

LATIN SQUARES WITH FORBIDDEN ENTRIES

JONATHAN CUTLER AND LARS-DANIEL ÖHMAN

ABSTRACT. An $n \times n$ array is avoidable if there exists a Latin square which differs from the array in every cell. The main aim of this paper is to present a generalization of a result of Chetwynd and Rhodes involving avoiding arrays with multiple entries in each cell. They proved a result regarding arrays with at most two entries in each cell, and we generalize their method to obtain a similar result for arrays with arbitrarily many entries per cell. In particular, we prove that if $m \in \mathbb{N}$, there exists an $N = N(m)$ such that if F is an $N \times N$ array with at most m entries in each cell, then F is avoidable.

1. INTRODUCTION

The study of avoiding given configurations in Latin squares was initiated by Häggkvist, who in 1989 asked which $n \times n$ arrays can be avoided in every cell by some $n \times n$ Latin square. While finding arrays which cannot be avoided has some merit, finding families which can be avoided seems to be a more difficult problem. Häggkvist [10] was the first to present such a positive result (Theorem 1 below). Throughout the paper, we shall use the notation $[n]$ for the set $\{1, 2, \dots, n\}$. Unless explicitly stated otherwise, an $n \times n$ Latin square uses symbols $[n]$. A *partial $n \times n$ column-Latin square* on $[n]$ is an array of n rows and n columns in which each cell is empty or contains one symbol from $[n]$ and every symbol appears at most once in each column.

Theorem 1. *Let $n = 2^k$ and P be a partial $n \times n$ column-Latin square on $[n]$ with empty last column. Then there exists an $n \times n$ Latin square, on the same symbols, which differs from P in every cell.*

We say that an $n \times n$ array is *avoidable* if there is an $n \times n$ Latin square which differs from the array in every entry. Chetwynd and Rhodes [5] proved that some partial Latin squares are avoidable:

Theorem 2. *If $k \geq 2$, then all $2k \times 2k$ and $3k \times 3k$ partial Latin squares are avoidable.*

In this paper, we generalize another result of Chetwynd and Rhodes [6] involving avoiding arrays with at most two entries in each cell:

Theorem 3. *Let $k > 3240$ and F be a $4k \times 4k$ array such that each cell contains at most two symbols and each symbol appears at most twice in every row and column. Then F is avoidable.*

This result depended on a lemma, which we shall generalize and show is an easy consequence of Hall's theorem, and the following result of Daykin and Häggkvist [7].

Theorem 4. *Let $0 \leq d < k$ and \mathcal{H} be an r -partite r -uniform hypergraph with minimum degree $\delta(\mathcal{H})$ and $|V(\mathcal{H})| = rk$. If*

$$\delta(\mathcal{H}) > \frac{r-1}{r} [k^{r-1} - (k-d)^{r-1}],$$

then \mathcal{H} has more than d independent edges.

The main result of this paper is as follows:

Theorem 5. *Let $m \in \mathbb{N}$. There exists a constant $c = c(m)$ such that if $k > c = c(m)$ and if F is a $2mk \times 2mk$ array on the symbols $[2mk]$ in which every cell contains at most m symbols and every symbol appears at most m times in every row and column, then F is avoidable.*

The proof yields a polynomial bound on $c(m)$, but we have not tried to make this bound best possible. In fact, the bound has leading term $8m^8$, and we believe any significant improvement on this, e.g., an absolute constant, will involve a different approach.

We note that any result involving avoidable arrays corresponds directly to a list coloring problem on bipartite graphs. An $n \times n$ array can be thought of as the complete bipartite graph $K_{n,n}$ with vertices in one class corresponding to the rows of the array and vertices in the other the columns. So, edges in this representation of $K_{n,n}$ correspond to cells of the array. We begin with the assumption that every edge, or cell, can receive colors $1, 2, \dots, n$. Then, a proper edge coloring of $K_{n,n}$ corresponds to a Latin square on the $n \times n$ array. This question of list colorings of bipartite graphs was answered in a general form by a theorem of Galvin [8].

