Transversals in Rectangles

G. H. J. van Rees*
Department of Computer Science, University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2
vanrees@cs.umanitoba.ca

Abstract

Let $L(m,n)$ be the largest integer such that, if each symbol in an $m \times n$ rectangle occurs at most $L(m,n)$ times, then the array must have a transversal. We improve the lower bound to $L(m,n) \geq \lfloor \frac{m(n-m+1)-1}{m-1} \rfloor$, for $m > 1$. Then we show that sporadically $L(m,n) < \lfloor \frac{mn-1}{m-1} \rfloor$ in the range $m \leq n \leq m^2 - 3m + 3$. Define $n_0(m)$ to be the smallest integer z such that if $n \geq z$ then $L(m,n) = \lfloor \frac{mn-1}{m-1} \rfloor$. We improve $n_0(m)$ from $O(m^3)$ to $O(m^{2.5})$. Finally, we determine $L(4,n)$ for all n.

1 Introduction

For positive integers m and n, where $2 \leq m \leq n$, an $m \times n$ rectangle consists of mn cells arranged in m rows and n columns in which each cell contains one symbol from some set, usually the integers from 1 to s. Let $L(m,n)$ be the largest integer such that if each symbol in an $m \times n$ rectangle occurs at most $L(m,n)$ times, then the array must have a transversal. Equivalently, $L(m,n)+1$ is the smallest integer such that there exists an $m \times n$ rectangle which has no transversal and the symbol that occurs most in the rectangle occurs $L(m,n)+1$ times. A *section* is a subset of the rectangle with one cell from each row and at most one cell from each column. A *transversal* is a section in which all the symbols are distinct. A *partial transversal* is a subset of the rectangle with at most one cell from each row and column in which the symbols are all distinct.

Stein and Szabó [4] gave the following prominent conjecture.

*research supported by NSERC grant 3558-02

Conjecture 1 $L(n-1,n) = n-1$

If the above conjecture were true then this old conjecture by Brualdi as cited by Dénes and Keedwell [3] about latin squares would also be true.

Conjecture 2 *In every latin square of order n there is a partial latin transversal of length $n-1$.*

See Stein and Szabó [4] for more background and examples.

Clearly, $L(m+1,n) \leq L(m,n) \leq L(m,n+1)$. We often say that an *array has maximum frequency* as a short hand for *the symbols in the array have maximum frequency.* We now give a short history on this problem.

In Alon [2] the following is proved using the Lovász Local Lemma.

Theorem 1.1 *If no symbol appears in more than $(n-1)/(4e)$ cells of an $n \times n$ array, then there is a transversal.*

It is straightforward to generalize this to the following theorem. For a proof, see [5].

Theorem 1.2 *If no symbol appears in more than $\frac{n-1}{2e(1+m/n)}$ cells of an $m \times n$ array, A, then there is a transversal in A.*

Stein and Szabó [4] proved the following upper bound.

Theorem 1.3 $L(m,n) < mn/(m-1)$.

We define $B(m,n)$ to be $\lfloor \frac{mn-1}{m-1} \rfloor$ and define $n_0(m)$ to be the smallest integer, z, such that if $n \geq z$ then $L(m,n) = B(m,n)$. Akbari et al. [1] proved the following theorem.

Theorem 1.4 *For $m \geq 2$, $n_0(m) \leq 2m^3 - 6m^2 + 6m - 1$.*

Stein and Szabó [4] proved the following two theorems.

Theorem 1.5 *If $n \leq 2m-2$, then $L(m,n) \leq n-1$.*

Theorem 1.6 $L(m,n) \geq n-m+1$.

In the next section, we improve the lower bound for $L(m,n)$ and prove a better bound for $n_0(m)$. Finally, we determine $n_0(4)$.

2 Results

Theorem 1.1 gives $L(n,n) \geq 2$ for $n \geq 23$ and $L(n,n) \geq 3$ for $n \geq 34$. We improve this with Theorem 2.1 which presents the best results known for square arrays.

Theorem 2.1 *$L(n,n) \geq 2$ for all $n \geq 3$. $L(n,n) \geq 3$ for $n = 4,5,6,7,8$.*

Proof In a square of order n, there are $n!$ sections. Let the maximum frequency in an $m \times n$ array be 2. There are essentially only 2 different ways a symbol with frequency 2 can appear in a square array. If we want to maximize the number of sections that contain 2 of the same symbol, then the symbol should occur exactly 2 times in cells that occupy distinct rows and distinct columns. Such a symbol will occur in $(n-2)!$ sections which can not be transversals. Since there are at most $\lfloor n^2/2 \rfloor$ symbols of frequency 2, there are at most $(\lfloor n^2/2 \rfloor)(n-2)!$ sections that are not transversals. If $n! > (\lfloor n^2/2 \rfloor)(n-2)!$ i.e. $n \geq 3$, then there is a section leftover that is a transversal.

Let us now consider an array whose symbols have maximum frequency 3. There are essentially 4 different ways a symbol with frequency 3 can appear in the array. If we want to maximize the number of sections that contain 2 or 3 of the same symbol, then the symbol should occur exactly 3 times in cells that occupy distinct rows and distinct columns. A pair of that symbol will occur in $(n-2)!$ sections which can not be transversals. Of those $(n-2)!$ sections $(n-3)!$ will be common to all 3 pairs of the trio of symbols. So there will be $3((n-2)! - (n-3)!) + (n-3)! = (n-3)!(3n-8)$ distinct sections that will not be transversals. Since there are at most $\lfloor n^2/3 \rfloor$ symbols of frequency 3, there are at most $(\lfloor n^2/3 \rfloor)(n-3)!(3n-8)$ sections that are not transversals. If $n! > (\lfloor n^2/3 \rfloor)(n-3)!(3n-8)$ i.e. $n(n-1)(n-2) > (3n-8)(\lfloor n^2/3 \rfloor)$, then there is a section leftover that is a transversal. Note that a symbol of frequency 1 and a symbol of frequency 3 can be replaced by 2 symbols of frequency 2 but this does not increase the number of sections that can not be transversals. If $n = 4$, there is another arrangement of 3 identical symbols in a 4×4 array that causes 4 sections not to be transversals but since the number of sections affected is the same as before, the argument does not change. This proves the second part of the theorem except for $n = 6$ or $n = 8$ where the inequality becomes an equality.

