Electronic Colloquium on Computational Complexity, Report No. 67 (2009)

On Parity Check $(0,1)$-Matrix over $\mathbb{Z}_p$

Nader H. Bshouty
Technion, Israel
bshouty@cs.technion.ac.il

Hanna Mazzawi
Technion, Israel
hanna@cs.technion.ac.il

August 18, 2009

Abstract

We prove that for every prime p there exists a $(0,1)$-matrix M of size $t_p(n,m)\times n$ where

$$t_p(n,m)=O\left(m+\frac{m\log\frac{n}{m}}{\log\min(m,p)}\right)$$

such that every m columns of M are linearly independent over $\mathbb{Z}_p$, the field of integers modulo p (and therefore over any field of characteristic p and over the real numbers field $\mathbb{R}$). In coding theory this matrix is a parity-check $(0,1)$-matrix over $\mathbb{Z}_p$ of a linear code of distance m. Using the Hamming bound (for $p<m$) and information theoretic argument (for $p\ge m$) it can be shown that the above bound is tight.

To reduce the number of random bits, we extend this result to a $(0,1)$-matrix of size $s_p(n,m,d)\times n$ where $s_p(n,m,d)=O(t(n,m))$ and each row in the matrix is a tensor product (Kronecker product) of a constant d $(0,1)$-vectors of size $n^{1/d}$. This reduces the number of random bits from $O(t_p(n,m)n)$ to $O(t_p(n,m)n^{\epsilon})$ for any constant $\epsilon>0$ and gives a deterministic subexponential time algorithm (in n) for constructing M.

This solves the following open problems:

- **Coin Weighing Problem**: Suppose that n coins are given among which there are at most m counterfeit coins of arbitrary weights. There is a non-adaptive algorithm that finds the counterfeit coins and their weights in $t(n,m)=O((m\log n)/\log m)$ weighings.

Previous algorithm, [CK08], solves the problem (with the same complexity) only for weights between n^{-a} and n^b for constants a and b and finds the counterfeit coins but not their weights.

- **Reconstructing Graph from Additive Queries**: Suppose that G is an unknown weighted graph with n vertices and m edges. There exists a non-adaptive algorithm that finds the edges of G and their weights in $O(t(n,m))$ additive queries.

Previous algorithms, [CK08, BM09], solves the problem only for weights between n^{-a} and n^b for constants a and b and finds the edges but not their weights.

- **Signature Coding Problem**: Consider n stations and at most m of them want to send messages from $\mathbb{Z}_p$ through an adder channel, that is, a channel that its output is the sum of the messages. Then all messages can be sent (encoded and decoded) with $O(t(n,m))$ transmissions. Previous algorithms, [BG07], run with the same number of transmissions only for messages in $\{0,1\}$.

Simple information theoretic arguments show that all the above bounds are tight.

ISSN 1433-8092

1 Introduction

A $t \times n$ $(0,1)$-matrix is called an *m-independent column $(0,1)$-matrix over $\mathbb{Z}_p$* if every m columns in the matrix are linearly independent over $\mathbb{Z}_p$. In coding theory this matrix is a parity-check $(0,1)$-matrix over $\mathbb{Z}_p$ of a linear code of distance m. Using the Hamming bound (for $p < m$) and information theoretic argument (for $p \geq m$) it can be shown that such a matrix must have at least

$$t = \Omega\left(m + \frac{m \log \frac{n}{m}}{\log \min(m,p)}\right)$$

rows. Using a straightforward probabilistic argument it is easy to show that an $O(m \log n) \times n$ m-independent column matrix exists. Simply take a random $(0,1)$-matrix over $\mathbb{Z}_p$ with such size and show that the probability that every m columns are independent over $\mathbb{Z}_p$ is greater than 0.

In this paper we close the gap between the lower and upper bound. We prove that there exists a $t_p(n,m) \times n$ m-independent column $(0,1)$-matrix with

$$t_p(n,m) = O\left(m + \frac{m \log \frac{n}{m}}{\log \min(m,p)}\right)$$

rows. We give a new analysis that shows that for any prime p and a random $(0,1)$-matrix M of size $t_p(n,m) \times n$, the probability that every m columns in M are independent over $\mathbb{Z}_p$, is greater than 0. Our proof is based on the following result from number theory: Given a prime p and any sequence of m elements $S = (a_1, a_2, \ldots, a_m) \in \mathbb{Z}_p^m$. The number of subsequences $T = (a_{i_1}, \ldots, a_{i_r})$, $1 \leq i_1 < i_2 < \cdots < i_r \leq m$ for which the sum of its elements is equal to 0 is at most $2^m / \min(m^{0.278}, p^{0.5})$.

One application is the (n,m)-coin weighing problem [D75, L75, C80, AS85, A86, A88, BG07, CK08]. Suppose that n coins are given among which there are at most m counterfeit coins of arbitrary weights. The goal is to find a non-adaptive algorithm that finds the counterfeit coins and their weights. We show that the above result implies that there exists a non-adaptive algorithm that finds the counterfeit coins and their weights in

$$t(n,m) = O\left(\frac{m \log n}{\log m}\right)$$

weighings. Previous algorithm in [CK08] solves the problem only for weights between n^{-a} and n^b for constants a and b and finds the counterfeit coins but not their weights.

We then show that there is an $O(t_p(n,m)) \times n$ m-independent column $(0,1)$-matrix over $\mathbb{Z}_p$ that each of its rows is a tensor product of a constant d $(0,1)$-vectors of size $n^{1/d}$. We show that if we randomly uniformly choose $O(t_p(n,m))$ d-tuples of $(0,1)$-vectors of size $n^{1/d}$, $(v_{i,1}, \ldots, v_{i,d})$, the probability that the matrix that its i-th row is $\otimes_j v_{i,j}$ has m linearly dependent columns is less than 1. This reduces the number of random bits for constructing the m-independent column $(0,1)$-matrix over $\mathbb{Z}_p$ from $O(n \cdot t_p(n,m))$ to $O(n^{1/d} \cdot t_p(n,m))$ which gives a deterministic subexponential time algorithm (in n) for constructing such matrix. This also (for $d = 2$) solves the following problem.

Consider the following problem of reconstructing weighted graphs using additive queries [G98, GK98, GK00, BGK05, RS07, CK08, BM09]: Let $G = (V, E, w)$ be a weighted hidden graph where $E \in V \times V$, $w : E \to \mathbb{R}$ and n is the number of vertices in V. Denote by m the size of E. Suppose that the set of vertices V is known and the set of edges E is unknown. Given a set of vertices $S \subseteq V$, an additive query, $Q(S)$, returns the sum of weights in the subgraph induces by S. That is,

$$Q(S) = \sum_{e \in E \cap (S \times S)} w(e).$$

Our goal is to exactly reconstruct the set of edges and find their weights using additive queries. See the many applications of this problem in [CK08].

Our result (for $d = 2$) implies that there exists a non-adaptive algorithm to find the edges of G and their weights using $O(t(n, m))$ additive queries. Previous algorithms in [CK08, BM09] solves the problem only for weights between n^{-a} and n^b for constants a and b and find the edges but not their weights.

Another application is the signature coding problem [BG07]. Consider n stations where m of the stations want to transmit messages in $\mathbb{Z}_p$ through an adder channel, that is, a channel that its output is the sum of the messages. Then all messages can be transmitted (encoded and decoded) in $O(t(n, m))$ transmissions. Previous algorithms run with the same transmission complexity in two stages: first it decides which of the stations are active, that is, stations that want to transmit messages (that is, messages in $\{0, 1\} \subset \mathbb{Z}_p$) and then, sequentially, asks each active station to send its message. Our algorithm is non-adaptive and can detect the active stations and their messages in one stage.

