Journal of Integer Sequences, Vol. 12 (2009),
Article 09.6.5

On Determining Paint by Numbers Puzzles with Nonunique Solutions

Ryan Mullen
Sacred Heart University
Fairfield, CT 06825
USA
mullenr@sacredheart.edu

Abstract

Paint by Numbers is a classic logic puzzle in which the squares of a $p \times n$ grid are to be colored in such a way as to display a picture. The decision on which squares to color is determined by sequences of numbers above each column and to the left of each row. The numbers describe how many consecutive squares are to be colored in that row or column, and multiple numbers represent multiple blocks of colored in squares (with at least one uncolored square inbetween blocks.) Certain natural questions arise. For a given $p \times n$ grid, how many possible sequences are in a single column or row? For a given grid, how many puzzles are there? How many of these have unique solutions? We will explore these questions as well as connections between Paint by Numbers puzzles, partition theory, and the Fibonacci sequence.

1 Introduction

Paint by Numbers (also called Nonograms, Picross, and many other names) is a logic puzzle which allows purely left brained thinkers to draw pictures of Matt Damon, Jim Morrison, a shootout at a poker table, or any picture that can be converted into a two toned highly pixelated version. The reward for completing the puzzle correctly is a completed drawing; however, solving the puzzle needs no greater artistic ability than staying inside the lines. Ueda and Nagao [4] have shown that determining if a solution is unique is NP-complete (a famous complexity class from computer science). Batenburg and Kosters [1] have shown that under certain assumptions puzzles can be solved in polynomial time. In addition Benton, Snow, and Wallach [2] compute the number of possible row and column sums.

The puzzle starts out with a $p \times n$ grid with sequences of numbers above each column and sequences of numbers alongside each row. The numbers tell you how many consecutive squares in the column or row need to be colored in. If there is more than one number in the sequence then there must be at least one uncolored (white) space between the colored (black) spaces. Figure 1 is an example of a small puzzle.

	1	2 1	5	2 1	1
1					
3					
3					
1					
5					

Figure 1: Example of a small puzzle

Initially we see that along the bottom and in the middle column we must have 5 black squares, and there is exactly room for 5. Thus we can fill in the following:

	1	2 1	5	2 1	1
1			■		
3			■		
3			■		
1			■		
5	■	■	■	■	■

Figure 2: Intermediate stage

Next we note that the top row is completed. That leaves only four remaining squares in which to fill columns 2 and 4 and that is just barely enough. See Figure 3.

	1	2 1	5	2 1	1
1			■		
3		■	■	■	
3		■	■	■	
1			■		
5	■	■	■	■	■

Figure 3: Finished picture

The puzzle is complete, and we have "drawn" a lonely tree.

If a sequence such as $3, 1, 4$ is listed alongside a row of the puzzle, then we know that there must be 3 black squares, at least one white, one black square, at least one white, and then 4 black squares. There may or may not be white squares before the 3 black and after the 4 black. The order is important as $4, 1, 3$ would describe a different coloring. Depending on how large the puzzle is, we may or may not be sure certain squares are colored. For example if $n = 10$ then we know exactly which squares are to be colored (see Figure 4).

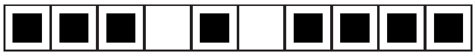

Figure 4: Forced coloring with 10 squares

If $n = 11$, then there are four possible configurations; however all four have the black squares in Figure 5.

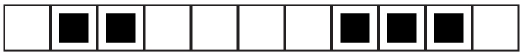

Figure 5: Forced coloring with 11 squares

Filling in the remaining black squares can be done using information from the columns and the other rows. If $n = 14$ or more, then we would not be able to fill in any of the squares in this row without using information from the rest of the puzzle because this sequence uses 10 squares and there would be at least 4 left over. That leaves the largest piece of 4 with two nonoverlapping possible locations. We define a *forceless* sequence to be a sequence such that no partial coloring can be determined without the aid of the rest of the puzzle (see Figure 6).

2 Number of Possible Sequences

For any $p \times n$ puzzle there are p sequences of natural numbers one to the left of each row and there are n sequences of natural numbers one above each column. For a row sequence $\{a_j\}_{j=1}^{i} = \mathbf{a_j}$ we define $l(\mathbf{a_j})$ to be the *length* of the sequence, which is the fewest number of columns needed for that sequence to fit in a puzzle. Since there must be at least one white square between each string of black squares the formula is

$$l(\mathbf{a_j}) = \sum_{j=1}^{i} a_j + i - 1.$$

Also define the norm of a row sequence as

$$\|\mathbf{a_j}\| = \max_{1 \le j \le i} a_j.$$

A row sequence $\mathbf{a_j}$ in a puzzle with n columns would be forceless if $\mathbf{a_j} \neq \{\}$ and

$$l(\mathbf{a_j}) + \|\mathbf{a_j}\| \le n.$$

Column sequences have corresponding definitions but we will focus on row sequences. For a given n, just how many sequences are allowable? Somewhat surprisingly the number of sequences follows a well known sequence. The Fibonacci sequence F_n (A000045) is determined recursively as

$$F_1 = F_2 = 1;$$
$$F_n = F_{n-1} + F_{n-2}.$$

Theorem 2.1. *For a puzzle with n columns, the number of row sequences that describe a coloring which fits in the puzzle is F_{n+2}.*

