On the Distances between Latin Squares and the Smallest Defining Set Size

Nicholas Cavenagh and Reshma Ramadurai

Department of Mathematics, The University of Waikato, Hamilton, New Zealand,
E-mail: nickc@waikato.ac.nz, reshmar@waikato.ac.nz

Received January 10, 2016; revised June 07, 2016

Published online 04 August 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/jcd.21529

Abstract: In this note, we show that for each Latin square \( L \) of order \( n \geq 2 \), there exists a Latin square \( L' \neq L \) of order \( n \) such that \( L \) and \( L' \) differ in at most \( 8\sqrt{n} \) cells. Equivalently, each Latin square of order \( n \) contains a Latin trade of size at most \( 8\sqrt{n} \). We also show that the size of the smallest defining set in a Latin square is \( \Omega(n^{3/2}) \). © 2016 Wiley Periodicals, Inc. J. Combin. Designs 25: 147–158, 2017

Keywords: Latin square; Latin trade; defining set; critical set; Hamming distance

1. INTRODUCTION

For each positive integer \( a \), we use the notation \([a]\) for the set of integers \( \{0, 1, 2, \ldots, a - 1\} \).

A partial Latin square of order \( n \) is an \( n \times n \) array, where each cell of the array is either empty or contains a symbol from \([n]\), such that each symbol occurs at most once per row and at most once per column. A Latin square is a partial Latin square in which no cell is empty, and hence each symbol occurs precisely once in each row and once in each column.

Indexing rows and columns by \([n]\), we may consider a partial Latin square to also be a set of ordered triples of the form \((i, j, L(i, j))\), where \(L(i, j)\) is the symbol in row \( i \) and column \( j \) (if occupied). The distance (or Hamming distance [8]) between two partial Latin squares \( L \) and \( L' \) of the same order is then defined to be \( |L \setminus L'| \).

We show the following.

Theorem 1.1. For each Latin square \( L \) of order \( n \), there exists a Latin square \( L' \neq L \) of order \( n \) such that \( |L \setminus L'| \leq 8\sqrt{n} \).
Note that it is trivial to obtain an upper bound of $2n$ in the above; simply swap two rows of $L$ to create $L'$. Theorem 1.1 however is the first such upper bound that is $o(n)$. This a step toward the possible truth of Conjecture 4.25 from [5]:

**Conjecture 1.1.** For each Latin square $L$ of order $n$,

$$\min(|L \setminus L'| \mid L' \text{ is a Latin square of order } n \text{ and } L' \neq L) = O(\log n).$$

We may also state Theorem 1.1 as a result about Latin trades. Given two distinct Latin squares $L$ and $L'$ of the same order $n$, $L \setminus L'$ is said to be a Latin trade with disjoint mate $L' \setminus L$. In terms of arrays, we say that two partial Latin squares are row balanced if corresponding rows contain the same set of symbols; column balanced is defined similarly. A Latin trade $T$ and its disjoint mate $T'$ are thus a pair of partial Latin squares that occupy the same set of cells, are disjoint and are both row and column balanced.

**Example 1.2.** In the example below, $d(L_1, L_2) = 18$ and $L \setminus L'$ is a Latin trade with disjoint mate $L' \setminus L$.

$$\begin{array}{cccccc}
1 & 2 & 0 & 6 & 3 & 4 \\
6 & 1 & 5 & 4 & 0 & 2 \\
0 & 5 & 4 & 2 & 1 & 3 \\
3 & 4 & 2 & 1 & 5 & 6 \\
4 & 3 & 1 & 5 & 6 & 0 \\
2 & 6 & 3 & 0 & 4 & 5 \\
5 & 0 & 6 & 3 & 2 & 1 \\
\end{array}$$

$$\begin{array}{cccccc}
2 & 1 & 0 & 6 & 3 & 4 \\
6 & 5 & 4 & 1 & 0 & 2 \\
0 & 2 & 5 & 4 & 1 & 3 \\
3 & 4 & 2 & 5 & 6 & 0 \\
4 & 3 & 1 & 2 & 5 & 6 \\
1 & 6 & 3 & 0 & 4 & 5 \\
5 & 0 & 6 & 3 & 2 & 1 \\
\end{array}$$

Theorem 1.1 thus implies:

**Theorem 1.3.** Each Latin square $L$ of order $n$ contains a Latin trade $T$ such that $|T| \leq 8\sqrt{n}$.

As Latin squares are precisely operation tables for quasigroups, our main result can also be considered in the context of Hamming distances of algebraic objects (see [8] for more detail on this topic).

