Probability: II

李昂生

Discrete Mathematics
U CAS
8 May, 2018

Outline

1. Warm up
2. Algorithms
3. Tail Inequalities
4. Random walk and expander
5. Eigenvalue
6. PageRank and Google Matrix

General view

- Understanding the principles
- Applications of the principles
- Enjoy randomness - powerful, useful, and beautiful

Linearity of expectation

For **any** random variables X and Y,

$$E[X + Y] = E[X] + E[Y]. \tag{1}$$

Basic properties

1. If $a_1, a_2, \cdots, a_n$ are some numbers whose average is c, then there exists an i such that $a_i \geq c$.
2. If X is a random variable which takes values from a finite set and $E[X] = \mu$, then

$$\Pr[X \geq \mu] > 0.$$

3. If $a_1, a_2, \cdots, a_n \geq 0$ are numbers whose average is c, then the fractions of a_i's that are $\geq k \cdot c$ is at most $\frac{1}{k}$.

Markov inequality

Let X be a positive random variable. Then

$$\Pr[X \geq k \cdot E[X]] \leq \frac{1}{k}. \qquad (2)$$

More properties

1. If $a_1, a_2, \cdots, a_n$ are numbers in the interval $[0, 1]$ whose average is ρ, then there are at least $\frac{\rho}{2}$ fraction of the a_i's that are at least $\geq \frac{\rho}{2}$.
2. If $X \in [0, 1]$ and $E[X] = \mu$, then for any $c < 1$,

$$\Pr[X \leq c\mu] \leq \frac{1-\mu}{1-c\mu}.$$

Variance

The variance of a random variable X is:

$$\begin{aligned}\mathrm{Var}[X] &= E[(X - E[X])^2] \\ &= E[X^2] - (E[X])^2. \end{aligned} \tag{3}$$

The standard deviation of X is:

$$\sigma(X) = \sqrt[2]{\mathrm{Var}[X]}. \tag{4}$$

Chebyshev inequality

If X is a random variable with standard deviation σ, then for every $k > 0$,

$$\Pr[|X - E[X]| > k \cdot \sigma] \leq \frac{1}{k^2}. \tag{5}$$

Proof. Applying Markov to $(X - E[X])^2$.

Variance property

If $X_1, X_2, \cdots, X_n$ are pairwise independent, then

$$\mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}[X_i]. \tag{6}$$

Max-Cut

Max-Cut Given an undirected graph $G = (V, E)$ with n vertices and m edges, the *maximum cut* problem, denoted Max-Cut, is to find a set $X \subset V$ such that the number of edges between X and the complement $\bar{X}$ of X, written $e(X, \bar{X})$, is maximised. Or, let $E(X, Y)$ be the set of all the edges of G with one endpoint in X and the other in Y. Then $e(X, Y) = |E(X, Y)|$. The problem is NP-hard.

Theorem 1

For any undirected graph $G = (V, E)$ with n vertices and m edges, there is a partition of the vertex set V into two sets A and B such that

$$e(A, B) \geq \frac{m}{2}.$$

Probabilistic Algorithm

Proof.

We define the cut (A, B) as follows:
Each vertex in V is independently and equiprobably assigned to either A or B.
Then for every edge $e = (x, y) \in E$, with probability $\frac{1}{2}$, the edge $e = (x, y)$ is in the cut (A, B).
Define random variable X_e by

$$X_e = \begin{cases} 1, & \text{if the edge } e \text{ is in the cut } (A, B), \\ 0, & otherwise. \end{cases} \quad (7)$$

Then for every edge e, $E[X_e] = \frac{1}{2}$.
By the linearity of expectation,

$$E[e(A, B)] = \sum_{e \in E} E[X_e] = \frac{m}{2}.$$

Deterministic Algorithm

The theorem implies that there is a cut of size at least $\frac{m}{2}$. Is there a polynomial time algorithm to find such a cut?
Algorithm $\mathcal{C}$:

(1) Let $e = (u, v) \in E$ be an edge of G.
– put u into A, written $u \searrow A$,
– put v into B, i.e., $v \searrow B$.