Proof. Let b_n be the number of sequences in a puzzle with n columns. For the case $n = 1$, there are 2 possible sequences: $\{\}$ and 1. Thus $b_1 = 2$. For the case $n = 2$, there are 3 possible sequences: $\{\}$, 1, and 2. Thus $b_2 = 3$. If $n > 2$, the number of sequences allowed is b_{n-1} plus the number of sequences of length n. If a sequence has length n and has i entries, then there must be $i-1$ white squares and $n-i+1$ black squares. To count the number of sequences, think of starting with a long string of $n-i+1$ black squares and then choosing $i-1$ places to cut the string of black squares and placing a white squares in the cuts. We can only cut between black squares and not in the middle of any square. Thus there are $n-i$ possible cutting points and $i-1$ cuts needed to be made, and the cutting order is irrelevent which gives $\binom{n-i}{i-1}$ sequences of i numbers of length n. Thus

$$\begin{aligned}
b_n &= b_{n-1} + \sum_{i=1}^{\lfloor \frac{n+1}{2} \rfloor} \binom{n-i}{i-1} \\
&= b_{n-1} + \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-i-1}{i} \\
&= b_{n-1} + 1 + \sum_{i=1}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-i-2}{i} + \binom{n-i-2}{i-1} \\
&= b_{n-1} + \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n-i-1}{i-1} + \sum_{i=1}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-i-2}{i-1} \\
&= 2b_{n-1} - b_{n-3}.
\end{aligned}$$

If we assume $b_{n-1} = b_{n-2} + b_{n-3}$ and use induction, we have

$$\begin{aligned}
b_n &= b_{n-1} + b_{n-2} + b_{n-3} - b_{n-3} \\
&= b_{n-1} + b_{n-2}.
\end{aligned}$$

So the sequence b_n has the same recursion as the Fibonacci sequence but is shifted since b_n starts with 2 and 3 rather than 1 and 1. So $b_n = F_{n+2}$. □

This proof also illustrates the closed form formula for the Fibonacci numbers

$$F_n = \sum_{i=1}^{\lfloor \frac{n+1}{2} \rfloor} \binom{n-i}{i-1}.$$

When counting the number of sequences of length n, we are almost counting the partitions of $n-i$ into $i-1$ pieces and to get all the partitions of $n-i$ we could simply sum over all possible numbers of pieces. If it were not for the seemingly trivial detail that some of the partitions could be counted numerous times, this would work. For example the partition $\{1,1,2\}$ of 5 would get counted 3 times since $(2,1,1),(1,2,1)$ and $(1,1,2)$ are different row sequences.

3 Forceless Sequences

In the process of solving a Paint by Numbers puzzle, the first thing to look for are sequences that force a coloring. But just how rare is a sequence with no forcing? A sequence would be forceless if in at least two possible colorings, the largest block of squares are nonoverlapping. In Figure 6 we see that the largest block of black squares, 4, has no overlapping between the extreme left and extreme right color placements. Thus $3,1,4$ would be a forceless row sequence in a puzzle with 14 columns. The goal of the remainder of the paper is to count the number of forceless sequences.

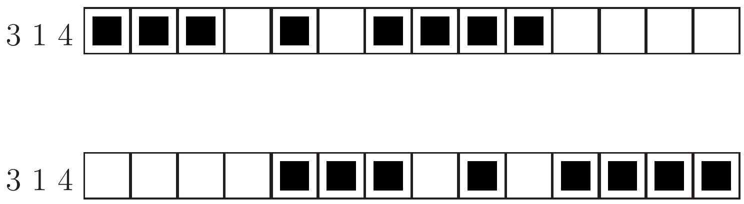

Figure 6: Extreme left and extreme right of color placement

In general, we get a forceless sequence in a puzzle with n columns if the row sequence $\{a_j\}$ satisfies

$$\|\mathbf{a}_j\| + l(\mathbf{a}_j) \leq n.$$

For the moment, let's consider sequences with a fixed length m. If this sequence is to be forceless, then its maximum norm is $n-m$. Let $A_{m,k}$ denote the sequences of length m and maximum norm k. To count all forceless sequences in a puzzle with n columns we need only sum $A_{m,n-m}$ over all possible lengths. That is, the number of forceless sequences, denoted G_n, is

$$G_n = \sum_{m=0}^{n-1} A_{m,n-m}.$$

To calculate $A_{m,k}$, it is easier to fix the maximum norm and let the length vary. We start with $k=1$. Since the maximum norm is one, the sequence must consist entirely of ones. Say there are r ones. The length of a sequence consisting entirely of ones is $l(\mathbf{a}_j) = r + r - 1 = 2r - 1$. Therefore if m is odd, there is one sequence with maximum norm 1, and if m is even, there are no such sequences. That is $A_{m,1} = 1$ for m odd, and $A_{m,1} = 0$ for m even.

Moving on to $k = 2$, there is only the sequence $\{1\}$ of length 1 and there is only the sequence $\{2\}$ of length 2. So $A_{1,2} = A_{2,2} = 1$. Considering $A_{m,2}$, we are counting the number of sequences of length m and maximum norm 2. There are two ways these sequences could come about. Either we addend a 1 to a sequence of length $m-2$ or addend a 2 to a sequence of length $m - 3$. Therefore $A_{m,2} = A_{m-2,2} + A_{m-3,2}$.

For a general k, similar reasoning shows $A_{1,k} = A_{2,k} = 1$. For $m \leq k$, the sequence $\{m\}$ has length m and max norm less than or equal to k, and is not obtained from addending a number to a previously counted sequence. However, all other sequences are obtained in this manner. The remaining sequences of length m are obtained by addending a 1 to the sequences of length $m-2$, addending a 2 to the sequences of length $m-3$,..., or addending a $m-2$ to the sequence of length 1. If $m > k$, then the sequence $\{m\}$ has a norm which is too large; thus that sequence would not be counted. So all sequences are obtained by addending a number to a previous sequence. The sequences of length m are obtained by addending a 1 to the sequences of length $m - 2$, addending a 2 to the sequences of length $m - 3$,..., or addending an m to the sequences of length $m - k - 1$. From this we obtain the following formula:

$$A_{m,k} = \begin{cases} 1 + \sum_{j=1}^{m-2} A_{j,k}, & m \leq k; \\ \sum_{j=m-k-1}^{m-2} A_{j,k}, & m > k. \end{cases}$$

Simple induction arguments can be used to obtain

$$A_{m,k} = \begin{cases} F_m, & m \leq k; \\ A_{m-1,k} + A_{m-2,k} - A_{m-k-2,k}, & m > k. \end{cases}$$

For $k = 1, \ldots, 9$, these sequences are in the On-Line Encyclopedia of Integer Sequences, $A_{m,1}$ (A000035), $A_{m,2}$ (A000931), $A_{m,3}$ (A013979), $A_{m,4}, \ldots, A_{m,9}$ (A013982),…,(A013987).

