

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

0			
1			
			2

0	1		
1			

A critical set PLS(4)

can not be completed uniquely

Fact 2.1. 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.

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

Proof.

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 $c = (1+2+\dots+\lfloor \frac{n}{2} \rfloor) + (1+2+\dots+\lfloor \frac{n-1}{2} \rfloor)$. Now, if n is even,

$$c = 1+2+\dots+\frac{n}{2} + 1+2+\dots+\frac{n-2}{2} = 2(1+2+\dots+\frac{n-2}{2}) + \frac{n}{2}$$

$$= 2 \cdot \frac{1}{2} \cdot \frac{n-2}{2} \cdot \frac{n}{2} + \frac{n}{2} = \frac{n^2-2n}{4} + \frac{n}{2} = \frac{n^2}{4}. \text{ On the other}$$

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.

Conjecture (Critical Set) 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)

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

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

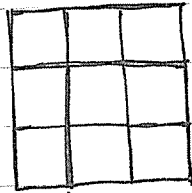
✓ Exercise 1.6. Find $\text{Spec}_C(4)$. (10 points)

(*) 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"?

plaintext $\xleftarrow{\text{Latin Square (Key)}}$ Ciphertext . How?

1. How to use "Latin squares as keys?"



Exercise 1.6' (10 points) Use your imagination!

Exercise 1.6'' (10 points)

How to use critical sets for sharing schemes?

Definition (Sharing schemes)

$\mathbb{P}$: participants

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

$K(\text{keys})$: Latin squares of order n .

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

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

$\downarrow$ reveals the key.

Lecture 4 - Continued

NO. 4
DATE

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 22. 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 23. If L is a Latin square of order n where n is odd, then $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes L$ contains no transversals.

Exercise 1.7. Prove Fact 23. (10 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 1.8. Give examples $L' \wedge$ ^{that} $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes L'$ contains a transverse when L' is a Latin square of even order. (10 points.)
 $\hat{n=2m, m \in \mathbb{N}}$

Ryser's Conjecture

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

Revised version of Ryser's Conjecture

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.

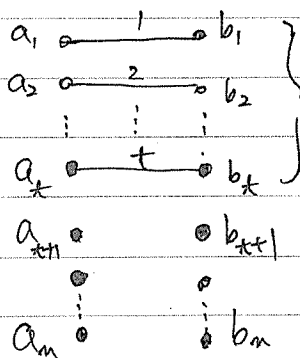
A.E.

Theorem (D. Woolbright and Brouwer)