When $n = 6$, $n! = (\lfloor n^2/3 \rfloor)(n-3)!(3n-8)$ and all symbols must occur exactly 3 times in three different rows and three different columns. Then every section contains 5 symbols, 4 symbols that are there exactly once and one symbol that occurs exactly twice. Without loss of generality, suppose symbol 1 occurs in cells $(1,1)$, $(2,2)$ and $(3,3)$. Let 2 be in cell $(4,4)$ of the array. Then symbol 2 can not occur again in row 4 or column 4. Further, symbol 2 can not occur in the bottom right 2 $\times 2$ of the array or there will be a section containing two 2's and three 1's which is a contradiction. So the symbol in $(4,4)$ can not occur again in the bottom right 3×3 of the array. This is true for any symbol in the bottom right 3×3 array. So all symbols in the bottom right 3×3 must be distinct. So if $(4,4)$ contains the symbol 2 and symbol 2 can not again be in row 4 or in column 4 or in the bottom right 3×3 corner of the array, then where are the other 2's? If they are in the top right 3×2 corner of the array then there is a section containing two 2's and two 1's which is a contradiction. Similarly, the symbol 2 can not occur in the bottom left 2×3 corner. So the symbol 2 must occur twice more in the top left 3×3 corner. This is true for all nine distinct symbols that are in the bottom right 3×3 corner. This puts 18 symbols in the top left 3×3 corner. This is a contradiction. The case for $n = 8$ is similar. □

Corollary 2.1.1 *$L(4,4) = 3$ and $L(5,5) = 3$.*

Proof Apply Theorem 2.1 and Theorem 1.5 with $n = 4$ to get the first result. Theorem 2.1 implies that $L(5,5) \geq 3$ and note that the 5×5 example from [4] implies that $L(5,5) \leq 3$. □

The key tool that Stein [4] and Akbari et al. [1] used to get their results was establishing the existence of near transversals. We call the $m-1$ positions in $m-1$ distinct rows and $m-1$ distinct columns containing $m-1$ distinct symbols, say $x_1, x_2, \ldots, x_{m-1}$ in an $m \times n$ array a *near transversal.* Since relabelling rows and/or columns and/or symbols does not affect the existence or non-existence of transversals, we will show that the near transversals occur in columns 1 to $2(m-2)$ and contain the symbols $1, 2, \ldots, (m-1)$. In fact, the near transversals are actually in some set of $2(m-1)$ columns scattered somewhere in A with symbols from some set of cardinality $m-1$. With this understanding we proceed.

Exploiting these near transversals in an array allows us to extend Theorem 1.6 to the following theorem. The proof uses the ideas in Akbari et al. [1] extensively.

Theorem 2.2 *$L(m,n) \geq \lfloor \frac{mn-1}{m-1} \rfloor - m$, $m > 1$.*

Proof The proof will be by induction on the number of rows in the array. The hypothesis is true for $m = 2$ and $m = 3$. We will assume that $L(m -$

1					x	x	...
	2						
		$\ddots$					
			$m-1$				
				1	x	x	...

Figure 1:

$1,n) \geq \lfloor\frac{(m-1)(n-m+2)-1}{m-2}\rfloor$ and prove that $L(m,n) \geq \lfloor\frac{m(n-m+1)-1}{m-1}\rfloor = \lfloor\frac{mn-1}{m-1}\rfloor - m$. Consider an $m \times n$ array, A, with maximum frequency $\lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$. We will assume that there is no transversal in A and that will lead to a contradiction.

Consider the first $m-1$ rows of the array. By mathematical induction we know that there is a transversal in these $m-1$ rows. Without loss of generality, we have $A(1,1) = 1, A(2,2) = 2, \ldots, A(m-1,m-1) = m-1$, These cells form a near transversal in A which we call t'. If $A(m,m)$ is adjoined with its symbol to t', we get a section, say t in A. If t is a transversal, we get a contradiction and we are done. So t is not a transversal and so the symbol in $A(m,m)$ must appear at least twice in t. Without loss of generality, let this symbol be 1. Because $A(1,1), A(2,2), \ldots, A(m-1,m-1)$ is a near transversal, $S = \{A(m,j), \text{ for } j > m\}$ must contain only elements from $1, 2, \ldots, m-1$ or otherwise we get the contradiction that there exists a transversal. Similarly, because the cells $A(2,2), A(3,3), \ldots, A(m,m)$ is also a near transversal, say t'', $A(1,j)$, for $j \geq m$ must also contain only elements from $1, 2, \ldots, m-1$. The situation is as shown in Figure 1. Let x represent a symbol that is a 1, 2, ... or $(m-1)$.

There are $2(n-m+1)$ symbols in the array of Figure 1 that are 1's or x's. But the symbol 1 can only appear with frequency $\lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$ which is the maximum frequency. Then, since $2(n-m+1) > \lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$, the x's can not all be 1's. So without loss of generality, let $A(1,m+1)$ contain a 2. Now the positions $A(1,m+1), A(3,3), A(4,4), \ldots, A(m,m)$ contain distinct symbols and form a near transversal, say t'''. For the same reasons as before, there must be x's in the range 1 to $m-1$ in the second row from columns $m+1$ to n. So there are at least $3(n-m+1)$ symbols that are 1, 2 or x. But symbols 1 and 2 can appear at most 2 times, the maximum frequency which is $2\lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$. Then, since $3(n-m+1) > 2\lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$, not all the x's can be 1's or 2's. So there is a symbol, larger than 1 or 2, say 3, that is in rows 1, 2 or m and in the columns 1,2, or $m+2$ of A. But there are several non-isomorphic possibilities of where this 3 could be and also we don't know if there is a near transversal that will force more x's into A. So we need to show that these near transversals exist very carefully.

