Combin. Des. March 31 → April 4,2

L.S. Critical Sets and partial Transversal

Definition (Critical set of L.S.)

A partial Latin square of order n (PLS(n)), C, is called a critical set of a Latin square L, if

(1) C can be completed uniquely to L, and

(2) Any proper subset of C can not be completed to a unique Latin square.

Here is an example.

0	1	2	3
1	2	3	0
2	3	0	1
3	0	1	2

A critical set PLS(4)

0			
1			
			2

0	1		
1			

can not be completed uniquely

Fact. Let C be a critical set of a Latin square of order n. Then, C has at most one empty row, one empty column or one missing elements in $\mathbb{Z}_n$.

Proof. If one of the three cases is true, then C can not be completed uniquely.

<u>Proposition 1</u> There exists a critical set C of a Latin square of order n with $|C| \sim \frac{n^2}{4}$.

<u>Proof.</u>

0	1	2	3
1	2	3	0
2	3	0	1
3	0	1	2

0	1	3	4	2
1	3	4	2	0
3	4	2	0	1
4	2	0	1	3
2	0	1	3	4

0	1	2	4	5	3
1	2	4	5	3	0
2	4	5	3	0	1
4	5	3	0	1	2
5	3	0	1	2	4
3	0	1	2	4	5

We may use the above pattern to find a critical set of size $c = (1+2+\cdots+\lfloor\frac{n}{2}\rfloor) + (1+2+\cdots+\lfloor\frac{n-1}{2}\rfloor)$. Now, if n is even,

$$c = 1+2+\cdots+\frac{n}{2}+1+2+\cdots+\frac{n-2}{2} = 2\left(1+2+\cdots+\frac{n-2}{2}\right)+\frac{n}{2}$$

$$= 2\cdot\frac{1}{2}\cdot\frac{n-2}{2}\cdot\frac{n}{2}+\frac{n}{2} = \frac{n^2-2n}{4}+\frac{n}{2} = \frac{n^2}{4}.$$ On the other hand, $c = \frac{n^2-1}{4}$ when n is odd. ∎

So far, no critical set of size less than $\lfloor\frac{n^2}{4}\rfloor$ has been constructed. Therefore, the following conjecture remains unsettled.

<u>Conjecture (Critical Set)</u> The size of a critical set of a $LS(n)$ is at least $\lfloor\frac{n^2}{4}\rfloor$.

But, we can not apply this idea to Sudoku. It has been proved that there exists critical sets of size less than $\lceil \frac{81}{4} \rceil$, in fact, far less than that. A recent result shows that a critical set for Sudoku square is at least of size 17 (?). The reason is very simple, since Sudoku squares do have extra properties, there are nine prescribed subsquares.

Definition (Spectrum of critical sets)

$$\mathrm{Spec}_C(n) = \{ |C| \mid C \text{ is a critical set of a L.S. of order } n \}.$$

e.g. $\mathrm{Spec}_C(2) = 1$, $\mathrm{Spec}_C(3) \supseteq \{2, 3\}$.

√ Exercise 2-1 Find $\mathrm{Spec}_C(4)$. (3 points)

(*) Critical sets can be applied in constructing Sharing Scheme which is an important topic cryptography. (Key: C, partition C into several parts.)

(**) The key is a Latin square of order about "50"?

plaintex $\xleftrightarrow{\text{Latin Square}}$ Ciphertex How?

key

1. How to use Latin squares as keys?

<u>Exercise 2-2.</u> (3 points) Use your imagination!

<u>Exercise 2-3</u> (3 points)

How to use critical sets for sharing schemes?

<u>Definition</u> (Sharing schemes)

$\mathbb{P}$: participants

$\forall\ a \in \mathbb{P}$, $s(a)$: share of a.

K (keys) : Latin squares of order n.

$A \subseteq \mathbb{P}$, Access structure

$\text{Comb.}(A) = \{$ combination of shares of $a \in A\}$

$\Downarrow$ reveals the key.

Transversal and Partial Transversal

Definition (Transversal)

A transversal of a Latin square of order n is a set of n entries, one from each row and each column, such that all the entries are distinct.

0	2	3	1
3	1	0	2
1	3	2	0
2	0	1	3

Definition (Partial Transversal)

A partial transversal of a Latin square (of order n) is a set of $m \leq n$ entries, no two of them are in the same row or the same column.