Simple information theoretic arguments show that all the above bounds are tight.

This paper is organized as follows. In Section 2 we prove some basic probability results that will be used throughout the paper. In Section 3 we give the m-independent column $(0, 1)$-matrix over $\mathbb{Z}_p$. In Section 4 and Section 5 we give the m-independent column $(0, 1)$-matrix over $\mathbb{Z}_p$ where each row is a tensor product of $(0, 1)$-vectors.

2 Basic Probability

In this section we give some preliminary results in probability theory that will be used in the sequel.

We denote by $\mathbb{R}$ the set of real numbers and by $\mathbb{Z}$ the set of integers. For a prime number p we denote by $\mathbb{Z}_p$ the field of integers modulo p. For any positive integer r, we denote by $[r]$ the set $\{1, 2, \ldots, r\}$. We will write $a =_p b$ for $a = b \mod p$.

Let X be a vector or a matrix, we denote by $wt(X)$ the Hamming weight of X, that is, the number of non-zero entries in X. For two vectors x and y the distance $dist(x, y)$ between x and y is the number of entries in x and y that differ, that is, $wt(x - y)$. For $\sigma \in \{0, 1\}$, we denote by σ^n the n-vector whose entries are all equal to σ. We also denote by $\sigma^{n \times m}$ the $n \times m$ matrix whose entries are all equal to σ.

The following three lemmas are well known from the literature over the field of real numbers. We give the proofs for any field $\mathbb{Z}_p$.

Lemma 1. *Let $a \in \mathbb{Z}_p^n \backslash \{0^n\}$. Then for a uniformly randomly chosen vector $x \in \{0,1\}^n$ we have*

$$\mathbf{Pr}_x[a^T x =_p 0] \leq 1/2.$$

Proof. Suppose w.l.o.g. that $a_1 \neq_p 0$. For any fixed $x_2, \ldots, x_n \in \{0,1\}$ we have $a^T x =_p a_1 x_1 + c$ for some $c \in \mathbb{Z}_p$. Now this takes the value c for $x_1 = 0$ and $c + a_1$ for $x_1 = 1$. Since $a_1 \neq 0$, one of the values c or $c + a_1$ is not equal to zero. □

Lemma 2. *Let $M \in \mathbb{Z}_p^{n\times n} \backslash \{0^{n\times n}\}$. Then for a uniformly randomly chosen vectors $x, y \in \{0,1\}^n$ we have*

$$\mathbf{Pr}_{x,y}[x^T M y =_p 0] \leq 3/4.$$

Proof. By Lemma 1, My has a non-zero entry with probability greater or equal to 1/2. Assuming $My \neq_p 0^n$, by Lemma 1 the probability that $x^T My \neq_p 0$ is greater or equal to 1/2. This implies the result. □

The following lemma was proved in the literature for the real number field using Littlewood-Offord Theorem [LO43, E45] (with $\beta = 1/2$). In this paper we prove it for any field $\mathbb{Z}_p$.

Lemma 3. *Let $a \in \mathbb{Z}_p^n \backslash \{0^n\}$ be a vector, where p is a prime number. Then for a uniformly randomly chosen vector $x \in \{0,1\}^n$ we have*

$$\mathbf{Pr}_x[a^T x =_p 0] \leq \max\left(\frac{1}{wt(a)^\beta}, \frac{1}{p^{1/2}}\right),$$

where $\beta = \frac{1}{2+\log 3} = 0.278943\cdots$.

Proof. Let $S = \{a_i \mid i \in [n]\}$ and $\alpha = \frac{\log 3}{2+\log 3}$. We take two cases:

- **Case 1**: The size of S is at most $wt(a)^\alpha$.

  Using the pigeon hole principle, there is an element $g \in \mathbb{Z}_p \backslash \{0\}$ that appears in a more than $wt(a)^{1-\alpha}$ times. Suppose w.l.o.g. that $a_1 = a_2 = \ldots = a_t = g$ where $t = \min(wt(a)^{1-\alpha}, p)$. For any fixed $x_{t+1}, x_{t+2}, \ldots, x_n \in \{0,1\}$ we have

  $$a^T x =_p g(x_1 + x_2 + \ldots + x_t) + c'.$$

  Therefore, $a^T x = 0$ implies

  $$x_1 + x_2 + \ldots + x_t =_p -c' g^{-1}.$$

  Since $t \leq p$, we have for $c = \sqrt{2/\pi} = 0.797885\cdots < 1$

  $$\mathbf{Pr}_x[a^T x =_p 0] \leq \frac{\binom{t}{\lfloor t/2 \rfloor}}{2^t} \leq \frac{c}{\sqrt{t}} \leq \frac{c}{\min\left(wt(a)^{\frac{1-\alpha}{2}}, p^{1/2}\right)} \leq \max\left(\frac{1}{wt(a)^\beta}, \frac{1}{p^{1/2}}\right).$$

- **Case 2**: The size of S is at least $wt(a)^\alpha$.
  For a set of elements $Q = \{q_1, q_2, \dots, q_r\} \subseteq \mathbb{Z}_p$ denote by

  $$\psi(Q) = |\{(q_1y_1 + q_2y_2 + \cdots + q_ry_r) \mod p \,|\, y_1, \dots, y_r \in \{0,1\}\}|,$$

  and

  $$\mathcal{A}(Q) = \{(q_1z_1 + q_2z_2 + \cdots + q_rz_r) \mod p \,|\, z_1, \dots, z_r \in \{-1,0,1\}\}.$$

  Since $|S| > wt(a)^\alpha$, we argue that there exists a set of entries $Q = \{a_{i_1}, a_{i_2}, \dots, a_{i_k}\}$ such that $k \geq \log_3 wt(a)^\alpha$ and $\psi(Q) = 2^k$. We prove this claim by showing how to find such set of entries. The process of finding the entries is iterative. At every iteration j we have a set of entries $Q_j = \{a_{i_1}, a_{i_2}, \dots, a_{i_j}\}$ of size j such that $\psi(Q_j) = 2^j$. It is easy to see that if $a_{i_{j+1}} \notin \mathcal{A}(Q_j)$ then $\psi(Q_j \cup \{a_{i_{j+1}}\}) = 2^{j+1}$. An element $a_{i_{j+1}} \in S$ can be added to Q_j as long as $|\mathcal{A}(Q_j)| \leq 3^j < |S|$. Therefore, we are able to find a set Q such that $|Q| \geq \log_3 |S|$ and $\psi(Q) = 2^{|Q|}$.

  Now, let W denote the set $[n]\backslash\{i_1, \dots, i_k\}$. For any fixed values for entries in W we have that

  $$a^T x =_p a_{i_1}x_{i_1} + a_{i_2}x_{i_2} + \cdots + a_{i_k}x_{i_k} + c'',$$

  where c'' is a constant. By the properties of Q, there is at most one $y \in \{0,1\}^k$ such that $a_{i_1}y_1 + a_{i_2}y_2 + \cdots + a_{i_k}y_k = -c''$. Therefore

  $$\mathbf{Pr}_x[a^T x =_p 0] \leq \frac{1}{2^k} \leq \frac{1}{2^{\log_3 wt(a)^\alpha}} = \frac{1}{wt(a)^{\alpha \log_3 2}} = \frac{1}{wt(a)^\beta}.$$

$\square$

Note that Lemma 3 is not true for non-prime p. Consider an even number p. Then with probability $1/2$ we have $(p/2)x_1 + \cdots + (p/2)x_n =_p 0$.

We now prove some properties of the rank of random (0,1)-matrices over $\mathbb{Z}_p$. Similar properties was proved for the field of real numbers in [K67, B01, BG08].