Let $B_n$ be the Latin square formed by the addition table for the integers modulo $n$; (i.e. $B_n(i, j) = i + j \mod n$). The upper bound in Conjecture 1.1 cannot be decreased, since it is known that any Latin trade in $B_n$ has size at least $e \log p + 3$, where $p$ is the least prime that divides $n$ ([4, 9]). It was recently shown in [14] that for each integer $n$, $B_n$ contains a Latin trade of size $5 \log_2 n$.
If Conjecture 1.1 above is true, it thus may be that the back circulant Latin square is the "loneliest" of all Latin squares; i.e. the Latin square with greatest minimum distance to any other Latin square. The smallest Latin trade, known as an intercalate, has size four and consists of a $2 \times 2$ subarray on two symbols (the disjoint mate is formed by swapping the symbols). It is shown in [13] that for any $\epsilon > 0$, almost all Latin squares of order $n$ possess at least $O(n^{3/2-\epsilon})$ intercalates. Thus we know that most Latin squares are not as "lonely" as the back circulant Latin square.

A defining set $L'$ for a Latin square $L$ of order $n$ is a subset $L' \subseteq L$ such that if $L''$ is a Latin square of order $n$ and $L' \subseteq L''$ then $L'' = L$. In other words, a defining set has unique completion to a Latin square of specified order. If $T$ is a Latin trade in a Latin square $L$ with disjoint mate $T'$, $(L \setminus T) \cup T'$ is a Latin square distinct from $L$. The following is immediate.

**Lemma 1.1.** If $D$ is a defining set for a Latin square $L$ and $T$ is a Latin trade such that $T \subseteq L$, then $T \cap D \neq \emptyset$.

It comes as no surprise then, that the new results on Latin trades in this paper yield a new result on defining sets.

A critical set for a Latin square of order $n$ is a minimal defining set; i.e. a defining set $D$ that is not the superset of any smaller defining set of the same order. A much studied open problem is to determine the smallest possible size of a critical set of order $n$ (equivalently, the smallest possible size of a defining set of order $n$), denoted by $scs(n)$. It is conjectured that the correct value for $scs(n)$ is equal to $\lceil n^2/4 \rceil$. (This has been verified computationally for $n \leq 8$ [1]). Defining sets of such size are known to exist for each $n \geq 1$ [6, 7]). Until quite recently, the best known lower bound for large $n$ was $scs(n) \geq n \lfloor (\log n)^{1/3}/2 \rfloor$ [2], which in turn improved results given in [10] and [12]. However, very recently this has been improved by Hatami and Qian [11]) who have shown that $scs(n) \geq 10^{-4}n^2$ for sufficiently large $n$.

Using the Latin trades constructed in Section 2 of this paper, we show that $scs(n) = \Omega(n^{3/2})$. Although this does not improve the result in [11], we include it as an interesting application of the theory used to prove Theorem 1.3.

**Theorem 1.4.** The size of the smallest defining set of any Latin square of order $n$ has size $\Omega(n^{3/2})$.

### 2. LATIN TRADES

Given a Latin square $L$ of order $n$, two distinct symbols $a, b \in [n]$ and some function $f = f(n) \in \mathbb{N}$ (we use $f(n) = \lceil 19\sqrt{n}/6 \rceil + 1$ in Section 3 and $f(n) = n$ in Section 4), we construct a coloured digraph $G(= G_{L,a,b,f})$ of order $n$ as follows. The vertices of $G$ are labeled with $[n]$ and correspond to the columns of $L$. Each directed edge will be colored green, black, or yellow. In what follows, a directed cycle of length $m \geq 1$ is a set of directed edges (of any color) of the form $\{(v_i, v_{i+1}) \mid i \in [m]\}$ where $v_0 = v_m$ and $v_1, v_2, \ldots, v_{m-1}$ are distinct vertices in $G$. Note that we include loops and directed circuits of length 2 are in our definition of a directed cycle.

Whenever $(r, c, a), (r, c', b) \in L$ for some fixed row $r$, we add a green edge from $c$ to $c'$ in the digraph $G$. There are no other green edges. We say that $\{(r, c, a), (r, c', b)\}$ is the partial Latin square associated with this green edge.
From the definition of a Latin square, the following is immediate.

**Lemma 2.1.** The green edges of the graph $G$ form a directed 2-factor of $G$.

For fixed $r, c$ and $r' \neq r$ (*), we define the following (infinite) sequence. Let $c_0 = c$ and $e_0 = a$ where $(r, c, a) \in L$. For each $i \geq 0$, let $e_{i+1}$ be the entry such that $(r', c_i, e_{i+1}) \in L$ and let $c_{i+1}$ be the column such that $(r, c_{i+1}, e_{i+1}) \in L$. This creates a “zig-zag” pattern, as shown below:

\[ c_0 \quad c_1 \quad c_2 \]
\[ r \quad e_0 \quad e_1 \quad e_2 \ldots \]
\[ r' \quad e_1 \quad e_2 \quad e_3 \ldots \]

For each integer $k$ we define $P_k(r, r', c)$ to be the following subset of $L$ of size $2k$:

\[ \{(r, c_i, e_i), (r', c_i, e_{i+1}) \mid i \in [k]\} \]

By finiteness, observe that $e_K = a$ for some $K$ such that $0 < K \leq n$; assume $K = K(r, r', c)$ is minimum with respect to this property.

