Maximal Crossword Grids

Kevin K. Ferland
kferland@bloomu.edu
Bloomsburg University, Bloomsburg, PA 17815

Robert W. Pratt
Rob.Pratt@sas.com
SAS Institute Inc., Cary, NC 27513

Abstract

The maximum number of clues in an $n \times n$ American-style crossword puzzle grid is explored. Grid constructions provided for all n are proved to be maximal for all even n. By using mixed integer linear programming, they are verified to be maximal for all odd $n \leq 49$. Further, for all $n \leq 30$, all maximal grids are provided.

1 Introduction

A crossword puzzle is constructed on a grid, each of whose squares is colored black or white. A *clue* is then assigned to each maximal linear segment of white squares; that segment is to contain the clue's *answer*. We consider only an American-style crossword puzzle grid, as dictated by the Basic Rules at `www.cruciverb.com`. Such a grid is square and constructed according to the following *structure rules*.

1. *Connectivity*: The centers of any two white squares in the grid can be joined by a path consisting of horizontal and vertical line segments that meet in the center of and pass through only white squares.
2. *Symmetry*: The grid looks the same if rotated 180 degrees.
3. *Three+*: Each clue's answer must be at least 3 characters long.

Given a grid, *cheater squares* are squares that may be switched to black or white, while following the structure rules, without changing the number of clues in the puzzle.

In [3], the 15×15 grid size used in Daily New York Times crossword puzzles is considered. There it is seen that 96 is the maximum number of clues and this is achievable by just two different grids, modulo cheater squares and symmetry. The

JCMCC **108** (2019), pp. 53-64

proof of those results in [3] is somewhat tedious and is presented in a link there. In this paper, we consider American-style crossword puzzle grids of all sizes, seek the maximum number of clues, and consider only grids that do not have cheater squares. For example, we establish here that 198 is the maximum number of clues for the 21×21 grid size used in Sunday New York Times crossword puzzles, and we describe the unique grid $G(21)$ achieving that maximum. See Figure 4.

For each $n \geq 3$, let $a(n)$ be the maximum number of clues in an $n \times n$ American-style crossword puzzle grid. Such a grid with $a(n)$ clues is said to be *maximal.* The values of $a(n)$ up to $n = 50$ appear as sequence A243826 in Sloane's On-Line Encyclopedia of Integer Sequences [1].

2 Grid Constructions

For $3 \leq n \leq 6$, the all-white grid uniquely achieves the maximum number of clues, and $a(n) = 2n$. For $n = 7, 9, 11$, we have verified by computer that Figure 1 displays all grids achieving the maximum number of clues. Specifically, $a(7) = 22$,

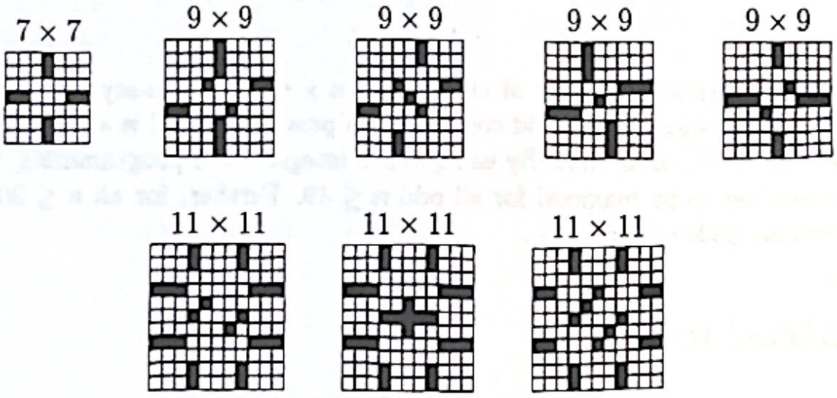

Figure 1: All 7×7, 9×9, and 11×11 Maximal Grids

$a(9) = 32$, and $a(11) = 50$. For each $n \geq 8$ with $n \neq 9, 11$, we now construct what will be shown to be a well-chosen grid $G(n)$.

