& 6

Approximate Counting and Mixing Time of Markov chains **Fall 2024**

Lecture 5&6: Spectral Graph Theory of MCs

Lecturer: Shayan Oveis Gharan *01/04/22*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

THroughout this lecture we study spectral properties of the Markov Kernel K for a reversible Markov chain.

5.1 Random Walk Operator

Recall that for a simple random walk on G says that if I am at vertex u, I choose an edge incident to u with probability proportional to its weight. For any adjacent nodes u and v, let

$$K(x, y) = \frac{w(\{x, y\})}{d_w(x)}.$$

For a function $f : V \to \mathbb{R}$, define

$$Kf(y) := \sum_x K(x, y)f(y).$$

We equip the linear space $\mathbb{R}^V$ with the following inner product: For two vectors $f, g : V \to \mathbb{R}$, define

$$\langle f, g\rangle_\pi = \mathbb{E}_{x\sim\pi}\, f(x)g(x) = \sum_x \pi(x)f(x)g(v).$$

When clear from context, we drop the subscript π from the norm. This naturally defines a norm, where for any such function f, $\|f\| = \sqrt{\langle f, f\rangle}$.

Fact 5.1. *K is self-adjoint with respect to the above inner product, i.e., for any two functions $f, g : V \to \mathbb{R}$,*

Proof.

$$\begin{aligned}\langle f, Kg\rangle &= \mathbb{E}_{x\sim\pi}\left[f(x)Kg(x)\right]\\ &= \mathbb{E}_{x\sim\pi}\left[f(x)\sum_{y\sim x}\frac{w(\{x, y\})}{d_w(x)}g(y)\right]\\ &= \sum_{\{x,y\}}\frac{w(\{x, y\})}{W}f(x)g(y)\end{aligned}$$

where W is the sum of weights of all edges. Similarly, you can verify that $\langle Pf, g\rangle$ is also equal to the RHS. □

Fact 5.2. *Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of the K. It holds that $\lambda_1 = 1$ and $\lambda_n \geq -1$.*

Proof. First, observe that the all-ones function $\mathbf{1}$ is an eigenfunction,

$$P\mathbf{1} = \mathbf{1}.$$

Second, we show $\lambda_i \leq 1$ for all i. For any eigenfunction $f : V \to \mathbb{R}$, with eigenvalue λ, i.e., $Pf = \lambda f$, we claim that $\lambda \leq 1$. Say $u = \text{argmax}_v |f(v)|$. Then,

$$\lambda f(u) = Pf(u) = \mathbb{E}_{\{u,v\}|u} f(v) \leq \mathbb{E}_{\{u,v\}|u} |f(v)| \leq \mathbb{E}_{\{u,v\}|u} |f(u)| = |f(u)|.$$

In the second inequality we used $|f(v)| \leq |f(u)|$ for all $v \in V$. So, we have $\lambda \leq 1$ as desired. Also, observe from the same inequality that $|\lambda| \leq 1$ as desired. □

5.2 Background on Spectral theorem for self-adjoint operators

For this section, we consider $\mathbb{C}^V$ as a vector space.

Definition 5.3 (Inner Product)**.** *We say $\langle .,. \rangle$ is an inner product if*

- *It is linear, i.e., for $f, g, h \in \mathbb{C}^V, a \in \mathbb{C}$, $\langle af + g, h \rangle = a\langle f, h \rangle + \langle g, h \rangle$.*
- *For any $\langle f, f \rangle \geq 0$, and*
- *For any $f, g \in \mathbb{C}^V$, we have $\langle f, g \rangle = \overline{\langle g, f \rangle}$.*

Definition 5.4 (Self-adjoint Operators)**.** *We say an operator P is self-adjoint with respect to an inner product $\langle .,. \rangle$, if for any $f, g \in \mathbb{C}^V$, $\langle Pf, g \rangle = \langle f, Pg \rangle$.*

Theorem 5.5. *Suppose $P : \mathbb{C}^V \to \mathbb{C}^V$ is a self-adjoint operator with respect to $\langle .,. \rangle$. Then, there are $n = dim(\mathbb{C}^V) = |V|$ real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ with corresponding orthonormal eigenfunctions $f_1, \ldots, f_n$ such that for any $1 \leq i < j \leq n$,*

$$\langle f_i, f_j \rangle = 0,$$

and for any i, $\|f_i\| = 1$.

