
@cref

Chapter 1 Gibbs Point Process

1.1 Independent sets in triangle-free graphs

Given a graph G , the *independence number* of graph G , denoted by $\alpha(G)$, is the size of the largest independent set, in which no two vertices are adjacent. Determining the independence number of a graph is one of the most pervasive and fundamental problems in graph theory. There is a huge amount of studies about bounding independence number from below and above in different problems. Besides, it is closely related to the Ramsey problems and so on.


Let's begin with a simple observation.

Observation 1.1

Let $d \in \mathbb{N}$ and G be an n -vertex graph with $\Delta(G) \leq d$. Then

$$\alpha(G) \geq \frac{n}{d+1}. \quad (1.1)$$

Proof Arbitrarily pick v_1 , then pick $v_2 \in G \setminus N(v_1)$ with $v_2 \neq v_1, \dots, v_i \in G \setminus \bigcup_{j=1}^{i-1} N(v_j)$, $v_i \neq v_j$, $j = 1, 2, \dots, i-1$. Finally, we can generate an independent set $I = \{v_1, v_2, \dots\}$, as $\Delta(G) \leq d$, v_i has at most d neighbors, then the size of I is at least $\frac{n}{d+1}$.


 **Exercise 1.1** Prove that for any n -vertex graph with $d(G) \leq d$, $\alpha(G) \geq \frac{n}{d+1}$.

Remark The lower bound $\frac{n}{d+1}$ in both Observation 1.1 and Exercise 4.1 is an optimal result. We can consider G as the vertex-disjoint union of cliques of size $d+1$.

Problem 1.1(Extremal problem) Given a collection F of all n -vertex graphs with average degree d , what is the minimum independence number of a graph G in F ? In this case, we can get $\min \alpha(G) = \frac{n}{d+1}$, $G \in F$.

Problem 1.2(Meta problem) It is natural to ask what if we forbid graphs that look like extremal structures, can we improve the bound? Since disjoint union of cliques have lots of triangles, whether the bound on $\alpha(G)$ can be improved if we add a triangle-free condition?

Ajtai-Komlós-Szemenédi [3] proved that any triangle-free graph G on n vertices with average degree d has an independent set of size at least $0.01 \frac{\log d}{d} n$. It improves the bound by a factor that is logarithmic in d . Later on, Shearer [22] improved the constant to 1, showing that such a graph has an independent set of size at least $f(d) \cdot n$ where $f(d) = \frac{d \log \frac{d-d+1}{d-1}}{(d-1)^2} = (1 + o(1)) \frac{\log d}{d}$, $f(0) = 1$, $f(1) = \frac{1}{2}$. Random graphs [21] show that for infinitely many d and n with $d = d(n) \rightarrow \infty$ as $n \rightarrow \infty$, there are n -vertex triangle-free graphs with average degree d and independence number $(2 - o(1)) \left(\frac{\log d}{d}\right) n$. Consequently, the results cannot be improved apart from the multiplicative constant.

 **Exercise 1.2** Use Shearer's bound to derive $R(3, k) \leq (1 + o(1)) \frac{k^2}{\log k}$.

There is a tight connection between the problem of determining $\alpha(G)$ and questions in Ramsey theory. More precisely, determining the minimum possible $\alpha(G)$ for a triangle-free G is equivalent to determining the Ramsey number $R(3, k)$, which is the minimum n such that every graph on n vertices contains either a triangle or an independent set of size k . A result of Kim [17] shows that $R(3, k) = \Theta\left(\frac{k^2}{\log k}\right)$. F. Pontiveros, Griffiths,

and Morris [19] proved that $R(3, k) \geq (1/4 + o(1)) \frac{k^2}{\log k}$. Reducing the gap between these bounds is still a major open problem in Ramsey theory.

Given a graph G , we define $\bar{\alpha}(G)$ to be the average size of all independent sets. We prove a lower bound on it in a triangle-free graph of maximum degree d .

Theorem 1.2

Let G be a triangle-free graph on n vertices with maximum degree d . Then

$$\bar{\alpha}(G) \geq (1 + o(1)) \frac{\log_2 d}{4d} n. \tag{1.2}$$

Idea The proof is based on Shearer’s method and a modification of Alon. The idea is to use the double-counting method: we pick an independent set I in G in a uniformly random way, and bound $\mathbb{E}|I|$ from two points of views: $v \in I$ or not, and how $N(v) \cap I$ looks like.

For the first view,

$$\mathbb{E}|I| = \sum_{v \in V(G)} \Pr(v \in I). \tag{1.3}$$

For the second view,

$$\mathbb{E}|I| \geq \frac{1}{d} \sum_{v \in V(G)} \sum_{u \in N(v)} \Pr(u \in I). \tag{1.4}$$

We shall see these two bounds go in opposite directions, and the desired bound on $\bar{\alpha}(G)$ follows.

Remark Spatial markov property drives the argument. In particular, whether $v \in I$ or not depends only on its "boundary condition".

Proof Let I be an independent set chosen uniformly at random in G . For every vertex $v \in G$, let $H = G - v - N(v)$. Fix a "boundary condition" by conditioning that $I \cap V(H) = S$, and define $X = N(v) \setminus N(S)$, $x = |X|$. Here x denotes the number of vertices in $N(v)$ that are suitable to be added to I .

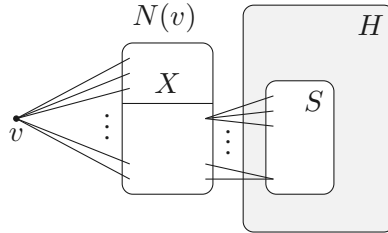


Figure 1.1

Note that $N(v)$ itself is an independent set due to the triangle-freeness of G and I is chosen uniformly at random, so v or any subset of X is equally likely to be included in I . Hence

$$\Pr(v \in I \mid I \cap V(H) = S) = \frac{1}{1 + 2^x}. \tag{1.5}$$

Observe that $\sum_{u \in N(v)} \Pr(u \in I) = \mathbb{E}|N(v) \cap I|$. We obtain a conditional expectation of $|N(v) \cap I|$ also from above two conditions,

$$\mathbb{E}(|N(v) \cap I| \mid I \cap V(H) = S) = \frac{x/2 \cdot 2^x}{1 + 2^x}. \tag{1.6}$$

So

$$\begin{aligned}
 2\mathbb{E}|I| &\geq \sum_{v \in V(G)} \Pr(v \in I) + \frac{1}{d} \sum_{v \in V(G)} \sum_{u \in N(v)} \Pr(u \in I) \\
 &\geq \sum_{v \in V(G)} (\Pr(v \in I) + \frac{1}{d} \mathbb{E}(|N(v) \cap I|)) \\
 &\geq \sum_{v \in V(G)} \min_{0 \leq x \leq d} \max\left\{ \frac{1}{1+2^x}, \frac{1}{d} \cdot \frac{x/2 \cdot 2^x}{1+2^x} \right\} \\
 &\geq (1+o(1)) \frac{\log_2 d}{2d} n.
 \end{aligned} \tag{1.7}$$

where in the third line we used the law of total probability and eqs. (1.5) and (1.6), and in the last line the fact that the function of x in eq. (1.7) achieves its minimum when its two terms equal.

1.2 Sampling with hard-core model

It is natural to ask what if we relax triangle-free in a different direction? Ajtai, Erdős, Komlós, and Szemerédi [2] relax being triangle-free to K_4 -free or any fixed size cliques free. In 1981, they conjectured that any K_t -free graph with maximum degree d has a lower bound on independence number.

Conjecture 1.3

Let $d \in \mathbb{N}$ and G be an n -vertex K_t -free graph with $\Delta(G) \leq d$. Then

$$\alpha(G) \geq \Omega\left(\frac{\log d}{d} n\right). \tag{1.8}$$

When $t \geq 4$, it's still an open problem. However, they [2] proved that there exists an absolute constant c_1 such that for any t -clique-free graph G on n vertices with average degree \bar{d} , $\alpha(G) \geq c_1 \frac{\log((\log \bar{d})/t)}{\bar{d}} n$. Besides, Shearer [23] improved the bound.

Theorem 1.4

Let $d \in \mathbb{N}$ and G be an n -vertex K_t -free graph with $\Delta(G) \leq d$. Then

$$\alpha(G) \geq \Omega\left(\frac{\log d}{d \cdot \log \log d} n\right). \tag{1.9}$$

Another direction is forbidding too many triangles instead of all triangles. On this direction, similar to the previous section, we gain a $\log d$ -factor improvement for graphs with few triangles, which is locally sparse.

Theorem 1.5

Let G be an n -vertex graph with $\Delta(G) = d$. If G contains at most $d^{2-\epsilon} n$ triangles, then

$$\alpha(G) \geq \Omega\left(\frac{\log d}{d} n\right). \tag{1.10}$$

The remainder of this section still focus on triangle-free graph. Let G be a triangle-free graph on n vertices with maximum degree d . Davies, Jenssen, Perkins, and Roberts [9] proved that the expected size of an independent set drawn uniformly at random from such a graph is at least $(1+o(1)) \frac{\log d}{d} n$, which is asymptotically tight.

Theorem 1.6

Let G be an n -vertex triangle-free graph with $\Delta(G) \leq d$. Then

$$\bar{\alpha}(G) \geq (1 + o(1)) \frac{\log d}{d} n. \quad (1.11)$$

The proof is based on the hard-core model from statistical physics which has received a lot of attention in many fields. Next, we will introduce some notions about this model.

For a graph $G = (V, E)$ and fugacity $\lambda > 0$, the hard-core model is defined on the family $\mathcal{I}(G)$ of all independent sets of G where $I \in \mathcal{I}$ has weight $w(I) = \lambda^{|I|}$.

Definition 1.7

The Partition function of the hard-core model on G is denoted by

$$P_G(\lambda) = \sum_{I \in \mathcal{I}} w(I) = \sum_{I \in \mathcal{I}} \lambda^{|I|}. \quad (1.12)$$

Definition 1.8

The hard-core distribution is given by

$$\Pr[I] = \frac{\lambda^{|I|}}{P_G(\lambda)} = \frac{\lambda^{|I|}}{\sum_{J \in \mathcal{I}} \lambda^{|J|}}. \quad (1.13)$$

Remark

- If $\lambda = 1$, the partition function is the number of independent sets and the hard-core distribution is the uniform distribution over all independent sets of G .
- Among all probability distributions over all independent sets of graph G with given mean size, the hard-core model distribution has the highest entropy.

Definition 1.9

Let I be an independent set drawn from the hard-core model with fugacity λ , the expected size is denoted by $\bar{\alpha}_G(\lambda)$.

Proposition 1.10

$\bar{\alpha}_G(\lambda)$ is the scaled log derivative of the partition function, which means $\bar{\alpha}_G(\lambda) = \lambda(\log P_G(\lambda))'$.

Proof

$$\begin{aligned} \mathbb{E}|I| &= \bar{\alpha}_G(\lambda) = \sum_{I \in \mathcal{I}} |I| \cdot \Pr[I] \\ &= \frac{\sum_{I \in \mathcal{I}} |I| \cdot \lambda^{|I|}}{P_G(\lambda)} = \frac{\lambda P'_G(\lambda)}{P_G(\lambda)} = \lambda(\log P_G(\lambda))'. \end{aligned}$$

The proposition has many applications. For instance, if $\bar{\alpha}_G(\lambda)$ has a lower bound, then we can get a lower bound on $P_G(\lambda)$, and by setting $\lambda = 1$, we can count the number of independent sets.

Definition 1.11

The occupancy fraction of G with fugacity λ is denoted by $\frac{\bar{\alpha}_G(\lambda)}{|G|}$.


To prove Theorem 1.6, the following result is helpful, which gives a lower bound on the occupancy fraction

for triangle-free graphs.


Theorem 1.12

Let G be a triangle-free graph on n vertices with maximum degree d . Then for any $\lambda > 0$,

$$\frac{1}{n} \bar{\alpha}_G(\lambda) \geq \frac{\lambda}{1+\lambda} \cdot \frac{W(d \log(1+\lambda))}{d \log(1+\lambda)}, \quad (1.14)$$

where for $z > 0$, $W(z)$ denotes the unique positive real number satisfying $W(z)e^{W(z)} = z$. 

Proposition 1.13

For any graph G , the expected size $\bar{\alpha}_G(\lambda)$ of an independent set is monotone increasing in λ . 

Proof It suffices to show $\bar{\alpha}'_G(\lambda) \geq 0$. For convenience we use P for $P_G(\lambda)$. I is a random independent set drawn from the hard-core model at fugacity λ .

Beacuse

$$\bar{\alpha}_G(\lambda) = \mathbb{E}|I| = \frac{\lambda P'}{P}$$

and

$$P'' = \sum_{I \in \mathcal{I}} |I|(|I| - 1)\lambda^{|I|-2} = \frac{\mathbb{E}|I|^2 - \mathbb{E}|I|}{\lambda^2} P,$$

we have

$$\begin{aligned} \bar{\alpha}'_G(\lambda) &= \left(\frac{\lambda P'}{P} \right)' = \frac{P'}{P} + \frac{\lambda P''}{P} - \frac{\lambda (P')^2}{P^2} \\ &= \frac{\mathbb{E}|I| + \mathbb{E}|I|^2 - \mathbb{E}|I| - (\mathbb{E}|I|)^2}{\lambda} \\ &= \frac{\text{Var}(|I|)}{\lambda} \geq 0. \end{aligned}$$

In order to get a lower bound $\bar{\alpha}(G) \geq (1 + o(1)) \frac{\log d}{d} n$, we use $\bar{\alpha}(G) = \bar{\alpha}_G(1) \geq \bar{\alpha}_G(\lambda)$, for any $0 < \lambda < 1$, and here we use $\lambda = \frac{1}{\log d}$.

Idea Similar to the proof of Theorem 1.2, the idea is to use the double-counting method: we pick an independent set I in G in a uniformly random way, and bound $\mathbb{E}|I|$ from two points of views.

$$\mathbb{E}|I| = \sum_{v \in V(G)} \Pr(v \in I). \quad (1.15)$$

$$\mathbb{E}|I| \geq \frac{1}{d} \sum_{v \in V(G)} \sum_{u \in N(v)} \Pr(u \in I). \quad (1.16)$$

Proof of Theorem 1.12 Let I be an independent set drawn uniformly at random from the hard-core model at fugacity λ . A vertex $v \in V(G)$ is *suitable* if $N(v) \cap I = \emptyset$. So $v \in I$ only if v is suitable.

$$\begin{aligned} \Pr(v \in I) &= \Pr(\{v \in I\} \cap \{v \text{ is suitable}\}) \\ &= \Pr(v \in I \mid v \text{ is suitable}) \cdot \Pr(v \text{ is suitable}). \end{aligned}$$

Claim: For any vertex v in G , $\Pr(v \in I \mid v \text{ is suitable}) = \frac{\lambda}{1+\lambda}$.

Proof Pair up choices of I by conditioning $I \cap (G - v - N(v)) = S$. There are two possibilities for I :

$I = S$ or $I = S \cup \{v\}$, so

$$\Pr(v \in I \mid v \text{ is suitable}) = \frac{\lambda^{|S|+1}}{\lambda^{|S|} + \lambda^{|S|+1}} = \frac{\lambda}{1 + \lambda}.$$

It remains to estimate $\Pr(v \text{ is suitable})$.

Define a random variable X_v to be the number of suitable neighbors of v . So v is suitable when none of the X_v suitable neighbors in I . For any suitable vertex $u \in N(v)$,

$$\Pr(u \notin I \mid u \text{ is suitable}) = \frac{1}{1 + \lambda}.$$

Note that $N(v)$ is a independent set due to the triangle-freeness, therefore,

$$\begin{aligned} \Pr(v \text{ is suitable}) &= \Pr(\text{none of } X_v \text{ suitable neighbors in } I) \\ &= \sum_{x=0}^d \left(\frac{1}{1 + \lambda} \right)^x \cdot \Pr(X_v = x) \\ &= \mathbb{E} \left(\frac{1}{1 + \lambda} \right)^{X_v}. \end{aligned}$$

Counting in eq. (1.15),

$$\begin{aligned} \mathbb{E}|I| &= \sum_{v \in V(G)} \Pr(v \in I) = \sum_{v \in V(G)} \Pr(v \in I \mid v \text{ is suitable}) \cdot \Pr(v \text{ is suitable}) \\ &= \sum_{v \in V(G)} \frac{\lambda}{1 + \lambda} \mathbb{E} \left(\frac{1}{1 + \lambda} \right)^{X_v}. \end{aligned}$$

Then the occupancy fraction

$$\begin{aligned} \frac{\mathbb{E}|I|}{n} &= \frac{\lambda}{1 + \lambda} \cdot \frac{1}{n} \sum_{v \in V(G)} \mathbb{E} \left(\frac{1}{1 + \lambda} \right)^{X_v} \\ &= \frac{\lambda}{1 + \lambda} \cdot \mathbb{E} \left(\frac{1}{1 + \lambda} \right)^X. \end{aligned}$$

where the random variable X is the number of suitable neighbors of a uniform random v and X has two layer of randomness – I, v . By Jensen's inequality, $\mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}(X))$ with $\varphi(X) = \left(\frac{1}{1+\lambda}\right)^X$, we get

$$\frac{\mathbb{E}|I|}{n} \geq \frac{\lambda}{1 + \lambda} \left(\frac{1}{1 + \lambda} \right)^{\mathbb{E}X} \quad (1.17)$$

Counting in eq. (1.16),

$$\mathbb{E}|I| \geq \frac{1}{d} \sum_{v \in V(G)} \sum_{u \in N(v)} \Pr(u \in I \mid u \text{ is suitable}) \cdot \Pr(u \text{ is suitable}).$$

Thus

$$\begin{aligned} \frac{\mathbb{E}|I|}{n} &\geq \frac{1}{d} \cdot \frac{\lambda}{1 + \lambda} \cdot \frac{1}{n} \sum_{v \in V(G)} \sum_{u \in N(v)} \Pr(u \text{ is suitable}) \\ &= \frac{1}{d} \cdot \frac{\lambda}{1 + \lambda} \mathbb{E}X. \end{aligned} \quad (1.18)$$

Combining eq. (1.17) and eq. (1.18), we have

$$\frac{1}{n} \bar{\alpha}_G(\lambda) \geq \frac{\lambda}{1 + \lambda} \max \left\{ \left(\frac{1}{1 + \lambda} \right)^{\mathbb{E}X}, \frac{\mathbb{E}X}{d} \right\} \quad (1.19)$$

$$\geq \frac{\lambda}{1 + \lambda} \cdot \frac{W(d \log(1 + \lambda))}{d \log(1 + \lambda)} \quad (1.20)$$

where in the last line we used the fact that the function in eq. (1.19) is optimized when its two terms are equal.

From Theorem 1.12, we have the following consequences:

- Since $\bar{\alpha}_G(\lambda)$ is monotone increasing in λ ,

$$\bar{\alpha}(G) = \bar{\alpha}_G(1) \geq \bar{\alpha}_G\left(\frac{1}{\log d}\right) = (1 + o(1)) \frac{\log d}{d} n.$$

- Recall that $\bar{\alpha}_G(\lambda) = \lambda[\log P_G(\lambda)]'$, then

$$\begin{aligned} \frac{1}{n} \log P_G(1) &= \int_0^1 \frac{\bar{\alpha}_G(t)}{t} dt \\ &\geq \frac{1}{d} \int_0^1 \frac{n}{1+t} \cdot \frac{W(d \log(1+t))}{\log(1+t)} dt \end{aligned} \quad (1.21)$$

$$= \frac{1}{d} \int_0^{W(d \log 2)} (1+u) du \quad (1.22)$$

$$= \left(\frac{1}{2} + o_d(1)\right) \frac{\log^2 d}{d},$$

where we used Theorem 1.12 in eq. (1.21) and we let $u := W(d \log(1+t))$ in eq. (1.22).

Using this counting result, we have the following corollary

Corollary 1.14

Let $d \in \mathbb{N}$ and G be a triangle-free graph on n vertices with $\Delta(G) \leq d$. Then the number of independent sets in G

$$i(G) \geq e^{(\frac{1}{2} + o_d(1)) \frac{\log^2 d}{d} n}.$$



1.3 Conclusion

This occupancy fraction method comes from statistical physics which studying of matter via probabilistic and statistical method. Our motivation is trying to see whether the (global) macroscopic properties of matter can be derived solely from their local microscopic interactions. Instead of keeping track of all particles, we gonna treat them as random distributed with certain local constraints

For our problems, we use the following table to give a summary.

	Global	Local
Independent set	$\alpha(G)$	For any edge $u \sim v$ whether vertex v can be chosen into the independent set depends on whether v is suitable or not
Sphere packing	Packing density	Centers of balls are not so close(the distance of centers of balls is at least $2r_d$ where r_d is the radius in d -dimension)

Chapter 2 Sphere Packing

In this chapter, we will set up the distribution that can be viewed as continuous version of hard-core model. Before setting on, we introduce the packing density firstly. Let P be a sphere packing of none overlapping identical spheres in R^d , we have the following definitions.

Definition 2.1

1. $B_R(0)$ denotes the radius- R ball in R^d centered at the origin;
2. r_d denotes the radius such that $B_{r_d}(0)$ has volume 1 (a unit ball);
3. $\theta(d) = \sup_P \lim_{R \rightarrow \infty} \frac{\text{vol}(\phi, B_R(0))}{\text{vol}(B_R(0))}$ denotes the sphere packing density in R^d .



In history, there are numerous results on sphere packing density. Obviously, $\theta(1)=1$. Since R^1 is a line and the ball in R^1 is a line segment. If we want to pack the $B_R(0)$ in R^1 , then we just need to put the line segment one next another.

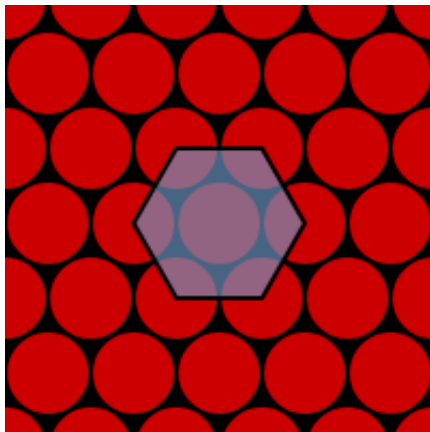
In 1910, Thue [24] proved the following theorem.

Theorem 2.2

In R^2 , $\theta(2) = \frac{\pi}{\sqrt{12}} = 0.9068 \dots$



Remark: The optimal arrangement is achieved by hexagonal packing. We can place the ball inside every hexagon. Figure 2.1 (1) is an illustration of hexagonal packing.



(1)



(2)

Figure 2.1: Illustration of hexagonal packing and orange packing

After more than one hundred years late, Hales [13] proved the the following theorem in 2005.

Theorem 2.3

In R^3 , $\theta(3) = \frac{\pi}{\sqrt{18}} = 0.7404 \dots$



Remark: The optimal arrangement is orange packing. Figure 2.1 (2) is an illustration of orange packing.

In 2017, Maryna S. Viazovska [26] showed an optimal arrangement in R^8 . After the sphere packing in

R^8 was solved, within a week, Viazovska, along with Cohn and three other mathematicians [8], successfully extended her method to cover R^{24} too. So it is natural to have the following problem.

Problem 2.1 What's the order of the main magnitude of $\theta(d)$ with $d \rightarrow \infty$?

2.1 Definitions

To define the partition function of Canonical Hard Sphere model and Grand Canonical Hard Sphere model respectively, we employ the following definitions.

Definition 2.4

Let S be a set and $k \in \mathbb{N}$,

1. $\binom{S}{k}$ denotes the set of all k -sets in S ;
2. S^k denotes all ordered k -tuples in S .



Definition 2.5

Let $S \subseteq R^d$ be a measurable set, $P_k(S) \subseteq \binom{S}{k}$ with $P_k(S) = \{(x_1, x_2, \dots, x_k) : d(x_i, x_j) > 2r_d\}$, i.e., $P_k(S)$ is the set of all size- k sphere packing in S .



Equipped with these two definitions, we focus on the definitions of partition function of Canonical Hard Sphere model and Grand Canonical Hard Sphere model.

Definition 2.6

The partition function of Canonical Hard Sphere model is denoted by

$$\hat{Z}_S(k) = \frac{1}{k!} \int_{S^k} I_{D(x_1, \dots, x_k)} dx_1, \dots, dx_k$$

where $D(x_1, \dots, x_k)$ is the event that $d(x_i, x_j) > 2r_d$.



Note that $\hat{Z}_S(k)$ is the volume of $P_k(S)$ and for a uniform random k -tuple X_k ,

$$\Pr(X_k \in P_k(S)) = \frac{\hat{Z}_S(k)}{|V|} = \frac{k!}{\text{vol}(S)^k} \cdot \hat{Z}_S(k),$$

where $|V| = \frac{\text{vol}(S)^k}{k!}$ is the total volume.

Definition 2.7

The partition function of Grand Canonical Hard Sphere model at S with fugacity λ is denoted by

$$\hat{Z}_S(\lambda) = \sum_{k=0}^{\infty} \lambda^k \hat{Z}_S(k).$$



Remark $Z_G(\lambda) = P_G(\lambda) = \sum_{k=0}^{\infty} i_k(G) \lambda^k$ where $i_k(G)$ denotes the number of independent set in G of size k .

Definition 2.8

Let P be a sphere packing with same radius and non-overlapping in R^d , the sphere packing density is

denoted by

$$\theta(d) = \sup_P \lim_{R \rightarrow \infty} \frac{\text{vol}(\phi, B_R(0))}{\text{vol}(B_R(0))}.$$



2.2 Known Results

The following are two results about the upper bound of $\theta(d)$. In 1978, Kabatianskii and Levenshtein [15] proved the following theorem.

Theorem 2.9

Let $d \in \mathbb{N}$, $\theta(d) \leq 2^{-0.599d}$.



In 2013, Venkatesh [25] showed the following results.

Theorem 2.10

For a fixed $d \in \mathbb{N}$, it holds that

1. For a sufficiently large d , $\theta(d) \geq 65963 \cdot 2^{-d}$;
2. Along a sparse sequence of $\dim\{d_i\}$, $\theta(d) = \Omega(d \cdot \log \log d_i \cdot 2^{-d_i})$.



Later on, in 2014, Cohn and Zhao [7] proved that

Theorem 2.11

Let $d \in \mathbb{N}$, there exists $c > 0$ such that in every R^d the following holds

$$\theta(d) \leq c \cdot 2^{-0.599d}.$$



Exercise 2.1 : Prove that $\theta(d) \geq 2^{-d}$.

Idea Consider a maximal packing in R^d .

In this section, we will prove a lower bound with a slightly small constant $c' \leq 65963$ which holds for every d . Before the proof, we recall some definitions of Hard-Sphere model over bounded measurable set $S \subseteq R^d$.

The first one is the partition function of Hard-Sphere model.

Definition 2.12

Let $S \subseteq R^d$ be a measurable set, the partition function of Hard-Sphere model with fugacity λ is denoted by

$$Z_S(\lambda) = \sum_{k=0}^{\infty} \lambda^k \hat{Z}_S(k)$$

where

$$\hat{Z}_S(k) = \frac{1}{k!} \int_{S^k} 1_{D(x_1, \dots, x_k)} dx_1 \dots dx_k$$

is the volume of size- k packing and $D(x_1, \dots, x_k)$ is the event that, for any $i, j \in [k]$ and $i \neq j$, $d(x_i, x_j) > 2r_d$.



By the definition of partition function of Hard-Sphere model, we have following observations.

Observation 2.13

1. Let S be a measurable set in R^d , then $\hat{Z}_S(0) = 1$;
2. If we sample a packing X according to Hard-Sphere model distribution over some region S , then $\Pr(|X| = k) = \frac{\lambda^k \cdot \hat{Z}_S(k)}{Z_S(\lambda)}$.

Next, we give the definition of expected packing density with $X \sim$ Hard Sphere model distribution over S .

Definition 2.14

Let $S \subseteq R^d$ be a measurable set with fugacity λ , then the expected packing density was denoted as $\alpha_S(\lambda) = \frac{\mathbb{E}_{S,\lambda}(X)}{\text{vol}(S)}$.

In this notation, there are two results about expected packing density. The first one was proved by Jenssen, Joos and Perkins [14] in 2019.

Theorem 2.15

Let $d \in N$, S is a bounded measurable set in R^d and $\lambda \geq 3^{-\frac{d}{2}}$, then

$$\theta(d) \geq \alpha_S(\lambda) \geq (1 + o(1)) \cdot \log\left(\frac{2}{\sqrt{3}}\right) \cdot d \cdot 2^{-d}.$$

Recently, Gil-Fernandéz, Kim, Liu and Pikhurko improves the bound by a factor of 2.4.

Theorem 2.16

For any $\epsilon > 0$, there exist $\delta > 0$ and d_0 such that for any $d > d_0$ and $\lambda \geq (\frac{1}{\sqrt{2}} - \delta)^d$, it holds that

$$\alpha_S(\lambda) \geq (\log \sqrt{2} - \epsilon) \cdot d \cdot 2^{-d}.$$

Before the proof of Theorem 2.16, we need following basic properties.

Lemma 2.17

The expected density $\alpha_S(\lambda)$ is the scaled log derivative of the partition functions which is

$$\alpha_S(\lambda) = \frac{\lambda}{\text{vol}(S)} \cdot (\log Z_S(\lambda))'.$$

Proof :

$$\begin{aligned} \alpha_S(\lambda) &= \frac{\mathbb{E}[|X|]}{\text{vol}(S)} = \frac{1}{\text{vol}(S)} \cdot \sum_{k=1}^{\infty} k \cdot \Pr(|X| = k) \\ &= \frac{1}{\text{vol}(S)} \cdot \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k \cdot \hat{Z}_S(k)}{Z_S(\lambda)} \\ &= \frac{\lambda}{\text{vol}(S)} \cdot \sum_{k=1}^{\infty} k \cdot \frac{\lambda^{k-1} \cdot \hat{Z}_S(k)}{Z_S(\lambda)} \\ &= \frac{\lambda}{\text{vol}(S)} \cdot \frac{Z'_S(k)}{Z_S(k)} \\ &= \frac{\lambda}{\text{vol}(S)} \cdot (\log Z_S(\lambda))'. \end{aligned}$$

Lemma 2.18

The expected density $\alpha_S(\lambda)$ is monotone increasing with λ .



Proof : By Lemma 2.17, we have $\alpha_S(\lambda) = \frac{\lambda}{\text{vol}(S)} \cdot (\log Z_S(\lambda))'$, then $\lambda \cdot \text{vol}(S) \cdot \alpha'_S(\lambda) = \text{Var}(|X|) > 0$.

Another key definition we need to define is some notation for points that are suitable to be added, which is so called the free volume.

Definition 2.19

The expected free volume of the hard sphere model on S is denoted by

$$FV_S = \frac{1}{\text{vol}(S)} \int_S \Pr(d(x_0, X) > 2r_d) dx_0$$

where X is the random packing sample in the hard sphere model.



Note that FV_S is the expected fraction of the volume at which a new sphere can be added to X . Now let see some basic properties about the free volume.

Lemma 2.20

Let S be a bounded measurable set in \mathbb{R}^d with positive volume, then

$$\alpha_S(\lambda) = \lambda \cdot FV_S.$$



Proof :

$$\begin{aligned} \alpha_S(\lambda) &= \frac{\mathbb{E}|X|}{\text{vol}(S)} \\ &= \frac{\mathbb{E}|X|}{\text{vol}(S)} \sum_{k=0}^{\infty} (k+1) \Pr(|X| = k+1) \\ &= \frac{1}{\text{vol}(S)} \sum_{k=0}^{\infty} (k+1) \frac{\lambda^{k+1} \hat{Z}_S(k+1)}{Z_S(\lambda)} \\ &= \frac{1}{\text{vol}(S) Z_S(\lambda)} \sum_{k=0}^{\infty} (k+1) \int_{S^{k+1}} \frac{\lambda^{k+1}}{k!} \mathbf{1}_{D(x_0, x_1, \dots, x_k)} dx_1 \dots dx_k dx_0 \\ &= \frac{\lambda}{\text{vol}(S) Z_S(\lambda)} \int_S \left[1 + \sum_{k=1}^{\infty} \int_{S^k} \frac{\lambda^k}{k!} \mathbf{1}_{D(x_0, x_1, \dots, x_k)} dx_1 \dots dx_k \right] dx_0 \\ &= \lambda \cdot FV_S. \end{aligned}$$

where in the last equality, we used

$$\frac{1}{Z_S(\lambda)} \int_S \left[1 + \sum_{k=1}^{\infty} \int_{S^k} \frac{\lambda^k}{k!} \mathbf{1}_{D(x_0, x_1, \dots, x_k)} dx_1 \dots dx_k \right] dx_0 = \int_S \Pr(d(x_0, X) > 2r_d) dx_0.$$

Recall that for independent set problems, we have two layers of randomness, one is that we sample I according to hard-core model, another one is that we sample uniform random vertex v . Now we can do the same two layers of randomness experiment. We sample X according to hard sphere model over S with fugacity λ , then we sample a uniform point v over S .

