MATHEMATICS OF COMPUTATION
Volume 78, Number 267, July 2009, Pages 1733–1747
S 0025-5718(09)02210-8
Article electronically published on January 23, 2009

RECONSTRUCTION OF MATRICES FROM SUBMATRICES

GÉZA KÓS, PÉTER LIGETI, AND PÉTER SZIKLAI

ABSTRACT. For an arbitrary matrix A of $n \times n$ symbols, consider its submatrices of size $k \times k$, obtained by deleting $n-k$ rows and $n-k$ columns. Optionally, the deleted rows and columns can be selected symmetrically or independently. We consider the problem of whether these multisets determine matrix A.

Following the ideas of Krasikov and Roditty in the reconstruction of sequences from subsequences, we replace the multiset by the sum of submatrices. For $k > cn^{2/3}$ we prove that the matrix A is determined by the sum of the $k \times k$ submatrices, both in the symmetric and in the nonsymmetric cases.

1. Introduction

1.1. **Problem statement.** Let Σ be an alphabet, and denote by $\Sigma^{n\times n}$ the set of $n \times n$ matrices over Σ. Call a matrix $B \in \Sigma^{k\times k}$ a *submatrix* of $A \in \Sigma^{n\times n}$ if B can be obtained by deleting $n-k$ rows and $n-k$ columns of A. If we delete rows and columns symmetrically, then B is a *principal submatrix* of A.

Denote by $M_k(A)$ and $M_k^{\mathrm{sym}}(A)$ the multisets of the $\binom{n}{k}^2$ submatrices and the $\binom{n}{k}$ principal submatrices of A of size $k \times k$, respectively. We consider the following two questions.

Problem 1. For a given n, what is the smallest k such that every $A \in \Sigma^{n\times n}$ is uniquely determined by $M_k(A)$, i.e. the map M_k is injective on $\Sigma^{n\times n}$?

Problem 2. For a given n, what is the smallest k such that every $A \in \Sigma^{n\times n}$ is uniquely determined by $M_k^{\mathrm{sym}}(A)$, i.e. the map M_k^{sym} is injective on $\Sigma^{n\times n}$?

Notice that if $k < \ell$, then $\biguplus_{B\in M_\ell(A)} M_k(B) = \binom{n-k}{\ell-k}^2 \cdot M_k(A)$ and therefore $M_\ell(A)$ determines $M_k(A)$. (Here $\biguplus$ is the multiset union symbol.) So, if M_k is injective, then M_ℓ also must be injective as well. Similarly, we can show that $M_\ell^{\mathrm{sym}}(A)$ determines $M_k^{\mathrm{sym}}(A)$. Hence, it is sufficient to find the smallest such k in both cases.

Beyond their theoretical interest, Problems 1 and 2 have a connection with the (vertex) graph reconstruction problem of Kelly [5] and Ulam [11]. For $\{0,1\}$ matrices, the two variants of the matrix reconstruction problem (in the symmetric and nonsymmetric case) are equivalent to the vertex reconstruction problems of ordered ordinary and bipartite graphs, respectively.

Received by the editor February 15, 2008 and, in revised form, August 8, 2008.

2000 *Mathematics Subject Classification.* Primary 05B20; Secondary 11B83.

The first and the third authors were supported in part by the Bolyai Grant of the Hungarian Academy of Sciences.

The third author was partially supported by the OTKA T-67867 grant.

Marcus and Tardos [8], Tardos [10], Pach and Tardos [9] also settled a series of conjectures and gave new proofs for related problems on 0-1 matrices and ordered graphs.

1.2. **Previous work.** The one-dimensional analogue of our problems is the reconstruction of sequences of length n from the multiset of subsequences of length k. The problem was raised first in an information-theoretic study of Kalashnik [4] about noisy deletion channels in which characters of a transmitted sequence are randomly (but not necessarily independently) omitted.

The best known lower bound is due to Dudik and Schulman [2] who proved that if $k < e^{c\sqrt{\log n}}$, then there exist distinct 0-1 sequences having the same multiset of the $\binom{n}{k}$ subsequences.

The best upper bounds are based on the ideas of Krasikov and Roditty [6]. Assuming $\Sigma = \{0,1\}$ — which can be done without loss of generality — they considered the coordinatewise *sum* of the subsequences of length k. Suppose that $(a_0, a_1, \ldots, a_{n-1})$ and $(b_0, b_1, \ldots, b_{n-1})$ are distinct 0-1 sequences such that their subsequences of length k give the same sum, and let $d_i = a_i - b_i$. Krasikov and Roditty showed that for every polynomial $p(x)$ with $\deg p < k$,

$$\sum_{i=0}^{n-1} p(i) \cdot d_i = 0. \tag{1.1}$$

(This observation links the problem to the famous Prouhet-Tarry-Escott problem as well.)

In order to obtain an upper bound, Krasikov and Roditty combined this fact with a result of Borwein, Erdélyi and Kós [1]: for every positive integer n, there exists a polynomial $p(x)$ such that $\deg p < \left(\frac{16}{7} + \varepsilon\right)\sqrt{n}$ and

$$p(0) > |p(1)| + |p(2)| + \ldots + |p(n)|. \tag{1.2}$$

Later, in [3], Foster and Krasikov showed that the constant $\frac{16}{7} \approx 2.286$ can be replaced by $2\sqrt{\log 2} \approx 1.665$.

It is easy to see that the relations (1.1) and (1.2) are mutually exclusive, as by permuting the sequences one can assume that $d_0 \neq 0$. Hence, if

$$k > \left(2\sqrt{\log 2} + \varepsilon\right)\sqrt{n},$$

then every 0-1 sequence of length n is determined by the sum of its $\binom{n}{k}$ subsequences of length k.

Contrary to the case of sequences, the reconstruction problem of matrices has not been extensively studied. We refer only to a result by Manvel and Stockmeyer [7], who proved that for $n \geq 5$, every matrix A of size $n \times n$ is reconstructible from $M_{n-1}^{\mathrm{sym}}(A)$.

1.3. **New results on matrices.** The answers in Problems 1 and 2 are obviously the same for all alphabets consisting of at least two symbols. From now on we assume, without loss of generality, that $\Sigma = \{0,1\}$.