0	1	2	3	4	5
1	2	0	4	5	3
2	0	1	5	3	4
3	4	5	0	1	2
4	5	3	1	2	0
5	3	4	2	0	1

Fact 1 If $L \perp M$, then both L and M contain transversals.

In fact, if L (resp. M) is an L.S. of order n, then L (resp. M) contains n disjoint transversals.

Fact 2: If L is a Latin square of order n where n is odd, then $\begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline \end{array} \otimes L$ contains no transversals.

Exercise 2-4 Prove Fact 2. (3 points)

- Determining whether a Latin square contains a transversal or not is a very difficult problem.
- This problem is equivalent to finding a rainbow perfect matching in an n-edge-colored $K_{n,n}$.

Exercise 2-5 Give examples L' that $\begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline \end{array} \otimes L'$ contains a transversal
(3 points)
when L' is a Latin square of even order. (10 points)
$n = 2m,\ m \in \mathbb{N}$

Ryser's Conjecture

For each Latin square of odd order, L, there exists a transversal.

Revised version of Ryser's Conjecture (Brualdi's Conjecture)

For each Latin square of order n, there exists a partial transversal which contains at least $n-1$ distinct entries (partial Transversal of size $\geq n-1$).

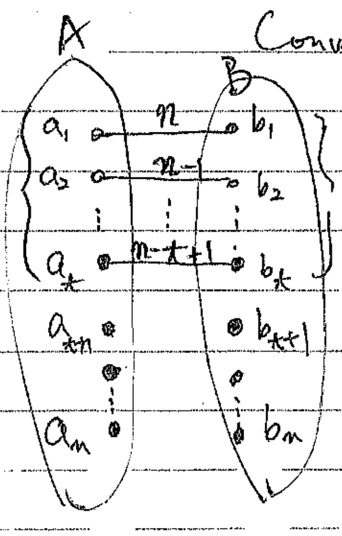
Theorem (P. Shor)

Let T_n be a partial transversal of maximum size in a Latin square of order n. Then $|T_n| \geq n - O((\ln n)^2)$ or $n - c \cdot (\ln n)^2$ where c is positive constant.

Theorem (D. Woolbright and A. E. Brouwer)

$|T_n| \geq n - \sqrt{n}$. (Bonus: 5 points for details.)

Proof.

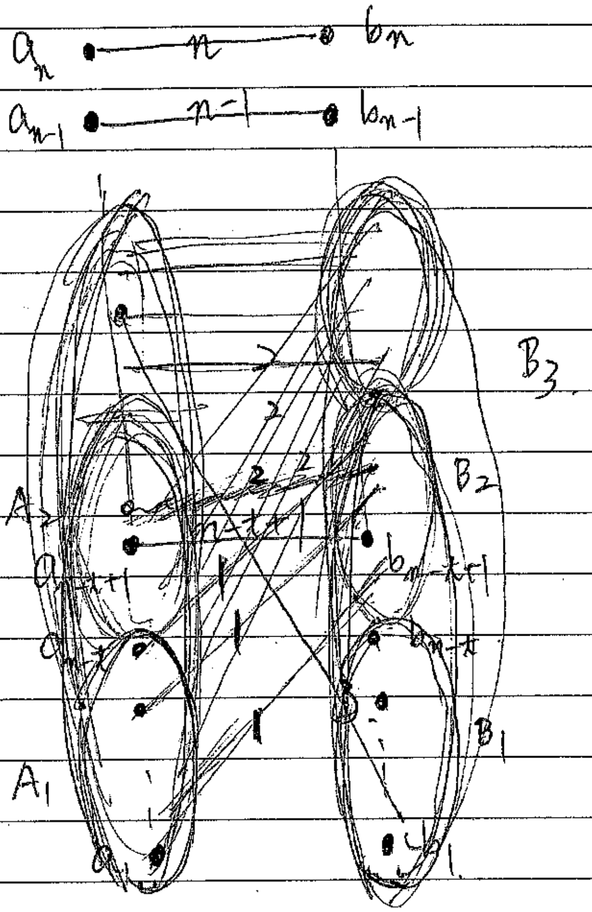


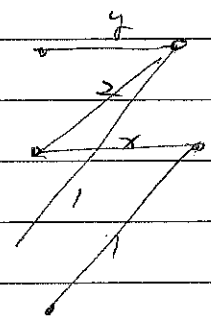
Convert an $LS(n)$ into an n-edge-colored $K_{n,n}$.

