Solution of the Intersection Problem for Latin Squares of Different Orders

Peter Dukes* and Jared Howell
Mathematics and Statistics
University of Victoria
Victoria, BC V8W 3R4
Canada
dukes@uvic.ca, jajh@uvic.ca

January 26, 2011

Abstract

A complete solution is obtained for the possible number of common entries between two latin squares of different given orders. This intersection problem assumes the entries of the smaller square are also entries of the larger, and that, for comparison, the smaller square is overlayed on the larger. However, these extra restrictions do not affect the solution, apart from one small example.

1 Introduction

A *partial latin square* L of *order* n and entry set E, $|E| = n$, is an $n \times n$ array in which

- every cell of L is either empty or contains an element of E, and
- every row and every column of L contains no repeated elements of E.

More importantly, a *latin square* is a partial latin square with no empty cells. Every row and every column is a permutation of E in this case.

*research supported in part by NSERC

A partial latin square L can be identified with a subset of the ordered triples $R \times C \times E$, where R indexes the rows, C indexes the columns, and $(r, c, e) \in L$ if and only if the (r, c)-entry in L is e. In this representation, no ordered pair appears in the same two coordinates more than once. For instance, one of the two latin squares with $R = C = E = \{1, 2\}$ is $\{(1,1,1), (1,2,2), (2,1,2), (2,2,1)\}$.

In what follows, the 'array' and 'ordered triple' representations are used interchangeably. The set of row indices of L is here denoted rows(L), and similarly for columns (cols) and entries (ents).

Given two partial latin squares L and L' on the same set of rows, columns and entries, regard their intersection $L \cap L'$ as the set of common entries in common cells, or, alternatively as the set of common triples.

Define the *intersection spectrum* for latin squares of order n as

$$I(n) = \left\{ |L \cap L'| : \begin{array}{c} L, L' \text{ are latin squares of order } n, \text{ rows}(L) = \text{rows}(L') \\ \text{cols}(L) = \text{cols}(L'), \text{ents}(L) = \text{ents}(L') \end{array} \right\}.$$

Since every row and column needs to be represented either zero or at least twice in the difference $L \setminus L'$, it follows that

$$I(n) \subseteq \Upsilon(n) := \{0, 1, 2, \ldots, n^2\} \setminus \{n^2 - i : i = 1, 2, 3, 5\}. \tag{1}$$

The complete determination of $I(n)$ is an important result due to Fu and Fu.

Theorem 1.1 ([3]).

$$I(n) = \begin{cases} \{1\} & \text{if } n = 1; \\ \{0, 4\} & \text{if } n = 2; \\ \{0, 3, 9\} & \text{if } n = 3; \\ \{0, 1, 2, 3, 4, 6, 8, 9, 12, 16\} & \text{if } n = 4; \\ \Upsilon(n) & \text{if } n \geq 5. \end{cases}$$

In [2], a two-parameter version of this problem was introduced. Let L_1 be a latin square of order n and L_2 a latin square of order $m \geq n$. Assume the sets of rows, columns and entries of L_2 contain the corresponding sets for L_1. In this way, L_1 and L_2 are partial latin squares on the same m rows, columns, and entries. For such squares, define $L_1 \cap L_2$ to be the intersection as partial latin squares, and let

$$I(n, m) = \left\{ |L_1 \cap L_2| : \begin{array}{c} L_1 \text{ order } n, L_2 \text{ order } m, \text{rows}(L_1) \subseteq \text{rows}(L_2), \\ \text{cols}(L_1) \subseteq \text{cols}(L_1), \text{ents}(L_1) \subseteq \text{ents}(L_2) \end{array} \right\}.$$

Example 1.2. The reader is encouraged to verify that $I(2,3) = \{0,1,3\}$. Probably the least obvious fact is that $2 \notin I(2,3)$, but two easy cases (either adjacent or diagonal intersecting cells) rule this out.