It is worth noting that for every fixed m, $A_{m,k}$ is an increasing sequence in k. In other words, if the length is fixed and the maximum norm is allowed to increase, more sequences would be obtained. Also for every fixed $k > 1$, $A_{m,k}$ is an increasing sequence in m. Also notice that if we fix a length m and consider the sequence in that row, the sequence is bounded. In fact, it is bounded by the mth Fibonacci number. This should by no means be surprising since m represents the length of sequence, and as we increase the maximum norm (largest piece size) we will eventually get all sequences of length m, which we have already decided is determined by the Fibonacci sequence.

To obtain G_n, we simply add all terms such that $m + k = n$; that is, add along diagonals in Table 1. Table 2 gives the values of G_n up to $n = 11$.

So far we have outlined a way of computing the number of forceless row sequences for any number of columns n. It would be nice to have a closed form formula to do this more efficiently. That goal seems unlikely to be met, however other questions can be answered about $\{G_n\}$. In the remainder of the paper we will discuss the growth rate of $\{G_n\}$ and how

$m, k \rightarrow$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	0	1	1	1	1	1	1	1	1	1
3	1	1	2	2	2	2	2	2	2	2
4	0	2	2	3	3	3	3	3	3	3
5	1	2	4	4	5	5	5	5	5	5
6	0	3	5	7	7	8	8	8	8	8
7	1	4	8	10	12	12	13	13	13	13
8	0	5	11	16	18	20	20	21	21	21
9	1	7	17	24	29	31	33	33	34	34
10	0	9	24	37	45	50	52	54	54	55

Table 1: $A_{m,k}$ for $1 \leq m, k \leq 10$

$$
\begin{aligned}
G_2 &= 1 \\
G_3 &= 0 + 1 = 1 \\
G_4 &= 1 + 1 + 1 = 3 \\
G_5 &= 0 + 1 + 1 + 1 = 3 \\
G_6 &= 1 + 2 + 2 + 1 + 1 = 7 \\
G_7 &= 8 \\
G_8 &= 15 \\
G_9 &= 20 \\
G_{10} &= 33 \\
G_{11} &= 47
\end{aligned}
$$

Table 2: G_n for small n

that growth relates to the growth rate of the Fibonacci sequence. In particular we will prove the following results.

Theorem 3.1. *For every n, $G_n \geq C_1 \frac{\phi^n}{n}$, where ϕ is the golden ratio.*

Theorem 3.2. *For every n, $G_n \leq C \frac{\ln n \phi^n}{n}$.*

But first we present an alternate method for computing the sequence $\{G_n\}$.

4 Generating Polynomial

We can also use polynomials to find the sequence $\{G_n\}$. Consider the polynomial

$$x^i \left(\sum_{j=1}^{k} x^j \right)^{i+1}. \tag{4.1}$$

For a fixed k, let us examine this polynomial as i goes from 0 to ∞. When $i = 0$, the coefficient of x^m in (4.1) is the number of ways to pick one number between 1 and k with sum m. In the context of Paint by Numbers puzzles, the coefficient of x^m is the number of sequences with norm k, length m, and one entry. When $i = 1$, the coefficient of x^{m-1} in $\left(\sum_{j=1}^{k} x^j\right)^{i+1}$ is the number of ways to pick two numbers between 1 and k with sum $m-1$. Again in our context, we have the number of sequences with norm k, length m, and two entries. Thus the coefficient of x^m in (4.1) is the number of sequences with norm k, length m, and two entries. For a general i, the coefficient of x^{m-i} in $\left(\sum_{j=1}^{k} x^j\right)^{i+1}$ is the number of sequences with norm k, length m, and $i+1$ entries. Hence the coefficient of x^m in (4.1) is the number of sequences with norm k, length m and $j+1$ pieces. If we sum over all possible numbers of pieces, we will have all sequences of length m and norm k. Therefore,

$$\sum_{m=1}^{\infty} x^m A_{m,k} = \sum_{j=0}^{\infty} x^j \left(\sum_{i=1}^{k} x^i \right)^{j+1}. \tag{4.2}$$

Now we compute the generating polynomial for G_n as follows:

$$\begin{aligned}
\sum_{n=1}^{\infty} G_n x^n &= \sum_{n=1}^{\infty} x^n \sum_{k=1}^{n-1} A_{n-k,k} \\
&= \sum_{k=1}^{\infty} \sum_{n=k+1}^{\infty} x^n A_{n-k,k} \\
&= \sum_{k=1}^{\infty} x^k \sum_{m=1}^{\infty} x^m A_{m,k} \\
&= \sum_{k=1}^{\infty} x^k \sum_{j=0}^{\infty} x^j \left(\sum_{i=1}^{k} x^i \right)^{j+1}.
\end{aligned}$$

5 Lower Bound for G_n

5.1 The Main Idea

Unfortunately G_n is not itself a linear recursion so finding a closed form formula for it is a daunting task. However, the columns (fix k and consider the sequence $A_{m,k}$ in m) are linear recursions. We will use this fact along with some clever observations to find asymptotic results for G_n.

$m,k \rightarrow$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	0	1	1	1	1	1	1	1	1	1
3	1	1	2	2	2	2	2	2	2	2
4	0	2	2	3	3	3	3	3	3	3
5	1	2	4	4	5	5	5	5	5	5
6	0	3	5	7	7	8	8	8	8	8
7	1	4	8	10	12	12	13	13	13	13
8	0	5	11	16	18	20	20	21	21	21
9	1	7	17	24	29	31	33	33	34	34
10	0	9	24	37	45	50	52	54	54	55

Table 3: $A_{m,k}$ for $1 \leq m, k \leq 10$, red numbers sum to G_{11}

First let us find a lower bound for G_n. When calculating G_{11}, we add the red numbers in Table 3. Now choose any column and add the red number and all numbers above it in that column. No matter which column you choose you will obtain a smaller result than G_{11}. This is the idea behind the following proposition.