(2) For every vertex $x \in V \setminus (A \cup B)$,
Case 2a If $e(x, A) > e(x, B)$, then

$$x \searrow B,$$

and
Case 2b. Otherwise. Then

$$x \searrow A.$$

Proof

At every step i, at which we decide a vertex x in A or B, we consider m_i edges. The algorithm $\mathcal{C}$ ensures that at least $\lceil \frac{m_i}{2} \rceil$ edges in the cut.
Therefore,

$$\begin{aligned} e(A,B) &\geq \sum_i \lceil \frac{m_i}{2} \rceil \\ &\geq \frac{m}{2}. \end{aligned}$$

Time complexity: $O(n)$.

$\frac{1}{2}$-approximation algorithm

Note that the maximum number of edges in the cut is at most m. We use OPT to denote the solution for the Max-Cut. Then

$$\mathrm{OPT} \le m$$

Our algorithm $\mathcal{C}$ outputs a cut (A, B) such that

$$e(A, B) \ge \frac{1}{2} \cdot \mathrm{OPT}.$$

This means that the algorithm $\mathcal{C}$ is a $\frac{1}{2}$-approximation algorithm for the Max-Cut problem.

Open Question Is there a polynomial time algorithm that gives approximation ratio better than $\frac{1}{2}$ for the Max-Cut problem?

Maximum Satisfiability

Assume the conjunctive norm form (**CNF**) of formula.
Given a CNF formula ϕ of n variables and m clauses, that is, ϕ is of the following form:

$$\phi : C_1 \wedge C_2 \wedge \cdots \wedge C_m,$$

where each C_i is a clause of the form:

$$z_1 \vee z_2 \vee \cdots \vee z_k,$$

in which each z_j is either a variable x or the negation $\neg y$ of a variable y, referred to as literal.
The question is to find an assignment for the n variables such that the number of satisfied clauses among the m clauses is maximised.
We use **MAX SAT** to denote the problem.
Clearly, it is NP-hard.

Probabilistic Algorithm

Consider a clause C of k variables of the form:

$$C = y_1 \vee y_2 \vee \cdots \vee y_k, \text{each } y_j \text{ is a literal.}$$

Suppose that for each variable x occurred in C, x is defined independently and randomly with equal probability to either 0 or 1. Then the probability that C is satisfied is $1 - \frac{1}{2^k}$.
Suppose that all the variables are assigned randomly with equal probability to either 0 or 1.
For every clause C, define random variable

$$X_C = \begin{cases} 1, & \text{if } C \text{ is satisfied,} \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Then, if C contains k literals, then

$$E[X_C] = 1 - \frac{1}{2^k}.$$

Proof

Let $X = \sum_C X_C$.
Then X is the random number of the satisfied clauses of ϕ.
Suppose that $k_1, k_2, \cdots, k_m$ are the number of literals of $C_1, C_2, \cdots, C_m$, respectively.
By the linearity of expectation,

$$E[X] = \sum_{i=1}^{m}(1 - \frac{1}{2^{k_i}}),$$

which can be computed independently from the random assignments.
Let $N_\phi = E[X]$.

- Generally, $E[X] \geq \frac{m}{2}$.
- If every clause has at least 2 literals, then $E[X] \geq \frac{3}{4}$.
- If every clause has at least 3 literals, then $E[X] \geq \frac{7}{8}$.

Deterministic Algorithm

Fix an ordering of all the variables of ϕ as

$$x_1, x_2, \cdots, x_n.$$

Consider x_1. There are two cases:

Case 1: $x_1 = 0$.

Let n_0 be the number of clauses of ϕ that are satisfied simply by $x_1 = 0$, and ϕ_0 be the formula obtained from ϕ by deleting the satisfied clauses and the literal x_1.

Case 2: $x_1 = 1$.

Let n_1 be the number of clauses of ϕ that are satisfied simply by $x_1 = 1$, and ϕ_1 be the formula obtained from ϕ by deleting the satisfied clauses and the literal x_1.

