

A Lower Bound for the Length of a Partial Transversal in a Latin Square, Revised Version

Pooya Hatami^a and Peter W. Shor^b

^a*Department of Mathematical Sciences,*

Sharif University of Technology, P.O.Box 11365-9415, Tehran, Iran

^b*Department of Mathematics,*

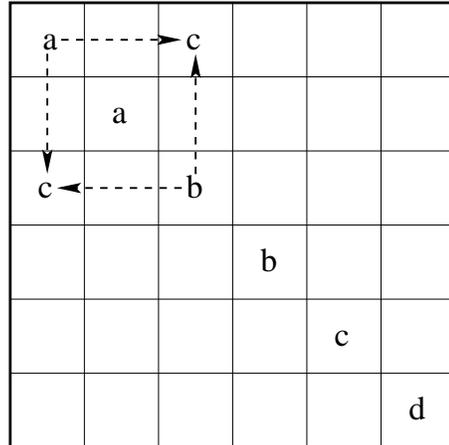
Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

It is proved that every $n \times n$ Latin square has a partial transversal of length at least $n - O(\log^2 n)$. The previous papers proving these results [including one by the second author] not only contained an error, but were sloppily written and quite difficult to understand. We have corrected the error and improved the clarity.

1 Introduction

A Latin square of order n is an $n \times n$ array of cells each containing one of n distinct symbols such that in each row and column every symbol appears exactly once. We define a partial transversal of length j as a set of n cells with exactly one in each row and column and containing exactly j distinct symbols (this differs from the usual definition in that $n - j$ extra cells are added). Koksma [5] showed that a Latin square of order n has a partial transversal of length at least $(2n + 1)/3$. This was improved by Drake [3] to $3n/4$ and then simultaneously by Brouwer *et al.* [1] and by Woolbright [8] to $n - \sqrt{n}$. This was in turn improved by Shor [7] to $n - 5.53 \log^2 n$ and then by Fu *et al.* [4], who optimized the parameters in [7] to slightly improve the constant. One of us (P.H.) discovered a bug in [7] that also affects [4]. This paper fixes the bug, which was caused by the reversal of inequality (26) in [7]. We still obtain an $n - O(\log^2 n)$ lower bound, albeit with a worse constant than in [7, 4]. This is well below Brualdi's conjecture of $n - 1$, and Ryser's of n for odd n [2, 6]. The proof in this paper is much the same as in [7] except for the last part of Section 4. The earlier part of the paper has been revised to improve the clarity of the presentation.



Square 1

Figure 1: Square 1: An example of the operator $\#$. In this case, we replace the cells (1,1) and (3,3) in the partial transversal on the diagonal with the cells (1,3) and (3,1) to obtain another partial transversal, also of length 4.

2 The Operation $\#$

Given a partial transversal T of length $n - k$, with $k \geq 2$, one can find another partial transversal of equal or greater length in the following manner: Choose two cells in T , say cells (i_1, j_1) and (i_2, j_2) , such that $T - \{(i_1, j_1), (i_2, j_2)\}$ contains $n - k$ distinct symbols. These two cells can either contain two distinct duplicated symbols, or two occurrences of the same symbol, provided this symbol appears in the transversal at least three times. Replace these two cells with the cells (i_1, j_2) and (i_2, j_1) . Since we chose cells containing duplicated symbols, the new partial transversal has length at least $n - k$, as each of the symbols in the original transversal is represented in one of the unchanged cells (see Square 1). We call this operation $\#$, a notation chosen for its shape.

We now give a motivating example of the use of the operation $\#$, by applying it to show that every Latin square of order 6 has a partial transversal of length at least 5. Consider a counterexample. Assume for now that the longest partial transversal has length 4. The square must thus have a partial transversal containing a multiset of symbols either of the form (a, a, b, b, c, d) or the form (a, a, a, b, c, d) . Let us analyze the case where it contains (a, a, b, b, c, d) (See Square 1). We assume that this partial transversal is on the diagonal, and call it T_0 . We can apply $\#$ to the cells (1, 1) and (3, 3) in T_0 to get a new partial transversal T_1 . By our hypotheses, the new cells (1, 3) and (3, 1) in T_1 must contain a symbol chosen from the set $\{c, d\}$. By symmetry, we only need to analyze two cases here: either both symbols are the same or there is one c and one d . We will analyze the case where they are both c 's. We can apply $\#$ to the cells (1, 3) and (5, 5) in T_1 to obtain a new partial transversal T_2 (as shown