Definition 2.21

Let

$$T := \{x \in B_{2r_d}(v) \cap S : d(x, y) > 2r_d, \forall y \in X \cap B_{2r_d}^c\}.$$



We call T the set of externally uncovered points, see Figure 2.2.

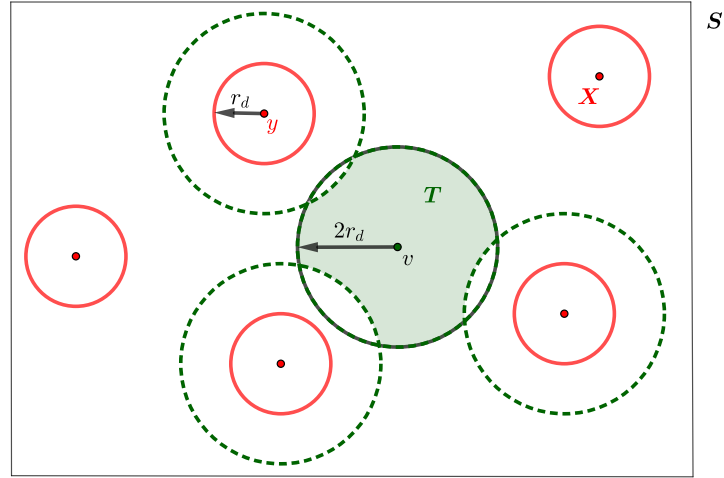


Figure 2.2

2.3 Proof of Theorem 2.7

Idea We will bound $\alpha_S(\lambda)$ in two ways.

- The first way is to use FV_S .
- The second way is to run separate Poisson point process on T .

Now, we do the first way of counting by the following lemma.

Lemma 2.22

- (i) $\alpha_S(\lambda) = \lambda \cdot \mathbb{E}\left[\frac{1}{Z_T(\lambda)}\right]$;
(ii) $\alpha_S(\lambda) = \lambda \cdot \mathbb{E}_{X,v}(e^{-\lambda \cdot \text{vol}(T)})$.

Proof :

(i) By Lemma 2.20, we have

$$\begin{aligned}
 \alpha_S(\lambda) &= \lambda \cdot FV_S \\
 &= \lambda \cdot \frac{1}{\text{vol}(S)} \int_S \Pr(d(v, X) > 2r_d) \cdot dv \\
 &= \lambda \cdot \mathbb{E}[\mathbf{1}_{\{T \cap X = \emptyset\}}] \\
 &= \lambda \cdot \mathbb{E}\left[\frac{1}{Z_T(\lambda)}\right].
 \end{aligned}$$

(ii) Recall that

$$\begin{aligned}
 Z_S(\lambda) &= \sum_{k \geq 0} \lambda^k \hat{Z}_S(k) \\
 &= \sum_{k \geq 0} \frac{\lambda^k}{k!} \int_{S^k} \mathbf{1}_{D(x_1, \dots, x_k)} dx_1 \dots dx_k \\
 &\leq \sum_{k \geq 0} \frac{\lambda^k}{k!} \text{vol}(S)^k \\
 &= e^{\lambda \cdot \text{vol}(S)}.
 \end{aligned}$$

By (i), $\alpha_S(\lambda) = \lambda \cdot \mathbb{E}[\frac{1}{Z_T(\lambda)}] \geq \lambda \cdot \mathbb{E}_{X,v}(e^{-\lambda \cdot \text{vol}(T)})$.

Lemma 2.22 (ii) is the first bound we obtain. Next, we do the second way of counting.

Lemma 2.23

$$\alpha_S(\lambda) \geq 2^{-d} \cdot \mathbb{E}[\alpha_T(\lambda) \cdot \text{vol}(T)].$$

Proof : As $\text{vol}(S \cap B_{2r_d}(v)) \leq 2^d$ for any $v \in S$, we have

$$\begin{aligned}
 \alpha_S(\lambda) &= \frac{1}{\text{vol}(S)} \cdot \mathbb{E}|X| \\
 &\geq 2^{-d} \cdot \mathbb{E}|X \cap B_{2r_d}(v)| \\
 &= 2^{-d} \cdot \mathbb{E}[\alpha_T(\lambda) \cdot \text{vol}(T)],
 \end{aligned}$$

where the last equality holds by Spatial Markov property.

Let $t = \text{vol}(T)$. By the two lemmas above, we have

$$\alpha_S(\lambda) \geq \max\{\lambda \cdot \mathbb{E}_{X,v}(e^{-\lambda t}), 2^{-d} \cdot \mathbb{E}(\alpha_T(\lambda) \cdot t)\}. \quad (2.1)$$

Note that the first bound $\lambda \cdot \mathbb{E}_{X,v}(e^{-\lambda t})$ is a decrease function of t , while the second bound $2^{-d} \cdot \mathbb{E}(\alpha_T(\lambda) \cdot t)$ is an increase function of t . Hence the first bound is large when t is small, and the second bound is large when t is big.

To bound $2^{-d} \cdot \mathbb{E}(\alpha_T(\lambda) \cdot t)$, we use the following lemma.

Lemma 2.24

For every $\beta > 0$, there exists k_0 such that for any integer $k \geq k_0$ and any $\lambda, t, d > 0$, if a measurable set $T \subseteq \mathbb{R}^d$ is of volume t and $k \leq \lambda t$, then we have

$$\alpha_T(\lambda) \cdot t \geq (1 - \beta) P_k \cdot k,$$

where $P_i = \Pr(\text{uniform independent } i \text{ points in } T \text{ are at pairwise distance at least } 2r_d)$.

To bound $\lambda \cdot \mathbb{E}_{X,v}(e^{-\lambda t})$, we first give some definitions and lemmas.

Definition 2.25

- For a measurable set $A \subseteq \mathbb{R}^d$, its symmetric rearrangement is

$$A^* = B_{\text{vol}(A)^{1/d} \cdot r_d}(0).$$

- For a measurable set $T \subseteq \mathbb{R}^d$, define

$$f(T) = \int_T \text{vol}(B_{2r_d}(u) \cap T) du.$$

Lemma 2.26

For any bounded measurable set $T \subseteq \mathbb{R}^d$,

$$f(T) \leq f(T^*).$$



This lemma can be proved by Riesz's rearrangement inequality.

Lemma 2.27

Let T be a measurable set in \mathbb{R}^d of volume $t \in [2^{d/2}, 2^d]$ and u be a uniform random point in T . Then

$$\mathbb{E}_u[\text{vol}(B_{2r_d}(u) \cap T)] \leq 2 \cdot 2^d (1 - t^{-2/d})^{d/2}.$$



Proof : Note that

$$\mathbb{E}_u[\text{vol}(B_{2r_d}(u) \cap T)] = \frac{f(T)}{\text{vol}(T)}.$$

By Lemma 2.26, we may assume that T is the ball of radius $\rho = t^{1/d} \cdot r_d$ around the center 0, that is $T = B_\rho(0)$.

Then

$$\begin{aligned} \mathbb{E}_u[\text{vol}(B_{2r_d}(u) \cap T)] &= \frac{1}{t} \int_T \left(\int_T \mathbf{1}_{\{d(u,v) \leq 2r_d\}} dv \right) du \\ &= \frac{2}{t} \int_T \int_T \mathbf{1}_{\{d(u,v) \leq 2r_d\}} \mathbf{1}_{\{\|v\| \leq \|u\|\}} dv du \\ &= 2 \cdot \max_{u \in B_\rho(0)} \int_T \mathbf{1}_{\{d(u,v) \leq 2r_d\}} \mathbf{1}_{\{\|v\| \leq \|u\|\}} dv. \end{aligned}$$

When u is on the boundary of T , the volume of the intersection is maximum and we bound it by the red ball in

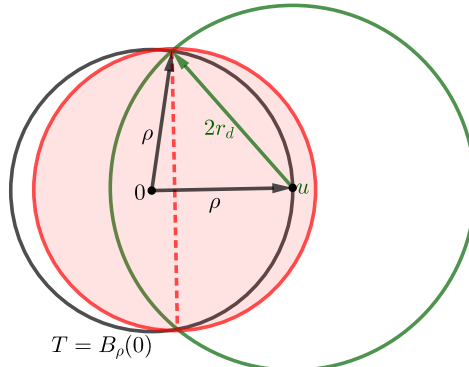


Figure 2.3

Figure 2.3. To be more precise, we may assume that $\rho \geq \sqrt{2}r_d$, otherwise $t \leq 2^{d/2}$. Then the radius of the red ball is $2 \cdot \sqrt{1 - t^{-2/d}} \cdot r_d$. Hence

$$\begin{aligned} \mathbb{E}_u[\text{vol}(B_{2r_d}(u) \cap T)] &\leq \max\{2^{d/2}, 2 \cdot \max_{\sqrt{2} \leq x \leq t^{1/d}} (2\sqrt{1 - x^{-2}})^d\} \\ &= 2 \cdot (2\sqrt{1 - t^{-2/d}})^d \\ &= 2 \cdot 2^d (1 - t^{-2/d})^{d/2}. \end{aligned}$$

Proof of Theorem 2.16. : Given $\epsilon > 0$, choose $\beta \gg \delta > 0$. Let $d \rightarrow \infty$. Since $\alpha_S(\lambda)$ is non-decreasing in λ , it is enough to consider $\lambda = (1/\sqrt{2} - \delta)^d$. Let S be a large ball in \mathbb{R}^d . We need to show that

$$\alpha_S(\lambda) \geq (\log \sqrt{2} - \epsilon) \cdot d \cdot 2^{-d}. \quad (2.2)$$

Then we do the two-step experiment, where we sample X according to hard sphere model over S with

fugacity λ , then we sample a uniform random point v over S . This generates an externally uncovered part $T = T(X, v)$, which depends on X and v .

Let $k = (\log \sqrt{2} - \epsilon/2) \cdot d$. Now we cut the points in S into two parts. For the part where the corresponding volume of T is small, we can use the first bound in eq. (2.1). For the other part, we can use the second bound in eq. (2.1).

For $X \subseteq S$, let

$$L = L(X) = \{u \in S : t(X, u) \leq k/\lambda\},$$

where $t(X, u)$ is the volume of $T(X, u)$. Using the first bound in eq. (2.1), we get

$$\alpha_S(\lambda) \geq \lambda \cdot \mathbb{E}_X \mathbb{E}_v e^{-\lambda t(X, v)}.$$

Then

$$\begin{aligned} \text{vol}(S) \cdot \alpha_S(\lambda) &\geq \lambda \mathbb{E}_X \left[\int_{v \in S} e^{-\lambda t(X, v)} dv \right] \\ &\geq \lambda \mathbb{E}_X \left[\int_{v \in L} e^{-\lambda t(X, v)} dv \right] \\ &\geq e^{\epsilon d/3} \cdot 2^{-d} \cdot \mathbb{E}_X [\text{vol}(L)]. \end{aligned}$$

This means we may assume $\mathbb{E}_X [\text{vol}(L)] \leq \text{vol}(S) \cdot e^{-\epsilon d/4}$, otherwise eq. (2.2) holds. By Markov's inequality, we have

$$\Pr_X(\text{vol}(L) \geq \text{vol}(S) \cdot e^{-\epsilon d/6}) \leq e^{-\epsilon d/12}.$$

That is, for a typical outcome X , t is relatively large.

That is, for a typical outcome X , t is relatively large, except for a very small of points in S . Now let's fix one such outcome X where the inequality section 2.3 holds. Take any X with $\text{vol}(L) \leq \text{vol}(S) e^{-\epsilon d/6}$. And for every $v \in S \setminus L$, by definition we have

$$t = t(X, v) \geq k/\lambda \geq (\sqrt{2} + \delta/3)^d,$$

we also have $t \leq 2^d$, as $T \subseteq B_{2r_d}(\cdot)$. This means $k \leq \lambda t$, by Lemma 2.24. So for every $v \in S \setminus L$, we have

$$\alpha_T(\lambda) t \geq (1 - \beta) P_k k.$$

Claim P_k is very close to 1 i.e. for every $\delta > 0, P_k \geq 1 - \delta$.

Idea Greedily pick k points. Each point takes up negligible (in particular exponentially small in d) portion of T (by Lemma 2.27).

Proof Let's consider the function $g(t) = (f(\tau))^{-d}$, where $\tau := t^{1/d}$ and $f(\tau) = \frac{\tau}{2\sqrt{1-\tau^{-2}}}$. We can observe that $f(\sqrt{2}) = 1$, and f is strictly increasing on $[\sqrt{2}, 2]$, since

$$f'(x) = \frac{x^2 - 2}{2\sqrt{1 - 1/x^2}(x^2 - 1)} > 0.$$

Then recall that for $v \notin L, t \geq (\sqrt{2} + \delta/3)^d$, that means

$$f(t^{1/d}) \geq f(\sqrt{2} + \delta/3) > 1$$

i.e. $g(t)$ is exponentially small in d .

The Lemma 2.27 says that $2g(t)$ upper bounds expected fraction of measure of T intersect ball of radius $2r_d$ center at a uniform point of T . Let $x_1, \dots, x_k \in T$ independent uniform points. Call x_i bad if

$$(E_1) : \text{vol}(B_{2r_d}(x_i) \cup T) \geq t/d^3$$

or

$$(E_2) : \text{within distance } 2r_d \text{ from } x_1, \dots, x_{i-1}.$$

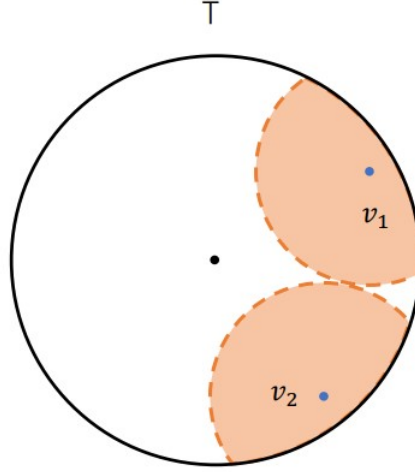


Figure 2.4

Therefore, in order to solve the above problem, it is sufficient to show $P_r(\text{exite at least one bad } x_i) = o(1)$; this is equivalent of showing that for each i , $P_r(x_i \text{ is the bad vertex}) = o(1/k)$.

This is simple because this probability $P_r(x_i \text{ is the bad vertex}) \leq P_r(E_1^c) + P_r(E_2^c|E_1^c)$. By Markov inequality, $P_r(E_1^c) = e^{-\Omega(d)}$ (we have shown that in Lemma 2.27). Meanwhile, $P_r(E_2^c|E_1^c) \leq \frac{i-1}{d^3}$, so we can get the conclusion above.

By Lemma 2.24, for every point $v \in S \setminus L$, we have $\alpha_T(\lambda)t \geq (1 - \beta)P_k k$. According to the Claim, we have $\alpha_T(\lambda)t \geq (1 - \beta)P_k k \geq (1 - 2\beta)k$.

Now using the second bound in eq. (2.1), we get

$$\begin{aligned}
 2^d \alpha_S(\lambda) &\geq E_{X,v}(\alpha_T(\lambda)t) \\
 &= \frac{1}{\text{vol}(S)} E_X \left[\int_{v \in S} \alpha_T(\lambda)t dv \right] \\
 &\geq \frac{1}{\text{vol}(S)} E_X \left[\int_{v \in S \setminus L} \alpha_T(\lambda)t dv \right] \\
 &\geq (1 - e^{-cd/12})(1 - e^{-cd/6})(1 - 2\beta)k \\
 &= (1 - o(1))(\ln(\sqrt{2}) - \epsilon/2)d
 \end{aligned} \tag{2.3}$$

where in eq. (2.3), We only take X such that $\text{vol}(L) \leq \text{vol}(S)e^{-cd/6}$ (this is by the inequality section 2.3). Finally, we can find out that $\alpha_S(\lambda) \geq (\ln(\sqrt{2}) - \epsilon)2^{-d}d$.

2.4 Example

The last example is about counting the number of the independent sets in G .

Definition 2.28

The number of independent sets in G is denoted by $i(G)$.



In 2000, Kahn [16] implied the above question, and in 2010, Yufei Zhao [27] solved it. We get the following theorem.

Theorem 2.29. [Kahn-Zhao]

For every n – vertex d – regular G ,


$$i(G) \leq i(K_{d,d})^{n/2d}. \quad (2.4) \quad \heartsuit$$

And then, Davies, Jessen, Perkins and Roberts [9] made an extension in 2017.

Theorem 2.30. [Davies-Jessen-Perkins-Roberts]

For every d – regular G and every $\lambda > 0$,

$$\bar{\alpha}_G(\lambda) \leq \bar{\alpha}_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1} \quad (2.5) \quad \heartsuit$$

 **Exercise 2.2** Show that theorem Davies-Jessen-Perkins-Roberts implies theorem Kahn-Zhao.

To proof theorem Davies-Jessen-Perkins-Roberts:

Idea We will draw this random independent set I according to the hard core model with fugacity λ on this graph G . Independently, we draw vertex v uniformly over all vertices in G . Let random value $X = X(I, v)$ counting the number of occupied neighbour $|I \cap N(v)|$. We write for $k \in \{0, 1, \dots, d\}$, $P_k = Pr(X = k)$.

Again we get a bound of occupancy fraction $\bar{\alpha}_G(\lambda)$ in two ways. For the first view, if v is in the independent set I or not. Only when v is suitable, i.e. $X = 0$, v is in X . For the second view, by looking at X , how $N(v)$ intersect I .

Our goal is that maximize $\bar{\alpha}_G(\lambda)$ via a linear program (involving P_0, P_1, \dots, P_d).

Proof For the first view, we have

$$\begin{aligned} \bar{\alpha}_G(\lambda) &= \frac{\mathbb{E}|I|}{|G|} = \frac{1}{|G|} \sum_{u \in V(G)} Pr(u \in I) = Pr(v \in I) \\ &= Pr(\{v \in I\} \cap \{X = 0\}) \\ &= Pr(v \in I | X = 0) Pr(X = 0). \end{aligned}$$

Recall

$$Pr(v \in I | X = 0) = Pr(v \in I | v \text{ is suitable}) = \frac{\lambda}{1 + \lambda}.$$

Then we can get

$$\bar{\alpha}_G(\lambda) = \frac{\lambda}{1 + \lambda} P_0.$$

For the second view,

$$\begin{aligned} \bar{\alpha}_G(\lambda) &= \frac{1}{|G|} \sum_{u \in V(G)} \frac{1}{d} |N(v) \cap I| = \frac{\mathbb{E}}{d} \\ &= \frac{1}{d} (P_1 + 2P_2 + \dots + dP_d). \end{aligned}$$

Then we have

$$\begin{aligned} \bar{\alpha}_G(\lambda) &= \frac{\lambda}{1 + \lambda} P_0 \\ &= \frac{1}{d} (P_1 + 2P_2 + \dots + dP_d). \end{aligned}$$

Next, we want too get maximum P_0 , such that

$$\text{C1 } P_0 + P_1 + \dots + P_d = 1.$$

$$\text{C2 } \frac{\lambda}{1 + \lambda} P_0 = \frac{1}{d} (P_1 + 2P_2 + \dots + dP_d).$$

C3 for every $2 \leq k \leq d$, $(d - k + 1)\lambda P_{k-1} \geq kP_k$.

Proof of C3 On the one hand, for an outcome (of $X = k$) J from the independent set I with $|J \cap N(v)| = k$, we have k ways from G to get J' with $|J' \cap N(v)| = k - 1$.

On the other hand, by the definition of hard-core model, we have

$$\lambda Pr(J') = Pr(J).$$

And, J' can be obtained from at most $d - k + 1$ many other choices of J with $|J \cap N(v)| = k$. That finish the proof.

Recall, we have $\bar{\alpha}_G(\lambda) = \frac{\lambda}{1+\lambda} P_0$, suffices to show that

$$P_0 \leq \frac{(1 + \lambda)^d}{2(1 + \lambda)^d + 1}.$$

Claim If a choice of (P_0, \dots, P_d) with maximum P_0 , then all equality hold in C3.

Proof of Claim Suppose not, say for some k , $(d - k + 1)\lambda P_{k-1} > kP_k$, then we can increase P_0 by small $\epsilon > 0$, move some mass (function of ϵ) from P_{k-1} to P_k and fix other P_i . We can check that these C1,C2,C3 still hold, and thne we get P_0 is not maximum. Contradiction!

Now that the equality in C3 holds, we have a system linear equations with $(d + 1)$ unknowns and $(d + 1)$ equalities constraints. Then it's full rank, so there exists a unique solution. One can check $K_{d,d}$ satisfies all C1-C3.

To solve it, we iterate C3 $(d - k + 1)\lambda P_{k-1} = kP_k$ for $2 \leq k \leq d$. We can show that

$$P_k = \frac{(d - 1)!}{k!(d - k)!} \lambda P_1. \quad (2.6)$$

Then plug eq. (2.6) into C2 we can show that

$$\frac{P_1}{d\lambda} = \frac{P_0}{(1 + \lambda)^d}. \quad (2.7)$$

Plug eq. (2.6) and eq. (2.6) into C1 then we can get the result we want.

Chapter 3 The Polynomial method

Usually, the idea of polynomial methods is to use the information of roots of polynomial to solve some combination problems.

Disc-Kakeya problem is an example of using the low degree polynomial can not have too many roots to show that certain combination structure can not be too small.

Definition 3.1

- We say a polynomial is a 0 – polynomial if the coefficient of all its monomials are 0.
- We say that a polynomial f vanishes on a set A , or we also say f is identically 0 on A , if $f(a) = 0$, for every $a \in A$.



The two definition are different. The difference is, f is polynomial on \mathbb{F}_p^n , i.e. n – variate polynomial in $\mathbb{F}_p[x_1, \dots, x_n]$, f could identically 0 everywhere (on \mathbb{F}_p^n) but not 0 – polynomial, e.g. $f(x) = x^p - x$ when $n = 1$.

Fact 3.2

For every non-zero polynomial f on \mathbb{F}_p^n with degree d , and every $\mathbf{a}, \mathbf{z} \in \mathbb{F}_p^n$ with $\mathbf{z} \neq \mathbf{0}$, let $L = \{\mathbf{a} + t\mathbf{z} : t \in \mathbb{F}_p\}$ be the corresponding line. Then the restriction of f on L , denote by $f_L(t)$, is a univariate polynomial (of t) with degree at most d and leading coefficient $f_d(\mathbf{z})$, where $f_d(\mathbf{z})$ is the homogeneous degree d part of f .



Proof For each monomials $\prod_{i=1}^n x_i^{r_i}$ of f has value $\prod_{i=1}^n (a_i + tz_i)^{r_i}$ at $\mathbf{a} + t\mathbf{z}$. To get t^d term, we need to choose tz_i term in each bracket. So the leading coefficient is $t^d \prod_{i=1}^n z_i^{r_i}$, where $\prod_{i=1}^n z_i^{r_i} = f_d(\mathbf{z})$ as require.

Lemma 3.3

If f is a non-zero polynomial on \mathbb{F}_p^n with degree $d < p$, then f cannot vanish on the \mathbb{F}_p^n .



Proof We use induction on n .

For the basic case $n = 1$, if f vanish on \mathbb{F}_p^n , then it has at least p roots which implies that the degree of f is at least p . Contradiction!

For general n , we suppose that f vanishes on \mathbb{F}_p^n . Think of f as a polynomial of x_1 with coefficient in $\mathbb{F}_p[x_2, \dots, x_n]$. By polynomial division algorithm, for each $a \in \mathbb{F}_p$, $f(x) = P(x_1, \dots, x_n)(x_1 - a) + Q(x_2, \dots, x_n)$, where $Q(x_2, \dots, x_n)$ is the remainder. As $Q(x_2, \dots, x_n) = f(a, x_2, \dots, x_n)$ vanishes on \mathbb{F}_p^{n-1} . By induction hypothesis, Q is a 0 – polynomial. And because for each $a \in \mathbb{F}_p$, $(x_1 - a)$ divides f , as a univariate polynomial of x_1 , f has at least p roots, which means its degree is at least p . Contradiction!

3.1 Schwartz–Zippel Lemma

In this section, we will state and prove Schwartz–Zippel Lemma, which is a very basic and powerful lemma.

Lemma 3.4. Schwartz–Zippel

Every non-zero polynomial $f(x_1, \dots, x_n)$ of degree d on \mathbb{F}_p^n has at most dp^{n-1} roots.



Proof For every $\mathbf{a}, \mathbf{z} \in \mathbb{F}_p^n$ and $\mathbf{z} \neq \mathbf{0}$, we consider a line

$$L = \{\mathbf{a} + t\mathbf{z} : t \in \mathbb{F}_p\},$$

and the restriction of f on L , denoted by $f_L(t)$. The Fact 3.2 implies that $f_L(t)$ is a univariate polynomial (of t) of degree at most d and leading coefficient $f_d(\mathbf{z})$, where $f_d(\mathbf{z})$ is the homogeneous degree d part of f . Since f_d is non-zero and $d < p$, by Lemma 3.3, there exists $\mathbf{z} \neq \mathbf{0}$ such that $f_d(\mathbf{z}) \neq 0$, which implies that $f_L(t)$ is a non-zero polynomial of degree d . Thus, $f_L(t)$ can have at most d roots on L , implying that the polynomial f can vanish on at most d points of the line L . How many lines the field should have in same direction \mathbf{z} ? We associate with each vector $\mathbf{a} \in \mathbb{F}_p^n$ the line $L_{\mathbf{a}} = \{\mathbf{a} + t\mathbf{z} : t \in \mathbb{F}_p\}$ in direction \mathbf{z} through \mathbf{a} . Then $L_{\mathbf{a}} \cap L_{\mathbf{b}} = \emptyset$ as long as $\mathbf{b} \notin L_{\mathbf{a}}$. Since $\mathbf{z} \neq \mathbf{0}$, each line $L_{\mathbf{a}}$ contains $|L_{\mathbf{a}}| = p$ points. Hence, we can partition \mathbb{F}_p^n into $\frac{p^n}{p} = p^{n-1}$ lines. Since the number of roots of f on each of the lines $L_{\mathbf{a}}$ is at most d , the total number of roots of f cannot exceed dp^{n-1} , as claimed.

Remark

- Schwartz–Zippel Lemma says that multivariate low-degree polynomials cannot have too many roots.
- The bound is sharp! (We can consider a example that f depends only on x_1).
- It is useful for polynomial identity testing.
- We can also bound the number of roots of f in a finite subset.

The following probabilistic version of Lemma 3.4 bounds the probability that a non-zero multivariate polynomial will have roots at randomly selected test points.

Lemma 3.5

Suppose that $f(x_1, \dots, x_n)$ is a nonzero polynomial of degree d over a field \mathbb{F} and $S \subseteq \mathbb{F}$ is a non-empty finite subset. Let r_1, \dots, r_n be random elements selected uniformly and independently from S . Then

$$\Pr[f(r_1, \dots, r_n) = 0] \leq \frac{d}{|S|}.$$



Proof Suppose that f is a non-zero polynomial. We use induction on n , the number of variables of f . The statement is true for $n = 1$ since the number of roots of f does not exceed its degree. Now let $n \geq 2$ and write f as a polynomial in x_1 , which means that

$$f(x_1, \dots, x_n) = \sum_{i=0}^d x_1^i \cdot f_i(x_2, \dots, x_n).$$

Since f is a nonzero polynomial, we take max i , then f_i is a nonzero polynomial and $\deg(f_i) \leq d - i$. Let r_1, \dots, r_n be random elements selected uniformly and independently from S . We define two events as follows:

$$A = \{f(r_1, \dots, r_n) = 0\}, \text{ and } B = \{f_i(r_2, \dots, r_n) = 0\}.$$

Then we shall upper bound

$$\begin{aligned} \Pr(A) &= \Pr(A \cap B) + \Pr(A \cap B^c) \\ &= \Pr(A \mid B) \Pr(B) + \Pr(A \mid B^c) \Pr(B^c) \\ &\leq \Pr(B) + \Pr(A \mid B^c) \\ &\leq \frac{d-i}{|S|} + \frac{i}{|S|} = \frac{d}{|S|}, \end{aligned}$$

where the first term is from the induction hypothesis and the second term is from the basic fact that $f(x_1, r_2, \dots, r_n)$ is a nonzero polynomial of degree i .

3.2 Testing polynomial identity & Existence of Perfect Matching

In this section, Let's see how to use the Schwartz–Zippel Lemma.

3.2.1 Polynomial identity testing

Problem 3.1 How do we test whether two given polynomials on \mathbb{F}_p^n are the same, i.e. $f = g$?

One can check the coefficients of all monomials, but there are too many terms. (The number of monomials will be n^d , if $\deg(f) = d$.)

Schwartz–Zippel Lemma can be used to design efficient *Probabilistic Algorithm*: Given two polynomial f and g of degree d on n variables. Suppose that $f \neq g$. Then $f - g$ is a nonzero polynomial of degree at most d .

- We pick *random* numbers in place of the variables and compute the value of the polynomial. That is, pick $\mathbf{r} = (r_1, \dots, r_n)$, where every r_i is random element selected uniformly and independently from \mathbb{F}_p . Then Lemma 3.5 tells us that

$$\Pr[(f - g)(\mathbf{r}) = 0] \leq \frac{d}{p}.$$

- Repeat the above process k times and we have

$$\Pr[(f - g) \text{ vanishes on all } k \text{ choices}] \leq \left(\frac{d}{p}\right)^k.$$

So if $\frac{d}{p} < \frac{1}{2}$, then we only need $k = O(\log \frac{1}{\epsilon})$ to make $\left(\frac{d}{p}\right)^k < \frac{1}{\epsilon}$.

- We only need to test $O(\log \frac{1}{\epsilon})$ to have probability at least $1 - \epsilon$ to test correctly. It means that we have two outcomes:
 - ★ If there exists some outcome $\neq 0$, then we get that $f \neq g$.
 - ★ If all outcomes = 0, then either $f = g$ or $f \neq g$. If $f \neq g$, the above probability should be at most ϵ .

3.2.2 Testing existence of Perfect matching

Problem 3.2 How do we test whether a given graph G has a perfect matching?

Idea The graph G has **NO** perfect matching \iff its corresponding polynomial is 0-polynomial.

Let us consider the special case of Problem 3.2 when the graph G is bipartite.

Definition 3.6. Edmonds matrix

Let $G = (U \cup V, E)$ with $U = \{u_1, \dots, u_n\}$ and $V = \{v_1, \dots, v_n\}$. Its Edmonds matrix A is the $n \times n$ matrix with variables entries corresponding to edges of G . That is

$$A = (a_{ij})_{n \times n}, \quad a_{ij} = \begin{cases} x_{ij}, & \text{if } u_i v_j \in E(G); \\ 0, & \text{otherwise.} \end{cases}$$



- $\det(A)$ is a polynomial with $e(G)$ many variables of degree at most n .
- We can identify the perfect matching of G with permutation $\pi \in S_n$:

$$\{(u_1, v_{\pi(1)}), \dots, (u_n, v_{\pi(n)})\}.$$

- Note that

$$\det(A) = \sum_{\pi \in S_n} \text{sgn}(\pi) a_{1,\pi(1)} \cdots a_{n,\pi(n)} = \sum_{\pi \in PM(G)} \text{sgn}(\pi) x_{1,\pi(1)} \cdots x_{n,\pi(n)},$$

where each term in sum $a_{1,\pi(1)} \cdots a_{n,\pi(n)}$ is nonzero only when π is a perfect matching.

Lemma 3.7

The graph G has no perfect matching $\iff \det(A)$ is 0-polynomial.



Exercise 3.1 Prove Lemma 3.7. (Hint: (\implies) is trivial, and (\impliedby) can try to prove contrapositive.)

3.3 Discrete Kakeya Problem

A famous unsolved problem in mathematics is the Kakeya conjecture in geometric measure theory. This conjecture is descended from the following question asked in 1917 by Japanese mathematician Soichi Kakeya: What is the smallest set in the plane in which one can rotate a needle around completely?

Obviously, one can rotate a unit needle inside a disk with radius $1/2$, which has area $\frac{\pi}{4}$. Note that a necessary condition for turning a unit-length needle inside a set X is that X must contain a unit-length segment of every direction. Such set is called a *Kakeya set*. Besicovitch proved that there exists a Kakeya set of measure zero. Wolff (1999) proposed a simpler finite field analogue of this problem:

Problem 3.3 How large is a Kakeya set in finite field setting?

Definition 3.8

A set $K \subseteq \mathbb{F}_p^n$ is a Kakeya set, if K contains a line in every direction, namely for any nonzero vector $z \in \mathbb{F}_p^n$ there exists a vector $\mathbf{a} \in \mathbb{F}_p^n$ such that the line

$$\{\mathbf{a} + tz : t \in \mathbb{F}_p\} \subseteq K.$$



Dvir (2009) used a surprisingly simple and elegant application of the polynomial method to prove the Kakeya conjecture on finite field.