**Lemma 2.2.** The partial Latin square $P_K(r, r', c)$ is a Latin trade of size $2K$.

**Proof.** Simply swap $r$ and $r'$ in each triple to form the disjoint mate. \qed

Any such Latin trade as in the previous lemma is called a row cycle trade.

We are now ready to define the black edges in the digraph $G$. Suppose that $e_k = b$ for some $k$ such that $0 < k < K$ and $k \leq f(n)$. Then add a black edge from $c_0$ to $c_{k-1}$ in the graph $G$. We say that $P_k(r, r', c = c_0)$ is the partial Latin square associated with a black edge.

Note that not every choice of $(r, c, a)$ and $r' \neq r$ (* see *) will result in a black edge; there are two possible obstacles. Firstly, it may happen that the sequence $e_0, e_1, e_2, \ldots$ does not contain the symbol $b$; equivalently, the Latin trade $P_K(r, r', c)$ does not include symbol $b$. Secondly, it is possible that $e_k = b$ implies that $k > f(n)$.

The following lemma is straightforward.

**Lemma 2.3.** Considering only green and black edges, the graph $G$ has neither loops nor multiple edges (i.e. no two edges sharing the same initial vertex $c$ and the same terminal vertex $c'$.)

It is our next aim to choose $a$ and $b$ in order to maximize the number of black edges in our digraph.

**Lemma 2.4.** Either the Latin square $L$ contains a Latin trade of size at most $2f(n)$ or the total number of black edges in digraphs of the form $G_{L,a,b,f}$ (where $a$ and $b$ are distinct symbols) is at least $n^2(n - 1)(f(n) - 1)$.

**Proof.** If the Latin square $L$ contains a Latin trade of size at most $2f(n)$ we are done. So in what follows, we assume that no such Latin trade exists.

Now, there are $n^2(n - 1)$ ways of choosing an element $(r, c, a) \in L$ and a row $r' \neq r$ for fixed $r, r'$, and $c$. Each such choice yields lists $c_0, c_1, \ldots c_{f(n)-1}$ and $e_0, e_1, \ldots e_{f(n)-1}$ (each with possible repeated elements) as above. If $f(n) \geq K$, then from Lemma 2.2 there exists a Latin trade of size $2K \leq 2f(n)$, a contradiction. Thus $f(n) < K$ and
the list $c_0, c_1, \ldots, c_{f(n)-1}$ has no repeated elements. Moreover, for each $i$ such that $0 < i < f(n)$, there is a black edge from $c_0$ to $c_i$ in the graph $G_{L,e_0,e_i+1,f(n)}$.

Thus there are a total of $n'(n-1)(f(n)-1)$ black edges in all of the graphs of the form $G_{L,a,b,f(n)}$ (where $a \neq b$). Note there is no overcounting here because the edges are directed and for each choice of $a$ in column $c$ and $b$ in column $c'$ there are unique rows $r$ and $r'$ such that $(r, c, a), (r', c', b) \in L$. □

Since there are $n(n-1)$ choices for the ordered pair of symbols $(a, b)$ (where $a \neq b$), the following corollary is immediate.

**Corollary 2.1.** There exist distinct symbols $a$ and $b$ such that the graph $G_{L,a,b,f}$ contains at least $n(f(n)-1)$ black edges.

We finally define yellow edges as follows. Whenever there is a green edge from $c_1$ to $c_2$, a black edge from $c_3$ to $c_2$ and a green edge from $c_3$ to $c_4$ (and $c_1 \neq c_4$), we add a yellow edge from $c_1$ to $c_4$. By Lemma 2.3, $c_1 \neq c_2$, $c_1 \neq c_3$, $c_2 \neq c_3$, $c_2 \neq c_4$, and $c_3 \neq c_4$. However $c_1 = c_4$ is possible; in such a case our yellow edge is a loop. The partial Latin square associated with a yellow edge is the union of the partial Latin squares associated with these two green edges and one black edge.

Our next aim is to show that any directed cycle in $G$ gives rise to a Latin trade in $L$.

A symbol cycle trade is a Latin trade containing only two symbols. For example:

\[
\begin{array}{ccc}
  a & b & \cdot \\
  c & d & \cdot \\
  a & b & \cdot \\
  b & a & \cdot \\
\end{array}
\]

Clearly the disjoint mate is formed by swapping symbols $a$ and $b$ within each row. In our graph, such a trade corresponds to a directed cycle of green edges. The following lemma is immediate.

