Latin squares and Sudoku

Peter J. Dukes

By now, virtually everyone has heard of Sudoku puzzles. In the standard puzzle, a 9×9 partially filled array must be completed with digits 1 to 9 so that every row, every column, and each of nine disjoint 3×3 'boxes' contains every digit exactly once. There are many ways to vary the puzzle, including different sizes (e.g. 4×4 mini-Sudoku or the 100×100 'Sudoku-zilla'), different boxes (6×6 with rectangular boxes, jigsaw Sudoku), puzzles with arithmetic constraints (Kakuro), and crossword-Sudoku hybrids using letters.

Here, we are interested in a relaxation which drops the condition on boxes. A *latin square* is an $n \times n$ array of n symbols (we often use the first n positive integers) such that every row and every column contains every symbol exactly once. For example, if our first row is filled $\boxed{1}\boxed{2}\boxed{3}\ldots\boxed{9}$ and we form eight more rows from different cyclic shifts, then the result is a 9×9 latin square. Note that it is not necessarily a filled Sudoku square unless we carefully arrange the shifted rows.

Problem 1 *On $\triangle ABC$ are given distinct points $L_1, \ldots, L_{10}$ on BC, $M_1, \ldots, M_{10}$ on AC, and $N_1, \ldots, N_{10}$ on AB. Show that there are 100 triangles $\triangle L_i M_j N_k$, no two of which share a common edge. Use a latin square.*

Latin squares have a rich history, appearing in art, agriculture, and statistics. Today, they are used in information-theoretic settings such as network routing, hash functions, and pseudo-random number generation. Research on the mathematics behind latin squares began with Euler in the 18th century and is still ongoing. Typical research involves counting or generating them, their existence (or non-existence) with various extra structure, and connections to other topics in combinatorics.

Regarding enumeration, the number $f(n)$ of $n \times n$ latin squares has been exactly determined for n up to 11; these values can be found at `http://oeis.org/A002860`. Asymptotically, it is known that $f(n)^{1/n^2} \sim e^{-2}n$.

As an example of extra structure, a *transversal* in a latin square is a set of cells in distinct rows and columns, and filled with distinct symbols; see Figure 1. Ryser's conjecture asserts that, for n odd, every $n \times n$ latin square has a transversal. This is still open, however.

1	5	3	4	**2**
3	1	2	**5**	4
5	**3**	4	2	1
2	4	**1**	3	5
4	2	5	1	3

Figure 1: A transversal

Problem 2 *Suppose $a_1, a_2, \ldots, a_n$ is a permutation of $\{1, 2, \ldots, n\}$, where n is even. For each i, put $b_i \equiv a_i + i \pmod{n}$ so that $1 \le b_i \le n$. Prove that $b_1, b_2, \ldots, b_n$ is not a permutation. Conclude that there is no transversal in a latin square of even size if its rows are cyclic shifts of each other.*

Two $n \times n$ latin squares are said to be *orthogonal* if, when superimposed, each of the n^2 possible pairings of symbols appears exactly once. A cute example for $n = 4$ is shown in Figure 2.

A♠	J♡	Q♣	K♢
J♣	A♢	K♠	Q♡
Q♢	K♣	A♡	J♠
K♡	Q♠	J♢	A♣

Figure 2: Orthogonal latin squares

It is easy to see that orthogonal latin squares each possess n disjoint transversals. Indeed, the transversals of one square are determined by the cells on which its orthogonal mate is constant. Orthogonal latin squares (more generally, sets of mutually orthogonal latin squares) are useful in the design of statistical experiments and are connected with other areas of discrete mathematics, such as finite geometries and extremal graph theory.

It is also possible to construct magic squares using orthogonal latin squares. Letting $N(n)$ denote the maximum size of a family of mutually orthogonal $n \times n$ latin squares, it is known that $N(n) \ge n^{1/14.8}$ for large n. It is presently a challenging open problem to find any improvement on the exponent.

Problem 3 *Show that $N(n) \le n - 1$ for $n \ge 2$, and that equality holds when n is prime.*

There are many additional topics worth exploring from here, but let us now return to where we started and focus on the Sudoku puzzle, at least in spirit. Define a *partial latin square* as an $n \times n$ array, each of whose cells is either empty or contains one of n symbols, and such that no symbol appears twice in any row or column. See Figure 3.

1		4	
	3		
4		2	

Figure 3: A partial latin square

Naturally, a *completion* of a partial latin square P is a latin square L where P agrees with L on its nonblank cells. Figure 4 shows that even very sparsely filled partial latin squares may admit no completion.

2				
	1			
		1		
			$\ddots$	
				1

Figure 4: One with no completion

In studying the completion question, it is worth beginning with some special partial latin squares. For $1 \le k \le n$, a $k \times n$ *latin rectangle* is a $k \times n$ array of symbols from an n-element set such that every row contains all symbols exactly once, and every column contains k distinct symbols. An example 3×6 latin rectangle is shown on the left of Figure 5.

3	4	5	6	1	2
6	1	4	5	2	3
2	6	1	4	3	5

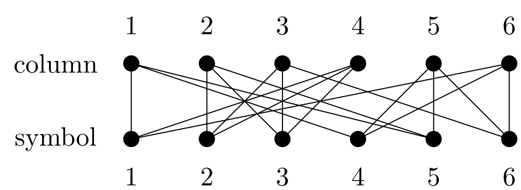

Figure 5: A latin rectangle and its graph of unused symbols

Theorem 1 *For $1 \le k \le n$, every $k \times n$ latin rectangle admits a completion to an $n \times n$ latin square.*

To prove Theorem 1, it is enough to show that a new row can be added when $k < n$. This latter fact comes from a special case of Phillip Hall's so-called 'marriage theorem', a major result in the early development of combinatorics.

