CS395T: Continuous Algorithms, Part XIV
Langevin algorithms

Kevin Tian

1 W_2^2 convergence of unadjusted Langevin

In Part XIII, we gave tools for understanding the convergence of the Langevin dynamics,

$$\mathrm{d}x_t = -\nabla V(x_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t. \tag{1}$$

For instance, we gave a simple coupling argument showing that when the target stationary distribution $\pi^\star \propto \exp(-V)$ (Theorem 1, Part XIII) is strongly logconcave, then the Langevin dynamics converge linearly in W_2^2 (Theorem 2, Part XIII). Moreover, using tools from Markov semigroup theory, we established that when the stationary distribution satisfies weaker functional inequalities such as Poincaré or log-Sobolev, the Langevin dynamics (1) actually converge under stronger error metrics such as χ^2 or D_{KL}. Unfortunately, these results do not immediately lead to implementable algorithms, because they only hold in continuous time.

Our goal in this lecture is now to give an introduction to convergence guarantees for discrete-time approximate implementations of the Langevin dynamics. In this and the following section, we will specifically focus on the *unadjusted Langevin algorithm* (ULA), which samples x_0 from a starting distribution π_0, and for a step size $\eta > 0$, iterates[1]

$$x^{(k+1)} \leftarrow x^{(k)} - \eta\nabla V(x^{(k)}) + \sqrt{2\eta}\xi^{(k)}, \text{ where } \xi_k \sim \mathcal{N}(\mathbb{0}_d, \mathbf{I}_d). \tag{2}$$

The motivation for considering (2), a forward Euler discretization of the Langevin dynamics, is that it only requires one query to ∇V, as opposed to running (1) which would require an unbounded number of queries. This is entirely analogous to the relationship between gradient descent (a discrete-time algorithm) and its continuous-time counterpart, gradient flow.

It is straightforward to check that the iteration (2) is equivalently induced by the SDE

$$\mathrm{d}x_t = -\nabla V(x_0)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \tag{3}$$

up to time $t = \eta$, initialized at $x_0 \leftarrow x^{(k)}$. In other words, rather than the position-dependent drift $\nabla V(x_t)$ typically used in the Langevin dynamics, ULA uses a constant drift $\nabla V(x_0)$. In this sense, the (discrete-time) ULA is simply an Euler discretization of the (continuous-time) Langevin dynamics, just as gradient descent is an Euler discretization of gradient flow (Part II).

Our strategy for analyzing the convergence of (2) under strong logconcavity, when the error metric is W_2^2, is then fairly straightforward. We first use rapid convergence of the continuous-time Langevin dynamics as in Theorem 2, Part XIII, and then bound the discretization error through a coupling argument. We introduce two standard helper claims which help in our analysis.

Lemma 1. *Let $\pi^\star \propto \exp(-V)$, where $V : \mathbb{R}^d \to \mathbb{R}$ is L-smooth. Then,*

$$\mathbb{E}_{x\sim\pi^\star}\left[\|\nabla V(x)\|_2^2\right] \le Ld.$$

Proof. **kjtian:** This lemma is Homework V, Problem 1. I will update when it is due. □

[1] In this lecture, for consistency with Part XIII, we use superscripts to denote an iteration count for ULA, to contrast with subscripts which are used to indicate the passage of time.

Lemma 2. *Let $\{x_t\}_{t\in[0,\eta]}$ follow (1), where $V : \mathbb{R}^d \to \mathbb{R}$ is L-smooth and $\eta \le \frac{1}{3L}$. Then,*

$$\mathbb{E}\left[\|x_\eta - x_0\|_2^2\right] \le 6\eta^2 \mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 12\eta d.$$

Proof. By using $\|a+b+c\|_2^2 \le 3\|a\|_2^2 + 3\|b\|_2^2 + 3\|c\|_2^2$, we have for any $t \in [0,\eta]$,

$$\begin{aligned}
\mathbb{E}\left[\|x_t - x_0\|_2^2\right] &= \mathbb{E}\left[\left\|-\int_0^t \nabla V(x_s)\mathrm{d}s + \sqrt{2}B_t\right\|_2^2\right] \\
&\le 3t^2\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 3\mathbb{E}\left[\left\|\int_0^t (\nabla V(x_s) - \nabla V(x_0))\mathrm{d}s\right\|_2^2\right] + 6\mathbb{E}\|B_t\|_2^2 \\
&\le 3t^2\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 3t\mathbb{E}\left[\int_0^t \|\nabla V(x_s) - \nabla V(x_0)\|_2^2\,\mathrm{d}s\right] + 6\mathbb{E}\|B_t\|_2^2 \\
&\le 3\eta^2\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 3\eta L^2\mathbb{E}\left[\int_0^t \|x_s - x_0\|_2^2\,\mathrm{d}s\right] + 6\eta d.
\end{aligned} \tag{4}$$

