Statistics 291: Lecture 10 (February 22, 2024)

Concentration for Langevin Dynamics

Instructor: Mark Sellke

Scribe: Alan Chung

1 Introduction

In the last lecture, we looked at fast mixing for Langevin Dynamics at high temperature on the sphere, and also for log-concave distributions on the full space (i.e., on $\mathbb{R}^N$). Today, we will look at mixing on $O(1)$ time scales for general β. There exist some ways to do this by considering some complicated set of equations, either using Dynamical Mean-Field theory of via Approximate Message Passing. In this lecture though, we will instead show concentration of the hamiltonian and overlap functions, $\frac{1}{N}H_n(X_t)$ and $R(X_s, X_t)$.

One might imagine that we should strive to show that for $t = O(1)$, that $\frac{1}{N}H_N(X_t)$ is a Lipschitz function of $(H_N, B_{[0,t]}, \ldots)$, but this turns out not to be true. In particular, considering the diffusion equation

$$dX_t = \sqrt{2}P^{\perp}_{X_t} dB_t + \left(\beta\nabla_{sph}H_N(X_t) - \left(\frac{N-1}{N}\right)X_t\right)dt,$$

the projection $P^{\perp}_{X_t}$ prevents things from being Lipschitz. Even if this projection term did not cause problems, we would still need a concentration result in an infinite-dimensional space for $B_{[0,t]}$, which also presents challenges. Actually, there are standard ways to fix this Lipschitz issue, one of which is- to consider "soft spherical dynamics," such as in the equation

$$dX_t = \sqrt{2}dB_t + (\beta\nabla_{sph}H_N(X_t) - \nabla V_N(X_t)),$$

where this V_N is a "confining potential" that helps project things back onto the sphere. In particular, such a potential might have the form

$$V_N(x) = \frac{\lambda}{N}(||x||^2 - N)^2 + \frac{||x||^{2p}}{N^{p-1}}.$$

However, it turns out that the process is still not Lipschitz if X_t is unusually large.

2 Discrete-time Langevin Dynamics for $H_{N,p}$

In the remainder of the lecture, we will consider the pure p-spin hamiltonian. For ease of notation, however, we will drop the subscript p and write this simply as H_N. The dynamics are given by

$$X^{(k+1)} = \text{Proj}\left(X^{(k)} + \beta\nabla_{sph}H_N(X^{(k)}) + \sqrt{2}g_k\right) \tag{1}$$

$$\text{Proj}(x) = \begin{cases} \frac{x\sqrt{N}}{||x||} & ||x|| \geq \frac{\sqrt{N}}{2} \\ 2x & ||x|| < \frac{\sqrt{N}}{2} \end{cases},$$

where (g_k) is a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables. Here, we are considering a step size of 1 for convenience, and also we note that the projection function is 2-Lipschitz.

Definition 2.1 (C-bounded hamiltonian)**.** We say (in the context of these lectures) that a hamiltonian H_N is C-bounded (of order 2) if

$$\sup_{||x||\le\sqrt{N}} \|H_N(x)\| \le CN, \qquad \sup_{||x||\le\sqrt{N}} \|\nabla H_N(x)\| \le C\sqrt{N}, \qquad \sup_{||x||\le\sqrt{N}} \left\|\nabla^2 H_N(x)\right\|_{op} \le C$$

Lemma 2.2. *Suppose that $H_N, \tilde{H}_N$ are C-bounded. Suppose that $X^{(k)}$ evolves according to eq.* (1) *with hamiltonian H_N and gaussian variables (g_k), and $\tilde{X}^{(k)}$ evolves similarly but according to hamiltonian $\tilde{H}_N$ and randomness $(\tilde{g}_k)$. If $X^{(0)} = \tilde{X}^{(0)}$, then*

$$\left\|X^{(k)} - \tilde{X}^{(k)}\right\| \le (10Cp(\beta+1))^k \left(\left\|G_N^{(p)} - \tilde{G}_N^{(p)}\right\| + \sum_{i=0}^{k-1} \|g_i - \tilde{g}_i\|\right)$$

Sketch of Proof. The proof follows by induction. We saw in the last lecture that if H_N if C-bounded, then

$$\left\|\nabla H_N(X^{(k)}) - \nabla H_N(\tilde{X}^{(k)})\right\| \le C\left\|X^{(k)} - \tilde{X}^{(k)}\right\|.$$

We also have that if $||y|| \le \sqrt{N}$, then

$$\left\|\nabla H_N(y) - \nabla \tilde{H}_N(y)\right\| \le p\left\|G_N^{(p)} - \tilde{G}_N^{(p)}\right\|.$$