In the language of graph theory, this theorem guarantees a perfect matching in any k-regular bipartite graph on $2n$ vertices. An example of such a graph with $k = 3$ and $n = 6$ is shown in the diagram. Lines, or 'edges' are drawn between each column and the symbols available for that column. The interested reader can find a perfect matching in the graph and use it to extend the given latin rectangle. More details on latin rectangles can be found in [6].

Problem 4 *Show that the number of different ways to complete an $(n-2) \times n$ latin rectangle is a power of two.*

It is also possible to complete smaller sub-rectangles under a mild hypothesis.

Theorem 2 (Ryser, 1956) *Let P be an $n \times n$ partial latin square in which the filled cells are precisely those of an $r \times s$ sub-array. Suppose that, in P, every symbol appears at least $r + s - n$ times. Then P admits a completion to an $n \times n$ latin square.*

The above can be reduced to the latin rectangle case by appending an $r \times (n - s)$ sub-array to the right of P.

Problem 5 *Let $n > m$ be positive integers. Use Theorem 1 to prove that an $m \times m$ latin square can be extended to an $n \times n$ latin square if and only if $n \geq 2m$.*

Problem 6 *Let P be any partial latin square. Prove that it is possible to erase at most three-quarters of the cells of P so that the resulting partial latin square P' admits a completion. Can the fraction of erased cells be lowered?*

We have seen that partial latin squares with as few as n filled cells can fail to admit completions. On the other hand, a celebrated result, previously known as Evans' conjecture, asserts that $n - 1$ or fewer filled cells can always be completed.

Theorem 3 (Smetaniuk, 1981) *Every $n \times n$ partial latin square with at most $n - 1$ filled cells admits a completion.*

The proof is a beautiful and surprisingly simple use of mathematical induction.

Problem 7 *Using Theorem 2, obtain an easy proof of the following weakening of Theorem 3: Every $n \times n$ partial latin square with at most $n/2$ filled cells admits a completion.*

As a sort of middle ground between Theorems 2 and 3, we may want to complete typical 'sparse' latin squares. For a positive real number ϵ, let us call an $n \times n$ latin square *ϵ-dense* if every row, column, and symbol is used at most ϵn times. For this, it is helpful to think of n as very large. Recall that our partial latin square with no completion in Figure 4 over-used one symbol; this sort of thing is excluded by our sparseness condition, and the completion question is back in play.

Thresholds on ϵ for the completion of ϵ-dense latin squares have been a topic of interest over the past few decades. Daykin and Häggkvist conjectured in [4] that all 1/4-dense partial latin squares can be completed. The first serious progress toward this conjecture was by Chetwynd and Häggkvist, who showed in [3] that, for sufficiently large even integers n, $\epsilon = 10^{-5}$ suffices to guarantee a completion. Gustavsson [5] obtained the threshold $\epsilon = 10^{-7}$ for all n. These technical proofs required long chains of substitutions. Recently, Bartlett [2] obtained completions with $\epsilon = 10^{-4}$ for large n using a neat idea involving 'negatively occurring' symbols. In fact, he showed that completion is possible for densities near 1/12, but under a strong additional assumption on the total number of filled cells.

Problem 8 *Justify why $\epsilon = 1/4$ is a barrier for the sparse completion problem.*

Very recently, work of the author, along with a key theorem in [1], has shown that all large partial latin squares which are about 4%-dense have a completion. The proofs are constructive in principle, yet very technical. And we are still a long way from the barrier of 25%. What's worse, it is more honest to say that the squares to which it applies are not just large, but colossal beyond comprehension! Still, it represents a step toward a better understanding of the completion problem. The work actually addresses a series of more general questions in graph theory, drawing upon techniques from different areas of mathematics, including probability and linear algebra.

Mathematics is truly exciting when insightful solutions meet natural or universal questions, especially those that might be casually pondered by a curious non-mathematician. Science and technology have fed a steady diet of such questions to the mathematician. But it is fair to say that puzzles such as Sudoku, with their simplicity and broad appeal, offer another source of 'natural' questions. In some cases, even seemingly innocent puzzles are pushing the frontiers of mathematics.

References

[1] B. Barber, D. Kühn, A. Lo, D. Osthus and A. Taylor, Clique decompositions of multipartite graphs and completion of Latin squares. Preprint at `http://arxiv.org/abs/1603.01043`.

[2] P. Bartlett, Completions of ϵ-dense partial latin squares, *J. Combin. Des.* 21 (2013), 447–463.

[3] A.G. Chetwynd and R. Häggkvist, Completing partial $n \times n$ latin squares where each row, column and symbol is used at most cn times. *Reports, Dept. of Mathematics, University of Stockholm*, 1985.

[4] D.E. Daykin and R. Häggkvist, Completion of sparse partial latin squares, in: Graph Theory and Combinatorics: Proceedings of the Cambridge Conference in Honor of Paul Erdos. Academic Press, London, pp. 127–132, 1984.

[5] T. Gustavsson, Decompositions of large graphs and digraphs with high minimum degree. Doctoral Dissertation, Department of Mathematics, Stockholm University, 1991.

[6] J.H. van Lint and R.M. Wilson, A course in combinatorics, 2nd ed. Cambridge University Press, Cambridge, 2001.

Peter J. Dukes
Department of Mathematics and Statistics
University of Victoria
Victoria, BC, Canada
`dukes@uvic.ca`