Lifting Redundancy from Latin Squares to Pandiagonal Latin Squares

Bart Demoen

Report CW 645, August 2013

KU Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

Lifting Redundancy from Latin Squares to Pandiagonal Latin Squares

Bart Demoen

Report CW 645, August 2013

Department of Computer Science, KU Leuven

Abstract

In the pandiagonal Latin Square problem, a square grid of size N needs to be filled with N types of objects, so that each column, row, and wrapped around diagonal (both up and down) contains an object of each type. This problem dates back to at least Euler. In its specification as a constraint satisfaction problem, one uses the *all_different* constraint. The known redundancy result about *all_different* constraints in the Latin Square problem is lifted to the pandiagonal Latin Square problem. This proof method's theoretical limits are established.

Lifting Redundancy from Latin Squares to Pandiagonal Latin Squares

Dedicated to Ajana Beke 18 August 2013

Bart Demoen
Department of Computer Science
KU Leuven

Abstract

In the pandiagonal Latin Square problem, a square grid of size N needs to be filled with N types of objects, so that each column, row, and wrapped around diagonal (both up and down) contains an object of each type. This problem dates back to at least Euler. In its specification as a constraint satisfaction problem, one uses the *all_different* constraint. The known redundancy result about *all_different* constraints in the Latin Square problem is lifted to the pandiagonal Latin Square problem. This proof method's theoretical limits are established.

1 Introduction

In one popular formulation, solving the Latin Square problem of size N requires filling out an $N \times N$ square grid with numbers from 1 to N in such a way that each row and column contains every number exactly once. In the *pandiagonal* Latin Square problem, the additional requirement is that all pandiagonals also have exactly the numbers 1 to N. A pandiagonal is a wrapped around diagonal, of which there are $2N$: N up going pandiagonals, and N down going ones[1]. The pandiagonal Latin Square problem can therefore be specified as a Constraint Satisfaction Problem with (a) N^2 variables $x_{ij}, i, j \in [1..N]$, representing the value assigned to the cell in row i and column j of the square, (b) N^2 *domain* constraints indicating that the domain of each variable x_{ij} is $[1..N]$, and (c) $4N$ *all_different* constraints of N variables each, one for the variables in each column, one for the variables in each row, one for the variables in each up pandiagonal, and one for the variables in each down pandiagonal. We refer to the set of $4N$ *all_different* constraints of the pandiagonal Latin Square problem as *PDLS(N)*. The $2N$ *all_different* constraints of the usual Latin Square problem will be referred to by *LS(N)*. The domain constraints are left implicit. Solutions to *PDLS(N)* are known as Knut Vik designs [7], and used for example in the setup of agricultural experiments.

We denote by *Rows(N)*, *Cols(N)*, *UpPDs(N)*, *DownPDs(N)* the sets of *all_different* constraints of the rows, columns, up pandiagonals, and down pandiagonals. We prove in Section 2 that any three *all_different* constraints are *redundant* if they belong to three different sets from *Rows(N)*, *Cols(N)*, *UpPDs(N)*, *DownPDs(N)*: the proof uses the result from [2] about the redundancy present in $LS(N)$. Section 3 discusses the limits of the proof method: using this method, no more than those triplets can be proven redundant for $N > 5$: $N = 5$ provides a nice exception and an even nicer picture.

We will use freely the analogy between (solving) a CSP consisting only of disequalities and (coloring) its constraint graph: [4] contains more details.

2 The main result

Solutions to $PDLS(N)$ exist if and only if N is not divisible by 2 or 3: this result goes back to [6] and even the work of the great L. Euler[5]. Since we most often are only interested in the case when there are solutions, we will assume N is odd and not divisible by 3, unless mentioned explicitly.
The key to our theorem is the insight that an *PDLS(N)* contains a number of *LS(N)* problems: for each of these *LS(N)* problems, we can use a result proven in [2] about redundancy of *all_different* constraints, and a recombination to

[1]Other sources use *left* and *right* broken diagonal instead.

the *PDLS(N)* level then finishes the proof. We start by showing how an *PDLS(N)* is composed of six useful *LS(N)*s.