a		c			
	a				
c		b			
			b		
		y		c	
					d

Square 2

a		c	d	x	x
	a				
c		b			
d			b		
		x		c	
			x		d

Square 3

Figure 2: Square 2: Another example of the operation $\#$. In this, we replace the cells (1,3) and (5,5) in the partial transversal indicated in bold with the cells containing symbols x and y . If this Latin square had no partial transversal of length greater than 4, we must have that $x \in \{b, d\}$ and $y \in \{a, d\}$. Square 3: After two more applications of $\#$, we know that if the Latin square had no partial transversal of length greater than 4, it would have to look like this, with the symbols x chosen from the set $\{a, b, c, d\}$.

in Square 2), and we discover that the symbols in (1, 5) and (5, 3) must be chosen from the set $\{a, b, d\}$.

Now, starting from T_0 again we can apply $\#$ to the cells (1, 1) and (4, 4) to obtain a partial transversal T_3 , and we discover that the cells (1, 4) and (4, 1) must both contain d . (See Square 3.) We can now apply $\#$ to the cells (1, 4) and (6, 6) in T_3 to obtain T_4 , and discover that the symbols in (1, 6) and (6, 4) must be chosen from the set $\{a, b, c\}$. We now know that our Latin square looks like Square 3, where the x 's are symbols from the set $\{a, b, c, d\}$. The first row contains five distinct symbols from the set $\{a, b, c, d\}$, a contradiction by the pigeonhole principle.

The case where the longest partial transversal has length less than 4, as well as the cases where the cell (3, 1) in Square 1 is a d instead of a c and the case where the diagonal is (a, a, a, b, c, d) , can be handled using a very similar analysis, which we will not present here.

We define a *partial Latin square* as an $n \times n$ square with some of its cells containing symbols (the others we call empty) such that no symbol appears twice in any row or column. A *partial transversal* of an $n \times n$ partial Latin square is a set of n non-empty cells, one from each row and column. We say this partial transversal *has length* j if it contains j distinct symbols. An $m \times m$ *subsquare* S' of an $n \times n$ partial Latin square S is the set of m^2 cells in some subset of m rows and some subset of m columns of S , where some nonempty cells in S may possibly be replaced by empty

cells in S' .

Consider a Latin square with a partial transversal of maximum length $n - k$, with $k \geq 2$. By applying $\#$ to this partial transversal, we will get other partial transversals, whose length must also be $n - k$ and whose set of symbols is the same as the first. Applying $\#$ repeatedly to these partial transversals, we eventually will obtain a set of such partial transversals closed under $\#$. All of these partial transversals contain the same set of $n - k$ distinct symbols, so by ignoring all cells except those in this set of partial transversals, we obtain a partial Latin square S containing $n - k$ symbols and a set \mathcal{T} of partial transversals of S closed under $\#$. We will call this pair (S, \mathcal{T}) a partial Latin square satisfying A_k . More formally, we have:

Definition. An $n \times n$ partial Latin square satisfying A_k is a partial Latin square, together with a nonempty set \mathcal{T} of partial transversals of S of length $n - k$. Each non-empty cell must appear in at least one of the partial transversals in \mathcal{T} . The set \mathcal{T} of partial transversals must form a connected graph under the operation $\#$, and must be closed under the operation $\#$.

For a partial Latin subsquare (S', \mathcal{T}') satisfying $A_{k'}$ of a partial Latin square (S, \mathcal{T}) satisfying A_k , in addition to the properties contained in the above definition, we also require an inheritance property. Namely, we require \mathcal{T}' to be a subset of the set \mathcal{T} restricted to S' , i.e., that

$$\mathcal{T}' \subseteq \{T \cap S' : T \in \mathcal{T}\}.$$

Note that Brualdi's conjecture (that all Latin squares of order n have a partial transversal of length at least $n - 1$) does not appear to rule out the existence of partial Latin squares satisfying A_2 , or A_k for $k > 2$.

