Combinatorics, Probability and Computing (2008) 17, 519–536. © 2008 Cambridge University Press
doi:10.1017/S0963548307008590 Printed in the United Kingdom

Orthogonal Latin Rectangles

ROLAND HÄGGKVIST[1] and ANDERS JOHANSSON[2]

[1]Matematiska Institutionen, Umeå Universitet, S-901 87 Umeå, Sweden
(e-mail: rolandh@math.umu.se)

[2]N-Institutionen, Högskolan i Gävle, S-801 76 Gävle, Sweden
(e-mail: ajj@hig.se)

Received 14 December 2005; revised 10 April 2007; first published online 3 June 2008

We use a greedy probabilistic method to prove that, for every $\epsilon > 0$, every $m \times n$ Latin rectangle on n symbols has an orthogonal mate, where $m = (1 - \epsilon)n$. That is, we show the existence of a second Latin rectangle such that no pair of the mn cells receives the same pair of symbols in the two rectangles.

1. Introduction

This paper was inspired by a problem posed by Anthony J. W. Hilton at the Thirteenth British Combinatorial Conference 1991 [6]. The problem is as follows:

> Let R be an $n \times 2n$ Latin rectangle on $2n$ symbols. A partial transversal T of size s of R is a collection of s cells, no two in the same row or column, and no two containing the same symbol. Is it true that R can be expressed as the union of $2n$ partial transversals of size n?
>
> An equivalent formulation: Call two $n \times 2n$ Latin rectangles R, S on the same set of symbols *orthogonal* if the pairs (r_{ij}, s_{ij}), for $i = 1, \ldots, n$ and $j = 1, \ldots, 2n$, are all distinct. Does every $n \times 2n$ Latin rectangle have an orthogonal mate?

The updated problem list from the British Combinatorial Conferences from number twelve and upwards can be found electronically as a link on the homepage of the British Combinatorial Conference.

While it is quite easy to see that every $n \times 4n$ Latin rectangle has an orthogonal mate, we know of no argument that solves the problem for $n \times 3n$, when $n = 100$, say. No doubt such an argument can be found eventually.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

In the current paper we use probabilistic methods to prove a statement far stronger than Hilton proposed, but only valid for large n, namely that for every $\varepsilon > 0$, every $(n - \varepsilon n) \times n$ Latin rectangle has an orthogonal mate for large enough n.

We know of no example of an $(n-1) \times n$ Latin rectangle without an orthogonal mate, but would not be too surprised if such an example could be constructed. This ties up with well-known conjectures and results concerning the length of partial transversals in Latin squares. Recall that Ryser [9], Brualdi [4, s. 103] and Stein [12] have conjectures (in particular, Stein has much stronger conjectures, one of which was refuted by Drisko [5]) which imply that every $(n-1) \times n$ Latin rectangle has a transversal of length $n-1$. In this context we also recall some standard results on the length of partial transversals in Latin squares, namely that every $n \times n$ Latin square has a partial transversal of length at least $n - \sqrt{n}$ (proved by Woolbright [13] and Brouwer, de Vries and Wieringa [3]), and $n - 5.53(\log n)^2$ (proved by Shor [11]).

1.1. The result once again

We consider $m \times n$ *Latin rectangles* on n symbols, n columns and m rows, *i.e.*, an assignment to any cell in an $m \times n$-table one of n symbols such that each symbol occurs exactly once in each row and at most once in each column.

Two Latin $m \times n$-rectangles, L and J, are *orthogonal* if the following holds. For any two colours α, β the colour classes $\mathsf{L}^{-1}(\alpha)$ and $\mathsf{J}^{-1}(\beta)$ intersect in at most one element. Equivalently, each colour class $\mathsf{L}^{-1}(\alpha)$ is a *transversal* of J and *vice versa*.

Theorem 1.1. *For every $\varepsilon > 0$ there is an $n_0 = n_0(\varepsilon)$ such that, to any $m \times n$ Latin rectangle J, $n \geqslant n_0$ and $m = n(1-\varepsilon)$, there is an orthogonal companion L.*

Remark. With some extra effort it is perhaps possible to prove Theorem 1.1 for all $\varepsilon = \omega(n^{-1/2})$. However, it will be clear from the proof that, in order to reach $\varepsilon \leqslant n^{-1/2}$, some new ideas must be found, if indeed the theorem is valid in this range.

The basic method is related to nibble methods used to colour graphs having 'near-disjoint' cliques. An orthogonal companion L of J can be thought of as an n-colouring of the graph, where the mn cells are vertices and where each row, column and J-colour class make up a clique, *i.e.*, a complete induced subgraph. The recent monograph of Reed and Molloy [10] contains many results in this area, *e.g.*, J. Kahn's result [8] on edge colourings of near-disjoint hypergraphs.

There are elements of this proof that we believe are new. First of all, we have a distinguished parallel class of cliques of size n – corresponding to the rows – that we colour with n colours. In other words we have no slack here.

In the proof, we construct an orthogonal companion in a random greedy manner by adding one row at a time. We use a process, q^t, $t \in [0, m]$, of 'fractional Latin rows' to guide the greedy extensions, so that $\mathsf{q}^t(t+1, \cdot, \cdot) \in \mathbb{R}^{\mathscr{K} \times \mathscr{S}}$ gives the expectation of the row, $t+1$, added at time t. We maintain the 'legality' of q^t by setting $\mathsf{q}^t(i,k,\gamma) = 0$ if the cell (i,k) belongs to a J-colour class or column already coloured with symbol γ.

By analysing the time evolution for certain statistics of the random process q^t, we deduce that, with positive probability, q^t can be legally maintained for all times $t = 0, 1, \ldots, m-1$, so that the Latin rectangle L is constructed at time m.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

1.2. Rows, columns, cells, diagonals and points

We will think of a vector f in Cartesian space $\mathbb{R}^A$ as a real-valued mapping f from the index set A. Pointwise relations extend to relations between vectors in the natural way, *e.g.*, $f \leqslant g$ means that $f(a) \leqslant g(a)$ for all $a \in A$.

Let $\mathscr{R} = [1, m] := \{1, 2, \ldots, m\}$ denote the set of *rows*, let $\mathscr{K}$ denote the set of *columns* and let $\mathscr{S}$ be the set of *symbols*. Thus, $|\mathscr{R}| = m = n(1 - \varepsilon)$ and $|\mathscr{K}| = |\mathscr{S}| = n$. We refer to elements of $\mathscr{R} \times \mathscr{K}$ as *cells*. Let J be the given Latin rectangle from Theorem 1.1. A *diagonal* is a set of cells assigned a common colour by J and the family of diagonals is denoted by $\mathscr{D}$. We refer to the elements of $\mathscr{X} := \mathscr{R} \times \mathscr{K} \times \mathscr{S}$ as *points*. If nothing else is stated, we assume that the variable i refers to a row, the variables k, l refer to columns and a variable γ refers to a symbol. The variables x and y will be preferred for points.

We have, for $\mathscr{P} = \mathscr{R}, \mathscr{K}, \mathscr{D}, \mathscr{S}$, mappings $\mathscr{X} \xrightarrow{\iota_{\mathscr{P}}} \mathscr{P}$ assigning to each point the unique row, column, diagonal or symbol to which it belongs. We usually say that a point $x \in \mathscr{X}$ *belongs to* the corresponding row, column, diagonal or symbol $\alpha \in \mathscr{P}$, when $\iota_{\mathscr{P}}(x) = \alpha$.

A *line* is a set of points with two of these coordinates fixed, *i.e.*, a line is the set of the form $(\iota_{\mathscr{P}} \times \iota_{\mathscr{Q}})^{-1}(\alpha, \beta)$ with $\alpha \in \mathscr{P}$ and $\beta \in \mathscr{Q}$. We introduce, for any pair of two distinct coordinates $\mathscr{P}$ and $\mathscr{Q}$, the mapping $\ell_{\mathscr{P}\mathscr{Q}}$, assigning to each point the corresponding line to which it belongs. More precisely, we let

$$\ell_{\mathscr{P}\mathscr{Q}}(x) := (\iota_{\mathscr{P}} \times \iota_{\mathscr{Q}})^{-1}(\iota_{\mathscr{P}} \times \iota_{\mathscr{Q}}(x)). \tag{1.1}$$

The collection of lines makes up a linear hypergraph on the points $\mathscr{X}$, *i.e.*, for every pair of distinct points $x, y \in \mathscr{X}$ there is at most one line containing them.

1.2.1. Latin rectangles. A *rectangle* can now be identified with a $\{0,1\}$-valued vector $\mathsf{L} \in \{0,1\}^{\mathscr{X}}$ in the obvious way: for a point $x = (i, k, \gamma)$, $\mathsf{L}(x) = 1$ if γ is the symbol assigned to cell (i, k) and $\mathsf{L}(x) = 0$ otherwise.