The second-to-last inequality was due to Cauchy-Schwarz, i.e. for $\{v_s\}_{s\in[0,t]} \subset \mathbb{R}^d$,

$$\left\|\int_0^t v_s \mathrm{d}s\right\|_2^2 = \int_0^t\int_0^t \langle v_s, v_{s'}\rangle\,\mathrm{d}s\mathrm{d}s' \le \int_0^t\int_0^t \left(\frac{1}{2}\|v_s\|_2^2 + \frac{1}{2}\|v_{s'}\|_2^2\right)\mathrm{d}s\mathrm{d}s' = t\int_0^t \|v_s\|_2^2\,\mathrm{d}s, \tag{5}$$

and the last inequality in (4) used our smoothness assumption. Therefore, the conclusion follows from a variant of Grönwall's inequality (Fact 1, Part II), which states that if $\{\Phi_t\}_{t\in[0,\eta]}$ satisfies the integral inequality $\Phi_t \le C_1 + C_2\int_0^t \Phi_s \mathrm{d}s$, then $\Phi_\eta \le C_1\exp(C_2\eta)$. We apply this to $\Phi_t := \mathbb{E}[\|x_t - x_0\|_2^2]$ and use the assumption on η, yielding the claim:

$$\mathbb{E}\left[\|x_\eta - x_0\|_2^2\right] \le \exp\left(3\eta^2L^2\right)\left(3\eta^2\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 6\eta d\right) \le 6\eta^2\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 12\eta d.$$

□

We can now analyze the discretization error of one step of the unadjusted Langevin algorithm.

Lemma 3. *Let $V : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and μ-strongly convex. Let $x_0 \sim \pi_0$, let $\{x_t\}_{t\in[0,\eta]}$ follow (3), and let π_η denote the law of x_η. Then, for $\eta \le \frac{\mu}{10L^2}$,*

$$W_2^2(\pi_\eta, \pi^\star) \le \left(1 - \frac{\mu\eta}{2}\right) W_2^2(\pi_0, \pi^\star) + \frac{32\eta^2L^2d}{\mu}.$$

Proof. We first introduce some simplifying notation. Let $\{\bar{x}_t\}_{t\in[0,\eta]}$ follow (1), starting from $\bar{x}_0 = x_0$, and with law $\bar{\pi}_t$ at time $t \in [0,\eta]$. Then the proof of Theorem 2, Part XIII shows that

$$W_2^2(\bar{\pi}_\eta, \pi^\star) \le \exp(-2\mu\eta)\, W_2^2(\pi_0, \pi^\star). \tag{6}$$

Next, applying Lemma 2 (with $\eta \leftarrow t$ for each $t \in [0,\eta]$), and using the coupling γ_η of π_η and $\bar{\pi}_\eta$ which share a copy of Brownian motion driving the respective SDEs, shows that

$$\begin{aligned}
W_2^2(\pi_\eta, \bar{\pi}_\eta) &\le \mathbb{E}_{(x_\eta,\bar{x}_\eta)\sim\gamma_\eta}\left[\|x_\eta - \bar{x}_\eta\|_2^2\right] = \mathbb{E}\left[\left\|\int_0^\eta (\nabla V(x_t) - \nabla V(x_0))\mathrm{d}t\right\|_2^2\right] \\
&\le \eta L^2\mathbb{E}\left[\int_0^\eta \|x_t - x_0\|_2^2\,\mathrm{d}t\right] \le 6\eta^4L^2\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 12\eta^3L^2d.
\end{aligned}$$

