The Thesis Committee for Yangxinyu Xie

Certifies that this is the approved version of the following Thesis:

Matrix Rigidity: A Survey

APPROVED BY

SUPERVISING COMMITTEE:

Anna Gál, Supervisor

Ngoc Mai Tran

Matrix Rigidity: A Survey

by

Yangxinyu Xie

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCES IN COMPUTER SCIENCE

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2022

Matrix Rigidity: A Survey

Yangxinyu Xie, M.S.C.S.
The University of Texas at Austin, 2022

Supervisor: Anna Gál

Since Valiant's establishment of matrix rigidity to analyse circuit complexity, various contributions to the bounds of matrix rigidity of special candidates has bloomed. In this thesis, we cover the following topics: existing explicit and semi-explicit rigidity lower bounds for various families of matrices, Paturi-Pudlák dimensions, rigid sets, connections between data structure lower bounds and rigidity lower bounds and non-rigidity results.

Table of Contents

Chapter 1

Motivation and Definition

The definition of matrix rigidity was first introduced by Valiant [Val77].

Definition 1.0.1. The **density** of a matrix A is the number of nonzero elements drawn from a field $\mathbb{F}$, denoted by $\text{dens}(A)$.

Definition 1.0.2. The **rigidity** of a matrix A is the function $\mathcal{R}_A^{\mathbb{F}}(r) : \{1, ..., n\} \to \{0, 1, ..., n^2\}$ defined by

$$\mathcal{R}_A^{\mathbb{F}}(r) := \min\{i | \exists\ B, \text{dens}(B) = i, \text{rank}(A + B) \leq r\}$$

Valiant motivated the definition of matrix rigidity from the analysis of circuit complexity and proved that if $\mathcal{R}_A^{\mathbb{F}}(\varepsilon n) = n^{1+\delta}$ for some $\varepsilon, \delta > 0$, then the corresponding linear transformation $x \to Ax$(which was called "linear program" in the original paper) cannot be computed by circuits of size $O(n \log\log n)$ and depth $O(\log n)$.

With slight modifications of the original proof of the previous statement, we have that if $\mathcal{R}_A^{\mathbb{F}}(n/\log\log n) = n^{1+\delta}$ for some $\delta > 0$, then the corresponding linear transformation $x \to Ax$ cannot be computed by circuits of size $O(n)$ and depth $O(\log n)$. In fact, the matrices satisfying

$$\mathcal{R}_A^{\mathbb{F}}(n/\log\log n) \geq n^{1+\delta}$$

are later called "Valiant-rigid" in recent papers [GT18, AC19, BHPT20].

Moreover, Valiant gave a non-constructive proof for the following; that is, a random matrix is rigid with high probability.

Theorem 1.0.1 ([Val77]). *1. For an infinite field $\mathbb{F}$, for all n, there exists a $n \times n$ matrix A such that $\mathcal{R}_A^{\mathbb{F}}(r) = (n-r)^2$.*

2. For a finite field $\mathbb{F}$ with c elements, for all n, there exists a $n \times n$ matrix A such that for all $r < n - \sqrt{2n \log_c 2 + \log_2 n}$,

$$\mathcal{R}_A^{\mathbb{F}}(r) \geq \frac{(n-r)^2 - 2n \log_c 2 - \log_2 n}{2 \log_c n + 1}$$

The main challenge is to find *explicit* "Valiant-rigid" matrices A satisfying

$$\mathcal{R}_A^{\mathbb{F}}(n/\log\log n) = n^{1+\delta}$$

which has remained a major barrier since it was first proposed in the 1970's. What do we mean by explicit? For a family of matrices $\{A_n\}$, we say that it is **explicit** if for each n, given indices $1 \leq i, j \leq n$, there is a Boolean circuit of size polynomial in n to compute A_{ij}. Explicit constructions of such matrices lead to explicit problems that cannot be solved in $O(n)$-size $O(\log n)$-depth circuits. In this survey, we will review the past attempts and perennial difficulties in tackling this challenge.

Chapter 2

Explicit Lower Bounds

In this chapter, we will cover the past techniques for proving explicit lower bounds of matrix rigidity. For explicit lower bounds, most of the proofs consist of two steps: first, we show that most sub-matrices, also called minors, of the given matrix M has large or full rank r; second, if $\mathcal{R}_A^{\mathbb{F}}(r)$ is small, we are likely to get a sub-matrix that remains intact. The first step implies that it is plausible to find rigid matrices with high regularity.

Definition 2.0.1 ([BCS97])**.** A matrix A is called **totally regular** if and only if every minor of A is invertible.

A relaxation of this definition will suffice for our purpose.

Definition 2.0.2. A matrix A is called **almost totally regular** if and only if every $r \times r$ minor of A has rank $\Omega(r)$.

Another notion of regularity based on expectation is introduced by Pudlak [Pud94].

Definition 2.0.3 ([Pud94])**.** Let A be an $n \times n$ matrix, $0 \leq \varepsilon, \delta, \eta \leq 1$. We say that A is $(\varepsilon, \delta, \eta)$–**densely regular**, if for every k with $\eta n \leq k \leq n$, there are nonempty sets of k elements subset $\mathcal{X}, \mathcal{Y} \subseteq \{1, ..., n\}^k$ such that for every $i, j = 1, ..., n$

$$\delta \mathbb{P}[i \in X] \leq k/n \quad \text{and} \quad \delta \mathbb{P}[j \in Y] \leq k/n$$

where $X \in \mathcal{X}, Y \in \mathcal{Y}$ are chosen with some probability distributions and such that for random $X \in \mathcal{X}, Y \in \mathcal{Y}$ the mean value of the rank of the matrix determined by X and Y is at least εk.

Again, a relaxation of this will also be sufficient to us.

Definition 2.0.4 ([Che05])**.** Let A be an $n \times n$ matrix. We say that A is **ε-densely regular**, if there is a constant $0 < \varepsilon < 1$ such that for every k with $0 \leq k \leq n$, a $k \times k$ minor of A picked uniformly at random has an expected rank at least εk.

However, as pointed out by Lokam [Lok00], any proof relying on the second step cannot produce a lower bound better than $\Omega((n^2/r)\log(n/r))$. Moreover, due to the existence of linear size superconcentrators, the first step is far from sufficient to show a desirable rigidity.

Proposition 2.0.1 ([Val77])**.** *For each n there is an $n \times n$ totally regular matrix A such that*

$$\mathcal{R}_A^{\mathbb{F}}\left(\frac{n \log\log\log n}{\log\log n}\right) \leq n^{1+O\left(\frac{1}{\log\log n}\right)}$$

Nevertheless, finding explicit matrices with high regularity is still a reasonable start.

2.1 Totally Regular Matrices

Recall that a matrix A is called **totally regular** if and only if every minor of A is invertible. The following combinatorial lemma, along with its corollary, says that if not many changes are made to the matrix A, then there

will be a large enough submatrix that remains intact. If A is totally regular, then the intact submatrix, or "untouched minor", will have full rank. With this result at hand, it remains to find families totally regular matrices, which include Cauchy matrices, Fourier transform matrices and generator matrices of asymptotically good codes [SSS97, Lok00]. The proofs of this section are mainly taken from [Lok09].

Lemma 2.1.1 ([SSS97])**.** *If fewer than*

$$\mu(n,r) = (n-r+1)(n-(r-1)^{1/r}n^{1-1/r})$$

entries of an $n \times n$ matrix A is marked, then there is an $r \times r$ submatrix that remains intact.

Proof. Think of A as the adjacency matrix of a bipartite graph $G_{n,n}$ which contains an edge (i,j) if and only if $A_{i,j}$ has not been marked. Hence, having an $r \times r$ submatrix that remains intact is equivalent to having a $K_{r,r}$ complete bipartite subgraph in $G_{n,n}$. Hence, we have that $G_{n,n}$ cannot have more than $n^2 - \mu(n,r)$ edges, or it will contain a $K_{r,r}$ complete bipartite subgraph (see [Juk11], Theorem 2.10). □

Corollary 2.1.2. *Let $r \geq \log^2 n$ and let n be sufficiently large. If fewer than*

$$\frac{n(n-r+1)}{2r}\log\frac{n}{r-1}$$

changes are made to an $n \times n$ matrix A, then there exists an $r \times r$ submatrix that remains intact.

It is easy to see that if any $r \times r$ minor has rank $\Omega(r)$, we have that the rigidity of A is $\Omega(\frac{n^2}{r}\log\frac{n}{r})$.

2.1.1 Cauchy Matrix

Definition 2.1.1. Let $x_1, ..., x_n, y_1, ..., y_n$ be elements of a field $\mathbb{F}_n$ with the property that

$$\prod_{i \neq j}(x_i - x_j) \neq 0, \quad \prod_{i \neq j}(y_i - y_j) \neq 0, \quad \prod_{i,j}(x_i + y_j) \neq 0$$

we define the **Cauchy matrix** by

$$C := \Big(\frac{1}{x_i + y_j}\Big)_{1 \leq i,j \leq n}$$

Hence, for every $1 \leq r \leq n$, each of its $r \times r$-submatrix has the determinant

$$\frac{\prod_{i \neq j}(x_i - x_j) \prod_{i \neq j}(y_i - y_j)}{\prod_{i,j}(x_i + y_j)}$$

which is nonzero. In other words, the Cauchy matrix is totally regular.

The following theorem is thus a direct result of Corollary 2.1.2.

Theorem 2.1.3 ([SSS97])**.** *Let $\mathbb{F}_n$ be a sequence of fields and let (C_n) be a sequence of Cauchy matrices where $C_n \in \mathbb{F}_n^{n \times n}$. Then if $\log^2 n \leq r \leq n/2$, we have*

$$\mathcal{R}_{C_n}^{\mathbb{F}_n}(r) \geq \Big(\frac{n^2}{4r} \log \frac{n}{r-1}\Big)$$

2.1.2 Fourier Transform Matrix

Another type of totally regular matrices is the **discrete Fourier transform matrices**. Hence, the following result is also a direct consequence of Corollary 2.1.2.

Theorem 2.1.4 ([Lok00, Lok09])**.** *Let $F = (\omega_i^{j-1})_{i,j=0}^{n-1}$, where ω is a primitive nth root of unity. Then, as n ranges over all prime numbers and $\log^2 n \leq r \leq$*

$n/2$,

$$\mathcal{R}_F(r) \geq \frac{n^2}{4(r+1)} \log \frac{n}{r}$$

2.1.3 Asymptotically Good Error Correcting Codes

The existence of asymptotically good error correcting codes is a corollary of the Tsfasman-Vladut-Zink Bound [TVZ82], whose statements and proofs are not presented here; we refer the reader to [Lok09] for more details. We simply use the following claim as given and show an explicit lower bound of matrix rigidity.

Claim 2.1.1. For infinitely many n, there exists a $[2n, n, d]$-code with $d \geq (1-\varepsilon)n$ where $\varepsilon = 2/(\sqrt{q}-1)$.

For a given n, let Γ be the $[2n, n, d]$-code as in Claim 2.1.1, whose generator matrix has the form $(I_n|A)$ where I_n is the $n \times n$ identity matrix.

Theorem 2.1.5 ([SSS97])**.** *Let A be an $n \times n$ matrix as defined above. Then, for* $\max(\log^2 n, \varepsilon n) \leq r \leq n/4$,

$$\mathcal{R}_A^{\mathbb{F}_q}(r) \geq \frac{n^2}{8r} \log \frac{n}{2r-1}$$

Proof. We first show that for all $2r \times 2r$ submatrix of A, the rank must be at least r. Suppose on the contrary, then let B be a $2r \times 2r$ submatrix of A with r dependent rows. Then a linear combination of these r rows of the generator matrix gives a code word of weight

$$r + n - 2r = n - r \leq (1-\varepsilon)n - 1 \leq d$$

where the first r comes from the identity matrix and the $n - 2r$ comes from the fact that these r rows are dependent. Hence, we reach a contradiction. By Corollary 2.1.2, we obtain the desired result. □

2.2 Densely Regular Matrices

Recall the definition of densely regular matrices:

Definition 2.2.1 ([Pud94])**.** Let A be an $n \times n$ matrix, $0 \leq \varepsilon, \delta, \eta \leq 1$. We say that A is $(\varepsilon, \delta, \eta)$–**densely regular**, if for every k with $\eta n \leq k \leq n$, there are nonempty sets of k elements subset $\mathcal{X}, \mathcal{Y} \subseteq \{1, ..., n\}^k$ such that for every $i, j = 1, ..., n$

$$\delta \mathbb{P}[i \in X] \leq k/n \quad \text{and} \quad \delta \mathbb{P}[j \in Y] \leq k/n$$

where $X \in \mathcal{X}, Y \in \mathcal{Y}$ are chosen with some probability distributions and such that for random $X \in \mathcal{X}, Y \in \mathcal{Y}$ the mean value of the rank of the matrix determined by X and Y is at least εk.

The following theorem shows are densely regular matrices are $\Omega(n^2/r)$ rigid. In fact, a similar proof can be applied to show that ε-densely regular matrices are $\Omega(n^2/r)$ rigid, which we omit here. Notice that this rigidity bound is weaker than that given by totally regular matrices. For the remainder of this section, we show that Vandermonde matrices and Hadamard matrices are densely regular, and thus $\Omega(n^2/r)$ rigid.

Theorem 2.2.1 ([Pud94])**.** *For every positive $\varepsilon, \delta, \eta, 0 < \varepsilon, \delta, \eta \leq 1$, any field $\mathbb{F}$, and any $n \times n$ $(\varepsilon, \delta, \eta)$–densely regular matrix A,*

$$\mathcal{R}_A^{\mathbb{F}}(r) = \Omega(n^2/r)$$

for $\varepsilon\eta n/2 \leq r \leq \varepsilon n/2$

Proof. Let $\varepsilon\eta n/2 \leq r \leq \varepsilon n/2$ and sets $\mathcal{X}, \mathcal{Y}$ be given and set $k = \lceil 2r/\varepsilon \rceil$. For random variables $X \in \mathcal{X}, Y \in \mathcal{Y},$, we have that for a coordinate (i, j),

$$\mathbb{P}[(i, j) \in X \times Y] \leq \frac{k^2}{\delta^2 n^2}$$

Let Z be the minimal set of coordinates changed to reduce the rank of A to r and let z be the number of coordinates in the set $Z \cap (X \times Y)$. Then, we have

$$\mathbb{E}[z] = \sum_{(i,j)\in Z} \mathbb{P}[(i,j) \in X \times Y] \leq |Z| \frac{k^2}{\delta^2 n^2}$$

Notice that on the other hand, the mean value of the rank of the matrix determined by X and Y is at least εk, which implies

$$\mathbb{E}[z] \geq \varepsilon k - r \geq 2r - r = r$$

Therefore, we have

$$r \leq \mathbb{E}[z] \leq |Z| \frac{k^2}{\delta^2 n^2}$$

and thus by plugging $k = \lceil 2r/\varepsilon \rceil$

$$|Z| \geq \frac{r\delta^2 n^2}{k^2} = \Omega(\frac{n^2}{r})$$

□

2.2.1 Vandermonde Matrix

Proposition 2.2.2 ([Pud94]). *Let $V = (x_i^{j-1})_{i,j=1}^n$ be a Vandermonde matrix with distinct x_i over some field. V is $(1, 1/2, 0)$-densely regular.*

Proof. Let $k \leq n$ be given and $l = \lfloor n/k \rfloor$. We define $\mathcal{X}$ as the set of sets of the form

$$\{a, a+l, a+2l, ..., a+(k-1)l\}$$

where $1 \leq a \leq l$. Let X be uniformly distributed on $\mathcal{X}$. This gives us

$$\mathbb{P}[i \in X] \leq \frac{1}{l} \leq \frac{2k}{n}$$

We also let $\mathcal{Y}$ be defined in the same way. This gives us $\delta = 1/2$. Notice that any minor of a Vandermonde matrix is full rank, which means $\varepsilon = 1$. □

By theorem 2.2.1, we prove the $\Omega(n^2/r)$ lower bound for Vandermonde matrices. An alternative proof is given by Shparlinsky, which we will present later (see Theorem 2.3.1.)

2.2.2 Hadamard Matrix

A matrix $H = (h_{i,j}) \in \mathbb{C}^{n\times n}$ is a (generalised) Hadamard matrix if $|h_{i,j}| = 1$ for all $i, j \in [n]$ and the rows of H are pairwise orthogonal. In this section, we present the lower bound by Razborov [KR98], whose proof took advantages of singular value decomposition and Frobenius norm. Before that, let's introduce some definitions.

Definition 2.2.2. A matrix $H = (h_{i,j}) \in \mathbb{C}^{n\times n}$ is called a **(generalised) Hadamard matrix** if $|h_{i,j}| = 1$ for all $i, j \in [n]$ and $HH^* = nI_n$ where H^* is the conjugate transpose of H and I_n is the $n \times n$ identity matrix.

Definition 2.2.3. The **Frobenius norm** of a matrix $A \in \mathbb{C}^{n\times n}$ is

$$||A||_F := \Big(\sum_{i,j} |a_{i,j}|^2\Big)^{1/2}$$

Definition 2.2.4. The **trace** of a matrix $A \in \mathbb{C}^{n\times n}$ is the sum of its eigenvalues:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i(A)$$

Definition 2.2.5. The i**th singular value** $\sigma_i(A)$ is defined by

$$\sigma_i(A) := \sqrt{\lambda_i(AA^*)}, 1 \le i \le n$$

where λ_i denotes the ith largest eigenvalue of AA^*.

We recall some fact about singular value decomposition and Frobenius norm. The proof can be found in chapter 2.4 in [GVL12].

Proposition 2.2.3. *For any matrix $A \in \mathbb{C}^{n\times n}$,*

- *there exists unitary matrices $U, V \in \mathbb{C}^{n\times n}$ such that*

$$U^*AV = \mathrm{diag}(\sigma_1(A), \sigma_2(A), ..., \sigma_n(A))$$

- $||A||_F^2 = \sigma_1^2(A) + \sigma_2^2(A) + ... + \sigma_n^2(A)$.

Proposition 2.2.4. *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then*

$$\frac{\mathrm{Tr}(A)^2}{||A||_F^2} \leq \mathrm{rank}(A)$$

Proof. Let $B = AA^*$. Notice that

$$\mathrm{Tr}(B) = \sum_{i=1}^{n} \lambda_i(B) = ||A||_F^2$$

Because A is symmetric,

$$\sum_{i=1}^{n} \lambda_i(B) = \mathrm{Tr}(B) = \sum_{i=1}^{n} \lambda_i(A^2) = \sum_{i=1}^{n} \lambda_i^2(A)$$

Moreover, B has only $\mathrm{rank}(B)$ =$\mathrm{rank}(A)$ non-zero eigenvalues, which are all positive. Assume without loss of generality $\lambda_1^2(A) \geq \lambda_2^2(A) \geq ... \geq \lambda_n^2(A)$. Then, $\sum_{i=1}^{n} \lambda_i^2(A) = \sum_{i=1}^{\mathrm{rank}(A)} \lambda_i^2(A)$. Hence, by Cauchy-Schwarz inequality, we have

$$||A||_F^2 = \sum_{i=1}^{\mathrm{rank}(A)} \lambda_i^2(A) \geq \frac{\left(\sum_{i=1}^{\mathrm{rank}(A)} \lambda_i(A)\right)^2}{\mathrm{rank}(A)} \geq \frac{\mathrm{Tr}(A)^2}{\mathrm{rank}(A)}$$

□

Proposition 2.2.5 ([KR98])**.** *Let H be an $n\times n$ generalised Hadamard matrix. Let G be a random $q\times n$ submatrix of H and let A be a random $q\times q$ submatrix of G. Then $\mathbb{E}[\mathrm{rank}(A)] \geq r/8$.*

Proof. Let $B = AA^*$. Then B is a positive definite symmetric matrix in $\mathbb{R}^{q\times q}$. Recall that $h_{i,j} = 1$, which implies all entries of B on the main diagonal equals to q. Thus $\mathrm{Tr}(B) = q^2$. By proposition 2.2.4, we obtain that $\mathrm{rank}(A) \leq r$ for some positive integer r implies

$$||B||_F^2 \geq \frac{\mathrm{Tr}(B)^2}{\mathrm{rank}(B)} \geq \frac{q^4}{r}$$

Let

$$\varepsilon_j = \begin{cases} 1 & \text{if the } j\text{th column of } H \text{ is in } H_0 \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\mathbb{E}[\varepsilon_{j_1}\varepsilon_{j_2}] = \begin{cases} \frac{q}{n} & \text{if } j_1 = j_2 \\ \frac{q(q-1)}{n(n-1)} & \text{if } j_1 \neq j_2 \end{cases}$$

Now, notice that $b_{i,j} = \sum_{k=1}^q a_{i,k}a_{j,k}^* = \sum_{k=1}^q g_{i,k}g_{j,k}^*\varepsilon_k$ and $b_{i,j}^* = \sum_{l=1}^q a_{i,l}^* a_{j,l} = \sum_{k=1}^q g_{i,k}g_{j,k}^*\varepsilon_l$

$$\begin{aligned} ||B||_F^2 &= \sum_{1\leq i,j\leq q} |b_{i,j}|^2 = \sum_{1\leq i,j\leq q} b_{i,j}b_{i,j}^* = \sum_{1\leq i\leq q, 1\leq j\leq n} \sum_{1\leq k,l\leq q} g_{i,k}g_{j,k}g_{i,l}^*g_{j,l}^*\varepsilon_k\varepsilon_l \\ &= \sum_{1\leq k,l\leq q} \Big(\varepsilon_k\varepsilon_l \sum_{1\leq i\leq q, 1\leq j\leq n} g_{i,k}g_{j,k}g_{i,l}^*g_{j,l}^*\Big) \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}[||B||_F^2] &= \sum_{1\leq k,l\leq q} \Big(\mathbb{E}[\varepsilon_k\varepsilon_l] \sum_{1\leq i\leq q, 1\leq j\leq n} g_{i,k}g_{j,k}g_{i,l}^*g_{j,l}^*\Big) \\ &= \frac{q(q-1)}{n(n-1)} \sum_{1\leq k,l\leq q} \sum_{1\leq i\leq q, 1\leq j\leq n} g_{i,k}g_{j,k}g_{i,l}^*g_{j,l}^* + \Big(\frac{q}{n} - \frac{q(q-1)}{n(n-1)}\Big) \sum_{1\leq i\leq q, 1\leq j\leq n} g_{i,k}g_{j,k}g_{i,l}^*g_{j,l}^* \\ &= \frac{q(q-1)}{n(n-1)} ||GG^*||_F^2 + \Big(\frac{q}{n} - \frac{q(q-1)}{n(n-1)}\Big) \sum_{k=1}^q \sum_{1\leq i\leq q, 1\leq j\leq n} g_{i,k}g_{i,k}^*g_{j,k}g_{j,k}^* \end{aligned}$$

Notice that $GG^* = nI_q$ where I_q is the $q \times q$ identity matrix.

$$\begin{aligned}
\mathbb{E}[||B||_F^2] &= \frac{q(q-1)}{n(n-1)}||GG^*||_F^2 + \Big(\frac{q}{n} - \frac{q(q-1)}{n(n-1)}\Big)\sum_{k=1}^{q}\sum_{1\le i\le q, 1\le j\le n} g_{i,k}g_{i,k}^*g_{j,k}g_{j,k}^* \\
&= \frac{q(q-1)}{n(n-1)}n^2q^2 + \Big(\frac{q}{n} - \frac{q(q-1)}{n(n-1)}\Big)nq^2 \\
&= q^2(q + (n-q)\frac{q-1}{n-1}) \le 2q^3
\end{aligned}$$

By Chebyshev's inequality, we have

$$\mathbb{P}[||B||_F^2 \ge \frac{q^4}{r}] \le \frac{r}{q^4}\mathbb{E}[||B||_F^2] \le \frac{2r}{q}$$

Thus, we have $\mathbb{P}[\text{rank}(A) \le r] \le \mathbb{P}[||B||_F^2 \ge \frac{q^4}{r}] \le 2r/q$. Choosing $r = q/4$, we have $\mathbb{P}[\text{rank}(A) \le q/4] \le 1/2$. Again, using Chebyshev's inequality,

$$\mathbb{E}[\text{rank}(A)] \ge \frac{q}{4}\mathbb{P}[\text{rank}(A) \ge \frac{q}{4}] = \frac{q}{4}(1 - \mathbb{P}[\text{rank}(A) \ge \frac{q}{4}]) \ge \frac{q}{8}$$

□

Corollary 2.2.6. *Let H be an $n \times n$ generalised Hadamard matrix, then H is $1/8$-densely regular.*

2.3 Averaging Argument

Another set of proofs utilises averaging argument. The averaging argument says that if we partition the matrix into equally sized parts, there is at least one part that won't have too many changes. For example, we can use this argument to select some number of rows that has small changes and then show that the remaining part of these rows has high rank. In this section, we show how this is used for Vandermonde matrices (Theorem 2.3.1) and Hadamard matrices (Corollary 2.3.5).

