JOURNAL OF COMBINATORIAL THEORY, Series A 33, 1–8 (1982)

A Lower Bound for the Length of a Partial Transversal in a Latin Square*,†

P. W. Shor

California Institute of Technology, Pasadena, California 91125

Communicated by the Managing Editors

Received June 26, 1981

It is proved that every $n \times n$ Latin square has a partial transversal of length at least $n - 5.53(\log n)^2$.

1. Introduction

A Latin square 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 positions are added). Koksma [3] showed that an $n \times n$ Latin square has a partial transversal of length at least $(2n+1)/3$. This was improved by Drake [2] to $3n/4$, and then simultaneously by Brouwer *et al.* [1] and by Woolbright [5] to $n - \sqrt{n}$. This paper will prove a lower bound of $n - c(\log n)^2$, where $c \approx 5.53$ (this bound is sharper than $n - \sqrt{n}$ for $n \geqslant 2{,}000{,}000$), and also give a recursive inequality which can be used to compute a bound sharper than $n - \sqrt{n}$ for much lower values of n. This is still well below Ryser's conjecture of $n - 1$, and n for odd n [4].

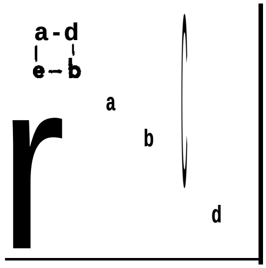
2. The Operation

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 duplicated symbols in T, in the 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. Replace

* This research was supported in part by a Richter Fellowship from Caltech.

† Present address: Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

0097-3165/82/040001-08$02.00/0



SQUARE 1

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 positions (see square 1). Call this operation $\neq$.

Consider a Latin square with a partial transversal of maximum length $n - k$, with $k \geqslant 2$. By applying $\neq$ to this partial transversal, we get other partial transversals, whose length must also be n -k and whose symbols must be the same as the first. Continuing in this manner, we obtain a set of partial transversals closed under #. All of these partial transversals contain only a certain set of n -k distinct symbols, so by ignoring all positions except those in this set of partial transversals, we obtain a partial Latin square (a partially tilled $n \times n$ array such that in each row and column no symbol appears more than once) with the following properties (we call this a partial Latin square satisfying A,):

A,: ***The partial Latin square contains only n k distinct symbols, and these symbols are exactly those contained in the cells of a set of partial transversals of length n − k generated from a single partial transversal by $\neq$ and closed under #.***

LEMMA. ***Given a partial Latin satisfying A, such that no subsquare satisfies A,, then no cell is contained in all partial transversals; i.e., given a cell (i, j) and a partial transversal T containing (i, j), by a sequence of operations $\neq$, one can obtain a partial transversal not containing (i, j) from T.***

Proof: Suppose there is a fixed position containing the symbol ***a.*** If ***a*** appears anywhere else in the partial Latin square, then by applying # to the transversal with two ***a's*** (there must be a transversal with the second ***a,*** since it is in the square, and this transversal must also contain the first ***a*** since all partial transversals do), one obtains a partial transversal without the fixed position, a contradiction. If ***a*** does not appear anywhere else in the partial Latin square, then by deleting the row and the column containing the ***a,*** one finds a subsquare satisfying A_k, a contradiction of the hypothesis. Thus no cell can be fixed.

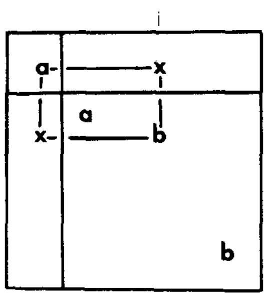
We have just proved that every cell in a transversal gets moved, so given a filled cell in the square, there is a partial transversal containing that cell and another cell with the same symbol. Choose any filled cell, say (1, 1), and choose a partial transversal through it that duplicates the symbol in it, say ***a.*** Now hold the position (1, 1) fixed, and consider the set of partial transversals containing two ***a's*** generated from that transversal by #, i.e., the set of partial transversals generated by # acting on the subsquare formed by deleting the first two and column. These will give an $(n-1) \times (n-1)$ partial Latin square satisfying A_{k-1}.

LEMMA. ***In this $(n-1) \times (n-1)$ square the transversals generated by # must have a fixed position.***

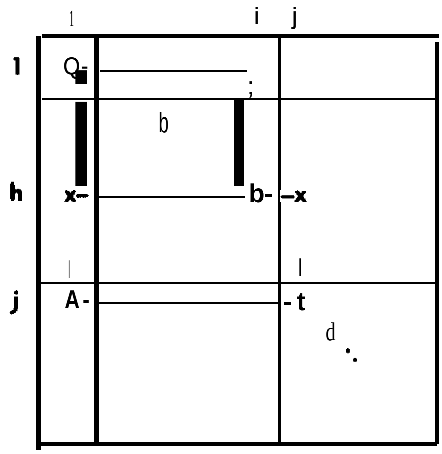
Proof Suppose they do not. Then every cell must be moved, so there is a transversal which duplicates the symbol in any positions of the original transversal. Thus by applying # to this transversal, deleting the fixed element a in (1, 1) and the cell in the ith column of the original transversal, we obtain a transversal of the original square with position (1, ***i)*** filled (see square 2). Since ***i*** was arbitrary,.this gives us n filled cells in the first row of the square, a contradiction since there are only $n-k$ distinct sumbols. Thus by fixing (1, 1), a certain number of other cells of the transversal must also become fixed.

