On maximal premature partial Latin squares

Anton Černý

Department of Mathematics and Computer Science,
Kuwait University,
P.O. Box 5969 Safat 13060, Kuwait
cerny@mcs.sci.kuniv.edu.kw

Abstract

A partial Latin square is premature if it has no completion, but it admits a completion after removing any of its symbols. This type of partial Latin square has been introduced by Branković, Horák, Miller and Rosa [Ars Combinatoria, to appear] where the authors showed that the number of empty cells in an $n \times n$ premature latin square is at least $3n - 4$. We improve this lower bound to $7n/2 - o(n)$.

1 Introduction

A *partial Latin square* is an $n \times n$ array partially filled by symbols — numbers from $\{1, 2, \ldots, n\}$ — such that each row and each column contains each symbol at most once. It is a *(complete) Latin square* if each cell of the array is filled by some number. A partial Latin square is *premature* if it cannot be completed to a Latin square, but such a completion exists after erasing the contents of any single one of its cells. Premature partial Latin squares were introduced in [1]. The property of being premature is quite close to that of being a critical set in a Latin square, which has several interesting applications, e.g. in design theory, group theory, graph theory or cryptography (a survey is given in [3]). However, as pointed out in [1], investigation of premature Latin squares requires different techniques from the ones used for critical sets. One of the natural problems, extensively studied in [1], is to characterize the spectrum of the size of premature Latin squares. The question as to how large a premature Latin square can be is of particular interest. The authors have shown in [1] that the size of a maximal premature Latin square of order n is asymptotic to n^2 while there are always at least $3n - 4$ empty cells. They further stated a conjecture, that there are always at least $n^{\frac{3}{2}}$ empty cells. Recently Branković and Miller ([2]) showed that if a premature partial Latin square contains a row (or a column) with $n-1$ full cells then it contains at least $4n - 10$ empty cells. We present here a slight improvement of the lower bound on the number of empty cells in any premature partial Latin square to $7n/2 - o(n)$, being still far below the non-linear conjecture.

2 Preliminaries

We will denote $[n] = \{1, 2, ..., n\}$ for $n \geq 1$. We will refer to the positions in an $n \times n$ array partially filled by elements of $[n]$ as *cells* and to entries in these cells as *symbols*. Such an array is a *partial Latin square* (*pls* for short) if it contains each symbol in each row and in each column at most once. By $S(i, j)$ we denote the symbol contained in the cell at position (i, j) of the *pls* S; we write $S(i, j) = \varepsilon$ if the cell is empty. A *(complete) Latin square* is a *pls* having all its cells filled.

Alternatively, a *pls* S can be described as a set of triples (i, j, k) where each pair (i, j), (j, k), (k, i) occurs at most once. Using our previous notation, $S = \{(i, j, k) \in [n]^3;\ S(i, j) = k\}$. The partial Latin squares obtained from S by permutation of entries in the triples will be called conjugates of S. Using conjugacy, properties of partial Latin squares expressed in terms of rows, columns and symbols can be translated to conjugate properties obtained by permuting the three terms.

An $n \times n$ *pls* S is *premature*, if it cannot be completed (i.e., as a set of triples, it is not a subset of any complete $n \times n$ Latin square), but any proper subset of S can be completed. The property of being premature is obviously preserved by conjugacy.

In the remaining text we assume that L is an arbitrary but fixed $n \times n$ premature Latin square, $n \geq 8$. We will denote by $C_{i,j}$ one (any) Latin square being a completion of the *pls* obtained from L by erasing the cell (i, j). We denote as r_i, c_j, s_k respectively the number of empty cells in row i, the number of empty cells in column j, and the number of occurrences of symbol k missing in L (i.e. $s_k = n-$ "*the number of occurrences of symbol k in L*"). We further let $E = \Sigma_i r_i = \Sigma_j c_j = \Sigma_k s_k = n^2 - |S|$, the total number of empty cells in L.