By the definition of N_ϕ,

$$\frac{1}{2}(n_0 + N_{\phi_0}) + \frac{1}{2}(n_1 + N_{\phi_1}) = N_\phi. \tag{9}$$

Proof

Therefore, either $n_0 + N_{\phi_0} \geq N_\phi$ or $n_1 + N_{\phi_1} \geq N_\phi$.
Case 1: If $n_0 + N_{\phi_0} \geq N_\phi$, then
– set $x_1 = 0$, and
– $\phi \leftarrow \phi_0$.
Case 2: Otherwise, then
– $x_1 = 1$, and
– $\phi \leftarrow \phi_1$.
In either case, repeat the procedure above, until we assigned a value for every variable x_i.
The assignment satisfies at least N_ϕ clauses.

Self-reducibility method

The method of the algorithm for the MAX SAT problem above is due to an important property of SAT, that is, the **self-reducibility property**.

This is a general idea for many algorithmic problems.

The method is referred to as **Self-reducibility method**.

The Chernoff bounds

Let $X_1, X_2, \cdots, X_n$ be mutually independent random variables over $\{0, 1\}$, and let $\mu = \sum_{i=1}^{n} E[X_i]$. Then for every $\delta > 0$,

(1)

$$\Pr[\sum_{i=1}^{n} X_i \geq (1+\delta)\mu] \leq [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}. \tag{10}$$

(2)

$$\Pr[\sum_{i=1}^{n} X_i \leq (1-\delta)\mu] \leq [\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}]^{\mu}. \tag{11}$$

For every $c > 0$,

$$\Pr[|\sum_{i=1}^{n} X_i - \mu| \geq c \cdot \mu] \leq 2 \cdot e^{-\min\{c^2/4, c/2\} \cdot \mu}.$$

Poisson Trials

Recall: Let $X_1, \cdots, X_n$ be independent Bernoulli trials such that for $1 \le i \le n$, $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$. Let $X = \sum_{i=1}^{n} X_i$, then X is said to have the binomial distribution.

Generally, let $X_1, \cdots, X_n$ be independent coin tosses such that for $1 \le i \le n$, $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Such coin tosses are referred to as *Poisson trials*.

Let $X = \sum_{i=1}^{n} X_i$, X_i are Poisson trials.

Clearly,

$$E[X] = \sum_{i=1}^{n} p_i = \mu \text{ (denoted)} \tag{12}$$

Questions

1) For a real number $\delta > 0$, what is the probability of $X > (1 + \delta)\mu$?
2) How large must δ be in order that the tail probability is less than a prescribed value ϵ?

The answer: **The Chernoff bounds**.

Moment Generating Function

For a random variable X, we call the quantity $E[e^{tX}]$ the *moment generating function of X*.
Because:

$$E[e^{tX}] = \sum_{k=0}^{\infty} \frac{E[X^k]}{k!} t^k, \tag{13}$$

where $E[X^k]$ is the k-th moment of X, for natural number k.
The idea to prove the Chernoff bounds is:
the moment generating function $+$ the Markov inequality.

Chernoff bound - lower bound

Theorem 2

Let $X_1, X_2, \cdots, X_n$ be independent Poisson trials such that for $1 \le i \le n$, $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$, where $0 < p_i < 1$. Then, for $X = \sum_{i=1}^{n} X_i$, $\mu = E[X] = \sum_{i=1}^{n} p_i$ and any $\delta > 0$,

$$\Pr[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}. \tag{14}$$

Concentration Theorem

Proof

For any positive real t,

$$
\begin{aligned}
& \Pr[X > (1+\delta)\mu] \\
= \;& \Pr[tX > t(1+\delta)\mu] \\
= \;& \Pr[\exp(tX) > \exp(t(1+\delta)\mu)] \\
< \;& \frac{E[\exp(tX)]}{\exp(t(1+\delta)\mu)},
\end{aligned}
$$

the last inequality is by the Markov Inequality.

