401-4634-24L: Diffusion Models, Sampling and Stochastic Localization

Lecture 2 – Geometric random walks

Lecturer: Yuansi Chen *Spring 2024*

Key concepts:

- Geometric random walks: ball walk and hit-and-run
- Markov chain and related notions
- Intuition about rapid mixing
- Conductance and isoperimetry
- Ball walk mixing time proof

The material of this lecture is mainly based on [Vem05].

2.1 Introduction

We study the fundamental problem of sampling from a uniform distribution on a convex body (bounded convex set with non-empty boundary) $\mathcal{K} \subseteq \mathbb{R}^n$. We use it as a running example for gaining intuitions about how sampling algorithms on continuous state-space and build up key proof techniques. To make the computational complexity discussion precise, we study specifically the problem of uniform sampling from convex bodies under the *membership oracle*, i.e. we assume access to an oracle which takes an input $x \in \mathbb{R}^n$, and tells us whether $x \in \mathcal{K}$. Then here is the quantitative formulation of the computational complexity problem we care about.

ϵ-approximate sampling problem. Given $\epsilon > 0$, $\mu \propto \mathbf{1}_{\mathcal{K}}$, how many queries of the membership oracle do we need to produce a sample $X \sim \nu$, such that

$$d_{\mathrm{TV}}(\mu, \nu) \leq \epsilon,$$

where the total variation distance is defined as

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{\text{measurable } A \subset \mathbb{R}^n} |\mu(A) - \nu(A)|.$$

2.1.1 Ball walk and hit-and-run

We introduce two ways for uniform sampling in a convex body.

Ball walk. The Ball walk is an iterative algorithm which tries to step to a random point within distance δ of the current point. At each iteration, at the current point x, it does

- Pick a uniform random point from the ball of radius δ centered at x, denoted as $\mathbb{B}(x, \delta)$
- If y is in $\mathcal{K}$, go to y; otherwise stay at x.

Hit-and-run. The Hit-and-run is also an iterative algorithm which picks a random point along a random line through the current point x. It does not need a "step-size" parameter. At each iteration, at the current point x, it does

- Pick a uniform random line ℓ through the current point x
- Go to a uniform random point on the chord $\ell \cap \mathcal{K}$.

Notes on the per-iteration cost. Each iteration of Ball walk can be implemented with one call of the membership oracle. Here we have ignored the cost of sample a random Gaussian vector, assuming it is much smaller than the cost of a query of membership oracle. To sample the first step of Hit-and-run, we can first generate a random standard Gaussian vector u, then the $u/\|u\|$ is a uniformly distributed unit direction. For the second step of Hit-and-run, using the membership oracle of $\mathcal{K}$, we can first find the two points a, b where the line ℓ intersects $\mathcal{K}$, which can be done approximately by a bisection search with a logarithmic overhead. Even if the a, b we found are not exact where the line ℓ intersects $\mathcal{K}$, we can easily use reject sampling to sample uniformly on the chord $\ell \cap \mathcal{K}$.

Given the above discussion on per-iteration cost, what matters for the ϵ-approximate sampling problem is the *mixing time*. The mixing time is the number of iterations that an algorithm takes to obtain an approximate sample, which we define formally after introducing the Markov chain basics.

Metropolized random walk with any proposal. Ball walk and Hit-and-run are not the only Markov chains that have the correct stationary measure. In fact, we can introduce infinite number of Markov chains that have the correct stationary measure using the Metropolis-Hastings filter (or the accept-reject step). The corresponding Markov chain works at follows: at each iteration, at the current point x, it does

- Pick a random point y according to the proposal kernel $\mathcal{P}_x(\cdot)$.
- Metropolis-Hastings filter: go to

$$\begin{cases} y & \text{with probability } \min\left\{1, \frac{\mu(dy)\mathcal{P}_y(dx)}{\mu(dx)\mathcal{P}_x(dy)}\right\} \\ x & \text{with the remaining probability} \end{cases}$$

Here the proposal kernel $\mathcal{P}_\cdot(\cdot)$ can be any kernel that satisfies the same definition of a transition kernel. We give it a new name to avoid confusion with the transition kernel after the Metropolis-Hastings filter.

2.2 Markov chain basics

Here we introduce a few basic notions about Markov chains.

