CPSC 661: Sampling Algorithms in Machine Learning 2/24/2021

Lecture 7

Lecturer: Andre Wibisono *Scribe: Argyris Oikonomou*

1 Outline

Today's lecture covers how to bound the conductance of a reversible Markov Chain using the notion of isoperimetry. A good reference on this topic is [6].

2 Preliminaries

First we are going to review the notion used throughout this section. Consider a reversible Markov Chain (MC) P with stationary distribution ν on state space $\mathcal{X}$. For a subset of state space $A \subset \mathcal{X}$, we denote by $\Phi(A)$ the ergodic flow of MC P, that is:

$$\Phi(A) = \int_A P_x(A^c)\, d\nu(x) = \Pr\left[X_1 \in A^c \wedge X_0 \in A \mid X_0 \sim \nu\right]$$

The conductance of MC P is

$$\phi = \inf_{A \subset \mathcal{X}} \frac{\Phi(A)}{\min\{\nu(A), 1 - \nu(A)\}}$$

3 Motivating Isoperimetry

Consider a reversible Markov Chain P with stationary distribution ν. What can we say about ν if we assume that P has large conductance?

Unfortunately, a bound on the conductance cannot help us infer anything about ν. More precisely Example 1 demonstrates that for any stationary distribution ν there exists a MC with high conductance.

Example 1. *For any distribution ν on state space $\mathcal{X}$, consider MC P with transition probability from state $x \in \mathcal{X}$ to state $y \in \mathcal{X}$ equal to $P_x(y) = \nu(y)$. Observe that we converge to the stationary distribution after one step regardless of our initial distribution. As an exercise one can prove that the conductance of P is $\frac{1}{2}$.*

On the other hand, there are stationary distributions that are hard to sample from as the one in Example 2.

Example 2. *Consider a reversible MC chain P with state space the blue area of the dumbbell graph illustrated in Figure 1. MC P has a uniform stationary distribution on the state space. If we only allow transition between points that are "close" (the definition of close is going to become formal later), then the conductance can be very small which makes sampling from the stationary distribution hard.*

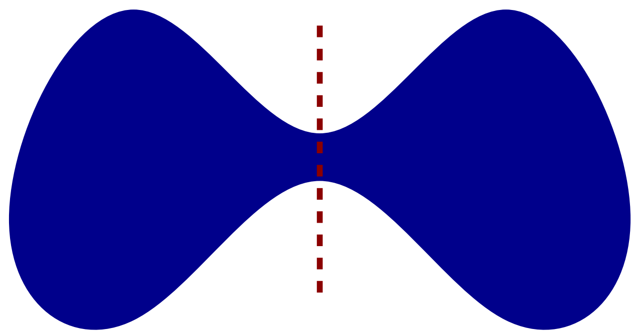

Figure 1: The MC of Example 2.

A formal definition of a local MC is defined in Definition 1.

Definition 1. *A MC P on a metric state space $\mathcal{X}$ w.r.t. distance d is* local *if for every pair $(x, y) \in \mathcal{X} \times \mathcal{X}$ such that $P_x(y) > 0$, then $d(x, y) \ll 1$ with high probability.*

Intuitively, a MC P is local if we can transition from a state only to states that are nearby. Now we are ready to introduce the notion of isoperimetry. The following informal example provides some intuition for the usefulness of isoperimetry that we are going to define later.

Example 3. *Consider a local MC P on a metric state space $\mathcal{X}$ as depicted in Figure 2, where we want to calculate the ergodic flow $\Phi(A)$ of A. By the locality assumption, we can only move from a point in A to a point in A^c only from points in A that are near the boundary between A and A^c (we denote this boundary by ∂A). Thus, if the value of $\Phi(A)$ is close to $\nu(A)$, ∂A has "volume" comparable to $\nu(A)$.*

Therefore if MC P has large conductance, then by the definition of conductance for any set $A \subset \mathcal{X}$ such that $\nu(A) \leq \frac{1}{2}$, $\Phi(A)$ has value close to $\nu(A)$. Thus, we can infer that $\nu(A)$ is comparable to $\nu(\partial A)$.