If S is an $n \times n$ partial Latin square satisfying A_k , we can construct an $(n + 1) \times (n + 1)$ partial Latin square S' satisfying A_k , by adding an extra row and column that consist entirely of empty cells except for the $(n + 1, n + 1)$ cell, which contains a new symbol. The partial transversals in \mathcal{T}' are those in \mathcal{T} augmented by the cell $(n + 1, n + 1)$. Together with the case analysis on Latin squares of size 6 sketched earlier, this observation implies that any partial Latin square satisfying A_2 must have size at least 7. In terminology we will be defining later in this paper, this means that

$$n_2 \geq 7. \tag{1}$$

We use the properties of a minimal Latin square satisfying A_k to obtain a set of inequalities, and then use these inequalities to derive our main result. We first prove a lemma:

Lemma 1 *Given a partial Latin square (S, \mathcal{T}) satisfying A_k such that no subsquare satisfies A_k , then no cell is contained in all partial transversals in \mathcal{T} . That is, given a non-empty cell (i, j) and a partial transversal containing (i, j) , by a sequence of operations $\#$, one can obtain a partial transversal in \mathcal{T} not containing (i, j) .*

Proof. Suppose there is a cell (i, j) contained in all partial transversals. We will call this a *fixed cell*. Let a be the symbol in this cell. If a appears anywhere else in the partial Latin square, there is a transversal containing both a 's (the second a appears in some partial transversal since every non-empty cell does, and this partial transversal must contain the first a since all partial transversals do). We can then apply $\#$ to this partial transversal to obtain a partial transversal without the fixed cell, a contradiction. We are left with the case where a does not appear anywhere else in the partial Latin square. Now, by deleting the row and column containing the a , one finds a subsquare satisfying A_k , a contradiction of the hypothesis. ■

We have just proved that no cell in a minimal square satisfying A_k is fixed, so given a non-empty cell in such a square, there is a partial transversal in \mathcal{T} containing both that cell and another cell with the same symbol (otherwise, the graph of partial transversals would not be connected by $\#$). We can choose any filled cell, say $(1, 1)$, and choose a partial transversal T_0 through it that duplicates the symbol in it, say a . Now, let $\mathcal{T}^* \subseteq \mathcal{T}$ be the set of partial transversals containing at least two a 's, including the one in cell $(1, 1)$. Consider the connected component of \mathcal{T}^* which is generated by a sequence of operations $\#$ starting with T_0 . This component corresponds precisely to the set of partial transversals generated by $\#$ starting from $T_0 - (1, 1)$ in the subsquare formed by deleting the first row and column. Taking this set of partial transversals gives an $(n - 1) \times (n - 1)$ partial Latin square satisfying A_{k-1} . Note that this subsquare may have some empty cells which were filled in the original $n \times n$ square.

Lemma 2 *In an $(n - 1) \times (n - 1)$ partial Latin square satisfying A_{k-1} induced as described above from an $n \times n$ partial Latin square satisfying A_k , the partial transversals generated by $\#$ must have a fixed cell, i.e., there is some cell that appears in all of these partial transversals.*

Proof. We assume without loss of generality that the partial transversal T_0 containing two a 's discussed above is the diagonal. Suppose that there are no fixed cells in the $(n - 1) \times (n - 1)$ partial Latin square. Then there must be a partial transversal T'_1 in this smaller square which contains the cell (i, i) as well as another cell with the same symbol. The partial transversal $T_1 := T'_1 \cup (1, 1)$ must also appear in the $n \times n$ square. Now, in T_1 , either $(1, 1)$ and (i, i) contain two different duplicated symbols, or there are at least three a 's in T_1 and $(1, 1)$ and (i, i) both contain a . In either case, we can apply $\#$ to T_1 , deleting the cells $(1, 1)$ and (i, i) and obtaining the cells $(1, i)$ and $(i, 1)$ (See Square 4). Since i was arbitrary, this gives us n filled cells in the first row and column of the square, a contradiction since there are only $n - k$ distinct symbols. ■

We now extend the analysis of Lemma 2 to prove the following.

