Understanding special Sudoku solutions through geometry and error correcting codes

Meagan Heckert
Southern Illinois University Edwardsville

April 29, 2007

1 Introduction

1.1 Sudoku History and Rules

Sudoku is a game in which you fill a square 9 x 9 grid with the numbers 1 through 9. Each number may only appear once in each row, column, and subsquare. The term row describes a horizontal line of entries in the grid like a row of a matrix. In the same sense, the term column refers to a vertical line of entries in the grid like those of a matrix. The term subsquare refers to a 3 x 3 sub-grid inside the large 9 x 9 grid. These subsquares are aligned in the large grid in such a way that there are three stacked vertically and three aligned horizontally. Sudoku puzzles will have some clues filled in and the object of the puzzle is to fill in the rest of the grid under the above constraints.

The history of Sudoku as we know it today is a long one. Sudoku sprouted from the concept of Latin Squares. Latin Squares are grids in which each symbol from the same set appear only once in each row and each column. Latin Square puzzles have been around long before Sudoku puzzles were established. Sudoku puzzles are a type of Latin Square that have the additional restriction that the symbols 1 through 9 may only appear once in each 3 x 3 sub-grid of a 9 x 9 grid. The first Sudoku puzzle appeared in an edition of Dell Pencil Puzzles and Word Games in 1979 and is said to have been created by Howard Garns, a retired architect. These puzzles gained popularity in Japan during 1984. Sudoku became successful when Wayne Gould, after seeing a Sudoku puzzle in Japan, created a computer program that creates Sudoku grids. Today there are published books of these puzzles and they appear in most daily newspapers. See [Delahye, 2006] for more history.

Sudoku is still a relatively new craze. You do not need to have an extravagant mathematical background in order to play this game. All you need are the logical and critical thinking skills to place the numbers 1 through 9 in designated squares. Completing Sudoku puzzles can be reasonably simple and have become something of an addiction for many people. Even though completing a Sudoku puzzle does not take very advanced math skills Sudoku has some very interesting mathematical qualities.

1.2 Defining Sudoku entries

For clarity, we will define how to describe each entry of a Sudoku grid. There are nine 3 x 3 *subsquares* in a 9 x 9 Sudoku grid. Each subsquare is bounded by the bold lines in the Sudoku grid. There are three *horizontal bands* and three *vertical bands* in each table. A horizontal band is a row of subsquares. In the following Sudoku grid, the red numbers make up a horizontal band.

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

A vertical band is a column of subsquares. In the following Sudoku grid, the green numbers make up a vertical band.

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

Within each subsquare there are *minirows* and *minicolumns*. The minirows are the 1 x 3 row in each subsquare and the minicolumns are the 3 x 1 columns in each subsquare. In the following Sudoku grid, the blue numbers are a minirow and the magenta numbers are a minicolumn.

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

We define a *broken row* as the three minirows in the same position of each subsquare in a vertical band. In the following Sudoku grid, the yellow numbers make up a broken row.

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

A *broken column* is the three minicolumns in the same position of each subsquare in a horizontal band. A *location* is the set of entries having the same position in each subsquare. For example, the bottom left entry of each subsquare would make up a location. In the following Sudoku grid, the red numbers make up a broken column and the green numbers are a location.

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

1.3 Project Plan

Using ideas from the article by Bailey, Cameron and Connelly, we will determine the number of symmetric Sudoku solutions up to an equivalence. A *symmetric Sudoku solution* is a Sudoku solution which has the additional property that each number occurs precisely once in each broken row, broken column, and location. Figure 1 is an example of a symmetric Sudoku solution. For the purposes of this paper, we consider two Sudoku solutions *equivalent* if one can be obtained from the other by renaming symbols. Bailey, Cameron, and Connelly considered a much more general type of symmetry in their article. They then show that up to these symmetries there exist only 2 solutions. This full result is beyond the scope of this senior project.

In this senior project, we prove that there are 104 symmetric solutions up to our definition of equivalence. To do this we will use the following general plan:

Section 2:

- we will fully coordinatize the Sudoku grid by $\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$,

Section 3:

- a set of locations of a symbol form a 1-error correcting code in a symmetric Sudoku solution,

- the set of all locations of each symbol partition $\mathbb{Z}_3\times\ \mathbb{Z}_3\times\ \mathbb{Z}_3\times\ \mathbb{Z}_3$,
- any perfect 1-error correcting code in $\mathbb{Z}_3\times\ \mathbb{Z}_3\times\ \mathbb{Z}_3\times\ \mathbb{Z}_3$ is an affine subspace of $\mathbb{Z}_3\times\ \mathbb{Z}_3\times\ \mathbb{Z}_3\times\ \mathbb{Z}_3$,

Section 4:

- we show each affine subspace is a translation of a special vector space called an allowable space,
- determine the eight allowable vector subspaces, and

Section 5:

- count how to partition V using those subspaces.

Applications for Sudoku in a high school setting will also be discussed.

2 Coordinatizing Sudoku entries

We will now coordinatize the Sudoku entries. Our coordinate system will be over the finite vector field $\mathbb{Z}_3$. Each entry in the grid has a unique coordinate $(x_1, x_2, x_3, x_4) \in \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$ as follows. We will start our numbering at zero and will number from top to bottom for the rows and left to right for the columns.

x_1 is the number of the horizontal band;
x_2 is the number of the minirow of the subsquare;
x_3 is the number of the vertical band;
x_4 is the number of the minicolumn of the subsquare.

Now we can describe each entry by its coordinates. For example, the red 2 in Figure 1 would have the coordinates $(0, 1, 2, 1)$. An example of a broken row would be $x_2 = x_3 = 0$ and a broken column would be $x_1 = x_4 = 0$.

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

Figure 1: Symmetric Sudoku solution

In Table 1, we can see some other ways to represent the definitions given before as subspaces of $\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$.