The lower bound by Dudik and Schulman can be applied to matrices as well. Suppose that the sequences $(a_1,\dots,a_n)$ and $(b_1,\dots,b_n)$ have the same $\binom{n}{k}$ subsequences, and consider

$$A=\begin{pmatrix} a_1 & a_2 & \dots & a_n\\ a_1 & a_2 & \dots & a_n\\ \vdots & \vdots & \ddots & \vdots\\ a_1 & a_2 & \dots & a_n\end{pmatrix} \quad\text{and}\quad B=\begin{pmatrix} b_1 & b_2 & \dots & b_n\\ b_1 & b_2 & \dots & b_n\\ \vdots & \vdots & \ddots & \vdots\\ b_1 & b_2 & \dots & b_n\end{pmatrix}.$$

These matrices obviously satisfy $M_k(A)=M_k(B)$ and $M_k^{\mathrm{sym}}(A)=M_k^{\mathrm{sym}}(B)$. Therefore, the smallest values of k in Problems 1 and 2 are greater than $e^{c\sqrt{\log n}}$.

In this paper we focus on the upper bound and generalize the ideas of Krasikov and Roditty. For an arbitrary matrix $A\in\{0,1\}^{n\times n}$, define the sums of submatrices (with multiplicities) of A as

$$S_k(A)=\sum_{B\in M_k(A)} B \quad\text{and}\quad S_k^{\mathrm{sym}}(A)=\sum_{B\in M_k^{\mathrm{sym}}(A)} B.$$

Replacing in Problems 1 and 2 the maps M_k and M_k^{sym} by S_k and S_k^{sym}, respectively, we ask the following simplified questions as well.

Problem 3. For a given n, what is the smallest k such that every $A\in\{0,1\}^{n\times n}$ is uniquely determined by $S_k(A)$, i.e. the map S_k is injective on $\{0,1\}^{n\times n}$?

Problem 4. For a given n, what is the smallest k such that every $A\in\{0,1\}^{n\times n}$ is uniquely determined by $S_k^{\mathrm{sym}}(A)$, i.e. the map S_k^{sym} is injective on $\{0,1\}^{n\times n}$?

Similarly to Problems 1 and 2, it is sufficient to ask the smallest possible values of k, since $S_k(A)=\frac{S_k\left(S_\ell(A)\right)}{\binom{n-k}{\ell-k}^2}$ and $S_k^{\mathrm{sym}}(A)=\frac{S_k^{\mathrm{sym}}\left(S_\ell^{\mathrm{sym}}(A)\right)}{\binom{n-k}{\ell-k}}$ for $k<\ell$.

In Section 2 we prove the analogues of equation (1.1) for matrices. Then, in Section 3 we prove the following results.

Result 1 (Theorem 3.1)**.** (a) If $k<\frac{n^{2/3}}{\sqrt[3]{2\log_2(n+1)}}$, then the map S_k is not injective on $\{0,1\}^{n\times n}$.

(b) If $k<\frac{n^{2/3}}{\sqrt[3]{\log_2(n+1)}}$, then the map S_k^{sym} is not injective on $\{0,1\}^{n\times n}$.

Result 2 (Theorem 3.2)**.** If n is sufficiently large and $k>38n^{2/3}$, then both S_k and S_k^{sym} are injective on $\{0,1\}^{n\times n}$.

From Result 2 we immediately obtain the following corollary:

Result 3. If n is sufficiently large and $k>38n^{2/3}$, then both M_k and M_k^{sym} are injective, and every matrix $A\in\Sigma^{n\times n}$ is uniquely determined by $M_k(A)$ as well as by $M_k^{\mathrm{sym}}(A)$.

The main tool, which is the analogue of relation (1.2), is proved in Section 3.3.

2. Rephrasing the reconstruction problem

In this section we generalize equation (1.1) for matrices. The generalizations are different for the symmetric and nonsymmetric cases.

2.1. **The nonsymmetric case.** In the case of nonsymmetric deletion, we prove the following fact.

Lemma 2.1. *Let $A, B \in \{0,1\}^{n\times n}$ be two arbitrary matrices and let $D = A - B = (d_{ij})_{1\le i,j\le n}$ be their difference. The following two statements are equivalent:*

(a) *$S_k(A) = S_k(B)$;*

(b) *if $p(x,y)$ is an arbitrary polynomial with real coefficients such that $\deg_x p < k$ and $\deg_y p < k$, then*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p(i,j)\cdot d_{ij} = 0. \tag{2.1}$$

The proof is a combination of the following two observations.

Lemma 2.2. *Define the polynomials $\beta_u(x) = \binom{x-1}{u-1}\binom{n-x}{k-u}$ for $1 \le u \le k$. Let $A, B \in \{0,1\}^{n\times n}$ be arbitrary matrices and let $D = A - B = (d_{ij})_{1\le i,j\le n}$. Then for each $1 \le u, v \le k$, the (u,v)th entry in the matrix $S_k(A) - S_k(B) = S_k(D)$ can be expressed as*

$$\big(S_k(D)\big)_{uv} = \sum_{i=1}^{n}\sum_{j=1}^{n}\beta_u(i)\beta_v(j)d_{ij}.$$

Proof.

$$\big(S_k(D)\big)_{uv} = \sum_{1\le i_1<i_2<\ldots<i_k\le n}\ \sum_{1\le j_1<j_2<\ldots\le j_k\le n}\begin{pmatrix} d_{i_1j_1} & \ldots & d_{i_1j_k}\\ \vdots & \ddots & \vdots\\ d_{i_kj_1} & \ldots & d_{i_kj_k}\end{pmatrix}_{uv}$$

$$= \sum_{i_u=1}^{n}\sum_{j_v=1}^{n}\ \sum_{i_1<\ldots<i_{u-1}<i_u}\ \sum_{i_u<i_{u+1}<\ldots<i_k\le n}\ \sum_{j_1<\ldots<j_{v-1}<j_v}\ \sum_{j_v<j_{v+1}<\ldots<j_k\le n} d_{i_uj_v}$$

$$= \sum_{i_u=1}^{n}\sum_{j_v=1}^{n}\binom{i_u-1}{u-1}\binom{n-i_u}{k-u}\binom{j_v-1}{v-1}\binom{n-j_v}{k-v}d_{i_uj_v} = \sum_{i=1}^{n}\sum_{j=1}^{n}\beta_u(i)\beta_v(j)d_{ij}.$$

□

Lemma 2.3. (*i*) *The polynomials $\beta_u(x)$ $(1 \le u \le k)$ form a basis of the linear space of all polynomials with degree less than k.*

(*ii*) *The polynomials $\beta_u(x)\beta_v(y)$ $(1 \le u, v \le k)$ form a basis of the linear space of all polynomials in two variables which have degree less than k in each variable.*

Remark. The first statement was also proved and used in [6].