For n even, $G(n)$ generalizes the examples displayed in Figure 2. The key to

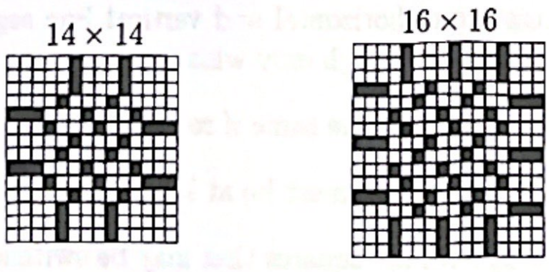

Figure 2: Maximal Grids $G(n)$ for $n \equiv 2 \pmod 4$ and $n \equiv 0 \pmod 4$

the construction is the repeated use of the 4×4 grid displayed in Figure 3 that

Figure 3: The Fundamental Pane

we call a *pane*. When $n \equiv 2 \pmod 4$, the rows and columns within 3 units of the boundary of the grid make up the *frame* of the grid, and the interior is filled by $(n-6)/4$ rows of $(n-6)/4$ panes. When $n \equiv 0 \pmod 4$, the rows and columns within 4 units of the boundary of the grid make up the frame of the grid, and the interior is filled by $(n-8)/4$ rows of $(n-8)/4$ panes.

For n odd, $G(n)$ depends on n mod 8 and generalizes the examples displayed in Figure 4. When $n \equiv 1 \pmod 8$, the top left pane starts at $(4,4)$, there are

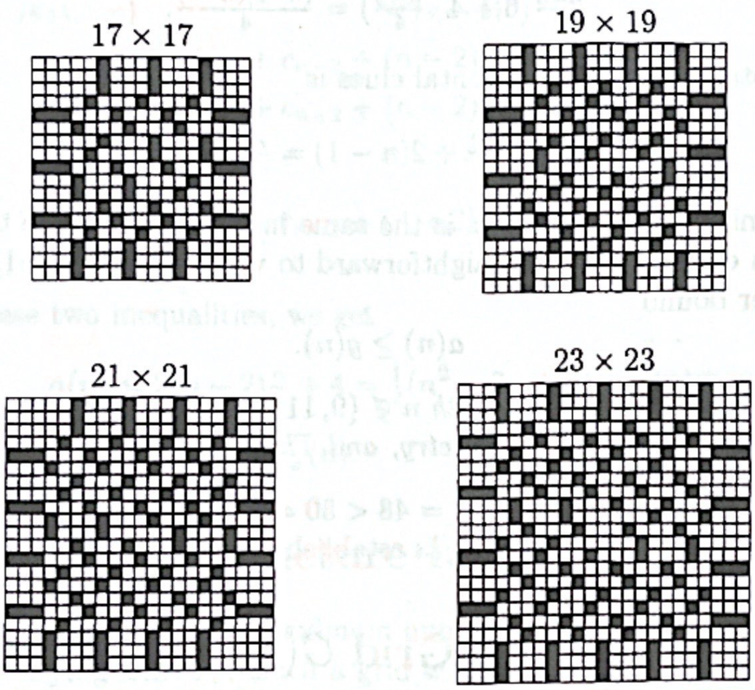

Figure 4: Maximal Grids $G(n)$ for Odd n, Based on n mod 8

$(n-9)/8$ rows of $(n-5)/4$ panes in the top half of the grid, the 180-degree rotation of this places the panes in the bottom half of the grid, and the remainder is considered the frame. Here the frame has a $3 \times n$ strip across the middle of the grid. When $n \equiv 3 \pmod 8$, the top left pane starts at $(5,4)$, and there are $(n-11)/8$ rows of $(n-7)/4$ panes in the top half of the grid. When $n \equiv 5 \pmod 8$, the top left pane starts at $(4,4)$, and there are $(n-13)/8$ rows of $(n-5)/4$ panes in the top half of the grid. When $n \equiv 7 \pmod 8$, the top left pane starts at $(3,3)$, and there are $(n-7)/8$ rows of $(n-7)/4$ panes in the top half of the grid. In each case, the bottom half of the grid is the 180-degree rotation of the top half, and the frame contains a horizontal strip across the middle of the grid.

To count the number of clues in $G(n)$, we define the following function.