Theorem 3.9

Let $K \subseteq \mathbb{F}_p^n$ be a Kakeya set. Then

$$|K| \geq \binom{n+p-1}{n}.$$



Remark

- If we think of n fixed and $p \rightarrow \infty$, then we get $\binom{n+p-1}{n} \sim \frac{p^n}{n!}$. In contrast to the Euclidean case, we will see that Kakeya set in finite field is dense.

Let us start with giving a lower bound on the dimension of low degree multivariate polynomial on \mathbb{F}_p^n .

Lemma 3.10

Given a set $A \subseteq \mathbb{F}_p^n$ with $|A| \leq \binom{n+d}{d}$, there exists a non-zero polynomial $f \in \mathbb{F}_p[x_1, \dots, x_n]$ of degree at most d vanishing on A .



Proof For every $f \in \mathbb{F}_p[x_1, \dots, x_n]$, we write its linear combination of monomials of degree at most d in x_1, \dots, x_n . That is

$$f(x_1, \dots, x_n) = \sum_{\alpha_1 + \dots + \alpha_n \leq d, \alpha_i \geq 0} c_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

We want to count the number of such monomials parts. Using the fact that the number of ways to distribute $n + d$ sweets to n children in a fair way is $\binom{n+d-1}{d-1} = \binom{n+d-1}{d}$ and we can show that the number of integer solutions to the equation

$$x_1 + \dots + x_n = d$$

under the condition that $x_i \geq 0$ for all $i = 1, \dots, n$, is $\binom{n+d-1}{d}$. Then by Pascal Triangle: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, we have

$$\sum_{i=0}^d \binom{n+i-1}{i} = \binom{n+d}{n}.$$

Thus, the number of distinct monomials of degree at most d is $\binom{n+d}{d}$. Next, we treat each coefficient c_α as unknowns and each $\in A$ as constraints. Then the number of unknowns is larger than the number of constraints (that is, the vector space $\mathbb{F}_p^{|A|}$ of all functions $g : A \rightarrow \mathbb{F}_p$ has dimension $|A| < \binom{n+d}{d}$). Thus, there exists a nonzero solution of this linear systems, meaning that the polynomial f vanishes on A .

Lemma 3.11

Let $f \in \mathbb{F}_p[x_1, \dots, x_n]$ be a nonzero polynomial of degree at most $d < p$ vanishing on a Kakeya set K . Then its degree- d part f_d is a nonzero polynomial vanishing everywhere on \mathbb{F}_p^n . In other words, if f vanishes on a Kakeya set K , then f is the 0-polynomial. ♥

Proof The argument is similar to that in the proof of Lemma 3.4. Let $\mathbf{z} \in \mathbb{F}_p^n \setminus \{\mathbf{0}\}$ be an arbitrary direction. As K is a Kakeya set, K contains a line $L = \{\mathbf{a} + t\mathbf{z} : t \in \mathbb{F}_p\}$ for some $\mathbf{a} \in \mathbb{F}_p^n$. Recall the restriction of f on L , denoted by $f_L(t)$. The Fact 3.2 implies that $f_L(t)$ is a univariate polynomial (of t) of degree at most d and leading coefficient $f_d(\mathbf{z})$, where $f_d(\mathbf{z})$ is the homogeneous degree d part of f . As f vanishes on K , f_L vanishes on L . But $\deg(f_L) = d < p = |L|$. Hence, f_L is 0-polynomial and $f_d(\mathbf{z}) = 0$. Obviously, we have $f_d(\mathbf{0}) = 0$. Then f_d vanishes on \mathbb{F}_p^n , as claimed.

Proof of Theorem 3.9. Take K Kakeya set and suppose that $|K| < \binom{n+p-1}{n}$. Then, by Lemma 3.10, there exists a nonzero polynomial f of degree at most $p - 1$ vanishing on K , which contradicts Lemma 3.11.

Remark

- Dvir [10] proved that every Kakeya set in \mathbb{F}_p^n is of size at least $\frac{p^n}{n!}$.
- Sarof and Sudan [20] show that there exists a Kakeya set of size $2^{-n+1}p^n + O_n(p^{n-1})$.
- Dvir, Kopparty, Saraf, and Sudan [11] give a lower bound that $|K| \geq (2 - \frac{1}{p})^{-n+1}p^n$ (★).
- Very recently, Bukh and Chao [5] improve the lower bound (★) by a factor of $2 - \frac{1}{p}$, thereby closing the factor-of-two gap in all dimensions.

Theorem 3.12. [Bukh–Chao]

The size of every Kakeya set $K \subseteq \mathbb{F}_p^3$ is

$$|K| \geq \left(2 - \frac{1}{p}\right)^{-(n-1)} p^n.$$



We begin by presenting a proof of a slightly weaker bound in dimension 3. Though this proof does *not* seem to generalize to the $n > 3$, it illustrates one of the ideas used in the general case.

Recall Dvir’s method: Consider a vector space of polynomials of degree less than p and show that subspace vanishing on a Kakeya set K must be trivial.

- Let $U = \{\sum_{\alpha_1+\dots+\alpha_n \leq p, \alpha_i \geq 0} c_\alpha x^\alpha : c_\alpha \in \mathbb{F}_p\}$.
- $U' \subseteq U$ vanishes on $K \implies U'$ is trivial.
- $|K| \geq \text{codim } U' = \dim U = \binom{n+p-1}{n}$.

Theorem 3.13. [Bukh–Chao]

Let $K \subseteq \mathbb{F}_p^3$ be a Kakeya set. Then

$$|K| \geq \frac{1}{4}(p^3 + p).$$



- Let

$$A = \{(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{Z}_{\geq 0}^3 : \alpha_1 + \alpha_2 + \alpha_3 < 2p, \alpha_1, \alpha_2 < p\},$$

and consider the vector space of polynomials with the monomials indexed by A ,

$$V = \left\{ \sum_{\alpha \in A} c_\alpha x^\alpha : c_\alpha \in \mathbb{F}_p \right\}.$$

- Considering V has some similarity to Green’s twist on corner-free set.
- Bukh and Chao do so by considering a larger vector space of polynomials and subspace that vanishes on a Kakeya set of high order.

Definition 3.14

A polynomial f on \mathbb{F}_p^3 vanishes at $z \in \mathbb{F}_p^3$ to order 2, if $f(z) = 0$ and $\nabla f(z) = 0$. (That is, $\frac{\partial f}{\partial x_1}(z) = 0$, $\frac{\partial f}{\partial x_2}(z) = 0$ and $\frac{\partial f}{\partial x_3}(z) = 0$.)



Lemma 3.15

Let K be a Kakeya set in \mathbb{F}_p^3 . If a polynomial $f \in V$ vanishes to order 2 at every point of K , then f is the 0-polynomial.



Proof Suppose that $f \in V$ is a nonzero polynomial of degree $d < 2p$ vanishing on K to order 2. Then f_d is nonzero. Let $z \in \mathbb{F}_p^n \setminus \{0\}$ be an arbitrary direction. As K is a Kakeya set, K contains a line $L = \{a + tz : t \in \mathbb{F}_p\}$ for some $a \in \mathbb{F}_p^n$. Recall the restriction of f on L , denoted by $f_L(t)$. The Fact 3.2

implies that $f_L(t)$ is a univariate polynomial (of t) of degree at most d and leading coefficient $f_d(z)$, where $f_d(z)$ is the homogeneous degree d part of f . Since f vanishes at every point of L to order 2, the polynomial f_L vanishes at all points of \mathbb{F}_p to order 2. Because $\deg(f_L) = d < 2p = 2|L|$, this implies that f_L is the 0-polynomial. So its leading coefficient $f_d(z) = 0$.

Let $g(x_1, x_2) = f_d(x_1, x_2, 1)$ be a polynomial of degree less than p in each of x_1 and x_2 . Then g vanishes on \mathbb{F}_p^2 . Write g_i as $g(x_1, x_2) = \sum_{i=0}^{p-1} x_1^i g_i(x_2)$ where $\deg(g_i) < p$ (as $\alpha_2 < p$). Since f vanishes identically on \mathbb{F}_p^2 , g vanishes on \mathbb{F}_p . But $\deg(g_i) < p$, this means that g_i is 0-polynomial, and this implies that f_d is zero as well, contrary to $\deg(f) = d$.

Proof of Theorem 3.13. Let $V' \subseteq V$ be a subspace vanishing to order 2 on K . Then $\text{codim}V' \leq 4|K|$. Lemma 3.15 implies that V' is trivial. Thus

$$4|K| \geq \text{codim}V' = \dim V = |A| = \sum_{\alpha_1, \alpha_2=0}^{p-1} (2p - \alpha_1 - \alpha_2) = p^3 + p^2.$$

3.4 Joint Theorem

In this section, we will show another important application of the polynomial method in discrete geometry.

Definition 3.16

Given a set of lines in \mathbb{R}^d , a joint formed by these lines is a point that lines on d given lines, whose directions are linearly independent.



For example: Suppose that x, y and z are three given lines in \mathbb{R}^3 . The origin is a joint.

Indeed, what we care about is that if we give a set of lines, how many joints are there? Namely, we want to bound the number of joints formed by a set of lines. The Joint Theorem gives a tight upper bound for any given a set of lines.

It has long been conjectured that the correct upper bound on the number of joints in \mathbb{R}^3 is $O(n^{3/2})$. Guth and Katz [Guth_2010] have settled the conjecture in the affirmative, showing that the number of joints in \mathbb{R}^3 is $O(n^{3/2})$. Lately, Kaplan, Sharir and Shustin [Kaplan_2010], and Quilodr an [Ren_2020] independently extended this result into \mathbb{R}^d . In 2020, Carbery and Iliopoulou [Anthony_2020] solved this in all filed \mathbb{F}^d . In what follows, $x \lesssim_d (y)$ is the same as $x = O_d(y)$.

Theorem 3.17. [Joint Theorem]

The number of joints formed by N lines in \mathbb{R}^d is $\lesssim_d N^{d/(d-1)}$.



Remark Take S hyperplanes in \mathbb{R}^d in general position. Then

- Any $d - 1$ of them intersect at a line and the number of lines is $\binom{s}{d-1}$.
- Any d of them intersect at a point (joint) and the number of joints is $\binom{s}{d} \gtrsim_d N^{d/(d-1)}$.

Idea Suppose a non-zero low-degree polynomial f is heavily incident to given lines. This means that f vanish on all these lines. Then we will derive some information about ∇f to get a contradiction.

Proof Let J be the set of joints formed by a set L of lines. This theorem immediately follows from the following claim. We see that if this claim is true, then take out one such "light" line at a time, each time we remove $\lesssim_d |J|^{1/d}$ joints. This implies $|J| \lesssim_d N |J|^{1/d}$, as required. Therefore, it suffices to show the following claim.

Claim There exists a line $l \in L$ containing $\lesssim_d |J|^{1/d}$ points.

Proof of Claim Suppose not, i.e., for any line $l \in L$, the number of joints on L is $\gtrsim_d |J|^{1/d}$. Let us take a non-zero polynomial f with minimal degree in \mathbb{R}^d that vanishes on J . Then the number of constraint is at most $|J|$. This implies that the degree of F is $\lesssim_d |J|^{1/d}$.

For each $l \in L$, consider the restriction f_l . Because

$$\text{deg} f_l \leq \text{deg} f \lesssim_d |J|^{1/d} < |J \cap l|,$$

f_l vanishes on the whole line l . This yields that f vanishes on all lines L . Thus, for each $p \in J$, let l_1, \dots, l_d be the set of d linearly independent lines going through P . Let v_i be the direction of l_i (See Figure 7.3).

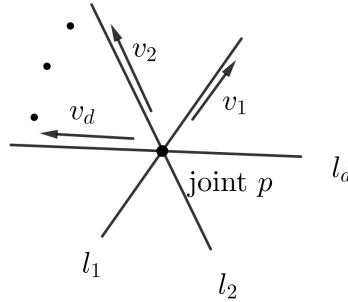


Figure 3.1

Since f vanishes on l_i , we have $\nabla f(p) \cdot v_i = 0$ (directional derivation). Because $\{v_i\}_{i \in d}$ linearly independent, we get $\nabla f(p) = \vec{0}$. This implies $\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{pmatrix}$ vanishes on J . In particular, all $\frac{\partial f}{\partial x_i}$ vanishes on J . Note that there is at least one $\frac{\partial f}{\partial x_i}$ such that it is not a 0-polynomial. Otherwise, f is a non-zero constant, which contradicts the fact that f vanished on J . Now, $\frac{\partial f}{\partial x_i}$ is a non-zero polynomial with a lower degree than f and vanishing on J , contradicting the choice of f .

There exists another different proof without taking derivative by Zhang. Before we launch out the proof, we first give a definition about ordinary or special joints on a line with respect to a polynomial f .

Definition 3.18

Let f be a polynomial. For each joint p in a given line l , take an affine linear map

$$T_0 : \begin{cases} l \mapsto x_d\text{-axis} \\ p \mapsto 0 \\ f \mapsto f \circ T_0^{-1} \end{cases}$$

Then $f \circ T_0^{-1}(0) = f(p) = 0$. We say that p is ordinary on l if the lowest homogeneous part of $f \circ T_0^{-1}$ is independent of x_d ; Otherwise we say it special.



This definition is independent of choices of T_0 , because two such maps $T_0 = T'_0 \circ T$ only differ by another map T that is a scaling when restricted to x_d -axis.

Exercise 3.2 Let x, y and $z - a$ be three polynomial in \mathbb{R}^3 , where a is a constant. What is the lowest homogeneous part? ($-xya$ is independent of z .)

Idea Instead of caring about polynomial vanish at a given point, Zhang rather looks at the Taylor series around that point.

Proof Let L be the set of N given lines and J a set of joints. We first take a non-zero polynomial f with degree $\lesssim_d |J|^{1/d}$ and vanishing on J . Then we shall see the following two statements.

(i) Every point p is special on at least one line passing through it.

(ii) There exist at most $\deg f$ many special joints on any line $l \in L$.

Next we use the double-counting method on the number of special joints. On the one hand, the number of special joints is at least $|J|$ by (i). On the other hand, by (ii), the number of special joints is at most $N \cdot \deg f \lesssim_d N |J|^{1/d}$, as required. So we just need to prove these two statements.

Let us first prove (i). For any $p \in J$, take an affine linear map

$$T_0 : \begin{cases} p \mapsto 0 \\ \text{lines through } p \mapsto x_i\text{-axis} \end{cases}$$

(See Figure 3.2). This implies that the lowest homogeneous part of $f \circ T_0^{-1}$ depends on some x_i , since otherwise, it is a constant and $f \circ T_0^{-1}(0) \neq 0$, a contradiction. Hence p is special on the corresponding line.

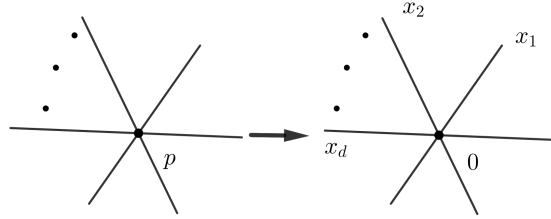


Figure 3.2

For (ii), we show the following claim.

Claim If p is special on l , for any $T_0 : \begin{cases} l \mapsto x_d\text{-axis} \\ p \mapsto 0 \end{cases}$, then $f \circ T_0^{-1}(x) = \sum_{\alpha} x^{(\alpha,0)} f_{\alpha}(x_d)$, where $\alpha = (\alpha_1, \dots, \alpha_{d-1})$. Further, for any α with $|\alpha| = \alpha_1 + \dots + \alpha_{d-1}$ and $f_{\alpha}(x_d)$ not a 0-polynomial, then $f_{\alpha}(x_d)$ vanished at $x_d = 0$.

Let us see how the claim prove (ii). For any map $T : \begin{cases} l \mapsto x_d\text{-axis} \\ p \mapsto (0, 0, \dots, 0, p_T) \end{cases}$, we have $f \circ T^{-1}(x) = \sum_{\alpha} x^{(\alpha,0)} f_{\alpha}(x_d - p_T)$. Then for any α with $|\alpha|$ minimal and $f_{\alpha} \neq 0$, p_T is a root of f_{α} . This implies that the number of special p on l is at most $\deg f_{\alpha} \leq \deg f \lesssim_d |J|^{1/d}$.

Proof of Claim Suppose for a contradiction that $f_{\alpha}(0) \neq 0$. Notice that $f_{\alpha}(0)$ is the constant term of f_{α} . By minimality of α , we get that $f_{\alpha}(0)x^{(\alpha,0)}$ is in lowest homogeneous part of $f \circ T_0^{-1}$. This implies that other terms in lowest parts are all independent of x_d . Otherwise, some of power of x_1, \dots, x_{d-1} is strict less than $|\alpha|$, contradicting the choice of α .

Chapter 4 Capset problem and Slice rank

Definition 4.1

A capset $A \subseteq \mathbb{F}_3^n$ is a set with no line $\{x, x+r, x+2r\}$, where $x, r \in \mathbb{F}_3^n$ and $r \neq 0$.



The capset problem is to bound the maximum size of a capset in \mathbb{F}_3^n .

Remark

- Meshulam extended Roth's Founer argument $\leq O(\frac{3^n}{n})$.
- Croot-lev-Path, 3AP-free set in \mathbb{Z}_4^n in exponexily small.
- Ellenberg-Gijwijt independently showed that a variation of *CLP* technique $\Rightarrow O(2.756^n)$.
- Lower bound: Edel showed the lower bound $\geq (2.2174)^n$.
- Tao: a symmetrical variation of *CLP*, which treats all 3 variables the same.

4.1 Slice rank

let us start with the basic two variable polynomial. Let X, Y be two finite sets and \mathbb{F} a field. Let $f : X \times Y \rightarrow \mathbb{F}$ be a two-variable function of rank one if there exist two single variable functions $u : X \rightarrow \mathbb{F}$ and $v : Y \rightarrow \mathbb{F}$ such that $f(x, y) = u(x)v(y)$ for any $(x, y) \in X \times Y$. In general, the rank of a two-variable function is the minimal number of rank one functions whose linear span contains it. i.e.

$$f(x, y) = \sum_{i=1}^k u_i(x)v_i(y)$$

Further, if $f : X \times Y \times Z \rightarrow \mathbb{F}$ is a three variable function, naturally,

$$f(x, y, z) = \sum_{i=1}^k u_i(x)v_i(y)w_i(z).$$

Instead, f has slice rank one if $f(x, y, z) = u(x)v(y, z)$ for each $(x, y, z) \in X \times Y \times Z$. In general, the slice rank of f is the minimal number of slice rank one functions needed to write f as a linear combination.

This means

$$f(x, y, z) = \sum_{i=1}^{r_1} u_i(x)v_i(y, z) + \sum_{i=r_1+1}^{r_2} u_i(y)v_i(x, z) + \sum_{i=r_2+1}^r u_i(z)v_i(x, y).$$

In \mathbb{R}^2 , if we have a diagonal matrix with centry $f(x, y)$, which means $f(x, y) \neq 0$ if and only if $x = y$, then rank of this function is the number of non-zero diagonal entries.

Frow now on, we use $sr(f)$ to denote the slice rank of f . Let $\mathbb{1}_S$ denote indicator function for a set S . This means $\mathbb{1}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$. If $S = a$, then we write $\mathbb{1}_a$ instead of $\mathbb{1}_{\{a\}}$.

Lemma 4.2

Let X be a finite set and $A \subseteq X$. Let \mathbb{F} be field and $f : X^3 \rightarrow \mathbb{F}$ three variable function such that $f(x, y, z) \neq 0$ if and only if $x = y = z$ and $x \in A$. Then $sr(f) = |A|$.



Proof Since $f(x, y, z) \neq 0$ if and only if $x = y = z \in A$, we have $f(x, y, z) = \sum_{a \in A} \mathbb{1}_a(x) \mathbb{1}_a(y) \mathbb{1}_a(z) f(a, a, a)$. This implies $sr(f) \leq |A|$. Therefore, it is left to show $sr(f) \geq |A|$. Suppose $sr(f) = r$ and write $f(x, y, z) = \sum_{i=1}^{r_1} u_i(x) v_i(y, z) + \sum_{i=r_1+1}^{r_2} u_i(y) v_i(x, z) + \sum_{i=r_2+1}^r u_i(z) v_i(x, y)$, without loss of generality assume $r_1 > 0$.

We shall construct a two variable function $g : Y \times Z \rightarrow \mathbb{F}$, that is, $g(y, z) = \sum_{x \in X} h(x) f(x, y, z)$, where $h : X \rightarrow \mathbb{F}$ and show the rank of g (i) $\geq |A| - r_1$; (ii) $\leq r - r_1$. This yields $r \geq |A|$, as desired.

Claim There exists a function $h : X \rightarrow \mathbb{F}$ such that $\sum_{x \in X} h(x) u_i(x) = 0$ and the number of zero entries in h is at most r_1 .

Proof of Claim Consider the vector space of function $h : X \rightarrow \mathbb{F}$ orthogonal to all u_i in V . This implies $\dim V \geq |X| - r_1$. Let $M = \begin{pmatrix} \dots & u_1 & \dots \\ \dots & u_2 & \dots \\ \dots & \dots & \dots \\ \dots & u_{r_1} & \dots \end{pmatrix}$. Then $Mh = 0$ for any $h \in V$. Now, we want to find h with at most r_1 zeros. We have $M_{r_1, |X|} h_{|X|, 1} = O_{r_1, 1}$. Then the number of variables is $|X|$ constraint is r_1 and freedom is at most $|X| - r_1$. This implies that there exists a $h \in V$ with at least $|X| - r_1$ non-zero entries.

Claim g is diagonal with diagonal entry $h(a) f(a, a, a)$ for any $a \in A$, out of which is at least $|A| - r_1$ non-zero.


Proof of Claim Recall $f(x, y, z) \neq 0 \Leftrightarrow x = y = z = a \in A$. This yields that $g(y, z) = 0$ for any $y \neq z$, and $g(y, z) = 0$ for any $y = z \notin A$.

This claim implies (i). For (ii),

$$g(y, z) = \sum_{x \in X} h(x) f(x, y, z) = \sum_{x \in X} h(x) \left[\sum_{i=1}^{r_1} u_i(x) v_i(y, z) + \sum_{i=r_1+1}^{r_2} u_i(y) v_i(x, z) + \sum_{i=r_2+1}^r u_i(z) v_i(x, y) \right]. \quad (4.1)$$


Since $h \perp u_1$, the first term of the above equality equals 0. The second term equals $\sum_{x \in X} h(x) \sum_{i=r_1+1}^{r_2} u_i(y) v_i(x, z)$, that is, $\sum_{i=r_1+1}^{r_2} u_i(y) \sum_{x \in X} h(x) v_i(x, z)$. This implies the rank of g is at most $r - r_1$.

Lemma 4.3. [Rank of diagonal hyper-matrices]

Let X be a finite set, there is $A \subset X$. Let \mathbb{F} be a field. Let $f : X^3 \rightarrow \mathbb{F}$, where $f(x, y, z) \neq 0$ if and only if $x = y = z \in A$, then the slice rank of f $sr(f) = |A|$. 


Often we want to (upper) bound the cardinality of a set A , with certain forbidden structure, i.e. s -free. Define a polynomial $f : A^3 \rightarrow \mathbb{F}$, s.t. it is non-vanishing with $f(x, y, z) \neq 0$ if and only if $x = y = z \in A$. Then there is slice rank of f $sr(f) = |A|$. Ellenberg and Gijswijt[12] gave a better upper bound.

Theorem 4.4. [Ellenberg-Gijswijt]

For any capset $A \in \mathbb{F}_3^n$, i.e., there is no line $\{x, x+r, x+2r\}$ with $r \neq 0$, then $|A| \leq O(2.756^n)$. 

To prove the theorem, we first give an observation for a necessary and sufficient condition on $x = y = z = 0$. Then we define a suitable f with utilizing the opposite direction of the observation.

Observation 4.5

For $x, y, z \in \mathbb{F}_3^n$, $x + y + z = 0$ if and only if $x = y = z = 0$ or x, y, z form a line $\{x, x+r, x+2r\}$ 

So for a capset A and for any $x, y, z \in A$, $x + y + z = 0$ if and only if $x = y = z = 0$. Recall that if $x \in \mathbb{F}_p$, $x \neq 0$, then $x^{p-1} = 1$. Then if $x + y + z \neq 0$, there exists $i \in [n]$ s.t. $x_i + y_i + z_i \neq 0$. And $1 - (x_i + y_i + z_i)^2 = 0$.

Definition 4.6

Define $f : A^3 \rightarrow \mathbb{F}_3$ to be $f(x, y, z) = \prod_{i=1}^n (1 - (x_i + y_i + z_i)^2)$.



Then $f(x, y, z) \neq 0$ if and only if $x = y = z \in A$. According to the rank of diagonal hyper-matrices Lemma, there is $sr(f) = |A|$. The last preparation before proof is offering a Claim about contraiting $sr(f)$ with the number of 0,1,2-vectors.

Claim $sr(f) \leq 3R$, where $R = \sum \frac{n!}{a!b!c!}$ denotes the number of 0,1,2-vectors v of length n , with $a, b, c \geq 0$, $a + b + c = n$ and $b + 2c \leq \frac{2n}{3}$ s.t. $v_1 + \dots + v_n \leq \frac{2n}{3}$.

Finally, we finish the proof of the Ellenberg-Gijswijt theorem.

Proof Note that $f(x, y, z) = \prod_{i=1}^n (1 - (x_i + y_i + z_i)^2)$ is a $3n$ -variable polynomial with total degree $\leq 2n$ and the degree of each variable (x_i, y_i, z_i) is no more than 2. Thus, f is a linear combination of monomials of the form $x^\alpha y^\beta z^\gamma = \prod_{j=1}^n x_j^{\alpha_j} \prod_{j=1}^n y_j^{\beta_j} \prod_{j=1}^n z_j^{\gamma_j}$, where α, β, γ are 0,1,2-vectors of length n and $|\alpha| + |\beta| + |\gamma| \leq 2n$. Then for any such $x^\alpha y^\beta z^\gamma$, by pigeonhole principle, one of the exponent is no more than $\frac{2n}{3}$, i.e. either $|\alpha| \leq \frac{2n}{3}$ or $|\beta| \leq \frac{2n}{3}$ or $|\gamma| \leq \frac{2n}{3}$.

Besides, we have $f = f_1 + f_2 + f_3$, with
$$\begin{cases} f_1(x, y, z) = \sum_{|\alpha| \leq \frac{2n}{3}} C_\alpha x^\alpha g_\alpha(y, z), \\ f_2(x, y, z) = \sum_{|\beta| \leq \frac{2n}{3}} C_\beta y^\beta g_\beta(x, z), \\ f_3(x, y, z) = \sum_{|\gamma| \leq \frac{2n}{3}} C_\gamma z^\gamma g_\gamma(x, y). \end{cases}$$
 Note that $sr(f) =$

$\sum_{i=1}^3 sr(f_i)$ where $sr(f_i)$ is no more than the number of 0, 1, 2-vectors of length n with coordinates sum up to $\frac{2n}{3}$. Let a, b, c be the number of 0s, 1s, 2s in such vectors. Then we have $R = \sum_{a,b,c} \frac{n!}{a!b!c!}$, where a, b, c satisfy $a, b, c \geq 0$, $a + b + c = n$ and $b + 2c \leq \frac{2n}{3}$ at the same time.

To estimate R , say $a = (\alpha + o(1))n$, $b = (\beta + o(1))n$, $c = (\gamma + o(1))n$. Stirling's formula tells us that $\frac{n!}{a!b!c!} = e^{h(\alpha, \beta, \gamma)n + o(n)}$, with $h(\alpha, \beta, \gamma) = \alpha \log \frac{1}{\alpha} + \beta \log \frac{1}{\beta} + \gamma \log \frac{1}{\gamma}$. Set $N = e^{n(X + o(1))}$, where $X = \max h(\alpha, \beta, \gamma)$, s.t. $\alpha, \beta, \gamma \geq 0$, $\alpha + \beta + \gamma = 1$, and $\beta + 2\gamma \leq \frac{2}{3}$. By Lagrange multiplier, the max attains

$$h(\alpha, \beta, \gamma) \approx 1.01345 \text{ when } \begin{cases} \alpha = \frac{32}{3(15 + \sqrt{33})} \\ \beta = \frac{4(\sqrt{33} - 1)}{3(15 + \sqrt{33})} \\ \gamma = \frac{(\sqrt{33} - 1)^2}{6(15 + \sqrt{33})} \end{cases} . \text{ And then we finish the proof.}$$

4.2 Sunflower

Definition 4.7

Let $k \in \mathbb{N}$, sets A_1, \dots, A_k form a k -sunflower, if they have common pairwise intersection, that is, \exists a set C , subject to, \forall distinct $i, j \in [k]$, $A_i \cap A_j = C$

**Conjecture 4.8. [Erdős- Szemerédi sunflower conjecture]**

$\forall k, \exists c = c(k) < 2$, subject to, $\forall A \subseteq [n]$ with no k -sunflower has size $|A| \leq C^n$, we will see a solution for 3-sunflower.



Naslund and Sawin[18] prove the following theorem for explicit bounds.

Theorem 4.9. [Naslund- Sawin]

$\forall A \subseteq 2^{[n]}$, a 3-sunflower free collection of subsets of $[n]$, then $|A| \leq 3n \sum_{k \leq \frac{n}{3}} \binom{n}{k} \leq \left(\frac{3}{2}\right)^{(1+o(1))n}$



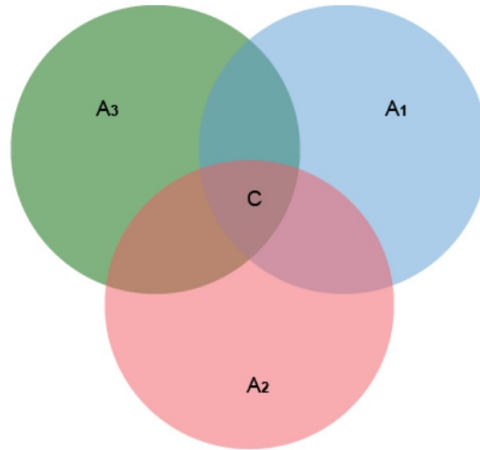


Figure 4.1

$$\frac{3}{2^{\frac{3}{3}}} \approx 1.8898 < 2$$

Proof Identify $S \in A$ is equal to $\mathbb{1}_S \in \{0, 1\}^n$ indicator function. So view A as a set of vectors in $\{0, 1\}^n$.
Need to construct $f : A^3 \rightarrow F$.

$f(x, y, z) \neq 0$ is equal to $x = y = z \in A$, then [keylemma] $\Rightarrow sr(f) = |A|$.

\forall distinct 3 sets, $x, y, z \in A$, 3 – sunflower – free $\Rightarrow Y \cap Z \setminus X \neq \emptyset$

$\Rightarrow \exists i, s.t. \{x_i, y_i, z_i\} = \{0, 1, 1\}$

$\Rightarrow z - (x_i + y_i + z_i) = 0$

Problematic off-diag case:

X, X, Y and $X \subseteq Y$,

$\Rightarrow \forall i, \{x_i, y_i, z_i\}$ can only be $\{0, 0, 0\}, \{0, 0, 1\}$ or $\{1, 1, 1\}$

To fix this, partition $A = \cup_{l=1}^n A_l$, where A_l consists of all sets of size l .

Take $l \max |A_l|$, so $|A| \leq n |A_l|$.

Now define $f : A_l^3 \rightarrow F_2$ as $f(x, y, z) = \prod_{i=1}^n (2 - (x_i + y_i + z_i))$

then $f(x, y, z) \neq 0 \Leftrightarrow x = y = z$

so $sr(f) = |A_l| \geq \frac{1}{n} |A|$

f is spanned by monomials. $x^\alpha y^\beta z^\gamma$, s.t. $|\alpha| + |\beta| + |\gamma|, \alpha, \beta, \gamma \in \{0, 1\}^n$

\Rightarrow one of them $\leq \frac{n}{3}$

group monomials by this smallest degree one.

$f = f_1 + f_2 + f_3$, where $f_1 = \sum_{|\alpha| \leq \frac{n}{3}} c_\alpha X^\alpha g_\alpha(y, z)$, $f_2 = \sum_{|\beta| \leq \frac{n}{3}} c_\beta X^\beta g_\beta(y, z)$, $f_3 = \sum_{|\gamma| \leq \frac{n}{3}} c_\gamma X^\gamma g_\gamma(y, z)$.

$sr(f_i) \leq \#\{0, 1\}^n$ vectors with $\leq \frac{n}{3} = \sum_{k \subseteq y_3} \binom{n}{k}$

$\Rightarrow \frac{1}{n} |A| \leq |A_l| = sr(f) \leq 3 \sum_{k \subseteq y_3} \binom{n}{k}$

In fact, we have a theorem by Alon[4] which is useful for lower bounding.