One ***PDLS(N)*** is worth at least six ***LS(N)***

We introduce the following six sets of *all_different* constraints:

- *row_col(N)* = *Rows(N)* ∪ *Cols(N)*
- *col_up(N)* = *Cols(N)* ∪ *UpPDs(N)*
- *col_down(N)* = *Cols(N)* ∪ *DownPDs(N)*
- *row_up(N)* = *Rows(N)* ∪ *UpPDs(N)*
- *row_down(N)* = *Rows(N)* ∪ *DownPDs(N)*
- *up_down(N)* = *UpPDs(N)* ∪ *DownPDs(N)*

For $N = 3$, the different problems are depicted in Figure 1. The cells are labeled with symbols instead of numbers, so as not to confuse the reader with the domain of the cells:

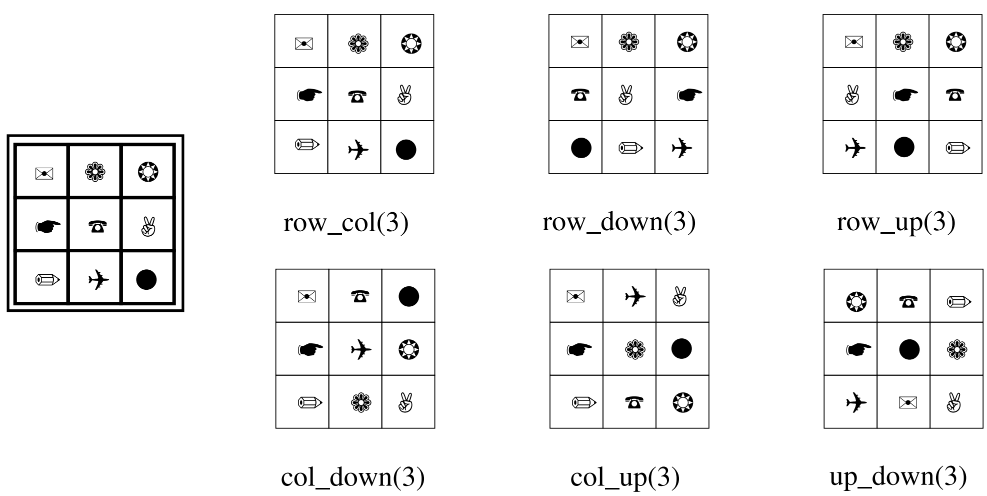

Figure 1: *PDLS(3)* (left with double border) and six of its *LS(3)*s

Clearly, each of these six sets is an $LS(N)$ for any N for which $LSDP(N)$ has a solution. Note that *up_down(N)* for even N is not, because an up and down pandiagonal always intersect in zero or two cells. It is also clear that $LSDP(N) = \mathit{row\_up(N)} \cup \mathit{col\_down(N)}$ and two more similar equalities can easily be established. We are now ready for the main theorem.

Sets of three redundant constraints in $LSDP(N)$

Theorem 1 Let $C \in \mathit{Cols(N)}$, $R \in \mathit{Rows(N)}$, $D \in \mathit{DownPDs(N)}$, and $U \in \mathit{UpPDs(N)}$. Then each of the sets $\{C, R, D\}$, $\{C, R, U\}$, $\{C, U, D\}$, $\{R, U, D\}$ is redundant.

