RANDOM WALKS ON GRAPHS

PIOTR NAYAR

Wharton Statistics Department
University of Pennsylvania

ABSTRACT.

1. Mixing via couplings

1.1. **Discrete time finite Markov chains.** Consider a finite set V with $|V| = n$ and a Markov kernel (or transition matrix) $P : V \times V \to \mathbb{R}$, i.e.,

$$P(x,y) \geq 0,\ x,y \in V \qquad \sum_{y \in V} P(x,y) = 1,\ x \in V.$$

The discrete time Markov chain associated with K with an initial distribution ν is a V-valued sequence $(X_n)_{n=0}^{\infty}$ whose law $\mathbb{P}_\nu$ is given by

$$\mathbb{P}_\nu[V_i = v_i,\ 0 \leq i \leq l] = \nu(x_0)P(x_0,x_1) \cdot \ldots \cdot P(x_{l-1},x_l), \qquad l = 0,1,\ldots$$

Consider the Markov chain started at x and set $\mathbb{P}_x = \mathbb{P}_{\delta_x}$. Then the law of X_l is given by $\mathbb{P}_x(X_l = y) = P^l(x,y)$, where P^l is defined recursively via

$$P^l(x,y) = \sum_{z \in V} P^{l-1}(x,z)P(z,y).$$

The kernel P defines an operator

$$(Pf)(x) = \sum_{y \in V} P(x,y)f(y).$$

Clearly, the lth power of this operator has kernel $P^l(x,y)$.

We also set $\mathbb{E}_\nu$ to be the expectation with respect to $\mathbb{P}_\nu$ and $\mathbb{E}_x$ to be the expectation with respect to $\mathbb{P}_x$. Let us also define

$$P_t^x(y) = P_t(x,y) = \mathbb{P}_x[X_t = y] = \mathbb{P}[X_t = y | X_0 = x].$$

1.2. **Gamblers ruin.** We start with a very simple example, called the *gamblers ruin*. Here $V = \{0,1,\ldots,n\}$, $|V| = n+1$ and the transition probabilities are given by

$$P(k,k+1) = P(k,k-1) = \frac{1}{2}, \quad k = 1,\ldots,n-1 \quad \text{and} \quad P(0,0) = P(n,n) = 1.$$

We can imagine a gambler playing the following simple round game. In each round he wins one dollar with probability $\frac{1}{2}$ and he with the same probability he looses one dollar. The game end when the player looses all his money, or reaches the amount of n dollars. Suppose

E-mail address: `nayar@mimuw.edu.pl`.

our player starts the game having k dollars. Let X_t, $t = 0, 1, \ldots$ be the (random) amount of money in player's pocket after time t. Let τ be the random time of ending the game, that is

$$\tau = \min\{t \geq 0 : \; X_t \in \{0, n\}\}.$$

Suppose we start with k dollars. What is a chance of winning the game? What is the expected playing time?

Fact 1. We have $\mathbb{P}_k[X_\tau = n] = \frac{k}{n}$ and $\mathbb{E}_k[\tau] = k(n-k)$.

Proof. Let $p_k = \mathbb{P}_k[X_\tau = n]$. Clearly we have

$$p_k = \frac{1}{2}p_{k-1} + \frac{1}{2}p_{k+1}, \quad 1 \leq k \leq n-1, \qquad p_0 = 0, \quad p_n = 1.$$

Thus $(p_k)_{k=0}^n$ is an arithmetic sequence with $p_0 = 0$ and $p_n = 1$. It follows that $p_k = k/n$.

To prove the second part let us set $T_k = \mathbb{E}_k[\tau]$. We have $T_0 = T_n = 0$. Moreover,

$$T_k = \frac{1}{2}(1 + T_{k-1}) + \frac{1}{2}(1 + T_{k+1}) = \frac{1}{2}(T_{k-1} + T_{k+1}) + 1, \qquad k = 1, \ldots, n-1.$$

Let $S_k = T_k + k^2$. For $1 \leq k \leq n-1$ we have

$$S_k - k^2 = T_k = \frac{1}{2}(T_{k-1} + T_{k+1}) + 1 = \frac{1}{2}(S_{k+1} - (k+1)^2 + S_{k-1} - (k-1)^2) + 1 = \frac{1}{2}(S_{k-1} + S_{k+1}) - k^2.$$

Thus, $S_k = \frac{1}{2}(S_{k-1} + S_{k+1})$. Since $S_0 = 0$ and $S_n = n^2$ we get $S_k = nk$. Thus $T_k = nk - k^2 = k(n-k)$. $\square$

1.3. **Coupon collecting.** A collector desires a complete set of n distinct coupons. The probability of getting a coupon k at a certain round is $1/n$. Let X_t be the number of different coupons accumulated at time t. Clearly, we have the following transition probabilities,

$$P(k, k+1) = \frac{n-k}{k}, \quad k = 1, \ldots, n-1, \qquad P(k, k) = \frac{k}{n}, \quad k = 1, \ldots, n.$$

Take

$$\tau = \min\{t \geq 0 : \; X_t = n\}.$$

How long will it take to collect all the coupons (starting with no coupons)?

Fact 2. We have $\mathbb{E}[\tau] = n \sum_{k=1}^n \frac{1}{k}$. Here $\mathbb{E} = \mathbb{E}_0$. In particular

$$\mathbb{E}[\tau] \leq n \ln n + n.$$

Proof. Let τ_k be the number of coupons accumulated when the collection first contained k distinct coupons. We have

$$\tau = (\tau_1 - \tau_0) + (\tau_2 - \tau_1) + \ldots + (\tau_n - \tau_{n-1}),$$

where $\tau_0 = 0$. Note that $\tau_k - \tau_{k-1}$ is a geometric random variable with success probability $p_k = \frac{n-k+1}{n}$, i.e.,

$$\mathbb{P}[\tau_k - \tau_{k-1} = l] = p_k(1-p_k)^{l-1}, \qquad l = 1, 2, \ldots.$$

Therefore,

$$\mathbb{E}[\tau_k - \tau_{k-1}] = \sum_{l=1}^{\infty} l p_k (1-p_k)^{l-1} = -p_k \frac{\mathrm{d}}{\mathrm{d}p_k} \sum_{l=0}^{\infty} (1-p_k)^l = -p_k \frac{\mathrm{d}}{\mathrm{d}p_k} \frac{1}{p_k} = \frac{1}{p_k} = \frac{n}{n-k+1}.$$

Thus,

$$\mathbb{E}[\tau] = \sum_{k=1}^{n} \mathbb{E}[\tau_k - \tau_{k-1}] = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{k=1}^{n} \frac{1}{k}.$$

Now the second part follows from the well known inequality $|\sum_{k=1}^{n} \frac{1}{k} - \ln n| \le 1$. □

We will soon need the following tail bound for τ in the coupon collector problem.

Fact 3. For any $c > 0$ we have $\mathbb{P}[\tau > \lceil n \ln n + cn \rceil] \le e^{-c}$.

Proof. Let

$$A_i = \{\text{coupon } i \text{ does not appear among the first } \lceil n \ln n + cn \rceil \text{ coupons}\}.$$

We have

$$\mathbb{P}[\tau > \lceil n \ln n + cn \rceil] = \mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] \le \sum_{i=1}^{n} \mathbb{P}[A_i] \le n\left(1 - \frac{1}{n}\right)^{\lceil n \ln n + cn \rceil} \le n e^{-\frac{n \ln n + cn}{n}} = e^{-c}.$$

□

1.4. **Stationary measure and the total variation distance.** Assume that our kernel P is strongly irreducible, i.e., there is i such that $P^i(x, y) > 0$ for every $x, y \in V$. This implies the existence of the unique stationary measure π. This means that

$$\pi(x) = \sum_{y \in V} \pi(y) P(y, x), \qquad \lim_{t \to \infty} P^t(x, y) = \pi(y).$$

In other words, $\lim_{t\to\infty} \mathbb{P}_x[X_t = y] = \pi(y)$. One can say that long time law of the chain is independent of the initial distribution and is given by the measure π.

Our goal is to bound the quantity $|\mathbb{P}_x[X_t \in A] - \pi(A)|$ for every $A \subseteq V$. In other words, we would like to estimate the rate of convergence to the stationary measure π. We need to work with the total variation distance,

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq V} |\mu(A) - \nu(A)|.$$

In the sequel we will need two simple lemmas.

Lemma 1. Let μ and ν be probability measures on V. Then

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in V} |\mu(x) - \nu(x)| = \sum_{x \in V:\ \mu(x) \ge \nu(x)} (\mu(x) - \nu(x)).$$

Proof. For the second equality note that

$$0 = \sum_{x} (\mu(x) - \nu(x)) = \sum_{x:\ \mu(x) \ge \nu(x)} (\mu(x) - \nu(x)) + \sum_{x:\ \mu(x) \le \nu(x)} (\mu(x) - \nu(x)).$$

Thus,

$$\begin{aligned} \frac{1}{2} \sum_{x} |\mu(x) - \nu(x)| &= \frac{1}{2} \sum_{x:\ \mu(x) \ge \nu(x)} (\mu(x) - \nu(x)) + \frac{1}{2} \sum_{x:\ \mu(x) \le \nu(x)} (\nu(x) - \mu(x)) \\ &= \sum_{x:\ \mu(x) \ge \nu(x)} (\mu(x) - \nu(x)). \end{aligned}$$

For the first part we define $B = \{x: \ \mu(x) \geq \nu(x)\}$. We have

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B).$$

Passing to the complement of A we also get

$$\nu(A) - \mu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B).$$

Thus,

$$|\mu(A) - \nu(A)| \leq \mu(B) - \nu(B) = \sum_{x: \ \mu(x) \geq \nu(x)} (\mu(x) - \nu(x)),$$

with equality for $A = B$. $\square$

Remark 1. From the above lemma we deduce the triangle inequality

$$\|\mu - \nu\|_{TV} \leq \|\mu - \eta\|_{TV} + \|\eta - \nu\|_{TV}.$$

1.5. **Couplings.**

Definition 1. A coupling of two measures μ and ν on V is a measure m on $V \times V$ such that $m(V \times \{x\}) = \nu(x)$ and $m(\{x\} \times V) = \mu(x)$. In other words, a coupling of μ and ν is a pair of random variables (X, Y) such that $X \sim \mu$ and $Y \sim \nu$.

Lemma 2. We have

$$\|\mu - \nu\|_{TV} = \inf \{\mathbb{P}[X \neq Y] : \ (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

Proof. We have

$$\begin{aligned}
\mu(A) - \nu(A) &= \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] = \mathbb{P}[X \in A, Y \notin A] + \mathbb{P}[X \in A, Y \in A] - \mathbb{P}[Y \in A] \\
&\leq \mathbb{P}[X \in A, Y \notin A] \leq \mathbb{P}[X \neq Y].
\end{aligned}$$

Changing the roles of μ and ν and taking the supremum with respect to A gives

$$\|\mu - \nu\|_{TV} \leq \inf \{\mathbb{P}[X \neq Y] : \ (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

To construct a coupling with $\|\mu - \nu\|_{TV} = \mathbb{P}[X \neq Y]$ take the random vector (X, Y) with

$$\mathbb{P}[X = Y = z] = \min\{\mu(z), \nu(z)\}, \qquad z \in V$$

and

$$\mathbb{P}[X = x, Y = y] = \frac{(\mu(x) - \nu(x)(\nu(y) - \mu(y)))}{\|\mu - \nu\|_{TV}} \mathbf{1}_{\mu(x) \geq \nu(x)} \mathbf{1}_{\nu(y) \geq \mu(y)}, \qquad x \neq y.$$

To check that this is indeed a coupling we write

$$\begin{aligned}
\mathbb{P}[X = x] &= \sum_y \mathbb{P}[X = x, Y = y] \\
&= \mathbb{P}[X = Y = x] + \sum_{y: \ \nu(y) \geq \mu(y)} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{TV}} (\nu(y) - \mu(y)) \mathbf{1}_{\mu(x) \geq \nu(x)} \\
&= \min\{\mu(x), \nu(x)\} + (\mu(x) - \nu(x)) \mathbf{1}_{\mu(x) \geq \nu(x)} = \mu(x).
\end{aligned}$$

Similarly $\mathbb{P}[Y=y]=\nu(y)$. Moreover,

$$\begin{aligned}\mathbb{P}[X\neq Y] &= 1-\mathbb{P}[X=Y]=1-\sum_x \min\{\mu(x),\nu(x)\}\\ &= 1-\frac{1}{2}\sum_x(\mu(x)+\nu(x)-|\mu(x)-\nu(x)|)=\frac{1}{2}\sum_x|\mu(x)-\nu(x)|=\|\mu-\nu\|_{TV},\end{aligned}$$

due to Lemma 1. $\square$

We now define two crucial quantities we would like to control,

$$d(t)=\max_{x\in V}\|P_t^x(\cdot)-\pi(\cdot)\|_{TV}, \qquad \tilde{d}(t)=\max_{x,y\in V}\|P_t^x(\cdot)-P_t^y(\cdot)\|_{TV}.$$

We now show the comparison between those two quantities.

Lemma 3. We have $d(t)\le\tilde{d}(t)\le 2d(t)$.

Proof. We first note that

$$\|P_t^x(\cdot)-P_t^y(\cdot)\|_{TV}\le\|P_t^x(\cdot)-\pi(\cdot)\|_{TV}+\|\pi(\cdot)-P_t^y(\cdot)\|_{TV}\le 2d(t).$$

Taking the supremum over x and y gives $\tilde{d}(t)\le 2d(t)$.

Now for the first inequality note that by stationarity of π we have

$$\pi(A)=\sum_y \pi(y)P_t(y,A), \qquad A\subseteq V.$$

Thus,

$$\begin{aligned}\|P_t^x(\cdot)-\pi(\cdot)\|_{TV} &= \max_{A\subseteq V}|P_t(x,A)-\pi(A)|=\max_{A\subseteq V}\left|\sum_y\pi(y)(P_t(x,A)-P_t(y,A))\right|\\ &\le \max_{A\subseteq V}\sum_y\pi(y)|P_t(x,A)-P_t(y,A)|\le\sum_y\pi(y)\max_{A\subseteq V}|P_t(x,A)-P_t(y,A)|\\ &= \sum_y\pi(y)\,\|P_t^x(\cdot)-P_t^y(\cdot)\|_{TV}\le\max_{x,y}\|P_t^x(\cdot)-P_t^y(\cdot)\|_{TV}=\tilde{d}(t).\end{aligned}$$

Therefore $d(t)\le\tilde{d}(t)$. $\square$

The following lemma shows monotonicity of d and $\tilde{d}$.