2.3.1 Vandermonde Matrix

Theorem 2.3.1 (Shparlinsky, see [Lok00]). *Let $V = (x_i^{j-1})_{i,j=1}^n$ be a Vandermonde matrix with distinct x_i over some field. Then*

$$\mathcal{R}_V(r) \geq \frac{(n-r)^2}{r+1}$$

Proof. Let r be given and let $s = \mathcal{R}_V(r)$. By averaging argument, we can select $r+1$ consecutive columns such that the total number of changes within these columns are at most $s(r+1)/(n-r)$. Then we select the rows that do not contain any changes in these these columns, which gives us at least $n - s(r+1)/(n-r)$ rows. Hence, we constructed a submatrix S of size $(r+1) \times (n - s(r+1)/(n-r))$. Because the rank of this submatrix is at most r, we have that there exists a nonzero vector g such that $Sg = 0$. In other words, we obtain a polynomial $\sum_{t=0}^{r} g_t x^t = 0$ with at least $n - s(r+1)/(n-r)$ roots. On the other hand, this polynomial can have at most r roots. Therefore,

$$r \geq n - s(r+1)/(n-r)$$

which gives $s \geq (n-r)^2/(r+1)$. □

2.3.2 Hadamard Matrix

In this section, we present three different proofs for Hadamard matrices. The first two relies on some results from spectral matrix theory of Hadamard matrices, while the last one is a simple proof with a clever use of the averaging argument. First we note that every non-trivial linear combination of a Hadamard matrix has many nonzero entries.

Proposition 2.3.2 (Alon, see [Juk11]). *Every non-trivial linear combination of any k rows of a Hadamard matrix $H = (h_{i,j}) \in \mathbb{C}^{n \times n}$ has at least n/k nonzero entries.*

Proof. Let A be a $k \times n$ submatrix of H and let $y = x^T A$ for some nonzero vector $x \in \mathbb{R}^k$. Let S be the set of the coordinates of the non-zero entries in y and let $s = |S|$. We need to show that $s \geq n/k$.

Assume without loss of generality that $x_1 = \max_{i \in [k]} |x_i|$. Let a^i denote the ith row of A. Because the rows of A are mutually orthogonal, we have

$$kx_1^2 n \geq \sum_{i=1}^{k} x_i^2 n = \sum_{i=1}^{k} \langle x_i a^i, x_i a^i \rangle = \langle \sum_{i=1}^{k} x_i a^i, \sum_{i=1}^{k} x_i a^i \rangle$$

Notice that $\sum_{i=1}^{k} x_i a^i = x^T A = y$, we have

$$\langle \sum_{i=1}^{k} x_i a^i, \sum_{i=1}^{k} x_i a^i \rangle = \langle y, y \rangle = \sum_{j=1}^{n} y_j^2 = \sum_{j=1}^{n} y_j^2 = \sum_{j \in S} y_j^2 = \sum_{j \in S} |y_j|^2$$

Using Cauchy-Schwarz inequality, we have

$$\sum_{j \in S} |y_j|^2 \geq \frac{1}{s} (\sum_{j \in S} |y_j|)^2 = \frac{1}{s} (\sum_{j=1}^{n} |y_j|)^2$$

On the other hand, because $|a_{i,j}| = 1$, we have

$$\begin{aligned} \sum_{j=1}^{n} |y_j| &\geq \sum_{j=1}^{n} y_j a_{1,j} = \sum_{j=1}^{n} \langle x, a^j \rangle a_{1,j} = \sum_{j=1}^{n} \sum_{i=1}^{k} x_i a_{i,j} a_{1,j} \\ &= \sum_{i=1}^{k} x_i \sum_{j=1}^{n} a_{i,j} a_{1,j} = \sum_{i=1}^{k} x_i \langle a^i, a^1 \rangle = x_1 \langle a^1, a^1 \rangle = x_1 n \end{aligned}$$

This gives us

$$kx_1^2 n \geq \frac{1}{s} (x_1 n)^2$$

Thus, $s \geq n/k$. □

Now we are ready to present to first proof.

Corollary 2.3.3 (Alon, see [Juk01])**.** *If $t > (1 - 1/r)n$, then every $r \times t$ sub-matrix H' of an $n \times n$ Hadamard matrix $H \in \mathbb{R}^{n \times n}$ has rank r.*

Proof. For the sake of contradiction, we assume the opposite that $\text{rank}(H') < r$. Hence, there exists a nonzero vector $x \in \mathbb{R}^r$ such that $x^t H' = 0$. Because $t > (1 - 1/r)n$, this contradicts with proposition 2.3.2 that any nonzero linear combination of these r rows of H has at least n/r nonzero entries. □

Corollary 2.3.4 (Alon, see [Juk01])**.** *If fewer than $(n/r)^2$ entries of an $n \times n$ Hadamard matrix $H \in \mathbb{R}^{n \times n}$ are changed, then the rank of the resulting matrix remains at least r.*

Proof. By averaging argument, we can choose (n/r) rows that has fewer than (n/r) changes in total. Therefore, the number of columns that remain intact in these (n/r) rows is greater than $(1 - 1/r)n$. By 2.3.3, we complete the proof. □

The second proof utilizes a clever observation of the rank of submatrices of Hadamard matrices, which is, in effect, the regularity of Hadamard matrices.

Lemma 2.3.5 ([Lok95])**.** *For any $u \times v$ submatrix H_0 if an $n \times n$ generalised Hadamard matrix H,* $\text{rank}(H_0) \geq uv/n$.

Proof. Let $A \in \mathbb{C}^{k \times k}$ for some $k > 0$. Let $\lambda_1(A)$ be the largest eigenvalue of AA^*. We thus have

$$\frac{||A||_F^2}{\lambda_1(A)} = \frac{\sum_{i=1}^{n} \lambda_i(A)}{\lambda_1(A)}$$

Notice that AA^* has exactly $\text{rank}(AA^*) = \text{rank}(A)$ nonzero entries, all of which are positive, which implies

$$\frac{||A||_F^2}{\lambda_1(A)} = \frac{\sum_{i \in [n], \lambda_i(A) > 0} \lambda_i(A)}{\lambda_1(A)} \leq \text{rank}(A)$$

On the other hand, H_0 is a submatrix of H, we have $\lambda_1(H_0) \leq \lambda_1(H)$. Thus

$$\frac{||H_0||_F^2}{\lambda_1(H_0)} \geq \frac{||H_0||_F^2}{\lambda_1(H)} = \frac{uv}{n}$$

Notice that the last equality follows from the fact that $H_0 H_0^* = vI_u$. Therefore, we have $\text{rank}(H_0) \geq ||H_0||_F^2/\lambda_1(H_0) \geq uv/n$. □

Theorem 2.3.6 ([dW06]). *If $r \leq n/2$, then $\mathcal{R}_H(r) \geq n^2/4r$.*

Proof. Let r be given and let $s = \mathcal{R}_H(r)$. By averaging argument, we can select $2r$ rows that has fewer than $2rs/n$ changes. If $2rs/n \geq n$, we have $s \geq n^2/(2r)$ and we are done. If $2rs/n < n$, we then have that by lemma 2.3.5, for the submatrix H_0 that contains the $n - 2rs/n$ intact columns of these $2r$ rows,

$$r \geq \text{rank}(H_0) \geq \frac{2r(n - 2rs/n)}{n}$$

which implies $s \geq n^2/4r$. □

The same bound can be proved with a much simpler argument for a special type of Hadamard matrices called the **Sylvester matrix**, which is recursively defined as follows:

- $S_1 := (1)$.
- $S_{2n} := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes S_n = \begin{bmatrix} S_n & S_n \\ S_n & -S_n \end{bmatrix}$

where $\otimes$ denotes the Kronecker product.

Theorem 2.3.7 ([Mid05]). *Let n be a power of 2. If $S \in \mathbb{F}^{n \times n}$ is a Sylvester matrix and $r \leq n/2$ is a power of 2, then*

$$\mathcal{R}_S(r) \geq \frac{n^2}{4r}$$

Proof. Let r be given and let $s = \mathcal{R}_S(r)$. Assume on the contrary that $s < n^2/4r$. If we divide S into $(n/2r)^2$ grids of size $2r \times 2r$, then by averaging argument, there exists a grid that has fewer than

$$s \cdot \frac{(2r)^2}{n^2} < \frac{n^2}{4r} \cdot \frac{(2r)^2}{n^2} = r$$

changes. Notice that each grid has full rank because it is exactly a Sylvester matrix of size $2r \times 2r$. Then this grid still has rank more than $2r - r = r$ after these r changes. Hence, the rank of S after these s changes will be more than r, which gives us a contradiction. □

Remark 2.3.1. Notice that this simple proof can be applied to any totally regular matrices, for example, the discrete Fourier transform matrices. In the next chapter, we will see that a similar argument is also useful for Hankel matrices (Section 3.2.2).

Chapter 3

Semi-Explicit Lower Bounds

As we saw in the last chapter, the purely combinatorial techniques fail to achieve the desired explicit lower bounds. In the past few decades, people began to explore various tools to construct rigid matrices, though not explicit. Some involves exploitation of algebraic structures [Lok00, Lok06, KLPS14], some introduces some randomness [GT18] and a more modern line of work applies probabilistically checkable proofs [AC19, BHPT20].

3.1 Exploiting Algebraic Structures

Algebraic arguments have been successful in yielding quadratic lower bounds $\Omega(n^2)$ for suitable target rank r. All such arguments exploit some notion of "independence." Moreover, it is unclear how to efficiently find these "independent" elements, thus making the construction only semi-explicit. As the following results rely on technically involved tools from algebra and the theory of field extensions, we omit the proofs here and refer the reader to the cited references.

The construction of Vandermonde matrices with $\Omega(n^2)$ rigidity exploits algebraic independence.

Definition 3.1.1 ([Mor96])**.** Let K be a field extension of $\mathbb{F}$, and let $t_1, ..., t_n \in K$. The set $\{t_1, ..., t_n\}$ is **algebraically independent** over $\mathbb{F}$ if $f(t_1, ..., t_n) \neq 0$

for all nonzero polynomials $f \in \mathbb{F}[x_1, ..., x_n]$.

Theorem 3.1.1 ([Lok00]). *Let $V = (x_i^{j-1})_{i,j=1}^n$ be a Vandermonde matrix where x_i are algebraically independent over $\mathbb{Q}$. Then*

$$\mathcal{R}_V(r) \geq \frac{n(n - cr^2)}{2}$$

where $c > 0$ is an absolute constant.

The following theorem constructs a matrix using entries from the complex field, with every product of any nr distinct entries being linearly independent over the rational field. However, it is unknown how we can find these numbers efficiently.

Theorem 3.1.2 ([Lok06]). *Let A be an $n \times n$ matrix over $\mathbb{C}$ and $0 \leq r \leq n$. Suppose all products of nr distinct entries of A are linearly independent over $\mathbb{Q}$. Then,*

$$\mathcal{R}_A(r) \geq n(n - 16r)$$

The next result could be obtained using elimination theory from algebraic geometry, but finding distinct primes efficiently remains a computational challenge. Intuitively, the roots of unity of orders of distinct primes can be viewed as somewhat "independent" of each other.

Theorem 3.1.3 ([KLPS14]). *Let $p_{i,j} > n^{4n^2}$ be distinct primes for $1 \leq i, j \leq n$. Let $\mathcal{K} = \mathbb{Q}[\zeta_{1,1}, ..., \zeta_{n,n}]$ be an extension field where $\zeta_{i,j} = e^{2\pi i/p_{i,j}}$. Let $A = (\zeta_{i,j}) \in \mathcal{K}^{n \times n}$. Then for any field L containing K, we have*

$$\mathcal{R}_A(r) \geq (n - r)^2$$

3.2 Random Hankel Matriches

We say that $H \in \mathbb{F}_2^{n\times n}$ is a **Hankel matrix** if $H_{ij} = h_{i+j-1}$ for some $h_1, h_2, ..., h_{2n-1}$. That is

$$H = \begin{bmatrix} h_1 & h_2 & ... & h_n \\ h_2 & h_3 & ... & h_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_n & h_{n+1} & ... & h_{2n-1} \end{bmatrix}$$

H is a **random Hankel matrix** if $h_1, h_2, ..., h_{2n-1}$ are independent uniformly random elements of $\mathbb{F}_2$. In this section we aim to show the following:

Theorem 3.2.1 ([GT18])**.** *Let H be a random Hankel matrix of size $n \times n$. Then, for every $r \in [\sqrt{n}, n/32]$, with probability $1 - o(1)$, the matrix H has rigidity $\Omega(\frac{n^3}{r^2 \log n})$.*

Notice that this bound improves the $\Omega(\frac{n^2}{r}\log\frac{n}{r})$ bound we have seen before when

$$r = o(\frac{n}{\log n \log\log n})$$

To do this, we first show the rigidity of a kind of generalizd Hankel matrices; and then we apply the averaging argument to finish up the proof.

3.2.1 Rigidity of k-Hankel Matrices

We first consider a generalization of the Hankel matrices introduced above. Let $m, k \in \mathbb{N}, 16 \le k \le m$. Let $A \in \mathbb{F}_2^{m\times m}$ be the k-Hankel random matrix, which is

$$\begin{bmatrix} a_1 & a_2 & ... & a_m \\ a_{k+1} & a_{k+2} & ... & a_{k+m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(m-1)k+1} & a_{(m-1)k+2} & ... & a_{(m-1)k+m} \end{bmatrix}$$

where $a_1, a_2, ..., a_{(m-1)k+m}$ are independent, uniformly random bits, and let $S \in \mathbb{F}_2^{m \times m}$ be some fixed matrix. We aim to show the following lemma

Lemma 3.2.2 ([GT18])**.** $\mathbb{P}_A[\text{rank}(S + A) \leq m/2] \leq 2^{-km/16}$.

Let $B = S + A$ and let B_i denote the i-th row of B. If $\text{rank}(B) \leq m/2$, then we can find a basis $B_{i_1}, B_{i_2}, ..., B_{i_{\text{rank}(B)}}$ of the row space spanned by B in the following constructive fashion:

1. Let i_1 be the index of the first nonzero row of B.
2. For each t, let i_t be the index of the first row of B that cannot be spanned by $B_{i_1}, ..., B_{i_{t-1}}$.

Let an index set $I = \{i_1, ..., i_r\}, r \leq m/2$ be given and set $J = [m] \setminus I$. We have that

$$\forall j \in J, B_j \in \text{span}\{B_i : i \in I, i < j\}$$

Let an arbitrary $j \in J$ be given. Notice that if we fix the random bits $a_1, ..., a_{(j-1)k}$, then the j-th row is completely undetermined because the first entry of the j-th row is $a_{(j-1)k+1}$.

Claim 3.2.1 ([GT18]). Let $I' = I \cap [j-1]$ and $p = |I'|$ and fix a vector $c \in \{0,1\}^p$. We have that

$$\mathbb{P}[B_j = \sum_{i \in I'} c_i B_i] = 2^{-m}.$$

Proof. Notice that for all $h \in [m], B_{j,h} = S_{j,h} + a_{(j-1)k+h}$ where $S_{j,h}$ is fixed but $a_{(j-1)k+h}$ is not. Hence, since $\sum_{i \in I'} c_i B_i$ is fixed, and $a_{(j-1)k+h}$ is uniformly chosen from $\{0,1\}$, we have that the h-th bit of this linear combination will be equal to the h-th bit of B_j with probability exactly $1/2$. Also notice that

each of these probabilities are independent since $a_{(j-1)k+h}$ are independently chosen. □

Let $\Delta = \lceil m/k \rceil$, then we can select an increasing sequence of $|J|/\Delta$ indices in J such that each two indices differ by at least Δ. Let $j_1, j_2, ..., j_t$ be such a sequence of indices where $t \geq |J|/\Delta$. For each $l \in [t]$, let E_l be the event that j_l-th row is spanned by the rows indexed by $I \cap [j_l - 1]$.

Claim 3.2.2 ([GT18]). For all $l \in [t], \mathbb{P}[E_l | E_1, E_2, ..., E_{l-1}] \leq 2^{-m/2}$.

Proof. Notice that $j_l \geq j_{l-1} + \Delta = j_{l-1} + \lceil m/k \rceil$. That is, $(j_l - 1)k \geq (j_{l-1} - 1)k + m$. On the other hand, given fixed bits $a_1, ..., a_{j_l - 1}$, we can determine the rows $B_{j_1}, ..., B_{j_{l-1}}$ but $B_{j_l} = (a_{(j_l-1)k+1}, ..., a_{(j_l-1)k+m})$. Hence, we have

$$\mathbb{P}[E_l | E_1, E_2, ..., E_{l-1}] \leq \mathbb{P}[E_l | a_1, ..., a_{j_l - 1}]$$

By claim 3.2.1, for a fixed linear combination, we have $\mathbb{P}[E_l | a_1, ..., a_{j_l-1}, \mathbf{c}] = 2^{-m}$. Let $I' = I \cap [j-1]$ and $p = |I'|$. Because there are 2^p different values for $\mathbf{c}$, and recall that $p \leq r \leq \text{rank}(B) \leq m/2$, by union bound, $\mathbb{P}[E_l | a_1, ..., a_{j_l-1}] \leq 2^p 2^{-m} \leq 2^{-m/2}$. □

We are now in a good shape to prove the following lemma.

Lemma 3.2.3 ([GT18])**.** *Let E be the event that*

$$\forall j \in J, B_j \in \textit{span}\{B_i : i \in I, i < j\}$$

for a given index set I. Then $\mathbb{P}[E] \leq 2^{-mk/8}$.

Proof. Recall that $(j_1, j_2, ..., j_t)$ is a sequence of indices where $t \geq |J|/\Delta$ and each two indices are Δ apart. For each $l \in [t]$, let E_l be the event that j_l-th row is spanned by the rows indexed by $I \cap [j_l - 1]$. Hence,

$$\mathbb{P}[E] \leq \mathbb{P}[E_1, E_2, ..., E_{t-1}, E_t] = \mathbb{P}[E_1]\mathbb{P}[E_2|E_1]...\mathbb{P}[E_t|E_1, E_2, ..., E_{t-1}] \leq \left(2^{-m/2}\right)^t$$

Notice that

$$t \geq |J|/\Delta \geq \frac{m/2}{\lceil m/k \rceil} \geq k/4$$

Therefore, $\mathbb{P}[E] \leq \left(2^{-m/2}\right)^t \leq 2^{-mk/8}$. □

Proof of lemma 3.2.2. We can simply apply union bound among all possible choices of I, which is less than 2^m. Hence, because we chose $k \geq 16, \mathbb{P}_A[\text{rank}(S+A) \leq m/2] \leq 2^m\mathbb{P}[E] \leq 2^{-km/16}$. □

3.2.2 Rigidity of Hankel Matrices

Now we are ready show the proof Theorem 3.2.1. The main idea of this proof is based on an averaging argument very similar to the one used to prove Theorem 2.3.7. For simplicity, we assume $m = 2r$ and $k = n/m$ are integers. Recall that we define $H \in \mathbb{F}_2^{n\times n}$ by $H_{ij} = h_{i+j-1}$ for random $h_1, h_2, ..., h_{2n-1}$. Let $S \in \mathbb{F}_2^{n\times n}$ be an s-sparse matrix with

$$s \leq \frac{n^3}{160r^2}\log(\frac{960r^2}{n})$$

Consider the following partition of H and S into $(n/m)^2$ submatrices:

- For each $i \in [n/m]$, let the subset of row indices

$$I_i := \{i, i + k, ..., i + (m - 1)k\}$$

- For each $j \in [n/m]$, let the subset of column indices

$$J_j := \{(j-1)m+1, (j-1)m+2, ..., jm\}$$

- Let H^{ij} be the submatrix of H indexed by I_i and J_j and S^{ij} be the submatrix of S indexed by I_i and J_j.

