

On the number of symbols that forces a transversal

Peter Keevash* Liana Yepremyan†

October 5, 2018

Abstract

Akbari and Alipour [1] conjectured that any Latin array of order n with at least $n^2/2$ symbols contains a transversal. For large n , we confirm this conjecture, and moreover, we show that $n^{399/200}$ symbols suffice.

1 Introduction

A Latin square of order n is an n by n square with cells filled using n symbols so that every symbol appears once in each row and once in each column. A transversal in a Latin square is a set of cells using every row, column and symbol exactly once. The study of transversals in Latin squares goes back to Euler in 1776; his famous ‘36 officers problem’ is equivalent to showing that there is no Latin square of order 6 that can be decomposed into 6 transversals (this was finally solved by Tarry in 1900). An even more fundamental question is whether a Latin square always has a transversal. A quick answer is ‘no’, as shown by the addition table of \mathbb{Z}_{2k} , but it remains open whether there is always a transversal when n is odd (a conjecture of Ryser [11]) or whether there is always a partial transversal size $n - 1$ (a conjecture of Brualdi [5] and Stein [12]). The best known positive result, due to Hatami and Shor [7], is that there is always a partial transversal size $n - O(\log^2 n)$.

Given the apparent difficulty of finding transversals in Latin squares, it is natural to ask if the problem becomes easier in Latin arrays with more symbols (now we fill a square with any number of symbols such that every symbol appears at most once in each row and at most once in each column). This problem was considered by Akbari and Alipour [1], who conjectured that any Latin array of order n with at least $n^2/2$ symbols contains a transversal. Progress towards this conjecture was independently obtained by Best, Hendrey, Wanless, Wilson and Wood [3] (who showed that $(2 - \sqrt{2})n^2$ symbols suffice) and Barát and Nagy [2] (who showed that $3n^2/4$ symbols suffice).

We will henceforth adopt the standard graph theory translation of the problem (see [4] for standard notation), where we consider a Latin array of order n as a properly edge-coloured complete bipartite graph $K_{n,n}$, with one part corresponding to rows, the other part to columns, and the colour of an edge is the symbol in the corresponding cell of the array. In this language, the Akbari–Alipour conjecture is that if there are at least $n^2/2$ colours then there is a rainbow perfect matching. Our main result confirms this conjecture for large n in a strong form.

Theorem 1.1. *Suppose the complete bipartite graph $K_{n,n}$ is properly edge-coloured using dn^2 colours, where n is sufficiently large and $d > n^{-1/200}$. Then there is a rainbow perfect matching.*

*Mathematical Institute, University of Oxford, Oxford, UK. E-mail: keevash@maths.ox.ac.uk.
Research supported in part by ERC Consolidator Grant 647678.

†Mathematical Institute, University of Oxford, Oxford, UK. E-mail: yepremyan@maths.ox.ac.uk.

Here the constant ‘200’ could be somewhat improved, but we have sacrificed some optimisations for the sake of readability of our proof, as in any case the best bound we can obtain seems far from optimal (it might even be true that $n^{1+o(1)}$ colours suffice!).

2 Proof

Here we give the proof of Theorem 1.1, assuming two lemmas that will be proved later in the paper. Consider the complete bipartite graph $K_{n,n}$ with parts A and B both of size n , and a proper edge-colouring using at least dn^2 colours, where n is sufficiently large and $d > n^{-1/200}$.

Choose uniformly random $A' \subseteq A$ and $B' \subseteq B$ with $|A'| = |B'| = n/2$ and let G be any subgraph of $K_{n,n}[A', B']$ obtained by including exactly one edge of each colour. By Chernoff bounds, with high probability (whp) $d(G) = e(G)(n/2)^{-2} \geq 0.99d$. We apply the following lemma, which will be proved in the next section, to find a pair (A_1, B_1) that satisfies Hall’s condition for a perfect matching ‘robustly’, so that it will still satisfy Hall’s condition after deleting small sets of vertices from each part. Note that as $d > n^{-1/200}$ we obtain $|A_1| = |B_1| > n^{0.7}/12$.