Assume that $|T_n| = t$ and they are arranged as "left".

Consider the color $1, 2, \cdots, n-t$ in turn.

Sketch of the proof $|T_n| \ge n-\sqrt{n}$.





(•) $1,2,\cdots,n-t$ (colors) are missing in $\langle A_1,B_1\rangle$

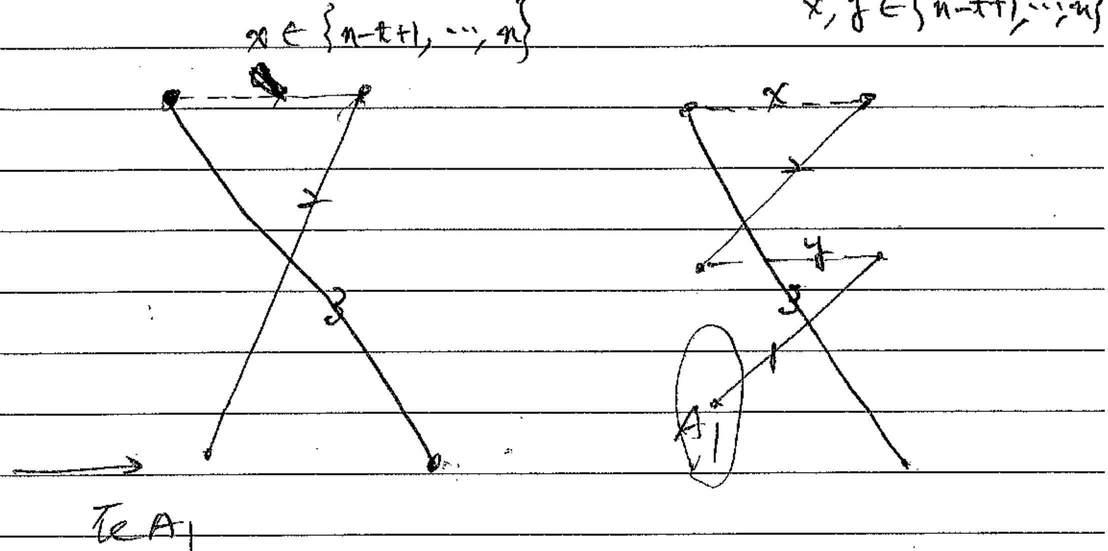
(•) A_1中的点有着"1"的边都往上连到B_2中的点.

(•) A_2中着 $2\sim n-t$ 颜色的点都不会连到B_1中的点。

(•) A_2中(有 $2(n-t)$ 边着颜色2), 由于不能连到B_1中, 至少有 $n-t$ 个会连到 B_2 的外面, 亦即 $B_3\setminus B_2$, 令 $A_3\setminus A_2$ 为对应的点集.

(•) A_3中着 $3\sim n-t$ 颜色的点都不会连到B_1.

如果 $A_3\setminus A_2$ 中的点有3以上的颜色出现在 $\langle A_2\setminus A_1, B_1\rangle$ 中，例如"3"，则有下列两种情况



由于 A_3 可以看颜色"3"出现在 B 中的位置而产生 $B_4\setminus B_3$，以下类推至 A_{n-t} 可以看颜色 $n-t$ 出现的位置而产生 $B_{n+1}\setminus B_n$，

因此 $n \ge (n-t+1)(n-t) \ge (n-t)(n-t)+(n-t)$

$$\ge (n-t)^2$$

$n-t \le \sqrt{n}$，$t \ge n-\sqrt{n}$。

(Bonus, 10 points) Give a talk about Pooya Hatami and P. Shor's work on partial transversal.

(JCT(A) 115, 1103 – 1113, 2008)

0	2	3	1	4	6	7	5
3	1	0	2	7	5	4	6
1	3	2	0	5	7	6	4
2	0	1	3	6	4	5	7
4	7	5	6	0	2	3	1
6	5	7	4	3	1	0	2
7	4	6	5	1	3	2	0
5	6	4	7	2	0	1	3

For example in finding transversal in an $LS(4t)$.

<u>Open problem</u>

1. Find an algorithm to determine the maximum size of a partial transversal.
2. Given an $LS(n)$ L, determine whether L has an orthogonal mate, i.e., find an M such that $L \perp M$ if L does have an orthogonal mate, otherwise, output "No"!