Theorem 4.10. [Alon, Combinatorial Nullstellensatz]

Let f be a non-zero n -variable polynomial on \mathbb{F}_p^n . Let S_1, S_2, \dots, S_n be subsets of \mathbb{F}_p . Suppose f has a term $x_1^{k_1} \dots x_n^{k_n}$ with non-zero coefficients and denote the degree of f equals $\sum_1^n k_i$. For any $i \in [n]$, $|S_i| > k_i$, which implies f cannot vanish on $S_1 \times \dots \times S_n$.



Proof

To prove the theorem, we induct on degree of f .

- Base case is degree $f = 0$, which is trivial.

- On inductive step, $\deg f > 0$, without loss of generality, assume $k_1 > 0$. Suppose f vanishes on the $\prod_{i=1}^n S_i$, pick an arbitrary $a \in S_1$ and use polynomial division to write $f(x) = (x_1 - a)g(x) + h(x_2, \dots, x_n)$. As $f(x)$ vanishes on $\{a\} \times S_2 \times \dots \times S_n$, $h(x_2, \dots, x_n)$ vanishes on $S_2 \times \dots \times S_n$, then $(x_1 - a)g(x)$ vanishes on $S_1 \times \dots \times S_n$, which implies $g(x)$ vanishes on $(S_1 \setminus \{a\}) \times \dots \times S_n$.
- f has $x_1^{k_1} \dots x_n^{k_n}$ with non-zero coefficients and $g(x)$ has $x_1^{k_1-1} \dots x_n^{k_n}$ term with non-zero coefficients and degree equals $\deg g$. A contradiction.

Combinatorial Nullstellensatz theorem is useful for lower bounding size of some set A . The strategy is to suppose $|A|$ is too small and that we want to find a low-degree polynomial f' which vanishes on a too large Cartesian product set, leading to a contradiction.

Definition 4.11

$A+B = \{a + b : a \in A, b \in B\}$ denotes the sum set of set A and B .



We have an important iniquation [Davenport, H.]

Theorem 4.12. [Cauchy-Davenport]

Let p be a prime and $A, B \subset \mathbb{F}_p$. Then $|A + B| \geq \min\{|A| + |B| - 1, p\}$.



Remark

- (i) Best possible: $A = \{0, 1, \dots, a - 1\}, B = \{0, 1, \dots, b - 1\}, A + B = \{0, 1, \dots, a + b - 2\}$
- (ii) p being a prime is necessary.

Divisibility barriers: Consider $Z_{2p}, A = B = \text{evens}, A + B = A = B$

Proof

If $|A| + |B| > P$, then $A + B = \mathbb{F}_p$. Indeed, $\forall c \in \mathbb{F}_p, C - B = \{c - b : b \in B\}, (C - B) \cup A \neq \emptyset$, then $\exists a \in A, b \in B$, subject to $c - a = b$ or $c = a + b$. Assume then $|A| + |B| \leq P$, then $|A| + |B| - 1 < P$. Suppose for contrary that $|A + B| \leq |A| + |B| - 2 < P - 1$, and there exists set $C \supseteq A + B$ of size $|A| + |B| - 2$. It is thought that we need to find 'low-deg' f vanishing on a large product set and it is natural product $= A \times B$, i.e., we want to get $f(a, b) = 0$, then $\prod_{c \in C} (x + y - c)$.

Let $f(x, y) = \prod_{c \in C} (x + y - c)$. The degree of $f = |C| = |A| + |B| - 2$. Then f vanishes on $A \times B$ by definition of f . It is left to check f has term $X^{|A|-1} Y^{|B|-1}$ with non-zero coefficient. It is true about the coefficient $\binom{|A|+|B|-2}{|A|-1} \neq 0$ in \mathbb{F}_p , because $|A| + |B| - 2 < P$ and P is a prime.

What about restricted sumset $A \hat{+} A = \{a + a' : a, a' \in A, a \neq a'\}$? We conjecture by Erdős-Heilbrow.

Theorem 4.13. [Dasilv- Hamidollne 94]

Let P prime, $A \subseteq \mathbb{F}_p$, then $A \hat{+} P = \mathbb{F}_p$ or $|A \hat{+} A| \geq 2|A| - 3$.



Exercise 4.1 Prove the theorem.

4.3 Covering cube by affine hyperplanes

An affine hyperplane is a set of vectors $H = \{x \in \mathbb{R}^n : \langle a, x \rangle = b\}$ with $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. How many affine hyperplanes we need to cover the hypercube (apart from origin: $\{0, 1\}^n \setminus \{\bar{0}\}$, where $\bar{0} = \{0, \dots, 0\}$)? If we have no further restrictions on the covering, then just two hyperplanes $H_0 = \{x \in \mathbb{R}^n : \langle e_1, x \rangle = 0\}$ and $H_1 = \{x \in \mathbb{R}^n : \langle e_1, x \rangle = 1\}$ are enough, where $e_1 = (1, 0, \dots, 0)$ is the first unit vector. So we require that the all-0 vector $\bar{0}$ remains uncovered?

Theorem 4.14. [Alon–Füredi, 1993]

Let H_1, \dots, H_m be hyperplanes in \mathbb{R}^n such that none of them contain $\bar{0}$ and they together cover all other vertices of the hypercube $\{0, 1\}^n \setminus \{\bar{0}\}$. Then $m \geq n$.



Let us recall that how do we use Theorem 4.10 to lower bound $|A|$?

Strategy: Suppose $|A|$ is too small. Then we can find a polynomial f with “low degree” (w.r.t some set S_i) such that f vanishes on too large product set $\prod_i S_i$.

Thought process:

- Here, the natural product set is $\{0, 1\}^n$, and we take each subset $S_i = \{0, 1\}$. Note that $|S_i| = 2$ implies that $x_1 x_2 \cdots x_n$ is a high order term and $\deg f = n$. We need to construct a polynomial f to vanish on $\{0, 1\}^n = \prod_i S_i$.
- For each $j \in [m]$, we write

$$H_j := \{x : \langle a_j, x \rangle = b_j\} \text{ for some } a_j, b_j \neq 0. \quad (\star)$$

We know that for every $x \in \{0, 1\}^n \setminus \{\bar{0}\}$, there exists some index $j \in [m]$ such that $x \in H_j$, i.e., $b_j - \langle a_j, x \rangle = 0$. This suggests that the polynomial $\prod_{j=1}^m (b_j - \langle a_j, x \rangle)$ vanishes on hypercube apart from $\bar{0}$.

- Then we need to add to above a polynomial

$$h = \begin{cases} -\prod_{j=1}^m b_j, & \text{if } x = \bar{0}; \\ 0, & \text{elsewhere in hypercube;} \end{cases}$$

and finally $g = (-\prod_{j=1}^m b_j) \cdot \prod_{i=1}^n (1 - x_i)$ as we desired.

Proof We write H_j as in (\star) . We consider the polynomial

$$f(x) = \prod_{i=1}^m (b_j - \langle a_j, x \rangle) - \left(\prod_{j=1}^m b_j\right) \prod_{i=1}^n (1 - x_i).$$

and $S_1 = \dots = S_n = \{0, 1\}$. Suppose $m < n$. Then $\deg f = \max\{m, n\} = n$ and the coefficient at $x_1 \cdots x_n$ is nonzero. But f vanishes on $\prod S_i$. This is a contradiction because Theorem 4.10 implies that there must be a point $x \in \{0, 1\}^n$ such that $f(x) \neq 0$.

Theorem 4.15. [Chevalley–Warning, 1935]

Let p be a prime, and let f_1, \dots, f_m be polynomials on \mathbb{F}_p^n with $\sum_{i=1}^m \deg(f_i) < n$. If there is one common root, then there is another common root.



We remark that the original statement is stronger: if there is one common root, then the number of common roots of f_1, \dots, f_m is divisible by p . In particular, if there is one common root, then there are at least p common roots.

Thought process:

- Here, the natural product set is \mathbb{F}_p^n , and we take each slice $S_i = \mathbb{F}_p$ for every $i \in [n]$.
- Suppose f_1, \dots, f_m have exactly one common root $c \in \mathbb{F}_p^n$. Then for every $c' \neq c$, there exists an index

$j \in [m]$ such that $f_j(c') \neq 0$ in \mathbb{F}_p . Then we can take a polynomial

$$f(x) = \prod_{j=1}^m (1 - f_j(x)^{p-1}) = \begin{cases} 0, & x \in \mathbb{F}_p^n \setminus c; \\ 1, & x = c. \end{cases}$$

- To make it vanish also at c , we need to add to it $-\prod_{i=1}^n (1 - (x_i - c_i)^{p-1})$.

Proof Take $S_1 = \dots = S_n = \mathbb{F}_p$ and

$$f(x) = \prod_{j=1}^m (1 - f_j(x)^{p-1}) - \prod_{i=1}^n (1 - (x_i - c_i)^{p-1}).$$

Then $\deg f = \max \left\{ (p-1) \sum_{j=1}^m \deg f_j, (p-1)n \right\} = (p-1)n$ and the coefficient at $x_1^{p-1} \dots x_n^{p-1}$ is nonzero. But f vanishes on $\prod S_i$. This is a contradiction because Theorem 4.10 implies that there must be a point $x \in \mathbb{F}_p^n$ such that $f(x) \neq 0$.

4.4 Finding regular subgraphs

Theorem 4.16. [Pyber, 1985]

Let $k \in \mathbb{N}$ and let G be an n -vertex graph with average degree $d(G) \geq 32k^2 \cdot \log n$. Then G contains a k -regular subgraph.



Theorem 4.17. [Pyber–Rödl–Szemerédi, 1995]

Let $k \in \mathbb{N}$ and let G be an n -vertex graph with average degree $d(G) \geq c_k \log \Delta(G)$. Then G contains a k -regular subgraph. In particular, there exists an n -vertex graph G with $d(G) \geq c \cdot \log \log n$ with no 3-regular subgraph.



The following sufficient condition for a graph to contain a regular subgraph was derived by Alon, Friedland and Kalai (1984) using the Combinatorial Nullstellensatz.

Theorem 4.18. [Alon–Friedland–Kalai, 1984]

Let p be a prime and let G be a loopless multigraph with $d(G) > 2p - 2$ and $\Delta(G) \leq 2p - 1$. Then G contains a p -regular subgraph.



A standard trick to view a graph algebraically is to identify G with its edge set $E(G)$.

- We will often identify $E(G)$ with elements (vectors) of $\{0, 1\}^{\binom{[n]}{2}}$ by associating a set with its characteristic vector.
- Also, we treat subgraph of G with elements (vectors) of $\{0, 1\}^{E(G)}$.
- (★) There exists a p -regular subgraph \Leftrightarrow a vector $x \in \{0, 1\}^m \setminus \{\bar{0}\}$ ($m = e(G)$) such that for every vertex $v \in V(G)$, $\sum_{v \in e} x_e \equiv 0 \pmod p$.

Thought process:

- Here, the natural product set is $\prod_{e \in E(G)} S_e$, where $S_e = \{0, 1\}$. Note that $|S_e| = 2$ implies that $\prod_{e \in E(G)} x_e$ is a high order term and $\deg f = m$. We need to construct a polynomial f to vanish on

$$\{0, 1\}^m = \prod_{e \in E(G)} S_e.$$

- Suppose no such vector of (★) exists. Then for every $x \in \{0, 1\}^m \setminus \{\bar{0}\}$, there exists a vertex $v \in V(G)$ such that $\sum_{e:v \in e} x_e \not\equiv 0 \pmod p$. It suggests that we can take a polynomial

$$g(x) = \prod_{v \in V(G)} \left(1 - \left(\sum_{e:v \in e} x_e \right)^{p-1} \right) = \begin{cases} 0, & x \in \{0, 1\}^m \setminus \{\bar{0}\}; \\ 1, & x = \bar{0}. \end{cases}$$

- To make it vanish also at $\{\bar{0}\}$, we need to add to it

$$h(x) = - \prod_{e \in E(G)} (1 - x_e) = \begin{cases} 0, & x \in \{0, 1\}^m \setminus \{\bar{0}\}; \\ -1, & x = \bar{0}. \end{cases}$$

Proof Associate each edge e of G with a variable x_e and let

$$f(x) = \prod_{v \in V(G)} \left(1 - \left(\sum_{e:v \in e} x_e \right)^{p-1} \right) - \prod_{e \in E(G)} (1 - x_e).$$

We have that $\deg f = \max\{n(p-1), m\} = m$ as $d(G) = \frac{2m}{n} > 2p-2$, and the coefficient at the highest order term $\prod_{e \in E(G)} x_e$ is nonzero. Theorem 4.10 implies that there must be a point $x \in \{0, 1\}^m$ for which $f(x) \neq 0$. But f vanishes on $\prod_{e \in E(G)} S_e$, a contradiction.

4.5 Finding zero-sum multisubset

Let $A = (a_{i,j})$ be an $n \times n$ matrix over a field \mathbb{F} . The *permanent* $\text{per}(A)$ of A is the sum

$$\text{per}(A) = \sum_{(i_1, i_2, \dots, i_n)} a_{1, i_1} a_{2, i_2} \cdots a_{n, i_n}$$

of $n!$ products, where (i_1, i_2, \dots, i_n) is a permutation of $(1, 2, \dots, n)$.

Lemma 4.19. [Permanent Lemma]

Let A be an $n \times n$ matrix over a field \mathbb{F} with $\text{per}(A) \neq 0$. Let $b = (b_1, \dots, b_n) \in \mathbb{F}^n$ and S_1, \dots, S_n be subsets of \mathbb{F} with $|S_i| = 2$ for each $i \in [n]$. Then there exists a vector $x \in \prod_{i=1}^n S_i$ such that $Ax - b$ has no zero coordinate, i.e., $(Ax)_i \neq b_i$ for every $i \in [n]$. ♥

Proof Exercise!

Theorem 4.20. [Erdős–Ginzburg—Ziv, 1961]

Let p be a prime and let A be a multiset of \mathbb{Z}_p with size $2p-1$. Then there exists a submultiset of size p whose element sum up to $0 \pmod p$. ♥

Proof We first order elements of A in non-decreasing order as

$$0 \leq a_1 \leq a_2 \leq \cdots \leq a_p \leq a_{p+1} \leq \cdots \leq a_{2p-1} \leq p-1.$$

We may assume that for any $i \in [p-1]$, $a_i \neq a_{i+p-1}$. Otherwise, for some $i \leq p-1$, we have $a_i = a_{i+p-1}$, then $a_i = a_{i+1} = \cdots = a_{i+p-1}$ which sum up to $pa_i = 0$ in \mathbb{Z}_p . Let $S_i = \{a_i, a_{i+p-1}\}$ for all $i \in [p-1]$ and $J = (1)_{(p-1) \times (p-1)}$ be all-1 matrix. Is easy to verify that $\text{per}(J) = (p-1)! \neq 0$. Let $b = (b_1, \dots, b_{p-1})$ be such that $B = \{(b_1, \dots, b_{p-1}) = \mathbb{Z}_p \setminus \{-a_{2p-1}\}\}$. Then Lemma 4.19 implies that there exists a vector $x \in \prod_{i=1}^{p-1} S_i$ such that $(Jx)_i \neq b_i$ for every $i \in [p-1]$. It means that $(Jx)_i = x_1 + \cdots + x_{p-1} \notin B$. Thus we

have $x_1 + \cdots + x_{p-1} + a_{2p-1} = 0$. As $x_i \in S_i$ and all S_i is disjoint, we get a multisubset of size p with sum up to 0.

Chapter 5 Pseudorandomness

Recall Szemerédi's regularity lemma, which partitions any (large) graph G into bounded number of parts such that almost all pairs of parts induces a random-like bipartite graph. We will take a look at the notion of pseudorandomness, also referred to in other contexts as quasirandomness, regularity, uniformity to describe objects that are random-like.

5.1 Quasirandom graphs

We will first take a look at quasirandom graphs, introduced in the 80s by Thomason and independently by Chung–Graham–Wilson. We shall define several properties that at the first glance seems irrelevant of one another but turns out to be equivalent in the sense of being random-like. One immediate application of this is that we have many different ways of checking whether a graph is quasirandom, as if a graph satisfies any one of the equivalent properties, then it satisfies all of them.

We need some notations before stating the equivalent quasirandom properties.

- G is an n -vertex graph with edge density $p \in (0, 1)$, i.e. $p = \frac{e(G)}{\binom{n}{2}}$.
- We let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of the adjacency matrix A of G , ordered by

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|.$$

- For every $u, v \in V(G)$, we write $d(u, v) = |N_G(u) \cap N_G(v)|$ for the codegree of u and v .
- Subgraph density of H in G :

$$\begin{aligned} t(H, G) &= \frac{\text{the number of label copies of } H}{|V(G)|^{|V(H)|}} \\ &= Pr(\text{a uniform random map } f \text{ induces a copy of } H). \end{aligned}$$

- Homomorphism density of H in G :

$$\begin{aligned} t(H, G) &= \frac{\text{the number of label homomorphisms of } H}{|V(G)|^{|V(H)|}} \\ &= Pr(\text{a uniform random map } f \text{ is a homomorphism}). \end{aligned}$$

- Induced subgraph density of H in G :

$$\begin{aligned} t_{\text{ind}}(H, G) &= \frac{\text{the number of induced label copies of } H}{|V(G)|^{|V(H)|}} \\ &= Pr(\text{a uniform random map } f \text{ induces an induced copy of } H). \end{aligned}$$

Theorem 5.1

Let $p \in (0, 1)$ and G be a d -regular n -vertex graph with $d = pn$. Then the following properties are equivalent.

- (1) [Induced Subgraph Count] For every graph H ,

$$t_{\text{ind}}(H, G) = p^{e(H)}(1 - p)^{e(\bar{H})} + o(1).$$

- (2) [Subgraph Count] For every graph H , $t(H, G) = p^{e(H)} + o(1)$.

- (3) [4-cycle Count] $t(C_4, G) \leq p^4 + o(1)$.
 (4) [Spectral gap] $|\lambda_2| = o(n)$.
 (5) [Discrepancy] For any $A, B \subseteq V(G)$, $e(A, B) = p|A||B| + o(n^2)$.
 (6) [Codegree] $\sum_{u, v \in V(G)} |d(u, v) - p^2 n| = o(n^3)$.



Proof We will prove [Induced Subgraph Count] \Rightarrow [Subgraph Count] \Rightarrow [4-cycle Count] \Rightarrow [Spectral gap] \Rightarrow [Discrepancy] \Rightarrow [Codegree] \Rightarrow [Induced Subgraph Count].

- (Induced Subgraph Count) \Rightarrow (Subgraph Count): **Exercise.**
- (Subgraph Count) \Rightarrow (4-cycle Count): By definitions.
- (4-cycle Count) \Rightarrow (Spectral gap):

IDEA: This amounts to write C_4 -count using trace of A^4 and the correct count of C_4 means the contribution from the non-trivial eigenvalues $\lambda_i, i \geq 2$, is negligible.

Exercise. For every $u, v \in V(G)$ and $k \in \mathbb{N}$, $(A^k)_{u,v}$, the u, v -th entry of the k -th power of the adjacency matrix A , is the number of u, v -walk of length k in G .

Then the trace of A^k

$$\text{tr}(A^k) = \sum_{i \in [n]} \lambda_i^k$$

counts the number of closed walks of length k in G . Among these walks, the non-degenerate ones are C_k , while the degenerate ones is easily seen to be negligible, at most $O(n^{k-1})$. Recall that for d -regular graphs, $\lambda_1 = d$. Splitting out the first term in $\text{tr}(A^4)$, we see that

$$p^4 n^4 + o(n^4) \geq t(C_4, G) n^4 + o(n^4) = \text{tr}(A^4) = \lambda_1^4 + \sum_{i \geq 2} \lambda_i^4 = p^4 n^4 + \sum_{i=2}^n \lambda_i^4,$$

implying that $|\lambda_i| = o(n)$ for all $i \geq 2$.

5.1.1 (Codegree) \Rightarrow (Induced Subgraph Count)

Proof Let H be a graph with vertex set $V(H) := \{v_1, v_2, \dots, v_s\}$. For $r \in [s]$, let $H_r := H[\{v_1, \dots, v_r\}]$ be an induced subgraph of H . Note that $H = H_s$. We shall use induction on $1 \leq r \leq s$, via building H_{r+1} from H_r , to show that G has the ‘correct’ count of induced copies of $H = H_s$. We use N_r to denote the number of labelled induced copies of H_r in G . Our aim is to show that

$$N_r = (1 + o(1)) n^r p^{e(H_r)} (1 - p)^{e(\overline{H_r})}. \quad (5.1)$$

The base case $r = 1$ trivially holds. Now assume that (5.1) holds for $1 \leq r < s$, we will show that it holds for $r + 1$.

Extension function. Let $\varepsilon \in \{0, 1\}^r$ be the vector recording the adjacencies of v_{r+1} to H_r in H_{r+1} . Namely, for any $j \in [r]$, we have $\varepsilon_j = 1$ if and only if $v_j v_{r+1} \in E(H_{r+1})$. Let $V^{(r)}$ be the set of ordered r -tuples in V and $n_{(r)} := |V^{(r)}|$. For any $w \in V^{(r)}$, define

$$X(w) = |\{v \in V(G) : v \notin w \text{ and } v \sim w_j \text{ if and only if } \varepsilon_j = 1 \text{ for all } j \in [r]\}|.$$

It is convenient to view things probabilistically. Let $\Omega := V^{(r)}$ and $\Omega^* := \{w \in \Omega : w \cong H_r\}$. Note that

$$N_{r+1} = \sum_{w \in \Omega^*} X(w). \quad (5.2)$$

Now we endow Ω with uniform probabilistic measure and let X be a random variable such that for any $w \in \Omega$,

$$\Pr[X = X(w)] = \frac{1}{n^{(r)}}.$$

Concentration of random variable X . To complete the proof, we need a concentration equality as follows.

$$\sum_{w \in \Omega^*} X(w) = |\Omega^*| \cdot \mathbb{E}[X] + o(n^{r+1}). \quad (5.3)$$

Assuming (5.3) for now, let us finish the proof first. Recalling the inductive hypothesis and the definition, we have that

$$|\Omega^*| = N_r = (1 + o(1))n^r p^{e(H_r)} (1 - p)^{e(\overline{H}_r)}.$$

Observe that

$$\mathbb{E}[X] = \frac{1}{|\Omega|} \sum_{w \in \Omega} X(w).$$

We count $\sum_{w \in \Omega} X(w)$ from the perspective of the $(r + 1)$ -tuple. For any u_{r+1} , the number of w attaching to u_{r+1} with respect to ε is $p^{|\varepsilon|}(1 - p)^{r - |\varepsilon|}n^r$, which implies that

$$\sum_{w \in \Omega} X(w) = (1 + o(1))p^{|\varepsilon|}(1 - p)^{r - |\varepsilon|}n^{r+1}.$$

Thus $\mathbb{E}[X] = (1 + o(1))p^{|\varepsilon|}(1 - p)^{r - |\varepsilon|}n$. Finally, combining (5.2) and (5.3), we derive

$$\begin{aligned} N_{r+1} &= \sum_{w \in \Omega^*} X(w) = |\Omega^*| \cdot \mathbb{E}[X] + o(n^{r+1}) \\ &= (1 + o(1))p^{|\varepsilon|}(1 - p)^{r - |\varepsilon|}n \cdot p^{e(H_r)}(1 - p)^{e(\overline{H}_r)}n^r + o(n^{r+1}) \\ &= (1 + o(1))p^{e(H_{r+1})}(1 - p)^{e(\overline{H}_{r+1})}n^{r+1}. \end{aligned}$$

The last equality holds as $|\varepsilon| + e(H_r) = e(H_{r+1})$ and $r - |\varepsilon| + e(\overline{H}_r) = e(\overline{H}_{r+1})$.

Proof of concentration equality. Use Cauchy-Schwarz inequality to prove the following lemma (Exercise).

Lemma 5.2

Let X be a random variable over finite set Ω with uniform measure. Let $\Omega^* \subseteq \Omega$, then

$$\sum_{w \in \Omega^*} X(w) = |\Omega^*| \cdot \mathbb{E}[X] \pm \sqrt{|\Omega^*| \cdot |\Omega| \text{Var}[X]}$$



By lemma above, it suffices to prove that

$$\sqrt{|\Omega^*| \cdot |\Omega| \text{Var}[X]} = o(n^{r+1}). \quad (5.4)$$

As $\text{Var}[X] = \mathbb{E}[X - \mathbb{E}[X]]^2 = \frac{1}{|\Omega|} \sum_{w \in \Omega} (X - \mathbb{E}[X])^2$, (5.4) is equivalent to

$$\begin{aligned} \sqrt{|\Omega^*| \cdot |\Omega| \text{Var}[X]} &= \sqrt{|\Omega^*| \sum_{w \in \Omega} (X - \mathbb{E}[X])^2} \\ &= \sqrt{|\Omega^*| \sum_{w \in \Omega} (X^2 - (\mathbb{E}[X])^2)} \\ &= o(n^{r+1}). \end{aligned}$$

Recall $|\Omega^*| = N_r = O(n^r)$ and it suffices to show

$$\sum_{w \in \Omega} X(w)^2 = |\Omega|(\mathbb{E}[X])^2 + o(n^{r+2}) = p^{2|\varepsilon|}(1 - p)^{2(r - |\varepsilon|)}n^{r+2} + o(n^{r+2}). \quad (5.5)$$

We shall approximate it by

$$T = \sum_{w \in \Omega} X(w)(X(w) - 1)$$

using double counting. Counting from the perspective of $\{u, v\}$ where both u and v attach to the same w with respect to ε , we have

$$T = \sum d_G(u, v)^{|\varepsilon|} d_{\overline{G}}(u, v)^{r-|\varepsilon|}.$$

To compute T , we need a claim as follows (Exercise).

Claim For any $u \neq v \in V$ and any integers $k, k' \geq 1$, we have

$$\sum_{u \neq v} d_G(u, v)^k d_{\overline{G}}(u, v)^{k'} = (1 + o(1)) p^{2k} (1 - p)^{2k'} n^{k+k'+2}.$$

Hint: Let $\delta_{uv} = d_G(u, v) - p^2 n$. Codegree condition actually implies that for any $k \in \mathbb{N}$, $\sum_{u \neq v} |\delta_{uv}|^k = o(n^{k+2})$.

In addition, $\overline{\delta_{uv}} = d_{\overline{G}}(u, v) - (1 - p)^2 n$ and $\sum_{u \neq v} |\overline{\delta_{uv}}|^k = o(n^{k+2})$.

This claim implies that

$$T = (1 + o(1)) p^{2|\varepsilon|} (1 - p)^{2(r-|\varepsilon|)} n^{r+2}.$$

Finally, we compute the difference and obtain

$$\sum_{w \in \Omega} X(w)^2 = T + \sum_{w \in \Omega} X(w) = T + O(n^{r+1}) = p^{2|\varepsilon|} (1 - p)^{2(r-|\varepsilon|)} n^{r+2} + o(n^{r+2})$$

as desired.

5.1.2 (Spectral gap) \Rightarrow (Discrepancy)

Definition 5.3. $[(n, d, \lambda)$ -graph]

An (n, d, λ) -graph is an n -vertex, d -regular graph whose adjacency matrix has eigenvalues $d = \lambda_1 \geq \dots \geq \lambda_n$ satisfying $\max\{|\lambda_2|, |\lambda_n|\} \leq \lambda$.



If we can prove the following lemma, then (Discrepancy) holds.

Lemma 5.4. [Expander mixing lemma]

If G is an (n, d, λ) -graph, then for any $S, T \subseteq V(G)$,

$$\left| e(S, T) - \frac{d}{n} |S| |T| \right| \leq \lambda \sqrt{|S| |T|}.$$



Proof Consider characteristic vectors 1_S and 1_T . Let A be the adjacency matrix of G and assume that $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of A and v_1, \dots, v_n be the corresponding eigenvectors which form an orthonormal basis. Thus, there exist constants $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ such that

$$1_S = \sum_{i \in [n]} \alpha_i v_i, 1_T = \sum_{i \in [n]} \beta_i v_i.$$

Note that $\lambda_1 = d$ and $v_1 = (1/\sqrt{n}, \dots, 1/\sqrt{n})^T$ is a corresponding eigenvector. We have

$$\begin{aligned} e(S, T) &= 1_S^T A 1_T = \left\langle \sum_{i \in [n]} \alpha_i v_i, \sum_{i \in [n]} \beta_i A v_i \right\rangle \\ &= \left\langle \sum_{i \in [n]} \alpha_i v_i, \sum_{i \in [n]} \beta_i \lambda_i v_i \right\rangle \\ &= \sum_{i \in [n]} \lambda_i \alpha_i \beta_i = d \alpha_1 \beta_1 + \sum_{i=2}^n \lambda_i \alpha_i \beta_i. \end{aligned}$$

To bound the last term, we first compute the values of α_1, β_1 . We have $\alpha_1 = \langle v_1, 1_S \rangle = \frac{|S|}{\sqrt{n}}$ and $\beta_1 = \frac{|T|}{\sqrt{n}}$. Thus, we have

$$e(S, T) - \frac{d|S||T|}{n} = \sum_{i=2}^n \lambda_i \alpha_i \beta_i.$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left| e(S, T) - \frac{d|S||T|}{n} \right| &\leq \lambda \sum_{i=2}^n |\alpha_i \beta_i| \leq \lambda \left(\sum_{i=2}^n \alpha_i^2 \right)^{1/2} \left(\sum_{i=2}^n \beta_i^2 \right)^{1/2} \\ &= \lambda \sqrt{(|S| - \alpha_1^2)(|T| - \beta_1^2)} = \lambda \sqrt{|S| \left(1 - \frac{|S|}{n}\right) |T| \left(1 - \frac{|T|}{n}\right)} \\ &\leq \lambda \sqrt{|S||T|}, \end{aligned}$$

the first equality holds since $|S| = \langle 1_S, 1_S \rangle = \langle \sum_{i \in [n]} \alpha_i v_i, \sum_{i \in [n]} \alpha_i v_i \rangle = \sum_{i \in [n]} \alpha_i^2$, as desired.

5.1.3 (Discrepancy) \Rightarrow (Codegree)

Proof We will prove a stronger statement that every vertex has small codegree deviation: for any u ,

$$\sum_{v: v \neq u} |d(u, v) - p^2 n| = o(n^2).$$

To get rid of the absolute value sign, we split $V(G) \setminus \{u\} = B^+ \cup B^-$ where $B^+ := \{v : d(u, v) > p^2 n\}$, let $A := N(u)$, so $|A| = pn$ and we have

$$\begin{aligned} \sum_{v: v \neq u} |d(u, v) - p^2 n| &= \sum_{v \in B^+} (d(u, v) - p^2 n) + \sum_{v \in B^-} (p^2 n - d(u, v)) \\ &= (e(A, B^+) - p^2 n |B^+|) + (p^2 n |B^-| - e(A, B^-)) \\ &= (e(A, B^+) - p|A||B^+|) + (p|A||B^-| - e(A, B^-)). \end{aligned}$$

Now applying (Discrepancy) to the two terms, we finish the proof.

For sparse graphs, the analogue of (Spectral gap) ($\lambda_2 = o(d)$) implies the analogue of (Discrepancy) ($|e(A, B) - p|A||B|| = o(n^2)$) by Lemma 5.1.2, while the analogue of (Discrepancy) does not necessarily imply the analogue of (Spectral gap). For example, consider an n -vertex graph G , which is the disjoint union of a random d -regular graph with $n - d - 1$ vertices and a K_{d+1} where $d = o(n)$. This graph has (Discrepancy) property, because for any $A, B \subseteq V(G)$, $|e(A, B) - p|A||B|| = o(dn)$. However, we obtain $\lambda_1 = \lambda_2 = d$ by the following claim.

Claim For a d -regular graph G , the multiplicity of the eigenvalue d is exactly the number of components of G .

Proof Assume that the number of components of G is k , its adjacency matrix A can be block diagonal with k blocks, denoted by A_1, \dots, A_k . We have $\det(xE - A) = \det(xE - A_1) \cdots \det(xE - A_k)$, so the multiplicity of d for A is the sum of the multiplicity of d of each A_i , i.e. the number of components of G .

Remark In retrospect, it is perhaps not that surprising now that the seemingly weaker property of [4-cycle Count] is equivalent to [Induced Subgraph Count]. Indeed, we've seen that [4-cycle count] \Rightarrow [Codegree]. In the proof of [Codegree] \Rightarrow [Induced Subgraph Count], we count H -subgraphs by building it up one vertex at a time, and define $H_1, H_2, \dots, H_s = H$. Let H_{r+1}^* be the graph obtained from H_{r+1} by adding a new vertex v_{r+1}^* , which is a copy of v_{r+1} . To count H_{r+1} , we need to control the variance: the number of H_{r+1}^* , which in view of the twins v_{r+1} and v_{r+1}^* , is governed by the [Codegree] property.