Lemma 2.1. *There is $G_1 = G[A_1, B_1]$ for some $A_1 \subseteq A'$ and $B_1 \subseteq B'$ with $|A_1| = |B_1| \geq d(G)^{60}|A'|/3$ such that G_1 has minimum degree at least $10^{-3}d|A_1|$, and for any $S \subseteq A_1$ or $S \subseteq B_1$ we have $|N_{G_1}(S)| \geq \min\{2|S|, 2|A_1|/3\}$.*

We define a random subgraph G^r of $K_{n,n}$ of ‘reserved colours’ as follows. Choose each colour independently with probability $p := n^{-0.32}$. Let G^r consist of all edges of all chosen colours. By Chernoff bounds, whp $|N_{G^r}(b) \cap A_1| = p|A_1| \pm (p|A_1|)^{2/3}$ and $|N_{G^r}(b) \setminus A_1| = p|A \setminus A_1| \pm (p|A \setminus A_1|)^{2/3}$ for all $b \in B$, and similarly with A and B interchanged. Let

$$G^* := (K_{n,n} \setminus G^r) \setminus (A_1 \cup B_1).$$

Then the minimum degree in G^* satisfies $\delta(G^*) \geq (1-p)(n - |A_1|) - (pn)^{2/3}$.

Let M_2^0 be a maximum size rainbow matching in G^* . Let $A_2^0 = V(M_2^0) \cap A$ and $B_2^0 = V(M_2^0) \cap B$. By a result of Gyarfas and Sarkozy [6, Theorem 2] we have $|A_2^0| = |B_2^0| \geq \delta(G^*) - 2\delta(G^*)^{2/3} \geq (1-2p)(n - |A_1|)$, as $pn = n^{0.68} \gg n^{2/3}$.

Next we construct G_1' from G_1 by deleting any edges using a colour used by M_2^0 and restricting to subsets $A_1' \subseteq A_1$, $B_1' \subseteq B_1$ with $|A_1'| = |B_1'| = |A_1| - n^{1/3}$ so that G_1' has minimum degree $\delta(G_1') \geq (10^{-3} - 2 \cdot 10^{-4})d|A_1|$. To see that this is possible, note that we delete at most n edges from G_1 , so each of A_1 and B_1 has at most $10^4n/d|A_1| \ll n^{1/3} \ll p|A_1| \ll d|A_1|$ vertices at which we delete more than $10^{-4}d|A_1|$ edges. Below we will define rainbow matchings M_2^i for $i \geq 1$ satisfying:

property (P_i) : $V(M_2^{i-1}) \subseteq V(M_2^i) \subseteq (A \setminus A_1') \cup (B \setminus B_1')$, $|M_2^i| = |M_2^{i-1}| + 1$ and M_2^i uses exactly two colours not used by M_2^{i-1} .

We regard (P_0) as being vacuously true. We will write $A_2^i = V(M_2^i) \cap A$, $B_2^i = V(M_2^i) \cap B$, $A_0^i = A \setminus (A_1' \cup A_2^i)$ and $B_0^i = B \setminus (B_1' \cup B_2^i)$. For all $i \geq 0$, let X_0^i be the set of vertices a in A_0^i such that at least $|B_1'|/2$ of the edges between a and B_1' have a colour used by M_2^i . Define Y_0^i similarly, interchanging A and B . We prove the following lemma in Section 4.

Lemma 2.2. *Given a rainbow matching M_2^i satisfying (P_i) at some step $i \geq 0$, either **(iterate)**: there exists a rainbow matching M_2^{i+1} satisfying (P_{i+1}) , or **(stop)**: both X_0^i and Y_0^i have size at most $p|A_1|/4$.*

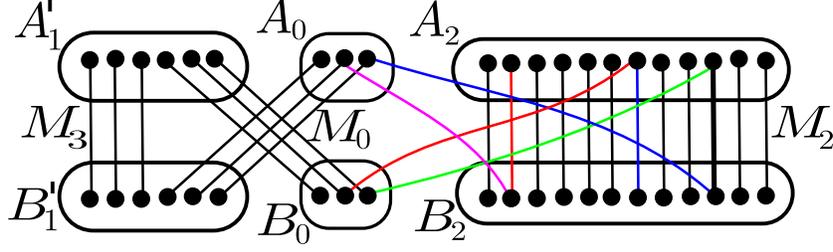


Figure 1: Proof by picture

We repeatedly apply Lemma 2.2 while condition **(iterate)** holds. After at most $|A_0|$ steps, condition **(stop)** must hold. We omit the superscript i in our notation for sets at the step where we stop; thus we stop at some rainbow matching M_2 where both X_0 and Y_0 have size at most $p|A_1|/4$. We write $A_2 = V(M_2) \cap A$, $B_2 = V(M_2) \cap B$, $A_0 = A \setminus (A'_1 \cup A_2)$ and $B_0 = B \setminus (B'_1 \cup B_2)$. The form of our intended rainbow matching is illustrated by the black and/or vertical edges in Figure 1 (the coloured diagonal edges illustrate the augmentation algorithm used in Section 4).