4 Where did Isoperimetry Originate from?

Isoperimetry deals with the problem of finding a region with the largest volume, among all regions of equal perimeter / area. In two dimensions the region that maximizes the area for a fixed perimeter is the circle. Now we are ready to provide a formal definition of isoperimetry.

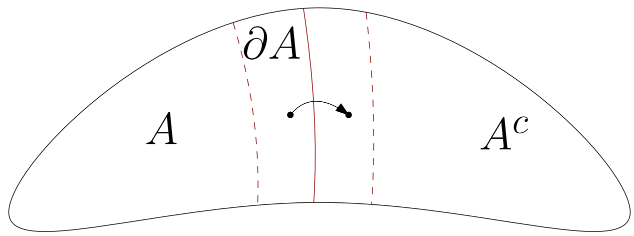

Figure 2: The MC of Example 3.

Definition 2. *Let $\mathcal{X}$ be a metric space equipped with distance metric $d(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (for example $\mathcal{X} = \mathbb{R}^n$ and $d(x,y) = ||x-y||_2$). Given two sets $S_1, S_2 \subset \mathcal{X}$, we define the distance between S_1 and S_2 as $d(S_1,S_2) = \inf\{d(x,y) : x \in S_1, y \in S_2\}$. Given a distribution ν on $\mathcal{X}$, we say that ν is* isoperimetric *with constant $\Psi > 0$ if for every partition (S_1, S_2, S_3) of $\mathcal{X}$:*

$$v(S_3) \geq \Psi \cdot d(S_1,S_2) \cdot \min\{v(S_1), v(S_2)\}$$

Informally speaking, isoperimetry means that for any two disjoint sets S_1, S_2 of $\mathcal{X}$ with fixed volume, the set $\mathcal{X}\backslash S_1 \backslash S_2$ must grow proportionally to the distance of S_1 and S_2 as depicted in Figure 3

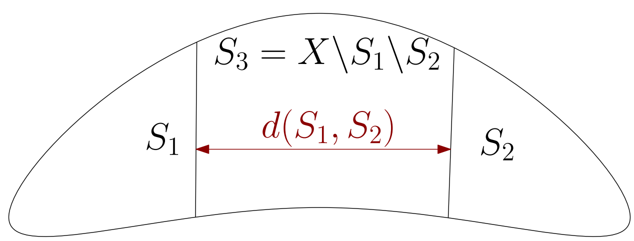

Figure 3: Intuition behind definition of isoperimetry.

The notion of isoperimetry with constant Ψ is equivalent to the following:

1. The differential formulation which states that:

   $$\forall S \subset \mathcal{X},\ \nu(\partial S) \geq \Psi \cdot \min\{\nu(S), \nu(S^c)\}$$

   where $\nu(\partial S) = \lim_{\epsilon \to 0} \frac{\nu(S_\epsilon) - v(S)}{\epsilon}$, and $S_\epsilon = \{x \in \mathcal{X} : d(x,S) \leq \epsilon\}$. Note that $\frac{\nu(S_\epsilon)-\nu(S)}{\epsilon}$ can be thought as the rate of growth of the boundary of S as demonstrated in Figure 4. (Note that $\nu(S_\epsilon) - \nu(S) = \nu(S_3)$ w.r.t. the sets defined in Figure 4 if $S = S_1$.) By considering sets S_1, S_2 and S_3 as depicted in Figure 4, we can see that isoperimetry implies the differential formulation.

2. the Poincare inequality, which states that:

   $$\forall g : \mathcal{X} \to \mathbb{R},\ \mathbb{E}_\nu \left[||\nabla g||^2\right] \geq \frac{\Psi^2}{4} \mathrm{Var}_\nu(g).$$

We will see more of this in a later lecture.

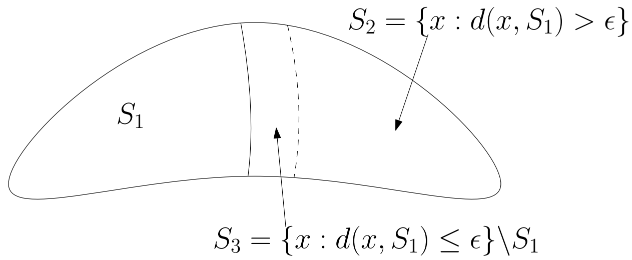

Figure 4: Isoperimetry implies differential formulation.