$$|T_n| \geq n - \sqrt{n}. \quad (\text{Bonus: 10 points for details.})$$

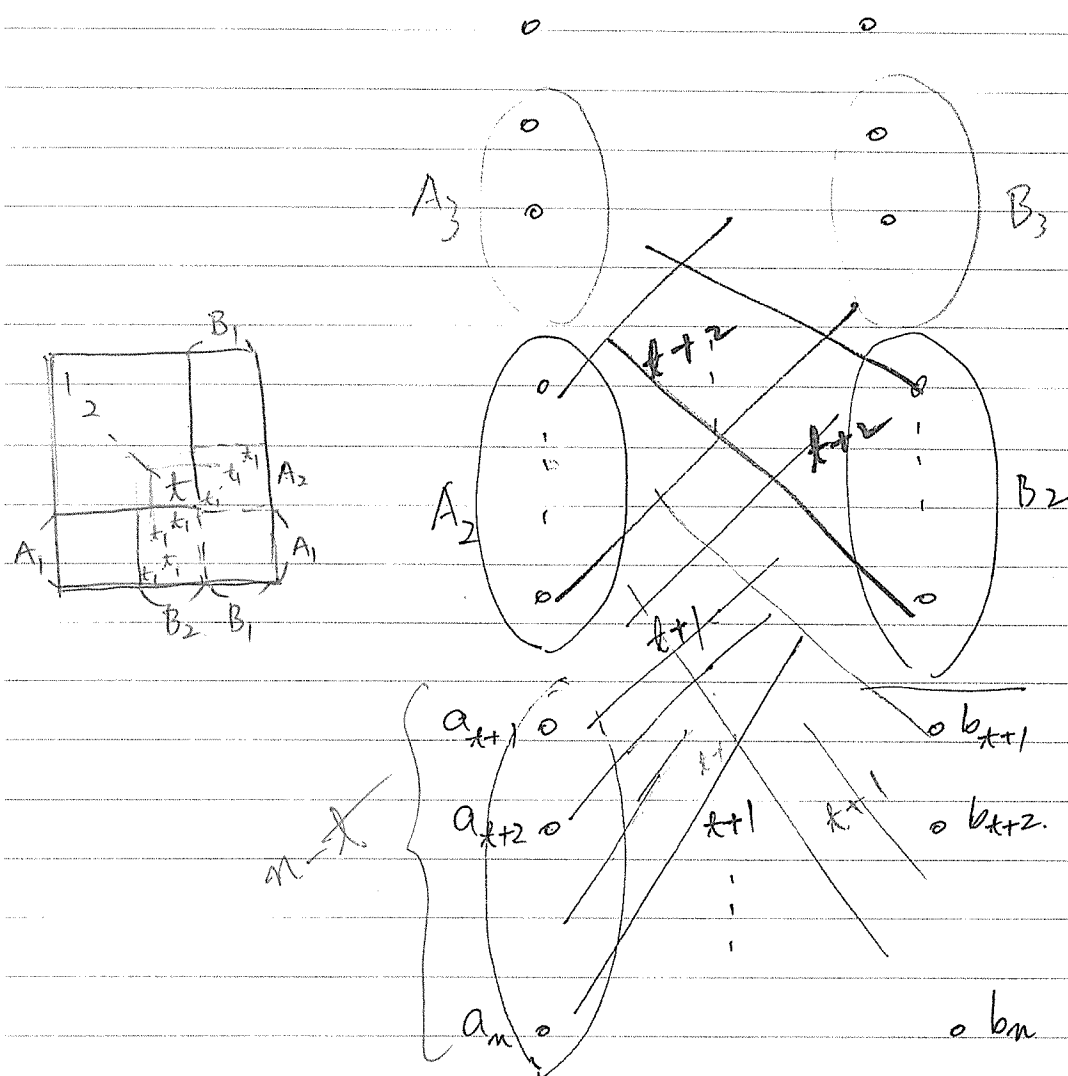
Proof. (Scratch)

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 $\underline{t+1}, \dots, \underline{n}$ in turn.



Sometimes later

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

$$\Rightarrow n-t \leq \sqrt{n}$$

$$\Rightarrow \underline{t \geq n - \sqrt{n}}$$

$$\bullet \quad t \geq \frac{3n}{4} \Rightarrow 4t \geq 3n \Rightarrow t \geq 3(n-t)$$

$$\bullet \quad \text{Let } A_1 = \{a_{t+1}, a_{t+2}, \dots, a_n\} \text{ and } B_1 = \{b_{t+1}, b_{t+2}, \dots, b_n\}.$$

• The color $t+1$ incident to a_i 's in A_1 can not be incident to b_i 's in B_1 , so does the other way around.

(Bonus, 30 points) Give a talk about P. Shor's work on partial transversal.

(JCT(A) 115, 1103-1113, 2008)

0	2	3	①	4	6	7	⑤
3	1	⑦	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 $LS(4\mathbb{Z})$.

Open problem

1. Find an algorithm to determine the maximum size of a partial transversal.
2. Given an $LS(n)$, 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"!

Definition (Transversal)

A transversal of a Latin square consists n cells, one from each row and no two from the same column. If the symbols of the cells (entries) are all distinct, then it's called a Latin transversal.

- Let L be a Latin square of order n , T be a transversal.

Then, $|T|$ is called the size of the transversal where $|T|$ denotes the number of distinct entries in T .

Conjecture (Ryser)

For each Latin square of odd order, L , L consists a Latin transversal. Subsequently, it's conjectured that for each Latin square of order n , L , L consists a transversal of size at least $n-1$.

The progress of proving the conjecture. (On order n .)

(1) $|T| \geq \frac{n}{2}$ is easy to see.?

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

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

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

$$|T| \geq \frac{3n}{4}. \quad (\text{Simpler method})$$

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

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

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

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

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

(6) P. W. Shor (1982, JCTA 33, 1-8)

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

(7) Our result (with 林士强), (2002, JCMCC 43, 57-64)

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

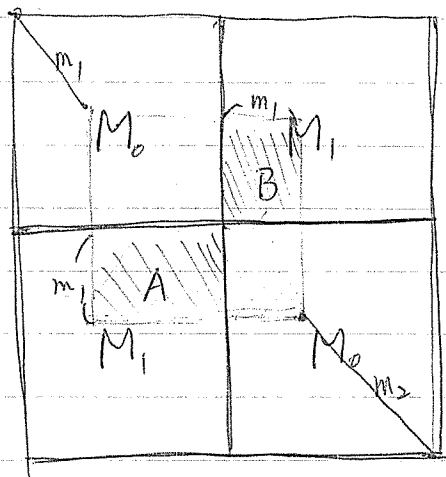
(8) Errata of (6), pooya Hatami and P. W. Shor,
(2008, JCTA 115, 1103-1113)

$$|T| \geq n - O(\ln n)^2. \quad \rightarrow \text{Find a bug in (6).}$$

Non-existence of a Latin transversal

Theorem 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.