Proposition 5.1. *For every k, $G_m \geq \sum_{j=1}^{m-k} A_{j,k}$.*

Proof. Fix $k \geq 1$. If $m - j \geq k$, then $A_{j,m-j} \geq A_{j,k}$. This is because for a fixed length j, as the maximum norm increases, so does the number of sequences. Therefore

$$\begin{aligned} G_m &= \sum_{j=1}^{m-k} A_{j,m-j} + \sum_{j=m-k+1}^{m-1} A_{j,m-j} \\ &\geq \sum_{j=1}^{m-k} A_{j,k} + 0. \end{aligned}$$

□

5.2 Linear recursions

When studying linear recursions such as the Fibonacci sequence, we notice that the ratios of consecutive terms seem to approach a fixed number (call it the *limiting ratio*). If the ratios of consecutive terms were a fixed number ρ then then mth term of the sequence would be ρ^m and the sequence would also have to satisfy the Fibonacci recursion

$$\rho^m = \rho^{m-1} + \rho^{m-2}.$$

If we factor out ρ^{m-2}, we the get characteristic equation for the Fibonacci recursion

$$\rho^2 = \rho + 1.$$

Solving for ρ, we obtain two roots $\phi = \frac{1+\sqrt{5}}{2}$ and $\lambda_1 = \frac{1-\sqrt{5}}{2}$. We find another closed form formula for the Fibonacci sequence by writing

$$F_n = b_0\phi^n + b_1\lambda_1^n,$$

and then using the fact that $F_1 = 1, F_2 = 1$ to solve for the constants b_0 and b_1, we obtain

$$F_n = \frac{\phi^n - \lambda_1^n}{\sqrt{5}}.$$

Notice since ϕ is the larger of the two roots (in absolute value), as n increases the Fibonacci numbers grow very closely to $\frac{\phi^n}{\sqrt{5}}$.

Just as with all linear recursions, for a fixed k, the terms of $A_{m,k}$ can be expressed in a closed form manner using the zeros of the recursion's characteristic polynomial. In this case the characteristic polynomial is

$$x^{k+2} - x^{k+1} - x^k + 1.$$

This polynomial has $k+2$ zeros (counting multiplicity), and for $k > 1$, one of these zeros is real and larger than 1. We will call this zero ϕ_k with multiplicity α_0. Label the remaining zeros as $\lambda_1, \ldots, \lambda_j$ with multiplicities $\alpha_1, \ldots, \alpha_j$ (most of these are complex numbers). Using linear recursion theory, we can write

$$A_{m,k} = \sum_{i=1}^{\alpha_0} b_{i,0} m^{i-1}\phi_k^m + \sum_{l=1}^{j}\sum_{i=1}^{\alpha_j} b_{i,l} m^{i-1}\lambda_l^m. \tag{5.1}$$

The coefficients $b_{i,l}$ can be determined using the first $k+2$ terms of the recursion. So in theory, nice formulas can be obtained for $A_{m,k}$; however computing the zeros and then solving the systems of linear equations can only be done approximately. Nontheless determining which zero is the largest in absolute value will go a long way in approximating $A_{m,k}$.

Proposition 5.2. *Let $k > 1$ and $\phi_k, \lambda_1, .., \lambda_j$ be the zeros of $f_k(z) = z^{k+2} - z^{k+1} - z^k + 1$, where ϕ_k is the real zero that is larger than 1. For all $1 \le i \le j$, $\phi_k > |\lambda_i|$. Furthermore, ϕ_k is a zero of multiplicity one.*

Proof. Since ϕ_k is a zero of $f_k(z)$, $(z - \phi_k)$ is a factor. A simple multiplication shows that

$$f_k(z) = (z - \phi_k)(z^{k+1} + (\phi_k - 1)z^k - \sum_{j=0}^{k-1} z^j\phi_k^{-j-1}).$$

Consider the function $g_k(z) = (\phi_k - 1)z^k - \sum_{j=0}^{k-1} z^j\phi_k^{-j-1}$. For all z such that $|z| = \phi_k$, we have

$$|g_k(z)| \le (\phi_k - 1)\phi_k^k + \sum_{j=0}^{k-1} \phi_k^j\phi_k^{-j-1}$$

$$
\begin{aligned}
&= \phi_k^{k+1} - \phi_k^k + \sum_{j=0}^{k-1} \phi_k^{-1} \\
&= \phi_k^{k+1} - \phi_k^k + k\phi_k^{-1} \\
&< \phi_k^{k+1}.
\end{aligned}
$$

Where the last inequality is justified because for all $k \geq 2$, $\phi_k \geq \phi_2 > \sqrt[3]{2} \geq \sqrt[k+1]{k}$. On the closed curve $|z| = \phi_k$, we have $|g_k(z)| \leq \phi_k^{k+1} = |z^{k+1}|$. Thus Rouche's theorem (see any complex analysis textbook, for example [3]) tells us that z^{k+1} and $z^{k+1} + g_k(z)$ have the same number of zeros inside the curve $|z| = \phi_k$. The polynomial z^{k+1} has $k+1$ zeros inside this curve. Therefore all $k+1$ zeros of $z^{k+1} + g_k(z)$ are inside this curve. That is, for all $1 \leq i \leq j$, $\phi_k > |\lambda_i|$ and ϕ_k has multiplicity one. □

In light of (5.1) and Proposition 5.2, we see that for large m the term $b_{1,0}\phi_k^m$ is the dominating factor in computing $A_{m,k}$. Furthermore

$$\lim_{m\to\infty} \frac{A_{m+1,k}}{A_{m,k}} = \phi_k.$$

Clearly, $\phi_k < \phi$ for all k; the sequences $A_{m,k}$ grow slower than the Fibonacci sequence. Also $\lim_{k\to\infty} \phi_k = \phi$; as k goes to infinity $A_{m,k}$ goes to the Fibonacci sequence. Furthermore, for $k_1 < k_2$, we have $\phi_{k_1} < \phi_{k_2}$. We can generalize the definition of ϕ_k for nonintegers $k > 1$ as well by defining ϕ_k to be the largest positive real zero of $f_k(z) = z^{k+2} - z^{k+1} - z^k + 1$. Now that we know which zero to focus on, we must find a way to approximate it. We begin the following subsection with a lower bound for ϕ_k.