5 Classes of Distributions with Isoperimetric Constant

Not every distribution has a meaningful isoperimetric inequality. For instance, in the MC in Example 2, if we consider the set that lies left to the dashed line as illustrated in Figure 1, then the differential formulation implies a very small isoperimetric constant.

One special class of distribution with good isoperimetric constants are the log-concave distributions. A non-trivial isoperimetric inequality on convex bodies is the following.

Theorem 1 ([3]). *Let $K \subset \mathbb{R}^n$ be a convex body with diameter D and let ν be the uniform distribution on K with isoperimetric constant Ψ. Then:*

$$\Psi \ge \frac{2}{D}.$$

Note that Theorem 1 implies that for any two disjoint sets S_1, S_2 of K:

$$v(K \backslash S_1 \backslash S_2) \ge \frac{2d(S_1, S_2)\min\{v(S_1), v(S_2)\}}{D}$$

where by $v(S)$ we denote the volume of the set S. The theorem is tight. For example we consider the convex set $K = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, 0 \le z \le h\}$, that is a cylinder of radius 1 and height $h \gg 1$. Observe that the diameter of this convex set is approximately h. If we consider sets $S_1 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, 0 \le z \le \frac{h-1}{2}\}$, $S_2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, \frac{h+1}{2} \le z \le h\}$ and $S_3 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, \frac{h-1}{2} \le z \le \frac{h+1}{2}\}$, that is S_1 is the first part of the cylinder until height $\frac{h-1}{2}$, S_2 is the second part of the cylinder from height $\frac{h-1}{2}$ to $\frac{h+1}{2}$ and S_3 is the part of the cylinder from height $\frac{h+1}{2}$ until h. Note that $v(S_1) = v(S_2) = \frac{\pi(h-1)}{2}$, $v(S_3) = \pi$ and $d(S_1, S_2) = 1$. Thus the isoperimetric constant is bounded by:

$$\Psi \le \frac{2}{h-1} \approx \frac{2}{h}$$

Theorem 1 can be generalised to include log-concave distributions.

Definition 3. *For a distribution ν, we denote by* $\mathrm{supp}(\nu)$ *the support of this distribution, that is* $\mathrm{supp}(\nu) = \{x \in \mathbb{R}^n : v(x) > 0\}$. *A distribution $\nu(x) \propto e^{-f(x)}$, is log-concave if f is convex on* $\mathrm{supp}(\nu)$.

Theorem 2 (Generalization of Theorem 1)**.** *Let ν be a log-concave distribution with isoperimetric constant Ψ, supported on a space with diameter D. Then:*

$$\Psi \geq \frac{2}{D}$$

Unfortunately, if the support of our distributions is not restricted, then Theorem 2 does not provide any meaningful bound. The following theorem provides a bound on the isoperimetric constant even on unbounded supports.

Theorem 3 ([4])**.** *Let ν be a log-concave distribution with isoperimetric constant Ψ. Then:*

$$\Psi \geq \frac{\log(2)}{M_1(\nu)} \geq \frac{1}{4M_1(\nu)} \geq \frac{1}{4\sqrt{\mathrm{Var}_\nu(x)}}$$

where $M_1(\nu)$ is the first moment of ν. More formally if $\mu = \mathbb{E}_\nu[\mathcal{X}]$, then $M_1(\nu) = \mathbb{E}_\nu\left[||\mathcal{X} - \mu||_2\right]$. The second inequality in the statement of the theorem is implied by Cauchy-Schwarz.

Of special interest, are isotropic distributions.

Definition 4. *A distribution ν on space $\mathcal{X}$ is* isotropic *if it has mean zero and covariance matrix equal to the identity matrix, that is $\mu = \mathbb{E}_\nu[x] = 0$ and $\Sigma = \mathbb{E}_\nu\left[(x-\mu)(x-\mu)^\top\right] = I$. Therefore, $Var_\nu(x) = Tr(I) = n$.*

Observe that for arbitrary distribution ν on space $\mathcal{X}$, its covariance matrix $\Sigma = \mathbb{E}_\nu\left[(x-\mu)(x-\mu)^\top\right]$ is always symmetric and positive semi-definite (that is the eigenvalues of Σ are all non-negative).