Since they occur frequently in what follows, intervals of integers are represented in the usual way:

$$[\alpha, \beta] := \{x \in \mathbb{Z} : \alpha \leq x \leq \beta\}.$$

From Ryser's theorem on completing partial latin squares (see Theorem 2.1 below), it is straightforward to see that $I(n,m) = [0, n^2]$ when $m \geq 2n$. In light of this and Theorem 1.1, the focus is on $I(n, n+k)$, $0 < k < n$. In [2], it was shown that the maximum intersection in this case is $n^2 - nk + k^2$, obtained when L_1 of order n is 'quasi-embedded' (i.e. completed after a minimum number of cells are emptied) in L_2 of order $n+k$. One has

$$I(n, n+k) \subseteq [0, n^2 - nk + k^2], \tag{2}$$

and equality is easily seen to hold in many cases. The results in [2] fall short of reducing the determination of all $I(n, n+k)$ to a finite problem. The difficult cases are observed to be small k for certain large intersections.

This article completely settles the intersection problem for latin squares of different orders.

Theorem 1.3. *Let $n > k$ be positive integers. Then*

$$I(n, n+k) = \begin{cases} \{0,1,3\} & \text{if } (n,k) = (2,1); \\ [0, n^2 - nk + k^2] & \text{otherwise.} \end{cases}$$

The reduction to a finite problem appears in Section 2, and is divided into two broad cases: $k > 2$ and $k \leq 2$. In the latter case, four constructions suffice. The proof of Theorem 1.3 is then completed by an easy computer search. The details of this search are given in Section 3.

As an aside, consider 'unrestricted' intersections between latin squares L_1, L_2 of different orders, where the condition

$$\text{ents}(L_1) \subset \text{ents}(L_2) \tag{3}$$

is dropped. To this end, define

$$I_u(n,m) = \{|L_1 \cap L_2| : L_1 \text{ order } n, L_2 \text{ order } m\}.$$

Then obviously

$$I(n,m) \subseteq I_u(n,m) \subseteq [0, n^2 - nk + k^2],$$

and by Theorem 1.3 I_u is settled except for one easy case. (Likewise, Theorem 1.1 is not substantially impacted by dropping condition (3).) Due to squares

$$\begin{array}{|cc|}\hline 1 & 2 \\ 2 & 1 \\ \hline \end{array} \text{ and } \begin{array}{|ccc|}\hline 3 & 2 & 4 \\ 2 & 4 & 3 \\ 4 & 3 & 2 \\ \hline \end{array},$$

it follows that $2 \in I_u(2,3)$. Compare with Example 1.2.

Corollary 1.4. *For positive integers $n > k$, $I_u(n, n+k) = [0, n^2 - nk + k^2]$.*

2 Proof for sufficiently large orders

Before moving on to details of the proof, it is worth mentioning a standard approach for obtaining some desired intersection. This is relevant both for theoretical constructions and small examples on computer. Recall the following important result on completing rectangles to latin squares.

Theorem 2.1 (Ryser's Theorem, [6]). *Suppose L is a partial latin square of order n in which a cell is filled if and only if it lies in the first r rows and s columns. If each symbol appears at least $r + s - n$ times in L, then L can be completed to a latin square of order n.*

In this context, if L is an $n \times n$ array on $n + k$ symbols, and such that each symbol appears at least $n - k$ times, then L can be completed to a latin square L_2 of order $n + k$. Constructing L from an initial square L_1 of order n by modifying x entries results in intersection value $|L_1 \cap L_2| = |L_1 \cap L| = n^2 - x$.

A key technique for entry modification is substitution along transversals.

Recall that a *transversal* in a (partial) latin square L is a subset of triples (r, c, e) which exhaust the rows, columns, and entries of the square. In a *partial transversal*, the requirement is weakened so that no row, column, or entry is repeated.

The existence of an orthogonal mate to L is enough to guarantee n disjoint transversals in L. It is also known that there are 4 disjoint transversals in certain latin squares of order six. See [1] for more information on orthogonality and transversals.

In any case, if $T_1, \ldots, T_k$ are disjoint (partial) transversals of L, then replacing $n-k$ of the entries in each transversal T_i with a new symbol ∞_i results in a Ryser-completeable partial latin square of order $n+k$. This proves the existence of a quasi-embedding and shows that $n^2 - nk + k^2 \in I(n, n+k)$. In fact, replacing up to all n entries on each of k full transversals settles the top piece of the intersection spectrum.