Equation	Description
$x_1 = x_2 = 0$	Top Row
$x_3 = x_4 = 0$	Left Column
$x_1 = x_3 = 0$	Top Left Subsquare
$x_1 = x_4 = 0$	Broken column
$x_2 = x_3 = 0$	Broken row
$x_2 = x_4 = 0$	Location (top left entries)

Table 1: Some subspaces of $\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$

Using these definitions and coordinates we will look at one special Sudoku solution, the symmetric Sudoku solutions.

3 Sudoku, Affine Geometry, and Error Correcting Codes

A subspace, W, of a vector space V is a subset $W \subseteq V$ that is closed under addition and scalar multiplication. For example, any line through the origin is a subspace of the vector space $\mathbb{R}^2$. An *n-affine space* A over a field, F, is the set of n-tuples $(a_1, a_2, ..., a_n)$, for $a_i \in F$ with addition defined coordinate-wise. An affine space differs from a vector space in that in a vector space all subspaces must contain point $(0, 0, 0, ..., 0)$, however, an affine space has no set "origin".

Definition 1 *Since A is a set of n-tuples, it has a natural vector space structure. A subset $A' \subseteq A$ is called an affine subspace of A if A' is a coset of a vector subspace of the vector space A.*

An *affine subspace* of A is a *coset* of a vector subspace of A, i.e., given any vector subspace $W \subseteq V$ and any $v \in V$, then $v + W$ is an affine subspace. The set $\{v + W\}_{v \in V}$ is called a *parallel class*. Some representatives of this class are shown in Figure 2, where A and V are in $\mathbb{R}^2$, W is the line going through the origin (black line). The other colored lines are $v + W$ for some values of v.

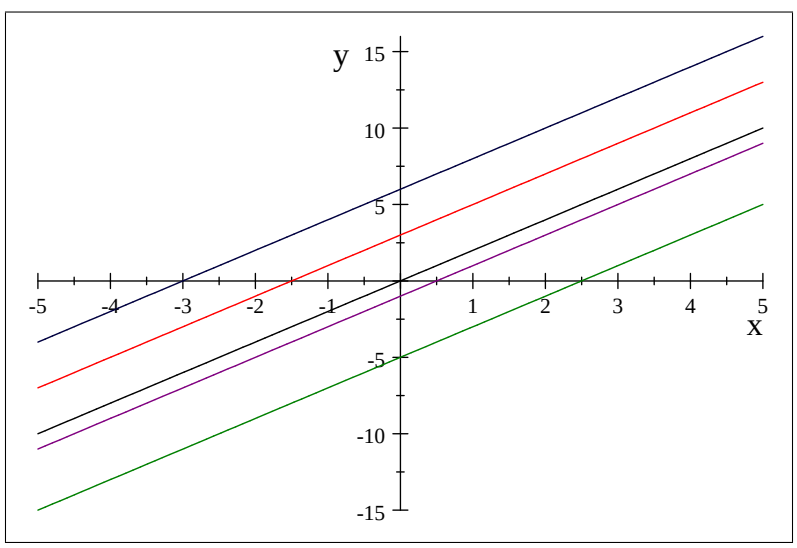

Figure 2: Five elements in a parallel class

Definition 2 *A transversal is a chosen coset representative for each affine subspace in the parallel class.*

If we take one representative from each affine subspace, those coset representatives form a transversal. Therefore, in Figure 2 a transversal is a choice of one coset representative for each of the different colored lines.

3.1 Sudoku and Affine Geometry

Now we will look at how affine geometry connects to Sudoku puzzles. For Sudoku puzzles, one of the affine subspaces is

$$R = \{(x_1, x_2, x_3, x_4) \,|x_1 = x_2 = 0\}\,. \tag{1}$$

We can describe the sixth row in multiple ways. One way is

$$\begin{aligned} R + (1,2,1,1) &= \{(x_1,x_2,x_3,x_4)|x_1 = x_2 = 0\} + (1,2,1,1) \\ &= \{(x_1,x_2,x_3,x_4)|x_1 = 1, x_2 = 2\}\,. \end{aligned}$$

Row 6 can also be described by

$$\begin{aligned} R + (1,2,0,0) &= \{(x_1,x_2,x_3,x_4)|x_1 = x_2 = 0\} + (1,2,0,0) \\ &= \{(x_1,x_2,x_3,x_4)|x_1 = 1, x_2 = 2\}\,. \end{aligned}$$

These two are both cosets but could not be in the same transversal since they are both representatives of the same affine subspace. So for the Sudoku solution in Figure 4 the rows, numbered 1 through 9 from top to bottom, can be described by a transversal where

$$\begin{aligned} R + (0,0,0,0) &= \text{row } 1, \\ R + (0,1,1,1) &= \text{row } 2, \\ R + (0,2,2,0) &= \text{row } 3, \\ R + (1,0,1,0) &= \text{row } 4, \\ R + (1,1,0,2) &= \text{row } 5, \\ R + (1,2,2,1) &= \text{row } 6, \\ R + (2,0,2,2) &= \text{row } 7, \\ R + (2,1,1,2) &= \text{row } 8, \end{aligned}$$

and

$$R + (2,1,0,1) = \text{row } 9.$$

So pick a symbol, say 1. We know that each symbol, 1, appears in each row precisely once by the Sudoku rules. Thus this symbol has nine distinct coordinates. Each of these coordinates is in each row coset precisely once and is a representative for each row coset. This set of row cosets forms a parallel class where $R = \{(x_1, x_2, x_3, x_4)|x_1 = x_2 = 0\}$. These coordinates form a transversal for this parallel class since the symbol, 1, is a representative for each row coset and thus a transversal. So for the Sudoku solution in Figure 3, the transversal of the nine rows defined by the symbol 1 is

$$\begin{aligned} &R + (0,1,0,1) \\ &R + (0,2,1,1) \\ &R + (0,0,2,2) \\ &R + (1,2,0,2) \\ &R + (1,0,1,2) \\ &R + (1,1,2,0) \\ &R + (2,0,0,0) \\ &R + (2,1,1,0) \\ &R + (2,2,2,1). \end{aligned}$$

Let S_i be a set of coordinates of the digit i. A filled in grid is a Sudoku solution if S_i forms a transversal for the rows, columns, and subsquares for all i at the same time. If S_i also forms a transversal simultaneously for broken columns, broken rows, and locations, then the S_is form a symmetric Sudoku solution.