Lemma 4. *Let $M \in \{0,1\}^{k \times m}$ be a matrix of rank $r = r(M) < m$. For a uniformly randomly chosen row vector $y \in \{0,1\}^m$ the rank of the matrix*

$$M' = \begin{pmatrix} M \\ y \end{pmatrix}.$$

over $\mathbb{Z}_p$ is r with probability at most $1/2$.

Proof. Denote by M_i the ith column of M. Let $M_{i_1}, M_{i_2}, \dots, M_{i_r}$ be any r linearly independent columns of M. Let M_j be any other column of the matrix, that is, $j \neq i_s$ for all $s \in [r]$. Then, there are unique constants $\alpha_1, \alpha_2, \dots, \alpha_r$ such that

$$M_j =_p \alpha_1 M_{i_1} + \alpha_2 M_{i_2} + \ldots + \alpha_r M_{i_t}.$$

Therefore,

$$\alpha_1 M_{i_1} + \alpha_2 M_{i_2} + \ldots + \alpha_t M_{i_t} - M_j =_p 0.$$

Let a be the m-vector, where $a_j = -1$, $a_{i_s} = \alpha_{i_s}$ for all $s \in [r]$ and all other entries are zeros. Then,

$$\mathbf{Pr}[r(M') = r] \le \mathbf{Pr}[a^T y =_p 0].$$

Now by Lemma 1 the result follows. $\square$

A better bound can be obtained when additional properties are known about the matrix. We prove the following:

Lemma 5. *Let $M \in \{0,1\}^{k\times m}$ be a matrix of rank $r(M) = r < m$. Suppose that every s columns of M are linearly independent. Then, for a uniformly randomly chosen row vector $y \in \{0,1\}^m$ the rank of the matrix*

$$M' = \begin{pmatrix} M \\ y \end{pmatrix}.$$

over $\mathbb{Z}_p$ is r with probability at most

$$\max\left(\frac{1}{s^\beta}, \frac{1}{p^{1/2}}\right).$$

Proof. Using the same notation as in the proof of Lemma 4. Observe that $wt(a) \ge s$. Otherwise, there are s linearly dependent columns in M. Therefore, by Lemma 3,

$$\mathbf{Pr}[r(M') = r] \le \mathbf{Pr}[a^T y =_p 0] \le \max\left(\frac{1}{s^\beta}, \frac{1}{p^{1/2}}\right).$$

$\square$

We will also make use of the following

Lemma 6. (Chernoff bound) *Let $X_1, \ldots, X_t$ be independent Poisson trials such that $X_i \in \{0,1\}$ and $\mathbf{E}[X_i] = p_i$. Let $P = \sum_{i=1}^t p_i$ and $X = \sum_{i=1}^t X_i$. Then*

$$\mathbf{Pr}\left[X \le (1-\lambda)P\right] \le e^{-\lambda^2 P/2}.$$

3 m-Independent Column $(0,1)$-Matrix

In this section we prove the existence of an m-independent column $(0,1)$-matrix over $\mathbb{Z}_p$ with optimal size.

The following lower bound follows from the Hamming bound and using an information theoretic argument. We give the proof for completeness.

Theorem 7. *Let p be any prime number. A $(0,1)$-matrix $M \in \{0,1\}^{k\times n}$ such that every m columns in M are linearly independent over $\mathbb{Z}_p$ must have at least*

$$k = \Omega\left(m + \frac{m\log\frac{n}{m}}{\log\min(m,p)}\right)$$

rows.

Proof. For $p < m$ and by the Hamming bound we have

$$p^k \geq \sum_{i=0}^{m}\binom{n}{i}(p-1)^i.$$

Therefore,

$$k \geq \frac{\log\sum_{i=0}^{m}\binom{n}{i}(p-1)^i}{\log p} \geq \frac{\log(p-1)}{\log p}m + \frac{\log\binom{n}{m}}{\log p} = \Omega\left(m + \frac{m\log\frac{n}{m}}{\log p}\right).$$

For $p > m$, notice that for every $v, u \in \{0,1,\ldots,m\}^n$ of weight equal to $m/2$ we have $Mv \neq_p Mu$. Otherwise, $M(v-u) =_p 0^k$ and the columns that corresponds to the (at most m) entries that are not zero in $v-u$ are linearly dependent. Since for every $v \in \{0,1,\ldots,m\}^n$ of weight at most $m/2$ we have $Mv \in \{0,1,\ldots,m^2/2\}^k$ we must have

$$\left(\frac{m^2}{2}+1\right)^k \geq \binom{n}{m/2}(m-1)^{m/2}.$$

Therefore,

$$k = \Omega\left(m + \frac{m\log\frac{n}{m}}{\log m}\right).$$

□

It is easy to prove the following (see the first part of the proof of Theorem 9).

Theorem 8. *For any prime p there exists a matrix $M \in \{0,1\}^{k\times n}$ such that*

$$k = O\left(m\log\frac{n}{m}\right),$$

and every m columns are linearly independent over $\mathbb{Z}_p$.

The following theorem closes the gap between the upper and the lower bound

Theorem 9. *For any prime p there exists a matrix $M \in \{0,1\}^{k\times n}$ such that*

$$k = O\left(m + \frac{m\log\frac{n}{m}}{\log\min(m,p)}\right),$$

and every m columns are linearly independent over $\mathbb{Z}_p$.

Proof. Let $t = m/\log^2 m$. We first prove the existence of a matrix $M^* \in \{0,1\}^{k_1 \times n}$ such that

$$k_1 = t + \log t + 2\log\binom{n}{t}$$

where every t columns are linearly independent. We use probabilistic method. We randomly uniformly choose k_1 $(0,1)$-vectors of size n to be the rows of the matrix. Denote by M_i the ith column of the matrix M^*. Now, let $M_{i_1}, M_{i_2}, \ldots, M_{i_t}$ be any t columns. Consider the matrix

$$M' = [M_{i_1}|M_{i_2}|\cdots|M_{i_t}]$$

and let $M'^{(j)}$ be the jth row of M' and $M'^{[j]}$ be the first j rows of M'. Consider the random variable $X_j \in \{0,1\}$ where $X_j = 1$ if and only if $r(M'^{[j-1]}) = t$ or the jth row $M'^{(j)}$ increases the rank of $M'^{[j-1]}$, i.e., $r(M'^{[j]}) = r(M'^{[j-1]}) + 1$. By Lemma 4,

$$\mathbf{Pr}[X_j = 0 | X_1, X_2, \ldots, X_{j-1}] \le 1/2.$$

Therefore, the probability that the rank of the matrix M' is smaller than t is bounded by

$$\begin{aligned}
\mathbf{Pr}[X_1 + \cdots + X_{k_1} \le t-1] &= \sum_{\xi_1+\cdots+\xi_{k_1} \le t-1, \xi_j \in \{0,1\}} \mathbf{Pr}[X_1 = \xi_1, X_2 = \xi_2, \ldots, X_{k_1} = \xi_{k_1}] \\
&\le \frac{\sum_{i=0}^{t-1}\binom{k_1}{i}}{2^{k_1-t+1}} \le t2^{t-1}\frac{\binom{k_1}{t}}{2^{k_1}} < \frac{\binom{n}{t}}{2\binom{n}{t}^2} \le \frac{1}{2\binom{n}{t}}.
\end{aligned}$$

Using union bound, the probability that there exists a set of t columns that are linearly independent is less than $1/2$. This implies the existence of M^*.