Theorem 6. *The list chromatic index of a bipartite graph equals the maximum degree of the bipartite graph.*

However, our question does not relate to this theorem in that we are interested not in subgraphs of $K_{n,n}$, but rather shortening the lists at each edge and seeing if there still is a proper edge coloring. Given Galvin's theorem, or even König's theorem, we must put some additional constraint on these lists, but this constraint is provided by the requirement that we can only remove a color at most m times from the edges incident to a single vertex. The shortening of the lists is done by our forbidden array F in that the entries in each cell are the forbidden colors on the corresponding edges. Thus, in this setting, we have the following corollary of Theorem 5.

Theorem 7. *Let $m \in \mathbb{N}$. Then there exists an $n_0 = n_0(m)$ such that if $n > n_0$ and the edges of $K_{2n,2n}$ are given lists which are each subsets of $[2n]$, contain at least $2n - m$ colors, and each color is deleted from the edges incident to a vertex at most m times, then there is a proper edge coloring of $K_{2n,2n}$ using the colors of the list at each edge.*

This paper will present the proof of the key lemma in the next section and then proceed to the proof of Theorem 5. We end with some discussions of the general problem and a conjecture involving the constant in Theorem 5. While an effort has been made to define all terminology, any undefined terms should be possible to find in [1].

2. THE LEMMA

In a $2n \times 2n$ array, a *generalized diagonal* is a set of $2n$ cells, one from each row and column, i.e., there is a rearrangement of rows and columns of the array such that this set is a diagonal. In this terminology, Chetwynd and Rhodes proved the following using case analysis:

Lemma 8. *Let E be a 4×4 array in which some cells contain symbol s and s appears at most twice in every row and column. Then there exists a generalized diagonal of E along which s does not appear.*

We will generalize this to $2k \times 2k$ arrays in which a symbol appears at most k times in each row and column. It is an immediate corollary of the classical theorem of König [9], and so we will state it without proof.

Lemma 9. *Let E be an $l \times l$ array with the symbol s appearing at most $l_2 = \lfloor \frac{l}{2} \rfloor$ times in each row and column. Then we can find a set M of l cells in E not containing s , such that no two of these cells appear in the same row or column.*

Note that this lemma gives that by permuting the rows and columns of M , the cells not containing s can be arranged to appear on the, say, upper right to lower left diagonal of E . This does not depend on the parity of l . Thus, we note that this step is not where the proof method of the main theorem breaks down for $mk \times mk$ arrays where m is odd.

3. PROOF OF THEOREM 5

Proof. Let $X = \{X_1, X_2, \dots, X_k\}$. For a $n \times n$ array A and ℓ such that $\ell|n$, we say that an $\ell \times \ell$ subarray of A is a *standard subsquare* if it is of the form $\{(i-1)\ell+1, \dots, i\ell\} \times \{(j-1)\ell+1, \dots, j\ell\}$ for $i, j \in [n/\ell]$. We divide F into the k^2 $2m \times 2m$ standard subsquares and label these subsquares with X_1, \dots, X_k in such a way that the $k \times k$ array on X is a Latin square. Let $S = [2mk]$. We shall partition S into k $2m$ -tuples S_1, \dots, S_k in such a way that we can form a Latin square on each X_i with the symbols in S_i . Further, let $X_i^1, X_i^2, \dots, X_i^k$ be the k $2m \times 2m$ standard subsquares of F labelled X_i . In order to ease our count and to be able to define a hypergraph to apply Theorem 4, we only consider a certain subset of $2m$ -tuples on S . Let $A_i = \{(i-1)k+1, \dots, ik\}$ for $i = 1, 2, \dots, 2m$, so that $S = \bigcup_{i=1}^{2m} A_i$. Then, let \mathbb{M} be the set of all $2m$ -tuples of the form $\{a_0, a_1, \dots, a_{2m-1}\}$ where $a_i \in A_i$ for $i = 1, 2, \dots, 2m$.