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

Figure 3: Sample Sudoku solution

Definition 3 *A Sudoku solution is called linear if, for each S_i,where $1 \leq i \leq 9$, its nine coordinates form an affine subspace in the affine space $\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$.*

Lemma 1 *In the affine space over $\mathbb{Z}_3$, a line has three points and the third point on the line through p_1 and p_2 is the midpoint $\frac{p_1+p_2}{2} = -(p_1 + p_2)$.*

3.2 Symmetric Sudoku Solutions

Definition 4 *Let A be a set. A code S is a set of codewords, where $S \subset A^n$.*

Definition 5 *The Hamming distance between two vectors is the number of coordinates at which the two vectors differ.*

Definition 6 *A code S is called an e-error correcting code if the Hamming distance between any two codewords $d(c_1, c_2) \geq 2e + 1$.*

Therefore a 1-error correcting code S has the property that given a p and q in S, the Hamming distance $d(p, q) \geq 3$.

Definition 7 *A code S is called perfect if for each p in A^n, there is exactly one codeword q in S so that $d(p, q) \leq 1$.*

For example, if you were to send a code over the internet and one bit was changed, then the receiver could detect this change and correct the error.

Lemma 2 *A 1-error correcting code $S \subset (\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3)$ is perfect if and only if $|S| = 9$.*
Proof. *Given any word $s \in S$, there are $1 + 4 \cdot 2 = 9$ words which can be obtained from s by making 0 or 1 errors. Call this set of words $E(s)$. For any $s_1, s_2 \in S$ such that $s_1 \neq s_2, E(s_1) \cap E(s_2) = \emptyset$. Now, if $|S| = 9$, then*

$$\sum_{s \in S} |E(c)| = 9 \cdot 9 = 81 = |\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3|.$$

So every word $w \in \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$ is at distance 0 or 1 from exactly one codeword. Thus, S, is perfect. On the other hand, suppose S is perfect. Then for any $w \in \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3, w \in E(s)$ for a unique $s \in S$. Therefore, since each $E(s)$ has 9 elements, we must have $|S| = 9$. ■

Consider the vector space $V = \mathbb{Z}_3$ x $\mathbb{Z}_3$ x $\mathbb{Z}_3$ x $\mathbb{Z}_3$. For each S_i, we will have nine codewords of length 4 for each of the nine symbols. Each codeword will describe a single digit's coordinates in a symmetric Sudoku solution. Therefore, if we choose the number 2 from a symmetric Sudoku solution, as in Figure 1, the nine codewords are as follows.

$$S_2 = \{(0,2,0,2), (0,0,1,0), (0,1,2,1), (1,1,0,0), (1,2,1,1), (1,0,2,2), (2,0,0,1), (2,1,1,2), (2,2,2,0)\}$$

Proposition 1 *A symmetric Sudoku solution corresponds to a partition of V into nine perfect codes.*

Proof. Say we have a symmetric Sudoku solution. We want to prove that this solution partitions V. In order for the solution to partition V then $S_i \cap S_j = \emptyset$, for all $i, j \in \{1,2,3,4,5,6,7,8,9\}$ and $\cup S_i = V$. Since we have a symmetric Sudoku solution we know that any two elements in S_i must have distance at least 3 in any S_i must have Hamming distance at least 3 and thus are perfect codes. If they did not have Hamming distance at least 3, then they would contradict the rules of being a symmetric Sudoku solution, i.e. more than one of the same symbol in a broken column, broken row, location, subsquare, row, or column.

We now look at the distance between codewords in the set S_2. Let $W = \{x_i = x_j = 0 | i, j \in \mathbb{Z}_3\}$ be a vector space where $W + (0, a, b, 0)$ is a row, column, subsquare, broken column, broken row, or location. Let i and j be any fourtuple coordinates and let a and b, where $a, b \in \mathbb{Z}_3$, be individual elements in a codeword. Then $p_i = a$ and $p_j = b$ would define a unique codeword p. Codewords cannot agree in more than one positions in order to be distinct. Therefore distinct codewords must have Hamming distance of at least three. For example, if we were to describe two symbols in the top row of a Sudoku grid, they would have the same x_1 and x_2 and thus would have a Hamming distance of two. Therefore these two symbols could not be in the same set of codewords. Since no two codewords can agree in two positions, the Hamming distance between the two codewords is at least 3. Therefore, S is a 1-error correcting code.

Now suppose that we do have a set of nine perfect codes that partition V. Since the S_is are perfect codes, then they have Hamming distance 3. Now,

Case 1 *Suppose there does not exist such a coordinate with $p_i = a$ and $p_j = b$. Since S contains nine elements, then there exists a pair of coordinates agreeing in at least two positions. This is a contradiction.*

Case 2 *Suppose there exists more than one coordinate with $p_i = a$ and $p_j = b$. If so then those coordinates would agree in at least positions a and b. This contradicts the fact that the Hamming distance between coordinates is 3.*

■

Therefore, since S_i partitions V, S_i fills all the positions in the grid and none of the symbols break any of the rules of being a symmetric Sudoku solution, i.e. each S_i has only one representative in each broken column, broken row, location, subsquare, row, and column.

There would be nine perfect codes which represent the nine coordinates for each symbol {1, 2, . . . , 9}. If each symbol's nine perfect codes create a transversal for each affine subspace, then we have a symmetric Sudoku solution.

The following technical result will be used to help prove Proposition 2.