Proof. (*i*) The number of polynomials $\beta_u(x)$ is k which matches the dimension of the linear space of polynomials with degree less than k. So it is sufficient to prove that polynomials $\beta_u(x)$ are linearly independent. Suppose that λ_u $(1 \le u \le n)$ are real numbers, not all zero. We have to show that

$$\sum_{u=1}^{k}\lambda_u\beta_u(x) \ne 0.$$

Let u_0 be the first index for which $\lambda_{u_0} \neq 0$. Substituting $x = u_0$, we have $\lambda_u = 0$ for $u < u_0$ and $\beta_u(u_0) = 0$ for $u > u_0$. Hence,

$$\sum_{u=1}^{k} \lambda_u \beta_u(u_0) = \lambda_{u_0} \beta_{u_0}(u_0) = \lambda_{u_0} \binom{n-u_0}{k-u_0} \neq 0.$$

Statement (ii) follows from statement (i). □

Proof of Lemma 2.1. (b) $\Rightarrow$ (a). For each pair $1 \leq u, v \leq k$ of indices, apply Lemma 2.2 and property (b) on the polynomial $p_{uv}(x,y) = \beta_u(x)\beta_v(y)$. Since the degree of p_{uv} is less than k in each variable,

$$\big(S_k(D)\big)_{uv} = \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_u(i)\beta_v(j) d_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{uv}(i,j) \cdot d_{ij} = 0.$$

This holds for each pair (u,v), so $S_k(D) = 0$ and $S_k(A) = S_k(B)$.

(a) $\Rightarrow$ (b). Let $p(x,y)$ be an arbitrary polynomial with $\deg_x p, \deg_y p < k$. By Lemma 2.3, there exist real numbers $\lambda_{u,v}$ $(1 \leq u, v \leq k)$ such that

$$p(x,y) = \sum_{u=1}^{k}\sum_{v=1}^{k} \lambda_{u,v} \beta_u(x)\beta_v(y).$$

Then, applying Lemma 2.2,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p(i,j) \cdot d_{i,j} = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\sum_{u=1}^{k}\sum_{v=1}^{k} \lambda_{u,v}\beta_u(i)\beta_v(j)\right) d_{i,j}$$
$$= \sum_{u=1}^{k}\sum_{v=1}^{k} \lambda_{u,v} \left(\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_u(i)\beta_v(j) d_{i,j}\right) = \sum_{u=1}^{k}\sum_{v=1}^{k} \lambda_{u,v} \cdot 0 = 0.$$

□

2.2. **The symmetric case.** In the symmetric case, the diagonal, the upper triangle and the lower triangle of the matrix $S_k^{\mathrm{sym}}(A)$ are determined only by the diagonal, the upper and lower triangle of A. For this reason, instead of a single family of equations like (2.1), we have three distinct families for the elements in, above and below the diagonal, respectively.

Lemma 2.4. *For arbitrary matrices* $A, B \in \{0,1\}^{n\times n}$ *and* $D = A - B = (d_{ij})_{1\leq i,j\leq n}$, *the following two statements are equivalent:*

(a) $S_k^{\mathrm{sym}}(A) = S_k^{\mathrm{sym}}(B)$*;*

(b) *for arbitrary polynomials* $p(x,y)$ *and* $q(x)$ *with degrees* $\deg p < k-1$ *and* $\deg q < k$*,*

$$\sum_{1\leq i<j\leq n} p(i,j) \cdot d_{ij} = 0, \qquad \sum_{1\leq j<i\leq n} p(i,j) \cdot d_{ij} = 0, \quad \textit{and} \quad \sum_{j=1}^{n} q(i) \cdot d_{ii} = 0.$$

In order to obtain a simpler necessary condition, we can replace these three families by a single one.

Corollary 2.5. *Let* $A, B \in \{0,1\}^{n\times n}$ *and* $D = A - B = (d_{ij})_{1\leq i,j\leq n}$. *If* $S_k^{\mathrm{sym}}(A) = S_k^{\mathrm{sym}}(B)$. *Then*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p(i,j) \cdot d_{ij} = 0$$

for every polynomial $p(x,y)$ *with total degree* $\deg p < k-1$.

Proof. Let $p(x,y)$ be an arbitrary polynomial with $\deg p < k-1$ and let $q(x) = p(x,x)$. Then $\deg q \le \deg p < k$ and

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p(i,j)\cdot d_{ij} = \sum_{i<j} p(i,j)\cdot d_{ij} + \sum_{i=j} q(i)\cdot d_{ii} + \sum_{j>i} p(i,j)\cdot d_{ij} = 0+0+0.$$

□

For proving Lemma 2.4, we show the analogues of Lemmas 2.2 and 2.3.

Lemma 2.6. *For every pair* $1 \le u < v \le k$, *define*

$$\gamma_{uv}(x,y) = \binom{x-1}{u-1}\binom{y-x-1}{v-u-1}\binom{n-y}{k-v}.$$

If $A, B \in \{0,1\}^{n\times n}$ *are arbitrary matrices and* $D = A - B = (d_{ij})_{1\le i,j\le n}$, *then for each* $1 \le u, v \le k$, *the* (u,v)th *entry of* $S_k^{\mathrm{sym}}(D)$ *is*

$$\left(S_k^{\mathrm{sym}}(D)\right)_{uv} = \begin{cases} \displaystyle\sum_{i<j} \gamma_{uv}(i,j)\cdot d_{ij} & \text{if } u<v; \\ \displaystyle\sum_{i=1}^{n} \beta_u(i)\cdot d_{ii} & \text{if } u=v; \\ \displaystyle\sum_{j<i} \gamma_{vu}(j,i)\cdot d_{ij} & \text{if } u>v. \end{cases}$$

Proof. If $u = v$, then

$$\begin{aligned}\left(S_k^{\mathrm{sym}}(D)\right)_{uv} &= \sum_{1\le i_1<i_2<\ldots<i_k\le n} \begin{pmatrix} d_{i_1i_1} & \ldots & d_{i_1i_k} \\ \vdots & \ddots & \vdots \\ d_{i_ki_1} & \ldots & d_{i_ki_k} \end{pmatrix}_{uu} \\ &= \sum_{i_u=1}^{n} \sum_{i_1<\ldots<i_{u-1}<i_u} \sum_{i_u<i_{u+1}<\ldots<i_k} d_{i_ui_u} \\ &= \sum_{i_u=1}^{n} \binom{i_u-1}{u-1}\binom{n-i_u}{k-u} d_{i_ui_u} = \sum_{i=1}^{n} \beta_u(i) d_{ii}.\end{aligned}$$