Let E be any of the k^2 $2m \times 2m$ standard subsquares of F . Let a *cell pair* of E be any two symbols which appear together in some cell of E . We then define \mathcal{C}_E to be the subset of \mathbb{M} containing all $2m$ -tuples which contain at least one cell pair. Then, since E has $4m^2$ cells, and each cell has at most m symbols,

$$|\mathcal{C}_E| \leq 4m^2 \binom{m}{2} k^{2m-2}. \quad (1)$$

Let a *doubled pair* of E be any two symbols x and y such that both x and y appear at least twice in some row or column of E . Then, let \mathcal{D}_E be the subset of \mathbb{M} containing all $2m$ -tuples which contain at least one doubled pair of E . Since there are at most $4m^3$ total symbols in E , at most $2m^3$ symbols can appear twice in E , and thus we have

$$|\mathcal{D}_E| \leq \binom{2m^3}{2} k^{2m-2}. \quad (2)$$

If we let $\mathcal{R}_E = \mathbb{M} \setminus \{\mathcal{C}_E \cup \mathcal{D}_E\}$, $R \in \mathcal{R}_E$ and $\{x, y\} \subset R$, then since we have excluded doubled pairs, we know that only one of x and y can appear more than twice in any row or column of E . So, if x appears at least 2 times, y appears at most once. Since we can note this of any pair of symbols in R , only one symbol in R can appear more than once in any row or column of E and all others appear at most once. Now, we know that at most $2m^3$ symbols appear at least twice in any row and column of E and we shall denote these symbols by $\{s_1^E, s_2^E, \dots, s_p^E\}$ for some p such that $0 \leq p \leq 2m^3$. Applying Lemma 9 for each of the s_i^E , we rearrange the rows and columns of E to get E_i so that the generalized diagonal given by the lemma is the upper right to lower left diagonal of E_i .

We call a pair of symbols $\{x, y\}$ *improper* for a 2×2 array if either diagonal of the array has x in one cell and y in the other. Otherwise, we say that the pair is *proper*. Now, we let \mathcal{I}_E be the set of all $2m$ -tuples in \mathbb{M} containing at least one improper pair for any standard 2×2 subsquares of each E_i , $1 \leq i \leq p$. For a fixed $2m \times 2m$ array with each cell containing at most m symbols, we partition the array into its standard 2×2 subsquares. We then count the number of improper pairs for each of the 2×2 subsquares of this $2m \times 2m$ array. For each diagonal of a fixed 2×2 array, there are at most m^2 improper pairs, and so there are at most $2m^2$ improper pairs per 2×2 subsquare. Since there are at most $2m^3$ E_i 's and m^2 2×2 standard subsquares of each E_i with at

most $2m^2$ improper pairs per 2×2 subsquare, we have

$$|\mathcal{I}_E| \leq 2m^3 \cdot m^2 \cdot 2m^2 \cdot k^{2m-2} = 4m^7 k^{2m-2}. \quad (3)$$

Finally, we need to get rid of all of the unusable $2m$ -tuples for our $2m \times 2m$ array E . So, we let $\mathcal{U}_E = \mathcal{C}_E \cup \mathcal{D}_E \cup \mathcal{I}_E$ and let $M \in \mathbb{M} \setminus \mathcal{U}_E$. If all $2m$ symbols of M appear less than twice in every row and column of E , then, since we have excluded improper pairs, only one symbol from M can appear on each diagonal for every 2×2 standard subsquare of E . Thus, we can label the 2×2 subsquares with m symbols Y_1, Y_2, \dots, Y_m in such a way that the Y_i s form an $m \times m$ Latin square. Since every pair of M forms a proper pair for every 2×2 standard subsquare of E in this case, we can partition M into pairs R_1, R_2, \dots, R_m arbitrarily and use the pair R_i to form a Latin square avoiding Y_i for all i , $1 \leq i \leq m$. Thus, we can form a Latin square avoiding E using the symbols of M .

Otherwise, since we have excluded doubled pairs, at most one symbol of M , say s , appears more than twice in some row or column of E . Therefore, we can rearrange E to get E' so that the generalized diagonal that does not contain s by Lemma 9 is the upper right to lower left diagonal of E' . Since any other symbol appears less than twice and we have gotten rid of the improper pairs, any other symbol of M , say t , can be used to complete the 2×2 standard subsquares containing s in E' , while the other $2m - 2$ can be used in any of the other 2×2 standard subsquares. If we again label the 2×2 standard subsquares with Y_1, \dots, Y_m in such a way that the Y_i s form a diagonal from upper right to lower left, and thus we can use symbols s and t to form Latin squares avoiding F in all of the Y_i s. We can then arbitrarily partition the remaining $2m - 2$ symbols of M into pairs R_2, \dots, R_m and use the symbols in R_i to form a 2×2 Latin square which avoids Y_i for all i , $1 \leq i \leq m$. Thus, M can be used to form a Latin square avoiding E . Also, note that we have

$$\begin{aligned} |\mathcal{U}_E| &\leq |\mathcal{C}_E| + |\mathcal{D}_E| + |\mathcal{I}_E| \\ &\leq 4m^2 \binom{m}{2} k^{2m-2} + \binom{2m^3}{2} k^{2m-2} + 4m^7 k^{2m-2}. \end{aligned} \quad (4)$$