A rectangle $\mathsf{L} \in \{0,1\}^{\mathscr{X}}$ is a *Latin rectangle orthogonal to* J exactly when the relations

$$\sum_{y \in \ell_{\mathscr{R}\mathscr{K}}(x)} \mathsf{L}(y) = 1, \qquad \sum_{y \in \ell_{\mathscr{R}\mathscr{S}}(x)} \mathsf{L}(y) = 1, \tag{L}$$

$$\sum_{y \in \ell_{\mathscr{K}\mathscr{S}}(x)} \mathsf{L}(y) \leqslant 1, \qquad \sum_{y \in \ell_{\mathscr{D}\mathscr{S}}(x)} \mathsf{L}(y) \leqslant 1, \tag{C}$$

hold for all $x \in \mathscr{X}$. The relations in (L) and (C) define a polytope $\mathfrak{L} \subset [0,1]^{\mathscr{X}}$ so that a Latin rectangle L is a $\{0,1\}$-valued element of this polytope. For our purposes, *rational Latin rectangles* orthogonal to J are vectors in $\mathfrak{L}$. The constraints in (L) are *local* to each row since they concern lines contained in rows. The constraints in (C) are then *central* constraints since they concern lines transversal to the rows.

1.3. The greedy Latin rectangle process

We now give a bird's eye view of the proof. The probabilistic terminology used regarding vector-valued random processes is made precise in Section 1.4 below. Our purpose is to construct an increasing random process, a *greedy rectangle process*, $\mathsf{L}^t \in \{0,1\}^{\mathscr{X}}$ of partial J-orthogonal Latin rectangles that proceed row-wise. Initially, $\mathsf{L}^0 \equiv 0$ and at each tick of the clock, *i.e.*, when $t \mapsto$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

$t+1$, we extend – if the situation allows it – the partial Latin rectangle L^t to a partial Latin rectangle L^{t+1} having the row $t+1$ added to the Latin rectangle. The time variable $t \in [0,m] := \{0,\dots,m\}$ thus corresponds to rows being added to the rectangle. The process is *successful* if L^m actually produces a full J-orthogonal Latin rectangle.

It is quite easy to see that such a greedy rectangle process should always be successful as long as $m \leqslant n/4$. To see this, note that the legal choices of the added row $t+1$ are, by the local constraints, given by matchings in a 'legality graph', which is the balanced bipartite graph consisting of those symbol–column pairs for row $t+1$ that are not in conflict with any previously added row due to central constraints. Moreover, each previously added row can exclude at most two symbols for the column k on row $t+1$: one symbol on the same column and one symbol on the same diagonal. Similarly, each added row excludes at most two possible columns for any symbol γ. Thus, since $t < m \leqslant n/4$ are previously added, the legality graph will have minimum degree at least $n-2t \geqslant n/2$ and a well-known degree condition based on Hall's theorem ensures the existence of a legal matching for row number $t+1$.

This naive argument can be extended significantly when the obtained legality graph is sufficiently random-like to ensure the existence of a perfect matching for degrees well below $n/2$. To achieve this, we need to introduce some probabilistic tools.

A central idea is to let the greedy rectangle process L^t be 'guided' by a Markov process $\mathsf{p}^t \in [0,1]^{\mathscr{X}}$, $t \in [0,m]$. We refer to $\mathsf{p}^t \in [0,1]^{\mathscr{X}}$ as a *state*. The initial state is the uniform vector $\mathsf{p}^0 \equiv \frac{1}{n}$. The relationship between the processes L^t and p^t is that, at time t, $\mathsf{p}^t(x)$ approximately gives the expectation of $\mathsf{L}^s(x)$, for points x belonging to rows that are coloured at time $s > t$.

Care must therefore be taken in the construction of p^t, so that L^t never violates the local and central constraints, (L) and (C). We defer the exact definition p^t to Section 1.5 below. We note here that the construction of p^t ensures that the central constraints (C) are never violated by L^t. If a cell $(t+1,k)$ in the active row is assigned the colour γ at time t then we remove the possibility that any cell in the same column or diagonal later gets colour γ. Hence, we must 'kill' all points y belonging to the central lines going through the point $(t+1,k,\gamma)$, that is, we set

$$\mathsf{p}^{t+1}(y) = \mathsf{p}^{t+2}(y) = \cdots = 0, \tag{1.2}$$

for all y belonging to such a central line.

Given $t \in [0,m]$, we define a region $\Gamma \subset [0,1]^{\mathscr{X}}$, where $\mathsf{p}^t \in \Gamma$ should be interpreted as stating that p^t is a 'good state'. The exact definition of Γ is deferred to Section 1.6 below, but we mention that Γ is defined by three collections of inequalities: the first group of inequalities, (A_x), bounds the size of the individual values $\mathsf{p}^t(x)$ while the second group, $(B_{\ell(x)})$, states that p^t should almost satisfy the local constraints (L).

Note that, for a fixed row $i \in \mathscr{R}$, the local constraints given by (L) define a polytope $\mathfrak{L}_i$ in $\mathbb{R}_+^{\iota_{\mathscr{R}}^{-1}(i)} \cong \mathbb{R}_+^{\mathscr{K}\times\mathscr{S}}$ which can be interpreted as the polytope of *rational* (*perfect*) *matchings* in the complete bipartite graph $K(\mathscr{K},\mathscr{S})$ and we refer to $\{0,1\}$-valued vectors in $\mathfrak{L}_i$ as *matchings* on that row.

Lemma 1.2. *If $\mathsf{p}^t \in \Gamma$ then, for each row $i \in \mathscr{R}$, there is a rational matching $\mathsf{q}_i^t \in \mathfrak{L}_i$ such that, for all $(k,\gamma) \in \mathscr{K}\times\mathscr{S}$,*

$$\mathsf{q}_i^t(i,k,\gamma) \leqslant (1+\mathfrak{b})\mathsf{p}^t(i,k,\gamma), \tag{1.3}$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

where $\mathfrak{b}$ *is an abbreviation of the asymptotic expression* $O\big(\sqrt{n^{-1/2}\log n}\big)$. (*See* (1.6) *below.*)

We prove this lemma in Section 2 using the Ford–Fulkerson theorem; in the proof, the third and last group of inequalities, $(C_{i,kl})$, which give a 'quasi-random' property of p^t, are central for the construction.

Now recall the well-known characterization by Birkhoff [2] stating that any rational matching $\mathsf{q}_i \in \mathfrak{L}_i$ can be expressed as a convex combination $q_i = \sum_M c_M M$ of matchings $M \in \mathfrak{L}_i$. By interpreting the convex coefficients c_M as probabilities, where we pick the matching M with probability c_M, Birkhoff's theorem can also be given the following formulation. Given any rational matching $\mathsf{q}_i \in \mathfrak{L}_i$, it is always possible to find a *random* matching $\mathsf{L}_i \in \mathfrak{L}_i$, such that the *expectation* of L_i equals q_i.

Therefore, modulo the precise definition of Γ and p^t, we can define the greedy Latin rectangle process L^t by iterating the following procedure for $t = 0, 1, \ldots, m-1$.

Extend. If $\mathsf{p}^t \in \Gamma$ then choose a rational matching q^t_{t+1} on row $t+1$ which satisfies (1.3). Then draw, using the random construction implied by Birkhoff's theorem, an extension L^{t+1} such that

$$\mathbb{E}_t\big[\mathsf{L}^{t+1}_{t+1}\big] = \mathsf{q}^t_{t+1}. \tag{1.4}$$

The new state p^{t+1} is then constructed from p^t, q_{t+1} and L^{t+1} according to the construction in (1.8) below.

Stop. If $\mathsf{p}^t \notin \Gamma$ then simply let $\mathsf{L}^s = \mathsf{L}^t$ and $\mathsf{p}^s = \mathsf{p}^t$ for all s, $t < s \leqslant m$. The greedy Latin rectangle is then said to be unsuccessful.

On account of the killing mechanism (1.2), the bound (1.3) and the property (1.4), the construction ensures that the central constraints are never violated by L^t. Thus, if p^t stays in Γ, the process produces an orthogonal companion L^m to J at time m and the probability of an unsuccessful rectangle process is the probability that p^t leaves Γ for some $t \in [0, m]$. The proof is thus concluded, if, after properly describing the construction of p^t and the definition of Γ and proving Lemma 1.2, we prove, in addition, the following lemma.

Lemma 1.3. *For all* $t = 1, 2, \ldots, m-1$ *we have that*

$$\mathbb{P}\{\mathsf{p}^t \in \Gamma\} = 1 - n^{-\omega(1)}. \tag{1.5}$$

Note that (1.5) implies that the probability that p^t stays inside Γ at all times t is of probability of order $1 - n^{-\omega(1)+1} = 1 - n^{-\omega(1)}$.