In the second-to-last inequality, we again used (5) and smoothness, and in the last inequality, we gained a factor of η by using Lemma 2 at each time $t \in [0,\eta]$. We further have, for the optimal coupling $\gamma \in \mathcal{C}(\pi_0, \pi^\star)$ realizing $W_2^2(\pi_0, \pi^\star)$,

$$\begin{aligned}
\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] &\le 2\mathbb{E}_{x^\star\sim\pi^\star}\left[\|\nabla V(x^\star)\|_2^2\right] + 2\mathbb{E}_{(x_0,x^\star)\sim\gamma}\left[\|\nabla V(x_0) - \nabla V(x^\star)\|_2^2\right] \\
&\le 2Ld + 2L^2\mathbb{E}_{(x_0,x^\star)\sim\gamma}\left[\|x_0 - x^\star\|_2^2\right] = 2Ld + 2L^2W_2^2(\pi_0, \pi^\star),
\end{aligned} \tag{7}$$

using Lemma 1. Combining the above displays and using $\eta L \le \frac{1}{10}$ implies

$$W_2^2(\pi_\eta, \bar{\pi}_\eta) \le 16\eta^3 L^2 d + 12\eta^4 L^4 W_2^2(\pi_0, \pi^\star). \tag{8}$$

Finally, because any three vectors $x_\eta \sim \pi_\eta, \bar{x}_\eta \sim \bar{\pi}_\eta, x_\eta^\star \sim \pi^\star$ satisfy

$$\left\|x_\eta - x_\eta^\star\right\|_2^2 \le (1+\mu\eta)\left\|\bar{x}_\eta - x_\eta^\star\right\|_2^2 + \left(1 + \frac{1}{\mu\eta}\right)\left\|x_\eta - \bar{x}_\eta\right\|_2^2,$$

we combine (6) and (8) to obtain the conclusion:

$$\begin{aligned}
W_2^2(\pi_\eta, \pi^\star) &\le (1+\mu\eta)\, W_2^2(\bar{\pi}_\eta, \pi^\star) + \left(1 + \frac{1}{\mu\eta}\right) W_2^2(\pi_\eta, \bar{\pi}_\eta) \\
&\le (1+\mu\eta)\exp(-2\mu\eta)\, W_2^2(\pi_0, \pi^\star) + \left(1 + \frac{1}{\mu\eta}\right)\left(16\eta^3 L^2 d + 12\eta^4 L^4 W_2^2(\pi_0, \pi^\star)\right) \\
&\le \left(1 - \frac{\mu\eta}{2}\right) W_2^2(\pi_0, \pi^\star) + \frac{32\eta^2 L^2 d}{\mu}.
\end{aligned}$$

□

By iterating upon Lemma 3, we obtain a convergence rate for the unadjusted Langevin algorithm in the W_2^2 error metric. As we will discuss in Section 3, this analysis can be slightly improved.

Theorem 1 (W_2^2 convergence of unadjusted Langevin)**.** *Let $V : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and μ-strongly convex, and let $\kappa := \frac{L}{\mu}$, $\epsilon \in (0,1)$. Let $x^{(0)} \leftarrow \text{argmin}_{x \in \mathbb{R}^d} V(x)$, and consider iterating the update (2) for $0 \le k < K$ with $\eta = \frac{\epsilon^2\mu}{128L^2 d}$. Then, if $\pi^{(K)}$ denotes the law of $x^{(K)}$,*

$$\mu W_2^2\left(\pi^{(K)}, \pi^\star\right) \le \epsilon^2, \text{ for } K \ge \frac{256\kappa^2 d}{\epsilon^2}\log\left(\frac{4d}{\epsilon^2}\right).$$

Proof. Let $\pi^{(k)}$ denote the law of $x^{(k)}$ for all $0 \le k \le K$, and recall that Lemma 5, Part II shows that $W_2^2(\pi^{(0)}, \pi^\star) \le \frac{2d}{\mu}$. Moreover, applying Lemma 3 with $\pi_0 \leftarrow \pi^{(k)}$ and $\pi_\eta \leftarrow \pi^{(k+1)}$ shows

$$W_2^2\left(\pi^{(k+1)}, \pi^\star\right) \le \left(1 - \frac{\mu\eta}{2}\right) W_2^2\left(\pi^{(k)}, \pi^\star\right) + \frac{32\eta^2 L^2 d}{\mu},$$