We will use a greedy algorithm to extend M_2 to a rainbow matching $M_2 \cup M_0$, where M_0 covers $A_0 \cup B_0$ and also uses some vertices of $A'_1 \cup B'_1$. We start by choosing these edges for vertices in $X_0 \cup Y_0$ using the reserved colours G^r . Recalling that $|A_1 \setminus A'_1| = n^{1/3} \ll p|A_1|$, the number of choices at each step is at least $p|A_1| - (p|A_1|)^{2/3} - n^{1/3} > 3p|A_1|/4$. Of these, the number forbidden by previous steps is at most $p|A_1|/4$ due to uses of vertices and at most $2 \cdot p|A_1|/4$ due to uses of colours, so this part of the algorithm can be completed.

Then we continue the greedy algorithm to choose edges covering the remainder of $A_0 \cup B_0$. To see that this is possible, we first note that $|A_0| = |B_0| \leq 2pn = 2n^{0.68} \ll d|A_1|$. At each step of the algorithm, by definition of X_0 and Y_0 there are at least $|A'_1|/2$ choices, of which at most $3|A_0|$ are forbidden due to a colour or a vertex used at a previous step, so the algorithm can be completed.

Finally, consider $G_3 = G[A_3, B_3]$ obtained from G'_1 by deleting all vertices covered by M_0 and all edges that share a colour with $(M_2 \setminus M_2^0) \cup M_0$. Then $|A_3| = |B_3| \geq |A'_1| - |A_0| > (1 - 10^{-3}d)|A_1|$. Recall that G'_1 has minimum degree $\delta(G'_1) \geq (10^{-3} - 2 \cdot 10^{-4})d|A_1|$. In passing to G_3 , any vertex loses at most $2|M_0| \leq 2n^{0.68} \ll d|A_1|$ edges due to vertex deletions, and at most $2|A_0| + |M_0| \leq 3n^{0.68} \ll d|A_1|$ edges due to colour deletions. We deduce $\delta(G_3) \geq 10^{-3}d|A_1|/2$.

We claim that G_3 has a perfect matching. To see this we check Hall's condition. Suppose for a contradiction there is $S \subseteq A_3$ with $|N(S)| < |S|$. By the minimum degree we have $|S| \geq 10^{-3}d|A_1|/2$. Now we cannot have $|S| \leq |A_1|/2$, as then $|N_{G_3}(S)| \geq \min\{2|S|, 2|A_1|/3\} - 2|A_0| > |S|$. However, letting $T = B_3 \setminus N(S)$ we have $N(T) \subseteq A_3 \setminus S$, so $|N(T)| = |A_3| - |S| < |B_3| - |N(S)| = |T|$. The same argument as for S gives $|T| > |A_1|/2$, contradiction. Therefore G_3 has a perfect matching M_3 . Now $M_2 \cup M_0 \cup M_3$ is a rainbow perfect matching in G , which completes the proof.

3 A robustly matchable pair

In this section we prove Lemma 2.1. Let G be a bipartite graph with parts A' and B' of size m and density $d(G) = e(G)m^{-2}$. We say G is (ε, δ) -dense if for any $A^* \subseteq A'$ and $B^* \subseteq B'$ with $|A^*| \geq \varepsilon m$ and $|B^*| \geq \varepsilon m$ we have $e_G(A^*, B^*) \geq \delta|A^*||B^*|$. We start by applying the following result of Peng,

Rödl and Ruciński [10, Theorem 1.3] with $\varepsilon = 1/10$, $c = 0.24$ and $c' = 1/50$.¹

Lemma 3.1. *Suppose $c, c' \in (0, 1)$ with $4c + c' \leq 1$. Then there are $A^* \subseteq A'$ and $B^* \subseteq B'$ with $|A^*| = |B^*| \geq d(G)^{2/\log_2(1+c\varepsilon)}m/2$ so that $G^* = G[A^*, B^*]$ is $(\varepsilon, c'd(G))$ -dense.*

Let $G_1 = G[A_1, B_1]$ be obtained from G^* by the following algorithm. Initially, $A_1 = A^*$ and $B_1 = B^*$. At any step of the algorithm, we update G_1 by deleting a vertex or set of vertices of one of the following types (choosing arbitrarily if there is a choice).