To see this, note that if $y = (\sqrt{N}, 0, 0, \dots, 0)$, then this bound is

$$\sqrt{p^2(g_{1,1,\dots,1} - \tilde{g}_{1,1,\dots,1})^2 + \sum_{i=2}^{N}\sum_{sym}(g_{1,1,\dots,i} - \tilde{g}_{1,1,\dots,i})^2} \le p\left\|G_N^{(p)} - \tilde{G}_N^{(p)}\right\|.$$

The triangle inequality then implies that

$$\begin{aligned}\left\|\nabla H_N(X^{(k)}) - \nabla \tilde{H}_N(\tilde{X}^{(k)})\right\| &\le \left\|\nabla H_N(X^{(k)}) - \nabla H_N(\tilde{X}^{(k)})\right\| + \left\|\nabla H_N(\tilde{X}^{(k)}) - \nabla \tilde{H}_N(\tilde{X}^{(k)})\right\| \\ &\le C\left\|X^{(k)} - \tilde{X}^{(k)}\right\| + p\left\|G_N^{(p)} - \tilde{G}_N^{(p)}\right\|.\end{aligned}$$

Then, using the fact that Proj is 2-Lipschitz, we have that

$$\|X^{(k)} - \tilde{X}^{(k)}\| \le 2\left((C\beta+1)\left\|X^{(k)} - \tilde{X}^{(k)}\right\| + p\beta\left|\|G_N^{(p)} - \tilde{G}_N^{(p)}\right\| + \sqrt{2}\left\|g_k - \tilde{g}_k\right\|\right).$$

Then using induction suffices for the proof. □

Before proving the main result, we first state the following:

Theorem 2.3 (Kirszbraun Extension Theorem)**.** *Let $\mathbb{R}^{d_1}, \mathbb{R}^{d_2}$ be Euclidean spaces. Suppose that $U \subseteq \mathbb{R}^{d_1}$ and that $\phi: U \to H_2$ is L-Lipschitz. Then, there exists an extension Φ of ϕ so that $\Phi: \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ that is L-Lipschitz, and so that*

$$clo\big(Convex\ Hull(\Phi(\mathbb{R}^{d_1})\big) = clo\big(Convex\ Hull(\Phi(U)\big)$$

Sketch of Proof. Assume WLOG that $L = 1$. It suffices to show this for the case where S is finite, and we extend the map to one additional point. This suffices if the underlying space is separable, since we just extend to a countable dense subset. Hence let $S = (x_1, x_2, \dots, x_m)$, and we extend to one more point $x \in \mathbb{R}^{d_1}$.

Given $(x_1, x_2, \dots, x_m)$ and $(\phi(x_1), \phi(x_2), \dots, \phi(x_m))$, let $y \in \mathbb{R}^{d_2}$ be the point minimizing

$$r = \max_{1\le i\le m} \frac{\|y - \phi(x_i)\|}{\|x - x_i\|} \overset{?}{\le} 1.$$

It suffices to show this latter inequality, since that would imply that the extension has Lipschitz constant 1. Firstly, we note that $y \in \text{Convex Hull}(\phi(x_1), \ldots, \phi(x_m))$. To see this, suppose that

$$r = \frac{\|y - \phi(x_i)\|}{\|x - x_i\|}$$

holds exactly on $1 \le i \le j$. Then, $y \in \text{Convex Hull}(\phi(x_1), \ldots, \phi(x_j))$. If this is not the case, we can move the point y slightly closer to the Convex Hull to minimize the objective further.

Hence, we write y as a convex combination $y = \sum_{i=1}^{m} p_i \phi(x_i)$. Then, consider i.i.d. random vectors Z, Z' so that $Z = x_i$ with probability p_i. Then note that $\mathbf{E}[\phi(Z)] = \mathbf{E}[\phi(Z')] = y$. Then, that ϕ is 1-Lipschitz implies that

$$\mathbf{E}\left[\left\|\phi(Z) - \phi(Z')\right\|^2\right] \le \mathbf{E}\left[\left\|Z - Z'\right\|^2\right].$$

The left hand side equals $2\mathbf{E}\left[\left\|\phi(Z) - y\right\|^2\right] = 2r^2\mathbf{E}\left[\|Z - x\|^2\right]$. The right hand side satisfies $\mathbf{E}\left[\left\|Z - Z'\right\|^2\right] = 2\mathbf{E}\left[\|Z - \mathbf{E}[Z]\|^2\right] \le 2\mathbf{E}\left[\|Z - x\|^2\right]$. Then this implies that $r \le 1$. □

3 Main Theorem

We first write the main theorem, then provide a sketch on how to prove it.

Theorem 3.1. *Fix $X^{(0)} \in \mathbb{S}^N$. For $k = O(1)$ and large N, there exists $A_k(G_N^{(p)}, g_0, g_1, \ldots, g_{k-1})$ so that*

(a) $\mathbb{P}\left(A_k(G_N^{(p)}, g_0, g_1, \ldots, g_{k-1}) = X^{(k)}\right) \ge 1 - e^{-N}$.