for each $0 \le k < K$. Recursing upon this guarantee yields

$$W_2^2\left(\pi^{(K)}, \pi^\star\right) \le \exp\left(-\frac{\mu\eta K}{2}\right) W_2^2\left(\pi^{(0)}, \pi^\star\right) + \frac{32\eta^2 L^2 d}{\mu} \cdot \frac{2}{\mu\eta},$$

where we summed a geometric sequence, and our choices of η, K give the claim. □

We remark that we use the more natural error metric μW_2^2 in Theorem 1 as opposed to W_2^2, as it is a scale-invariant quantity in the strong logconcavity parameter μ, and is directly comparable to $D_{\text{KL}}(\cdot\|\pi^\star)$ via Talagrand's transportation inequality, i.e. Lemma 13, Part XIII), which states

$$\begin{gathered}
\frac{\mu}{2} W_2^2(\pi, \pi^\star) \le D_{\text{KL}}\left(\pi\|\pi^\star\right), \\
\text{if } \pi^\star \text{ satisfies a log-Sobolev inequality with constant } \frac{1}{\mu}.
\end{gathered} \tag{9}$$

We also showed in Section 5.2, Part XIII that μ-strong logconcavity implies such a log-Sobolev inequality holds. In the following section, we give an alternative analysis of (2) which shows that we can directly achieve bounds on $D_{\text{KL}}(\pi^{(K)}\|\pi^\star)$, strengthening Theorem 1 as implied by (9).

2 D_{KL} convergence of unadjusted Langevin

Our goal in this section is to give a discrete-time analog of Lemma 10, Part XIII developed by [VW19], which shows rapid convergence of $D_{\mathrm{KL}}(\pi_t \| \pi^\star)$ along the Langevin dynamics when $\pi^\star$ satisfies a log-Sobolev inequality. As in Section 1, the simplest way to measure discretization error is in the W_2^2 metric, as we have already developed such tools (e.g. Lemma 2). We will use Talagrand's transportation inequality (9) to relate these W_2^2 errors back to the function value of interest, i.e. $D_{\mathrm{KL}}(\cdot \| \pi^\star)$. We again start by analyzing the change in KL divergence of the law of an iterate after one step of ULA, which runs the Euler-discretized SDE (3) for time η.

Lemma 4. *Let $V : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and suppose $\pi^\star \propto \exp(-V)$ satisfies a log-Sobolev inequality with constant $\frac{1}{\mu}$. Let $x_0 \sim \pi_0$, let $\{x_t\}_{t\in[0,\eta}$ follow (3), and let π_η denote the law of x_η. Then for $\eta \le \frac{\mu}{10L^2}$,*

$$D_{\mathrm{KL}}(\pi_\eta \| \pi^\star) \le \left(1 - \frac{\mu\eta}{2}\right) D_{\mathrm{KL}}(\pi_0 \| \pi^\star) + 9\eta^2 L^2 d.$$

Proof. Throughout this proof, let $\pi_{0t} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\ge 0}$ be the density corresponding to the joint law of (x_0, x_t), for all $t \in [0, \eta]$. We also use the notation $\pi_{0|t}(x_0 \mid x_t)$ to mean the conditional distribution of x_0 given x_t, and similarly define $\pi_{t|0}(x_t \mid x_0)$, such that

$$\pi_{0t}(x_0, x_t) = \pi_0(x_0)\pi_{t|0}(x_t \mid x_0) = \pi_t(x_t)\pi_{0|t}(x_0 \mid x_t). \tag{10}$$

Our first step is to derive a continuity equation (in the sense of Lemma 6, Part XIII) for the SDE (3). By using the Fokker-Planck equation (Proposition 3, Part XIII), we have that

$$\frac{\partial}{\partial t}\pi_{t|0}(x \mid x_0) = \nabla \cdot \left(\nabla V(x_0)\pi_{t|0}(x \mid x_0)\right) + \Delta \pi_{t|0}(x \mid x_0).$$