**Lemma 2.5.** Let $T$ be the union of partial Latin squares associated with the edges of a green directed cycle of length $k$. Then $T$ is a Latin trade of size $2k$.

We now wish to consider directed cycles with edges of colours green, black, or yellow. For expediency we say that a partial Latin square associated with a directed edge of color $s$ is coloured $s$. For each colored partial Latin square $P$, we define the mate $P'$ of $P$ as follows. (The mates of partial Latin squares will ultimately define a disjoint mate of a Latin trade.) Firstly, the mate $P'$ of $P$ has the same set of occupied cells of $P$ but is disjoint from $P$. For a green partial Latin square $P = \{(r, c, a), (r, c, b)\}$, its mate $P' = \{ (r, c, b), (r, c, a) \}$. For a black partial Latin square $P$, the mate $P'$ of $P$ is formed by swapping the rows of $P$ and then the symbols $a$ and $b$. For a yellow partial Latin square $P$, the mate $P'$ of $P$ is formed by swapping the symbols in each column with two symbols and swapping $a$ with $b$ in columns with only one symbol.
Examples of mates for green, black, and yellow partial Latin squares are exhibited below, with symbols from \( P' \) given as subscripts.

\[
\begin{array}{c|c}
\text{Green} & a_1 & b_1 & a_2 & b_2 & a_3 & b_3 \\
\text{Black} & 1_a & 1_b & 2_a & 2_b & 3_a & 3_b \\
\text{Yellow} & a & b & b & a & 2 & 1
\end{array}
\]

Observe the following lemma.

**Lemma 2.6.** If \( P' \) is the mate of a partial Latin square \( P \) associated with a green, black, or yellow edge, then:

- \( P \) and \( P' \) are row-balanced;
- If we remove \( a \) from \( P \) and \( b \) from \( P' \) in the first column and if we remove \( b \) from \( P \) and \( a \) from \( P' \) in the final column then \( P \) and \( P' \) become column-balanced.

(Here the first and final columns correspond to the initial and terminal vertices, respectively, of the colored edge.)

What we have then, are partial Latin squares paired with mates that behave almost like Latin trades except for at the initial and terminal vertices. However, if each terminal vertex of one edge coincides with an initial vertex of another edge (i.e. we have a directed cycle) and if the partial Latin squares are disjoint, it is clear we have a Latin trade. (This is demonstrated in Example 1.2: there is a green edge from the first column to the second column, a black edge from the second to the fourth column, another black edge from the fourth to the seventh column and finally a green edge from the seventh to the first column.) This is true even if our directed cycle has one edge; i.e. is a yellow loop.

**Corollary 2.2.** Suppose there is a directed cycle \( C \) in \( G \) (with edges of any colour) and that the partial Latin squares associated with the edges of \( C \) are pairwise disjoint. Then there is a Latin trade of size at most \( 2g + 2bf(n) + 2y(f(n) + 1) \), where \( g, b, \) and \( y \) are the number of green, black, and yellow edges, respectively, in the cycle \( C \).

**Proof.** Let \( T \) be the union of the partial Latin squares associated with the edges of \( C \) and let \( T' \) be the union of the mates of these partial Latin squares. Since \( T \) is a subset of the Latin square \( L \), \( T \) is certainly a partial Latin square (i.e. no symbols are repeated in rows or columns). Clearly \( T \) and \( T' \) occupy the same set of cells and by construction are row balanced. Let \( e_1 \) and \( e_2 \) be directed edges of \( C \) such that column \( c \) is the terminal vertex of \( e_1 \) and the initial vertex of \( e_2 \). (Note that \( e_1 = e_2 \) if \( C \) is a loop.) Then there is a cell in column \( c \) in the partial Latin square corresponding to \( e_1 \) where symbol \( b \) is replaced with symbol \( a \) in the mate, and vice versa in the partial Latin square corresponding to \( e_2 \). It follows that \( T \) and \( T' \) are column balanced. Thus \( T \) is a Latin trade with disjoint mate \( T' \).

We are nearly ready to prove Theorem 1.1 —we just have to consider the case when the associated partial Latin squares are not disjoint.

**Theorem 2.1.** Let \( C \) be a directed cycle of minimum length in \( G \). Then there is a Latin trade of size at most \( 2g + 2bf(n) + 2y(f(n) + 1) \), where \( g, b \) and \( y \) are the number of green, black, and yellow edges, respectively, in the cycle \( C \).
Proof. Let $C$ be a directed cycle in $G$ and consider only the partial Latin squares that are associated with edges of $G$. Firstly, if a green partial Latin square intersects either a black or yellow partial Latin square, two directed edges have the same initial or terminal vertex, a contradiction.