5.3 Proof of Theorem 3.1

Proposition 5.3. *For every $k \geq 8$, $\phi_k \geq \phi\left(\frac{\sqrt{5}\phi^{k+1}-(k+2)}{\sqrt{5}\phi^{k+1}-(k+1)}\right)$.*

Proof. Let $f_k(x) = x^2 + x - 1 - x^{k+2}$. The largest zero of $f_k(x)$ will be the reciprocal of ϕ_k. We use a linear approximation to f_k at ϕ^{-1} and the following:

$$
\begin{aligned}
f_k(\phi^{-1}) &= -\phi^{-k-2} \\
f_k'(\phi^{-1}) &= \sqrt{5} - (k+2)\phi^{-k-1} \\
f_k''(x) &= 2 - (k+2)(k+1)x^k.
\end{aligned}
$$

So the linear approximation is

$$f_k(x) \approx (\sqrt{5} - (k+2)\phi^{-k-1})(x - \phi^{-1}) - \phi^{-k-2}.$$

For $k \geq 8$, $2 - (k+1)(k+2)\phi^{-k} > 0$. Thus $f_k(x)$ is concave up at $x = \phi^{-1}$ and the linear aproximation is an underestimate. Now because the linear approximation is an underestimate and $f_k(\phi^{-1})$ is negative, the approximation of ϕ_k^{-1},call it $\widetilde{\phi_k^{-1}}$, using Newton's method, will be larger than ϕ_k^{-1}. Indeed,

$$\tilde{\phi}_k^{-1} = \phi^{-1} + \frac{\phi^{-k-2}}{\sqrt{5} - (k+2)\phi^{-k-1}}$$

$$
\begin{aligned}
&= \frac{\phi^{-1}(\sqrt{5}-(k+2)\phi^{-k-1})+\phi^{-k-2}}{\sqrt{5}-(k+2)\phi^{-k-1}} \\
&= \phi^{-1}\left(\frac{\sqrt{5}-(k+1)\phi^{-k-1}}{\sqrt{5}-(k+2)\phi^{-k-1}}\right) \\
&= \phi^{-1}\left(\frac{\phi^{k+1}\sqrt{5}-(k+1)}{\phi^{k+1}\sqrt{5}-(k+2)}\right).
\end{aligned}
$$

Since $\phi_k^{-1} < \widetilde{\phi}_k^{-1}$, we have

$$
\phi_k > \phi\left(\frac{\phi^{k+1}\sqrt{5}-(k+2)}{\phi^{k+1}\sqrt{5}-(k+1)}\right).
$$

□

Since G_n is bounded below by the sum of the first few terms of a linear recursion using Proposition 5.1 and in turn those terms are approximately exponential, we can approximate G_n by summing a geometric series. In order to provide the details, a lemma is in order.

Lemma 5.4. *Let $r = \log_\phi m$. For all $m > 2$, $\phi_r^m \geq C\phi^m$.*

Proof. If $m \geq 50$, then $r \geq 8$, we can use Proposition 5.3 to see that

$$
\begin{aligned}
\ln\left(\frac{\phi_r}{\phi}\right)^m &\geq m\ln\left(\frac{\sqrt{5}m\phi-r-2}{\sqrt{5}m\phi-r-1}\right) \\
&= m\ln\left(1-\frac{1}{\sqrt{5}m\phi-r-1}\right) \\
&\geq -2m\left(\frac{1}{\sqrt{5}m\phi-r-1}\right) \\
&\geq -2m\left(\frac{1}{\sqrt{5}m\phi-m}\right) \\
&= \left(\frac{-2}{\sqrt{5}\phi-1}\right).
\end{aligned}
$$

In the third step, we observed that for all $-\frac{1}{2} < x < 0$, $\ln(1+x) > 2x$. Exponentiating, we have

$$
\left(\frac{\phi_r}{\phi}\right)^m \geq e^{\left(\frac{-2}{\sqrt{5}\phi-1}\right)}.
$$

If $m < 50$ then we can bound $\phi_r^m \geq \left(\frac{\phi_1}{\phi}\right)^{50}\phi^m$, and the result is obtained with $C = \min\left\{e^{\left(\frac{-2}{\sqrt{5}\phi-1}\right)}, \left(\frac{\phi_1}{\phi}\right)^{50}\right\}$. □

We are now ready to prove Theorem 3.1.

Proof. Let $r = \lfloor \log_\phi n \rfloor$. Using Proposition 5.1 with $k = r$, we have $G_n \geq \sum_{m=1}^{n-r} A_{m,r}$. Since $A_{m,r}$ is a linear recursion with limiting ratio ϕ_r, $A_{m,r} \geq K_1 \phi_r^m$ for some constant K_1 which is independent of m. Therefore,

$$\begin{aligned} G_n \geq & \sum_{m=1}^{n-r} A_{m,r} \\ \geq & \sum_{m=1}^{n-r} K_1 \phi_r^m \\ = & K_2 \phi_r^{n-r} \\ \geq & K_2 \phi_{\log_\phi n}^{n-r} \\ \geq & K_2 \phi_{\log_\phi n}^{n-\log_\phi n-1} \\ \geq & K_2 C \phi^{n-\log_\phi n-1} \\ = & \frac{C_1 \phi^n}{n}, \end{aligned}$$

where in the 6th step we used Lemma 5.4. ∎

6 Upper Bound for G_n

6.1 The Main Idea

Let's refer to Table 4 to try to obtain an upper bound for G_n. Again the values in red are summed to get G_{11}. The important thing to notice for an upper bound is that roughly the first half (starting in the upper right corner) of the numbers to be summed are the Fibonacci sequence. So if we go to the column which starts with the same number of terms from the Fibonacci sequence and sum down, we will have an upper bound. How far down should we sum? Well G_{11} is the sum of 10 terms so using the first 10 terms in that column (underlined numbers in Table 4) will definitely give an upper bound.