Example 4. *For an isotropic distribution ν, Theorem 3 implies that the isoperimetric constant on distribution ν is at least $\frac{1}{4\sqrt{n}}$.*

Theorem 3 can be stated as:

Theorem 4 ([4])**.** *Let ν be a log-concave distribution with isoperimetric constant Ψ, covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$. Then:*

$$\Psi = \Omega\left(\frac{1}{\sqrt{\sum_{i=1}^n \lambda_i}}\right) = \Omega\left(\frac{1}{\sqrt{\mathrm{Tr}(\Sigma)}}\right)$$

Kannan, Lovasz, and Simonovits believed that their dependence on the sum of the eigenvalues of the covariance matrix of Σ is not of the right order and made the following conjecture.

Conjecture 1 ([4])**.** *Let ν be a log-concave distribution with isoperimetric constanct Ψ, covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$. Then:*

$$\Psi = \Omega\left(\frac{1}{\sqrt{\lambda_1}}\right) = \Omega\left(\frac{1}{\sqrt{\|\Sigma\|_{\mathrm{op}}}}\right)$$

In particular, for isotropic distribution ν, the conjecture states that the isoperimetric constant should be $\Omega(1)$. Significant progress has been made in the recent years towards the conjecture. In the following theorems we mention some of the most recent results.

Theorem 5 ([5])**.** *Let ν be a log-concave distribution with isoperimetric constant Ψ, covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$. Then:*

$$\Psi = \Omega\left(\frac{1}{\left(\sum_{i=1}^n \lambda_i^2\right)^{1/4}}\right) = \Omega\left(\frac{1}{\sqrt{\|\Sigma\|_{\mathrm{HS}}}}\right)$$

where $||\Sigma||_{HS} = \sqrt{\sum_{i=1}^n \lambda_i^2}$ is the Hilbert-Schmidt norm of Σ. In particular, for ν isotropic, the isoperimetric constant at least $\frac{1}{n^{1/4}}$.

The most recent improvement almost resolves the conjecture.

Theorem 6 ([1])**.** *Let ν be a log-concave distribution with isoperimetric constant Ψ, covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$. Then:*

$$\Psi = \Omega\left(\frac{1}{n^{o(1)}\sqrt{\lambda_1}}\right)$$

In particular, for ν isotropic, the isoperimetric constant at least $\frac{1}{n^{o(1)}}$.

Now we turn our attention on strongly log-concave functions.

6 Strongly Log-Concave Distribution

First we need to introduce what it means for a function to be strongly-convex.

Definition 5. *For a convex set $\mathcal{X}$ and a metric $\|\cdot\|$, a function $f : \mathcal{X} \to \mathbb{R}$ is α-*strongly convex *if for every $x, y \in \mathcal{X}$:*

$$\frac{f(x) + f(y)}{2} - f\left(\frac{x+y}{2}\right) \geq \frac{\alpha}{8}||y - x||^2$$

Equivalently, if f is differentiable, then f is α-strongly convex if for every $x, y \in \mathcal{X}$:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}||y - x||^2$$

where by $\nabla f(x)$ we define the vector of partial derivatives of f, $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right)$.

Equivalently, if f is twice differentiable, then f is α strongly convex if for every $x \in \mathcal{X}$:

$$\nabla^2 f(x) \succeq \alpha I$$

where by $\nabla^2 f(x)$ we define the Hessian $n \times n$ matrix with entry on the (i,j) cell equal to $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$. Note that the Hessian matrix is symmetric. By $\nabla^2 f(x) \succeq \alpha I$ we mean that all the eigenvalues of $\nabla^2 f(x)$ for $x \in \mathcal{X}$ are at least α.

More on convex functions on a future lecture.

Definition 6. *A distribution $\nu(x) \propto e^{-f(x)}$, is α-*strongly log-concave *(SLC) if f is α-strongly convex on* $\operatorname{supp}(\nu)$. *If ν is 0-strongly log-concave, then we say ν is log-concave.*

Next we provide an example of a very famous strongly log-concave function.