We define a $(2m + 1)$ -partite $(2m + 1)$ -uniform hypergraph

$$\mathcal{H} = \mathcal{H}(X, A_1, A_2, \dots, A_{2m})$$

in which edge $X_i a_1 a_2 \dots a_{2m}$ for $X_i \in X$ and $a_i \in A_i$ is present if and only if the $2m$ -tuple $\{a_1, a_2, \dots, a_{2m}\}$ is not in $\mathcal{U}_{X_i^q}$ for any q with $1 \leq q \leq k$. We shall show that \mathcal{H} has a set of k independent edges, which corresponds to a partition of S , say S_1, \dots, S_k , into k $2m$ -tuples such that the symbols in S_i can be used to form a Latin square which avoids each subsquare labelled X_i .

To show that \mathcal{H} has an independent set of at least k edges, we shall find a lower bound for $\delta(\mathcal{H})$ and then apply the theorem of Daykin and Häggkvist, i.e. Theorem 4. We do this by first finding a lower bound on the degree of vertices in X , and then for those in the A_i s. First of all, for any i with $1 \leq i \leq k$, we have, using (4),

$$\begin{aligned} d_{\mathcal{H}}(X_i) &\geq |\mathbb{M}| - \sum_{q=1}^k |\mathcal{U}_{X_i^q}| \\ &\geq k^{2m} - k \left[4m^2 \binom{m}{2} k^{2m-2} + \binom{4m^3}{2} k^{2m-2} + 4m^7 k^{2m-2} \right] \\ &= k^{2m} - \left[4m^2 \binom{m}{2} + \binom{4m^3}{2} + 4m^7 \right] k^{2m-1}. \end{aligned} \quad (5)$$

We now consider $s \in S$ and the minimal degree over such vertices in \mathcal{H} . Let $\mathcal{C}_E(s)$ be the subset of \mathbb{M} which contains the $2m$ -tuple M if $M \supset \{s, t\}$ where s and t appear together in some cell of E for some $t \in [2mk]$. Likewise, let $\mathcal{D}_E(s)$ be the subset of \mathbb{M} which contains the $2m$ -tuple M if $M \supset \{s, t\}$ where s and t appear at least twice in some row or column of E for some $t \in [2mk]$ and $\mathcal{I}_E(s)$ be the subset of \mathbb{M} which contains the $2m$ -tuple M if $M \supset \{s, t\}$ where s and t are an improper pair for some 2×2 standard subsquare of E_i for $i \in \{1, 2, \dots, p\}$ and for some $t \in [2mk]$. We now note that \mathcal{H} cannot contain any edge containing both s and any member of $\mathcal{C}_{X_i^q}(s) \cup \mathcal{D}_{X_i^q}(s) \cup \mathcal{I}_{X_i^q}(s)$. However, contrary to what is implicitly claimed in [6], there are other edges which cannot be present in \mathcal{H} , i.e., those which contain a bad pair for a given X_i^q not involving s . Let $\mathcal{C}_E(\bar{s})$ be the subset of \mathbb{M} which contains the $2m$ -tuple M if $s \in M$ and $M \supset \{u, v\}$ where $u, v \neq s$ and u and v appear together in some cell of E . Similarly, let $\mathcal{D}_E(\bar{s})$ be the subset of \mathbb{M} which contains the $2m$ -tuple M if $s \in M$ and $M \supset \{u, v\}$ where $u, v \neq s$ and $\{u, v\}$ is a doubled pair of E . Lastly, let $\mathcal{I}_E(\bar{s})$ be the subset of \mathbb{M} containing the $2m$ -tuple M if $s \in M$ and $M \supset \{u, v\}$ where $u, v \neq s$ and $\{u, v\}$ is an improper pair for E . Then,