Lemma 4. The functions $d(t)$ and $\tilde{d}(t)$ are non-increasing.

Proof. Let us define the action of P on the space of measures,

$$(\mu P)(x)=\sum_{y\in V}\mu(y)P(y,x).$$

Note that

$$\|\mu P-\nu P\|_{TV}\le\|\mu-\nu\|_{TV}.$$

Indeed,

$$\|\mu P - \nu P\|_{TV} = \frac{1}{2}\sum_x |(\mu P)(x) - (\nu P)(x)| = \frac{1}{2}\sum_x \left|\sum_y \mu(y)P(y,x) - \nu(y)P(y,x)\right|$$
$$\leq \frac{1}{2}\sum_{x,y} p(y,x)|\mu(y)-\nu(y)| = \frac{1}{2}\sum_y |\mu(y)-\nu(y)| = \|\mu-\nu\|_{TV}.$$

We have $P^x_{t+1}(\cdot) = \mu P^x_t(\cdot)$. It follows that

$$\left\|P^x_{t+1}(\cdot) - P^y_{t+1}(\cdot)\right\|_{TV} = \|P^x_t(\cdot)P - P^y_t(\cdot)P\|_{TV} \leq \|P^x_t(\cdot) - P^y_t(\cdot)\|_{TV}$$

and thus $\tilde{d}(t)$ is non-increasing.

Note also that stationarity means $\pi P = P$. Thus

$$\left\|P^x_{t+1}(\cdot) - \pi\right\|_{TV} = \|P^x_t(\cdot)P - \pi P\|_{TV} \leq \|P^x_t(\cdot) - \pi\|_{TV}.$$

This shows that $d(t)$ is non-increasing. □

Lemma 5. We have $\tilde{d}(s+t) \leq \tilde{d}(s)\tilde{d}(t)$. In particular, for any integers $l, t \geq 0$ we have $d(lt) \leq \tilde{d}(lt) \leq \tilde{d}(t)^l$.

Proof. Fix $x, y \in V$ and set (X_s, Y_s) to be the optimal coupling of $P^x_s(\cdot)$ and $P^y_s(\cdot)$, i.e.,

$$\|P^x_s(\cdot) - P^y_s(\cdot)\|_{TV} = \mathbb{P}[X_s \neq Y_s].$$

We have

$$P_{s+t}(x,\omega) = \sum_z P_s(x,z)P_t(z,\omega) = \sum_z \mathbb{P}[X_s = z]P_t(z,\omega) = \mathbb{E}P_t(X_s,\omega).$$

Similarly, $P_{s+t}(y,\omega) = \mathbb{E}P_t(Y_s,\omega)$. Thus,

$$P_{s+t}(x,\omega) - P_{s+t}(y,\omega) = \mathbb{E}[P_t(X_s,\omega) - P_t(Y_s,\omega)].$$

We get

$$\|P_{s+t}(x,\cdot) - P_{s+t}(y,\cdot)\|_{TV} = \frac{1}{2}\sum_\omega |\mathbb{E}[P_t(X_s,\omega) - P_t(Y_s,\omega)]| = \mathbb{E}\,\|P_t(X_s,\cdot) - P_t(Y_s,\cdot)\|_{TV}$$
$$\leq \tilde{d}(t)\mathbb{E}\mathbf{1}_{X_s\neq Y_s} = \tilde{d}(t)\mathbb{P}[X_s \neq Y_s]$$
$$= \tilde{d}(t)\,\|P_s(x,\cdot) - P_s(y,\cdot)\|_{TV} = \tilde{d}(t)\tilde{d}(s).$$

Thus $\tilde{d}(s+t) \leq \tilde{d}(s)\tilde{d}(t)$. The second assertion follows easily. □

1.6. **Mixing times.** Let us define the *mixing time* with parameter $\varepsilon > 0$,

$$t_{mix}(\varepsilon) = \min\{t \geq 0 : \; d(t) \leq \varepsilon\}.$$

Moreover, we set

$$t_{mix} = t_{mix}(1/4).$$

Note that

$$d(lt_{mix}(\varepsilon)) \leq \tilde{d}(lt_{mix}(\varepsilon)) \leq \tilde{d}(t_{mix}(\varepsilon))^l \leq (2d(t_{mix}(\varepsilon)))^l \leq (2\varepsilon)^l.$$

Taking $\varepsilon = 1/4$ we get $d(lt_{mix}) \leq 2^{-l}$. We have $2^{-l_0} \leq \varepsilon$ for $l_0 = \lceil \ln(1/\varepsilon)\rceil$. Thus,

$$t_{mix}(\varepsilon) \leq l_o t_{mix} = \lceil \ln(1/\varepsilon)\rceil\, t_{mix}.$$

Definition 2. A coupling of two Markov chains $\mathcal{C}_1$ and $\mathcal{C}_2$ on V is a process $(X_t, Y_t)_{t\geq 0}$ where $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ have the same distribution as $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.

Before we proceed further, let us give one simple example of a coupling of two Markov chains.

Claim 1. Let $V = \{0, 1, \ldots, n\}$ and consider the Gambler's ruin chain. If $0 \le x \le y \le n$ then $P_t(x, n) \le P_t(y, n)$.

This is an intuitive statement, but probably not so trivial to prove. Let us give a very simple explanation using couplings.

Proof. We construct a coupling of Gambler's ruin chains $\mathcal{C}_x$ and $\mathcal{C}_y$ started at x and y, respectively, in the following fashion. Start with a vector (x, y). At each step toss a fair coin and move both coordinates up if heads and both coordinates down if tails. The only restriction is that if one of the coordinates gets value 0 or n, we no longer perform moves for this particular coordinate. This is process $(X_t, Y_t)_{t\ge 0}$ is clearly a coupling of $\mathcal{C}_x$ and $\mathcal{C}_y$. Moreover, by construction we always have $X_t \le Y_t$. Thus

$$P_t(x, n) = \mathbb{P}(X_t = n) \le \mathbb{P}(Y_t = n) = P_t(y, n).$$

$\square$

If (X_t, Y_t) is a coupling of two Markov chains then we can define a modified coupling as follows. Define $\tau_c = \min\{t \ge 0 : \; X_t = Y_t\}$. Then the chain

$$(\tilde{X}_t, \tilde{Y}_t) = \begin{cases} (X_t, Y_t) & t \le \tau \\ (X_t, X_t) & t \ge \tau \end{cases}$$

is also a coupling of those two chains. We won't prove this statement because of two reasons. First is that it is completely obvious. The other is that writing down the proof requires some time and board/paper space and we would like to save those resources for more explanatory material. We will always assume that our couplings have the above form and will call them *standardized chains.*

The following lemma shows the relation between couplings and mixing times.

Lemma 6. Let τ_c be defined as above and consider a standardized coupling. We have

$$\|P_t^x(\cdot) - P_t^y(\cdot)\|_{TV} \le \mathbb{P}_{x,y}[\tau_c > t].$$

In particular

$$d(t) \le \max_{x,y} \mathbb{P}_{x,y}[\tau_c > t].$$

Proof. From Lemma 2 we have

$$\|P_t^x(\cdot) - P_t^y(\cdot)\|_{TV} \le \mathbb{P}_{x,y}[X_t \ne Y_t] = \mathbb{P}_{x,y}[\tau_c > t].$$

$\square$

Now we use the above theory to upper bound mixing times for several important and basic Markov chains.

Example 1 (Lazy random walk on a circle)**.** Consider an n-point discrete circle (cyclic group of cardinality n). We can identify it with the set $V = \{0, \ldots, n-1\}$. The transition matrix is the following,

$$P(k, k) = 1/2, \quad k = 0, \ldots, n-1, \qquad P(k, k+1) = P(k, k-1) = 1/4, \quad k = 0, \ldots, n-1,$$

where $(n-1) + 1 = 0$ and $0 - 1 = n - 1$ (in other words $+$ means adding modulo n).

Let us consider the above chain $\mathcal{C}_x$ started at x and $\mathcal{C}_y$ started at y. We construct the following coupling. Start X_t at x and Y_t at y. At each step toss a fair coin. If heads move X_t clockwise (i.e., add 1) with probability $1/2$ and counter-clockwise (subtract 1) with probability $1/2$. If tails do the same for Y_t. After X_t and Y_t meet, perform the same walk for both chains.

Clearly, we have construct a standardized coupling of $\mathcal{C}_x$ and $\mathcal{C}_y$. Let $D_t \in \{0, 1, \ldots, n\}$ be the clockwise distance between X_t and Y_t. Note that both chains meet when D_t reaches value 0 or n. It is easy to observe that D_t is a Gambler's ruin chain. Thus, from Lemma 6, Fact 1 and Markov inequality we have

$$d(t) \leq \max_{x,y\in\mathbb{Z}_n} [\tau_c > t] \leq \frac{\max_{x,y\in\mathbb{Z}_n} \mathbb{E}_{x,y}[\tau_c]}{t} \leq \frac{1}{t}\max_k (k(n-k)) = \frac{n^2}{4t}.$$

Thus, $t_{mix} \leq n^2$. One can show indeed $t_{mix} \approx n^2$ up to universal constants.

Example 2 (Random walk on the hypercube)**.** Let us consider a set $V = \{0,1\}^n$ and define the following Markov chain on V. We toss a fair coin. If heads, we do nothing. If tails, we pick a random coordinate (with probability $1/d$) and flip the bit on this direction.

Let us consider the above chain $\mathcal{C}_x$ started at x and $\mathcal{C}_y$ started at y. Here is our coupling. As always, start X_t at x and Y_t at y. Pick a random coordinate (with probability $1/d$) and replace this coordinate with a random bit (the same random bit for both chains). One can easily check that this is indeed a standardized coupling of $\mathcal{C}_x$ and $\mathcal{C}_y$. The number of coordinates replaced with random bits after time t is a coupon collector process. Thus,

$$d(\lceil n\ln n + cn\rceil) \leq \mathbb{P}[\tau > n\ln n + cn] \leq e^{-c}, \qquad C > 0.$$

Thus,

$$t_{mix}(\varepsilon) \leq \lceil n\ln n + n\log(1/\varepsilon)\rceil \leq n\ln n + n\log(1/\varepsilon) + 1.$$

Example 3 (Lazy random walk on the discrete d-torus $\mathbb{Z}_n^d$)**.** The definition of this chain is natural. We first choose a random coordinate and the perform a move of a discrete circle random walk. We couple $\mathcal{C}_x$ and $\mathcal{C}_y$ as follows. We firs pick a random coordinate. If the positions of X_t and Y_t agree on that coordinate, perform a usual step of a discrete circle chain and move both X_t and Y_t. Otherwise we randomly choose one of the processes to move and the other one to stay. It is again easy to see that we have constructed a standardized coupling.

Let

$$X_t = (X_t^{(1)}, \ldots, X_t^{(d)}), \qquad Y_t = (Y_t^{(1)}, \ldots, Y_t^{(d)})$$

and

$$\tau_i = \min\{t \geq 0 : \ X_t^{(i)} = Y_t^{(i)}\}.$$

Note that if $X_1, \ldots, X_n$ are i.i.d. random variables with mean μ and if τ is independent of the above sequence, then

$$\mathbb{E}\left[\sum_{i=1}^{\tau} X_i\right] = \mu\mathbb{E}[\tau].$$

Indeed,

$$\mathbb{E}\left[\sum_{i=1}^{\tau} X_i\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]\mathbb{P}[\tau = n] = \sum_{n=1}^{\infty} n\mu\mathbb{P}[\tau = n] = \mu\mathbb{E}[\tau].$$

Note that τ_i has precisely the above structure, where τ is equal to τ_c of discrete circle random walk and $X_1, l \dots, X_n$ are waiting times for picking ith coordinate. Thus, each X_i has geometric distribution with mean d. Thus,

$$\mathbb{E}_{x,y}[\tau_i] = d\mathbb{E}_{x,y}[\tau_{\text{circle}}] \leq \frac{dn^2}{4}.$$

Let τ_c be the standard "couple time" for our chain. We get

$$\mathbb{E}_{x,y}[\tau_c] \leq \mathbb{E}\left[\sum_{i=1}^{d} \tau_i\right] \leq \frac{d^2n^2}{4}.$$

Thus,

$$d(t) \leq \mathbb{P}_{x,y}[\tau_c > t] \leq \frac{1}{t}\mathbb{E}_{x,y}[\tau_c] \leq \frac{d^2n^2}{4t}.$$

For $t_0 = d^2n^2$ we get $d(t_0) \leq 1/4$ and thus $t_{mix} \leq d^2n^2$. Therefore $t_{mix}(\varepsilon) \leq d^2n^2 \lceil \ln(1/\varepsilon) \rceil$.

To get an optimal dependence on d we have to work a little more. Note that for any x, y we have

$$\mathbb{P}_{x,y}[\tau_i > dn^2] \leq \frac{\mathbb{E}_{x,y}\tau_i}{dn^2} \leq \frac{1}{4}.$$

Thus for any integer $k \geq 1$ we have $\mathbb{P}_{x,y}[\tau_i > kdn^2] \leq \frac{1}{4^k}$ (since we can consider k periods of length dn^2 of waiting for coupling the ith coordinate). We get

$$\mathbb{P}_{x,y}[\tau_c > kdn^2] \leq \sum_{i=1}^{n} \mathbb{P}_{x,y}[\tau_i > kdn^2] \leq \frac{d}{4^k}.$$

The right hand side is smaller than $1/4$ for $k \geq \left\lceil \frac{1}{2}\log_2 d + 1 \right\rceil$. From Lemma 6 we get

$$t_{mix} \leq \left\lceil \frac{1}{2}\log_2 d + 1 \right\rceil dn^2.$$

Example 4 (Metropolis chain on graph colourings)**.** Let G be a graph with maximal degree Δ. Suppose we have a list of q colors and we want to colour the vertex set of our graph according to the rule that any two neighbours must have different color. We call such a colouring proper. Now we run a Markov chain on the set of all proper colourings. We first choose a vertex v uniformly at random. Then we choose one of the q colors at random and we update the color of v if this leads to a proper colouring.