We are going to prove that there are $m-1$ near transversals, $T_1 = t'', T_2 = t''', T_3, \ldots, T_{m-1}$ each consisting of $m-1$ cells, such that for every $1 \leq i \leq m-1$, the conditions in the following table hold:

Table 1

1: $A(i,i) = i$.
2: Each row, except row i, has a cell in T_i.
3: For every j, $i < j < m$, cell (j,j) is included in T_i
4: The symbols in T_i are $1, 2, \ldots, m-1$.
5: The cells in T_i lie in the first $m+i-1$ columns. Clearly, T_1 and T_2 satisfy Conditions 1-5.

Let t, t', t'' and S be as before. Assume that S has k distinct symbols in it from 1 to k. Without loss of generality let $A(m,m)$ contain 1, $A(m,m+1)$ contain 2,..., $A(m,m+k-1)$ contain k. The other cells $A(m,j)$, $j \geq m+k$ contain x's.

For $1 \leq i \leq k$, define T_i to be $\{\{(j,j)|1 \leq j \leq m-1, j \neq i\} \cup \{m, m+i-1\}\}$. Clearly, T_1, T_2, ..., T_k satisfy Conditions 1-5.

Suppose that $T_1, T_2, \ldots T_p$ for some $p \geq k$ have been constructed. The construction for T_{p+1}, for $p \leq m-2$ will now be shown. Let $X_i = \{(i,j)|T_i$ has no cell in column $j\}$. $X_1, X_2, \ldots, X_p$ are disjoint. Let X be their union. All the symbols in X and S are less than or equal to p. Since the cells in X are pairwise disjoint, we have $|X \cup S| = (p+1)(n-m+1)$. The symbols in $X \cup S$ are either 1, 2, ..., p, or x.

Since the symbols $1, 2, \ldots, p$ can appear at most $p\lfloor \frac{m(n-m+1)-1}{m-1} \rfloor$ times and since $(p+1)(n-m+1) > p\lfloor \frac{m(n-m+1)-1}{m-1} \rfloor$, then there is a cell $(r,s) \in X \cup S$ such that $A(r,s) = t > p$. Since T_r has no cells in row r and since T_r has no cell from column s, if $t \geq m$ then $T_r \cup \{(r,s)\}$ is a transversal. So $t \leq m-1$. We may assume that $A(r,s) = p+1$ because symbols t and $p+1$, rows t and $p+1$, and columns t and $p+1$ can be interchanged. These interchanges have the effect of interchanging cells $A(t,t)$ and $A(p+1, p+1)$ along with their contents. Since $T_{p+1}, T_{p+2}, \ldots, T_{m-1}$ all have entries $A(i,i) = i$ for $p \leq i \leq m$, Conditions 1-5 are preserved. Since $A(r,s) = p+1$, $T_{r+1} = T_r \cup \{r,s)\} \backslash \{(p+1, p+1)\}$ is a near transversal that does not violate Conditions 1-5. Condition 5 is not violated because at most one new column is used when T_{p+1} is added to $T_1, T_2, \ldots, T_p$. In this way, $T_1, T_2, \ldots, T_{m-1}$ can be constructed.

Let us now count the 1's, 2's, ..., $(m-1)'s$ and x's in all the rows of the array. There are $m(n-m+1)$ of them. The maximum number of times that the symbols 1,2,..., $m-1$ can occur in the array is $(m-1)\lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$ times. Since $m(n-m+1) > (m-1)\lfloor\frac{m(n-m+1)-1}{m-1}\rfloor$, we have at least one x that is not 1, 2, ..., or $m-1$. This is a contradiction. □

Note that the difference between the upper and lower bounds is now at most m.

We now tackle upper bounds. The next result shows that $L(m,n) < B(m,n)$, at least sporadically, in the range $3 \le m \le n \le m^2 - 3m + 3$.

Theorem 2.3 *Let $m \ge 3$. Then $L(m,n) < B(m,n)$ for $n = m + a(m-1) + b$ where $a = 0,1,\dots,m-3$ and $b = 0,1,\dots,m-3-a$.*

Proof Let a and b be as in the statement of the theorem and consider an $m \times (m + a(m-1) + b)$ rectangle with m symbols in it in which $m-1$ of the symbols occur $\lfloor\frac{mn-1}{m-1}\rfloor = m+1+b+am$ times in the rectangle and one symbol occurs $b+1$ times in the rectangle. Note that these m symbols fill the array. Call the first $m-1$ symbols $1,2,\dots,m-1$ and call the last one m. Place m in positions $(1,1),(2,2),\dots(b+1,b+1)$. Place symbol i, $i = 1,2,3,\dots b+1$ throughout row i except for position (i,i). Also place symbol i in the last $a+2$ rows of column i. Note that $a+2+b+1 \le m$ so that we do not place two symbols in the same cell. Place the rest of the symbols anywhere. Since there are only m symbols in the $m \times n$ rectangle, one of the symbols m must be in the transversal, say the one in position (i,i). But then the i^{th} symbol can not be in the transversal and the transversal can not contain m distinct symbols, i.e. this rectangle contains no transversal.
□

The construction in Theorem 2.3 for $n \le 2m-2$ is superceded by the construction in Theorem 1.5. The largest n for which the construction in Theorem 2.3 works is $n = m^2 - 3m + 3$. This leaves a large gap to the Akbari et al. [1] result which proved that $L(m,n) = \lceil\frac{mn-1}{m-1}\rceil$ for $n \ge 2m^3 - 6m^2 + 6m - 1$. Before we reduce this gap, we prove a useful lemma.