Proof.



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, $\begin{matrix} m_1 < m_2, \\ m_1 + m_2 = m. \end{matrix}$ Then,

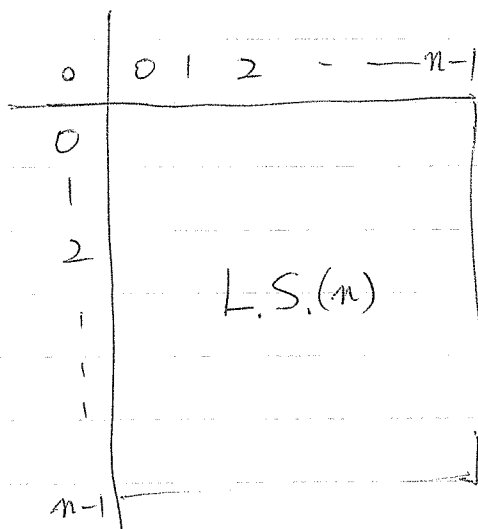
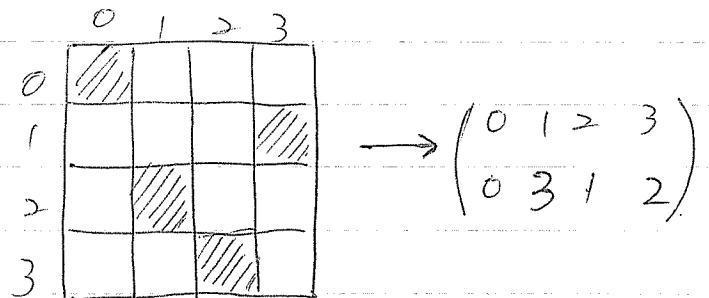
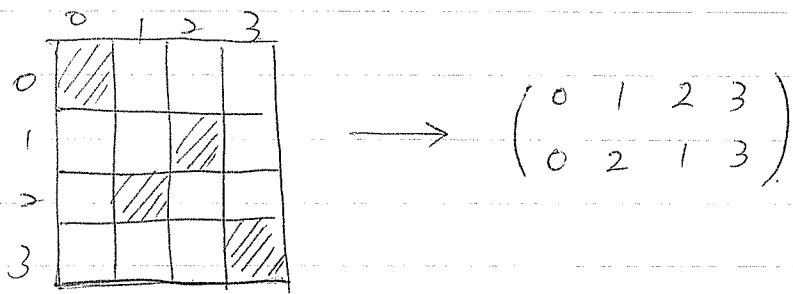
the entries of $m + \mathbb{Z}_m$ must be in A or B . Then, at most $2m$, can

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

be selected, which is smaller than m . ▣

- (*) A transversal in a Latin square can be associated with a bijection of a quasigroup $\langle G, \circ \rangle$, or a permutation of G . (Note that G is finite.)



- (*) Let $f = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix}$ be a permutation of G .

If $g = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1+b_1 & a_2+b_2 & \dots & a_n+b_n \end{pmatrix}$ is also a permutation, then

f is called a complete permutation, also a complete

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

← Finite abelian group

← Cayley Table C

Each transversal of C can be associated with a permutation of $\mathbb{Z}_3$, $f(0)=0$, $f(1)=2$ and $f(2)=1$.

	0	1	2
0	0	1	2
1	2	0	1
2	1	2	0

$f(0)=1$, $f(1)=2$ and $f(2)=0$

(••) If the elements (entries) $a_i + f(a_i)$, $1 \leq i \leq n$ form a permutation of G of order n, then the n cells contain n distinct elements and the Transversal is a Latin Transversal.

(•••) In a group, we may add the same element to obtain another Latin Transversal if we already have one.

$$g(a_i) = a_i + f(a_i)$$

$$g_1(a_i) = (b_1) + f(a_i) \quad (b_1 = a_i + c, \quad b_j = a_i + c_j)$$

(*) It is interesting to know which Cayley table of an abelian group consists of a Latin transversal.