Note that it is easy to verify that this chain satisfies the detailed balance condition with π being the uniform measure on proper colourings. Note also that in order to have this chain transitive we need to assume that $q > \Delta + 1$.

We shall prove the following theorem.

Theorem 1. Let G be a graph with n vertices and maximal degree Δ. If $q > 3\Delta$ we have

$$t_{mix}(\varepsilon) \leq \frac{1}{1 - \frac{3\Delta}{q}} n \left(\ln n + \log\left(\frac{1}{\varepsilon}\right)\right).$$

Proof. Let $\mathcal{C}_x$ be random walks started at *any* colourings x (not necessarily proper). However, we keep the updating rule unchanged. We construct X_t as follows. We choose a vertex v uniformly at random and then one of the q colors uniformly at random. We update X_t whenever our color is admissible. The main point of this procedure is that we use the same vertex an color for *every* starting colouring x. This is called the grand coupling.

For a colouring x let $x(v)$ be the color of the vertex v. Let us introduce the following distance,

$$D(x,y) = |\{v : \ x(v) \neq y(v)\}|.$$

Let $X_t^{(y)}$ be the Markov chain started at colouring y. Note that $X_t^{(x)} = X_t$ and $X_t^{(y)} = Y_t$.

We first consider the case when $D(x,y) = 1$, that is, the colourings differ at only one vertex v_0. Let $\mathcal{N}$ be the set of colors appearing in the neighbourhood of v_0 in x (equivalently, in y). We want to compute the probability that after one step the chains will meet. Note that in order for this to happen we have to choose vertex v_0 and a colour outside of $\mathcal{N}$. Thus,

$$\mathbb{P}(D(X_1^x, X_1^y) = 0) = \frac{1}{n} \cdot \frac{q - |\mathcal{N}|}{q} \geq \frac{q-\Delta}{nq}.$$

Unfortunately, D can go up to 2. For this to happen we have to choose one of the neighbours of v_0 and we have to choose one of the colors from the set $\{X_0^{(x)}(v_0), X_0^{(y)}(v_0)\}$. Thus,

$$\mathbb{P}(D(X_1^x, X_1^y) = 2) \leq \frac{\Delta}{n} \cdot \frac{2}{q}.$$

It follows that

$$\mathbb{E}D(X_1^x, X_1^y) \leq 1 + \frac{2\Delta}{nq} - \frac{q-\Delta}{nq} = 1 - \frac{q-3\Delta}{nq} \in (0,1).$$

Suppose now x, y are colouring with $D(x,y) = r > 1$. Let $x = z_0, z_1, \ldots, z_r = y$ be a path between colourings (again, not necessarily proper), that is, $D(z_{k-1}, z_k) = 1$, $k = 1, \ldots, r$. Since D is a metric, we get

$$\mathbb{E}D(X_1^x, X_1^y) \leq \sum_{k=1}^{r} \mathbb{E}D(X_1^{(z_{k-1})}, X_1^{(z_k)}) \leq r \left(1 - \frac{q-3\Delta}{nq}\right) = D(x,y)\left(1 - \frac{q-3\Delta}{nq}\right).$$

Similarly,

$$\begin{aligned}\mathbb{E}\left[D(X_t^x, X_t^y) \mid X_{t-1}^x = x_{t-1}, X_{t-1}^y = y_{t-1}\right] &= \mathbb{E}\left[D(X_1^{x_{t-1}}, X_1^{y_{t-1}})\right] \\ &\leq D(x_{t-1}, y_{t-1})\left(1 - \frac{q-3\Delta}{nq}\right).\end{aligned}$$

In other words,

$$\mathbb{E}\left[D(X_t^x, X_t^y) \mid X_{t-1}^x, X_{t-1}^y\right] \leq D(X_{t-1}^x, X_{t-1}^y)\left(1 - \frac{q-3\Delta}{nq}\right).$$

Taking expectation,

$$\mathbb{E}\left[D(X_t^x, X_t^y)\right] \leq \mathbb{E}D(X_{t-1}^x, X_{t-1}^y)\left(1 - \frac{q-3\Delta}{nq}\right).$$

Iterating we get

$$\mathbb{E}\left[D(X_t^x, X_t^y)\right] \leq D(x,y)\left(1 - \frac{q-3\Delta}{nq}\right)^t.$$

Thus, by Markov inequality

$$\mathbb{P}(X_t^x \neq X_t^y) = \mathbb{P}(D(X_t^x, X_t^y) \geq 1) \leq \mathbb{E}D(X_t^x, X_t^y) \leq D(x,y)\left(1 - \frac{q-3\Delta}{nq}\right)^t \leq n e^{-\frac{t}{n} \cdot \frac{q-3\Delta}{q}}.$$

The last expression is smaller than ε for

$$t = \frac{1}{1 - \frac{3\Delta}{q}} n \left(\ln n + \log\left(\frac{1}{\varepsilon}\right)\right).$$

This holds for all colourings, in particular for proper colourings. □

2. Expander graphs

Definition 3. Let $G = (V, E)$ be a graph. Define

$$\Gamma(S) = \{v \in V \setminus S \mid \exists u \in S \ v \sim u\}.$$

We say that G is a (n, d, c)-expander when $|V| = n$, G is d-regular and for any $S \subset V$ with $|S| \leq n/2$ we have $|\Gamma(S)| \geq c|S|$.

Fact 4. Any two vertices x, y in a (n, d, c)-expander can be connected with a path of length at most $\left\lceil 2\left(\log_{1+c}\left(\frac{n}{2(d+1)}\right) + 1\right)\right\rceil$.

Proof. Let

$$N_k(x) = |\{z \in V \mid d(x, z) \leq k\}|.$$

Clearly, $N_1(x) = d+1$ for any $x \in V$. Moreover, $N_2(x) \geq d+1+c(d+1) = (1+c)(d+1)$. By induction we get $N_k(x) \geq (1+d)(1+c)^{k-1}$. Thus, $N_k(x) > n/2$ if $k_0 > 1+\log_{1+c}\left(\frac{n}{2(1+d)}\right)$. As a consequence, there is a vertex z in the intersection $N_{k_0}(x) \cap N_{k_0}(y)$. The assertion follows by triangle inequality $d(x, y) \leq d(x, z) + d(z, y)$. □

Let A be the adjacency matrix of G, that is

$$a_{uv} = \begin{cases} 1 & u \sim v \\ 0 & \text{otherwise} \end{cases}.$$

Let

$$e(S, T) = \{(u, v) \mid u \in S, v \in T\}.$$

Assume that $S \cap T = \emptyset$. Then $e(S, T) \leq |S| \cdot |T|$.

Fact 5. Let $G = (V, E)$ be a d-regular graph on n vertices. Let λ_2 be the second largest eigenvalue of its adjacency matrix A. Then for every partition $V = S \cup T$ we have

$$e(S, T) \geq \frac{d - \lambda_2}{n} \cdot |S| \cdot |T|.$$

Proof. If $\mathbf{1} = (1, \ldots, 1)$ then $A\mathbf{1} = d\mathbf{1}$ (this is due to the fact that G is d-regular). The matrix A is symmetric, thus it has eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. We claim that $\lambda_1 = d$. Indeed, for any vector $x \in \mathbb{R}^n$ we have

$$x^T(dI - A)x = d\sum_{v \in V} x_v^2 - 2\sum_{\{u,v\} \in E} x_u x_v = \sum_{\{u,v\} \in E} (x_u - x_v)^2 \geq 0.$$

Here the notation $\{u, v\} \in E$ means that the edge (u, v) is the same as (v, u) and is counted once. It follows that all the eigenvalues of A are not greater than d. Note also that from the above inequality it follows that G is connected if and only if $\lambda_2 < d$.

For any $x \perp \mathbf{1}$ we have $x^T A x \leq \lambda_2 |x|^2$. Take $x = \mathbf{1}_S - \frac{|S|}{n}\mathbf{1}$. We have

$$(d - \lambda_2)|x|^2 \leq x^T(dI - A)x = \sum_{\{u,v\} \in E} (x_u - x_v)^2 = e(S, T).$$

To finish the proof it suffices to observe that

$$|x|^2 = \langle x, x\rangle = \left\langle \mathbf{1}_S - \frac{|S|}{n}\mathbf{1}, \mathbf{1}_S - \frac{|S|}{n}\mathbf{1}\right\rangle = |S| - 2\frac{|S|^2}{n} + \frac{|S|^2}{n} = \frac{1}{n}|S|(n - |S|).$$

□

Fact 6. Let λ_2 be the second eigenvalue of the adjacency matrix A of a regular graph G. Then G is a $\left(n, d, \frac{d-\lambda}{2d}\right)$-expander.

Proof. Note that for d-regular graphs $|e(S, S^c)| \le d|\Gamma(S)|$. Thus, for $|S| \le n/2$ we have

$$|\Gamma(S)| \ge \frac{1}{d}|e(S, S^c)| \ge \frac{d - \lambda_2}{dn}|S|(n - |S|) \ge \frac{d - \lambda_2}{2dn}|S|.$$

□

Fact 7. Let P be the transition matrix of a reversible Markov chain. Let $\pi_\star = \min_{x\in\Omega} \pi(x)$. Then

$$t_{mix}(\varepsilon) \le \log\left(\frac{1}{\varepsilon\pi_\star}\right)\frac{1}{\gamma_\star},$$

where

$$\gamma_\star = 1 - \max\left\{|\lambda| \mid \lambda - \text{eigenvalue of } P \text{ with } \lambda \ne 1\right\}.$$

Proof. Note that $(\sqrt{\pi(x)})_{x\in V}$ is an eigenvector of P with eigenvalue 1. Indeed,

$$\sum_y \sqrt{\frac{\pi(x)}{\pi(y)}}P(x, y)\sqrt{\pi(y)} = \sqrt{\pi(x)}.$$

Let D_π be a diagonal matrix with $D_\pi(x, x) = \pi(x)$. Take $A = D_\pi^{1/2} P D_\pi^{-1/2}$. Note that A is symmetric. Indeed,

$$A(x, y) = \sqrt{\frac{\pi(x)}{\pi(y)}}P(x, y) = \sqrt{\frac{\pi(y)}{\pi(x)}}P(y, x),$$

which is true due to the detailed balance condition $\pi(x)P(x, y) = \pi(y)P(y, x)$. Let $(\varphi_x)_x$ be the orthonormal basis formed by eigenvectors $(\mu_x)_x$ of A. Take $f_x = D_\pi^{-1/2}\varphi_x$. Note that

$$Pf_x = PD_\pi^{-1/2}\varphi_x = D_\pi^{-1/2}(D_\pi^{1/2}PD_\pi^{-1/2})\varphi_x = D_\pi^{-1/2}A\varphi_x = D_\pi^{-1/2}\mu_x\varphi_x = \mu_x f_x.$$

Moreover,

$$\langle \varphi_x, \varphi_y\rangle = \left\langle D_\pi^{1/2} f_x, D_\pi^{1/2} f_y\right\rangle = \langle f_x, f_y\rangle_\pi.$$

Thus, the vectors $(f_x)_x$ form an orthonormal basis with respect to the scalar product $\langle\cdot,\cdot\rangle_\pi$. Moreover, they are eigenvectors for the matrix P.

Take $\delta_y(x) = \mathbf{1}_{x=y}$. We have

$$\delta_y = \sum_z \langle \delta_y, f_z\rangle_\pi f_z = \sum_z f_z(y)\pi(y) f_z.$$

Recall that

$$(Pf)(x) = \sum_y f(y)P(x, y).$$

Thus,

$$P(\delta_y)(x) = \sum_z \delta_y(z)P(x, z) = P(x, y).$$

Also,

$$P_t(x,y) = P^t(\delta_y)(x) = P^t\left(\sum_z f_z(y)\pi(y)f_z\right)(x) = \sum_z f_z(y)\pi(y)P^t\left(f_z\right)(x)$$
$$= \sum_z f_z(y)\pi(y)\lambda_z^t(x)f_z(x).$$

Let us order the vectors $(f_x)_x$ such that $f_1(x) = \sqrt{\pi(x)}$ is the eigenvector with eigenvalue 1. We have $f_1 = \sqrt{\pi}^{-1}\varphi_1 = \mathbf{1}$. Define $\lambda_\star = 1 - \gamma_\star$. Then

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| = \left|\sum_{j=2}^n f_j(x)f_j(y)\lambda_j^t\right| \le \sum_{j=2}^n |f_j(x)f_j(y)|\,\lambda_\star^t \le \lambda_\star^t\left(\sum_{j=2}^n f_j(x)^2 \sum_{j=2}^n f_j(y)^2\right)^{1/2}.$$

Note that

$$\pi(x) = \langle \delta_x, \delta_x\rangle_\pi = \left\langle \sum_{j=1}^n f_j(x)\pi(x)f_j, \sum_{j=1}^n f_j(x)\pi(x)f_j\right\rangle_\pi$$
$$= \pi(x)^2\sum_{j=1}^n f_j(x)^2 \ge \pi(x)^2\sum_{j=2}^n f_j(x)^2.$$

We get

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \le \frac{\lambda_\star^t}{\sqrt{\pi(x)\pi(y)}} \le \frac{(1-\gamma_\star)^t}{\pi_\star} \le \frac{e^{-\gamma_\star t}}{\pi_\star}.$$

Therefore,

$$\left\|P^t(x,\cdot) - \pi\right\|_{TV} = \sum_{y:\ P^t(x,y)<\pi(y)} (\pi(y) - P^t(x,y)) = \sum_{y:\ P^t(x,y)<\pi(y)} \pi(y)\left(1 - \frac{P^t(x,y)}{\pi(y)}\right)$$
$$\le \max_y \left|1 - \frac{P^t(x,y)}{\pi(y)}\right| \le \frac{e^{-\gamma_\star t}}{\pi_\star}.$$

□

Remark 2. For a lazy version of a random walk, i.e., $\tilde{P} = \frac{1}{2}P + \frac{1}{2}I$, all the eigenvalues $\tilde{\lambda}_i = \frac{1}{2}\lambda_i + \frac{1}{2}$ are non-negative. Indeed, it suffices to show that $I + P$ is positive semi-definite. To show this we observe that under the detailed-balance condition we have

$$\langle (I+P)f, f\rangle_\pi = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 P(x,y)\pi(x).$$

Then $\tilde{\gamma_\star} = 1 - \tilde{\lambda}_2$.