1.4. Probabilistic preliminaries and asymptotic notation

1.4.1. Asymptotic notation. We will use the standard asymptotic notation, $O(\cdot)$, $o(\cdot)$, $\omega(\cdot)$, $\Omega(\cdot)$, *etc.*, where all are interpreted as asymptotic estimates relative the limit $n \to \infty$. That is,

- $f = O(g)$ if and only if $\limsup_{n\to\infty} |f/g| < \infty$,
- $f = \omega(g)$ if and only if $\liminf_{n\to\infty} |f/g| = \infty$,
- $f = \Omega(g)$ if and only if $\limsup_{n\to\infty} |g/f| < \infty$,
- $f = o(g)$ if and only if $\limsup_{n\to\infty} |f/g| = 0$, and
- $f = \Theta(g)$ if and only if $\limsup_{n\to\infty}(|f/g| + |g/f|) < \infty$.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

Such asymptotic expressions are often used to estimate the components of vectors and, of course, if we have such a local quantity expressed in terms of some asymptotic expression, then all implicit constants in the asymptotic expression are assumed to be independent of the particular point, row, column, and so on, at which the local quantity is defined.

Since some asymptotic expressions are extensively used, we also introduce the following *abbreviations* of asymptotic expressions:

$$\mathfrak{a} := O\left(\frac{\log n}{\sqrt{n}}\right), \quad \mathfrak{b} := O\left(n^{-1/4}(\log n)^{1/2}\right), \quad \mathfrak{p} := O\left(\frac{1}{n}\right). \tag{1.6}$$

Note that $\sqrt{\mathfrak{a}} = \mathfrak{b}$.

1.4.2. Probabilistic terminology and notation. The proof will use a dynamic probabilistic method, so we introduce some concepts and terms from probability theory. We will construct a *filtered probability space* $(\Omega, \mathbb{P}\{\cdot\}, \mathscr{F}^t)$ with a discrete time variable t taking values in $[0,m] := \{0,1,\dots,m\}$ and we can assume that the probability space we work with is finite. The finite algebras $\{\mathscr{F}^t\}$, $t \in [0,m]$, form an increasing sequence of subsets of 2^Ω with $\mathscr{F}^0 = \{\emptyset, \Omega\}$. In our case, $\mathscr{F}^t$ captures the random operations used for adding the first t rows to a Latin rectangle. A random variable is *determined at time t* if it is $\mathscr{F}^t$-measurable.

We will work with vector-valued random variables without explicitly noting this: a (vector-valued) random variable is a mapping $X : \Omega \to \mathbb{R}^A$ from Ω to a Cartesian space $\mathbb{R}^A$. For our purposes, a *stochastic process* is a function $X : \Omega \times [0,m] \to \mathbb{R}^A$. We write $X(\omega, t)$ simply as X^t. Functions defined for $t \in [0,m]$ will generally have the variable t as a superscript. We suppress the dependence on $\omega \in \Omega$ for the random entities.

A process X^t is *adapted* if the value of X^t is determined at time t, and we usually assume this to be the case without further notice. A process X^t is *previsible* if the value of X^t is determined at time $t-1$.

The expectation operator $\mathbb{E}[\cdot]$ and the probability $\mathbb{P}\{\cdot\}$ refers to the unconditional probability, while the temporal conditional expectation is denoted by $\mathbb{E}_t[f] = \mathbb{E}[f \mid \mathscr{F}^t]$ and the conditional probability by $\mathbb{P}_t\{A\} = \mathbb{P}\{A \mid \mathscr{F}^t\}$.

The various expectation operators apply to *vectors*, so that, if F is a random vector taking values in a Cartesian space $\mathbb{R}^A$, we mean by $\mathbb{E}[F]$ the vector in $\mathbb{R}^A$ given by $\mathbb{E}[F](a) = \mathbb{E}[F(a)]$, $a \in A$.

An adapted process X^t is a *martingale* (*super-martingale or sub-martingale*) if $\mathbb{E}_t[X^{t+1}] = X^t$ ($\mathbb{E}_t[X^{t+1}] \leqslant X^t$ or $\mathbb{E}_t[X^{t+1}] \geqslant X^t$).

A *stopping time* is a random time $\tau : \Omega \to [0,m] \cup \{\infty\}$ such that the event $\{\tau \leqslant t\} \in \mathscr{F}^t$. We will work with *vectors of random times*, *i.e.*, mappings $\tau : \Omega \times A \to [0,m] \cup \{\infty\}$.

Given a vector-valued process $X^t \in \mathbb{R}^A$, $t \in [0,m]$, then $X^{t\wedge\tau}$ is the process whose value at $a \in A$ at time t is $X^t(a)$ if $t \leqslant \tau(a)$ and $X^{\tau(a)}(a)$ otherwise. (We use $s \wedge t$ as a shorthand for $\min\{s,t\}$.)

If X^t is adapted and τ is a vector of stopping times, then $X^{t\wedge\tau}$ is adapted. We say that an adapted process X^t is *stopped* at a vector of stopping times τ if $X^t = X^{t\wedge\tau}$. If X^t is a super-martingale and τ is a vector of stopping times, then the process $X^{t\wedge\tau}$ is a super-martingale.

Before proving Lemma 1.2 and Lemma 1.3 in the following sections, we first proceed to define the process p^t rigorously as well as the good set Γ.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

1.5. The construction of p^t

For ease of notation, we define a deterministic vector $\tau_c : \mathscr{X} \to [0, m]$ of *colouring times*, so that, for a point $x = (i, k, \gamma)$, $\tau_c(x) = i - 1$ is the time when the row i is added to the Latin rectangle process.

1.5.1. The killing. Consider a point $x = (i, k, \gamma) \in \mathscr{X}$ such that $\tau_c(x) \neq t$. For a central line $\ell = \ell_{\mathscr{K}\mathscr{S}}$ or $\ell = \ell_{\mathscr{D}\mathscr{S}}$, let $\ell^t(x)$ be the unique point on the active row $t+1$ lying in the line $\ell(x)$, $x \in \mathscr{X}$. If any of $\mathsf{L}^{t+1} \circ \ell^t_{\mathscr{K}\mathscr{S}}(x)$ and $\mathsf{L}^{t+1} \circ \ell^t_{\mathscr{D}\mathscr{S}}(x)$ take the value 1, then the point x is *killed* at time t.

The two 'projections' of x, $\ell^t_{\mathscr{K}\mathscr{S}}(x)$ and $\ell^t_{\mathscr{D}\mathscr{S}}(x)$, are distinct points and belong to a common local line $\ell_{\mathscr{R}\mathscr{S}}(\ell^t_{\mathscr{K}\mathscr{S}}(x)) = \ell_{\mathscr{R}\mathscr{S}}(\ell^t_{\mathscr{D}\mathscr{S}}(x))$ in the row $t+1$. The indicators $\mathsf{L}^{t+1} \circ \ell^t_{\mathscr{K}\mathscr{S}}(x)$ and $\mathsf{L}^{t+1} \circ \ell^t_{\mathscr{D}\mathscr{S}}(x)$ can therefore never take the value 1 simultaneously, and we can write

$$\mathsf{K}^{t+1}(x) = \mathsf{L}^{t+1} \circ \ell^t_{\mathscr{K}\mathscr{S}}(x) + \mathsf{L}^{t+1} \circ \ell^t_{\mathscr{D}\mathscr{S}}(x), \tag{1.7}$$

for the indicator $\mathsf{K}^{t+1}(x) \in \{0, 1\}$ of the event that the point x is killed.

1.5.2. The construction of p^t. Initially, we set $\mathsf{p}^0(x) \equiv \frac{1}{n}$ for all $x \in \mathscr{X}$. We define the global stopping time T marking exit from Γ, *i.e.*,

$$T := \min\{t \in [0, m-1] \,:\, \mathsf{p}^t \notin \Gamma\} \cup \{\infty\}.$$

For $t > T$ the greedy process is thus in effect 'stopped' and the greedy random colouring has failed if $T < \infty$.

Therefore, for $t < T$ and $x \in \mathscr{X}$ such that $\tau_c(x) > t$, define

$$\mathsf{p}^{t+1}(x) := \mathsf{p}^t(x) \cdot \frac{1 - \mathsf{K}^{t+1}(x)}{\mathbb{E}_t[1 - \mathsf{K}^{t+1}(x)]}, \tag{1.8}$$

and for $t \leqslant T$ and $t \geqslant \tau_c(x)$, let $\mathsf{p}^{t+1}(x) = \mathsf{p}^t(x)$.