First, we show that any operator on $\mathbb{C}^V$ has an eigenvalue:

Lemma 5.6. *Let $W \subseteq \mathbb{C}^V$ be a function/vector space and let $P : W \to W$ be a linear operator. Then, P has an eigenvalue, $Pf = \lambda f$ for a function $f \in W$.*

Proof. Let $f \in W$ be an arbitrary non-zero function. Consider

$$f, Pf, P^2 f, \ldots P^n f$$

where $n = \dim(W)$ is the dimension of our function space. It follows that there exists $c_0, \ldots, c_n \in \mathbb{C}$ such that

$$c_0 f + c_1 Pf + \ldots c_n P^n f = \mathbf{0}$$

Consider the polynomial $p(x) = c_0 + c_1 x + \cdots + c_n x^n$. By the fundamental theorem of algebra this polynomial has n roots over the complex domain, $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. So, we can write

$$p(x) = c_n \prod_{i=1}^{n} (x - \lambda_i).$$

Consider the operator $p_P : W \to W$ where for $g \in W$ we have

$$p_P g = c_0 g + c_1 P g + \cdots + c_n P^n g,$$

and notice that by definition $p_P f = \mathbf{0}$. Consequently,

$$p_P f = \left(c_n \prod_{i=1}^{n} (P - \lambda_i I) \right) f = \mathbf{0}$$

Consider the following sequence of vectors:

$$f_0 = f, \quad f_1 = Pf - \lambda_1 f, \quad f_2 = Pf_1 - \lambda_2 f_1, \quad \ldots, \quad f_n = Pf_{n-1} - \lambda_n f_{n-1}$$

But, $c_n f_n = p_P f = \mathbf{0}$. Since $f \neq \mathbf{0}$, there must be an $0 \leq i < n$ such that $f_i \neq 0$ but $Pf_i - \lambda f_i = \mathbf{0}$. But then we get f_i is an eigenfunction with eigenvalue λ_i. □

Now, we show that for self-adjoint operators this eigenvalue is real:

Fact 5.7. *Let P be a self-adjoint operator $\mathbb{C}^V \to \mathbb{C}^V$ with an eigenvalue λ, i.e., $Pf = \lambda f$ Then, $\lambda \in \mathbb{R}$. In particular, if $P : \mathbb{R}^V \to \mathbb{R}^V$ then we can assume f is real, i.e., $f \in \mathbb{R}^V$.*

Proof. Recall for any function $f \in \mathbb{C}^V$, $\|f\|^2 = \langle f, f \rangle \geq 0$. So,

$$\|Pf\|^2 = \langle Pf, Pf \rangle = \langle P^2 f, f \rangle = \lambda^2 \|f\|^2 .$$

Since $\|Pf\|^2, \|f\|^2 \geq 0$ we must have, $\lambda^2 \geq 0$ and therefore $\lambda \in \mathbb{R}$.

To see the second conclusion suppose $f = ig + h$ for $g, h \in \mathbb{R}^V$. Then,

$$Pf = P(ig + h) = iPg + Ph = \lambda f = i\lambda g + \lambda h$$

Since $Pg, Ph, \lambda g, \lambda h \in \mathbb{R}^V$ we must have $Pg = \lambda g$ and $Ph = \lambda h$. □

Proof of Theorem 5.5. We prove this by induction on the dimension of the function space. By above lemmas P has a real eigenvalue λ with corresponding eigenfunction $f \in \mathbb{C}^V$ and without loss of generality assume $\|f\| = 1$.

Let $W = \{g : \langle g, f \rangle = 0\}$ be the set of vectors orthogonal to f. For any $g \in W$, we have

$$\langle Pg, f \rangle = \langle g, Pf \rangle = \langle g, \lambda f \rangle = \lambda \langle g, f \rangle = 0.$$

Therefore, P maps $W \to W$. So, since W has one less dimension, by the IH P has $n - 1$ real eigenvalues $\lambda_2, \ldots, \lambda_n$ with corresponding orthonormal eigenfunctions $f_2, \ldots, f_n \in W$. □

5.3 Functional Analysis of Markov Chains

Consider a Markov chain on a state space Ω with a (reversible) Markov kernel K. Let Φ be a convex function that we choose later.

For a function $f : \Omega \to \mathbb{R}$, define

$$\mathrm{D}_\pi^\Phi(f) = \mathbb{E}_\pi \left[\Phi \circ f\right] - \Phi(\mathbb{E}_\pi f)$$

Observe that, if Φ is convex, then by the Jensen's inequality,

$$\mathrm{D}_\pi^\Phi(f) \geq 0$$

for all functions f. Furthermore, it is equal to zero if f is a constant function.