2.1. **Ramanujan graphs.** Let A be a adjacency matrix of a graph G and let $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n$ be the eigenvalues of A. Note that

$$x^T(dI \pm A)x = \sum_{\{u,v\}\in E} (x_u \pm x_v)^2 \ge 0.$$

It follows that for every $i = 1, \ldots, n$ we have $\lambda_i \in [-d, d]$. We already know that $\lambda_1 = d$. Moreover, from the above identity we see that $\lambda_n = -d$ when G is bipartite. Indeed, then

$G = S \cup T$ and setting $x_u = 1$ for $u \in S$ and $x_u = -1$ for $u \in T$ gives $x^T(dI + A)x = 0$, that is

$$d = -\frac{x^T Ax}{|x|^2} \leq -\min_{x \neq 0} \frac{x^T Ax}{|x|^2} = -\lambda_n \leq d.$$

Thus $\lambda_n = -d$. For bipartite graphs d and $-d$ will be called trivial eigenvalues.

Fact 8. The adjacency matrix of a bipartite graph G has symmetric spectrum.

Proof. Suppose that $G = S \cup T$ and $|S| = k$, $|T| = l$. The matrix A can be written in the form

$$A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix},$$

where B is a $k \times l$ matrix. Suppose that a vector $[u, v] \in \mathbb{R}^k \times \mathbb{R}^l$ is an eigenvector of A with eigenvalues λ,

$$A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Bv \\ B^T u \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix},$$

that is, $Bv = \lambda u$, $B^T u = \lambda v$. We the have

$$A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \cdot \begin{bmatrix} u \\ -v \end{bmatrix} = \begin{bmatrix} -Bv \\ B^T u \end{bmatrix} = -\lambda \begin{bmatrix} u \\ -v \end{bmatrix},$$

Thus, the vector $[u, -v]$ is an eigenvector with eigenvalue $-\lambda$. $\square$

A bipartite graph G is a good expander if all non-trivial eigenvalues are contained in a small symmetric interval around 0. The following theorem shows that, for big graphs, this interval cannot be much smaller than $2\sqrt{d-1}$.

Theorem 2 (Alon-Boppana, 1986)**.** For every $\varepsilon > 0$ there exists N such that any d-regular graph on N vertices has a non-trivial eigenvalue λ with $|\lambda| \geq 2\sqrt{d-1} - \varepsilon$.

So, for an infinite family of graphs, having all the non-trivial eigenvalues in the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ is the best what we could hope for. This motivates the following definition.

Definition 4. A d-regular graph having all the non-trivial eigenvalues in the symmetric interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ is called Ramanujan. Infinite family of d-regular Ramanujan graphs is called Ramanujan family.

We are not going to prove the above theorem. Our goal is to show the existence of Ramanujan families of d-regular bipartite graphs.

We now introduce the concept of 2-lifts. For a graph $G = (V, E)$ we define the 2-lift G_s, where $s \in \{\pm 1\}^{|E|}$, as follows. Take two copies of G_1, G_2 of G and set $V(G_s) = V(G_1) \cup V(G_2)$. Assume that $e = (a, b)$ is an edge in G. Let a_1, b_1 and a_2, b_2 be corresponding vertices in G_1 and G_2. If $s(e) = 1$ then we add to $E(G_s)$ the edges (a_1, b_1) and (a_2, b_2). If $s(e) = 1$ then we add to $E(G_s)$ the edges (a_1, b_2) and (a_2, b_1).

There are $2^{|E|}$ lifts and for every 2-lift we have $\deg_G(a) = \deg_{G_s}(a_1) = \deg_{G_s}(a_2)$. It is easy to see that the following is true.

Fact 9. Suppose that G is a bipartite d-regular graph. Then for every s the graph G_s is d-regular and bipartite.

Let A be an adjacency matrix and let s be a signing. We define the signed adjacency matrix A_s as follows. For every edge $e = (u, v)$ multiply the entry a_{uv} by the corresponding sign $s(e)$.

Fact 10 (Bilu-Linial, 2006)**.** For ant s the eigenvalues of $A(G_s)$ are the union of the set of eigenvalues of A and the eigenvalues of A_s.

Proof. We have

$$A(G_s) = \begin{bmatrix} A_1 & A_2 \\ A^2 & A_1 \end{bmatrix},$$

where A_1 is the adjacency matrix of $V(G_1)$ with edges corresponding to positive lifts and A_2 corresponds to edges going form G_1 to G_2.

Clearly $A = A_1 + A_2$ and $A_s = A_1 - A_2$. Let v be an eigenvector of A with eigenvalue λ. Then,

$$\begin{bmatrix} A_1 & A_2 \\ A_2 & A_1 \end{bmatrix} \cdot \begin{bmatrix} v \\ v \end{bmatrix} = \begin{bmatrix} A_1 v + A_2 v \\ A_2 v + A_1 v \end{bmatrix} = \begin{bmatrix} Av \\ Av \end{bmatrix} = \lambda \begin{bmatrix} v \\ v \end{bmatrix}.$$

So, (v, v) is an eigenvector of $A(G_s)$ with eigenvalues λ. Moreover, if u is an eigenvector of A_s with eigenvalue μ then

$$\begin{bmatrix} A_1 & A_2 \\ A_2 & A_1 \end{bmatrix} \cdot \begin{bmatrix} u \\ -u \end{bmatrix} = \begin{bmatrix} A_1 u - A_2 u \\ A_2 u - A_1 u \end{bmatrix} = \begin{bmatrix} A_s u \\ -A_s u \end{bmatrix} = \mu \begin{bmatrix} u \\ -u \end{bmatrix}.$$

Thus, $(u, -u)$ is an eigenvector of $A(G_s)$ with eigenvalue μ. Let $v_1, \ldots, v_n$ be orthonormal eigenvectors of A and $u_1, \ldots, u_n$ be orthonormal eigenvectors of $A(G_s)$. Since $(v_i, v_i) \perp (u_j, -u_j)$, we produced $2n$ orthonormal eigenvector. Thus, these is a complete basis of eigenvectors of $A(G_s)$. ☐

Example 5. Show that the complete bipartite graph $K_{d,d}$ has all non-trivial eigenvalues equal 0.

To construct a Ramanujan family it is enough to prove the following theorem.

Theorem 3. For every d-regular bipartite graph G there is a signing s such that all the eigenvalues of A_s are not greater than $2\sqrt{d-1}$.

We then run our construction as follows:

1. Take $G^{(1)} = K_{d,d}$.
2. Having a bipartite graph $G^{(i)}$ with all the eigenvalues in the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$, choose (using Theorem 3) a signing s such that $A_s(G^{(i)})$ has all the eigenvalues not greater than $2\sqrt{d-1}$. Then by Fact 10 we get that all the eigenvalues of $A(G_s^{(i)})$ are not greater than $2\sqrt{d-1}$. From Fact 8 we know that the spectrum of $G_s^{(i)}$ is symmetric. From Fact 9 we know that $G_s^{(i)}$ is bipartite. Thus, the spectrum of $G_s^{(i)}$ is contained in $[-2\sqrt{d-1}, 2\sqrt{d-1}]$.
3. Take $G^{(i+1)} = G_s^{(i)}$ and iterate the procedure.

Let M be a symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. We define the characteristic polynomial

$$\chi_M(x) = \det(xI - M) = \prod_i (x - \lambda_i).$$

Let A be the adjacency matrix of graph G. For $k \leq m$ and $t \in \{\pm 1\}^k$ define

$$p_t(x) = \frac{1}{2^{m-k}} \sum_{s:\ s_1=t_1,\dots,s_k=t_k} \chi_{A_s}(x).$$

Note that for $t \in \{\pm 1\}^m$ we have $p_t = \chi_{A_t}$. We also define

$$p_\emptyset(x) = \frac{1}{2^m} \sum_{s \in \{\pm 1\}^m} \chi_{A_s}(x).$$

We would like to find $t \in \{\pm 1\}^m$ such that the roots of p_t are not greater than $2\sqrt{d-1}$. We will start slowly by showing that this is true for $p_\emptyset$.

2.2. **Properties of $p_\emptyset$.**

Fact 11. For any graph G we have

$$p_\emptyset(x) = \sum_i x^{n-2i}(-1)^i m_i,$$

where m_i is the number of matchings (subsets of E that touch each vertex at most once) in G of size i. Here $m_0 = 1$.

Proof. Recall that

$$\det(A) = \sum_{\sigma-\text{permutation}} (-1)^{|\sigma|} a_{1,\sigma(1)} a_{2,\sigma(2)} \dots a_{n,\sigma(n)},$$

where $|\sigma|$ is the parity of σ.

Note that

1. Permutations that hit any off-diagonal non-edge do not contribute to the above sum.
2. permutation that hit a_{ij} but not a_{ji} cancel in expectation.
3. We are left with permutations with $n-2i$ entries on the diagonal (contributing x^{n-2i}) and i terms of the form $a_{ij} \cdot a_{ji}$ with $a_{ij} = a_{ji} = 1$ (this corresponds to i element matchings; we have m_i such matchings). These contributions correspond to permutations of parity $(-1)^i$, since each pair $(i, \sigma(i)), (\sigma(i), i)$ contribute one transposition.

□

Fact 12. Let G be a graph. Then the polynomial $\mu_G(x) = \sum_i x^{n-2i}(-1)^i m_i$ is real rooted.

Proof. We consider a more general setting where the edges of a complete graph have weights $W(e) \geq 0$. We first assume $W(e) > 0$ for all $e \in E$. Let M_i be the set of all matchings having i edges. Take

$$m_i(G) = m_i^{(W)}(G) = \sum_{M \in M_i} \prod_{e \in M} W(e).$$

We have

$$m_i(G) = m_i(G \setminus \{v\}) + \sum_{u \sim v} W(u,v) m_{i-1}(G \setminus \{u,v\}).$$

Take

$$\mu_G(x) = \mu_G^{(W)}(x) = \sum_i x^{n-2i}(-1)^i m_i^{(W)}(G).$$

Using the recurrence relations gives

$$\mu_G(x) = x\mu_{G\setminus\{v\}}(x) - \sum_{u\sim v} W(u,v)\mu_{G\setminus\{u,v\}}(x).$$

We show by induction that μ_G has n different real roots $a_1 < a_2 \ldots < a_n$ and for every v the roots $b_1 < \ldots < b_{n-1}$ of $\mu_{G\setminus\{v\}}$ satisfy

$$a_1 < b_1 < a_2 < b_2 < \ldots < b_{n-1} < a_n.$$

The base of the induction is trivial. In this case we take $n = 2$ and then $\mu_G(x) = x^2 - W$, where W is the weight on the only edge of G. For $v \in G$ we have $\mu_{G\setminus\{v\}}(x) = x^2$, so the assertion holds.

To do the induction step fix v and assume that for every u the zeroes of $\mu_{G\setminus\{u,v\}}$ interlace the zeroes of $\mu_{G\setminus\{v\}}$. By Darboux principle the zeroes $c_1 < \ldots < c_{n-2}$ of the sum

$$s(x) = \sum_{u\sim v} W(u,v)\mu_{G\setminus\{u,v\}}(x)$$

also interlaces the zeroes $b_1 < \ldots < b_{n-1}$ of $\mu_{G\setminus\{v\}}$. Consider the sign of the right hand side for $x = b_1, \ldots, b_{n-1}$. We get

$$\mu_G(b_{n-1}) < 0, \quad \mu_G(b_{n-2}) > 0, \qquad \ldots .$$

However, $\lim_{x\to\infty} \mu_G(x) = \infty$, so there is a zero of μ_G to the right of b_{n-1}. Assume $\mu_g(b_1) > 0$, that is $s(b_1) < 0$. Note that s is a monic polynomial of degree $N-2$ and we have determined all the zeroes of s. If follows that N is odd. Since the degree of μ_G is odd, $\lim_{x\to-\infty} \mu_G(x) = -\infty$ and thus there is a zero of μ_G to the left of b_1. The same holds when $\mu_g(b_1) < 0$. Thus, there we have found N zeroes of μ_G: $N-1$ interlace the numbers b_i, plus two additional zeroes. So, the zeroes of μ_G interlace the numbers b_i.

If $W(e)$ are non-negative but not necessarily positive, then by passing to the limit we deduce that

$$a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le b_{n-1} \le a_n$$

and that all the zeroes of μ_G are real. □

Fact 13. Let G be a graph with maximum degree Δ. Then the polynomial $\mu_G(x) = \sum_i x^{n-2i}(-1)^i m_i$ satisfy $\mu_G(y) > 0$ for $y > 2\sqrt{\Delta - 1}$.

Proof. Given a connected graph G we construct a set $\mathcal{E}_G$ (not necessarily unique) of pairs $(G', G' \setminus \{v\})$, where G' is a vertices induced subgraph of G, in the following way:

1. Fix $v_0 \in G$. A pair $(G \setminus \{v_0\}, G \setminus \{v_0, u\})$ is a member of $\mathcal{E}_G$ if $u \sim v_0$. Since G is connected, there is at least one element in $\mathcal{E}_G$ of this type.
2. If $(H, H \setminus \{v\}) \in \mathcal{E}_G$ then $(H \setminus \{u\}, H \setminus \{v, u\}) \in \mathcal{E}_G$ whenever $u \sim v$ and $u \in H \setminus \{v\}$.
3. If $(H, H \setminus \{v\}) \in \mathcal{E}_G$ but v is isolated, then choose any vertex $u \in H \setminus \{v\}$ such that u had some neighbour in G not belonging to $H \setminus \{v\}$ and choose $(H \setminus \{u\}, H \setminus \{v, u\})$ to be a member of $\mathcal{E}_G$.