However, we can find a better upper bound. Towards the end of the column with underlined numbers, the underlined numbers (approximation of G_{11}) are much larger than the numbers in red (G_{11}). How can we get around using those extraordinarily large numbers? The largest summand used to determine G_{11} is 11. So instead of adding 20, 31, and 50, replace each of those with 11 (notice also that leaving the 20 alone and replacing 31 and 50 with 7 and 7 would give another upper bound). In fact, we need not be concerned about which column we choose. We are finding an upper bound so that we can approximate the first segment of summands with Fibonacci numbers. So for a general G_n, the strategy is to locate the largest summand, $g_{\max}$, or at least find a summand so that all the summands to the left and below are smaller. Starting from the upper right, replace all summands in G_n up to $g_{\max}$ with the Fibonacci sequence and replace the remaining summands with $g_{\max}$.

$m, k \rightarrow$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	$\underline{1}$	1	1	1	1
2	0	1	1	1	1	$\underline{1}$	1	1	1	1
3	1	1	2	2	2	$\underline{2}$	2	2	2	2
4	0	2	2	3	3	$\underline{3}$	3	3	3	3
5	1	2	4	4	5	$\underline{5}$	5	5	5	5
6	0	3	5	7	7	$\underline{8}$	8	8	8	8
7	1	4	8	10	12	$\underline{12}$	13	13	13	13
8	0	5	11	16	18	$\underline{20}$	20	21	21	21
9	1	7	17	24	29	$\underline{31}$	33	33	34	34
10	0	9	24	37	45	$\underline{50}$	52	54	54	55

Table 4: $A_{m,k}$ for $1 \leq m, k \leq 10$, underlined numbers are an upper bound for G_{11}

But how do we find $g_{\max}$ for an arbitrary n? We need to determine which column contains the largest summand for G_n. In order to do this we must first find an upper bound for ϕ_k.

6.2 Upper Bound for ϕ_k

Proposition 6.1. *For all k,*

$$\phi_k^{-1} \geq \frac{1 + k(k+2)\phi^{-k-1} - \sqrt{5 - k(k+2)\phi^{-2k-2} + 2\phi^{-k}(k(\phi-3)-2)}}{-2 + (k+1)(k+2)\phi^{-k}}. \tag{6.1}$$

Proof. As in Proposition 5.3, we will come up with this estimate using a degree 2 Taylor polynomial approximation, $p_k(x)$ to $f_k(x)$ centered at ϕ^{-1}. The concavity of the approximation is $p_k''(x) = f_k''(\phi^{-1})$, a constant. The concavity of $f_k(x)$ is given by $f_k''(x) = 2-(k+1)(k+2)x^k$. If $x > \phi^{-1}$, then the concavity of f_k is less than the concavity of p_k. Therefore the slopes of p_k are increasing at a faster rate. Now $f_k(\phi^{-1}) = p_k(\phi^{-1}) < 0$, and p_k increases faster than f_k, so the zero of p_k, $\widetilde{\phi}_k^{-1}$, will be less than the zero of f_k, ϕ_k^{-1}. The second degree Taylor polynomial centered at ϕ^{-1} is given by

$$\begin{aligned} p_k(x) &= \frac{f_k''(\phi^{-1})}{2}x^2 + x\left(f_k'(\phi^{-1}) - f_k''(\phi^{-1})\phi^{-1}\right) + \left(f_k(\phi^{-1}) - f_k'(\phi^{-1})\phi^{-1} + \frac{f_k''(\phi^{-1})}{2}\phi^{-2}\right) \\ &= ax^2 + bx + c, \end{aligned}$$

where

$$\begin{aligned} a &= 1 - \frac{(k+1)(k+2)}{2}\phi^{-k}, \\ b &= 1 + k(k+2)\phi^{-k-1}, \\ c &= -1 - \frac{k(k+1)}{2}\phi^{-k-2}. \end{aligned}$$

The approximation $p_k(x)$ will have two roots the positive one is the right hand side of (6.1). To find the zeros of p_k, we use the quadratic formula and much omitted algebra. □

k	min n such that $A_{n-k,k} < A_{n-k-1,k+1}$
1	2
2	7
3	13
4	24
5	44
6	75
7	125
8	207
9	339
10	551
11	901
12	1462

Table 5: Locating $g_{\max}$

The upper bound for ϕ_k (as well as the lower bound) appear to be artificially complicated. They are only bounds, after all. Why not find simpler looking bounds? For example, why not use

$$\phi_k^{-1} \geq \frac{1 + k(k+2)\phi^{-k-1} - \sqrt{5}}{-2 + (k+1)(k+2)\phi^{-k}} \tag{6.2}$$

to obtain an upper bound for ϕ_k? We will see later that every term in the bounds from Propositions 5.3 and 6.1 is necessary for later estimates to work out.

6.3 Determing How to Split the Sum

It turns out that when finding an upper bound for G_n, locating the largest summand, $g_{\max}$ is not crucial. We will estimate G_n by splitting the sum into two pieces. We can split the sum anywhere as long as the summand before (left and below) the split is larger than any other summand before the split. In other words, we need to find a number k_s such that, $A_{n-k_s,k_s} > A_{n-k,k}$ for all $k < k_s$. In Table 4 we see that k_s for approximating G_{11} could be 1,2, or 3. Nonetheless, one can still find $g_{\max}$ for some n. Using a computer algebra system, we obtain Table 5. We can use Table 5 for $n \leq 1462$ to determine which summand of G_n is the largest. For example let $n = 200$. We can see from Table 5 that $A_{n-k,k} < A_{n-k-1,k+1}$ for all $k \leq 7$, but $A_{n-8,8} \not< A_{n-9,9}$. So the eighth summand of G_{200} is the largest one. Looking at Table 5 we observe that the second column seems to grow exponentially. This is verified in the following proposition.