It is very annoying that when one is trying to prove that all $m \times n$ arrays with fixed maximum frequency have a transversal, there are so many distributions of the frequencies of the symbols in the array to be considered. The next lemma reduces the number to be considered considerably. Let f_i be the number of times the symbol i appears in an array.

Lemma 2.4 *If every $m \times n$ array whose symbols have a maximum frequency f and whose symbols have a frequency distribution in which $f_i + f_j > f$ for any two symbols i and j, has a transversal then any $m \times n$ array whose symbols have a maximum frequency f must have a transversal.*

Proof Consider an array, A, in which it is not true that $f_i + f_j > f$ for any two symbols i and j. Form a new array, A' as follows: Replace any two symbols i and j with the property $f_i + f_j \leq f$ with a new symbol i'. Repeat this until we get an A' that does have the sum of the frequencies of any two symbols to be greater than f. We know, from the hypothesis, that A' has a transversal. If we now consider those positions in A', it is easy to see that they also must be a transversal in A. □

Again we exploit the use of near transversals, to improve the bound on $n_0(m)$. The proof of the following theorem will closely match the proof of Theorem 2.2. We find a set of near transversals as we did before but we use different inequalities to get our result.

Theorem 2.5 *For $m \geq 2$, $n_0(m) \leq (m-1)(2+m\sqrt{m+3})/2$.*

Proof We will prove this by induction on m. It is true for $m = 2, 3$. We assume the hypothesis is true for $m-1$. To prove the theorem, we assume that there exists an $m \times n$ array where $n \geq (m-1)(2+m\sqrt{m+3})/2$ and where the elements of the array have maximum frequency of $B(m,n) = B = \lfloor \frac{mn-1}{m-1} \rfloor$ that does not have a transversal. Consider an $m \times n$ array, A with maximum symbol frequency, $B = \lfloor \frac{mn-1}{m-1} \rfloor$. Then there must be at least m distinct symbols. Consider the symbol with the lowest frequency in the array A. Let it occur in position (m, m). As in the proof of Theorem 2.2, consider the array that is formed when row m and column m are deleted from A. We get t, t', t'' and S exactly as before as depicted in Figure 1. Again x represents a symbol that is a 1,2,..., or $(m-1)$.

There are $2(n-m+1)$ 1's and x's in rows 1 and m. Recall that 1 is the symbol that occurs the fewest times. There is another symbol that appears B times. So symbol 1 occurs at most $(mn-B)/(m-1)$ times. Since there are at least $2(n-m+1)$ 1's and x's in rows 1 and m and at most $(mn-B)/(m-1)$ 1's in A, if $2(n-m+1) > (mn-B)/(m-1)$, all the x's in rows 1 and m could not be 1's. But is this inequality true? The answer will come shortly, so let us just take it as true. So there is a symbol that is not 1 in either row 1 or row m in some column c where $c > m$. Also let $T_1 = t''$.

Just as in the proof of Theorem 2.2, we will need to prove that there are

$m-1$ near transversals, $T_1, T_2, \ldots T_{m-1}$ each consisting of $m-1$ cells, such that for every $1 \le i \le m-1$, the conditions in Table 1 hold. Assume that S has k distinct symbols in it from 1 to k. Without loss of generality let $A(m,m)$ contain 1, let $A(m,m+1)$ contain 2,..., and let $A(m,m+k-1)$ contain k. The other cells $A(m,j), j \ge m+k$ contain x's. For $1 \le i \le k$, define T_i to be $\{(j,j) | 1 \le j \le m-1, j \ne i\} \cup \{m, m+i-1)\}$. Clearly, $T_1 = t'', T_2 = t''', \ldots, T_k$ satisfy Conditions 1-5.

Suppose that $T_1, T_2, \ldots, T_p$ for some $p \ge k$ has been constructed. The construction for T_{p+1}, for $p \le m-2$ will now be shown. Let $X_i = \{(i,j) | T_i$ has no cell in column $j\}$. $X_1, X_2, \ldots, X_p$ are disjoint. Let X be their union. The symbols in $X \cup S$ are $1, 2, \ldots, p$. If we don't know which symbol it is, we use the symbol x as a place holder. Since the cells in X are pairwise disjoint, we have $|X \cup S| = (p+1)(n-m+1)$.

But how many times can the symbols from 1 to p occur in A? An upper bound would be pB but symbol 1 can not occur the maximum number of times permitted as it is the least frequent symbol. So if symbols 2 to p occur the maximum number of times allowed, they would account for $(p-1)B$ occurrences. Then there are $mn-(p-1)B$ places in A for the other elements in A. There are at least $m-p$ other elements and 1's frequency (the least frequent in A) is optimized if its frequency is the same (or almost the same) as all the others; i.e., $\lfloor \frac{mn-yB}{m-y} \rfloor$. So if $(p+1)(n+m-1) > (p-1)B + \lfloor \frac{mn-yB}{m-y} \rfloor$, then there is a cell $(r,s) \in X \cup S$ such that $A(r,s) = t > p$. Again, we delay proving the inequality and delay giving the definition of y.

Since T_r has no cells in row r and since T_r has no cell from column s, if $t \ge m$ then $T_r \cup \{(r,s)\}$ is a transversal. So $t \le m-1$. We may assume that $A(r,s) = p+1$. Since $T_{p+1}, T_{p+2}, \ldots, T_{m-1}$ all have entries $A(i,i) = i$ for $p \le i \le m$, Conditions 1-5 are preserved. Since $A(r,s) = p+1$, $T_{r+1} = T_r \cup \{(r,s)\} \backslash \{(p+1, p+1)\}$ is a near transversal that does not violate Conditions 1-4. Condition 5 is not violated because at most one new column is used when T_{r+1} is added to $T_1, T_2, \ldots, T_r$.