Proof - continued

Consider $E[\exp(tX)]$. By the independency of X_i's, and hence $\exp(tX_i)$'s,

$$\begin{aligned} E[\exp(tX)] &= E[\exp(t\sum_{i=1}^{n} X_i)] \\ &= E[\prod_{i=1}^{n} \exp(tX_i)] \\ &= \prod_{i=1}^{n} E[\exp(tX_i)]. \end{aligned}$$

This gives

$$\Pr[X > (1+\delta)\mu] < \frac{\prod_{i=1}^{n} E[\exp(tX_i)]}{\exp(t(1+\delta)\mu)}. \tag{15}$$

Proof - continued

By definition,

$$e^{tX_i} = \begin{cases} e^t, & \text{with probability } p_i, \\ 1, & \text{with probability } 1 - p_i. \end{cases} \tag{16}$$

Therefore,

$$E[e^{tX_i}] = p_i e^t + 1 - p_i = 1 + p_i(e^t - 1).$$

For $x = p_i(e^t - 1)$, we use the inequality $1 + x < e^x$ to obtain:

Proof - continued

$$
\begin{aligned}
\Pr[X > (1+\delta)\mu] &< \frac{\prod_{i=1}^{n} \exp(p_i(e^t - 1))}{\exp(t(1+\delta)\mu)} \\
&= \frac{\exp(\sum_{i=1}^{n} p_i(e^t - 1))}{\exp(t(1+\delta)\mu)} \\
&= \frac{\exp((e^t - 1)\mu)}{\exp(t(1+\delta)\mu)}.
\end{aligned}
$$

Proof -continued

Let $f = \frac{\exp((e^t - 1)\mu)}{\exp(t(1+\delta)\mu)}$.
Set $f' = 0$. Solving the equation, we obtain

$$t = \ln(1 + \delta).$$

For this choice of t,

$$f = \left[\frac{e^{\delta}}{(1 + \delta)^{(1+\delta)}} \right]^{\mu}.$$

Summary of the proof

1. We studied the random variable e^{tX} rather than X
2. The expectation of the product of the e^{tX_i} turns into the product of their expectations due to independence
3. We pick a value of t to obtain the best possible upper bound.

The approach above works for the sum of other distributions.

Significance

- Usually, $\mu = \Theta(n)$
- For $\delta > 0$ such that

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} = \frac{1}{2},$$

then,

$$\left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} = \frac{1}{2^{\Theta(n)}},$$

which is **exponentially decreasing to 0**.

Chernoff bound - upper bound

Theorem 3

Let $X_1, X_2, \cdots, X_n$ be independent Poisson trials such that for $1 \le i \le n$, $\Pr[X_i = 1] = p_i$, $0 < p_i < 1$. Then, for $X = \sum_{i=1}^{n} X_i$, $\mu = E[X] = \sum_{i=1}^{n} p_i$ and δ with $0 < \delta < 1$,

$$\Pr[X < (1-\delta)\mu] < \exp(-\mu\frac{\delta^2}{2}). \tag{17}$$

Proof

As before,

$$\begin{aligned} & \Pr[X < (1-\delta)\mu] \\ = & \Pr[-X > -(1-\delta)\mu] \\ = & \Pr[\exp(-tX) > \exp(-t(1-\delta)\mu)], \end{aligned}$$

for any positive real t.
By Markov and the same argument as before,

$$\Pr[X < (1-\delta)\mu] < \frac{\prod_{i=1}^{n} E[\exp(-tX_i)]}{\exp(-t(1-\delta)\mu)}. \qquad (18)$$

Proof -continued

Computing $E[\exp(-tX_i)]$, we have

$$\Pr[X < (1-\delta)\mu] < \frac{\exp(\mu(e^{-t}-1))}{\exp(-t(1-\delta)\mu)}.$$

Set $t = \ln \frac{1}{1-\delta}$, we have

$$\Pr[X < (1-\delta)\mu] < \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^{\mu}. \tag{19}$$

For $\delta \in (0, 1]$,

$$(1-\delta)^{(1-\delta)} > \exp(-\delta + \delta^2/2),$$

using the Mclaurin expansion for $\ln(1-\delta)$.

Martingales

There is a Martingale theory dealing with the case that X_i are not totally independent, with similar bounds.
Powerful and useful in theoretical computer science.