Lemma 3 *The minimum of $(n_0^2 + n_1^2 + n_2^2)$ is 27 and occurs at $(3,3,3)$, subject to the condition that $n_0 + n_1 + n_2 = 9$, where $n_0, n_1, n_2 \in \{1,2,3,4,5,6,7,8,9\}$.*

Proof. Since there are nine possibilities for the value of n_0, n_1, and n_2, where n_0, n_1, n_2 are the integers 1 through 9, we have the following given equation

$$n_0 + n_1 + n_2 = 9. \tag{2}$$

Solving (2) for n_1 we get

$$n_1 = 9 - n_0 - n_2. \tag{3}$$

Let $x = n_0$ and $y = n_2$. Then substituting 3 into the contribution equation

$$f(x, y) = n_0^2 + n_1^2 + n_2^2$$

we get

$$f(x, y) = 2x^2 - 18x + 2xy - 18y + 2y^2 + 81. \tag{4}$$

To find the minimum of this function, we have to take the first partial derivative with respect to x and with respect to y. We get

$$f_x = 4x - 18 + 2y$$

and

$$f_y = 4y - 18 + 2x.$$

We set the first partial derivatives equal to zero. Thus

$$4x - 18 + 2y = 0 \tag{5}$$

and

$$4y - 18 + 2x = 0. \tag{6}$$

Now we want to solve these two equations, 5 and 6, simultaneously. We take equation 5 and solve for y getting

$$y = -2x + 9. \tag{7}$$

We substitute this value of y into equation 6. Simplifying we get

$$x = 3. \tag{8}$$

To get a value for y we substitute this value of x into equation 7 to get

$$y = 3. \tag{9}$$

Since there are two variables we must now use the second derivative test to determine if this function does have a minimum value. For this we will take the second derivative with respect to both variable of each of the first partial derivatives. We get

$$\begin{aligned} f_{xx} &= 4, \\ f_{xy} &= 2, \\ f_{yx} &= 2, \end{aligned}$$

and

$$f_{yy} = 4.$$

These values give us the following matrix.

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

To tell if there is a minimum value we look at the determinant, D, of this matrix. Where

$$D = f_{xx} f_{yy} - f_{xy} f_{yx}.$$

Substituting in the values of the second partial derivatives we get

$$\begin{aligned} D &= (4)(4) - (2)(2). \\ &= 12. \end{aligned}$$

Since $D = 3$ is greater than zero, by the second derivative test we have a minimum at this point. To find our minimum value we put these values of x and y from equations 8 and 9 into our original function 4. Thus we get

$$\begin{aligned} f(x,y) &= 9(3) - 3^2 + 9(3) - 3^2 - (3)(3) \\ &= 27. \end{aligned}$$

Therefore the minimum value of $f(x,y)$ is 27. ■

Proposition 2 *Any perfect 1-error correcting code in* $(\mathbb{Z}_3)^4$ *is an affine subspace.*

Proof. Let H be a perfect code. By the definition of a perfect code $|H| = 9$, meaning H contains nine vectors, where for $p, q \in H$, $d(p,q) \geq 3$, or p and q agree in no more than one coordinate and have a Hamming distance of at least three. Since any two vectors in H have Hamming distance of at least three,

$$\sum_{p,q \in H} d(p,q) \geq 9 \cdot 8 \cdot 3 = 216.$$

This is because the first vector, say p, has a possibility of being one of nine vectors. Thus the second vector, say q, has the possibility of being from the eight other vectors not including p. Since the distance between any two vectors, is three, the total Hamming distance, $\sum d(p,q)$, is greater than or equal to 216.

We can also look at the total Hamming distance by concentrating on each coordinate individually. To do this we will choose one coordinate, say the first position, and let n_0, n_1, n_2 be the number of vectors in H, having the entries 0, 1, 2 respectively, in that first position. We will then look at the contribution of this coordinate to the overall sum of the Hamming distance. Each coordinate would have the following contribution

$$n_0(9 - n_0) + n_1(9 - n_1) + n_2(9 - n_2) \tag{10}$$

which simplifying becomes

$$81 - (n_0^2 + n_1^2 + n_2^2) \leq 81 - 27 = 54.$$

We obtain the number 27 by Lemma 3.

Since each coordinate's contribution is 54 and there are four coordinates in a vector, our total sum is once again $4 \cdot 54 \leq 216$. This indicates that equality does in fact hold. We can conclude that since equality holds each pair of vectors must have distance exactly three, in other words, have agreement in exactly one coordinate position.

Now we look at finding a third point on an affine line. Let p, q be vectors in H. Without the loss of generality, we will let the first coordinate of p and q agree, say $p = (a, b_1, c_1, d_1)$ and $q = (a, b_2, c_2, d_2)$. Since these vectors can only agree in the first position, then $b_1 \neq b_2$, $c_1 \neq c_2$, and $d_1 \neq d_2$. Since we are in $\mathbb{Z}_3$ the remaining elements are $-(b_1 + b_2)$, $-(c_1 + c_2)$, and $-(d_1 + d_2)$. Since each vector can only agree in one position, there is a unique element r of H that also has the same first coordinate a and with the other remaining second, third, and fourth elements given above. This vector must be $r = (a, -(b_1 + b_2), -(c_1 + c_2), -(d_1 + d_2)) = -(p + q)$. The points p, q, and $-(p + q)$ make up an affine line. Thus H makes up an affine subspace since it has these three points on the affine line. ■

Finding the third point on an affine line is similar to determining where to place the symbols in a Sudoku solution. If you know two placements of a certain symbol in a horizontal band of a symmetric Sudoku solution then you know the placement of the third symbol in that horizontal band. For example, in the following Sudoku grid, if we know the placement of two of the symbol 7 in the middle horizontal band then we can determine the position of the other symbol 7 in the middle horizontal band. Since the first given (in red) symbol 7 is in the 0 minirow and 2 minicolumn of the left subsquare of the middle horizontal band and the second given symbol 7 is in the 1 minirow and 0 minicolumn of the right subsquare of the middle horizontal band, we know that the third symbol 7 (in black) must be in the 2 minirow and the 1 minicolumn of the middle subsquare of the middle horizontal band.