In this way, $T_1, T_2, \ldots$ and T_{m-1} can be constructed assuming that the inequality in each case is true.

Let us now count the 1's, 2's, ..., $(m-1)'s$ and x's in all the rows of the array. There are $m(n-m+1)$ of them. The maximum number of times that the symbols $1, 2, \ldots, m-1$ can occur is if symbol 1 (the least frequent symbol) occurs $(mn-(m-2)B)/2$ times and symbols 2 to m occur B times each. If $m(n-m+1) > (m-2)B + (mn-(m-2)B)/2$ which is true for $n > 2(m-1)^2$ which is true by the hypothesis, we have at least one x that is not 1, 2, ..., or $m-1$. Contradiction of the definition of x.

We now show that the inequalities are satisfied under the conditions of the theorem. Depending on k we may have as many as $m-2$ inequalities. The first inequality is slightly different than the others. The first inequality, $2(n-m+1) > (mn-B)/(m-1)$ will be true if the second inequality $3(n-m+1) > B+(mn-B)/(m-1)$ is true. In general the y^{th} inequality for $y = 2, 3, \ldots, m-1$ is $(y+2)(n-m+1) > yB + (mn - yB)/(m-y)$. To show that inequalities $2, 3, \ldots, m-2$ are true, consider the y^{th} inequality for $2 \leq y \leq m-1$ where B has been substituted by its formula.

We have

$$(y+2)(n-m+1) > y\lfloor\frac{mn-1}{m-1}\rfloor + \lfloor\frac{mn - y\lfloor\frac{mn-1}{m-1}\rfloor}{m-y}\rfloor$$

This will be true if:

$$(y+2)(n-m+1) > y\lfloor\frac{mn-1}{m-1}\rfloor + \frac{(mn - y\lfloor\frac{mn-1}{m-1}\rfloor)}{m-y}$$

or

$$(y+2)(n-m+1) > \frac{y(m-y-1)\lfloor\frac{mn-1}{m-1}\rfloor}{m-y} + \frac{mn}{m-y}.$$

This will be true if:

$$(y+2)(n-m+1) > \frac{y(m-y-1)\frac{mn-1}{m-1}}{m-y} + \frac{mn}{m-y}.$$

The above implies that:

$$n > \frac{(m-1)^2(m-y)(y+2) - y(m-y-1)}{y^2 + (2-2m)y + (m^2-m)}.$$

This will be true if:

$$n > \frac{(m-1)^2(m-y)(y+2)}{y^2 + (2-2m)y + (m^2-m)} = \frac{(m-1)^2(-y^2 + (m-2)y + 2m)}{y^2 + (2-2m)y + (m^2-m)}.$$

Consider the right hand side of the inequality as a function in y; i.e., $f(y)$. Clearly, this is a quadratic in y over a quadratic in y. To find the maximum value of this function over the appropriate range of y we find the critical points. Since the discriminant of the denominator of $f(y)$ is $(2-2m)^2 - 4(1)(m^2-m) = 4-4m < 0$ for $m > 2$, the denominator has complex roots. Then to find the critical points of $f(y)$ we set the numerator of the derivative , f', to 0 and get

$$(-(2-2m)-m+2)y^2 + 2(-m^2+m-2m)y + (m^2-m)(m-2) - 2m(2-2m) = 0$$

which simplifies to

$$y^2 - 2(m+1)y + (m^2 + m - 2) = 0.$$

Solving we get

$$y = (m+1) \pm \sqrt{m+3}.$$

The positive root is never in the range of y so we only use the negative root. Substituting the negative root into the function we get that $f(y)$ has maximum value $(m-1)(2+m\sqrt{m+3})/2$. So if $n > (m-1)(2+m\sqrt{m+3})/2$ then all the inequalities will be true and we get a contradiction that y will be less than p and bigger than p at the same time. □

We give a table comparing the value $2m^3 - 6m^2 + 6m - 1$, $\lceil (m-1)(2+m\sqrt{m+3})/2 \rceil$, and the first value of n, for which all y inequalities are true and stay true for larger n. This number is an upper bound for $n_0(m)$.

n	$2m^3 - 6m^2 + 6m - 1$	$\lceil (m-1)(2+m\sqrt{m+3})/2 \rceil$	$n_0(m) \leq$
3	17	10	8
4	55	19	18
5	129	33	31
6	251	50	49
7	433	73	70
8	687	100	96
9	1025	133	125
10	1459	172	166
11	2001	216	213
12	2663	267	266

There is still a gap in the leading term, n^2 to $n^{2.5}$, from where the largest known value of n where $L(m,n) \neq B(m,n)$ to the smallest value of n_0 where we know that $L(m,n) = B(m,n)$ for all $n > n_0$.

We now show that the asymptotic result is true for even smaller n when $m = 4$. Even though we later prove the case $n = 11$, it is included here for those people who dislike computer proofs.

Theorem 2.6 $L(4,n) = \lfloor \frac{4n-1}{3} \rfloor$ *for* $n \geq 11$.

Proof Consider a $4 \times n$ array A with maximum frequency $B(m,n)$ where $n \geq 11$. Since there must be at least one symbol of frequency $B(4,n) = \lfloor \frac{4n-1}{3} \rfloor$, then a least frequent symbol occurs at most $\lfloor (4n - \lfloor \frac{4n-1}{3} \rfloor)/3 \rfloor$

1				2	x	...
	1			x	x	...
		2				
			3			

Figure 2:

times. In order to get a contradiction, we assume that there exists an array which has no transversal and whose elements have maximum frequency of $\lfloor \frac{4n-1}{3} \rfloor$.