We prove the following lemma.

Lemma 7. Let G be a connected graph of maximal degree Δ. Then for every $(H, H \setminus \{v\}) \in \mathcal{E}_G$ and $x \ge 2\sqrt{\Delta - 1}$ we have

$$\frac{\mu_H(x)}{\mu_{H\setminus\{v\}}(x)} \ge \sqrt{\Delta - 1} \qquad \text{and} \qquad \mu_H(x) > 0.$$

Proof. We prove this fact by induction over $\mathcal{E}_G$. It is easy to verify the base of induction, i.e., graphs H with $|H| = 1$. For the induction step we use the identity

$$\mu_H(x) = x\mu_{H\setminus\{v\}}(x) - \sum_{u\sim v,\ u\in H\setminus\{v\}} \mu_{H\setminus\{u,v\}}(x).$$

This is the same as

$$\mu_H(x) = x\mu_{H\setminus\{v\}}(x) - \sum_{u\sim v:\ (H\setminus\{v\}, H\setminus\{u,v\})\in\mathcal{E}_G} \mu_{H\setminus\{u,v\}}(x).$$

Now, the cardinality of the set $u \sim v :\ (H \setminus \{v\}, H \setminus \{u, v\}) \in \mathcal{E}_G$ is at most $\Delta - 1$. This is because according to steps 1. and 2. above, the vertex v has been removed because either v was a neighbour of a removed point (and the the degree of v decreased by 1) or v has been removed as a result of step 3, i.e., there was some other vertex w that has been remover even earlier with $w \sim v$ and thus the degree of v is also at most $\Delta - 1$. We get, by induction hypothesis,

$$\begin{aligned}\frac{\mu_H(x)}{\mu_{H\setminus\{v\}}(x)} &= x - \sum_{u\sim v:\ (H\setminus\{v\}, H\setminus\{u,v\})\in\mathcal{E}_G} \frac{\mu_{H\setminus\{u,v\}}(x)}{\mu_{H\setminus\{v\}}(x)} \\ &\geq x - (\Delta - 1)\cdot\frac{1}{\sqrt{\Delta-1}} > 2\sqrt{\Delta-1} - (\Delta-1)\cdot\frac{1}{\sqrt{\Delta-1}} = \sqrt{\Delta-1}.\end{aligned}$$

Clearly $\mu_H(x) > 0$. □

We can assume that G is connected. If not, observe that if G has two components S, T then the matrix A_s has two blocks B_{s_1}, C_{s_2}, where $s = (s_1, s_2)$ and

$$\sum_s \chi_{A_s}(x) = \sum_{s_1,s_2} \chi_{B_{s_1}}(x)\chi_{C_{s_2}}(x) = \sum_{s_1} \chi_{B_{s_1}}(x) \sum_{s_2} \chi_{C_{s_2}}(x),$$

so we can consider the roots of $\sum_{s_1} \chi_{B_{s_1}}(x)$ and $\sum_{s_2} \chi_{C_{s_2}}(x)$ separately.

If G is connected and $\Delta = 1$ then the assertion holds trivially. Assume that $\Delta \geq 2$ and $x > 2\sqrt{\Delta-1}$. Then

$$\begin{aligned}\frac{\mu_G(x)}{\mu_{G\setminus\{v_0\}}(x)} &= x - \sum_{u\sim v_0:\ (H\setminus\{v_0\}, G\setminus\{u,v_0\})\in\mathcal{E}_G} \frac{\mu_{G\setminus\{u,v_0\}}(x)}{\mu_{G\setminus\{v_0\}}(x)} \\ &\geq x - \Delta\cdot\frac{1}{\sqrt{\Delta-1}} > 2\sqrt{\Delta-1} - \frac{\Delta}{\sqrt{\Delta-1}} = \frac{\Delta-2}{\sqrt{\Delta-1}} \geq 0,\end{aligned}$$

where we have used the Lemma. Since $\mu_{G\setminus\{v_0\}}(x) > 0$, it follows that $\mu_G(x) > 0$. □

2.3. **From $p_\emptyset$ to p_t.** We know that $p_\emptyset$ has all its roots not greater than $2\sqrt{d-1}$. Suppose we know such a thing for some p_t with $t \in \{pm1\}^k$, $k \leq m$. Now,

$$p_t(x) = \frac{1}{2}p_{t,+}(x) + \frac{1}{2}p_{t,-}(x).$$

Let $M(p)$ be the maximal root of a polynomial p. We are going to use the following fact.

Lemma 8. Let p, q be monic real-rooted polynomials of the same degree. Suppose that for every $t \in [0, 1]$ the polynomial $tp + (1 - t)q$ is also real rooted. Then for every $\lambda \in [0, 1]$ the root $M((1 - \lambda)p + \lambda q)$ lies between $M(p)$ and $M(q)$.

Proof. Assume without loss of generality that $M(p) \leq M(q)$. Our goal is to show that $M(p) \leq M((1-\lambda)p + \lambda q) \leq M(q)$. Since $p > 0$ to the right of $M(p)$ and $q > 0$ to the right of $M(q)$, we get that $(1-\lambda)p + \lambda q > 0$ to the right of $M(q)$, so $M((1-\lambda)p + \lambda q) \leq M(q)$. Suppose our assertion fails, that is, $(1-\lambda)p + \lambda q$ has no root in $[M(p), M(q)]$. This means that $(1-\lambda)p + \lambda q > 0$ on $[M(p), M(q)]$. In particular $q(M(p)) > 0$. As a consequence q has at least 2 zeroes in $[M(p), M(q)]$ (counting multiplicities). Let $F(q)$ be the smallest of those zeroes.

Note that for small t the polynomial $p_t = (1-t)p + tq$ has no zeroes to the right of $M(p)$. To see this take $I_0 = [M(p), F(q)]$, $I_1 = [F(q), M(q)]$ and $I_2 = (M(q), \infty)$. On I_2 the combination p_t is strictly positive. Since $q(M(p)) > 0$ we see that $p_t > 0$ on $[M(p), F(q))$. The same holds true at the point $F(q)$, due to the positivity of p. Thus $p_t > 0$ on I_0. Now, on I_1 we have $p \geq \varepsilon$ for some $\varepsilon > 0$. Thus, for sufficiently small t we have $p_t > 0$ there.

Changing t from $t \approx 0$ to $t = 1$ we see that the number of roots of p_t in the interval $(M(p), \infty)$ increases from 0 to at least 2. Let us consider the open set $U = \{\mathrm{Re} z > M(p)\} \subset \mathbb{C}$. Since $p_t(z)$ are real rooted, the number of roots of $p_t(z)$ in U has a discontinuity (jumps from 0 to more than 0). This contradicts Hurwitz theorem. □

From this fact it follows that if $\lambda p_{t,+} + (1-\lambda)p_{t,-}$ is real rooted for every $t \in [0,1]$, then either $M(p_{t,+}) \leq M(\frac{1}{2}p_{t,+} + \frac{1}{2}p_{t,-}) = M(p_t)$, or $M(p_{t,-}) \leq M(p_t)$. So, there is a sign s_{k+1} such that $M(p_{t,s_{k+1}}) \leq 2\sqrt{d-1}$. Iterating gives $s \in \{\pm 1\}^m$ with $M(p_s) \leq 2\sqrt{d-1}$.

To finish the proof it suffices to show the following lemma.

Lemma 9. For any $\lambda \in [0,1]$ and $t \in \{\pm 1\}^k$ the polynomial $\lambda p_{t,+} + (1-\lambda)p_{t,-}$ is real rooted.

Note that
$$p_s(x-d) = \chi_{A_s}(x-d) = \det(xI - (dI + A_s))$$
and
$$dI + A_s = \sum_{s_{(i,j)}=-1} (\delta_i - \delta_j)(\delta_i - \delta_j)^T + \sum_{s_{(i,j)}=-1} (\delta_i + \delta_j)(\delta_i + \delta_j)^T.$$
Thus,
$$\sum_{s \in \{\pm 1\}^m} \prod_{s_i=+1} \theta_i \prod_{s_i=-1} (1-\theta_i) p_s(x-d) = \mathbb{E} \det \left(xI - \sum_{e \in E} u_e u_e^T \right),$$
where for $e = e_k = (i,j)$ we have
$$u_e = \begin{cases} \delta_i + \delta_j & \text{w.p. } \theta_k \\ \delta_i - \delta_j & \text{w.p. } 1-\theta_k \end{cases}.$$

Note that this covers the case of $\lambda p_{t,+} + (1-\lambda)p_{t,-}$. It suffices to take $\theta_k \in \{0,1\}$ for edges corresponding to the coordinates of t, $\theta_k = \lambda$ for the edge corresponding to $(+,-)$ and $\theta_k = 1/2$ for other edges.

We shell prove the following theorem.

Theorem 4. For any independent random vectors $u_1, \ldots, u_m$ the expected characteristic polynomial
$$\det \left(xI - \sum_{i=1}^m u_i u_i^T \right)$$
is real rooted.

3. Real stable polynomials

We would like to prove the following fact.

Proposition 1. Let $A_1, \ldots, A_m$ be $d \times d$ rank 1 matrices. Take $A = A_1 + \ldots + A_m$. Then the characteristic polynomial $p_A(z) = \det(zI - A)$ can be written in the form

$$p_A(z) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det\left(zI + \sum_{i=1}^{m} z_i A_i\right)\Big|_{z_1=\ldots=z_m=0}.$$

Let us also define

$$\mu[A_1, \ldots, A_m](z) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det\left(zI + \sum_{i=1}^{m} z_i A_i\right)\Big|_{z_1=\ldots=z_m=0}.$$

We need the following lemma.

Lemma 10. Let A, B be $d \times d$ matrices and assume that A is rank 1. Then the function $t \mapsto \det(B + tA)$ is affine linear.

Proof. **Sylvester identity.** Suppose X is a $m \times n$ matrix and Y is a $n \times m$ matrix. Then

$$\det(I_m + XY) = \det(I_n + YX).$$

To prove this, let us first observe that we have the identity

$$\begin{pmatrix} I_n & -Y \\ X & I_m \end{pmatrix} \cdot \begin{pmatrix} I_n & Y \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ X & XY + I_m \end{pmatrix}.$$

We have

$$\det\left(\begin{pmatrix} I_n & -Y \\ X & I_m \end{pmatrix} \cdot \begin{pmatrix} I_n & Y \\ 0 & I_m \end{pmatrix}\right) = \det\begin{pmatrix} I_n & 0 \\ X & XY + I_m \end{pmatrix} = \det(XY + I_m)$$

Since $\det(AB) = \det(BA)$, the left hand side is the same as

$$\det\left(\begin{pmatrix} I_n & Y \\ 0 & I_m \end{pmatrix} \cdot \begin{pmatrix} I_n & -Y \\ X & I_m \end{pmatrix}\right) = \det\begin{pmatrix} I_n + YX & 0 \\ X & I_m \end{pmatrix} = \det(I_n + YX).$$

We have proved the Sylvester identity.

To prove the lemma it suffices to consider the case when B is invertible. Then

$$\det(B + tA) = \det(B)\det(I + tB^{-1}A).$$

Note that $B^{-1}A$ is rank 1, so it suffices to consider the case $B = I$. Any rank 1 matrix can be written in the form $A = u \cdot v^T$. Then

$$\det(I + tA) = \det(I_d + tuv^\star) = \det(I_1 + tv^\star u) = 1 + tv^\star u.$$

∎

Proof of Proposition 1. From the above lemma we have

$$p(z_1, \ldots, z_m) = \det(zI + \sum_{i=1}^{m} z_i A_i) = \sum_{i_1 < i_2 \ldots < i_k} a_{i_1, \ldots i_k} z_{i_1} \ldots z_{i_k},$$

Now, any affine multi-linear polynomial $p : \mathbb{R}^m \to \mathbb{R}$ satisfies

$$p(t_1, \ldots, t_m) = \prod_{i=1}^{m} \left(1 + t_i \frac{\partial}{\partial z_i}\right) p(z_1, \ldots, z_m)\Big|_{z_1=\ldots=z_m=0}.$$

It suffices to use this fact for $t_1 = \ldots = t_m = -1$. □

We now need the following crucial definition.

Definition 5. We say that a polynomial $p\mathbb{C}^m \to \mathbb{C}$ is *stable* if it has no zeroes in the set $\{\operatorname{Im} z_1 > 0, \ldots, \operatorname{Im} z_m > 0\}$. If additionally p has real coefficients then it is called *real stable*.

Remark 3. Note that $p : \mathbb{C} \to \mathbb{C}$ is real stable if and only if it has real coefficients and is real rooted.

Fact 14. Let $A_1, \ldots, A_m$ be positive semi-definite Hermitian matrices. Then

$$p(z, z_1, \ldots, z_m) = \det(zI + \sum_{i=1}^{m} z_i A_i)$$

is real stable.

Proof. For real inputs the values of this polynomial are real (the determinant of a Hermitian matrix is real). Suppose p has a root $(z, z_1, \ldots, z_m)$. This means that the corresponding matrix is singular and therefore there is a vector v such that $(zI + \sum_{i=1}^{m} z_i A_i)v = 0$. Suppose $z, z_1, \ldots, z_m$ have positive imaginary parts. Then

$$0 = \operatorname{Im}\left(\left\langle (zI + \sum_{i=1}^{m} z_i A_i)v, v \right\rangle\right) = (\operatorname{Im} z)|v|^2 + \sum_{i=1}^{m} (\operatorname{Im} z_i) \langle A_i v, v\rangle > 0,$$

since the first term is strictly positive and other terms are non-negative (because A_i are positive semi-definite). This is a contradiction. □

Fact 15. Suppose that $p(z_1, \ldots, z_m)$ is real stable. Then

(a) for any $t \in \mathbb{R}$ the polynomial $q(z_1, \ldots, z_{m-1}) = p(z_1, \ldots, z_{m-1}, t)$ is real stable (or identically zero),
(b) if $t \in \mathbb{R}$ then $(1 + t\frac{\partial}{\partial z_m})p(z_1, \ldots, z_m)$ is real stable,
(c) if σ is a permutation then $p(z_{\sigma(1)}, \ldots, z_{\sigma(n)})$ is real stable.

