COUNTING CLUES IN CROSSWORDS

Kevin K. Ferland

Bloomsburg University, Bloomsburg, PA 17815

kferland@bloomu.edu

Abstract

We consider different ways to count the number of clues in American-style crossword puzzle grids. One yields a basic parity result for symmetric square grids. Another works efficiently even for non-symmetric grids that are already numbered. We further discuss the upper limit on the number of clues in a crossword puzzle with no 2-letter answers, and open questions are given. As a bonus, a mathematically-themed crossword puzzle is included!

1 Introduction

Crossword puzzles are judged, in part, by the number of clues they contain. This is the case since those with fewer clues tend to be harder to construct and to solve, and those with lots of clues have more short answers that are easier to fill in. We consider ways to count the number of clues in a crossword puzzle grid and further combinatorial results.

Figure 1 displays an example of the type of crossword puzzle that appears as a 15×15 New York Times crossword puzzle.

It serves as both a mathematically-themed puzzle for your solving enjoyment and an example on which we shall practice our counting techniques. The grid itself is a square array each of whose elemental squares is colored black or white. A *clue* is then assigned to each maximal linear segment of white squares, which is to contain the clue's answer. Moreover, for each such maximal linear segment, its top left-most white square is assigned a number to serve as the label for that clue. Clues with horizontal answers are then listed by their numbers as the Across clues, and clues with vertical answers are similarly listed as the Down clues.

Specifically, we consider American-style crossword puzzle grids, as dictated by the Basic Rules listed under Resources at www.cruciverb.com. Such a grid must be square and obey the following structure rules. *Connectivity*: Between any two white squares there is a path of white squares, each step of which is horizontal or vertical. *Symmetry*: The grid looks the same when rotated 180 degrees. And, *Three+*: Each clue's answer must be at least 3 characters long.

ACROSS

- 1 25% of the data
- 9 Toe starter
- 15 Type of statistical estimator
- 16 Mathematically inspired artist
- 17 First impression of Superman
- 18 Sigourney Weaver sequel
- 19 Baroque stringed instruments played upright
- 20 Charles' ex
- 22 Upstate NY engineering college
- 23 Additions to buildings
- 24 Tiger's ex
- 25 Seventh letter in Athens
- 26 RI college named after theologian
- 27 Product of fire
- 30 Like a productive fire
- 31 ___ Lingus
- 32 Solves
- 34 ___ Mobile
- 35 Type of value or vector for a matrix
- 36 Sketch comedy show based on a magazine
- 39 101 to a computer
- 41 Type of court
- 44 Number of birth mothers Diego has
- 45 1040 producer
- 48 Collection of elements
- 49 Dorm VIPs
- 50 Only male since JDB with a tennis grand slam
- 51 See 2-Down
- 53 Informal negative of 37-Down
- 54 NSF program for college students
- 55 See 2-Down
- 57 Axiom
- 58 Tense
- 60 Like the twos, sometimes
- 62 Coach to Alcindor and Walton
- 63 (Re)turn a Jewish creature to mud
- 64 Most protected
- 65 Judge again

DOWN

- 1 Shakes
- 2 See you later, with 51-Across and 55-Across
- 3 Start of the grid completer
- 4 An Iranian toman is ten of these
- 5 Diet colas
- 6 "It ___", said Jesus in John 6:20
- 7 Reflux problem above the esophagus
- 8 Money or Izzard
- 9 One who is not nice
- 10 Java or Capri
- 11 Area of math
- 12 End of the grid completer
- 13 Like a softened consonant
- 14 Clooney, Edwards, and Margulies
- 21 Cat eater from Melmac
- 27 Basic course in math
- 28 Deduce
- 29 QVC rival
- 33 No-win situation
- 34 Using your DRF to play the horses from afar
- 36 Arresters of those that are AWOL
- 37 Exist
- 38 A vector product
- 39 Signs of anger or confusion
- 40 Location of the Cathedral of San Lorenzo
- 42 Snarls
- 43 Values
- 45 Snapchat posts
- 46 Louis XIV had one, at first
- 47 South China or East China
- 52 Chemist's tipping point
- 53 A Seville parting
- 56 Real housewife Leakes
- 57 Var. list entered into a function
- 59 JFK predecessor
- 61 Nutritional stat. to be met

Figure 1: Crossword Puzzle with 74 Clues. Solution appears on 7.

2 Fundamental Grid Counting

Let Row be the total number of Across clues, Col the number of Down clues, and Clues the total number of clues. Starting with 1, we number the rows of a grid from top to bottom and the columns from left to right. Thus, position (i, j) in the grid is the square in row i and column j . We denote the number of clues in row i by r_i and the number in column i by c_i . By definition,

$$\begin{aligned} \text{Row} &= \sum_{i=1}^n r_i = r_1 + r_2 + \cdots + r_n \\ \text{Col} &= \sum_{i=1}^n c_i = c_1 + c_2 + \cdots + c_n \text{ and} \\ \text{Clues} &= \text{Row} + \text{Col} \end{aligned} \tag{2.1}$$

Thus, (2.1) is the exhaustive way to count the clues in any grid, even absent the structure rules. However, when additional structure is present, we can do better.