Theorem 1 *In a partial Latin square satisfying A_k such that no subsquare satisfies A_k , there are at least $n_{k-1} + k$ filled cells in each row and column, where n_{k-1} is the size of the smallest subsquare satisfying A_{k-1} .*

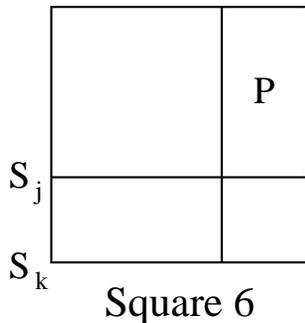


Figure 4: The region P in the proof of Theorem 2

we find that the $(j, 1)$ cell and the (i, j) cell are filled. (Again, if both cells (j, j) and $(i, 1)$ contain c , there are three c 's in the partial transversal.) Thus, we know that the $(j, 1)$ cell is filled. Since there are at least $k - 1$ symbols in the $(1, i)$ cells, $2 \leq i \leq m + 1$, which are not in the subsquare satisfying A_{k-1} , we can apply the same process to obtain $k - 1$ filled cells in the first column in or below the $(m + 2)$ nd row. This gives at least $m + k$ filled cells in the first column, since the first $m + 1$ cells are also filled. Now, $m \geq n_{k-1}$, because m is the size of a subsquare satisfying A_{k-1} , and n_{k-1} was the size of the minimal such subsquare. Since $(1, 1)$ was an arbitrary cell in our original partial transversal, this argument shows that at least $n_{k-1} + k$ cells are filled in each row and column. ■

If we let $n_k = n$ be the size of the original partial Latin square satisfying A_k , then this Theorem shows that

$$n_k \geq n_{k-1} + 2k, \tag{2}$$

since the larger square has $n_k - k$ distinct symbols, of which at least $n_{k-1} + k$ appear in each row and column.

3 An Inequality

Let S_k be a square satisfying A_k such that no subsquare satisfies A_k . It was shown in Section 2 that there must be a subsquare satisfying A_{k-1} . Choose S_{k-1} to be the smallest subsquare of S_k satisfying A_{k-1} and, recursively, S_m to be the smallest subsquare of S_{m+1} satisfying A_m , until the sequence ends at S_2 . Let n_j be the size of S_j .

Theorem 2 *In S_k , as defined above, for all $j < k$,*

$$(n_k - n_j)(2n_j + n_{k-1} - 2n_k + 2k - j) \leq n_j(n_j - n_{j-1} - 2j). \tag{3}$$

Proof. Consider Square 6. We will count the number of filled cells in the rectangle P in two different ways. First, there are $n_k - n_j$ columns in P , and since

each column of S_k has at least $n_{k-1} + k$ filled cells, we have at least $n_{k-1} + k - (n_k - n_j)$ filled cells in each column of P , and at least $(n_k - n_j)(n_{k-1} + n_j - n_k + k)$ filled cells in P .

We will call the symbols in S_j *old* symbols and those in S_k and not in S_j *new* symbols. There are $n_j - j$ old symbols and $n_k - k - n_j + j$ new symbols. There are n_j rows in P . In each row of S_j there are at least $n_{j-1} + j$ old symbols. Since there are only $n_j - j$ distinct old symbols, this leaves at most $n_j - j - (n_{j-1} + j)$ old symbols in each row of P , for a total of at most $n_j(n_j - n_{j-1} - 2j)$ filled cells containing old symbols in P .

There are $n_k - k - n_j + j$ new symbols, and $n_k - n_j$ columns in P . Thus, there are at most $(n_k - n_j)(n_k - k - n_j + j)$ filled cells containing new symbols in P . Adding the number of cells with old and with new symbols in P , we get a upper bound for the number of filled cells in P . Setting this upper bound greater than or equal to the lower bound, and simplifying, we obtain the inequality (3) above. ■

4 The Main Result

Suppose we have a Latin square with no partial transversal of length more than $n - l$. By the previous sections, we have a sequence $n_2 < n_3 < \dots < n_l$ satisfying the Inequalities (1), (2), (3), from the previous section. Reiterating these inequalities, we have that if $2 \leq i \leq l$ and $1 \leq j < k \leq l$, then

$$n_2 \geq 7, \tag{1}$$

$$n_i \geq n_{i-1} + 2i, \tag{2}$$

$$(n_k - n_j)(2n_j + n_{k-1} - 2n_k + 2k - j) \leq n_j(n_j - n_{j-1} - 2j). \tag{3}$$

We will now derive the inequality $k \leq 11.053 \log^2 n_k$ from the Inequality (3).