Proof. (a) Let $\Omega = \{\operatorname{Im} z_1 > 0, \ldots, \operatorname{Im} z_{m-1} > 0\}$. For any $\varepsilon > 0$ the function $p_\varepsilon(z_1, \ldots, z_{m-1}) = p(z_1, \ldots, z_{m-1}, t + i\varepsilon)$ has no roots on Ω. Since $p_\varepsilon \to p_0$ when $\varepsilon \to 0$, the function p_0 cannot have roots in Ω by Hurwitz theorem (or p_0 is identically zero).

(b) Fix $z_1, \ldots, z_{m-1}$ and define $q(z) = p(z_1, \ldots, z_{m-1}, z)$. If $\operatorname{Im} z_i > 0$ for $i = 1, \ldots, m-1$ then q is stable. Our goal is to prove that $q + tq'$ is also stable. Let $q(z) = c\prod_{i=1}^{d}(z - w_i)$. Then

$$q(z) + tq'(z) = q(z)\left(1 + \sum_{i=1}^{d} \frac{t}{z - w_i}\right).$$

Note that $\operatorname{Im} w_i \leq 0$. If $\operatorname{Im} z > 0$ then

$$\operatorname{Im}\left(\frac{1}{z - w_i}\right) = \operatorname{Im}\left(\frac{\bar{z} - \bar{w}_i}{|z - w_i|^2}\right) = \frac{1}{|z - w_i|^2}\operatorname{Im}(\bar{z} - \bar{w}_i) < 0.$$

In particular, the number $1+\sum_{i=1}^{d} \frac{t}{z-w_i}$ has non-zero imaginary part and therefore is non-zero. Thus, z is not a root of $q(z) + tq'(z)$.

Part (c) is trivial. □

Corollary 1. *From the Fact 14 and Fact 15 we see that for any positive semi-definite Hermitian matrices $A_1, \ldots, A_m$ the polynomial $\mu[A_1, \ldots, A_m](z)$ is real stable and thus real rooted.*

To finish the construction of Ramanuja family it suffices to show the following proposition.

Proposition 2. Let $A_1, \dots, A_m$ be independent rank one matrices. Then

$$\mathbb{E}\mu[A_1, \dots, A_m] = \mu[\mathbb{E}A_1, \dots, \mathbb{E}A_m].$$

We need a lemma.

Lemma 11. Let A be invertible $n \times n$ matrix.

(a) If $u, v \in \mathbb{C}^n$ then

$$\det(A + uv^\star) = \det(A)(1 + v^\star A^{-1} u).$$

(b) If B is a $n \times n$ matrix then

$$\partial_t \det(A + tB)\big|_{t=0} = \det(A)\operatorname{Tr}(A^{-1}B).$$

Proof. For (a) note that due to Sylvester identity we have

$$\det(A + uv^\star) = \det(A)\det(I_n + A^{-1}uv^\star) = \det(A)\det(I_1 + v^\star A^{-1}u) = \det(A)(1 + v^\star A^{-1}u).$$

For (b) we again write

$$\det(A + tB) = \det(A)\det(I + tA^{-1}B),$$

so it suffices to show that $\partial_t \det(I + tA)|_{t=0} = \operatorname{Tr}(A)$, which is obvious. $\square$

Lemma 12. For every deterministic square $n \times n$ matrix A and every random vector $v \in \mathbb{C}^n$ we have

$$\mathbb{E}\det(A - vv^\star) = (1 - \partial_t)\det(A + t\mathbb{E}vv^\star)\Big|_{t=0}.$$

Proof. From Lemma 12(a) we have

$$\begin{aligned}\mathbb{E}\det(A - vv^\star) &= \mathbb{E}\det(A)(1 - v^\star A^{-1}v) = \mathbb{E}\det(A)(1 - \operatorname{Tr} A^{-1}vv^\star) \\ &= \det(A) - \det(A)\mathbb{E}\operatorname{Tr}(A^{-1}vv^\star) = \det(A) - \det(A)\operatorname{Tr}(A^{-1}\mathbb{E}vv^\star).\end{aligned}$$

Also,

$$(1-\partial_t)\det(A + t\mathbb{E}vv^\star)\Big|_{t=0} = \det(A) - \partial_t \det(A + t\mathbb{E}vv^\star)\Big|_{t=0} = \det(A) - \det(A)\operatorname{Tr}(A^{-1}\mathbb{E}vv^\star),$$

by Lemma 12(b). $\square$

We are ready to give a proof of Proposition 2.

Proof of Proposition 2. We show by induction on k that for every matrix M we have

$$\mathbb{E}\det\left(M - \sum_{i=1}^k v_i v_i^\star\right) = \left(\prod_{i=1}^k (1 - \partial_{z_i})\right)\det\left(M + \sum_{i=1}^m z_i A_i\right)\Big|_{z_1 = \ldots = z_k = 0}.$$

To pass from $k-1$ to k we observe that

$$\begin{aligned}
\mathbb{E}\det\left(M-\sum_{i=1}^{k} v_i v_i^\star\right) &= \mathbb{E}_{v_1,\dots,v_{k-1}}\mathbb{E}_{v_k}\det\left(M-\sum_{i=1}^{k-1} v_i v_i^\star - v_k v_k^\star\right)\\
&= \mathbb{E}_{v_1,\dots,v_{k-1}}(1-\partial_{z_k})\det\left(M-\sum_{i=1}^{k-1} v_i v_i^\star + z_k\mathbb{E}v_k v_k^\star\right)\Big|_{z_k=0}\\
&= (1-\partial_{z_k})\mathbb{E}_{v_1,\dots,v_{k-1}}\det\left(M+z_k\mathbb{E}v_k v_k^\star-\sum_{i=1}^{k-1} v_i v_i^\star\right)\Big|_{z_k=0}\\
&= (1-\partial_{z_k})\left(\prod_{i=1}^{k-1}(1-\partial_{z_i})\right)\det\left(M+z_k\mathbb{E}v_k v_k^\star+\sum_{i=1}^{k-1} z_i\mathbb{E}v_i v_i^\star\right)\Big|_{z_k=0}\\
&= \left(\prod_{i=1}^{k}(1-\partial_{z_i})\right)\det\left(M+\sum_{i=1}^{m} z_i A_i\right)\Big|_{z_1=\ldots=z_k=0}.
\end{aligned}$$

□

4. Towards Kadison-Singer conjecture

4.1. M-S-S Theorem.

Our first goal is to prove the following theorem.

Theorem 5 (Marcus, Spielman, Srivastava)**.** Let $m, d \geq 1$ and let $v_1, \dots, v_d \in \mathbb{C}^d$ be independent random vector (finitely supported). Assume that $\mathbb{E}\left(\sum_{i=1}^m v_i v_i^\star\right) = I$ and $\mathbb{E}\|v_i\|^2 \leq \varepsilon$. Then $\|\sum_{i=1}^m v_i v_i^\star\|_{op} \leq (1+\sqrt{\varepsilon})^2$ with positive probability.

Equivalently,

Theorem 6 (Marcus, Spielman, Srivastava)**.** Let $m, d \geq 1$ and let $A_1, \dots, A_m$ be independent Hermitian positive semi-definite rank 1 random matrices (finitely supported). Take $A = A_1 + \ldots + A_m$ and assume that $\mathbb{E}A = I$ and $\mathbb{E}\operatorname{Tr} A_i \leq \varepsilon$. Then $\|A\|_{op} \leq (1+\sqrt{\varepsilon})^2$ with positive probability.

The above theorem is an easy consequence of the following two propositions.

Proposition 3. Let $A_1, \dots, A_m$ be independent rank 1 Hermitian matrices. Take $A = A_1 + \ldots + A_m$. Then $M(p_A) \leq M(\mathbb{E}p_A)$ with positive probability.

Proposition 4. Let $A_1, \dots, A_m$ be independent rank 1 positive semi-definite Hermitian matrices. Let $A = A_1 + \ldots + A_m$ and assume that $\mathbb{E}A = I$ and $\mathbb{E}\operatorname{Tr} A_i \leq \varepsilon$. Then $M(\mathbb{E}p_A) \leq (1+\sqrt{\varepsilon})^2$.

Proof of Proposition 3. We have

$$\begin{aligned}
\mu[A_1,\dots,A_{j-1},\mathbb{E}A_j,\dots,\mathbb{E}A_m] &= \mathbb{E}_{A_1,\dots,A_m}\mu[A_1,\dots,A_m]\\
&= \mathbb{E}_{A_j}\mu[A_1,\dots,A_{j-1},A_j,\mathbb{E}A_{j+1},\dots,\mathbb{E}A_m].
\end{aligned}$$

Here $A_1, \dots, A_{j-1}$ are deterministic, but are treated as random with Dirac delta distributions. Thus, the polynomial

$$p = \mu[A_1,\dots,A_{j-1},\mathbb{E}A_j,\dots,\mathbb{E}A_m]$$

is a convex combination of polynomials

$$p(\omega) = \mu[A_1,\dots,A_{j-1},A_j(\omega),\mathbb{E}A_{j+1},\dots,\mathbb{E}A_m].$$

Thus, $p = \sum_\omega \lambda(\omega)p(\omega)$. All such combinations $\sum_\omega \nu(\omega)p(\omega)$ are real stable, because they are of the form

$$\mu[A_1, \dots, A_{j-1}, \mathbb{E}_\nu A_j, \mathbb{E}A_{j+1}, \dots, \mathbb{E}A_m].$$

Iterating The Lemma 8 it is easy to show that

$$M(\sum_\omega \lambda(\omega)p(\omega)) \in \operatorname{conv}((M(p(\omega)))_\omega).$$

Thus,

$$M(\mu[A_1, \dots, A_j, \mathbb{E}A_{j+1}, \dots, \mathbb{E}A_m]) \le M(\mu[A_1, \dots, A_{j-1}, \mathbb{E}A_j, \dots, \mathbb{E}A_m])$$

with positive probability. Iterating gives

$$M(\mu[A_1, \dots, A_m]) \le M(\mu[\mathbb{E}A_1, \dots, \mathbb{E}A_m])$$

with positive probability. This is $M(p_A) \le M(\mathbb{E}p_A)$ with positive probability. □

4.2. **Proof of Proposition 4.**

Definition 6. Let $p(z_1, \dots, z_m)$ be a real stable polynomial. We say that a point $x = (x_1, \dots, x_m) \in \mathbb{R}^m$ lies above the roots of p if p has no roots in the octant $\{(y_1, \dots, y_m) \in \mathbb{R}^m \mid y_i \ge x_i, i = 1, \dots, m\}$.

We will prove the following proposition.

Proposition 5. Let $A_1, \dots, A_m$ be Hermitian deterministic positive semi-definite $d \times d$ matrices and let $A_1 + \ldots + A_m = I_d$. Assume that $\operatorname{Tr} A_i \le \varepsilon$ for every $i = 1, \dots, m$ and some $\varepsilon > 0$. Define

$$p(z_1, \dots, z_m) = \det\left(\sum_{i=1}^m z_i A_i\right).$$

Then $((1+\sqrt{\varepsilon})^2, \dots, (1+\sqrt{\varepsilon})^2)$ lies above the roots of $\left(\prod_{i=1}^m (1-\partial_{z_i})\right) p$.

We first show that Proposition 5 implies Proposition 4.

Proposition 5 implies Proposition 4. We have

$$\begin{aligned}
\mathbb{E}p_A(z) &= \left(\prod_{i=1}^m (1-\partial_{z_i})\right) \det\left(zI_d + \sum_{i=1}^m z_i \mathbb{E}A_i\right)\Big|_{z_1=\ldots=z_m=0} \\
&= \left(\prod_{i=1}^m (1-\partial_{z_i})\right) \det\left(\sum_{i=1}^m (z+z_i) \mathbb{E}A_i\right)\Big|_{z_1=\ldots=z_m=0} \\
&= \left(\prod_{i=1}^m (1-\partial_{z_i})\right) \det\left(\sum_{i=1}^m z_i \mathbb{E}A_i\right)\Big|_{z_1=\ldots=z_m=z}.
\end{aligned}$$

Let $\tilde{A}_i = \mathbb{E}A_i$. Note that $\tilde{A}_i$ satisfy the assumptions of Proposition 5. Thus, if real z satisfy $z \ge (1+\sqrt{\varepsilon})$, then $\mathbb{E}p_A(z) \ne 0$ because $((1+\sqrt{\varepsilon})^2, \dots, (1+\sqrt{\varepsilon})^2)$ lies above the roots of $\left(\prod_{i=1}^m (1-\partial_{z_i})\right) \det\left(\sum_{i=1}^m z_i \tilde{A}_i\right)$. Thus $M(\mathbb{E}p_A) \le (1+\sqrt{\varepsilon})^2$.

□

Lemma 13. Let $p(z)$ be a real stable polynomial of one variable. Define $\Phi_p = p'/p = (\log p)'$. Then

$$(-1)^k \Phi_p^{(k)}(x) > 0, \qquad \text{for } \ x > M(p).$$

Proof. If $y_1 \leq \ldots \leq y_d$ are the roots of p then $\Phi_p(x) = \sum_{i=1}^d \frac{1}{x-y_i}$. Thus, for $x > y_d$ we have

$$(-1)^k \Phi_p^{(k)}(x) = \sum_{i=1}^d \frac{1}{(x-y_i)^k} > 0.$$

□

Lemma 14. Let $p(z_1, \ldots, z_m)$ be a real stable polynomial and take $1 \leq i, j \leq m$. Define $\Phi_p^{(i)} = \frac{\partial}{\partial z_i} \log p$. Then

$$(-1)^k \frac{\partial^k}{\partial x_j^k} \Phi_p^{(i)}(x) \geq 0, \qquad k \geq 0,$$

for $x = (x_1, \ldots, x_m)$ above the roots of p.

