CS294 Markov Chain Monte Carlo: Foundations & Applications	**Fall 2009**

Lecture 17: November 3rd

Lecturer: Alistair Sinclair *Scribes:*

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.*

17.1 Volume of a Convex Body

Our main goal in this lecture is to analyze the mixing time of the (heat bath) Ball Walk on a convex body K, as defined in the previous lecture. In this walk, from any point $x \in K$, we move to a new point chosen uniformly at random in $B(x,\delta) \cap K$, where $B(x,\delta)$ denotes the ball of radius δ centered at x, and $\delta = \Theta(\frac{1}{\sqrt{n}}$. Recall that this walk has stationary measure $\mu(x) = \frac{\ell(x)}{L}$, where $L = \int_K \ell(x)dx$ is the normalizing factor and the density $\ell(x)$ is defined by

$$\ell(x) = \frac{\mathrm{Vol}(B(x,\delta) \cap K)}{\mathrm{Vol}(B(x,\delta))}.$$

We will throughout assume a bound on the curvature of K given by the *Curvature Assumption*, which together with the value of δ (as we saw in the last lecture) ensures that $\ell(x) \geq 0.4$ for all $x \in K$. This in turn means that we can effectively implement the ball walk, and also that its stationary distribution is not too far from uniform. (In particular, given samples from μ we can produce uniform random samples by rejection sampling with expected constant overhead.)

17.2 Mixing Time

We will analyze the mixing time by examining how the variance of a measurable function φ decays as the Ball Walk evolves. Recall that in the case of discrete Markov chains, we defined the Poincaré constant

$$\alpha := \inf_{\varphi} \frac{\mathcal{E}_\pi(\varphi,\varphi)}{\mathrm{Var}_\pi(\varphi)} = \frac{\sum_{x,y}(\varphi(x)-\varphi(y))^2\pi(x)P(x,y)}{\sum_{x,y}(\varphi(x)-\varphi(y))^2\pi(x)\pi(y)},$$

where the infimum is over all non-constant functions $\varphi : \Omega \to \mathbf{R}$. The quantity α has a natural interpretation as the ratio of "local variation" to "global variation" of any function φ.

We then showed that the following holds for any lazy discrete Markov chain:

$$\tau_x(\epsilon) \leq O\big(\frac{1}{\alpha}(\log(\pi(x)^{-1}) + \log(\epsilon^{-1})\big). \tag{17.1}$$

Thus the mixing time is controlled by α, modulo a dependence on the starting state x. It turns out that a version of (17.1) (suitably modified) extends to the continuous setting of our Ball Walk. To state this, define

$$\begin{aligned} h_\varphi(x) &:= \frac{1}{2}\int_K (\varphi(x)-\varphi(y))^2 P(x,dy) \\ \alpha &:= \inf_\varphi \frac{\int_K h_\varphi(x)\mu(dx)}{\int_K \varphi(x)^2\mu(dx) - (\int_K \varphi(x)\mu(dx))^2} \equiv \inf_\varphi \frac{\int_K h d\mu}{\int_K \varphi^2 d\mu - (\int_K \varphi d\mu)^2}, \end{aligned}$$

where the infimum is over non-constant, measurable, real-valued functions φ.[1] Note that this is exactly the analog of α as defined in the discrete case; the numerator is a "local variation" and the denominator is just the "global variation" (or variance) of φ.

We then have the following claim,

Claim 17.1 *Suppose that the Ball Walk is started in an initial distribution that is uniform in a ball $B(x,\delta) \subseteq K$. Then the mixing time satisfies*

$$\tau_x(\epsilon) \le O\Big(\frac{1}{\alpha}\left(n\log(D/2\delta) + \log\epsilon^{-1}\right)\Big)$$

where $D = \sup_{x,y\in K}\|x-y\|_2$ is the diameter of the convex body K.

Note that this has essentially the same form as (17.1). The only significant difference is that the dependence on the initial distribution is replaced by the term $n\log(D/2\delta)$, reflecting the fact that we begin at the uniform distribution in a δ-ball.[2]

We shall prove this general Claim in the next lecture. For now we just assume it. As a result, our analysis of the mixing time of the Ball Walk will boil down to the following:

Theorem 17.2 *For the Ball Walk, under the curvature assumption and with $\delta = \frac{c}{\sqrt{n}}$ for a suitable constant c, we have*

$$\alpha \ge \frac{C\delta^2}{D^2 n}$$

for some universal constant C.