Transition kernel. A Markov chain on $\mathbb{R}^n$ is specified via its *transition kernel*, which is a function $\mathcal{T}_\cdot(\cdot) : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}^n) \to [0, 1]$ satisfying

- For $x \in \mathbb{R}^n$, $\mathcal{T}_x(\cdot)$ is a probability measure on $\mathbb{R}^n$ (for simplicity, we always assume density w.r.t. Lebesgue measure).
- For any Borel measurable set $B \in \mathcal{B}(\mathbb{R}^n)$, $x \mapsto \mathcal{T}_x(B)$ is a measurable function.

Measurability issues aside, in words, $\mathcal{T}_x(B)$ describes the probability of going from x to the set B in one step of the Markov chain.

With the transition kernel and an initial measure μ_0, we can define the distribution of the Markov chain after k steps, denoted by μ_k, recursively as follows, for $k \geq 0$

$$\mu_{k+1}(dx) := \int \mu_k(dy)\mathcal{T}_y(dx).$$

By a slight abuse of notation, we also use $\mathcal{T}$ as an operator which maps a measure to a new one as the above equation defines. And we use $\mu_0\mathcal{T}^k$ to denote the measure at the k-th step of the Markov chain. Additionally, we often use the random variable notation to denote the Markov chain

$$X_0 \to X_1 \to \cdots \to X_k \to X_{k+1}$$

where $X_0 \sim \mu_0$, and $X_k \sim \mu_k$, for all $k \geq 0$.

Reversible. We say a Markov chain $\mathcal{T}$ is *reversible* w.r.t. μ if for all $(x, dy) \in \mathbb{R}^n \times \mathcal{B}(\mathbb{R}^n)$,

$$\mu(dx)\mathcal{T}_x(dy) = \mu(dy)\mathcal{T}_y(dx).$$

The above condition is sometimes also called the detailed balance condition.

Stationary measure. A measure μ is called a *stationary* measure of a Markov chain $\mathcal{T}$ if one steps from it gives back the measure, i.e., for any measurable set B,

$$\int_{x\in\mathbb{R}^n} \mu(dx)\mathcal{T}_x(B) = \mu(B).$$

Irreducible. We say a Markov chain is *irreducible* if any state can be reached in a finite number of steps. Formally, for any two measurable sets A, B with $\mu(A) > 0, \mu(B) > 0$ and any initialization μ_0, if $\mu_0(A) > 0$, then there exists an integer $k > 0$ such that $(\mu_0 \mathcal{T}^k)(B) > 0$. In the case of sampling from a convex body, since the convex body is connected, both Ball walk and Hit-and-run are irreducible.

Lazy. We say a Markov chain is *lazy* if $\mathcal{T}_x$ puts at least mass $\frac{1}{2}$ at x. Any Markov chain can be made lazy by a coin tossing step: if it is head, then stay at the current state; otherwise move as it does before. For simplicity, we always assume our Markov chain is made lazy, because it facilitates simpler proofs while it affects the mixing by at most a constant two factors.

Is the stationary measure unique? In the case of sampling from convex body, it is noted in Theorem 2.1 of [Vem05] that if μ is stationary w.r.t. a lazy Markov chain, then it is the unique stationary measure.

ϵ-mixing time. Given a Markov chain $\mathcal{T}$, $\epsilon > 0$ and an initial measure μ_0, the mixing time in total variation distance is

$$t_{\text{mix}}(\epsilon, \mu_0) = \inf \left\{ k \in \mathbb{N} \mid d_{\text{TV}} \left(\mu_0 \mathcal{T}^k, \mu \right) \leq \epsilon \right\}.$$

2.3 Guess the mixing time

Before we dive into the modern techniques for proving the mixing time of a Markov chain, let's try to gain some intuition about the mixing time of Ball walk in simple scenarios.

Consider the problem of using Ball walk to sample from the unit cube $\mathcal{K} = [-1, 1]^n$. Let's set the initial measure μ_0 to be point mass at 0. We would like to answer the two following questions intuitively

1. Can we have a rough estimate of the mixing time of Ball walk for small enough step-size $\delta > 0$?

2. What is the largest step-size we can choose?

Fix ϵ to be a very small number, say 0.0001. Let X_k denote the random variable at the k-th step of Ball walk. For the first question, we give a rough lower bound of the mixing time by considering the variance propagation. This argument was used in [Lov99] to obtain a mixing lower bound for Hit-and-run.