Consider sparse graphs ($d = o(n)$) and see what still holds. By expanding lemma, the [Eigenvalue](Spectral gap, $|\lambda_2| = o(d)$) implies the [Discrepancy] ($|e(A, B) - \frac{d}{n}|A||B|| = o(dn)$), but the converse implication is no longer true for sparse graphs. Let us consider an example graph which is union of a d -regular genuinely random graph on $n - d - 1$ vertices and a K_{d+1} . It satisfies the discrepancy property but not the eigenvalue property.

What if we impose more symmetry to exclude such examples? In fact, a theorem that [Discrepancy] implies [Eigenvalue] for all Caylay graphs was proved by Kohayakawa, Rödl and Schacht.

Definition 5.5. [Disc(δ)]

Let $0 < \delta \leq 1$, we say a d -regular graph G satisfies $\text{Disc}(\delta)$ if for any disjoint subset U, V of $V(G)$, we have

$$e_G(U, V) = (1 \pm \delta) \frac{d}{n} |U||V|.$$



Definition 5.6. [EIG(ϵ)]

For a d -regular graph G and its eigenvalues $d = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, we say it satisfies $\text{EIG}(\epsilon)$ if for any $i \geq 2$, we have $|\lambda_i| \leq \epsilon d$.



Definition 5.7. [Caylay graph]

Let Γ be an abelian group and let $A \subseteq \Gamma \setminus \{0\}$ be symmetric (i.e. $A = -A$). The Caylay graph $\text{Cay}(\Gamma, A)$ is the graph with vertex set Γ and edge set $\{xy : x - y \in A\}$.



For example, if $\Gamma = \mathbb{Z}_N$, $A = \{-1, 1\}$, then $\text{Cay}(\Gamma, A) = C_N$.

Now the previous theorem could be stated more formally as following, and we shall see a proof by Gowers.

Theorem 5.8. [K-R-S/G]

For any $\epsilon > 0$, there exists $\delta > 0$ and n_0 such that the following holds. Let $G = \text{Cay}(\Gamma, A)$ be a Caylay graph for some abelian group Γ with $|\Gamma| = n \geq n_0$ and a symmetric $A \subseteq \Gamma \setminus \{0\}$. Then $\text{Disc}(\delta) \Rightarrow \text{EIG}(\epsilon)$.



5.2 Basics on Caylay graphs

Let Γ be a finite abelian group. A character χ of Γ is a group homomorphism from Γ to S^1 , where S^1 is the multiplicative group of complex numbers. In another word, $\chi : \Gamma \rightarrow S^1$ has a property that for any $a, b \in \Gamma$, $\chi(a + b) = \chi(a)\chi(b)$. For example, all 1 function is a trivial character. If $|\Gamma| = n$, then there are n characters $\chi_1, \chi_2, \dots, \chi_n$ pairwise orthogonal, i.e.

$$\langle \chi_i, \chi_j \rangle = \sum_{g \in \Gamma} \chi_i(g) \overline{\chi_j(g)}.$$

- Consider the space of functions from Γ to \mathbb{C} , then the characters $\chi_1, \chi_2, \dots, \chi_n$ form an orthonormal basis of this space.
- If $|\{\chi(g) : g \in \Gamma\}| = m$, then χ takes value in m -th root of unity.
- Eigenvalues of Cayley graphs $\text{Cay}(\Gamma, A)$ are Fourier coefficient of A .

Theorem 5.9

Let $G = \text{Cay}(\Gamma, A)$ with Γ finite abelian and $A \subset \Gamma \setminus \{0\}$ symmetric. Then every character χ is an eigenvector of the adjacency matrix of G with eigenvalue

$$\hat{A}(\chi) = \langle A, \chi \rangle = \sum_{a \in A} \chi(a).$$



Proof Let M be the adjacency matrix of G , it suffices to show that

$$M \cdot \chi = \langle A, \chi \rangle \cdot \chi = \left(\sum_{a \in A} \chi(a) \right) \cdot \chi.$$

Fix a coordinate $g \in \Gamma$,

$$\begin{aligned} (M \cdot \chi)_g &= \sum_{b \in g+A} \chi(b) \\ &= \sum_{a \in A} \chi(g+a) \\ &= \sum_{a \in A} \chi(g) \cdot \chi(a) = \left(\sum_{a \in A} \chi(a) \right) \cdot \chi(g). \end{aligned}$$

Lemma 5.10. [Chernoff type concentration]

Let $p_1, \dots, p_n \in [0, 1]$ and $p = \frac{1}{n} \sum_{i=1}^n p_i$. Let X_i be centered Bernoulli random variables, i.e.

$$X_i = \begin{cases} 1 - p_i & \text{with probability } p_i \\ -p_i & \text{with probability } 1 - p_i \end{cases} \quad \text{and } X = \sum_{i=1}^n X_i. \text{ Then for any } a > 0,$$

$$\Pr(|X| > a) < 2e^{-\frac{2a^2}{n}}.$$



Definition 5.11

Given two functions f and $h : \Gamma \rightarrow \mathbb{C}$, let $f * h : \Gamma \rightarrow \mathbb{C}$ be their convolution defined as: for any $a \in \Gamma$,

$$f * h(a) = \sum_{g \in \Gamma} f(a-g)h(g).$$



Idea In order to prove Theorem 5.8, it suffices to prove the contraposition. More precisely, we prove that if $\lambda_1 = d = |A|, \lambda_2 \geq \epsilon d$, then we can find two sets X, Y such that $e(X, Y)$ is abnormal. We shall find such X, Y randomly by choosing elements $g \in \Gamma$ with probability depending on χ , a nontrivial character, whose corresponding eigenvalue is large.

Proof Write $\chi = c + i \cdot s = e^{i\theta_\chi}$, for any $g \in \Gamma, c(g) = \text{Re}(\chi(g)) = \cos(\theta_\chi), s(g) = \text{Im}(\chi(g)) = \sin(\theta_\chi)$. By the orthogonality of character, we have $\chi \perp 1$, then

$$0 = \langle 1, \chi \rangle = \sum_{g \in \Gamma} (c(g) + i \cdot s(g)).$$

we can get $\sum_{g \in \Gamma} c(g) = 0, \sum_{g \in \Gamma} s(g) = 0$.

Since A is symmetric, $\forall a \in A, \exists -a \in A$, then $s(a) = -s(-a) \Leftrightarrow s(a) + s(-a) = 0$, we can get

$$\sum_{a \in A} s(a) = 0.$$

$$\begin{aligned} \langle A, \chi \rangle &= \sum_{a \in A} \chi(a) \\ &= \sum_{a \in A} (c(a) + i \cdot s(a)) \\ &= \langle A, c \rangle. \end{aligned}$$

Set probability vector $p = \frac{1+c}{2}$. Define $-X$ to be the random set obtained as follows: $\forall g \in \Gamma$, g is included in $-X$ with probability $p(g) = \frac{1+c(g)}{2}$ independently. Define Y in the same way.

The goal is to bound the deviation $|e(X, Y) - \frac{d}{n}|X||Y||$.

Let us first get a hold on sizes of X, Y . Each element g appears independently with probability $p(g)$, thus $|\Gamma| = n$,

$$\mathbb{E}|Y| = \mathbb{E}|X| = \sum_{g \in \Gamma} \frac{1+c(g)}{2} = \frac{n}{2}.$$

Since each element is chosen independently and Lemma 5.2, we have

$$\begin{aligned} \mathbb{P}\left(|X| = \left(\frac{1}{2} + o(1)\right)n\right) &= 1 - o(1). \\ \mathbb{P}\left(|Y| = \left(\frac{1}{2} + o(1)\right)n\right) &= 1 - o(1). \end{aligned}$$

By linearity of expectation,

$$\begin{aligned} \mathbb{E}|X \cap Y| &= \sum_{g \in \Gamma} p(-g) \cdot p(g) \\ &= \sum_{g \in \Gamma} p(g)^2 \\ &= \frac{1}{4} \sum_{g \in \Gamma} (1+c(g))^2 \\ &= \frac{1}{4}n + \frac{1}{4} \sum_{g \in \Gamma} c(g)^2 \\ &= \frac{1}{4}n + \frac{1}{4} \cdot \frac{n}{2} = \frac{3n}{8}. \end{aligned}$$

By the Lemma 5.2, we have

$$\begin{aligned} \mathbb{P}\left(|X \cap Y| = \left(\frac{3}{8} + o(1)\right)n\right) &= 1 - o(1). \\ \mathbb{P}\left(|X \cup Y| = \left(\frac{5}{8} + o(1)\right)n\right) &= 1 - o(1). \end{aligned}$$

$$\begin{aligned}
 e(X, Y) &= \sum_{a \in A} \sum_{g \in \Gamma} X(g-a)Y(g) \\
 &= \sum_{a \in A} \left(\sum_{g \in \Gamma} (-X)(a-g)Y(g) \right) \\
 &= \sum_{a \in A} (-X) * Y(a) \\
 &= \sum_{g \in \Gamma} A(g) \cdot (-X) * Y(g) \\
 &= \langle A, (-X) * Y \rangle.
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}(e(X, Y)) &= \mathbb{E}\langle A, (-X) * Y \rangle \\
 &= \sum_{a \in A} \sum_{g \in \Gamma} \mathbb{E}((-X)(a-g)) \mathbb{E}(Y(g)) \\
 &= \sum_{a \in A} \left(\sum_{g \in \Gamma} p(a-g)p(g) \right) \\
 &= \sum_{a \in A} p * p(a) \\
 &= \sum_{a \in A} \frac{1+c}{2} * \frac{1+c}{2}(a) \\
 &= \left\langle A, \frac{1+c}{2} * \frac{1+c}{2} \right\rangle \\
 &= \frac{1}{4} \sum_{a \in A} \sum_{g \in \Gamma} (1+c(a-g)) \cdot (1+c(g)) \\
 &= \frac{1}{4} \sum_{a \in A} \sum_{g \in \Gamma} (1+c(a-g)c(g)) \\
 &= \frac{1}{4} dn + \frac{1}{4} \langle A, c * c \rangle.
 \end{aligned}$$

Claim $c * c(g) = \begin{cases} \frac{n}{2}c(g) & m > 2 \\ nc(g) & m = 2 \end{cases}.$

$$\begin{aligned}
 \left| \mathbb{E}(e(X, Y)) - \frac{1}{4}dn \right| &\geq \frac{1}{8}n |\langle A, c \rangle| \\
 &\geq \frac{1}{8}n\varepsilon d.
 \end{aligned}$$

Recall $0 \leq e(X, Y) \leq dn = |A|n$, $\eta = \eta(X, Y) = e(X, Y) - \frac{dn}{4}$, we can get

$$-\frac{1}{4}dn \leq \eta \leq \frac{3}{4}dn.$$

Write $q = \Pr(|\eta| \leq \frac{\varepsilon dn}{16})$,

$$\frac{1}{8}\varepsilon dn \leq |\mathbb{E}(\eta)| \leq \mathbb{E}(|\eta|) \leq q \frac{\varepsilon dn}{16} + (1-q) \frac{3dn}{4}.$$

we get

$$q = \Pr(|\eta| \leq \frac{\varepsilon dn}{16}) \leq \frac{1 - \varepsilon/6}{1 - \varepsilon/12} \leq 1 - \frac{\varepsilon}{12}.$$

thus there exist choices of X, Y such that $|\eta| \geq \frac{\varepsilon dn}{16}$ and $|X|, |Y|$ are as expected. That is,

$$|X| = |Y| = \left(\frac{1}{2} + o(1)\right)n.$$

$$|X \cap Y| = \left(\frac{3}{8} + o(1)\right)n.$$

$$|X \cup Y| = \left(\frac{5}{8} + o(1)\right)n.$$

$$|\eta| = |e(X, Y) - \frac{dn}{4}| \geq \frac{\varepsilon dn}{16}.$$

Apply $\text{Disc}(\delta)$ on $X \cap Y, X \cup Y, X - Y, Y - X$, finally we can get

$$\frac{\varepsilon dn}{16} \leq |e(X, Y) - \frac{dn}{4}| < \frac{5\delta dn}{16}.$$

By setting $\delta < \frac{\varepsilon}{5}$, we get a contradiction.

5.3 Quasirandomness for Hypergraphs

We restrict attention to bipartite graphs.

Theorem 5.12

Let G be a bipartite graph with vertex sets X and Y , each of size N . Suppose that G has pN^2 edges. Then the following properties of G are equivalent.

- (i) [C_4 Count] The number of labelled 4-cycles in G is at most $p^4N^4 + o(N^4)$.
- (ii) [Discrepancy] For any $X' \subseteq X, Y' \subseteq Y$, $|e(X', Y') - p|X'||Y'||| = o(N^2)$.



Now we define a notion of quasirandomness for subsets of \mathbb{Z}_N .

Theorem 5.13. [Chung and Graham [6]]

Let A be a subset of \mathbb{Z}_N of size pN . Then the following properties are equivalent.

- (i) The number of quadruples $(a, b, c, d) \in A^4$ such that $a + b = c + d$ is at most $p^4N^3 + o(N^3)$.
- (ii) For any arithmetic progression X in \mathbb{Z}_N , $|A \cap X| = p|X| + o(N)$.



◆ Connection between quasirandom subsets of \mathbb{Z}_N and quasirandom graphs.

Take $A \subseteq \mathbb{Z}_N$ with $|A| = pn$. Consider the following Cayley Sum graph. Define a bipartite graph G with vertex sets $X = Y = \mathbb{Z}_N$ by letting $(x, y) \in XY$ be an edge if and only if $x + y \in A$. Note that $e(G) = pN^2$. Take a C_4 in G , say (x_1, x_2, y_1, y_2) , then $x_1 + y_1, x_1 + y_2, x_2 + y_1, x_2 + y_2$ all belong to A , and moreover that $(x_1 + y_1) + (x_2 + y_2) = (x_1 + y_2) + (x_2 + y_1)$. So there is an N -to-one correspondence between the 4-cycles from Theorem 5.12 (i) and the quadruples from Theorem 5.13 (i). Thus, the set A is quasirandom if and only if the corresponding graph G is quasirandom.

◆ What about quasirandom hypergraphs, how to define them?

Consider 3-uniform hypergraphs. When a set is quasirandom, the number of 3-AP (arithmetic progression) is as expected. But there are quasirandom set whose 4-AP count deviates a lot from expected. To detect whether a set has abnormal count of 4-AP. Gowers introduced a high order quasirandomness.

Definition 5.14. (quadratic uniformity)

Let $A \subseteq \mathbb{Z}_N$ of size pN . We say A is quadratically uniform if the number of octuples $(x, x + a, x + b, x + c, x + a + b, x + a + c, x + b + c, x + a + b + c)$ in A^8 is at most $p^4 N^4 + o(N^4)$. ♣

Remark Since quadruples (a, b, c, d) with $a + b = c + d$ are in one-to-one correspondence with quadruples of the form $(x, x + a, x + b, x + a + b)$, this definition is a natural generalization of property (i) of Theorem 5.13.

To define quasirandom properties for 3-uniform hypergraphs, we consider Cayley Sum hypergraph H with vertex sets $X = Y = Z = \mathbb{Z}_N$. The triple $(x, y, z), x \in X, y \in Y, z \in Z$ forms an edge of H if and only if $x + y + z \in A$, where $A \subseteq \mathbb{Z}_N$.

It turns out a counterpart of C_4 in 3-uniform hypergraphs is octahedron, where an octahedron is a set of eight 3-edges of the form $(x_i, y_j, z_k) : i, j, k \in \{1, 2\}$, with $x_1, x_2 \in X, y_1, y_2 \in Y, z_1, z_2 \in Z$. Equivalently, an octahedron is $K_{2,2,2}^{(3)}$, a complete tripartite subhypergraph with two vertices from each vertex set of H .

Another way to look at it is that the dual of 2-dimensional cube (4-cycle) is 4-cycle and the dual of 3-dimensional cube is octahedron (see Figure 5.1).

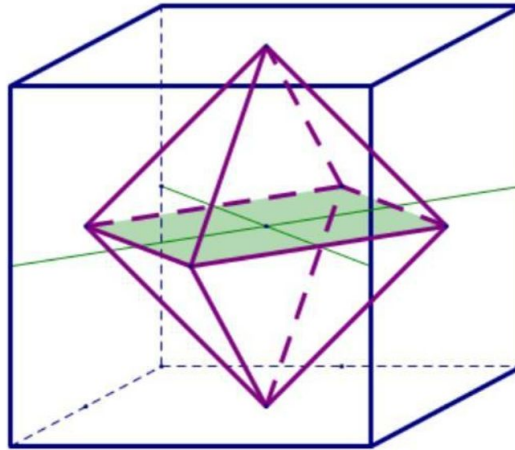


Figure 5.1: the dual of 3-dimensional cube

Definition 5.15. (Octahedron count)

Let H be a tripartite 3-uniform hypergraph with N vertices in each partite sets and pN^3 edges. Then H is quasirandom if it contains at most $p^8 N^6 + o(N^6)$ octahedra. ♣

Remark A subset A of \mathbb{Z}_N gives rise to an quasirandom 3-graph H (Caley Sum hypergraph) if and only if A is quadratically uniform.

◆ What about Discrepancy for 3-uniform hypergraphs?

A natural way is the following.

Definition 5.16. (vertex-uniformity)

Let H be a 3-uniform hypergraph with partite sets X, Y, Z each of size N and suppose that H has pN^3 edges. Then H is vertex-uniform if, for any $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$,

$$|e(X', Y', Z') - p|X'||Y'||Z'|| = o(N^3).$$



A quasirandom 3-graph must be vertex-uniform, but the reverse is not true. Here is a simple example. Let

X, Y, Z be three sets of size N and let G be a random tripartite graph with vertex sets X, Y and Z and density $1/2$. Let H be the hypergraph consisting of all triangles in G . Then the edge density of H is $1/8$, but the number of octahedra in H is about $2^{-12}N^6$ rather than $8^{-8}N^6$ as it should have if H is quasirandom.

Definition 5.17. (edge-uniformity)

Let H be a 3-uniform hypergraph with partite sets X, Y, Z each of size N and suppose that H has pN^3 edges. Then H is edge-uniform if, for every $t \in [0, 1]$ and every tripartite graph G with vertex sets X, Y, Z and tN^3 triangles, the number of triangles in G are edges in H is $ptN^3 \pm o(N^3)$.



- The discrepancy property says that a bipartite graph does not significantly correlate with graphs induced by sets of vertices (that is, complete bipartite graphs on subsets of the vertex sets).
- Edge-uniformity says that a 3-uniform hypergraph does not correlate with 3-uniform hypergraphs induced by sets of edges.

Chapter 6 The Spectral method

6.1 Spectral theorem and Hoffman's bound

Spectral methods use linear algebra to derive information about graphs, usually via studying certain real matrices (often symmetric) associated to graphs. For example there are adjacency matrix, Laplacian matrix and transition matrix.

Let $G = (V, E)$ be an n -vertex graph. We can view every set of vertex $X \subseteq V$ as an $\{0, 1\}^n$ -vector. Edge sets correspond to quadratic forms defined by some matrices. For example, in the Expander Mixing Lemma, we have seen that if there are two sets X and Y , then we can take the adjacency matrix and compute

$$\begin{aligned} X^T AY &= e(X, Y) \\ &= \sum_{u,v \in V} X_u A_{uv} Y_v \end{aligned}$$

When dealing with matrix M , we usually use it in two ways. The first way is to view them as linear operator $x \mapsto Mx$ for $x \in \mathbb{R}^v$ and $Mx \in \mathbb{R}^v$. The second way is to define a quadratic form $x \mapsto x^T Mx \in \mathbb{R}$.

Theorem 6.1. [Spectral theorem]

Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric $n \times n$ matrix. Then it has n real eigenvalues (not necessary distinct) $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and n orthonormal real eigenvectors $\{x_1, \dots, x_n\} \in \mathbb{R}^n$ where x_i is eigenvector of λ_i . ♥

We will prove the theorem later.

Equipped with this theorem, we can already prove something meaningful. Let's try to prove one of the basic results on Hoffman's bound on independence number. Before getting into it, let's introduce the first matrix which is the most natural one which is the adjacency matrix.

Let A be the adjacency matrix of G . $A_{u,v} = 1$ if and only if u is adjacent to v . The following are some basics.

Theorem 6.2

If G is an n -vertex and d -regular graph, then

- $\mathbf{1}$ is an eigenvector for adjacency matrix A with eigenvalue d ;
 - $d = \lambda_1 \geq \dots \geq \lambda_n$;
 - $\lambda_n < 0$;
 - $\text{trac}(A) = 0 = \sum_{i \in [n]} \lambda_i$;
 - the multiplicity of $\lambda_1 = d$ is the number of connected components in G .
- ♥

For adjacency matrix A as an operator, take some $x \in \mathbb{R}^v$,

$$Ax = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ & \ddots & \vdots \\ a_{n1} & & a_{nn} \end{pmatrix} \begin{pmatrix} x_{v_1} \\ \vdots \\ x_{v_n} \end{pmatrix}. \text{ Let's take a look at a specific example}$$

$$(Ax)_v = \begin{pmatrix} a_{v1} & \dots & a_{vn} \end{pmatrix} \begin{pmatrix} x_{v_1} \\ \vdots \\ x_{v_n} \end{pmatrix} = \sum_{u \in N(v)} x_u \text{ where } (Ax)_v \text{ is the sum of weights at } N(v).$$

Clearly, $x^T Ax = e(X, X) = 2e(X)$. Next, let's take a look at an easy example. K_n is a complete graph. It is easy to check that the eigenvalues of K_n is $\{n-1, -1, \dots, -1\}$ and the multiplicity of λ_1 is 1.

🔥 **Exercise 6.1** The d -regular graph has d as eigenvalue.

Theorem 6.3. [Hoffman's bound]

For every n -vertex d -regular graph G , let $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of its adjacency matrix. Then $\alpha(G) \leq \left(\frac{-\lambda_n}{d-\lambda_n}\right)n$.



Remark The Hoffman's bound is tight. $K_{d,d}$ is an example of the tightness. Since in the adjacency matrix of $K_{d,d}$, we have $\lambda_1 = d$, $\lambda_n = -d$ and $\alpha(K_{d,d}) = d$.

Before the proof of Hoffman's bound, we need some facts. By Theorem 6.1, we have orthonormal eigenvectors $\{v_1, \dots, v_n\}$ for the adjacency matrix A of n -vertex and d -regular graph G . For any set of vertices $X \subseteq V$, $X = \sum_{i \in V} a_i v_i$. Here we want to know the meaning of coefficient a_i . Some coefficients have special meaning. For example, a_1 has a special meaning. Let G be an n -vertex and d -regular graph.

Then $v_1 = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. We can obtain a_1 by taking $\langle x, v_1 \rangle = \langle \sum a_i v_i, v_1 \rangle = a_1$. On the other hand, $\langle x, v_1 \rangle = \langle x, \frac{1}{\sqrt{n}} \mathbf{1} \rangle = \frac{1}{\sqrt{n}} |X|$. So the meaning of the coefficient is the size of the set. Next, we want to know the meaning of $\sum a_i^2$. Note that $|X| = \langle X, X \rangle = \langle \sum a_i v_i, \sum a_i v_i \rangle = \sum a_i^2$.

Proof Take an independent set I in G . We have $I = \sum a_i v_i$, $a_1 = \frac{1}{\sqrt{n}} |I|$ and $\sum a_i^2 = |I|$. Given by the adjacency matrix,

$$\begin{aligned}
 0 &= 2e(I) = I^T A I \\
 &= \langle I, A I \rangle \\
 &= \langle \sum a_i v_i, A(\sum a_i v_i) \rangle \\
 &= \langle \sum a_i v_i, \sum a_i (A v_i) \rangle \\
 &= \langle \sum a_i v_i, \sum \lambda_i a_i v_i \rangle \\
 &= \sum_{i=1}^n \lambda_i a_i^2 \\
 &= d \cdot a_1^2 + \sum_{i \geq 2} \lambda_i a_i^2 \\
 &\geq d \cdot a_1^2 + \sum_{i \geq 2} \lambda_n a_i^2 \\
 &= (d - \lambda_n) a_1^2 + \left(\sum_{i=1}^n a_i^2 \right) \lambda_n \\
 &\geq (d - \lambda_n) \frac{|I|^2}{n} + |I| \lambda_n.
 \end{aligned}$$

Now, we have $0 \geq (d - \lambda_n) \frac{|I|^2}{n} + |I| \lambda_n$ and $|I| \leq \left(\frac{-\lambda_n}{d-\lambda_n}\right)n$.


Corollary 6.4

Let G be an n -vertex and d -regular graph G with $\lambda_1 \geq \dots \geq \lambda_n$. Then $\chi(G) \geq \frac{n}{\alpha(G)} \geq \frac{\lambda_1 - \lambda_n}{-\lambda_n}$.



Remark $\chi(G) \geq \frac{\lambda_1 - \lambda_n}{-\lambda_n}$ also holds for irregular graph.


Definition 6.5

The Rayleigh quotient of x with respect to M is $\frac{x^T M x}{x^T x} = R_M(x)$. 

If x is an eigenvector of M with eigenvalue λ , then the Rayleigh quotient $R_M(x) = \frac{x^T M x}{x^T x} = \frac{x^T \lambda x}{x^T x} = \lambda$.

- The variational characterisation of eigenvalues.

Theorem 6.6. (Courant-Fisher)

Let M be a real $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$. Then $\lambda_k = \min_{k\text{-dim}, V} \max_{x \in V, x \neq 0} \frac{x^T M x}{x^T x} = \min_{k\text{-dim}, V} \max_{x \in V, x \neq 0} R_M(x)$. 

Proof By Theorem 6.1, we can take orthonormal eigenvectors $\{v_1, \dots, v_n\}$ and denote the corresponding eigenvalues as $\{\lambda_1, \dots, \lambda_n\}$. We divide the proof into following two cases.

- $\lambda_k \geq \min_{k\text{-dim}, V} \max_{x \in V, x \neq 0} R_M(x)$;
- $\lambda_k \leq \min_{k\text{-dim}, V} \max_{x \in V, x \neq 0} R_M(x)$.

Case 1: We need to find some k -dimensional space V such that $\lambda_k \geq \max_{x \in V, x \neq 0} R_M(x)$. We consider V as

a space which spans from $\{v_1, \dots, v_k\}$ and take non-zero $x \in V$ and $x = \sum_{i=1}^k a_i v_i$. We have

$$\begin{aligned} x^T M x &= x^T M \left(\sum_{i=1}^k a_i v_i \right) \\ &= \left\langle \sum_{i=1}^k a_i v_i, \sum_{i=1}^k a_i \lambda_i v_i \right\rangle \\ &= \sum_{i=1}^k \lambda_i a_i^2 \\ &\leq \lambda_k \sum_{i=1}^k a_i^2 \end{aligned}$$

and

$$\begin{aligned} x^T x &= \left\langle \sum_{i=1}^k a_i v_i, \sum_{i=1}^k a_i v_i \right\rangle \\ &= \sum_{i=1}^k a_i^2. \end{aligned}$$

Hence, we have $x^T M x = \sum_{i=1}^k a_i^2 \geq \frac{x^T M x}{\lambda_k}$ and $\lambda_k \geq \frac{x^T M x}{x^T x}$.

Case 2: We need to show that, for every k -dimensional V , there exists a non-zero $x \in V$ with $R_M(x) \geq \lambda_k$.

Fix an arbitrary k -dimensional V , let U be a space which spans from $\{v_k, \dots, v_n\}$ and $\dim(U) = n - k + 1$.

As $\dim(U) + \dim(V) > n$, there exists non-zero $x \in U \cap V$. Since $x \in U$, $x = \sum_{i=k}^n a_i v_i$. We also have

$x^T x = \sum_{i=k}^n a_i^2$. Now,

$$\begin{aligned}
x^T M x &= \left\langle \sum_{i=k}^n a_i v_i, \sum_{i=k}^n \lambda_i a_i v_i \right\rangle \\
&= \sum_{i=k}^n \lambda_i a_i^2 \\
&\geq \lambda_k \sum_{i=k}^n a_i^2 = \lambda_k x^T x.
\end{aligned}$$

Hence, we have $x^T M x \geq \lambda_k x^T x$ and $\lambda_k \leq \frac{x^T M x}{x^T x}$.

We will see that the Courant-Fisher theorem has lots of quick applications. For example, we can extend Hoffman's bound on independence number to irregular graphs. Previously, we only deal with d -regular graphs.

Remark Courant-Fisher theorem means that we can view eigenvalues as optima of min-max optimisation problem in which the cost function is the Rayleigh quotient.

Usually, the important eigenvalues that we care about are λ_1 , λ_2 and λ_n , where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. For λ_n , we have already seen in Hoffman's bound that λ_n has link to the independence number. For λ_1 and λ_2 , they are both important in Expander Mixing Lemma where we use to derive some equivalence on the quasirandom properties.

By the Courant-Fisher theorem, we have the following corollary.

Corollary 6.7

$$\begin{aligned}
\lambda_1 &= \min_{x \neq 0} R_M(x) = \min_{x \neq 0} \frac{x^T M x}{x^T x}. \\
\lambda_n &= \max_{x \neq 0} R_M(x) = \max_{x \neq 0} \frac{x^T M x}{x^T x}. \\
\lambda_2 &= \min_{x \neq 0, x \perp v_1} R_M(x) = \min_{x \neq 0, x \perp v_1} \frac{x^T M x}{x^T x},
\end{aligned}$$

where v_1 is the eigenvector of λ_1 .



6.2 Laplacian matrix

Definition 6.8. (Laplacian matrix)

Given an n -vertex d -regular graph G , we define

$$L = dI - A = \begin{pmatrix} d & & & \\ & d & & \\ & & \ddots & -1 \\ & & & d \end{pmatrix}$$

as the Laplacian matrix of G , where I is the identity matrix, A is the adjacency matrix and $L_{uv} = -1$ if u and v are adjacent.



Fact 6.9.

- $L \cdot \mathbf{1} = 0$;
- If v_i is an eigenvector of A with eigenvalue α_i , then v_i is also an eigenvector of L with eigenvalue $d - \alpha_i$.



Definition 6.10. (Normalised Laplacian matrix)

Given an n -vertex d -regular graph G , we define $N = \frac{1}{d} \cdot L = I - \frac{1}{d} \cdot A$ as the normalised Laplacian matrix of G .



Note that this definition makes that the spectral radius is independent of d .

For a general graph, the Laplacian matrix is $L = D - A$, where $D = \begin{pmatrix} d(v_1) & & & 0 \\ & d(v_2) & & \\ & & \ddots & \\ 0 & & & d(v_n) \end{pmatrix}$.

Convention:

A denotes the adjacency matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$.

L denotes the Laplacian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

N denotes the normalised Laplacian matrix with eigenvalues $\nu_1 \leq \nu_2 \leq \dots \leq \nu_n$.

6.3 Hoffman's bound for irregular graphs

Theorem 6.11. [Godsil-Newman]

Let I be an independent set in G and let $d(I)$ be the average degree of vertices in I , then $|I| \leq \frac{\lambda_n - d(I)}{\lambda_n} \cdot n$, where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ are eigenvalues of Laplacian of G .



Remark If G is d -regular, then it implies Hoffman's bound. This is because $\lambda_n = d - \alpha_n$ by Fact 6.2 ($\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$ are eigenvalues of adjacency matrix A) and then

$$|I| \leq \frac{\lambda_n - d}{d - \alpha_n} \cdot n = \frac{-\alpha_n}{d - \alpha_n} \cdot n$$

Proof By the Courant-Fisher theorem, we have $\lambda_n = \max_{x \neq 0} R_L(x) = \max_{x \neq 0} \frac{x^T L x}{x^T x}$. Hence, $\lambda_n \geq R_L(x) = \frac{x^T L x}{x^T x}$

for arbitrary $x \neq 0$. Let $x = \mathbf{1}_I - \frac{|I|}{n} \cdot \mathbf{1}$, where $\mathbf{1}_I$ is the indicator vector of I . So $x \perp \mathbf{1}$. We have

$$\begin{aligned} x^T L x &= (\mathbf{1}_I - \frac{|I|}{n} \cdot \mathbf{1})^T L (\mathbf{1}_I - \frac{|I|}{n} \cdot \mathbf{1}) \\ &= \mathbf{1}_I^T L \mathbf{1}_I \\ &= \mathbf{1}_I^T (D - A) \mathbf{1}_I \\ &= \sum_{u,v} (\mathbf{1}_I)_u (D - A)_{u,v} (\mathbf{1}_I)_v \\ &= \sum_{v \in I} d(v) - 2e(I) \\ &= |I| \cdot d(I), \end{aligned}$$

where the second equation holds as $L \cdot \mathbf{1} = 0$ and the last equation comes from $e(I) = 0$. Moreover, $x^T x = (\mathbf{1}_I - \frac{|I|}{n} \cdot \mathbf{1})^T (\mathbf{1}_I - \frac{|I|}{n} \cdot \mathbf{1}) = |I|(1 - \frac{|I|}{n})$ by the following exercise. Then $\lambda_n \geq \frac{|I| \cdot d(I)}{|I|(1 - \frac{|I|}{n})}$. Hence,

$$|I| \leq \frac{\lambda_n - d(I)}{\lambda_n} \cdot n.$$

Remark The reason to choose centered characteristic function of I instead of $\mathbf{1}_I$ is because $x^T Lx$ remain the same, shifting by $\frac{|I|}{n} \cdot \mathbf{1}$ minimises the norm $x^T x$.