- i. $v \in A_1$ or $v \in B_1$ with $d_{G_1}(v) \leq 10^{-3}d(G)|A^*|$,
- ii. $S \subseteq A_1$ or $S \subseteq B_1$ with $|S| < \varepsilon|A^*|$ and $|N_{G_1}(S)| \leq 2|S|$.

Whenever we delete some vertices from A_1 or B_1 we delete an arbitrary set of the same size from the other, so that we always maintain $|A_1| = |B_1|$. We stop if no deletion is possible or if we have deleted at least $2\varepsilon|A^*|$ vertices from each side.

We claim that the latter option is impossible. Indeed, then without loss of generality we deleted $\varepsilon|A^*|$ vertices from A^* of type (i) or (ii) as above (at least half of the deleted vertices are deleted for a reason other than maintaining equal part sizes). Let $D^A = D_i^A \cup D_{ii}^A$ be the deleted vertices in A^* according to deletions of type (i) or (ii). Note that $|D^A| < 3\varepsilon|A^*|$, and $|N_{G_1}(D_{ii}^A)| \leq 2|D_{ii}^A| < 6\varepsilon|A^*|$. Let $B_0 = B_1 \setminus N_{G_1}(D_{ii}^A)$, so $|B_0| > (1 - 9\varepsilon)|A^*| = \varepsilon|A^*|$. Now $e_{G^*}(D^A, B_0) \leq |D_i^A| \cdot 10^{-3}d(G)|A^*| < \frac{d(G)}{50}|D^A||B_0|$ contradicts $(\varepsilon, d(G)/50)$ -density of G^* , which proves the claim.

Thus the algorithm stops with $|A_1| = |B_1| > (1 - 3\varepsilon)|A^*| \geq d(G)^{60}m/3$ (using $2/\log_2(1.024) < 60$), minimum degree at least $10^{-3}d(G)|A^*|$ and $|N_{G_1}(S)| \geq 2|S|$ for any $S \subseteq A_1$ or $S \subseteq B_1$ with $|S| < \varepsilon|A^*|$. Furthermore, for any $S \subseteq A_1$ or $S \subseteq B_1$ with $|S| \geq \varepsilon|A^*|$, by $(\varepsilon, d(G)/50)$ -density of G^* we have $|N_{G_1}(S)| \geq |B_1| - \varepsilon|A^*| \geq 2|A_1|/3$. This proves Lemma 2.1.

4 Augmentation algorithm

In this section we prove Lemma 2.2, which will complete the proof of Theorem 1.1. Let us drop the superscript i from $M_2^i, A_0^i, B_0^i, X_0^i$ and Y_0^i . We suppose condition **(stop)** of the lemma does not hold and show that condition **(iterate)** must hold. Without loss of generality, assume $|X_0| > p|A_1|/4$. We will iteratively construct $R = R^A \cup R^B \subseteq M_2$, where we think of R^A and R^B as ‘reachable’ from A_0 and B_0 . At some point R^A and R^B will intersect, which will allow us to extend M_2 . Let $\theta := n^{-0.66}$, and note that $\theta|A_1| > n^{0.04}/3$.

Algorithm 4.1. Let $R^A = R^B = \emptyset$ and let C be the set of colours not used by M_2 . At step $j \geq 1$:

- i. if $R^A \cap R^B \neq \emptyset$ stop, otherwise let R_j^A be the set of all $uv \in M_2$ where $v \in B_2 \setminus V(R^A)$ such that at least $\theta|A_1|$ edges in G^* from v to A_0 use a colour in C , let C_j^A be the set of colours used by R_j^A , update R^A by adding R_j^A and C by adding C_j^A ,
- ii. if $R^A \cap R^B \neq \emptyset$ stop, otherwise let R_j^B be the set of all $uv \in M_2$ where $u \in A_2 \setminus V(R^B)$ such that at least $\theta|A_1|$ edges in G^* from u to B_0 use a colour in C , let C_j^B be the set of colours used by R_j^B , update R^B by adding R_j^B and C by adding C_j^B .