By the mixing time definition, for the Markov to mix in k steps, the distribution of X_k is ϵ-close to that of the target measure $\mu \propto \mathbf{1}_{\mathcal{K}}$. A necessary condition is that the

covariance matrix of X_k is close to the covariance matrix of $\mu \propto \mathbf{1}_{\mathcal{K}}$ which is $\frac{2}{3}\mathbb{I}_n$. Then consider the variance propagation in the first coordinate of $X_{\cdot 1}$. Note that before we reject any sample, we have

$$\mathbb{E}\left[X_{(k+1)1} - X_{k1} \mid X_{k1}, X_{k2}, \cdots, X_{kn}\right] = 0$$
$$\mathbb{E}\left[\left(X_{(k+1)1} - X_{k1}\right)^2 \mid X_{k1}, X_{k2}, \cdots, X_{kn}\right] = \frac{c\delta^2}{n}$$

where c is upper bounded by a universal constant. The second equation follows because the total variance (the sum of variance across n direction) is of order δ^2 and there is symmetry in the choice of directions in Ball walk. Then we have

$$\begin{aligned}
\mathbb{E}\left[(X_{k1})^2\right] &= \sum_{\ell=0}^{k-1} \mathbb{E}\left[(X_{(\ell+1)1})^2 - (X_{\ell 1})^2\right] + \underbrace{\mathbb{E}[X_{\ell 0}^2]}_{=0} \\
&= \sum_{\ell=0}^{k-1} \mathbb{E}\left[\left(X_{(\ell+1)1} - X_{\ell 1}\right)^2 + 2\left(X_{(k+1)1} - X_{k1}\right) X_{k1}\right] \\
&= \frac{ck\delta^2}{n}.
\end{aligned}$$

Hence, for the above variance to be of order 1, k needs to be at least as large as

$$\frac{n}{\delta^2}. \tag{2.1}$$

For the second question, intuitively, we want to choose a step-size that is as large as possible to reduce the mixing time. However, the Metropolis-Hastings step in Ball walk prevents us from choosing a too large step-size. Suppose we choose $\delta \geq \sqrt{n}$, the proposal in Ball walk would roughly cover the entire cube. Then according to the curse of dimension discussion about the rejection sampling algorithm in Lecture 1, the acceptance probability would be less than $\frac{1}{2^n}$, which prevents rapid mixing in high dimension settings.

Furthermore, we observe that at the corner of the cube, say $x = (1, 1, \cdots, 1)$, the acceptance probability is always less than $\frac{1}{2^n}$ no matter how small the step-size is. Luckily, we can choose a small step-size so that such points with small acceptance probability have limited measure. Once, we ignore these points with small acceptance probability, as long as their measure is less than $\epsilon/2$, we can still achieve ϵ-mixing. Consider the inner cube $[-(1-\sigma), (1-\sigma)]^n$, with $\sigma \in (0, 1)$. We want to make sure that

- Points in the inner cube have least a constant acceptance rate.
- The probability measure of the inner cube is at least $1 - \epsilon/2$.

For the second point to hold, we require the volume ratio to be

$$\frac{(1-\sigma)^n}{1} \approx 1 - \sigma n \geq 1 - \epsilon/2,$$

where we used the approximate $(1-\sigma)^n \approx 1 - \sigma n$ for small enough σ. Hence, σ needs to be smaller than $\frac{\epsilon}{2n}$. For the first point to hold, we need the step-size to be at most $c\sqrt{n}\sigma$ (see Figure 2.1). Finally, the largest step-size we can choose is

$$\delta \leq \frac{c\epsilon}{\sqrt{n}}. \tag{2.2}$$

Combining Eq. (2.2) with Eq. (2.1), we obtain a rough lower estimate of Ball walk's mixing time on the unit cube

$$t_{\text{mix}} \gtrsim \frac{n^2}{\epsilon^2}.$$

Hence, we guess that the mixing time of Ball walk for general convex bodies after an