🔗 **Exercise 6.2** Let $S \subset V$ be a set of size $s|V|$ and let $f_t = \mathbf{1}_S - t \cdot \mathbf{1}$. Then $\|f_t\|^2$ is minimised when $t = s$, i.e. when $f_t \perp \mathbf{1}$ and $\|\mathbf{1}_S - s \cdot \mathbf{1}\|^2 = s(1 - s)|V|$.

6.4 Why this definition of Laplacian?

Motivation: Given a graph G , one type of problem we care about is to find a cut between X and $X^C = V \setminus X$, and then see how many edges between them. Sometimes we want to find a sparse cut, sometimes we want to find a dense one. For example, one of the basic theorem says that you can always find a cut with at least half of the number of edges of the graph G , which can be proved by probability method and take a random bi-partition. In that problem, our aim is to try to cut many edges, i.e. find a dense cut. Meanwhile, people are also interested in sparse cuts. For example, theoretical computer scientists care about finding sparse cuts. Why do they care about this? Because sometimes we can use divide-conquer idea to solve a problem. If there is too much work to run the algorithm on the whole graph, then we want to find a sparse cut and cut the graph into two parts. Then we repeatedly try to find sparse cuts, and finally find very small components. Then we just run the algorithm on each small component. Not only improve the running time, but we also don't need much space to store the information. Because the cut is very sparse, sometime we can place them together, and the solution will be ok. These are some big motivations.

Now, we represent a cut in terms of linear algebra, using the following homogeneous quadratic polynomial. We define a degree-2 homogeneous polynomial for graph $G = (V, E)$:

$$\sum_{uv \in E} (x_u - x_v)^2,$$

which measures smoothness of x . The value of this polynomial is small when no big jump over edges. When $x \in \{0, 1\}^V$ is a Boolean vector and let X be a subset of V with indicator vector x , then $e(X, X^C) = \sum_{uv \in E} (x_u - x_v)^2$, i.e. the cut value. Note that every homogeneous quadratic polynomial can be written as $x^T Mx$ for some matrix M . For d -regular graph, $\sum_{uv \in E} (x_u - x_v)^2 = x^T (dI - A)x = x^T Lx$, whose both sides equal to $\sum_{v \in V} dx_v^2 - 2 \sum_{uv \in E} x_u x_v$. From this, we immediately get the following.

Proposition 6.12

The Laplacian L is singular and positive semidefnite.

Note that this proposition is true for all graphs.

At the end, let's conclude what does Laplacian matrix do? Both as operator and as quadratic form. Recall

that for d -regular graph $G = (V, E)$, the Laplacian $L = \begin{pmatrix} d & & & \\ & d & & \\ & & \ddots & -1 \\ & & & d \end{pmatrix}$, where $L_{uv} = -1$ if u and v are

adjacent. Take $x \in \mathbb{R}^V$, then $(Lx)_v = dx_v - \sum_{u \in N(v)} x_u = d(x_v - \frac{1}{d} \sum_{u \in N(v)} x_u)$. Moreover, the quadratic form $x^T Lx = \sum_{uv \in E} (x_u - x_v)^2$ represents the cut value when $x \in \{0, 1\}^V$. When $x \in \mathbb{R}^V$, optimising $x^T Lx$ can be viewed as a relaxation of the cut problem.

6.5 Spectrum of the normalised Laplacian

In this section, we study spectrum of the normalised Laplacian. We start with regular graphs and prove that the spectrum of normalised Laplacian for regular graphs is between 0 and 2. Later, we will see how we can normalise the Laplacian for irregular graphs and they also have spectrum from 0 and 2.

Consider an n -vertex d -regular graph $G = (V, E)$ with normalised Laplacian $N = I - \frac{1}{d} \cdot A$. Recall that the quadratic form $x^T N x = \sum_{v \in V} x_v^2 - \frac{1}{d} \sum_{uv \in E} 2x_u x_v = \frac{1}{d} (\sum_{v \in V} dx_v^2 - \sum_{uv \in E} 2x_u x_v) = \frac{1}{d} \sum_{uv \in E} (x_u - x_v)^2$ and the Rayleigh quotient of x is $R_N(x) = \frac{\sum_{uv \in E} (x_u - x_v)^2}{d \sum_{v \in V} x_v^2}$.

Theorem 6.13

Let $G = (V, E)$ be an n -vertex d -regular graph and let $N = I - \frac{1}{d}A$ be its normalised Laplacian with eigenvalues $\nu_1 \leq \nu_2 \leq \dots \leq \nu_n$. Then

- (i) $\nu_1 = 0$.
- (ii) $\nu_k = 0$ if and only if G has at least k connected components. In particular, the number of components in G equals the multiplicity of the eigenvalue 0.
- (iii) $\nu_n \leq 2$. The equality holds if and only if G has a bipartite component.



Proof

(i) By the Courant-Fisher theorem, $\nu_1 = \min_{x \neq 0} R_N(x) = \min_{x \neq 0} \frac{\sum_{uv \in E} (x_u - x_v)^2}{d \sum_{v \in V} x_v^2} \geq 0$. On the other hand, $N \cdot \mathbf{1} = 0$.

Hence, $\nu_1 = 0$.

(ii) (\Rightarrow) Suppose $\nu_k = 0$. We want to show the number of components is at least k . By the Courant-Fisher theorem, $\nu_k = \min_{\dim U = k} \max_{x \in U, x \neq 0} R_N(x) = \min_{\dim U = k} \max_{x \in U, x \neq 0} \frac{\sum_{uv \in E} (x_u - x_v)^2}{d \sum_{v \in V} x_v^2} = 0$. Thus, there exists a k -

dimension space U such that for any $x \in U, x \neq 0$, we have $R_N(x) = \frac{\sum_{uv \in E} (x_u - x_v)^2}{d \sum_{v \in V} x_v^2} = 0$. This means that if u and v are adjacent, then $x_u = x_v$. This further implies that x is constant over any connected component. Hence, the dimension of U ($= k$) is at most the number of components of G .

(\Leftarrow) Suppose the number of component is at least k . Consider the subspace U spanned by vectors that are constant on each component, then the dimension of U equals the number of components ($\geq k$). For every $x \in U$ and $x \neq 0$, $R_N(x) = 0$. Then by the Courant-Fisher theorem, $\nu_k = 0$.

(iii) By the Courant-Fisher theorem,

$$\begin{aligned} \nu_n &= \max_{x \neq 0} R_N(x) \\ &= \max_{x \neq 0} \frac{x^T N x}{x^T x} \\ &= \max_{x \neq 0} \frac{\sum_{uv \in E} (x_u - x_v)^2}{d \sum_{v \in V} x_v^2}. \end{aligned}$$

Moreover,

$$\begin{aligned}
 x^T N x &= \frac{1}{d} \sum_{uv \in E} (x_u - x_v)^2 \\
 &= \sum_{v \in V} x_v^2 - \frac{2}{d} \sum_{uv \in E} x_u x_v \\
 &= x^T x - \frac{2}{d} \sum_{uv \in E} x_u x_v \\
 &= 2x^T x - \frac{2}{d} \sum_{uv \in E} x_u x_v - x^T x \\
 &= 2x^T x - \frac{1}{d} \sum_{uv \in E} (x_u + x_v)^2.
 \end{aligned}$$

So $R_N(x) = 2 - \frac{\sum_{uv \in E} (x_u + x_v)^2}{d \sum_{v \in V} x_v^2} \leq 2$. Hence, $\nu_n \leq 2$.


If $\nu_n = 2$, then we must have

$$\sum_{uv \in E} (x_u + x_v)^2 = 0, \tag{6.1}$$

which means that if u and v are adjacent, then $x_u = -x_v$. Define $S = \{v : x_v > 0\}$, $T = \{v : x_v < 0\}$. Then $S \cup T$ sends no edge to $V \setminus (S \cup T)$, for otherwise such edge contributes positively to $\sum_{uv \in E} (x_u + x_v)^2$, contradict (6.1). Also, $N(S) \subseteq T$ and $N(T) \subseteq S$ by (6.1). Thus, $S \cup T$ induces union of bipartite component in G .

6.6 Examples about eigenvalues of Laplacian

Example 1: Complete graph K_n .

 **Exercise 6.3** For complete graph K_n , the Laplacian of K_n has

eigenvalue	eigenvector	multiplicity
0	$\mathbf{1}$	1
n	$\forall x : x \perp \mathbf{1}$	$n - 1$

Example 2: Star S_n .

Consider a star S_n with core vertex v_1 and its neighbours v_2, \dots, v_n . The Laplacian of S_n has

eigenvalue	eigenvector	multiplicity
0	$\mathbf{1}$	1
1	$\delta_{v_i} - \delta_{v_{i+1}} \ (2 \leq i \leq n - 1)$	$n - 2$
n	$\begin{pmatrix} -(n-1) \\ 1 \\ \vdots \\ 1 \end{pmatrix}$	1

where $\delta_v = \mathbf{1}_v$ is the indicator vector for a single vertex v .

Lemma 6.14

Let G be an n -vertex graph with two degree-1 vertices a and b both adjacent to another vertex c , then $x = \delta_a - \delta_b$ is an eigenvector of G with eigenvalue 1. ♥

🚩 **Exercise 6.4** Prove the above lemma. Check $(Lx)_v = x_v$.

We can apply the above lemma to all adjacent pairs of star S_n . So $v_2 - v_3, v_3 - v_4, \dots, v_{n-1} - v_n$ are all eigenvectors with eigenvalue 1 and they are all linearly independent. Now consider the largest eigenvalue λ_n of

Laplacian L of S_n . Recall that $L = \begin{pmatrix} n-1 & & & -1 \\ & 1 & & \\ & & \ddots & \\ -1 & & & 1 \end{pmatrix}$, then $tr(L) = 2n - 2 = \sum_{i=1}^n \lambda_i = n - 2 + 0 + \lambda_n$

Hence $\lambda_n = n$. Moreover, $y = \begin{pmatrix} -(n-1) \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ is the eigenvector with eigenvalue n . It is easy to check that y is orthogonal to $\mathbf{1}$ and $\delta_{v_i} - \delta_{v_{i+1}}$ ($2 \leq i \leq n-1$).

Example 3: Hypercube Q_d .

Definition 6.15

Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, $G_1 \times G_2$ is a graph with

$$V(G_1 \times G_2) = V_1 \times V_2,$$

$$E(G_1 \times G_2) = \{(x, y) \sim (x', y), \text{ where } xx' \in E_1\} \cup \{(x, y) \sim (x, y'), \text{ where } yy' \in E_2\}.$$
♣

Theorem 6.16

Let $G_1 = (V_1, E_1)$ be a graph with Laplacian eigenvalues $\lambda_1^{(1)} \leq \dots \leq \lambda_n^{(1)}$ and eigenvectors x_1, \dots, x_n . Let $G_2 = (V_2, E_2)$ be a graph with Laplacian eigenvalues $\lambda_1^{(2)} \leq \dots \leq \lambda_m^{(2)}$ and eigenvectors y_1, \dots, y_m . Then $G_1 \times G_2$ has eigenvalues $\lambda_i^{(1)} + \lambda_j^{(2)}$, for any $i \in [n]$ and $j \in [m]$, with eigenvectors $z_{i,j}$, where $(z_{i,j})_{(u,v)} = (x_i)_u \cdot (y_j)_v$. ♥

Let Q_d be a d -dimension hypercube. Consider Laplacian L_{Q_d} . When $d = 1$, Q_1 has

eigenvalue	eigenvector
0	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
2	$\begin{pmatrix} 1 \\ -1 \end{pmatrix}$


If Q_{d-1} has eigenvalue λ with eigenvector x , then $Q_d = Q_1 \times Q_{d-1}$ has

eigenvalue	eigenvector
λ	$\begin{pmatrix} x \\ x \end{pmatrix}$
$\lambda + 2$	$\begin{pmatrix} x \\ -x \end{pmatrix}$


Hence, Q_d has eigenvalues $2i$ for each $0 \leq i \leq d$ with multiplicity $\binom{d}{i}$. We can identify eigenvectors of L_{Q_d} with $V(Q_d)$: take $v \in V(Q_d) = \{0, 1\}^d$ and the corresponding eigenvector $x^{(v)}$ is for any $u \in V(Q_d)$, $x_u^{(v)} = (-1)^{v^T u}$.

6.7 Isoperimetry and the second eigenvalue

Definition 6.17

The boundary of a set $S \subset V$ is $\partial(S) = E_G(S, V \setminus S) = E_G(S, S^c)$. 


Definition 6.18

The isoperimetric number of $S \subset V$ is $\theta(S) = \frac{|\partial S|}{|S|}$, basically the average degree out of S . 

Note that $\theta(S) = \frac{|\partial S|}{|S|} = R_L(\mathbf{1}_S)$, where $\mathbf{1}_S$ is the characteristic function of the set S .

Definition 6.19


The isoperimetric number, or Cheeger constant of a graph G with $|G| = n$ is

$$h(G) = \min_{0 < |S| \leq n/2} \theta(S) = \min \left\{ \frac{|\partial S|}{|S|} : S \subset V(G), 0 < |S| \leq \frac{n}{2} \right\}. $$

Trivially, $h(G) > 0$ if and only if G is connected. The Cheeger constant $h(G)$ measures how small the "bottleneck" in G is. If "bottleneck" is large, then it is difficult to cut off the graph, then the graph is an expansion graph.

Theorem 6.20

For any graph G with $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ of Laplacian, we have

$$h(G) \geq \frac{\lambda_2}{2}. $$

Proof 1 It suffices to show that for any $S \subset V(G)$ of size sn , where $0 < s \leq \frac{n}{2}$, $n = |V(G)|$, we have $\theta(S) \geq (1-s)\lambda_2$.

By Courant-Fischer, we have that

$$\lambda_2 = \min_{\substack{x \neq 0 \\ x \perp \mathbf{1}}} R_L(x) = \min_{\substack{x \neq 0 \\ x \perp \mathbf{1}}} \frac{x^T Lx}{x^T x}.$$

Then

$$\forall x \perp \mathbf{1}, x \neq 0, \frac{x^T Lx}{x^T x} \geq \lambda_2. \tag{6.2}$$

We shall use the characteristic function of the set $S(\mathbf{1}_S)$, but first we center it to be orthogonal to $\mathbf{1}$.

So we will use

$$x = \mathbf{1}_S - s\mathbf{1}, \langle x, \mathbf{1} \rangle = 0$$

in eq. (6.2). As $L \cdot \mathbf{1} = 0$,

$$x^T Lx = \mathbf{1}_S^T L \mathbf{1}_S = e(S, S^c) = |\partial S|.$$

So we have

$$\frac{x^T Lx}{x^T x} = \frac{|\partial S|}{|S|(\mathbf{1} - s)} = \frac{|\partial S|}{s(1-s)n} \geq \lambda_2$$

Proof 2 We want to show that for any S with $|S| \leq \frac{n}{2}$, we have $\theta(S) \geq \frac{\lambda_2}{2}$. By Courant-Fischer we have

$$\lambda_2 = \min_{\substack{\dim U=2 \\ x \neq 0 \\ x \in U}} \max R_L(x) \quad (6.3)$$

We'll take $U = \text{Span of } \mathbf{1}_S \text{ and } \mathbf{1}_{S^c}$. Since $\mathbf{1}_S \perp \mathbf{1}_{S^c}$, we have $\dim U = 2$.

Suppose $z \neq 0, z \in U$ such that

$$\max_{\substack{x \neq 0 \\ x \in U}} R_L(x) = R_L(z).$$

We have then eq. (6.3) is saying $\lambda_2 \leq R_L(z)$.

Lemma 6.21

Let M be a positive semidefinite matrix and x, y be two orthogonal vectors. Then

$$R_M(x + y) \leq 2 \max \{R_M(x), R_M(y)\}.$$



Assuming this lemma holds and L is positive semidefinite matrix. Write $z = \alpha x + \beta y$ where $\alpha, \beta \in \mathbb{R}$ and $x = \mathbf{1}_S, y = \mathbf{1}_{S^c}$, then we have

$$\begin{aligned} \lambda_2 &\leq R_L(z) = R_L(\alpha x + \beta y) \\ &\leq 2 \max \{R_L(\alpha x), R_L(\beta y)\} \end{aligned}$$

As multiplying by a scalar does not change Rayleigh quotient, i.e. $R_L(\alpha x) = R_L(x)$. And we also have $R_L(x) = R_L(\mathbf{1}_S) = \theta(S)$. We can get

$$\begin{aligned} \lambda_2 &\leq 2 \max \{R_L(\alpha x), R_L(\beta y)\} \\ &= 2 \max \{R_L(x), R_L(y)\} \\ &= 2\theta(S). \end{aligned}$$

Proof of Lemma 6.21 Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be eigenvalues of M with corresponding eigenvectors v_1, v_2, \dots, v_n which are basis.

We write $x = \sum_{i=1}^n a_i v_i, y = \sum_{i=1}^n b_i v_i$, then

$$\begin{aligned} R_M(x + y) &= \frac{(\sum_{i=1}^n (a_i + b_i) v_i)^T M (\sum_{i=1}^n (a_i + b_i) v_i)}{\|x + y\|^2} \\ &= \frac{\sum_{i=1}^n \lambda_i (a_i + b_i)^2}{\|x + y\|^2} \\ &\leq \frac{\sum_{i=1}^n \lambda_i \cdot 2(a_i^2 + b_i^2)}{\|x\|^2 + \|y\|^2} \\ &= \frac{2R_M(x)\|x\|^2 + 2R_M(y)\|y\|^2}{\|x\|^2 + \|y\|^2} \\ &\leq 2 \max \{R_M(x), R_M(y)\}. \end{aligned}$$

Cheeger constant $h(G)$ is often useful when dealing with vertex subsets. For edge expansion, it's more convenient to work with a notion called conductance. For this, we need to define normalized Laplacian for general graphs.

Recall for d -regular graphs, $L = dI - A = D - A, N = I - \frac{1}{d}A$. It turns out for general graphs,

$$\begin{aligned} N &= D^{\frac{1}{2}} L D^{\frac{1}{2}} \\ &= D^{\frac{1}{2}} (D - A) D^{\frac{1}{2}} \\ &= I - D^{\frac{1}{2}} A D^{\frac{1}{2}}. \end{aligned}$$

Definition 6.22

For a d -regular graph $G = (V, E)$, the conductance (or edge expansion) of a set $S \subseteq V$ is

$$\phi(S) = \frac{\partial S}{d|S|} = R_N(\mathbf{1}_S) = \frac{\sum_{uv \in E} (x_u - x_v)^2}{d \sum_{v \in V} x_v^2}$$

i.e. $\phi(S)$ is the average fraction of neighbours (of vertices of S) lying outside S .

The conductance (edge expansion) of a cut is

$$\phi(S, S^c) = \max\{\phi(S), \phi(S^c)\}.$$

The conductance (edge expansion) of G is

$$\phi(G) = \min_{1 \leq |S| \leq \frac{n}{2}} \phi(S) = \min_S \phi(S, S^c).$$



If there exists a polynomial time approximation with constant-factor, then there is an approximation ratio. Find sparse cut has many applications. The first is clustering problem and image segmentation. The second is divide and conquer algorithm.

Fiedler's algorithm (1970s) is particularly useful when applied to x =eigenvectors of the second eigenvalue of normalized Laplacian ν_2 .

First step is sorting vertices so that $x_{v_1} \leq x_{v_2} \leq \dots \leq x_{v_n}$. This step runs in $O(|V| \log |V|)$.

Second step is finding k make $\phi(\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_n\})$ minimize. Next we only need to check the how long it takes. Let $e_k = e(\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_n\})$, $e_{k+1} = e(\{v_1, \dots, v_{k+1}\}, \{v_{k+2}, \dots, v_n\})$, so $e_{k+1} = e_k - a + b$, where $a = |N(v_{k+1}) \cap \{v_1, \dots, v_k\}|, b = |N(v_{k+1}) \cap \{v_{k+2}, \dots, v_n\}|$. Thus the k^{th} step of finding requires $d(v_{k+1})$, the total time of finding is $\sum d(v_{k+1}) = O(|E|)$.

So Fiedler algorithm runs in $O(|E| + |V| \log |V|)$.

Theorem 6.23. (Cheeger's inequality)

Let $G = (V, E)$ be a regular graph and ν_i be eigenvalues of its normalized Laplacian N . Then

$$\frac{\nu_2}{2} \leq \phi(G) \leq \sqrt{2\nu_2} = \sqrt{2R_N(x)}.$$



In Theorem 6.20, we have $h(G) \geq \frac{\lambda_2}{2}$. So for d -regular graph and normalized Laplacian N , we have $\phi(G) \geq \frac{\nu_2}{2}$.

For the other part, we shall see an analysis of Fiedler's algorithm by Trevisan and show that Fiedler's algorithm produce the cut $\phi(S, S^c) \leq \sqrt{2\nu_2}$.

Trevisan's idea Define a clever distribution to cut randomly and show in expectation (w.r.t. this distribution), the cut is sparse (in terms of ν_2).

For random $t \in [0, 1]$, t^2 is uniformly distributed in $[0, 1]$, i.e. the probability density function is $f(x) = 2x$.

Theorem 6.24. (robust version)

Let $x \perp \mathbf{1}$ and let the cut (S, S^c) be the minimizer in Fiedler's algorithm. Then

$$\phi(S, S^c) \leq \sqrt{2R_N(x)}.$$



Remark This robust version can be used on approximate eigenvectors (i.e. vectors with small Rayleigh quotient) and produce a sparse cut.

Consider the first non-negative vector, we have the following lemma:

Lemma 6.25

Let $y \in R_{\geq 0}^V$, $0 \leq y_{v_1} \leq \dots \leq y_{v_n} = 1$. Then $\exists t > 0$ such that

$$\phi(S_t) := \phi(\{v : y_v \geq t\}) \leq \sqrt{2R_N(y)}.$$



Proof As multiplying by a scalar does not change Rayleigh quotient, i.e. $R_N(cy) = R_N(y)$, we may assume $y_{v_n} = 1$. Let $t \in [0, 1]$ be a random variety with probability density function $f(x) = 2x$, so that t^2 is uniformly distributed in $[0, 1]$. It suffices to prove

$$\frac{\mathbb{E}_t |\partial S_t|}{\mathbb{E}(d|S_t|)} \leq \sqrt{2R_N(y)} \quad (6.4)$$

Indeed,

$$\begin{aligned} \text{eq. (6.4)} &\Leftrightarrow \mathbb{E} |\partial S_t| \leq \sqrt{2R_N(y)} \mathbb{E}(d|S_t|) \\ &\Leftrightarrow \mathbb{E} (|\partial S_t| - \sqrt{2R_N(y)} d|S_t|) \leq 0. \end{aligned}$$

Then we can get: \exists a choice of $t \in [0, 1]$ such that

$$\begin{aligned} |\partial S_t| - \sqrt{2R_N(y)} d|S_t| &\leq 0 \\ \Leftrightarrow \frac{|\partial S_t|}{d|S_t|} &= \phi(S_t) \leq \sqrt{2R_N(y)} \end{aligned}$$

as desired.

Then let's see the denominator:

$$\begin{aligned} \mathbb{E}(d|S_t|) &= d \sum_{v \in V} \mathbb{P}(v \in S_t) \\ &= d \sum_{v \in V} \mathbb{P}(t \leq y_v) \\ &= d \sum_{v \in V} \mathbb{P}(t^2 \leq y_v^2) \\ &= d \sum_{v \in V} y_v^2. \end{aligned}$$

Next we analyse the numerator.

$$\begin{aligned}
 \mathbb{E}|\partial S_t| &= \sum_{uv \in E} \mathbb{P}(uv \in \partial S_t) \\
 &= \sum_{uv \in E} \mathbb{P}(y_u \leq t \leq y_v) \\
 &= \sum_{uv \in E} \mathbb{P}(y_u^2 \leq t^2 \leq y_v^2) \\
 &= \sum_{uv \in E} |y_u^2 - y_v^2| \\
 &= \sum_{uv \in E} |y_u - y_v|(y_u + y_v) \\
 &\leq \sqrt{\sum_{uv \in E} (y_u - y_v)^2} \sqrt{\sum_{uv \in E} (y_u + y_v)^2} \text{ (Cauchy - Schwarz inequality)} \\
 &\leq \sqrt{\sum_{uv \in E} (y_u - y_v)^2} \sqrt{2d \sum_{v \in V} y_v^2}.
 \end{aligned}$$

Therefore, we have

$$\frac{\mathbb{E}|\partial S_t|}{\mathbb{E}(d|S_t|)} \leq \frac{\sqrt{\sum_{uv \in E} (y_u - y_v)^2} \sqrt{2d \sum_{v \in V} y_v^2}}{d \sum_{v \in V} y_v^2} = \sqrt{\frac{2 \sum_{uv \in E} (y_u - y_v)^2}{\sum_{v \in V} y_v^2}} = \sqrt{2R_N(y)}.$$

Lemma 6.26

Let $x \perp \mathbf{1}$. Then there exists $y \in R_{\geq 0}^V$ such that

(i) $|\{y_i > 0\}| \leq \frac{n}{2}$;

(ii) $R_N(y) \leq R_N(x)$;

(iii) Cuts considered in Fiedler's algorithm with input y are the same as those with input x .



Proof of Theorem 6.24. : Let $x \perp \mathbf{1}$ and let the cut (S, S^c) be the minimizer in Fiedler's algorithm. Take y from Lemma 6.26 and let S_t be the set returned from Lemma 6.25. We have

$$\phi(S_t) = \phi(\{v : y_v \geq t\}) \stackrel{\text{lem6.25}}{\leq} \sqrt{2R_N(y)} \stackrel{\text{lem6.26(ii)}}{\leq} \sqrt{2R_N(x)}.$$

On the other hand, $\phi(S_t) \stackrel{\text{lem6.26(i)}}{=} \phi(S_t, S_t^c) \stackrel{\text{lem6.26(iii)}}{\geq} \phi(S, S^c)$. (as (S, S^c) is the minimizer.) Combines the above inequalities, we have done.

It reminds to show that Lemma 6.26.

Proof of Lemma 6.26. : First, we use the standard trick so-called "shifting": for every $c \in \mathbb{R}$, we just take the shifting $x + c\mathbf{1}$, and observe the change of the Rayleigh quotient of x . Note that for every $c \in \mathbb{R}$, we claim that

$$R_N(x + c\mathbf{1}) \leq R_N(x).$$

In order to see it, just analyze the formula of the Rayleigh quotient $R_N(x) = \frac{x^T N x}{\|x\|_2^2}$. For every regular graph, the all-1 vector is a eigenvector of its normalised Laplacian operator, $N \cdot \mathbf{1} = 0$, so both numerator are same ("shifting" part vanishing). We only need to compare denominator. $\|x + c\mathbf{1}\|^2 = \|x\|^2 + \|c\|^2$, as $x \perp \mathbf{1}$. Thus denominator is increasing.

Let $m = \text{median value of entries of } x$. We define $z = x - m\mathbf{1}$, so $R_N(z) \leq R_N(x)$. Note that z has at

most $\frac{n}{2}$ positive entries and at most $\frac{n}{2}$ negative entries. Write $z = z^+ - z^-$,

$$z_v^+ =: \begin{cases} x_v, & \text{positive } v, \\ 0, & \text{otherwise.} \end{cases} \quad z_u^- =: \begin{cases} -x_u, & \text{negative } u, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to check that cuts in Fiedler's algorithm with input z^+ or z^- same as those with input x .

Left to show $R_N(z) \geq \min\{R_N(z^+), R_N(z^-)\}$.

$$R_N(z) = \frac{\sum_{uv \in E} (z_u - z_v)^2}{\|z\|^2} = \frac{\sum_{uv \in E} [(z_u^+ - z_v^+) - (z_u^- - z_v^-)]^2}{\|z^+\|^2 + \|z^-\|^2}.$$

Claim $[(z_u^+ - z_v^+) - (z_u^- - z_v^-)]^2 \geq (z_u^+ - z_v^+)^2 + (z_u^- - z_v^-)^2$.

With this claim,

$$\begin{aligned} R_N(z) &\geq \frac{\sum_{uv \in E} [(z_u^+ - z_v^+)^2 + (z_u^- - z_v^-)^2]}{\|z^+\|^2 + \|z^-\|^2} = \frac{R_N(z^+)\|z^+\|^2 + R_N(z^-)\|z^-\|^2}{\|z^+\|^2 + \|z^-\|^2} \\ &\geq \min\{R_N(z^+), R_N(z^-)\}. \end{aligned}$$

6.8 Application for isoperimetry in hypercube

Recall.

- Q_d : d -dim hypercube.

L denotes the Laplacian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

N denotes the normalised Laplacian matrix with eigenvalues $\nu_1 \leq \nu_2 \leq \dots \leq \nu_n$.

eigenvalue ($0 \leq i \leq d$)	multiplicity
$2i$	$\binom{d}{i}$
$\frac{2i}{d}$	$\binom{d}{i}$

- $\theta(G) \geq \frac{\lambda_2}{2}$.

Theorem 6.27. (Harper, 1976)

The isoperimetric number of Q_d is

$$\theta(Q_d) \geq 1.$$



Proof We know that $\lambda_2 = 2$ and $\theta(Q_d) \geq \frac{\lambda_2}{2} = 1$.

The lower bound is tight: consider a copy of Q_{d-1} inside.

- Conductance:

$$\phi(Q_d) \geq \frac{\nu_2}{2} = \frac{1}{d}.$$

Using multicommodity flow method,

Theorem 6.28. (Babai-Szegedy, 1992)

If G is connected, edge-transitive with diameter D , then

$$\phi(G) \geq \frac{1}{D}.$$



Note that $\text{diam}(Q_d)=d$, thus [Babai-Szegedy] implies that $\phi(Q_d) \geq \frac{1}{d}$.

6.9 Normalised Laplacian for general graph

- For a general graph G , the Laplacian matrix of G is $L = D - A_G$, where $D = \begin{pmatrix} d(v_1) & & 0 \\ & d(v_2) & \\ 0 & & \ddots \\ & & & d(v_n) \end{pmatrix}$.

- Normalised Laplacian for general graph G is $N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$,

$$\text{where } D = \begin{pmatrix} \frac{1}{\sqrt{d(v_1)}} & & 0 \\ & \frac{1}{\sqrt{d(v_2)}} & \\ 0 & & \ddots \\ & & & \frac{1}{\sqrt{d(v_n)}} \end{pmatrix}.$$

Recalling that for a d -regular graph G ,

- $N = I - \frac{1}{d} A$;
- $\nu_i \in [0, 2]$ for all eigenvalues ν_i of the normalised Laplacian matrix of G ;
- Multiplications with eigenvalues equal to 0 are equal to the number of connected components;
- $\nu_n = 2 \Leftrightarrow$ there exists a bipartite component.

For irregular graph, the above “nice” properties still holds. Let us mention some details in previous proof of regular graph. For d -regular graph,

$$R_N(x) = \frac{x^T N x}{x^T x} = 2 - \frac{\sum_{uv \in E} (x_u + x_v)^2}{d \sum_{v \in V} x_v^2}.$$

If the concerned graph is irregular, the above formula suggests that

$$2 - \frac{\sum_{uv \in E} (x_u + x_v)^2}{\sum_{v \in V} \mathbf{d}(\mathbf{v}) x_v^2} = \frac{\sum_{uv \in E} (x_u - x_v)^2}{\sum_{v \in V} d(v) x_v^2} = \frac{x^T L x}{x^T D x} = \frac{x^T L x}{(x^T D^{\frac{1}{2}})(D^{\frac{1}{2}} x)}.$$

As for numerator, we need

$$x^T L x = (x^T D^{\frac{1}{2}}) D^{-\frac{1}{2}} L D^{-\frac{1}{2}} (D^{\frac{1}{2}} x).$$

Then

$$\frac{x^T L x}{(x^T D^{\frac{1}{2}})(D^{\frac{1}{2}} x)} = R_{D^{-\frac{1}{2}} L D^{-\frac{1}{2}}}(D^{\frac{1}{2}} x) = \frac{x^T L x}{x^T x}.$$

Some properties.