Proof

We give the proof only for $\{C, R, D\}$: the proofs for the other sets are similar.

$$LSDP(N) \setminus \{C, R, D\} = \\ (\mathit{col\_up}(N) \setminus C) \cup (\mathit{col\_down}(N) \setminus \{C, D\}) \cup (\mathit{row\_down}(N) \setminus \{R, D\})$$

so we have a union of three $LS(N)$ problems minus some individual *all_different* constraints: one misses one *all_different* constraint, the others miss two. Since in [2], it was proven that any single *all_different* constraint of the $LS(N)$ problem is redundant, we know that $(\mathit{col\_up}(N) \setminus C)$ implies C and we can make the following derivation. The rewritten term is underlined:

$$\underline{(\mathit{col\_up}(N) \setminus C)} \cup (\mathit{col\_down}(N) \setminus \{C, D\}) \cup (\mathit{row\_down}(N) \setminus \{R, D\}) \Longrightarrow \\ (\mathit{col\_up}(N) \cup \underline{(\mathit{col\_down}(N) \setminus \{D\})} \cup (\mathit{row\_down}(N) \setminus \{R, D\}) \Longrightarrow \\ (\mathit{col\_up}(N) \cup \mathit{col\_down}(N) \cup \underline{(\mathit{row\_down}(N) \setminus \{R\})} \Longrightarrow \\ (\mathit{col\_up}(N) \cup \mathit{col\_down}(N) \cup \mathit{row\_down}(N)) = LSDP(N)$$

So we can conclude: $LSDP(N) \setminus \{C, R, D\} \Longrightarrow LSDP(N)$ meaning that $\{C, R, D\}$ is a redundant set of constraints in $LSDP(N)$. ■

The proof works because the six $LS(N)$ subproblems of the original $PDLS(N)$ problem are enough to leave always at least one of the subproblems with only one missing constraint.

Unfortunately, using this proof method, the result cannot be made stronger.

3 The limits of the Proof Method

Theorem 1 hinges on the identification of a sufficient number of $LS(N)$ subsets of $PDLS(N)$, and we have found some useful $LS(N)$ subsets already: *row_col(N)*, *col_up(N)*, *col_down(N)*, *row_up(N)*, *row_down(N)*, and *up_down(N)*. So it seems worthwhile to investigate the existence of other $LS(N)$ subsets of $PDLS(N)$, in the hope that a similar application of the proof renders different redundant sets of *all_different* constraints. But do other $LS(N)$ subsets of $PDLS(N)$ exist ? Lemma 1 prepares the ground for a negative answer to this question, Theorem 2 proves it.

One *PDLS(N)* is worth at most six *LS(N)* for $N > 5$

The above question boils down to finding subgraphs of the $PDLS(N)$ disequality graph that are isomorphic to the $LS(N)$ disequality graph. For the sake of the argument, we restrict ourselves to N for which $PDLS(N)$ has a solution.

The disequality graph PDG of $PDLS(N)$ consists of $4N$ N-cliques: no larger clique can exist, because otherwise there would be no solution. It is tempting to think that the only N-cliques are the $4N$ ones explicitly given, however, for $N = 5$, this is not true as shown in Figure 2: the dotted lines indicate a 3-clique, the full lines just one edge.

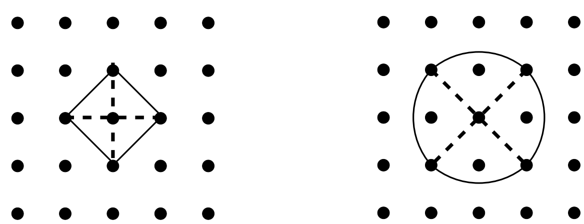

Figure 2: *Two new (symmetric) 5-cliques in* $PDLS(5)$

By an exhaustive search programmed in hProlog [3] we found out that $PDLS(7)$ contains no new 7-cliques, so we first formulated as a conjecture and later proved

Lemma 1 For $N > 5$, and such that $PDLS(N)$ has a solution, $PDLS(N)$ contains exactly $4N$ N-cliques.[2]

<u>Proof</u> The proof structure is as follows: suppose K is a clique, let Q be the maximum number of nodes K has in common with any row, column,

[2]Note that instead of $N > 5$, we could have written $N \geq 7$.

up or down diagonal. Because of symmetry considerations, we can restrict ourselves to the case that this is in row 1. We show that for any $Q < N$, K has size strictly smaller than N.