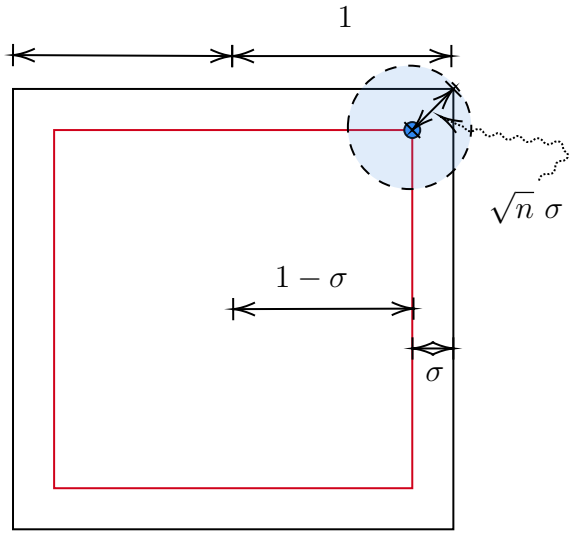

Figure 2.1. Illustration of the inner cube and the largest step-size to maintain at least constant acceptance rate inside the inner cube.

appropriate normalization is also $\frac{n^2}{\epsilon^2}$. Of course, this is just a guess. It has taken people a while to make the above arguments rigorous. Remark that in the breakthrough work by Dyer, Frieze and Kannan [DFK91], where the first polynomial algorithm for convex body volume computation was introduced, roughly n^{19} membership oracle queries were needed to sample a single point from a convex body. See Table 1 in [JLLV21] for the progress in reducing the complexity for sampling and volume computation throughout the years.

2.4 Conductance-based mixing proof framework

In this section, we introduce the conductance-based mixing proof framework, which is one of two major mixing proof techniques we discuss in this course. A typical proof of the mixing time upper bound via conductance is done in two steps

1. First, having a conductance lower bound implies mixing time upper bound
2. Second, a conductance lower bound follows from
   - target measure satisfying isoperimetry
   - Markov chain having large transition overlap for close states.

We introduce the two steps formally in the following two subsections.

2.4.1 Conductance

Flow. Given a Markov chain $\mathcal{T}$, its *flow* of a subset B w.r.t. the measure μ is defined as

$$Q(B) := \int_{x\in B} \mu(dx)\mathcal{T}_x(B^c).$$

It is not hard to see that $Q(B) = Q(B^c)$ when μ is the stationary measure of $\mathcal{T}$, because

$$\begin{aligned} Q(B) &= \int_{x\in B} \mu(dx)\mathcal{T}_x(B^c) \\ &= \int_{x\in B}\int_{y\in B^c} \mu(dx)\mathcal{T}_x(dy) \\ &\overset{(i)}{=} \int_{x\in B}\int_{y\in B^c} \mu(dy)\mathcal{T}_y(dx) \\ &= Q(B^c) \end{aligned}$$

where (i) follows from the detailed balance condition.

Conductance. We define the conductance of Markov chain $\mathcal{T}$ by

$$\Phi := \inf_{0<\mu(B)\le\frac{1}{2}} \frac{Q(B)}{\mu(B)}.$$

The quantity $\frac{Q(B)}{\mu(B)}$ has a natural interpretation as the probability that one step of the Markov chain initialized randomly according to $\mu\mathbf{1}_S$ in S goes to the set S^c. Intuitively, if the conductance Φ is large, then leaving any set S is not hard and so no set S is a bottleneck for the Markov chain. As a consequence, the Markov chain will mix rapidly. The following result by Lovász and Simonovits [LS93] makes the above intuition rigorous.

Warm initialization. We say a measure μ_0 is M-warm with respect to μ if

$$\frac{\mu_0(x)}{\mu(x)} \leq M, \forall x \in \mathbb{R}^n.$$

Theorem 2.4.1 (Conductance to mixing [LS93]). *Let $\epsilon > 0$. Consider a Markov chain $\mathcal{T}$ on $\mathbb{R}^n$. Assume it is lazy and reversible w.r.t. μ. Suppose it has conductance Φ. Then starting from an M-warm initialization μ_0, with μ_k denoting the measure at k-th step of the Markov chain, then*

$$d_{TV}(\mu_k, \mu) \leq M \left(1 - \frac{\Phi^2}{2}\right)^k.$$

According to the above theorem, the mixing time of a Markov chain with conductance Φ is upper bounded by

$$\frac{2}{\Phi^2} \log\left(\frac{M}{\epsilon}\right),$$

when started from an M-warm initialization.