If $u < v$, then

$$\begin{aligned}\left(S_k^{\mathrm{sym}}(D)\right)_{uv} &= \sum_{1\le i_1<i_2<\ldots<i_k\le n} \begin{pmatrix} d_{i_1i_1} & \ldots & d_{i_1i_k} \\ \vdots & \ddots & \vdots \\ d_{i_ki_1} & \ldots & d_{i_ki_k} \end{pmatrix}_{uv} \\ &= \sum_{i_u<i_v} \sum_{1\le i_1<\ldots<i_{u-1}<i_u} \sum_{i_u<i_{u+1}<\ldots<i_{v-1}<i_v} \sum_{i_v<i_{v+1}<\ldots<i_k\le n} d_{i_ui_v} \\ &= \sum_{i_u<i_v} \binom{i_u-1}{u-1}\binom{i_v-i_u-1}{v-u-1}\binom{n-i_v}{k-v} d_{i_ui_v} = \sum_{i<j} \gamma_{uv}(i,j)\cdot d_{ij}.\end{aligned}$$

The case $u > v$ can be proved similarly. □

Lemma 2.7. *The polynomials* $\gamma_{uv}(x,y)$ $(1 \le u < v \le k)$ *form a basis of the linear space of polynomials in two variables with total degree less than* $k-1$.

Proof. Again, the number of polynomials matches the dimension which is $\frac{1}{2}k(k-1)$, so it is sufficient to prove linear independence.

Let λ_{uv} $(1 \le u < v \le k)$ be real numbers, not all zero; we have to show that

$$\sum_{1\le u<v\le k} \lambda_{uv}\gamma_{u,v}(x,y) \neq 0.$$

Let (u_0, v_0) be the first pair of indices in lexicographical order, for which $\lambda_{u_0,v_0} \neq 0$. This means that $\lambda_{uv} = 0$ in every case when $u < u_0$, or $u = u_0$ and $v < v_0$. Substituting $x = u_0$, $y = v_0$, we have $\gamma_{u,v}(u_0, v_0) = 0$ for $u > u_0$ and $v-u > v_0-u_0$. The only case when $\lambda_{uv}\gamma_{uv}(u_0, v_0) \neq 0$ is $u = u_0$, $v = v_0$. Therefore,

$$\sum_{1\le u<v\le k} \lambda_{uv}\gamma_{uv}(u_0,v_0) = \lambda_{u_0v_0}\gamma_{u_0v_0}(u_0,v_0) = \lambda_{u_0v_0}\binom{n-v_0}{k-v_0} \neq 0. \qquad \square$$

Proof of Lemma 2.4. Similarly to the nonsymmetric case, Lemma 2.4 follows from Lemma 2.6 and Lemma 2.7. □

3. Proofs of the results

3.1. **Lower bounds.** By simple applications of the pigeonhole principle, we can obtain lower bounds for the smallest values of k in Problems 3 and 4.

Theorem 3.1 (Result 1)**.** (*i*) *If* $k < \frac{n^{2/3}}{\sqrt[3]{2\log_2(n+1)}}$, *then there exist matrices* $A, B \subset \{0,1\}^{n\times n}$ *such that* $A \neq B$ *but* $S_k(A) = S_k(B)$.

(*ii*) *If* $k < \frac{n^{2/3}}{\sqrt[3]{\log_2(n+1)}}$, *then there exist matrices* $A, B \subset \{0,1\}^{n\times n}$ *such that* $A \neq B$ *but* $S_k^{\mathrm{sym}}(A) = S_k^{\mathrm{sym}}(B)$.

Proof. (*i*) For an arbitrary matrix $A \in \{0,1\}^{n\times n}$, $S_k(A)$ is the sum of $\binom{n}{k}^2$ submatrices, so each entry in $S_k(A)$ is a nonnegative integer, not exceeding $\binom{n}{k}^2$. Hence,

$$\left|\left\{S_k(A):\ A \in \{0,1\}^{n\times n}\right\}\right| \le \left(\binom{n}{k}^2 + 1\right)^{k^2}$$
$$\le (n^{2k}+1)^{k^2} \le (n+1)^{2k^3} < 2^{n^2} = \left|\{0,1\}^{n\times n}\right|.$$

There are fewer possible values of $S_k(A)$ than 0-1 matrices, so the map S_k cannot be injective.

(*ii*) Similarly to the nonsymmetric case, each entry of $S_k^{\mathrm{sym}}(A)$ is at most $\binom{n}{k}$ and therefore

$$\left|\left\{S_k^{\mathrm{sym}}(A):\ A \in \{0,1\}^{n\times n}\right\}\right| \le \left(\binom{n}{k} + 1\right)^{k^2} \le (n+1)^{k^3} < 2^{n^2} = \left|\{0,1\}^{n\times n}\right|.$$

□

Remark. With more careful computation the conditions can be improved to $k < \left(\sqrt[3]{\frac{3}{2}} - \varepsilon\right)\frac{n^{2/3}}{\sqrt[3]{\log_2 n}}$ and $k < \left(\sqrt[3]{3} - \varepsilon\right)\frac{n^{2/3}}{\sqrt[3]{\log_2 n}}$, respectively, without any change in the order of magnitude.

3.2. **Upper bounds.** As mentioned in the Introduction, we prove the following upper bound for the smallest k for which the maps S_k and S_k^{sym} are injective.

Theorem 3.2 (Result 2). *If n is sufficiently large, $k \geq 38n^{2/3}$, and $A, B \in \{0,1\}^{n\times n}$ are two distinct matrices, then $S_k(A) \neq S_k(B)$, and $S_k^{\mathrm{sym}}(A) \neq S_k^{\mathrm{sym}}(B)$.*

The main tool for proving the theorem is the following result.