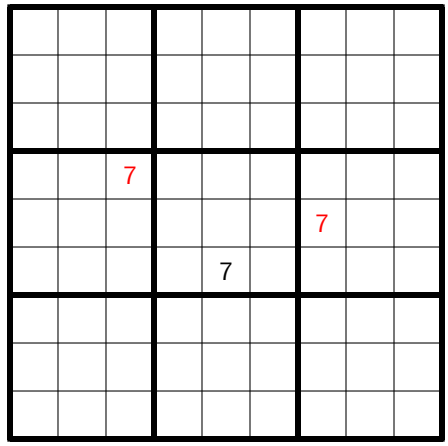

Figure 4: Finding the placement of a symbol in a symmetric Sudoku puzzle

4 Allowable Subspaces and Ratios

A perfect code is still a perfect code after a translation. Hence any perfect code is a coset of a vector subspace.

Definition 8 *A subspace is allowable if it is a perfect code of* $(\mathbb{Z}_3)^4$.

Now we must determine which subspaces are allowable. A perfect code contains nine vectors, thus an allowable subspace must also contain nine vectors and thus have a dimension of 2. In order for such a subspace to be 2-dimentional it must have exactly two basis elements. Since the subspace is a perfect code the distance between any two vectors have distance 3.
Proof.

Lemma 4 *The determinant* $det[m_k, m_l, p, q] = (a_i b_j - b_i a_j)$ *, where* $i, .j, m, l$ *are any of {1, 2, 3, 4}, with* $i \neq j$.
Proof. *The definition for the determinant of a square matrix of order 2 or greater says that*

$$\det(A) = |A| = \sum_{j=1}^{n} a_{j1} C_{j1} = a_{11} C_{11} + a_{21} C_{21} + \cdots + a_{n1} C_{n1} \tag{11}$$

where we expand along the first column. Let A be the matrix 13. Since A has order 4, we will have to find the 3 x 3 minors of the matrix. Let m_k *be arbitrary and* $m_l = (0, 0, 0, 1)$. *The minors are as follows.*

$$M_{11} = \begin{bmatrix} 0 & a_2 & b_2 \\ 0 & a_3 & b_3 \\ 1 & a_4 & b_4 \end{bmatrix}$$

$$M_{21} = \begin{bmatrix} 0 & a_1 & b1 \\ 0 & a_3 & b_3 \\ 1 & a_4 & b_4 \end{bmatrix}$$

$$M_{31} = \begin{bmatrix} 0 & a_1 & b_1 \\ 0 & a_2 & b_2 \\ 1 & a_4 & b_4 \end{bmatrix}$$

$$M_{41} = \begin{bmatrix} 0 & a_1 & b_1 \\ 0 & a_2 & b_2 \\ 0 & a_3 & b_3 \end{bmatrix}$$

Next we must find the determinants of these 3x3 matrices. To do this we add the first and second columns of A to form fourth and fifth columns. Then the determinant is the sum of the three diagonals going from top to bottom and subtracting the three diagonals from the bottom to the top. Doing this for all the matrices M_{ij} we obtain the following determinants.

$$\begin{aligned} |M_{11}| &= a_3b_2 - a_2b_3 \\ |M_{21}| &= a_1b_3 - a_3b_1 \\ |M_{31}| &= a_1b_2 - a_2b_1 \\ |M_{41}| &= 0 \end{aligned}$$

Now we will need to find the cofactors, where the cofactor $C_{ij} = (-1)^{i+j}M_{ij}$. Thus the cofactors are as follows.

$$\begin{aligned} C_{11} &= a_3b_2 - a_2b_3 \\ C_{21} &= -a_1b_3 + a_3b_1 \\ C_{31} &= a_1b_2 - a_2b_1 \\ C_{41} &= 0 \end{aligned}$$

If $k = 1$, then $a_{1j} = 0$ unless $j = 1$. So the $\det A = a_{11}C_{11} = a_3b_2 - a_2b_3$. Similarly if $k = 2, \det A = a_{21}C_{21} = -a_1b_3 + a_3b_1$, if $k = 3, \det A = a_{31}C_{31} = a_1b_2 - a_2b_1$, if $k = 4, \det A = a_{41}C_{41} = 0$. ■

Lemma 5 *Now let $m_l = (0, 1, 0, 0)$. Then*

$$\begin{aligned} M_{11} &= \begin{bmatrix} 1 & a_2 & b_2 \\ 0 & a_3 & b_3 \\ 0 & a_4 & b_4 \end{bmatrix} \\ M_{21} &= \begin{bmatrix} 0 & a_1 & b1 \\ 0 & a_3 & b_3 \\ 0 & a_4 & b_4 \end{bmatrix} \\ M_{31} &= \begin{bmatrix} 0 & a_1 & b_1 \\ 1 & a_2 & b_2 \\ 0 & a_4 & b_4 \end{bmatrix}. \end{aligned}$$

Again using the cofactors we get for $k = 1, \det A = a_{11}C_{11} = a_3b_4 - b_3a_4,$ if $k = 2, \det A = 0,$ if $k = 3, \det A = a_{31}C_{31} = a_4b_1 - b_4a_1$.Now let $m_l = (1, 0, 0, 0)$. Then

$$M_{31} = \begin{bmatrix} 1 & a_1 & b_1 \\ 0 & a_2 & b_2 \\ 0 & a_4 & b_4 \end{bmatrix}.$$

Again using the cofactors we get $k = 3, \det A = a_{31}C_{31} = a_2b_4 - b_2a_4$.