The next part of the proof follows the proof of Theorem 2.5 very closely. Let symbol 1 be one of the least frequent symbols in A. So symbol 1 occurs at most $\lfloor (4n - \lfloor \frac{4n-1}{3} \rfloor)/3 \rfloor$ times. Let the symbol be 1 and, without loss of generality, let 1 be in $A(1,1)$. Then consider A with row 1 and column 1 deleted. This new array is $3 \times (n-1)$ and has maximum frequency $\lfloor \frac{4n-1}{3} \rfloor$. Since $L(3, n-1) = \lfloor \frac{3n-4}{2} \rfloor$ for $n - 1 > 4$, and since $\lfloor \frac{4n-1}{3} \rfloor \leq \lfloor \frac{3n-4}{2} \rfloor$ for $n \geq 11$, we know that the $3 \times (n-1)$ array has a transversal. If this transversal has $A(1,1)$ attached to it, the positions form a section in the $4 \times n$ array. If the section is a transversal we are done. If not, there must be two symbol the same in the section. Without loss of generality let the symbol be 1 and let it occur in $A(1,1)$ and $A(2,2)$. Also let $A(3,3)$ hold symbols 2 and $A(4,4)$ hold symbol 3. Then $A(1,j)$ and $A(2,j)$ for $j = 5, 6, \ldots, m$ hold an x which may be a 1, 2 or 3.

In A, the first two rows have $n-3$ positions filled in. Could the entries all be 1's? Since $2(n-3) > \lfloor (4n - \lfloor \frac{4n-1}{3} \rfloor)/3 \rfloor$ for $n \geq 11$, all the x's are not 1's. So without loss of generality, let $A(1,5)$ hold symbol 2 as shown in Figure 2.

Because $A(1,5), A(2,2)$ and $A(4,3)$ is a near transversal there are at least $n-3$ x's and 2's in the third row of A. Now symbol 2 may be the most frequent symbol in the array so the sum of the frequencies of symbol 1 and 2 is at most $\lfloor (4n - \lfloor \frac{4n-1}{3} \rfloor)/3 \rfloor + \lfloor \frac{4n-1}{3} \rfloor$. Since $3(n-3) > \lfloor (4n - \lfloor \frac{4n-1}{3} \rfloor)/3 \rfloor + \lfloor \frac{4n-1}{3} \rfloor$ for $n \geq 12$, the x's can not all be 1's and 2's in the first three rows. So there is a 3 somewhere in the first 3 rows and first 6 columns of A. This 3 will be in a transversal if we consider just the array of the first 3 rows. Then the last row must contain $n-4$ x's to prevent the transversal in the $4 \times n$ array. In the case of $n = 11$, the x's could just all be 1's and 2's in the first 3 rows. But $A(2,6)$ is either a 1 or a 2. If it is a 1 then it forces an x into $A(1,2)$ or if it is a 2 it forces an x into $A(3,2)$. Then there must be a 3 in the first 3 rows and first 6 columns of A and as before to prevent a transversal there are $n-4$ x's forced into the fourth

1				2	x	x	...
	1			x	x	x	...
x		2			x	x	...
?	?	?	3	?	?	x	...

Figure 3:

row. Without loss of generality, we have Figure 3.

Note that at least two of the ?'s are x's; i.e., symbols 1, 2 or 3. If $n \geq 18$, then by Theorem 2.5 we would be done as there would be too many x's that would have to be 1, 2 or 3's. For $11 < n < 18$, we get no contradiction but we do get a lower bound on the frequency of the symbols larger than 3 in A. In Figure 3, there are at most 12 positions in A which could hold a symbol larger than 3. If these positions were filled with two or more symbols greater than 3, then the sum of the frequency of these two symbols would be less than $B(4,n) = \lfloor \frac{4n-1}{3} \rfloor$. In this situation, Lemma 2.4 implies that we need only consider the case when there is one symbol in A greater than 3, say 4. Symbol 4 can only occur in either the empty cells or in three of the "?" occupied cells in Figure 3. Symbol 4 and symbols 1, 2 and 3 fill these 12 cells. Let f_i be the frequency of symbol i in the array. Then $f_1 + f_2 + f_3 + f_4 = 4n$. Since f_2 and f_3 are less than or equal to $\lfloor \frac{4n-1}{3} \rfloor$, $f_1 + f_4 \geq 4n - 2\lfloor \frac{4n-1}{3} \rfloor$. Since $f_4 \geq f_1$. Then $f_4 \geq 2n - \lfloor \frac{4n-1}{3} \rfloor$. For $n \geq 11$, $f_4 \geq 8$.

Because there are no transversals in A, putting symbol 4 into a cell of A forces some of the symbols not to be in some row beyond column 6 and forces other symbols to be in that row beyond column 6. For instance, if $A(1,3) = 4$, then there is a near transversal $A(1,3) = 4, A(2,2) = 1$ and $A(4,4) = 3$. Then the x's in the third row beyond column 6 must be 1's and/or 3's but not 2's. Table 1 shows the near transversals and their implication for the symbols in the rows beyond column 6.

So, let $A(2,3)$ contain a 4. Then result 5 says row 3 beyond column 6 contains only 3's. But these new 3's are also in near transversals (with $A(1,5)$ and $(2,2)$) that force x's into $A(4,1), A(4,3)$ and $A(4,6)$. Also, these new 3's form near transversals with $A(2,3) = 4$ and $A(1,5) = 2$ forcing row 4 beyond row 6 not to have any 1's. Further. these new 3's are in near transversal with $A(2,3) = 4$ and $A(1,1) = 1$ forcing row 4 beyond column 6 to not contain any 2's. So row 4 beyond column 6 is all 3's. The new 3's in row 4 beyond column 6 form near transversals with $A(3,3) = 2$ and $A(2,2) = 1$ to force an x in $A(1,4)$; with $A(3,3) = 2$ and $A(1,1) = 1$ to force an x in $A(2,4)$; with $A(1,5)=2$ and $A(2,2)=1$ to force an x in $A(3,4)$. Now