Lemma 3.3. *For sufficiently large n, for an arbitrary nonempty set $H \subset \{1, 2, \dots, n\}^2$ there exists a point $\mathbf{a} = (a_1, a_2) \in H$ and a polynomial $p(x, y)$ such that $\deg p < 37.5n^{2/3}$ and*

$$p(a_1, a_2) > \sum_{(x,y)\in H,\ (x,y)\neq(a_1,a_2)} |p(x, y)|. \tag{3.1}$$

We prove this lemma in Section 3.3.

Proof of Theorem 3.2. Let $D = A - B = (d_{ij})_{1\leq i,j\leq n}$ and apply Lemma 3.3 on the set $H = \big\{(i, j) \in \{1, 2, \dots, n\}^2 :\ d_{ij} \neq 0\big\}$. By the lemma, there exists a point $(a_1, a_2) \in H$ and a polynomial $p(x, y)$ such that $\deg p < k - 1$ and relation (3.1) holds. Then

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{n} p(i, j)d_{ij}\right| = \left|\sum_{(i,j)\in H} p(i, j)d_{ij}\right| \geq p(a_1, a_2) - \sum_{(x,y)\in H\setminus\{(a_1,a_2)\}} |p(x, y)| > 0.$$

Hence, $\sum_{i=1}^{n}\sum_{j=1}^{n} p(i, j)d_{ij} \neq 0$. By Lemmas 2.1 and 2.4 this implies $S_k(A) \neq S_k(B)$ and $S_k^{\mathrm{sym}}(A) \neq S_k^{\mathrm{sym}}(B)$, respectively. □

3.3. **Proof of Lemma 3.3.**

Lemma 3.4. *For arbitrary real numbers $A, M > 0$ there exists a polynomial $f(x)$ with real coefficients with the following properties:*

(a) $f(0) = M$,

(b) $|f(x)| \leq \min\left(M, \dfrac{1}{x^2}\right)$ *for all* $x \in (0, A]$ *and*

(c) $\deg f < \sqrt{\pi}\sqrt{A}\sqrt[4]{M} + 2$.

Remark. This lemma and this polynomial come from a previous paper [1], but the proof has been arranged in a different way to make generalizations easier, such as in Lemma 3.5.

Proof. Let $k = \left\lceil \frac{\sqrt{\pi}}{2}\sqrt{A}\sqrt[4]{M}\right\rceil + 1$ and consider the Chebyshev polynomial $T_k(x)$. Let $u_0 = \cos\frac{\pi}{2k}$ which is the largest root and $u_1 = \cos\frac{\pi}{k}$ which is the largest local minimum (see Figure 1).

The polynomial we seek will be constructed as

$$f(x) = cg^2(x); \qquad g(x) = \frac{-T_k\big(u_0 - \frac{1+u_0}{A}x\big)}{x}, \qquad c = \frac{M}{g^2(0)}.$$

Obviously, $f(0) = M$ and $\deg f = 2(k - 1) < \sqrt{\pi}\sqrt{A}\sqrt[4]{M} + 2$, so properties (a) and (c) hold.

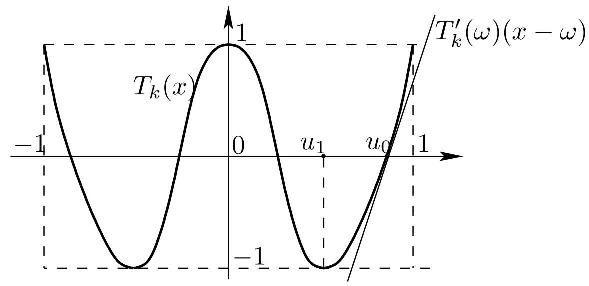

FIGURE 1. Construction of polynomial in Lemma 3.4

To estimate $g(0)$, notice that

$$T_k'(\cos t) = -\frac{\big(T_k(\cos t)\big)'}{\sin t} = -\frac{\big(\cos kt\big)'}{\sin t} = \frac{k \sin kt}{\sin t}$$

for all $0 < t < \pi$. Then

$$g(0) = \frac{1+u_0}{A} T_k'(u_0) = \frac{1+\cos\frac{\pi}{2k}}{A} \cdot \frac{k \sin\frac{\pi}{2}}{\sin\frac{\pi}{2k}} > \frac{2-\frac{1}{2}\left(\frac{\pi}{2k}\right)^2}{A} \cdot \frac{k}{\frac{\pi}{2k}} = \frac{\frac{4}{\pi}k^2 - \frac{\pi}{4}}{A} > \sqrt{M},$$

therefore $c = \frac{M}{g^2(0)} < 1$.

For all $x \in (0, A]$, we have $u_0 - \frac{1+u_0}{A}x \in [-1, 1]$ and $|T_k\big(u_0 - \frac{1+u_0}{A}x\big)| \le 1$. Hence,

$$|f(x)| = c \left(\frac{T_k\big(u_0 - \frac{1+u_0}{A}x\big)}{x} \right)^2 \le \frac{c}{x^2} < \frac{1}{x^2}. \tag{3.2}$$

In the interval $[u_1, u_0]$, by the convexity of the function $T_k(x)$, we have $|T_k(x)| \le T_k'(u_0)(x - u_0)$. For $x \in [-1, u_1]$ we have $T_k'(u_0)(x - u_0) < -1$. Therefore $|T_k(x)| \le |T_k'(u_0)| \cdot (u_0 - x)$ holds in the entire interval $[-1, u_0]$. Then, for all $x \in (0, A]$,

$$|f(x)| = c \left(\frac{T_k\big(u_0 - \frac{1+u_0}{A}x\big)}{x} \right)^2 \le c \left(\frac{|T_k'(u_0)| \cdot \frac{1+u_0}{A}x}{x} \right)^2 = cg^2(0) = M. \tag{3.3}$$

Estimates (3.2) and (3.3) together provide property (b). □

Lemma 3.5. *For arbitrary real numbers $A, B, M \ge 1$, there exists a polynomial $f(x)$ with real coefficients such that*

(a) $f(0) = M$,

(b) $|f(x)| < \min\left(4M, \frac{1}{x^2}\right)$ *for all* $x \in [-A, B]$, $x \ne 0$ *and*

(c) $\deg f < 7\sqrt{ABM} + 2$.

