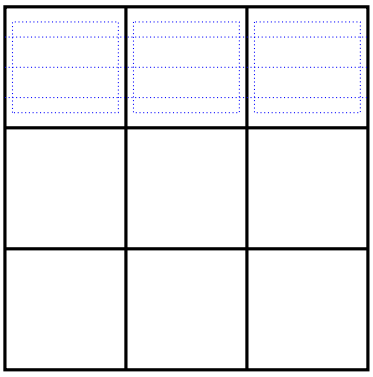

Figure 4.2: What C_1 looks like thus far.

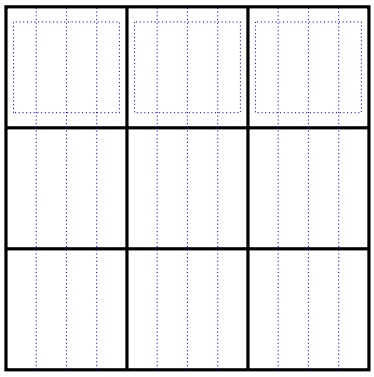

Figure 4.3: What C_2 looks like thus far.

But, since we know that RS_1 is single covered in $C_1 \Delta C_2$, $CS_1, CS_2, CS_3 \notin$

$C_1\Delta C_2$ leading to $CS_1, CS_2, CS_3 \in C_1$. It would follow that C_1 would then contain triple covered cells. According to the double coverage proposition, C_1 would not be a configuration leading to a contradiction. Therefore C_1 can not contain b_1, b_2, and b_3, instead C_1 must contain CS_1, CS_2, and CS_3. Since RS_1 is single covered in $C_1\Delta C_2$, $CS_1, CS_2, CS_3 \notin C_1\Delta C_2$. So, CS_1, CS_2, and CS_3 are also elements in C_2. Since C_2 is a configuration, C_2 has to have double covered cells. Since it does not contain RS_1, $\{b_1, b_2, b_3\} \in C_2$. See Figure 4.4 for what C_1 looks like and 4.5 for what C_2 looks like. But, $\{b_1, b_2, b_3\} \notin C_1\Delta C_2$, it would follow that $\{b_1, b_2, b_3\} \subset C_1$ as well. This leads to C_1

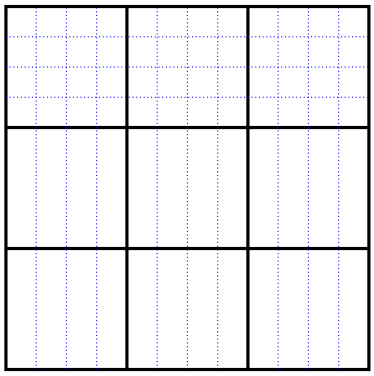

Figure 4.4: What C_1 looks like thus far.

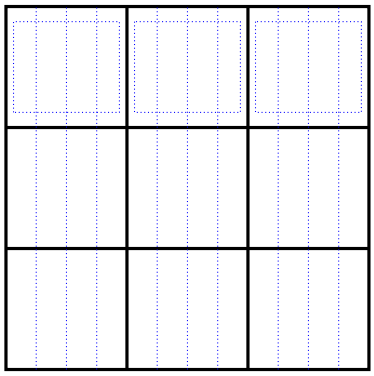

Figure 4.5: What C_2 looks like thus far.

having triple covered cells, see Figure 4.6.

By the double coverage proposition, C_1 is not a configuration, leading to a contradiction.

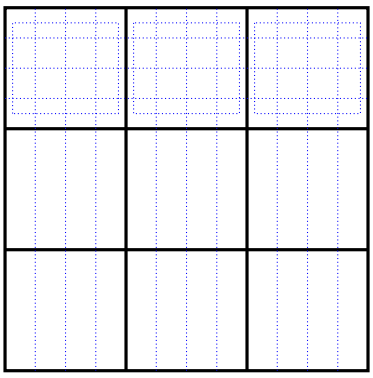

Figure 4.6: Configuration 1 with triple covered cells.

Thus, $C_1\Delta C_2$ cannot contain any single covered cells. □

Combining these two previous lemmas leaves us with only one option for the

coverage of cells in the symmetric difference of configuration sets, they must be double covered.

