Chapter 23

Log-Sobolev Inequalities

We introduce log-Sobolev inequalities for analyzing mixing time of random walks, and see that they provide the optimal bound for the down-up walks on strongly log-concave distributions.

23.1 Analyzing Mixing Time Using Log-Sobolev Inequalities

In this section, we first define variance and relative entropy, and then define spectral gap and log-Sobolev constants, and then see their uses in bounding mixing time, and finally some intuition about these definitions. The presentation in this section is based on [CGM21, AJK+21, BT06, MT06].

Variance and Entropy

In Chapter 6, when we analyze the mixing time of random walks, we upper bound the total variation distance $d_{\mathrm{TV}}(p_t, \pi)$ by upper bounding $\|D^{-1/2}(p_t - \pi)\|_2$ (see Problem 6.20). This can be understood as bounding the variance of $f := p_t/\pi$, the density of p_t with respect to π at time $t \geq 0$.

Definition 23.1 (π-Variance)**.** *Let $f : [n] \to \mathbb{R}$ be a function and π be a probability distribution on $[n]$. The variance of f with respect to π is defined as*

$$\mathrm{Var}_\pi[f] := \mathbb{E}_\pi[f^2] - \left(\mathbb{E}_\pi[f]\right)^2,$$

where $\mathbb{E}_\pi[f] = \sum_{i \in [n]} \pi(i) f(i)$.

There are other ways to measure the closeness of two probability distributions. A well-known measure is the relative entropy between the two distributions.

Definition 23.2 (KL-Divergence)**.** *Let p and q be probability distributions on $[n]$ such that $q(i) = 0$ implies $p(i) = 0$ for $1 \leq i \leq n$. The Kullback-Liebler divergence, or relative entropy, between p and q is defined as*

$$\mathcal{D}_{\mathrm{KL}}(p \parallel q) = \sum_{i=1}^{n} p(i) \log \frac{p(i)}{q(i)},$$

where we follow the convention that $0 \log 0 = 0$. Check that $\mathcal{D}_{\mathrm{KL}}(p, q) \geq 0$ by Jensen's inequality.

Pinsker's inequality shows that KL-divergence can be used to upper bound the total variation distance.

Theorem 23.3 (Pinsker's Inequality)**.** *For any two probability distributions p, q on $[n]$,*

$$d_{\mathrm{TV}}(p,q)^2 \leq 2\mathcal{D}_{\mathrm{KL}}(p \parallel q).$$

So, to analyze mixing time, it is also natural to consider the relative entropy betwen p_t and π.

Definition 23.4 (π-Entropy)**.** *Let $f : [n] \to \mathbb{R}$ be a function and π be a probability distributions on $[n]$. Define*

$$\mathrm{Ent}_\pi[f] := \mathbb{E}_\pi[f \log f] - \mathbb{E}_\pi\big[f \log(\mathbb{E}_\pi[f])\big].$$

Check that $\mathrm{Ent}_\pi[\frac{p}{\pi}] = \mathcal{D}_{\mathrm{KL}}(p \parallel \pi)$ for a probability distribution p.

Spectral Gap and Log-Sobolev Constants

We only consider reversible Markov chains in this course, which include transition matrices of random walks on undirected graphs.

Definition 23.5 (Reversible Markov Chain)**.** *Let $P \in \mathbb{R}^{n \times n}$ be the transition matrix of a Markov chain whose stationary distribution is π. We say P is reversible if for all $i, j \in [n]$,*

$$\pi(i) \cdot P(i,j) = \pi(j) \cdot P(j,i)$$

Let $\Pi := \mathrm{diag}(\pi)$. Then the reversible condition can be stated as ΠP being a symmetric matrix.

The following definition should be understood as the quadratic form of the Laplacian matrix when we consider random walks on undirected graphs.