(1) $|T| \geq \frac{n}{2}$ (Trivial case)

(2) K. K. Koksma (1969, J.C.T. 7, 94–95)

$|T_n| \geq \frac{2n+1}{3}$ for $n \geq 7$.

(3) D. A. Drake (1977, J.S.P.I. 1, 143–149)

$|T_n| \geq \frac{3n}{4}$. (Simpler method)

(4) S. M. P. Wang (1978, Ph. D. Thesis OSU)

$|T_n| \geq (9n-15)/11$.

(5) A. E. Brouwer et. al. (1978, Nieuw Archief voor Wiskunde (3) 26, 330–332)

D. E. Woolbright (1978, JCTA 24, 235–237)

$|T_n| \geq n - \sqrt{n}$.

(6) P. W. Shor (1982, JCTA 33, 1–8)

$|T_n| \geq n - 5.53 (\log_e n)^2 = n - 5.53 (\ln n)^2$.

(7) Our result (with 林士強), (2002, JCMCC 43, 57–64)

$|T| \geq n - 5.518 (\ln n)^2$. (Using Calculus.)

(8) Errata of (6), Pooya Hatami and P. W. Shor, (2008, JCTA 115, 1103–1113)

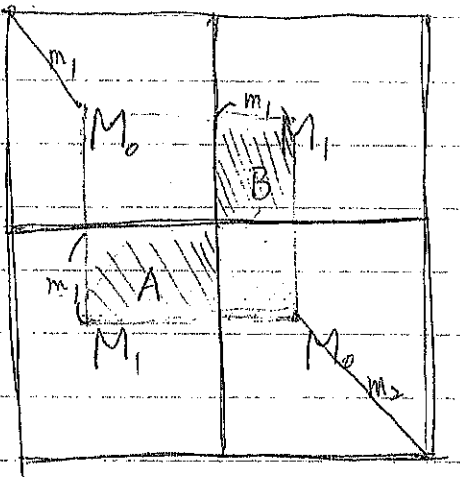
$|T_n| \geq n - O(\ln n)^2$. Find a bug in (6).

(For reference)

Non-existence of a Latin transversal

<u>Theorem</u> Let L be a Latin square of order 2 and M be a Latin square of order m (odd). Then, $L \otimes M$ consists no transversals.

<u>Proof.</u>



M_0 : based on $\mathbb{Z}_m$

M_1 : based on $m + \mathbb{Z}_m$

The entries in $\mathbb{Z}_m$ must occur in M_0's, say m_1 in upper left corner and m_2 in right lower corner, $m_1 < m_2$, $m_1 + m_2 = m$. Then, the entries of $m + \mathbb{Z}_m$ must be in A or B. Then, at most $2m_1$ can be selected, which is smaller than m. ∎

e.g.

0	1	2	3	4	5
1	2	0	4	5	3
2	0	1	5	3	4
3	4	5	0	1	2
4	5	3	1	2	0
5	3	4	2	0	1

JOURNAL OF COMBINATORIAL THEORY, Series A 24, 235–237 (1978)

For reference 1

Note

An $n \times n$ Latin Square Has a Transversal with at Least $n - \sqrt{n}$ Distinct Symbols

DAVID E. WOOLBRIGHT

Department of Mathematics, Auburn University, Auburn, Alabama 36830

Communicated by the Managing Editors

Received January 28, 1977

1. INTRODUCTION

A *latin square* is an $n \times n$ array such that in each row and column each of the integers 1, 2, 3,..., n occurs exactly once. A *transversal* is a collection of n cells, no two of which are in the same row or column. If T is a transversal then by $|T|$ is meant the number of distinct symbols which occur in T. The number $|T|$ is called the *size* of the transversal. Koksma [2] has shown that for all latin squares there is a transversal T such that $|T| \geqslant (2n+1)/3$. Recently Drake [1] has shown $|T| \geqslant (3n)/4$ for at least one transversal T. The purpose of this paper is to substantially improve this lower bound by showing that every $n \times n$ latin square has at least one transversal containing at least $n - \sqrt{n}$ distinct symbols.

2

An $n \times n$ latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols. Let S be an $n \times n$ latin square. It is well known that S can be represented as an n-coloring of the complete bipartite graph as follows: Let $K_{n,n}$ be the complete bipartite graph with vertex classes $A = \{a_1, a_2, ..., a_n\}$ and $B = \{b_1, b_2, ..., b_n\}$ and let $c_1, c_2, ..., c_n$ be n distinct colors. Color edge (a_i, b_j) with c_k if and only if cell (i, j) in S contains the symbol k. Then a transversal T of S is equivalent to n parallel edges in $K_{n,n}$ and the n parallel edges in $K_{n,n}$ will have $|T|$ distinct colors.