Proposition 6.2. *There exists a $K \in \mathbb{N}$ such that for all $k \geq K$, if $n \geq \phi^{k+2} 4\sqrt{5}\ln\phi$, then $\phi_k^n \leq \phi_{k+1}^{n-1}$.*

Proof. For any given k, $\phi_k^n = \phi_{k+1}^{n-1}$ if $n = \frac{\ln \phi_{k+1}}{\ln \frac{\phi_{k+1}}{\phi_k}}$. Since $\phi_{k+1} > \phi_k$, if $n \geq \frac{\ln \phi_{k+1}}{\ln \frac{\phi_{k+1}}{\phi_k}}$, then $\phi_k^n \leq \phi_{k+1}^{n-1}$. So we only need find an upper bound for $\frac{\ln \phi_{k+1}}{\ln \frac{\phi_{k+1}}{\phi_k}}$. Let us start by finding a lower bound for $\ln \frac{\phi_{k+1}}{\phi_k}$. Indeed,

$$\begin{aligned}
\ln \frac{\phi_{k+1}}{\phi_k} &= \ln\left(1 + \frac{\phi_{k+1}-\phi_k}{\phi_k}\right) \\
&\geq \frac{1}{2}\left(\frac{\phi_{k+1}-\phi_k}{\phi_k}\right) \\
&\geq \frac{1}{2\phi}\left(\phi_{k+1}-\phi_k\right).
\end{aligned}$$

Using Propositions 6.1 and 5.3, and setting

$$A_k = \sqrt{5 - k(k+2)\phi^{-2k-2} + 2\phi^{-k}(k(\phi-3)-2)},$$

we have

$$\begin{aligned}
\phi_{k+1}-\phi_k &\geq \phi\left(\frac{\sqrt{5}\phi^{k+2}-(k+3)}{\sqrt{5}\phi^{k+2}-(k+2)}\right) - \frac{-2+(k+1)(k+2)\phi^{-k}}{1+k(k+2)\phi^{-k-1}-A_k} \\
&= \phi\left(1 - \frac{1}{\sqrt{5}\phi^{k+2}-(k+2)}\right) + \frac{2\phi^k-(k+1)(k+2)}{\phi^k+k(k+2)\phi^{-1}-A_k\phi^k} \\
&= -\frac{1}{\sqrt{5}\phi^{k+2}-(k+2)} + \frac{2\phi^k-(k+1)(k+2)+\phi^{k+1}+k(k+2)-A_k\phi^{k+1}}{\phi^k+k(k+2)\phi^{-1}-A_k\phi^k} \\
&= -\frac{1}{\sqrt{5}\phi^{k+2}-(k+2)} + \frac{\phi^{k+1}(\sqrt{5}-A_k)-(k+2)}{\phi^k+k(k+2)\phi^{-1}-A_k\phi^k} \\
&= \frac{-\frac{\phi^k+k(k+2)\phi^{-1}-A_k\phi^k}{\sqrt{5}\phi^{k+2}-(k+2)} + \phi^{k+1}(\sqrt{5}-A_k)-(k+2)}{\phi^k+k(k+2)\phi^{-1}-A_k\phi^k} \\
&:= \frac{-B_k+C_k}{\phi^k+k(k+2)\phi^{-1}-A_k\phi^k} \\
&= \frac{B_k-C_k}{-\phi^k-k(k+2)\phi^{-1}+A_k\phi^k}.
\end{aligned}$$

Let's investigate B_k and C_k closer by computing the limit of each as $k \to \infty$. We see

$$\begin{aligned}
\lim_{k\to\infty} B_k &= \lim_{k\to\infty} \frac{-A_k\phi^k+\phi^k+k(k+2)\phi^{-1}}{\sqrt{5}\phi^{k+2}-(k+2)} \\
&= \frac{1-\sqrt{5}}{\sqrt{5}\phi^2}.
\end{aligned}$$

In addition,

$$\begin{aligned}
\lim_{k\to\infty} C_k &= \lim_{k\to\infty}\left(\phi^{k+1}(\sqrt{5}-A_k)-(k+2)\right) \\
&= \lim_{k\to\infty} \frac{(\sqrt{5}\phi^{k+1}-(k+2))^2-(\phi^{k+1}A_k)^2}{\phi^{k+1}A_k+\sqrt{5}\phi^{k+1}-(k+2)}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{k\to\infty} \left(\frac{5\phi^{2k+2} - 2\phi^{k+1}\sqrt{5}(k+2) + (k+2)^2}{\phi^{k+1}A_k + \sqrt{5}\phi^{k+1} - (k+2)} \right. \\
&\qquad \left. - \frac{\phi^{2k+2}(5 - k(k+2)\phi^{-2k-2} + 2\phi^{-k}(k\phi - 3k - 2))}{\phi^{k+1}A_k + \sqrt{5}\phi^{k+1} - (k+2)} \right) \\
&= \lim_{k\to\infty} \frac{\phi^{k+1}(2 - 2\sqrt{5}) + (k+2)(2k+2)}{\phi^{k+1}A_k + \sqrt{5}\phi^{k+1} - (k+2)} \\
&= \frac{1-\sqrt{5}}{\sqrt{5}}
\end{aligned}
$$

So $\lim_{k\to\infty}(B_k - C_k) = \frac{\sqrt{5}-1}{\phi\sqrt{5}}$. Therefore there exists a $K \in \mathbb{N}$ such that for all $k > K$, $B_k - C_k \geq \frac{1}{2}\frac{\sqrt{5}-1}{\phi\sqrt{5}}$.

