82

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 $\{1, 2, \ldots, n\}$ 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 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 clements 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 $t(L) = \max\{|P| \mid P$ 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.

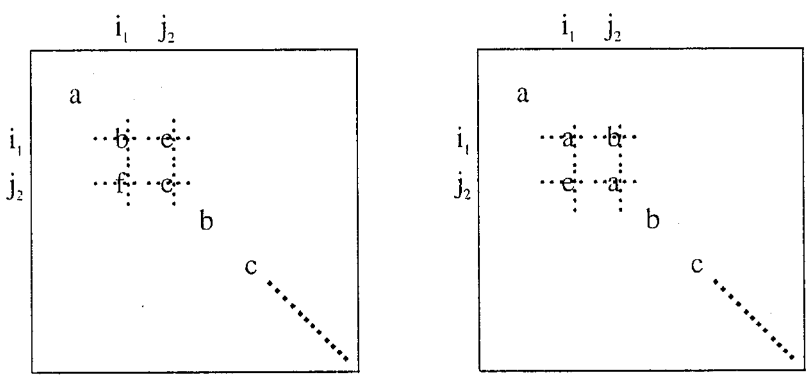

Figure 1. Operation $\sharp$

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-l}. 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 hilled. 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 $j = 1, 2, \ldots, k$. 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*

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

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 $n_j \geq \frac{q}{p}n_k$ for some j and k. Then

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

From (1) and (2)

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

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

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

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:

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

Now if

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

then

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

And if

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

by (4), we have

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

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

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

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

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

Thus

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

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

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

Thus,

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

Since

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

we conclude that

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

□

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

Now assume that $\frac{q}{p} = x$, 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 $f(x) = \frac{1}{2(\ln(4-\frac{2}{x})(\ln\frac{1}{x})})$. Now consider $\{x \,|\, \frac{2}{3} < x < 1\}$.

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

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 $\frac{q}{p}$) 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, Oberwollach, July 24–29, 1967.

[2] K. K. Koksma, A lower bound for the order of a partial transversal in a latin squai, *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 $n - \sqrt{n}$ (listinct 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.