■

Lemma 6 *The vectors $p = (a_1, a_2, a_3, a_4)$, and $q = (b_1, b_2, b_3, b_4)$ span an allowable subspace of V if and only if the four ratios a_i/b_i, for $i = 1, 2, 3, 4$ are distinct, where $\pm 1/0 = \infty$ is one ratio that must appear, and the indeterminate form 0/0 does not appear.*

Proof. We need to show that if p and q span an allowable subspace then the four ratios are distinct, where $\pm 1/0 = \infty$ is one ratio that must appear, and the indeterminate form $0/0$ does not appear. Suppose the span of $\{p, q\}$ is allowable. Since V is an allowable subspace then it is also a perfect code. Then two coordinates can agree in at most one position. Suppose there exist d_1, d_2, d_3, d_4 such that

$$d_1p + d_2q + d_3m_k + d_4m_l = 0, \tag{12}$$

where $k, l \in \{1, 2, 3, 4\}$. Let $d_1 p + d_2 q = (c_1, c_2, c_3, c_4)$. We can also think of m_k and m_l as (1,0,0,0) and (0,1,0,0), respectively. This turns equation 12 into

$$(c_1, c_2, c_3, c_4) + d_3(1, 0, 0, 0) + d_4(0, 1, 0, 0) = 0.$$

We know that (c_1, c_2, c_3, c_4) can only contain one zero coordinate, since the origin is in the span$\{p, q\}$. This implies that $d_i = 0$ therefore the set $\{m_k, m_l, p, q\}$ is linearly independent. Let the zero coordinate be one that is already 0 for both m_k and m_l (for this example c_3 or c_4). We can also look at this in matrix form. We would have the following matrix.

$$\begin{array}{c} \begin{array}{cccc} m_k & m_l & p & q \end{array} \\ \begin{bmatrix} 0 & 0 & a_1 & b_1 \\ 0 & 0 & a_2 & b_2 \\ 1 & 0 & a_3 & b_3 \\ 0 & 1 & a_4 & b_4 \end{bmatrix} \end{array} \tag{13}$$

Since the columns of the matrix are linearly independent, the determinant is nonzero. We know from Lemma 4 that the det$[m_k, m_l, p, q] = (a_i b_j - b_i a_j)$, where $i, .j$ are any of $\{1,\ 2,\ 3,\ 4\}$, with $i \neq j$.

Since $|A| = a_1 b_2 - a_2 b_1 \neq 0$, then the rows in A are linearly independent. This means that $\frac{a_1}{b_1} \neq \frac{a_2}{b_2}$, thus the ratios are distinct and the ratio $\frac{0}{0}$may not appear.

Since $\{m_k, m_l, p, q\}$ are linearly independent, the determinant must be nonzero. Thus $a_i b_j - b_i a_j \neq 0$ and the ratios are distinct and $\frac{0}{0}$ does not appear. There are three choices for a_k and three choices for b_k, namely 0, 1, or 2. Therefore the possibilities for $\frac{a_k}{b_k}$ are as follows:

$$\frac{1}{0}; \frac{0}{1}; \frac{1}{1}; \frac{1}{2}.$$

To prove the converse of this lemma, reverse the argument. ■

From the above Lemma we can conclude that there are only eight vector subspaces formed when a basis for an allowable subspace is put into row-reduced echelon form.

Lemma 7 *The eight vector subspaces formed when a basis for an allowable subspace is put into row-reduced echelon form are the following.*

$$\left\{ \left\{ \begin{array}{c} 1011 \\ or \\ 1022 \end{array} \right\} \ and \ \left\{ \begin{array}{c} 0112 \\ or \\ 0121 \end{array} \right\} \right\} \ or \ \left\{ \left\{ \begin{array}{c} 1012 \\ or \\ 1021 \end{array} \right\} \ and \ \left\{ \begin{array}{c} 0111 \\ or \\ 0122 \end{array} \right\} \right\}$$

Proof. Given an allowable subspace, we know that it must be two dimensional and be spanned by $\langle a, b \rangle$ with $\frac{a_i}{b_i}$ distinct and $\frac{0}{0}$ may not appear. If we are given any two such vectors, say $\langle 0, 2, 1, 1 \rangle$ and $\langle 1, 1, 0, 1 \rangle$, these vectors can be put into matrix form.

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

Putting this matrix into row-reduced echelon form we get,

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

This is one combination of our eight possibilities. Without going through all the other vectors we can see why these are in fact the only eight possibilities. Since the vectors are to be put in to row-reduced echelon form, we know the first two positions of the top vector have to be 1 and then 0. Likewise the first two positions of the bottom vector have to be 0 and then 1. Thus the start of our matrix is of the following form.

$$\begin{bmatrix} 1 & 0 & a & b \\ 0 & 1 & c & d \end{bmatrix} \tag{14}$$

We also know that $a, b, c, d \in \mathbb{Z}_3$ and since the vectors can only have one position being 0,otherwise the ratios would not be distinct, the only possibilities for a, b, c and d are ± 1. If we determine what a, b,and c are then we only have one choice for d, since the ratios are distinct. Since d is determined by a, b, and c, there are only eight possible combinations for the vector subspaces. ■

5 Proof of Main Theorem

We have determined thus far that for symmetric Sudoku solutions:

- Any symmetric Sudoku solution is linear.
- In a symmetric Sudoku solution, the positions of each symbol form a coset of one of the eight allowable subspaces.

The following two Lemmas will aid in the proof of the main theorem.