Now, we have a matrix M^* that every $m/\log^2 m$ columns are linearly independent. We add k_2 uniformly randomly chosen rows to the matrix, where

$$k_2 = \frac{m\log q + \log\binom{n}{m}}{\log q - 1}, \tag{1}$$

where

$$q = \min(t^\beta, p^{1/2}).$$

Again, given any m columns $M_{j_1}, M_{j_2}, \ldots, M_{j_m}$. By Lemma 5 the probability that the matrix

$$M'' = [M_{j_1}|M_{j_2}|\cdots|M_{j_m}]$$

has rank smaller than m is bounded by

$$\frac{\sum_{i=0}^{m-1}\binom{k_2}{i}}{q^{k_2-m+1}} < \frac{2^{k_2}}{2q^{k_2-m}} \le \frac{1}{2\binom{n}{m}}. \tag{2}$$

Therefore, the probability that there exists m columns of M that are linearly independent is bounded by $1/2$.

This, together with the fact that

$$k_1 + k_2 = O\left(m + \frac{m \log \frac{n}{m}}{\log \min(m, p)}\right),$$

implies the result. $\square$

The following corollary solves the coin weighing problem and the signature coding problem.

Corollary 1. *There exists a matrix $M \in \{0,1\}^{k \times n}$ where*

$$k = O\left(m + \frac{m \log \frac{n}{m}}{\log m}\right),$$

and for every two distinct vectors $x, y \in \mathbb{R}^n$ such that $wt(x) \le m$ and $wt(y) \le m$ we have $Mx \ne My$.

Proof. Choose a prime $2m < p$. By Theorem 9 there exists a $k \times n$ matrix M such that every $2m$ columns are linearly independent over $\mathbb{Z}_p$, and therefore, over $\mathbb{R}$. For any two vector $x, y \in \mathbb{R}^n$ such that $wt(x) \le m$, $wt(y) \le m$ and $x \ne y$ we have that $0 < wt(x - y) \le 2m$. Therefore,

$$M(x - y) \ne 0^k,$$

and

$$Mx \ne My.$$

$\square$

4 $(0,1)$-Matrices with Rows that are Tensor Product of Two Vectors

In this section we show that there is an m-independent column $t \times n$ $(0,1)$-matrix that its rows are tensor product of two $(0,1)$-vectors in $\{0,1\}^{\sqrt{n}}$ and

$$t = O\left(m + \frac{m \log \frac{n}{m}}{\log \min(m, p)}\right).$$

In the next section we extend this result to $(0,1)$-matrix that its rows are tensor product of d $(0,1)$-vectors in $\{0,1\}^{n^{1/d}}$.

The following theorem follows from Case 1 in the proof of Theorem 11.

Theorem 10. *Let $p < n^\gamma$ be a prime number for some constant $\gamma > 1$. There exists a set of $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\}$ where $x_i, y_i \in \{0,1\}^n$ and*

$$k = O\left(m \log \frac{n^2}{m}\right),$$

such that: for any matrix $A \in \mathbb{Z}_p^{n\times n}\backslash\{0^{n\times n}\}$ with $wt(A) \le m$, there exists an i such that $x_i^T A y_i \ne_p 0$.

We now prove the following:

Theorem 11. *Let $p < n^\gamma$ be a prime number for some constant $\gamma > 1$. There exists a set of $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\}$ where $x_i, y_i \in \{0,1\}^n$ and*

$$k = O\left(m + \frac{m \log \frac{n^2}{m}}{\log \min(m,p)}\right),$$

such that: for any matrix $A \in \mathbb{Z}_p^{n\times n}\backslash\{0^{n\times n}\}$ with $wt(A) \le m$, there exists an i such that $x_i^T A y_i \ne_p 0$.

Proof. First notice that when m is constant then

$$O\left(m + \frac{m \log \frac{n^2}{m}}{\log \min(m,p)}\right) = O\left(m \log \frac{n^2}{m}\right),$$

and Theorem 10 implies the result. Therefore we may assume that $m = \omega(1)$. Note also that we may assume that $m < n^2/2$. Otherwise, we can just take the all the n^2 pairs (e_i, e_j) where $\{e_i\}_{i\in[n]}$ is the standard basis.

We divide the set of matrices

$$\mathcal{A} = \{A \mid A \in \mathbb{Z}_p^{n\times n}\backslash\{0^{n\times n}\} \text{ and } wt(A) \le m\}$$

into three (non-disjoint) sets:

- $\mathcal{A}_1$: The set of all non-zero matrices $A \in \mathbb{Z}_p^{n\times n}$ such that $wt(A) \le m/\log m$.
- $\mathcal{A}_2$: The set of all non-zero matrices $A \in \mathbb{Z}_p^{n\times n}$ such that $m \ge wt(A) > m/\log m$ and there are at least $\sqrt{\frac{m}{\log m}}$ non-zero rows.
- $\mathcal{A}_3$: The set of all non-zero matrices $A \in \mathbb{Z}_p^{n\times n}$ such that $m \ge wt(A) > m/\log m$ and there are at least $\sqrt{\frac{m}{\log m}}$ non-zero columns.

Note that for any matrix A of weight $wt(A) > d = m/\log m$, either A has more than $\sqrt{d}$ non-zero rows or more than $\sqrt{d}$ non-zero columns. Therefore, $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$.

Using the probabilistic method, we give three sets S_1, S_2 and S_3 of vector pairs, such that for every $j = 1, 2, 3$ and $A \in \mathcal{A}_j$ there exists a pair of vectors $(x, y) \in S_j$ such that $x^T Ay \neq_p 0$ and

$$|S_1| + |S_2| + |S_3| = O\left(m + \frac{m \log \frac{n^2}{m}}{\log \min(m, p)}\right).$$

Case 1: $A \in \mathcal{A}_1$.
By Lemma 2 for randomly chosen vectors $x, y \in \{0,1\}^n$ and $A \in \mathcal{A}_1$ we have

$$\mathbf{Pr}[x^T Ay =_p 0] \leq 3/4.$$

Randomly uniformly choose

$$k_1 = c\left(m + \frac{m \log \frac{n^2}{m}}{\log \min(m, p)}\right) > c\left(m + \frac{m \log \frac{n^2}{m}}{\log m}\right) = c\left(\frac{m \log n^2}{\log m}\right),$$

vectors $x_i, y_i \in \{0,1\}^n$ where $c = 3(2+\gamma)$. Then, the probability that for all x_i, y_i we have $x_i^T Ay_i =_p 0$ is bounded by

$$\mathbf{Pr}[\forall i \in [k_1] : x_i^T Ay_i =_p 0] \leq \left(\frac{3}{4}\right)^{k_1}.$$

Therefore, by union bound, the probability that there exists a matrix A of weight smaller than $m/\log m$ such that $x_i^T Ay_i =_p 0$ for all $i \in [k_1]$ is

$$\begin{aligned}
\mathbf{Pr}[\exists A \in \mathcal{A}_1, \forall i \in [k_1] : x_i^T Ay_i =_p 0] &\leq \binom{n^2}{\frac{m}{\log m}} p^{\frac{m}{\log m}} \left(\frac{3}{4}\right)^{k_1} \\
&< n^{2\frac{m}{\log m}} n^{\gamma \frac{m}{\log m}} \left(\frac{1}{2}\right)^{k_1/3} \\
&< n^{(2+\gamma)\frac{m}{\log m}} \left(\frac{1}{2}\right)^{k_1/3} \\
&< n^{(2+\gamma)\frac{m}{\log m}} 2^{-(c/3)\frac{m \log n^2}{\log m}} \\
&= n^{-(2+\gamma)\frac{m}{\log m}} < 1.
\end{aligned}$$

This implies the result.