Remarks on the conductance-based mixing time bound [LS93]. First, readers who are familiar with the mixing time analysis of a discrete-state reversible Markov chain (say on a graph), might know that the spectral gap (the difference between the first and second eigenvalues, denoted as $1-\lambda$) of the transition kernel of a Markov chain determines its mixing time. For a discrete-state Markov chain, the transition kernel is simply a square matrix of row size the number of states. The mixing time is of the order $1/(1-\lambda)$. While discussing the eigenvalues of a reversible transition kernel on $\mathbb{R}^n$ is technically more difficult, one can still imagine a notation of spectral gap for a symmetric operator. Second, the spectral gap and the conductance is closely related via the famous Cheeger's inequality

$$2\Phi \geq 1 - \lambda \geq \frac{\Phi^2}{2},$$

see e.g. [JS88]. Third, one shortcoming of Theorem 2.4.1 is that in many continuous-state sampling, the warmness M can only be made as small as 2^n. For example, using the uniform measure on $\mathbb{B}(0,1)$ as an initialization for the uniform measure on $\mathbb{B}(0,2)$ results in a warmness parameter

$$M = \frac{\text{vol}(\mathbb{B}(0,2))}{\text{vol}(\mathbb{B}(0,1))} = 2^n.$$

As a result, the $\log(M)$ factor in the mixing time bound causes an extra n dimension dependency overhead in the mixing time. An alternative approach in the literature

is to directly show entropy decay rather than variance decay, which results in mixing time proportional to $\log\log(M)$ rather than $\log(M)$. See e.g. [CDWY20] and average conductance frameworks in [LK99, GMT06]. Finally, in practice, lower bounding the conductance for all measurable sets might be difficult. Since approximate sampling is desired anyway, the following weaker notion of conductance allows us to ignore a small part of the state space where bounding the conductance is difficult.

s-conductance. For $0 < s < \frac{1}{2}$, the s-conductance of a Markov chain is defined as

$$\Phi_s := \inf_{s<\mu(B)\leq\frac{1}{2}} \frac{Q(B)}{\mu(B) - s}$$

Note that any sets of measure less than s is no longer taken into account in the s-conductance definition and for any other measurable sets, we are allowed to ignore a part of measure s. Lovász and Simonovits [LS93] proved the following mixing time upper bound using s-conductance

Theorem 2.4.2 (s-conductance to mixing [LS93])**.** *Let $\epsilon > 0$. Consider a Markov chain $\mathcal{T}$ on $\mathbb{R}^n$. Assume it is lazy and reversible w.r.t. μ. Suppose it has s-conductance Φ_s. Then starting from an M-warm initialization μ_0, with μ_k denoting the measure at k-th step of the Markov chain, then*

$$d_{TV}(\mu_k, \mu) \leq Ms + M\left(1 - \frac{\Phi_s^2}{2}\right)^k.$$

As long as we can prove s-conductance for $s \leq \frac{\epsilon}{2M}$, the mixing time is upper bounded by

$$\frac{c}{\Phi_s^2} \log\left(\frac{M}{\epsilon}\right),$$

which is similar to the case of Theorem 2.4.1.

The proofs of Theorem 2.4.1 and Theorem 2.4.2 are omitted. Interested readers are referred to the original paper [LS93] and the survey [Vem05].

2.4.2 Isoperimetry and transition overlap

In this subsection, we show a geometric argument by [LS93] which reduces conductance lower bounds to bounding two quantities which are more accessible in applications.

Isoperimetry. Given the Euclidean distance d on $\mathbb{R}^n$, we say a measure μ has an *isoperimetric constant* ψ, if for any measurable partition S_1, S_2, S_3 of $\mathbb{R}^n$, we have

$$\mu(S_3) \geq \psi \cdot d(S_1, S_2)\mu(S_1)\mu(S_2),$$

where $d(S_1, S_2) = \inf_{x \in S_1, y \in S_2} d(x, y)$. Remark that the isoperimetric constant is an intrinsic property of the target measure μ, and it has nothing to do with the transition kernel $\mathcal{T}$. It allows us to isolate the geometry property of the target measure from the algorithmic details of a Markov chain. To gain some intuition about the isoperimetric constant, take S_1, S_2 to partition $\mathbb{R}^n$ and S_3 being the infinitesimal small boundary between S_1 and S_2. Then $\frac{\mu(S_3)}{d(S_1,S_2)}$ is roughly the surface area of the boundary between S_1 and S_2. The isoperimetric constant is a lower bound on the surface area-to-volume ratio among all possible 2-partitions. Having a large isoperimetric constant indicates that the measure cannot be cut into halves with a small bottleneck in the boundary (see Figure 2.2).