Proof. The cases $i = j$ or $k = 0$ follow from Lemma 13. We can therefore take $i \neq j$ and $k \geq 1$. Without loss of generality we can assume $i = 1$ and $j = m = 2$. Let $x = (x_1, x_2)$ be above the roots of $p(z_1, z_2)$. We have

$$(-1)^k \frac{\partial^k}{\partial x_j^k} \Phi_p^{(i)}(x) = (-1)^k \frac{\partial^k}{\partial x_j^k} \frac{\partial}{\partial x_1} \log p = \frac{\partial}{\partial x_1} \left[(-1)^k \frac{\partial^k}{\partial x_j^k} \log p \right].$$

It is therefore enough to show that $(-1)^k \frac{\partial^k}{\partial x_j^k} \log p$ is non-decreasing in $x_1 \in \mathbb{R}$.

Fix x_1 and consider $p_{x_1}(x_2) = p(x_1, x_2)$, which has roots $y_1(x_1), \ldots, y_d(x_1)$. Thus,

$$p(x_1, x_2) = c(x_1) \prod_{i=1}^d (x_2 - y_i(x_1)).$$

For a *generic* x_1 the number d is constant and the functions $y_1(x_1), \ldots, y_d(x_1)$ are smooth in a neighbourhood of x_1. Then

$$(-1)^k \frac{\partial^k}{\partial x_j^k} \log p = -(k-1)! \sum_{i=1}^d \frac{1}{(x_2 - y_i(x_1))^k}.$$

Note that $x_2 > y_i(x_1)$ for any i, since (x_1, x_2) lies above the roots of p. It is therefore enough to show that $x_1 \mapsto y_i(x_1)$ is non-increasing.

Suppose by contradiction that there is a generic x_0 and i such that $y_i'(x_0) = \alpha > 0$. Since $(x_0, y_i(x_0))$ is a root of p, for small $\varepsilon > 0$ the is a root of p of the form $z_\varepsilon = (x_0 + \varepsilon i, y_i(x_0 + \varepsilon i))$, where the function $y_i(x_0 + \varepsilon i)$ is smooth. We have $y_i(x_0 + \varepsilon i) \sim y_i(x_0) + \varepsilon \alpha i$. This implies that $\operatorname{Im}(x_0 + \varepsilon i) > 0$ and $\operatorname{Im}(y_i(x_0 + \varepsilon i)) > 0$, for $\varepsilon > 0$. Thus, the root z_ε lies in $\{\operatorname{Im} z_1 > 0, \operatorname{Im} z_2 > 0\}$, contradicting the real stability of p. □

Lemma 15. Let $q(z_1, \ldots, z_m)$ be a real stable polynomial and let $x = (x_1, \ldots, x_m)$ be above the roots of q. Assume that for some $1 \leq j \leq m$ we have $\Phi_q^j(x) < 1$. Let $y = (y_1, \ldots, y_m)$ be such that $y_k \geq x_k$ for any $1 \leq k \leq m$. Then y is above the roots of $q - \partial_j q$.

Proof. Clearly, y is above the roots of q. By monotonicity of Φ_q^j above the roots of q (Lemma 14) we get $\Phi_q^j(y) \leq \Phi_q^j(x) < 1$. Thus

$$q(y) - \partial_j q(y) = q(y) \left(1 - \Phi_q^j(y)\right) \neq 0.$$

□

Lemma 16. Let $q(z_1, \dots, z_m)$ be a real stable polynomial and let $x = (x_1, \dots, x_m)$ be above the roots of q. Assume that for some $1 \le j \le m$ and $\delta > 0$ we have $\Phi_q^j(x) + \frac{1}{\delta} \le 1$. Then $x + \delta e_j$ is above the roots of $q - \partial_j q$ and

$$\Phi^i_{q-\partial_j q}(x + \delta e_j) \le \Phi^i_q(x), \qquad 1 \le i \le m. \tag{1}$$

Proof. Since $\Phi_q^j < 1$, we can apply the above lemma with $y = x + \delta e_j$ and deduce that $x + \delta e_j$ is above the roots of $q - \partial_j q$.

For the second part observe that $q - \partial_j q = q(1 - \Phi_q^j)$ and thus

$$\log(|q - \partial_j|) = \log(|q|) + \log(|1 - \Phi_q^j|).$$

This implies

$$\Phi^i_{q-\partial_j} = \Phi^i_q - \frac{\partial_i \Phi_q^j}{1 - \Phi_q^j}.$$

Therefore (1) is equivalent with

$$-\frac{\partial_i \Phi_q^j(x + \delta e_j)}{1 - \Phi_q^j(x + \delta e_j)} \le \Phi_q^i(x) - \Phi_q^i(x + \delta e_j).$$

Since $t \mapsto \Phi_q^i(x + t e_j)$ is convex (Lemma 14) we get

$$\Phi_q^i(x) - \Phi_q^i(x + \delta e_j) \ge -\delta \partial_j \Phi_q^i(x + \delta e_j) = -\delta \partial_i \Phi_q^j(x + \delta e_j).$$

By Lemma 14 we have $-\delta \partial_i \Phi_q^j(x + \delta e_j) > 0$. Thus it is enough to show

$$\frac{1}{1 - \Phi_q^j(x + \delta e_j)} \le \delta.$$

But this is evident since $\Phi_q^j(x) \le 1 - 1/\delta$ implies

$$\frac{1}{1 - \Phi_q^j(x + \delta e_j)} \le \frac{1}{1 - \Phi_q^j(x)} \le \delta.$$

□

The above lemma is strong enough to apply it inductively and obtain the following corollary.

Corollary 2. *Let $q(z_1, \dots, z_m)$ be a real stable polynomial and let $x = (x_1, \dots, x_m)$ be above the roots of q. Assume that for some $\delta > 0$ and every $1 \le j \le m$ we have $\Phi_q^j(x) + \frac{1}{\delta} \le 1$. Then $x + (\delta, \dots, \delta)$ is above the roots of $\prod_{i=1}^m (1 - \partial_i) q$.*

Proof of Theorem 5. Note that for any $t > 0$ the point $(t, \dots, t)$ is above the roots of p. Indeed, if $x_1, \dots, x_m \ge t$ then $\sum_{i=1}^m x_i A_i$ is non-singular since

$$\sum_{i=1}^m x_i A_i - tI = \sum_{i=1}^m (x_i - t) A_i$$

is positive semi-definite.

Now, for $x_1, \dots, x_m > 0$ we have (using Lemma 11)

$$\begin{aligned}\Phi_p^j(x_1, \dots, x_m) &= \frac{\partial}{\partial y} \log\det\left(\sum_{i=1}^m x_i A_i + y A_j\right)\Big|_{y=0} \\ &= \frac{\det\left(\sum_{i=1}^m A_i + yA_j\right) \operatorname{Tr}\left(A_j \left(\sum_{i=1}^m x_i A_i\right)^{-1}\right)}{\det\left(\sum_{i=1}^m A_i + yA_j\right)} \\ &= \operatorname{Tr}\left(A_j \left(\sum_{i=1}^m x_i A_i\right)^{-1}\right).\end{aligned}$$

Thus,

$$\Phi_p^j(t, \dots, t) = \operatorname{Tr}\left(A_j \left(\sum_{i=1}^m t A_i\right)^{-1}\right) = \frac{1}{t}\operatorname{Tr}(A_j I) = \frac{\operatorname{Tr} A_j}{t} \le \frac{\varepsilon}{t}.$$

To apply Corollary 2 we want to find δ such that $\frac{\varepsilon}{t} + \frac{1}{\delta} \le 1$. Then $(t+\delta, \dots, t+\delta)$ will lie above the roots of $\prod_{i=1}^m (1 - \partial_i)p$. Thus, we are to optimize $t + \delta$ under the constrain $\frac{\varepsilon}{t} + \frac{1}{\delta} \le 1$. Take $\frac{\varepsilon}{t} + \frac{1}{\delta} = 1$. Then

$$t + \delta = t + \frac{t}{t-\varepsilon} = (t - \varepsilon) + \frac{\varepsilon}{t-\varepsilon} + \varepsilon + 1 \ge 2\sqrt{\varepsilon} + \varepsilon + 1 = (1 + \sqrt{\varepsilon})^2,$$

with equality for $t = \varepsilon + \sqrt{\varepsilon}$. □

4.3. **Weaver-type bounds.** In this section we are going to prove the following theorem.

Theorem 7 (Generalized Weaver bound)**.** Let $m, r, d \ge 1$ be integers and let $w_1, \dots, w_m$ in $\mathbb{C}^d$ be such that $|w_i|^2 \le A$. Suppose that for any unit vector u in $\mathbb{C}^d$ we have

$$\sum_{i=1}^m |\langle u, w_i\rangle|^2 = B.$$

Then there is a partition of $\{1, \dots, m\}$ into sets $S_1, \dots, S_r$ such that for any unit vector in $\mathbb{C}^d$ we have

$$\sum_{i \in S_j} |\langle u, w_i\rangle|^2 \le \left(\sqrt{A} + \sqrt{B/r}\right)^2.$$

Taking $A = 1$ and $B = 18$ we deduce the following theorem.

Theorem 8 (KS_r Weaver conjecture)**.** For any $r \ge 2$ and any vectors $w_1, \dots, w_m$ such that $|w_i| \le 1$ and

$$\sum_{i=1}^m |\langle u, w_i\rangle|^2 = 18, \qquad |u| = 1, u \in \mathbb{C}^d$$

there is a partition $\{S_1, \dots, S_r\}$ of $\{1, \dots, m\}$ with

$$\sum_{i \in S_j} |\langle u, w_i\rangle|^2 \le 16.$$

The following fact gives several equivalent conditions for the so-called B-tight frames.

Fact 16. Let $w_1, \dots, w_m$ be vectors in $\mathbb{C}^d$ and let $B > 0$. Then the following conditions are equivalent,

(a) $\sum_{i=1}^{m} |\langle u, w_i\rangle|^2 = B$ for any $u \in \mathbb{C}^d$ with $|u| = 1$,
(b) $\sum_{i=1}^{m} |\langle u, w_i\rangle|^2 = B|u|^2$ for any $u \in \mathbb{C}^d$,
(c) $\sum_{i=1}^{m} \langle u, w_i\rangle \langle w_i, v\rangle = B\langle u, v\rangle$ for any $u, v \in \mathbb{C}^d$
(d) $\sum_{i=1}^{m} \langle u, w_i\rangle w_i = Bu$ for any $u \in \mathbb{C}^d$,
(e) $\sum_{i=1}^{m} w_i w_i^\star = BI_d$.

Proof. Clearly (a) and (b) are equivalent. The equivalence between (b) and (c) follows from polarization principle (note that $\langle w_i, v\rangle = \overline{\langle v, w_i\rangle}$). The equivalence of (c) and (d) is clear. The equivalence between (d) and (e) follows from the fact that $w_i^\star u = \langle u, w_i\rangle$. □

Remark 4. From (d) it follows that a B-tight frame is a basis of $\mathbb{C}^d$. Thus, we always have $m \geq d$.

Example 6. Let $w_j = (\cos(2\pi j/m), \sin(2\pi j/m))$, $j = 1, \ldots, m$, be vectors in $\mathbb{C} \approx \mathbb{R}^2 \subset \mathbb{C}^2$. We show that this system is a $(m/2)$-tight frame in $\mathbb{C}^2$. It is easy to check that for $u = (x, y) \in \mathbb{C}^2$ we have

$$\begin{aligned}\sum_{i=1}^{m} |\langle u, w_i\rangle|^2 &= \sum_{j=1} \left(x\cos\left(\frac{2\pi j}{m}\right) + y\sin\left(\frac{2\pi j}{m}\right)\right)\left(\bar{x}\cos\left(\frac{2\pi j}{m}\right) + \bar{y}\sin\left(\frac{2\pi j}{m}\right)\right)\\ &= \frac{m}{2}(|x|^2 + |y|^2) = \frac{m}{2}|u|^2.\end{aligned}$$

Example 7. Let $e_1, \ldots, e_m \in \mathbb{C}^m$ be the standard basis. Let $d < m$ and let H be a d-dimensional subspace of $\mathbb{C}^m$. Let $P_H : \mathbb{C}^m \to H$ be the orthogonal projection. Consider the vectors $P_H e_1, \ldots, P_H e_m$ in $H \approx \mathbb{C}^d$. Since $I_{\mathbb{C}^m} = \sum_{i=1}^{m} e_i e_i^\star$ we get $u = \sum_{i=1}^{m} \langle u, e_i\rangle e_i$. Now, if $u \in H$ then

$$u = P_H u = \sum_{i=1}^{m} \langle u, e_i\rangle P_H e_i = \sum_{i=1}^{m} \langle P_H u, e_i\rangle P_H e_i = \sum_{i=1}^{m} \langle u, P_H e_i\rangle P_H e_i,$$

which means that $I_H = \sum_{i=1}^{m}(Pe_i)(Pe_i)^\star$. This means that $P_H e_1, \ldots, P_H e_m$ is a 1-tight frame on H.