Case 2: $A \in \mathcal{A}_2$.
We start by proving the following two lemmas

Lemma 12. *Let $U \subset \mathbb{Z}_p^n$ be the set of all non-zero vectors with weight smaller than $m^{3/4}$. For any constant $C > (1+\gamma)16/\log e$ and*

$$k_2 = C\left(m + \frac{m\log\frac{n^2}{m}}{\log\min(m,p)}\right) > C\left(m + \frac{m\log\frac{n^2}{m}}{\log m}\right) = C\left(\frac{m\log n^2}{\log m}\right), \tag{3}$$

there exists a multiset of (0,1)-vectors $Y = \{y_1, y_2, \ldots, y_{k_2}\}$ such that for every $u \in U$ the size of the multiset

$$Y_u = \{i \,|\, u^T y_i \neq_p 0\}$$

is at least $k_2/4$.

Proof. By Lemma 1 for a randomly chosen vector $y \in \{0,1\}^n$ and any $u \in U$ we have

$$\mathbf{Pr}[u^T y =_p 0] \leq 1/2.$$

Therefore, if we randomly uniformly choose the vectors of Y, then the expected size of Y_u is greater than $k_2/2$ for any $u \in U$. Using Chernoff bound (Lemma 6) we have that

$$\mathbf{Pr}[|Y_u| < k_2/4] \leq e^{-\frac{k_2}{16}}$$

Therefore, the probability that there exists $u \in U$ such that $|Y_u| < k_2/4$ is

$$\begin{aligned}
\mathbf{Pr}[\exists u \in U : |Y_u| < k_2/4] &\leq \frac{|U|}{e^{\frac{C}{16}\left(\frac{m\log n^2}{\log m}\right)}} \\
&\leq \frac{\sum_{i=0}^{m^{3/4}} \binom{n}{i}(p-1)^i}{n^{\frac{C\log e}{8}\left(\frac{m}{\log m}\right)}} \\
&\leq \frac{n^{m^{3/4}} n^{\gamma m^{3/4}}}{n^{\frac{C\log e}{8}\left(\frac{m}{\log m}\right)}} \\
&\leq n^{(1+\gamma)\left(m^{3/4} - 2\frac{m}{\log m}\right)} < 1.
\end{aligned}$$

This implies the result. $\square$

Note that the constant C will be determined later in the proof. Now for the next lemma, define for non-negative integer r, $\iota(r) = \min(r,p)$ if $r > 0$ and $\iota(0) = 1$.

Lemma 13. *Let $m_1, m_2, \ldots, m_{k_2}$ be integers in $[m] \cup \{0\}$ such that*

$$m_1 + m_2 + \cdots + m_{k_2} = \ell \geq k_2.$$

Then

$$\prod_{i=0}^{k_2} \iota(m_i) \geq \min(m,p)^{\lfloor(\ell - k_2)/(m-1)\rfloor}.$$

Proof. We first proof that when $1 < m_1 \le m_2 < m$ then

$$\iota(m_1 - 1)\iota(m_2 + 1) \le \iota(m_1)\iota(m_2). \tag{4}$$

We have four cases: When $p \le m_1 - 1$ then (4) gives $p^2 \le p^2$. When $p = m_1$ then (4) gives $(p-1)p \le p^2$. When $m_1 < p \le m_2$ then (4) gives $(m_1 - 1)p \le m_1 p$. When $p \ge m_2 + 1$ then (4) gives $(m_1 - 1)(m_2 + 1) < m_1 m_2$. In all cases the inequality is true.

Also when $m_1 = 0$ and $1 < m_2 < m$ then $\iota(m_1 + 1)\iota(m_2 - 1) = \min(m_2 - 1, p) \le \min(m_2, p) = \iota(m_1)\iota(m_2)$. Therefore the optimal value of $\iota(m_1)\iota(m_2)\cdots\iota(m_t)$ is obtained when for every $0 < i < j \le k_2$ we either have $m_i \in \{1, m\}$ or $m_j \in \{1, m\}$. This is equivalent to: all $m_i \in \{1, m\}$ except at most one. This implies that at least $\lfloor (\ell - k_2)/(m-1) \rfloor$ of the m_is are equal to m. $\square$

Now let U be the set of vectors defined in Lemma 12. Let $A \in \mathcal{A}_2$. Since $wt(A) \le m$ there are at most $m^{1/4}$ rows in A with weight greater than $m^{3/4}$. Therefore, there are at least

$$q = \sqrt{\frac{m}{\log m}} - m^{1/4}$$

rows in A that are in U. Let A_U be $q \times n$ matrix that its rows are any q rows in A that are in U. Let $Y = \{y_1, y_2, \ldots, y_{k_2}\}$ be the set we proved its existence in Lemma 12 (see (3)). Note that

$$\sum_i wt(A_U y_i) \ge \frac{qk_2}{4}$$

Since $wt(A_U y_i) \le q$ for all $i \in [k_2]$, by Lemma 13 we have

$$\prod_i \iota(wt(A_U y_i)) \ge (\min(q, p))^{\lfloor \frac{\frac{qk_2}{4} - k_2}{q-1} \rfloor} \ge (\min(q, p))^{c_1 k_2} \ge (\min(m, p))^{c_2 k_2},$$

where c_1 and c_2 are constants. If we randomly choose $x_1, x_2, \ldots, x_{k_2}$ then by Lemma 3, we have

$$\begin{aligned}
\mathbf{Pr}[\forall i \in [k_2] : x_i^T A y_i \ne_p 0] &\le \prod_i \frac{1}{\min((wt(Ay_i))^\beta, p^{1/2})}, \\
&\le \prod_i \frac{1}{\iota(wt(Ay_i))^\beta}, \\
&\le \prod_i \frac{1}{\iota(wt(A_U y_i))^\beta} \\
&= \left(\frac{1}{\prod_i \iota(wt(A_U y_i))} \right)^\beta \\
&\le \frac{1}{(\min(m, p))^{\beta c_2 k_2}} \\
&= (\min(m, p))^{-c_3 k_2},
\end{aligned}$$

where c_3 is a constant. Therefore, the probability that there exists a matrix $A \in \mathcal{A}_2$ such that for all x_i, y_i we have $x_i^T A y_i =_p 0$ is

$$\begin{aligned}
\mathbf{Pr}[\exists A \in \mathcal{A}_2, \forall i \in [k_2] : x_i^T A y_i =_p 0] &\leq \frac{|\mathcal{A}_2|}{(\min(m,p))^{c_3 k_2}} \\
&\leq \frac{\binom{n^2}{m} p^m}{(\min(m,p))^{c_3 k_2}} \\
&\leq \frac{\left(\frac{n^2}{m}\right)^m (ep)^m}{\left(\frac{n^2}{m}\right)^{c_3 C m} \min(m,p)^{c_3 C m}}.
\end{aligned}$$

For $p < m$, since $m < n^2/2$, the above is less than 1 for $c_3 C > 3$. For $p \geq m$ we get

$$\frac{\left(\frac{n^2}{m}\right)^m (ep)^m}{\left(\frac{n^2}{m}\right)^{c_3 C m} \min(m,p)^{c_3 C m}} = \frac{\left(\frac{n^2}{m}\right)^m (ep)^m}{n^{2 c_3 C m}} \leq \frac{n^{(2+2\gamma)m}}{n^{c_3 C m}} < 1,$$

for $C > (2+2\gamma)/c_3$. Thus, the results follows.