Observe that H^{ij} is exactly a random k-Hankel matrix we analyzed in the last section. Now, by averaging argument, there muse be some i, j such that S^{ij} is s'-sparse with $s' \leq s \cdot (m/n)^2$.
Let $\mathcal{E}^{ij}$ denote the event that $H^{ij} = S^{ij} + R'$ for some R' with $\text{rank}(R') \leq r$. The next observation is that we can write $H = S + R$ with $\text{rank}(R) \leq r$ only if there is a submatrix H^{ij} such that $\mathcal{E}^{ij}$ occurs. Now we can apply union bound with lemma 3.2.2

$$\begin{aligned}\mathbb{P}[\exists i, j : \mathcal{E}^{ij} \text{ occurs}] &\leq \sum_{i,j} \mathbb{P}[\mathcal{E}^{ij}] \leq \sum_{i,j} \sum_{T:s'\text{-sparse}} \mathbb{P}_A[\text{rank}(T+A) \leq m/2] \\ &\leq \binom{n}{m}^2 \cdot \binom{m^2}{\leq s'} \cdot 2^{-km/16} < n^2 \cdot \left(\frac{6m^2}{s'}\right)^{s'} \cdot 2^{-n/16}\end{aligned}$$

Using the assumption that

$$s \leq \frac{n^3}{160r^2} \log(\frac{960r^2}{n})$$

we can show that $\mathbb{P}[\exists i, j : \mathcal{E}^{ij} \text{ occurs}] = o(1)$, which concludes the proof.

3.3 Towards Efficient Constructions of Rigid Matrices

The complexity class FNP is the function-problem extension of the decision-problem class NP. Formally, a relation $R(x, y)$ is in FNP if there exists a non-deterministic polynomial-time Turing machine M such that for any input

x, $M(x)$ outputs y such $R(x, y) = 1$ or rejects if no such y exists. Recent work by Alman and Chen introduces a surprising application of probabilistically checkable proofs to construct rigid matrices in FNP.

Theorem 3.3.1 ([AC19], as stated in [BHPT20]). *There is a constant* $0 < \delta < 1$ *such that for all* $0 < \varepsilon < 1$*, there is an FNP-machine that for infinitely many* n*, on input* 1^{2^n} *outputs an* $2^n \times 2^n$ *matrix* M *with*

$$\mathcal{R}_M(2^{n^{1/4-\varepsilon}}) \geq \delta \cdot 2^{2n}$$

More recently, Bhangale, Harsha, Paradise and Tal improved this result using rectangular probabilistically checkable proofs [BHPT20].

Theorem 3.3.2 ([BHPT20]). *There is a constant* $0 < \delta < 1$ *such that there is an FNP-machine that for infinitely many* n*, on input* 1^{2^n} *outputs an* $2^n \times 2^n$ *matrix* M *with*

$$\mathcal{R}_M(2^{n/\Omega(\log n)}) \geq \delta \cdot 2^{2n}$$

For technical details of the proofs, we refer the reader to the cited references. We remark that these constructions still fail to give us "Valiant"-rigid matrices, which would require $\mathcal{R}_A^{\mathbb{F}}(n/\log\log n) = n^{1+\delta}$ for some $\delta > 0$.

Chapter 4

Paturi-Pudlák Dimensions

In this chapter, we first trace the origin of Paturi-Pudlák Dimensions. This was motivated by the pioneering results by Friedman [Fri93], who gave the first nontrivial lower bound on the rigidity of a matrix over finite fields. The matrix of interest was the generating matrix of a linear code, which can be viewed as a linear subspace. In [PP06], Paturi and Pudlák extended the ideas of Friedman into two notions called inner dimension and outer dimension of linear subspaces, which were later referred to as *Paturi-Pudlák Dimensions*. Lastly, we formalize the notion of row-rigidity, which will be useful in later chapters, and discuss its relation to the original notion of rigidity we have been familiar so far.

4.1 Friedman's result

To appreciate the development of Paturi-Pudlák Dimensions, we first present the original theorem and proof in Friedman's paper [Fri93]. Many people claim that the following theorem implies an $\Omega(\frac{n^2}{r}\log_q \frac{n}{r})$ lower bound of matrix rigidity. However, it is not directly obvious from the way this theorem was presented. After we introduce the formal definitions of Paturi-Pudlák Dimensions in the next section, we will show a clear argument for the $\Omega(\frac{n^2}{r}\log_q \frac{n}{r})$ lower bound.

Theorem 4.1.1 ([Fri93]). *For any constant $C_1 > 0$ there is a constant $C_2 > 0$ such that the following holds. Let $\mathbb{F}$ be a finite field of q elements. Let A be an $n \times n$ matrix such that the first $n/2$ rows are the basis of a linear error-correcting code in $\mathbb{F}^n$ of minimum distance $\geq C_1 n$. If B is any $n \times n$ matrix over $\mathbb{F}$ with at most s non-zero entries in each row, where $s \leq n/C_2$, then we have*

$$\operatorname{rank}(A + B) \geq \frac{n}{C_2 s}(\log_q s + \log_q(q-1))$$

Proof. Let $A_{n/2}$ denote the first $n/2$ rows of A and $B_{n/2}$ the first $n/2$ rows of B. We set $D_{n/2} \in \mathbb{F}_q^{n/2 \times n}$ by $D_{n/2} = A + B$ and let r denote the rank of $_{n/2}$. Let S denote the linear space spanned by all vectors $w \in \mathbb{F}_q^{n/2}$ such that

$$w \cdot D_{n/2} = 0$$

We see that S is a subspace of $\mathbb{F}_q^{n/2}$ with dimension $n/2 - r$. The first part of the proof applies a packing argument, as shown in the following claim.

Claim 4.1.1. Suppose t is an integer such that the size of a Hamming sphere of radius $t/2$ in $\mathbb{F}_q^{n/2}$ is at least q^r. Then there is a vector $w \in S$ with weight at most t.

Proof. Suppose on the contrary that the weight of w is greater than t for all $w \in S$. Let l denote the size of a Hamming sphere of radius $t/2$ in $\mathbb{F}_q^{n/2}$. Then

$$|S| \cdot l > |S| \cdot q^r = q^{n/2-r} \cdot q^r = q^{n/2}$$

However, we know that there are at most $q^{n/2}$ points in $\mathbb{F}_q^{n/2}$. That is

$$q^{n/2} \geq |S| \cdot l$$

Hence, we complete the proof by contradiction. ■

Now, let $w \in S$ be a vector of weight t as defined in the claim above. We then have

$$0 = w \cdot D_{n/2} = w \cdot A_{n/2} + w \cdot B_{n/2}$$

Because all the rows in $A_{n/2}$ are independent, we have $w \cdot A_{n/2} \neq 0$ and thus $w \cdot B_{n/2} \neq 0$. Since each row of B has at most s non-zero entries and w has weight t, we have that the weight of $w \cdot B_{n/2}$ is at most ts. On the other hand, since the code represented by $A_{n/2}$ has minimum distance $C_1 n$, we have that the weight of $w \cdot A_{n/2}$ is at least $C_1 n$. Therefore, we must have

$$ts \geq C_1 n$$

Take $t_0 = \lceil C_1 n/s \rceil$, we then have the size of a Hamming sphere of radius $t/2$ in $\mathbb{F}_q^{n/2}$ is at most q^t because the weight of w must be greater than t_0 to achieve $0 = w \cdot D_{n/2}$. Then

$$q^r \geq \binom{n/2}{t_0/2}(q-1)^{t_0/2}$$

which is

$$r \geq \log_q \left[\binom{n/2}{t_0/2}(q-1)^{t_0/2}\right] \geq \log_q \binom{n/2}{t_0/2} + \frac{t_0}{2}\log_q(q-1) \geq \frac{t_0}{2}\frac{n}{s}\log_q s$$

Choosing $C_2 \sim 1/C_1$, we have

$$r \geq \log_q \left[\binom{n/2}{t_0/2}(q-1)^{t_0/2}\right] \geq \frac{n}{C_2 s}(\log_q s + \log_q(q-1))$$

□

4.2 Strong Rigidity and Paturi-Pudlák Dimensions

Before diving into Paturi-Pudlák Dimensions, we first introduce a convenient notion called **sparsity**. It is essentially the same as the definition of

density, which we introduced in the first chapter. However, "sparsity" is used in the papers we cited in the chapter.

Definition 4.2.1 (Sparsity, [PP06, DGW19])**.** A vector $v \in \mathbb{F}^n$ is **s-sparse** if the number of non-zero coordinates in v is at most s. A matrix $A \in \mathbb{F}^{m \times n}$ is **s-sparse** if it has at most s nonzero entires. We say the matrix A is **s-row sparse** if each of its row is s-sparse.

After presenting the theorem shown in the last section, Friedman introduced the following notion of strong rigidity.

Definition 4.2.2 (Strong Rigidity for Subspaces, [Fri93])**.** Let $V \subseteq \mathbb{F}^n$ be a subspace. We say that V is **(s,t)-strongly rigid** if for any subspace $U \subseteq \mathbb{F}^n$ generated by s-sparse vectors and with $\dim(U) \leq \dim(V)$,

$$\dim(V \cap U) \leq \dim(V) - t$$

Notice that Theorem 4.1.1 shows that the row-space of A is (k,t)-strongly rigid with $t = n \log_q k/(C_2 k)$. To refine this notion of strong rigidity, Paturi and Pudlák introduced inner dimension in [PP06].

Definition 4.2.3 (Inner Dimension, [PP06])**.** Let $V \subseteq \mathbb{F}^n$ be a subspace, and s be a positive integer less than n. We defined the **inner dimension $d_V(s)$ of V** by

$$d_V(s) := \max_U \{\dim(V \cap U) : U \subseteq \mathbb{F}^n, \dim(U) \leq \dim(V), \\ U \text{ is a subspace generated by } s\text{-sparse vectors}\}$$

Notice that a subspace V is (s,t)-strongly rigid if $d_V(s) \leq \dim(V) - t$, or equivalently, $t \leq \dim(V) - d_V(s)$. We emphasize this because there was a typo about this remark in the original [PP06] paper. Paturi and Pudlák also introduced a related concept called outer dimension.

Definition 4.2.4 (Outer Dimension, [PP06])**.** Let $V \subseteq \mathbb{F}^n$ be a subspace, and s be a positive integer less than n. We defined the **outer dimension** $D_V(s)$ **of V** by

$$D_V(s) := \max_U \{\dim(U) : U \subseteq \mathbb{F}^n, V \subseteq U, \\ U \text{ is a subspace generated by } s\text{-sparse vectors}\}$$

The following simple bound gives a first connection between the inner and outer dimension.

Proposition 4.2.1 ([PP06])**.** *Let $V \subseteq \mathbb{F}^n$ be a subspace and s be a positive integer less than n. Then,*

$$d_V(s) + D_V(s) \geq 2\dim(V)$$

Proof. Let $V \subseteq \mathbb{F}^n$ be a subspace such that $V \subseteq U, U$ is s-sparse and $\dim(U) = D_V(s)$. Let $m = \dim(V)$ and W be an m-dimensional subspace of $\mathbb{F}^n$ such that $W \subseteq U$. Hence, $\dim(V \cap W) \leq d_V(s)$ and thus

$$\begin{aligned} 2\dim(V) &= \dim(V) + \dim(W) \\ &= \dim(V \cap W) + \dim(V \cup W) \\ &\leq d_V(s) + \dim(U) = d_V(s) + D_V(s) \end{aligned}$$

□

In the following theorem, we see that linear codes can be used to show nontrivial lower bounds of $D_V(s)$ and upper bounds of $d_C(s)$, which generalizes the proof technique invented by Friedman, as we saw in Theorem 4.1.1.

Theorem 4.2.2 ([PP06])**.** *Let C be an $[n, k, d]$ linear code over $\mathbb{F}_2$. Then for $s \leq d/2$,*

$$D_C(s) \geq k + \frac{d}{2s}\log(\frac{2sk}{d})$$
$$d_C(s) \leq k - \frac{d}{2s}\log(\frac{2sk}{d})$$

Before we get into the proof, we need the following auxiliary lemma.

Lemma 4.2.3. *Let C be an $[n, k, d]$ linear code over $\mathbb{F}_2$. Then for $s \leq d/2$, then there exists a $[D_C(s), k, d/s]$-code.*

Proof. Consider the subspace $W \subseteq \mathbb{F}_2^n$ with $\dim(W) = D_C(s), C \subseteq W$ and W is s-sparse. Let $D = D_C(s)$ and $\{w_1, ..., w_D\}$ be a basis of W where each $w_i, i \in [D]$ is s-sparse. Hence, for any $x \in C$, we have that there exists a $y \in \mathbb{F}_2^D$ such that

$$x = \sum_{i=1}^{D} y_i w_i$$

Let E be the set of all such y for all x. Then, we have $\dim(E) = \dim(C) = k$. Let $y' \in E$ be a nonzero vector with minimum weight. Because $x' = \sum_{i=1}^{D} y'_i w_i$ has weight at least d and each w_i is s-sparse, we have that at least d/s coordinates of y' is nonzero. Hence, we obtain that E is a $[D_C(s), k, d/s]$-code. □

Proof of Theorem 4.2.2. Using the sphere packing bound on the $[D = D_C(s), k, d/s]$-code E we just constructed, we have that the Hamming balls of radius at most $d/2s$ at each vector in B do not intersect with each other. Hence,

$$\sum_{j=1}^{d/2s} \binom{D}{j} \leq 2^{D-k}$$

Notice that

$$\sum_{j=1}^{d/2s} \binom{D}{j} \geq \binom{D}{d/2s} \geq \binom{k}{d/2s} \geq (2sk/d)^{d/2s}$$

Hence,

$$D - k \geq \frac{d}{2s} \log(\frac{2sk}{d}) \tag{4.2.0.1}$$

Let $U \subseteq \mathbb{F}_2^n$ be a subspace with $\dim(U) = k$, U is s-sparse and $\dim(C \cap U) = d_C(s)$. Then $F = C \cap U$ is simply a $[n, d_C(s), d]$ code. Applying Lemma 4.2.3 again, we obtain a $[D_F(s), d_C(s), d/s]$-code. Hence, because $D_F(s) \leq \dim(U) = k$, equation 4.2.0.1 implies

$$k - d_C(s) \geq D_F(s) - d_C(s) \geq \frac{d}{2s} \log(\frac{2sk}{d})$$

which means $d_C(s) \leq k - (d/2s)\log(2sk/d)$. □

To see how this result relate back to Theorem 4.1.1 by Friedman, we define $\rho_A(s)$ for a matrix $A \in \mathbb{F}^{m \times n}$, where m, n are integers with $m \leq n$:

$$\rho_A(s) := \min_B \{\text{rank}(A + B) : B \text{ is } s\text{-row sparse}\}$$

Proposition 4.2.4. *Let $A \in \mathbb{F}^{m \times n}$ and $0 < s \leq n$ be given. Let V be the row space of A. Then*

$$\text{rank}(A) - d_V(s) = \dim(V) - d_V(s) \leq \rho_A(s)$$

This proposition, combined with Theorem 4.2.2, connects all the dots. If $A \in \mathbb{F}^{m \times n}$ is the generator matrix of an $[n, k, d]$-code, then essentially we have for any s-row sparse matrix B,

$$\text{rank}(A+B) \geq \rho_A(s) \geq \text{rank}(A) - d_V(s) \geq k - d_V(s) \geq k - k - \frac{d}{2s}\log(\frac{2sk}{d}) = \frac{d}{2s}\log(\frac{2sk}{d})$$

which is basically the result in Theorem 4.1.1.

Proof. Let B be the matrix that matches $\rho_A(s)$, i.e., $\text{rank}(A - B) = \rho_A(s)$. Let U be the row space of B and let W be the row space of $A - B$. Then we have $\dim(W) = \text{rank}(A - B) = \rho_A(s)$ and thus

$$\dim(V \cup U) \leq \dim(U) + \dim(W) = \dim(U) + \rho_A(s)$$

Hence, we obtain

$$\begin{aligned}\operatorname{rank}(A) - d_V(s) &= \dim(V) - d_V(s) = \dim(V \cap U) + \dim(V \cup U) - \dim(U) - d_V(s) \\ &\leq \dim(V \cap U) + \dim(U) + \rho_A(s) - \dim(U) - d_V(s) \\ &= \dim(V \cap U) - d_V(s) + \rho_A(s)\end{aligned}$$

Because B is s-row sparse, we have that U is s-sparse. Thus, $d_V(s) \geq \dim(V \cap U)$. Hence, we obtain $\operatorname{rank}(A) - d_V(s) \leq \rho_A(s)$. □

4.3 Row Rigidity

Now we come back to the question raised earlier: how does Proposition 4.2.4 give us the $\Omega(\frac{n^2}{r}\log\frac{n}{r})$ lower bound of matrix rigidity? We are now ready to show a clear argument.

Corollary 4.3.1 (Presented in [Lok09])**.** *Let $A \in \mathbb{F}_q^{k\times n}$ be the generator matrix of an asymptotically good $[n, k = \Omega(n), d = \Omega(n)]$-linear code C. Then for $0 < r < k/2$,*

$$\mathcal{R}_A(r) = \Omega(\frac{n^2}{r}\log\frac{n}{r})$$

The following proof adapts the argument given in Lokam's survey [Lok09] and Golovnev's lecture notes [Gol20].

Proof. As we discussed above, by Proposition 4.2.4 and Theorem 4.2.2, for any s-row sparse matrix B,

$$\operatorname{rank}(A + B) \geq \rho_A(s) \geq \frac{d}{2s}\log(\frac{2sk}{d})$$

This means that for any s-row sparse matrix B, to achieve $\operatorname{rank}(A + B) \leq r$, we must have

$$s = \Omega(\frac{n}{r}\log\frac{n}{r})$$

Now, suppose we have a t-sparse matrix $D \in \mathbb{F}_q^{k \times n}$, that is, D has at most t non-zero entries overall. Then, by averaging argument, we have that for each of the $k/2$ sparsest rows of D, each row has at most $t/(k/2)$ entries. In other words, if we write D' as the submatrix of D consisting these $k/2$ sparsest rows, then D' is $t/(k/2)$-row sparse. In order to achieve $\text{rank}(A+B) \leq r < k/2$, on these $k/2$ sparsest rows of D, we must have

$$t/(k/2) \geq s$$

which translates to

$$t \geq s \cdot \frac{k}{2} = \Omega(\frac{n^2}{r} \log \frac{n}{r})$$

because $k = \Omega(n)$. □

In the preceding proof, we see a linkage between $\rho_A(s)$ to the original definition of matrix rigidity. Inspired by this connection, we introduce the definition of *row rigidity*, which will turn out useful in later chapters.

Definition 4.3.1 ([APY09, AC15, DGW19])**.** Let m, n be positive integers and $A \in \mathbb{F}^{m \times n}$. We say that A is (r, s)**-row rigid** if for every $m \times n$ matrix R with $\text{rank}(R) \leq r$, $A + R$ contains a row with at least s non-zero entries.

That is, A is (r, s)-row rigid if one cannot decrease the rank of A to r by altering fewer than s entries in each row of A.

Dvir, Golovnev and Weinstein introduced the definition of **strong row rigidity** for matrices, which as we will see from Lemma 6.1.3 later, is equivalent to small inner dimensions.

Definition 4.3.2 (Strong Row Rigidity, [DGW19])**.** A matrix $A \in \mathbb{F}^{m \times n}$ is said to be (r, s)**-strongly row rigid** if for any invertible matrix $C \in \mathbb{F}^{n \times n}$, we have $A \times C$ is (r, s)-row rigid.

Notice that in the few definitions above, m and n are not necessarily equal. In fact, Alon, Panigrahy and Yekhanin extended the study of rigidity of non-square matrices, where we allow the number of rows m to be larger than the number of columns n [APY09]. We will explain more when we study rigid sets in the next chapter. For now, we first show a more sophisticated way, than the averaging argument we saw above, to construct a rigid matrix given a row-rigid matrix. We first rewrite what it means to be rigid and strongly rigid for matrices.

Definition 4.3.3 (Rigidity and Strong Rigidity for Matrices, [DGW19])**.** A matrix $A \in \mathbb{F}^{m\times n}$ is said to be (r, s)**-rigid** if for any s-sparse matrix $B \in \mathbb{F}^{m\times n}$, we have $A + B$ has rank at least r. A matrix $A \in \mathbb{F}^{m\times n}$ is said to be (r, s)**-strongly rigid** if for any invertible matrix $C \in \mathbb{F}^{n\times n}$, we have $A \times C$ is (r, s) rigid.

Before we state the result, we need the definition of locally decodable codes.

Definition 4.3.4 (Locally Decodable Codes)**.** A linear code $C : \mathbb{F}^m \to \mathbb{F}^n$ is said to be (t, δ, ε)**-locally decodable** if there exists a randomised decoding algorithm $\mathcal{A}$ such that for all $m \in \mathbb{F}^m$ and all $w \in \mathbb{F}^n$ such that $\text{dist}(C(m), w) \leq \delta$:

1. For every index $i \in [m]$

$$\mathbb{P}[\mathcal{A}(w, i) = m_i] \geq 1 - \varepsilon,$$

where the probability is taken over the random coin tosses of the algorithm $\mathcal{A}$.

2. $\mathcal{A}$ makes at most t queries to w.

We abuse the notation and write $C \in \mathbb{F}^{m \times n}$ as its generating matrix.