Next, suppose two black partial Latin squares intersect; say $P_{k}(r_1, r_2, c)$ and $P_{k}(r_3, r_4, c')$. Then clearly $\{r_1, r_2\} \cap \{r_3, r_4\} \neq \emptyset$. If either $r_1 = r_3$ or $r_2 = r_4$ then two directed edges in $C$ start at the same vertex or terminate at the same vertex (respectively), a contradiction. Otherwise if $r_1 = r_4$ and $r_2 = r_3$, then $P_{k}(r_1, r_2, c) \cup P_{k}(r_3, r_4, c')$ forms a row cycle Latin trade of size $2(k + \ell)$ and we are done.

Next, suppose that $r_3 = r_2$ and $r_1 \neq r_4$ and the associated partial Latin squares intersect. (The case when $r_1 = r_4$ and $r_2 \neq r_3$ is equivalent.) Removing the two black edges from $C$ creates two directed paths; let $D$ be the directed path that starts at $c_k$ (i.e. the final column of $P_{k}(r_1, r_2, c)$) and terminates at $c'$. However, since $r_3 = r_2$ there is a green edge from $c'$ to $c_k$; together these form a directed cycle with length shorter than $C$, a contradiction.

Next, suppose two yellow partial Latin squares intersect. Let the black partial Latin squares that are subsets of these yellow partial Latin squares be $P_{k}(r_1, r_2, c)$ and $P_{k}(r_3, r_4, c')$, respectively. The cases $r_1 = r_3$ or $r_2 = r_4$ lead to contradictions as above. The case $r_1 = r_4$ and $r_2 = r_3$ implies a trade of size $2(k + \ell)$, similarly to above. This leaves the case $r_2 = r_3$ and $r_1 \neq r_4$ (equivalent to the case $r_1 = r_4$ and $r_2 \neq r_3$). Again, there is a green edge from $c'$ to $c_k$, which combined with a directed path from $C$, forms a directed cycle that is shorter than $C$.

Finally, suppose a black partial Latin square $P_{k}(r_1, r_2, c)$ intersects with a yellow partial Latin square (containing the black partial Latin square $P_{k}(r_3, r_4, c')$). If $P_{k}(r_1, r_2, c) = P_{k}(r_3, r_4, c')$, then there is a green edge from $c_k$ to $c'$, which combined with a directed path from $C$ forms a directed cycle that is shorter than $C$. Otherwise if $\{r_1, r_2\} = \{r_3, r_4\}$ there is a trade of size $2(k + \ell)$ within these rows. The cases $r_1 = r_3$ or $r_2 = r_4$ lead to contradictions as above. If $r_1 = r_4$ and $r_2 \neq r_3$, let $c''$ be the column of $P_{k}(r_3, r_4, c')$ that contains $b$ in row $r_4$. Then there is a directed path in $C$ from $c''$ to $c$ and a green edge from $c$ to $c''$, creating a cycle shorter than $C$. Finally, if $r_2 = r_3$ and $r_1 \neq r_4$, let $c'''$ be the column of $P_{k}(r_3, r_4, c')$ that contains $b$ in row $r_3$. Then there is a directed path in $C$ from $c'''$ to $c'$ and a green edge from $c'$ to $c'''$, again causing a contradiction.

\[\square\]

3. AN UPPER BOUND ON THE DISTANCE BETWEEN LATIN SQUARES

In this section, we give a proof of Theorem 1.1. To ultimately show the existence of Latin trades, we first make some observations about drawing edges between parallel line paths without edges crossing. In what follows, for the sake of simplicity of explanation, we embed two directed paths $P$ and $Q$ (of orders $p$ and $q$, respectively) in the Euclidean plane so that $P$ lies entirely within the line $y = 0$ and $Q$ lies entirely within the line $y = 1$ and the vertices have integer coordinates with each edge directed from left to right. (Really we simply need $P$ and $Q$ to be drawn as parallel line segments, the above specificity avoids any ambiguity). The following lemma is easy to show for example by induction.

Lemma 3.1. At most $p + q - 1$ straight line segments can be drawn between vertices of $P$ and vertices of $Q$ without any edges crossing.
Proof. Our proof is by induction on \( p + q \). The result is clearly true for \( p + q = 2 \). Next, assume the result holds whenever \( p + q = k \) for some integer \( k \geq 2 \). Then adding one vertex to either path we can add at least one line segment that does not cross existing edges; the result follows.\( \square \)

For our purposes we need something more specific.

Lemma 3.2. Let \( p, q \geq 3 \). If more than \( 2(p + q - 2) \) straight line segments are drawn between vertices of \( P \) and vertices of \( Q \), then there exists two edges that cross such that the edges do not use vertices adjacent in \( P \) or adjacent in \( Q \).