Case 3: $A \in \mathcal{A}_3$.
Let $S_2 = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{k_2}, y_{k_2})\}$ be the set of vectors we proved their existence in Case 2. Define $S_3 = \{(y_1, x_1), (y_2, x_2), \ldots, (y_{k_2}, x_{k_2})\}$. We argue that S_3 is the desired set we are looking for. For any $A \in \mathcal{A}_3$ we have that $A^T \in \mathcal{A}_2$. Therefore, there exist i such that

$$x_i^T A^T y_i \neq 0.$$

Note that

$$0 \neq x_i^T A^T y_i = (x_i^T A^T y_i)^T = y_i^T A x_i.$$

Thus,

$$0 \neq y_i^T A x_i.$$

This implies the result. $\square$

Now we can prove our main result

Corollary 2. *There exists a matrix $M \in \{0,1\}^{k \times n}$ where*

$$k = O\left(m + \frac{m \log \frac{n}{m}}{\log \min(m,p)}\right),$$

every row of M is a tensor product of two vectors $x, y \in \{0,1\}^{\sqrt{n}}$ and every m columns of M are linearly independent over $\mathbb{Z}_p$.

In particular the same matrix is $(0,1)$-matrix with m-independent columns over any field of characteristic p and over the real field $\mathbb{R}$.

Proof. Assume n is a perfect square. Let $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}$ be the set we found in Theorem 11 with vectors in $\{0,1\}^{\sqrt{n}}$. Define the matrix M where the ith row is $x_i \otimes y_i$. We argue that every m columns of M are linearly independent. Suppose on the contrary that there is a set of columns $M_{i_1}, M_{i_2}, \dots, M_{i_m}$ that are linearly dependent. Then, there are constants $\alpha_1, \dots, \alpha_m$ that are not all equal to 0 such that

$$\alpha_1 M_{i_1} + \alpha_2 M_{i_2} + \cdots + \alpha_m M_{i_m} = 0^k.$$

Define the following matrix $A \in \mathbb{Z}_p^{\sqrt{n}\times\sqrt{n}}$: For every column's index i_j let the entry (u, v) of the matrix be equal to α_j where $u = \lfloor (i_j - 1)/\sqrt{n} \rfloor + 1$ and $v = (i_j - 1 \mod \sqrt{n}) + 1$. All other entries are zero. It is easy to see that

$$x_i^T A y_i$$

equals the ith entry of the vector $\alpha_1 M_{i_1} + \alpha_2 M_{i_2} + \cdots + \alpha_m M_{i_m}$. Therefore we get that

$$x_i^T A y_i = 0,$$

for all $i \in [k]$. Since $A \neq 0^{n\times n}$ and $wt(A) \leq m$ we get a contradiction. □

Corollary 3. *There exists a set $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}$ where $x_i, y_i \in \{0,1\}^n$ and*

$$k = O\left(\frac{m \log n}{\log m}\right)$$

where for any matrix $A \in \mathbb{R}^{n\times n}$ such that $wt(A) \leq m$ and $A \neq 0^{n\times n}$, there exists an i such that $x_i^T A y_i \neq 0$.

Proof. Again, we argue that the set S found in Theorem 11 is the desired set. Let A be a matrix, let $A^{(i)}$ denote the ith row. Define the n^2-vector

$$A^v = [A^{(1)} | A^{(2)} | \cdots | A^{(n)}].$$

Then, for any $x, y \in \{0,1\}^n$ we have

$$x^T A y = (x \otimes y)^T A^v.$$

Define the matrix M where the ith row is $x_i \otimes y_i$. In the previous corollary we showed that every m columns of M are linearly independent over $\mathbb{R}$. Now, suppose that there exists a matrix A such that $wt(A) \leq m$ and for all $i \in [k]$ we have

$$x_i^T A y_i = 0.$$

Since $x^T A y = (x \otimes y)^T A^v$ and $x_i^T A y_i = 0$ for all $i \in [k]$ we get that

$$M A^v = 0^k.$$

This is a contradiction since $wt(A^v) = wt(A) \leq m$ and every m columns of M are linearly independent. □

Consider the following problem of reconstructing weighted graphs using additive queries [G98, GK98, GK00, BGK05, RS07, CK08, BM09]: Let $G = (V, E, w)$ be a weighted hidden graph where $E \in V \times V$, $w : E \to \mathbb{R}$ and n is the number of vertices in V. Denote by m the size of E. Suppose that the set of vertices V is known and the set of edges E is unknown. Given a set of vertices $S \subseteq V$, an additive query, $Q(S)$, returns the sum of weights in the subgraph induces by S. That is,

$$Q(S) = \sum_{e \in E \cap (S \times S)} w(e).$$

Our goal is to exactly reconstruct the set of edges and find their weights using additive queries.

Consider a variable x_i for each node $v_i \in V$. Define for each subset of vertices $V' \subseteq V$ a $\{0,1\}$-vector $a_{V'}$ where $a_{V'i} = 1$ if and only if $v_i \in V'$. Consider the matrix A_G where $A_G[i,j] = w((v_i, v_j))$ if and only if $(v_i, v_i) \in E$ and $A_G[i,j] = 0$ otherwise. It is easy to see that

$$a_{V'}^T A_G a_{V'} = 2 \cdot Q(V').$$

So the Q oracle is equivalent to the assignment oracle of the function $f_{A_G}(x) = x^T A_G x$ over the domain $\{0,1\}^n$. The problem now is to reconstruct a symmetric matrix A using the assignment oracle to $f_A(x) = x^T A x$ over the domain $x \in \{0,1\}^n$.

Grebinski and Kucherov, [G98, GK00], show that for any symmetric matrix A one can turn this oracle to an oracle to $f_A(x, y) = x^T A y$ in 5 queries. The following, [P], shows that 4 queries are sufficient: Let $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ and define $x \wedge y = (x_1 y_1, x_2 y_2, \ldots, x_n y_n)$, $x \vee y = (x_1 + y_1 - x_1 y_1, x_2 + y_2 - x_2 y_2, \ldots, x_n + y_n - x_n y_n)$ and $\bar{x} = (1 - x_1, 1 - x_2, \ldots, 1 - x_n)$. Then

$$x^T A y =$$

$$\frac{(x \vee y)^T A (x \vee y) + (x \wedge y)^T A (x \wedge y) - (x \wedge \bar{y})^T A (x \wedge \bar{y}) - (\bar{x} \wedge y)^T A (\bar{x} \wedge y)}{2}.$$

We now prove

Corollary 4. *There exists a non-adaptive algorithm that uses*

$$k = O\left(\frac{m \log n}{\log m}\right)$$

additive queries and reconstruct any weighted hidden graph with at most m edges.

Proof. From Corollary 3 it follows that that there exists a set $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\}$ where $x_i, y_i \in \{0,1\}^n$ and

$$k = O\left(\frac{m \log n}{\log m}\right)$$

where for any matrix $A \in \mathbb{R}^{n\times n}$ such that $wt(A) \le 4m$ and $A \ne 0^{n\times n}$, there exists an i such that $x_i^T A y_i \ne 0$. Now we use (x_i, y_i) to find $z_i = x_i^T A_G y_i$. We claim that the answers $(z_i)_i$ uniquely determines A_G. Otherwise, there are two weighted graphs $G \ne G'$ with at most m edges such that for all i, $x_i^T A_G y_i = x_i^T A_{G'} y_i$. This implies that for every i, $x_i^T (A_G - A_{G'}) y_i = 0$. Since $1 \le wt(A_G - A_{G'}) \le 4m$ we get a contradiction. □