Theorem 4.3.2 (From Row Rigidity to General Rigidity, [DGW19]). *Let $A \in \mathbb{F}^{m \times n}$ be a rectangular matrix, $E \in \mathbb{F}^{l \times m}$ a $(t, \delta, 3/4)$-linear locally decodable code and matrix $B := E \cdot A$. Then,*

1. *If $A \in \mathbb{F}^{m \times n}$ is $(r, s+1)$-row rigid, then $B \in \mathbb{F}^{l \times n}$ is $(r, (\delta sl)/t)$-rigid.*
2. *If $A \in \mathbb{F}^{m \times n}$ is $(r, s+1)$-strongly row rigid, then $B \in \mathbb{F}^{l \times n}$ is $(r, (\delta sl)/t)$-strongly rigid.*

Let $\mathrm{dist}(u, v)$ to denote the Hamming distance between two vectors u, v. We need a lemma from locally decodable code to prove this theorem, which we state without a proof.

Lemma 4.3.3 ([GKST02, DS07]). *Let $C \in \mathbb{F}^{m \times n}$ be a $(t, \delta, 3/4)$-linear locally decodable code and let R be a set of rows of C with $|R| \geq (1-\delta)m$. For any $i \in [n]$, there exists a set of t rows in R which spans the ith standard basis vector e_i.*

Proof of Theorem 4.3.2. Let $A \in \mathbb{F}^{m \times n}$ be $(r, s+1)$-row rigid and suppose that B is not $(r, (\delta sl)/t)$-rigid. Then we have $B = D + S$ where $D \in \mathbb{F}^{l \times n}$ has rank at most r and $S \in \mathbb{F}^{l \times n}$ has density $\mathrm{dens}(S) \leq (\delta sl)/t$. Let S' be the set of row of S that are s/t-sparse. By averaging argument, we have that $|S'| \geq (1-\delta)l$. Let D' be the corresponding rows in D. Because A is $(r, s+1)$-row rigid, some rows A_i has Hamming distance at least $(s+1)$ from the space generated by D.

On the other hand, by Lemma 4.3.3, there exist t rows in D' and S' which

spans A_i. This means that A_i has a Hamming distance at most $t \cdot (s/t) = s$ from the row space of l'. Therefore, by contradiction, we must have B is $(r, (\delta sl)/t)$-rigid.

Let $A \in \mathbb{F}^{m \times n}$ be $(r, s+1)$-strongly row rigid, then for all invertible matrix $T \in \mathbb{F}^{n \times n}$ such that $A \times T$ is $(r, s+1)$-row rigid. Notice that

$$E \times (A \times T) = (E \times A) \times T = B \times T$$

where $B \in (r, (\delta sl)/t)$. As we have just shown, $B \times T$ is $(r, (\delta sl)/t)$-rigid. Since T is arbitrary, we have that B is $(r, (\delta sl)/t)$-strongly rigid. □

The following corollary shows that we can obtain rigid square matrices from rigid non-square matrices.

Corollary 4.3.4 (Non-square Matrices to Square Matrices, [DGW19])**.** *For every constant $\alpha > 0$, and an $(r, s+1)$-row rigid matrix $A \in \mathbb{F}^{m \times n}$, we can construct a square matrix $B \in \mathbb{F}^{l \times l}, l = m^{O(1/\alpha)}$, which is*

$$(r, \frac{l}{n} \cdot \frac{s}{(\log m)^{1+\alpha}})\text{-row rigid and } (r, \frac{l^2}{n} \cdot \frac{s}{(\log m)^{1+\alpha}})\text{-rigid.}$$

To prove this corollary we use the following result from locally decodable code without proof. This lemma allows us to construct linear locally decodable codes.

Lemma 4.3.5 ([Dvi11])**.** *For every $\alpha, \varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ and an explicit family of $((\log n)^{1+\alpha}, \delta, \varepsilon)$-linear locally decodable codes $C \in \mathbb{F}^{m \times n}$ for $m = n^{O(1/\alpha)}$.*

Proof of Corollary 4.3.4. Let $l = m^{O(1/\alpha)}$ be a multiple of n and δ some constant. We also let $C \in \mathbb{F}^{l \times m}$ be a $((\log m)^{1+\alpha}, \delta, 3/4)$-linear locally decodable

code. Hence, we can construct a square matrix $B \in \mathbb{F}^{l \times l}$ by putting side by side (l/n) copies of $C \times A$.

It remains to show that B is rigid. By Theorem 4.3.2, we have $C \times A$ is $(r, (\delta sl)/(\log m)^{1+\alpha})$-rigid. Hence,

$$\mathcal{R}_B(r) \geq \frac{l}{n} \cdot \frac{\delta sl}{(\log m)^{1+\alpha}}$$

and by averaging over l rows, B must be $(r, (\delta sl)/[n \cdot (\log m)^{1+\alpha}])$-row rigid. Setting $\delta = 1$ or some suitable constant, we get the desired result. $\square$

Chapter 5

Rigid Sets

Recall from the definition of row-rigidity from last chapter:

Definition 5.0.1. Let m, n be positive integers and $A \in \mathbb{F}^{m \times n}$. We say that A is (r, s)**-row rigid** if for every $m \times n$ matrix R with $\text{rank}(R) \leq r$, $A + R$ contains a row with at least s non-zero entries.

That is, A is (r, s)-row rigid if one cannot decrease the rank of A to r by altering fewer than s entries in each row of A. In this chapter we view rigidity from a different perspective. Let's think of the matrix $A \in \mathbb{F}^{m \times n}$ as a subset S of $\mathbb{F}^n$, where each row of A is an element of this set S. So far we have seen the trade-off between the values of *dimensions* (rank) and *distance* (rigidity) that can be obtained by explicit sets of *size* n (i.e. $m = n$ for the size of the matrix). Alon, Panigrahy and Yekhanin initiated the study of **rigid sets** to investigate the trade-off between the values of *size* and *distance*, when the value of *dimension* is fixed [APY09]. Let's introduce the definition of rigid sets.

Definition 5.0.2. For $x \in \mathbb{F}_2^n$, and a linear subspace $U \subseteq \mathbb{F}_2^n$, we define the **Hamming distance from** x **to** U by

$$\text{dist}(x, U) = \min_{u \in U} |x + u|$$

where $|v|$ denotes the Hamming weight of v.

Definition 5.0.3 (Rigid Sets, [APY09])**.** A set $S \subseteq \mathbb{F}_2^n$ is called (n, k, d)**-rigid** if for every linear subspace $U \subseteq \mathbb{F}_2^n, \dim(U) = k$, we have

$$\max_{s \in S} \operatorname{dist}(s, U) \geq d$$

Let $A \in \mathbb{F}_2^{m \times n}, m = |S|$ be the matrix whose rows are the elements of S. Notice that A is (k, d)-row rigid if and only if S is (n, k, d)-rigid. In this setting, we no longer insist on $m = n$ or $m = O(n)$, but aims to get m as small as possible as a function of d, with the goal of achieving $m = O(n) + d^c$ for any constant c. However, we are quite far from this goal:

Theorem 5.0.1 ([APY09], [SY11])**.** *For every $0 \leq d \leq O(n)$, there exists an explicit set $(n, n/2, d)$-rigid set $S \subseteq \mathbb{F}_2^n$ of size $2^{O(d)}n/d$.*

Before diving into the proof, we need some notation and auxilliery lemmas.

Notation 5.0.4. Let $I \subseteq [n]$ be a set of coordinates. For a vector $x \in \mathbb{F}^n$, we write $x|_I$ to denote the vector x restricted to the coordinates in I. Similarly, for a linear subspace $U \subseteq \mathbb{F}^n$, we write $U|_I$ to denote the linear subspace U restricted to the coordinates in I.

Lemma 5.0.2. *Let $U \subseteq \mathbb{F}_2^n$ be a linear subspace with $\dim(U) = k$, then*

$$\mathbb{P}_{x \in \{0,1\}^n}[x \in U] \leq \frac{1}{2^{n-k}}$$

Proof. Let $I \subseteq [n]$ be the set of coordinates such that $U|_I = \mathbb{F}_2^n$ and $J = [n] \setminus I$. Hence, we note that a vector $x = x|_I + x|_J \in U$ is uniquely determined by $x|_I$ because $x|_J$ is the zero vector. Hence, for a random vector $x \in \{0, 1\}^n$, there is at most 2^{k-n} chance that x is in U. □

Lemma 5.0.3. *For every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all linear subspaces $U \subseteq \mathbb{F}_2^n, \dim(U) \leq (1-\varepsilon)n$, there exists a point $x \in \{0,1\}^n$ such that*

$$\operatorname{dist}(x, U) \geq \delta n$$

Proof. Let U be given. We note that for a random vector $x \in \{0,1\}^n$,

$$\mathbb{P}[\operatorname{dist}(x, U) \leq \delta n] = \mathbb{P}[\exists I \subseteq [n], |I| = (1-\delta)n \text{ such that } x|_I \in L|_I]$$

For a fixed set I with $|I| = (1-\delta)n$, we have by Lemma 5.0.2, we have

$$\mathbb{P}[x|_I \in U|_I] = \frac{1}{2^{(1-\delta)n-(1-\varepsilon)n}} = \frac{1}{2^{(\varepsilon-\delta)n}}$$

Hence, by union bound on all possible set I of size $(1-\delta)n$, the probability

$$\mathbb{P}[\operatorname{dist}(x, U) \leq \delta n] \leq \binom{n}{\delta n} \frac{1}{2^{(\varepsilon-\delta)n}}$$

is negligible when δ is sufficiently smaller than ε. □

Proof of Theorem 5.0.1. As a consequence of Lemma 5.0.3, let δ be the constant that for all linear subspace $U \subseteq \mathbb{F}_2^n, \dim(U) = n/2$, there exists a point p in $\mathbb{F}_2^n$ that is more than δn-far from U.

To obtain S, we first split the coordinates into cn/d disjoint sets $Z_1, Z_2, ..., Z_{\delta n/d}$, each of size d/δ. For each set Z_i, we let W_i be the set of all binary vectors $x_i \in \mathbb{F}_2^n$ with support on this set Z_i. That is, each x_i has some value 1 on some coordinates in set Z_i and has 0 on every other coordinates. Let $S = \bigcup_i W_i$ consist of all these vectors. Hence,

$$|S| = 2^{O(d)} n/d$$

and every vector in $\mathbb{F}_2^n$ is the sum of at most $\delta n/d$ vectors in S.

Let a linear subspace $U \subseteq \mathbb{F}_2^n, \dim(U) = n/2$ be given. Suppose every vector

in S is at most d-far from U. That is, any vector in S is the sum of one vector in U and at most d unit vectors. Because every vector v in $\mathbb{F}_2^n$ is the sum of at most $\delta n/d$ vectors in S, v must also be the sum of a vector in U and at most $d \times (\delta n/d) = \delta n$ unit vectors. Hence, no point p will be more than δn-far from U, which gives us a contradiction. □

5.1 Strong Rigid Sets

As an attempt to break the $2^{O(d)}n/d$-barrier, Alon and Cohen introduced *U-polynomials* and show that explicitly constructing rigid sets can be reduced to explicitly constructing a small hitting set [AC15]. In fact, Alon and Cohen used a stronger notion than rigid sets, called *strong rigid set*, and show that small-biased sets are strong rigid sets:

Definition 5.1.1 (Strong Rigid Sets, [AC15])**.** A set $S \subseteq \mathbb{F}_2^n$ is called **strong** (n, k, d)**-rigid** if for every linear subspace $U \subseteq \mathbb{F}_2^n, \dim(U) = k$, we have

$$\mathbb{E}_{s\sim S}[\text{dist}(s, U)] \geq d$$

where $\sim$ denotes a uniformly random sample.

Definition 5.1.2 (Small Biased Sets, [NN93])**.** We say that a set $S \subseteq \mathbb{F}_2^n$ is ε-biased if for every nonzero $\alpha \in \mathbb{F}_2^n$,

$$\left|\mathbb{E}_{s\sim S}[(-1)^{\langle \alpha, s\rangle}]\right| \leq \varepsilon$$

Theorem 5.1.1 ([AC15])**.** *For every $0 \leq d \leq cn$ for some suitable constant $0 < c < 1$. If $S \subseteq \mathbb{F}_2^n$ is an $\exp(-d)$-biased set, then S is $(n, n/2, d)$-strong rigid.*

Via probabilistic methods, there exists ε-biased sets in $\mathbb{F}_2^n$ with size $O(n/\varepsilon^2)$. Unfortunately, we do not know an explicit construction of small biased sets that would have the desired size, i.e. $m = O(n) + d^c$. Nonetheless, the construction by Alon et al [ABN$^+$92] yields a small biased set of size $n \cdot \exp(d)$, which matches that in Theorem 5.0.1.

5.1.1 U polynomials

Let's first introduce U-polynomials.

Definition 5.1.3 ([AC15])**.** For a subspace $U \subseteq \mathbb{F}_2^n$, the **U-polynomial** $p_{U,\rho} :\subseteq \mathbb{F}_2^n \to \mathbb{R}$ is defined as

$$p_{U,\rho}(x) = \frac{1}{W_\rho(U)} \cdot \sum_{u \in U} \rho^{|u|} \cdot (-1)^{\langle u,x \rangle}$$

where $W_\rho(U) = \sum_{u \in U} \rho^{|u|}$ is the **weight enumerator** of U with parameter $\rho \in (0,1)$.

The following theorem implies that to explicitly construct an (n,k,d)-rigid set, it suffices to explicitly construct a set S such that for every $U \subset \mathbb{F}_2^n$ with dimension $n-k$, there exists $s \in S$ such that $p_U(s) \leq 2^{-\Omega(d)}$.

Theorem 5.1.2 ([AC15])**.** *Let $U \subseteq \mathbb{F}_2^n$ be a linear subspace. Then, for any parameter $\rho \in (0,1)$ and any point $x \in \mathbb{F}_2^n$,*

$$\operatorname{dist}(x,U) \geq \Big(\log \frac{1+\rho}{1-\rho}\Big)^{-1} \cdot \log \frac{1}{p_{U^\perp,\rho}(x)}$$

In particular, we have

$$\operatorname{dist}(x,U) = \Omega(\log \frac{1}{p_{U^\perp,\rho}(x)})$$

5.1.2 Fourier analysis

We will need a few definitions and tools from Fourier analysis (without proof) first.

Definition 5.1.4. We define the **inner product** $\langle \cdot, \cdot \rangle$ on pairs of function $f, g : \mathbb{F}_2^n \to \mathbb{R}$ by

$$\langle f, g \rangle = 2^{-n} \sum_{x \in \mathbb{F}_2^n} f(x)g(x)$$

Definition 5.1.5 (Fourier Expansion)**.** Every function $f : \mathbb{F}_2^n \to \mathbb{R}$ can be uniquely expressed as a multilinear polynomial,

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(x)$$

where $\chi_\alpha(x) = (-1)^{\langle \alpha, x \rangle}$. This expression is called the **Fourier expansion** of f, and the real number $\hat{f}(\alpha) = \langle f, \chi_\alpha \rangle$ is called the **Fourier coefficient** of f on S. Collectively, the coefficients are called the **Fourier spectrum** of f.

Definition 5.1.6 (Noise Operator)**.** For $0 \geq \rho \geq 1$ and $f : \mathbb{F}_2^n \to \mathbb{R}$, we define the **noise operator** with parameter $\rho, T_\rho(f) : \mathbb{F}_2^n \to \mathbb{R}$ on the function f by

$$T_\rho(f)(x) = \sum_{y \in \mathbb{F}_2^n} \left(\frac{1-\rho}{2}\right)^{|y|} \cdot \left(\frac{1+\rho}{2}\right)^{n-|y|} \cdot f(x+y)$$

Proposition 5.1.3. $\widehat{T_\rho(f)}(\alpha) = \rho^{|\alpha|}\hat{f}(\alpha)$.

Definition 5.1.7. Let the parameter $\rho \in (0, 1)$ and the linear subspace $U \subseteq \mathbb{F}_2^n$ be given. The function $\text{energy}_{U,\rho} : \mathbb{F}_2^n \to \mathbb{R}$ is defined as

$$\text{energy}_{U,\rho}(x) = \frac{1}{W_\rho(U)} \cdot \sum_{u \in U} \rho^{|u+x|}$$

Notice that $\text{energy}_{U,\rho}(x) \in (0, 1]$ and $\text{energy}_{U,\rho}(x) = 1$ if and only if $x \in U$.

Theorem 5.1.4 (MacWilliam's Theorem [MS77])**.** *Let $U \subseteq \mathbb{F}_2^n$ be a linear subspace with $\dim U = k$. Then, for any parameter $\rho \in (0,1)$,*

$$W_\rho(U^\perp) = \frac{(1+\rho)^n}{2^k} \cdot W_{\frac{1-\rho}{1+\rho}}(U)$$

Now we are ready to prove Theorem 5.1.2.

Proof of Theorem 5.1.2. We use $\mathbf{1}_U : \mathbb{F}_2^n \to \{0,1\}$ to denote the **indicator function** for U, i.e., $\mathbf{1}_U = 1$ if and only if $x \in U$. Then,

$$\begin{aligned} T_\rho(\mathbf{1}_U)(x) &= \sum_{y \in \mathbb{F}_2^n} \Big(\frac{1-\rho}{2}\Big)^{|y|} \cdot \Big(\frac{1+\rho}{2}\Big)^{n-|y|} \cdot \mathbf{1}_U(x+y) \\ &= \Big(\frac{1+\rho}{2}\Big)^n \cdot \sum_{y \in \mathbb{F}_2^n} \Big(\frac{1-\rho}{1+\rho}\Big)^{|y|} \cdot \mathbf{1}_U(x+y) \\ &= \Big(\frac{1+\rho}{2}\Big)^n \cdot \sum_{u \in U} \Big(\frac{1-\rho}{1+\rho}\Big)^{|x+u|} \\ &= \Big(\frac{1+\rho}{2}\Big)^n \cdot W_{\frac{1-\rho}{1+\rho}}(U) \cdot \mathrm{energy}_{U,\frac{1-\rho}{1+\rho}}(x) \end{aligned} \tag{5.1.2.1}$$

Note that because U is a subspace, we have for $\alpha \notin U^\perp$, $\chi_\alpha(x) = 1$ for exactly half of the time and $\chi_\alpha(x) = -1$ for exactly the other half,

$$\langle \mathbf{1}_U, \chi_\alpha \rangle = \frac{1}{2^n}\Big(\sum_{x \in U} f(x)\chi_\alpha(x) + \sum_{x \notin U} f(x)\chi_\alpha(x)\Big) = \frac{1}{2^n}\sum_{x \in U} f(x)\chi_\alpha(x) = 0$$

As for $\alpha \in U^\perp$,

$$\langle \mathbf{1}_U, \chi_\alpha \rangle = \frac{1}{2^n}\Big(\sum_{x \in U} f(x)\chi_\alpha(x) + \sum_{x \notin U} f(x)\chi_\alpha(x)\Big) = \frac{1}{2^n}\sum_{x \in U} \chi_\alpha(x) = \frac{1}{2^n} \cdot 2^k = 2^{k-n}$$

Therefore,

$$\widehat{\mathbf{1}_U}(\alpha) = \begin{cases} 2^{k-n}, & \alpha \in U^\perp \\ 0, & \text{otherwise} \end{cases}$$

By Proposition 5.1.3,

$$\begin{aligned} T_\rho(\mathbf{1}_U)(x) &= \sum_{\alpha\in\mathbb{F}_2^n} \widehat{T_\rho(\mathbf{1}_U)}(\alpha)\cdot\chi_\alpha \\ &= \sum_{\alpha\in\mathbb{F}_2^n} \rho^{|\alpha|}\widehat{\mathbf{1}_U}(\alpha)\cdot\chi_\alpha \\ &= \sum_{\alpha\in U^\perp} \rho^{|\alpha|}\widehat{\mathbf{1}_U}(\alpha)\cdot\chi_\alpha \\ &= 2^{k-n}\sum_{\alpha\in U^\perp} \rho^{|\alpha|}\cdot\chi_\alpha \end{aligned}$$

By Definition 5.1.3, we have

$$\begin{aligned} T_\rho(\mathbf{1}_U)(x) &= 2^{k-n}\sum_{\alpha\in U^\perp} \rho^{|\alpha|}\cdot\chi_\alpha \\ &= 2^{k-n}\cdot W_\rho(U^\perp)\cdot p_{U^\perp,\rho}(x) \end{aligned}$$

By MacWilliam's Theorem, 5.1.4, we have

$$\begin{aligned} T_\rho(\mathbf{1}_U)(x) &= 2^{k-n}\cdot W_\rho(U^\perp)\cdot p_{U^\perp,\rho}(x) \\ &= 2^{k-n}\cdot\frac{(1+\rho)^n}{2^k}\cdot W_{\frac{1-\rho}{1+\rho}}(U)\cdot p_{U^\perp,\rho}(x) \\ &= \Big(\frac{1+\rho}{2}\Big)^n\cdot W_{\frac{1-\rho}{1+\rho}}(U)\cdot p_{U^\perp,\rho}(x) \end{aligned} \tag{5.1.2.2}$$

Combining Equation 5.1.2.1 and 5.1.2.2, we have

$$\mathrm{energy}_{U,\frac{1-\rho}{1+\rho}}(x) = p_{U^\perp,\rho}(x)$$

Let $d = \mathrm{dist}(x, U)$. Then there exists $w \in U$ such that $|x + w| = d$. By Definition 5.1.7, we have

$$W_{\frac{1-\rho}{1+\rho}}(U)\cdot\mathrm{energy}_{U,\frac{1-\rho}{1+\rho}}(x) = \sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u+x|}$$

Because U is a subspace,

$$\sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u+x|} = \sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u+x+w|}$$

Using triangle inequality, we have $|u+x+w| \leq |u|+|x+w|$. Thus,

$$\begin{aligned}\sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u+x+w|} &\geq \sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u|+|x+w|}\\ &= \Big(\frac{1-\rho}{1+\rho}\Big)^{d}\sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u|}\\ &= \Big(\frac{1-\rho}{1+\rho}\Big)^{d}\cdot W_{\frac{1-\rho}{1+\rho}}(U)\end{aligned}$$

In summary, we obtain

$$p_{U^\perp,\rho}(x) = \mathrm{energy}_{U,\frac{1-\rho}{1+\rho}}(x) = \frac{1}{W_{\frac{1-\rho}{1+\rho}}(U)}\cdot\sum_{u\in U}\Big(\frac{1-\rho}{1+\rho}\Big)^{|u+x+w|} \geq \Big(\frac{1-\rho}{1+\rho}\Big)^{d}$$

which concludes the proof. □

5.1.3 Random Sets

As we discussed before, to explicitly construct an (n,k,d)-rigid set, it suffices to explicitly construct a set S such that for every $U \subset \mathbb{F}_2^n$ with dimension $n-k$, there exists $s \in S$ such that $p_U(s) \leq 2^{-\Omega(d)}$. The following proposition suggests that a random set S has such property with high probability.