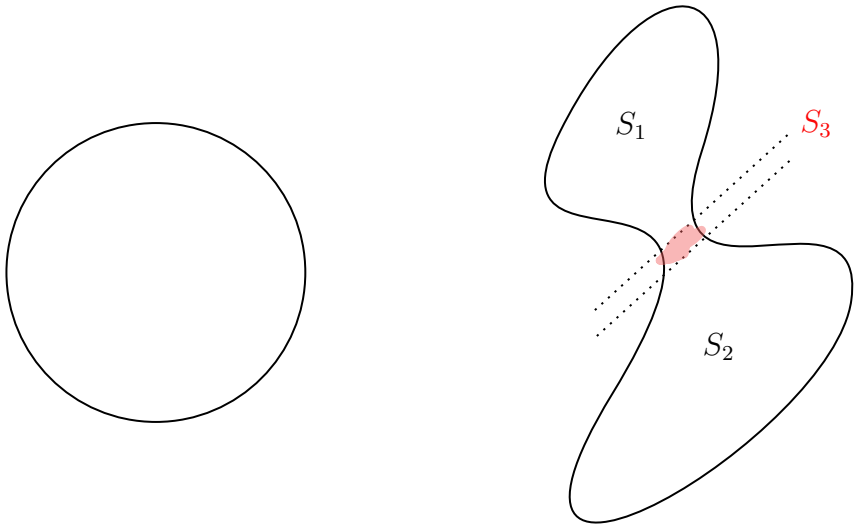

Figure 2.2. Illustration of isoperimetric constant. Left: a measure with large isoperimetric constant. No matter how the partition looks like, the surface area-to-volume ratio is always large. Right: a measure with a small isoperimetric constant according to the partition S_1, S_2, S_3.

Transition overlap. We say that a Markov chain $\mathcal{T}$ has large *transition overlap* for any pair of points within Δ distance if

$$d_{\mathrm{TV}}\left(\mathcal{T}_x, \mathcal{T}_y\right) \leq \frac{1}{2}, \forall x, y \in \mathbb{R}^n \text{ such that } d(x,y) \leq \Delta. \tag{2.3}$$

Remark that unlike the conductance which is a global property of the transition kernel as it requires going through all subsets and considering the flow, the transition overlap is a local property of the transition kernel. It only requires us to check the property for any pair of nearby points (x, y).

The following theorem establishes a lower bound on the conductance via isoperimetry and transition overlap.

Theorem 2.4.3 (Conductance from isoperimetry and transition overlap)**.** *Consider a Markov chain $\mathcal{T}$, lazy and reversible w.r.t. μ. Suppose μ has isoperimetric constant ψ, and Eq. (2.3) holds with Δ, then the conductance of $\mathcal{T}$ is lower bound as*

$$\Phi \geq \min\left\{\frac{1}{8}, \frac{\Delta\psi}{64}\right\}.$$

See Proposition 2 in Kevin Tian's Ball walk notes for a proof. A corollary for s-conductance can also be obtained similarly, which only requires transition overlap over a set of measure $1-s$. See [DCWY19] for a detailed proof for s-conductance.

Remark that one of the reasons that this conductance framework for proving mixing time is popular is that it separates the intrinsic property of the target measure (isoperimetry) and the local property of the Markov chain (transition overlap). Furthermore, the transition overlap only requires us to consider any pair of points, which is often a more tractable problem than considering all partitions.

2.5 Application to Ball walk.

Our target measure is $\mu \propto \mathbf{1}_{\mathcal{K}}$, where $\mathcal{K}$ is a convex body in $\mathbb{R}^n$. We make a simplifying assumption that μ is *isotropic*, meaning that $\mathbb{E}_{X\sim\mu} X = 0$ and $\mathrm{Cov}_{X\sim\mu} = \mathbb{I}_n$.

Isoperimetry for convex body. Kannan, Lovász and Simonovits (KLS) [KLS95] conjectured that the isoperimetric constants of all convex bodies should be lower-bounded by $c\left\|\mathrm{Cov}_\mu\right\|_{\mathrm{op}}^{-\frac{1}{2}}$, where c is a universal constant. This is a deep result in probability theory which has close connections to many other conjectures such as Bourgain's slicing problem (see e.g. the survey [LV18]). Recently, there has been tremendous process on the KLS conjecture. The current state-of-the-art bound of the isoperimetric constant is as follows.