Random Walk

- To understand the dynamics of physical systems
- To understand the operations, interactions and communications that occur in networks
- To understand virus spreading in networks
- To understand the evolution of systems in nature and society
- To understand the role of randomness

Expanders vs Randomness

Advanced topic and research directions: on the basis of randomness

- Communication networks
- Pseudo random generator
- Randomness
- Derandomisation
- UPATH is Log space
- PageRank

Conventions

For simplicity, we assume that the graphs are:

- regular
- selfloop
- parallel edges

Theory is possible for general graphs without these assumptions.

Inner product

$\langle u, v \rangle$

- $\langle xu + yv, w \rangle = x\langle u, w \rangle + y\langle v, w \rangle$
- $\langle v, u \rangle = \overline{\langle u, v \rangle}$, $\bar{z}$ is the complex conjugation of z
- For all u, $\langle u, u \rangle \geq 0$, with 0 only if $u = 0$
- $\langle u, v \rangle = 0$ means u, v are orthogonal, written $u \perp v$
- If $u^1, u^2, \cdots, u^n$ satisfy $u^i \perp u^j$ for all $i \neq j$, then they are linearly independent.

Parseval's identity: If $u^1, u^2, \cdots, u^n$ form an orthonormal basis for C^n, then for every v, if $v = \sum_i \alpha_i u^i$, then

$$\langle v, v \rangle = \sum_{i=1}^{n} |\alpha_i|^2. \tag{20}$$

Hilbert space: Vector spaces with inner product.

Dot product

- For $u, v \in \mathbb{F}^n$, $u \odot v = \sum_{i=1}^{n} u_i v_i$
- $S \subset \mathbb{F}^n$, $S^\perp = \{u : \; u \perp S\}$
- $u \perp v$, if $u \odot v = 0$, $u \perp S$, if for all $v \in S$, $u \perp v$.
- $\dim(S) + \dim(S^\perp) = n$
- $u \in \mathbb{F}^n$, $u^\perp = \{v : \; v \perp u\}$, and $\dim(u^\perp) = n - 1$.

Random subsum principle

For every non-zero $u \in \mathrm{GF}(2^n)$,

$$\Pr_{v \in_{\mathrm{R}} \mathrm{GF}(2^n)}[u \odot v = 0] = \frac{1}{2}. \tag{21}$$

Eigenvectors and eigenvalues

If A is a real, symmetric matrix, for λ and v, if $Av = \lambda v$, then

$$\lambda\langle v, v\rangle = \langle Av, v\rangle = \overline{\langle v, Av\rangle} = \overline{\langle v, \lambda v\rangle} = \bar{\lambda}\langle v, v\rangle$$

This implies that:

$$\lambda = \bar{\lambda}$$

so λ is a real.

Norms

It is a function of the following form:

$$|| \; || : \mathbb{F}^n \to \mathbb{R}^{\geq 0} \tag{22}$$

A norm satisfies the following properties:

(i) $||v|| = 0 \iff v = 0$

(ii) $||\alpha v|| = |\alpha| \cdot ||v||$, where α is a real scale.

(iii) $||u + v|| \leq ||u|| + ||v||$.

L_p-norm

L_p-norm of v, $p \geq 1$,

$$||v||_p = (\sum_{i=1}^{n} |v_i|^p)^{1/p}$$

$p = 2$, L_2-norm, the Euclidean norm

$$||v||_2 = (\sum_{i=1}^{n} |v_i|^2)^{1/2}$$

$p = 1$, L_1-norm

$$||v||_1 = \sum_{i=1}^{n} |v_i|$$

$p = \infty$, L_∞-norm

$$||v||_\infty = \max_i |v_i|.$$

Hölder inequality

For every p, q, if $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\|u\|_p \cdot \|v\|_q \geq \sum_{i=1}^{n} |u_i v_i|. \tag{23}$$