Definition 23.6 (Dirichlet Form)**.** *Let $P \in \mathbb{R}^{n \times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. For two vectors $f, g \in \mathbb{R}^n$, the Dirichlet form is defined as*

$$\mathcal{E}_P(f,g) := \langle (I-P)f, g\rangle_\pi = g^T(\Pi - \Pi P)f = \frac{1}{2}\sum_{1 \leq i,j \leq n} \pi(i) \cdot P(i,j) \cdot \big(g(i) - g(j)\big) \cdot \big(f(i) - f(j)\big),$$

where $\Pi := \mathrm{diag}(\pi)$. The last equality can be seen by thinking of $\Pi - \Pi P$ as the Laplacian matrix of the underlying undirected graph with an edge weight $\Pi(i) \cdot P(i,j)$ for each pair $i, j \in [n]$.

Remark 23.7 (Laplacian?)**.** *The matrix $I - P$ is often called the Laplacian in the literature, but we resist not to do so as it is not consistent with our convention (e.g. $I - P$ may not be symmetric).*

For random walks on an undirected graph G, the adjacency matrix of G is ΠP, and note that $I - P$ has the same spectrum as the normalized *Laplacian matrix of G.*

The Dirichlet form is sometimes called the energy of the function f, which can be thought of as a measure of the *local* variation of f along the edges of the underlying graph. On the other hand, the variance in Definition 23.1 can be thought of as a measure of the *global* variation of f. Then the spectral gap can be interpreted as a lower bound on the local variance by the global variance, and this perspective is useful in designing approximation algorithms on expander graphs (e.g. [BRS11]).

Definition 23.8 (Variational Characterization of Spectral Gap)**.** *Let $P \in \mathbb{R}^{n \times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Define*

$$\lambda(P) := \inf\left\{ \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)} \;\middle|\; f : [n] \to \mathbb{R},\ \mathrm{Var}_\pi[f] \neq 0 \right\}.$$

Exercise 23.9. *Show that* $\lambda(P) = 1 - \alpha_2(P)$ *where* $\alpha_2(P)$ *is the second largest eigenvalue of* P*.*

We note that the spectral gap is sometimes called the Poincaré constant. The log-Sobolev constant replaces the variance of f in the denominator of the spectral gap by the π-entropy of f in Definition 23.4.

Definition 23.10 (Log-Sobolev Constant [DSC96])**.** *Let* $P \in \mathbb{R}^{n\times n}$ *be the transition matrix of a reversible Markov chain whose stationary distribution is* π*. Define the log-Sobolev constant of* P *as*

$$\alpha(P) := \inf\left\{ \frac{\mathcal{E}_P(f,f)}{\mathrm{Ent}_\pi(f^2)} \;\middle|\; f : [n] \to \mathbb{R}_{\geq 0},\ \mathrm{Ent}_\pi[f^2] \neq 0 \right\}.$$

The modified log-Sobolev constant is introduced by Bobkov and Tetali [BT06].

Definition 23.11 (Modified Log-Sobolev Constant [BT06])**.** *Let* $P \in \mathbb{R}^{n\times n}$ *be the transition matrix of a reversible Markov chain whose stationary distribution is* π*. Define the modified log-Sobolev constant of* P *as*

$$\rho(P) := \inf\left\{ \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)} \;\middle|\; f : [n] \to \mathbb{R}_{\geq 0},\ \mathrm{Ent}_\pi[f] \neq 0 \right\}.$$

These definitions may not look intuitive. In the following, we will first state the results of using spectral gap and log-Sobolev constants to bounding mixing time, and then we will provide some intuitions about these definitions.

Bounding Mixing Time by Log-Sobolev Constants

We have already seen in Chapter 6 that we can upper bound the mixing time by lower bounding the spectral gap. The following is a generalization of Theorem 6.16.

Theorem 23.12 (Mixing Time by Spectral Gap)**.** *Let* $P \in \mathbb{R}^{n\times n}$ *be the transition matrix of a reversible Markov chain whose stationary distribution is* π*. The* ϵ*-mixing time in Definition 6.15 is*

$$\tau_\epsilon(P) \lesssim \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\epsilon}\right),$$

where $\pi_{\min} := \min_{i\in[n]} \pi(i)$ *(which is* $\frac{1}{n}$ *when* π *is the uniform distribution).*

The significance of the log-Sobolov constant is a much better dependence on $1/\pi_{\min}$. The following result is proved by Diaconis and Saloff-Coste [DSC96].