Example 5. *Let ν be a Gaussian distribution $\mathcal{N}(0, \Sigma)$, where Σ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$. Then ν is $\frac{1}{\lambda_1}$-SLC.*

Proof. The density function of $\mathcal{N}(0, \Sigma)$ is:

$$\nu(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}x^\top \Sigma^{-1} x}.$$

Then

$$f(x) = -\log \nu(x) = \frac{1}{2} x^\top \Sigma^{-1} x + \frac{1}{2} \log \det(2\pi\Sigma).$$

It is sufficient to prove that $f(x)$ is $\frac{1}{\lambda_1}$-strongly convex. Note $\nabla f(x) = \Sigma^{-1} x$ and $\nabla^2 f(x) = \Sigma^{-1}$. Note that Σ^{-1} has eigenvalues $\frac{1}{\lambda_n} \geq \frac{1}{\lambda_{n-1}} \geq \ldots \geq \frac{1}{\lambda_1} > 0$, which implies that $\Sigma^{-1} \succeq \frac{1}{\lambda_1} I$, which in turn implies that f is $\frac{1}{\lambda_1}$-strongly convex. $\square$

In Figure 5 we plot the density function of $\mathcal{N}(0, 10)$, (0.1-SLC distribution) and in Figure 6 we plot the density function of $\mathcal{N}(0, 1)$, (1-SLC distribution). Note that SLC distributions have more concentrated peaks. The following theorem provides a connection between how strongly log-concave a function is, and the quality of its isoperimetric constant.

Theorem 7 ([2])**.** *if ν is α-SLC, then we can bound its isoperimetric constant Ψ:*

$$\Psi \geq \log(2)\sqrt{\alpha} \geq \frac{\sqrt{\alpha}}{4}$$

An important property of Theorem 7 is that it is dimension-free.

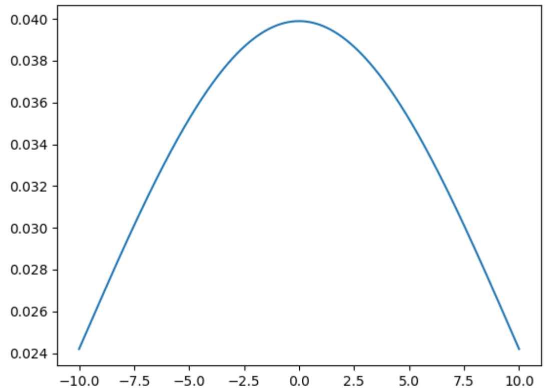

Figure 5: Density function of $\mathcal{N}(0, 10)$.

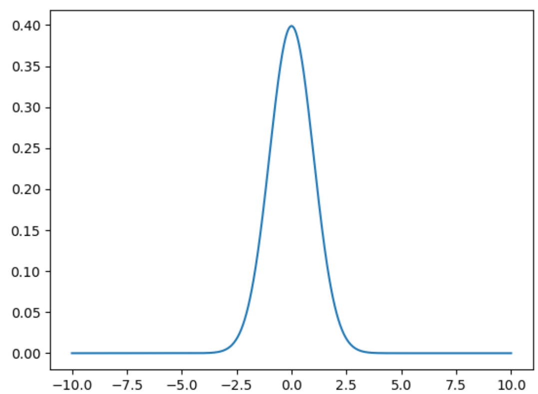

Figure 6: Density function of $\mathcal{N}(0, 1)$.

References

[1] Y. Chen. An almost constant lower bound of the isoperimetric coefficient in the kls conjecture, 2021.

[2] B. Cousins and S. S. Vempala. Gaussian cooling and $O^*(n^3)$ algorithms for volume and gaussian volume. *SIAM J. Comput.*, 47(3):1237–1273, 2018.

[3] M. Dyer and A. Frieze. Computing the volume of convex bodies: A case where randomness provably helps. 1991.

[4] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discret. Comput. Geom.*, 13:541–559, 1995.

[5] Y. T. Lee and S. S. Vempala. The Kannan-Lovász-Simonovits conjecture. *arXiv preprint*, abs/1807.03465, 2018.

[6] S. Vempala. Geometric random walks: a survey. *Combinatorial and Computational Geometry*, pages 573–612, 2005.