Combinatorics, Probability and Computing (2019), **28**, pp. 675–695
doi:10.1017/S096354831800055X

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Restricted completion of sparse partial Latin squares

Lina J. Andrén[1,†], Carl Johan Casselgren[2,*,‡] and Klas Markström[3,§]

[1]University Library, Mälardalen University, SE-721 23 Västerås, Sweden, [2]Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden and [3]Department of Mathematics and Mathematical Statistics, Umeå University, SE-901 87 Umeå, Sweden

[*]Corresponding author. Email: carl.johan.casselgren@liu.se

(Received 7 September 2016; revised 29 October 2018; first published online 20 February 2019)

Abstract

An $n \times n$ partial Latin square P is called α-dense if each row and column has at most αn non-empty cells and each symbol occurs at most αn times in P. An $n \times n$ array A where each cell contains a subset of $\{1, \dots, n\}$ is a $(\beta n, \beta n, \beta n)$-array if each symbol occurs at most βn times in each row and column and each cell contains a set of size at most βn. Combining the notions of completing partial Latin squares and avoiding arrays, we prove that there are constants $\alpha, \beta > 0$ such that, for every positive integer n, if P is an α-dense $n \times n$ partial Latin square, A is an $n \times n$ $(\beta n, \beta n, \beta n)$-array, and no cell of P contains a symbol that appears in the corresponding cell of A, then there is a completion of P that avoids A; that is, there is a Latin square L that agrees with P on every non-empty cell of P, and, for each i, j satisfying $1 \leqslant i, j \leqslant n$, the symbol in position (i, j) in L does not appear in the corresponding cell of A.

2010 MSC Codes: Primary 05B15; Secondary 05C15

1. Introduction

Consider an $n \times n$ array A where each cell contains a subset of the symbols in $[n] = \{1, \dots, n\}$. If no cell in A contains a set of size larger than m_1, and if no symbol occurs more than m_2 times in any row or more than m_3 times in any column, then A is an (m_1, m_2, m_3)-*array (of order n)*. A $(1, 1, 1)$-array is usually called a *partial Latin square* (or PLS), and such an array with no empty cell is a *Latin square*. The cell in position (i, j) of A is denoted by $(i, j)_A$, and the set of symbols in cell $(i, j)_A$ is denoted by $A(i, j)$. By slight abuse of notation, if L is a (partial) Latin square, then $L(i, j)$ usually denotes the symbol in cell $(i, j)_L$, that is, $L(i, j) = k$. Moreover, the symbol $L(i, j)$ is called an *entry* of cell $(i, j)_L$.

An $n \times n$ partial Latin square P is called α-*dense* if each row and column contains at most αn non-empty cells and each symbol appears at most αn times in P. An $n \times n$ partial Latin square P is *completable* if there is an $n \times n$ Latin square L such that $L(i, j) = P(i, j)$ for each non-empty

[†]Part of the work done while the author was a postdoctoral researcher at the Mittag-Leffler Institute. Research supported by a postdoctoral grant from the Mittag-Leffler Institute.

[‡]Part of the work done while the author was a postdoctoral researcher at the Mittag-Leffler Institute. Research supported by a postdoctoral grant from the Mittag-Leffler Institute.

[§]Part of the work done while the author was visiting the Mittag-Leffler Institute. Research supported by the Mittag-Leffler Institute.

cell $(i,j)_P$ of P; L is also called a *completion* of P. Similarly, an $n \times n$ array A is *avoidable* if there is an $n \times n$ Latin square L such that for each i,j satisfying $1 \leqslant i,j \leqslant n$, $L(i,j)$ does not appear in cell $(i,j)_A$ of A; we also say that L *avoids* A.

The problem of completing partial Latin squares is classical in combinatorics and there is a wealth of results in the literature. Let us here just mention a few classical and recent results. In general, it is an NP-complete problem to determine if a partial Latin square is completable [16]. Thus it is natural to ask if particular families of partial Latin squares are completable. A classical result due to Ryser [28] states that if $n \geqslant r, s$, then every $n \times n$ partial Latin square whose non-empty cells form an $r \times s$ Latin rectangle Q is completable if and only if each of the symbols $1, \ldots, n$ occurs at least $r + s - n$ times in Q. Another classical result is Smetaniuk's proof [29] of Evans' conjecture [20] that every $n \times n$ partial Latin square with at most $n - 1$ entries is completable. This was also independently proved by Andersen and Hilton [2]. Adams, Bryant and Buchanan [1] characterized which partial Latin squares with two filled rows and columns are completable and by results of Casselgren and Häggkvist [10], and Kuhl and Schroeder [24], all partial Latin squares of order at least 6 with all entries in one fixed column or row, or containing a prescribed symbol, is completable. Building on techniques by Chetwynd and Häggkvist [12] and Gustavsson [21], Bartlett [7] proved that every ε-dense partial Latin square is completable, provided that $\varepsilon < 9.8 \times 10^{-5}$. This was recently improved upon in [6] where it was proved that the same conclusion holds under the assumption that $\varepsilon < 1/25$.

The problem of avoiding arrays was first posed by Häggkvist [22]. He also found the first (non-trivial) family of avoidable arrays: If $n = 2^k$ and P is a $(1, n, 1)$-array of order n with empty last column, then P is avoidable. In his original paper [22] Häggkvist also conjectured that there is a constant $c > 0$ such that for every positive integer n, every (cn, cn, cn)-array is avoidable. Andrén [3] established that Häggkvist's conjecture holds for arrays of even order and the case of odd order arrays was settled by Andrén, Casselgren and Öhman [4] confirming Häggkvist's conjecture in the affirmative. Related results appear in [9, 15, 17]; in particular, in [9] it is proved that it is NP-complete to decide if an array with at most two symbols per cell is avoidable, even if only two distinct symbols occur in the array.

Much of the research on avoiding arrays has focused on avoiding arrays that contain at most one symbol in each cell, so-called *single entry arrays*. Most notably, by results of Chetwynd and Rhodes [14], Cavenagh [11] and Öhman [26], all partial Latin squares of order at least 4 are avoidable. In [13], [9] and [25] some families of avoidable and unavoidable arrays are given.

In this paper we combine the notions of completing partial Latin squares and avoiding arrays and consider the problem of completing a partial Latin square subject to the condition that the completion should avoid a given array as well. There are some previous results in this direction: Öhman [27] determined for which pairs P, A, where P is a partial Latin square of order n with entries only from two distinct symbols, and A is a single entry array of order n with entries only from the same two distinct symbols, there is a completion of P that avoids A. Denley and Kuhl [19] proved that if P is an $n \times n$ partial Latin square and Q is an $n \times n$ partial Latin square that avoids P, then there is a completion of P that avoids Q if $n = 4t$, P contains at most $t - 1$ non-empty cells and $t \geqslant 9$.

Note further that the problem of determining if a given partial Latin square P has a completion L which avoids a given array A is certainly NP-complete in the general case, since it contains both the problem of completing partial Latin squares and avoiding arrays as special cases.

The main result of this paper is the following proposition, which is proved by combining techniques developed by Bartlett [7] and Andrén, Casselgren and Öhman [4].

Theorem 1.1. *There are constants $\alpha > 0$ and $\beta > 0$, such that for every positive integer n, if P is an $n \times n$ α-dense partial Latin square, A is an $n \times n$ $(\beta n, \beta n, \beta n)$-array, and no cell of P contains a symbol that occurs in the corresponding cell of A, then there is a completion of P that avoids A.*

In this paper we also consider random partial Latin squares and arrays. Let $\mathcal{P}(n, p)$ denote the probability space of all $n \times n$ partial Latin squares P where each cell $(i, j)_P$ independently is empty with probability $1 - p$ and contains symbol s with probability p/n, $s = 1, \ldots, n$. Further, for $i = 1, \ldots, n$, we empty any cell $(i, j_1)_P$ in row i that contains the same entry as another cell $(i, j_2)_P$ in row i, where $j_2 > j_1$; similarly for columns.

Using our main result we prove the following proposition on random arrays and random partial Latin squares.

Corollary 1.2. *Let P be a random PLS distributed as $\mathcal{P}(n, p)$, and let A be a random $n \times n$ array where each cell $(i, j)_A$ of A is assigned a set $A(i, j)$ of size $m = m(n)$ by choosing each set independently and uniformly at random from all m-subsets of $[n]$, and where any entry of A that occurs in the corresponding cell of P is removed. There are constants ρ_1 and ρ_2 such that if $p < \rho_1$ and $m \leqslant \rho_2 n$, then with probability tending to 1, as $n \to \infty$, there is a completion of P that avoids A.*

This result is deduced from our Theorem 1.1, and it also holds if we take P to be a given (deterministic) α-dense PLS and A a random array, or P a random PLS and A a given $(\beta n, \beta n, \beta n)$-array.

The rest of the paper is organized as follows. In Section 2 we introduce some terminology and notation and also outline the proof of Theorem 1.1. Section 3 contains the proof of a slightly reformulated version of Theorem 1.1. In Section 4 we prove Corollary 1.2, and in Section 5 we give some concluding remarks; in particular, we give an example indicating what numerical values of α and β in Theorem 1.1 might be best possible. In the beginning of Section 3 we shall present numerical values of α and β for which our main theorem holds, provided that n is large enough.

2. Terminology, notation and outline of the proof of Theorem 1.1

If L is a Latin square, A is an array, and L does not avoid A, then the cells $(i, j)_L$ such that $L(i, j) \in A(i, j)$ are the *conflict cells of L with A* (or just the *conflicts* of L). If P is a PLS, then the cells $(i, j)_L$ that correspond to non-empty cells in P are the *prescribed cells of L with P* (or just the *prescribed cells*).