For the rest of the paper, we will make an assumption that will exclude the singular case when all non-empty cells of L are in one row plus one column only. In this case L may not have some properties common to other premature squares. In such a premature square the row and the column may contain at most $n-1$ full cells each, hence there are at least $n^2 - 2n + 2 > \frac{7}{2}n$ empty cells in L. Therefore we will assume that for each position (i, j) there is a position (i', j') with $i' \neq i$, $j' \neq j$ and $L(i', j') \neq \varepsilon$.

3 The lower bound

3.1 Basic facts and the lower bound

For proving our result we will need two easy-to-prove lemmas. Let us denote by $RS(i)$ the set of all symbols from $[n]$ not occurring in row i of L and by $CS(j)$ the set of all symbols from $[n]$ not occurring in column j. We will call a row (a column) containing exactly m empty cells an m-row (m-column).

Lemma 1 *$RS(i) \cap CS(j) \neq \emptyset$ for each $i, j \in [n]$.*

Proof. Let $L(i, j) = \varepsilon$ and let $L(i', j') \neq \varepsilon$ for some $i' \neq i, j' \neq j$. Then $C_{i',j'}(i, j)$ belongs to the intersection. If $L(i, j) \neq \varepsilon$ then $C_{i,j}(i, j)$ belongs to the intersection. ■

Corollary 2 *The only symbol not occurring in a 1-row does not occur in L at all.*

From Lemma 1 we easily obtain a conjugate assertion of Theorem 2.3 from [1], based on a slightly different proof argument.

Corollary 3 *If $n \geq 3$ then $\sum_k \mathbf{s}_k^2 \geq n^2$.*

Proof. A symbol k occurs in $\mathbf{s}_k^2$ different intersections $RS(i) \cap CS(j)$. The inequality follows from the fact implied by Lemma 1 that the total number of symbol occurrences in all intersections $RS(i) \cap CS(j)$ is at least n^2. ■

The next lemma is a generalization of the assertions used in [1] for $m = 1, 2$.

Lemma 4 *If $j_1, j_2, ..., j_m$ are positions of all free cells in some m-row i, then the columns $j_1, j_2, ..., j_m$ together contain at least n empty cells.*

Proof. Let k be any symbol. Let us first assume that each of the columns $j_1, j_2, ..., j_m$ contains a single full cell with symbol k and there are no other full cells except in row i. Then $m \geq 2$, otherwise all full cells are in one row and one column only. Hence there are at least $2(n-1) \geq n$ free cells in columns $j_1, j_2, ..., j_m$. Let the former assumption be not true. We will prove that the symbol k is missing in at least one of the columns $j_1, j_2, ..., j_m$. If k occurs in row i at position j_r then k is placed in row i of C_{i,j_r} to some cell j_s which was originally free in L. Hence column j_s does not contain k. If k is not contained in row i then there is a cell (i', j') being outside of row i and either outside columns $j_1, j_2, ..., j_m$, or filled with a symbol different from k. In either case k may appear in row i of $C_{i',j'}$ in one of the columns $j_1, j_2, ..., j_m$ only; this column in L cannot contain k. ■

Corollary 5 *If some row contains $n-1$ symbols then its only empty cell belongs to an empty column.*

We are now able to state our result; the principal part of the proof will be contained in Section 3.2

Theorem 6 *Each premature partial Latin square contains at least $7n/2 - o(n)$ empty squares.*

Proof. We will distinguish the following cases.
1. $\min_i r_i \geq 4$ or $\min_j c_j \geq 4$. In this case $E \geq 4n > 7n/2$.
2. $\min_i r_i = 3$ and $\min_j c_j \leq 3$ (or the conjugate case $\min_i r_i \leq 3$ and $\min_j c_j = 3$). We apply Lemma 4 to a column with at most 3 empty cells. The corresponding 3 rows contain at least n free cells and the remaining rows at least $3(n-3)$ free cells, hence $E \geq 4n - 9 \geq 7n/2 - 5$.
3. $\min_i r_i = 1$, $\min_j c_j \leq 2$. Then L contains at least $4n - 10$ empty cells, as proved in [2].
4. $\min_i r_i = \min_j c_j = 2$. The lower bound will be proved in Section 3.2. ■

3.2 Case $\min_i r_i = \min_j c_j = 2$

In this section we assume that there are no 1-rows or 1-columns but there is at least one 2-row and at least one 2-column.