We first prove the following lemma.

Lemma 3 *Either*

$$n_j \leq \frac{4}{5}n_k$$

or

$$n_j - n_{j-1} \geq \frac{1}{3}(n_k - n_{j-1}).$$

Proof. Letting

$$d_k := n_k - n_{k-1}, \quad d_j := n_j - n_{j-1}, \tag{4}$$

we have, from (3)

$$d_j - 2j \geq \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k + 2k - j). \tag{5}$$

The direction of the inequality lets us remove the lower order terms, giving

$$d_j \geq \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k). \quad (6)$$

Now, we assume that $n_j \geq \frac{4}{5}n_k$. This gives

$$d_k = n_k - n_{k-1} \leq n_k - n_j \leq \frac{1}{5}n_k, \quad (7)$$

$$n_k + d_k \leq \frac{6}{5}n_k, \quad (8)$$

$$n_k + d_k \leq \frac{3}{2}n_j. \quad (9)$$

Combining (6) and (9) gives

$$d_j \geq \frac{1}{2}(n_k - n_j), \quad (10)$$

By the definition of d_j , we have

$$n_j - n_{j-1} \geq \frac{1}{2}n_k - \frac{1}{2}n_j, \quad (11)$$

so

$$\frac{3}{2}n_j - \frac{3}{2}n_{j-1} \geq \frac{1}{2}(n_k - n_{j-1}). \quad (12)$$

and

$$n_j - n_{j-1} \geq \frac{1}{3}(n_k - n_{j-1}), \quad (13)$$

completing the proof. ■

Now, suppose that $n_k < \frac{5}{4}n_j$, so

$$\frac{1}{3}(n_k - n_{j-1}) \leq n_j - n_{j-1}, \quad (14)$$

giving

$$n_k - n_j \leq \frac{2}{3}(n_k - n_{j-1}). \quad (15)$$

Since (15) holds for all j where $j < k$, and $n_k < \frac{5}{4}n_j$, by induction we get,

$$1 \leq n_k - n_{k-1} \leq \left(\frac{2}{3}\right)^{k-j} (n_k - n_{j-1}), \quad (16)$$

or

$$k - j \leq \log_{3/2}(n_k - n_{j-1}). \quad (17)$$

Now, suppose in addition that

$$k - j - 1 > \log_{3/2} \frac{n_j}{4}, \quad (18)$$

then

$$\log_{3/2} \frac{n_j}{4} < k - j - 1 \leq \log_{3/2}(n_k - n_j), \quad (19)$$

$$\frac{n_j}{4} < n_k - n_j \quad (20)$$

$$\frac{5}{4}n_j < n_k, \quad (21)$$

a contradiction.

If $k - j \geq \lfloor \log_{3/2} n_j \rfloor$, then since $\lfloor \log_{3/2} n_j \rfloor > \log_{3/2} n_j - \log_{3/2} 4 + 1$, we have that (18) holds, implying that $n_k \geq \frac{5}{4}n_j$.

We now let $k_4 = 2$, and

$$k_i = k_{i-1} + \lfloor \log_{3/2}(n_{k_{i-1}}) \rfloor. \quad (22)$$

By induction, we obtain that for $l \geq k_i$,

$$n_l \geq \left(\frac{5}{4}\right)^{i+1}, \quad (23)$$

where the base case follows from $n_2 \geq 7 > (5/4)^5$.

We know that $n_{k_{i-1}} < n_k$ for $k_{i-1} < k$. So from (22), if $k_{i-1} < k$, we have

$$k_i \leq k_{i-1} + \lfloor \log_{3/2} n_k \rfloor. \quad (24)$$

We can now prove the following Lemma. We will specify the exact value of c later.