An *intercalate* in an $n \times n$ Latin square L is a set

$$C = \{(r_1, c_1)_L, (r_1, c_2)_L, (r_2, c_1)_L, (r_2, c_2)_L\}$$

of cells in L such that $L(r_1, c_1) = L(r_2, c_2)$ and $L(r_1, c_2) = L(r_2, c_1)$. If in addition

$$|\{L(r_1, c_1), L(r_1, c_2)\} \cap \{1, \ldots, \lfloor n/2 \rfloor\}| = 1,$$

then C is called a *strong intercalate*.

If

$$C = \{(r_1, c_1)_L, (r_1, c_2)_L, (r_2, c_1)_L, (r_2, c_2)_L\}$$

is an intercalate in L with $L(r_1, c_1) = s_1$ and $L(r_1, c_2) = s_2$, then a *swap on C* is the operation $L \mapsto L'$, where L' is a Latin square with

$$L'(r_1, c_1) = L'(r_2, c_2) = s_2, \; L'(r_1, c_2) = L'(r_2, c_1) = s_1,$$

and $L'(i, j) = L(i, j)$ for all other (i, j). The intercalate C is called *allowed with respect to A* (or just *allowed*) if performing a swap on it yields a Latin square L' in which none of the cells in

$$\{(r_1, c_1)_{L'}, (r_1, c_2)_{L'}, (r_2, c_1)_{L'}, (r_2, c_2)_{L'}\}$$

is a conflict cell of L' with A.

Let T be some set of cells from a Latin square L. If there is a Latin square L' satisfying

- $L'(i, j) = L(i, j)$ if $(i, j)_L \notin T$, and
- $L'(i, j) \neq L(i, j)$ for some $(i, j)_L \in T$,

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

then we say that L' is obtained from L *by performing a trade on* T. We will also refer to the set T as a trade. Note that a swap on an intercalate may be seen as performing a trade on the intercalate.

A *generalized diagonal* $\mathcal{D}$, or simply a *diagonal*, in an array A of order n is a set of n cells in A, such that no two cells of $\mathcal{D}$ are in the same row or column of A. The *main diagonal* in A is the diagonal $\{(i,i)_A : i \in [n]\}$. A *transversal* of a Latin square L of order n is a diagonal $\mathcal{D}$ in L such that that $\{L(r,c) : (r,c)_L \in \mathcal{D}\} = [n]$.

For the proof of Theorem 1.1, we need some previous results. The following is due to Brègman [8] (see also [5], p. 22).

Theorem 2.1. *If* $A = [A(i,j)]$ *is an* $n \times n$ $(0,1)$*-matrix with row sum* r_i *on the ith row, then the permanent* $\operatorname{per}(A)$ *of* A *satisfies*

$$\operatorname{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A(i,\sigma(i)) \leqslant \prod_{1 \leqslant i \leqslant n} (r_i!)^{1/r_i},$$

where S_n *is the symmetric group of order* n.

By a simple correspondence between (0, 1)-matrices and bipartite graphs, we get the following corollary.

Corollary 2.2. *If* B *is a balanced bipartite graph on* $2n$ *vertices and* $d_1, \ldots, d_n$ *are the degrees of the vertices in one part of* B*, then the number of perfect matchings in* B *is at most* $\prod_{1 \leqslant i \leqslant n} (d_i!)^{1/d_i}$.

We also need some definitions on list edge colouring. Given a graph G, assign to each edge e of G a set $\mathcal{L}(e)$ of colours (positive integers). Such an assignment $\mathcal{L}$ is called a *list assignment* for G and the sets $\mathcal{L}(e)$ are referred to as *lists* or *colour lists*. If all lists have equal size k, then $\mathcal{L}$ is called a *k-list assignment*. Usually, we seek a proper edge colouring φ of G, such that $\varphi(e) \in \mathcal{L}(e)$ for all $e \in E(G)$. If such a colouring φ exists then G is $\mathcal{L}$*-colourable* and φ is called an $\mathcal{L}$*-colouring*. Let $\chi'_L(G)$ denote the minimum integer t such that G is $\mathcal{L}$-colourable whenever $\mathcal{L}$ is a t-list assignment. We let $\chi'(G)$ denote the chromatic index of G, *i.e.* the minimum integer t such that G has a proper t-edge colouring.

Note further that the main result of this paper can be formulated as a theorem on list edge colouring of balanced complete bipartite graphs.

Instead of proving Theorem 1.1 we will prove the following theorem, which is easily seen to imply Theorem 1.1.

Theorem 2.3. *There are constants* $\alpha > 0$, $\beta > 0$ *and* n_0*, such that, for every positive integer* $n \geqslant n_0$*, if* P *is an* α*-dense partial Latin square of order* n*,* A *is a* $(\beta n, \beta n, \beta n)$*-array of order* n*, and no entry of* P *appears in the corresponding cell of* A*, then there is a completion* L *of* P *that avoids* A.

The proof of Theorem 2.3 combines techniques from [4] and [7]. In particular, the last part of the proof is an extension of the technique developed by Bartlett for completing α-dense PLS.

Below we outline the proof of Theorem 2.3.

Step I. Find a 'starting Latin square' L_0 of order n, such that each cell in L_0 except at most $3n + 7$ is in $\lfloor n/2 \rfloor$ strong intercalates.

Step II. Given A and P, find a pair of permutations (ρ, θ) so that if A' and P' denote the arrays obtained from A and P, respectively, by applying ρ to the rows of A and P and θ to the columns of A and P, then P' and A' satisfy certain 'sparsity' conditions with respect to L_0. These conditions will be articulated more precisely below.

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

Step III. Define an $n \times n$ PLS R such that a cell of R is non-empty if and only if the corresponding cell of L_0 is a conflict cell with A' and the corresponding cell of P' is empty. We shall also require that P and R together form a PLS, and that each symbol in R does not appear in too many cells in R. Let $\hat{P}$ be the PLS obtained by putting P' and R together.

Step IV. Apply our modified variant of the technique by Bartlett [7] to construct from L_0 a Latin square L_q that is a completion of $\hat{P}$ (and thus P') and which avoids A'.

The above construction yields a Latin square L_q that is a completion of P' and which avoids A'. However, in order to obtain a Latin square S_q from L_q that is a completion of P and which avoids A, we can just apply the inverses of the permutations ρ and θ to the rows and columns of L_q, respectively. Hence, it suffices to prove that there is a Latin square L_q as above.

3. Proof of Theorem 2.3

In the proof of Theorem 2.3 we shall verify that it is possible to perform Steps I–IV described in Section 2 to obtain the Latin square L_q. We will not specify the value of n_0 in the proof, but rather assume that n is large enough whenever necessary. Since the proof of the theorem will contain a finite number of inequalities that are valid if n is large enough, this suffices for proving Theorem 2.3.

The proof of Theorem 2.3 involves a number of other functions and parameters,

$$\alpha, \beta, c(n), f(n), d, k, \varepsilon,$$

and a number of inequalities that they must satisfy. For the reader's convenience, explicit choices for which the proof holds are presented here:

$$\alpha = \frac{1}{100000}, \quad \beta = \frac{1}{100000}, \quad k = \frac{1}{500}, \quad \varepsilon = \frac{1}{10000},$$
$$d = \frac{1}{20}, \quad c(n) = \left\lfloor \frac{n}{35000} \right\rfloor, \quad f(n) = \left\lfloor \frac{n}{17500} \right\rfloor.$$

We remark that since the numerical values of α and β are nowhere near what we expect to be optimal, we have not put any effort into choosing optimal values for these parameters.

Proof of Theorem 2.3. Let P be an $n \times n$ α-dense PLS and A an $n \times n$ $(\beta n, \beta n, \beta n)$-array such that no cell of A contains a symbol that occurs in the corresponding cell of P.

Step I. Below we shall define the *starting Latin square* L_0. This Latin square was used in [4] and [7] and also appears in the original paper by Chetwynd and Häggkvist [12] on completing sparse partial Latin squares.

We shall give the explicit construction assuming that n is even. For the case when n is odd, one can modify the construction in the even case by swapping on some intercalates and using a transversal; the details are given in Lemma 2.1 in [7].

So suppose that $n = 2r$.

Definition. Let M_{11} be the cyclic Latin square of order r (*i.e.* the Latin square corresponding to the addition table of the cyclic group of order r). Note that $M_{11}(i,j) = j - i + 1$, taking $j - i + 1$ modulo r. The $r \times r$ array M_{12} is defined from M_{11} by setting $M_{12}(i,j) = M_{11}(i,j) + r$, $1 \leqslant i, j \leqslant r$. Let $M_{21} = M_{12}^{\mathrm{T}}$ and $M_{22} = M_{11}^{\mathrm{T}}$, where M^{T} is the transpose of M, defined in the obvious way:

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

$$M_{11} = \begin{array}{|c|c|cccc|c|}
\hline
1 & 2 & 3 & \cdots & & r-1 & r \\
\hline
r & 1 & 2 & \cdots & & r-2 & r-1 \\
\hline
r-1 & r & 1 & \cdots & & r-3 & r-2 \\
\vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\
3 & 4 & 5 & \cdots & & 1 & 2 \\
\hline
2 & 3 & 4 & \cdots & & r & 1 \\
\hline
\end{array}\,.$$

Now we define the $2r \times 2r$ Latin square M by letting

- M_{11} be the $r \times r$ subarray in its upper left corner,
- M_{12} be the $r \times r$ subarray in its upper right corner,
- M_{21} be the $r \times r$ subarray in its lower left corner, and
- M_{22} be the $r \times r$ subarray in its lower right corner:

$$M = \begin{array}{|c|c|}
\hline
M_{11} & M_{12} \\
\hline
M_{21} & M_{22} \\
\hline
\end{array}\,.$$

Every cell in M belongs to a large number of strong intercalates.

