CS 7535 Markov Chain Monte Carlo Methods **Fall 2017**

Lecture 22: November 9

Lecturer: Santosh Vempala *Scribe: Matthew Fahrbach*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the instructor.*

22.1 Volume Computation

We study approximation algorithms for volume computation using Markov chain Monte Carlo methods. Many of these techniques can be extended to estimating integrals of log-concave functions.

Definition 22.1 *The* volume *of a set $K \subseteq \mathbb{R}^n$ is defined as the multi-dimensional integral*

$$\mathrm{Vol}(K) = \int_K \mathrm{d}\boldsymbol{x}.$$

We first give the exact volume of a few highly-structured sets in high-dimensional geometry.

Example 22.2 *For any $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^n$, the volume of the* simplex $S = \mathrm{Conv}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_n\})$ *is*

$$\mathrm{Vol}(S) = \frac{1}{n!}\left|\det\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{x}_0 & \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}\right)\right|.$$

Example 22.3 *The volume of the n-dimensional* ball *of radius r is*

$$\mathrm{Vol}(rB_n) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} r^n,$$

where $\Gamma(z)$ denotes the gamma function. More generally, the volume of an n-dimensional ellipsoid

$$E = \left\{\boldsymbol{x} \in B_n : \boldsymbol{x}^\intercal A^{-1} \boldsymbol{x} \le 1\right\} = A^{1/2} B_n,$$

for a positive definite matrix A, is

$$\mathrm{Vol}(E) = \sqrt{\det(A)}\mathrm{Vol}(B_n).$$

22.1.1 Hardness of Exact Computation

To motivate approximation algorithms for volume computation, we present a hardness result that reduces computing the permanent of a matrix to computing volume.

Definition 22.4 *A* rational polytope P *is a set*

$$P = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \geq \boldsymbol{b}\},$$

where $A \in \mathbb{Q}^{m \times n}$ *is a rational matrix and* $\boldsymbol{b} \in \mathbb{Q}^m$ *is a rational vector.*

Theorem 22.5 ([DF88]) *It is* **#P-hard** *to compute the volume of a rational polytope.*

For approximate volume computation algorithms, we consider the case where K is a convex body, as this seems to be the most general case that avoids being **#P-hard**. Analogously, we consider integrating log-concave functions since they naturally generalize the indicator function for a convex body.

To analyze such algorithms in greater generality, we give as input a membership oracle for the convex body instead of a list of constraints defining the body (e.g. rational polytopes). The complexity of such algorithms is then defined by the number of oracle calls and the number of arithmetic operations. Similarly, for integrating log-concave functions we use a function-evaluation oracle.

Definition 22.6 *A* well-guaranteed membership oracle *for a convex body* K *parameterized by a point* $\boldsymbol{x}_0 \in K$ *and* $r, R \in \mathbb{R}_{\geq 0}$ *such that*

$$\boldsymbol{x}_0 + rB_n \subseteq K \subseteq RB_n$$

takes an input $\boldsymbol{x} \in \mathbb{R}^n$ *and returns the answer to the question* $\boldsymbol{x} \in K$*.*

Definition 22.7 *An* approximation algorithm for computing the volume of a convex body K *takes as input a well-guaranteed membership oracle for* K *and an* $\varepsilon > 0$*. It then outputs a value* $V \in \mathbb{R}$ *such that*

$$(1-\varepsilon)\mathrm{Vol}(K) \leq V \leq (1+\varepsilon)\mathrm{Vol}(K).$$

The following integration problem generalizes approximating the volume of a convex body.