By the definition of p^{t+1} above, the process p^t is *stopped* both at the global stopping time T and at the deterministic stopping time vector $\tau_c : \mathscr{X} \to [0, m]$, *i.e.*, $\mathsf{p}^t = \mathsf{p}^{t \wedge \tau_c \wedge T}$.

1.5.3. Some properties of p^t. For $t < T$, *i.e.*, as long as $\mathsf{p}^t \in \Gamma$, we can, by Definition 1 below, assume that

$$\mathsf{p}^t \leqslant \mathfrak{p} \tag{1.9}$$

with the asymptotic abbreviation $\mathfrak{p} = O(1/n)$ as in (1.6). (We understand that a relation like (1.9) holds with the same implicit constant at all points $x \in \mathscr{X}$.)

Since the vector $1 - \mathsf{K}^{t+1} \in \{0, 1\}^{\mathscr{X}}$ indicates survival, the definition ensures that all points $x \in \mathscr{X}$ such that $\tau_c(x) > t+1$ and $\mathsf{p}^{t+1}(x) > 0$ can be used to extend L^{t+1}. It follows that L^t is indeed a process of legal (partial) J-orthogonal Latin rectangles.

Note that, from (1.3), (1.7) and (1.9), we have that

$$\begin{aligned} \mathbb{E}_t[1 - \mathsf{K}^{t+1}] &= 1 - \mathsf{q} \circ \ell^t_{\mathscr{K}\mathscr{S}} - \mathsf{q} \circ \ell^t_{\mathscr{D}\mathscr{S}} \\ &= 1 - \mathsf{p} \circ \ell^t_{\mathscr{K}\mathscr{S}} - \mathsf{p} \circ \ell^t_{\mathscr{D}\mathscr{S}} - \mathfrak{b}\,\mathfrak{p} \\ &= (1 + \mathfrak{p})^{-1}. \end{aligned} \tag{1.10}$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

Hence

$$\mathsf{p}^t(x) \leqslant \mathsf{p}^{t+1}(x) \leqslant \mathsf{p}^t(x) \cdot (1 + \mathfrak{p}), \tag{1.11}$$

unless x is *killed*, *i.e.*, unless $\mathsf{p}^{t+1}(x) = 0$.

By the construction (1.8), the process p^t is a *martingale*, so that

$$\mathbb{E}_t[\mathsf{p}^{t+1}] = \mathsf{p}^t. \tag{1.12}$$

Relations (1.11) and (1.12) will be essential in deriving the concentration results upon which the proof is founded.

1.6. The definition of Γ

The process p^t is controlled by keeping a set of local inequalities alive through the iterations. These local relations make up the notion of 'goodness' that we rely on throughout the arguments. First, for $x = (i, k, \gamma) \in \mathscr{X}$, let the inequality (A_x) be defined by

$$\mathsf{p}^t(x) \leqslant 1.1\, \varepsilon^{-2}\, \frac{1}{n}. \tag{A_x}$$

Secondly, for a local line $\ell = \ell_{\mathscr{R}\mathscr{K}}, \ell_{\mathscr{R}\mathscr{S}}$ and $x \in \mathscr{X}$, we say that $(B_{\ell(x)})$ holds if

$$\left(1 - \frac{\log n}{\sqrt{n}}\right) \leqslant \sum_{y \in \ell(x)} \mathsf{p}^t(y) \leqslant \left(1 + \frac{\log n}{\sqrt{n}}\right). \tag{$B_{\ell(x)}$}$$

Finally, for $k \neq l \in \mathscr{K}$ and $i \in \mathscr{R}$ let $(C_{i,kl})$ be the 'quasi-random' inequality

$$\sum_{\gamma} \mathsf{p}^t(i, k, \gamma)\, \mathsf{p}^t(i, l, \gamma) \leqslant \frac{1}{n}\left(1 + \frac{\log n}{\sqrt{n}}\right). \tag{$C_{i,kl}$}$$

Definition 1. We say that the state $\mathsf{p}^t \in [0,1]^{\mathscr{X}}$ is *good* if (A_x), $(B_{\ell_{\mathscr{R}\mathscr{K}}(x)})$ and $(B_{\ell_{\mathscr{R}\mathscr{K}}(x)})$ hold at all $x \in X$ and, in addition, $(C_{i,kl})$ hold for all $k, l \in \mathscr{K}$ and $i \in \mathscr{R}$. We write Γ for the region $\Gamma \subset [0,1]^{\mathscr{X}}$ of good states.

It is trivial to check that the initial state, $\mathsf{p}^0 = 1/n$, is an element of Γ.

1.6.1. Asymptotic relaxations. The precise formulations of (A_x), $(B_{\ell(x)})$ and $(C_{i,kl})$ are needed to make Γ well defined, but that precision is otherwise not critical. What we actually will use in the computations below are the following less precise asymptotic statements

$$\mathsf{p}^t(x) = \mathfrak{p}, \tag{A}$$

$$\sum_{\ell(x)} \mathsf{p}^t = 1 \pm \mathfrak{a}, \tag{B}$$

and

$$\sum_{\gamma} \mathsf{p}^t(i, k, \gamma)\, \mathsf{p}^t(i, l, \gamma) \leqslant (1 + \mathfrak{a})\, \frac{1}{n}. \tag{C}$$

It should be noted that the states p^t, if they eventually leave Γ at the time T, will stay quite close to Γ. This is due to the fact that we stop p^t at time $t = T$, with the previous state p^{T-1}

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

being a good state. As a consequence, we can use the somewhat relaxed bounds (A), (B) and (C) in our arguments, *without considering if we are conditioning on $t < T$ or not*, since these bounds then hold for all times t. In particular, we can assume the relaxed bounds when we later show that p^t with high probability stays inside Γ. (For definiteness, one may choose to replace the implicit constants in (A), (B) and (C) with explicit constants slightly larger or smaller than those used in the definition of Γ.)

In order to see that p^T is close to Γ in this sense, note that from (1.11) it is clear that the left-hand side of (A) can only increase with a fraction $1 + \mathfrak{p}$ at a time. Similarly, the left-hand side of (C) can increase by at most a factor $(1 + \mathfrak{p})^2 = 1 + \mathfrak{a}$ at time. Finally, from the fact that at most two terms (see the argument preceding (3.10) below) in the sums in (B) can be killed at a time, it also follows that the left-hand side of (B) can only change by a fraction $1 \pm \mathfrak{p}$ at a time.

2. The proof of Lemma 1.2

The reason behind introducing the set Γ is Lemma 1.2 stated in the Introduction. This lemma shows the existence of q^t and hence lets us define the extension L^{t+1} of L^t as long as $\mathsf{p}^t \in \Gamma$. We now proceed to prove this lemma.

We only have to look at one fixed row $i \in \mathscr{R}$ at a time. Given a good state $\mathsf{p}^t \in \Gamma$, let $p \in [0,1]^{\mathscr{K} \times \mathscr{S}}$ be given by

$$p(k,\gamma) := \mathsf{p}^t(i,k,\gamma) / \sum_l \mathsf{p}^t(i,l,\gamma), \tag{2.1}$$

so that $p = (1 \pm \mathfrak{a})\, \mathsf{p}^t_i$ by (B) and p is normalized at each symbol, *i.e.*,

$$\sum_k p(k,\gamma) = 1. \tag{2.2}$$

We assume that p^t satisfies the inequalities (A), (B) and (C), which translates to the following set of inequalities:

$$p(k,\gamma) = \mathfrak{p}, \tag{2.3}$$

$$\sum_{\gamma \in \mathscr{S}} p(k,\gamma) = 1 + \mathfrak{a}, \tag{2.4}$$

$$\sum_{\gamma \in \mathscr{S}} p(k,\gamma) p(l,\gamma) \leqslant (1 + \mathfrak{a})/n, \tag{2.5}$$

for all values of $k, l \in \mathscr{K}$ and $\gamma \in \mathscr{S}$.

We want to show that for some $\eta = \mathfrak{b}$ there exists a rational matching, *i.e.*, a vector $q \in [0,1]^{\mathscr{K} \times \mathscr{S}}$ such that, for all k and γ, $\sum_{k'} q(k',\gamma) = \sum_{\gamma'} q(k,\gamma') = 1$, that in addition satisfies

$$q(k,\gamma) \leqslant (1 + \eta) \cdot p(k,\gamma), \quad \forall (k,\gamma) \in \mathscr{K} \times \mathscr{S}.$$

Such a rational matching can also be defined as a flow on a directed graph from a source s connected to all vertices $k \in \mathscr{K}$ to a sink t connected to all vertices $\gamma \in \mathscr{S}$. The flow should take the value 1 on each edge sk, $k \in \mathscr{K}$, and each edge γt, $\gamma \in \mathscr{S}$. On the remaining edges, of the form $k\gamma$, we prescribe the capacities $c_{k\gamma} = (1 + \eta) p(k,\gamma)$. The Ford–Fulkerson theorem says that

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

it is enough to show that, for all pairs of nonempty sets $A \subsetneq \mathscr{S}$ and $B \subsetneq \mathscr{K}$, we have

$$2n - |A| - |B| + (1+\eta) \sum_{\substack{\gamma \in A \\ k \in B}} p(k,\gamma) \geqslant n, \tag{2.6}$$

since the left-hand side is the capacity for a cut in a flow defining a rational matching, where the capacity of edge $k\gamma$ equals $(1+\eta) \cdot p(k,\gamma)$.