Lemma 3.1. *Each cell $(i, j)_M$ in M belongs to exactly r distinct strong intercalates.*

Proof. Without loss of generality, we assume that $1 \leqslant i, j \leqslant r$. It is easy to verify that for every $l \in \{1, \dots, r\}$,

$$\{(i, j)_M, (i, r+l)_M, (r+l+j-i, j)_M, (r+l+j-i, r+l)_M\}$$

is a strong intercalate in M. Hence each cell $(i, j)_M$ is in at least r strong intercalates, and since a strong intercalate is uniquely determined by two cells, it follows from the definition of M that each cell is in at most r strong intercalates. □

The case when $n = 2r + 1$ is not as elegant; as mentioned above, using the Latin square M one can construct a Latin square M' of order $2r + 1$ such that all but at most $3n + 7$ cells are in $\lfloor n/2 \rfloor$ strong intercalates. In particular, there is a row and column in M' where no cell belong to at least $\lfloor n/2 \rfloor$ strong intercalates. The full proof appears in [7] and therefore we omit the details here.

We define $L_0 := M$ when n is even, and $L_0 := M'$ when n is odd.

Step II. Let A' be an $n \times n$ $(\beta n, \beta n, \beta n)$-array, P' an $n \times n$ α-dense PLS and L a Latin square. If the following conditions hold, then L is *well-behaved* with respect to A' and P' (or just *well-behaved* when A' and P' are clear from the context):

(a) all cells in L, except for $3n + 7$, belong to at least $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates;
(b) each row of L contains at most $c(n)$ conflicts with A';
(c) each column of L contains at most $c(n)$ conflicts with A';
(d) for each symbol $s \in [n]$ there are at most $c(n)$ cells in L that contain s and that are conflicts with A';

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

(e) for each symbol $s \in [n]$ there are at most $c(n)$ cells in L that contain s and satisfy that the corresponding cell in P' is non-empty;

(f) for each pair of symbols $s_1, s_2 \in [n]$ there are at most $c(n)$ cells in L with entry s_1 such that s_2 belongs to the corresponding cell in A'.

We shall prove that there is a pair of permutations (ρ, θ) such that if ρ is applied to the rows of the given arrays A and P, and θ is applied to the columns of A and P, then the resulting arrays A' and P', respectively, satisfy that the starting Latin square L_0 is well-behaved with respect to A' and P'.

If J is a subset of cells of an array S, and S' is the array obtained from S by applying ρ to the rows of S and θ to the columns of S, then $\rho(\theta(J))$ denotes the set of cells in S' that J are mapped to under ρ and θ.

Following [4], we shall for convenience in fact prove that there are permutations σ, τ such that if S is the Latin square obtained from L_0 by applying σ to the rows and τ to the columns of L_0, then L_0, S, A and P satisfy the following:

(a′) all cells in S except for $3n+7$ are in at least $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates;

(b′) for a collection $J_1, \dots, J_{3n}$ of $3n$ given n-sets of cells in L_0, each J_i satisfies that the corresponding n-set $\sigma(\tau(J_i))$ of cells in S has at most $c(n)$ conflicts with A;

(c′) for a collection $J_1, \dots, J_n$ of n given n-sets of cells in L_0, each J_i satisfies that the corresponding n-set $\sigma(\tau(J_i))$ of cells in S contains at most $c(n)$ prescribed cells;

(d′) for a collection $J_1, \dots, J_n$ of n given n-sets of cells in L_0 and each symbol $s \in \{1, \dots, n\}$, each J_i satisfies that the corresponding n-set $\sigma(\tau(J_i))$ of cells in S contains at most $c(n)$ cells such that s is in the corresponding cell of A.

It is straightforward to deduce that if the above conditions hold, then if we let P' and A' denote the arrays obtained from P and A, respectively, by applying the inverses of σ and τ to the rows and columns, respectively, of P and A, then L_0 is well-behaved with respect to P' and A'; if (a′) holds, then clearly (a) is true for L_0, P' and A' as well; and if (b′) is true, then by taking the $3n$ n-sets in (b′) to be the sets of the cells in a particular row or column, or containing a particular symbol, we deduce that (b), (c) (d) hold for L_0, A' and P'. That (e) and (f) are true is deduced similarly from the fact that (c′) and (d′) hold.

Now, let L_0 be the starting Latin square defined above, and let σ and τ be two permutations chosen independently and uniformly at random from all $n!$ permutations of $\{1, \dots, n\}$. Let S denote a random Latin square obtained from L_0 by applying σ to the rows of L_0 and τ to the columns of L_0.

Lemma 3.2. *If*

$$\left(\frac{2\beta}{\varepsilon - 2\beta}\right)^{\varepsilon - 2\beta} \left(\frac{1}{1 - 2\varepsilon + 4\beta}\right)^{1/2 - \varepsilon + 2\beta} < 1,$$

and $\varepsilon > 2\beta$, *then the probability that* S *fails condition* (a′) *tends to* 0 *as* $n \to \infty$.

Proof. We bound the number of pairs (σ, τ) such that there is at least one cell, except the $3n+7$ excluded, which does not belong to at least $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates.

There are at most n^2 cells that can belong to too few allowed strong intercalates in S; choose such a cell $(r', c')_S$. Next, we fix τ by choosing one out of $n!$ possible permutations for τ. Assume that $c' = \tau(c)$.

With τ fixed, we now count how many ways σ can be chosen so that the cell $(r', c')_S$ belongs to less than $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates.

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

There are n choices for a row r in L_0 so that $\sigma(r) = r'$. This choice partitions the rows of L_0 into two sets: the set Q of rows r^* for which $\{(r, c)_{L_0}, (r, c^*)_{L_0}, (r^*, c)_{L_0}, (r^*, c^*)_{L_0}\}$ is a strong intercalate in L_0 for some $c^* \neq c$, and its complement $\bar{Q}$. Note that $|Q| = \lfloor n/2 \rfloor$.

Note further that choosing the row r in L_0 so that $\sigma(r) = r'$, determines the value of $s = L_0(r, c)$. When row r and thus $S(r', c')$ is fixed, there are at most βn columns c_1 such that $S(r', c') \in A(r', c_1)$. Furthermore, at most βn columns c_2 satisfy $S(r', c_2) \in A(r', c')$. Consequently, if there are less than $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates containing $(r', c')_S$ in S, then there have to be at least $\varepsilon n - 2\beta n$ strong intercalates in S containing $(r', c')_S$ that are not allowed because swapping on them would cause a conflict in another row than r'. (Note that $(\varepsilon - 2\beta) > 0$ by assumption.) The number of ways of choosing σ so that in S at least $(\varepsilon - 2\beta)n$ of the strong intercalates containing $(r', c')_S$ satisfy this condition can be estimated in the following way. Let W be the set of rows in S to which σ maps Q. There are $\binom{n-1}{\lfloor n/2 \rfloor}$ ways of choosing W. After choosing W we can now choose how σ acts on $\bar{Q} \setminus \{r\}$ in any of the at most $(\lceil n/2 \rceil)!$ possible ways. Next, we choose a subset $V \subseteq Q$ of size $\lceil (\varepsilon - 2\beta)n \rceil$. If we set $p(n) = \lceil (\varepsilon - 2\beta)n \rceil$, then this can be done in at most $\binom{\lfloor n/2 \rfloor}{p(n)}$ ways.

Now we define a bipartite graph G_1 with parts $Q = \{r_1, \ldots, r_{\lfloor n/2 \rfloor}\}$ and $W = \{r'_1, \ldots, r'_{\lfloor n/2 \rfloor}\}$. Include an edge between r_i and r'_j in G_1 if and only if

- $r_i \notin V$, or
- $r_i \in V$ and $\sigma(r_i) = r'_j$ implies that the strong intercalate

  $$\{(r', c')_S, (r', \tau(c_q))_S, (r'_j, c')_S, (r'_j, \tau(c_q))_S\}$$

  is not allowed in S because swapping on it yields a conflict in row r'_j, where c_q is the unique column such that

  $$\{(r, c)_{L_0}, (r, c_q)_{L_0}, (r_i, c)_{L_0}, (r_i, c_q)_{L_0}\}$$

  is a strong intercalate in L_0.

A perfect matching in G_1 corresponds to choosing σ so that at least $(\varepsilon - 2\beta)n$ strong intercalates in S containing $(r', c')_S$ are not allowed because swapping on them yields conflicts on other rows than r'.

The degree of a vertex in V is at most $2\beta n$, because the symbols $L(r, c)$ and $L(r, c_q)$ each occur at most βn times in columns $\tau(c_q)$ and $\tau(c) = c'$ in A, respectively. The degree of a vertex in $Q \setminus V$ is $\lfloor n/2 \rfloor$. Hence, by Corollary 2.2, there are at most

$$(\lfloor 2\beta n \rfloor !)^{p(n)/\lfloor 2\beta n \rfloor} \, (\lfloor n/2 \rfloor !)^{(\lfloor n/2 \rfloor - p(n))/\lfloor n/2 \rfloor}$$

perfect matchings in G_1.