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

For example, when $n \equiv 0 \pmod 4$, the number of clues in the first four rows is

$$3 + 4(\tfrac{n-8}{4} + 1) = n - 1,$$

and the number of clues in rows 5 through $n - 4$ is

$$\tfrac{n-8}{4}(6 + 4 \cdot \tfrac{n-8}{4}) = \tfrac{(n-2)(n-8)}{4}.$$

Hence, the total number of horizontal clues is

$$\tfrac{(n-2)(n-8)}{4} + 2(n-1) = \tfrac{n^2-2n+8}{4}.$$

Since the number of vertical clues is the same in this case, we see that $G(n)$ has $g(n)$ clues. In each case, it is straightforward to verify Theorem 2.1, and we thus have the lower bound

$$a(n) \geq g(n). \tag{2.1}$$

Theorem 2.1. *For each $n \geq 8$ with $n \notin \{9, 11, 15\}$, $G(n)$ is an $n \times n$ crossword grid satisfying Connectivity, Symmetry, and Three+ that has $g(n)$ clues.*

Note that $g(9) = a(9)$ and $g(11) = 48 < 50 = a(11)$, as is reflected in Figure 1. Also, $g(15) = 94 < 96 = a(15)$, as was established in [3].

3 Optimality of the Grid $G(n)$

For all n, the number of clues in any row or column is at most $(n + 1)/4$. To see this, think of appending a black square to each row and column, and note that each clue's answer requires at least three white squares and one black square. Thus, in a grid with n rows and n columns, the total number of clues is at most

$$2n\left\lfloor \frac{n+1}{4} \right\rfloor. \tag{3.1}$$

Moreover, this upper bound holds for any $n \times n$ grid satisfying Three+ and depends on neither Connectivity nor Symmetry.

Theorem 3.1. *For each even $n \geq 8$, $a(n) = g(n)$, whence $G(n)$ is maximal. Moreover, $g(n)$ is the maximum number of clues in any $n \times n$ grid satisfying Three+.*

Proof. When $n \equiv 2 \pmod 4$, the upper bound in (3.1) agrees with $g(n)$ and Theorem 2.1 now gives us $a(n) = g(n)$. So assume that $n \equiv 0 \pmod 4$ and we have a grid with $a(n)$ clues.

Let r_i be the number of clues in row i, and let c_j be the number of clues in column j. So we have all $r_i, c_j \leq \frac{n}{4}$. Let b_i be the number of black squares in row i. By Three+, squares $(3,3)$, $(3,n-2)$, $(n-2,3)$, and $(n-2,n-2)$ are white (as are all squares closer to the corners of the grid than these, since cheater squares are not allowed).

Observe that $r_i \leq b_i + 1$ for all i. Hence, $-b_3 \leq 1 - r_3$ and $-b_{n-2} \leq 1 - r_{n-2}$. If square $(3,j)$ is black, then $c_j \leq \frac{n}{4} - 1$. Moreover, if square $(n-2,j)$ is also black, then $c_j \leq \frac{n}{4} - 2$. It follows that

$$\begin{aligned}\sum_{j\in\{1,\ldots,n\}} c_j &= c_3 + c_{n-2} + \sum_{j\notin\{3,n-2\}} c_j \\ &\leq c_3 + c_{n-2} + (n-2)\tfrac{n}{4} - (b_3 + b_{n-2}) \\ &\leq c_3 + c_{n-2} + (n-2)\tfrac{n}{4} + 2 - r_3 - r_{n-2}.\end{aligned}$$

Similarly,

$$\sum_{i\in\{1,\ldots,n\}} r_i \leq r_3 + r_{n-2} + (n-2)\tfrac{n}{4} + 2 - c_3 - c_{n-2}.$$

By adding these two inequalities, we get

$$a(n) \leq 2(n-2)\tfrac{n}{4} + 4 = \tfrac{1}{2}(n^2 - 2n + 8) = g(n).$$

From (2.1), it follows that $a(n) = g(n)$. □

4 Relaxing the Structure Rules

For each $n \geq 3$, let $u(n)$ be the maximum number of clues in an $n \times n$ crossword puzzle grid satisfying Three+. Such a grid with $u(n)$ clues that does not satisfy all of the structure rules is said to be *maximal weak*. Theorem 3.1 tells us that for even $n \geq 8$, $u(n) = a(n) = g(n)$. However, for odd n, equality does not hold, and $u(n)$ serves as an upper bound for $a(n)$ that is not tight.