Therefore, averaging over $x_0 \sim \pi_0$, we have

$$\begin{aligned}
\frac{\partial}{\partial t}\pi_t(x) &= \int \left(\nabla \cdot \left(\nabla V(x_0)\pi_{t|0}(x \mid x_0)\right) + \Delta \pi_{t|0}(x \mid x_0)\right)\pi_0(x_0)\mathrm{d}x_0 \\
&= \int \left(\nabla \cdot \left(\nabla V(x_0)\pi_{0t}(x_0, x)\right) + \Delta \pi_{0t}(x_0, x)\right)\mathrm{d}x_0 \\
&= \nabla \cdot \left(\pi_t(x)\int \pi_{0|t}(x_0 \mid x)\nabla V(x_0)\mathrm{d}x_0\right) + \Delta\pi_t(x) \\
&= \nabla \cdot \left(\pi_t(x)\mathbb{E}_{x_0\sim\pi_{0|t}}\left[\nabla V(x_0) \mid x_t = x\right]\right) + \Delta\pi_t(x) \\
&= \nabla \cdot \left(\pi_t(x)\nabla \log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\right) + \nabla \cdot \left(\pi_t(x)\mathbb{E}_{x_0\sim\pi_{0|t}}\left[\nabla V(x_0) - \nabla V(x) \mid x_t = x\right]\right).
\end{aligned} \tag{11}$$

Comparing to Eq. (16), Part XIII, we see that the continuity equations differ only by a term that looks like $\mathbb{E}_{x_0\sim\pi_{0|t}}[\nabla V(x_0) - \nabla V(x) \mid x_t = x]$. At this point, our proof is very similar to Lemma 10, Part XIII, except we use the tools from Section 1 to bound the discretization error. Concretely,

$$\begin{aligned}
\frac{\partial}{\partial t}D_{\mathrm{KL}}(\pi_t \| \pi^\star) &= \frac{\partial}{\partial t}\left(\int \pi_t(x)\log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\mathrm{d}x\right) \\
&= \int \log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\nabla \cdot \left(\pi_t(x)\nabla\log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\right)\mathrm{d}x \\
&+ \int \log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\nabla \cdot \left(\pi_t(x)\mathbb{E}_{x_0\sim\pi_{0|t}}\left[\nabla V(x_0) - \nabla V(x) \mid x_t = x\right]\right)\mathrm{d}x \\
&= -\int \left\|\nabla \log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\right\|_2^2 \pi_t(x)\mathrm{d}x \\
&- \int \left\langle \nabla\log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right), \mathbb{E}_{x_0\sim\pi_{0|t}}\left[\nabla V(x_0) - \nabla V(x) \mid x_t = x\right]\right\rangle \pi_t(x)\mathrm{d}x \\
&\le -\frac{1}{2}\int \left\|\nabla \log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\right\|_2^2 \pi_t(x)\mathrm{d}x \\
&+ \frac{1}{2}\int \left\|\mathbb{E}_{x_0\sim\pi_{0|t}}\left[\nabla V(x_0) - \nabla V(x) \mid x_t = x\right]\right\|_2^2 \pi_t(x)\mathrm{d}x \\
&\le -\frac{1}{2}\int \left\|\nabla \log\left(\frac{\pi_t(x)}{\pi^\star(x)}\right)\right\|_2^2 \pi_t(x)\mathrm{d}x + \frac{1}{2}\mathbb{E}_{(x_0,x)\sim\pi_{0t}}\left[\left\|\nabla V(x_0) - \nabla V(x)\right\|_2^2\right],
\end{aligned} \tag{12}$$

where the second line again used $\frac{\partial}{\partial t}\int \pi_t(x)\mathrm{d}x = \frac{\partial}{\partial t}1 = 0$ and substituted (11), the fourth line used integration by parts, the sixth line used $\langle a, b\rangle \le \frac{1}{2}\|a\|_2^2 + \frac{1}{2}\|b\|_2^2$, and the last line used Jensen's inequality. Now, by plugging in the log-Sobolev inequality (in the form of Lemma 8, Part XIII) into (12), as well as our bound from Lemma 2,

$$\frac{\partial}{\partial t} D_{\mathrm{KL}}\left(\pi_t\|\pi^\star\right) \le -\mu D_{\mathrm{KL}}\left(\pi_t\|\pi^\star\right) + 3\eta^2 L^2 \mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] + 6\eta L^2 d.$$

Moreover, using the bound (7) with Talagrand's transportation inequality (9) shows

$$\mathbb{E}\left[\|\nabla V(x_0)\|_2^2\right] \le 2Ld + 2L^2 W_2^2(\pi_0, \pi^\star) \le 2Ld + \frac{4L^2}{\mu} D_{\mathrm{KL}}\left(\pi_0\|\pi^\star\right).$$