Claim 4.2. $|R_1^A| \geq |A_1|/4$.

To see this, we consider the number X of edges in G^* with colour in C between X_0 and B_2 . We have $X \leq |R_1^A||X_0| + |B_2|\theta|A_1|$ by definition of R_1^A . Also, by definition of X_0 , every vertex in X_0 has

¹This result follows from their proof; they state the case $c = 1/8$, $c' = 1/2$ and use $\log_2(1 + \varepsilon/8) \geq \varepsilon/6$.

at least $(1-2p)|B_2| - (|M_2| - |B_1'|/2) \geq |A_1|/3$ edges in G^* to B_2 with colour in C , so $X \geq |X_0| \cdot |A_1|/3$. As $|X_0| \geq p|A_1|/4$ and $p^{-1}\theta n = n^{0.66} \ll pn \ll |A_1|$, we deduce $|R_1^A| \geq |A_1|/3 - |X_0|^{-1}|B_2|\theta|A_1| \geq |A_1|/3 - 4p^{-1}\theta n \geq |A_1|/4$, as claimed.

Claim 4.3. For $j \geq 1$, we have $|R_j^B| \geq |R^A| - 3pn$ and $|R_{j+1}^A| \geq |R^B| - 3pn$.

To see this, we first note that as $R^A \cap R^B = \emptyset$, any vertex in B_0 has at least $(1-2p)|A_2| - |R^B| - |M_2 \setminus (R^A \cup R^B)| = |R^A| - 2p|A_2|$ edges in G^* to $A_2 \setminus V(R^B)$ with colour in C . Double-counting such edges as in the previous claim gives $|B_0|(|R^A| - 2p|A_2|) \leq |R_j^B||B_0| + |B_2|\theta|A_1|$, so $|R_j^B| \geq |R^A| - 2p|A_2| - |B_0|^{-1}|B_2|\theta|A_1| \geq |R^A| - 3pn$. The proof of the second inequality is similar, so the claim holds.

Claim 4.4. The algorithm terminates at some step $j = j^+ < \log n$.

To see this, we show inductively that if $R^A \cap R^B = \emptyset$ at step j then $|R_j^B| \geq f(j)|A_1|/3$ and $|R_j^A| \geq (f(j) - 2^{-3})|A_1|/3$ where $f(j) = 2^{j-4} + 2^{-1}$. First note that $f(1) = 5/8$ and for $j \geq 2$ we have $\sum_{k=1}^{j-1} (f(k) - 2^{-3}) = 2^{-4}(2^j - 1) + (j-1)(2^{-1} - 2^{-3}) \geq f(j) - 3/16$. At step 1 we have $|R_1^A| \geq |A_1|/4 > (f(1) - 2^{-3})|A_1|/3$ and $|R_1^B| \geq |R^A| - 3pn > 0.21|A_1| > f(1)|A_1|/3$. Supposing the statement at step $j-1 \geq 1$, we have $|R_j^A| \geq (\sum_{k=1}^{j-1} |R_k^B|) - 3pn \geq (\sum_{k=1}^{j-1} f(k))|A_1|/3 - 3pn \geq (f(j) - 1/16)|A_1|/3 - 3pn \geq (f(j) - 2^{-3})|A_1|/3$ and $|R_j^B| \geq (\sum_{k=1}^j |R_k^A|) - 3pn \geq (\sum_{k=1}^j f(k) - 2^{-3})|A_1|/3 - 3pn \geq (f(j+1) - 3/16)|A_1|/3 - 3pn \geq f(j)|A_1|/3$. Thus the required bounds hold by induction. While $R^A \cap R^B = \emptyset$ we deduce $(2f(j) - 2^{-3})|A_1|/3 < |M_2| < n$, so $j^+ < \log n$, as claimed.