Notation 5.1.8. Let $\mathcal{P}_k$ denote the class of all U-polynomials $p_{U,\rho}$ with $\dim U = k$.

Proposition 5.1.5. *Let the parameter $\rho \in (\sqrt{2}-1, 1)$ and the linear subspace $U \subseteq \mathbb{F}_2^n, \dim U = n/2$ be given. Then, with high probability, for a random set $S \subset \mathbb{F}_2^n$ of size $O(n)$, the following holds: for every $p_{U,\rho} \in \mathcal{P}_{n/2}$,*

$$p_{U,\rho}(s) \leq 2^{\Omega(n)}$$

for at least half of the elements $s \in S$.

Before diving the proof, we need an auxiliary lemma.

Lemma 5.1.6. *Let the parameter $\rho \in (0,1)$ and the linear subspace $U \subseteq \mathbb{F}_2^n, \dim U = n/2$ be given. Then*

$$W_\rho(U) \geq \Big(\frac{1+\rho}{\sqrt{2}}\Big)^n$$

Proof. Because U is a linear subspace with $\dim U = n/2$, there are exactly $2^{n/2}$ cosets $x+U$ of subspace U. For each coset, as $\rho < 1$,

$$\sum_{w \in x+U} \rho^{|w|} \leq \sum_{u \in U} \rho^{|u|} = W_\rho(U)$$

On the other hand, using binomial expansion, we have

$$(1+\rho)^n = \sum_{w \in \mathbb{F}_2^n} \rho^{|w|} = \sum_x \sum_{w \in x+U} \rho^{|w|} \leq \sum_x W_\rho(U) = 2^{n/2} W_\rho(U)$$

which concludes the proof. □

Proof of Proposition 5.1.5. Let $p_{U,\rho} \in \mathcal{P}_{n/2}$ be given. Then,

$$\begin{aligned}\mathbb{E}_{x \sim \mathbb{F}_2^n}[p_{U,\rho}(x)] &= \mathbb{E}_{x \sim \mathbb{F}_2^n}\left[\frac{1}{W_\rho(U)} \cdot \sum_{u \in U} \rho^{|u|}(-1)^{\langle u,x\rangle}\right] \\ &= \frac{1}{W_\rho(U)} \cdot \sum_{u \in U} \Big(\rho^{|u|} \cdot \mathbb{E}_{x \sim \mathbb{F}_2^n}[(-1)^{\langle u,x\rangle}]\Big)\end{aligned}$$

Notice that $\mathbb{E}_{x \sim \mathbb{F}_2^n}[(-1)^{\langle u,x\rangle}] = 1$ if $u = 0$ and $\mathbb{E}_{x \sim \mathbb{F}_2^n}[(-1)^{\langle u,x\rangle}] = 0$ otherwise. By Lemma 5.1.6,

$$\mathbb{E}_{x \sim \mathbb{F}_2^n}[p_{U,\rho}(x)] = \frac{1}{W_\rho(U)} \cdot \sum_{u \in U} \Big(\rho^{|u|} \cdot \mathbb{E}_{x \sim \mathbb{F}_2^n}[(-1)^{\langle u,x\rangle}]\Big) = \frac{1}{W_\rho(U)} \leq \Big(\frac{\sqrt{2}}{1+\rho}\Big)^n$$

Because $\rho > \sqrt{2}-1$, $\sqrt{2}/(1+\rho) < 1$. Hence, there exists a constant α such that

$$\mathbb{E}_{x \sim \mathbb{F}_2^n}[p_{U,\rho}(x)] \leq \Big(\frac{\sqrt{2}}{1+\rho}\Big)^n < 2^{\alpha n}$$

By Markov's inequality, we have

$$\mathbb{P}_{x\sim\mathbb{F}_2^n}[p_{U,\rho}(x) > 2^{\alpha n/2}] \leq 2^{-\alpha n/2}$$

For some $m \leq n$, let $x_1, ..., x_m$ be independent and uniformly random vectors in $\mathbb{F}_2^n$. Let $\mathcal{E}$ be the event that there exists some subset $S \subseteq [n], |S| = m/2$ such that for all $i \in S$, $p_{U,\rho}(x_i) > 2^{\alpha n/2}$. We have

$$\begin{aligned}\mathbb{P}_{x_1,\ldots,x_m\sim\mathbb{F}_2^n}[\mathcal{E}] &\leq \binom{m}{m/2} \cdot (\mathbb{P}_{x\sim\mathbb{F}_2^n}[p_{U,\rho}(x) > 2^{-\alpha n/2}])^{m/2} \\ &\leq \binom{m}{m/2} \cdot 2^{-\alpha nm/4} \leq 2^m \cdot 2^{-\alpha nm/4}\end{aligned}$$

Set $m = O(n)$, we have that $\mathbb{P}_{x_1,\ldots,x_m\sim\mathbb{F}_2^n}[\mathcal{E}] = 2^{-\Omega(n^2)}$. □

5.2 Small Biased Sets Are Rigid

We are now ready to prove Theorem 5.1.1 using U-polynomials.

1st proof of Theorem 5.1.1. Let d be given. Let $S \subseteq \mathbb{F}_2^n$ be an ε-biased set with $\varepsilon = \exp(-d)$ and $|S| = O(n/\varepsilon^3)$. Let $U \subseteq \mathbb{F}_2^n$ be a linear subspace with $\dim U = n/2$. By Lemma 5.1.6, for $\rho \in (\sqrt{2} - 1, 1)$ and suitable constant d, we have

$$W_\rho(U) \geq \left(\frac{1+\rho}{\sqrt{2}}\right)^n \geq \exp(d) = \frac{1}{\varepsilon}$$

Now,

$$\begin{aligned}
\mathbb{E}_{x\sim S}[p_{U,\rho}(x)] &= \mathbb{E}_{x\sim S}\left[\frac{1}{W_\rho(U)}\cdot\sum_{u\in U}\rho^{|u|}(-1)^{\langle u,x\rangle}\right] \\
&= \frac{1}{W_\rho(U)}\cdot\sum_{u\in U}\Big(\rho^{|u|}\cdot\mathbb{E}_{x\sim S}[(-1)^{\langle u,x\rangle}]\Big) \\
&= \frac{1}{W_\rho(U)}\cdot\Big(1+\varepsilon\sum_{u\in U,u\neq 0}\rho^{|u|}\Big) \\
&< \frac{1}{W_\rho(U)}\cdot\Big(1+\varepsilon\frac{1}{W_\rho(U)}\Big) \\
&= \frac{1}{W_\rho(U)}+\varepsilon \\
&\leq 2\varepsilon
\end{aligned}$$

Because $\log(1/x)$ is a convex function, by Jensen's Inequality,

$$\mathbb{E}_{x\sim S}[\log\frac{1}{p_{U,\rho}(x)}] \geq \log\frac{1}{\mathbb{E}_{x\sim S}[p_{U,\rho}(x)]} > \log\frac{1}{2\varepsilon}$$

Because U is a linear subspace with $\dim U = n/2$, its dual $U^\perp$ is also a linear subspace with $\dim U = n/2$. Hence, by Theorem 5.1.2,

$$\mathbb{E}_{x\sim S}[\mathrm{dist}(x,U)] = \mathbb{E}_{x\sim S}[\Omega\Big(\log\frac{1}{p_{U^\perp,\rho}(x)}\Big)] = \Omega\Big(\log\frac{1}{\varepsilon}\Big)$$

In summary, S is a $(n, n/2, d)$-strong rigid set with $|S| = O(n/\varepsilon^3) = n\cdot 2^{\Theta(d)}$. □

Alon and Cohen also gave two alternative proofs for Theorem 5.1.1, one using bias reduction and the other using expander graphs, which we discuss below.

5.2.1 Bias Reduction

The second proof relies on the Parity Lemma [NN93].

Definition 5.2.1 (Statistical Distance)**.** Let $x \sim X, y \sim Y$ be two random variables from two distributions X, Y on the same support S. We define the statistical distance between X and Y by

$$\mathrm{SD}(X, Y) := \max_{A \subseteq S} \left| \mathbb{P}[x \in A] - \mathbb{P}[y \in A] \right|$$

This following lemma says that the projection of a small biased set onto a small set of coordinates is close to the uniform distribution, which we state without proof.

Lemma 5.2.1 (The Parity Lemma [NN93])**.** *Let $S \subseteq \mathbb{F}_2^n$ is an ε-biased set. Let $T \subseteq [n]$ be an nonempty set of size k. Denote S_T the projection of S on the index set T. Then,*

$$\mathrm{SD}(S_T, \mathrm{U}_k) \leq \varepsilon \cdot 2^{k/2}$$

The name "bias-reduction" is exemplified by the following lemma.

Lemma 5.2.2 ([AC15])**.** *Let $S \subseteq \mathbb{F}_2^n$ is an ε-biased set. For every integer $c \geq 1$, let $c \cdot S$ denote the set $S + ... + S$ for c times. We have $c \cdot S$ is an ε^c-biased set.*

Proof. For every nonzero $\alpha \in \mathbb{F}_2^n$,

$$\begin{aligned}
\left| \mathbb{E}_{s \sim c \cdot S}[(-1)^{\langle \alpha, s \rangle}] \right| &= \left| \mathbb{E}_{s_1, ..., s_c \sim S}[(-1)^{\langle \alpha, s_1 + ... + s_c \rangle}] \right| \\
&= \left| \mathbb{E}_{s_1, ..., s_c \sim S}[\prod_{i=1}^{c} (-1)^{\langle \alpha, s_i \rangle}] \right| \\
&= \prod_{i=1}^{c} \left| \mathbb{E}_{s_i \sim S}[(-1)^{\langle \alpha, s_i \rangle}] \right| \leq \varepsilon^c
\end{aligned}$$

□

Here we present the second proof, the main idea is to "reduce the bias enough so as to cancel the exponential loss incurred by the Parity Lemma" [AC15].

2nd proof of Theorem 5.1.1. Let $U \subseteq \mathbb{F}_2^n$ be a subspace with $\dim U = n/2$. Then,

$$\mathbb{P}_{x \in \mathbb{F}_2^n}[\mathrm{dist}(x, U) > n/10] > 0.6$$

Let S be an 2^{-cd}-biased set where we choose constant c that satisfies

$$2^{-cn/20+n/2} < 0.1$$

Let $S' = (n/20d) \cdot S$. Then, Lemma 5.2.2, we have that S' is $2^{-cn/20}$-biased. By the Parity Lemma 5.2.1, we have

$$\mathrm{SD}(S', \mathrm{U}_n) \leq 2^{-cn/20} 2^{n/2} < 0.1$$

where U_n is the uniform distribution on $\mathbb{F}_2^n$. This gives us

$$\mathbb{P}_{x \sim S'}[\mathrm{dist}(x, U) > n/10] > 0.6 - 0.1 = 0.5$$

By Markov's inequality, this implies

$$\mathbb{E}_{x \sim S'}[\mathrm{dist}(x, U)] > n/20$$

Equivalently, let $k = (n/20d)$ because $S' = k \cdot S$,

$$n/20 < \mathbb{E}_{x \sim S'}[\mathrm{dist}(x, U)] = \mathbb{E}_{x_1, \ldots, x_k \sim S}[\mathrm{dist}(\sum_{i=1}^{k} x_i, U)]$$

Notice

$$\mathrm{dist}(\sum_{i=1}^{k} x_i, U) = |\sum_{i=1}^{k} x_i + u|$$

for some u. For all $i \in [k]$, let u_i be such that $|s_i + u_i| = \text{dist}(s_i, U)$, then

$$\text{dist}(\sum_{i=1}^{k} x_i, U) = |\sum_{i=1}^{k} x_i + u| \leq |\sum_{i=1}^{k} x_i + \sum_{i=1}^{k} u_i| \leq \sum_{i=1}^{k} |x_i + u_i| = \sum_{i=1}^{k} \text{dist}(s_i, U)$$

Then because each x_i are independent

$$\begin{aligned}
\mathbb{E}_{x\sim S}[\text{dist}(x, U)] &= \frac{\sum_{i=1}^{k} \mathbb{E}_{x\sim S}[\text{dist}(x, U)]}{k} \\
&= \frac{\mathbb{E}_{x_1,\ldots,x_k\sim S}[\sum_{i=1}^{k} \text{dist}(x_i, U)]}{k} \\
&\geq \frac{\mathbb{E}_{x_1,\ldots,x_k\sim S}[\text{dist}(\sum_{i=1}^{k} x_i, U)]}{k} \\
&\geq \frac{\mathbb{E}_{x\sim S'}[\text{dist}(x, U)]}{k} \\
&> \frac{n/20}{k} = \frac{n/20}{n/20d} = d
\end{aligned}$$

Hence, S is a $(n, n/2, d)$-strong rigid set. □

5.2.2 Expander Graphs

Again, we introduce some backgrounds for expander graphs before proving Theorem 5.1.1.

Definition 5.2.2. If every vertex of a graph G has degree k, then G is said to be k**-regular**.

Let $G = (V.E)$ be an undirected D-regular on n vertices. Let A_G be the **normalised adjacency matrix** of G. That is, for any $u, v \in V$,

$$(A_G)_{u,v} = \frac{\text{number of edges between } u \text{ and } v}{D}$$

Definition 5.2.3. A D-regular graph G on n vertices is called a (n, D, λ)**-expander** if the absolute value of the second largest eigenvalue is λ.

Lemma 5.2.3 ([AC88]). *Let $G = (V, E)$ be an (n, D, λ)-expander. Then for any set $S \subseteq V$ with size $S = \alpha n$,*

$$\left| e(S) - \frac{1}{2} D\alpha^2 n \right| \leq \frac{1}{2} \lambda D\alpha(1-\alpha)n$$

Theorem 5.2.4 ([AR94]). *Let $S \subseteq \mathbb{F}_2^n$ be an ε-biased set. Let $G_S = (V, E)$ be a subgraph of the Boolean cube on $\mathbb{F}_2^n$ such that $V = \mathbb{F}_2^n$ and an edge connects a pair $u, v \in V$ if $u + v \in S$. Then G_S is a $(2^n, |S|, \varepsilon)$-expander.*

Lemma 5.2.5 ([AC15]). *Let $S \subseteq \mathbb{F}_2^n$ be an ε-biased set. Then for any subspace $U \subseteq \mathbb{F}_2^n$ with $\dim U = k$,*

$$\left| \frac{S \cap U}{S} \right| \leq \frac{1}{2^{n-k}} + \varepsilon$$

Proof. Let $G_S = (V, E)$ be constructed as in Theorem 5.2.4. Then G_S is a $(2^n, |S|, \varepsilon)$-expander. Let $U \subset V = \mathbb{F}_2^n$ be a subspace with with $\dim U = k$. Then, for any vertex $u \in U$, we have the degree of u in the induced subgraph of G_S on U is

$$\Big| \{ s \in S | u + s \in U \} \Big| = \Big| \{ s \in S | s \in U \} \Big| = |U \cap S|$$

Hence,

$$e(U) = \frac{1}{2} \sum_{u \in u} \deg(u) = \frac{1}{2} \sum_{u \in u} |U \cap S| = \frac{1}{2} |U| \cdot |U \cap S|$$

By Lemma 5.2.3, we have

$$|U| \cdot |U \cap S| = 2e(U) \leq |S| \big(\frac{|U|}{2^n} \big)^2 2^n + \varepsilon \cdot |S| \cdot |U|$$

Hence,

$$\frac{|U \cap S|}{|S|} \leq \frac{|U|}{2^n} + \varepsilon$$

□

3rd proof of Theorem 5.1.1. Let $U \subset \mathbb{F}_2^n$ be a subspace with with $\dim U = n/2$. We partition the n unit vectors of $\mathbb{F}_2^n$ into $8d$ disjoint sets

$$B_1, ..., B_{8d}$$

each of size $n/8d$. For each subset $I \subset [8d]$ with $|I| = 2d$, we define

$$U_I = \text{span}(U \cup \bigcup_{i \in I} B_i)$$

Hence, we have that for each I, $\dim U_I \leq 3n/4$ and for every vector x, $\text{dist}(x, U_I) \leq 2d$ for some I. Let $S \subseteq \mathbb{F}_2^n$ be an ε-biased set. By Lemma 5.2.5, for every I, we have

$$|S \cap U_I| \leq |S| \cdot (\frac{1}{2^{n-3n/4}} + \varepsilon) = |S| \cdot (\frac{1}{2^{n/4}} + \varepsilon)$$

Because there are at most

$$\binom{8d}{2d} < 120^d$$

such sets I, which covers all x such that $\text{dist}(x, U) \leq 2d$. Thus there are at most

$$120^d |S| \cdot (\frac{1}{2^{n/4}} + \varepsilon)$$

vectors x in S such that $\text{dist}(x, U) \leq 2d$. Set $\varepsilon = 1/(120^d \cdot 4)$, then

$$120^d |S| \cdot (\frac{1}{2^{n/4}} + \varepsilon) \leq 120^d |S| \cdot (\frac{1}{2^{n/4}} + \frac{1}{120^d \cdot 4}) \leq |S| \cdot (\frac{2^{7d}}{2^{n/4}} + 1/4) \leq \frac{|S|}{2}$$

for $c \leq 1/28$ and $d \leq cn$. Hence, we have that at most half of the vectors in S are at most $2d$-far from U, which gives us

$$\mathbb{E}_{x \sim S}[\text{dist}(x, U)] \geq d$$

and S is a $(n, n/2, d)$-strong rigid set. □

Chapter 6

Linear Data Structures and Rigidity

In the last chapter, we saw the $2^{O(d)}n$-barrier for constructing explicit rigid sets. Meanwhile, we are also stuck at proving strong lower bounds for linear data structures. A recent line of work found that such difficulty is not an accident [DGW19, RR20]. They show that sufficiently strong lower bounds for linear data structures would imply new bounds for rigid matrices and rigid matrices directly correspond to hard query sets for the linear data structures. In this chapter, we present the main idea on how such a link is established. For simplicity, we omit the discussion of explicitness in the chapter, we refer the readers to the referenced papers for further details.

6.1 Linear Data Structures and Row Rigidity

We first present the idea by Dvir, Golovnev and Weinstein [DGW19], who found a connection between linear data structures lower bounds and row rigidity via establishing a new result on the Paturi-Pudlák dimensions. Before we dive in, we first re-introduce the definition of sparsity.

Definition 6.1.1 (Sparsity, [DGW19])**.** A vector $v \in \mathbb{F}^n$ is **s-sparse** if the number of non-zero coordinates in v is at most s. A matrix $A \in \mathbb{F}^{m\times n}$ is **s-sparse** if it has s nonzero entires. A is **s-row sparse** if each of its row is s-sparse. A subspace $V \in \mathbb{F}^m$ is **s-sparse** if it is the *column space* of a s-row

sparse matrix B.

Notice that the sparsity defined for subspaces are a bit unconventional. This leads to the inner dimensions and out dimensions defined with respect to *column* spaces, while in Definition 6.1.2 and 6.1.2, the inner dimensions and out dimensions defined with respect to *row* spaces. Dvir, Golovnev and Weinstein [DGW19] suggested this new definition and noted such a difference is essential in the context of this work: because the strong row rigidity (Definition 6.1.3) and linear data structures (Definition 6.1.4) are defined in a particular way, we need to work with the column space later for Lemma 6.1.3 and 6.1.4 in order to establish the link between linear data structure lower bounds and matrix rigidity (Theorem 6.1.5).

Definition 6.1.2 (Inner Dimension and Outer Dimension, [DGW19])**.** Let $V \subseteq \mathbb{F}^n$ be a subspace, and s be a positive integer less than n. We define

$$d_V(s) := \max_U \{\dim(V \cap U) : U \subseteq \mathbb{F}^n, \dim(U) \leq \dim(V), U \text{ is } s\text{-sparse}\}$$
$$D_V(s) := \max_U \{\dim(U) | U \subseteq \mathbb{F}^n, V \subseteq U, U \text{ is } s\text{-sparse}\}$$

The main building block to connect data structure lower bounds and row rigidity is the following theorem, which says that every matrix either has small outer dimension or contains a matrix of small inner dimension.

Theorem 6.1.1 ([DGW19])**.** *Let m, n, t, k be positive integers and let $0 < \varepsilon < 1$. If $M \in \mathbb{F}^{m \times n}$ is a matrix whose column space $V \subseteq \mathbb{F}^m$ has an outer dimension*

$$D_V(tk + n\varepsilon^k) \geq \frac{n}{1-\epsilon}$$

then for some $n' \geq n\varepsilon^k$, A contains a submatrix $M' \in \mathbb{F}^{m \times n'}$ whose column space $U \subseteq \mathbb{F}^m$ has an inner dimension

$$d_U \leq \text{rank}(B) - \varepsilon n'$$

To prove this result, we first need an auxiliary lemma:

Lemma 6.1.2 ([DGW19])**.** *Let m, n, k be positive integers. A matrix $M \in \mathbb{F}^{m\times n}$, whose column space $V \subseteq \mathbb{F}^m$ has an inner dimension $d_V \geq \text{rank}(M)-r$, can be decomposed as*

$$M = A \cdot B + M' \cdot C$$

where $M' \in \mathbb{F}^{m\times r}$ is a submatrix of M, $A \in \mathbb{F}^{m\times n}$ is t-row sparse, $B \in \mathbb{F}^{n\times n}, C \in \mathbb{F}^{r\times n}$.