Proof. Without loss of generality, we can assume $A \ge B$. Let k be the smallest odd integer which is not less than $\frac{7}{2}\sqrt{ABM}$ and consider the Chebyshev polynomial $T_k(x)$. Let $\omega = \arccos\frac{A-B}{A+B}$ and let $u_0 = \cos\omega_0$ be the largest root of $T_k(x)$ in the interval $[-1, \frac{A-B}{A+B}]$. Since k is odd, $u_0 \ge 0$. Similarly to Lemma 3.4, the polynomial we seek will be constructed as

$$f(x) = cg^2(x), \qquad g(x) = \frac{T_k\left(u_0 + \dfrac{1+u_0}{A}x\right)}{x}, \qquad c = \frac{M}{g^2(0)}.$$

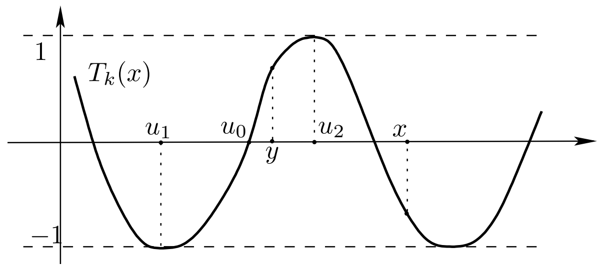

FIGURE 2. Construction of polynomial in Lemma 3.5

Again, properties (a) and (c) are obvious. For all $x \in [-A, B]$ we have $u_0 + \frac{1+u_0}{A} x \in [-1, 1]$ and therefore $|g(x)| \le \frac{1}{|x|}$.

Since $\omega \le \omega_0 \le \min\left(\omega + \frac{\pi}{k}, \frac{\pi}{2}\right)$,

$$\sin\omega_0 < \sin\omega + \frac{\pi}{k} \le \sqrt{1 - \left(\frac{A-B}{A+B}\right)^2} + \frac{\pi}{\frac{7}{2}\sqrt{ABM}} = \frac{2\sqrt{AB}}{A+B} + \frac{\frac{2}{7}\pi}{\sqrt{ABM}} < 3\sqrt{\frac{B}{A}},$$

$$|g(0)| = \frac{1+u_0}{A} |T_k'(u_0)| \ge \frac{1}{A} \cdot \frac{k}{\sin\omega_0} > \frac{\frac{7}{2}\sqrt{ABM}}{A \cdot 3\sqrt{\frac{B}{A}}} > \sqrt{M}$$

and

$$|f(x)| = \frac{M}{g^2(0)} g^2(x) < 1 \cdot \left(\frac{1}{x}\right)^2 = \frac{1}{x^2}. \tag{3.4}$$

To finish proving property (b) we show that

$$\left|\frac{T_k(x)}{x - u_0}\right| < 2|T_k'(u_0)| \tag{3.5}$$

for all $x \in [-1, 1]$, $x \ne u_0$. Let $u_1 = \cos\omega_1$ and $u_2 = \cos\omega_2$ be the two neighboring local extrema of $T_k(x)$ around u_0 (see Figure 2). Consider an arbitrary point $x \in [-1, 1]$, $x \ne u_0$. If $T_k(x) = 0$, then inequality (3.5) is trivial. Otherwise, choose the point $y = \cos\vartheta \in [u_1, u_2]$ such that x and y lie on the same side of u_0 and $|T_k(y)| = |T_k(x)|$. Then $0 < |y - u_0| \le |x - u_0|$ and by Cauchy's mean value theorem, there exists $\xi \in (\omega_2, \omega_1)$ such that

$$\left|\frac{T_k(x)}{x-u_0}\right| \le \left|\frac{T_k(y) - T_k(u_0)}{y - u_0}\right| = \left|\frac{\cos k\vartheta - \cos k\omega_0}{\cos\vartheta - \cos\omega_0}\right|$$
$$= \left|\frac{-k\sin k\xi}{-\sin\xi}\right| < \frac{k}{\sin\omega_2} = \frac{\sin\omega_0}{\sin\omega_2} \cdot |T_k'(u_0)|.$$

Since $\omega \le \omega_0 \le \frac{\pi}{2}$ and $\omega_2 = \omega_0 - \frac{\pi}{2k}$,

$$\frac{\sin\omega_2}{\sin\omega_0} > \frac{\sin\omega_0 - \frac{\pi}{2k}}{\sin\omega_0} = 1 - \frac{\frac{\pi}{2k}}{\sin\omega_0} \ge 1 - \frac{\frac{\pi}{2k}}{\sin\omega} \ge 1 - \frac{\frac{\pi}{7\sqrt{ABM}}}{\frac{2\sqrt{AB}}{A+B}} > \frac{1}{2}$$

and inequality (3.5) follows.

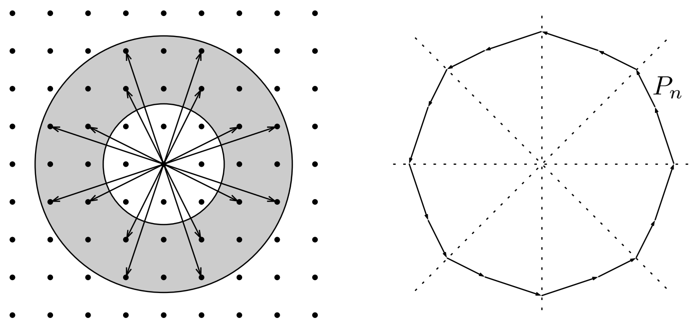

FIGURE 3. Construction of P_n in Lemma 3.6

Applying inequality (3.5) to polynomial $g(x)$,

$$|g(x)| = \frac{1+u_0}{A} \cdot \left| \frac{T_k\big(u_0 + \frac{1+u_0}{A}x\big)}{\frac{1+u_0}{A}x} \right| < \frac{1+u_0}{A} \cdot 2|T_k'(u_0)| = 2g(0)$$

and

$$|f(x)| = M\left(\frac{g(x)}{g(0)}\right)^2 < 4M. \tag{3.6}$$

Inequalities (3.4) and (3.6) prove property (b). □

Lemma 3.6. *For sufficiently large n there exists a convex lattice polygon P_n with the following properties:*

(a) *P_n contains a square of size $n \times n$ in its interior, with horizontal and vertical sides;*

(b) *the side lengths of P_n lie in the interval $[n^{1/3}, 2n^{1/3}]$;*

(c) *the sides of P_n do not contain any lattice point other than the vertices.*

Proof. Denote by $N(R)$ the number of lattice points (x, y) on the disk $x^2+y^2 < R^2$ which are visible from the origin (i.e. x and y are relatively prime). It is well known that

$$\lim_{R\to\infty} \frac{N(R)}{R^2} = \frac{6}{\pi}.$$

Let $R_1 = n^{1/3}$ and $R_2 = 2n^{1/3}$ and consider the lattice vectors (x, y) where x and y are relatively prime integers and $R_1^2 \le x^2 + y^2 < R_2^2$. Choose these vectors to be the sides of P_n; i.e. sort the vectors by direction and arrange them such that they form a convex polygon (see Figure 3). Obviously, properties (b) and (c) hold.