Another example

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

$\mathbb{Z}_6$

Does $\nearrow$ consist a Latin transversal?

Yes, how; No, why?

How about $\mathbb{Z}_{2m}$ in general?



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

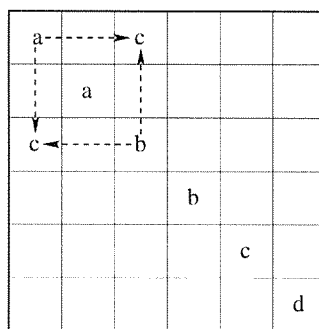
It is proved that every $n \times n$ Latin square has a partial transversal of length at least $n - O(\log^2 n)$. The previous papers proving these results (including one by the second author) not only contained an error, but were sloppily written and quite difficult to understand. We have corrected the error and improved the clarity. © 2008 Published by Elsevier Inc.

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).



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 $k \geq 2$, 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 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

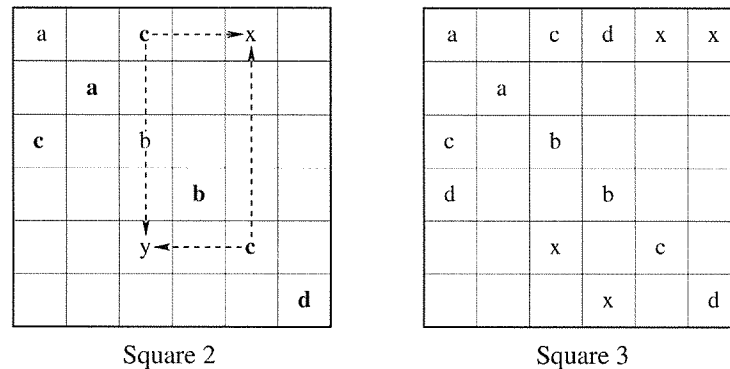


Fig. 2. Square 2: Another example of the operation $\#$. In this, we replace the cells $(1, 3)$ and $(5, 5)$ in the partial transversal indicated in bold with the cells containing symbols x and y . If this Latin square had no partial transversal of length greater than 4, we must have that $x \in \{b, d\}$ 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 $k \geq 2$. 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 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

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

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

$$n_2 \geq 7. \quad (1)$$

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. $\square$

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.*

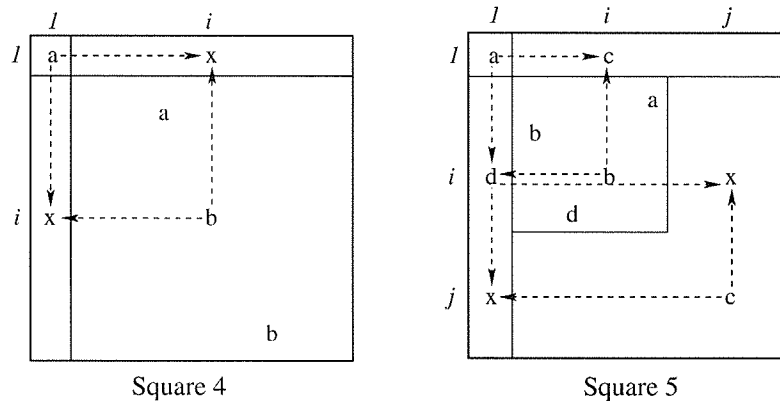


Fig. 3. Square 4: Illustration for the proof of Lemma 2. By applying # as shown, we find the first row contains only non-empty cells, a contradiction. Square 5: Illustration for the proof of Theorem 1. If an element c lies in the first row above the small square and also on the diagonal below and to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

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 , $2 \leq i \leq m+1$. Applying #, we find that there is a symbol in cells $(1, i)$ and $(i, 1)$, for $2 \leq i \leq m+1$. 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 $(1, i)$, $2 \leq i \leq m+1$, 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 $\mathcal{T}$ 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, $2 \leq i \leq m + 1$, 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. $\square$

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

$$n_k \geq n_{k-1} + 2k, \quad (2)$$

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 .

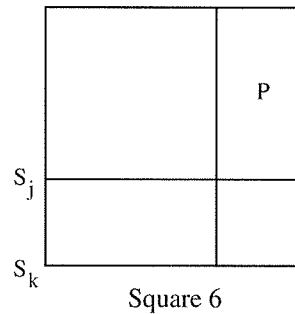
Theorem 2. In S_k , as defined above, for all $j < k$,

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

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. $\square$

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)–(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

$$n_2 \geq 7, \quad (1)$$

$$n_i \geq n_{i-1} + 2i, \quad (2)$$

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

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

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

we have, from (3)

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

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 $n_j \geq \frac{4}{3}n_k$. 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. $\square$

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). $\square$

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

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, Mathematisches Forschungsinstitut 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.