5 $(0,1)$-Matrices with Rows that are Tensor Product of Vectors

In this section we show that there is an m-independent column $t \times n$ $(0,1)$-matrix that its rows are tensor product of d $(0,1)$-vectors in $\{0,1\}^{n^{1/d}}$ and

$$t = O\left(m + \frac{m \log \frac{n}{m}}{\log \min(m,p)}\right).$$

A d-dimensional matrix A of size $n_1 \times \cdots \times n_d$ over a field F is a map $A : \prod_{i=1}^d [n_i] \to F$. We denote by $F^{n_1\times\cdots\times n_d}$ the set of all d-dimensional matrices A of size $n_1 \times \cdots \times n_d$. We write $A_{i_1,\ldots,i_d}$ for $A(i_1,\ldots,i_d)$. The zero map is denoted by $0^{n_1\times\cdots\times n_d}$. For $I_j \subseteq [n_j]$, the matrix $B = (A_{i_1,i_2,\ldots,i_d})_{i_1\in I_1, i_2\in I_2,\ldots,i_d\in I_d}$ is the $|I_1|\times\cdots\times|I_d|$ matrix where $B_{j_1,\ldots,j_d} = A_{\ell_1,\ldots,\ell_d}$ where ℓ_i is the j_ith smallest number in I_i. When $I_j = [n_j]$ we just write j and when $I_j = \{\ell\}$ we just write $j = \ell$. For example $(A_{i_1,i_2,\ldots,i_d})_{i_1, i_2=\ell, i_3\in I_2,\ldots,i_d\in I_d} = (A_{i_1,i_2,\ldots,i_d})_{i_1\in[n_1], i_2\in\{\ell\}, i_3\in I_2,\ldots,i_d\in I_d}$

When $n_1 = n_2 = \cdots = n_d = n$ then we denote $F^{n_1\times\cdots\times n_d}$ by $F^{\times_d n}$ and $0^{n_1\times\cdots\times n_d}$ by $0^{\times_d n}$. For d-dimensional matrix A we denote by $wt(A)$ the number of points in $\prod_{i=1}^d [n_i]$ that are mapped to non-zero elements in F. For d-dimensional matrix A of size $n_1 \times \cdots \times n_d$ and $x_i \in F^{n_i}$ we define

$$A(x_1,\ldots,x_d) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} A_{i_1,i_2,\ldots,i_d} x_{1i_1} \cdots x_{di_d}.$$

The vector $v = A(\cdot, x_2, \ldots, x_d)$ is n_1-dimensional vector that its i_1 entry is

$$\sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} A_{i_1,i_2,\ldots,i_d} x_{2i_2} \cdots x_{di_d}.$$

We first prove the following:

Theorem 14. *Let $p < n^\gamma$ be a prime number for some constant γ. There exists a set $S = \{(x_{11},\ldots,x_{1d}), (x_{21},\ldots,x_{2d}), \ldots, (x_{k1},\ldots,x_{kd})\}$ where $x_{ij} \in \{0,1\}^n$ and*

$$k = O\left(m + \frac{m \log \frac{n^d}{m}}{\log \min(m,p)}\right),$$

such that: for any d-dimensional matrix $A \in \mathbb{Z}_p^{\times_d n} \backslash \{0^{\times_d n}\}$ with $wt(A) \le m$, there exists an i such that

$$A(x_{i1}, \ldots, x_{id}) \neq_p 0.$$

Proof. We divide the set of matrices

$$\mathcal{A} = \{A \mid A \in \mathbb{Z}_p^{\times_d n} \backslash \{0^{\times_d n}\} \text{ and } wt(A) \le m\}$$

into $d + 1$ (non-disjoint) sets:

- $\mathcal{A}_0$: The set of all matrices $A \in \mathbb{Z}_p^{\times_d n} \backslash \{0^{\times_d n}\}$ such that $wt(A) \le m/\log m$.
- $\mathcal{A}_j, j = 1, \ldots, d$: The set of all matrices $A \in \mathbb{Z}_p^{\times_d n}$ such that $m \ge wt(A) > m/\log m$ and there are at least

  $$\left(\frac{m}{\log m}\right)^{1/d}$$

  non-zero elements in

  $$I_j = \{i_j \mid \exists (i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d)\ A_{i_1, i_2, \ldots, i_d} \neq 0\}.$$

Note that $I = \{(i_1, \ldots, i_j) \mid A_{i_1, \ldots, i_j} \neq 0\} \subseteq I_1 \times I_2 \times \cdots \times I_d$ and therefore either $|I| = wt(A) \le m/\log m$ or there is j such that $|I_j| > (m/\log m)^{1/d}$. Therefore, $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_d$.

Using the probabilistic method, we give $d+1$ sets of d-tuples of vectors $S_0, S_1, \ldots, S_d$ such that for every $j \in \{0\} \cup [d]$ and $A \in \mathcal{A}_j$ there exists a d-tuple of vectors $(x_1, \ldots, x_d) \in S_j$ such that $A(x_1, \ldots, x_d) \neq_p 0$ and

$$|S_0| + |S_1| + \cdots + |S_d| = O\left(m + \frac{m \log \frac{n^d}{m}}{\log \min(m, p)}\right).$$

Case 1: $A \in \mathcal{A}_0$.
As in the proof of Lemma 2 it can be shown that for randomly chosen vectors $x_{i1}, \ldots, x_{id} \in \{0,1\}^n$ and $A \in \mathcal{A}_0$ we have

$$\mathbf{Pr}[A(x_{i1}, x_{i2}, \ldots, x_{id}) =_p 0] \le \frac{2^d - 1}{2^d}.$$

Randomly uniformly choose

$$k_1 = c\left(m + \frac{m \log \frac{n^d}{m}}{\log \min(m, p)}\right)$$

d-tuples of $(0,1)$-vectors $x_i = (x_{i1}, \ldots, x_{id}) \in (\{0,1\}^n)^d$ where c is a constant. The probability that for all x_i we have $A(x_i) =_p 0$ is bounded by

$$\mathbf{Pr}[\forall i \in [k_1] : A(x_i) =_p 0] \le \left(\frac{2^d - 1}{2^d}\right)^{k_1}.$$

Therefore, by union bound, the probability that there exists a matrix A of weight smaller than $m/\log m$ such that $A(x_i) =_p 0$ for all $i \in [k_1]$ is

$$\mathbf{Pr}[\exists A \in \mathcal{A}_0, \forall i \in [k_1] : A(x_i) =_p 0] \leq \binom{n^d}{\frac{m}{\log m}} p^{\frac{m}{\log m}} \left(\frac{2^d - 1}{2^d}\right)^{k_1} < 1,$$

for some constant c. This implies the result.

Case 2: $A \in \mathcal{A}_1$.
We start by proving the following two lemmas

Lemma 15. *Let $U \subset \mathbb{Z}_p^{\times_{d-1} n}$ be the set of all non-zero $d-1$-dimensional matrices with weight smaller than $m^{d/(d+1)}$. Then there is a constant c_0 such that for any constant $C > c_0$ and*

$$k_2 = C\left(m + \frac{m \log \frac{n^d}{m}}{\log \min(m, p)}\right)$$

there exists a multiset of $d-1$-tuple of (0,1)-vectors $Y = \{y_1, y_2, \ldots, y_{k_2}\} \subseteq (\{0,1\}^n)^{d-1}$ such that for every $A \in U$ the size of the multiset

$$Y_A = \{i \mid A(y_i) \neq_p 0\}$$

is at least $\frac{k_2}{2^d}$.