Let T be a transversal in S such that $|T| = t$ is a maximum and let P be the collection of n edges in $K_{n,n}$ that correspond to T. By renaming the vertices in A and B we can assume without loss of generality that $P = \{(a_i, b_i) \mid i \in 1, 2,..., n\}$ and that $\pi = \{(a_i, b_i) \mid i \in 1, 2,..., t\}$ is a collec-

t ?

0097-3165/78/0242–0235$02.00/0

tion of t edges with pairwise distinct colorings. Since T was a transversal of maximal size in S, π is a maximal collection of parallel edges in $K_{n,n}$ with pairwise distinct colorings. We can assume that $c_1, c_2, \ldots, c_{n-t}$ are the colors which do not occur in π.

Let $A_1, A_2, \ldots, A_k$ and $B_1, B_2, \ldots, B_k$ be two sequences of sets of vertices in A and B with the following properties:

(1) $A_1 = \{a_n, a_{n-1}, \ldots, a_{t+1}\}$, and $B_1 = \{b_n, b_{n-1}, \ldots, b_{t+1}\}$,

(2) $A_i \subset A_{i+1}$ and $B_i \subset B_{i+1}$,

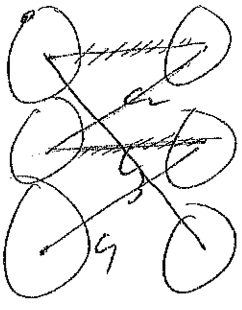
(3) $| A_{i+1} \backslash A_i | = | B_{i+1} \backslash B_i | = n - t$,

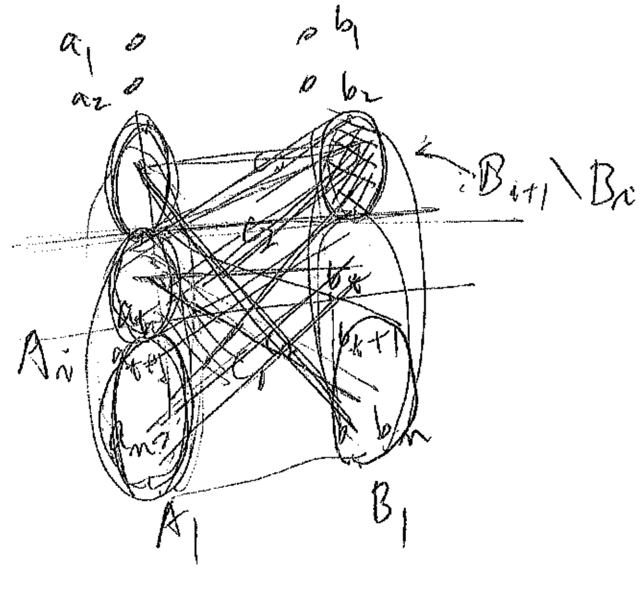
(4) the $n - t$ edges with vertices in $B_{i+1} \backslash B_i$ which are colored c_i all have vertices in A_i, and

(5) any edge (a, b) such that $a \in A_i$ and $b \in B_1$ is not colored c_j if $i \leqslant j \leqslant n - t$.