Lemma 4.3. *If $C = C_1 \Delta C_2$ where C_1 and C_2 are configurations then C has exactly double covered cells.*

Proof. Assume that $C = C_1 \Delta C_2$ contained single covered cells. By Lemma 4.2 then C is not a symmetric difference, which is a contradiction. Next assume that C contains triple covered cells, then by Lemma 4.1 C would not be a symmetric difference, which is a contradiction. Thus, C contains only cells that are double covered. □

Now that we know that the symmetric differences of two configuration sets have only double covered cells we can argue that the symmetric difference leads to a configuration type.

Theorem 4.4. *If $C = C_1 \Delta C_2$ then C is either configuration type 1-7 or contains configuration type 1.*

Proof. Assume that $C = C_1 \Delta C_2$. By Lemma 4.3 that means all cells in C are double covered. By Theorem 2.8 that means that C is a configuration or contains configuration type 1. □

From Lemma 4.4 we know that if we choose to remove only one element from the intersection of two configuration sets, that subset will contain a configuration set.

Theorem 4.5. *If C_1 and C_2 are configurations, and $e \in (C_1 \cap C_2)$, then there exists a configuration $C_3 \subset (C_1 \cup C_2) - e$.*

Proof. Assume that C_1 and C_2 where $e \in (C_1 \cap C_2)$. Notice that $C_1 \Delta C_2 \subset (C_1 \cup C_2)$. By Theorem 4.4 we know that the set $C_1 \Delta C_2$ is a configuration. Thus, there exists a subset, more specifically either $C_1 \Delta C_2$ or configuration type 1 in $C_1 \cup C_2 - \{e\}$. □

Through these lemmas and theorems we are able to build up to our epic conclusion of proving axiom all of the circuit axioms of a matroid hold true for the collection of all configuration sets.

Theorem 4.6. *The set C of configurations in a Sudoku puzzle form the set of Circuits in a matriod.*

Proof. First, by definition of configurations, C is nonempty. So, circuit axiom 1 is met.
Secondly, by definition of configurations, C is minimal. So, circuit axiom 2 is met.
Finally, by Theorem 4.5 circuit axiom 3 is met.
Thus, the set of configurations of a Sudoku puzzle form the set of Circuits in a matriod.
□

Now, we can official say that the elements of a Sudoku puzzle do form the ground set of a matroid. This means that we can begin to describe the collection of subsets that will answer our main question of this paper, what is the minimal amount of elements needed to verify an entire puzzle. As foreshadowed in previous chapters, we will next describe what the basis of our matroid must look like.

4.2 Bases for Sudoku Matroid

Now we know that the a Sudoku puzzle forms a matroid we can describe, to an extent, the other components of a matroid whose ground set are the elements of a Sudoku puzzle. First, we can describe what the independent sets of the matroid will look like. By the previously stated definition, we know that these independent sets cannot contain a circuit, thus they can not contain a configuration set. More importantly, we can now describe what the bases of these matroids will look like. From chapter 3 we know that a basis is the maximal independent set of the matroid, meaning that we are looking for subsets in our independent set that are not properly contained in any other subset. In this case, that would mean that we are looking for the largest independent set where the addition of any element from its complement would lead to that subset containing a configuration set. Furthermore, we have to show that these subsets are minimal, meaning that if we remove any element from that subset we would no longer be able to verify all the elements in that subsets complement. If we are able to find one subset that has these two properties then we have found a basis. First, we need to formally define what a basis is in terms of verification.

Theorem 4.7. *If $A \subset K$ that are minimal with the respect to the property $A \to (K - A)$ then A forms the bases of the Sudoku matroid M.*

Proof. From 4.6 we know that if C is a configuration then C is also a circuit. Since Bases are maximal independent sets, they cannot contain a configuration. Let $B \subset K$, where

B does not contain a configuration but B is a maximal independent set. Then $B \cup x$ does for all $x \in (K - B)$. By our definition of configurations, then $B \to (K - B)$. Moreover, if $e \in B$, then consider $B - e$. We know that $B - e$ still does not contain a configuration, and since B does not either, it follows that $B - e \not\to e$. So, $B - e \not\to K - (B - e)$, for any $e \in B$. Thus, B is minimal with respect to the property that $B \to (K - B)$. Therefore the bases are the minimal verification sets of K. □

We know from chapter 3 that by axiom $B2'$, all bases have the same cardinality. It is our claim that the size of a basis is of 21. In order for us to prove that we will look at one subset of that size, show that it is a full verification set and that it meets both requirements of a basis.