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

Proof. Consider a transversal with a fixed position, say (1, 1), and consider the subsquare generated by this as above. Now, say ***m*** cells of this transversal move and are in rows and columns ***2-m*** + 1. By the same



SQUARE 2



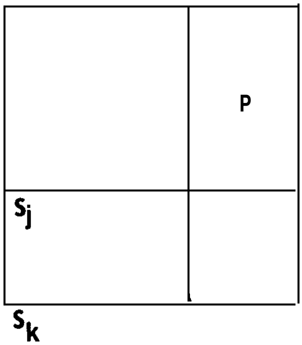
SQUARE 3

reasoning as in the above lemma there is a transversal with a duplicated symbol in column i, for all i, $2 \leqslant i \leqslant m + 1$. Applying #, we find that there is a symbol in positions $(1, i)$, for $2 \leqslant i \leqslant m + 1$. Similarly, the cells $(i, 1)$, $2 \leqslant i \leqslant m + 1$, are filled. There are $m - (k \quad 1)$ symbols in the small square, leaving $k - 1$ symbols in $(1, i)$, $2 \leqslant i \leqslant m + 1$, which are not in the small square (see square 3). Suppose one of these symbols, say c, is in the $(1, i)$ position. There is a c in the original transversal. Since it is not in the small square, which contains all the moving positions, it must be in one of the fixed positions. Say the c is in (j, j). Moreover, there is a transversal of the small square with a duplicate letter, say b, in the ith column, say in (h, i).

Apply # to remove the (1, 1) and the (k, i) positions and till the $(1, i)$ and $(k, 1)$ positions. Now, c in the (j, j) position and the symbol in the $(k, 1)$ position are both duplicates, so by applying # again we can fill the $(j, 1)$ and (k, j) positions. Thus, the $(j, 1)$ position is tilled. Since there are at least $k - 1$ symbols in the $(1, i)$ cells, $2 \leqslant i \leqslant m + 1$, which are not in the small square, we can apply the same process to obtain $k - 1$ tilled positions in the first column below the $(m + 2)$nd row. This gives at least $m + k$ filled positions in the first column, since the first $m + 1$ positions are tilled. Now, $m \geqslant n_{k-1}$, because m is the size of a subsquare satisfying A,, and n_{k-1} was the minimal such subsquare.

3. AN INEQUALITY

Let S_k be a square satisfying A, such that no subsquare satisfies A,. Choose S_{k-1} to be the smallest subsquare of S_k satisfying A_{k-1}, S_{k-2} the



SQUARE 4

smallest subsquare of S_{k-1} satisfying A_{k-2}, and, in general, S_m the smallest subsquare of S_{m+}, satisfying A_{m}, until the sequence ends at S_2, Let n_j be the size of S_j.

THEOREM. *In S_k, as defined above, for all $j < k$,*

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

Proof. Consider square 4. We will count the number of filled cells in the rectangle P in two different ways. First, there are $n_k \quad n_j$ columns in P, and since each column of S_k has $n_{k-1} + k$ filled positions, and there are $n_k - n_j$ columns in P, we have $n_{k-1} + n_j - n_k + k$ filled positions in each column of P, and at least $(n_k - n_j)(n_{k-1} + n_j - n_k + k)$ filled positions in P.

We will call the symbols in S_j ***old*** symbols and those 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 there are at least $n_{j-1} +$ j old symbols in S. 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, or at most $n_j(n_j - n_{j-1} - 2j)$ 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 $(n, - n_j)(n_k$ -k $- n_j +$ j) new symbols in P. Adding the number of old and new symbols, we get a maximum for the number of symbols in P. Setting this maximum greater than or equal to the minimum, we obtain the inequality stated above.

4. THE MAIN RESULT

We will now derive the inequality $k \leqslant 5.53(\log n_k)^2$ from the inequality obtained in the previous theorem. We have a sequence $n_2, n_3, \ldots, n_k$ with the inequality (1) holding between any four elements n_{i-1}, n_i, n_{j-1}, n_j with $3 \leqslant i < j \leqslant k$. From (1),

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

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

Let

$$n_k - n_{k-1} = d_k, \qquad n_j - n_{j-1} = d_j. \tag{3}$$

Then

$$d_j - 2j \leqslant \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k + 2k - j). \tag{4}$$

Now we will assume

$$n_j \geqslant \tfrac{4}{5} n_k, \tag{5}$$

$$d_k = n_k - n_{k-1} \leqslant n_k - n_j \leqslant \tfrac{1}{5} n_k, \tag{6}$$

$$n_k + d_k \leqslant \tfrac{6}{5} n_k, \tag{7}$$

$$n_j \geqslant \tfrac{2}{3}(n_k + d_k). \tag{8}$$

Thus, from (4),

$$d_j \geqslant \frac{n_k - n_j}{n_j}(2n_j - d_k - n_k). \tag{9}$$

Applying (8), we get

$$d_j \geqslant \tfrac{1}{2}(n_k - n_j), \tag{10}$$

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

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

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

Thus, we have shown the following.