Given an arbitrary pair of subsets A and B as above, we shall prove that, for some $\eta = \mathfrak{b}$ (with the implicit constant independent of A and B), the left-hand side in (2.6) is not strictly less than n if we assume that p satisfies the inequalities (2.3), (2.4) and (2.5).

Hence, assume that (2.6) does not hold and proceed to derive a contradiction. It must obviously then be the case that $|A| + |B| > n$ and both A and B must be non-empty. Let $a := |A|/n$, $b := |B|/n$. Then, $a, b > 0$ and

$$a + b > 1. \tag{2.7}$$

For $\gamma \in \mathscr{S}$ and $S \subset \mathscr{K}$, let

$$p(S,\gamma) := \sum_{k \in S} p(k,\gamma),$$

and define

$$x := b - \frac{1}{|A|} \sum_{\gamma \in A} p(B,\gamma) \quad \text{and} \quad y := b - \frac{1}{n-|A|} \sum_{\gamma \in \mathscr{S} \setminus A} p(B,\gamma). \tag{2.8}$$

It is easy to see that, by dividing both sides in (2.6) by n and substituting for a, b and x, the assumption that (2.6) does not hold is equivalent to

$$(1-a)(1-b) + ab\eta - ax(1+\eta) < 0. \tag{2.9}$$

In order to contradict (2.9), it is enough to show that

$$ax \leqslant ab\,\mathfrak{b}, \tag{2.10}$$

since we can take η equal to, say, two times the $\mathfrak{b}$-function on the right-hand side of (2.10).

We claim that it follows from (2.7) and (2.3)–(2.5) that

$$|ax + (1-a)y| \leqslant ab \cdot \mathfrak{a}, \tag{2.11}$$

and that

$$ax^2 + (1-a)y^2 \leqslant ab \cdot \mathfrak{a}\,. \tag{2.12}$$

Postponing the proofs for the two relations (2.11) and (2.12) until later, we proceed to show that they imply (2.10). We divide into two cases depending on the value of a. If $a \leqslant 1/2$ then (2.12) gives

$$ax^2 \leqslant ab\,\mathfrak{a} = ab^2\,\mathfrak{a},$$

where we use that (2.7) implies that $b > 1/2$ so that $\mathfrak{a} = b\,\mathfrak{a}$. Multiplying both sides with a and taking the square root gives (2.10), since $\sqrt{\mathfrak{a}} = \mathfrak{b}$.

In the case $a > 1/2$, then (2.12) gives that

$$(1-a)y^2 \leqslant ab\,\mathfrak{a},$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

and since $1 - a < b$ by (2.7) we can multiply the left-hand side by $1 - a$ and the right-hand side by b. This gives

$$(1-a)^2 y^2 \leqslant ab^2\,\mathfrak{a} = a^2 b^2\,\mathfrak{a},$$

where the last inequality is due to $a > 1/2$. Taking the square root, substituting in (2.11) and using the triangle inequality, we get

$$a|x| \leqslant ab\,\mathfrak{a} + (1-a)|y| \leqslant ab\,\mathfrak{a} + ab\,\mathfrak{b} = ab\,\mathfrak{b},$$

and (2.10) is proved. □

(It follows that we can take η to be four times the square root of the maximum $\mathfrak{a}$-function found in (2.11) and (2.12).)

What remains now is to show that (2.11) and (2.12) follows from (2.7) and the goodness assumptions (2.3), (2.4) and (2.5). Note that, by (2.7),

$$ab\,\mathfrak{a} \geqslant (1-b)b\,\mathfrak{a} \geqslant (1/2)\,\min\{b, 1-b\}\,\mathfrak{a} = \min\{b, 1-b\}\,\mathfrak{a}\,.$$

Consequently, it is enough to show (2.11) and (2.12) for the right hand side $\beta\,\mathfrak{a}$, $\beta \in \{b, 1-b\}$, instead of $ab\,\mathfrak{a}$.

Moreover, since for each γ, $p(B,\gamma) = 1 - p(\mathscr{K} \setminus B, \gamma)$, the value x and y will both change only in sign if we interchange B with $\mathscr{K} \setminus B$ in (2.8). Thus, if we do not use the assumption (2.9) (and, *e.g.*, (2.7)) about B, we may freely interchange B and $\mathscr{K} \setminus B$. This means that we only have to consider the case $\beta = b$.

We first show that $|ax + (1-a)y| \leqslant b\,\mathfrak{a}$. We have that $|A|(b-x) + (n-|A|)(b-y)$ equals $\sum_\gamma p(B,\gamma)$ which, by (2.4), is of order $|B|(1+\mathfrak{a})$. Dividing by n gives

$$a(b-x) + (1-a)(b-y) = b + b\,\mathfrak{a} \iff |ax + (1-a)y| = b\,\mathfrak{a}\,.$$

In order to see that $ax^2 + (1-a)y^2 \leqslant b\,\mathfrak{a}$, we let $z_{kl} = \sum_{\gamma \in \mathscr{S}} p(k,\gamma) \cdot p(l,\gamma)$ denote the left-hand side of (2.5) above. Note that

$$\sum_\gamma p(B,\gamma)^2 = \sum_{\substack{k,l \in B \\ k \neq l}} z_{kl} + \sum_{\substack{\gamma \\ k \in B}} p^2(k,\gamma). \tag{2.13}$$

The last sum in (2.13) above is of the order $|B|\,\mathfrak{p}$ by (2.3) and the first sum on the right-hand side is less than $|B|^2\,(1+\mathfrak{a})/n$ by (2.5). Furthermore, the Cauchy–Schwarz inequality implies that the left-hand side of (2.13) is greater than

$$|A| \left(\sum_{\gamma \in A} p(B,\gamma) \right)^2 + (n - |A|) \left(\sum_{\gamma \notin A} p(B,\gamma) \right)^2 = |A|(b-x)^2 + (n-|A|)(b-y)^2.$$

Thus, we have

$$|A|(b-x)^2 + (n-|A|)(b-y)^2 \leqslant |B|^2\,(1+\mathfrak{a})/n + |B|\,\mathfrak{p},$$

and dividing by n and expanding the squares gives the relation

$$ab^2 + (1-a)b^2 - 2b\,(ax + (1-a)y) + ax^2 + (1-a)y^2 \leqslant b^2 + b^2\,\mathfrak{a} + b\,\mathfrak{a},$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

which is equivalent to

$$ax^2 + (1-a)y^2 \leqslant b^2\,\mathfrak{a} + b\,\mathfrak{a} + 2b\,(ax + (1-a)y).$$

The sought statement follows if we use (2.11) to estimate the last term. □

3. The proof of Lemma 1.3

For the parameter n tending to infinity, we stipulate that an event has *very high probability* if it holds with probability having asymptotic order $1 - n^{-\omega(1)}$. An inequality of the form $X^t \leqslant f(t)$, for all $t \in [0, m]$, that holds with very high probability is said to be *stable*. Note that, since we only consider $n^{O(1)}$ different times t, a stable inequality holds with very high probability *simultaneously* for all $t \in [0, m]$.

In order to conclude the proof of the lemma, it suffices to show that, for every possible value of x, i, k, l, γ and ℓ,

$$\text{inequalities } (A_x), (B_{\ell(x)}) \text{ and } (C_{i,kl}) \text{ defining } \Gamma \text{ are all stable.} \tag{3.1}$$

Since the definition of Γ considers $O(n^4)$ such inequalities, the probability that $\mathsf{p}^t \notin \Gamma$ is then shown to be of order $O(n^{-\omega(1)+4}) = n^{-\omega(1)}$.

In the first subsection, we state and prove a more general 'concentration lemma', Lemma 3.1, and then we prove, in three separate subsections, the stability of $(B_{\ell(x)})$, $(C_{i,kl})$ and (A_x), where we regularly invoke Lemma 3.1. Of the three, proving the stability of (A_x) involves the most complex argument, but it should be noted that analysing the structure of the linear hypergraph given by the lines is also an essential component for deriving the other two statements.