(b) *A_k is $O_{k,\beta}(1)$-Lipschitz.*

(c) $\left\|A_k(G_N^{(p)}, g_0, g_1, \ldots, g_{k-1})\right\| \le \sqrt{N}$.

The idea behind the proof is that we first define ϕ on C-bounded H_N as in Theorem 2.2, and then extend to all H_N using the Kirszbraun Extension Theorem, which implies that the extension is Lipschitz, and then we use concentration.

Corollary 3.2. *Fix $j, k \in \mathbb{Z}_+$, and $c = c(j, k, \beta, \epsilon)$. Then*

$$\mathbb{P}\left(\left|R\left(X^{(j)}, X^{(k)}\right) - \mathbf{E}\left[R\left(X^{(j)}, X^{(k)}\right)\right]\right| \ge \epsilon\right) \le e^{-cN}.$$

Proof Sketch. We note that $R(X^{(j)}, X^{(k)}) = \frac{1}{N}\left\langle X^{(j)}, X^{(k)}\right\rangle$ is $2/\sqrt{N}$-Lipschitz on

$$\{(X^{(j)}, X^{(k)}) : \|X^{(j)}\|, \|X^{(k)}\| \le \sqrt{N}\}.$$

Then, consider the following composition of functions

$$\left(G_N^{(p)}, g_0, \ldots, g_{k-1}\right) \mapsto \left(A_j(\ldots), A_k(\ldots)\right) \mapsto R(A_j(\ldots), A_k(\ldots)).$$

Because both functions are Lipschitz, then the composed function has Lipschitz constant that is the product of the Lipschitz constants of the two individual functions, which is $O_{k,\beta}\left(\frac{1}{\sqrt{N}}\right)$. The result then follows through Lipschitz concentration. □

Corollary 3.3. *Fix $k \in \mathbb{Z}_+$. Then for $c = c(k, \beta, \epsilon)$, we have*

$$\mathbb{P}\left(\left|\frac{1}{N}H_N(X^{(k)}) - \mathbf{E}\left[\frac{1}{N}H_N(X^{(k)})\right]\right| \ge \epsilon\right) \le e^{-cN}.$$

Proof Sketch. For C-bounded H_N, the function $X^{(k)} \mapsto \frac{1}{N} H_N(X^{(k)})$ is $C/\sqrt{N}$−Lipschitz. Hence, $\{\frac{1}{N} H_N(X^{(k)}\}$ has $O_{k,\beta}(1/\sqrt{N})$ Lipschitz modification $E_k\left(G_N^{(p)}, g_0, \ldots, g_{k-1}\right)$. The result then follows through Lipschitz concentration. □

For the next corollary, define the response $Y(k, j) = R(X^{(k)}, g_j)$ for $j < k$. This captures how the noise at previous time steps affects the future dynamics.

Corollary 3.4.

$$\mathbb{P}\left(\left|Y(k, j) - \mathbf{E}[Y(k, j)]\right| \geq \epsilon\right) \leq e^{-cN}.$$

Proof Sketch. The idea is that $R(X^{(k)}, g_j)$ is $O(1/\sqrt{N})$-Lipschitz on the set

$$\hat{S} = \left\{(G_N^{(p)}, g_0, \ldots, g_{k-1}), \quad H_N \text{ is } C\text{-bounded and } \|g_j\| \leq 2\sqrt{N}\right\}.$$

Then the proof follows similarly to the previous ones. □

4 Analytical Solutions for Spin Glasses

The paper [CK93] presents an analytical solution for a large range spin-glass model. In particular, for some version of Langevin Dynamics, they define $R(s, t) = \text{plim}_{N\to\infty} R(X_s, X_t)$, $Y(s, t) = \text{plim}_{N\to\infty} R(X_s, B_t)$, $s > t$. Then, these quantities satisfy the following set of differential equations:

$$\begin{aligned}
&Y(s, s) = R(s, s) = 1\\
&\partial_s Y(s, t) = -\mu(s)Y(s, t) + \beta^2 p(p-1)\int_t^s Y(u, t)Y(s, u)R(s, u)^{\rho-2} du\\
&\partial_s R(s, t) = -\mu(s)R(s, t) + \beta^2 p(p-1)\int_0^s R(u, t)Y(s, u)R(s, u)^{p-2} du + \beta^2 p\int_0^t R(s, u)^{p-1} Y(t, u) du\\
&\mu(s) = \frac{1}{2} + \beta^2 p^2 \int_0^s R(s, u)^{p-1} Y(s, u) du
\end{aligned}$$

This was proven formally in [ADG04].

References

[ADG04] Gerard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-kurchan equations for dynamics of spin-glasses, 2004. 4

[CK93] L. F. Cugliandolo and J. Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173–176, July 1993. 4