LEMMA. *Either*

$$n_j \leqslant \tfrac{4}{5} n_k \tag{14}$$

or

$$n_j - n_{j-1} \geqslant \tfrac{1}{3}(n_k - n_{j-1}). \tag{15}$$

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

$$n_j - n_{j-1} \geqslant \tfrac{1}{3}(n_k - n_{j-1}),$$
$$n_k - n_j \leqslant \tfrac{2}{3}(n_k - n_{j-1}). \tag{16}$$

By induction, we get, since this holds for all j, k, with j $< k$ and $n_j \geqslant \frac{4}{5} n_k$,

$$1 \leqslant n_k \quad n_{k-1} \leqslant (\tfrac{2}{3})^{k-j}(n_k - n_{j-1}), \tag{17}$$
$$(\tfrac{3}{2})^{k-j} \leqslant (n_k - n_{j-1}), \tag{18}$$
$$k \quad j \leqslant \log_{3/2}(n_k - n_{j-1}). \tag{19}$$

Now, if

$$k - j \geqslant \log_{3/2} \frac{n_{j-1}}{4}, \tag{20}$$

then

$$\log,,,\left(\frac{n_{j-1}}{4}\right) \leqslant \log_{3/2}(n_k - n_{j-1}), \tag{21}$$
$$\frac{n_{j-1}}{4} \leqslant n_k - n_{j-1}, \tag{22}$$
$$\tfrac{5}{4} n_{j-1} \leqslant n_k. \tag{23}$$

So if $k -$ j $\geqslant \log_{3/2}(n_j)$, then $\frac{5}{4} n_j \leqslant n_k$.

Now let $k_4 = 2$, and

$$k_i = k_{i-1} + \log_{3/2}(n_{k_{i-1}}). \tag{24}$$

Using the above lemma and induction, we obtain

$$n_j \geqslant (\tfrac{5}{4})^{i+1} \qquad \text{for } j \geqslant k_i. \tag{25}$$

By induction we have, from (24) and (25),

$$k_i \geqslant \sum_{j=1}^{i} \log_{3/2}\left(\frac{5}{4}\right)^j, \tag{26}$$
$$k_i \geqslant \frac{1}{2} j(j+1) \log_{3/2} \frac{5}{4} > \frac{i^2}{2} \frac{\log(5/4)}{\log(3/2)}. \tag{27}$$

Thus,

$$n_k \geqslant \left(\frac{5}{4}\right)^{1+i}, \qquad \text{where} \quad i = \left(2\,\frac{\log(3/2)}{\log(5/4)}\,k\right)^{1/2}, \tag{28}$$

$$\log n_k \geqslant \left(2 \log \frac{3}{2} \log \frac{5}{4}\right)^{1/2} k^{1/2}, \tag{29}$$

$$k \leqslant \frac{(\log n)^2}{2 \log(3/2) \log(5/4)}. \tag{30}$$

This means there is a transversal of length at least $n_k - k$, or

$$n - \frac{(\log n)^2}{2 \log(3/2) \log(5/4)}, \tag{31}$$

or

$$n - 5.53(\log n)^2.$$

The inequality (32) is clearly not the best result that can be derived from the inequality (1). For instance, replacing $\frac{4}{5}$ in (5) with another constant will improve the constant in (32) slightly. In fact, even for relatively small values, inequality (1) implies results better than $n - \sqrt{n}$. Note that in the proof of (1) we showed $n_k - n_{k-1} \geqslant 2k$. This implies by induction $n_k > k^2$, giving partial transversal of length $n - \sqrt{n}$. The inequality (1), however, cannot imply anything better than $n - (\log, n)$, since the sequence $n_k = 2^k$ satisfies (1).

REFERENCES

1. A. E. BROUWER, A. J. DE VRIES, AND R. M. A. WIERINGA, A lower bound for the length of partial transversals in a Latin square, *Nieuw Arch. Wisk.* (3) 24 (1978), 330–332.
2. D. A. DRAKE, Maximal sets of Latin squares and partial transversals, *J. Statist. Plann. Inference* I (1977), 143-149.
3. K. K. K OKSMA, A lower bound for the order of a partial transversal in a Latin square, *J. Combin. Theory Ser. A* 7 (1969), 94-95.
4. H. J. RYSER, Neuere Probleme der Kombinatorik, *in* "Vorträge über Kombinatorik." pp. 69-9 1, a collection of papers presented at the Mathematischen Forschungsinstitute, Oberwolfach, July 24-29, 1967.
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.