So the probability that S fails condition (a′) is at most

$$\frac{n^2 n! n \binom{n-1}{\lfloor n/2 \rfloor} \lceil n/2 \rceil ! \binom{\lfloor n/2 \rfloor}{p(n)} (\lfloor 2\beta n \rfloor !)^{p(n)/\lfloor 2\beta n \rfloor} \, (\lfloor n/2 \rfloor !)^{(\lfloor n/2 \rfloor - p(n))/\lfloor n/2 \rfloor}}{(n!)^2}$$
$$\leqslant \frac{n^3 \, (\lfloor 2\beta n \rfloor !)^{p(n)/\lfloor 2\beta n \rfloor} \, (\lfloor n/2 \rfloor !)^{(\lfloor n/2 \rfloor - p(n))/\lfloor n/2 \rfloor}}{p(n)!(\lfloor n/2 \rfloor - p(n))!}.$$

By applying Stirling's formula, this expression tends to zero as $n \to \infty$, if

$$\left(\frac{2\beta}{\varepsilon - 2\beta}\right)^{\varepsilon - 2\beta} \left(\frac{1}{1 - 2\varepsilon + 4\beta}\right)^{1/2 - \varepsilon + 2\beta} < 1,$$

which holds by assumption. □

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

Lemma 3.3. *Let*

$$J = \{(r_1, c_1)_{L_0}, \ldots, (r_n, c_n)_{L_0}\}$$

be a set of n cells in L_0 and denote

$$J' = \{(r'_1, c'_1)_S, \ldots, (r'_n, c'_n)_S\},$$

where $\sigma(r_i) = r'_i$ and $\tau(c_i) = c'_i$, $i = 1, \ldots, n$. Then the following hold for some positive constants C and a:

(i) *the probability that J' has at least $c(n)$ conflicts with A is at most*

$$Cn^a\left(\frac{\beta(n-c(n))}{c(n)}\right)^{c(n)}\left(\frac{n}{n-c(n)}\right)^n,$$

(ii) *the probability that J' contains at least $c(n)$ prescribed cells is at most*

$$Cn^a\left(\frac{\alpha(n-c(n))}{c(n)}\right)^{c(n)}\left(\frac{n}{n-c(n)}\right)^n,$$

(iii) *for a given symbol s, the probability that J' contains at least $c(n)$ cells such that the corresponding cell in A contains s is at most*

$$Cn^a\left(\frac{\beta(n-c(n))}{c(n)}\right)^{c(n)}\left(\frac{n}{n-c(n)}\right)^n.$$

Proof. We first prove (i). We estimate the number of pairs (σ, τ) such that at least $c(n)$ cells from J' are conflict cells with A. There are $n!$ ways of choosing the permutation σ. Fix such a permutation σ and suppose that $\sigma(r_i) = r'_i$, $i = 1, \ldots, n$.

Let K be a subset of J such that $|K| = c(n)$ and all cells in K are mapped to conflict cells by (σ, τ). Such a set K can be chosen in $\binom{n}{c(n)}$ ways. The number of ways of choosing τ so that $(r'_i, c'_i)_S$ is a conflict cell whenever $(r_i, c_i)_{L_0} \in K$ can be estimated by considering a bipartite graph G_2 as follows: the parts of G_2 are J and $\{1, \ldots, n\}$ and there is an edge between $(r_i, c_i)_{L_0} \in J$ and $j \in \{1, \ldots, n\}$ if

- $(r_i, c_i)_{L_0} \notin K$, or
- $(r_i, c_i)_{L_0} \in K$ and $L_0(r_i, c_i) \in A(r'_i, j)$.

Note that if $(r_i, c_i)_{L_0} \in K$ then the degree of $(r_i, c_i)_{L_0}$ in G_2 is at most βn, because the symbol $L_0(r_i, c_i)$ occurs at most βn times in row r'_i in A. If $(r_i, c_i)_{L_0} \notin K$, then the degree of $(r_i, c_i)_{L_0}$ is n.

A perfect matching in G_2 corresponds to a choice of τ so that all cells in K are mapped to conflict cells of S. By Corollary 2.2, the number of perfect matchings in G_2 is at most

$$(\lfloor\beta n\rfloor!)^{c(n)/\lfloor\beta n\rfloor}(n!)^{(n-c(n))/n}.$$

So the probability that J' has at least $c(n)$ conflicts with A is at most

$$\begin{aligned}&\frac{n!\binom{n}{c(n)}(\lfloor\beta n\rfloor!)^{c(n)/\lfloor\beta n\rfloor}(n!)^{(n-c(n))/n}}{(n!)^2}\\ &\quad = \frac{(\lfloor\beta n\rfloor!)^{c(n)/\lfloor\beta n\rfloor}(n!)^{(n-c(n))/n}}{c(n)!(n-c(n))!}\\ &\quad \leqslant Cn^a\left(\frac{\beta(n-c(n))}{c(n)}\right)^{c(n)}\left(\frac{n}{n-c(n)}\right)^n,\end{aligned}$$

where C and a are some positive constants.

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

The proof of (ii) is almost identical to the proof of (i), the only difference being that one uses the property that each row in P has at most αn non-empty cells, instead of the property that each symbol occurs at most βn in each row of A. The details are omitted.

The proof of (iii) is also almost identical to the proof of (i) above except that one uses the property that a fixed symbol s occurs at most βn times in each row of A. Here as well the details are omitted. □

Lemma 3.4. *If*

$$\alpha < \frac{c(n)}{n-c(n)}\left(\frac{n-c(n)}{n}\right)^{n/c(n)}, \quad \beta < \frac{c(n)}{n-c(n)}\left(\frac{n-c(n)}{n}\right)^{n/c(n)},$$

then the probability that S fails condition (b′), (c′) *or* (d′) *tends to* 0 *as* $n \to \infty$.

Proof. Let $J_1, \ldots, J_{3n}$ be $3n$ given n sets of cells in L_0. By part (i) of Lemma 3.3, the probability that J_i has at least $c(n)$ conflicts with A is at most

$$p_1 = Cn^a\left(\frac{\beta(n-c(n))}{c(n)}\right)^{c(n)}\left(\frac{n}{n-c(n)}\right)^n,$$

where C and a are some positive constants. Since $3np_1 \to 0$ as $n \to \infty$, it follows that the probability that S fails condition (b′) tends to zero as $n \to \infty$. That the probability that S fails condition (c′) or (d′) tends to zero can be proved similarly using parts (ii) and (iii) of Lemma 3.3. □

We conclude from the preceding lemmas that there are permutations (σ, τ) such that if S is obtained from L_0 by applying σ to the rows of L_0, and τ to the columns of L_0, then S satisfies (a′), (b′), (c′) and (d′). Hence, if we let A' and P' denote the arrays obtained from A and P, respectively, by applying σ^{-1} to the rows and τ^{-1} to the columns, then L_0 is well-behaved with respect to A' and P'.

Step III. By the preceding step, we may assume that the starting Latin square L_0 is well-behaved with respect to the array A' and the PLS P' defined above. We shall define a PLS R such that a cell in R is non-empty if and only if the corresponding cell of L_0 is a conflict cell with A' and the corresponding cell of P' is empty. We shall also require that R and P' together form a PLS.

Consider a bipartite graph G_3, where the rows and columns of L_0 are the vertices of the partite sets of G_3, and the conflict cells of L_0 defines the edge set of G_3, that is, there is an edge between two vertices in G_3 if the corresponding cell of L_0 is a conflict with A'.

We want to find a proper n-colouring of $E(G_3)$ satisfying that if R is the PLS corresponding to this edge colouring of G_3 (by taking the partite sets of G_3 to be the rows and columns of R, and the coloured edges of G_3 as the non-empty cells of R), then R contains at most $c(n)$ entries in each row and column and each symbol in R is used at most $f(n)$ times. If, in addition, P and R together form a PLS, then each row and column in this PLS is used at most $\alpha n + c(n)$ times, and each symbol is used at most $\alpha n + f(n)$ times.

We may assume that there is no conflict cell in L_0 such that the corresponding cell in P' is non-empty, because then we just remove this cell from the set of conflict cells. We define a list assignment $\mathcal{L}$ for G_3 for every symbol (colour) $c \in \{1, \ldots, n\}$ and every edge $e = ij$ including c in $\mathcal{L}(e)$ if and only if $c \notin A'(i, j)$ and c does not appear in row i or column j in P'. Clearly,

$$\mathcal{L}(e) \geqslant n - \beta n - 2\alpha n,$$

for every edge e of G_3. Our goal is to find an $\mathcal{L}$-colouring ϕ of $E(G_3)$ such that each colour appears on at most $f(n)$ edges. Such a colouring of G_3 corresponds to a PLS R satisfying the conditions stipulated above.

The maximum degree in G_3 is $c(n)$, because each row and column in L_0 contains at most $c(n)$ conflict cells (by conditions (b) and (c) above). Now, the required colouring ϕ of $E(G_3)$ can be obtained greedily: suppose that we have constructed a partial colouring of the edges of G_3 and let e be some hitherto uncoloured edge of G_3. The number of colours that have been used at least $f(n)$ times in the hitherto constructed colouring is at most $nc(n)/f(n)$. Moreover, there are at most $2c(n)$ distinct colours that are used on edges which are adjacent to e. Hence, we can select a colour for e from its list so that the resulting colouring is proper if

$$n - \beta n - 2\alpha n - 2c(n) - \frac{nc(n)}{f(n)} \geqslant 1,$$

which holds by assumption. We conclude that the required colouring ϕ exists and thus also the required PLS R.

Let $\hat{P}$ be the PLS obtained by putting P' and R together. The PLS $\hat{P}$ satisfies the following:

- (a″) $\hat{P}$ contains at most $\alpha n + c(n)$ entries in each row or column;
- (b″) each symbol is used at most $\alpha n + f(n)$ times in $\hat{P}$.

Furthermore, since L_0 is well-behaved with respect to A' and P', it satisfies the following conditions with respect to A' and $\hat{P}$:

- (c″) each cell in L_0 (except for $3n + 7$) belongs to at least $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates;
- (d″) each row and column of L_0 contains at most $\alpha n + c(n)$ prescribed cells;
- (e″) for each symbol s, there are at most $2c(n)$ prescribed cells in L_0 with entry s;
- (f″) for each pair of symbols s_1, s_2, there are at most $c(n)$ cells in L_0 with entry s_1 such that s_2 appears in the corresponding cell in A'.

Step IV. Let $\hat{P}$ be the PLS obtained in the previous step, and A', P' and L_0 as above. In this section, all prescribed cells of a Latin square are taken with respect to $\hat{P}$.

Let L be a Latin square obtained from the starting Latin square L_0 by performing a sequence of trades. We say that a cell $(i, j)_L$ in L is *L-disturbed* if $(i, j)_L$ appears in a trade which is used for obtaining L from L_0, or if $(i, j)_{L_0}$ is one of the original at most $3n + 7$ cells in L_0 that do not belong to at least $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates in L_0. Moreover, for a constant $d > 0$, we say that a row or column r or symbol s is *d-overloaded* if more than dn entries in row or column r or with symbol s has been involved in the trades that have transformed L_0 into L.