The algorithm terminates by finding some edge $ab \in R^A \cap R^B$ where $a \in A_2$ and $b \in B_2$. Given two colours c and c' in C , we say that c is earlier than c' if c was added to C before c' . We start by applying the definition of R^A and R^B to find edges a_0b and ab_0 of G^* with $a_0 \in A_0$ and $b_0 \in B_0$ where the colours of a_0b and ab_0 are distinct, in C and earlier than that of ab . We modify M_2 to obtain M_2' by deleting ab and adding a_0b and ab_0 . Thus we obtain a larger matching, but M_2' may not be rainbow, due to repeating the colours of a_0b and ab_0 . While the current matching M_2' is not rainbow, we apply the following ‘trace back’ algorithm (similar to that of [6]).

Algorithm 4.5. At step $j \geq 1$ we have at most two ‘active’ edges, which are edges of M_2' having some colour in C shared with some edge that is still present from M_2 . At step 1 these are a_0b and ab_0 . If there is an active edge at step j , we choose one arbitrarily, call it a_jb_j , and let $a_j'b_j'$ be the edge of M_2 of the same colour $c \in C$. By construction of C , one of a_j' or b_j' , say a_j' , has at least $\theta|A_1| - 1$ edges to B_0 or A_0 using an earlier colour than c in C distinct from the colour of the second active edge (if it exists). We modify M_2' by deleting $a_j'b_j'$ and adding some such edge $a_j'b_0^j$ where $b_0^j \in B_0$ is distinct from all previous choices. We say that a_jb_j is no longer active. We make $a_j'b_0^j$ active if its colour is shared with some edge that is still present from M_2 .

Algorithm 4.5 is illustrated in Figure 1: the thick black edge represents the edge $ab \in R^A \cap R^B$, at step 1 the green and blue diagonals are active, at step 2 the blue diagonal is active, at step 3 the red diagonal is active, at step 4 the pink diagonal is active, at step 5 there are no active edges so the algorithm terminates. To see that the algorithm succeeds, note that there are at most $4 \log n$ steps of replacing an active edge by another, and each choice has at least $\theta|A_1| > n^{0.04}/3 > 4 \log n$ options. Thus we obtain a rainbow matching $M_2^{i+1} = M_2'$ in G^* satisfying property (P_{i+1}) . This proves Lemma 2.2, and so completes the proof of our theorem.

Acknowledgement. We thank an anonymous referee for a very careful reading of our paper and pointing out some significant corrections.

Postscript. The Akbari–Alipour conjecture for large n was proved independently and simultaneously by Montgomery, Pokrovskiy and Sudakov [9]. Our proof is much simpler than theirs, and gives a better bound on the number of symbols required, whereas their proof applies in a much more general setting, and so has several further applications. Results similar to those in [9] (but not including the Akbari–Alipour conjecture) were obtained by Kim, Kühn, Kupavskii and Osthus [8].

References

- [1] S. Akbari and A. Alipour, Transversals and multicolored matchings, *J. Combin. Designs* 12:325–332, 2004.
- [2] J. Barát and Z.L. Nagy, Transversals in generalized Latin squares, arXiv:1701.08220.
- [3] D. Best, K. Hendrey, I.M. Wanless, T.E. Wilson and D. Wood, Transversals in Latin arrays with many distinct symbols, arXiv:1612.09443.
- [4] B. Bollobás, *Modern graph theory*, Springer-Verlag, 1998.
- [5] R.A. Brualdi and H.J. Ryser, Combinatorial matrix theory, *Cambridge University Press*, 1991.
- [6] A. Gyárfás and G. Sárközy, Rainbow matchings and partial transversals of Latin squares, arXiv:1208.5670.
- [7] P. Hatami and P.W. Shor, A lower bound for the length of a partial transversal in a Latin square, *J. Combin. Theory Ser. A* 115:1103–1113, 2008.
- [8] J. Kim, D. Kühn, A. Kupavskii and D. Osthus, Rainbow structures in locally bounded colourings of graphs, arXiv:1805.08424.
- [9] R. Montgomery, A. Pokrovskiy and B. Sudakov, Decompositions into spanning rainbow structures, arXiv:1805.07564.
- [10] Y. Peng, V. Rödl and A. Ruciński, Holes in graphs, *Electron. J. Combin.* 9:R1, 2002.
- [11] H. Ryser, Neuere Probleme in der Kombinatorik, *Vorträge über Kombinatorik*, Oberwolfach, pp. 69–91, 1967.
- [12] S. K. Stein, Transversals of Latin squares and their generalizations, *Pacific J. Math.* 59:567–575, 1975.