Lemma 8 *Given any allowable subspace X, there exist exactly 4 allowable subspaces X' such that span$\{X, X'\}$ is 3-dimensional.*

Proof. If A_i and A_j are the same set of vectors, their span will be 2-dimensional. Let

$$X = span\left\{(1, 0, a_1, a_2), (0, 1, b_1, b_2)\right\}$$

and

$$X' = span\left\{(1, 0, -a_1, a_2), (0, 1, -b_1, b_2)\right\}.$$

To get these two vector spaces to form a 3-dimensional space, they must be multiplied as follows.

$$a\begin{pmatrix}1\\0\\a_1\\a_2\end{pmatrix} + b\begin{pmatrix}1\\0\\b_1\\b_2\end{pmatrix} - a\begin{pmatrix}1\\0\\-a_1\\a_2\end{pmatrix} - b\begin{pmatrix}1\\0\\-b_1\\b_2\end{pmatrix} = \begin{pmatrix}0\\0\\2aa_1 + 2bb_1\\0\end{pmatrix}$$

Therefore these cosets are linearly dependent since in order for

$$2aa_1 + 2bb_1 = 0$$

a and b do not have to both be zero. For example, $a = -1$ and $b = 1$. Now if we change the diagonals of the last two positions of the coordinate we get the following.

$$X = span\left\{(1, 0, a_1, a_2), (0, 1, b_1, b_2)\right\}$$

and

$$X' = span\left\{(1, 0, -a_1, a_2), (0, 1, b_1, b_2)\right\}.$$

We still have to multiply by a, b, $-a$, and $-b$ in order to eliminate the 1 terms. After doing so we get the following.

$$a\begin{pmatrix}1\\0\\a_1\\a_2\end{pmatrix} + b\begin{pmatrix}1\\0\\b_1\\b_2\end{pmatrix} - a\begin{pmatrix}1\\0\\-a_1\\a_2\end{pmatrix} - b\begin{pmatrix}1\\0\\b_1\\-b_2\end{pmatrix} = \begin{pmatrix}0\\0\\2aa_1\\2bb_2\end{pmatrix}$$

In this case, the cosets form a 4-dimensional space and the cosets are linearly independent since to get the vector equation to equal zero, we must have the trivial solution where $a = b = 0$. This argument also work for the opposite diagonal. If

$$X = span\left\{(1, 0, a_1, a_2), (0, 1, b_1, b_2)\right\}$$

and

$$X' = span\left\{(1,0,-a_1,-a_2),(0,1,-b_1,-b_2)\right\}$$

then again multiplying by a, b, $-a$, and $-b$ we get

$$a\begin{pmatrix}1\\0\\a_1\\a_2\end{pmatrix}+b\begin{pmatrix}1\\0\\b_1\\b_2\end{pmatrix}-a\begin{pmatrix}1\\0\\-a_1\\-a_2\end{pmatrix}-b\begin{pmatrix}1\\0\\-b_1\\-b_2\end{pmatrix}=\begin{pmatrix}0\\0\\2aa_1+2bb_1\\2aa_2+2bb_2\end{pmatrix}.$$

Since

$$a_1b_2 = -a_2b_1$$

then if b_1 and b_2 have the same sign then $a_1 = -a_2$ and like wise if a_1 and a_2 have the same sign then $b_1 = -b_2$. In either case $a = b = 0$. This again gives us a 4-dimensional space. Therefore there are only 4 combinations that will give you a 3-dimensional space. If

$$X = span\left\{(1,0,a_1,a_2),(0,1,b_1,b_2)\right\}$$

then the four compatible allowable subspaces are

$$\begin{aligned}X_1' &= span\left\{(1,0,-a_1,a_2),(0,1,-b_1,b_2)\right\},\\ X_2' &= span\left\{(1,0,a_1,-a_2),(0,1,b_1,-b_2)\right\},\\ X_3' &= span\left\{(1,0,-a_1,-a_2),(0,1,b_1,b_2)\right\},\end{aligned}$$

and

$$X_4' = span\left\{(1,0,a_1,a_2),(0,1,-b_1,-b_2)\right\}.$$

■

Theorem 9 *There are 104 symmetric Sudoku solutions up to equivalence.*

Proof. We look at how the eight vector subsets can partition V. If $\{S_i\}_{i\in I}$ is a partition of V, then $S_i \cap S_j = \emptyset$ and $\cup S_i = V$. Thus for S_i, the set of positions for the symbol i, there is an allowable subspace A_i and an element $a_i \in V$ such that $S_i = A_i + a_i$. There are three possibilities for partitions of V.

There are two types of linear symmetric Sudoku solutions: parallel and nonparallel. A *parallel* Sudoku solution is a solution that has all nine affine subspaces being parallel and thus cosets are of the same vector subspace. Otherwise the linear Sudoku solution is of *nonparallel type.*

Case 1 All A_i are the same therefore all the S_i are cosets of one of the above mentioned 2-dimensional vector subspaces. This partition would give a solution that is of "parallel type". This gives 8 solutions since A can be any of the 8 allowable subspaces up to equivalences: meaning the symbols could be switched but would still be the same eight allowable subspaces.

Case 2 $\exists i,j$ such that A_i and A_j are two distinct vectors and the span of $\langle A_i, A_j\rangle = V$, i.e. for all $v \in V$ there exists an $x, y \in A_i, A_j$ such that $rx + sy = v$. Since $A_i + a_i$ and $A_j + a_j$ are two sets in a partition of V, they are disjoint, meaning

$$A_i + a_i \cap A_j + a_j = \emptyset.$$

On the other hand $a_i \in V$ and the span of $\langle A_i, A_j\rangle = V$. So there exists $x_i \in A_i$ and $y_i \in A_j$ and $r_i, s_i \in \mathbb{Z}_3$ such that

$$a_i = r_ix_i + s_iy_i.$$

Similarly

$$a_j = r_jx_j + s_jy_j.$$

So

$$\begin{aligned}A_i + a_i &= A_i + r_ix_i + s_iy_i\\ &= A_i + s_iy_i\end{aligned}$$

since $r_i x_i \in A_i$. Also

$$\begin{aligned} A_j + a_j &= A_j + r_j x_j + s_j y_j \\ &= A_j + r_j x_j \end{aligned}$$

since $s_j y_j \in A_j$. Now $r_j x_j \in A_i$, so

$$r_j x_j + s_i y_i \in A_i + s_i y_i = A_i + a_i.$$

But $s_i y_i \in A_j$, so

$$s_i y_i + r_j x_j \in A_j + r_j x_j = A_j + a_j.$$

This contradicts the fact that $A_i + a_i \cap A_j + a_j = \emptyset$. Thus Case 2 cannot happen.