For $n = 7, 11$, we have verified by computer that Figure 5 displays the only grids achieving $u(n)$ clues. Specifically, $u(7) = 24$, and $u(11) = 54$. For $n \geq 9$ with

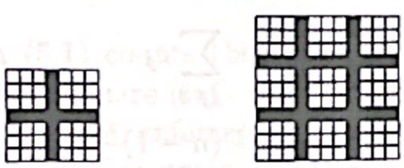

Figure 5: The Unique 7×7 and 11×11 Maximal Weak Grids

$n \neq 11$, we construct grids $H_i(n)$, for $i = 1, 2, 3, 4$. When $n \equiv 1 \pmod 4$, the

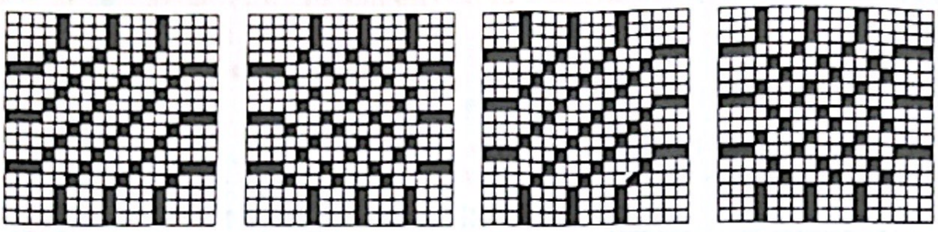

Figure 6: All 17×17 Maximal Weak Grids H_1, H_2, H_3, H_4

grids are shown in Figure 6. Note that H_1 satisfies Symmetry and not Connectivity, and H_2 satisfies Connectivity and not Symmetry. When $n \equiv 3 \pmod 4$, we construct only the grids H_3 and H_4 as shown in Figure 7. For each odd $n \geq 9$ with

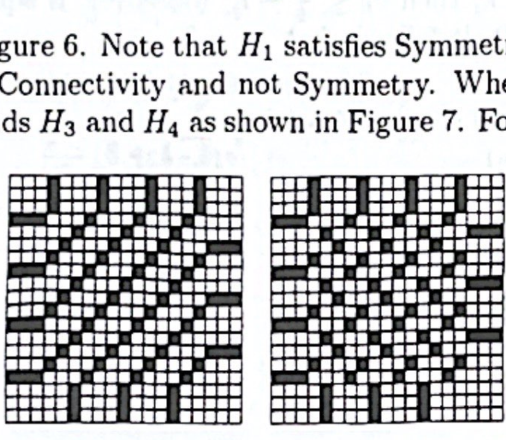

Figure 7: All 19×19 Maximal Weak Grids H_3 and H_4

$n \neq 11$ and each i, it is straightforward to verify that $H_i(n)$ has $\frac{1}{2}(n^2 - 2n + 5)$ clues. Moreover, when $n \equiv 1 \pmod 4$, we have the following result, which gives a decent upper bound for $a(n)$.

Theorem 4.1. *For $n \geq 9$ with $n \equiv 1 \pmod 4$, $u(n) = \frac{1}{2}(n^2 - 2n + 5)$, whence each $H_i(n)$ is maximal weak.*

Proof. Assume that $n \equiv 1 \pmod 4$ and we have a grid with $u(n)$ clues. Let r_i be the number of clues in row i, and let c_j be the number of clues in column j. So we have all $r_i, c_j \leq \frac{n-1}{4}$. Let b_i be the number of black squares in row i. By Three+, square $(3, 3)$ is white, as are all squares closer to the corner of the grid than this.