Again it should be noted, as in Section 1.6, that, since we stop p^t at time T, we can use the bounds (A), (B) and (C) in our arguments to derive the stability of $(B_{\ell(x)})$, $(C_{i,kl})$ and (A_x), without considering if we are conditioning on $t < T$ or not. It should be clear that we at no point make the assumption that the process p^t is unstopped. In particular, the conditions and conclusions of Lemma 3.1 work for stopped processes as well as 'live' processes.

3.1. A concentration result

The following lemma is a consequence of the *Azuma–Hoeffding inequality*: see, *e.g.*, [1]. In order to make the subsequent invocations transparent, we put the lemma in a suitable form and make no attempt to derive the best-possible result.

Lemma 3.1. *Let ξ and $\alpha^0, \alpha^1, \dots, \alpha^{m-1}$ be positive numbers such that $(\xi\sqrt{n} + a)\log n = o(1)$, where $a := \sum_{t=0}^{m-1} \alpha^t$. Let $X^t \geqslant 0$, $t \in [0, m]$, be a positive process such that*

$$|X^{t+1} - \mathbb{E}_t[X^{t+1}]| \leqslant \xi\, \max\{X^t, X^0\}, \tag{3.2}$$

and

$$\mathbb{E}_t[X^{t+1}] - X^t \leqslant \alpha^t\, \max\{X^t, X^0\}. \tag{3.3}$$

Then, for all $t \in [0, m]$,

$$\mathbb{P}\{X^t \leqslant (1+\Phi)\, X^0\} = 1 - n^{-\omega(1)}, \tag{3.4}$$

for any $\Phi = \Omega((\xi\sqrt{n} + a)\log n)$. Furthermore, if X^t is a martingale *(in which case $a = 0$), then*

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

the reverse inequality

$$X^t \geqslant (1-\Phi)\, X^0 \tag{3.5}$$

is stable as well.

Proof of Lemma 3.1. Note that the inequalities (3.2) and (3.3) are still valid if we stop the process at any stopping time τ. If we take τ to be the first time that $X^t > 2 \cdot X^0$ then, since

$$X^{s+1} \leqslant (1+\xi+\alpha^s)\, X^s = (1+o(1))\, X^s,$$

we can assume that the stopped process $X^t = X^{t\wedge\tau} \leqslant 3X^0$. It is therefore enough to prove the stability of (3.4) with the additional assumption that $X^t \leqslant 3 \cdot X^0$ for all t.

Consider the martingale

$$M^t := \sum_{s=0}^{t-1} \big(X^{s+1} - \mathbb{E}_s[X^{s+1}]\big),$$

and the previsible process

$$A^t := \sum_{s=0}^{t-1} \big(\mathbb{E}_s[X^{s+1}] - X^s\big).$$

The identity

$$X^t = X^0 + A^t + M^t \tag{3.6}$$

is called the Doob decomposition of X^t. We refer to the terms of A^t as *drifts* of X^t and to the terms in M^t as *deviations*.

On account of the bound $X^t \leqslant 3X^0$, we obtain from (3.3) that $A^t \leqslant a\, 3X^0$, and, from (3.2), that $|M^{t+1} - \mathbb{E}_t[M^{t+1}]| \leqslant \xi\, 3X^0$. Since M^t is a martingale, the Azuma–Hoeffding inequality implies that

$$\mathbb{P}\{|M^t| > \lambda\} < \exp\{-\lambda^2/4(\xi\sqrt{n}\, 3X^0)^2\}$$

if

$$\lambda = \Theta(\xi\sqrt{n}(\log n)) \cdot X^0.$$

It is then also seen that $|M^t| > \lambda$ is an event of probability of order $n^{-\omega(1)}$.

Hence, the probability $X^t - X^0 > \Phi X^0$ where

$$\Phi X^0 = \Omega(a \log n) X^0 + \Omega(\xi\sqrt{n} \log n) X^0$$

is of order $n^{-\omega(1)}$; in the display above the first term is greater than $A^t \leqslant 3aX^0$ and the second term bounds M^t within very high probability. This proves (3.4). The stability of (3.5) follows in the same manner from the stability of the event $M^t > -\lambda$. (In this case $A^t = 0$, since X is a martingale.) □

3.1.1. An estimate of ξ for a certain type of sums. The inequalities we are dealing with have a common form, and we will repeatedly use the formula in (3.8) below in order to to estimate the deviation parameter ξ used in Lemma 3.1.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

The processes we consider are sums

$$X^t = \sum_{i \in \mathscr{J}} X_i^t, \qquad X_i^t \geqslant 0,$$

for some index set $\mathscr{J}$. The terms have uniform bounds

$$0 \leqslant X_i^t \leqslant \mathfrak{m},$$

for some asymptotic expression $\mathfrak{m}$. It will also be the case that each term $X_i^t \geqslant 0$ in the sum changes moderately in the following sense: for $\xi_+, \xi_- > 0$, we have for all $t \in [0, m]$ and $i \in \mathscr{J}$ that

$$(1 - \xi_-)X_i^t \leqslant X_i^{t+1} \leqslant (1 + \xi_+)X_i^t, \tag{3.7}$$

unless the term X_i^t is *killed*, *i.e.*, unless $X_i^{t+1} = 0$ but $X_i^t > 0$. We assume that the maximum number of such terms killed is furthermore given by $\mathscr{N}$.

The following lemma is then immediate.

Lemma 3.2. *With X, $\mathfrak{m}$, $\mathscr{N}$, ξ_+ and ξ_- as above, the conclusions of Lemma 3.1 hold with*

$$\xi = \frac{\mathscr{N}\,\mathfrak{m}}{X^0} + \xi_- + \xi_+. \tag{3.8}$$

3.2. The stability of $(B_{\ell(x)})$

Let $\ell(x)$ be any local line, *i.e.*, $\ell(x) = \ell_{\mathscr{R}\mathscr{K}}(x)$ or $\ell(x) = \ell_{\mathscr{R}\mathscr{S}}(x)$. For $t \in [0, m]$, let

$$X^t := \sum_{y \in \ell(x)} \mathsf{p}^t(y). \tag{3.9}$$

Our objective is to show that $(1 - \Phi) \leqslant X^t \leqslant (1 + \Phi)$ with very high probability, where $\Phi = \log n/\sqrt{n}$. By the martingale property (1.12), the drift, a, for X^s is zero and we have that $X^0 = 1$. Unless $\mathsf{p}^{t+1}(y) = 0$, we have, by (1.11), that $\mathsf{p}^t(y) \leqslant \mathsf{p}^{t+1}(y) \leqslant (1 + \mathfrak{p})\,\mathsf{p}^t(y)$. Thus, in the notation from Lemma 3.2, we have $\xi_- = 0$ and $\xi_+ = \mathfrak{p}$. Moreover, each term in (3.9) is smaller than $\mathfrak{p}$ (by (A)), so we have $\mathfrak{m} = \mathfrak{p}$ and hence, by Lemma 3.2, $\xi = \mathscr{N}\ \mathfrak{p} + 0 + \mathfrak{p}$.

Thus, in order to show that $\xi = \mathfrak{p}$, it only remains to show that $\mathscr{N}$ – the maximum number of terms 'killed' – is of order $O(1)$. We claim that $\mathscr{N} \leqslant 2$. The number of terms killed is $\sum_{y \in \ell(x)} \mathsf{K}^{t+1}(y)$, where $\mathsf{K}^{t+1} = \mathsf{L}^{t+1} \circ \ell^t_{\mathscr{K}\mathscr{S}} + \mathsf{L}^{t+1} \circ \ell^t_{\mathscr{D}\mathscr{S}}$. Moreover, the maps $y \to z = \ell^t_{\mathscr{K}\mathscr{S}}(y)$ and $y \to z = \ell^t_{\mathscr{D}\mathscr{S}}(y)$ map $y \in \ell(x)$ one-to-one into a corresponding local line $z \in \ell(\ell^t_{\mathscr{K}\mathscr{S}}(y))$ and $z \in \ell(\ell^t_{\mathscr{D}\mathscr{S}}(y))$. Since L^{t+1} is Latin this means that

$$\mathscr{N} \leqslant \sum_y \mathsf{L}^{t+1} \circ \ell^t_{\mathscr{K}\mathscr{S}}(y) + \sum_y \mathsf{L}^{t+1} \circ \ell^t_{\mathscr{D}\mathscr{S}}(y) \leqslant 1 + 1 = 2. \tag{3.10}$$

□

3.3. Proof of stability for $(C_{i,kl})$

For a fixed $i \in \mathscr{R}$ and $k, l \in \mathscr{K}$, $k \neq l$, we consider the sum $X^t := \sum_\gamma X^t_\gamma$ where, for $\gamma \in \mathscr{S}$,

$$X^t_\gamma := \mathsf{p}^t(ik\gamma) \cdot \mathsf{p}^t(il\gamma). \tag{3.11}$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

We have $X^0 = 1/n$ and $X_\gamma \leqslant \mathfrak{p}^2$ by (A). Our objective is to prove that, with very high probability, $X^t \leqslant (1+\Phi)X^0$ where $\Phi = \log n/\sqrt{n}$. By Lemma 3.1, it is enough to show that $\xi = \mathfrak{p}$ and $a = \mathfrak{p}$. These bounds are proved in (3.12) and (3.14) below.