Theorem 2.5.1 (Klartag [Kla23]). *There exists a universal constant $c > 0$ such that for any uniform distribution μ on a convex body, its isoperimetric constant satisfies*

$$\psi \geq \frac{1}{c\sqrt{\log(n)}} \left\|\mathrm{Cov}_\mu\right\|_{op}^{-\frac{1}{2}}.$$

Given our simplifying assumption of isotropicity, the isoperimetry part only costs us a logarithmic dependency on n.

Transition overlap of Ball walk. In order to bound the transition overlap (2.3) of Ball walk, we observe by triangular inequality that

$$d_{\mathrm{TV}}\left(\mathcal{T}_x, \mathcal{T}_y\right) \leq d_{\mathrm{TV}}\left(\mathcal{T}_x, \mathcal{P}_x\right) + d_{\mathrm{TV}}\left(\mathcal{P}_x, \mathcal{P}_y\right) + d_{\mathrm{TV}}\left(\mathcal{T}_y, \mathcal{P}_y\right).$$

It suffices to bound two kinds of terms

- $d_{\mathrm{TV}}\left(\mathcal{P}_x, \mathcal{P}_y\right)$, namely proposal overlap for two close points. Since the proposal is uniform in a ball, it is just calculating the volume of the intersection of two n-dimensional balls of radius δ as a function of distance between their centers. Precise formula for a section of an n-dimensional ball exists. Intuitively, Δ needs to be of order $\delta/\sqrt{n}$ for the intersection to be of order constant.
- $d_{\mathrm{TV}}\left(\mathcal{T}_x, \mathcal{P}_x\right)$, namely acceptance rate. We have seen in Section 2.3 that the acceptance rate gets bad when one is close to the boundary of the convex body. The main strategy is to shave off a small piece from the boundary of the convex body and ignore that piece via s-conductance technique.

Local conductance. To bound the acceptance rate, we define the *local conductance* at x,

$$\ell(x) := \frac{\mathrm{vol}(\mathbb{B}(x,\delta) \cap \mathcal{K})}{\mathrm{vol}(\mathbb{B}(x,\delta))}.$$

It is also the probability of accepting a step, as we observe that for the next state from x takes the form

$$\begin{cases} x & \text{with probability } 1 - \frac{1}{2}\ell(x) \\ \text{a uniform draw from } \mathbb{B}(x,\delta) \cap \mathcal{K} & \text{with probability } \frac{1}{2}\ell(x). \end{cases}$$

We can also define the region of high local conductance as

$$\mathcal{K}_\delta = \left\{ u \in \mathcal{K} : \ell(u) \geq \frac{3}{4} \right\}.$$

For points inside $\mathcal{K}_\delta$, the acceptance rate is constant. The following lemma shows that it is possible to make K_δ to have large measure.

Lemma 1 (Lemma 5.3 in [Vem05], originally from [KLS97]). *Suppose that $\mathcal{K}$ is a convex body containing a unit ball in $\mathbb{R}^n$. Then*

- *$\mathcal{K}_\delta$ is a convex set,*
- $\mathrm{vol}(\mathcal{K}_\delta) \geq (1 - 2\delta\sqrt{n})\,\mathrm{vol}(K)$.

Remark that if $\mathcal{K}$ is isotropic, Theorem 4.1 in [KLS95] ensures that $\mathcal{K}$ contains a unit ball. Then we have the following implication: if we set $\delta \leq s/(2\sqrt{n})$, at least a measure $(1-s)$ part of $\mathcal{K}$ has large acceptance rate.

All pieces together. We combine all pieces together to get a mixing time upper bound of Ball walk

1. Choose $s \approx \frac{\epsilon}{2M}$ so that the first term in Theorem 2.4.2 is smaller than $\frac{\epsilon}{2}$.
2. Choose $\delta \approx \frac{s}{\sqrt{n}}$ so that $\mathcal{K}_\delta$ has measure at least $(1-s)$.
3. Ball walk satisfies transition overlap for $\Delta \approx \frac{\delta}{\sqrt{n}}$.
4. The isoperimetry for isotropic convex body follows from Theorem 2.5.1, $\psi \approx \log^{-\frac{1}{2}}(n)$.
5. The conductance lower bound follows from the s-conductance variant of Theorem 2.4.3, $\Phi_s \approx \Delta\psi$.
6. Finally, the mixing time is

$$\begin{aligned} &\frac{1}{\Phi_s^2}\log\left(\frac{M}{\epsilon}\right) \\ \approx\ &\frac{n\log n}{\delta^2}\log\left(\frac{M}{\epsilon}\right) \\ \approx\ &\frac{M^2}{\epsilon^2}n^2\log n\cdot\log\left(\frac{M}{\epsilon}\right). \end{aligned}$$