$$\begin{aligned} d_{\mathcal{H}}(s) &\geq k^{2m} - \left(\sum_{i,q} |\mathcal{C}_{X_i^q}(s)| + \sum_{i,q} |\mathcal{D}_{X_i^q}(s)| + \sum_{i,q} |\mathcal{I}_{X_i^q}(s)| \right) \\ &\quad - \left(\sum_{i,q} |\mathcal{C}_{X_i^q}(\bar{s})| + \sum_{i,q} |\mathcal{D}_{X_i^q}(\bar{s})| + \sum_{i,q} |\mathcal{I}_{X_i^q}(\bar{s})| \right). \end{aligned} \quad (6)$$

Symbol s occurs at most m times in each of the $2mk$ rows of F , so s can occur with t in at most $2m^2k$ of the X_i^q . Thus

$$\sum_{i,q} |\mathcal{C}_{X_i^q}(s)| \leq 2m^2k \cdot k^{2m-2} = 2m^2k^{2m-1}. \quad (7)$$

If a symbol s appears twice in some row or column of a $2m \times 2m$ standard subsquare of F , then at most $2m^3 - 1$ symbols also occur twice in some row or column of that same subsquare. Further, since s appears at most m times in each of the $2mk$ rows, and thus a total of at most $2m^2k$ times, it can occur twice in at most m^2k of the X_i^q . Thus,

$$\sum_{i,q} |\mathcal{D}_{X_i^q}(s)| \leq (2m^3 - 1)m^2k \cdot k^{2m-2} = (2m^5 - m^2)k^{2m-1}. \quad (8)$$

As noted above, there are at most $2m^2k$ cells containing s . If s occurs in some 2×2 standard subsquare of a particular $2m \times 2m$ standard subsquare of F , then, since there are at most m symbols in the cell diagonal from s , it occurs as an improper pair with at most m symbols. Furthermore, the $2m \times 2m$ standard subsquare is rearranged at most $2m^3$ times, since we rearrange once for each symbol appearing twice in some row or column of the standard subsquare, and all such improper pairs are counted. Thus,

$$\sum_{i,q} |\mathcal{I}_{X_i^q}(s)| \leq m \cdot 2m^3 \cdot 2m^2k \cdot k^{2m-2} = 4m^6k^{2m-1}. \quad (9)$$

We are left with bounding $\mathcal{C}_{X_i^q}(\bar{s})$, $\mathcal{D}_{X_i^q}(\bar{s})$ and $\mathcal{I}_{X_i^q}(\bar{s})$. But the same arguments that led to the inequalities (1), (2) and (3) give similar bounds for these, with the difference being that we have fixed one symbol, i.e., s , in each of the $2m$ -tuples in these sets. Thus, we see that for any fixed i and q ,

$$|\mathcal{C}_{X_i^q}(\bar{s})| \leq 4m^2 \binom{m}{2} k^{2m-3}, \quad (10)$$

$$|\mathcal{D}_{X_i^q}(\bar{s})| \leq \binom{2m^3}{2} k^{2m-3}, \quad (11)$$

and

$$|\mathcal{I}_{X_i^q}(\bar{s})| \leq 4m^7 k^{2m-3}. \quad (12)$$

Combining inequalities, (6), (7), (8) and (9) with (10), (11) and (12) after summing the last three over all i and q , we see that

$$\begin{aligned} d_{\mathcal{H}}(s) &\geq k^{2m} - [2m^2 k^{2m-1} + (2m^5 - m^2) k^{2m-1} + 4m^6 k^{2m-1}] \\ &\quad - k^2 \left[4m^2 \binom{m}{2} k^{2m-3} + \binom{2m^3}{2} k^{2m-3} + 4m^7 k^{2m-3} \right] \\ &= k^{2m} - \left[m^2 + 4m^2 \binom{m}{2} + 2m^5 + \binom{2m^3}{2} + 4m^6 + 4m^7 \right] k^{2m-1}. \end{aligned}$$

Using (5), and noting that $4m^2 \binom{m}{2} + \binom{2m^3}{2} + 4m^7 < m^2 + 4m^2 \binom{m}{2} + 2m^5 + \binom{2m^3}{2} + 4m^6 + 4m^7$ for $m \geq 1$, we see then that

$$\delta(\mathcal{H}) \geq k^{2m} - \left[m^2 + 4m^2 \binom{m}{2} + 2m^5 + \binom{2m^3}{2} + 4m^6 + 4m^7 \right] k^{2m-1}.$$