We show that if $A_1, A_2, \ldots, A_k$ and $B_1, B_2, \ldots, B_k$ are sequences with properties (1), (2), (3), (4), and (5) and if $k < n - t + 1$ then we can always construct sets A_{k+1} and B_{k+1} such that the sequences $A_1, A_2, \ldots, A_k, A_{k+1}$ and $B_1, B_2, \ldots, B_k, B_{k+1}$ also have properties (1), (2), (3), (4), and (5). If $k < n - t + 1$ construct A_{k+1} and B_{k+1} as follows: Consider the $k(n - t)$ edges with vertices in A_k which are colored c_k. None of these edges have a vertex in B_1 because of (5) so at least $n - t$ of these edges must have a vertex in $B - B_k$. Choose $n - t$ of these edges and let $B_{k+1} \backslash B_k$ be the vertices of these edges which are elements of B. Let $A_{k+1} = A_k \cup \{a_i \mid (a_i, b_i) \in \pi$ and $b_i \in B_{k+1} \backslash B_k\}$. Since the sequences $A_1, A_2, \ldots, A_{k+1}$ and $B_1, B_2, \ldots, B_{k+1}$ satisfy (1), (2), (3), and (4) we only have to verify property (5). Now if (a, b) is any edge in π with $b \in B_{j+1} \backslash B_j$, $1 \leqslant j \leqslant k$, there is an edge (a^*, b) colored c_j and $a^* \in A_j$. As a consequence, if (a, b) is in π with $b \in B_{j+1} \backslash B_j$, $i \leqslant j \leqslant k$, there exist edges $(a = x_1, b = y_1)$, $(x_2, y_2), \ldots, (x_s, y_s)$ in π such that the edges $(x_2, y_1), (x_3, y_2), \ldots, (x_s, y_{s-1}), (z, y_s)$ are colored $c_j = c_{i_1}, c_{i_2}, \ldots, c_{i_s}$, where $j = i_1 > i_2 > \cdots > i_s \geqslant 1$ and $z \in A_1$. Now suppose $e = (a, c)$ is an edge colored c_m, $k + 1 \leqslant m \leqslant n - t$, with $a \in A_{k+1}$ and $c \in B_1$. By (5) $a \notin A_k$ and so $a \in A_{k+1} \backslash A_k$. If (a, b) is the edge in π containing a, then $b \in B_{k+1} \backslash B_k$. Hence by the above remarks there exist edges $(a = x_1, b = y_1)$, $(x_2, y_2), \ldots, (x_s, y_s)$ in π such that the edges $(x_2, y_1), (x_3, y_2), \ldots, (x_s, y_{s-1})$, (z, y_s) are colored $c_k = c_{i_1}, c_{i_2}, \ldots, c_{i_s}$, where $k = i_1 > i_2 > \cdots > i_s \geqslant 1$ and $z \in A_1$. But then if we remove the s edges $(x_1, y_1), (x_2, y_2), \ldots, (x_s, y_s)$ from π and replace them with the $s + 1$ edges $e = (a, c), (x_2, y_1), (x_3, y_2), \ldots, (x_s, y_{s-1}), (z, y_s)$ we have a set of $t + 1$ parallel edges with mutually distinct colorings. This contradicts the fact that the maximum number of parallel edges with mutually distinct colorings is t and so the extended sequences we have constructed must satisfy (5).

$k \leq n-t$





THEOREM. *Every $n \times n$ latin square has a transversal containing at least $n - \sqrt{n}$ distinct symbols.*

Proof. The sequences A_1 and B_1 as defined above satisfy properties (1), (2), (3), (4), and (5) and therefore can be extended to the sequences $A_1, A_2, \ldots, A_{n-t+1}$ and $B_1, B_2, \ldots, B_{n-t+1}$ satisfying these same conditions. Since $| A_{i+1} \backslash A_i | = n - t$ we have $n \geqslant (n-t)(n-t+1)$, $n \geqslant (n-t)^2 + (n-t)$, $n \geqslant (n-t)^2$, $\sqrt{n} \geqslant n - t$, $| T | = t \geqslant n - \sqrt{n}$.

REFERENCES

1. D. DRAKE, Maximal sets of latin squares and partial transversals, to appear.
2. K. K. KOKSMA, A lower bound for the order of a partial transversal in a latin square, *J. Combinatorial Theory Ser. A* 7 (1969), 94–95.

For reference 2

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Combinatorial Theory, Series A 115 (2008) 1103–1113

Journal of Combinatorial Theory
Series A

www.elsevier.com/locate/jcta

A lower bound for the length of a partial transversal in a Latin square

Pooya Hatami [a], Peter W. Shor [b]

[a] *Department of Mathematical Sciences, Sharif University of Technology, PO Box 11365-9415, Tehran, Iran*
[b] *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Received 29 November 2006

Available online 18 March 2008

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.

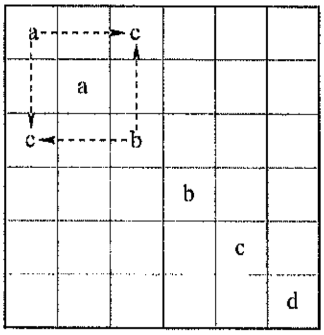
Keywords: Latin square; Partial transversal; Brualdi's conjecture

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 $n - O(\log^2 n)$

E-mail address: shor@math.mit.edu (P.W. Shor).