Example 5.1. *Let $f = \nu/\pi$ be the ratio of a probability distribution ν over Ω to π. For, $\Phi(x) = \frac{1}{2}|x-1|$, we have*

$$\mathrm{D}_\pi^\Phi(f) = \mathbb{E}_\pi \frac{1}{2}\left|\frac{\nu}{\pi}(x) - 1\right| - \frac{1}{2}\left|\mathbb{E}_\pi \frac{\nu}{\pi} - 1\right| = \sum_x \frac{1}{2}|\nu(x) - \pi(x)| - 0 = \|\nu - \pi\|_{TV}$$

Example 5.2 (Variance)**.** *Suppose $\Phi(x) = x^2$. Then,*

$$\mathrm{D}_\pi^\Phi(f) = \mathbb{E}_\pi[f^2] - (\mathbb{E}\, f)^2 = \|f - \mathbb{E}\, f\|_\pi^2 = \mathrm{Var}(f)$$

It turns out that that this gives upper-bound on the total variation distance: For a probability distribution ν, $f = \nu/\pi$, by Cauchy-Schwartz inequality we have

$$\begin{aligned}\|\nu - \pi\|_{TV} &= \mathbb{E}_\pi \frac{1}{2}|\frac{\nu}{\pi}(x) - 1| \\ &\le \frac{1}{2}\sqrt{\mathbb{E}_\pi(\frac{\nu}{\pi}(x) - 1)^2} \\ &= \frac{1}{2}\sqrt{\mathbb{E}_\pi(\frac{\nu}{\pi})^2 - \mathbb{E}_\pi \frac{\nu}{\pi}} \\ &= \frac{1}{2}\sqrt{\mathbb{E}_\pi(\frac{\nu}{\pi})^2 - \left(\mathbb{E}_\pi \frac{\nu}{\pi}\right)^2} \\ &= \frac{1}{2}\sqrt{\mathrm{D}_\pi^\Phi(f)}\end{aligned}$$

where we used that $\mathbb{E}_\pi \frac{\nu}{\pi} = \sum_x \nu(x) = 1$.

Example 5.3 (Entropy)**.** *Now, suppose $\Phi(x) = x \log x$. Then,*

$$\mathrm{D}_\pi^\Phi(f) = \mathbb{E}_\pi[f \log f] - \mathbb{E}_\pi[f] \log \mathbb{E}_\pi[f]$$

In this case, $\mathrm{D}_\pi^\Phi(\nu\|\pi) = \sum_x \nu(x) \log \frac{\nu(x)}{\pi(x)}$, is the KL-divergence of ν, ϕ where we again used that $\mathbb{E}_\pi \frac{\nu}{\pi} = 1$. In this case it follows by the Pinsker's inequality that

$$\|\nu - \pi\|_{TV} \le \sqrt{\frac{1}{2}\mathrm{D}_\pi^\Phi(\frac{\nu}{\pi})}$$

5.4 Contraction

Lemma 5.8 (Data Processing Inequality)**.** *For any non-negative function $f : \Omega \to \mathbb{R}_{\ge 0}$ and for any convex Φ, we have*

$$\mathrm{D}_\pi^\Phi(Kf) \le \mathrm{D}_\pi^\Phi(f)$$

Proof. First, observe that

$$\mathbb{E}_\pi Kf = \langle Kf, \mathbf{1}\rangle = \langle f, K\mathbf{1}\rangle = \langle f, \mathbf{1}\rangle = \mathbb{E}_\pi f$$

where we used that K is self-adjoint. So, we have $\Phi(\mathbb{E}_\pi Kf) = \Phi(\mathbb{E}_\pi f)$. So, to prove the statement it is enough to show that

$$\mathbb{E}_\pi[\phi \circ Kf] \le \mathbb{E}_\pi[\phi \circ f]$$

First by non-negativity of D, for any x, we have

$$0 \le \mathrm{D}_{K(x,.)}^\Phi(f) = \mathbb{E}_{K(x,.)}\, \phi \circ f - \phi(\mathbb{E}_{K(x,.)}\, f)$$

Averaging out w.r.t., π we get

$$\begin{aligned} 0 &\leq \mathbb{E}_{x\sim\pi}\,\mathbb{E}_{y\sim K(x,.)}\,\phi\circ f(y) - \mathbb{E}_{x\sim\pi}\,\phi(\mathbb{E}_{K(x,.)}\,f) \\ &= \mathbb{E}_{y\sim\pi}\,\phi\circ f(y) - \mathbb{E}_{x\sim\pi}\,\phi(Kf(x)) \\ &= \mathbb{E}_{\pi}[\phi\circ f] - \mathbb{E}_{\pi}[\phi\circ Kf] \end{aligned}$$

as desired. □

We say that a Markov Kernel K exhibits a contraction if for any non-negative function $f : \Omega \to \mathbb{R}_{\geq 0}$,

$$\mathrm{D}_{\pi}^{\Phi}(Kf) \leq (1-\epsilon)\,\mathrm{D}_{\pi}^{\Phi}(f).$$

It follows that if K exhibits a $(1-\epsilon)$ contradiction w.r.t. Φ, then for any t and any state x,

$$\mathrm{D}_{\pi}^{\Phi}\left(K^t\frac{1_x}{\pi(x)}\right) \leq (1-\epsilon)^t\,\mathrm{D}_{\pi}^{\Phi}(\frac{1_x}{\pi(x)})$$