Case 3 There are three or more distinct A_i, i.e. $A_i \neq A_j \neq A_k$. If there are more than three distinct A_i then without the loss of generality the span$\{A_i, A_j\}$ has dimension 3, the span$\{A_i, A_k\}$ has dimension 3, and the span$\{A_j, A_k\}$ has dimension 3. Now, $A_i = span\,\{(1, 0, a_1, a_2), (0, 1, b_1, b_2)\}$. From the proof of Lemma 8, we see that the only compatible subspaces A_j must have either the signs of a_1 and a_2 switched, the signs of b_1 and b_2 switched, the signs of a_1 and b_1 switched, or the signs of a_2 and b_2 switched. Similarly, the subspace A_k must have one of these sets of signs switched. But then A_j and A_k will not be compatible, because A_k will not be obtained from A_j by switching one set of these signs. Thus this case may not happen.

Case 4 A_i and A_j are distinct and span $\langle A_i, A_j \rangle$ is three dimensional. Let Y be a three-dimensional vector subspace and let span $Y = \langle A_i, A_j \rangle$. Since Y is 3-dimensional it has 27 elements and has distinct cosets of Y in V. Let the span of $(X,\ X')$be 3-dimensional and $V = Y \cup (Y + v_1) \cup (Y + v_2)$. Since $X \cap X'$ is a vector space then the dimension of $(X \cap X') = 0$ *or* 1. The dimension formula states that the $\dim(X + X') = \dim(X) + \dim(X') - \dim(X \cap X')$,thus since $\dim(X) = \dim(X') = 2$, then $\dim(X \cap X')$ must equal 1. Suppose

$$Y = X \cup X + x_1 \cup X' + x_2.$$

If $x_2 \in X$, then

$$\begin{aligned} X &= (X + x_2) \cap (X' + x_2) \\ &= X \cap X' \neq \emptyset. \end{aligned}$$

Thus $x_2 \notin X$. So $X + x_2 \neq X' + x_2$. Since $X + x_2$ is a set of 9 points, then $(X + x_2) \cap (X' + x_2) = (X \cap X') + x_2$ which is 1-dimensional. If we have X and $(X + x_1)$ then $(X + x_2)$ is either equal to one of the partitions X or $(X + x_1)$ or is disjoint. If

$$X + x_2 \neq X$$

and

$$X + x_2 \neq X' + x_2$$

then

$$X + x_1 = X + x_2.$$

This is a contradiction.

We know that from Lemma 7, there are 4 possible compatible subspaces of each allowable subspace X such that the *span*$\{X, X'\}$ is 3-dimensional. Let the three cosets of Y partition V. We now show that cosets of only one of X or X' may appear in each of the three cosets of Y.

From this analysis we can see that from our 8 allowable subspaces there are 4 possible choices for another subspace that will produce a span that is 3-dimensional. Therefore there are $8 \cdot 4/2 = 16$ possible choices for the parallel type subspaces. (We divide by 2 to eliminate duplicates that are in a different order.) There are $\binom{3}{2}$ ways to partition the space. This is because we have three cosets of Y, namely $Y, (Y + v_1)$, and $(Y + v_2)$. For each of these three cosets of Y, Y can be partitioned in two ways, by either X or X'. We only choose two of the three possible cosets of Y to partition Y. Thus there are 6 ways of doing this, $2 \cdot \binom{3}{2}$. Therefore there are $6 \cdot 16 = 96$ possible Sudoku solutions of this type. There are also 8 parallel type solutions for a total of $8 + 96 = 104$ total possible symmetric Sudoku solutions. ■

6 Conclusions

The attempt to generalize these results to other more general Sudoku grids was not obtained as stated in the proposal. After a closer examination of the series of lemmas and proofs, we discovered that in order to go up a class, such as to have a 16x16 grid containing 4x4 subgrids, would allow for more ratios and subspaces that are not as friendly to handle. Many of the proofs hinge on the fact that we are in $\mathbb{Z}_3$ and would not work as well in a larger group.

I will be teaching math in high school and I am always looking for new ways to apply topics to real life situations to help my students learn. When I started this project, my hope was to be able to pull out some of the math involved in this project that could be used as an example for the high school setting. However, after completing this project, I realize that the mathematics used in this project were above the skills of most high school students.

Nonetheless, there are some ways in which to use Sudoku to enhance learning at the high school level. In order to complete a Sudoku puzzle, one must use problem solving strategies and logic skills. Students could work on a puzzle and then discuss the strategies that they used and how well they worked for them. This could then lead into a discussion on problem solving strategies for word problems. When a class is talking about patterns in problems, symmetric Sudoku solutions could be show to see if the students could find all the patterns that occur for each of the numbers.

Sudoku swept our nation quickly, but I predict that it will not leave as quickly as it came in.

References

[Bailey] Bailey, R., Cameron, and Connelly. Sudoku, gerechte designs, resolutions, affine space, spread, reguli, and Hamming codes. to appear: Amer. Math Monthly

[Delahye, 2006] Delahye, J. The Science behind SUDOKU. June 2006: Scientific American. Vol 294.6. p 80-87.

[Frie, 1989] Friedberg, Stephen H., Arnold J. Insel, and Lawrence E. Spence. Linear Algebra: Second Edition. Prentice Hall: 1989.

[Lar, 2004] Larson, R., Bruce H. Edwards, and David C. Falvo. Elementary Linear Algebra. 5th ed. New York: Houghton, 2004.

[Rot, 2000] Rotman, J. A first course in Abstract Algebra: Second Edition. Prentice Hall: 2000.

[Lint, 1982] van Lint, J. H. Introduction to Coding Theory. Springer-Verlag. New York: 1982.