Proof. Assume, for the sake of contradiction, that no such two edges exist. Properly color the vertices of \( P \) with colors \( c_1 \) and \( c_2 \) and the vertices of \( Q \) with colors \( c'_1 \) and \( c'_2 \) (i.e. within each path vertices of the same color are not adjacent). Let \( X_{ij} \) be the number of vertices of color \( c_i \) or \( c'_i \). Then \( X_{ij} + X_{i'j} + X_{i'j'} + X_{ij'} = 2(p + q) \). Let \( Y_{ij} \) be the set of line segments between vertices of colors \( c_i \) and \( c'_j \). From the previous lemma, \( |Y_{ij}| \leq X_{ij} - 1 \). Thus there are at most \( 2(p + q - 2) \) line segments, a contradiction.\( \square \)

We now explain why we need the previous lemmas.

Lemma 3.3. Let \( P \) and \( Q \) be directed paths of green edges embedded in the Euclidean plane as above, where \( p \geq 3 \) and \( q \geq 3 \). If there exist more than \( 2(p + q - 2) \) black edges between vertices of \( P \) and vertices of \( Q \), then there exists a directed cycle in \( G \) on the vertices of \( P \) and \( Q \) such that the number of edges that are either black or yellow is at most 2.

Proof. Let the vertices of \( P \) be \( 1, 2, \ldots, p \) and let the vertices of \( Q \) be \( 1', 2', \ldots, q' \) where each directed edge is from \( i \) to \( i + 1 \) in \( P \) or from \( i' \) to \( (i + 1)' \) in \( Q \) for some \( i \). From the previous lemma, there exists a black edge on vertices \( i \) and \( (i + 1)' \) and if there exist more than \( 2(p + q - 2) \) black edges between vertices of \( P \) and vertices of \( Q \), then there exists a directed cycle in \( G \) on the vertices of \( P \) and \( Q \) such that the number of edges that are either black or yellow is at most 2.

We now explain why we need the previous lemmas.

Theorem 3.1. For \( n \geq 2 \), each Latin square contains a Latin trade of size at most \( 8\sqrt{n} \).

Proof. If \( n < 16, 8\sqrt{n} > 2n \) and since any Latin square of order \( n \geq 2 \) has a Latin trade of size \( 2n \) we may assume henceforth that \( n \geq 16 \). Let \( b = 4, k = 4/3, \) and \( d = 19/6 \). Suppose, for the sake of contradiction, there exists a Latin square \( L \) of order \( n \) such that every Latin trade in \( L \) has size greater than \( 2b\sqrt{n} \). We consider the directed colored graph \( G = G_{L,a,b,f(a)} \) as defined in the previous section where \( f(n) = \lfloor d\sqrt{n} \rfloor + 1 \). From Lemma 2.5, each directed green cycle in \( G \) has length greater than \( b\sqrt{n} \geq 16 \). We now partition the green edges into directed paths so that each path has order at most \( k\sqrt{n} \). It is clear that we can minimize the number of such paths by ensuring that each cycle contributes at most one path of order less than \( \lfloor k\sqrt{n} \rfloor \).

Since each cycle has length greater than \( b\sqrt{n} \), the number of directed green cycles is less than \( \sqrt{n}/b \). Thus the number of paths of order less than \( \lfloor k\sqrt{n} \rfloor \) is at most \( \sqrt{n}/b \). Also the total number of paths of order equal to \( \lfloor k\sqrt{n} \rfloor \) is at most \( \sqrt{n}/k \). Thus the total number of paths in our partition of the green edges is at most \( \sqrt{n}(b + k)/bk \). In turn, the total number of pairs of paths is less than \( n(b + k)^2/2b^2k^2 \).

Journal of Combinatorial Designs DOI 10.1002/jcd
By Corollary 2.1, $G$ has at least $dn^{3/2}$ black edges. Suppose there are at least $4k\sqrt{n}$ black edges between two of the paths of order at least three in the partition. Then by Lemma 3.3, there is a directed cycle in $G$ using at most $2k\sqrt{n}$ green edges and at most two edges that are either black or yellow. Thus by Theorem 2.1 there is a Latin trade of size at most $2(k - 1)\sqrt{n} + 2d\sqrt{n} + 4 = 2\sqrt{n}(k + d - 1) + 4$. Since $b = k + d - 1/2$ and $n \geq 16$, we are done in this case.

Thus there are less than $4k\sqrt{n}$ black edges between each pair of directed green paths of order at least three. There are also at most $4k\sqrt{n}$ black edges between a path of order at most two and any other path. If there is a black edge using two vertices from the same directed green path, then we create a directed cycle with at most $k\sqrt{n}$ edges and thus we are done.