Lemma 5.9. *If K exhibits $1-\epsilon$ contraction w.r.t., any of the Variance, Entropy or linear distance functions, then K mixes in*

$$\frac{\log(\mathrm{D}_{\pi}^{\Phi}(1_x/\pi(x))}{\epsilon}$$

Proof. Let $g = K^t\frac{1_x}{\pi(x)}$. Then, by reversibility for any y,

$$g(y) = K^t\frac{1_x}{\pi(x)}(y) = \frac{K^t(y,x)}{\pi(x)} = \frac{K^t(x,y)}{\pi(y)}$$

Therefore,

$$\begin{aligned} \|K^t(x,.)-\pi\|_{TV} &\lesssim O(1)\sqrt{\mathrm{D}_{\pi}^{\Phi}(\frac{K^t(x,.)}{\pi})} \\ &= \sqrt{\mathrm{D}_{\pi}^{\Phi}(g)} \\ &= \sqrt{\mathrm{D}_{\pi}^{\Phi}(K^t f)} \\ &\leq (1-\epsilon)^{t/2}\sqrt{\mathrm{D}_{\pi}^{\Phi}(\frac{1_x}{\pi(x)})} \leq 1/2 \end{aligned}$$

The last inequality follows by letting $t = \frac{2}{\epsilon}\ln \mathrm{D}_{\pi}^{\Phi}(\frac{1_x}{\pi(x)})$. □

Notice that

$$\begin{aligned} \mathrm{D}^{\frac{1}{2}|x-1|}(1_x) &= \frac{1}{2}|1-\pi(x)| \\ \mathrm{D}^{x^2}(1_x) &= (\frac{1}{\pi(x)}-1) \\ \mathrm{D}^{x\log x}(1_x) &= \log\frac{1}{\pi(x)} \end{aligned}$$

This implies that we get

$$\frac{\log 1/\pi(x)}{\epsilon}$$

-mixing time with a variance contraction whereas a $\frac{\log\log 1/\pi(x)}{\epsilon}$-mixing time with an entropy contraction.

5.5 Dirichlet Form and Poincaré Constant

Definition 5.10. *The Dirichlet form of the Markov Kernel K with respect to two functions f, g is defined as follows:*

$$\mathcal{E}_K(f,g) = \frac{1}{2}\langle (I-K)g, f\rangle_\pi = \mathbb{E}_{x\sim\pi}\sum_y K(x,y)(f(x)-f(y))(g(x)-g(y)).$$

The matrix $I-K$ is the well-known *normalized Laplacian matrix.* In particular, since all eigenvalues of K are in the range $[-1,+1]$ it follows that the eigenvalues of $I-K$ are in the range $[0,2]$ and thus $I-K$ is a PSD matrix. This can be seen immediately by writing the quadratic form w.r.t. an arbitrary function f. In particular, in the special case that $f=g$ the above equation is

$$\mathcal{E}_P(f,f) = \frac{1}{2}\,\mathbb{E}_{x\sim\pi}\sum_y K(x,y)(f(x)-f(y))^2. \tag{5.1}$$

This Dirichlet form is also called the *Local Variance*, capturing how much the function f squared varies along edges.

Some simple facts about the Dirichlet forms:

1. $\mathcal{E}_K(f,f)\geq 0$.
2. For any constant c, we have $\mathcal{E}_K(cf,cf) = c^2\mathcal{E}_K(f,f)$.
3. For any constant c, we have $\mathcal{E}_K(f+c,f+c) = cE_K(f,f)$.

(The same facts hold for variance as well.)

The Poincaré constant of K is defined as

$$\lambda(K) := \inf_{f\geq 0}\frac{\mathcal{E}_K(f,f)}{\mathrm{Var}(f)}$$

Note that the smallest eigenvalue of $I-K$, namely $\lambda_1(I-K)=0$ as $(I-K)\mathbf{1}=0$. So, by the variational characterization of eigenvalues, the second smallest eigenvalue satisfies

$$\lambda_2(I-K) = \min_{f:\langle f,\mathbf{1}\rangle=0}\frac{\langle (I-K)f,f\rangle}{\langle f,f\rangle} = \min_{f\neq\text{constant}}\frac{\langle (I-K)f,f\rangle}{\mathrm{Var}(f)} = \min_{f\neq\text{constant}}\frac{\mathcal{E}_K(f,f)}{\mathrm{Var}(f)} \tag{5.2}$$

Here the middle equality follows from the fact that the Dirichlet form is invariant over shift; so we can replace a (non-constant f) with $f-\mathbb{E}f$. It thus follows that $\lambda(K)=\lambda_2(I-K)$; this simply follows from the fact that both Var and $\mathcal{E}_K$ are shift-invariant, so one can assume $f\geq 0$.