An immediate consequence of this gives our main result:

Corollary 17.3 *For the Ball Walk started in an initial distribution that is uniform in a ball $B(x,\delta) \subseteq K$, under the curvature assumption and with $\delta = \frac{c}{\sqrt{n}}$, we have*

$$\tau_{mix} = O\big(\frac{D^2 n}{\delta^2} n\log(D/2\delta)\big) = O(D^2 n^3 \log(D\sqrt{n})) = O(n^4\log n).$$

Note that we cannot expect to do much better than this, for the following reason. In one step of the Ball Walk, the expected distance traveled parallel to any fixed axis is $O(\delta/\sqrt{n}) = O(1/n)$. So in particular, the expected number of steps needed to explore the entire "longest" axis of the body (whose length is the diameter, D), is about $O(D^2 n^2)$. This is only about a factor of n less than the bound we obtain.

We will first give a "big picture" proof of Theorem 17.2 before proceeding to the technical details. The proof is by contradiction, and is in three parts, following the exposition in [Jer]. Define $\alpha^* = \frac{C\delta^2}{D^2 n}$ and assume for contradiction that

$$\frac{\int_K h\,d\mu}{\int_K \varphi^2 d\mu} < \alpha^* \tag{17.2}$$

for some measurable function φ with $\int \varphi d\mu = 0$. (Note that we can always assume w.l.o.g. that the expectation $\int \varphi d\mu = 0$ simply by adding an appropriate constant to φ.) We will find successively smaller subsets

[1] From now on, where no confusion arises, we will omit the subscript φ from h and also suppress the variable of integration by abreviating expressions like $\int h(x)\mu(dx)$ to $\int h d\mu$. Thus, for example, $\int \varphi d\mu$ is just the μ-expectation $\int_K \varphi(x)\mu(dx)$ of φ.

[2] The previous value $\log \pi(x)^{-1}$ would clearly have no sensible meaning in the continuous setting.

of K for which (17.2) (or a similar condition) continues to hold. Eventually we will end up with a set that is so small that (17.2) cannot hold. Thus we get a contradiction.

Here are the three stages of the proof:

Proof Outline:

1. There exists a convex subset $K_1 \subseteq K$ that also satisfies inequality (17.2) and is "needle-like", i.e.,

$$K_1 \subseteq [0, \epsilon]^{n-1} \times [0, D]$$

   for an arbitrarily small $\epsilon > 0$.

2. We subdivide K_1 into sections of length $\eta = c' \frac{\delta}{\sqrt{n}}$ along its long axis. We show that some region K_0 consisting of a pair of adjacent such sections satisfies

$$\frac{\int_{K_0} h \, d\mu}{\int_{K_0} (\varphi - \bar{\varphi})^2 d\mu} \leq \frac{1}{10}, \tag{17.3}$$

   where $\bar{\varphi} = \frac{1}{\mu(K_0)} \int_{K_0} \varphi d\mu$ is the expectation of φ on K_0. Thus K_0 also satisfies (17.2) but with a weaker bound of $\frac{1}{10}$ rather than α. Here we can take $c' > 0$ to be as small as we like.

3. We show that (17.3) is a contradiction, since K_0 is so small that its local and global variation cannot differ by much (since the reachable regions for any pair of points in K_0 have large intersection).

Proof: We prove the theorem in the three steps outlined above. Part 1 uses a classical argument due to Payne & Weinberger together with an induction on dimension. Part 2 is mostly technical and will be omitted. Part 3 involves a direct calculation.

Proof of Part 1: We call a dimension "fat" if the size of the projection of K onto it is $> \epsilon$. We proceed by induction on the number of fat dimensions. Given a set K_j with $j \geq 2$ fat dimensions that satisfies (17.2) for some φ with $\int_{K_j} \varphi d\mu = 0$, we construct a convex set K_{j-1} which has only $j-1$ fat dimensions for which φ still satisfies (17.2). By repeated application, we will end up with a set K_1 which has only one fat dimension, as required.

So let $K_j \subseteq [0, D]^j \times [0, \epsilon]^{n-j}$, and let S be the projection of K onto any two of the fat dimensions. We need a basic fact about convex sets S in two dimensions:

Fact: For any convex set $S \subseteq \mathbf{R}^2$, there exists a point $x \in S$ such that every line through x cuts S into two pieces each of size $\geq \frac{1}{3}\text{Area}(S)$.