Definition 22.8 *An* approximation algorithm for integrating a log-concave function $f(\boldsymbol{x})$ *takes as input a function-evaluation oracle and an* $\varepsilon > 0$*. It then outputs a value* $F \in \mathbb{R}$ *such that*

$$(1-\varepsilon)\int_{\mathbb{R}^n} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \leq F \leq (1+\varepsilon)\int_{\mathbb{R}^n} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.$$

22.2 Deterministic Approximation Algorithms

We begin by proposing some simple algorithmic approaches for approximating volume, all of which suffer from the curse of dimensionality:

- For some high-dimensional grid, use the divide-and-conquer paradigm to count the number of cubes fully contained in K. This fails because the number of cubes grows exponentially in the dimension.
- Find an enclosing cube, ball, or ellipsoid using binary search since we have an expression for this larger volume. This fails because the ratio of the two volumes grows exponentially in the dimension.
- For polytopes P, decompose P into simplices and compute their individual volumes. This fails because the number of simplices grows exponentially in the dimension.

Next we present two hardness of approximation results for deterministic algorithms. We give a proof of the second, simpler theorem.

Theorem 22.9 ([BF87, Elekes86]) *For any deterministic algorithm that uses at most n^a oracle calls and computes on input K the values $A, B \in \mathbb{R}$ such that*

$$A \leq \mathrm{Vol}(K) \leq B,$$

there exists a convex set K such that

$$\frac{B}{A} \geq \left(\frac{10n}{a \log n}\right)^{n/2}.$$

Theorem 22.10 *A deterministic 2^{cn}-approximation algorithm requires $2^{(1-c)n}$ oracle calls.*

Proof: Assume without loss of generality that $K \subseteq B_n$. Query the points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ and consider

$$\mathrm{Conv}\left(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}\right) \subseteq K.$$

We will show that

$$\mathrm{Vol}\left(\mathrm{Conv}\left(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}\right)\right) \leq \frac{m}{2^n}\mathrm{Vol}\left(B_n\right).$$

For each $i \in [m]$, draw the ball B^i whose diameter is the chord from the origin to $\boldsymbol{x}_i$. It follows that

$$\mathrm{Vol}\left(B^i\right) \leq \frac{1}{2^n}\mathrm{Vol}\left(B_n\right).$$

We then claim

$$\mathrm{Conv}\left(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}\right) \subseteq \bigcup_{i=1}^{m} B^i.$$

To prove this claim, assume there exists some $\boldsymbol{y} \in \mathrm{Conv}\left(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}\right)$ such that $\boldsymbol{y} \notin B_i$ for all $i \in [m]$. Then all points $\boldsymbol{x}_i$ and the origin lie on one side of a separating hyperplane whose normal vector is defined by $\boldsymbol{y}$. Thus, $\boldsymbol{y} \notin \mathrm{Conv}\left(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}\right)$, a contradiction.

Therefore, using a union bound we have

$$\begin{aligned}\mathrm{Vol}\left(\mathrm{Conv}\left(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}\right)\right) &\leq \sum_{i=1}^{m} \mathrm{Vol}\left(B^i\right) \\ &\leq \frac{m}{2^n}\mathrm{Vol}\left(B_n\right),\end{aligned}$$

which proves the desired result. ■

22.3 Randomized Approximation Algorithms

Given the hardness of deterministic approximations, we turn to randomized algorithms—in particular Markov chain Monte Carlo. The first randomized approach one may try is to enclose K in a ball B_n and then sample points uniformly at random from B_n to estimate the volume of K as

$$\mathrm{Vol}(K) \approx \frac{\text{number of points in } K}{\text{total number of points sampled}}\mathrm{Vol}\left(B_n\right).$$

This fails to be practical, however, because exponentially samples may be needed (e.g. when K is a cube). Nonetheless, this sampling-based approach coupled with the notion of self-reducibility sets the stage for a randomized polynomial-time approximation algorithm.

Theorem 22.11 ([DFK91]) *For any convex body K and for all $\varepsilon, \delta > 0$, there exists a randomized polynomial-time algorithm that computes an ε-relative error approximation of* $\mathrm{Vol}(K)$ *with probability at least* $1-\delta$ *in time* $\mathrm{poly}(n, \log(R/r), 1/\varepsilon, \log(1/\delta))$.