The Length of a Partial Transversal in a Latin Square

Hung-Lin Fu* and Shyh-Chung Lin

Department of Applied Mathematics
Nation Chiao Tung University
Hsin Chu, Taiwan, R.O.C.

Chin-Mei Fu

Department of Mathematics
Tankang University
Tamsui, Taipei Hsein, Taiwan, R.O.C.

Abstract

A latin square of order n is an $n \times n$ array of cells containing one of the n elements in $\{1, 2, \dots, n\}$ such that in each row and each column each element appears exactly once. A partial transversal P of a latin square L is a set of n cells such that no two are in the same row and the same column. The number of distinct elements in P is referred to as the length of P , denoted by $|P|$, and the maximum length of a partial transversal in L is denoted by $t(L)$. In this paper, we study the technique used by Shor which shows that $t(L) \geq n - 5.53(\ln)^2$ and we improve the lower bound slightly by using a more accurate evaluation.

1 Introduction

A *latin square* of order n is an $n \times n$ array of cells containing one of n distinct symbols such that in each row and column every symbol appears exactly one. A *transversal* of a latin square of order n is a set of n cells, one in each row, one in each column, and no two of them contain the same symbol. If we simply select a set of n cells in a latin square of order n , one in each row and one in each column, then we have a partial transversal P . The

*Research supported by National Science Council of the Republic of China (NSC-85-2121-M-009-010).

number of distinct elements in P is referred to as the length of P , denoted by $|P|$. We are interested in finding a partial transversal in a given latin square which has the maximum length. For convenience, let $t(L) = \max\{|P| \mid P \text{ is a partial transversal in } L\}$. In 1967, Ryser[1] conjectured that $t(L) \geq n - 1$ if L is a latin square of even order n and $t(L) = n$ otherwise. In 1969, Koksma[2] showed that $t(L) > (2n + 1)/3$, then Drake[3] improved this lower bound to $3n/4$ in 1977. In 1978, this lower bound was increased to $n - \sqrt{n}$ by Brouwer *et al.* [4] and Wollbright[5] independently. Later in 1982, Shor[6] gave a better bound for $n \geq 2,000,000$, namely, $n - 5.53(\ln n)^2$. So far, this is considered as the best known lower bound. In this paper, we apply an elementary argument using calculus to obtain a better lower bound $n - 5.51(\ln n)^2$.

2 Improvement of Shor's lower bound

First, we introduce the key idea of Shor's approach in finding the lower bound for $t(L)$. Let L be a latin square of order n and P be a partial transversal which has length at most $n - 2$. For otherwise, there is no room for improvement. Then there exist at least two cells of P , (i_1, j_1) and (i_2, j_2) , such that $L(i_1, j_1) = L(i', j')$ and $L(i_2, j_2) = L(i'', j'')$, for some (i', j') and (i'', j'') in $P \setminus \{(i_1, j_1), (i_2, j_2)\}$. Clearly, if we let $P' = (P \setminus \{(i_1, j_1), (i_2, j_2)\}) \cup \{(i_1, j_2), (i_2, j_1)\}$, then $|P'| \geq |P|$. The operation obtained above will be called the operation $\sharp$. See Figure 1.

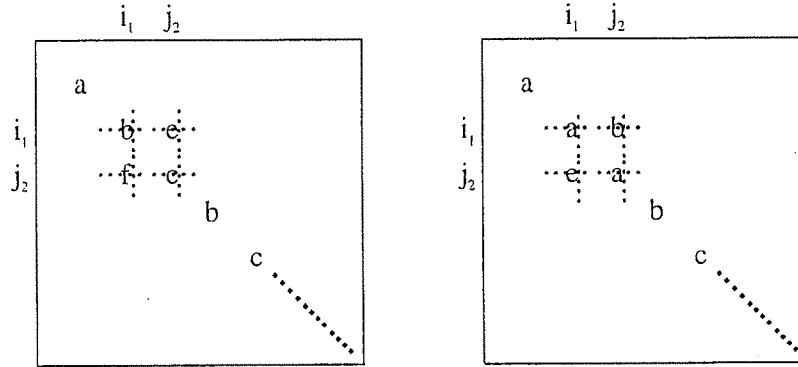


Figure 1. Operation $\sharp$

Now if we start with a partial transversal P of maximum length $n - k$, then by applying operation $\sharp$ to the partial transversal, and then to another partial transversal, and so on, we obtain a set of partial transversals closed under $\sharp$, i.e., no other cells can be added by using operation $\sharp$.