Theorem 23.13 (Mixing Time by Log-Sobolev Constant [DSC96])**.** *Let* $P \in \mathbb{R}^{n\times n}$ *be the transition matrix of a reversible Markov chain whose stationary distribution is* π*. Then*

$$\tau_\epsilon(P) \lesssim \frac{1}{\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{\epsilon}\right).$$

Bobkov and Tetali [BT06] proved a similar result for modified log-Sobolev constant.

Theorem 23.14 (Mixing Time by Modified Log-Sobolev Constant [BT06]). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Then*

$$\tau_\epsilon(P) \lesssim \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{\epsilon}\right).$$

Bobkov and Tetali also proved that

$$2\lambda(P) \geq \rho(P) \geq 4\alpha(P),$$

and so the lower bounds on these constants are increasingly difficult to obtain. The modified log-Sobolev constant has the advantage that it provides the same upper bound on the mixing time, while it is always at least as large as the log-Sobolev constant.

Intuition from Continuous Time Random Walks

The definitions of the spectral gap and the modified log-Sobolev constant come quite naturally from continuous time random walks. We would not be able to introduce continuous time random walks properly, so we just state the definition.

Definition 23.15 (Continuous Time Random Walks). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. For any $t \geq 0$, the transition matrix, or the heat kernel, is defined as*

$$H_t = e^{-t(I-P)} = \sum_{k=0}^{\infty}\frac{t^k(P-I)^k}{k!}.$$

Let $p_0 \in \mathbb{R}^n$ be an initial distribution. Then $p_t^T = p_0^T H_t$ is the distribution at time t.

As discussed earlier, we will consider $f_t := p_t/\pi$ and keep track of how fast it converges to $\vec{1}$.

Exercise 23.16 (Change of Density). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Let $f_t(i) = p_t(i)/\pi(i)$ for all $i \in [n]$ be the density of p_t with respect to π at time $t \geq 0$. For any initial distribution p_0 and all $t \geq 0$, show that*

$$f_t = H_t f_0.$$

Furthermore, for any $i \in [n]$, show that

$$\frac{df_t(i)}{dt} = \big((P-I)f_t\big)(i).$$

It turns out that the change of variance is exactly the Dirichlet form.

Lemma 23.17 (Change of Variance). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Let $f_t = p_t/\pi$. Then*

$$\frac{d}{dt}\mathrm{Var}_\pi(f_t) = -2\mathcal{E}_P(f_t, f_t).$$

Proof. Note that $\mathrm{Var}_\pi(f_t) = \mathbb{E}_\pi[f_t^2] - 1$ by Definition 23.1, and so

$$\frac{d}{dt}\mathrm{Var}_\pi(f_t) = \sum_{i=1}^{n} \pi(i) \cdot \frac{d}{dt} f_t(i)^2 = 2\sum_{i=1}^{n} \pi(i) \cdot f_t(i) \cdot \big((P-I)f_t\big)(i) = -2\mathcal{E}_P(f_t, f_t).$$

□

So we can understand the spectral gap in Definition 23.8 is *defined* to ensure that

$$\frac{d}{dt}\mathrm{Var}_\pi(f_t) = -2\mathcal{E}_P(f_t, f_t) \leq -2\lambda(P) \cdot \mathrm{Var}_\pi(f_t) \quad\implies\quad \frac{d}{dt}\log\big(\mathrm{Var}_\pi(f_t)\big) \leq -2\lambda(P).$$

By integrating on both sides, we see that the variance is exponentially decreasing as

$$\log\big(\mathrm{Var}_\pi(f_t)\big) - \log\big(\mathrm{Var}_\pi(f_0)\big) \leq -2\lambda(P) \cdot t \quad\implies\quad \mathrm{Var}_\pi(f_t) \leq \mathrm{Var}_\pi(f_0) \cdot e^{-2\lambda(P)\cdot t}.$$

Note that the initial variance $\mathrm{Var}_\pi(f_0) \leq 1/\pi_{\min}$, and this implies Theorem 23.12 for continuous time random walks. This argument can be adapted for discrete time random walks. This is a good exercise to work out; see [MT06] for a solution.