It is also instructive to compare with the contraction inequalities:

Lemma 5.11. *For any function $f:V\to\mathbb{R}$, $\mathrm{Var}(f)-\mathrm{Var}(Kf)=\mathcal{E}_{K^2}(f,f)$. Consequently, for $\Phi=x^2$,*

$$\inf_{f\neq constant} 1-\frac{\mathrm{D}^\Phi_\pi(Kf)}{\mathrm{D}^\Phi_\pi(f)} = \inf_{f\neq constant} 1-\frac{\mathrm{Var}(Kf)}{\mathrm{Var}(f)} = \inf_{f\neq constant}\frac{\mathcal{E}_{K^2}(f,f)}{\mathrm{Var}(f)} = \lambda_2(I-K^2) = .$$

Proof. We can write

$$\mathrm{Var}(Kf) = \langle Kf,Kf\rangle - (\mathbb{E}\,Kf)^2 = \langle K^2f,f\rangle - \langle Kf,\mathbf{1}\rangle^2 = \langle K^2f,f\rangle - (\mathbb{E}\,f)^2$$

(since $\langle Kf, \mathbf{1}\rangle = \langle f, K\mathbf{1}\rangle = \langle f, \mathbf{1}\rangle$.) Therefore,

$$\mathrm{Var}(f) - \mathrm{Var}(Kf) = \langle f, f\rangle - \langle K^2 f, f\rangle = \langle (I - K^2) f, f\rangle = \mathcal{E}_{K^2}(f, f).$$

□

In short, note that Poincaré constant is slightly different from the contraction ratio that we studied in the previous section, but they are closely related as one relates to eigenvalues of $I - K$ and the other to eigenvalues of $I - K^2$.

5.6 MLS constants and the Continuous time chain

Let $T_1, T_2, \dots$ be a sequence of independent and identically distributed exponential random variables of rate 1. That is, each T_i takes values in $[0, \infty)$ and has an exponential distribution

$$\mathbb{P}[T_i \le t] = \begin{cases} 1 - e^{-t} & \text{if } t \ge 0, \\ 0 & \text{o.w.} \end{cases}$$

T_i's determine times of transitions of the chain. We start with Y_0 to denote the starting state of our chain; At transition times $T_1 + \dots T_k$ for any $k \ge 1$ we make a jump (according to the discrete chain). Let $S_k = T_1 + \dots + T_k$. In other words if $X_0 = Y_0$ and $X_1, \dots,$ is the sequence of steps of our discrete chain then for any $t \ge 0$,

$$Y_t = X_k \text{ for } t \in [S_k, S_{k+1}).$$

Let Y_0 be distributed according ν. Let N_t be a Poisson random variable with rate t, then

$$Y_t \sim \sum_x \nu(x) \sum_{k=0}^{\infty} \mathbb{P}[N_t = k]\, K^k(x, .)$$

Therefore, the transition probability of the continuous chain H_t is

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{P}[N_t = k]\, K^k &= \sum_{k=0}^{\infty} \frac{e^{-t} t^k}{k!} K^k \\ &= e^{-t} \sum_{k=0}^{\infty} \frac{(tK)^k}{k!} \\ &= \exp(-t(I - K)) =: H_t \end{aligned}$$

The contraction of the Entropy, i.e., $\mathrm{D}_\pi^{x \log x}$ is often called the **Modified log-Sobolev constant**:

$$\xi(K) := \inf_{f \ge 0} \frac{\mathcal{E}_K(f, \log f)}{\mathrm{D}_\pi^{x \log x}(f)}$$

Note that in the numerator we naturally want to write $\mathrm{D}^{x \log x}(Kf)$. The Dirichlet form naturally appears if we instead the rate of decrease of the entropy with respect to the continuous time chain (as opposed to the discrete chain).

Lemma 5.12. *For every distribution f on Ω we have,*

$$\frac{d}{dt} \mathrm{D}_\pi(K^t f) = -\mathcal{E}_K(f_t, \log f_t).$$

Proof. Let $f_t = H_t f$. Then, as usual $\mathbb{E} f_t = \langle 1, H_t f\rangle = 1$. So, $\mathrm{D}(f_t) = \mathbb{E} f_t \log f_t$. It follows that,

$$\begin{aligned}\frac{d}{dt}\mathrm{D}_\pi(f_t) &= \sum_x \pi(x)\frac{d}{dt}H_t f(x)\log H_t f(x)\\ &= -\sum_x \pi(x)(I-K)H_t f(x)\cdot \log H_t f(x) - \sum_x \pi(x)(I-K)f(x)\\ &= -\sum_x \pi(x)(I-K)H_t f(x)\cdot \log H_t f(x) = -\mathcal{E}_K(f_t, \log f_t).\end{aligned}$$

The second identity uses that

$$\frac{d}{dt}H_t = -e^{-t}\sum_{k=0}^{\infty}\frac{(tK)^k}{k!} + e^{-t}\sum_{k=1}^{\infty}K\cdot\frac{(tK)^{k-1}}{(k-1)!} = -(I-K)H_t$$

The third identity uses that $\langle 1, (I-K)f\rangle = \langle (I-K)1, f\rangle = 0$ as $(I-K)1 = 0$. □

We don't write the details but one can similarly bound the mixing time of the continuous time chain uses the MLS constant and the contraction of entropy.