Result	Near transversal	Row	Symbol is not
1	A(1,2)=4 A(3,3)=2 A(4,4)=3	2	1
2	A(1,3)=4 A(2,2)=1 A(4,4)=3	3	2
3	A(1,4)=4 A(2,2)=1 A(3,3)=2	4	3
4a	A(2,1)=4 A(3,3)=2 A(4,4)=3	1	1
4b	A(2,1)=4 A(1,5)=2 A(4,4)=3	3	1
5a	A(2,3)=4 A(1,1)=1 A(4,4)=3	3	2
5b	A(2,3)=4 A(1,5)=2 A(4,4)=3	3	1
6	A(2,4)=4 A(1,1)=1 A(3,3)=2	4	3
7a	A(3,2)=4 A(1,1)=1 A(4,4)=3	2	2
7b	A(3,2)=4 A(1,5)=2 A(4,4)=3	2	1
8	A(3,4)=4 A(1,5)=2 A(2,2)=1	4	3
9a	A(3,5)=4 A(1,1)=1 A(4,4)=3	2	2
9b	A(3,5)=4 A(2,2)=1 A(4,4)=3	1	2
10a	A(4,1)=4 A(2,2)=1 A(3,3)=2	1	3
10b	A(4,1)=4 A(2,2)=1 A(1,5)=2	3	3
11	A(4,2)=4 A(1,1)=1 A(3,3)=2	2	3
12	A(4,3)=4 A(1,5)=2 A(2,2)=1	3	3
13a	A(4,5)=4 A(1,1)=1 A(3,3)=2	2	3
13b	A(4,5)=4 A(2,2)=1 A(3,3)=2	1	3
14a	A(4,6)=4 A(1,1)=1 A(3,3)=2	2	3
14b	A(4,6)=4 A(2,2)=1 A(3,3)=2	1	3
14c	A(4,6)=4 A(1,5)=2 A(2,2)=1	3	3

Table 1: Effect of near transversals beyond column 6

there is just enough room for 8 4's in the array. So all the empty positions in the array must be 4's. Then result 4a says $A(2,1) = 4$ implies that row 1 beyond column 6 can not contain 1's and result 9b says $A(3,5) = 4$ implies that row 1 beyond column 6 can not contain 2's. So row 1 beyond column 6 contains 3's. Then rows 1, 2 and 3 beyond column 6 contains $3(n-6)$ 3's. Since $3(n-6) > \lfloor\frac{4n-1}{3}\rfloor$ for $n \geq 11$, we get a contradiction. So $A(2,3)$ contains an x. There are now 11 positions that could hold a 4 in the array.

Suppose $A(3,2)$ contains a 4, then result 7 says that row 2 beyond column 6 contains only 3's. These 3's are in near transversals which force x's into $A(4,2)$, $A(4,5)$ and $A(4,6)$. Then there are only 10 positions for the 4's. Now these new 3's with $A(3,2)$ and $A(1,1)$ force row 4 beyond column 6 not to contain any 2's. Also these new 3's with $A(3,2)$ and $A(1,5)$ force row 4 beyond column 6 not to contain 1's. So row 4 beyond column 6 contains only 3's. But, if $A(1,4)$, $A(2,4)$ or $A(3,4)$ were 4's then they would be in transversals so they must contain x's. But now we have room for only seven 4's in our array and this is a contradiction. So $A(3,2)$ must contain an x. There is only room for 10 4's in our array.

Let us now assume that $A(2,1) = 4$, then result 4 implies that rows 1 and 3 beyond column 6 and do not contain 1's. If we further assume that $A(3,5) = 4$ then result 9 implies that rows 1 and 2 beyond column 6 do not contain 2's. So row 1 beyond column 6 contains 3's. So the new 3's are in near transversals forcing $A(4,1), A(4,5)$ and $A(4,6)$ to contain x's. So there are 9 spots for the 4's. Two of the new 3's along with $A(2,1) = 4$ and $A(3,3) = 2$ force row 4 beyond column 6 to not contain 1's. Also one of the new 3's along with $A(3,5) = 4$ and $A(2,2) = 1$ force row 4 beyond column 6 to not contain 2's. So row 4 beyond column 6 must contain 3's. But then if $A(1,4), A(2,4)$ or $A(3,4)$ contained 4's they would be in transversals which would be a contradiction. So they contain x's. But that would leave only 6 spots in the array for the 4's which is a contradiction.

So if $A(2,1) = 4$ then $A(3,5) = x$ and there are 9 spots for the 4's. Let us now assume that $A(1,3) = 4$. Result 2 implies that row 3 beyond column 6 can not contain 2's. We already know it can not contain 1's so it must contain 3's. Any one of these 3's are in near transversal that forces $A(4,1), A(4,3)$ and $A(4,6)$ to contain x's. There are 8 spots for the 4's now. Assume that $A(4,5) = 4$. By result 13b, we know that row 1 beyond column 6 can not contain 3's. Since we already know that it can not contain 1's it must contain 2's. Now a new 2 and a new 3 in different columns along with $A(4,5) = 4$ a $A(2,2) = 1$ form a transversal so $A(4,5)$ must contain an x. This is a contradiction as there are only 7 spots for the 4's in the array. So $A(1,3)$ can not be a 4 and must be an x.

So we have $A(2,1) = 4$ with $A(1,3) = x$ and $A(3,5) = x$ with 8 spots for the 4's. We also have the row 1 beyond column 6 and row 3 beyond column 6 can not contain 1's. We now know that $A(1,2), A(1,4), A(2,4)$ and $A(3,4)$ are 4's. So result 1 gives that row 2 beyond column 6 can not be 1's also. Now 3 of the 5 non-determined spots in row 4 must be 4's. If $A(4,6) = 4$, then result 14 gives that the rows 1, 2, 3 beyond column 6 can not be 3's and so must be 2's. This is too many 2's so A(4,6)=x. We get the same contradiction if we assume that $A(4,1) = A(4,2) = 4$ using results 10 and 11, So only at most 1 of $A(4,1)$ and $A(4,2)$ can be a 4. Similarly, using results 12 and 13, only one of $A(4,3)$ and $A(4,5)$ can be a 4. Hence there can only be at most seven 4's in the array. Contradiction.