Proof. By the definition of inner dimension (Definition 6.1.2), we have that there exists an s-sparse subspace $U \subset \mathbb{F}^n$ with $\dim(U) \leq \dim(V)$ and

$$\dim(V \cap U) \geq \text{rank}(M) - r$$

Let $A \in \mathbb{F}^{m\times n}$ be a t-row sparse matrix generating U. In order to generate the row space V, it suffices to extend A with at most k column vectors from M; we let M' be the matrix stacked with these column vectors. We then have that there exists some matrices $B \in \mathbb{F}^{n\times n}, C \in \mathbb{F}^{r\times n}$ satisfying $M = A \cdot B + C \cdot M'$. □

We are now ready to use this expansion to prove Theorem 6.1.1.

Proof of Theorem 6.1.1. Let $0 < \varepsilon < 1$. Suppose such an M' does not exist. Consider the following expansion of $M \in \mathbb{F}^{m\times n}$: starting with $i = 0, M_0 = M$, for $M_i \in \mathbb{F}^{m\times n_i}$, as long as V_i, the column space of M_i, satisfies

$$d_{V_i} \geq \text{rank}(M_i) - n\varepsilon^{i+1}$$

we use the previous lemma to get

$$M_i = A_i \cdot B_i + M_{i+1} \cdot C_i$$

where $M_{i+1} \in \mathbb{F}^{m \times n\varepsilon^{i+1}}$ is a submatrix of M_i, $A_i \in \mathbb{F}^{m \times n\varepsilon^i}$ is t-row sparse, $B_i \in \mathbb{F}^{n\varepsilon^i \times n\varepsilon^i}, C_i \in \mathbb{F}^{n\varepsilon^{i+1} \times n\varepsilon^i}$. Then for a given positive integer k, we then obtain an expansion of M as the following

$$\begin{aligned}
M = M_0 &= A_0 \cdot B_0 + M_1 \cdot C_0 \\
&= A_0 \cdot B_0 + (A_1 \cdot B_1 + M_2 \cdot C_1) \cdot C_0 \\
&= A_0 \cdot B_0 + A_1 \cdot B_1 \cdot C_0 + M_2 \cdot C_1 \cdot C_0 \\
&= A_0 \cdot B_0 + A_1 \cdot B_1 \cdot C_0 + A_2 \cdot B_2 \cdot C_1 \cdot C_0 + M_3 \cdot \cdot C_2 \cdot C_1 \cdot C_0 \\
&= A_0 \cdot B_0 + \sum_{i=1}^{k-1} A_i \cdot B_i \Big(\prod_{j=i-1}^{0} C_j \Big) + M_k \Big(\overset{=}{\prod_{j=k-1}} C_j \Big) \\
&= \begin{bmatrix} A_0 & A_1 & \dots & A_{k-1} & M_k \end{bmatrix} \cdot \begin{bmatrix} D_0 \\ D_1 \\ \vdots \\ D_{k-1} \\ D_k \end{bmatrix}
\end{aligned}$$

where

$$D_i = \begin{cases} B_0 & \text{for } i = 0 \\ B_i \Big(\prod_{j=i-1}^{0} C_j \Big) & \text{for } 1 \le i < k \\ \prod_{j=k-1}^{0} C_i & \text{for } i = k \end{cases}$$

Let

$$A := \begin{bmatrix} A_0 & A_1 & \dots & A_{k-1} & M_k \end{bmatrix}, \quad D = \begin{bmatrix} D_0 \\ D_1 \\ \vdots \\ D_{k-1} \\ D_k \end{bmatrix}$$

Then $M = A \times D$. Recall that for all i, A_i is t-sparse and $M_k \in \mathbb{F}^{n\varepsilon^k \times n}$. This means that A has at most $kt + n\varepsilon^k$ non-zero entries on each row. The number of rows of D is

$$\sum_{i=0}^{k} n\varepsilon^i < \frac{n}{1-\varepsilon}$$

This implies that M can be generated by a $kt + n\varepsilon^k$-row sparse matrix whose column subspace has dimension less than $n/(1-\varepsilon)$. This contradicts the fact that

$$D_V(tk + n\varepsilon^k) \geq \frac{n}{1-\epsilon}$$

and concludes the proof. □

6.1.1 Rigidity and Inner Dimension

Let's recall the definition of strong row rigidity, which as we saw earlier, implies the general notion of rigidity (see Theorem 4.3.2.)

Definition 6.1.3 (Strong Row Rigidity, [DGW19])**.** A matrix $A \in \mathbb{F}^{m\times n}$ is said to be (r, s)**-strongly row rigid** if for any invertible matrix $C \in \mathbb{F}^{n\times n}$, we have $A \times C$ is (r, s)-row rigid.

The following lemma shows that the definition of strong row rigidity of rectangular matrices is equivalent to small inner dimensions. In particular, we limit our attention to matrices with more rows than columns, i.e. $m > n$.

Lemma 6.1.3 ([DGW19])**.** *Let matrix $A \in \mathbb{F}^{m\times n}$ have rank n and let $V \subseteq \mathbb{F}^m$ be its columns space. Then the following are equivalent:*

1. *A is (r, s)-strongly row rigid.*
2. *$d_V(s) \leq \operatorname{rank}(A) - r$.*
3. *V is not contained in a subspace of the form $E \cup F$ where $E, F \subseteq \mathbb{F}^m$ are subspaces with $\dim(E) \leq n, \dim(F) < r$ and E is s-sparse.*

Proof. $(1 \Rightarrow 2)$: Suppose $d_V(t) > \operatorname{rank}(A) - r$. By the definition of inner dimension 6.1.2, there exists a subspace $U \subseteq \mathbb{F}^m, \dim(U) \leq \dim(V) =$

$\text{rank}(A) = n, U$ is s-sparse and $\dim(U \cap V) > \text{ rank}(A) - r$. In other words, there exists a subspace $W \in \mathbb{F}^m$ with $\dim(W) < r$ such that $V = U \cup W$. Let $C \in \mathbb{F}^{m \times n}$ be a s-row sparse basis matrix of U and $B \in \mathbb{F}^{m \times n}$ a basis matrix of W. There exists an invertible matrix $T \in \mathbb{F}^{n \times n}$ such that

$$A = C \times T + B$$

This implies $A \times T^{-1} = C + B \times T^{-1}$ is not (r, s)-row rigid because $\text{rank}(A \times T^{-1} - C) = \text{rank}(B \times T^{-1}) < r$. This means that A is not (r, s)-strongly row rigid and leads to a contradiction.

$(2 \Rightarrow 3)$: Because $d_V(s) \leq \text{ rank}(A) - r$, for all subspaces $E \subseteq \mathbb{F}^m$ such that $\dim(E) \leq n = \dim(V)$ and E is s-sparse, we have $\dim(V \cap E) \leq n - r$. That is, for any subspace $F \subseteq \mathbb{F}^m$ with $V \subseteq E \cup F$, we have

$$\begin{aligned} n - r \geq \dim(V \cap E) &= \dim(V) + \dim(E) - \dim(V \cup E) \\ &\geq \dim(V) + \dim(E) - \dim(E \cup F) \\ &= \dim(V) + \dim(E) - \dim(E) - \dim(F) + \dim(E \cap F) \\ &= n - \dim(F) + \dim(E \cap F) \end{aligned}$$

This implies $\dim(F) \geq r$.

$(3 \Rightarrow 1)$: Take any invertible $T \in \mathbb{F}^{n \times n}$, if we write

$$A \times T = C + B$$

where $C \in \mathbb{F}^{m \times n}$ be a s-row sparse and $B \in \mathbb{F}^{m \times n}$. Let E be the columns space of C and F the columns space of B. Because T is invertible and, $\text{rank}(A \times T) = \text{rank}(A) = \dim(V)$. As we have just seen, $\dim(F) \geq r$, hence $\text{rank}(B) > r$ and thus A must be (r, s)-strongly row rigid. $\square$

6.1.2 Linear Data Structures and Outer Dimension

A **linear data structure** problem with m queries over a field $\mathbb{F}$ and an input database $x \in \mathbb{F}^n$ is defined by a matrix $V \in \mathbb{F}^{m \times n}$. Each row V_i is a query of V and the answer to the ith query os given by $\langle V_i, x \rangle = (Vx)_i \in \mathbb{F}$. An (s, t) linear data structure $\mathcal{D}$ for the problem V in the cell-probe model is a pair $\mathcal{D} = (P, Q)$ where $P \in \mathbb{F}^{s \times n}$ is a processing map that encodes the database x into s memory cells and $Q \in \mathbb{F}^{m \times s}$ is a query map that answers every query of V by probing at most t memory cells. For simplicity, we give the following definition of linear data structure.

Definition 6.1.4 (Linear Data Structure, [DGW19])**.** We say that an (s, t) linear data structure $\mathcal{D} = (Q, P)$ computes a matrix $V \in \mathbb{F}^{m \times n}$ if V can be decomposed as

$$V = Q \cdot P$$

where $Q \in \mathbb{F}^{m \times s}$ is t-row sparse and $P \in \mathbb{F}^{s \times n}$.

The following lemma shows that if a matrix cannot be computed by a linear data structure, then it must have large outer dimension.

Lemma 6.1.4 ([DGW19])**.** *Let matrix $A \in \mathbb{F}^{m \times n}$ have rank n and let $V \subseteq \mathbb{F}^m$ be its columns space. Then there is an (s, t) linear data structure computing A if and only if $D_V(s) \leq s$.*

Proof. Suppose the (s, t) linear data structure $\mathcal{D} = (Q, P)$ compute A. Then we have

$$A = Q \cdot P$$

where $Q \in \mathbb{F}^{m \times s}$ is t-row sparse and $P \in \mathbb{F}^{s \times n}$. Let U be the column space of A. Then U is t-row sparse and $V \subseteq U$. This implies $D_V(s) \leq \dim(U) \leq s$.

On the other hand, suppose $D_V(s) \leq s$, then there exists a t-sparse subspace $U \subset \mathbb{F}^m$ with $V \subset U, \dim(U) \leq s$. Let $Q \in \mathbb{F}^{m \times s}$ be a t-row sparse matrix whose columns generate U. Then as $V \subset U$, each column of A can be written as a linear combination of the columns of Q. In other words, there exists a matrix $P \in \mathbb{F}^{s \times n}$ such that $A = Q \cdot P$. □

6.1.3 Data Structure Lower Bounds Imply Strong Row Rigidity

We are now ready to prove the main result in [DGW19].

Theorem 6.1.5 ([DGW19])**.** *Let $\varepsilon > 0$ be a constant. If the linear map given by a matrix $M \in \mathbb{F}^{m \times n}$ cannot be solved by an*

$$\Big(\frac{n}{1-\varepsilon}, (t-1) \cdot \frac{\log(n/t)}{\log(1/\varepsilon)}\Big)$$

linear data structure, then M contains an $(\varepsilon n', t)$-row rigid submatrix $M \in \mathbb{F}^{m \times n'}$ for some $n' \geq t$.

Proof. Let V be the column space of M. Since $M \in \mathbb{F}^{m \times n}$ cannot be solved by the specified linear data structure, by Lemma 6.1.4, we have

$$D_V((t-1) \cdot \frac{\log(n/t)}{\log(1/\varepsilon)}) > \frac{n}{1-\varepsilon}$$

Set

$$k = \frac{\log(n/t)}{\log(1/\varepsilon)}$$

By Theorem 6.1.1, M contains a submatrix $M' \in \mathbb{F}^{m \times n'}$, whose column space is V' with

$$d_{V'}(t) \leq \operatorname{rank}(M') - \varepsilon n'$$

for

$$n' \geq n\varepsilon^k = n\varepsilon^{\frac{\log(n/t)}{\log(1/\varepsilon)}} = n \cdot \frac{t}{n} = t$$

By Lemma 6.1.3, we have that M' is $(\varepsilon n', t)$-row rigid. □

6.2 Linear Data Structures and Rigid Sets

In this section, we discuss the connection between data structure lower bounds and rigid sets, established by [RR20].

6.2.1 Systematic Linear Data Structure Model

The task of the **inner product problem** is to preprocess a vector $v \in \mathbb{F}_2^n$ to compute inner products. The queries are specified by a set $Q \subseteq \mathbb{F}_2^n$, called **query set**, and the data structure computes the inner product $\langle q, v \rangle$ of v and any **query** $q \in Q$. During preprocessing, a **systematic linear data structure model** stores v and k extra bits, which are the evaluations of k linear functions $\langle a_1, v \rangle, ..., \langle a_k, v \rangle$ where $a_1, ..., a_k \in \mathbb{F}_2^n$. To compute the answer on query q, the data structure outputs a linear combination of these k bits and any d entries from v.

Definition 6.2.1 (Systematic Linear Model)**.** Let the systematic linear data structure model be defined above. For a set $Q \subseteq \mathbb{F}_2^n$, we define the **time** $T(Q, k)$ by

$$T(Q,k) := \max_{v \in \mathbb{F}_2^n} \Big(\min\{d | \text{ can compute } \langle q, v \rangle \forall q \in Q \text{ as a linear combination of } k \text{ extra bits and any } d \text{ bits of } v\} \Big)$$

where we are only allowed to output a linear function of k precomputed linear functions of v along with any d bits of v.

The following theorem states the equivalence of rigid sets and the data structure lower bound of the systematic linear model.

Theorem 6.2.1 ([RR20])**.** *A set $Q \subseteq \mathbb{F}_2^n$ is (n, k, d)-rigid if and only if $T(Q, k) \geq d$.*

Proof. ($\Leftarrow$) Suppose that Q is not (n, k, d)-rigid. That is, there is a k-dimensional subspace $U \subseteq \mathbb{F}_2^n$ such that $\text{dist}(q, U) < d$ for all $q \in Q$. Let $v \in \mathbb{F}_2^n$ be the input data and $\{b_1, ..., b_k\}$ be a basis of U. Let the data structure store $\langle b_1, v\rangle, ..., \langle b_k, v\rangle$. Then, there exists a $q_u \in U$ such that

$$\text{dist}(q, q_u) < d$$

Because q_u is a linear combination of $\{b_1, ..., b_k\}$, we have $T(Q, k) < d$.
($\Rightarrow$) Suppose that Q is (n, k, d)-rigid. Let $\{e_1, ..., e_n\}$ be the standard basis and $t = T(Q, k)$ be the query time. Let the evaluations of k linear functions $\langle a_1, v\rangle, ..., \langle a_k, v\rangle$, where $a_1, ..., a_k \in \mathbb{F}_2^n$, be given and let $U = \text{span}(a_1, ..., a_k)$. Because Q is (n, k, d)-rigid, there exists a $q^* \in Q$ such that

$$\text{dist}(q^*, U) \geq d$$

Let q^* be the query and assume that we can access bits $v_{i_1}, ..., v_{i_t}$ of v. Let

$$V = \text{span}(a_1, ..., a_k, e_{i_1}, ..., e_{i_t})$$

Then

$$\text{dist}(q^*, U) \leq \text{dist}(q^*, V) + t$$

It remains to show that $\text{dist}(q^*, V) = 0$, which would imply $d \leq \text{dist}(q^*, U) \leq t = T(Q, k)$. Suppose that $\text{dist}(q^*, V) \geq 1$, there exists a vector $y \in \mathbb{F}_2^n$ such that $\langle y, q^*\rangle = 1$ and $\langle y, x\rangle = 0$ for all $x \in V$. This implies

$$\langle y + v, x\rangle = \langle y, x\rangle + \langle v, x\rangle = \langle v, x\rangle$$

However,

$$\langle q^*, y + v\rangle \neq \langle q^*, v\rangle$$

which implies that the output on query q^* will err either on v or $y + v$. $\square$

6.2.2 Linear Data Structure Model

A **linear data structure model**, on the other hand, stores s bits, which are the evaluations of s linear functions $\langle a_1, v\rangle, ..., \langle a_s, v\rangle$ where $a_1, ..., a_s \in \mathbb{F}_2^n$. To compute the answer on query q, the data structure outputs a linear combination of these s bits. Notice that the systematic linear model is different, as the query algorithm for the systematic model is not charged for accessing the k precomputed bits. The **time** $LT(Q, s)$ for linear model is defined to be

$$LT(Q,s) := \max_{v\in\mathbb{F}_2^n}\Big(\min\{d|\text{ can compute } \langle q, v\rangle \forall q \in Q \text{ as a linear combination of any } d \text{ bits chosen from the } s \text{ stored bits}\}\Big)$$

In the linear model, we can simply add the n bits of v by $\langle e_i, v\rangle, i \in [n]$. Taking into account the k precomputed bits that the systematic model can access without charge, we obtain the desired $LT(Q, n + k) \leq d + k$. This idea leads to the following proposition, which gives a simple comparison between the linear model and the systematic linear model.

Proposition 6.2.2. *If $T(Q, k) \leq d$, then $LT(Q, n + k) \leq d + k$.*

The following lemma, whose proof is similar to that of Theorem 5.0.1, says that we can find rigid sets contained in a $2k$-dimensional space based on a given rigid sets in the n-dimensional space.

Lemma 6.2.3. *Let $S \subseteq \mathbb{F}_2^n$ be (n, k, d)-rigid of size m. Then there exists a set $S' \subseteq \mathbb{F}_2^{2k}$ of size at most $m \cdot \lceil n/2k\rceil$ that is $(2k, k, dk/n)$-rigid.*

Proof. Let $r = n/2k$. Without loss of generality, assume that r is an integer. We first split the coordinates into r blocks $Z_1, Z_2, ..., Z_r$. Let $S_i, i \in [r]$ be

the set obtain from S by projecting each vector $v \in S$ to the ith block. Let $S' = \bigcup_i S_i$. We show that S' is $(2k, k, dk/n)$-rigid. Suppose not. Then there is a subspace $V \subseteq \mathbb{F}_2^{2k}$ such that $\text{dist}(v, V) < dk/n$ for all $v \in S'$. Because every vector $s \in S$ is the sum of at most r vectors in S', there is a subspace $U \subseteq \mathbb{F}_2^n$, where each $u \in U$ is a vector of r copies of a vector $v \in V$, such that for all $s \in S$, $\text{dist}(s, U) \leq (dk/n) \cdot r = (dk/n) \cdot (n/2k) < d$, which is a contradiction. □

The following theorem states the equivalence of rigid sets and the data structure lower bound of the linear model.

Theorem 6.2.4 ([RR20])**.** *Let $k = LT(Q, 3n/2)$ and $Q \subseteq \mathbb{F}_2^n$ of size m be a query set. Then there exists a $(k, k/2, k^2/(4n))$-rigid set $Q' \subset \mathbb{F}_2^n$ with size $m \cdot \lceil n/k \rceil$ if $k \geq 2\sqrt{n}$.*

Proof. Because $k = LT(Q, 3n/2)$ and $k \leq n$, we have that $LT(Q, n+k/2) \geq k$. By Proposition 6.2.2, $T(Q, k/2) \geq k/2$. By Theorem 6.2.1, we have Q is $(n, k/2, k/2)$ rigid. By Lemma 6.2.3, we have that there is a set Q' of size at most $m \cdot \lceil n/k \rceil$ and is $(k, k/2, k^2/(4n))$-rigid. □

Chapter 7

Non-Rigidity

To our surprise, many families of matrices we saw before are now known to be non-rigid, that is, they fail to meet the following goal suggested in the first chapter: $\mathcal{R}_A^{\mathbb{F}}(n/\log\log n) = n^{1+\delta}$ for some $\delta > 0$. This includes some generating matrices of a good error correcting code (Dvir), many families of Hadamard matrices [AW15, DL20], and Kronecker products of many smaller matrices [DL20, Alm21, Kiv21]. In this chapter we present the main tools used in these proofs, and for simplicity, we restrict ourselves to the binary field $\mathbb{F}_2$ throughout this chapter. All the theorems in this chapter can be generalized to finite fields with slightly different parameters, but the proofs can be a bit more technical.

7.1 Error Correcting Codes

Based on our observations in previous sections, it is tempting to conjecture that the generating matrices of error correcting codes have high rigidity. This conjecture, however, is false. We remark that this is not the end for error correcting codes. As we saw in the first chapter, a random matrix has high rigidity, meanwhile, as we show below, a random generating matrix G is a good code with high probability. The observation is that for any nonzero vector $v \in \{0,1\}^k$, the vector vG has entries that are distributed uniformly and independently in $\{0,1\}^n$. Let $\mathrm{Vol}(d,n)$ denote the volume of the Hamming

ball with vectors of length n and Hamming weight at most d.

Proposition 7.1.1. *If Vol$(d-1,n) < 2^{n-k}$ then there exists linear code of dimension k and distance at least d in $\{0,1\}^n$.*

Proof. Let G be a generating matrix whose entries are chosen uniformly at random. Let $v \in \{0,1\}^k$ be a nonzero vector. Then the probability that the vector vG has at most $d-1$ entries is thus $(\mathrm{Vol}(d-1,n)/2^n)$. As there are $2^k - 1$ nonzero vectors, as long as

$$(2^k - 1) \cdot \frac{\mathrm{Vol}(d-1,n)}{2^n} < 1$$

there must exists a linear code of dimension k and distance at least d in $\{0,1\}^n$. □

Hence, if $n >> d$, then $\mathrm{Vol}(d-1,n)/2^{n-k} << 1$, which means a random generator matrix G is a good code with high probability. Therefore, it is still possible that some kind of generating matrix of a good code has high rigidity. The following only says that simply seeking generating matrices of a good code is not enough; that is, some specific generating matrix of a good code can have low rigidity.

Theorem 7.1.2 (Dvir)**.** *For every sufficiently small constant $\varepsilon > 0$, all sufficiently large k, and every $d \in [k/4]$, there exist an $n = O(k)$ and a k-by-n matrix M such that every non-zero linear combination of the rows of M has Hamming weight $(1/2 \pm \varepsilon)n$ but the matrix M has rigidity at most $O(kn/d)$ with respect to rank $10d\log(k/d)$.*

Proof. Let $\varepsilon > 0$ be given. Let $k >> d$ be chosen later. Let $r = 10d\log(k/d)$, $m = k + r$, $c = O(\log(1/\varepsilon))$ and $n = \Omega(m/\varepsilon^2))$. Let H be a random m-by-n matrix with each entry set to 1 with probability $p = c/d$ independently.

Next, choose sufficiently large r such that for a random k-by-r matrix G' with entires uniformly chosen in $\{0,1\}$ and any linear combination of its rows yields a vector of weight at least d with high probability. Let $G = [I_k|G']$, where I_k is the identity matrix. Then the code generated by G has distance at least d. Denoting the top k rows of H by S and the remaining rows by H' we get $GH = IS + G'H' = S + G'H'$.