We conclude this section by recalling a classical theorem which relates the mixing time of the continuous and discrete time chains.

Theorem 5.13 (Thm 20.3 Peres book)**.** *Let K be a half-lazy (not necessarily reversible) Markov chain. For any k and state x we have,*

$$\left\|K^{4k}(x,.) - \pi\right\|_{TV} \le \left\|e^{-k(I-K)}(x,.) - \pi\right\|_{TV} + \epsilon_k$$

where $\epsilon_k \to 0$ as $k \to \infty$.

Lemma 5.14 (Bobkov-Tetali'03)**.** *For every reversible Markov chain, we have $\xi(K) \le O(\lambda(K))$.*

So, in principal MLSI constant may not necessarily give the faster mixing time if is much smaller than $\lambda(K)$.

5.7 Concentration Inequalities

For a function $f : \Omega \to \mathbb{R}$, let

$$v(f) := \max_{x\in\Omega}\sum_{y\in\Omega} K(x,y)(f(x)-f(y))^2$$

This function quantifies how Lipschitz f is w.r.t. the underlying graph of K. For example, if f is 1-Lipschitz, i.e,. $|f(x) - f(y)| \le 1$ for all x, y with $K(x,y) \neq 0$, then $v(f) \le 1$.

Theorem 5.15 (Modified Log-Sobolev Implies Concentration)**.** *If K is reversible, then for any $\epsilon \ge 0$ and $f : \Omega \to \mathbb{R}$, we have*

$$\mathbb{P}\left[|f(x) - \mathbb{E}_\pi f| \ge \epsilon\right] \le 2\exp\left(-\frac{\xi(K)\epsilon^2}{2v(f)}\right)$$

where $\xi(K)$ is the MLS of K.

Before proving this let us explain some applications for this inequality: Let Ω be the space of all spanning trees of a given graph G and π be the uniform distribution. Consider the following neighborhood structure:

two spanning trees T, T' are neighbors if $|T \oplus T'| = 2$, i.e., if they differ in exactly two edges. In future lectures we will see a metropolis filter supported on this neighborhood structure with MLS constant $\xi(K) \geq 1/n$ (the statement holds more generally for matroids). This implies concentration inequality for any Lipschitz function of spanning trees. For example, let $f(T)$ be the number of even degree vertices in T. Observe that for any two neighbor trees T, T', $|f(T) - f(T')| \leq 4$. So, $v(f) \leq 4$. This implies that

$$\mathbb{P}\left[|f(T) - \mathbb{E} f| \geq c\sqrt{n}\right] \leq 2\exp(-c^2/8),$$

i.e., the number of even degree vertices in a random spanning tree is tightly concentrated around its expectation. Such events has been used in analyzing algorithms for TSP.

Let me explain the main idea of the proof:

Definition 5.16. *A random variable $X \in \mathbb{R}$ is said to be sub-Gaussian with variance proxy α^2 if its moment generating function satisfis*

$$\mathbb{E}\, e^{t(X - \mathbb{E} X)} \leq e^{t^2\alpha^2/2}$$

for all $t \in \mathbb{R}$.

It follows that if X is sub-Gaussian then $\mathbb{P}[X - \mathbb{E} X \geq \epsilon] \leq e^{-\frac{\epsilon^2}{2\sigma^2}}$. So, to prove the theorem it is enough to show that the random variable $f(x)$ is sub-Gaussian with $\sigma^2 = \frac{\xi(K)}{2v(f)}$

The proof has two steps: The first step is the Herbst argument which says the following:

Lemma 5.17 (Herbst argument)**.** *Let X be a random variable such that for all $t \geq 0$,*

$$D^{x \log x}(e^{tX}) \leq \frac{t^2\xi^2}{2}\,\mathbb{E}[e^{tX}]$$

Then, for all $t \geq 0$

$$\log \mathbb{E} \exp(t(X - \mathbb{E}[X])) \leq \frac{1}{2}t^2\xi^2$$

i.e., X is sub-Guassian.

Then, we observe that

$$\frac{D(e^{tf})}{t^2\,\mathbb{E}\, e^{tf}} \leq \frac{\mathbb{E}_K(tf, e^{tf})}{t^2\xi(K)\,\mathbb{E}\, e^{tf}}$$

The last step is to use (5.3) to bound the RHS by $\frac{v(f)}{2\xi(K)}$ as desired. Below we give a self-contained proof in case you are interested.

We start by bounding the expectation of the moment generating function.