Since $r_i \leq b_i + 1$ for all i, we have $-b_3 \leq 1 - r_3$. If square $(3, j)$ is black, then $c_j \leq \frac{n-1}{4} - 1$. It follows that

$$\begin{aligned}
\sum_{j \in \{1,\ldots,n\}} c_j &= c_3 + \sum_{j \neq 3} c_j \\
&\leq c_3 + (n-1)\tfrac{n-1}{4} - b_3 \\
&\leq c_3 + (n-1)\tfrac{n-1}{4} + 1 - r_3.
\end{aligned}$$

Similarly,

$$\sum_{i \in \{1,\ldots,n\}} r_i \leq r_3 + (n-1)\tfrac{n-1}{4} + 1 - c_3.$$

Hence, we get

$$u(n) \le 2(n-1)\tfrac{n-1}{4} + 2 = \tfrac{1}{2}(n^2 - 2n + 5). \qquad \square$$

5 Mathematical Optimization

To verify optimality of $G(n)$ for an initial string of odd n, we use the OPTMODEL procedure in SAS/OR [4] and formulate the problem of computing $a(n)$ using mixed integer linear programming (MILP) [5] with three sets of binary decision variables. Let $w_{i,j} = 1$ if square (i, j) is white, and $w_{i,j} = 0$ otherwise. Let $h_{i,j,k} = 1$ if a horizontal answer starts in square (i, j) and has length $k \ge 3$, and $h_{i,j,k} = 0$ otherwise. Let $v_{i,j,k} = 1$ if a vertical answer starts in square (i, j) and has length $k \ge 3$, and $v_{i,j,k} = 0$ otherwise. Let $N = \{1, \ldots, n\}$. Over the set $N \times N \times (N \setminus \{1, 2\})$ we want to maximize

$$\sum_{i,j,k} (h_{i,j,k} + v_{i,j,k}) \tag{5.1}$$

subject to

$$w_{i,j} = \sum_{\substack{j_2,k: \\ j_2 \le j \le j_2+k-1}} h_{i,j_2,k} \qquad i \in N, j \in N \tag{5.2}$$

$$w_{i,j} = \sum_{\substack{i_2,k: \\ i_2 \le i \le i_2+k-1}} v_{i_2,j,k} \qquad i \in N, j \in N \tag{5.3}$$

$$\sum_k h_{i,j,k} \le 1 - w_{i,j-1} \qquad i \in N, j \in N \setminus \{1\} \tag{5.4}$$

$$\sum_k v_{i,j,k} \le 1 - w_{i-1,j} \qquad i \in N \setminus \{1\}, j \in N \tag{5.5}$$

$$w_{i,j} = w_{n-i+1,n-j+1} \qquad i \in N, j \in N \tag{5.6}$$

$$h_{i,j,k} = h_{n-i+1,n-j-k+2,k} \qquad i \in N, j \in N, k \in \{3, \ldots, n-j+1\} \tag{5.7}$$

$$v_{i,j,k} = v_{n-i-k+2,n-j+1,k} \qquad i \in N, j \in N, k \in \{3, \ldots, n-i+1\} \tag{5.8}$$

$$w_{i,j} \in \{0,1\} \qquad i \in N, j \in N \tag{5.9}$$

$$h_{i,j,k} \in \{0,1\} \qquad i \in N, j \in N, k \in \{3, \ldots, n-j+1\} \tag{5.10}$$

$$v_{i,j,k} \in \{0,1\} \qquad i \in N, j \in N, k \in \{3, \ldots, n-i+1\} \tag{5.11}$$

The objective function (5.1) counts the total number of clues (or answers). Constraint (5.2) enforces that square (i, j) is white if and only if some horizontal answer contains it. Constraint (5.3) enforces that square (i, j) is white if and only if some vertical answer contains it. Constraint (5.4) forces square $(i, j-1)$ to be black if some horizontal answer starts in square (i, j). Constraint (5.5) forces square $(i-1, j)$ to be black if some vertical answer starts in square (i, j). Constraints (5.6), (5.7), and (5.8) enforce Symmetry for the white squares, horizontal answers, and vertical answers, respectively. Constraints (5.9), (5.10), and (5.11) enforce integrality for all variables. Three+ is enforced implicitly by the indexing of the $h_{i,j,k}$ and $v_{i,j,k}$ variables; whenever these variables appear, $k \geq 3$.