Let us denote as $\mathbf{R} = \{i \in [n]; r_i = 2\}$ the set of all 2-rows and as $\mathbf{C} = \{j \in [n]; c_j = 2\}$ (the set of all 2-columns).

Lemma 7 *If there are at most $n/2$ 2-rows or at most $n/2$ 2-columns then $E \geq 7n/2 - 6$.*

Proof. If there are at most $n/2$ different 2-rows then the remaining rows contain at least 3 free cells each. Applying Lemma 4 to one 2-column we get two rows containing together at least n empty cells. Consequently, $E \geq n + 3(n/2 - 2) + 2(n/2) = 7n/2 - 6$. The assertion for columns is obtained using conjugacy properties. ∎

For the rest of Part 3.2 we will assume that there are at least $n/2+1 \geq 5$ different 2-rows. and at least $n/2 + 1 \geq 5$ different 2-columns. Then using Lemma 1 we get the following property of the square L.

Lemma 8 *Either one of the sets $\bigcup_{i \in \mathbf{R}} RS(i)$, $\bigcup_{j \in \mathbf{C}} CS(j)$ contains at most 4 different symbols, or $\bigcap_{i \in \mathbf{R}} RS(i) \cap \bigcap_{j \in \mathbf{C}} CS(j) \neq \emptyset$.*

Proof. All our conclusions will be based on Lemma 1 using the fact that, for $i \in \mathbf{R}$ and $j \in \mathbf{C}$, $|RS(i)| = |CS(j)| = 2$. Only the following five situations are possible (i_1, i_2 denote pairwise different indices from $\mathbf{R}$, j_1, j_2, j_3 denote pairwise different indices from $\mathbf{C}$, and a, b, c, d denote pairwise different symbols):
1. For some j_1, j_2 there exist a, b, c, d such that $CS(j_1) = \{a, b\}$ and $CS(j_2) = \{c, d\}$. Each $RS(i)$ then contains one symbol from $CS(j_1)$ and one symbol from $C(j_2)$, hence $\bigcup_{i \in \mathbf{R}} R(i) \subset \{a, b, c, d\}$.
2. For some j_1, j_2, j_3 there exist a, b, c such that $CS(j_1) = \{a, b\}$, $CS(j_2) = \{a, c\}$, $CS(j_3) = \{b, c\}$. Then $\bigcup_{i \in \mathbf{R}} RS(i) \subset \{a, b, c\}$.
3. For some j_1, j_2, j_3 there exist a, b, c, d such that $CS(j_1) = \{a, b\}$, $CS(j_2) = \{a, c\}$, $CS(j_3) = \{a, d\}$ and the situation 1. does not occur (therefore $a \in \bigcap_{j \in \mathbf{C}} CS(i)$). Then $a \in \bigcap_{i \in \mathbf{R}} RS(i)$.
4. There exist a, b, c such that for each $j_1 \in \mathbf{C}$ either $CS(j_1) = \{a, b\}$ or $CS(j_1) = \{a, c\}$. Then $\bigcup_{j \in \mathbf{C}} CS(j) \subset \{a, b, c\}$.
5. There exist a, b such that for each $j_1 \in \mathbf{C}$, $CS(j_1) = \{a, b\}$. Then $\bigcup_{j \in \mathbf{C}} CS(j) \subset \{a, b\}$. ∎

In our considerations we will concentrate on the relative position of the free cells in different 2-rows (or different 2-columns). We will distinguish the subcases listed in the following proposition.

Proposition 9 *One of the following assertions is true:*
1. There are at least two 2-rows such that no two out of the four free cells in these rows are in the same column.
2. There are at least two 2-columns such that no two out of the four free cells in

these columns are in the same row.
3. There are three columns containing all free cells of all 2-rows but none of the three columns contains a free cell in each 2-row.
4. There are three rows containing all free cells of all 2-columns but none of the three rows contains a free cell in each 2-column.
5. There is a column that contains a free cell in each 2-row and there is a row that contains a free cell in each 2-column.