We obtain an order n^2 mixing time for Ball walk from a warm initialization (M is constant). Wait! Didn't we mention that the warmness parameter in continuous-state sampling is typically exponential? Would the dependency on M make the bound very bad? Right. Let's discuss two ways to get around this issue.

- [KLS97] introduced a variant of Ball walk called speedy walk and provide a more careful analysis of the acceptance rate to improve the warmness dependency.

- Sometimes it is possible to deal with a problem with large warmness via considering a sequence of problems with small warmness. For example, consider the case of volume computation of a convex body $\mathcal{K}$ satisfying $\mathbb{B}(0,1) \subseteq \mathcal{K} \subseteq \mathbb{B}(0,R)$. One can define a sequence of convex bodies $\mathcal{K}_i = \mathcal{K} \cap \mathbb{B}(0, 2^{i/n})$, for $i = 1, \cdots, m := n \log R$. Then

$$\mathbb{B}(0,1) = \mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_m = \mathcal{K}.$$

  And these convex bodies satisfy

$$\mathrm{vol}(K_i) \leq 2\,\mathrm{vol}(K_{i-1}).$$

  As a consequence, the uniform distribution on $\mathcal{K}_{i-1}$ serves as a 2-warm initialization for the uniform distribution on $\mathcal{K}_i$. In order to sample from $\mathcal{K}$, we can consider a sequence of sampling problems on $\mathcal{K}_i$, each equipped with a constant warm initialization. Then the polynomial dependency on M is no longer an issue.

2.6 Discussion

In the study of sampling from a uniform distribution on a convex body, we realize that, from a geometric perspective, the main obstacle of mixing resides in

- Covariance Cov_μ. It affects the isoperimetric constant and step-size choices. It becomes relatively minor after the isotropic simplifying assumption. In general, *rounding* is the main technique to deal with it (see Section 7 in [Vem05]).

- Isoperimetry ψ. In the case of sampling from an isotropic convex body, we are lucky that ψ only has logarithmic dependency on n. In general, especially for multimodal target measurable, ψ might be exponentially small in n. Intuitively, it would create a bottleneck for Markov chains making local moves. Would it always be a main obstacle for mixing?

- Corners. According to our proof, Ball walk has to take small step-sizes mainly to ensure that a large measure of points in $\mathcal{K}$ have good acceptance. And the main places where the acceptance rate is bad are at the corners of the convex body. What if we assume that the target measure is smooth and have no corners? Would there be better Markov chains and would the mixing time become better?

Bibliography

[CDWY20] Yuansi Chen, Raaz Dwivedi, Martin J Wainwright, and Bin Yu. Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients. *J. Mach. Learn. Res.*, 21:92–1, 2020.

[DCWY19] Raaz Dwivedi, Yuansi Chen, Martin J Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42, 2019.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.

[GMT06] Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 11:1–26, 2006.

[JLLV21] He Jia, Aditi Laddha, Yin Tat Lee, and Santosh Vempala. Reducing isotropy and volume to KLS: an $O(n^3\psi^2)$ volume algorithm. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 961–974, 2021.

[JS88] Mark Jerrum and Alistair Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of permanent resolved. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 235–244, 1988.

[Kla23] Bo'az Klartag. Logarithmic bounds for isoperimetry and slices of convex sets. *arXiv preprint arXiv:2303.14938*, 2023.

[KLS95] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3):541–559, 1995.

[KLS97] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

[LK99] László Lovász and Ravi Kannan. Faster mixing via average conductance. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 282–287, 1999.

[Lov99] László Lovász. Hit-and-run mixes fast. *Mathematical programming*, 86(3):443–461, 1999.

[LS93] László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms*, 4(4):359–412, 1993.

[LV18] Yin Tat Lee and Santosh S Vempala. The Kannan-Lov\'asz-Simonovits conjecture. *arXiv preprint arXiv:1807.03465*, 2018.

[Vem05] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.