So we have that $A(2,1), A(3,2)$ and $A(2,3)$ are all x's and there are 9 possible places in the array for the at least 8 4's. Now assume that $A(3,5) = 4$. Result 9 gives that rows 1 and 2 beyond column 6 can not contain 2's. If we further assume that $A(1,2) = 4$ then result 1 gives that row 2 beyond column 1 contain no 1's so it must contain only 3's. Now these

3's in row 2 along with $A(3,3) = 2$ and $A(1,2) = 4$ force that row 4 beyond column 6 can not contain 1's. Also the 3's in row 2 along with $A(1,1) = 1$ and $A(3,5) = 4$ force row 4 beyond column 6 to not contain 2's. So row 4 beyond 6 must contain 3's. Now $A(4,7) = 3, A(3,3) = 2, A(2,2) = 1$ force $A(1,4) = x$. Also $A(4,7) = 3, A(3,3) = 2$ and $A(1,1) = 1$ force $A(2,4) = x$. But this is a contradiction as there are only 7 spots for the 4's in the array. So $A(1,2)$ can not contain a 4 but must contain a x.

So if $A(3,5) = 4$ then $A(1,2) = x$ and there are 8 places for the 4's. So $A(1,3) = A(1,4) = A(2,4) = A(3,4) = 4$ and we have that rows 1, 2, 3 beyond column 6 can not be 2's using results 2 and 9. We now repeat the argument from two paragraphs before about 4's in row 4 to get the same kind of contradiction as before.

So now we have $A(3,5) = A(2,1) = A(2,3) = A(3,2) = x$ and there are 10 places to put the 8 4's with 3 of the 4's being in row 4. So we know that $A(1,3) = A(1,4) = A(2,4) = A(3,4) = 4$. By results 1 and 2 we know that the cells in row 2 beyond column 6 can not be 1's and the cells in row 3 beyond column 6 can not be 2's. Then, if $A(4,6) = 4$, result 14 implies that there are no 3's in rows 1, 2 and 3 beyond column 6. Then in row 2 beyond column 6 we have 2's and and in row 3 beyond column 6 we have 1's. Then $A(2,7) = 2$, $A(3,8) = 1$, $A(4,4) = 3$ and $A(1,3) = 4$. This is a contradiction so $A(4,6) = x$. Similarly, one of $A(4,1)$ and $A(4,2)$ is an x and one of $A(4,3)$ and $A(4,5)$ is an x. So there is space in A for only 7 4's. This is our final contradiction. So our assumption that there is an array with symbols having a maximum frequency of $\lfloor \frac{4n-1}{3} \rfloor$ and with no transversal is false. □

We will use the near transversal concept, one more time to get a backtrack program to show that if an array has maximum frequency of $f = B(4,n)$ then it must have a transversal. Consider the situation in figure 3. Those x's could be filled in with 1's, 2's or 3's in a backtrack manner. As you fill them in then more near transversals are formed forcing more positions that must contain an x. The process stops in one of two ways. Either you run out of x's and the algorithm is stuck or the frequency of 1, 2 or 3 or the sum of the frequencies of 1, 2, 3, x becomes too large. In these cases, we backtrack. When $m = 4$ and n=8, 9, 10 and 11, this was done and the algorithm never got stuck. These results and the fact that $L(4,7)$ equals $8 < 9 = B(4,7)$ prove the following theorem.

Theorem 2.7 $n_0(4) = 8$.

Then, using the table from Stein and Szabó [4] and the previous theorem we have now determined $L(4,n)$.

Theorem 2.8

n	4	5	6	7	$n \geq 8$
$L(4,n)$	3	4	5	8	$B(4,n)$

3 Conclusion

We have shown that the difference between the upper bound and the lower bound for $L(m,n)$ is at most m. Then we have shown that $n_0(m)$ is less than or equal to $\lfloor \frac{m(n-m+1)-1}{m-1} \rfloor$. The last thing we did was to determine $L(4,n)$ for all n. We now conjecture how $L(m,n)$ behaves for $m > 4$. Our conjecture is modified from Stein's conjectures. The most interesting area for $L(m,n)$ is from $n = m$ to $n = 2m - 2$ where we know little for $m > 4$. There seems to be a discontinuity at $n = 2m - 1$. For $n \geq 2m - 1$, we think that either $L(m,n) = B(m,n)$ or for those values of n in Theorem 2.3, $L(m,n) = B(m,n) - 1$. This conjecture is true for $n = 2,3$ and has now been proven for $n = 4$ also.

4 Acknowledgements

The author would like the thank the referee for his very careful reading of the paper and all his suggestions.

References

[1] S. Akbari, O. Etsami, H. Mahini, M. Mahmoody & A. Sharifi, *Latin transversls in long rectangular arrays*, Discrete Math. **306** (2006) no., 23, 3011–3013.

[2] N. Alon, *Probabilistic proofs of existence of rare events*, Springer Lecture Notes in Mathematics No. 1376 (J. Lindenstrauss and V. D. Milman Eds.), Springer-Verlag (1988), 186-201.

[3] J. Dénes and A. D. Keedwell,*Latin Squares and Their Applications*, Akademiai Kiado, Budapest, (1974) 103.

[4] S. K. Stein & S. Szabó, *The number of distinct symbols in sections of rectangular arrays*, Discrete Math. **306** (2006), no. 2, 254–261.

[5] G. H. J. van Rees, *Lovász Local Lemma*, vanrees@cs.umanitoba.ca/ ∼ vanrees.