- The point is that the map $x \mapsto D^{\frac{1}{2}} x$ is bijective. Note that $\nu_1 \leq \dots \leq \nu_n$ for $N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$.

$$\nu_k = \min_{\dim U=k} \max_{\substack{x \in U \\ x \neq 0}} R_N(x) = \min_{\dim U=k} \max_{\substack{x \in U \\ x \neq 0}} R_N(D^{\frac{1}{2}} x) = \dots = \frac{x^T L x}{x^T D x}.$$

- $d^{\frac{1}{2}} = \begin{pmatrix} \sqrt{d(v_1)} \\ \sqrt{d(v_2)} \\ \vdots \\ \sqrt{d(v_n)} \end{pmatrix}$ is an eigenvector for 0.

Check: $N \cdot d^{-\frac{1}{2}} = D^{-\frac{1}{2}}L(D^{-\frac{1}{2}} \cdot d^{\frac{1}{2}}) = D^{-\frac{1}{2}}L \cdot \mathbf{1} = 0$.

• [C-F]

$$\nu_2 = \min_{\substack{x \perp d^{-\frac{1}{2}} \\ x \neq 0}} \frac{x^T N x}{x^T x} = \min_{\substack{y \perp d \\ y \neq 0}} \frac{y^T N y}{y^T D y}.$$

• Define *volume* of $S = \text{vol}(S) = \sum_{v \in S} d(v)$.

$$\phi(S) = \frac{|\partial S|}{\text{vol}(S)}, \phi(S, S^c) = \max\{\phi(S), \phi(S^c)\}.$$

• $\phi(G) = \min_{|S| \leq \frac{n}{2}} \phi(S) = \min_S \phi(S, S^c)$.

Theorem 6.29. (Cheeger's inequality)

Let $G = (V, E)$ be a graph and ν_i be eigenvalues of its normalized Laplacian N . Then

$$\frac{\nu_2}{2} \leq \phi(G) \leq \sqrt{2\nu_2}.$$



• Recall Spectral theorem: Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric $n \times n$ matrix. Then M has n real eigenvalues (not necessary distinct) $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and n orthonormal real eigenvectors $\{\nu_1, \dots, \nu_n\} \in \mathbb{R}^n$ where ν_i is the eigenvector of λ_i .

Corollary 6.30. (Eigenvalue decomposition)

Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric $n \times n$ matrix with eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$. Then

$$M = V \Lambda V^T = \sum_{k=1}^n \lambda_k v_k v_k^T, \text{ where } V = (v_1, \dots, v_n), \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$



Proof Note that $MV = M(v_1, \dots, v_n) = (\lambda_1 v_1, \dots, \lambda_n v_n) = V\Lambda$. Then $M = V\Lambda V^T$. (Note that $V^{-1} = V^T$.)

Perron–Frobenius theorem: If $A = (a_{ij})$ is an $n \times n$ real matrix with non-negative entries $a_{ij} \geq 0$ and irreducible, then there is a real eigenvalue r of A such that

$$\min_i \sum_j a_{ij} \leq r \leq \max_i \sum_j a_{ij}$$

and any other eigenvalue λ satisfies $|\lambda| \leq r$. A matrix is *reducible* if there is a subset $I \subseteq [n]$ such that $a_{ij} = 0$ for all $i \in I$ and $j \notin I$. In particular, an adjacency matrix of a graph is irreducible if and only if the graph is connected.

Theorem 6.31. (Perron–Frobenius symmetric version)

Let $G = (V, E)$ be a connected graph with adjacency matrix A and eigenvalues $\alpha_1 \geq \dots \geq \alpha_n$. Then

- (i) The largest eigenvalue α_1 has a strictly positive eigenvector.
- (ii) For every $i \in [n]$, $\alpha_i \in [-\alpha_1, \alpha_1]$.
- (iii) $\alpha_1 > \alpha_2$.



Lemma 6.32

Let $G = (V, E)$ be a connected graph with adjacency matrix A . If x is a non-negative eigenvector, then x has to be a positive vector.



Proof Suppose that the conclusion is not true. Then there exists $uv \in E$ such that $x(v) = 0$ and $x(u) > 0$. Note that $0 < \sum_{u \in N(v)} x(u) = (Ax)_v = (\alpha x)_v = 0$, contradict!

Proof of Theorem 6.31 :

- For (i). Let A be a real $n \times n$ adjacency matrix with eigenvalues $\alpha_1, \dots, \alpha_n$ and corresponding eigenvectors x_1, \dots, x_n . Let $y \in \mathbb{R}_{\geq 0}^V$ such that $y(v) = |x_1(v)|$ (Thus $y^T y = x_1^T x_1$). We claim that y is an eigenvector of α_1 . Then by Lemma 6.32, $y \in \mathbb{R}_{> 0}^V$ is a positive eigenvector.

$$\alpha_1 = x_1^T A x_1 = \sum_{u,v} x_1(u) A(u,v) x_1(v) \leq \sum_{u,v} |x_1(u)| A(u,v) |x_1(v)| = y^T A y = R_A(y) \leq \alpha_1.$$

That tells us that all of “the equal” above hold. Then y is an eigenvector of α_1 .

- For (ii). It suffices to show that $\alpha_n \geq -\alpha_1$. Let $z \in \mathbb{R}_{\geq 0}^V$ such that $z(v) = |x_n(v)|$.

$$|\alpha_n| = |x_n^T A x_n| \leq \left| \sum_{u,v} x_n(u) A(u,v) x_n(v) \right| \leq \sum_{u,v} |x_n(u)| A(u,v) |x_n(v)| = z^T A z \leq \alpha_1.$$

- For (iii). Let $\omega \in \mathbb{R}_{\geq 0}^V$ such that $\omega(u) = |x_2(u)|$. Similar analysis in (ii), we have

$$\alpha_2 = x_2^T A x_2 \leq \omega^T A \omega \leq \alpha_1.$$

If $\alpha_2 = \alpha_1$, then the above equality holds. Thus ω is a non-negative vector of α_1 , by Lemma 6.32, ω is a positive vector. Then x_2 has no zero entry. As G is connected and $x_2 \perp x_1$ (assume that x_1 is positive by Part (i)), there exists $uv \in E$ such that $x_2(u) > 0$ and $x_2(v) < 0$. Then such u, v contributes negative to $x_2^T A x_2$. So $x_2^T A x_2 < \omega^T A \omega$. It means that $\alpha_2 < \alpha_1$.

Corollary 6.33

For every connected G with adjacency matrix A , if a positive vector x is an eigenvector of A , then the corresponding eigenvalue is α_1 . ♥

Proof Let $Ax = \alpha_i x$ for some $i \in [n]$. By Theorem 6.31, there exists $y \in \mathbb{R}_{> 0}^V$ which is a eigenvector of α_1 . As A is a symmetric matrix, then

$$\alpha_i x^T y = \langle Ax, y \rangle = \langle x, Ay \rangle = x_1^T x^T y.$$

Note that $x^T y > 0$, and we have $\alpha_i = \alpha_1$.

6.10 Random Walk

- A *random walk* on a graph G is a random process that starts at a vertex in G , and in each step, a neighbor is uniformly and randomly selected to move.
- Let $(P)_{ij} = p(i, j)$ be a *transition matrix* of a random walk on an undirected connected graph $G = (V, E)$ with n vertices $V = \{1, \dots, n\}$. That is, $p(i, j) = 1/d(i)$ if $\{i, j\} \in E$, and $p(i, j) = 0$ otherwise (in other words, $p(i, j) = \Pr[i \rightarrow j]$).
- For every d -regular graph G with adjacency matrix A , $P = \frac{1}{d}A$.
- For each general graph G with adjacency matrix A ,

$$P = D^{-1}A = D = \begin{pmatrix} \frac{1}{d(v_1)} & \cdots & \frac{1}{d(v_1)} \\ \vdots & & \vdots \\ \frac{1}{d(v_i)} & \cdots & \frac{1}{d(v_i)} \\ \vdots & & \vdots \\ \frac{1}{d(v_n)} & \cdots & \frac{1}{d(v_n)} \end{pmatrix} \text{ where } D = \begin{pmatrix} d(v_1) & & & 0 \\ & d(v_2) & & \\ & & \ddots & \\ 0 & & & d(v_n) \end{pmatrix}.$$

- Consider an initial distribution $p_0 = \begin{pmatrix} p_0(v) \\ \vdots \end{pmatrix}$, where $p_0(v) = \Pr[\text{A random walk starting at } v]$ and write p_t for the distribution at time t . We will see that

$$p_t^T = p_0^T \cdot P^t \Leftrightarrow (P^T)^t \cdot p_0 = p_t.$$

It suffices to show that $p_{t+1}^T = p_t^T \cdot P$. We often write $(p_{t+1})_v = p_{t+1}(v)$

$$\begin{aligned} p_{t+1}(v) &= \Pr[\text{arrive at } v \text{ at time } t+1] = \sum_{u \in N(v)} \Pr[\text{arrive at } u \text{ at time } t] \cdot \Pr[u \rightarrow v] \\ &= \sum_{u \in N(v)} p_t(u) P(u, v) = (p_t^T \cdot P)_v = (P^T \cdot p_t)_v. \end{aligned}$$

- $p^t(i, j) = \Pr(\text{start at } i \text{ and arrive at } j \text{ at time } t) = \delta_i^T \cdot P^t = p_t^T.$

Definition 6.34

Stationary distribution π for a random walk with transition matrix P is

$$P^t \cdot \pi = \pi.$$



Exercise 6.5

Let

$$\pi = \frac{d}{\sum_{i=1}^n d(v_i)} = \frac{d}{\mathbf{1}^T \cdot d}, \text{ where } d = \begin{pmatrix} d(v_1) \\ d(v_2) \\ \vdots \\ d(v_n) \end{pmatrix}.$$

It means that at stationary distribution, every vertex is visited with probability proportional to its degree. If G is regular, then $\pi = \frac{1}{n} \cdot \mathbf{1}$.

Here, the question we care about is that: does random walk converge to stationary distribution? If yes, how fast? We shall see a connection between this and spectral gap of the transition matrix P . So let us first look at the spectral of P .

Recall that $P = D^{-1}A$ in general is not symmetry, so we don't have orthogonal eigenvectors for P . However, P is similar to the symmetry matrix $\tilde{A} = D^{1/2}AD^{1/2}$. (Define X to be similar with Y if $X = Q^{-1}YQ$. Similar matrices have the same set of eigenvalues.) Here, we can write $P = D^{1/2}\tilde{A}D^{1/2}$ or $\tilde{A} = D^{1/2}PD^{1/2}$.

Lemma 6.35

Given a connected graph G , let \tilde{A} be the normalised adjacency matrix and P be the transition matrix, then the vector x is the eigenvector of \tilde{A} with corresponding eigenvalue α if and only if $y = D^{-1/2}x$ is the eigenvector of P with corresponding eigenvalue α .



Proof The proof is simple. So we only give the proof of the necessity, and the sufficiency is left to the reader. Since $P = D^{1/2}\tilde{A}D^{1/2}$, we obtain $Py = D^{1/2}\tilde{A}D^{1/2}D^{-1/2}x = D^{-1/2}\alpha x = \alpha y$, and we are done.

Lemma 6.36

The degree vector d is a Perron vector of P^T with eigenvalue 1.



Proof $d^T P = d^T D^{-1}A = \mathbf{1}^T A = d^T.$

Corollary 6.37

The spectral of P is in $[-1, 1]$.

**Definition 6.38**

A lazy random walk is a normal random walk with 1/2 of the time staying put and 1/2 of the time doing the same as random walk.



The transition matrix of a lazy random walk is $\tilde{P} = P/2 + I/2$, and the spectral of \tilde{P} is in $[0, 1]$.

Remark

- Lazy random walk converges on all graphs.
- All eigenvalue non-negative, convergence (of the random walk) only depends on α_2 or the spectral gap $1 - \alpha_2$.

Next we shall prove that for any non-bipartite connected graph G , the random walk (with any initial distribution P_0) converges to the stationary distribution π .

Remark

- Not true for bipartite graphs because of the parity issue.
- Can consider instead lazy random walk for bipartite graph.

We shall see that in fact the rate of convergence is exponential when having spectral gap.

Definition 6.39

Mixing matrix of a random walk is $\mu = \lim_{t \rightarrow \infty} \sup \max_{i,j} |P^t(i, j) - \pi(j)|^{1/t}$.

**Theorem 6.40**

Let G be a graph with transition matrix P and eigenvalues $\alpha_1 \geq \dots \geq \alpha_n$. Then any starting vertex i and any other vertex j , we have $|P^t(i, j) - \pi(j)| \leq \sqrt{\frac{\pi(j)}{\pi(i)}} \cdot \mu^t$, where $t > 0$ and $\mu = \max\{|\alpha_2|, |\alpha_n|\}$.

More generally, for any $S \subseteq V(G)$, $|P \left(\begin{array}{c} \text{start at } i \\ \text{end in } S \text{ at time } t \end{array} \right) - \pi(S)| \leq \sqrt{\frac{\pi(S)}{\pi(i)}} \cdot \mu^t$.



If G is connected, then $\alpha_2 < \alpha_1$, if G is non-bipartite, then $\alpha_n > -1$.

Corollary 6.41

For every connected and non-bipartite G , mixing rate is $\mu \leq \max\{|\alpha_2|, |\alpha_n|\}$.



Proof of Theorem 6.40 : Recall $P = D^{-1}A = D^{-1/2}\tilde{A}D^{1/2}$, where $\tilde{A} = D^{-1/2}AD^{-1/2}$. As \tilde{A} is symmetry, we can write it as $\tilde{A} = \sum_{k=1}^n \alpha_k v_k v_k^T$, where v_1, \dots, v_k are orthogonal with eigenvalue $\alpha_1 \geq \dots \geq \alpha_n$. Now

$$P^t = (D^{-1/2}\tilde{A}D^{1/2})^t = D^{-1/2}\tilde{A}^t D^{1/2} = \sum_{k=1}^n \alpha_k^t D^{-1/2} v_k v_k^T D^{1/2}.$$

So

$$P^t(i, j) = e_i^T P^t e_j = \sum_{k=1}^n \alpha_k^t e_i^T D^{-1/2} v_k v_k^T D^{1/2} e_j = \sqrt{\frac{d(v_j)}{d(v_i)}} \sum_{k=1}^n \alpha_k^t e_i^T v_k v_k^T e_j.$$

Let us look at the first term in sum. That is, $\alpha_1^t e_i^T v_1 v_1^T e_j = \sqrt{\pi(i)} \sqrt{\pi(j)}$. This implies the first term

of above equality is $\pi(j)$, and then $|P^t(i, j) - \pi(j)| \leq \sqrt{\frac{\pi(j)}{\pi(i)}} \left| \sum_{k=2}^n \alpha_k^t e_i^T v_k v_k^T e_j \right|$. What is left is to show $\left| \sum_{k=2}^n \alpha_k^t e_i^T v_k v_k^T e_j \right| \leq \mu^t$, where $\mu = \max\{|\alpha_2|, |\alpha_n|\}$. Then

$$\left| \sum_{k=2}^n \alpha_k^t e_i^T v_k v_k^T e_j \right| \leq \mu^t \sum_{k=1}^n |e_i^T v_k| \cdot |v_k^T e_j| \leq \mu^t \sqrt{\sum_{k=1}^n |e_i^T v_k|^2} \sqrt{\sum_{k=1}^n |v_k^T e_j|^2} = \mu^t \|e_i\| \|e_j\| = \mu^t,$$

where the second inequality holds by Cauchy Schwartz inequality, and the last equality holds, as v_1, \dots, v_n are orthogonal. This completes the proof.

Remark Usually by considering the lazy random walk (so $\tilde{P} = P/2 + I/2$, and the spectral of \tilde{P} is in $[0, 1]$), the mixing rate is only related to α_2 or the spectral gap $1 - \alpha_2$.

Exercise 6.6 In the above proof, show $v_1 = \pi^{1/2}$, $v_1(i) = \sqrt{\pi(i)}$, i.e., $\pi^{1/2}$ is a Perron eigenvector of P .

- A more quantitative version, mixing time, measures how many steps needed to get close to stationary distribution? What is a good measure of distance between distributions?
- Natural candidate: Euclidean l_2 - norm, it means $\|x\|_2 = (\sum x_i^2)^{1/2}$. But it is not ideal here: Consider $S = [\frac{n}{2}]$, $x = \frac{2}{n} \cdot \mathbf{1}_S, y = \frac{2}{n} \cdot \mathbf{1}_{S^c}$. $\|x - y\|_2 = \sqrt{\sum (\frac{2}{n})^2} = \frac{2}{\sqrt{n}} \rightarrow 0, n \rightarrow \infty$. But x, y as distribution are very different.
- Better one: a scaled l_1 - norm.

Definition 6.42

The total variation distance between x, y is $\|x - y\|_{TV} = \max_{S \subseteq V} |\sum_{v \in S} x(v) - \sum_{v \in S} y(v)|$. Which is the maximal difference of probabilities of events with respect to x and y .



Exercise 6.7 $\|x - y\|_{TV} = \frac{1}{t} \|x - y\|_1$. In the above example, $\|x - y\|_{TV} = 1$. In general, distributions with disjoint support have distance 1.

Definition 6.43

A random walk mixes at time t if

$$\|P_t - \Pi\|_{TV} < \frac{1}{4} \tag{*}$$

Call such t the mixing time.



Remark The constant $\frac{1}{4}$ is not that important. Any small constant will do.

It means that: for most of $i \in V, \frac{\Pi(i)}{2} \leq P_t(i) \leq 2\Pi(i)$

It has an useful equivalent form: $\forall i \in V, |P_t(i) - \Pi(i)| \leq \frac{\Pi(i)}{2}$.

Consider the simpler d -regular $G = (V, E)$ case do lazy random walk, no need to worry about α_n .

Recall $1 \geq \alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n \leftarrow \tilde{A}$

$$\tilde{L} = N = I - \tilde{A} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

$$\gamma_2 = 1 - \alpha_2 > 0$$

Assume there exists spectral gap $|P^t(i, j) - \Pi(j)| \leq \sqrt{\frac{d(j)}{d(i)}} \cdot \mu^t$

We want to show $|P^t(i, j) - \Pi(j)| \leq \frac{\Pi(j)}{2}$

So we want to show $|P^t(i, j) - \Pi(j)| \leq \sqrt{\frac{d(j)}{d(i)}} \cdot \mu^t \leq \frac{d(j)}{2 \cdot \sum_{i=1}^n d(i)}$

Let $\mu = \alpha_2 = 1 - \gamma_2$,

$$e^{-\gamma_2 t} \approx (1 - \gamma_2)^t \leq \sqrt{\frac{d(j)}{d(i)}} \cdot e^{-\gamma_2 t} < \frac{d(j)}{2 \cdot \sum_{i=1}^n d(i)}, \text{ and } \sum_{i=1}^n d(i) = 2e(G)$$

$$\Leftrightarrow e^{-\gamma_2 t} \geq \frac{4e(G)}{\sqrt{d(i)-d(j)}}.$$

If G is d -regular, $\frac{4e(G)}{\sqrt{d(i)-d(j)}} = 2n$. Then $t \geq \frac{1}{\gamma_2} \cdot \log \frac{4e(G)}{\sqrt{d(i)-d(j)}} = \frac{1}{\gamma_2} \cdot \log(2n)$.

Remark Sometimes, the $\log n$ term can be avoided as we use α_2 to bound all $\alpha_i, i \geq 2$.

If there is eigenvalue decay, then we can take advantage to improve the bound.

Example.

Q_d is a d -regular graph, where $n \geq 2^d, \gamma_2 = \frac{2}{d}$, the above bound on mixing time yields

$$t = O\left(\frac{\log n}{\gamma_2}\right) = O(d^2).$$

But the mixing time for hypercube is known to be $\Theta(d \log d)$.

Then I want to mention that conductance and mixing time

$$\frac{1}{\Phi(G)} \leq t_{mix} \leq \frac{\log n}{\Phi(G)^2}$$

Where $\Phi(G) = \min_{\substack{|S| \leq \frac{n}{2} \\ |S| \geq \frac{n}{2}}} \frac{|\partial S|}{d|S|}$, G is a d -regular graph.

The second inequality was proved by *Lovász – Simonvits*.

The first inequality is nature.

$$\mathbb{E}(\# \text{ steps to cross}) \sim \frac{1}{\mathbb{P}(\text{cross})} = \frac{1}{\Phi(G)}$$

This offers a probabilistic way to define an expander :if the random walk on it is fast mixing.

Application.

Make use of the following fact:

"Random walk on expanders resembles independent sampling."

From computational aspect ,it is used in error reduction for probabilistic algorithm.

Suppose we have a probabilistic algorithm A ,

A :uses K random bits $\rightarrow \begin{cases} \text{right answer with probability } \frac{2}{3} \\ \text{wrong answer with probability } \frac{1}{3} \end{cases}$

To boost the probability, we can run it t times and take the majority answer.

$$\mathbb{P}(\text{we get the wrong answer}) \leq (1 - c)^t, c > 0$$

The cost is $t \cdot K$ random bits.

More economic: $K + O(t)$

We can consider $G = (V, E)$,with constant degree expander and $V = \{0, 1\}^K$.

We need initial K random bits, and do t steps random walk.

Then t random vertices(t strings of random k - bit).

Chapter 7 Applications of Concentration of measure

7.1 Concentration of measure

Let's see some basic inequalities:

- (Markov) Let X be a non-negative random variable. Then

$$\forall a > 0, \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}.$$

- (Chebyshev) Let X be a random variable with finite $\mathbb{E}X = \mu, \text{Var}X = \sigma^2$. Then

$$\forall k > 0, \mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Let's start with the simplest setup. Consider sum of Rademacher random variable X_i :

$$S_n = X_1 + \dots + X_n,$$

where X_i is independent of random variable .

Trivially: $-n \leq S_n \leq n$.

Typically: We shall see that S concentrates sharply within a window of width $O(S_n)$.

Intuition: It is very rare that all independent random variable. X_i team up to go in the same direction.

For instance: $\mathbb{P}(S_n = n) = \mathbb{P}(\text{all } X_i = 1) = 2^{-n}$ exponentially unlikely.

General phenomenon: Assuming boundedness and sufficient independence. Then we get concentration of measure. Usually of subgaussian nature

$$i.e. \mathbb{P}(\lambda\sigma \text{ away from mean } \mu) \leq c_1 \cdot e^{-c_2\lambda^2}$$

Simple scenario:

$$S_n = X_1 + \dots + X_n, X_i, i.i.d r.v.$$

$$-n \leq S_n \leq n, X_i = \begin{cases} 1, & \text{with probability } \frac{1}{2}; \\ -1, & \text{with probability } \frac{1}{2}. \end{cases}$$

But typically the value of S_n is sharply concentrated in a small window with width $O(S_n)$.

Theorem 7.1. [Chernoff]

Let X_1, \dots, X_n be i.i.d, Rademacher random variables and $S_n = X_1 + \dots + X_n$. Then for any $a > 0$,

$$\mathbb{P}(|S_n| > a) < 2 \exp\left(\frac{-a^2}{2n}\right).$$



The idea to prove it is applying Markov inequality to exponential moment.

Proof By symmetry, it suffices to show $\mathbb{P}(S_n > a) < e^{-\frac{a^2}{2n}}$. Consider the exponential moment of each X_i . For any $\lambda > 0$,

$$\mathbb{E}(e^{\lambda X_i}) = \frac{e^\lambda + e^{-\lambda}}{2} = \cosh(\lambda) \leq e^{\frac{\lambda^2}{2}},$$

where the least inequality is obtained by comparing Taylor series. Note that independence of X_i 's implies that

$$\mathbb{E}(e^{\lambda S_n}) = \mathbb{E}(e^{\lambda \sum_{i=1}^n X_i}) = \mathbb{E}\left(\prod_{i=1}^n e^{\lambda X_i}\right) \stackrel{\text{indep.}}{=} \prod_{i=1}^n \mathbb{E}e^{\lambda X_i} \leq e^{\frac{\lambda^2 n}{2}}.$$

Then

$$\mathbb{P}(S_n > a) = \mathbb{P}(e^{\lambda S_n} > e^{\lambda a}) < \frac{\mathbb{E}(e^{\lambda S_n})}{e^{\lambda a}} \leq e^{\frac{\lambda^2 n}{2} - \lambda a},$$

where the first inequality holds as Markov inequality. It is optimised when $\lambda = \frac{a}{n}$.

There are many extensions. For example, we can replace each X_i with some bounded random variable. Another direction is that we can replace “sum” with other functions $f(X_1, \dots, X_n)$ under some restrictions like Lipschitz condition.

Theorem 7.2. [Hoeffding’s inequality]

et X_1, \dots, X_n be independent random variables, where $X_i \in [a_i, b_i]$ and let $S_n = \sum_{i=1}^n X_i$ and $\sigma^2 = \sum_{i=1}^n |b_i - a_i|^2$. Then

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq \lambda\sigma) \leq Ce^{-c\lambda^2}.$$

Remark Note that that is equal to $\mathbb{P}(|S_n - \mathbb{E}S_n| \geq a) \leq C \exp\left(\frac{ca^2}{\sum_{i=1}^n |b_i - a_i|^2}\right)$ (let $a = \lambda\sigma$).

Definition 7.3. [Martingale]

A sequence of random variables X_1, \dots, X_n is a martingale if $\mathbb{E}|X_n| < \infty$, and $\mathbb{E}(X_{n+1}|X_n, \dots, X_1) = X_n$.

Example.

1. Random walk on \mathbb{Z} where X_n denotes the position at time n .
2. Gambler’s fortune where X_n denotes the total fortune at time n .

Theorem 7.4. [Azuma–Hoeffding’s inequality]

et X_0, X_1, \dots, X_n be martingales with $|X_i - X_{i-1}| \leq c_i$. Then for any $a > 0$,

$$\mathbb{P}(|X_n - X_0| \geq a) \leq 2 \exp\left(\frac{-2a^2}{\sum_{i=1}^n c_i^2}\right).$$

Using Azuma–Hoeffding’s inequality, we can get the large deviation for Lipschitz functions.

Definition 7.5. [bounded difference]

function $f : \Omega_1 \times \dots \times \Omega_n \rightarrow \mathbb{R}$ has bounded difference with parameter $c_1, \dots, c_n \in \mathbb{R}^n$ if for any $i \in [n]$, and any $x_i, x'_i \in \Omega_i$, we have $|f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(\dots, x'_i, \dots)| \leq c_i$.

Theorem 7.6. (McDiarmid’s inequality)

Let x_1, \dots, x_n be independent random variables with x_i taking values in ω_i and let $f : \omega_1 \times \dots \times \omega_n \rightarrow \mathbb{R}$ with bounded difference (c_1, \dots, c_n) , then we have, $\forall a > 0$, $P(|f(x_1, \dots, x_n) - \mathbb{E}f(x_1, \dots, x_n)| \geq a) < 2e^{-\frac{2a^2}{\sum c_i^2}}$.

We shall see some variation. A common product space is $\Omega_i = \{0, 1\} \Rightarrow \{0, 1\}^n$. The following is a large deviation inequality for Lipschitz functions on a slice of Boolean Cube.

Lemma 7.7. [Kwan, Sudakov, Tran]

Suppose $g : \{0, 1\}^n \rightarrow \mathbb{R}$ satisfies the bounded difference condition with parameter (c_1, \dots, c_n) and $\xi \in \{0, 1\}^n$ is a random variable uniformly distributed in $\binom{[n]}{k}$. Then $\forall t > 0$, $P(|g(\xi) - \mathbb{E}g(\xi)| \geq t) \leq 2e^{-\frac{t^2}{8\sum c_i^2}}$.



Proof We may assume without loss of generality that $c_1 \geq c_2 \geq \dots \geq c_n$. Consider the Doob martingale $Z_i = \mathbb{E}[g(\xi) | \xi_1, \dots, \xi_i]$, so $Z_0 = \mathbb{E}g(\xi)$ and $Z_n = Z_{n-1} = g(\xi)$. Let $L(x_1, \dots, x_i)$ be the conditional distribution of ξ given $\xi_1 = x_1, \dots, \xi_i = x_i$.

We claim that

$$|\mathbb{E}(g(L(x_1, \dots, x_{i-1}, 0))) - \mathbb{E}(g(L(x_1, \dots, x_{i-1}, 1)))| \leq 2c_i,$$

for any feasible choice of $x_1, \dots, x_{i-1} \in \{0, 1\}$. The claim implies that $|Z_i - Z_{i-1}| \leq 2c_i$ and the conclusion follows from Azuma–Hoeffding bound.

If $\xi \sim L(x_1, \dots, x_{i-1}, 0)$ changes ξ_i to 1 and then randomly choose one of the ones among ξ_{i+1}, \dots, ξ_n and change it to 0. We thereby obtain the distribution $L(x_1, \dots, x_{i-1}, 1)$. This provides a coupling between $L(x_1, \dots, x_{i-1}, 0)$ and $L(x_1, \dots, x_{i-1}, 1)$ that differs in only two coordinates i and $j \geq i$, as $c_j \leq c_i$ this implies the desired bound.

Lemma 7.8. [Kim, Liu, Tran]

Suppose $f : \{0, 1, \dots, q-1\}^n \rightarrow \mathbb{R}$ satisfies the bounded difference condition with parameter (c_1, \dots, c_n) and η is drawn uniformly at random from $\{0, 1, \dots, q-1\}^n$ subject to $wt(\eta) = k$ ($wt(\eta)$ = the number of non-zero coordinates). Then

$$\mathbb{P}(|f(\eta) - \mathbb{E}f(\eta)| \geq t) \leq 2 \exp\left(-\frac{t^2}{68 \sum c_i^2}\right) \text{ for all } t \geq 0.$$



Next, we shall use Lemma 7.7 and the following standard fact about subgaussian random variables to prove Lemma 7.8.

Lemma 7.9. Subgaussian properties

Let X be a random variable with mean 0. Then the following properties are equivalent:

(i) There exists $K_1 > 0$ such that the tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/K_1}, \forall t \geq 0.$$

(ii) There exists $K_2 > 0$ such that the moment generating function of X satisfies

$$\mathbb{E}e^{\lambda X} \leq e^{K_2 \lambda^2}, \forall \lambda \geq 0.$$

In particular, for (i) \Rightarrow (ii), we can take $K_2 = 2K_1$ and for (ii) \Rightarrow (i), we can take $K_1 = 4K_2$.



Proof of Lemma 7.8 Let $\xi \in \{0, 1\}^n$ be a random vector uniformly distributed in $\binom{[n]}{k}$ and u be drawn uniformly from $[q-1]^n = \{1, \dots, q-1\}^n$, independent from ξ . Then the conclusion is that the distribution of η coincides with the distribution of $u * \xi = (u_1 \xi_1, \dots, u_n \xi_n)$.

By Lemma 7.9, it suffices to show that

$$\mathbb{E}_u \mathbb{E}_\xi e^{\lambda(f(u*\xi) - \mathbb{E}_{u\xi} f(u*\xi))} \leq e^{17\|c\|^2 \lambda^2}.$$

Fix an instance of u . Note that the function $f(u * \cdot)$ has the bounded different condition with parameter

$c = (c_1, \dots, c_n)$. Then by Lemma 7.7 with $f(u * \cdot)$ playing the role of $g(\cdot)$ and Lemma 7.9, we get

$$\mathbb{E}_\xi e^{\lambda(f(u*\xi) - \mathbb{E}f(u*\xi))} \leq e^{16\|c\|^2\lambda^2}.$$

Thus,

$$\begin{aligned} \mathbb{E}_u \mathbb{E}_\xi e^{\lambda(f(u*\xi) - \mathbb{E}f(u*\xi))} &= e^{-\lambda \mathbb{E}_{u,\xi} f(u*\xi)} \mathbb{E}_u e^{\lambda \mathbb{E}_\xi (f(u*\xi) - \mathbb{E}_\xi f(u*\xi))} \\ &= e^{16\|c\|^2\lambda^2} \mathbb{E}_u e^{\lambda(\mathbb{E}_\xi f(u*\xi) - \mathbb{E}_{u,\xi} f(u*\xi))}. \end{aligned}$$

As $g(\cdot) := \mathbb{E}_\xi f(\cdot * \xi)$ has the bounded different condition with parameter c , by McDiarmid's inequality,

$$\mathbb{P}(|g(u) - \mathbb{E}_u g(u)| \geq t) \leq 2e^{-2t^2/\|c\|^2}.$$

By Lemma 7.9, we have $\mathbb{E}_u e^{\lambda(g(u) - \mathbb{E}g(u))} \leq e^{\|c\|^2\lambda^2}$.

7.2 Applications.

We shall see a geometric application of the large deviation inequality over a slice. The question to consider is that given a metric space (X, d) , how can we bound the volume of intersection of two balls. We will prove a result providing natural sufficient condition on the metric space (X, d) guaranteeing exponential decay on intersection volume.

Definition 7.10


A metric space (X, d) has exponential growth at radius r with rate c . If $\forall a \in X$ and $\forall t < r$,

$$\frac{\text{vol}(B(a, r-t))}{\text{vol}(B(a, r))} \leq 2e^{-ct}$$

which $B(a, r-t)$ is a ball centered at a with radius $r-t$. 