Since P is of maximum length $n - k$, we obtain a partial latin square of order n and only $n - k$ distinct elements are used as entries. We shall call this partial latin square A_k , i.e.,

A_k : The partial latin square contains only $n - k$ distinct elements, and these elements are exactly those contained in the cells of a set of partial transversals obtained by $\sharp$ and closed under $\sharp$.

Then the following lemma is not difficult to see.

Lemma 2.1.[6] *Given a partial latin square L satisfying A_k such that no proper subsquare satisfies A_k , then no cell is contained in all partial transversals.*

The above lemma shows that every cell in the partial latin square L gets moved. Without loss of generality, we may let $(1, 1)$ be a filled cell and $L(1, 1) = a$. Now if we fix the cell $(1, 1)$ and consider the set of partial transversals generated by $\sharp$ acting on the subsquare formed by deleting the row and the column containing $(1, 1)$ cell, then we have an $(n - 1) \times (n - 1)$ partial latin square L' satisfies A_{k-1} . Note that we have $n - k$ elements in the partial latin square L' of order $n - 1$. Now it is not difficult to see that L' must have at least one fixed position. Suppose not. Since $(1, 1)$ can be moved (by Lemma 2.1, and for each filled (i, j) in L' can also be moved, we conclude that $(1, i)$ must be filled. By the fact that i is an arbitrary index and each column contains at least one filled cell of L' , the first row of L must be completed. This implies that L contains n distinct elements. But, by assumption, L contains exactly $n - k$ elements. Hence, we conclude that L' contains some fixed positions.

Now we have the following result which was obtained in [6].

Lemma 2.2.[6] *Let L be a partial latin square of smallest order satisfying A_k . Then there are at least $n_{k-1} + k$ filled cells in each row and each column, where n_{k-1} is the order of smallest subsquare satisfying A_{k-1} .*

Let L_k be a partial latin square of smallest order which satisfies A_k . Then define L_i , $2 \leq i < k$, recursively as a smallest subsquare of L_{i+1} which satisfies A_i , and the order of L_j is n_j $j = 1, 2, \dots, k$. Again the following inequality was obtained by Shor.

Lemma 2.3.[6] *In L_k , as defined above,*

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

for all $j < k$.

Now we can derive a lower bound for $t(L)$ using the inequality in Lemma 2.3 and the ratio $\frac{n_j}{n_k}$.

Theorem 2.4. *Let L be a latin square of order n . Then*

$$t(L) \geq n - \frac{(\ln n)^2}{2 \ln \frac{4q-2p}{q} \ln \frac{p}{q}}, \quad \text{where } \frac{1}{2} < \frac{q}{p} < 1.$$

Proof. We shall use the notations mentioned above. By Lemma 2.3, we have the following inequality.

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

Since $n_k - n_{k-1} = d_k$ and $n_j - n_{j-1} = d_j$,

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

and then

$$d_j - 2j = n_j - n_{j-1} - 2j \geq \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k + 2k - j). \quad (1)$$

Consider n_j and n_k where $j < k$, and $1/2 < \frac{p}{q} < 1$. Then either there exist j and k such that $n_j \geq \frac{p}{q}n_k$ or $n_j \leq \frac{p}{q}n_k$ for all j and k . First if $n_j \geq \frac{p}{q}n_k$ for some j and k . Then

$$\begin{aligned} d_k = n_k - n_{k-1} &\leq n_k - n_j \leq \frac{p-q}{p}n_k, \\ n_k + d_k &\leq \frac{2p-1}{p}n_k, \\ n_k + d_k &\leq \frac{2p-q}{p}(\frac{p}{q}n_j), \\ n_j &\leq \frac{p}{2p-q}(n_k + d_k). \end{aligned} \quad (2)$$

From (1) and (2)

$$\begin{aligned} d_j &\geq \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k), \\ d_j &\geq (n_k - n_j)(2 - \frac{d_k + n_k}{n_j}), \\ d_j &\geq (n_k - n_j)(2 - \frac{2p-q}{q}), \\ d_j &= (n_j - n_{j-1}) \geq (\frac{3q-2p}{q}(n_k - n_j)). \end{aligned}$$