In this step we describe a modified variant of the machinery developed in [7] for completing sparse partial Latin squares. The main difference is that we have to make sure that no trades will cause any ‘new’ conflict cells with A'. In particular, the intercalates that we will swap on will be allowed with respect to A'. Another difference is that all symbols used in the trade created by Lemma 3.5 below (our version of Lemma 2.2 in [7]) are not d-overloaded. Apart from these differences, the proofs in this section are almost identical to the ones in [7], so in general, proofs are sketched, rather than given in full detail. Also, we omit many verifications which can be done exactly as in [7] (or [4] in some cases).

We will define a sequence of Latin squares $L_0, \ldots, L_q$, where L_i is obtained from L_{i-1}, $i = 1, \ldots, q - 1$, by performing some trade T_i. The trade T_i will contain (at least) one prescribed cell $(r, c)_{L_{i-1}}$ such that $L_{i-1}(r, c) \neq \hat{P}(r, c)$, $L_i(r, c) = \hat{P}(r, c)$, and, furthermore, all conflict cells of L_i will be prescribed cells (r', c') such that $\hat{P}(r', c') \neq L_{i-1}(r', c')$, that is, the trade T does not create any ‘new’ conflict cells.

In the following we shall refer to the ‘lower half’ and ‘upper half’ of an array L; by these expressions we mean the subarray of L consisting of the first $\lfloor n/2 \rfloor$ rows of L and the subarray consisting

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

of the last $\lceil n/2 \rceil$ rows of L, respectively. We also assume that if n is odd, then the row and column of L_0 where no cells are in at least $\lfloor n/2 \rfloor$ strong intercalates are the last row and column of L_0, respectively.

The following lemma is essentially a strengthened variant of Lemma 2.2. in [7].

Lemma 3.5. *Let L_0, $\hat{P}$ and A' be as above. Suppose that L is an $n \times n$ Latin square obtained from L_0 by performing some sequence of trades on L_0, and that at most kn^2 cells in L are L-disturbed, for some constant $k > 0$.*

Let $\{t_1, \ldots, t_a\}$ be a set of a symbols from L.

If

$$\left\lfloor \frac{n}{2} \right\rfloor - 2\varepsilon n - 6dn - 5\frac{k}{d}n - 4\alpha n - 8c(n) - 3a - 3\beta n > 6,$$

then for any row r_1 of L and all but at most

- $2(k/d)n + \alpha n + c(n) + a$ *choices of* c_1, *and*
- $a + 1 + 4c(n) + 2\beta n + 4(k/d)n + 2\alpha n + 2dn$ *choices of* c_2,

there is a trade on a set of cells T such that, if we let L' denote the Latin square obtained from L by performing this trade on T, then L' satisfies the following:

- *the trade T uses only symbols that are not d-overloaded;*
- *no prescribed cells of L are in T;*
- *L and L' differ on at most 16 cells (i.e. T uses at most 16 cells);*
- *no cell with entry $\{t_1, \ldots, t_a\}$ in L is in T;*
- $L'(r_1, c_1) = L(r_1, c_2)$ *and* $L'(r_1, c_2) = L(r_1, c_1)$;
- *if there is a conflict of L' with A', then the corresponding cell of L is also a conflict with A'.*

Proof. Consider a given row r_1. We choose a column c_1 in L, such that the following properties hold.

- Column c_1 is not d-overloaded, and the symbol $s_1 = L(r_1, c_1)$ is not overloaded. This eliminates at most $2(k/d)n$ choices.
- The cell $(r_1, c_1)_L$ is not a prescribed cell. This eliminates at most $\alpha n + c(n)$ choices.
- The symbol s_1 is not one of $\{t_1, \ldots, t_a\}$. This eliminates at most a choices.

Summing up, we have at least

$$n - 2\frac{k}{d}n - \alpha n - c(n) - a$$

choices for c_1; by assumption this expression is greater than zero, so we fix such a column c_1.

Next, we choose a column c_2 in L so that the following properties hold.

- $c_2 \neq c_1$ and $s_2 = L(r_1, c_2) \notin A'(r_1, c_1)$ and $s_1 \notin A'(r_1, c_2)$. This excludes at most $1 + 2\beta n$ choices for c_2.
- Column c_2 is not d-overloaded, and the symbol $s_2 = L(r_1, c_2)$ is not d-overloaded. This eliminates at most $2(k/d)n$ choices.
- The cell $(r_1, c_2)_L$ is not a prescribed cell. This eliminates at most $\alpha n + c(n)$ choices.

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

- The cell $(r_3, c_1)_L$ in column c_1 in L containing s_2 is not L-disturbed, and the cell $(r_4, c_2)_L$ in column c_2 in L containing s_1 is not L-disturbed. Since neither the column c_1 nor the symbol s_1 is d-overloaded, this excludes at most $2dn$ choices. We also require that the cells $(r_3, c_1)_L$ and $(r_4, c_2)_L$ are not prescribed, which excludes an additional at most $3c(n) + \alpha n$ choices.
- The rows r_3, r_4 are not d-overloaded. This eliminates at most $2(k/d)n$ choices.
- $s_2 \notin \{t_1, \ldots, t_a\}$. This excludes at most a choices.

Summing up, we have at least

$$n - 4c(n) - 2\beta n - 4\frac{k}{d}n - 2\alpha n - 2dn - a - 1$$

choices for c_2; by our assumptions this expression is greater than zero, and so we fix such a column c_2 in L.

Case 1. *Both of the rows r_3 and r_4 lie either in the upper half or in the lower half of the Latin square L (and thus in L_0).*

We may assume that $r_3 \neq r_4$, since otherwise we may swap on the intercalate consisting of all hitherto considered cells, and we are done. Assuming $r_3 \neq r_4$, we now proceed as follows.

For the trade in Case 1, we shall construct two disjoint allowed strong intercalates

$$C_1 = \{(r_3, c_1)_L, (r_3, c_4)_L, (r_2, c_1)_L, (r_2, c_4)_L\}$$

and

$$C_2 = \{(r_4, c_2)_L, (r_4, c_3)_L, (r_2, c_2)_L, (r_2, c_3)_L\},$$

containing the cells $(r_3, c_1)_L$ and $(r_4, c_2)_L$, respectively. Since these two cells are not L-disturbed, they agree with L_0, and the corresponding cells in L_0 are both in at least $\lfloor n/2 \rfloor - \varepsilon n$ allowed strong intercalates in L_0, and since they lie in 'the same half' of L_0, there are at least $\lfloor n/2 \rfloor - 2\varepsilon n$ such pairs of allowed strong intercalates in L_0 containing a common row r_2. We further require the following.

- None of the cells $(r_2, c_1)_L, (r_2, c_2)_L, (r_2, c_3)_L, (r_2, c_4)_L, (r_3, c_4)_L$, or $(r_4, c_3)_L$ are L-disturbed. Because none of the rows r_3, r_4, the columns c_1, c_2 or the symbols s_1, s_2 are overloaded, this excludes at most $6dn$ choices. Note that this condition ensures that all cells of C_1 and C_2 have the same entry in L as the corresponding cells of L_0.
- None of the cells above are prescribed. This excludes at most $4(\alpha n + c(n)) + 4c(n)$ choices.
- Neither $s_3 = L(r_2, c_1)$ or $s_4 = L(r_2, c_2)$ is in $\{t_1, \ldots, t_a\}$. This eliminates at most $2a$ choices.
- The symbols s_3 and s_4 are not d-overloaded. This excludes at most $2(k/d)n$ choices.
- $s_1 \notin A'(r_2, c_1)$ and $s_2 \notin A'(r_2, c_2)$. This eliminates at most $2\beta n$ choices.

Summing up, we have at least

$$\left\lfloor \frac{n}{2} \right\rfloor - 2\varepsilon n - 6dn - 4\alpha n - 8c(n) - 2a - 2\frac{k}{d}n - 2\beta n$$

choices for the required intercalates C_1 and C_2. Since this expression is greater than zero, we choose two such disjoint intercalates, C_1 and C_2.

By swapping on C_1 and C_2 we obtain a Latin square $L^{(1)}$. Note that the set

$$\{(r_1, c_1)_{L^{(1)}}, (r_1, c_2)_{L^{(1)}}, (r_2, c_1)_{L^{(1)}}, (r_2, c_2)_{L^{(1)}}\}$$

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

is an allowed intercalate in $L^{(1)}$ and by swapping on this intercalate we obtain the required Latin square L'. This completes the proof of the lemma in Case 1.

Case 2. *One of rows r_3 and r_4 occur in the upper half and the other one in the lower half of the Latin square L.*

Suppose without loss of generality that r_3 lies in the lower half of L and that r_4 lies in the upper half of L. We will construct several intercalates for the trade in Case 2. To begin with we construct an allowed strong intercalate

$$C_3 = \{(r_4, c_2)_L, (r_4, c_3)_L, (r_2, c_2)_L, (r_2, c_3)_L\},$$

containing the cell $(r_4, c_2)_L$ such that the following holds.

- None of the cells $(r_2, c_1)_L, (r_2, c_2)_L, (r_2, c_3)_L, (r_4, c_3)_L$ are L-disturbed. Because neither row r_4, nor columns c_1, c_2, nor symbols s_1, are d-overloaded, this eliminates at most $4dn$ choices.
- If we let $(r_2, c_4)_L$ denote the cell in row r_2 containing s_2, then $(r_2, c_4)_L$ and $(r_3, c_4)_L$ are not L-disturbed. This excludes at most $2dn$ choices.
- The symbols $s_3 = L(r_2, c_1)$, $s_4 = L(r_2, c_2)$ and $s_5 = L(r_3, c_4)$ are not d-overloaded, nor are row r_2 or column c_4, and these new cells are disjoint from the ones previously included in our trade. This eliminates at most $5(k/d)n + 2$ choices.
- None of the cells above are prescribed. This eliminates at most $4(\alpha n + c(n)) + 4c(n)$ choices.
- None of the symbols s_3, s_4, s_5 is in $\{t_1, \ldots t_a\}$. This eliminates at most $3a$ choices.
- $s_1 \notin A'(r_2, c_1)$, $s_2 \notin A'(r_2, c_2) \cup A'(r_3, c_4)$. This eliminates at most $3\beta n$ choices.