0097-3165/$ – see front matter © 2008 Published by Elsevier Inc.
doi:10.1016/j.jcta.2008.01.002



Square 1

Fig. 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.

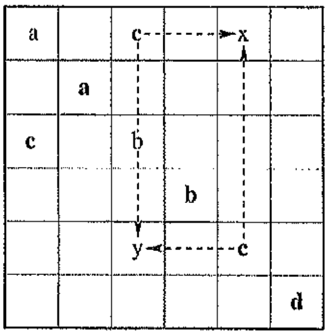
lower bound, albeit with a worse constant than in [4,7]. This is well below Brualdi's conjecture of $n - 1$, and Ryser's of n for odd n [2,6]. The proof in this paper is much the same as in [7] except for the last part of Section 4. The earlier part of the paper has been revised to improve the clarity of the presentation.

2. Operation

Given a partial transversal T of length $n-k$, with $k \geqslant 2$, 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 in Fig. 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 in Square 2 in Fig. 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 in Fig. 2.) 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



Square 2

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

Square 3

Fig. 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 $\{a,b,c,d\}$.

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 d instead of 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 non-empty 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 $k \geqslant 2$. 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 non-empty 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

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

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

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

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 in $\mathcal{T}$ 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.*

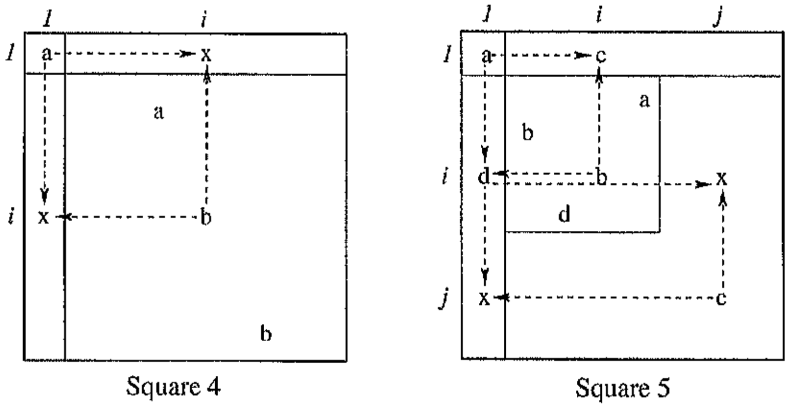


Fig. 3. Square 4: Illustration for the proof of Lemma 2. By applying # as shown, we find the first row contains only non-empty cells, a contradiction. Square 5: Illustration for the proof of Theorem 1. If an element c lies in the first row above the small square and also on the diagonal below and to the right of the small square, we find a non-empty cell x in the first column, below the small square, as shown.

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 in Fig. 3.) 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}.*

Proof. Consider a partial transversal T_0, which we will assume is along the diagonal, and a cell within it, say $(1,1)$, containing a duplicated symbol. Now, hold this cell fixed, and consider the $(n-1)\times(n-1)$ partial Latin square satisfying A_{k-1} generated by the operation # as above. Let us assume that m cells of T_0 are not fixed, and are in rows and columns 2 through $m+1$. Note that this $m\times m$ subsquare satisfies A_{k-1}. By the same reasoning as in the above lemma, there is a transversal with a duplicated symbol in cell (i,i), for all i, $2\leqslant i\leqslant m+1$. Applying #, we find that there is a symbol in cells $(1,i)$ and $(i,1)$, for $2\leqslant i\leqslant m+1$. There are only $m-(k-1)$ symbols in the $m\times m$ subsquare satisfying A_{k-1} containing rows and columns 2 through $m+1$, leaving $(k-1)$ symbols in $(1,i)$, $2\leqslant i\leqslant m+1$, which are not in the subsquare satisfying A_{k-1}. (See Square 5 in Fig. 3.) Note that some of these symbols may appear in the $m\times m$ subsquare in the original partial Latin square, but they do not appear in the set T of partial transversals associated with this subsquare.

Suppose one of these $k-1$ symbols, say c, is in the $(1,i)$ cell. There is at least one c in the original partial transversal T_0, and since it is not in the subsquare, it must be in cell (j,j), for some $j > m+1$. Moreover, there is a partial transversal of the small square with a duplicate letter in cell (i,i), say b. (Note this letter could be a, the same as in cell $(1,1)$, in which case there are three a's in the corresponding partial transversal of the $n \times n$ square). We can now apply # to remove the $(1,1)$ and the (i,i) cells, and we find that the $(1,i)$ and $(i,1)$ cells are filled. Now, the c in the (j,j) cell and the symbol in the $(i,1)$ are both duplicates, so by applying # again 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, $2 \leqslant i \leqslant m+1$, 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, $m \geqslant n_{k-1}$, 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. $\square$

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

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

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 $j<k$,*

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

Proof. Consider Square 6 in Fig. 4. 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 an upper bound for the number of filled cells in P.