The perimeter of P_n is at least

$$\big(N(R_2) - N(R_1)\big) \cdot R_1 > \left(\frac{6}{\pi} - \varepsilon\right) R_2^2 R_1 - \left(\frac{6}{\pi} + \varepsilon\right) R_1^3 = \left(\frac{18}{\pi} - 5\varepsilon\right) n > 4\sqrt{2}n.$$

By the symmetry of P_n, property (a) follows. □

Lemma 3.7. *Let ℓ be an arbitrary line intersecting P_n and let ℓ_1 and ℓ_2 be the two supporting lines of P_n, parallel to ℓ; denote the distance between ℓ and ℓ_i by*

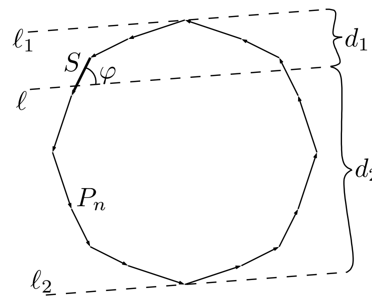

FIGURE 4. Estimating $\min(d_1, d_2)$ in Lemma 3.7

d_i $(i = 1, 2)$. *Assume that ℓ has a common point with a side S of P_n such that the angle between ℓ and S is $\varphi = \arcsin n^{-1/3}$ (see Figure 4). Then*

$$\min(d_1, d_2) < 15n^{1/3}.$$

Proof. Without loss of generality, we can assume that $d_1 \leq d_2$. Consider the side vectors of P_n which lie completely or partially between the lines ℓ and ℓ_1. Translating these vectors to start from the origin, the endpoints lie in a region D which is bounded by two concentric circular arcs of radii $R_1 = n^{1/3}$ and $R_2 = 2n^{1/3}$ and two radii of the same circles. The central angle of the arcs is 2φ (see Figure 5).

Drawing a unit square around the endpoints of the vectors, these squares do not overlap and they lie in a region denoted by D' in Figure 5. The central angle of this region is less than 4φ and its area is less than $\big((R_2+1)^2 - (R_1-1)^2\big) \cdot 4\varphi < 15n^{1/3}$. Therefore, the number of sides of P_n which have at least one endpoint between the lines ℓ and ℓ_1 is less than $15n^{1/3}$. Since the side lengths of P_n do not exceed $2n^{1/3}$ and the angles between ℓ and the mentioned sides do no exceed $\arcsin n^{-1/3}$, this implies $d_1 < 15n^{1/3}$. □

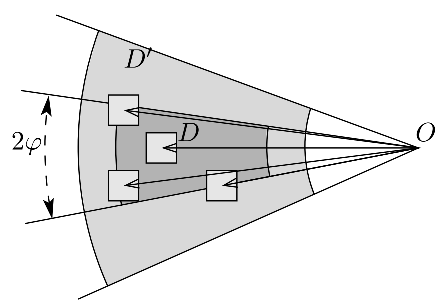

FIGURE 5. Regions D and D'

Proof of Lemma 3.3. Let $H \subset \{1, 2, \ldots, n\}^2$ be an arbitrary nonempty set. Translate the polygon P_n, provided by Lemma 3.6, to polygon P'_n such that the set H is contained in P'_n and at least one point of H lies on the boundary of P'_n. By the choice of the side vectors, any side of P'_n may contain at most two lattice points; if a side contains two lattice points, they must be the two endpoints. Since set

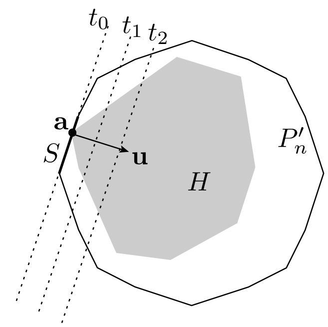

FIGURE 6. Construction of the first polynomial in Lemma 3.3

H cannot contain all vertices of the polygon P'_n, there is a side S which contains exactly one element of H. Let $\mathbf{a} = (a_1, a_2)$ be this element.

The desired polynomial $p(x, y)$ will be constructed as a product of two polynomials p_1 and p_2. To construct the first polynomial, rotate the side vector of S by 90 degrees such that it points inside P'_n; let this vector be $\mathbf{u} = (u_1, u_2)$; by the construction of P_n, the coordinates u_1 and u_2 are relatively prime integers and $n^{1/3} \le |\mathbf{u}| \le 2n^{1/3}$ (see Figure 6).

Let $f_1(t)$ be the polynomial provided by Lemma 3.4 for $M = 19$ and $A = 2n^{4/3}$ and define

$$g_1(x, y) = u_1(x - a_1) + u_2(y - a_2), \qquad p_1(x, y) = f_1\big(g_1(x, y)\big).$$

For each integer k, let t_k be the line where $g_1(x, y) = k$. Line t_0 is the extension of side S and the distance between lines t_k and t_{k+1} is $1/|\mathbf{u}|$ for every k. Since the diameter of set H is at most $\sqrt{2}n$ and $|\mathbf{u}| \le 2n^{1/3}$, we have $g_1(H) \subset \big\{0, 1, 2, \dots, \big[2\sqrt{2}n^{4/3}\big]\big\}$.

To construct the second polynomial, take a unit vector $\mathbf{v}$ which encloses an angle $\varphi = \arcsin n^{-1/3}$ with u. Let ℓ be the line through (a_1, a_2) which is perpendicular to v and let ℓ_1 and ℓ_2 be the two supporting lines of the set H, parallel to ℓ. Let d_i be the distance between ℓ and ℓ_i $(i = 1, 2)$. We can assume $d_1 \le d_2$. Moreover, by Lemma 3.7, we have $d_1 < 15n^{1/3}$ and $d_2 \le \sqrt{2}n$ since the diameter of H is at most $\sqrt{2}n$ (Figure 7).