We add $\frac{3q-2p}{q}(n_j - n_{j-1})$ to both sides of the above inequality, then

$$\frac{4q-2p}{q}(n_j - n_{j-1}) \geq \frac{3q-2p}{q}(n_k - n_{j-1}),$$

and hence

$$n_j - n_{j-1} \geq \frac{3q-2p}{4q-2p}(n_k - n_{j-1}). \quad (3)$$

Therefore

$$\begin{aligned} (n_k - n_{j-1}) - (n_j - n_{j-1}) &\leq (n_k - n_{j-1}) - \frac{3q-2p}{4q-2p}(n_k - n_{j-1}), \\ n_k - n_j &\leq \frac{q}{4q-2p}(n_k - n_{j-1}). \end{aligned}$$

Since $n_j \leq \frac{q}{p}n_k$, hence for each $k > j' > j$, $n_{j'} \geq \frac{q}{p}n_k$. By induction:

$$\begin{aligned} 1 \leq n_k - n_{k-1} &\leq \left(\frac{p}{4q-2p}\right)^{k-j}(n_k - n_{j-1}), \\ \left(\frac{4q-2p}{q}\right)^{k-j} &\leq n_k - n_{j-1}, \\ k-j &\leq \log_{\frac{4q-2p}{q}}(n_k - n_{j-1}). \end{aligned} \quad (4)$$

Now if

$$k-j \geq \log_{\frac{4q-2p}{q}}\left(\frac{p-q}{q}n_{j-2}\right),$$

then

$$\begin{aligned} \log_{\frac{4q-2p}{q}}(n_k - n_{j-1}) &\geq \log_{4q-2p} q \left(\frac{p-q}{q}n_{j-1}\right), \\ n_k - n_{j-1} &\geq \frac{p-1}{q}n_{j-1}, \\ n_{j-1} &\leq \frac{q}{p}n_k. \end{aligned}$$

And if

$$k-j \geq \log_{\frac{4q-2p}{q}}(n_j),$$

by (4), we have

$$\begin{aligned} k-j-1 &\geq \log_{\frac{4q-2p}{q}}(n_k - n_j), \\ k-j &\geq \log_{\frac{4q-2p}{q}}(n_k - n_j) + 1, \\ k-j &\geq \log_{\frac{4q-2p}{q}}(n_k - n_j) \left(\frac{4q-2p}{q}\right). \end{aligned}$$

So we have

$$\begin{aligned} n_j &\leq \frac{4q-2p}{q}(n_k - n_j), \\ \frac{5q-2p}{q}n_j &\leq \frac{4q-2p}{q}n_k, \\ n_k &\leq \frac{5q-2p}{4q-2p}n_j. \end{aligned}$$

Since $\frac{5q-2p}{4q-2p} \geq \frac{p}{q}$,

$$n_j \leq \frac{q}{p} n_k.$$

Now we obtain a recurrence relation of k_i where k_i is the index of n .
Let $k_4 = 2$ and

$$k_i = k_{i-1} + \lceil \log_{\frac{4q-2p}{q}} (n_{k_{i-1}}) \rceil. \quad (5)$$

This implies that

$$k_i \geq k_{i-1} + \log_{\frac{4q-2p}{q}} (n_{k_{i-1}}).$$

Since $n_{k_i} \geq (\frac{p}{q})^{i+1}$, then from (5), by direct counting, we have

$$\begin{aligned} k^i &\geq \sum_{j=1}^i \log_{\frac{4q-2p}{q}} \left(\frac{p}{q}\right)^j = \sum_{j=1}^i j \log_{\frac{4q-2p}{q}} \left(\frac{p}{q}\right), \\ k^i &\geq \frac{1}{2} i(i+1) \log_{\frac{4q-2p}{q}} \left(\frac{p}{q}\right) \geq \frac{i^2}{2} \frac{\ln \frac{p}{q}}{\ln \frac{4q-2p}{q}}. \end{aligned}$$

Thus

$$\begin{aligned} n_k &\geq \left(\frac{p}{q}\right)^{i+1}, \quad \text{where } i = \lfloor (2 \frac{\ln \frac{4q-2p}{q}}{\ln \frac{p}{q}} k)^{1/2} \rfloor, \\ \ln n_k &\geq (2 \ln \frac{4q-2p}{q} \ln \frac{p}{q})^{1/2} k^{1/2}, \\ k &\leq \frac{(\ln n)^2}{2 \ln \frac{4q-2p}{q} \ln \frac{p}{q}}. \end{aligned} \quad (6)$$