Since there are at least $\lfloor n/2 \rfloor - \varepsilon n$ strong intercalates in L_0 containing $(r_4, c_2)_{L_0}$, we have at least

$$\left\lfloor \frac{n}{2} \right\rfloor - \varepsilon n - 6dn - 5\frac{k}{d}n - 2 - 4\alpha n - 8c(n) - 3a - 3\beta n$$

choices for the required intercalate C_3. By assumption this expression is greater than zero, and we choose such an intercalate C_3.

Now, note that since r_4 lies in the upper half of L, r_2 lies in the lower half of L. Since r_3 also lies in the lower half of L, and none of the cells $(r_3, c_4)_L$, $(r_2, c_4)_L$, $(r_3, c_1)_L$ and $(r_2, c_1)_L$ are L-disturbed, and $L(r_3, c_1) = L(r_2, c_4) = s_2$, it follows that in L_0 there are at least $\lfloor n/2 \rfloor - 2\varepsilon n$ pair of allowed disjoint strong intercalates

$$C_4^{L_0} = \left\{(r_2, c_1)_{L_0}, (r_2, c_6)_{L_0}, (r_6, c_1)_{L_0}, (r_6, c_6)_{L_0}\right\}$$

and

$$C_5^{L_0} = \left\{(r_3, c_4)_{L_0}, (r_3, c_5)_{L_0}, (r_5, c_4)_{L_0}, (r_5, c_5)_{L_0}\right\}$$

containing $(r_2, c_1)_{L_0}$ and $(r_3, c_4)_{L_0}$, respectively, and such that $L_0(r_6, c_1) = L_0(r_3, c_5)$.

We choose such a pair

$$C_4 = \{(r_2, c_1)_L, (r_2, c_6)_L, (r_6, c_1)_L, (r_6, c_6)_L\}$$

and

$$C_5 = \{(r_3, c_4)_L, (r_3, c_5)_L, (r_5, c_4)_L, (r_5, c_5)_L\}$$

of intercalates in L such that the following holds.

- None of the cells in these intercalates are L-disturbed. Because the columns c_1, c_4, rows r_2, r_3 and symbols s_3, s_5 are not d-overloaded, this eliminates at most $6dn$ choices.

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

- None of the cells in these intercalates are prescribed. This eliminates at most $4(\alpha n + c(n)) + 4c(n)$ choices.
- The symbol $s_6 = L(r_6, c_1) \notin \{t_1, \ldots, t_a\}$, and it is not overloaded. This eliminates at most $a + (k/d)n$ choices.
- $s_6 \notin A'(r_3, c_1) \cup A'(r_2, c_4)$ and $s_6 \notin \{s_1, s_2, s_3, s_4\}$. This eliminates at most $2\beta n + 6$ choices.

Thus we have at least

$$\left\lfloor \frac{n}{2} \right\rfloor - 2\varepsilon n - 6dn - 4\alpha n - 8c(n) - a - \frac{k}{d}n - 6 - 2\beta n$$

choices for the required intercalates C_4 and C_5 in L, and by assumption this expression is greater than zero.

By swapping on the disjoint intercalates C_3, C_4 and C_5 we obtain a Latin square $L^{(1)}$. Note that the set

$$\{(r_2, c_1)_{L^{(1)}}, (r_2, c_4)_{L^{(1)}}, (r_3, c_1)_{L^{(1)}}, (r_3, c_4)_{L^{(1)}}\}$$

is an intercalate in $L^{(1)}$ and by swapping on this intercalate we obtain a Latin square $L^{(2)}$, in which the set

$$\{(r_1, c_1)_{L^{(2)}}, (r_1, c_2)_{L^{(2)}}, (r_2, c_1)_{L^{(2)}}, (r_2, c_2)_{L^{(2)}}\}$$

is an intercalate; by swapping on this intercalate we finally obtain the required Latin square L'. Moreover, it can be verified that L' contains no conflicts with A' that were not present in L. This completes the proof in Case 2. □

Of course the analogous statement for columns is true as well.

Lemma 3.6. *Let L_0, $\hat{P}$ and A' be as above. Suppose that L is an $n \times n$ Latin square obtained from L_0 by performing some sequence of trades on L_0, and that at most kn^2 cells of L are L-disturbed, for some $k > 0$.*

Let $\{t_1, \ldots, t_a\}$ be a set of a symbols from L.

If

$$\left\lfloor \frac{n}{2} \right\rfloor - 2\varepsilon n - 6dn - 5\frac{k}{d}n - 4\alpha n - 8c(n) - 3a - 3\beta n > 6,$$

then for any column c_1 of L and all but at most

- *$2(k/d)n + \alpha n + c(n) + a$ choices of r_1, and*
- *$a + 1 + 4c(n) + 2\beta n + 4(k/d)n + 2\alpha n + 2dn$ choices of r_2,*

there is a trade on a set of cells T such that if we let L' denote the Latin square obtained from L by performing this trade, then L' satisfies the following:

- *the trade T uses only symbols that are not d-overloaded;*
- *no prescribed cells of L are in T;*
- *L and L' differs on at most 16 cells (i.e. T uses at most 16 cells);*
- *no cell with entry $\{t_1, \ldots, t_a\}$ in L is in T;*

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

- *$L'(r_1, c_1) = L(r_2, c_1)$ and $L'(r_2, c_1) = L(r_1, c_1)$;*
- *if there is a conflict of L' with A', then the corresponding cell of L is also a conflict with A'.*

The two above lemmas are used for exchanging the content of two cells in a Latin square; in the case of Lemma 3.5, the cells are in positions (r_1, c_1) and (r_1, c_2), respectively. When using this lemma below, we shall refer to the cell in position (r_1, c_1) as the 'first cell' and the cell in position (r_1, c_2) as the 'second cell', and similarly for Lemma 3.6.

The two above lemmas can be used for proving the following, which essentially is a variant of Lemma 2.3 in [7].

Lemma 3.7. *Let L_0, $\hat{P}$ and A' be as above, and L be a Latin square obtained from L_0 by performing some sequence of trades on L_0. Assume that at most kn^2 cells of L are L-disturbed, where $k > 0$. Suppose that L has some prescribed cells where L and $\hat{P}$ do not agree. In particular, for each symbol s_i, assume that at most $2c(n) + 2d(n)$ cells with symbol s_i are prescribed in L, and assume further that at most $4(c(n) + d(n) + \alpha n + f(n))$ cells in L with symbol s_i are L-disturbed. Let $(r_1, c_1)_L$ be a cell of L such that*

$$L(r_1, c_1) = s_1 \quad \text{and} \quad \hat{P}(r_1, c_1) = s_2, \quad s_1 \neq s_2.$$

If

$$n - 2\left(4\frac{k + 64/n^2}{d}n + 3 + 6c(n) + 2\beta n + 4\frac{k}{d}n + 2\alpha n + 2f(n) + 4dn\right) > 1,$$

then there is a trade on a set of cells T in L, such that if we let L' denote the Latin square obtained from L by performing this trade on T, then the following holds:

- *$L'(r_1, c_1) = s_2$;*
- *L' and L disagree on at most 69 cells;*
- *besides $(r_1, c_1)_L$, L and L' disagree on at most two prescribed cells;*
- *if L and L' disagree on a prescribed cell $(r, c)_L$ (where $r \neq r_1$ or $c_1 \neq c$), then $L'(r, c)$ is not d-overloaded and $L(r, c) \neq \hat{P}(r, c)$;*
- *the trade T contains exactly two cells with entry s_1 in L, and at most four cells with entry s_2;*
- *except s_1 and s_2 the trade T contains only cells with symbols that are not d-overloaded;*
- *if there is a conflict of L' with A', then the corresponding cell of L is also a conflict with A'.*

Proof. We shall construct a trade from which we obtain L' from L, where L' and $\hat{P}$ agree on the cell in position (r_1, c_1). We will accomplish this by four successive applications of Lemmas 3.5 and 3.6, similarly to how Lemma 2.2 in [7] is applied in that paper. In our application of Lemmas 3.5 and 3.6 we will avoid the symbols $\{s_1, s_2\}$; so $a = 2$ in the application of these lemmas.

Let $(r_1, c_3)_L$ and $(r_3, c_1)_L$ be the cells in row r_1 and column c_1, respectively, that contains s_2. We want to choose a cell $(r_4, c_4)_L$ such that $L(r_4, c_4) = s_1$, and if r_2 and c_2 are the row and column, respectively, satisfying that $L(r_4, c_2) = s_2$ and $L(r_2, c_4) = s_2$, then the following holds.

- The cells $(r_4, c_4)_L, (r_4, c_2)_L, (r_2, c_4)_L$ are not prescribed cells. This eliminates at most $4c(n) + 4dn$ choices.
- The cell $(r_4, c_4)_L$ is not L-disturbed and $s_2 \notin A'(r_4, c_4)$. This eliminates at most $4(c(n) + d(n) + \alpha n + f(n)) + c(n)$ choices.
- $s_2 \notin A'(r_3, c_2) \cup A'(r_2, c_3)$ and $s_1 \notin A'(r_4, c_1) \cup A'(r_1, c_4)$. This excludes at most $4\beta n$ choices.