- let $N > Q \geq 4$: every node in another row has exactly 3 neighbors in row 1, through its column, up diagonal, and down diagonal; since $Q > 3$, the Q nodes in row 1 cannot have any common neighbor in any other row; therefore, K has size Q, which is strictly smaller than N

- let $Q = 3$: one can check that if N is not a multiple of 3, the 3 nodes in row 1 can have at most two common neighbors and that when they exist, they are in two different rows; for any N, this results in a 5-clique, as shown in Figure 2, but 5 is strictly smaller than N

- let $Q = 2$: we number the two nodes as 1 and 2; we use the notation *cu* as an abbreviation for the node where the column of node 1 intersects with the up diagonal of node 2, and likewise for other combinations of c, u and d; the common neighbors of the given nodes can thus be written as *cu*, *cd*, *uc*, *dc*, *ud*, *du*, and we have a total of 8 potential nodes in K; however, *cu* and *cd* are in the same column as node 1, so they cannot be both in K, since otherwise Q would be 3; likewise for *uc* and *dc*; so, K has at most 6 nodes, which is strictly less than N

- finally, if $Q = 1$, K cannot have 2 connected nodes, meaning K has size 1, and that is strictly smaller than N ∎

Let G be the disequality graph of an $LS(N)$-subset of $PDLS(N)$. G consists of $2N$ cliques of size N, and is a subgraph of PDG. As a consequence of Lemma 1, for $N > 5$, G contains only elements of *Rows(N)*, *Cols(N)*, *UpPDs(N)* and *DownPDs(N)*. Moreover, because of $LS(N)$, every node of G is in exactly two N-cliques, a property we will use later.

The following theorem essentially says that G contains either the whole of *Rows(N)*, or none of it - and similarly for the three other sets.

Theorem 2 Let G be an $LS(N)$-subset of $PDLS(N)$, and $N > 5$, then

$$G \cap \mathit{Rows(N)} \neq \emptyset \implies \mathit{Rows(N)} \subset G$$

and the same holds for *Cols(N)*, *UpPDs(N)* and *DownPDs(N)*.

<u>Proof</u> First some notation to refer to the cliques of $PDLS(N)$: for $1 \leq i \leq N$

- $row(i)$ and $col(i)$ refer to the i^{th} row and column
- $down(i)$ and $up(i)$ refer to the down (or up) pandiagonal going through the i^{th} cell in the first row.

Because of symmetry considerations, we can assume, without loss of generality, that $row(1)$ is in G. Since the first cell of the first row (let us name it $cell_1$) must be in two cliques in G, and because of the symmetry between up and down pandiagonals (at this point in the proof), we need to consider only the following two cases:

1. $\underline{col(1)\ is\ in\ G}$: then $up(1)$ neither $down(1)$ can be in G, otherwise $cell_1$ would be in more than two N-cliques; now suppose $up(j)$ is in G for a particular $j \neq 1$: $up(j)$ intersects $col(1)$ in the j^{th} cell (name it $cell_2$), and it intersects also $row(1)$ in the j^{th} cell (name it $cell_3$); the three nodes $cell_1$, $cell_2$ and $cell_3$ are in a 3-clique; but an $LS(N)$ disequality graph has the property that any three cells in a 3-clique, are also in an N-clique, name it C: C differs from $row(1)$ and $col(1)$, so now $cell_1$ is in $row(1)$, $col(1)$ and C, 3 N-cliques, which cannot be true in an $LS(N)$; so, no $up(j)$ can be in G, and by symmetry the is true for the $down(j)$; this leaves only all other rows and columns to make G, meaning that *Rows(N)* $\subset G$
2. $\underline{down(1)\ is\ in\ G}$: the argument is similar to the previous one

So, if $row(1) \in G$ then *Rows* $\in G$ or $G \cap$ *Rows(N)* $\neq \emptyset \implies$ *Rows(N)* $\subset G$. By symmetry this holds also for *Cols(N)*, *UpPDs(N)* and *DownPDs(N)*. ■

As a consequence of Lemma 2, we cannot hope to extend the reach of the proof method of Theorem 1, because there are no other $LS(N)$ subsets of $PDLS(N)$ besides the six used in Theorem 1.