$p = q = 2$: Cauch-Schwarz

L_1- and L_2-norms

For every vector $v \in \mathbb{R}^n$,

$$\frac{|v|_1}{\sqrt{n}} \leq ||v||_2 \leq |v|_1. \tag{24}$$

Notations: Adjacent matrix

- G: d-regular, n vertices,
- p: a column vector, a distribution over the vertices of G

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ p_n \end{pmatrix} \tag{25}$$

  where $p_1 + p_2 + \cdots p_n = 1$.
- A_{ij}: $\frac{n_{ij}}{d}$, where n_{ij} the number of edges between i and j.
- A: the adjacent matrix. It is normalised, symmetric, stochastic

Notations: Adjacent matrix

- $q = Ap$: the distribution of a random walk in G from distribution p.
- $A^l e^i$: the distribution of l-step random walk from node i
- 1:
  The uniform distribution is:

$$\mathbf{1} = \begin{pmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \cdots \\ \frac{1}{n} \end{pmatrix} \tag{26}$$

- $1^{\perp}$: $\{v : \ v \perp 1\}$
- $v \perp 1 \iff \sum v_i = 0$.

$\lambda(A)$

Define

$$\begin{aligned}\lambda(A) &= \lambda(G) \\ &= \max\{||Av||_2 : \; ||v||_2 = 1, v \perp 1\}.\end{aligned} \tag{27}$$

Suppose that

$$\lambda_1, \lambda_2, \cdots, \lambda_n$$

are the eigenvalues of A with orthogonal eigenvectors

$$v^1, v^2, \cdots, v^n$$

respectively, that are listed such that:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|. \tag{28}$$

$|\lambda_i| \leq 1$

For λ and v such that $Av = \lambda v$. Then $\lambda = \frac{\langle v, Av \rangle}{\langle v, v \rangle}$.
By definition,

$$\langle v, Av \rangle = \sum_{i=1}^{n} a_{ii} v_i^2 + 2 \sum_{i<j, i \sim j} a_{ij} v_i v_j$$

For $i < j$, $i \sim j$:

$$a_{ij}(v_i - v_j)^2 = a_{ij} v_i^2 - 2 a_{ij} v_i v_j + a_{ij} v_j^2$$

Summing up all such i, j's:

$$\sum_{i<j, i \sim j} a_{ij}(v_i - v_j)^2 = \sum_{i=1}^{n} (1 - a_{ii}) v_i^2 - 2 \sum_{i<j, i \sim j} a_{ij} v_i v_j$$

Proof - I

$$\begin{aligned}
\langle v, Av\rangle &= \sum_{i=1}^{n} a_{ii}v_i^2 + \sum_{i=1}^{n}(1-a_{ii})v_i^2 - \sum_{i<j, i\sim j} a_{ij}(v_i - v_j)^2 \\
&= \sum_{i=1}^{n} v_i^2 - \sum_{i<j, i\sim j} a_{ij}(v_i - v_j)^2. \qquad (29)
\end{aligned}$$

Noting that $\sum_{i=1}^{n} v_i^2 \geq 2\sum_{i<j} a_{ij}v_iv_j$, we have

$$-\sum_{i=1}^{n} v_i^2 \leq \sum_{i=1}^{n} v_i^2 - \sum_{i<j, i\sim j} a_{ij}(v_i - v_j)^2 \leq \sum_{i=1}^{n} v_i^2.$$

So that

$$-1 \leq \lambda \leq 1.$$

By definition, $A\mathbf{1} = \mathbf{1}$. So $\lambda_1 = 1$, and $\mathbf{1}$ is the eigenvector of $\lambda_1 = 1$. By the choice of the eigenvectors, $\mathbf{1}^{\perp} = \mathrm{Span}\{v^2, \cdots, v^n\}$.