Finally we return to estimating $\phi_{k+1} - \phi_k$ as follows:

$$
\begin{aligned}
\phi_{k+1} - \phi_k &\geq \frac{\frac{\sqrt{5}-1}{2\phi\sqrt{5}}}{-\phi^k - k(k+2)\phi^{-1} + A\phi^k} \\
&\geq \frac{\frac{\sqrt{5}-1}{2\phi\sqrt{5}}}{(\sqrt{5}-1)\phi^k} \\
&= \frac{1}{2\sqrt{5}\phi^{k+1}}.
\end{aligned}
$$

All the pieces fit together to provide:

$$
\begin{aligned}
\frac{\ln \phi_{k+1}}{\ln \frac{\phi_{k+1}}{\phi_k}} &\leq \frac{\ln \phi_{k+1}}{\frac{1}{2\phi}(\phi_{k+1} - \phi_k)} \\
&\leq \frac{\ln \phi_{k+1}}{\frac{1}{2\phi} \cdot \frac{1}{2\sqrt{5}\phi^{k+1}}} \\
&\leq \phi^{k+2} 4\sqrt{5} \ln \phi
\end{aligned}
$$

provided $k > K$. □

Note that if we had used (6.2) to obtain our upper bound on ϕ_k, that would make $A_k = \sqrt{5}$ for all k and $\lim_{k\to\infty} C_k$ would not exist.

6.4 Proof of Theorem 3.2

The previous proposition would be perfect for our purposes if it were not for the fact that the estimate only works beyond an unspecified term K. In fact, we know the lower bound used in the proof is only valid for $k > 8$. Computer algebra systems can be used to verify that $K = 12$. However, Table 5 can be used for small k. Proposition 6.2 not only verifies that the growth of the terms in the second column of Table 5 are exponential but it also allows us to locate a splitting point, k_s, for the sum determining G_n even for large n.

We are now ready to prove Theorem 3.2

Proof. Let $k_s = \lfloor \log_\phi \frac{n}{8\phi^2\sqrt{5}\ln\phi} \rfloor$. For all $12 \le k \le k_s$, we have

$$\begin{aligned}
\phi^{k+2} 4\sqrt{5}\ln\phi \le & \phi^{k_s+2} 4\sqrt{5}\ln\phi \\
\le & \frac{n}{2} \\
\le & n - k_s \\
\le & n - k.
\end{aligned}$$

Therefore by Proposition 6.2, we have for all $12 \le k \le k_s$,

$$\phi_k^{n-k} \le \phi_{k+1}^{n-k-1}. \tag{6.3}$$

Also by Table 5 we have for all $k < \min(12, k_s)$,

$$\phi_k^{n-k} \le \phi_{k+1}^{n-k-1}. \tag{6.4}$$

Putting (6.3) and (6.4) together we have for all $k \le k_s$,

$$\phi_k^{n-k} \le \phi_{k+1}^{n-k-1}. \tag{6.5}$$

Now we split the sum determing G_n and apply (6.5) to obtain

$$\begin{aligned}
G_n &= \sum_{k=1}^{k_s} A_{n-k,k} + \sum_{k=k_s+1}^{n-1} A_{n-k,k} \\
&\le \sum_{k=1}^{k_s} A_{n-k,k} + \sum_{k=1}^{n-k_s-1} f_k \\
&\le \sum_{k=1}^{k_s} A_{n-k,k} + f_{n-k_s+1} \\
&\le K_1 \sum_{k=1}^{k_s} \phi_k^{n-k} + f_{n-k_s+1} \\
&\le K_1 k_s \phi^{n-k_s} + K_1 \phi^{n-k_s+1} \\
&= K_1 \phi^{n-k_s}(k_s + \phi) \\
&\le K_1 \phi^{n-\log_\phi \frac{n}{8\phi^2\sqrt{5}\ln\phi}} \left(\log_\phi \frac{n}{8\phi^2\sqrt{5}\ln\phi} + 1 + \phi \right) \\
&\le K_1 8\phi^2\sqrt{5}\frac{\phi^n}{n}(\ln n),
\end{aligned}$$

where $K_1 = \max_{k \le k_s}\{b_k | A_{m,k} < b_k \phi_k^m\}$. □

7 Conclusion

Putting Theorem 3.1 and Theorem 3.2 together we have

$$\frac{\phi^n}{n} \lesssim G_n \lesssim \frac{\ln n \phi^n}{n}.$$

Since the total number of sequences in a puzzle with n columns is $F_{n+2} \approx \phi^{n+2}$, the ratio of forceless sequences to total sequences satisfies

$$\frac{1}{\phi^2 n} \lesssim \frac{G_n}{F_{n+2}} \lesssim \frac{\ln n}{\phi^2 n}.$$

So we see in the limit, as the number of rows goes to infinity, the ratio of forceless sequences to total sequences goes to zero. We can conclude that for very large puzzles there are very few forceless sequences.

This paper dealt only with an individual row sequence isolated from the rest of the puzzle. Many questions about how the row and column sequences interact are left to be answered. For example, given an arbitrary list of row sequences, which column sequences are permitted to create a puzzle with a solution? Certainly, one must obtain the same value when adding all the terms from the row sequences as one does when adding all the terms from the column sequences, but there are many other more subtle connections. One could also extend the puzzle to higher dimensions and more colors. There is much more to be learned about these puzzles. The title of the paper itself begs more research to determine just when there is a unique solution to a Paint by Numbers puzzle.

References

[1] K. J. Batenburg and W. A. Kosters, Solving Nonograms by combining relaxations, *Pattern Recognition* **42** (2009), 1672–1683.

[2] Jessica Benton, Rion Snow, and Nolan Wallach, A combinatorial problem associated with nonograms, *Linear Algebra Appl.* **412** (2006), 30–38.

[3] Ruel V. Churchill and James Ward Brown, *Complex Variables and Applications*, McGraw-Hill Book Co., 1984.

[4] Nobuhisa Ueda and Tadaaki Nagao, NP-completeness results for Nonogram via Parsimonious Reductions, Technical Report TR96-0008, Department of Computer Science, Tokyo Institute of Technology, 1996.

2000 *Mathematics Subject Classification*: Primary 05A20.
Keywords: Nonograms.

(Concerned with sequences A000035, A000045, A000931, A013979, A013982, A013983, A013984, A013985, A013986 and A013987.)

Received July 9 2009; revised version received August 31 2009. Published in *Journal of Integer Sequences*, September 1 2009.

Return to Journal of Integer Sequences home page.