Lemma 10 *If 1. or 2. of Proposition 9 is true then* $E \geq 4n - 8 \geq 7n/2 - 4$.

Proof. If 1. is true then Lemma 4 applied twice (once for each of the two rows) implies existence of two disjoint pairs of columns each containing at least n free cells. Each of the remaining columns contains at least 2 free cells, hence $E \geq 2n + 2(n - 4) = 4n - 8$. The assertion for the case 2. is obtained using conjugacy. ■

Lemma 11 *If 3. or 4. of Proposition 9 is true then* $E \geq 7n/2 - 6$.

Proof. Let 3. be true and let the indices of the three columns be j_1, j_2, j_3. Since there are at least three 2-rows, Lemma 4 implies that each two of the columns contain together at least n empty cells. Hence $c_{j_1} + c_{j_2} \geq n$, $c_{j_2} + c_{j_3} \geq n$, $c_{j_3} + c_{j_1} \geq n$ and, consequently, $c_{j_1} + c_{j_2} + c_{j_3} \geq 3n/2$. Each of the remaining columns contains at least 2 free cells. Therefore $E \geq 3n/2 + 2(n - 3) = 7n/2 - 6$. ■

For the rest of Part 3.2, we will assume that 5. of Proposition 9 is true since this is the only case when our lower bound on E remains to be proved. We will denote by i_0 the index of the row that contains a free cell in each 2-column and by j_0 the index of the column that contains a free cell in each 2-row. We will denote by a the symbol whose existence is guaranteed by the following Lemma 12.

Lemma 12 *There exists a symbol* $a \in \bigcap_{i \in \mathbf{R}} RS(i) \cap \bigcap_{j \in \mathbf{C}} CS(i)$.

Proof. Since there are at least 5 2-columns, $RS(i_0)$ contains at least 5 elements. For a similar reason $CS(j_0)$ contains at least 5 elements as well. Lemma 8 implies existence of a symbol $a \in \bigcap_{i \in \mathbf{R}} R(i) \cap \bigcap_{j \in \mathbf{C}} C(i)$. ■

Lemma 13 *There is at most one pair of 2-rows having the free cells in two common columns.*

Proof. Let there be two such pairs of 2-rows. Both pairs have their empty cells in the column j_0. Either four or three occurrences of the symbol a are missing in the two pairs of rows depending on whether the rows in the pairs are, or are not, pairwise different. Consider the completion of L after discarding a symbol outside the two pairs of rows. In this completion, four occurrences of the symbol a must be placed in at most three different columns in the former case, while three occurrences must be placed in only two different columns in the latter case. ■

Lemma 14
(a) If there are two different 2-rows where the same pair of symbols is missing then $E \geq 7n/2 - 8$.
(b) If two 2-rows have their free cells in two common columns then $E \geq 7n/2 - 8$.

Proof. (a) By conjugacy, Lemma 13 implies that there is at most one pair of 2-rows where the same pair of symbols is missing. Let $i_1, i_2 \in \mathbf{R}$ be the only two different rows with equal pair of missing symbols. There exist $|\mathbf{R}| - 2$ symbols different from a, each missing in one 2-row different from i_1, i_2. Lemma 13 implies the existence of at least $(|\mathbf{R}|-2)/2$ columns different from j_0 having free cells in the rows from $\mathbf{R} - \{i_1, i_2\}$. If none of these columns contains any of the $|\mathbf{R}| - 2$ symbols (we know that no one contains a) then each of these columns contains at least $|\mathbf{R}| - 1$ free cells. Moreover, the column j_0 contains at least $|\mathbf{R}|$ free cells and there are $n-(|\mathbf{R}|-2)/2-1 = n-|\mathbf{R}|/2$ additional columns containing at least 2 free cells each. Therefore (since $|\mathbf{R}| \geq n/2+1$) we get $E \geq (|\mathbf{R}|-1)(|\mathbf{R}|-2)/2+|\mathbf{R}|+2(n-|\mathbf{R}|/2) \geq n(n+12)/8 \geq 7n/2 - 8$. Let, on the other hand, some column j_1 contain a free cell in a row $i_3 \in \mathbf{R} - \{i_1, i_2\}$ and at the same time a symbol b, which is missing in a row $i_4 \in \mathbf{R} - \{i_1, i_2\}$. The symbol b cannot be missing in row i_3. Consider the Latin square C being the completion of L after b has been removed from row i_3. Then $C(i_3, j_0) = b$, otherwise b would be in position (i_3, j_1) and column j_1 already contains b. Consequently, $C(i_4, j_0) = a$. We obtain a contradiction, since one of the values $C(i_1, j_0), C(i_2, j_0)$ must be a.