We first show that with high probability, the matrix GH generates a good code in which all non-zero codewords have weight $(1/2 \pm \varepsilon)n$. Let x be a nonzero vector in $\{0,1\}^k$. Observe that xG can be seen as a vector whose entries, except for the first k ones, are distributed uniformly and independently in $\{0,1\}^m$. Let $v = xGH$. By the definition of H, each entry in v is a sum of at least d independent random variables that are each nonzero with probability $p = c/d$. Hence, each entry in v is nonzero with probability $1/2 \pm \exp(-\Omega(c))$. A detailed Proof of this claim is presented in Section 0.1 in the appendix.

Now we show that GH does not have rigidity $2p \cdot kn = 2c \cdot kn/d$ with respect to rank $r = 10d\log(k/d)$. Observe that, with high probability, the matrix S has weight at most $2p \cdot kn = 2c \cdot kn/d$. On the other hand, G' is an k-by-r matrix, which implies that $G'H'$ has rank at most $r = 10d\log(k/d)$. Hence, $GH = S + G'H'$ does not have rigidity $2c \cdot kn/d$ with respect to rank $r = 10d\log(k/d)$. □

7.2 Preliminaries

Before getting into more non-rigidity results, we first introduce some results that are useful for the rest of the chapter.

7.2.1 Chernoff Bound

Theorem 7.2.1 (Chernoff Bound, [Che81], See [CCG+06], Theorem 2.4)**.** *Let $X_1, ..., X_n$ be independent random variables with $\mathbb{P}[X_i = 1] = p_i, \mathbb{P}[X_i = 0] = 1 - p_i$ Let $X = X_1 + ... + X_n$ and $\mu = E[X]$. Then for every real number $a > 0$,*

$$\mathbb{P}[X \leq \mu - \lambda] \leq e^{-\lambda^2/2\mu}, \mathbb{P}[X \geq \mu + \lambda] \leq e^{-\lambda^2/2(\mu+\lambda/3)}$$

Lemma 7.2.2. *Let $0 < \varepsilon < 1/2$. Then*

$$\sum_{i=0}^{(1/2-\varepsilon)n} \binom{n}{i} = \sum_{i=(1/2+\varepsilon)n}^{n} \binom{n}{i} \leq 2^{(1-\Omega(\varepsilon^2))n}$$

Proof. We present a proof via probabilistic method. Let $x \in \{0,1\}^n$ be picked uniformly random. For $i \in n$, let $Y_i = \mathbf{I}\{x_i = 1\}$ be the indicator random variable of whether x_i equals 1. Then Y_i is a Bernoulli random variable with $\mathbb{P}[Y_i = 1] = 1/2$. Let $Y = \sum_{i=0}^{n-1} Y_i$. Notice that Y is the number of 1's in a given vector x and

$$\sum_{i=0}^{(1/2-\varepsilon)n} \binom{n}{i} = 2^n \mathbb{P}[Y \leq (1/2 - \varepsilon)n].$$

By Chernoff Bound,

$$\sum_{i=0}^{(1/2-\varepsilon)n} \binom{n}{i} = 2^n \mathbb{P}[Y \leq (1/2 - \varepsilon)n] \leq 2^n e^{-\frac{(\varepsilon n)^2}{2(1/2)n}} = e^{-\varepsilon^2 n} \leq 2^{(1-\Omega(\varepsilon^2))n}$$

Via symmetry, we conclude the proof. □

7.2.2 Binary Entropy

Definition 7.2.1. Let $0 \leq \varepsilon \leq 1$. Then the binary entropy function H is defined as

$$H(\delta) = -\delta \log_2 \delta - (1 - \delta) \log_2 (1 - \delta)$$

Lemma 7.2.3. *For for sufficiently small $\delta, H(\delta) \leq \Theta(\delta \log_2(1/\delta))$.*

Proof. Notice that for all real $s > 0, s - 1 \geq \ln s$. Equivalently, $1 - 1/s \leq \ln s$. Letting $s = 1 - \delta$, we have $\ln(1 - \delta) \geq -\delta/(1 - \delta)$. Thus,

$$-\ln(1-\delta) \leq \frac{\delta}{1-\delta}$$

Now,

$$\begin{aligned} H_2(\delta) &= -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta) \\ &= \delta \log_2(1/\delta) + (1-\delta)(-\log_2(1-\delta)) \\ &= \delta \log_2(1/\delta) + \frac{(1-\delta)(-\ln(1-\delta))}{\ln 2} \\ &\leq \delta \log_2(1/\delta) + \frac{\delta}{\ln 2} = \Theta(\delta \log_2(1/\delta)) \end{aligned}$$

for sufficiently small δ. □

We state the following lemma without proof.

Lemma 7.2.4 (Volume of a Hamming Ball, See [GRS12] Proposition 3.3.1)**.** *For $0 \leq \delta \leq 1/2$,*

$$\binom{n}{\delta n} \leq \sum_{i=0}^{\delta n} \binom{n}{i} \leq 2^{nH(\delta)}$$

Combining the last two lemmas we immediately get the following corollary.

Corollary 7.2.5. *For for sufficiently small δ,*

$$\sum_{i=0}^{\delta n} \binom{n}{i} \leq 2^{\Theta(\delta \log_2(1/\delta))n}$$

7.3 Polynomial Methods

Proving non-rigidity of a target matrix M via polynomial methods consists of the following two main steps.

1. Construct a low rank matrix M' approximating the target matrix M. To do this, we first find a low degree polynomial p, whose corresponding matrix M' approximates the target matrix M. Then we observe that the low degree of p would imply the low rank of M'.

2. Each row of the matrix $M - M'$ will have a small number of non-zero entries.

We illustrate these two steps by proving the non-rigidity of Walsh-Hadamard Matrices [AW17] and matrices of the form $M_{x,y} = f(x + y)$ for arbitrary function $f : \{0,1\}^n \to \{0,1\}$ [CLP17]. Using more technically involved polynomials, Dvir and Liu show that generalised Hadamard matrices also fail to be rigid, but we omit this result in our discussion and refer the readers to their paper [DL20]. Before we get into the results, we introduce a lemma, which states that for a given polynomial p, the rank of the truth table matrix M, defined by

$$M_{x,y} := p(x, y)$$

for any $x, y \in \{0,1\}^n$, is at most the number of monomials of p.

Lemma 7.3.1 ([AW17])**.** *Let $p : \{0,1\}^{2n} \to \{0,1\}$ be a polynomial with m monomials and let M be a $2^n \times 2^n$ be the truth table matrix of p. Then the rank of M is at most m.*

Proof. Let $a_1, ..., a_m, b_1, ..., b_m : \{0,1\}^n \to \{0,1\}$ be monomials such that $p(x, y) = \sum_{i=1}^{m} a_i(x)b_i(y)$ is the monomial expansion of p. For each $1 \leq i \leq m$,

we can define a vector $\vec{a_i}$ by $\vec{a_i}_x = a_i(x)$ for all $x \in \{0,1\}^n$ and similarly a vector $\vec{b_i}$ by $\vec{b_i}_y = b_i(y)$ for all $y \in \{0,1\}^n$. Then we see that

$$M = \sum_{i=1}^{m} \vec{a_i} \otimes \vec{b_i}$$

where $\otimes$ denotes the outer product. Thus, $\text{rank}(M) \leq m$. □

7.3.1 Walsh-Hadamard Matrices

In this section, we let $H = ((-1)^{\langle x,y \rangle})_{x,y \in \{0,1\}^n}$ be the Walsh-Hadamard matrix. The following theorem implies that a Walsh-Hadamard matrix is not rigid enough.

Theorem 7.3.2 ([AW17]). *For every sufficiently small $\varepsilon > 0$, and for all n, we have*

$$\mathcal{R}_H(2^{(1-\Omega(\varepsilon^2))n}) \leq 2^{O(1+\varepsilon \log(1/\varepsilon))n}$$

We first introduce a few tools from polynomial methods for the first step. That is, we introduce a polynomial that approximates the Walsh-Hadamard matrices.

Proposition 7.3.3 (Low-degree polynomial approximating Walsh-Hadamard matrices, [AW17]). *For every $0 < \varepsilon < 1/2$, there is a multilinear polynomial $p : \{0,1\}^{2n} \to \{0,1\}$ with at most $2^{(1-\Omega(\varepsilon^2))n}$ monomials, such that for all $x, y \in \{0,1\}^n$, with $\langle x, y \rangle \in [2\varepsilon n, (1/2 + \varepsilon)n]$,*

$$p(x,y) = (-1)^{\langle x,y \rangle}$$

To to prove this proposition, we need an auxiliary lemma from [AW15], which we state without proof.

Lemma 7.3.4 (Polynomial Interpolation, [AW15]). *For any integers n, r, k with $n \geq r + k$ and any $c_1, \ldots, c_r \in \{0, 1\}$, there is a multivariate polynomial $q : \{0, 1\}^n \to \{0, 1\}$ of degree $r - 1$ with integer coefficients such that $q(z) = c_i$ for all $z \in \{0, 1\}^n$ with Hamming weight $|z| = k + i, 1 \leq i \leq r$.*

Proof of Proposition 7.3.3. Set

$$k = 2\varepsilon n - 1, r = (\frac{1}{2} - \varepsilon)n + 1, c_i = (-1)^{k+i}$$

By Lemma 7.3.4, we have a multivariate polynomial $q : \{0, 1\}^n \to \{0, 1\}$ with degree $\leq (1/2-\varepsilon)n$ such that for $|z| \in [2\varepsilon n, (1/2+\varepsilon)n], q(z) = c_i = (-1)^{k+i} = (-1)^{|z|}$. Define the multilinear polynomial p by setting $p(x, y) = q(\langle x, y\rangle)$. Notice that as a result of the inner product $\langle x, y\rangle$, each monomial of p contains x_i if and only if it also contains y_i. The number of monomials in p is thus the same as the number of monomials of q, who has degree $\leq (1/2 - \varepsilon)n$. By Lemma 7.2.2, the number of monomials in p is thus upper bounded by

$$\sum_{i=0}^{(1/2-\varepsilon)n} \binom{n}{i} \leq 2^{(1-\Omega(\varepsilon^2))n}$$

□

Simply using the truth table matrix of the polynomial p in the last proposition isn't quite enough. We introduce a few more tools to improve our approximation without increasing the rank too much. Since changing one column or one row can only change the rank of a matrix by 1, the following lemma is immediate.

Lemma 7.3.5 ([AW17]). *Let M' be a matrix of rank r. Let M be a matrix which is equal to M' except in at most k columns and l rows. Then the rank of M is at most $r + k + l$.*

Note that the number of vectors $v \in \{0,1\}^n$ with $|v| \notin [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$ is at most

$$\sum_{i=0}^{(1/2-\varepsilon)n} \binom{n}{i} + \sum_{i=(1/2+\varepsilon)n}^{n} \binom{n}{i} \leq 2 \cdot 2^{n-\Omega(\varepsilon^2 n)}$$

using Lemma 7.2.2. Applying Lemma 7.3.5 with $k = l = 2 \cdot 2^{n-\Omega(\varepsilon^2 n)}$, we get the following corollary.

Corollary 7.3.6 ([AW17])**.** *Let $\varepsilon \in (0, 1/100)$. Let T be any $2^n \times 2^n$ matrix and let M be a $2^n \times 2^n$ matrix of rank $r \geq 2\varepsilon n$, indexed by n-bit vectors. There is a $2^n \times 2^n$ matrix M' of rank at most $r + 4 \cdot 2^{n-\Omega(\varepsilon^2 n)}$ such that $M'(x,y) = T(x,y)$ on all $x, y \in \{0,1\}^n$ where at least one of the following holds:*

- $|x| \notin [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$,
- $|y| \notin [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$, *or*
- $M(x,y) = T(x,y)$.

Combining all the tools together, we are now ready to wrap up the first step, as shown in the next corollary.

Corollary 7.3.7 ([AW17])**.** *For every sufficiently small $\varepsilon > 0$, there is a matrix M' with rank at most $2^{n-\Omega(\varepsilon^2 n)}$, such that*

$$M'_{x,y} = (-1)^{\langle x,y \rangle}$$

for all $x, y \in \{0,1\}^n$ where at least one of the following holds:

- $|x| \notin [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$,

- $|y| \notin [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$, *or*
- $\langle x, y\rangle \in [2\varepsilon n, (1/2+\varepsilon)n]$.

Proof. By Proposition 7.3.3, we obtain a multilinear polynomial $p : \{0,1\}^{2n} \to \{0,1\}$ with at most $2^{n-\Omega(\varepsilon^2 n)}$ monomials, such that for all $x, y \in \{0,1\}^n$, with $\langle x, y\rangle \in [2\varepsilon n, (1/2+\varepsilon)n]$,

$$p(x,y) = (-1)^{\langle x,y\rangle}$$

Let M be the truth table matrix of p. Then by Lemma 7.3.1, we have that $\text{rank}(M) \leq 2^{n-\Omega(\varepsilon^2 n)}$. Now, apply Corollary 7.3.6, with $T = H_n$, we obtain the desired matrix M' with rank at most $2^{n-\Omega(\varepsilon^2 n)} + 4 \cdot 2^{n-\Omega(\varepsilon^2 n)} = 2^{n-\Omega(\varepsilon^2 n)}$. □

For the second step, we aim to show that each row of the matrix $M - M'$ will have a small number of non-zero entries, which is presented in the next lemma.

Lemma 7.3.8 ([AW17])**.** *Let $\varepsilon \in (0, 1/100)$. For every vector $x \in \{0,1\}^n$ with $|x| \in [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$, there are at most $2^{O(\varepsilon \log(1/\varepsilon)n)}$ such $y \in \{0,1\}^n$ satisfying*

- $|y| \in [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$, *and*
- $\langle x, y\rangle \leq 2\varepsilon n$

Proof. Let $x \in \{0,1\}^n$ by a vector with $|x| \in [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$. Fix $k \in [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$ and $s \leq 2\varepsilon n$. We count the number of $y \in \{0,1\}^n$ with $|y| = k$ and $\langle x, y\rangle = s$. Notice that a vector y satisfies these two properties if and only if

- there is some set $S \subseteq [n]$ with $|S| = s$ and $x_i = y_i = 1$ for all $i \in S$.
- there is some set $T \subseteq ([n] \setminus S)$ with $x_j = 0, y_j = 1$ for all $j \in T$.

In other words, with fixed x, k, s, we have at most

$$\binom{|x|}{s}\binom{n-|x|}{k-s}$$

choices of y. Now, counting all possible choices of k, s, we have the number of y meeting the statement in the lemma is

$$\begin{aligned}\sum_{k=(1/2-\varepsilon)n}^{(1/2+\varepsilon)n}\sum_{s=0}^{2\varepsilon n}\binom{|x|}{s}\binom{n-|x|}{k-s} &\leq \sum_{k=(1/2-\varepsilon)n}^{(1/2+\varepsilon)n}\sum_{s=0}^{2\varepsilon n}\binom{(1/2+\varepsilon)n}{s}\binom{(1/2+\varepsilon)n}{k-s} \\ &\leq 4\varepsilon^2 n^2\binom{(1/2+\varepsilon)n}{2\varepsilon n}\binom{(1/2+\varepsilon)n}{(1/2-3\varepsilon)n} \\ &\leq 2^{O(nH(\frac{4\varepsilon}{1/2+\varepsilon}))} = 2^{O(\varepsilon\log(1/\varepsilon)n)}\end{aligned}$$

by applying Lemma 7.2.3 and Lemma 7.2.4. □

Notice that M' and H_n can only differ in indices (x, y) where $|x|, |y| \in [(1/2-\varepsilon)n, (1/2+\varepsilon)n]$ and $\langle x, y\rangle \leq 2\varepsilon n$. So for each fixed x, the number of y's that would cause M' and H_n to differ is at most $2^{O(\varepsilon\log(1/\varepsilon)n)}$ as shown in the lemma above. Therefore, M' and H_n differ in at most $2^n \cdot 2^{O(\varepsilon\log(1/\varepsilon)n)} = 2^{O(1+\varepsilon\log(1/\varepsilon))n}$ entries. In other words, we reach the following conclusion

$$\mathcal{R}_{H_n}(2^{(1-\Omega(\varepsilon^2))n}) \leq 2^{O(1+\varepsilon\log(1/\varepsilon))n}$$

which complete the proof for Theorem 7.3.2.

7.3.2 Matrices $M_{x,y} = f(x+y)$

In this subsection, we show the following result following the two steps outlined at the beginning of the section.

Theorem 7.3.9 ([DE19]). *Let $f : \{0,1\}^n \to \{0,1\}$ and set matrix $M \in \{0,1\}^{2^n \times 2^n}$ with $M_{x,y} = f(x+y)$ for $x, y \in \{0,1\}^n$. Then for all any $\varepsilon > 0$, there exists an $\varepsilon' > 0$ such that for sufficiently large n,*

$$\mathcal{R}_M(2^{(1-\Omega(\varepsilon^2))n}) \leq 2^{O(1+\varepsilon \log(1/\varepsilon))n}$$

where $N = 2^n$.

We first introduce some definitions. Let $\mathcal{F}(n)$ be the family of functions $f : \{0,1\}^n \to \{0,1\}$. Let $\mathcal{M}_d(n)$ be the family of monomials of degree at most d, i.e.

$$\mathcal{M}_d(n) = \{x_1^{a_1} \dots x_n^{a_n} | a_i \in \{0,1\}, \sum_{i=1}^{n} a_i \leq d\}$$

and $m_d(n) = |\mathcal{M}_d(n)|$. The following lemma, which we state without proof, says that the rank of a matrix $M_{x,y} = p(x+y)$ is low if p does not have a degree that is too high.

Lemma 7.3.10 (Croot-Lev-Pach Lemma, [CLP17]). *Let $p : \{0,1\}^n \to \{0,1\}$ be a polynomial of degree at most d and set matrix $M \in \{0,1\}^{2^n \times 2^n}$ with $M_{x,y} = p(x+y)$ for $x, y \in \{0,1\}^n$. Then*

$$\text{rank}(M) \leq 2 \cdot m_{\lfloor d/2 \rfloor}(n)$$

The next lemma says that any function can be approximated by a polynomial with sufficiently high degree.

Lemma 7.3.11 (Polynomials Approximates any Function, [DE19]). *Let $f : \{0,1\}^n \to \{0,1\} \in \mathcal{F}(n)$. Then for all $d \leq n$, there exists a polynomial $p : \{0,1\}^n \to \{0,1\}$ of degree at most d satisfying*

$$|\{x \in \mathbb{F}_2^n | f(x) \neq p(x)\}| \leq 2^n - m_d(n)$$

Proof. We view $\mathcal{F}(n)$ as a linear space V. Naturally, the space of polynomials of degree at most d, which we denote as W, is a linear subspace of V. Notice that for any vector v in V, we can decompose it as

$$v = w + u$$

a vector $w \in W$ and $u \in V \setminus W$ and u has weight at most $\dim(V) - \dim(W)$. To translate back the context of functions and polynomials, we have that

$$\begin{aligned} |\{x \in \mathbb{F}_2^n | f(x) \neq p(x)\}| &\leq \dim(V) - \dim(W) \\ &= \dim(V) - \dim(\mathcal{M}_d(n)) = 2^n - m_d(n) \end{aligned}$$

□

We are now ready to prove the Theorem 7.3.9.

Proof of Theorem 7.3.9. Let $d = (1-\varepsilon)n$ and $f : \{0,1\}^n \to \{0,1\}$ be given. Let M denote the $2^n \times 2^n$ matrix with $M_{x,y} = f(x+y)$. Fix $x \in \{0,1\}^n$. Using Lemma 7.3.11, we can find a polynomial p of degree at most d with

$$\begin{aligned} |\{y \in \mathbb{F}_2^n | f(x+y) \neq p(x+y)\}| \leq 2^n - m_d(n) = 2^n - (2^n - m_{\varepsilon n}(n)) = m_{\varepsilon n}(n) \\ = \sum_{i=0}^{\varepsilon n} \binom{n}{i} \leq 2^{\Theta(\varepsilon \log_2(1/\varepsilon))n} \end{aligned}$$

by Corollary 7.2.5. Let L denote the $2^n \times 2^n$ matrix with $L_{x,y} = p(x+y)$. Then M and L differ in at most $2^{\Theta(\varepsilon \log_2(1/\varepsilon))n}$ entries for a fixed row indexed by x. Thus, they differ in at most $2^n \cdot 2^{\Theta(\varepsilon \log_2(1/\varepsilon))n} = 2^{\Theta(1+\varepsilon \log_2(1/\varepsilon))n}$ entries in total. Using the Croot-Lev-Pach Lemma 7.3.10 and Lemma 7.2.2, we have

$$\operatorname{rank}(L) \leq m_{\lfloor d/2 \rfloor}(n) = m_{\lfloor (1-\varepsilon)n/2 \rfloor}(n) = \sum_{i=0}^{(1/2-\varepsilon)n} \binom{n}{i} \leq 2^{(1-\Omega(\varepsilon^2))n}$$

We thus conclude the proof. □

7.4 Kronecker Products

A very recent line of work has found that if a matrix M can be written as Kronecker products of many matrices, then M is not rigid [DL20, Alm21, Kiv21]. The key definition for these work is the row-column rigidity.

Definition 7.4.1 (Row-Column Rigidity, [DL20], [Alm21])**.** For a matrix $A \in \mathbb{F}^{n\times n}$ and a target rank $0 \leq r \leq n$, we define the **row-column rigidity** $\mathcal{R}^{rc}_A(r)$ as the minimal number t such that there exists a matrix $B \in \mathbb{F}^{n\times n}$ with at most t non-zero entries at each row and column, and $\text{rank}(A+B) \leq r$.

It is easy to see that if a matrix A is not row-column rigidity, then simply by changing sufficient number of entries in each row, A won't be rigid in the general notion. Notice that the row-column rigidity of a permutation matrix or a diagonal matrix is 1. The following lemma says that if two matrices A, B are not row-column rigid, then the product $A \cdot B$ would also fail to be row-column rigid. This is the key observation that allows us to study the non-rigidity of matrices via matrix factorization.