When n is odd, row $\frac{n+1}{2}$ is the center row, and column $\frac{n+1}{2}$ is the center column. When n is even, neither a center row nor a center column are present, but center lines serve as references for us. For any grid that satisfies the *Symmetry* rule, we can focus our counting on rows at or above the center and columns at or left of the center, and we have a faster way to count Clues.

$$\begin{aligned} \text{Row} &= \begin{cases} 2 \sum_{i=1}^{n/2} r_i & \text{if } n \text{ is even} \\ r_{(n+1)/2} + 2 \sum_{i=1}^{(n-1)/2} r_i & \text{if } n \text{ is odd} \end{cases} \\ \text{Col} &= \begin{cases} 2 \sum_{i=1}^{n/2} c_i & \text{if } n \text{ is even} \\ c_{(n+1)/2} + 2 \sum_{i=1}^{(n-1)/2} c_i & \text{if } n \text{ is odd} \end{cases} \end{aligned} \tag{2.2}$$

For example, for the grid in Figure 1, we have

$$\begin{aligned} \text{Row} &= 2 + 2 (2 + 2 + 2 + 3 + 3 + 4 + 2) = 38 \\ \text{Col} &= 2 + 2 (2 + 2 + 1 + 3 + 3 + 3 + 3) = 36 \text{ and} \\ \text{Clues} &= 38 + 36 = 74 \end{aligned}$$

Of course, using (2.2) is a more efficient way to count Clues than using (2.1). Moreover, when n is odd, by considering the cases in which the center square $(\frac{n+1}{2}, \frac{n+1}{2})$ is black or white, we get the following consequence of the *Symmetry* rule and (2.2).

Lemma 1 ([2]). *Suppose n is odd and we have an $n \times n$ crossword grid satisfying the Symmetry rule. Then, $r_{(n+1)/2}$, the number of clues in the center row, and $c_{(n+1)/2}$, the number of clues in the center column, have the same parity. That is, they are either both even or both odd.*

Moreover, we have the following consequence of (2.2) and Lemma 1.

Corollary 1 ([2]). *The number of clues in any $n \times n$ crossword grid satisfying the Symmetry rule is always even.*

For general rectangular grids the preceding parity result need not hold, as can be seen by considering an 8×7 crossword grid. If it is all white, then there are 15 clues. If it has black squares only in positions $(4, 4)$ and $(5, 4)$, then there are 18 clues. Thus non-square grids can have either an odd or even number of clues.

3 Numbering the Grid

To simplify our discussion we regard our grid as having an additional column of black squares on the left (in column 0) and an additional row of black squares on the top (in row 0). Let the set H contain each white square that has a black square immediately to its left, and let the set V contain each white square that has a black square immediately on its top. In fact, $\text{Row} = |H|$, and H is the set of white squares assigned a number for an Across clue. Also, $\text{Col} = |V|$, and V is the set of white squares assigned a number for a Down clue. Of course, numbers are assigned to the squares of $H \cup V$ from left to right in rows 1 through n . Note that the highest number used for a clue is thus $|H \cup V|$ and $\text{Clues} = |H| + |V|$. Additionally, in a given published crossword puzzle with all of its clues, we know where we can always read off $|H \cup V|$.

Lemma 2. *The highest clue number $|H \cup V|$ is that of the last Across clue.*

Proof. The last row containing white squares in the grid must contain a clue's answer, and that clue must be an Across clue. The right-most such answer must be the last and highest numbered Across clue. $\square$

The following lemma says that, in an already numbered grid, we can determine the number of clues by adding to the number of the last Across clue the number of white squares that have a black square both immediately to its left and immediately on its top. Recall that row 0 and column 0 are considered black.

Lemma 3. $\text{Clues} = |H \cup V| + |H \cap V|$.

Proof. The squares in $H \cap V$ are assigned the numbers that label both Across and Down clues. Since $|H \cup V| = |H| + |V| - |H \cap V|$, we thus conclude that

$$\text{Clues} = |H| + |V| = |H \cup V| + |H \cap V|$$

$\square$

For example, for the grid in Figure 1, noting the last clue number, and counting the number of white squares with both a black square to the left and on top gives

$$\begin{aligned} |H \cup V| &= 65 \\ |H \cap V| &= |\{1, 9, 27, 34, 36, 39, 45, 53, 57\}| = 9 \text{ and} \\ \text{Clues} &= 65 + 9 = 74 \end{aligned}$$

confirming our earlier count of 74 clues.

4 Maximizing Clues

In the submission criteria for the major publishers of 15×15 American style crossword puzzles, an upper limit on the number of clues is imposed that is 78, generally. That number is merely a matter of taste and undoubtedly arose from experience, as a means to keep puzzles from being too easy or uninteresting with lots of 3-letter answers. Of course, there is a mathematical limit to the number of clues, and our counting of Clues sheds light on this.