(b) The assertion is obtained from (a) by conjugacy when symbols are replaced by columns. ■

Let us now adopt the last two assumptions valid till the end of the current Part 3.2 (we assume so far that $|\mathbf{R}| \geq n/2+1 \geq 5$, $|\mathbf{C}| \geq n/2+1 \geq 5$ and *5.* of Proposition 9 is true). We will further assume that in no two different 2-rows the same pair of symbols is missing and that no two 2-rows have their free cells in the same pair of columns.

Denote by x the number of 2-columns sharing a free cell with some 2-row. Each such column has only one of its free cells in some 2-row, the other one is in row i_0, hence the number of 2-rows sharing a free cell with some 2-column is x as well. The free cells of all 2-rows are placed in at least $n/2+2$ columns (including the column j_0). Since there are at least $n/2+1$ different 2-columns, $x \geq 3$. Let $i_1, \ldots, i_x$ be the indices of pairwise different 2-rows and $j_1, \ldots, j_x$ the indices of pairwise different 2-columns such that, for $r = 1, \ldots, x$, $L(i_r, j_r) = \varepsilon$. Let $RS(i_r) = \{a, b_r\}$, $CS(j_r) = \{a, c_r\}$, hence the symbols $b_1, \ldots, b_x$ and $c_1, \ldots, c_x$ are pairwise distinct.

Lemma 15 *For* $r = 1, \ldots, x$, $b_r = c_r$.

Proof. We will use the fact that $x \geq 3$. If $b_1 \neq c_1, b_1 \neq c_2, b_1 \neq c_3$ then $C_{i_1,j_2}(i_1, j_0) = C_{i_1,j_3}(i_1, j_0) = b_1$ and $L(i_1, j_2) = C_{i_1,j_2}(i_1, j_1) = c_1 = C_{i_1,j_3}(i_1, j_1) = L(i_1, j_3)$ — a contradiction. Therefore $b_1 \in \{c_1, c_2, c_3\}$. Analogously, $b_2 \in \{c_1, c_2, c_3\}$ and $b_3 \in \{c_1, c_2, c_3\}$, therefore $\{b_1, b_2, b_3\} = \{c_1, c_2, c_3\}$. Assume $b_1 = c_2$ (the case $b_1 = c_3$ leads to an analogous contradiction). Let the row i_3 contain b_1 in the position

(i_3, j'). Then $C_{i_3,j'}(i_3, j_0) = b_1$, therefore $C_{i_2,j'}(i_1, j_0) = a$, therefore $C_{i_2,j'}(i_1, j_1) = b_1$ yielding a contradiction, since b_1 is not missing in the column j_1. The other equalities follow in a similar way. ■

Lemma 16 *The subsquare consisting of rows $i_1, \ldots, i_x$ and columns $j_1, \ldots, j_x$ does not contain any of the symbols $b_1, \ldots, b_x$.*

Proof. Without loss of generality, let $L(i_1, j_2) = b_3$. Consider the Latin square $C = C_{i_1,j_2}$. Since $a \neq b_3 \neq b_1$, the only possibility is $C(i_1, j_0) = b_3$, therefore $C(i_3, j_0) = a$, therefore $C(i_3, j_3) = b_3$, therefore $C(i_0, j_3) = a$, therefore $C(i_0, j_1) = b_1$, therefore $C(i_1, j_1) = a$, therefore $C(i_1, j_2) = b_1$, yielding a contradiction, since b_1 is not missing in the column j_2. ■