Proof of Theorem 7. For $i = 1, \ldots, m$ and $k = 1, \ldots, r$ we define $w_{i,k}$ by

$$w_{i,1} = \begin{pmatrix} w_i \\ 0^d \\ \vdots \\ 0^d \end{pmatrix}, \quad w_{i,1} = \begin{pmatrix} 0^d \\ w_i \\ \vdots \\ 0^d \end{pmatrix}, \quad \ldots \quad \begin{pmatrix} 0^d \\ 0^d \\ \vdots \\ w_i \end{pmatrix}$$

Let $v_1, \ldots, v_m$ be independent random vectors such that each v_i is distributed uniformly on the r-element set $\{w_{i,k}\sqrt{r/B}\}_{k=1}^{r}$. Clearly,

$$\mathbb{E}[v_i v_i^\star] = \frac{1}{B}\begin{pmatrix} w_i w_i^\star & 0^{d\times d} & \ldots & 0^{d\times d} \\ 0^{d\times d} & w_i w_i^\star & \ldots & 0^{d\times d} \\ \vdots & \vdots & \ddots & \vdots \\ 0^{d\times d} & 0^{d\times d} & \ldots & w_i w_i^\star \end{pmatrix}.$$

By our assumption $\sum_{i=1}^m \mathbb{E}[v_i v_i^\star] = I_{\mathbb{C}^{rd}}$. Also, $\mathbb{E}|v_i|^2 = \frac{r}{B}|w_i|^2 \le \frac{rA}{B}$. From Theorem 6 we know that with positive probability

$$\left\| \sum_{i=1}^m v_i v_i^\star \right\|_{op} \le \left(1 + \sqrt{\frac{rA}{B}}\right)^2.$$

Fix such a realization of $v_1, \ldots, v_m$. Define $S_k = \{1 \le i \le m : \; v_i = w_{i,k}\sqrt{r/B}\}$. We have

$$\left\| \sum_{i=1}^m v_i v_i^\star \right\|_{op} = \left\| \sum_{k=1}^r \left(\sum_{i \in S_k} \frac{r}{B} w_{i,k} w_{i,k}^\star \right) \right\|_{op}.$$

Thus,

$$\left(\sqrt{A} + \sqrt{\frac{B}{r}}\right)^2 \ge \left\| \sum_{k=1}^r \left(\sum_{i \in S_k} w_{i,k} w_{i,k}^\star \right) \right\|_{op} = \max_{k=1,\ldots,r} \left\| \sum_{i \in S_k} w_{i,k} w_{i,k}^\star \right\|_{op} = \max_{k=1,\ldots,r} \left\| \sum_{i \in S_k} w_i w_i^\star \right\|_{op}.$$

□

4.4. **Paving conjecture.**

Theorem 9. Let $r, m \ge 1$ be integers and let $P = (p_{ij})_{i,j=1}^n$ be an orthogonal projection $m \times m$ matrix, that is $P^2 = P$ and $P^\star = P$, such that $p_{ii} \le \varepsilon$ (it will be justified in the proof that p_{ii} is a non-negative real number). Then there is a partition $I = Q_1 + \ldots + Q_r$, where each matrix Q_j is diagonal of the form

$$Q_j = \begin{pmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_m \end{pmatrix}, \qquad \varepsilon_1, \ldots, \varepsilon_m \in \{0, 1\},$$

such that $\|Q_j P Q_j\|_{op} \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right)^2$.

Proof. Let $V = \mathrm{im}(P)$. We have

$$|Pe_i|^2 = \langle Pe_i, Pe_i \rangle = \langle e_i, P^\star P e_i \rangle = \left\langle e_i, P^2 e_i \right\rangle = \langle e_i, Pe_i \rangle = p_{ii} \le \varepsilon.$$

Now, for $v \in V$

$$\sum_{i=1}^m |\langle v, Pe_i \rangle|^2 = \sum_{i=1}^m |\langle Pv, e_i \rangle|^2 = \sum_{i=1}^m |\langle v, e_i \rangle|^2 = |v|^2.$$

Applying Theorem 7 with $B = 1$, $A = \varepsilon$ and $w_i = Pe_i$ (on the space V) we get a partition $\{S_1, \ldots, S_r\}$ of $\{1, \ldots, m\}$ such that

$$\sum_{i \in S_j}^m |\langle v, Pe_i \rangle|^2 \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right)^2 |v|^2, \qquad v \in V.$$

Define Q_i be the matrix of $\mathrm{Proj}_{\mathbb{C}^{S_j}}$. For any $u \in \mathbb{C}^m$ and any j we get

$$|Q_j P u| \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right) |Pu| \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right) |u|.$$

Thus, $\|Q_i P\|_{op} \le \sqrt{\varepsilon} + \frac{1}{\sqrt{r}}$. To finish the proof it suffices to observe that

$$\|Q_jPQ_j\|_{op} = \|Q_jPPQ_j\|_{op} = \|Q_jP(Q_jP)^\star\|_{op} = \|Q_jP\|_{op}^2 \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right)^2,$$

since for any square matrix A we have $\|AA^\star\|_{op} = \|A\|_{op}^2$. Recall also that in fact always

$$\|AA^\star\|_{op} = \|A\|_{op}^2 = \|A^\star\|_{op}^2.$$

□

Theorem 10. Let $r, m \ge 1$ be integers and let $P = (p_{ij})_{i,j=1}^n$ be an orthogonal projection $m \times m$ matrix, such that $p_{ii} \le \varepsilon$ and $\|P\|_{op} \le A$. Then there is a partition $I = Q_1 + \ldots + Q_r$ into diagonal projection matrices such that $\|Q_jPQ_j\|_{op} \le \left(\sqrt{\varepsilon} + \frac{A}{\sqrt{r}}\right)^2$.

Proof. By homogeneity we can assume that $A = 1$. We then write $P = \sum_{i=1}^m \lambda_i u_u u_i^\star$, where $\lambda_i \in [0,1]$ and u_i form an orthonormal basis for $\mathbb{C}^m$. We consider the embedding $\mathbb{C}^m \hookrightarrow \mathbb{C}^m \oplus \mathbb{C}^M$, where $M \ge 0$. If $M \ge M(m, \varepsilon)$, one can find orthonormal vectors $v_1, \ldots, v_m$ in $\mathbb{C}^M$ with disjoint support, such that each of the r_i have coefficients of magnitude at most $\varepsilon^{1/2}$. Define

$$\tilde{u}_i = \sqrt{\lambda_i} u_i \oplus \sqrt{1-\lambda_i} v_i, \qquad i = 1, \ldots, m.$$

This is an orthonormal system in $\mathbb{C}^{M+m}$. Define also

$$P' = \sum_{i=1}^m \tilde{u}_i \tilde{u}_i^\star,$$

which is an orthogonal projection on $\mathbb{C}^m \oplus \mathbb{C}^M$. The matrix P' is an $(m+M) \times (m+M)$ matrix whose top left $m \times m$ minor is equal to P. The bottom right minor of P' is equal to $\sum_{i=1}^m (1-\lambda_i) v_i v_i^\star$ and, since v_i have disjoint supports, has a block structure with blocks corresponding to $v_i v_i^\star$. Thus, the diagonal entries of this $M \times M$ minor are squares of entries of vectors v_i and so are not greater than ε. Since the diagonal entries of P are by assumption not greater than ε, we have $P'_{ii} \le \varepsilon$ for every $i = 1, \ldots, m+M$.

We are now is a position to apply Theorem 9. Thus, there is a partition $I_{\mathbb{C}^m \oplus \mathbb{C}^M} = Q'_1 + \ldots Q'_r$ onto diagonal projections $Q'_1, \ldots, Q'_r$ such that

$$\left\|Q'_j P' Q_j\right\|_{op} \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right)^2, \qquad j = 1, \ldots, r.$$

Let Q_j be the top left $m \times m$ minor of Q'_j. Then $I_{\mathbb{C}^m} = Q_1 + \ldots + Q_r$ and

$$\|Q_jPQ_j\|_{op} \le \left(\sqrt{\varepsilon} + \frac{1}{\sqrt{r}}\right)^2, \qquad j = 1, \ldots, r.$$

since Q_jPQ_j is the top left corner of $Q'_jP'Q'_j$. Indeed, if we write

$$Q'_j = \left[\begin{array}{c|c} Q_j & 0 \\ \hline 0 & D \end{array}\right], \qquad P' = \left[\begin{array}{c|c} P & A \\ \hline A^\star & B \end{array}\right],$$

then

$$Q'_jP'Q'_j = \left[\begin{array}{c|c} Q_jPQ_j & Q_jAD \\ \hline DA^\star Q_j & DBD \end{array}\right].$$

□

Theorem 11. Let $r, m \geq 1$ be integers and let $P = (p_{ij})_{i,j=1}^n$ be a Hermitian $m \times m$ matrix with $p_{ii} = 0$, whose eigenvectors lie in $[-A, B]$, $A, B > 0$. and $\|P\|_{op} \leq A$. Then there is a partition $I = Q_1 + \ldots + Q_r$ into diagonal projection matrices such that the eigenvalues of $Q_j P Q_j$ lie in $[-A, \frac{2}{\sqrt{r}}\sqrt{A(A+B)} + \frac{A+B}{r}]$.

Proof. We use Theorem 10 with $\tilde{P} = P + AI$, $\tilde{A} = A + B$ and $\varepsilon = A$. There is a partition $I_{\mathbb{C}^m} = Q_1 + \ldots + Q_r$ such that the eigenvalues of $Q_j(P + AI)Q_j$ lie in $[0, A + \frac{2}{\sqrt{r}}\sqrt{A(A+B)} + \frac{A+B}{r}]$.

Suppose now that λ is an eigenvalues of $Q_j P Q_j$, that is, there is a vector u such that $Q_j P Q_j u = \lambda u$. In particular u is in the image of Q_j and thus there is v such that $Q_j u = Q_j Q_j v = Q_j v = u$. We have

$$Q_j(P + AI)Q_j u = Q_j P Q_j u + A Q_j^2 u = Q_j P Q_j u + Au = (\lambda + A)u.$$

Thus $\operatorname{spec}(Q_j(P + AI)Q_j) = A + \operatorname{spec}(Q_j P Q_j)$. The assertion follows. $\square$

Lemma 17. If Q is a diagonal projection and P is a Hermitian matrix with $\operatorname{spec}(P) \in [a, b]$ for some a, b, then $\operatorname{spec}(QPQ) \subseteq [a, b]$.

Proof. Let $\lambda_{\max}(P)$ be the maximal eigenvalue of P. Then

$$\lambda_{\max}(QPQ) = \sup_{u \neq 0:\ |u| \leq 1} \langle u, Q_j P Q_j u \rangle = \sup_{u \neq 0:\ |u| \leq 1} \langle Qu, PQu \rangle .$$

Note that $|Qu| \leq |u| \leq 1$, so

$$\sup_{u \neq 0:\ |u| \leq 1} \langle Qu, PQu \rangle \leq \sup_{v:\ |v| \leq 1} \langle v, Pv \rangle = \lambda_{\max}(P).$$

Considering $-P$ instead of P we prove that that the minimal eigenvalue of P satisfies $\lambda_{\min}(QPQ) \geq -A$. $\square$

Theorem 12. Let $r, m \geq 1$ be integers and let $P = (p_{ij})_{i,j=1}^n$ be a Hermitian $m \times m$ matrix with $p_{ii} = 0$, $i = 1, \ldots, m$ and $\|A\|_{op} \leq A$. Then there is a partition $I = Q_1 + \ldots + Q_{r^2}$ into diagonal projection matrices such that $\|Q_j P Q_j\|_{op} \leq A(2\sqrt{2}r^{-1/2} + 2r^{-1})$.

Proof. According to Theorem 11 there is a partition $I = Q_1, \ldots, Q_r$ such that the spectrum of $Q_j P Q_j$ is contained in $[-A, \frac{2\sqrt{2}A}{\sqrt{r}} + \frac{2A}{r}]$. From Lemma 17 the spectrum of this matrix is also contained in $[-A, A]$. Now we can apply Theorem 11 again to the matrix $-Q_j P Q_j$ (note that it has zero diagonal) and thus for any j get a partition $I = Q_{j1} + \ldots + Q_{jr}$ such that the spectrum of $-Q_{jk} Q_j P Q_j Q_{jk}$ is contained in $[-A, \frac{2\sqrt{2}A}{\sqrt{r}} + \frac{2A}{r}]$. Thus, again using Lemma 17,

$$\operatorname{spec}(Q_{jk} Q_j P Q_j Q_{jk}) \subseteq \left[-\frac{2\sqrt{2}A}{\sqrt{r}} - \frac{2A}{r}, A \right] \cap \left[-A, \frac{2\sqrt{2}A}{\sqrt{r}} + \frac{2A}{r} \right].$$

As a consequence

$$\|(Q_{jk} Q_j P Q_j Q_{jk}\|_{op} \leq \frac{2\sqrt{2}A}{\sqrt{r}} + \frac{2A}{r}.$$

We can therefore use the partition $I = \sum_{j,k=1}^r Q_{jk} Q_j$. $\square$

Theorem 13. Let P be a $m \times m$ matrix with vanishing diagonal and $\|P\|_{op} \leq A$. Then there is a partition $I = Q_1 + \ldots + Q_{r^4}$ such that $\|Q_j P Q_j\|_{op} \leq (4\sqrt{2}r^{-1/2} + 4r^{-1})A$.

Proof. We apply Theorem 12 to Hermitian matrices $\frac{1}{2}(P+P^\star)$ and $\frac{1}{2}i(P-P^\star)$ (with operator norm bounded by A). This gives partitions $\{Q_j\}_{j=1}^{r^2}$ and $\{Q'_j\}_{j=1}^{r^2}$ such that

$$\left\| Q_j \frac{P \pm P^\star}{2} Q_j \right\|_{op} \leq (2\sqrt{2}r^{-1/2} + 2r^{-1})A.$$

Using Lemma 17 we get

$$\left\| Q'_j Q_j \frac{P \pm P^\star}{2} Q_j Q'_j \right\|_{op} \leq (2\sqrt{2}r^{-1/2} + 2r^{-1})A.$$

The assertion follows by triangle inequality. □

References

[1] David A. Levin, Yuval Peres, Elizabeth L. Wilmer, *Markov Chains and Mixing Times*.
[2] Terence Tao, Real stable polynomials and the Kadison-Singer problem, https://terrytao.wordpress.com/
[3] A. W. Marcus, D. A. Spielman, N. Srivastava, Ramanujan graphs and the solution of the Kadison-Singer problem, Proc. ICM, Vol III (2014), 375-386.
[4] A. W. Marcus, D. A. Spielman, N. Srivastava, Interlacing families I: bipartite Ramanujan graphs of all degrees, Ann. of Math. 182-1 (2015), 307-325.
[5] A. W. Marcus, D. A. Spielman, N. Srivastava, Interlacing families II: mixed characteristic polynomials and the Kadison-Singer problem, Ann. of Math. 182-1 (2015), 327-350.