Proof - II

Given v, with $v \perp 1$, $||v||_2 = 1$.
Let $v = \alpha_2 v^2 + \cdots + \alpha_n v^n$ with $\alpha_2^2 + \cdots + \alpha_n^2 = 1$.

$$Av = \alpha_2 Av^2 + \cdots + \alpha_n Av^n = \alpha_2 \lambda_2 v^2 + \cdots + \alpha_n \lambda_n v^n$$

$$||Av||_2^2 = \alpha_2^2 \lambda_2^2 + \cdots + \alpha_n^2 \lambda_n^2$$

Since $\lambda_2^2 \geq \cdots \geq \lambda_n^2$,

$$\max ||Av||_2^2 = \lambda_2^2.$$

Therefore

$$\lambda = \lambda(G) = |\lambda_2|.$$

Spectral gap

We call $1 - \lambda(G)$ the *spectral gap of G*.

Lemma 4

Let G be an n-vertex regular graph and p a probability distribution over G's vertices. Then,

$$||A^l p - 1||_2 \leq \lambda^l.$$

Proofs consist of the following items:
1) By definition of $\lambda = \lambda(G)$, for every $v \perp 1$,

$$||Av||_2 \leq \lambda ||v||_2.$$

Proofs - I

2) If $v \perp 1$, then so is Av.

$$\langle 1, Av\rangle = \langle A^{\mathrm{T}} 1, v\rangle = \langle 1, v\rangle = 0.$$

Note $A = A^{\mathrm{T}}$, and $A1 = 1$.
3) $A: \ 1^{\perp} \to 1^{\perp}$, and
A **shrinks** each $v \in 1^{\perp}$ by at least λ factor in L_2 norm.
4) By 3), A^l shrinks each $v \in 1^{\perp}$ by at least λ^l factor, giving

$$\lambda(A^l) \leq \lambda^l.$$

Proofs - II

5) Let $p = \alpha 1 + p'$, $p' \perp 1$, Since $p' \perp 1$, $\sum p'_i = 0$. But $\sum p_i = 1$, so $\alpha = 1$.

$$A^l p = A^l(1 + p') = A^l 1 + A^l p' = 1 + A^l p'.$$

$$\begin{aligned}
||A^l p - 1||_2 &= ||A^l p'||_2 \\
&\leq ||A^l||_2 \cdot ||p'||_2 \\
&\leq \lambda^l \cdot ||p'||_2 \\
&\leq \lambda^l \cdot ||p||_2 \\
&\leq \lambda^l \cdot |p|_1 = \lambda^l.
\end{aligned}$$

The third inequality uses $||p||_2^2 = ||1||_2^2 + ||p'||_2^2$.

Log space algorithm for connectivity in expanders

Suppose that λ is a constant significantly smaller than 1.
By Lemma 4 above, let $l = O(\log n)$.
Then $\lambda^l \approx 0$. Therefore

$$A^l p \approx 1.$$

This means that for any two nodes i, j, the distance between i and j is within $O(\log n)$.
According to this property, we are able to design a log space algorithm to decide, for any two vertices, whether or not, they are connected.
The algorithm simply enumerates all the paths from i of length $O(\log n)$, to see if there is a path passes j. The enumeration of all the paths can be done in log space.

Randomized log space (RL, for short) for connectivity

Lemma 5

*(**RL**) If G is a regular connected graph with self-loop at each vertex, then*

$$\lambda(G) \leq 1 - \frac{1}{4dn^2}. \quad (30)$$

Let $u \perp 1$, $||u||_2 = 1$. We show that $||Au||_2 \leq 1 - \frac{1}{4dn^2}$.
Let $v = Au$. It suffices to show that $1 - ||v||_2^2 \geq \frac{1}{2dn^2}$.
Since $||u||_2 = 1$,

$$1 - ||v||_2^2 = ||u||_2^2 - ||v||_2^2.$$

Considering $\sum_{i,j} A_{ij}(u_i - v_j)^2$, we have

Proofs - I

$$
\begin{aligned}
\sum_{i,j} A_{ij}(u_i - v_j)^2 &= \sum_{i,j} A_{ij}u_i^2 - 2\sum_{i,j} A_{ij}u_i v_j + \sum_{i,j} A_{ij} v_j^2 \\
&= \sum_{i=1}^{n} u_i^2 - 2\langle Au, v\rangle + \sum_{j=1}^{n} v_j^2 \\
&= ||u||_2^2 - 2\langle Au, v\rangle + ||v||_2^2 \\
&= ||u||_2^2 - 2||v||_2^2 + ||v||_2^2 \\
&= ||u||_2^2 - ||v||_2^2 \\
&= 1 - ||v||_2^2.
\end{aligned}
$$