Lemma 17 *If some row contains some of the symbols $b_1, \ldots, b_x$ in column j_s, $1 \leq s \leq x$, then the symbol b_s is missing in this row.*

Proof. Without loss of generality, let $L(i, j_1) = b_2$, $i \notin \mathbf{R}$. Consider the Latin square $C = C_{i,j_1}$. Then $C(i_0, j_1) = b_2$, therefore $C(i_0, j_2) = a$, therefore $C(i_2, j_2) = b_2$, therefore $C(i_2, j_0) = a$, therefore $C(i_1, j_0) = b_1$, therefore $C(i_1, j_1) = a$, therefore $C(i, j_1) = b_1$, therefore b_1 is missing in row i. ■

Corollary 18 *No row from $\mathbf{R}$ contains in any of the columns $j_1, \ldots, j_x$ any of the symbols $b_1, \ldots, b_x$. No column from $\mathbf{C}$ contains in any of the rows $i_1, \ldots, i_x$ any of the symbols $b_1, \ldots, b_x$.*

Corollary 19 *Each of the symbols $b_1, \ldots, b_x$ is missing in at least $x - 1$ of the rows not belonging to $\mathbf{R}$ and in at least $x - 1$ of the columns not belonging to $\mathbf{C}$.*

Proof. Corollary 18 implies that symbols $b_1, \ldots, b_x$ appear in columns from $\mathbf{C}$ outside rows from $\mathbf{R}$. The assertion follows from Lemma 17. ■

Lemma 20 $\min(|\mathbf{R}|, |\mathbf{C}|) \leq (n + x - 1)/2 < 3n/4$.

Proof. There are $|\mathbf{R}| - x$ different 2-rows not having an empty cell in any of the columns from $\mathbf{C}$. Their empty cells must occur (besides j_0) in $|\mathbf{R}| - x$ additional columns, since no two 2-rows have their empty cells in the same two columns. Hence $|\mathbf{R}| - x \leq n - |\mathbf{C}| - 1$, and, consequently, $\min(|\mathbf{R}|, |\mathbf{C}|) \leq (|\mathbf{R}| + |\mathbf{C}|)/2 \leq (n + x - 1)/2$. ■

Lemma 21 $E \geq 7n/2 - o(n)$.

Proof. Let e.g. $|\mathbf{C}| = \min(|\mathbf{R}|, |\mathbf{C}|)$. Lemma 4 applied to row j_1 implies that column j_0 contains at least $n - 2$ free cells. The 2-columns contain 2 free cells each, while all the remaining columns contain at least 3 free cells each. Using Lemma 20 we obtain $E \geq n - 2 + 2|\mathbf{C}| + 3(n - |\mathbf{C}| - 1) = 4n - |\mathbf{C}| - 5 \geq 7n/2 - x/2 - 9/2$. On the other hand we may use Corollary 19 for another estimation yielding $E \geq n - 2 + 2|\mathbf{C}| + x(x - 1) \geq n - 2 + 2x + x(x - 1) = n - 2 + x^2 + x$. Hence $E \geq \min_{3 \leq x \leq n/2} \max(7n/2 - x/2 - 9/2, n - 2 + x^2 + x) = 7n/2 - (\sqrt{40n - 31} + 33)/8 = 7n/2 - o(n)$. ■

Acknowledgment

The author would like to thank Peter Horák for introduction to the problem and for stimulating discussions.

References

[1] L. Branković, P. Horák, M. Miller and A. Rosa, *Premature partial latin squares*, Ars Combinatoria, to appear.

[2] L. Branković and M. Miller, *The spectrum of premature partial latin squares*, Proceedings of the Tenth Australasian Workshop on Combinatorial Algorithms (AWOCA'99), Perth, Australia, July 25-27, 1999, 168–175.

[3] A.D. Keedwell, *Critical sets for latin squares, graphs and block designs: a survey*, Congressus Numerantium **113**(1996), 231–245.

(Received 19/10/99)