Lemma 2.2 ([2]). *For $n > 6$, $I(n, n+k) \supseteq [n^2 - nk, n^2 - nk + k^2]$.*

The condition $n > 6$ in Lemma 2.2 ensures that k disjoint transversals exist, though in fact the conclusion holds unless $(n,k) = (2,1)$ or $(6,5)$.

2.1 $k > 2$

This section borrows two further techniques from [2]: permuting rows and imposing latin subsquares. However, these get combined in somewhat more generality.

It is clear what is meant by saying that one latin square is a *subsquare* of another. A related notion is a partial latin square L° whose only empty cells form a $t \times t$ sub-array, and the rows and columns of that sub-array in L° avoid precisely the same set of t entries. This object is usually called a (partial) latin square having a *hole* of size t. For instance, the square of order 5 below has a hole of size 2.

		3	4	5
		4	5	3
3	4	5	1	2
4	5	2	3	1
5	3	1	2	4

Obviously, if M is a $t \times t$ latin square on the missing rows, columns, and entries, then $L = L^\circ \cup M$ is a latin square having M as a subsquare.

Define $I(n, n+k; t)$ to be the set of all intersection sizes $|L_1^\circ \cap L_2^\circ|$, where

- L_1° is a latin square of order n with a hole of size t;
- L_2° is a latin square of order $n+k$ with a hole of size t; and
- the holes of L_1° and L_2° are on the same rows, columns and entries.

By the above remark,

$$I(t) + I(n, n+k; t) \subseteq I(n, n+k). \quad (4)$$

Of course, the sum of two sets of integers is to be interpreted in the usual way:

$$A + B = \{a + b : a \in A, b \in B\}.$$

Now, a key technical lemma combines transversals and permutations to obtain a mesh of values in $I(n, n+k; t)$.

Lemma 2.3. *Let $t \geq 2$ and suppose $n \geq 3t$. Then $I(n, n+k; t)$ contains 0, $n^2 - nk + k^2 - t^2$, and has no gap of size greater than $2(n-t)$.*

Proof: By a result of Heinrich, [4], the hypotheses guarantee that there exist orthogonal latin squares of order n having common holes of order t. For ease of reference, the holes are assumed in the upper-left corner.

By orthogonality, one of these partial latin squares – call it L° – has n disjoint partial transversals $T_1, \ldots, T_n$, with (say) the first $n-t$ being full transversals and the last t being of size $n-t$.

Consider the first k (partial) transversals $T_1, \ldots, T_k$. Replace the $n-k$ entries in rows $k+1$ through n in each T_i, $i = 1, \ldots, k$, with new symbol ∞_i.

It must be verified that this is possible. Observe that if $k < t$ then $k \leq n-t$ and $T_1, \ldots, T_k$ are all full. On the other hand, if $k \geq t$, then $n-k \leq n-t$. In either case, we have enough available entries on $T_1, \ldots, T_k$ for the above replacement.

Each symbol, old and new, appears at least $n-k$ times. So by Ryser's Theorem, extend to a partial latin square L_2° of order $n+k$. (To be precise, one must fill the upper-left hole with a subsquare of order t, Ryser-complete, and remove the subsquare.)

Now, return to L° and apply certain row-permutations π, leaving rows 1 through t invariant (but not necessarily fixed). There are between zero and $n-2$ fixed rows, or all n rows fixed. Call such a partial latin square L_1°. By construction, L_1° and L_2° have common holes of size t.

If no rows are fixed by π, then $|L_1^\circ \cap L_2^\circ| = 0$. If all rows are fixed, then $|L_1^\circ \cap L_2^\circ|$ achieves its maximum at $n^2 - nk + k^2 - t^2$. Otherwise, count $|L_1^\circ \cap L_2^\circ|$ row by row. Each row of L_1° that is fixed by π contributes $n-t$, $n-k$, $n-k-t$, or n to the intersection with L_2°, depending (respectively) on whether that row is among the first t rows, last $n-k$ rows, both, or neither. Varying the number and location of fixed rows (with $n-1$ fixed rows disallowed) prevents gaps of size greater than $2(n-t)$ in $I(n, n+k; t)$. □

The above constructions go a long way toward solving the problem.