Proof. As above for a randomly chosen vector $y \in (\{0,1\}^n)^{d-1}$ and any $A \in U$ we have

$$\mathbf{Pr}[A(y) =_p 0] \leq \frac{2^{d-1} - 1}{2^{d-1}}.$$

Therefore, if we randomly uniformly choose the vectors of Y, then the expected size of Y_u is greater than $k_2/2^{d-1}$ for any $A \in U$. Using Chernoff bound (Lemma 6) we have that

$$\mathbf{Pr}[|Y_A| < k_2/2^d] \leq e^{\frac{k_2}{2^{d+2}}}.$$

Therefore, the probability that there exists $A \in U$ such that $|Y_A| < k_2/2^d$ is

$$\mathbf{Pr}[\exists A \in U : |Y_A| < k_2/2^d] \leq \frac{|U|}{e^{\frac{k_2}{2^d}}} \leq \frac{\sum_{i=0}^{m^{d/(d+1)}} \binom{n^{d-1}}{i}(p-1)^i}{e^{\frac{k_2}{2^d}}} < 1,$$

for some constant c_0 and all $C > c_0$. This implies the result. $\square$

Now let U be the set of $d-1$-dimensional matrices defined in Lemma 15. Let $A \in \mathcal{A}_1$. Since $wt(A) \leq m$ there are at most $m^{1/(d+1)}$ $d-1$-dimensional matrices $(A_{i_1,i_2,\ldots,i_d})_{i_1=j,i_2,\ldots,i_d}$ with weight greater than $m^{d/(d+1)}$. Therefore, there are at least

$$q = \left(\frac{m}{\log m}\right)^{1/d} - m^{1/(d+1)}$$

indices j such that $(A_{i_1,i_2,\ldots,i_d})_{i_1=j,i_2,\ldots,i_d} \in U$. Let U' contains q indices j such that $(A_{i_1,i_2,\ldots,i_d})_{i_1=j,i_2,\ldots,i_d} \in U$. Let A_U be the matrix $(A_{i_1,i_2,\ldots,i_d})_{i_1\in U',i_2,\ldots,i_d}$. Let $Y = \{y_1, y_2, \ldots, y_{k_2}\}$ be the set we proved its existence in Lemma 15. Note that

$$\sum_i wt(A_U(\cdot, y_i)) \geq \frac{qk_2}{2^d}.$$

Since $wt(A_U(\cdot, y_i)) \leq q$ for all $i \in [k_2]$ and by Lemma 13 we have

$$\prod_i \iota(wt(A_U(\cdot, y_i))) \geq (\min(q,p))^{\left\lfloor \frac{\frac{qk_2}{2^d} - k_2}{q-1} \right\rfloor} = (\min(q,p))^{c_1 k_2} = (\min(m,p))^{c_2 k_2},$$

where c_1 and c_2 are constants. If we randomly choose $x_1, x_2, \ldots, x_{k_2}$ then by Lemma 3, we have

$$\begin{aligned}
\mathbf{Pr}[\forall i \in [k_2] : A(x_i, y_i) \neq 0] &\leq \prod_i \frac{1}{\iota(wt(A(\cdot, y_i)))^\beta}, \\
&\leq \prod_i \frac{1}{\iota(wt(A_U(\cdot, y_i)))^\beta} \\
&= \left(\frac{1}{\prod_i \iota(wt(A_U(\cdot, y_i)))} \right)^\beta \\
&\leq \frac{1}{(\min(m,p))^{\beta c_2 k_2}} \\
&= (\min(m,p))^{-c_3 k_2}
\end{aligned}$$

where c_3 is a constant. Therefore, the probability that there exists a matrix $A \in \mathcal{A}_1$ such that for all x_i, y_i we have $A(x_i, y_i) =_p 0$ is

$$\mathbf{Pr}[\exists A \in \mathcal{A}_1, \forall i \in [k_2] : A(x_i, y_i) =_p 0] \leq \frac{|\mathcal{A}_1|}{(\min(m,p))^{c_3 k_2}} \leq \frac{\binom{n^d}{m} p^m}{(\min(m,p))^{c_3 k_2}} < 1,$$

for constant some constant C. Thus, the results follows.

□

Now we get our main result

Corollary 5. *There exists a matrix $M \in \{0,1\}^{k \times n}$ where*

$$k = O\left(m + \frac{m \log \frac{n}{m}}{\log \min(m,p)} \right),$$

every row of M is a tensor product of d vectors $x_1, \ldots, x_d \in \{0,1\}^{n^{1/d}}$ and every m columns of M are linearly independent over $\mathbb{Z}_p$.

References

[A86] M. Aigner. Search problems on graphs. *Discrete Applied Mathematics*, V. 14 , I. 3, pp. 215 - 230 (1986).

[A88] M. Aigner. Combinatorial search. John Wiley and Sons, 1988.

[AS85] M. Aigner and M. Schughart. Determining defectives in a linear order. *J. Statist. Plan. Inform.* 12 pp. 359-368 (1985).

[B01] B. Bollobás, Random Graphs, second ed., Cambridge Stud. Adv. Math., vol. 73, Cambridge Univ. Press, Cambridge, 2001.

[BG07] E. Biglieri and L. Györfi. Multiple Access Channels Theory and Practice Volume 10 NATO Security through Science Series - D: Information and Communication Security, April 2007.

[BG08] L. Bruneau and F. Germinet. On the singularity of random matrices with independent entries *Proc. Amer. Math. Soc.* 137 (2009), 787-792.

[BGK05] M. Bouvel, V. Grebinski, G. Kucherov. Combinatorial Search on Graphs Motivated by Bioinformatics Applications: A Brief Survey. WG 2005, pp. 16-27, 2005.

[BM09] N. H. Bshouty and H. Mazzawi. Reconstructing Weighted Graphs with Minimal Query Complexity. ALT 2009.

[C80] C. Christen. A Fibonaccian algorithm for the detection of two elements. Publ. 341, Dept. d'IRO, Univ. Montreal. (1980)

[CK08] S. S. Choi and J. H. Kim. Optimal query complexity bounds for finding graphs. STOC 2008, pp. 749-758.

[D75] A. G. Djackov. On a search model of false coins. In Topics in Information Theory (Colloquia Mathematica Societatis Janos Bolyai 16, Keszthely, Hungary). Budapest, Hungary: Hungarian Acad. Sci., pp. 163170, (1975).

[DH93] D. Du and F. K. Hwang. Combinatorial group testing and its application, Volume 3 of Series on applied mathematics. World Science, 1993.

[E45] P. Erdös. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51, 898–902, 1945.

[G98] V. Grebinski, On the power of additive combinatorial search model. In Proceedings of the Fourth Annual International Computing and Combinatorics Conference (COCOON'98), vol. 1449 of LNCS, pp. 194-203. Springer, 1998.

[GK98] V. Grebinski and G. Kucherov. Reconstructing a hamiltonian cycle by querying the graph: Application to DNA physical mapping. *Discrete Applied Mathematics*, 88: 147–165, (1998).

[GK00] V. Grebinski and G. Kucherov. Optimal reconstruction of graphs under the additive model. Algorithmica 28(1): 104-124, (2000)

[K67] J. Komlós, On the determinant of matrices, *Studia. Sci. Math. Hungar.*. 2, 7-21 (1967).

[L65] B. Lindström. On a combinatorial problem in number theory. *Canad. Math. Bull.*, 8: 477-490, 1965.

[L71] B. Lindström. On Möbius functions and a problem in combinatorial number theory. *Canad. Math. Bull.*, 14(4): 513-516, 1971.

[L75] B. Lindström. Determining subsets by unramified experiments. In J.N. Srivastava, editor, *A Survey of Statistical Designs and Linear Models*, pp. 407-418. North Holland, Amsterdam, 1975.

[LO43] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. III. *Mat. Sbornik*, 12, 277–285, 1943.

[P] P. Long. Private Communication.

[MR95] R. Motwani and P. Raghavan. Randomized algorithms. Cambridge University Press 1995.

[RS07] L. Reyzin, N. Srivastava. Learning and Verifying Graphs Using Queries with a Focus on Edge Counting. ALT 2007, LNAI 4754, 285-297, 2007.

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. Journal of the ACM, 27:701-717, 1980.

[S28] E. Sperner, Ein Satz ber Untermengen einer endlichen Menge. Math. Z. 27, 544-548, 1928.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de