Example 1.
Let $A = \{r_2, r_3, r_5, r_6, r_8, r_9, c_2, c_3, c_5, c_6, c_8, c_9, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9\}$. We have to verify all elements in $K - A = \{r_1, r_4, r_7, c_1, c_4, c_7\}$ which can be seen below:

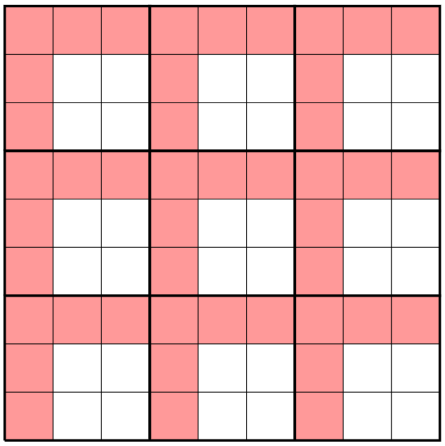

Figure 4.7: The elements in $K - A$ are seen in red.

First, we will verify the unverified rows. We will focus on verifying r_1 which is an element in RB_1. Without loss of generality assume that there is a repeated 1 in r_1, specifically in cells $(1, 2, 1)$ and $(1, 5, 2)$. Since $b_3 \in A$ we know that b_3 is consistent, meaning there is a 1 in b_3. Since $r_2 \in A$ we also have that r_2 is consistent so, without loss of generality, we can place a 1 in $(2, 8, 3)$.

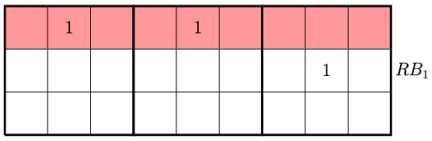

Figure 4.8: The row block, specifically RB_1, where a repeated 1 appears in r_1, as seen in red.

Now we have placed a 1 in all three of our consistent blocks in this specific row block, but there is no 1 in r_3, it follows that r_3 is unverified. However, r_3 is in A, leading to a contradiction. So there must be a 1 somewhere in r_3. This means that there cannot be a repeated 1 in r_1. So, r_1 is consistent. We can follow this same reasoning for the remaining two unverified rows by looking at their specific row blocks and following the same logic as above. Also, by the symmetry of a Sudoku Puzzle, we can verify the unverified three columns not in A by looking at their respective column blocks instead of row blocks and following the same logic as above. This means that we can verify all the unverified rows and columns in $K - A$ leading to $A \rightarrow K - A$. So, A is a verification set of $(K - A)$ implying A is consistent.

Now, that we know that this subset of size 21 can verify the entire puzzle, let us see what would happen if we were to remove any one element from our collection. First, we will start with removing a row from our subset. Without loss of generality let $r_2 \notin A$ so, $(K - A) = \{r_1, r_2, r_4, r_7, c_1, c_4, c_7\}$. This would mean that we would have the following set:

This would mean that now when we look at RB_1 there are two missing rows from that row block, and the method that we used to verify r_1 above would no longer hold. Since nothing else in our puzzle has changed, it can be assumed that the remaining elements in $K - A$ can be verified using their respective row blocks and column blocks. This would mean that all we have left to verify is r_1 and r_2.

We will try to construct a correct Sudoku puzzle where we position the numbers $1 - 9$ in each cell, where in every row, column, and block there were no repeated numbers in the respective elements, see Figure 4.10

Now, we can select two cells that are in the same column but one cell is in r_1 while the other is in r_2 and swap the positioned numbers to get the puzzle, see Figure 4.11

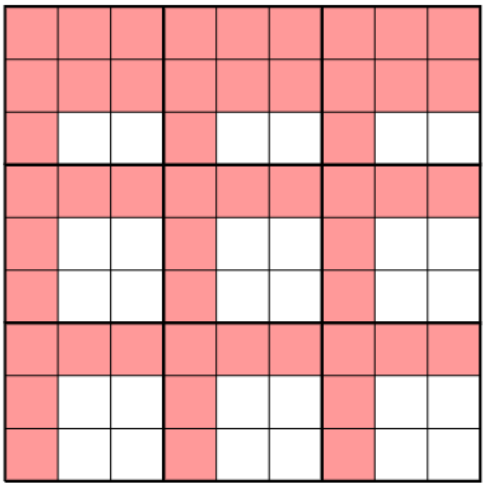

Figure 4.9: The elements in $(K - A - r_2)$ as seen in red.