Proposition 2.4. *For $n > 20$ and $2 < k < n$, $I(n, n+k) = [0, n^2 - nk + k^2]$.*

Proof: Pick $t = \lfloor \frac{n}{3} \rfloor > 6$. It can be easily verified that $2(n-t) \leq t^2 - 6$. By (4), Theorem 1.1 and Lemma 2.3, it follows that $I(n, n+k)$ contains all but possibly four values $n^2 - nk + k^2 - j$, $j = 1, 2, 3, 5$. But Lemma 2.2 covers these for $k > 2$. □

Actually, values of $n \leq 20$ yield partial results too. Often only a few intersection values are left undecided. For instance, it turns out that for $n = 15$ (and any $k > 2$) the entire interval is obtained with $t = 5$.

Note that this method also settles $k = 2$, except for the intersection value $n^2 - nk + k^2 - 5 = n^2 - 2n - 1$, and $k = 1$, except for 3 values for each n. These low values of k are treated next.

2.2 $k \leq 2$

As indicated earlier, the proof of Proposition 2.4 applies for $k \leq 2$, except that the interval in Lemma 2.2 misses a few intersection values. For $k = 1$, these are $n^2 - n - 4$, $n^2 - n - 2$, and $n^2 - n - 1$.

Lemma 2.5. *For $n \geq 6$, $n^2 - n - 4 \in I(n, n+1)$.*

Proof: As in the proof of Lemma 2.3, there exists a latin square L_1° of order n having a hole of order 2 and a disjoint (full) transversal T. Replace all n entries of T with new symbol ∞, and complete to a latin square L_2° of order $n+1$. (This is also known as a 'prolongation' along T.) It follows that $|L_1^\circ \cap L_2^\circ| = n^2 - n - 4$. Applying (4) with disjoint subsquares of order 2 results in the desired intersection of latin squares. □

Lemma 2.6. *For $n \geq 6$, $n^2 - n - 2 \in I(n, n+1)$.*

Proof: Take a latin square L of order n having a 2×2 subsquare and a partial transversal T of order $n-1$ that intersects the subsquare in exactly one entry. (This exists either from Heinrich's Theorem, or alternatively from a prolongation along 2 transversals in a square of order $n-2$.) Replace all $n-1$ entries of T with new symbol ∞, and Ryser-complete to L_2. Then $|L \cap L_2| = n^2 - n + 1$. Now, turn the 2×2 subsquare in L to produce L_1. This reduces the intersection by exactly 3, yielding $|L_1 \cap L_2| = n^2 - n - 2$, as desired. □

The remaining value requires a similar yet slightly more intricate construction.

Lemma 2.7. *For $n \geq 9$, $n^2 - n - 1 \in I(n, n+1)$.*

Proof: Take a latin square L_1 of order n having a 3×3 subsquare on entries $\{1, 2, 3\}$ and a partial transversal T of order $n-1$ that intersects the subsquare in exactly 2 cells. Replace all entries of T with new symbol ∞, and Ryser-complete to L_2. With minor changes it may be assumed that the subsquare in L_1 and corresponding entries of L_2 are

$$\begin{array}{|ccc|}\hline 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ \hline \end{array} \quad \text{and} \quad \begin{array}{|ccc|}\hline \infty & 2 & 3 \\ 2 & \infty & 1 \\ 1 & 3 & 2 \\ \hline \end{array},$$

respectively. Note that these squares have intersection $5 = 3^2 - 4$. It follows that $|L_1 \cap L_2| = n^2 - (n-3) - 4 = n^2 - n - 1$. □

In summary, Lemmas 2.5, 2.6, 2.7 and the construction in Section 2.1 completely settle the case $k = 1$ for large n.

Proposition 2.8. *For $n > 20$, $I(n, n+1) = [0, n^2 - n + 1]$.*

The only presently outstanding intersection value for $k = 2$ is $n^2 - 2n - 1$. This is handled with another hole of size 2.