To apply Theorem 4 with $r = 2m + 1$ and $d = k - 1$, we need $\delta(\mathcal{H}) > \frac{2m}{2m+1}(k^{2m} - 1)$ in order to insure the existence of k independent edges. In other words, we need

$$\begin{aligned} k^{2m} - \left[m^2 + 4m^2 \binom{m}{2} + 2m^5 + \binom{2m^3}{2} + 4m^6 + 4m^7 \right] k^{2m-1} \\ > \frac{2m}{2m+1}(k^{2m} - 1). \end{aligned}$$

This certainly holds if

$$k > (2m + 1) \left[m^2 + 4m^2 \binom{m}{2} + 2m^5 + \binom{2m^3}{2} + 4m^6 + 4m^7 \right].$$

□

4. CONCLUDING REMARKS

It has been noted by Pebody [11] that for any n the following is an unavoidable $n \times n$ array on $[n]$ where each symbol appears at most $\lfloor \frac{n}{3} \rfloor + 1$ times in every row and column. We set $k = \lfloor \frac{n}{3} \rfloor + 1$.

$1, \dots, k$		
	$k + 1, \dots, 2k$	
		$2k + 1, \dots, n$

The first and second row and column of the above array have width k , while the third row and column have width $n - 2k < k$. The nonempty cells above are either $n \times n$ or $(n - 2k) \times (n - 2k)$ arrays, where each entry contains all the symbols in that cell. For $n = 2$ this is the canonical 2×2 unavoidable (single entries) array.

As mentioned above, the method employed in Theorem 5 does not yield anything near what we believe to be the best bounds possible. Since we have yet to find a counterexample to the following statement, we rather brashly present it as a conjecture. Note that this may be seen as an elaboration of a conjecture of Häggkvist [10].

Conjecture 10. *Let $m \geq 2$ and let F be a $cm \times cm$ array where each cell contains at most m symbols, and each symbol appears at most m times in every row and column. If $c \geq 3$ then F is avoidable.*

If true, the conjecture is sharp for all $n = cm$, by the above example of Pebody's. For $m = 1$, F is a partial Latin square, and the following holds (see [2], [3] and [5]).

Theorem 11. *Let P be an $n \times n$ partial Latin square, $n \geq 4$. Then P is avoidable.*

This is not true for $n = 3$, as there are examples of unavoidable 3×3 partial Latin squares. This might be taken as an indication that we should have $c > 3$ instead of $c \geq 3$ in Conjecture 10.

REFERENCES

- [1] B. Bollobás, *Modern Graph Theory*. Springer Verlag, New York, 1998.
- [2] N.J. Cavenagh, *Avoidable partial Latin squares of order $4m + 1$* , *Ars Combinatoria*, to appear.
- [3] N.J. Cavenagh and L-D. Öhman, *Partial Latin squares are avoidable*, Research reports in Mathematics **2** (2006), Dept. of Mathematics and Mathematical Statistics, Umeå University.
- [4] A.G. Chetwynd and S.J. Rhodes, Chessboard squares. *Discrete Math.* **141** (1995) 47–59.
- [5] A.G. Chetwynd and S.J. Rhodes, Avoiding partial Latin squares and intricacy. *Discrete Math.* **177** (1997) 17–32.
- [6] A.G. Chetwynd and S.J. Rhodes, Avoiding multiple entry arrays. *J. Graph Theory* **25** (1997) 257–266.
- [7] D.E. Daykin and R. Häggkvist, Degrees giving independent edges in a hypergraph. *Bull. Austral. Math. Soc.* **23** (1981) 103–109.
- [8] F. Galvin, The list chromatic index of a bipartite multigraph. *J. Combin. Theory, Ser. B* **63** (1995) 153–158.
- [9] D. König, Über graph und ihre anwendung auf determinantentheorie und mengenlehre. *Math. Annalen* **77** (1916) 453–465.
- [10] R. Häggkvist, A note on Latin squares with restricted support. *Discrete Math.* **75** (1989) 253–254.
- [11] L. Pebody, personal communication.