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

- The cells $(r_4, c_1)_L, (r_2, c_3)_L, (r_3, c_2)_L, (r_1, c_4)_L$ are all valid choices for the first cell to be changed in an application of Lemma 3.5 or 3.6. Since these lemmas are applied four consecutive times this excludes at most
$$4\left(2\frac{k+64/n^2}{d}n+\alpha n+c(n)+2\right)$$
choices. In particular, this implies that none of these cells are prescribed or contain a d-overloaded symbol.

Thus we have at least
$$n-12c(n)-8d(n)-4\alpha n-4f(n)-4\beta n-4\left(2\frac{k+64/n^2}{d}n+2\right)$$
choices for such a cell $(r_4, c_4)_L$ containing symbol s_1. We note that this expression is greater than zero by assumption, so we can indeed make the choice.

Next, we want to choose a symbol s_3 in row r_1 and column c_3, such that the following holds.

- The cells with symbol s_3 in row r_1 and column c_3 are both valid choices for the second cell to be exchanged in an application of Lemma 3.5 or 3.6; this eliminates at most
$$2\left(3+4c(n)+2\beta n+4\frac{k+64/n^2}{d}n+2\alpha n+2dn\right)$$
choices.
- $s_3 \notin A'(r_1, c_3) \cup A'(r_2, c_4)$. This eliminates at most $2\beta n$ choices.

Thus we have at least
$$n-2\left(3+4c(n)+2\beta n+4\frac{k+64/n^2}{d}n+2\alpha n+2dn\right)-2\beta n$$
choices for the symbol s_3. By assumption, this expression is greater than zero, so we can indeed choose such a symbol s_3.

Similarly, we want to choose a symbol s_4 in row r_3 and column c_1 such that the following holds.

- The cells with symbol s_4 in row r_3 and column c_1 are both valid choices for the second cell to be exchanged in an application of Lemma 3.5 or 3.6; this eliminates at most
$$2\left(3+4c(n)+2\beta n+4\frac{k+64/n^2}{d}n+2\alpha n+2dn\right)$$
choices.
- $s_4 \notin A'(r_4, c_2) \cup A'(r_3, c_1)$. This eliminates at most $2\beta n$ choices.

Hence, we have precisely the same number of choices for the symbol s_4 as for s_3.

Now, by applying Lemmas 3.5 and 3.6 to the cells $(r_1, c_4)_L$, and $(r_2, c_3)_L$, and the cells in column c_3 and row r_1 containing symbol s_3, we may exchange the content of cells $(r_1, c_4)_L$, and $(r_2, c_3)_L$; and similarly for the cells $(r_4, c_1)_L$, $(r_3, c_2)_L$, and symbol s_4.

Hence, by four successive applications of Lemmas 3.5 and 3.6 we obtain a Latin square $L^{(1)}$, such that the sets
$$\left\{(r_3, c_1)_{L^{(1)}}, (r_4, c_1)_{L^{(1)}}, (r_3, c_2)_{L^{(1)}}, (r_4, c_2)_{L^{(1)}}\right\}$$
and
$$\left\{(r_1, c_3)_{L^{(1)}}, (r_1, c_4)_{L^{(1)}}, (r_2, c_3)_{L^{(1)}}, (r_2, c_4)_{L^{(1)}}\right\}$$

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

are disjoint intercalates. By swapping on these intercalates we obtain a Latin square $L^{(2)}$, where the set

$$\{(r_1,c_1)_{L^{(2)}},(r_1,c_4)_{L^{(2)}},(r_4,c_1)_{L^{(2)}},(r_4,c_4)_{L^{(2)}}\}$$

is an intercalate. By swapping on this intercalate we obtain the required Latin square L'. □

We will take care of all the prescribed cells of L_0 by successively applying Lemma 3.7; using this lemma one can construct the Latin squares $L_0, L_1, \ldots, L_q$, where L_i is constructed from L_{i-1} by an application of Lemma 3.7, and L_q is an completion of $\hat{P}$, where $q \leqslant n(\alpha n + c(n))$. Thus, in L_i one more prescribed cell has the same entry as the corresponding cell in $\hat{P}$, compared to L_{i-1}.

Except for the cell $(r_1,c_1)_L$ in Lemma 3.7, an application of Lemma 3.7 will possibly change the content of two other prescribed cells. However, it follows that if this is the case, then in L' each such prescribed cell contains a symbol that is not d-overloaded. Moreover, for each symbol s, L_0 has at most $2c(n)$ prescribed cells containing s. Thus for each $i = 1, \ldots, q$, any symbol s in L_i occurs in at most $2c(n) + 2dn$ prescribed cells. Furthermore, each application of Lemma 3.7 to a prescribed cell $(r_1,c_1)_L$ with $L(r_1,c_1) = s$ constructs a trade T with exactly two cells containing symbol s. Hence, a symbol s is used at most $2(2c(n) + 2dn)$ times in a trade where a prescribed cell has entry s.

Note further that at most $\alpha n + f(n)$ cells $(r',c')_{\hat{P}}$ in $\hat{P}$ have entry s, and a trade T constructed by an application of Lemma 3.7 for obtaining a Latin square L' such that $L'(r',c') = s$ uses four cells with entry s.

Except for the cells mentioned in the preceding two paragraphs, any other cells involved in a trade created by an application of Lemma 3.7 contain symbols that are not d-overloaded. Hence, at most

$$4(c(n) + dn + \alpha n + f(n))$$

distinct cells with a given symbol s are used in trades for constructing L_q from L_0.

Thus as long as (3.1), $kn^2 \geqslant 69n(\alpha n + c(n))$, and all the other conditions in the proof of Theorem 2.3 hold, it follows that we can apply the last lemma iteratively for constructing the sequence $L_0, \ldots, L_q$ of Latin squares, where L_q is a completion of $\hat{P}$ that avoids A'. This completes the proof of Theorem 2.3. □

4. Random partial Latin squares and arrays

In this section we prove Corollary 1.2. So let P be a random PLS from the probability space $\mathcal{P}(n,p)$ defined in the Introduction, and let A be a random array where each cell $(i,j)_A$ of A is a set $A(i,j)$ of size $m = m(n)$ obtained by choosing each set uniformly at random from all m-subsets of $[n]$. Assume further that no entry of A occurs in the corresponding cell of P. We need to prove that there are constants ρ_1 and ρ_2 such that, if $p < \rho_1$ and $m \leqslant \rho_2 n$ and where, for any cell of A containing an entry that occurs in the corresponding cell of P, we remove that entry from A, then with probability tending to 1, there is a completion of P that avoids A. We will use simple first moment calculations as in [4].

Let X_{ij} be the indicator random variable for the event that symbol i occurs at least βn times in row j of A, and set

$$X = \sum_{1 \leqslant i,j \leqslant n} X_{ij}.$$

Similarly, let Y_{ij} be the indicator random variable for the event that symbol i occurs at least βn times in column j of A, and set

$$Y = \sum_{1 \leqslant i,j \leqslant n} Y_{ij}.$$

Then we have

$$\mathbb{P}[X>0] \leqslant \mathbb{E}[X] \leqslant n^2 \frac{\binom{n}{\lceil \beta n \rceil}\binom{n-1}{m-1}^{\lceil \beta n \rceil}\binom{n}{m}^{n^2-\lceil \beta n \rceil}}{\binom{n}{m}^{n^2}} \leqslant n^2 \frac{(n)_{\lceil \beta n \rceil}}{(\lceil \beta n \rceil)!}\rho_2^{\lceil \beta n \rceil} \tag{4.1}$$

where $(n)_k$ is the usual falling factorial. By applying Stirling's formula, we see that the right-hand side of (4.1) tends to 0 as $n \to \infty$, provided that $\rho_2 < \beta/e$, where e is the base of the natural logarithm. Proceeding similarly, if $\rho_2 < \beta/e$, then $\mathbb{P}[Y > 0] \to 0$ as $n \to \infty$. Thus it follows that if $\rho_2 < \beta/e$, then the probability that A is a $(\beta n, \beta n, \beta n)$-array tends to 1 as $n \to \infty$.

Using calculations as above, it is straightforward to verify that if $\rho_1 \leqslant \alpha/e$, then with probability tending to 1 as $n \to \infty$, P is α-dense.

Hence, by Theorem 1.1, the probability that there is a completion of P that avoids A tends to 1 as $n \to \infty$. This concludes the proof of Corollary 1.2.

Remark. Note that the proof of Corollary 1.2 is valid if we take P to be a random PLS and A to be a given (deterministic) $(\beta n, \beta n, \beta n)$-array which the completion of P should avoid; or, if we take P to be a given α-dense PLS and A a random array. Furthermore, the proof of Corollary 1.2 is valid if $\rho_1 < \alpha/e$ and $\rho_2 < \beta/e$. Thus if we can get better bounds on α and β for which Theorem 1.1 holds, then we also get a better bound on ρ_1 and ρ_2.

5. Concluding remarks

We have proved that there are constants α and β such that every α-dense PLS can be completed to a Latin square L that avoids a given $(\beta n, \beta n, \beta n)$-array, provided that the PLS avoids the array. Let us now briefly indicate what the best possible values of α and β might be.

In [18] it is conjectured that if $\alpha \leqslant 1/4$, then any α-dense PLS is completable, and in [22] it is conjectured that if $\beta \leqslant 1/3$, then any $(\beta n, \beta n, \beta n)$-array is avoidable. In [30], for any $\gamma > 0$, examples of $(1/4 + \gamma)$-dense partial Latin squares that are not completable are given; from the perspective of avoiding arrays, an example by Pebody shows for any $\gamma > 0$, there are unavoidable $(\beta n, \beta n, \beta n)$-arrays with $\beta \geqslant 1/3 + \gamma$ (see *e.g.* [17]).