By (1.11), each term X_γ^{t+1} will either increase with a factor at most $(1+\mathfrak{p})^2 = 1+\mathfrak{p}$ or be killed. Furthermore, at most four terms can be killed, since, by the computation already done in (3.10), at most two points in each of the cells (i,k) and (i,l) are killed. Hence, with ξ as in Lemma 3.2,

$$\xi \leqslant \frac{4\,\mathfrak{p}^2}{1/n} + 0 + \mathfrak{p} = \mathfrak{p}\,. \tag{3.12}$$

In order to calculate the drift a of X^t, fix $\gamma \in \mathscr{S}$ and let $K_k = \mathsf{K}^{t+1}(i,k,\gamma)$ and $r_k = \mathbb{E}_t[K_k]$. Note that $r_k = \mathsf{q}^t \circ \ell_{\mathscr{K}\mathscr{S}}(i,k,\gamma) + \mathsf{q}^t \circ \ell_{\mathscr{D}\mathscr{S}}(i,k,\gamma) = \mathfrak{p}$. Then

$$\begin{aligned}\mathbb{E}_t[X_\gamma^{t+1}] &= X_\gamma^t \cdot \frac{\mathbb{E}_t[(1-K_k)(1-K_l)]}{\mathbb{E}_t[(1-K_k)]\mathbb{E}_t[(1-K_l)]} \\ &= X_\gamma^t \cdot \frac{1-r_k-r_l+\mathbb{E}_t[K_kK_l]}{1-r_k-r_l+r_kr_l}.\end{aligned} \tag{3.13}$$

But the event $K_kK_l \neq 0$ can happen only when the diagonal through the cell (i,k) intersects row $t+1$ in the column number l and *vice versa*, and this can be the case for at most two values of t. For all other values of t, the diagonals and columns through cells (i,k) and (i,l) intersect row $t+1$ at four disjoint positions. For these t, at most one cell in row $t+1$ is coloured by any colour γ, so at most one of the indicators K_k and K_l equals one making $K_kK_l \equiv 0$. Hence, for all but at most two values of t and *uniformly for all* γ, it holds that $\mathbb{E}_t[K_kK_l] = 0$, and from (3.13) it is clear that $\mathbb{E}_t[X_\gamma^{t+1}] \leqslant X_\gamma^t$ for these t.

Also, for the two possible exceptional values of t, when $\mathbb{E}_t[K_kK_l]$ is positive, we clearly have that

$$\mathbb{E}_t[K_kK_l] \leqslant \min\{r_k, r_l\} \leqslant \mathfrak{p},$$

even assuming the worst-possible correlation. From (3.13), we deduce that $\mathbb{E}_t[X^{t+1}] \leqslant X^t + \alpha \max\{X^0, X^t\}$, with $\alpha = \mathfrak{p}$. Putting this together gives that $a = \sum_{s=0}^{t-1} \alpha^s$, as in Lemma 3.1, is bounded by

$$a \leqslant (t-2)\cdot 0 + 2\,\mathfrak{p} = \mathfrak{p}\,. \tag{3.14}$$

□

3.4. Proving that (A_x) is stable

The derivation of the stability of (A_x) is a bit more involved. The important part is perhaps the identity (3.21) below, which makes it possible to relate the growth of the product

$$\prod_{s=0}^{t}(1-\mathsf{p}^s \circ \ell^s)^{-1}$$

to that of a sum

$$S^t = \sum_{s=0}^{t-1} \mathsf{p}_\ell^s \circ \ell^s$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

with suitable concentration properties. Our aim is to prove that, with very high probability,

$$\frac{\mathsf{p}^t(x)}{\mathsf{p}^0(x)} \leqslant (1+o(1))\left(1-(1+o(1))\,\frac{t}{n}\right)^{-2}. \tag{3.15}$$

Since $t/n \leqslant (1-\varepsilon) = 1-\Omega(1)$, the inequality (3.15) implies (A_x) for large enough n.

Let ℓ denote a central parallel class, *i.e.*, ℓ is either of $\ell_{\mathscr{K}\mathscr{S}}$ or $\ell_{\mathscr{D}\mathscr{S}}$. Define the vector of stopping times $\tau_\ell = \tau_\ell(x)$, $x \in \mathscr{X}$, giving the time that the central line $\ell(x)$ is 'killed', *i.e.*, let

$$\tau_\ell(x) := \inf\left\{t \,:\, \sum_{y\in\ell(x)} \mathsf{L}^t(y) = 1\right\} \cup \{\infty\}. \tag{3.16}$$

Also, let

$$\hat{t}(x,t) := t \wedge (\tau_{\ell_{\mathscr{K}\mathscr{S}}}(x) \wedge \tau_{\ell_{\mathscr{D}\mathscr{S}}}(x) - 1) \wedge T \wedge \tau_c(x).$$

Note that the value of $\hat{t}(x,t)$ is determined at time t for all t, and that $\mathsf{p}^{\hat{t}}(x) = \mathsf{p}^{\hat{t}(x,t)}(x)$ is an adapted process which is increasing in t. Moreover, we have $\mathsf{p}^t(x) = \mathsf{p}^{\hat{t}}(x)$ unless $t \geqslant \tau_{\ell_{\mathscr{K}\mathscr{S}}}(x) \wedge \tau_{\ell_{\mathscr{D}\mathscr{S}}}(x)$ and $\tau_{\ell_{\mathscr{K}\mathscr{S}}}(x) \wedge \tau_{\ell_{\mathscr{D}\mathscr{S}}}(x) < T \wedge \tau_c(x)$, in which case $\mathsf{p}^t(x) = 0$.

From the definition (1.8), we know that

$$\sup_{s\leqslant t} \frac{\mathsf{p}^s}{\mathsf{p}^0} = \prod_{s=0}^{\hat{t}-1} \frac{1}{1-\mathsf{q}^s\circ\ell^s_{\mathscr{K}\mathscr{S}} - \mathsf{q}^s\circ\ell^s_{\mathscr{D}\mathscr{S}}} = \prod_{s=0}^{\hat{t}-1} \frac{(1-\mathsf{p}^s\circ\ell^s_{\mathscr{K}\mathscr{S}})(1-\mathsf{p}^s\circ\ell^s_{\mathscr{D}\mathscr{S}})}{1-\mathsf{q}^s\circ\ell^s_{\mathscr{K}\mathscr{S}} - \mathsf{q}^s\circ\ell^s_{\mathscr{D}\mathscr{S}}} \cdot \prod_{\ell=\ell_{\mathscr{K}\mathscr{S}},\ell_{\mathscr{D}\mathscr{S}}} \prod_{s=0}^{\hat{t}-1} (1-\mathsf{p}^s\circ\ell^s)^{-1}. \tag{3.17}$$

Note that the first factor to the right is negligible: by (A) and the estimate in (1.10), it is of order $(1+\mathfrak{b}\,\mathfrak{p}+\mathfrak{p}^2)^{\hat{t}} = (1+\mathfrak{b})$. The aim will thus be to show that

$$\prod_{s=0}^{\hat{t}-1}(1-\mathsf{p}^s\circ\ell^s)^{-1} = \left(1-(1+o(1))\,\frac{\hat{t}}{n}\right)^{-1}, \tag{3.18}$$

where $\ell = \ell_{\mathscr{K}\mathscr{S}}$ or $\ell = \ell_{\mathscr{D}\mathscr{S}}$.