Connectivity could be violated by a solution to our current formulation, such as the 7×7 grid on the left-hand side of Figure 8, whose clue count is $24 > 22 = a(7)$.

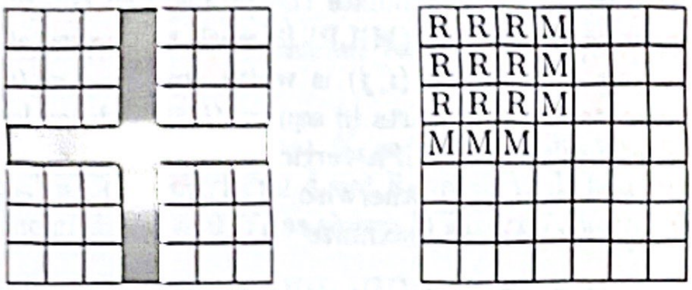

Figure 8: Maximal Disconnected 7×7 Grid, with Regions and Moats

To enforce Connectivity, we could introduce auxiliary binary and flow variables, with additional constraints that send one unit of flow from a white "source" square to each other white square. Since that approach makes the problem artificially large, we use a more sophisticated and faster *cutting-plane* method [2]. The idea is to relax Connectivity and generate new constraints as needed, only when they are violated by a potential solution.

For example, the disconnected solution in Figure 8 has four connected regions, and the squares in one such region are marked with an R in the grid on the right-hand side. The neighbors of these squares in the rest of the grid form a *moat*, $M = \{(1,4),(2,4),(3,4),(4,1),(4,2),(4,3)\}$, each of whose squares is marked with an M on the right-hand side of Figure 8. To prevent this particular connected region from arising in a solution, we introduce the additional constraints[1]

$$w_{i,j} \leq \sum_{(i_2,j_2) \in M} w_{i_2,j_2} \quad \text{for } (i,j) \in R. \tag{5.12}$$

The logical rule expressed here is that if square $(i,j) \in R$ is white, then some square $(i_2,j_2) \in M$ must also be white for a white path to connect (i,j) to its symmetric partner $(n-i+1, n-j+1)$, and R therefore cannot form a connected region in a solution. By adding such a constraint for each[2] connected component R (with moat M) arising in a possible solution, we cut off all disconnected solutions without cutting off any connected solutions.

Since the number of such constraints is large (the number of possible connected components grows exponentially with n), we generate them dynamically, rather

[1] We actually introduce instead the weaker "aggregated" constraint $\sum_{(i,j) \in R} w_{i,j} \leq |R| \sum_{(i_2,j_2) \in M} w_{i_2,j_2}$ because doing so turns out to speed up the MILP solver calls.

[2] We do not generate such a constraint for a connected component that has 180-degree rotational symmetry with respect to the full grid; doing so would cut off connected solutions.

than all at once in the initial formulation. Explicitly, the cutting-plane approach is to repeatedly call our MILP solver, check Connectivity of each resulting potential solution with our network solver, and add any violated constraints (5.12) to the formulation, terminating once a connected solution is obtained. Typically, only a few iterations of this loop are needed, with only a tiny percentage of constraints generated.

Cheater squares are eliminated using a secondary objective to maximize the number of white squares

$$\sum_{i,j} w_{i,j}, \tag{5.13}$$

subject to an *objective cut*

$$\sum_{i,j,k} (h_{i,j,k} + v_{i,j,k}) \geq a(n), \tag{5.14}$$

where $a(n)$ is the maximum number of clues already computed. The steps are:

1. Apply the cutting-plane method to maximize objective function (5.1).
2. Add the objective cut (5.14) to the formulation.
3. For all i, j such that $w_{i,j} = 1$, fix $w_{i,j}$.
4. Maximize the secondary objective function (5.13).
5. Unfix all $w_{i,j}$.

All desired solutions are found as follows. Once one optimal solution is found, we solve a sequence of related problems to find all optimal solutions. For each solution s, let $W_s = \{(i,j) : w_{i,j} = 1\}$, the set of white squares. Any solution that differs from s and has no cheater squares must have at least one white square that is black in solution s. The following constraint enforces this condition and excludes solution s:

$$\sum_{(i,j)\in(N\times N)\setminus W_s} w_{i,j} \geq 1. \tag{5.15}$$

To find all solutions, first solve once and include the objective cut (5.14). Then include constraint (5.15) for each new solution and solve again, until the problem becomes infeasible.