Lemma 7.4.1 ([DL20])**.** *For matrices $A, B \in \mathbb{F}^{n\times n}$,*

$$\mathcal{R}^{rc}_{A\cdot B}(r+s) \leq \mathcal{R}^{rc}_A(r) \cdot \mathcal{R}^{rc}_B(s)$$

Proof. We can decompose A, B as sums of low rank matrices and sparse matrices. That is, we write $A = L_A + S_A, B = L_B + S_B$ where $\text{rank}(L_A) \leq r, \text{rank}(L_B) \leq s$ and S_A has at most $\mathcal{R}^{rc}_A(r)$ at each row and column and S_B has at most $\mathcal{R}^{rc}_B(s)$ at each row and column. Then,

$$AB = (L_A + S_A)(L_B + S_B) = L_A(L_B + S_B) + S_A L_B + S_A S_B$$

where $\text{rank}(L_A(L_B + S_B)) \leq \text{rank}(L_A) \leq r, \text{rank}(S_A L_B) \leq \text{rank}(L_B) \leq s$. It remains to show that $S_A S_B$ is sparse. For a fixed row i of $S_A S_B$, since there are at most $\mathcal{R}_A^{rc}(r)$ nonzero entries of the ith row of S_A and at most $\mathcal{R}_B^{rc}(s)$ nonzero entries of each row of S_B, $S_A S_B$ will have at most $\mathcal{R}_A^{rc}(r) \cdot \mathcal{R}_B^{rc}(s)$ entries in the ith row. The rest of the proof follows a similar argument. □

The main goal of this section is to prove the following.

Theorem 7.4.2 ([Alm21, Kiv21]). *For matrices $M_1, \ldots, M_n \in \{0,1\}^{2\times 2}$, and sufficiently small $\varepsilon > 0$, the Kronecker product $M := \bigotimes_{\ell=1}^{n} M_\ell, M_\ell \in \mathbb{F}^{2\times 2}$ has*

$$\mathcal{R}_M^{rc}(2^{(1-\Omega(\varepsilon^2))n}) \leq 2^{\Theta(2\varepsilon \log_2(1/(2\varepsilon)))n}$$

and

$$\mathcal{R}_M(2^{(1-\Omega(\varepsilon^2))n}) \leq 2^{(1+\Theta(2\varepsilon \log_2(1/(2\varepsilon))))n}$$

We remark that Alman and Kivva used different matrix factorizations to prove the previous theorem. We give yet another factorization in the proof presented below.

7.4.1 LPU Factorization

In this section we introduce some basic facts on LPU factorization and Kronecker products.

Theorem 7.4.3 (LPU factorization, See [HJ12] Theorem 3.5.11). *Let M be a square matrix. Then there exists a weighted permutation matrix P, a lower triangular matrix L and an upper triangular matrix U such that $M = LPU$.*

Lemma 7.4.4 (Extended Mixed-product Property).

$$(L_1 \times P_1 \times U_1) \otimes (L_2 \times P_2 \times U_2) = (L_1 \otimes L_2)(P_1 \otimes P_2)(U_1 \otimes U_2)$$

Proof. We use the mixed-product property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$.

$$\begin{aligned}(L_1 \times P_1 \times U_1) \otimes (L_2 \times P_2 \times U_2) &= ((L_1 \times (P_1 \times U_1)) \otimes (L_2 \times (P_2 \times U_2)) \\ &= (L_1 \otimes L_2)((P_1 \times U_1) \otimes (P_2 \times U_2)) \\ &= (L_1 \otimes L_2)(P_1 \otimes P_2)(U_1 \otimes U_2)\end{aligned}$$

□

Using Extended Mixed-product Property, we immediately obtain the following:

Theorem 7.4.5. *Let $M = \bigotimes_{i=1}^{n} M_i$ for any matrices $M_1 = L_1 \times P_1 \times U_1, \ldots, M_n = L_n \times P_n \times U_n \in \mathbb{F}^{q \times q}$. Then*

$$M = L \times P \times U$$

where $L = \bigotimes_{i=1}^{n} L_i, P = \bigotimes_{i=1}^{n} P_i, U = \bigotimes_{i=1}^{n} U_i$.

7.4.2 Non-Rigidity of U

For $i \in \{1, ..., n\}$, let $U_i \in \mathbb{F}^{2\times 2}$ be upper triangular matrices. Let $U = \bigotimes_{i=0}^{n-1} U_i$. Notice that U is also upper triangular. In this section, we show that U is not rigid.

Theorem 7.4.6. *For $i \in \{1, ..., n\}$, let $U_i \in \mathbb{F}^{2\times 2}$ be upper triangular matrices. Let $U = \bigotimes_{i=1}^{n} U_i$. Then for sufficiently small $\varepsilon > 0$ and sufficiently large n, we have*

$$\mathcal{R}_U^{rc}(2^{(1-\Omega(\varepsilon^2))n}) \leq 2^{\Theta(2\varepsilon \log_2(1/(2\varepsilon)))n}$$

Let $V^{(1)} \in \mathbb{F}^{2\times 2}$ be an all one upper triangular matrix:

$$V^{(1)} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

For any integer $n > 1$, we let $V^{(n)} = V^{(1)} \bigotimes V^{(n-1)}$. The following lemma reveals exactly when $V^{(n)}_{x,y} = 1$ for $x, y \in \{0,1\}^n$.

Lemma 7.4.7. *For $x, y \in \{0,1\}^n$, we have $V^{(n)}_{x,y} = 1$ if and only if for all $i \in \{0, ..., n-1\}, x_i \leq y_i$.*

Proof. We prove this by induction. Since $V^{(1)}$ is the all one upper triangular matrix, the base case is true. For for any $x \in \{0,1\}^n$, let $x' = x_1 x_2 ... x_{n-1}$ denote the last $n-1$ bits of x and $y' = y_1 y_2 ... y_{n-1}$ denote the last $n-1$ bits of y. Inductively, since $V^{(n)} = V^{(1)} \bigotimes V^{(n-1)}$, we have that

$$V^{(n)}_{x,y} = V^{(1)}_{x_0,y_0} \cdot V^{(n-1)}_{x',y'}$$

which equals 1 if and only if $V^{(1)}_{x_0,y_0} = V^{(n-1)}_{x',y'} = 1$. By inductive hypothesis, $x_0 \leq y_0$ and for all $i, x'_i \leq y'_i$. □

We are now ready to prove Theorem 7.4.6.

Proof of Theorem 7.4.6. We first prove the non-rigidity of $V^{(n)}$. There are two main parts of this proof. Firstly, let W be a $2^n \times 2^n$ matrix with $W_{x,y} = 1$ if and only if one of the following holds:

- $|x| \notin [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n]$, or
- $|y| \notin [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n]$.

As we have seen before in the proof of Corollary 7.3.7, we have $\text{rank}(W) \leq 2^{(1-\Omega(\varepsilon^2))n}$. Fix a row $x \in [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n]$. We now bound the number of $y \in [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n]$ such that $V^{(n)}_{x,y} = 1$. By Lemma 7.4.7, $V^{(n)}_{x,y} = 1$ if and only if $x_i \leq y_i$ for all $i \in \{0, ..., n-1\}$. So for the indices i such that

$x_i = 1$, we must have $y_i = 1$. Notice that $|y| \leq (1/2+\varepsilon)n$ and $|x| \geq (1/2-\varepsilon)n$. This means that the number of possible y's given a fixed x is at most

$$\sum_{i=0}^{2\varepsilon n} \binom{n}{i} \leq 2^{\Theta(2\varepsilon \log_2(1/(2\varepsilon)))n}$$

by Corollary 7.2.5. As a last step, notice that our argument only depends on the upper-triangular property of $V^{(1)}$, but not the actual values in the upper-triangular part of $V^{(1)}$. Thus, the same argument applies to U. □

Combining Theorem 7.4.5, Lemma 7.4.1 and Theorem 7.4.6, we reach Theorem 7.4.2.

With more involved techniques, Dvir and Liu observe that infinitely many matrices in a family of Fourier transform matrices can be written as Kronecker products of generalised Hadamard matrices, which leads to the discovery that many Fourier transform matrices are not rigid. With the observation that circulant matrices can be factorized into products of Fourier transform matrices and diagonal matrices, Dvir and Liu show that circulant matrices fail to be rigid too. For more discussion, we refer the reader to [DL20, BK21].

Appendix

0.1 How to get an almost fair coin?

Observe a sequence of bits $x = x_1x_2...x_d$. Each bit x_i is 1 with probability $c/d, c \leq 0.5d$ and is 0 otherwise. Then the parity of this sequence has only exponential bias. Concretely,

$$\mathbb{P}[\chi(x) = 1] = \frac{1}{2} \pm \exp(\Omega(c))$$

This claim turns out extremely easy to show with a basic introduction of *finite, irreducible, ergodic Markov chains*.

0.1.1 Crash Course Markov Chains

Let S be a finite state space with positive integers and T be a subset of $[0, \infty)$. A **stochastic process** is a collection of random variables $\{X_n : n \in T\}$ which take values from S. Let $\{X_n, n = 0, 1, 2, ...\}$ be a stochastic process that takes on values from S. If $X_n = i$, we say that the process is in state i at time n. Suppose that

$$\mathbb{P}[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0] = \mathbb{P}[X_{n+1} = j | X_n = i] = p_{i,j}$$

for all states $i_0, i_1, ..., i, j \in S$ and all $n \geq 0$. Such a stochastic process is called a **finite Markov chain**. The **transition matrix P**$= (p_{i,j})$ is a $|S| \times |S|$ matrix of transition probabilities

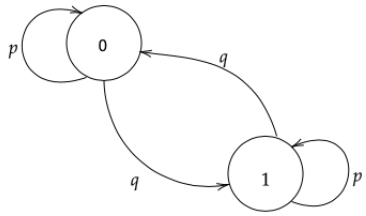
$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \\ p_{i0} & p_{i1} & p_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

In our example, we have two states, 0 and 1. Let $p = 1 - c/d$ and $q = c/d$. If we are at state 0 during time n, i.e. $\chi(x_1x_2...x_n) = 0$, then we will

remain in state 0 with probability p and transition to state 1 with probability 1. Thus, we obtain our transition matrix

$$\mathbf{Q} = \begin{bmatrix} p & q \\ q & p \end{bmatrix}$$

We say that a finite Markov chain is **irreducible** if and only if its graph representation is a strongly connected graph. Apparently, Q is irreducible, whose graph representation is shown below.



0.1.1.1 Periodicity

The **period** of a state i is the largest common devisor of the set $\{n : p_{i,i}(n) > 0, n \geq 1\}$. We write $d(i) = \gcd\{n : p_{i,i}(n) > 0, n \geq 1\}$. We call state i **periodic** if $d(i) > 1$ and **aperiodic** if $d(i) = 1$. We notice that in $\mathbf{Q}$, $p_{i,i} > 0$ for $i = 0, 1$ in our example. This means that in each step, we can get back to the last state with positive probability. In this case, our Markov chain is *aperiodic*.

0.1.1.2 Recurrence

Let $f_{i,i}(n) = \mathbb{P}[X_n = i, X_k \neq i \text{ for } 0 < k < n | X_0 = i]$ and let $f_{i,i}$ be the probability that given $X_0 = i$, $X_n = i$ for some $n > 0$. That is,

$$f_{i,i} = \sum_{n=1}^{\infty} f_{i,i}(n)$$

State i is said to be **recurrent** if $f_{i,i} = 1$; on the other hand, we say that state i is **transient** if $f_{i,i} < 1$. For a recurrent state, the **mean recurrence time** μ_i is define as

$$\mu_i = \sum_n n f_{i,i}(n)$$

A recurrent state i is called **positive recurrent** if $\mu_i < \infty$. Notice that a recurrent state i can have infinite mean recurrence time, in this case, we call such a state **null recurrent**. A state is said to be **ergodic** if it is positive recurrent and aperiodic. A Markov chain is ergodic if all its states are ergodic.

In our example, if we start from state 0, then in the next coin flip, we either get 0 with probability p, or get 1 with probability q. If we get a 1, we must wait until the next time to flip a 1 to get back to 0. That is,

$$\mu_0 = \sum_n n f_{0,0}(n) = p + q \sum_{k=0}^{\infty} (k+1) p^k q$$

Similarly, we obtain

$$\mu_1 = \sum_n n f_{1,1}(n) = p + q \sum_{k=0}^{\infty} (k+1) p^k q$$

Notice that $\sum_{k=0}^{\infty} k p^k q$ is the expectation of a geometric distribution with probability q and $\sum_{k=0}^{\infty} p^k q$ is the sum of the probability mass of the same geometric distribution. Thus, as $\sum_{k=0}^{\infty}(k+1)p^k q = \sum_{k=0}^{\infty} k p^k q + \sum_{k=0}^{\infty} p^k q = 1/q + 1$,

$$\mu_0 = \mu_1 = p + q \cdot (1/q + 1) = 2$$

Hence, both states of our Markov chain are positive recurrent. In fact, this Markov chain is *ergodic*.

0.1.1.3 Stationary distribution

A stationary distribution of a Markov chain is a probability distribution $\bar{\pi}$ such that $\bar{\pi}\mathbf{P} = \bar{\pi}$. In fact, any finite, irreducible, and ergodic Markov chain has a unique **stationary distribution** $\bar{\pi} = (\pi_0, \pi_1, ..., \pi_n)$, with $\pi_0 + \pi_1 + ... + \pi_n = 1$ (Theorem 7.7 in [MU17]). Hence, for our example,

$$\bar{\pi}\mathbf{Q} = (\pi_0, \pi_1) \begin{bmatrix} p & q \\ q & p \end{bmatrix} = \bar{\pi}, \pi_0 + \pi_1 = 1$$

gives us $\pi_0 = \pi_1 = 1/2$. This tells us that *if you flip a biased coin for an infinite number of times, you can get a fair coin.* In other words,

$$\mathbb{P}[\chi(x_1 x_2 ...) = 1] = 1/2$$

How fast does the probability converge to $1/2$ with respect to the number of coin tosses? In fact, this convergence is geometric.

Theorem 0.1.1 (Theorem 12.5 in [MU17])**.** *Let $\bar{\pi}_i^n$ represent the distribution of the state of the chain starting at state i after n steps. Let $\mathbf{P}$ be the transition matrix for a finite, irreducible, aperiodic Markov chain. Let m_j be the smallest entry in the jth column of the matrix, and let $m = \sum_j m_j$. Then for all i and n,*

$$||\bar{\pi}_i^n - \bar{\pi}|| \leq (1 - m)^n.$$

Now we are ready to conclude our motivating example. Since $c \leq 0.5d$, we have $q < p$. Thus, if we take d coin tosses,

$$||\bar{\pi}_0^n - \bar{\pi}|| \leq (1 - 2q)^d = (1 - 2 \cdot \frac{c}{d})^d \leq \exp(-2c)$$

Bibliography

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on information theory*, 38(2):509–516, 1992.

[AC88] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.

[AC15] Noga Alon and Gil Cohen. On rigid matrices and u-polynomials. *computational complexity*, 24(4):851–879, 2015.

[AC19] Josh Alman and Lijie Chen. Efficient construction of rigid matrices using an np oracle. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1034–1055, Nov 2019.

[Alm21] Josh Alman. Kronecker products, low-depth circuits, and matrix rigidity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 772–785, 2021.

[APY09] Noga Alon, Rina Panigrahy, and Sergey Yekhanin. Deterministic approximation algorithms for the nearest codeword problem. In Irit Dinur, Klaus Jansen, Joseph Naor, and José Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Al-*

gorithms and Techniques, pages 339–351, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[AR94] Noga Alon and Yuval Roichman. Random cayley graphs and expanders. *Random Structures & Algorithms*, 5(2):271–284, 1994.

[AW15] Josh Alman and Ryan Williams. Probabilistic polynomials and hamming nearest neighbors. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 136–150. IEEE, 2015.

[AW17] Josh Alman and Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 641–652, New York, NY, USA, 2017. Association for Computing Machinery.

[BCS97] Peter Bürgisser, Michael Clausen, and Mohammad Amin Shokrollahi. *Linear Complexity*, pages 305–349. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

[BHPT20] Amey Bhangale, Prahladh Harsha, Orr Paradise, and Avishay Tal. Rigid matrices from rectangular pcps or: Hard claims have complex proofs. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 858–869. IEEE, 2020.

[BK21] László Babai and Bohdan Kivva. Matrix rigidity depends on the target field. In *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[CCG+06] Fan Chung, Fan RK Chung, Fan Chung Graham, Linyuan Lu, Kian Fan Chung, et al. *Complex graphs and networks*. Number 107. American Mathematical Soc., 2006.

[Che81] Herman Chernoff. A note on an inequality involving the normal distribution. *The Annals of Probability*, pages 533–535, 1981.

[Che05] Mahdi Cheraghchi. On matrix rigidity and the complexity of linear forms. 2005.

[CLP17] Ernie Croot, Vsevolod F Lev, and Péter Pál Pach. Progression-free sets in are exponentially small. *Annals of Mathematics*, pages 331–337, 2017.

[DE19] Zeev Dvir and Benjamin L. Edelman. Matrix rigidity and the croot-lev-pach lemma. *Theory of Computing*, 15(8):1–7, 2019.

[DGW19] Zeev Dvir, Alexander Golovnev, and Omri Weinstein. Static data structure lower bounds imply rigidity. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pages 967–978, New York, NY, USA, 2019. Association for Computing Machinery.

[DL20] Zeev Dvir and Allen Liu. Fourier and circulant matrices are not rigid. *Theory OF Computing*, 16(20):1–48, 2020.

[DS07] Zeev Dvir and Amir Shpilka. Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits. *SIAM Journal on Computing*, 36(5):1404–1434, 2007.

[Dvi11] Zeev Dvir. On matrix rigidity and locally self-correctable codes. *computational complexity*, 20(2):367–388, 2011.

[dW06] Ronald de Wolf. Lower bounds on matrix rigidity via a quantum argument. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming*, pages 62–71, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[EGH14] Yuli Eidelman, Israel Gohberg, and Iulian Haimovici. *Separable type representations of matrices and fast algorithms*. Springer, 2014.

[Fri93] Joel Friedman. A note on matrix rigidity. *Combinatorica*, 13(2):235–239, 1993.

[GKST02] Oded Goldreich, Howard Karloff, Leonard J Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings 17th IEEE Annual Conference on Computational Complexity*, pages 175–183. IEEE, 2002.

[Gol20] Alexander Golovnev. Lecture notes on matrix rigidity. *https://golovnev.org/rigidity/*, 2020.

[GRS12] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. *Draft available at http://www. cse. buffalo. edu/ atri/courses/coding-theory/book*, 2012.

[GT18] Oded Goldreich and Avishay Tal. Matrix rigidity of random toeplitz matrices. *computational complexity*, 27(2):305–350, 2018.

[GVL12] Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU press, 2012.

[HJ12] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[Juk01] Stasys Jukna. *Orthogonality and Rank Arguments*, pages 191–204. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[Juk11] Stasys Jukna. *Extremal Combinatorics: With Applications in Computer Science*, pages 197–212. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[Kiv21] Bohdan Kivva. Improved upper bounds for the rigidity of kronecker products. In *46th International Symposium on Mathematical Foundations of Computer Science*, 2021.

[KLPS14] Abhinav Kumar, Satyanarayana V Lokam, Vijay M Patankar, and MN Sarma. Using elimination theory to construct rigid matrices. *computational complexity*, 23(4):531–563, 2014.

[KR98] B. S. Kashin and A. A. Razborov. Improved lower bounds on the rigidity of hadamard matrices. *Mathematical Notes*, 63(4):471–475, 1998.

[Lok95] Satyanarayana V. Lokam. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 6–15, Oct 1995.

[Lok00] Satyanarayana V. Lokam. On the rigidity of vandermonde matrices. *Theoretical Computer Science*, 237(1):477 – 483, 2000.

[Lok06] Satyanarayana V. Lokam. Quadratic lower bounds on matrix rigidity. In Jin-Yi Cai, S. Barry Cooper, and Angsheng Li, editors, *Theory and Applications of Models of Computation*, pages 295–307, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[Lok09] Satyanarayana V. Lokam. Complexity lower bounds using linear algebra. *Found. Trends Theor. Comput. Sci.*, 4(1–2):1–155, January 2009.

[Mid05] Gatis Midrijanis. Three lines proof of the lower bound for the matrix rigidity, 2005.

[Mor96] Patrick Morandi. *Transcendental Extensions*, pages 173–224. Springer New York, New York, NY, 1996.

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*, volume 16. Elsevier, 1977.

[MU17] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[PP06] R. Paturi and P. Pudlák. Circuit lower bounds and linear codes. *Journal of Mathematical Sciences*, 134(5):2425–2434, 2006.

[Pud94] P. Pudlák. Communication in bounded depth circuits. *Combinatorica*, 14(2):203–216, 1994.

[RR20] Sivaramakrishnan Natarajan Ramamoorthy and Cyrus Rashtchian. Equivalence of Systematic Linear Data Structures and Matrix Rigidity. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 35:1–35:20, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[SSS97] M.A. Shokrollahi, D.A. Spielman, and V. Stemann. A remark on matrix rigidity. *Information Processing Letters*, 64(6):283 – 285, 1997.

[SY11] S. Saraf and S. Yekhanin. Noisy interpolation of sparse polynomials, and applications. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 86–92, 2011.

[TVZ82] M. A. Tsfasman, S. G. Vladutx, and Th. Zink. Modular curves, shimura curves, and goppa codes, better than varshamov-gilbert bound. *Mathematische Nachrichten*, 109(1):21–28, 1982.

[Val77] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In Jozef Gruska, editor, *Mathematical Foundations of Computer Science 1977*, pages 162–176, Berlin, Heidelberg, 1977. Springer Berlin Heidelberg.

Vita

Yangxinyu Xie was born in Shaoyang, Hunan, China on 15 February 2000. He received the Bachelor of Science degree in Computer Science and Mathematics from the University of Texas at Austin in May, 2021. He continued his study in computer science for a master's degree at University of Texas at Austin in August, 2021.

Permanent address: 2317 Speedway, Stop D9500
Austin, Texas 78712

This thesis was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.