Bobkov and Tetali used the same logic to define the modified log-Sobolev constant, by using relative entropy in place of variance.

Lemma 23.18 (Change of Relative Entropy)**.** *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Let $f_t = p_t/\pi$. Then*

$$\frac{d}{dt}\mathrm{Ent}_\pi(f_t) = -\mathcal{E}_P(f_t, \log f_t).$$

Proof. Note that $\mathbb{E}_\pi[f] = 1$ and thus $\mathrm{Ent}_\pi[f] = \mathbb{E}_\pi[f\log f]$ by Definition 23.4, and hence

$$\frac{d}{dt}\mathrm{Ent}_\pi(f_t) = \sum_{i=1}^{n} \pi(i) \cdot \frac{d}{dt} f_t(i)\log f_t(i) = \sum_{i=1}^{n} \pi(i) \cdot \big(1 + \log f_t(i)\big) \cdot \big((P-I)f_t\big)(i) = -\mathcal{E}_P(f_t, \log f_t),$$

where in the last equality we use that $\sum_{i=1}^{n} \pi(i)((P-I)f_t)(i) = \langle \pi, (P-I)f_t\rangle = 0$. □

So the modified log-Sobolev constant is *defined* to ensure that

$$\frac{d}{dt}\mathrm{Ent}_\pi(f_t) = -\mathcal{E}_P(f_t, \log f_t) \leq -\rho(P) \cdot \mathrm{Ent}_\pi(f_t) \quad\implies\quad \mathrm{Ent}_\pi(f_t) \leq \mathrm{Ent}_\pi(f_0) \cdot e^{-\rho(P)\cdot t}.$$

Crucially, the initial relative entropy is

$$\mathrm{Ent}_\pi[f_0] = \sum_{i=1}^{n} p_0(i) \log\frac{p_0(i)}{\pi(i)} \leq \log\frac{1}{\pi_{\min}},$$

and this implies Theorem 23.14 for continuous time random walks by Pinsker's inequality in Theorem 23.3.

I have not found a general proof of Theorem 23.14 for discrete time random walks, and it was mentioned in [MT06] that "there seems to be no discrete-time analog" of it.

In the combinatorial applications that we will see, however, there are direct proofs of the exponential decreasing of the relative entropy, and thus the mixing time bound in Theorem 23.14 holds.

23.2 Log-Sobolev Constant for Strongly Log-Concave Distribution

It is already difficult to prove a lower bound on the spectral gap, and so there are very few known result on proving a lower bound on the log-Sobolev constants. This is starting to change after the resolution of the matroid expansion conjecture. Not only do the techniques from high dimensional expanders provide a direct way to establish a lower bound on the spectral gap, recent developments extend the techniques further to establish a lower bound on the modified log-Sobolev constant. The first result in this direction is by Cryan, Guo and Mousa [CGM21].

Theorem 23.19 (Modified Log-Sobolev Constant for Strongly Log-Concave Distribution [CGM21]). *Let μ be a d-homogeneous strongly log-concave distribution. Then the modified log-Sobolev constant of the down-up walk $P_d^{\triangledown}$ in Definition 20.5 is*

$$\rho(P_d^{\triangledown}) \geq \frac{1}{d}.$$

The proof in [CGM21] is very nice, but we will not present it here. Rather, we will present a recent generalization for fractionally log-concave distributions in the next chapter. We just note here that the proof in [CGM21] shows that the relative entropy is exponentially decreasing after one step of the down-up walk such that

$$\mathcal{D}_{\mathrm{KL}}(P_d^{\triangledown} p \parallel \pi) \leq \Big(1 - \frac{1}{d}\Big) \cdot \mathcal{D}_{\mathrm{KL}}(p \parallel \pi),$$

and using Pinsker's inequality as in the previous section gives the optimal mixing time analysis for the down-up walk on matroid bases.