To reduce the solution set modulo dihedral symmetry, we modify the loop by introducing multiple constraints (5.15) for each solution s. By Symmetry, the eight dihedral group elements reduce to just four: identity (= 180-degree rotation), 90-degree rotation (= 270-degree rotation), horizontal reflection (= vertical reflection), and diagonal reflection (= antidiagonal reflection). Each time a solution s is found, we exclude not just s but the (up to four) images of s under multiplication by each of these four group elements.

6 Computer Results

For odd $n \notin \{11, 15\}$ with $9 \leq n \leq 49$, we have verified by computer that $a(n) = g(n)$. Moreover, for $8 \leq n \leq 30$ with $n \neq 9$ and $n \bmod 4 \neq 3$, we have verified that $G(n)$ is the unique maximal grid, modulo symmetry. For $n \in \{19, 23, 27\}$, our program has also yielded all maximal grids besides $G(n)$, as displayed in Figures 9 through 11.

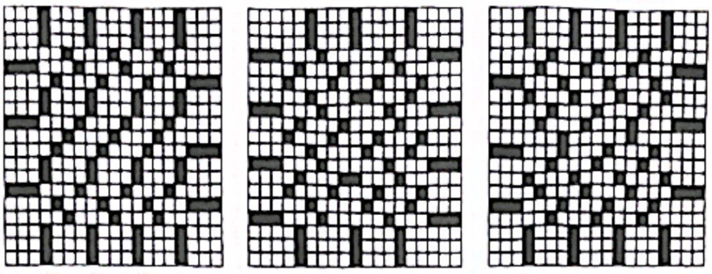

Figure 9: All Maximal 19×19 Grids Besides $G(19)$

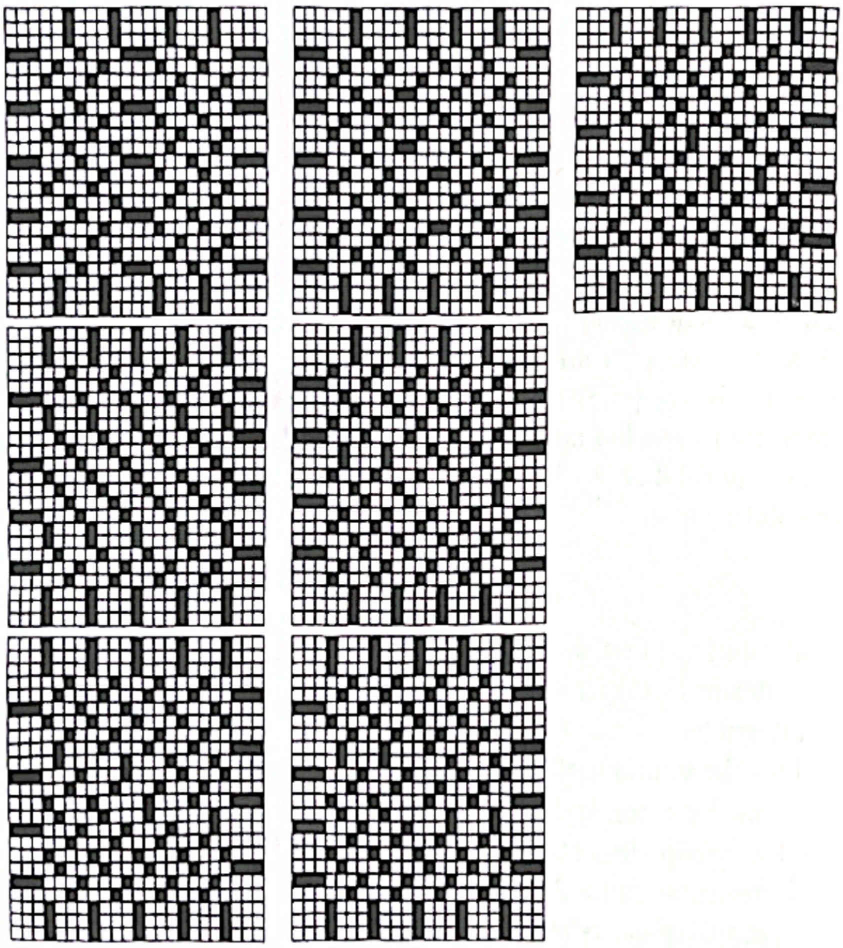

Figure 10: All Maximal 23×23 Grids Besides $G(23)$

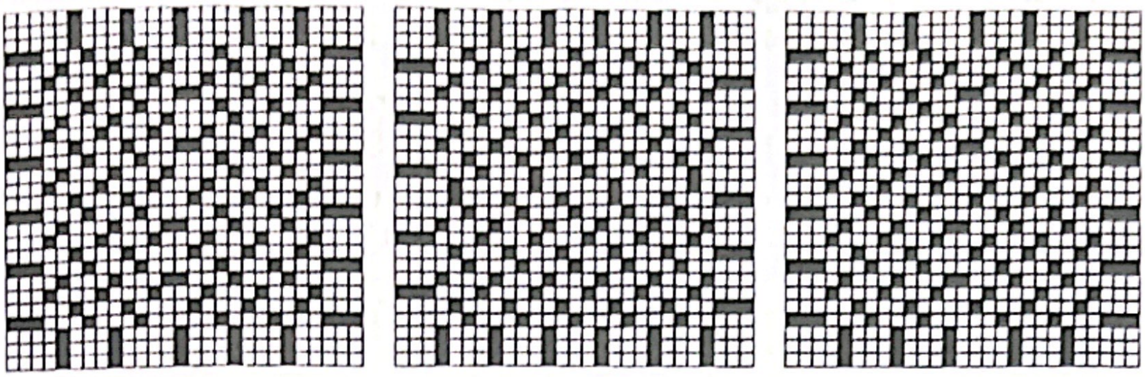

Figure 11: All Maximal 27 × 27 Grids Besides $G(27)$

For odd $n \neq 11$ with $9 \leq n \leq 49$, we have verified by computer that $u(n) = \frac{1}{2}(n^2 - 2n + 5)$. Moreover, for odd $n \leq 31$, we have verified that the grids $H_i(n)$ are the only maximal weak grids (when they are defined), modulo symmetry.