Consider the hyperplanes through x and orthogonal to S. They partition $\mathbf{R}^n$ into H^+ and H^-. Now because $\int_{K_j} \varphi d\mu = 0$ we have that

$$\int_{K_j \cap H^+} \varphi d\mu + \int_{K_j \cap H^-} \varphi d\mu = 0$$

and so we can find a hyperplane so that

$$\int_{K_j \cap H^+} \varphi d\mu = \int_{K_j \cap H^-} \varphi d\mu = 0$$

(If $\int_{K_j \cap H^+} \varphi d\mu$ is positive then after rotating the hyperplane 180 degrees it will be negative because before the rotation $\int_{K_j \cap H^-} \varphi d\mu$ was negative. Now just use the intermediate value theorem.)

(a) (b)

Figure 17.1:

Since K_j satisfies (17.2), at least one of $K_j \cap H^+$ and $K_j \cap H^-$ must satisfy it also. Moreover, by the above Fact, the projection of both of these bodies onto the plane of S has area at most $\frac{2}{3}\text{Area}(S)$. Now replace K_j by the appropriate half and repeat this procedure until $\text{Area}(S) \le \frac{1}{2}\epsilon^2$.

Finally, the width of a convex set of area A in $\mathbf{R}^2$ is $\le \sqrt{2A}$, so $\text{width}(S) \le \epsilon$ and the number of fat dimensions is reduced by one. Note that the above argument allows us to achieve any desired value $\epsilon > 0$.

Proof of Part 2 This step is rather technical and is omitted. For the details, see [Jer]. In this part we use the value of δ.

Proof of Part 3 Set $\delta' = \delta - \epsilon\sqrt{n}$. Let K_0 be the set that we are left with after Part 2. K_0 is contained in some prism of dimension $[0, \frac{c'\delta}{\sqrt{n}}] \times [0, \epsilon]^{n-1}$ (see figure 17.1(a)), where ϵ and c' are arbitrarily small positive values. Let z_1 and z_2 be the midpoints of the ($(n-1)$-dimensional) end faces of this prism.

We focus on the set $I := B(z_1, \delta') \cap B(z_2, \delta') \cap K$.

Fact 1: Any point in I can be reached in one step of ball walk by every point in the prism (and thus by every point in K_0, see figure 17.1(b)).

This Fact follows easily from our definition of $\delta' = \delta - \epsilon\sqrt{n}$.

Fact 2: $\text{Vol}(I) \ge \frac{1}{5}\text{Vol}(B(0, \delta))$.

This Fact follows from the Curvature Assumption, in similar fashion to our argument in the previous lecture that showed $\text{Vol}(B(x, \delta) \cap K) \ge \frac{2}{5}\text{Vol}(B(x, \delta))$. Here we are working with the intersection of two such balls whose centers, z_1 and z_2, are very close compared to their radii δ' (recall that both ϵ and η can be chosen very small). This causes us to lose at most a factor of 2 in the argument.

We will now show that, for any function φ,

$$\frac{\int_{K_0} h\, d\mu}{\int_{K_0} (\varphi - \bar{\varphi}) d\mu} > \frac{1}{10}$$

which will give us our desired contradiction. From Fact 1 and the definition of h and the ball walk transition density $P(x, dy)$:

$$\int_{K_0} h d\mu \ge \frac{1}{2} \int_{K_0} \frac{\mu(dx)}{\text{Vol}(B(x, \delta) \cap K)} \int_I (\varphi(x) - \varphi(y))^2 dy$$

Now applying Fubini's Theorem to interchange the order of integration, this becomes

$$\int_I dy \int_{K_0} \frac{(\varphi(x) - \varphi(y))^2}{\text{Vol}(B(x, \delta) \cap K)} \mu(dx) \ge \frac{1}{2\text{Vol}(B(0, \delta))} \int_I dy \int_{K_0} (\varphi(x) - \varphi(y))^2 \mu(dx)$$

Now, using the fact that $\varphi(y)$ is constant in the inner integral, and that $\int(\varphi - c)^2 d\mu$ is minimized by taking $c = \frac{1}{\mu(K_0)}\int_{K_0}\varphi d\mu$, i.e., the μ-expectation of φ, the above expression is bounded below by

$$\frac{1}{2\mathrm{Vol}(B(0,\delta))}\int_I dy \int_{K_0}(\varphi - \bar{\varphi}(y))^2 d\mu \geq \frac{1}{10}\int_{K_0}(\varphi - \bar{\varphi}(y))^2 d\mu,$$

where in this final inequality we have used Fact 2.

This completes the proof. ∎

References

[Jerrum] M.R. Jerrum, "Volume of a convex body," Draft lecture notes, available at http://homepages.inf.ed.ac.uk/chap6.ps.