We study the Unadjusted Langevin Algorithm (ULA): 

\[ x_{k+1} = x_k - \nabla f(x_k) + \sqrt{2} z_k \]

where \( \epsilon > 0 \) is step size and \( z_k \sim \mathcal{N}(0, I) \) is independent Gaussian. ULA is a discretization of the Langevin dynamics in continuous time:

\[ dX_t = -\nabla f(X_t) \, dt + \sqrt{2} \, dW_t \]

where \((W_t)_{t \geq 0}\) is the standard Brownian motion in \( \mathbb{R}^n \).

We say \( \nu \) satisfies log-Sobolev inequality (LSI) with constant \( \alpha > 0 \) if for all probability distribution \( \rho \):

\[ H_\nu(\rho) \leq \frac{1}{\alpha} J_\nu(\rho). \]

Here \( H_\nu(\rho) \) is the KL divergence (relative entropy):

\[ H_\nu(\rho) = \int_{\mathbb{R}^n} \rho(x) \log \frac{\rho(x)}{\nu(x)} \, dx \]

and \( J_\nu(\rho) \) is the relative Fisher information:

\[ J_\nu(\rho) = \int_{\mathbb{R}^n} \rho(x) \| \nabla \log \frac{\rho(x)}{\nu(x)} \|^2 \, dx. \]

If \( \nu = e^{-f} \) is \( \alpha \)-strongly log-concave (if \( f \) is \( \alpha \)-strongly convex), then \( \nu \) satisfies \( \alpha \)-LSI. But LSI is more general than strong log-concavity.

We recall that when \( \nu \) satisfies \( \alpha \)-LSI, along the Langevin dynamics in continuous time, KL divergence converges exponentially fast:

\[ H_\nu(\rho_t) \leq e^{-\alpha t} H_\nu(\rho_0). \]

We prove a similar convergence guarantee along ULA in discrete time up to the biased limit, when \( \nu \) satisfies LSI and smoothness.

We say \( \nu = e^{-f} \) is \( \alpha \)-smooth if \( \nabla f \) is \( \alpha \)-Lipchitz (\( -LI \leq \nabla^2 f \leq LI \)). But note we do not assume \( f \) is convex.

**Theorem:** Assume \( \nu \) satisfies \( \alpha \)-LSI and is \( \alpha \)-smooth. Then ULA with step size \( 0 < \epsilon \leq \frac{1}{2L} \) satisfies:

\[ H_\nu(\rho_k) \leq e^{-\alpha k} H_\nu(\rho_0) + \frac{c^2 \epsilon L^2}{\alpha}. \]

Suppose we start from \( x_0 \sim \rho_0 = \mathcal{N}(x^*, \frac{1}{L} I) \) where \( x^* \) is a stationary point for \( f \) (\( \nabla f(x^*) = 0 \)), so \( H_\nu(\rho_0) = \tilde{O}(n) \). The theorem above implies the following iteration complexity for ULA.

**Corollary:** Assume \( \nu \) satisfies \( \alpha \)-LSI and is \( \alpha \)-smooth. For \( \delta > 0 \), to reach \( H_\nu(\rho_k) \leq \delta \), it suffices to run ULA with step size \( \epsilon = \Theta\left(\frac{\frac{1}{L} \epsilon^2}{\delta \alpha^2}n^2\right) \) for the number of iterations:

\[ k = \tilde{O}\left(\frac{n L^2}{\epsilon^2 \delta \alpha^2}n^2\right). \]

This is the same complexity as previous results for ULA under strong log-concavity, but our result holds under more general condition (LSI).

We show KL divergence decreases by a constant factor in each step of ULA, with an additional \( O(\epsilon^2) \) error term. Iterating this bound yields the result above with \( O(\epsilon) \) bias.

**Lemma:** Assume \( \nu \) satisfies \( \alpha \)-LSI and is \( \alpha \)-smooth. Then ULA with step size \( 0 < \epsilon \leq \frac{1}{2L} \) satisfies:

\[ H_\nu(\rho_{k+1}) \leq e^{-\alpha \epsilon} H_\nu(\rho_k) + \epsilon^2 n L^2. \]

**Proof Idea:**

1. We compare one step of ULA with the Langevin dynamics.
2. We use Talagrand’s inequality to bound the difference.

We can show when \( \nu \) satisfies \( \alpha \)-LSI, Rényi divergence converges exponentially fast along the Langevin dynamics:

\[ R_{\epsilon, \nu}(\rho) \leq e^{-\frac{\epsilon^2 n L^2}{c^2}} R_{\epsilon, \nu}(\rho_0). \]

The theorem case \( q = 1 \) recovers KL divergence: \( \lim_{q \to 1} R_{\epsilon, \nu}(\rho) = H_\nu(\rho) \). Rényi divergence is a family of generalization of KL divergence which is stronger (\( q \to \nu \) is increasing). It has fundamental applications in statistics, physics, computer science (e.g., for differential privacy).

Rényi divergence of order \( q > 0 \) (\( q \neq 1 \)) of \( \rho \) with respect to \( \nu \) is:

\[ R_{\epsilon, \nu}(\rho) = \frac{1}{q - 1} \log \int_{\mathbb{R}^n} \rho(x)^q \nu(x)^{1-q} \, dx \]

We show Rényi divergence converges exponentially fast along ULA to the biased limit \( \nu \).

iteration complexity is determined by the bias \( R_\nu(\rho_0) \).