For $\ell = \ell_{\mathscr{K}\mathscr{S}}$ or $\ell = \ell_{\mathscr{D}\mathscr{S}}$, define recursively the adapted process $\mathsf{p}^t_\ell \in [0,1]^{\mathscr{X}}$, $t \in [0,m]$, as follows. Let $\mathsf{p}^0_\ell(x) = \mathsf{p}^0(x) = \frac{1}{n}$ and set

$$\mathsf{p}^{t+1}_\ell(x) := \begin{cases} (1-S^{t+1}(x))\cdot\mathsf{p}^{t+1}(x) & t \leqslant (\tau_\ell(x)-1)\wedge T\wedge\tau_c(x), \\ \mathsf{p}^t_\ell(x) & \text{otherwise,} \end{cases} \tag{3.19}$$

where

$$S^t(x) := \sum_{s=0}^{t-1} \mathsf{p}^s_\ell\circ\ell^s(x).$$

The definition of p^t_ℓ in (3.19) implies, for all $y \in \ell(x)$, that either $\mathsf{p}^{t+1}_\ell(y) = 0$, $\mathsf{p}^{t+1}_\ell(y) = \mathsf{p}^t_\ell(y)$ or else

$$\mathsf{p}^{t+1}_\ell(y) = \frac{(1-\mathsf{p}^t\circ\ell^s(y))\,\mathsf{p}^t_\ell(y)}{1-\mathsf{q}^s\circ\ell^s_{\mathscr{K}\mathscr{S}}(y) - \mathsf{q}^s\circ\ell^s_{\mathscr{D}\mathscr{S}}(y)}.$$

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

Therefore,

$$\mathsf{p}_\ell^t(y)\cdot(1-\mathfrak{p}) \leqslant \mathsf{p}_\ell^{t+1}(y) < \mathsf{p}_\ell^t(y)\cdot(1+\mathfrak{p}), \tag{3.20}$$

except for the case when the term $\mathsf{p}_\ell^t(y)$ is killed. Note also that killing of $\mathsf{p}_\ell^t(y)$ occurs exactly when $\tau_{\ell'}(y) = t+1 < \tau_\ell(y)$, where $\ell'(y)$ is the *complementary* central line, *i.e.*, $\ell' = \ell_{\mathscr{K}\mathscr{S}}$ if $\ell = \ell_{\mathscr{D}\mathscr{S}}$ and *vice versa*.

If $t = \hat{t}$ then $\mathsf{p}^t = \mathsf{p}_\ell^t/(1-S^t)$ and we deduce the identity

$$\prod_{s=0}^{\hat{t}-1}(1-\mathsf{p}^s\circ\ell^s)^{-1} = \prod_{s=0}^{\hat{t}-1}\frac{1-S^s}{1-S^{s+1}} = \frac{1}{1-S^{\hat{t}}}. \tag{3.21}$$

In order to prove the stability of (3.15) it is therefore enough to show that, for all x, with very high probability

$$S^{t_0}(x) \leqslant (1+o(1))\,(t_0/n), \tag{3.22}$$

where the time $t_0 \in [0,m]$ is fixed but arbitrary.

So fix $t_0 \in [0,m]$ and let $x \in \mathscr{X}$ be arbitrary. In order to ease the notation, we from now on suppress the dependency of $x \in \mathscr{X}$ for most quantities. Let $\ell^{<t_0}(x) := \{y \in \ell(x) \,:\, \tau_c(y) < t_0\}$. Define for $t \in [0,t]$ the variable $X^t \in \mathbb{R}_+^{\mathscr{X}}$ by

$$X^t(x) := \sum_{y\in\ell^{<t_0}(x)} \mathsf{p}_\ell^t(y).$$

Since $\mathsf{p}_\ell^t = \mathsf{p}_\ell^{t\wedge\tau_c}$, we have that $X^{t_0} = S^{t_0}$. Since $|\ell^{<t_0}| = t_0$ and $\mathsf{p}_\ell^0 = 1/n$, we have $X^0 = t_0/n$. Moreover, on account of (3.20), we have $X^t \leqslant (1+t\,\mathfrak{p})X^0$ and thus the sought bound, (3.22), clearly follows, if $t_0 = o(n)$. We may therefore assume that t_0 is greater than, say, $n^{2/3}$ and hence that $X^0 \geqslant n^{-1/3}$.

Thus, we can conclude the proof by showing that the total drift, a, of X^t satisfies $a \leqslant \mathfrak{b}$ and that the deviation parameter, ξ, satisfies $\xi = O(n^{-2/3})$. Then, from Lemma 3.1, we can deduce that

$$S^{t_0} = X^{t_0} \leqslant \big(1+O(n^{-1/6}\,\log n)\big)\,(t_0/n),$$

with very high probability.

From the definition (3.19), we know that $\mathbb{E}_t[\mathsf{p}_\ell^{t+1}/\mathsf{p}_\ell^t]$ equals 1 in the case when $t \geqslant \tau_\ell \wedge T \wedge \tau_c$, and that otherwise it equals

$$\begin{aligned}\mathbb{P}_t\{\tau_\ell = t+1\}\cdot 1 + \mathbb{P}_t\{\tau_\ell > t+1\}\cdot\frac{1-S^{t+1}}{1-S^t}\cdot\mathbb{E}_t[\mathsf{p}^{t+1}/\mathsf{p}^t \mid \tau_\ell > t+1] \\ = \mathsf{q}^t\circ\ell^t + (1-\mathsf{q}^t\circ\ell^t)\cdot\frac{1-\mathsf{p}^t\circ\ell^t}{1-\mathsf{q}^t\circ\ell^t} = 1+\mathsf{q}^t\circ\ell^t-\mathsf{p}^t\circ\ell^t,\end{aligned}$$

where we have used that $\frac{1-S^{t+1}}{1-S^t} = 1-\mathsf{p}^t\circ\ell^t$ and that $\mathbb{E}_t[\mathsf{p}^{t+1}/\mathsf{p}^t \mid \tau_\ell > t+1]$ equals

$$\frac{\mathsf{q}^t\circ\ell'^t}{1-\mathsf{q}^t\circ\ell^t}\cdot 0 + \left(1-\frac{\mathsf{q}^t\circ\ell'^t}{1-\mathsf{q}^t\circ\ell^t}\right)\cdot\frac{1}{1-\mathsf{q}^t\circ\ell^t-\mathsf{q}^t\circ\ell'^t} = \frac{1}{1-\mathsf{q}^t\circ\ell^t},$$

where ℓ' is the complementary line.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press

Hence, for all t,

$$\mathbb{E}_t\left[\mathsf{p}_\ell^{t+1}\right] = \begin{cases} \mathsf{p}_\ell^t \cdot (1 + \mathsf{q}^t \circ \ell^t - \mathsf{p}^t \circ \ell^t) & t < \tau_c \wedge T \wedge \tau_\ell \\ \mathsf{p}_\ell^t & \text{otherwise} \end{cases} \tag{3.23}$$
$$\leqslant \mathsf{p}_\ell^t \cdot (1 + \mathfrak{p}\,\mathfrak{b}).$$

From (3.23) we can estimate the drift term

$$a \leqslant t_0\,\mathfrak{b}\,\mathfrak{p} = \mathfrak{b}\,.$$

Note that only one cell in the row $t+1$ is coloured with the colour γ common to all points in $\ell(x)$, and that this cell can lie on at most one crossing central line $\ell'(y)$ intersecting the given central line $\ell(x)$. This means that only one term in the sum $\sum_{\ell'<t_0} \mathsf{p}_\ell^t$ above can be killed at a time. Putting this and the estimate (3.20) into the formula (3.8) gives

$$\xi \leqslant 1 \cdot \mathfrak{p}\,/X^0 + \mathfrak{p} + \mathfrak{p} = O(n^{-2/3}).$$

□

References

[1] Alon, N. and Spencer, J. (1992) *The Probabilistic Method*, Wiley.
[2] Birkhoff, G. (1946) Three observations on linear algebra. *Univ. Nac. Tucumán. Rev. Ser. A* **5** 147–151.
[3] Brouwer, A. E., de Vries, A. J. and Wieringa, R. M. A (1978) A lower bound for the length of a partial transversal in a Latin square. *Nieuw Archief vor Wieskunde* (3) **XXVI** 330–332.
[4] Dénes, J. and Keedwell, A. D., eds (1974) *Latin Squares and their Applications*, Academic Press, New York.
[5] Drisko, A. A. (1998) Transversals in row-Latin rectangles. *J. Combin. Theory Ser. A* **84** 181–195.
[6] Hilton, A. J. W. (1994) Problem BCC 13.20. *Discrete Math.* **125** 407–417.
[7] Jensen, T. R. and Toft, B. (1995) *Graph Colouring Problems*, Wiley, New York.
[8] Kahn, J. (1996) Asymptotically good list-colourings. *J. Combin. Theory. Ser. A* **73** 1–59.
[9] Koksma, K. K. (1969) A lower bound for the order of a partial transversal in a Latin square. *J. Combin. Theory* **7** 94–95.
[10] Molloy, M. S. and Reed, B. (2002) *Graph Colouring and the Probabilistic Method*, Springer.
[11] Shor, P. W. (1982) A lower bound for the length of a partial transversal in a Latin square. *J. Combin. Theory Ser. A* **33** 1–8.
[12] Stein, S. K. (1975) Transversals of Latin squares and their generalizations. *Pacific J. Math.* **59** 567–575.
[13] Woolbright, D. E. (1978) A $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols. *J. Combin. Theory Ser. A* **24** 235–237.

https://doi.org/10.1017/S0963548307008590 Published online by Cambridge University Press