To understand this algorithm, let $B_n \subseteq K \subseteq RB_n$ without loss of generality. Then let $K_i = 2^{i/n} B_n \cap K$ so that $K_0 = B_n$ and $K_m = K$, for m large enough. Observe that the volume of K can be written as the telescoping product

$$\mathrm{Vol}(K) = \mathrm{Vol}\left(B_n\right) \prod_{i=1}^{m} \frac{\mathrm{Vol}(K_i)}{\mathrm{Vol}(K_{i-1})}.$$

Therefore, we can reduce the problem of computing volume to estimating the ratio $\mathrm{Vol}(K_{i-1})/\mathrm{Vol}(K_i)$.

Suppose for now that we can sample a point $\boldsymbol{x} \sim K_i$ uniformly at random, and let the random variable

$$Y_i = \begin{cases} 1 & \text{if } \boldsymbol{x} \in K_{i-1} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly we have

$$\mathrm{E}[Y] = \frac{\mathrm{Vol}(K_{i-1})}{\mathrm{Vol}(K_i)},$$

which we can estimate by sampling k points independently from K_i uniformly at random and letting

$$W_i = \frac{1}{k} \sum_{j=1}^{k} Y_{i,j}.$$

Using the product formula above for $\mathrm{Vol}(K)$, we let the random variable

$$V = \mathrm{Vol}(B_n) \prod_{i=1}^{m} \frac{1}{W_i},$$

so that $\mathrm{E}[V] = \mathrm{Vol}(K)$. For such a sampling-based approximation to be efficient, we need to show that the probability distribution of V is concentrated around its mean.

We start with the following lemma, which we use to bound the variance of our estimator.

Lemma 22.12 *For all $i \in [m]$, we have*

$$\mathrm{Vol}\left(V_i\right) \leq 2\mathrm{Vol}\left(V_{i-1}\right).$$

Proof: By the definition of K_i, we have

$$\begin{aligned} K_i &= 2^{i/n} B_n \cap K \\ &\subseteq 2^{1/n} \left(2^{(i-1)/n} B_n \cap K\right) \\ &= 2^{1/n} K_{i-1}, \end{aligned}$$

which completes the proof. ■

We use Chebyshev's inequality to analyze the concentration of V. Observing that

$$\Pr\left(|V - \mathrm{E}[V]| \geq \varepsilon \mathrm{E}[V]\right) \leq \frac{\mathrm{Var}(V)}{\varepsilon^2 \mathrm{E}[V]^2},$$

it suffices to show that

$$\frac{\mathrm{Var}(V)}{\mathrm{E}[V]^2} \le \frac{\varepsilon^2}{2}.$$

Using the independence of the estimators W_i, we have

$$\begin{aligned}\frac{\mathrm{Var}(V)}{\mathrm{E}[V]^2} &= \frac{\mathrm{E}\left[V^2\right] - \mathrm{E}[V]^2}{\mathrm{E}[V]^2} \\ &= \frac{\mathrm{E}\left[V^2\right]}{\mathrm{E}[V]^2} - 1 \\ &= \frac{\prod_{i=1}^m \mathrm{E}\left[W_i^2\right]}{\prod_{i=1}^m \mathrm{E}\left[W_i\right]^2} - 1 \\ &= \prod_{i=1}^m \left(1 + \frac{\mathrm{Var}\left(W_i\right)}{\mathrm{E}\left[W_i\right]^2}\right) - 1.\end{aligned}$$

Therefore, it further suffices to show that

$$\frac{\mathrm{Var}\left(W_i\right)}{\mathrm{E}\left[W_i\right]^2} \le \frac{\varepsilon^2}{4m},$$

as it implies that

$$\begin{aligned}\frac{\mathrm{Var}(V)}{\mathrm{E}[V]^2} &= \prod_{i=1}^m \left(1 + \frac{\mathrm{Var}\left(W_i\right)}{\mathrm{E}\left[W_i\right]^2}\right) - 1 \\ &\le \left(1 + \frac{\varepsilon^2}{4m}\right)^m - 1 \\ &\le \exp\left(\frac{\varepsilon^2}{4}\right) - 1 \\ &\le \frac{\varepsilon^2}{2}.\end{aligned}$$