Lemma 2.9. *For $n \geq 6$, $n^2 - 2n - 1 \in I(n, n+2)$.*

Proof: This mirrors the proof of Lemma 2.5. Take a latin square L_1° of order n having a hole of order 2 and a two disjoint (full) transversals T_1, T_2. Replace $n-1$ entries of T_1 with new symbol ∞_1, $n-2$ entries of T_2 with ∞_2, and complete by Ryser's Theorem to L_2° of order $n+1$ with a common hole. It follows that $|L_1^\circ \cap L_2^\circ| = n^2 - (n-1) - (n-2) - 4 = n^2 - 2n - 1$ and disjoint subsquares in the hole complete the construction. □

So $k = 2$ is now finished for large n.

Proposition 2.10. *For $n > 20$, $I(n, n+2) = [0, n^2 - 2n + 4]$.*

3 Computation for small orders

It remains to determine $I(n, n+k)$ for positive integers $k < n \leq 20$. Although techniques from the previous section yield partial results for $n \leq 20$, it seems reasonable to simply compute all $I(n, n+k)$ directly. This is guided by a hunch that equality holds in (2) provided $(n, k) \neq (2, 1)$, and that examples of each allowable intersection are plentiful. Indeed, this is the result of our computation.

Proposition 3.1. *For $3 \leq n \leq 20$ and $1 \leq k < n$, $I(n, n+k) = [0, n^2 - nk + k^2]$.*

The computation has an outer loop on n and k. For fixed (n, k), initialize latin squares M_1^0 and M_2^0 as disjoint of orders n and $n+k$, respectively. (This is easy.) Store the pair (M_1^0, M_2^0) for intersection zero. Put `unachieved` $:= [1, n^2 - nk + k^2]$.

Repeat the following: Identify the longest gap in `unachieved`, and the intersection value i preceding this longest gap. Suppose (M_1^i, M_2^i) has been stored for intersection i. Repeatedly construct latin squares L_1, L_2 of orders n and $n+k$ by applying the algorithm of Jacobson and Matthews [5] with respective seeds M_1^i, M_2^i. Although the limiting distribution of L_1 (likewise L_2) is uniform, we stop the latin square algorithm relatively soon, so that $|M_j^i \cap L_j|$ is large. Finally, if $h := |L_1 \cap L_2| \in$ `unachieved`, delete it and store (L_1, L_2) as (M_1^h, M_2^h) for intersection h.

This implentation in C on an AMD Athlon (1.1 GHz, 256 Mb RAM, running Cygwin) took roughly 22 hours and 17 minutes to construct all

$$\sum_{n=3}^{20} \sum_{k=1}^{n-1} n^2 - nk + k^2 + 1 = 34101$$

pairs of latin squares achieving all intersection values. A complete file of these intersection values, measuring 6.1 Mb, is available for download at `http://www.math.uvic.ca/faculty/dukes/int-ls-data.gz`.

The proof of Theorem 1.3 now follows directly from Example 1.2 and Propositions 2.4, 2.8, 2.10, and 3.1.

Once again, it should be stressed that most of the direct constructions can be easily obtained without computer. Those cases for which the computer is useful probably admit nicer computational techniques. However, for ease of presentation we choose to draw the line at $n = 20$ and outline the simple algorithm above.

References

[1] J. Dénes and A.D. Keedwell, *Latin Squares and Their Applications*, English Universities Press, London, 1974.

[2] P. Dukes and E. Mendelsohn, Quasi-embeddings and intersections of latin squares of different orders. *Australasian J. Combin.* 43 (2009), 197–209.

[3] C.M. Fu and H.-L. Fu, The intersection problem of Latin squares, Graphs, designs and combinatorial geometries (Catania, 1989), *J. Combin. Inform. System Sci.* 15 (1990), 89–95.

[4] K.E. Heinrich, Latin squares with and without subsquares, in: *Latin Squares: New Developments in the Theory and Applications* (J. Dénes and A.D. Keedwell; eds.) North-Holland, Amsterdam, 1991.

[5] M.T. Jacobson and P. Matthews, Generating uniformly distributed random Latin squares. *J. Combin. Des.* 4 (1996), 405–437.

[6] H.J. Ryser, A combinatorial theorem with an application to Latin rectangles. *Proc. Amer. Math. Soc.* 2 (1951), 550–552.