Lemma 4 *For $c \geq 1/2$, either*

$$\frac{1}{c} \log_{3/2} n_k \geq k^{\frac{1}{2}}, \quad (25)$$

or

$$c \log_{5/4} n_k > k^{\frac{1}{2}}. \quad (26)$$

Proof. If $\log_{3/2} n_k \geq ck^{\frac{1}{2}}$, we have (25). Otherwise suppose that

$$\log_{3/2} n_k < ck^{\frac{1}{2}}. \quad (27)$$

Then from (24), for all $k_{i-1} < k$ we have

$$k_i < k_{i-1} + ck^{\frac{1}{2}}. \quad (28)$$

Let j be the minimum integer such that $k_j \geq k$. Summing (28) over i gives

$$k \leq k_j < k_4 + (j - 4)ck^{\frac{1}{2}}, \quad (29)$$

which rearranges to

$$j > \frac{k - k_4}{ck^{1/2}} + 4. \quad (30)$$

This shows that for minimum j such that $k_j \geq k$ we have

$$j > \frac{k - 2}{ck^{1/2}} + 4 > \frac{1}{c}k^{\frac{1}{2}}. \quad (31)$$

Using (23) with $i = j - 1$ and (31) we obtain

$$n_{k_{j-1}} \geq \left(\frac{5}{4}\right)^j > \left(\frac{5}{4}\right)^{\frac{1}{c}k^{1/2}}. \quad (32)$$

We know that $n_k > n_{k_{j-1}}$ (because $k > k_{j-1}$). Then

$$\log_{5/4} n_k > \frac{1}{c}k^{\frac{1}{2}}, \quad (33)$$

giving (26). ■

We can now make the left-hand side of the two equations in Lemma 4 equal by setting $c = \sqrt{\log \frac{5}{4} / \log \frac{3}{2}}$. This gives

$$\frac{1}{\sqrt{\log \frac{5}{4} \log \frac{3}{2}}} \log n_k > k^{1/2}, \quad (34)$$

from which follows:

Theorem 3 *Every Latin square has a partial transversal of length at least*

$$n - 11.053 \log^2 n. \quad (35)$$

Here $11.053 \approx (\log \frac{5}{4} \log \frac{3}{2})^{-1}$. No serious attempt has been made to optimize this constant.

As was remarked in [7] the Inequality (3) cannot imply anything better than $n - \log_2 n$, since the sequence $n_k = 2^k$ satisfies (3). Let us take the opportunity to remark that Inequalities (1–3) cannot in fact achieve a bound better than $n - O(\log^2 n)$. If we let $\kappa = \lfloor k^{\frac{1}{2}} \rfloor$ and $\gamma = k - \kappa^2$, then the sequence

$$n_k = \beta b^\kappa - \alpha a^{3\kappa - \gamma}$$

for sufficiently large $b \gg a \gg 1$ and $\beta \gg \alpha$ satisfies these inequalities and has growth rate of $n_k = e^{O(k^{1/2})}$.

5 Acknowledgments

The authors would like to thank the referees for careful reading of this paper and exceptionally helpful comments. The second author is indebted to Herb Ryser, who introduced him to this problem over 27 years ago.

References

- [1] A. E. Brouwer, A. J. de Vries, and R. M. A. Wieringa. A lower bound for the length of partial transversals in a Latin square. *Nieuw Arch. Wisk.*, 24(3):330–332, 1978.
- [2] R. A. Brualdi and H. J. Ryser. *Combinatorial Matrix Theory*. Cambridge Univ. Press, 1992.
- [3] D. A. Drake. Maximal sets of Latin squares and partial transversals. *J. Statist. Plann. Inference*, 1:143–149, 1977.
- [4] H. L. Fu, S. C. Lin, and C. M. Fu. The length of a partial transversal in a Latin square. *J. Comb. Math. Combin. Comput.*, 43:57–64, 2002.
- [5] K. K. Koksma. A lower bound for the order of a partial transversal in a Latin square. *J. Combinatorial Theory Ser. A*, 7:94–95, 1969.
- [6] H. J. Ryser. Neuere Probleme der Kombinatorik. In *Vorträge über Kombinatorik*, pages 69–91, Oberwolfach, Germany, July 24–29 1967. Mathematisches Forschungsinstitute Oberwolfach.
- [7] P. W. Shor. A lower bound for the length of a partial transversal in a Latin square. *J. Combinatorial Theory Ser. A*, 33(1):1–8, 1982.
- [8] D. E. Woolbright. An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct elements. *J. Combinatorial Theory Ser. A*, 24:235–237, 1978.