Thus the total number of black edges is less than $4k\sqrt{n}$ times the number of pairs of directed paths. Given the above lower bound of $dn^{3/2}$ on the number of black edges, we have:

\[
\frac{2(b + k)^2}{b^2k} \geq d,
\]

a contradiction given the above values of $b$, $d$, and $k$. □

4. A LOWER BOUND ON THE SIZE OF A DEFINING SET

In this section, we give a proof of Theorem 1.4. To that end, it suffices to prove the following.

Theorem 4.1. The size of the smallest defining set of any Latin square of order $n$ is $\Omega(n^{3/2})$.

Proof. Suppose for the sake of contradiction, there exists a Latin square $L$ of order $n$ with a defining set $D'$ such that $|D'| \leq cn^{3/2}$ where $0 < c < 1/\sqrt{40}$. If we can show that $L \setminus D'$ contains a Latin trade, we are done by Lemma 1.1. It thus suffices to show that $L \setminus D$ contains a Latin trade for any $D$ such that $D' \subseteq D \subseteq L$ and $|D| = [cn^{3/2}]$. We let $f(n) = n$ and define graphs $G_{L,a,b,f(n)}$ with colored edges as in Section 2, with the proviso that only edges corresponding to partial Latin squares that do not include elements of $D$ are included. By Theorem 2.1 it suffices to show that the graph $G$ contains a colored cycle for some ordered pair of symbols $(a, b)$. (We set $f(n) = n$ as we simply need to show the existence of a Latin trade; its size to us is irrelevant.) We let $\mathcal{B}$ be the total number of black edges in any graph of the form $G_{L,a,b,n}$.

We first obtain a lower bound for $\mathcal{B}$. Let $x_i$ be the number of elements in row $i$ of $D$. Then $\sum_{i=0}^{n-1} x_i = [cn^{3/2}]$. Consider the first two rows of $L$. Let $k = x_0 + x_1$. The elements in these rows partition into row-cycles (or possibly just one large row-cycle), as in Lemma 2.2. Rearrange the columns so that any columns in the same row-cycle form a consecutive set of integers. Since each row cycle gives a Latin trade, each row cycle contains an element of $D$. Rearrange the columns further so that the first column in each row-cycle contains an element of $D$. Finally, rearrange the columns within each trade so that the element in cell $(1, c)$ is the same as the element in cell $(0, c + 1)$, unless $c + 1$ belongs to a different row-cycle to $c$. 

Journal of Combinatorial Designs DOI 10.1002/jcd
Let \( c_1, c_2, \ldots, c_e \) be the columns containing elements of \( D \) where \( k/2 \leq e \leq k \). (The lower bound arises in the extreme case that each column contains two elements from \( D \) in rows 1 and 2.) For each \( 1 \leq i < e \) we define block \( B_i \) to be the set of columns strictly between \( c_i \) and \( c_{i+1} \) (not including \( c_i \) and \( c_{i+1} \)). Let \( b_i = c_{i+1} - c_i - 1 \) and \( b_e = (n - 1) - c_e \). Observe that \( \sum_{i=1}^{e} b_i = n - e \geq n - k \).

We illustrate the above notation in the example below. Elements of \( D \) are given in bold and underlined; \( x_0 = x_1 = 2, k = 4, e = 3, c_1 = 0, c_2 = 3, c_3 = 5, b_1 = 2, b_2 = 1, \) and \( b_3 = 4 \).

\[
\begin{array}{cccccccccccc}
0 & 7 & 2 & 3 & 4 & 5 & 6 & 1 & 8 & 9 \\
7 & 2 & 0 & 4 & 5 & 6 & 1 & 8 & 9 & 3 \\
\end{array}
\]

Observe that any pair of distinct columns within a block gives rise to two black edges. For example, in the above the columns in \( B_1 \) give rise to an edge from the second to the third column in \( G_{L,0,7,n} \), and an edge from the third to the second column in \( G_{L,0,7,n} \). Thus the number of black edges arising from rows 0 and 1 is equal to

\[
2 \sum_{i=1}^{e} \begin{pmatrix} b_i \\ 2 \end{pmatrix} = e - n + \sum_{i=1}^{e} b_i^2 \\
\geq e - n + e ((n - e)/e)^2 \\
= 2e - 3n + n^2/e \\
\geq k - 3n + n^2/k \\
= (x_0 + x_1) - 3n + \frac{n^2}{(x_0 + x_1)}.
\]

So, considering all pairs of rows,

\[
B \geq c(n - 1)n^{3/2} - \frac{3}{2} n^2(n - 1) + n^2 \sum_{i<j} \frac{1}{x_i + x_j}. 
\]

The last sum in the above expression is minimized when all \( x_i \)'s are equal, i.e. \( x_i = \lceil cn^{3/2} \rceil /n \leq cn^{1/2} + 1 \), so:

\[
B \geq c(n - 1)n^{3/2} - \frac{3}{2} n^2(n - 1) + \frac{n^3(n - 1)}{4(cn^{1/2} + 1)}. \tag{1}
\]

Next, we calculate an upper bound for \( B \); if this is less than the previous expression we are done. Let \( z_i \) be the number of times symbol \( i \) appears in \( D \). Then clearly \( \sum_{i=0}^{n-1} z_i = \lceil cn^{3/2} \rceil \).