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

Figure 4.10: A consistent Sudoku puzzle, where r_1, r_2 are seen in red.

Notice that r_1, r_2 are inconsistent but all other elements in our set remain consistent and are unable to detect these newly inconsistent rows. Removing any row or column from our original subset would lead to the same outcome, where within a row block or column block we can perform this swap between cells leaving all other elements verified but unable to verify those elements. If we apply this logic to our subset A, it would mean that $A \not\rightarrow (K - A - r_2)$ and A is no longer a verification set. This outcome, of having two

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

Figure 4.11: Cells in r_1, c_3 and r_2, c_3 swapped as seen in yellow.

unverified rows or columns in a row block or column block respectively would happen if we remove any rows or columns from A. Thus, we cannot remove any row or column from A.

Finally, we have to see what will happen if we try to remove any blocks from our set A. Without loss of generality let $K - A = \{r_1, r_4, r_7, c_1, c_4, c_7, b_1\}$. This means that now we have a row block and a column block with an inconsistent block and row or column, respectively. Meaning that the remaining four row blocks and column blocks can still verify the unverified elements in their respective subsets. So, the only elements we have to verify are r_1, c_1 and b_1. However, this would mean that cell $(1, 1, 1)$ is uncovered by the elements in $(A - b_1)$. By Lemma 2.1 that would mean that the elements in $(A - b_1) \not\twoheadrightarrow \{r_1, c_1, b_1\}$. Thus, r_1, c_1 and b_1 remain unverified and $(A - b_1)$ is not a full verification set.

Lastly, we need to show that if we were to add any element to A then A would contain a configuration set. Notice that the only elements that we can add are additional rows or additional columns. However, if we add any row or column to A, we can quickly see that then A would contain a configuration of type 1, leading to A containing a circuit. Thus, A is a maximal independent set. Therefore, A is a basis of the matroid.

This example shows that there does exist a subset of size 21 that can verify all the unverified elements in its complement and if one tries to remove an element from that subset then the verification of all elements in its complement is no longer possible. By the Axiom $(B2)'$ this means that all bases have to be of the same size of 21. Which means

we have finally answered our overarching question, how many elements do you need to check to know that your Sudoku puzzle is correct? One has to check a combination of 21 rows, columns, and blocks. However, as explained in the proof above, there are some elements you have to be sure to check. For example, one has to check at least 2 rows and 2 columns in every row and column set. Further one has to make sure that no cell is left uncovered by that subset at the end of all their checks. One may be wondering if there are more tools like the ones stated in above that could be useful in distinguishing a subset as being a verification set to ones that are not. Luckily, throughout this research we did find a few that may be useful.

4.2.1 Properties of Verification Sets

Being able to look at only the elements of a subset to know if that subset is a full verification set would be very useful. However, throughout this research we were only able to find a few tools to aid us in that journey. First, we already know two properties about verification sets, and to that extent basses, from our previous example of a basis. We will formally define those lemma's.

Lemma 4.8. *If A contains any uncovered cells then A is not a full verification set.*

Proof. Assume that $A = K$. That means that all rows, columns, and blocks are verified. However, if we choose to replace the number positioned in cell $(1, 1, 1)$ then that row, column, and block that covered that cell become inconsistent. Notice that all other rows, columns, and blocks remain verified and unable to detect the newly unverified row, column, and block. Thus, uncovered the newly inconsistent row, column, and block cannot be verified by the remaining elements. Meaning, $(A - r_1 - c_1 - b_1) \not\Rightarrow \{r_1, c_1, b_1\}$ and is therefore not a full verification set. □

Lemma 4.9. *If A is a set and does not contain at least two verified rows and two verified columns in every row set and column set, then S is not a full verification set.*

Proof. Let $A = K$. First let us swap two cells that belong to two different rows in the same row set but are in the same column. Without loss of generality lets swap cells $(1, 3, 1)$ and $(2, 3, 1)$. When we do that only r_1 and r_2 become inconsistent. Notice that all other rows, columns, and blocks remain consistent and unable to detect the newly

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

Figure 4.12: Completed puzzle with cell $(1, 1, 1)$ seen in yellow and the newly unverified row, column, and block as seen in red.