Proposition 5.18. *For any $t \geq 0$,*

$$\mathbb{E}_\pi\, e^{tf} \leq \exp\left(t \cdot \mathbb{E}_\pi f + t^2\frac{v(f)}{2\xi(K)}\right)$$

Proof of Theorem 5.15.

$$\begin{aligned}
\mathbb{P}[f(x) > \mathbb{E}_\pi f + \epsilon] &= \mathbb{P}\left[e^{tf(x)} > e^{t\,\mathbb{E} f + t\epsilon}\right] \\
&\leq \frac{\mathbb{E}\, e^{tf}}{e^{t\,\mathbb{E} f + t\epsilon}} && \text{(Markov's Inequality)} \\
&\leq \exp\left(t^2\frac{v(f)}{2\xi(K)} - t\epsilon\right) \leq \exp\left(\frac{\xi(K)\epsilon^2}{2v(f)}\right) && \text{(Proposition 5.18)}
\end{aligned}$$

where the last inequality follows by letting $t = \frac{\xi(K)\epsilon}{v(f)}$ □

It remains to prove Proposition 5.18.

Lemma 5.19. *For any $t \geq 0$,*

$$\frac{d}{dt}\frac{\log \mathbb{E}_\pi e^{tf}}{t} \leq \frac{v(f)}{2\xi(K)}$$

Having this to prove the proposition it is enough to integrate from 0 to t we get

$$\log \mathbb{E}_\pi e^{tf} - \lim_{s\to 0}\frac{\log \mathbb{E}_\pi e^{sf}}{s} \leq \frac{tv(f)}{2\xi(K)}$$

where by the Hopital rule, the above limit is exactly $\mathbb{E}_\pi f$. Re-arranging we get

$$\frac{\log \mathbb{E}_\pi e^{tf}}{t} \leq \frac{tv(f)}{2\xi(K)} + \mathbb{E} f$$

So,

$$\mathbb{E}_\pi e^{tf} \leq \exp\left(\frac{t^2 v(f)}{2\xi(K)} + t\,\mathbb{E} f\right)$$

which finishes the proof of Proposition 5.18.

Proof. Proof of Lemma 5.19

$$\begin{aligned}
\frac{d}{dt}\frac{\log \mathbb{E}_\pi e^{tf}}{t} &= \frac{\mathbb{E}_\pi[fe^{tf}]}{t\,\mathbb{E}_\pi e^{tf}} - \frac{\log \mathbb{E}_\pi e^{tf}}{t^2} \\
&= \frac{\mathbb{E}_\pi[tfe^{tf}] - \mathbb{E}_\pi e^{tf} \cdot \log \mathbb{E}_\pi e^{tf}}{t^2\,\mathbb{E} e^{tf}} \\
&= \frac{\mathrm{D}^\Phi(e^{tf})}{t^2\,\mathbb{E} e^{tf}} \leq \frac{\mathcal{E}_K(tf, e^{tf})}{\xi(K)t^2\,\mathbb{E} e^{tf}}
\end{aligned}$$

So, to prove the lemma it is enough to show that $\frac{2}{t}\mathcal{E}_K(tf, e^{tf}) \leq t \cdot \mathbb{E} e^{tf} v(f)$

$$\begin{aligned}
\frac{2}{t}\mathcal{E}_K(tf, e^{tf}) &= \mathbb{E}_{x\sim\pi}\sum_y K(x,y)(f(x)-f(y))(e^{tf(x)} - e^{tf(y)}) \\
&= \mathbb{E}_{x\sim\pi}\sum_y K(x,y)(f(x)-f(y))^2 \cdot \frac{e^{tf(x)} - e^{tf(y)}}{f(x)-f(y)} \\
&\leq \mathbb{E}_{x\sim\pi}\left(\sum_y K(x,y)(f(x)-f(y))^2\right) \cdot \max_{y\in\Omega}\frac{e^{tf(x)} - e^{tf(y)}}{f(x)-f(y)} \\
&\leq v(f)\,\mathbb{E}_{x\sim\pi} e^{tf(x)} \cdot \max_y \frac{1 - e^{-(tf(x)-tf(y))}}{f(x)-f(y)} \\
&\leq t \cdot v(f) \cdot \mathbb{E} e^{tf} \cdot \max_{z\in\mathbb{R}} \frac{1-e^{-z}}{z} && \text{(renaming } z = t(f(x)-f(y))\text{)} \\
&\leq t \cdot v(f) \cdot \mathbb{E} e^{tf} && (5.3)
\end{aligned}$$

The last equation uses that $1 - z \leq e^{-z}$ for all $z \in \mathbb{R}$. $\square$