Consider the green edges in \( G_{L,0,1,n} \). Let \( z_0 + z_1 = \ell \). Let \( m \) be the number of green edges missing in the graph; observe that \( \ell/2 \leq m \leq \ell \). We know that there are no directed cycles of green edges, otherwise we will be done by Lemma 2.5. Thus there are \( m \) green paths; let their orders be: \( g_1, g_2, \ldots, g_m \).

There are less than \( 3m \leq 3\ell \) vertices not belonging to a green path of order at most three. Thus there are less than \( 3n\ell \) black edges that are not incident with a green path of order at least three. (Note we must avoid directed cycles of black edges of length two,
so each pair of vertices has at most one directed black edge.) Next, there are less than $2 \sum_{i<j}(g_i + g_j)$ black edges between green paths of order at least three, otherwise there exists a directed cycle (and thus from Theorem 3.3 a Latin trade).

In total, the number of black edges in $G_{L,0,1,n}$ is less than:

$$3n\ell + 2 \sum_{i<j} (g_i + g_j) \leq 3n\ell + 2(m-1)(n-m)$$

$$\leq 3n\ell + 2(\ell-1)(n-\ell/2)$$

$$< 5n\ell + \ell = (z_0 + z_1)(5n + 1).$$

Hence summing over all pairs of symbols,

$$B \leq (5n + 1) \sum_{i \neq j} (z_i + z_j) = 2(n-1)(5n+1) \sum_i z_i$$

$$\leq 2(n-1)(5n+1)(cn^{3/2} + 1).$$

However, combining this upper bound for $B$ with the lower bound in (1) creates a contradiction for large enough $n$. □

We remark that the bound in the previous theorem is asymptotic. For small orders greater than eight, linear bounds (e.g. [10, 12]) are still the best known. To give an idea of the limitations of our methods in the main results of this paper, consider the following graph $G$ on vertex set $N(n = m^2)$. For each $k \in [m]$, let $G_k = [km, km + 1, \ldots, k(m + 1) - 1]$ be a directed green path in $G$. Add a directed black edge from $i$ to $j$ whenever $i - j$ is divisible by $m$ and $i < j$. Such a graph has $\Omega(n^{3/2})$ black edges yet is free of directed cycles.

REFERENCES

[1] R. Bean, The size of the smallest uniquely completable set in order 8 latin squares, J Combin Math Combin Comput 52 (2005), 159–168.

[2] N. J. Cavenagh, A superlinear lower bound for the size of a critical set in a Latin square, J Combin Designs 15 (2007), 269–282.

[3] N. J. Cavenagh, The theory and application of latin bitrades: a survey, Math Slovac 58 (2008), 1–28.

[4] N. J. Cavenagh, The size of the smallest Latin trade in the back circulant Latin square, Bull Inst Combin Appl 38 (2003), 11–18.

[5] N. J. Cavenagh, Latin trades and critical sets in Latin squares, PhD Thesis, University of Queensland, 2003.

[6] J. Cooper, D. Donovan, and J. Seberry, Latin squares and critical sets of minimal size, Australasian J Combin 4 (1991), 113–120.

[7] D. Donovan and J. Cooper, Critical sets in back circulant latin squares, Aequationes Math 52 (1996), 157–179.

[8] A. Drápal, Hamming distance of groups and quasi-groups, Discrete Math 235 (2001), 189–197.

[9] A. Drápal and T. Kepka, On a distance of groups and Latin squares, Comment Math Univ Carolin 30 (1989), 621–626.
[10] C-M. Fu, H-L. Fu, and C. A. Rodger, The minimum size of critical sets in latin squares, J Stat Plann Inference 62 (1997), 333–337.

[11] H. Hatami and Y. Qian, Teaching dimension, VC dimension and critical sets in Latin squares, (submitted), arxiv.org/abs/1606.00032.

[12] P. Horak, R. E. L. Aldred, and H. Fleischner, Completing latin squares: critical sets, J Combin Designs 10 (2002), 419–432.

[13] B. D. McKay and I. M. Wanless, Most Latin squares have many subsquares, J Combin Theory Ser A 86 (1999), 323–347.

[14] M. Szabados, Distances of group tables and Latin squares via equilateral triangle dissections, J Combin Theory Ser A 123 (2014), 1–7.