Therefore, we only need to prove

$$
\sum_{i,j} A_{ij}(u_i - v_j)^2 \geq \epsilon = \frac{1}{2dn^2}.
$$

Proofs - II

By the choice of u, $\sum u_i = 0$, and $\sum u_i^2 = 1$. So there exist i, j such that $u_i u_j < 0$.
Let $u^+ = \max_i\{u_i\}$, and $u^- = \min_i\{u_i\}$. If both $u^+ < \frac{1}{\sqrt{n}}$ and $u^- > -\frac{1}{\sqrt{n}}$ hold, then $\sum_{i=1}^{n} u_i^2 < 1$.
Since $||u||_2 = 1$, either $u^+ \geq \frac{1}{\sqrt{n}}$ or $u^- \leq -\frac{1}{\sqrt{n}}$. Let i and j be such that $u_i = u^+$ and $u_j = u^-$. Then:

$$u_i - u_j \geq \frac{1}{\sqrt{n}}. \tag{31}$$

Proofs - III

Because G is connected, there is a path P between i and j. Suppose that the path P is labelled by $1, 2, \cdots, D+1$. Then:

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \\
\leq \; & u_1 - u_{D+1} \\
= \; & (u_1 - v_1) + (v_1 - u_2) + (u_2 - v_2) + \cdots + (v_D - u_{D+1}) \\
\leq \; & |u_1 - v_1| + |v_1 - u_2| + \cdots + |v_D - u_{D+1}| \\
\leq \; & \sqrt{(u_1 - v_1)^2 + (v_1 - u_2)^2 + \cdots + (v_D - u_{D+1})^2} \cdot \sqrt{2D+1}.
\end{aligned}
$$

Proofs - IV

Therefore,

$$(u_1 - v_1)^2 + (v_1 - u_2)^2 + \cdots + (v_D - u_{D+1})^2 \geq \frac{1}{n(2D+1)}.$$

Since $A_{ii}, A_{ii+1} \geq \frac{1}{d}$,

$$\begin{aligned}
\sum_{i,j} A_{ij}(u_i - v_j)^2 &\geq \frac{1}{d} \cdot [(u_1 - v_1)^2 + (v_1 - u_2)^2 + \cdots + (v_D - u_{D+1})^2] \\
&\geq \frac{1}{dn(2D+1)} \\
&\geq \frac{1}{2dn^2}.
\end{aligned}$$

Random walk lemma

Lemma 6

(Random walk lemma) Let G be a d-regular n-vertex graph with all vertices having a self-loop. Let s be a vertex in G. Let $l > \Omega(dn^2 \log n)$, and X_l be the distribution of the vertex of the lth step in a random walk from s. Then for every t,

$$\Pr[X_l = t] > \frac{1}{2n}.$$

Proofs - 1

By the previous lemma,

$$||A^l p - 1||_2 \leq (1 - \frac{1}{4dn^2})^{\Omega(dn^2 \log n)} < \frac{1}{n^\alpha}$$

for some constant α.
Choose α such that for $q = A^l p$,

$$|q - 1|_1 \leq \sqrt{n} \cdot ||q - 1||_2 < \frac{1}{n^2}.$$

Then for every i,

$$|q_i - \frac{1}{n}| < \frac{1}{n^2}$$

So that

$$-\frac{1}{n^2} < q_i - \frac{1}{n} < \frac{1}{n^2}$$

Proofs - 2

Therefore, the probability that $X_l = t$ is:

$$\begin{aligned} q_i &> \frac{1}{n} - \frac{1}{n^2} \\ &\geq \frac{1}{2n}. \end{aligned}$$

Run the l-step random walks for $t = O(n \log n)$ many times, then with high probability, every vertex is visited, if the graph is connected.

This gives a randomized log space, written **RL**, algorithm to decide the connectivity of two vertices.

谢谢！