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

Figure 4.13: Cells $(1, 3, 1)$ and $(2, 3, 1)$ as seen in yellow and the rows r_1, r_2 as seen in red.

inconsistent rows. This means that $(A - r_1 - r_2) \not\Rightarrow \{r_1, r_2\}$ and is therefore not a full verification set. □

Throughout this paper so far we have focused heavily on the cell coverage instead of the elements themselves. However, relying on cell coverage to determine if a subset could be a full verification set seems more daunting then just looking at the elements. Instead, focusing on the elements themselves would seem to be the easier path in determining if a subset could be a full verification set. We have seen above how a subset must contain a minimum number of rows and columns. Specifically, one has to have at least 2 rows and 2 columns from every row set and column set. Next we can ask ourselves, what about the blocks? Is there a minimum number of consistent blocks that a subset should have in order to possible be a full verification set? The answer, unsurprisingly, is yes.

Lemma 4.10. *If A is a full verification set, then A must contain at least 3 blocks*

Proof. Let A be a full verification set that contains all rows, columns and all blocks. If we take a completed Sudoku puzzle, we are able to find four cells in different blocks but where each cell is in the same row as another cell and a column of a different cell. From

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

Figure 4.14: Cells in $(4, 1, 4), (4, 7, 6), (9, 1, 7)$ and $(9, 7, 9)$ as seen in yellow.

here we can rearrange the Notice that all rows, columns, and all other blocks remain

verified. The only way to verify these newly inconsistent blocks is if one of those blocks is forced to be consistent. We know from Lemma 4.9 that all full verification sets must contain at least two rows and two columns in every row set and column set. This means that when we look at all the possible combinations of four blocks that match this pattern, further called a *block quad*, A would have to contain at least three blocks in order to have one verified block for every combination. □

Lemma 4.10 directly leads us to knowing the placement of the verified blocks if a subset only has three verified blocks as elements.

Lemma 4.11. *If A is a full verification set with only three verified blocks, then those verified blocks are in different row blocks and column blocks.*

Proof. Let A have two verified blocks in either the same row block or column block. Without loss of generality let two verified blocks be in RB_1. Then the other verified block must be somewhere in RB_2 or RB_3, without loss of generality let it occur somewhere in RB_3. Next, we can either have these verified blocks in the same column block or different column blocks. First, if I choose to have the verified blocks occur in the same column block then there would be a block quad without a verified block in it meaning A would not be a full verification set. So, I have to have my verified blocks occur in different column blocks. However, there still exists a block quad without a verified block in it. Thus, the verified blocks must occur in different row blocks as well as different column blocks. □

It may be surprising to the reader to learn that a subset having just three verified blocks is not enough to say that a subset is a full verification set. Starting with Lemma 4.11 helps us know that each subset has to have at least 3 blocks. However, we can increase the minimum amount of required blocks in a full verification set to be four.

Lemma 4.12. *Let A be a full verification set, then A must contain at least 4 blocks*

Proof. Assume that A is a full verification set. From Lemma 4.10 we know that A must contain at least three blocks. Assume that A contains all rows, columns, and only 3 blocks. By Lemma 4.11 that means that these blocks must be in different row blocks and column blocks. There are only 6 possible ways we can arrange our verified blocks within this criteria. Without loss of generality, let us assume that $\{b_1, b_6, b_8\} \in A$.

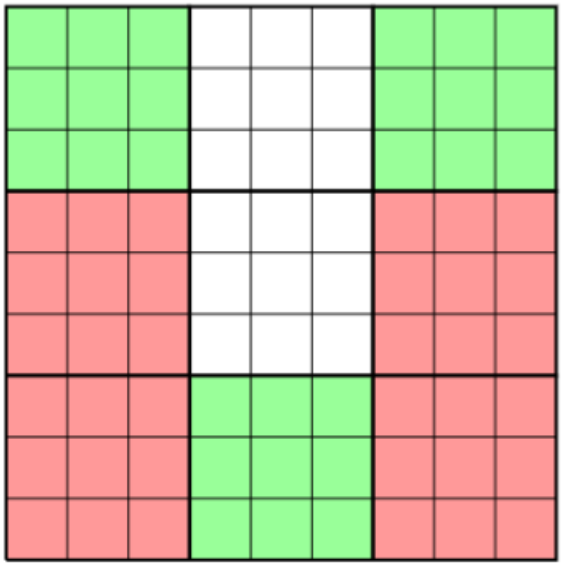

Figure 4.15: consistent blocks b_1, b_3, b_8 as seen in green and blocks in block quad as seen in red.