The special case $N = 5$

$PDLS(5)$ has one hundred new 5-cliques. Each of them belongs to one of four types: Figure 2 shows representatives of two of these types. We expected that they could be combined – just as the sets *Rows(5)*, *Cols(5)*, *UpPDs(5)*, and *DownPDs(5)* – into $LS(5)$ subsets of $PDLS(5)$. A lucky first try gave us the pleasing combination in Figure 3:

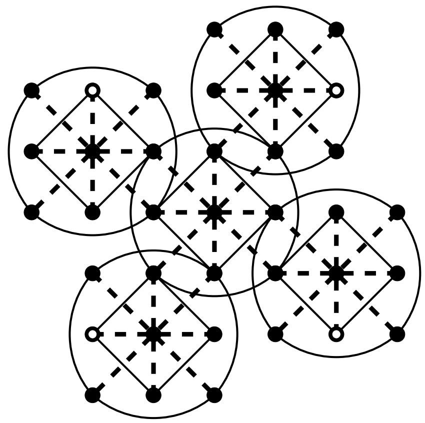

Figure 3: *A new $LS(5)$ in $PDLS(5)$*

Four nodes at the corners of the $PDLS(5)$ square in Figure 3 have been marked with a white dot. The nodes that are not in that square should be mapped to the inside in a cyclic way, while keeping their edges: each node is involved in two 5-cliques having a different shape, and two 5-cliques either have no nodes, or one node in common. It turns out that the four new types of 5-cliques can be composed two by two, to form siz different $LS(5)$ subsets of $PDLS(5)$, just enough to extend slightly the applicability of Theorem 1 for $N = 5$ in the obvious way.

It remains to prove whether Theorem 2 holds for $N = 5$, or in other words, whether the 20 original 5-cliques can be combined with the 100 new 5-cliques.

4 Discussion

We initially hoped that lifting the results from [2] to the $PDLS(N)$ problem would give an optimal result. Theorem 2 proves that we succeeded in the best possible lift, but the result is far from optimal, i.e. redundant sets with size larger than 3 exist for several N: this was experimentally established. It means that the naive composition of redundancy is weak in this case, even though it did the job completely for the Sudoku problem (see [1]). Clearly, more fundamental work is needed. [4] is a first start from the disequality graph perspective. Also, the question about the maximal size of redundant sets of *all_different* constraints in $PDLS(N)$ remains open.

Acknowledgement

We thank Michael Codish for bringing this problem to our attention.

References

[1] B. Demoen and M. Garcia de la Banda. Redundant Sudoku rules. *Theory and Practice of Logic Programming*, pages 1–15, Online 18 August 2013. doi: 10.1017/s1471068412000361.

[2] B. Demoen and M. Garcia de la Banda. Redundant disequalities in the Latin Square problem. *Constraints*, pages 1–8, Online 18 August 2013. doi: 10.1007/s1060101391471.

[3] B. Demoen and P.-L. Nguyen. So many WAM Variations, so little Time. In *Computational Logic, First International Conference*, volume 1861 of *LNAI*, pages 1240–1254. Springer, July 2000.

[4] B. Demoen and P.-L. Nguyen. The classification of graphs with a redundant inequality. CW Report 633, Department of Computer Science, KU Leuven, January 2013.

[5] L. Euler. Recherches sur une nouvelle espèce de carrés de magiques. *Verhandelingen uitgegeven door het zeeuwsch Genootschap der Wetenschappen te Vlissingen, Middelburg*, 9:85–239, 1782. URL `http://www.math.dartmouth.edu/~euler/pages/E530.html`.

[6] A. Hedayat. A complete solution to the existence and nonexistence of Knut Vik designs and orthogonal Knut Vik designs. *Journal of Combinatorial Theory, Series A*, 22(3):331 – 337, 1977.

[7] K. Vik. Bedømmelse av feilen på forsøksfelter med og uten malestokk. *Meldinger fra Norges Landbrukshøgskole*, 4:129–181, 1924.