7 Open Questions

Our first conjecture holds for $n \leq 49$.

Conjecture 7.1. *For each odd $n \geq 17$, $a(n) = g(n)$, whence $G(n)$ is maximal.*

Our second conjecture holds for $n \leq 30$.

Conjecture 7.2. *For $n \geq 10$ with $n \bmod 4 \neq 3$, $G(n)$ is the unique maximal grid, modulo symmetry.*

When $n \bmod 4 = 3$, how many different American-style crossword grids have $a(n)$ clues? Besides $G(n)$, what are they?
Our third conjecture holds for $n \leq 49$.

Conjecture 7.3. *For each $n \geq 15$ with $n \equiv 3 \pmod 4$, $u(n) = \frac{1}{2}(n^2 - 2n + 5)$, whence $H_3(n)$ and $H_4(n)$ are maximal weak.*

Our fourth conjecture holds for $n \leq 31$.

Conjecture 7.4. *For odd $n \geq 9$ with $n \bmod 4 \neq 3$, H_1, H_2, H_3, H_4 are the only maximal weak grids, modulo symmetry.*

References

[1] *The On-Line Encyclopedia of Integer Sequences*, published electronically at https://oeis.org/A243826, 2014.

[2] G. B. Dantzig, R. Fulkerson, and S. M. Johnson, *Solution of a large-scale traveling salesman problem*, Operations Research **2** (1954), 393–410.

[3] K. K. Ferland, *Record crossword puzzles*, The American Mathematical Monthly **121** (2014), no. 6, 534–536.

[4] SAS Institute Inc., *SAS/OR 14.2 User's Guide: Mathematical Programming*, SAS Institute Inc., Cary, NC, 2016.

[5] L. A. Wolsey, *Integer programming*, Wiley, New York, 1998.