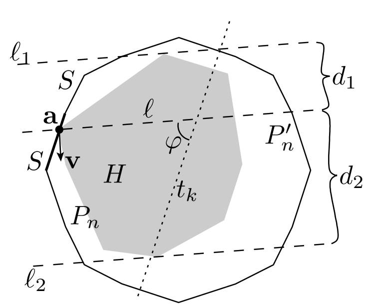

FIGURE 7. Construction of the second polynomial in Lemma 3.3

Let $f_2(t)$ be the polynomial by Lemma 3.5 for parameters $A = \max(d_1, 1)$, $B = \max(d_2, 1)$ and $M = 1$ and define

$$g_2(x, y) = v_1(x - a_1) + v_2(y - a_2), \qquad p_2(x, y) = f_2\big(g_2(x, y)\big).$$

For an arbitrary integer k, consider the lattice points on line t_k. The lattice points are distributed uniformly; the distance between the consecutive pairs is $|u|$. Hence, the values $g_2(x, y)$ on these points form an arithmetic progression lying in the interval $[-d_1, d_2]$ with difference $|u|/\sin\varphi \geq 1$.

Since $|f_2(t)| \leq \min(4, 1/t^2)$ in the interval $[-d_1, d_2]$, this implies

$$\sum_{(x,y)\in H\cap t_k} |p_2(x, y)| < 2 \sum_{h=0}^{[\max(d_1,d_2)]} \min\left(4, \frac{1}{h^2}\right) < 8 + \frac{\pi^2}{3}.$$

Now let

$$p(x, y) = p_1(x, y) \cdot p_2(x, y).$$

Then

$$p(a_1, a_2) = f_1(0) \cdot f_2(0) = 19 \cdot 1 = 19$$

and

$$\sum_{(x,y)\in H,\ (x,y)\neq(a_1,a_2)} |p(x, y)| = \sum_{k=1}^{[2\sqrt{2}n^{4/3}]} \sum_{(x,y)\in H\cap t_k} |p_1(x, y)| \cdot |p_2(x, y)|$$

$$= \sum_{k=1}^{[2\sqrt{2}n^{4/3}]} |f_1(k)| \sum_{(x,y)\in H\cap t_k} |p_2(x, y)| < \sum_{k=1}^{[2\sqrt{2}n^{4/3}]} \frac{1}{k^2}\left(8 + \frac{\pi^2}{3}\right) < \frac{\pi^2}{6}\left(8 + \frac{\pi^2}{3}\right)$$

$$< 19 = p(a_1, a_2),$$

so the polynomial $p(x, y)$ satisfies (3.1).

The degree of the polynomial is

$$\deg p = \deg p_1 + \deg p_2 < \left(\sqrt{\pi}\sqrt{2n^{4/3}}\sqrt[4]{19} + 2\right) + \left(7\sqrt{15n^{1/3} \cdot \sqrt{2}n} + 2\right) < 37.5n^{2/3}. \qquad \square$$

4. Summary

We proved that if $k > cn^{2/3}$, then every matrix $A \in \{0, 1\}^{n\times n}$ is uniquely determined by $M_k(A)$ and $M_k^{\mathrm{sym}}(A)$. To prove this, we simplified the problem, replacing the multisets by the sums $S_k(A)$ and $S_k^{\mathrm{sym}}(A)$. We also showed that the smallest values of k, for which $S_k(A)$ or $S_k^{\mathrm{sym}}(A)$ determines the matrix A, is between $\Omega\left(\frac{n^{2/3}}{\sqrt[3]{\log n}}\right)$ and $\mathcal{O}\big(n^{2/3}\big)$. These results indicate that the exponent $2/3$ is sharp in the simplified problem, pointing out the limitations of the presented method.

References

1. P. Borwein, T. Erdélyi, G. Kós, *Littlewood-type problems on* $[0, 1]$. Proc. London Math. Soc. **79**, No. 1 (1999), 22–46. MR1687555 (2000c:11111)
2. M. Dudik, L. J. Schulman, *Reconstruction from subsequences.* J. Combin. Theory Ser. A **103**, No. 2 (2003), 337–348. MR1996071 (2005a:05005)

3. W. H. Foster, I. Krasikov, *Inequalities for real-root polynomials and entire functions.* Adv. Appl. Math. **29**, No. 1 (2002), 102–114. MR1921546 (2004a:39055)
4. L. O. Kalashnik, *The reconstruction of a word from fragments.* Numer. Math. and Comp. Tech., Akad. Nauk. Ukrain. SSR Inst. Mat. IV (1973), 56–57. MR0485573 (58:5399)
5. P. J. Kelly, *On isometric transformations.* Ph.D. Thesis, University of Wisconsin (1942).
6. I. Krasikov, Y. Roditty, *On a reconstruction problem for sequences.* J. Combin. Theory Ser. A **77** No. 2 (1997), 344–348. MR1429086 (97m:05186)
7. B. Manvel, P. K. Stockmeyer, *On reconstruction of matrices.* Mathematics Magazine **44**, No. 4 (1971), 218–221. MR0295937 (45:4998)
8. A. Marcus, G. Tardos, *Excluded permutation matrices and the Stanley-Wilf conjecture.* J. Combin. Theory Ser. A **107**, No. 1 (2004), 153–160. MR2063960 (2005b:05009)
9. J. Pach, G. Tardos, *Forbidden patterns and unit distances.* SCG '05: Proc. 21st Annual Symposium on Computational Geometry, Pisa, Italy (2005), 1–9.
10. G. Tardos, *On 0-1 matrices and small excluded submatrices.* J. Combin. Theory Ser. A **111**, No. 2 (2005), 266-288. MR2156213 (2006e:05176)
11. S. M. Ulam, *A collection of mathematical problems.* Interscience Tracts in Pure and Applied Mathematics **8** (1960). MR0120127 (22:10884)

Mathematical Institute, Loránd Eötvös University, Pázmány P. s. 1/c, Budapest, Hungary H-1117; Computer and Automation Research Institute, Kende u. 13–17, Budapest, Hungary H-1111

E-mail address: kosgeza@cs.elte.hu

Department of Computer Algebra and Department of Computer Science, Loránd Eötvös University, Pázmány P. s. 1/c, Budapest, Hungary H-1117; Alfréd Rényi Institute of Mathematics, Reáltanoda u. 13-15, Budapest, Hungary H-1053

E-mail address: turul@cs.elte.hu

Mathematical Institute, Loránd Eötvös University, Pázmány P. s. 1/c, Budapest, Hungary H-1117

E-mail address: sziklai@cs.elte.hu