The middle inequalities use the fact that $1 + x \le e^x \le 1 + 2x$, for $0 \le x \le 1$. Letting $\mathrm{E}[Y_i] = p$ be the ratio of consecutive volumes, we have

$$\begin{aligned}\frac{\mathrm{Var}(W_i)}{\mathrm{E}[W_i]^2} &= \frac{k^{-1}\mathrm{Var}(Y_i)}{\mathrm{E}[Y_i]^2} \\ &= \frac{1}{k}\left(\frac{1}{p} - 1\right) \\ &\le \frac{1}{k},\end{aligned}$$

since $p \ge 1/2$ by Lemma 22.12. Thus, it suffices to sample $k = O(m/\varepsilon^2)$ points per phase. Moreover, because we have $m = O(n \log R)$, the total sample complexity of the algorithm is

$$O\left(\frac{m^2}{\varepsilon^2}\right) = O\left(\frac{n^2 \log^2 R}{\varepsilon^2}\right).$$

22.3.1 Markov Chain Monte Carlo Algorithms

To use this approximation scheme, it remains to design an efficient algorithm for uniformly sampling from a convex body K. Several classic Markov chain Monte Carlo algorithms, all of which require a membership oracle, have been studied extensively: grid walk, ball walk, and hit-and-run [Vem05]. Several new ideas for sampling from polytopes include the Dikin walk, geodesic walks, and Riemannian Hamiltonian Monte Carlo.

Given some initial point $\boldsymbol{x}_0 \in K$, the BallWalk(δ) Markov chain tries to iteratively step to a random point within distance δ of the current point. The state space is the entire set K, and this Markov chain is lazy to guarantee that the stationary distribution is unique.

BallWalk(δ):

- Pick a uniform random point $\boldsymbol{y}$ from the ball of radius δ centered at the current point $\boldsymbol{x}$.
- If $\boldsymbol{y}$ is in K, go to $\boldsymbol{y}$; else stay at $\boldsymbol{x}$.

A problem with BallWalk(δ) is that it requires a "warm start" close to the uniform distribution. Otherwise, it can take exponentially long to escape from corners (e.g. a high-dimensional cube). Under the additional assumption that K is isotropic, the KLS conjecture [LV16] implies that the mixing time of this Markov chain is $\widetilde{O}(n^{2.5})$ from a warm start.

The HitAndRun Markov chain picks a random point along a random line through the current point, and it does not need a step-size parameter like BallWalk(δ).

HitAndRun:

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

Unlike the ball walk Markov chain, the mixing time of HitAndRun is $\widetilde{O}(n^2 R^2)$, regardless of the initial distribution [LV06].

References

[BF87] Imre Bárány and Zoltán Füredi, "Computing the volume is difficult," *Discrete & Computational Geometry* **2.4** (1987), pp. 319–326.

[Cou17] Benjamin Cousins, "Efficient high-dimensional sampling and integration," *Dissertation, Georgia Institute of Technology* (2017).

[DF88] Martin Dyer and Alan Frieze, "On the complexity of computing the volume of a polyhedron," *SIAM Journal on Computing* **17.2** (1988), pp. 967–974.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan, "A random polynomial-time algorithm for approximating the volume of convex bodies," *Journal of the ACM* **38.1** (1991), pp. 1–17.

[Ele86] György Elekes, "A geometric inequality and the complexity of computing volume," *Discrete & Computational Geometry* **1.1** (1986), pp. 289–292.

[LV16] Yin Tat Lee and Santosh Vempala, "Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion," *Preprint, arXiv 1612.01507* (2016).

[LV06] László Lovász and Santosh Vempala, "Hit-and-run from a corner," *SIAM Journal on Computing* **35.4** (2006), pp. 985–1005.

[Vem05] Santosh Vempala, "Geometric random walks: A survey," *Combinatorial and Computational Geometry* **5.2** (2005), pp. 573–612.