We say that a point (α, β) is *feasible* if, for every pair (P, A), where P is an $n \times n$ α-dense PLS and A an $n \times n$ $(\beta n, \beta n, \beta n)$-array such that no entry of P occurs in the corresponding cell of A, it is possible to complete P into a Latin square that avoids A. A point that is not feasible is *infeasible*. So the above examples show that the points $(0, 1/4 + \gamma)$ and $(1/3 + \gamma, 0)$ are infeasible. Hence, the points outside the lines $(1/3, t)$ and $(t, 1/4)$ are infeasible.

Using a combination of the mentioned constructions we can generate arbitrarily large examples of α-dense partial Latin squares that cannot be completed to avoid a given (β, β, β)-array, provided that $\alpha + \beta = 1/3 + \gamma$, as follows.

For simplicity, assume that $n = 3r + 2$. Let A be an $(r + 1) \times (r + 1)$ array in which each cell contains the set $\{1, \ldots, r + 1\}$, let B be an $(r + 1) \times (r + 1)$ array in which each entry is $\{r + 2, \ldots, 2r + 2\}$, and let C be an $r \times r$ array in which each cell contains the set $\{2r + 2, \ldots, 3r + 2\}$. Define E_1 to be the $n \times n$ array containing A in the upper left $(r + 1) \times (r + 1)$ corner, B in the intersection of rows $r + 2, \ldots, 2r + 2$ and columns $r + 2, \ldots, 2r + 2$, and C in the lower right $r \times r$ corner:

$$E_1 = \begin{array}{|c|c|c|}\hline A & & \\ \hline & B & \\ \hline & & C \\ \hline \end{array}\,.$$

https://doi.org/10.1017/S096354831800055X Published online by Cambridge University Press

The array E_1 is an unavoidable $(\beta n, \beta n, \beta n)$-array for, asymptotically, $\beta = 1/3$; see *e.g.* [17].

(1) We define three sets S_1, S_2, S_3 by setting

$$S_1 = \{r+2\} \cup \{2r+3, \ldots, 3r+2\}, S_2 = \{1, \ldots, r+1\}, S_3 = \{r+3, \ldots, 2r+2\}.$$

(2) Following [30], for each set S_i we construct an $|S_i| \times |S_i|$ single entry array L_i with symbols from S_i such that each symbol occurs precisely once in each row and column, and with the property that the cells of L_i are the union of $|S_i|$ disjoint S_i-transversals $T_{i,j}$, $1 \leqslant j \leqslant |S_i|$, where an S_i-*transversal* is a generalized diagonal in L_i where each symbol in S_i occurs exactly once. For convenience, define $T_{3,r+1} = \emptyset$.

We now define an $n \times n$ PLS E_2 with L_1 in the position held by A in E_1, L_2 in the position held by B in E_1, and L_3 in the position held by C in E_1.

(3) Next, for each integer t satisfying $1 \leqslant t \leqslant r+1$, define an $n \times n$ array E_{1t} from E_1 by setting $E_{1t}(p, c) = \emptyset$ for each position (p, c) of E_1 which corresponds to a non-empty cell $(p, c)_{E_2}$ of E_2 such that $(p, c)_{E_2} \in \bigcup_i \bigcup_{j=1}^{t} T_{i,j}$. We retain the content of any other cell of E_1.

(4) We now define a PLS E_t^1 from E_2 by retaining the entry of each cell in $\bigcup_i \bigcup_{j=1}^{t} T_{i,j}$, and removing the entry of each cell in E_2 which does not belong to this set.

(5) It follows that E_t^1 is a t/n-dense PLS, and E_{1t} is a $(\beta n - t, \beta n - t, \beta n - t)$-array.

Now, the PLS E_t^1 cannot be completed to a Latin square which avoids E_{1t}. This follows from the fact that each cell in E_t^1 contains a symbol which does not occur in the corresponding cell of E_1, and outside the support of E_t^1 (*i.e.* the non-empty cells of E_t^1), the array E_{1t} agrees with E_1, so any Latin square which is a completion of E_t^1 that avoids E_{1t} would also avoid E_1.

Consider a line ℓ in the $\alpha\beta$-plane from $(1/3, 0)$ to $(0, 1/3)$. The pairs (E_{1t}, E_t^1) imply that each point outside the region bounded by ℓ and the α- and β-axes is infeasible. In fact, combined with the examples by Wanless, we know that the set of feasible points is a subset of the region bounded by ℓ, the line $(1/4, t)$ and the α- and β-axes.

It would be interesting to obtain more information on the structure of set of feasible points, but we expect that methods other than those used in this paper will be needed for this. Specifically, we would like to pose the following.

Problem 5.1. Is the set of feasible points (α, β) a convex set?

Both of the conjectured boundary points $(0, 1/4)$ and $(1/3, 0)$ are also boundary points for certain linear programming relaxations of the completion and avoidance problems [23]. So, it might be possible to use a relaxation of the combined problem to provide a convex domain which gives a tighter bound for the set of feasible points than that given by our construction.

Further, given that the constructions which give our bounds for the set of feasible points are highly structured and that our proof for Corollary 1.2 relies on our main result Theorem 1.1, it is not unreasonable to expect that the best possible parameters in Corollary 1.2 are larger than those which even an optimal version of Theorem 1.1 would give. Here it would be interesting both to see if Corollary 1.2 can be improved and if some upper bounds on the possible values of ρ_1 and ρ_2 can be proved.

Conflict of interest. None.

References

[1] Adams, P., Bryant, D. and Buchanan, M. (2008) Completing partial Latin squares with two filled rows and two filled columns. *Electron. J. Combin.* **15** #R56.

[2] Andersen, L. D. and Hilton, A. J. W. (1983) Thank Evans! *Proc. London Math. Soc.* **47** 507–522.

[3] Andrén, L. J. (2010) On Latin squares and avoidable arrays. Doctoral thesis, Umeå University.
[4] Andrén, L. J., Casselgren, C. J. and Öhman, L.-D. (2013) Avoiding arrays of odd order by Latin squares. *Combin. Probab. Comput.* **22** 184–212.
[5] Asratian, A. S., Denley, T. M. J. and Häggkvist, R. (1998) *Bipartite Graphs and Their Applications*, Cambridge University Press.
[6] Barber, B., Kühn, D., Lo, A., Osthus, D. and Taylor, A. (2017) Clique decompositions of multipartite graphs and completion of Latin squares. *J. Combin. Theory Ser. A* **151** 146–201.
[7] Bartlett, P. (2013) Completions of ε-dense partial Latin squares. *J. Combin. Designs* **21** 447–463.
[8] Brègman, L. M. (1973) Certain properties of nonnegative matrices and their permanents. *Dokl. Akad. Nauk SSSR* **211** 27–30.
[9] Casselgren, C. J. (2012) On avoiding some families of arrays. *Discrete Math.* **312** 963–972.
[10] Casselgren, C. J. and Häggkvist, R. (2013) Completing partial Latin squares with one filled row, column and symbol. *Discrete Math.* **313** 1011–1017.
[11] Cavenagh, N. (2010) Avoidable partial Latin squares of order $4m + 1$. *Ars Combinatoria* **95** 257–275.
[12] Chetwynd, A. G. and Häggkvist, R. (1984) Completing partial $n \times n$ Latin squares where each row, column and symbol is used at most cn times. Research report, Department of Mathematics, Stockholm University.
[13] Chetwynd, A. G. and Rhodes, S. J. (1995) Chessboard squares. *Discrete Math.* **141** 47–59.
[14] Chetwynd, A. G. and Rhodes, S. J. (1997) Avoiding partial Latin squares and intricacy. *Discrete Math.* **177** 17–32.
[15] Chetwynd, A. G. and Rhodes, S. J. (1997) Avoiding multiple entry arrays. *J. Graph Theory* **25** 257–266.
[16] Colbourn, C. J. (1984) The complexity of completing partial Latin squares. *Discrete Appl. Math.* **8** 25–30.
[17] Cutler, J. and Öhman, L.-D. (2006) Latin squares with forbidden entries. *Electron. J. Combin.* **13** #R47.
[18] Daykin, D. E. and Häggkvist, R. (1984) Completion of sparse partial Latin squares. In *Graph Theory and Combinatorics: Proceedings of the Cambridge Combinatorial Conference in Honour of Paul Erdős*, Academic Press, pp. 127–132.
[19] Denley, T. and Kuhl, J. (2012) Constrained completion of partial Latin squares. *Discrete Math.* **312** 1251–1256.
[20] Evans, T. (1960) Embedding incomplete Latin squares. *Amer. Math. Monthly* **67** 958–961.
[21] Gustavsson, T. (1991) Decompositions of large graphs and digraphs with high minimum degree. Doctoral thesis, Stockholm University.
[22] Häggkvist, R. (1989) A note on Latin squares with restricted support. *Discrete Math.* **75** 253–254.
[23] Häggkvist, R. Personal communication.
[24] Kuhl, J. S. and Schroeder, M. (2016) Completing partial Latin squares with one nonempty row, column, and symbol. *Electron. J. Combin.* **23** #P2.23.
[25] Markström, K. and Öhman, L.-D. (2009) Unavoidable arrays. *Contrib. Discrete Math.* **5** 90–106.
[26] Öhman, L.-D. (2011) Partial Latin squares are avoidable. *Ann. Combin.* **15** 485–497.
[27] Öhman, L.-D. (2011) Latin squares with prescriptions and restrictions. *Austral. J. Combin.* **51** 77–87.
[28] Ryser, H. J. (1951) A combinatorial theorem with an application to Latin rectangles. *Proc. Amer. Math. Soc.* **2** 550–552.
[29] Smetaniuk, B. (1981) A new construction for Latin squares, I: Proof of the Evans conjecture. *Ars Combinatoria* **11** 155–172.
[30] Wanless, I. (2002) A generalization of transversals for Latin squares. *Electron. J. Combin.* **2** #R12.

Cite this article: Andrén LJ, Casselgren CJ and Markström K (2019). Restricted completion of sparse partial latin squares. *Combinatorics, Probability and Computing* **28**, 675–695. https://doi.org/10.1017/S096354831800055X