Combining the above two displays and using our bound on η finally yields

$$\begin{aligned} &\frac{\partial}{\partial t} D_{\mathrm{KL}}\left(\pi_t\|\pi^\star\right) \le -\mu D_{\mathrm{KL}}\left(\pi_t\|\pi^\star\right) + 9\eta L^2 d + \frac{12\eta^2 L^4}{\mu} D_{\mathrm{KL}}\left(\pi_0\|\pi^\star\right) \\ \implies &\frac{\partial}{\partial t}\left(\exp\left(\mu t\right) D_{\mathrm{KL}}\left(\pi_t\|\pi^\star\right)\right) \le \exp\left(\mu t\right)\left(9\eta L^2 d + \frac{12\eta^2 L^4}{\mu} D_{\mathrm{KL}}\left(\pi_0\|\pi^\star\right)\right). \end{aligned}$$

The conclusion then follows from integrating and using our choice of η:

$$\begin{aligned} D_{\mathrm{KL}}\left(\pi_\eta\|\pi^\star\right) &\le \exp\left(-\mu\eta\right)\left(D_{\mathrm{KL}}\left(\pi_0\|\pi^\star\right) + \eta\exp(\mu\eta)\left(9\eta L^2 d + \frac{12\eta^2 L^4}{\mu} D_{\mathrm{KL}}\left(\pi_0\|\pi^\star\right)\right)\right) \\ &\le \left(1 - \frac{\mu\eta}{2}\right) D_{\mathrm{KL}}\left(\pi_0\|\pi^\star\right) + 9\eta^2 L^2 d. \end{aligned}$$

□

At this point, the same recursion as used in Theorem 1 (with slightly different parameters), using the one-step guarantee in Lemma 4 rather than Lemma 3, yields our desired convergence rate.

Theorem 2 (D_{KL} convergence of unadjusted Langevin)**.** *Let $V : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and suppose $\pi^\star \propto \exp(-V)$ satisfies a log-Sobolev inequality with constant $\frac{1}{\mu}$, and let $\kappa := \frac{L}{\mu}$, $\epsilon \in (0, 1)$. Let $x^{(0)} \sim \pi_0$, and consider iterating the update (2) for $0 \le k < K$ with $\eta = \frac{\epsilon^2\mu}{72L^2 d}$. Then, if $\pi^{(K)}$ denotes the law of $x^{(K)}$,*

$$D_{\mathrm{KL}}\left(\pi^{(K)}\|\pi^\star\right) \le \frac{\epsilon^2}{2} \text{ for } K \ge \frac{144\kappa^2 d}{\epsilon^2}\log\left(\frac{4D_{\mathrm{KL}}\left(\pi^{(0)}\|\pi^\star\right)}{\epsilon^2}\right).$$

Proof. As in the proof of Theorem 1, applying Lemma 4 for K iterations yields

$$\begin{aligned} D_{\mathrm{KL}}\left(\pi^{(K)}\|\pi^\star\right) &\le \exp\left(-\frac{\mu\eta K}{2}\right) D_{\mathrm{KL}}\left(\pi^{(0)}\|\pi^\star\right) + 9\eta^2 L^2 d \cdot \frac{2}{\mu\eta} \\ &\le \exp\left(-\frac{\mu\eta K}{2}\right) D_{\mathrm{KL}}\left(\pi^{(0)}\|\pi^\star\right) + \frac{\epsilon^2}{4} \le \frac{\epsilon^2}{2}. \end{aligned}$$

□

As discussed at the end of Section 1, the assumptions made in Theorem 2 are actually weaker than those in Theorem 1, since strong logconcavity implies a log-Sobolev inequality (but not the other way around). Moreover, Theorem 2 implies Theorem 1 up to constants, via (9). The reason for the scaling $\frac{\epsilon^2}{2}$ in Theorem 2 is that Pinsker's inequality then shows $D_{\mathrm{TV}}(\pi^{(K)}, \pi^\star) \le \epsilon$ as well.

3 The frontier

Source material

Portions of this lecture are based on reference material in [Che24], as well as the author's own experience working in the field.

References

[Che24] Sinho Chewi. *Log-Concave Sampling*. 2024.

[VW19] Santosh S. Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*, pages 8092–8104, 2019.