Corollary 23.20 (Optimal Mixing Time for Sampling Matroid Bases [CGM21]). *The mixing time of the down-up walk in Chapter 20 for sampling uniform random matroid bases of size d is*

$$\tau_\epsilon(P_d^{\triangledown}) \lesssim d\Big(\log d + \log\log n + \log\frac{1}{\epsilon}\Big).$$

Near-Linear Time Algorithm for Random Spanning Trees

One immediate consequence of Corollary 23.20 is that the mixing time of the down-up walk for sampling uniform random spanning trees is at most $O(n \log n)$. To design a near-linear time algorithm, one needs to implement each iteration in the down-up walk efficiently, but it is not known how to do so.

Fortunately, the trick in [ALO+21] is to consider the down-up walk on the *dual* matroid. Given a graph $G = (V, E)$, the rank of the dual matroid is $|E| - |V| + 1 \leq |E|$, so the mixing time of the down-up walk on the dual matroid $O(|E| \log \frac{|E|}{\epsilon})$ by Corollary 23.20. The resulting algorithm is as follows. Let T_0 be an arbitrary spanning tree. In iteration $t \geq 0$, sample a uniform random edge $e \in E - T_t$, and then sample a uniform random edge f in the unique cycle in $T_t + e$ and set $T_{t+1} := T_t + e - f$, and repeat. This algorithm has been studied by Russo, Teixeira and Francisco, and they show that each iteration can be implemented in amortized $O(\log |E|)$ time using the cut-link trees data structures.

Theorem 23.21 (Near-Linear Time Algorithm for Sampling Random Spanning Trees [ALO+21]). *Given a graph $G = (V, E)$, there is an algorithm to sample a random spanning tree in G with distribution ϵ-close to the uniform distribution and running time $O(|E| \log |E| \log \frac{|E|}{\epsilon})$.*

The problem of designing a fast algorithm for sampling a uniform random spanning tree is well-studied. The previous best known algorithm is by Schild with almost-linear running time $O(m^{1+o(1)})$. This is based on a line of work that simulated another Markov chain for generating random spanning trees, using techniques from Laplacian solvers and electrical flows. The algorithm by Schild is very sophisicated and complicated, and so Theorem 23.21 is a dramatic simplification based on better analysis of mixing time.

Concentration Inequality for Strongly Log-Concave Distribution

One main application of log-Sobolev inequalities is to prove concentration inequalities [BLM13, VH14]. The following result is a consequence of Theorem 23.19 for strongly log-concave distributions.

Theorem 23.22 (Concentration of Strongly Log-Concave Distributions [CGM21])**.** *Let μ be a d-homogeneous strongly log-concave distribution with support $\Omega \subseteq \{0,1\}^n$. For any observable function $f : \Omega \to \mathbb{R}$ and $a \geq 0$,*

$$\Pr_{x\sim\mu}\big[|f(x) - \mathbb{E}_\pi f| \geq a\big] \leq 2\exp\Big(-\frac{a^2}{2d\cdot\nu(f)}\Big),$$

where $\nu(f)$ is the maximum of one-step variances

$$\nu(f) := \max_{x\in\Omega}\Big\{\sum_{y\in\Omega} P_d^{\triangledown}(x,y)\cdot\big(f(x)-f(y)\big)^2\Big\}.$$

The proof of Theorem 23.22 follows from the Herbst argument (see [BLM13]). The reader is referred to [CGM21] for the proof. For a c-Lipschitz function under the graph distance in the bases exchange graph, $\nu(f) \leq c^2$ and thus Theorem 23.22 generalizes the concentration result for strongly Rayleigh distributions in Theorem 16.18.

23.3 References

[AJK+21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: Modified log-sobolev inequalities for fractionally log-concave polynomials and the ising model. *CoRR*, abs/2106.04105, 2021. 217, 225, 226, 227

[ALO+21] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant, and Thuy-Duong Vuong. Log-concave polynomials IV: approximate exchange, tight mixing times, and near-optimal sampling of forests. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 408–420. ACM, 2021. 222

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013. 223

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 ieee 52nd annual symposium on foundations of computer science*, pages 472–481. IEEE, 2011. 218

[BT06] Sergey G Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006. 217, 219, 220

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-sobolev inequalities for strongly log-concave distributions. *Annals of Probability*, 49(1):506–525, 2021. 217, 222, 223

[DSC96] Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996. 219

[MT06] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006. 217, 221

[VH14] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014. 223