On the other hand, if there exists no j such $n_j \geq \frac{p}{q} n_k$, then

$$\begin{aligned} n_k &\geq \frac{p}{q} n_{k-1} \\ n_{k-1} &\geq \frac{p}{q} n_{k-2} \end{aligned}$$

and so on. Hence we have

$$n_k \geq \left(\frac{p}{q}\right)^{k-2} n_2.$$

Thus,

$$\begin{aligned} \left(\frac{p}{q}\right)^{k-2} &\leq \frac{n}{n_2}, \\ k-2 &\leq \log_{\frac{p}{q}} \frac{n}{n_2}, \\ k &\leq \log_{\frac{p}{q}} \frac{n}{n_2} + 2. \end{aligned} \quad (7)$$

Since

$$n - k \geq \min\left\{n - \frac{(\ln n)^2}{2 \ln \frac{4p-2q}{q} \ln \frac{p}{q}}, n - \log_{\frac{p}{q}} \frac{n}{n_2} + 2\right\},$$

we conclude that

$$t(L) \geq n - \frac{(\ln n)^2}{2 \ln \frac{4p-2q}{q} \ln \frac{p}{q}}.$$

□

We note here that the lower bound for $t(L)$ obtained by Shor can be obtained by letting $\frac{p}{q} = \frac{4}{5}$.

Now assume that $\frac{p}{q} = x$, then $t(L) \geq n - \frac{(\ln n)^2}{2 \ln(4-2/x) \ln(1/x)}$. In order to obtain a better lower bound for $t(L)$, we have to minimize $f(x)$, where $f(x) = \frac{1}{2(\ln(4-2/x) \ln \frac{1}{x})}$. Now consider $\{x | \frac{2}{3} < x < 1\}$.

$$\begin{aligned} f'(x) &= \frac{1}{\ln(4-2/x) \ln x} \left[\frac{1}{(4-2/x)x \ln(4-2/x)} + \frac{1}{2 \ln x} \right], \text{ and} \\ f'(x) &= -\frac{\frac{1}{4}}{\ln(4-2/x)^3 (\ln x) x^4 (4-2/x)^2} - \frac{\frac{1}{2}}{(4-2/x)^2 (\ln x)^2 x^3 (4-2/x)} \\ &\quad - \frac{\frac{1}{2}}{\ln(4-2/x)^2 (\ln x) x^3 (4-2/x)} - \frac{\frac{1}{2}}{(4-2/x)^2 (\ln x) x^4 (4-2/x)^2} \\ &\quad - \frac{\frac{1}{\ln(4-2/x) (\ln x)^3 x^2}}{\ln(4-2/x) (\ln x)^3 x^2} - \frac{\frac{1}{(4-2/x) (\ln x)^2 x^2}}{(4-2/x) (\ln x)^2 x^2}. \end{aligned}$$

By direct counting, we obtain a local minimum at $x \approx 0.793921$. Subsequently $f(x) \approx 5.518427$ which gives the local minimum. Thus we have

Proposition 2.5. *Let L be a latin square of order n , then $t(L) \geq n - 5.518427(\ln n)^2$.*

Finally, we remark here that the idea of improving the lower bound using the above technique (changing $\frac{p}{q}$) has been mentioned in Shor's paper, but as we know no one has tried that so far. In this paper, we complete the research in this direction and we conclude that his method can only improve the lower bound for $t(L)$ a little bit further, and the result obtained in this paper should be the best possible.

References

- [1] H. J. Ryser, Neuere Probleme der Kombinatorik, in "Vorträge über Kombinatorik", pp. 69-91, a collection of papers presented at the Mathematischen Forschungsinstitute, Oberwolfach, July 24-29, 1967.

- [2] K. K. Koksma, A lower bound for the order of a partial transversal in a latin square, *J. Theory Ser.* **7** (1969), 94–95.
- [3] D. A. Drake, Maximal sets of Latin square and partial transversal, *J. Statist. Plann. Inference* **1** (1977), 143–149.
- [4] A. K. Brouwer, A. J. de Vries, and R. M. A. Wieringa, A lower bound for the length of partial transversal in a latin square, *Nieuw Arch. Wisk.* (3) **24** (1978) 330–332.
- [5] D. E. Woolbright, An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ (distinct symbols), *J. Combin. Theory Ser. A* **24** (1978), 235–237.
- [6] P. W. Shor, A lower bound for the length of a partial transversal in a latin square, *J. Combin. Theory Ser. A* **33** (1982), 1–8.