Definition 7.11

For $a, b \in X$, let $\ell_{a,b} : X \rightarrow \mathbb{R}$ be

$$\ell_{a,b}(x) = d(x, b) - d(x, a). $$

Given $r, k \in \mathbb{N}$ and $\alpha > 0$, we say that the metric space (X, d) is (r, k) -dispersed with constant α if $\forall a, b \in X$ with $d(a, b) = k$ and any $0 \leq i \leq \alpha k$,

$$\mathbb{E}_{x \sim S(a, r-i)}[\ell_{a,b}(x)] \geq 2\alpha k$$

which $S(a, r-i)$ are all points of distance $r-i$ from a .

Recall a real-valued random variable X is K -subgaussian if

$$\forall t \geq 0, \mathbb{P}(|x| \geq t) \leq 2e^{-t^2/K}.$$


Theorem 7.12. (Kim, Liu, Tran)

Let (X, d) be a finite metric space with d taking values in $\mathbb{N} \cup \{0\}$ and let $k, r \in \mathbb{N}$. Suppose

(A1) (X, d) has exponential growth at radius r with rate $c > 0$.

(A2) (X, d) is (r, k) -dispersed with a constant $\alpha > 0$

(A3) $\forall a, b \in X$ with $d(a, b) = k$ and $\forall 0 \leq i \leq \alpha k$, $\ell_{a,b}(x) - \mathbb{E}\ell_{a,b}(x)$ is K -subgaussian, where x is drawn uniformly from $S(a, r-i)$. Then $\forall a, b \in X$ with $d(a, b) = k$,

$$\frac{\text{vol}(B(a, r) \cap B(b, r))}{\text{vol}(B(a, r))} \leq 2e^{-\Omega_{c,\alpha}(k+k^2/K)}. $$

Proof Let $T = B(a, r) \cap B(b, r)$ and $\eta \sim B(a, r)$. Then

$$\frac{\text{vol}(B(a, r) \cap B(b, r))}{\text{vol}(B(a, r))} = \mathbb{P}(\eta \in T) = \mathbb{P}(d(\eta, b) \leq r).$$

By (A1), we have $\mathbb{P}(d(\eta, a) \leq r - \alpha k) \leq 2e^{-\Omega(k)}$. Thus,

$$\begin{aligned} \mathbb{P}(\eta \in T) &\leq \mathbb{P}(\eta \in T | d(\eta, a) > r - \alpha k) \cdot \mathbb{P}(d(\eta, a) > r - \alpha k) + \mathbb{P}(d(\eta, a) \leq r - \alpha k) \\ &\leq \sum_{i=0}^{\alpha k} \mathbb{P}(d(\eta, b) \leq r | d(\eta, a) = r - i) \mathbb{P}(d(\eta, a) = r - i) + 2e^{-\Omega(k)} \\ &\leq \max_{0 \leq i \leq \alpha k} \mathbb{P}(d(\eta, b) \leq r | d(\eta, a) = r - i) + 2e^{-\Omega(k)}. \end{aligned}$$

Fix an $i \in [0, \alpha k]$ and let $x \sim S(a, r - i)$. Note that, conditioning on $d(\eta, a) = r - i$, η and x are identically distributed, we can get that

$$\begin{aligned} \mathbb{P}(d(\eta, b) \leq r | d(\eta, a) = r - i) &= \mathbb{P}(d(\eta, b) - d(\eta, a) \leq i | d(\eta, a) = r - i) \\ &= \mathbb{P}(d(x, b) - d(x, a) \leq i) \\ &= \mathbb{P}(\ell_{a,b}(x) \leq i). \end{aligned}$$

By (A2), we have $\mathbb{E}\ell_{a,b}(x) \geq 2\alpha k$. Consequently, we have $i - \mathbb{E}\ell_{a,b}(x) \leq i - 2\alpha k \leq -\alpha k$.

Since by (A3) which show that $\ell_{a,b}(x) - \mathbb{E}\ell_{a,b}(x)$ is K -subgaussian, then we have

$$\begin{aligned} \mathbb{P}(\ell_{a,b}(x) \leq i) &= \mathbb{P}(\ell_{a,b}(x) - \mathbb{E}\ell_{a,b}(x) \leq i - \mathbb{E}\ell_{a,b}(x)) \\ &\leq \mathbb{P}(\ell_{a,b}(x) - \mathbb{E}\ell_{a,b}(x) \leq -\alpha k) \\ &\leq 2e^{-\Omega(k^2/K)}. \end{aligned}$$

7.3 Intersection volume

We first consider intersection of two balls in \mathbb{R}^3 . Take $\varepsilon > 0$ sufficiently small and consider two unit balls A and B whose centers are of distance ε apart, then $\text{vol}(A \cap B) \geq 99\% \text{vol}(A)$. But this is not longer true in high dimension.

Proposition 7.13

Let $\varepsilon > 0$. Then there exists $n_0 = n_0(\varepsilon)$ such that the following holds for all $n \geq n_0$.

Let A and B be two unit balls whose centers are of distance ε apart, then

$$\frac{\text{vol}(A \cap B)}{\text{vol}(A)} < 1\%.$$

Before the proof of Proposition 7.13, we need the following basic fact.

Fact 7.14

The volume of radius- r ball in \mathbb{R}^n is

$$\text{vol}_n(r) \sim \frac{1}{\sqrt{n\pi}} \left(\frac{2\pi e}{n} \right)^{n/2} r^n.$$

Proof of Proposition 7.13

It is easy to see that $r' < r - \varepsilon/10$. Then $\text{vol}(A \cap B) < \text{vol}_n(r')$, and then we are done because

$$\frac{\text{vol}(A \cap B)}{\text{vol}(A)} < \frac{\text{vol}_n(r')}{\text{vol}_n(r)} = \left(\frac{r'}{r} \right)^n < \left(1 - \frac{\varepsilon}{10r} \right)^n \rightarrow 0$$

as $n \rightarrow \infty$.

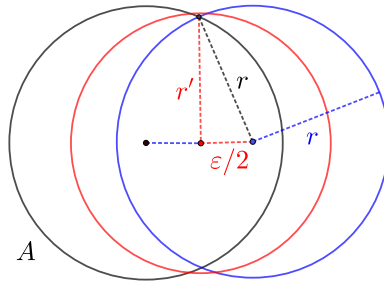


Figure 7.1

- This idea of using a ball of smaller radius to bound the intersection in the continuous \mathbb{R}^n setting fails in many discrete settings.

Example. Take the discrete cube $\{0, 1\}^n$ endowed with the Hamming metric and let $k, r \in \mathbb{N}$ with $2k \leq r$. Consider the two radius- r Hamming balls A and B centered at $a = 0^n$ and $b = 1^{2k}0^{n-2k}$ respectively. Take a mid-point c of a and b , say by symmetry $c = 1^k0^{n-k}$. Then the point $x = 0^k1^r0^{n-r-k}$ lies in the intersection $A \cap B$, but it is of Hamming distance $r + k$ from the chosen mid-point c .

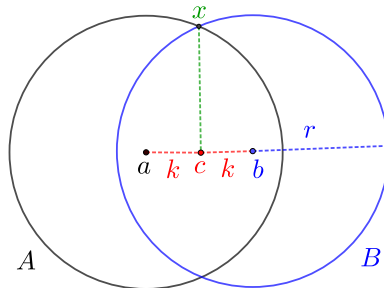


Figure 7.2

Recall the result Theorem 7.12 from last time. It gives natural sufficient condition for (X, d) to guarantee intersection volume is small.

The Hamming space satisfies the conditions of Theorem 7.12 as follows.

Theorem 7.15

Let $0 < p < \frac{q-1}{q}$ and let $k \in \mathbb{N}$. Consider $X = \{0, 1, \dots, q-1\}^n$ endowed with the Hamming metric Δ . Then (X, Δ) satisfies the conditions (A1) – (A3) of Theorem 7.12 as follows.

- (A1) (X, Δ) has exponential growth at radius pn with rate $c = \Omega_{p,q}(1)$.
- (A2) (X, Δ) is (pn, k) -dispersed with constant $\alpha = \frac{1}{2} \left(1 - \frac{pq}{q-1}\right) > 0$.
- (A3) For any $a, b \in X$ with $\Delta(a, b) = k$ and any $0 \leq i \leq \alpha k$, $\ell_{a,b}(\mathbf{x}) - \mathbb{E}\ell_{a,b}(\mathbf{x})$ is $400k$ -subgaussian, where $\ell_{a,b}(\mathbf{x}) = \Delta(\mathbf{x}, a) - \Delta(\mathbf{x}, b)$ and \mathbf{x} is drawn uniformly from $S(a, pn - i)$.

Consequently, for every $a, b \in X$,

$$\frac{\text{vol}(B(a, r) \cap B(b, r))}{\text{vol}(B(a, r))} = 2e^{-\Omega_{p,q}(\Delta(a,b))}.$$



- The conditions (A1) – (A3) can be easily verified. In particular, (A3) follows from the concentration of

measure on a slice.

This exponential decay on intersection volume for balls in Hamming space has applications on, for example, list-decodability of random codes, and improvements on Gilbert-Varshamov type bounds.

Next we give a unified proof of improvements on Gilbert-Varshamov type bounds on various models of error correction codes, which we now discuss in details.

An *error correcting code* (ECC) is an encoding scheme that transmits messages as binary numbers, in such a way that the message can be recovered even if some bits are erroneously flipped. They are used in practically all cases of message transmission, especially in data storage where ECCs defend against data corruption.

By encoding messages with codewords that are pairwise far apart, we can recover the message even if some bits are corrupted by noise. For example, suppose there are two persons, Alice and Bob. Bob is asking Alice a question, and is waiting Alice's answer "YES" or "NO". Suppose there is some noise that could corrupt up to 2 bits of the binary strings. We denote the encode "YES" by, say "000100", and "No" by "111111". Suppose the code that Bob receive is $x := 011100$. So how do we know what Alice says, YES or NO? It is very simple, we can look at their Hamming distances. It is easy to see that $\Delta(x, YES) = 2$ and $\Delta(x, NO) = 3$. So we can know that Alice said "YES". In general, if the noise could corrupt up to t bits, then as long as all messages are encoded by strings with distance at least $2t + 1$.

Definition 7.16

Given positive integers n and d , we denote by $A(n, d)$ the maximum number of messages (or codewords) in $\{0, 1\}^n$ with minimum distance d .



Theorem 7.17. (Gilbert-Varshamov bound)

$$A(n, d + 1) \geq \frac{2^n}{\text{vol}(n, d)},$$

where $\text{vol}(n, d) = \sum_{i=0}^d \binom{n}{i}$ is the volume of radius- d ball.



We can prove Gilbert-Varshamov bound by the following Turán's theorem.

Theorem 7.18. (Turán)

Let G be an N -vertex D -regular graph. Then $\alpha(G) \geq \frac{N}{D+1}$.



The bound is tight. (consider K_{D+1})

Exercise 7.1 Prove Gilbert-Varshamov bound using Turán's theorem.

We can improve Gilbert-Varshamov bound, if we get a better bound on independence number. Ajtai-Komlós-Szemerédi [1] proved that if G is K_3 -free or locally sparse, then $\alpha(G) \geq c \frac{N}{D} \log D$.

Theorem 7.19

Let G be an N -vertex graph with maximum degree D and minimum degree at least $D/2$. Let $K \in [D]$ and let $\Gamma \subseteq G$ be a subgraph induced by the neighborhood of an arbitrary vertex. Suppose there is a partition $V(\Gamma) = B \cup I$ such that

(1) For any $u \in B$, $\deg_{\Gamma}(u) \leq D/k$; and

(2) $|I| \leq D/k$.

Then $\alpha(G) \geq (1 - o_k(1)) \frac{N}{D} \log k$, and the number of the independent sets in G is at least $e^{(\frac{1}{8} + o_k(1)) \frac{N}{D} \log^2 k}$.



To apply the above theorem to improve Gilbert-Varshamov bound, we use Theorem 7.15 to check that certain auxiliary graph is locally sparse.

7.4 Johnson–Lindenstrauss Lemma

Motivation. In this digital era, lots of data are transmitted as we speak. Many data (such as images, videos) can be represented by high dimensional vectors. To speed up computation, it is of great practical importance to try to reduce the dimension.

Applications.

- Clustering.
- Regression analysis.
- ...

The basic task we want to do is to tell distinct vectors apart. Using points in \mathbb{R}^d to represent data and the Euclidean distance between points measures their “similarity”.

Given $x \in \mathbb{R}^d$, we write $\|x\| := \|x\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$. (ℓ_2 -norm)

Euclidean distance between x and y is $\|x - y\|$.

The Johnson–Lindenstrauss Lemma was first introduced in the paper “Extensions of Lipschitz mappings into a Hilbert Space” by William B. Johnson and Joram Lindenstrauss published 1984 in Contemporary Mathematics. The Lemma is as follows.

Lemma 7.20. [Johnson, Lindenstrauss]

Given $0 < \varepsilon < 1$ and a set X of n points in \mathbb{R}^d . Then there exists a linear map $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$, where $m = O(\frac{\log n}{\varepsilon^2})$ such that for every $u, v \in X$,

$$\|f(u) - f(v)\| = (1 \pm \varepsilon)\|u - v\|.$$



The Lemma states that after fixing an error level, one can map a collection of points from one Euclidean space (no matter how high its dimension m is) to a smaller Euclidean space while only changing the distance between any two points by a factor of $1 \pm \varepsilon$. The dimension of the image space is only dependent on the error and the number of points. Given that the dimension is very large, one can achieve significant dimension reduction, which has applications in data analysis and computer science.

Idea. Project points randomly to a low dimensional subspace. (Random projection trick is a powerful technique behind compressive sensing and matrix completion.)

Note that there is a naive way to choose m coordinates out of d uniformly at random. We can easily see this way failing by a simple example. Just take $u = (1, 0, \dots, 0) \in \mathbb{R}^d$ and $v = (0, 1, 0, \dots, 0) \in \mathbb{R}^d$. To preserve distance between u and v , there m random coordinates need to include the first one or the second coordinate, which is quite unlikely if $d \gg n$.

Lemma 7.21. [Distribution Johnson–Lindenstrauss Lemma]

Given $0 < \varepsilon, \delta < 1$, there exists a constant C such that the following holds. Let A be an $m \times d$ random matrix, in which each entry is a normal random variable $\sim \frac{1}{\sqrt{m}}N(0, 1)$ independent of others, where $m \leq C \cdot \varepsilon^{-2} \log \frac{1}{\delta}$. Then for every $x \in \mathbb{R}^d$,

$$\mathbb{P}(\|Ax\| = (1 \pm \varepsilon)\|x\|) \geq 1 - \delta.$$



To get the original Johnson–Lindenstrauss Lemma, we use Lemma 7.21 to all $\binom{n}{2}$ pairwise distances in X : let $A = f, f(u) - f(v) = Au - Av = A(u - v)$. Choose $\delta = \frac{\delta'}{\binom{n}{2}}$, where $m = O(\varepsilon^{-2} \log \frac{1}{\delta}) = O(\varepsilon^{-2} \log n)$.

Proof of Lemma 7.21 Fix $x \in \mathbb{R}^d$, we want to show that with high probability, $\|Ax\| \approx \|x\|$.

First we show that

$$\mathbb{E}\|Ax\|^2 = \|x\|^2.$$

Indeed, let $g = (g_1, \dots, g_d) \in \mathbb{R}^d$, where $g_i \sim \frac{1}{\sqrt{m}}N(0, 1) \sim N(0, \frac{1}{m})$, and let

$$A_{m \times d} = \begin{pmatrix} \dots & g^{(1)} & \dots \\ \vdots & \vdots & \vdots \\ \dots & g^{(m)} & \dots \end{pmatrix},$$

where $g^{(i)} \in \mathbb{R}^d$. Note that $\|Ax\|^2 = \sum_{i=1}^m \langle g^{(i)}, x \rangle^2 = m \langle g, x \rangle^2$, then

$$\mathbb{E} \langle g, x \rangle = \mathbb{E} \sum_{i=1}^d g_i x_i = \sum_{i=1}^d x_i \mathbb{E} g_i = 0.$$

Thus

$$\mathbb{E} \langle g, x \rangle^2 = \mathbf{Var} \langle g, x \rangle = \sum_{i=1}^d x_i^2 \mathbf{Var}[g_i] = \frac{1}{m} \|x\|^2.$$

We have done as $\mathbb{E}\|Ax\|^2 = \mathbb{E}(m \langle g, x \rangle^2) = \|x\|^2$.

In particular, $\|Ax\|^2 = \sum_{i=1}^m \langle g^{(i)}, x \rangle^2$, where $\langle g^{(i)}, x \rangle \sim \frac{1}{\sqrt{m}}\|x\|\dot{N}(0, 1)$. So $\|Ax\|^2$ is a χ^2 -random variable with m degree of freedom. By concentration of χ -squared distribution (better than Chernoff bound), we have

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon \mathbb{E}X) \leq 2e^{-\varepsilon^2 m/8} \leq \delta,$$

where the last inequality holds by taking $m = O(\log \frac{1}{\delta} / \varepsilon^2)$.

7.4.1 Applications to regression analysis

Setup. Given many data points $a_1, a_2, \dots, a_n \in \mathbb{R}^d$.

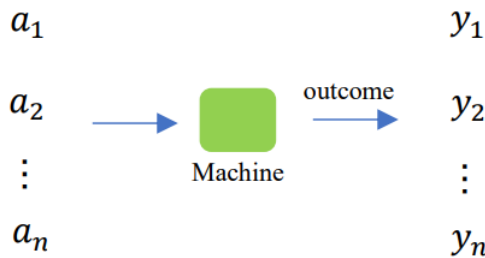


Figure 7.3

We want to figure out the relation between the data and the outcome $y_1, y_2, \dots, y_n \in \mathbb{R}^n$. We write

$$A_{n \times d} = \begin{pmatrix} \cdots & a_1 & \cdots \\ \cdots & a_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & a_n & \cdots \end{pmatrix}, \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n.$$

Whether $\exists x \in \mathbb{R}^d$ s.t. $Ax \approx y$?

Goal. (Least square regression)

$$\min_{x \in \mathbb{R}^d} \|Ax - y\|^2.$$

Denote an optimal solution of the above original problem by x^* . Instead, we only need to solve the following sketched problem.

$$\min_{x \in \mathbb{R}^d} \|\Pi Ax - \Pi y\|^2.$$

Let \tilde{x}^* be an optimal solution of the above sketched problem.

Claim If for any vector of the form $Ax - y$, we have $\|\Pi Ax - \Pi y\|^2 = (1 \pm \varepsilon)\|Ax - y\|^2$, then \tilde{x}^* gives a good approximate for x^* .

Proof Our goal is to obtain $\|A\tilde{x}^* - y\|^2 \leq (1 + \varepsilon)\|Ax^* - y\|^2$.

Note that for any $x \in \mathbb{R}^d$, we have $\|\Pi A\tilde{x}^* - \Pi y\|^2 \leq \|\Pi Ax - \Pi y\|^2$. It is easy to see that

$$\|A\tilde{x}^* - y\|^2 \leq (1 + \varepsilon)\|\Pi A\tilde{x}^* - \Pi y\|^2 \leq (1 + \varepsilon)\|\Pi Ax^* - \Pi y\|^2 \leq (1 + \varepsilon)^2\|Ax^* - y\|^2.$$

The first and last inequalities hold since the property of Π , while the second holds since the optimality of \tilde{x}^* .

Next, we want to obtain the condition of the above claim. Lemma 7.21 can preserve a single vector's length, but there are infinitely many vectors of the form $Ax - y$, the union bound cannot work. Thus, we use another tool.

Idea: We construct an ε -net for the subspace spanned by col^n of A and y , apply union bound over the ε -net.

Theorem 7.22

Let $U \subseteq \mathbb{R}^n$ be a d -dimension linear subspace of \mathbb{R}^n and let $\Pi \in \mathbb{R}^{m \times n}$ be the matrix from Lemma 7.21, where $m = O(\frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2})$. Then with probability at least $1 - \delta$, for any $v \in U$, we have

$$\|\Pi v\|^2 = (1 \pm \varepsilon)\|v\|^2.$$



We want to show how Theorem 7.22 implies Least square regression. Theorem 7.22 implies that we can find a projection $\Pi \in \mathbb{R}^{m \times n}$ with $m = O(\frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2})$ such that for any vector of the form $Ax - y$, $\|\Pi Ax - \Pi y\|^2 = (1 \pm \varepsilon)\|Ax - y\|^2$. Take U to be a space spanned by col^n of A and y . Note that the dimension of U is $d + 1$, we can use Theorem 7.22, as desired.

To prove Theorem 7.22, first observe that as Π is linear, it suffices to consider unit vectors $v \in U$. Write S_U for the unit sphere $S_U = \{v \in U : \|v\|^2 = 1\}$, we shall find an ε -net N_ε for S_U , that is, for any $v \in S_U$, there exists $x \in N_\varepsilon$ such that $\|v - x\|^2 \leq \varepsilon$.

Lemma 7.23

For $0 < \varepsilon < 1$, there exists an ε -net $N_\varepsilon \subseteq S_U$ with $|N_\varepsilon| \leq (\frac{4}{\varepsilon})^d$ such that for any $v \in S_U$, we have

$$\min_{x \in N_\varepsilon} \|x - v\|^2 \leq \varepsilon.$$



Proof Sketch. Iteratively pick x_1, x_2, \dots such that x_i 's pairwise distance is at least ε . Let N_ε be the

maximal set of such x_i . Note that for distinct x_i and x_j , we have $B(x_i, \varepsilon/2)$ is disjoint with $B(x_j, \varepsilon/2)$ and $B(x_i, \varepsilon/2) \subseteq B(0, 1 + \varepsilon/2)$. Thus, $|N_\varepsilon| \cdot \text{vol}_d(\varepsilon/2) \leq \text{vol}_d(1 + \varepsilon/2)$. Recall that $\text{vol}_d(r) = c \cdot r^d$, we have $|N_\varepsilon| \leq (\frac{1+\varepsilon/2}{\varepsilon/2})^d \leq (\frac{4}{\varepsilon})^d$.

Proof of Theorem 7.22 Let N_ε be an ε -net for S_U . By Lemma 7.21 and union bound for $m = O(\frac{\log |N_\varepsilon|/\delta}{\varepsilon^2}) = O(\frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2})$, there is a linear projection $\Pi \in \mathbb{R}^{m \times n}$ such that with probability at least $1 - \delta$, for any $x \in N_\varepsilon$, $\|\Pi x\|^2 = (1 \pm \varepsilon)\|x\|^2$. We need to show for any $v \in S_U$, we have $\|\Pi v\|^2 = (1 \pm \varepsilon)\|v\|^2$.

Claim For any $v \in S_U$, there exists a sequence of points $x_0, x_1, x_2, \dots \in N_\varepsilon$ such that $v = x_0 + c_1 x_1 + c_2 x_2 + \dots$ for $|c_i| \leq \varepsilon^i$.

Thus, we have

$$\begin{aligned} \|\Pi v\| &= \|\Pi x_0 + c_1 \Pi x_1 + c_2 \Pi x_2 + \dots\| \\ &\leq \|\Pi x_0\| + c_1 \|\Pi x_1\| + c_2 \|\Pi x_2\| + \dots \\ &\leq (1 + \varepsilon) + \varepsilon(1 + \varepsilon) + \varepsilon^2(1 + \varepsilon) + \dots \\ &\leq 1 + O(\varepsilon). \end{aligned}$$

The first equality holds by the triangle inequality and the second equality holds since the choice of Π and $\|x_i\|^2 = 1$ for any i . Similarly, $\|\Pi v\|^2 \geq 1 - O(\varepsilon)$. We obtain that $\|\Pi v\|^2 = (1 \pm \varepsilon)\|v\|^2$ since $\|v\|^2 = 1$.

Bibliography

- [1] M. Ajtai, J. Komlós, and E. Szemerédi. “A dense infinite Sidon sequence”. In: *European J. Combin.* 2 (1981), pp. 1–11. ISSN: 0195-6698. DOI: [10.1016/S0195-6698\(81\)80014-5](https://doi.org/10.1016/S0195-6698(81)80014-5). URL: [https://doi.org/10.1016/S0195-6698\(81\)80014-5](https://doi.org/10.1016/S0195-6698(81)80014-5).
- [2] M. Ajtai et al. “On Turán’s theorem for sparse graphs”. In: *Combinatorica* 1.4 (1981), pp. 313–317. ISSN: 0209-9683. DOI: [10.1007/BF02579451](https://doi.org/10.1007/BF02579451). URL: <https://doi.org/10.1007/BF02579451>.
- [3] Miklós Ajtai, János Komlós, and Endre Szemerédi. “A note on Ramsey numbers”. In: *J. Combin. Theory Ser. A* 29.3 (1980), pp. 354–360. ISSN: 0097-3165. DOI: [10.1016/0097-3165\(80\)90030-8](https://doi.org/10.1016/0097-3165(80)90030-8). URL: [https://doi.org/10.1016/0097-3165\(80\)90030-8](https://doi.org/10.1016/0097-3165(80)90030-8).
- [4] Alon and N. “Combinatorial Nullstellensatz”. In: *Combinatorics Probability Computing* (1999).
- [5] Boris Bukh and Ting-Wei Chao. *Sharp density bounds on the finite field Kakeya*. 2021. arXiv: [2108.00074](https://arxiv.org/abs/2108.00074) [math.CO].
- [6] F. R. K. Chung and R. L. Graham. “Quasi-random subsets of Z_n ”. In: *J. Combin. Theory Ser. A* 61.1 (1992), pp. 64–86. ISSN: 0097-3165. DOI: [10.1016/0097-3165\(92\)90053-W](https://doi.org/10.1016/0097-3165(92)90053-W). URL: [https://doi.org/10.1016/0097-3165\(92\)90053-W](https://doi.org/10.1016/0097-3165(92)90053-W).
- [7] Henry Cohn and Yufei Zhao. “Sphere packing bounds via spherical codes”. In: *Duke Math. J.* 163.10 (2014), pp. 1965–2002. ISSN: 0012-7094. DOI: [10.1215/00127094-2738857](https://doi.org/10.1215/00127094-2738857). URL: <https://doi.org/10.1215/00127094-2738857>.
- [8] Henry Cohn et al. “The sphere packing problem in dimension 24”. In: *Ann. of Math. (2)* 185.3 (2017), pp. 1017–1033. ISSN: 0003-486X. DOI: [10.4007/annals.2017.185.3.8](https://doi.org/10.4007/annals.2017.185.3.8). URL: <https://doi.org/10.4007/annals.2017.185.3.8>.
- [9] Ewan Davies et al. “On the average size of independent sets in triangle-free graphs”. In: *Proceedings of the American Mathematical Society* 146.1 (July 2017), pp. 111–124. ISSN: 1088-6826. DOI: [10.1090/proc/13728](https://doi.org/10.1090/proc/13728). URL: <http://dx.doi.org/10.1090/proc/13728>.
- [10] Zeev Dvir. “On the size of Kakeya sets in finite fields”. In: *J. Amer. Math. Soc.* 22.4 (2009), pp. 1093–1097. ISSN: 0894-0347. DOI: [10.1090/S0894-0347-08-00607-3](https://doi.org/10.1090/S0894-0347-08-00607-3). URL: <https://doi.org/10.1090/S0894-0347-08-00607-3>.
- [11] Zeev Dvir et al. “Extensions to the method of multiplicities, with applications to Kakeya sets and mergers”. In: *SIAM J. Comput.* 42.6 (2013), pp. 2305–2328. ISSN: 0097-5397. DOI: [10.1137/100783704](https://doi.org/10.1137/100783704). URL: <https://doi.org/10.1137/100783704>.
- [12] J. S. Ellenberg and D. Gijswijt. “On large subsets of F_q^n with no three-term arithmetic progression”. In: *Annals of Mathematics* 185.1 (2016).
- [13] Thomas C. Hales. “A proof of the Kepler conjecture”. In: *Ann. of Math. (2)* 162.3 (2005), pp. 1065–1185. ISSN: 0003-486X. DOI: [10.4007/annals.2005.162.1065](https://doi.org/10.4007/annals.2005.162.1065). URL: <https://doi.org/10.4007/annals.2005.162.1065>.
- [14] Matthew Jenssen, Felix Joos, and Will Perkins. “On the hard sphere model and sphere packings in high dimensions”. In: *Forum Math. Sigma* 7 (2019), Paper No. e1, 19. DOI: [10.1017/fms.2018.25](https://doi.org/10.1017/fms.2018.25). URL: <https://doi.org/10.1017/fms.2018.25>.

- [15] G. A. Kabatiansky and V. I. Levenshtein. “On Bounds for Packings on a Sphere and in Space”. In: *Problems of Information Transmission* 14.1 (1978), pp. 1–17.
- [16] Jeff Kahn. “An entropy approach to the hard-core model on bipartite graphs”. In: *Combin. Probab. Comput.* 10.3 (2001), pp. 219–237. ISSN: 0963-5483. DOI: [10.1017/S0963548301004631](https://doi.org/10.1017/S0963548301004631). URL: <https://doi.org/10.1017/S0963548301004631>.
- [17] Jeong Han Kim. “The Ramsey number $R(3, t)$ has order of magnitude $t^2 / \log t$ ”. In: *Random Structures Algorithms* 7.3 (1995), pp. 173–207. ISSN: 1042-9832. DOI: [10.1002/rsa.3240070302](https://doi.org/10.1002/rsa.3240070302). URL: <https://doi.org/10.1002/rsa.3240070302>.
- [18] Eric Naslund and Will Sawin. *Upper bounds for sunflower-free sets*. 2017. DOI: [10.1017/fms.2017.12](https://doi.org/10.1017/fms.2017.12).
- [19] Gonzalo Fiz Pontiveros, Simon Griffiths, and Robert Morris. *The triangle-free process and the Ramsey number $R(3, k)$* . 2018. arXiv: [1302.6279](https://arxiv.org/abs/1302.6279) [math.CO].
- [20] Shubhangi Saraf and Madhu Sudan. “An improved lower bound on the size of Kakeya sets over finite fields”. In: *Anal. PDE* 1.3 (2008), pp. 375–379. ISSN: 2157-5045. DOI: [10.2140/apde.2008.1.375](https://doi.org/10.2140/apde.2008.1.375). URL: <https://doi.org/10.2140/apde.2008.1.375>.
- [21] James B. Shearer. “A note on the independence number of triangle-free graphs”. In: *Discrete Math.* 46.1 (1983), pp. 83–87. ISSN: 0012-365X. DOI: [10.1016/0012-365X\(83\)90273-X](https://doi.org/10.1016/0012-365X(83)90273-X). URL: [https://doi.org/10.1016/0012-365X\(83\)90273-X](https://doi.org/10.1016/0012-365X(83)90273-X).
- [22] James B. Shearer. “A note on the independence number of triangle-free graphs. II”. In: *J. Combin. Theory Ser. B* 53.2 (1991), pp. 300–307. ISSN: 0095-8956. DOI: [10.1016/0095-8956\(91\)90080-4](https://doi.org/10.1016/0095-8956(91)90080-4). URL: [https://doi.org/10.1016/0095-8956\(91\)90080-4](https://doi.org/10.1016/0095-8956(91)90080-4).
- [23] James B. Shearer. “On the independence number of sparse graphs”. In: *Random Structures Algorithms* 7.3 (1995), pp. 269–271. ISSN: 1042-9832. DOI: [10.1002/rsa.3240070305](https://doi.org/10.1002/rsa.3240070305). URL: <https://doi.org/10.1002/rsa.3240070305>.
- [24] A. Thue. “Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene”. In: *Norske Videnskabs-Selskabets Skrifter* 1 (Jan. 1910).
- [25] Akshay Venkatesh. “A note on sphere packings in high dimension”. In: *Int. Math. Res. Not. IMRN* 7 (2013), pp. 1628–1642. ISSN: 1073-7928. DOI: [10.1093/imrn/rns096](https://doi.org/10.1093/imrn/rns096). URL: <https://doi.org/10.1093/imrn/rns096>.
- [26] Maryna S. Viazovska. “The sphere packing problem in dimension 8”. In: *Ann. of Math. (2)* 185.3 (2017), pp. 991–1015. ISSN: 0003-486X. DOI: [10.4007/annals.2017.185.3.7](https://doi.org/10.4007/annals.2017.185.3.7). URL: <https://doi.org/10.4007/annals.2017.185.3.7>.
- [27] Yufei Zhao. “The number of independent sets in a regular graph”. In: *Combin. Probab. Comput.* 19.2 (2010), pp. 315–320. ISSN: 0963-5483. DOI: [10.1017/S0963548309990538](https://doi.org/10.1017/S0963548309990538). URL: <https://doi.org/10.1017/S0963548309990538>.