We can arrange numbers within the cells of the other blocks in such a way that those blocks will be inconsistent but the all of the rows and columns will remain consistent, By the rotational symmetry of the puzzle, this argument can also be replicated if A contained the other pairings of verified blocks such as: b_2, b_6, b_7, b_3, b_4, b_8, and b_2, b_4, b_9. Next, we can replicate this argument when A contains the verified blocks down the diagonal, those being either b_1, b_5, b_9 and b_2, b_5, b_7. Without loss of generality, we will let $\{b_1, b_5, b_9\} \subset A$ and lets arrange the numbers withing the cells of the other unverified blocks, make them inconsistent while maintaining the consistency of all the rows and columns. Thus, no combination of three verified blocks would lead to A being a full verification set. Therefore, A must contain at least 4 blocks. □

These lemmas can help one rule out subsets that could not be a full verification set. However, the reader may notice that these lemmas require a minimum of 15 elements in a subset in order for one to hope that it is a full verification set. Unfortunately, these are all the lemmas we found that describe what these full verification sets must look like. We know from the previous section that a basis has a size of 21. This means that there could be more lemmas and tools out left to find and prove that would aid someone in their quest to describe what these basis look like. Beyond the search to describe the bases

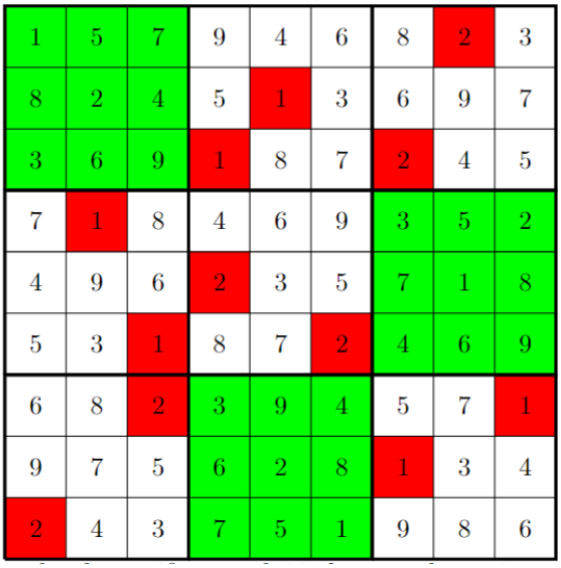

Figure 4.16: The blocks b_1, b_6, b_8 are seen in green and the cells that force the other blocks to be inconsistent are seen in red.

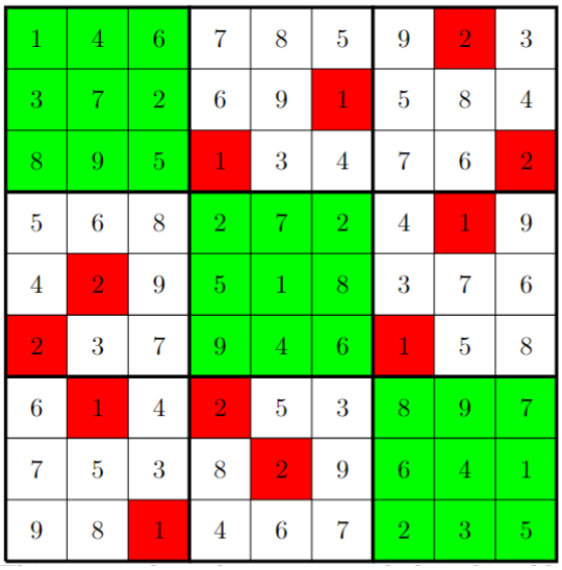

Figure 4.17: The blocks b_1, b_5, b_9 are seen in green and the cells that force the other blocks to be inconsistent are seen in red.

and full verification sets there appears to be more questions that one could focus on.