For each $n \geq 3$, let $a(n)$ be the maximum possible number of clues in an $n \times n$ American-style crossword puzzle grid. In [2], it is shown that $a(15) = 96$, and this is only achievable by two different grids. That result is also discussed in [1]. In [3], all grid sizes are considered, including the 21×21 grid size for Sunday New York Times crossword puzzles. It is shown there that $a(21) = 198$, and the grid in Figure 2 is the unique grid achieving that. The values of $a(n)$ up to $n = 50$, determined in [3], appear as sequence

Figure 2: Maximal 21×21 crossword grid

A243826 in Sloane's On-Line Encyclopedia of Integer Sequences [4].

For all n , the number of clues in any row or column of an $n \times n$ crossword grid satisfying *Three+* is an integer that can be at most $\frac{n+1}{4}$. We see this, while regarding row 0 and column 0 to be black, because each clue's answer requires at least three white squares and one black square. Since there are n rows and n columns satisfying this, we have

$$a(n) \leq 2n \left\lfloor \frac{n+1}{4} \right\rfloor = \begin{cases} \frac{1}{2}n^2 & \text{if } n \bmod 4 = 0 \\ \frac{1}{2}(n^2 - n) & \text{if } n \bmod 4 = 1 \\ \frac{1}{2}(n^2 - 2n) & \text{if } n \bmod 4 = 2 \\ \frac{1}{2}(n^2 + n) & \text{if } n \bmod 4 = 3 \end{cases} \quad (4.1)$$

Moreover, the upper bound for $a(n)$ in (4.1) holds for any $n \times n$ grid satisfying *Three+* and depends on neither *Connectivity* nor *Symmetry*.

In [3], for each $n \geq 12$, an $n \times n$ American-style crossword puzzle grid, satisfying each of the structure rules, is constructed with $g(n)$ clues, where

$$g(n) = \begin{cases} \frac{1}{2}(n^2 - 2n + 8) & \text{if } n \bmod 4 = 0 \\ \frac{1}{4}(2n^2 - 5n + 10 + (n \bmod 8)) & \text{if } n \bmod 4 = 1 \\ \frac{1}{2}(n^2 - 2n) & \text{if } n \bmod 4 = 2 \\ \frac{1}{4}(2n^2 - 5n + 8 - (n \bmod 8)) & \text{if } n \bmod 4 = 3 \end{cases}$$

That establishes the following lower bound for $a(n)$.

$$a(n) \geq g(n) \quad (4.2)$$

How good are these upper and lower bounds on $a(n)$? In [2], it is shown that $g(15) = 94 < 96 = a(15)$. Hence, the lower bound in (4.2) is not always tight. However, in the case that $n \bmod 4 = 2$, we see that $g(n)$ agrees with the upper bound for $a(n)$ in (4.1). That proves half of the following result. The other half (when n is divisible by 4) is proven in [3] by using a careful accounting of the black squares in the grid.

Theorem 1 ([3]). *For each even $n \geq 8$, $a(n) = g(n)$. Moreover, $g(n)$ is the maximum number of clues in any $n \times n$ grid satisfying *Three+*.*

The following conjecture is seen by computer in [2] to hold for $n \leq 49$.

Conjecture 2 ([3]). *For each odd $n \geq 17$, $a(n) = g(n)$.*

5 Possible Extensions

Suppose we change *Three+* to *Four+* and no longer allow answers with just 3 characters.

Question 3. *What is the maximum number of clues possible in an American-style crossword puzzle grid satisfying Four+?*

Equation (2.2) and Lemma 3 clearly holds for rectangular grids as well.

Question 4. *What is the maximum possible number of clues in an $m \times n$ crossword grid satisfying the structure rules?*

In [2], the two 15×15 grids that have the maximum possible number of clues have different numbers of black squares.

Question 5. *In an American-style crossword puzzle grid with the maximum possible number of clues, what can be said about the number of black squares?*

Conversely, as a much larger problem, given a number of black squares what is the maximum possible number of clues in a grid?

References

- [1] “Cross Purposes”. In: *Futility Closet* (2015), published electronically at <https://www.futilitycloset.com/2015/02/09/cross-purposes-3>.
- [2] K. K. Ferland. “Record Crossword Puzzles”. In: *The American Mathematical Monthly* 121.6 (2014), pp. 534–536.
- [3] K. K. Ferland and R. W. Pratt. “Maximal Crossword Grids”. In: *Journal of Combinatorial Mathematics and Combinatorial Computing* (To Appear).
- [4] *The On-Line Encyclopedia of Integer Sequences*. published electronically at <https://oeis.org/A243826>. 2014.

Answer

QU	ARTILE		MIS	LE
UN	BIASED		ESCHER	
IT	SABIRD		ALIENS	
VI	OLS	DIANA		RIT
EL	S	ELIN		ETA
RW	ASH	FED		AER
SE	TTL	S	OLD	S
	EIGEN		MADTV	
F	I	VE	PROBATE	
UN	A	IRS	SET	RAS
R	G	MEET		AIN
RE	U	AGAIN		ADAGE
ON	EDGE		TERRIBLE	
WO	DEN		ENDGOLEM	
SA	FEST		REASSESS	