5.8 Cheeger's Inequality

Given a graph $G = (V, E)$ with a the MC kernel K. For a set $S \subseteq V$ define

$$\phi(S) = \frac{\mathcal{E}_K(\mathbf{1}_S, \mathbf{1}_S)}{\pi_0(S)} = \frac{\frac{1}{2}\pi_1(E(S, \overline{S}))}{\pi_0(S)} = \frac{\frac{1}{2}\mathbb{E}_{u\sim\pi_0}\mathbb{E}_{\{u,v\}|u}\mathbb{P}\left[|\{u,v\}\cap S| = 1\right]}{\pi_0(S)} = \mathbb{E}_{u\sim\pi_0(S)}\mathbb{P}_{\{u,v\}|u}\left[v \notin S\right]$$

where $E(S, \overline{S})$ is the set of edges in the cut $(S, \overline{S})$. In other words, $\phi(S)$ is the probability that a walk started at a vertex of S chosen with probability proportional $\pi_0(.)$ leaves S in one step.

Lemma 5.20 (Cheeger's Inequality)**.** *Given a graph $G = (V, E)$, a set $S \subseteq V$ with $\pi_0(S) \leq 1/2$. Then,*

$$\frac{1}{2}\min_{f:S\to\mathbb{R}_{\geq 0}}\frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)} \leq \min_{T\subseteq S}\phi(T) \leq \min_{f:S\to\mathbb{R}_{\geq 0}}\sqrt{\frac{2\mathcal{E}(f,f)}{\mathrm{Var}(f)}}$$

Proof. First, we prove the left side. Fix a set $T \subseteq S$ whit minimum conductance. Let $f = \mathbf{1}_T$. Then, since $\pi_0(T) \leq \pi_0(S) \leq 1/2$,

$$\mathrm{Var}(\mathbf{1}_T) = \pi_0(T) - \pi_0(T)^2 \geq \pi_0(T)/2.$$

This proves the left inequality.

Next, we prove the harder direction. Fix a non-zero function $f : S \to \mathbb{R}_{\geq 0}$, we find a set $T \subseteq \mathrm{supp}(f)$ such that

$$\phi(T) \leq 2\sqrt{\frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)}}$$

Perhaps after renormalization, assume $f \leq 1$. For a threshold $t \geq 0$, define $S_t = \{v : f^2(v) \geq t\}$. Choose a threshold $t \sim [0, 1]$ uniformly at random. Then,

$$\begin{aligned}
\mathbb{E}_t\,\pi_1(E(S_t, \overline{S_t})) &= \mathbb{E}_t\,\mathbb{E}_{\{u,v\}\sim\pi_1}\mathbb{P}\left[(f^2(u) < t \wedge f^2(v) > t) \vee (f^2(v) < t \wedge f^2(u) > t)\right] \\
&= \frac{1}{2}\mathbb{E}_{\{u,v\}\sim\pi_1}|f^2(u) - f^2(v)| \\
&\leq \frac{1}{2}\mathbb{E}_{\{u,v\}\sim\pi_1}|f(u) - f(v)| \cdot |f(u) + f(v)| \\
&\underset{Cauchy-Schwarz}{\leq} \sqrt{\frac{1}{2}\mathbb{E}_{\{u,v\}\sim\pi_1}(f(u) - f(v))^2} \cdot \sqrt{\frac{1}{2}\mathbb{E}_{\{u,v\}\sim\pi_1}(f(u) + f(v))^2} \\
&\leq \sqrt{\mathcal{E}(f,f)} \cdot \sqrt{\mathbb{E}_{\{u,v\}\sim\pi_1} f(u)^2 + f(v)^2}
\end{aligned}$$

Furthermore, notice

$$\mathbb{E}_{\{u,v\}\sim\pi_1} f(u)^2 + f(v)^2 = \mathbb{E}_{\{u,v\}\sim\pi_1}\mathbb{E}_{u|\{u,v\}} 2f(u)^2 = \mathbb{E}_{u\sim\pi_0}\mathbb{E}_{\{u,v\}|u} 2f(u)^2 = 2\,\mathbb{E} f^2.$$

On the other hand,

$$\mathbb{E}_t\,\pi_0(S_t) = \mathbb{E}_t\,\mathbb{E}_{u\sim\pi_0}\mathbb{P}\left[t < f(u)^2\right] = \mathbb{E} f(u)^2.$$

Putting these together there must exist a value of t, say t^* such that

$$\phi(S_{t^*}) \leq \frac{\mathbb{E}_t\,\pi_1(E(S_t, \overline{S_t})}{\mathbb{E}_t\,\pi_0(S_t)} \leq \frac{\sqrt{2\mathcal{E}(f,f)}}{\sqrt{\mathbb{E} f^2}} \leq \sqrt{\frac{2\mathcal{E}(f,f)}{\mathrm{Var}(f)}}$$

where the last inequality uses that $\mathrm{Var}(f) = \mathbb{E} f^2 - (\mathbb{E} f)^2 \leq \mathbb{E} f^2$. □