Chapter 5

Conclusion

This research leads to various other questions in other interesting topics. In this chapter we will explore one consequence from the research we have done, specifically how we can represent these matroids. From Lemma 4.4 we know that the symmmetric difference of two configuration sets contains a configuration set. This directly means that we can describe this matroid as a binary matrix. We know that this is possible based off of the following theorem.

Theorem 5.1 ([Oxl92], Theorem 9.1.2)**.** *The following statements are equivalent for a matroid M:*

(i) M is binary

(ii) If C is a circuit and C^* is a cocircuit, then $|C \cap C^*|$ is even.

(iii) if C_1 and C_2 are distinct circuits, then $C_1 \Delta C_2$ contains a circuit.

(iv) If C_1 and C_2 are circuits, then $C_1 \cap C_2$ is a disjoint union of circuits.

(v) The symmetric difference of any set of cicuits is either empty or contains a circuit.

(vi) The symmetric difference of any set of circuits is a disjoint union of circuits.

(vii) if B is a basis and C is a circuit, then $C_{e \in C-B} C(e, B)$.

(viii) M has a basis B such that if C is a circuit, then

$C_{e \in C-B} C(e, B)$.

The proof to this theorem can be found in [Oxl92]. In particular we are using the equivalence statement 3 to 1 to say that this matroid is a binary matroid. We will look at an example of how we can take the basis in the previous chapter and describe it as a binary matrix.

Example 1. Take our example of our basis,
$A = \{r_2, r_3, r_5, r_6, r_8, r_9, c_2, c_3, c_5, c_6, c_8, c_9, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9\}$, from chapter 4 and convert it to a matrix. In order to create this matrix we will have the columns of the matrix be be the ground set while the rows of the matrix will be the elements in our basis. More specifically we will arrange our columns to where all of the elements in the complement appear at the end of our matrix. We can position a 1 into the intersection of a row and a column that represents that the element represented by that row is required to verify the element represented by that column. Therefore, if an element is in our basis, it is assumed that that element is verified by itself. See Figure 5.1.

One should notice that the first 21 columns we have the identity matrix. While the columns that represent the elements in our compliment require five elements represented by 1's positioned in the intersection of specific rows.

This could mean that further researcher into the connection between Sudoku puzzles and matroid theory may lead to more interesting results. One particular research question can include investigating coloring maps that follow the rules of our Sudoku matroid.

Sudoku puzzles is a game that anyone can play. With some basic deductive reasoning skills and some patience, one can spend a few hours completing copious amounts of puzzles and feeling an overwhelming sense of accomplishment. What the everyday player may not realize is that they are falling right into the hands of matroid theory. With the help of matroid theory, Sudoku puzzles can be seen in a whole new light. With a little additional knowledge, any player can check their completed puzzle to know that they are correct.

	r_2	r_3	r_5	r_6	r_8	r_9	c_2	c_3	c_5	c_6	c_8	c_9	b_1	b_2	b_3	b_4b	b_5	b_6	b_7	b_8	b_9	r_1	r_4	r_7	c_1	c_4	c_7
r_2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
r_3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
r_5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
r_6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
r_8	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
r_9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
c_2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
c_3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
c_5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
c_6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
c_8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
c_9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
b_1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0
b_2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
b_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1
b_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0
b_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0
b_6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
b_7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
b_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0
b_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1

Figure 5.1: Binary matrix that represents the basis in Chapter 4.

Bibliography

[GM12] Gary Gordon and Jennifer McNulty. *Matroids: A Geometric Introduction*. Cambridge University Press, 2012.

[Hyu14] Tony Hyuen. Sudoku matroids and graph colouring verification. http://matroidunion.org/?p=1070, 2014. Last accessed 1 February 2022.

[Je7] Emil Jeřábek. Verifying the correctness of a sudoku solution. https://mathoverflow.net/questions/129143/verifying-the-correctness-of-a-sudoku-solution, 2017. Last accessed 16 September 2017.

[Mak21] Sudoku creator maki kaji, who spread the joy of puzzles, has died. https://www.npr.org/2021/08/17/1028412532/sudoku-creator-maki-kaji-dies, 2021. Last accessed 1 February 2022.

[Oxl92] James Oxley. *Matroid Theory*. Oxford Univeristy press, 1992.