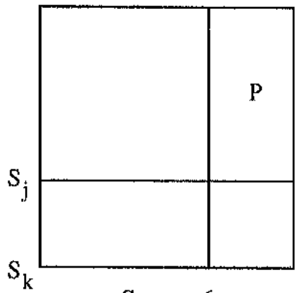


Fig. 4. The region P in the proof of Theorem 2.

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 $n_2 < n_3 < \cdots < n_l$ satisfying the inequalities (1)–(3) from the previous section. Reiterating these inequalities, we have that if $2 \leqslant i \leqslant l$ and $1 \leqslant j < k \leqslant l$, then

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

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

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

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

We first prove the following lemma.

Lemma 3. *Either*

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

or

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

Proof. Letting

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

we have, from (3)

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

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

$$d_j \geqslant \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k). \tag{6}$$

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

$$d_k = n_k - n_{k-1} \leqslant n_k - n_j \leqslant \frac{1}{5}n_k, \tag{7}$$

$$n_k + d_k \leqslant \frac{6}{5}n_k, \tag{8}$$

$$n_k + d_k \leqslant \frac{3}{2}n_j. \tag{9}$$

Combining (6) and (9) gives

$$d_j \geqslant \frac{1}{2}(n_k - n_j). \tag{10}$$

By the definition of d_j, we have

$$n_j - n_{j-1} \geqslant \frac{1}{2}n_k - \frac{1}{2}n_j, \tag{11}$$

so

$$\frac{3}{2}n_j - \frac{3}{2}n_{j-1} \geqslant \frac{1}{2}(n_k - n_{j-1}) \tag{12}$$

and

$$n_j - n_{j-1} \geqslant \frac{1}{3}(n_k - n_{j-1}) \tag{13}$$

completing the proof. □

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

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

giving

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

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

$$1 \leqslant n_k - n_{k-1} \leqslant \left(\frac{2}{3}\right)^{k-j}(n_k - n_{j-1}) \tag{16}$$

or

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

Now, suppose in addition that

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

then

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

$$\frac{n_j}{4} < n_k - n_j, \tag{20}$$

$$\frac{5}{4}n_j < n_k, \tag{21}$$

a contradiction.

If $k - j \geqslant \lfloor \log_{3/2} n_j \rfloor$, 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 $n_k \geqslant \frac{5}{4}n_j$.

We now let $k_4 = 2$, and

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

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

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

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

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

$$k_i \leqslant k_{i-1} + \lfloor \log_{3/2} n_k \rfloor. \tag{24}$$

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

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

$$\frac{1}{c}\log_{3/2} n_k \geqslant k^{\frac{1}{2}} \tag{25}$$

or

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

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

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

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

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

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

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

which rearranges to

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

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

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

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

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

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

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

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

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

from which follows:

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

$$n - 11.053 \log^2 n. \tag{35}$$

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 $n_k = e^{O(k^{1/2})}$.

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, R.M.A. Wieringa, A lower bound for the length of partial transversals in a Latin square, Nieuw Arch. Wiskd. 24 (3) (1978) 330–332.
[2] R.A. Brualdi, 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 (1977) 143–149.
[4] H.L. Fu, S.C. Lin, C.M. Fu, The length of a partial transversal in a Latin square, J. Combin. Math. Combin. Comput. 43 (2002) 57–64.
[5] K.K. Koksma, A lower bound for the order of a partial transversal in a Latin square, J. Combin. Theory Ser. A 7 (1969) 94–95.

[6] H.J. Ryser, Neuere Probleme der Kombinatorik, in: Vorträge über Kombinatorik, Oberwolfach, Matematisches Forschungsinstitute Oberwolfach, Germany, 24–29 July 1967, pp. 69–91.

[7] P.W. Shor, A lower bound for the length of a partial transversal in a Latin square, J. Combin. Theory Ser. A 33 (1) (1982) 1–8.

[8] D.E. Woolbright, An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct elements, J. Combin. Theory Ser. A 24 (1978) 235–237.