A Review of Diffusion-Based Sampling Methods: Langevin Algorithms, Discretization, and the Schrödinger Bridge Formulation

Yuexin Liao

September 10, 2025

Contents

1 Introduction

Sampling from a target probability measure is central to statistical inference, optimization, and computational physics. In this note, we analyze two complementary approaches:

(i) **Langevin-based methods:** Continuous-time Langevin diffusion

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

with $f : \mathbb{R}^n \to \mathbb{R}$ smooth, and its discretization via the Unadjusted Langevin Algorithm (ULA)

$$X_{k+1} = X_k - h\,\nabla f(X_k) + \sqrt{2h}\,\xi_k, \quad \xi_k \sim \mathcal{N}(0, I_n).$$

We rigorously analyze the convergence of the continuous process in both the Wasserstein metric and the χ^2-divergence, and then derive quantitative mixing time bounds for ULA.

(ii) **Schrödinger bridge formulation:** An optimal transport-based approach to construct, via change-of-measure arguments and optimal control, a finite-time diffusion process whose terminal distribution exactly equals the target measure.

Both methods build upon the theory of diffusion processes and stochastic differential equations (SDEs), yet they diverge in their goals: while ULA approximates the target distribution asymptotically (with discretization errors that vanish as $h \to 0$), the Schrödinger bridge yields exact sampling at a prescribed finite time.

2 Preliminaries and Notation

2.1 Target Measure and Smoothness Assumptions

Let μ be a probability measure on $\mathbb{R}^n$ with density

$$\mu(\,\mathrm{d}x) = \frac{e^{-f(x)}}{Z}\,\mathrm{d}x,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is assumed to be twice continuously differentiable. We impose:

- **L-smoothness:** There exists $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \le L\,\|x - y\|\,, \quad \forall x, y \in \mathbb{R}^n.$$

- **μ-strong convexity:** There exists $\mu > 0$ such that

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\,\|y - x\|^2\,, \quad \forall x, y \in \mathbb{R}^n.$$

Under these assumptions, μ satisfies a log-Sobolev inequality as well as Talagrand's transportation inequality:

$$\mu\, W_2^2(\nu, \mu) \le D_{\mathrm{KL}}(\nu\|\mu), \quad \forall \text{ probability measures } \nu.$$

2.2 Brownian Motion and the Fokker–Planck Equation

We define the standard Brownian motion in $\mathbb{R}^n$, $\{B_t\}_{t\ge 0}$, as a continuous process satisfying:

(a) $B_0 = 0$,

(b) Almost sure continuity,

(c) Independent increments: For $0 = t_0 < t_1 < \cdots < t_k$, the increments $B_{t_{i+1}} - B_{t_i}$ are independent,

(d) Gaussian increments: $B_t - B_s \sim \mathcal{N}(0, (t-s)I_n)$ for $0 \leq s < t$.

A drift-diffusion process in $\mathbb{R}^n$ governed by

$$\mathrm{d}X_t = a(X_t, t)\,\mathrm{d}t + b(X_t, t)\,\mathrm{d}B_t, \tag{1}$$

has a time-evolving density μ_t satisfying the Fokker–Planck equation:

$$\partial_t \mu_t(x) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\Big[a_i(x,t)\mu_t(x)\Big] + \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\Big[D_{ij}(x,t)\mu_t(x)\Big], \tag{2}$$

with diffusion matrix

$$D(x,t) = \frac{1}{2}\, b(x,t) b(x,t)^\top.$$

For the Langevin diffusion (see below), the Fokker–Planck equation becomes a deterministic PDE describing the evolution of probability density.

2.3 Metrics and Divergence Measures

Definition 2.1 (Wasserstein-2 Distance)**.** *For probability measures ν, μ on $\mathbb{R}^n$ with finite second moments,*

$$W_2(\nu, \mu) = \left(\inf_{\gamma \in \mathcal{C}(\nu,\mu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 \;\mathrm{d}\gamma(x,y) \right)^{1/2}.$$

Definition 2.2 (χ^2-divergence)**.** *For probability measures ν, μ with $\nu \ll \mu$, the χ^2-divergence is defined as*

$$\chi^2(\nu \| \mu) = \int_{\mathbb{R}^n} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) - 1 \right)^2 \mu(\,\mathrm{d}x) = \mathrm{Var}_\mu\!\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right).$$

Definition 2.3 (Poincaré Inequality)**.** *A measure μ satisfies a Poincaré inequality with constant C_P if for every smooth function g,*

$$\mathrm{Var}_\mu(g) \leq C_P\, \mathbb{E}_\mu \|\nabla g\|^2.$$

3 Convergence Analysis of Continuous Langevin Diffusion

3.1 Langevin Diffusion and Its Stationarity

The continuous-time Langevin diffusion is defined by

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t. \tag{3}$$

Its Fokker–Planck equation is

$$\partial_t \mu_t = \nabla \cdot \Big(\mu_t \nabla f\Big) + \Delta \mu_t. \tag{4}$$

Assuming that f is sufficiently smooth and $\mu \propto e^{-f}$, one verifies that μ is a stationary measure of (3) since plugging μ into (4) yields

$$\nabla \cdot (\mu \nabla f) + \Delta \mu = 0,$$

by direct differentiation and using the fact that $\mu(x) = ce^{-f(x)}$.

3.2 Exponential Convergence in Wasserstein Distance

Theorem 3.1 (Contraction in W_2)**.** *Let $\{X_t\}_{t\geq 0}$ evolve according to* (3) *with initial distribution μ_0, and assume that f is μ-strongly convex. Denote by μ_t the law of X_t. Then,*

$$W_2^2(\mu_t, \mu) \leq e^{-2\mu t} W_2^2(\mu_0, \mu).$$

Proof. Construct a synchronous coupling: let (X_0, Y_0) be an optimal coupling of (μ_0, μ) and evolve both processes using the same Brownian motion. Strong convexity implies

$$\frac{d}{dt} \|X_t - Y_t\|^2 \leq -2\mu \|X_t - Y_t\|^2,$$

so by Grönwall's inequality,

$$\|X_t - Y_t\|^2 \leq e^{-2\mu t} \|X_0 - Y_0\|^2.$$

Taking expectations yields the desired contraction. □

3.3 Convergence in χ^2-divergence under Poincaré Inequality

Theorem 3.2. *Let μ_t be the law of X_t in* (3) *with stationary measure $\mu \propto e^{-f}$, and assume μ is m-strongly log-concave and satisfies a Poincaré inequality with constant C_P. Then,*

$$\chi^2(\mu_t \| \mu) \leq \exp\Big(-\frac{2t}{C_P}\Big) \chi^2(\mu_0 \| \mu).$$

Proof. Differentiate $\chi^2(\mu_t \| \mu)$ with respect to t and use the Fokker–Planck equation (4). Integration by parts shows that

$$\frac{d}{dt} \chi^2(\mu_t \| \mu) \leq -\frac{2}{C_P} \chi^2(\mu_t \| \mu),$$

and applying Grönwall's inequality completes the proof. □

4 Discretization: The Unadjusted Langevin Algorithm (ULA)

4.1 The ULA Update and Its Interpretation

The Unadjusted Langevin Algorithm is the Euler discretization of (3). Starting from $X_0 \sim \mu_0$, the ULA update is

$$X_{k+1} = X_k - h \nabla f(X_k) + \sqrt{2h}\, \xi_k, \quad \xi_k \sim \mathcal{N}(0, I_n). \tag{5}$$

The interpretation is that for small h, ULA approximates the continuous diffusion, with the discretization error vanishing as $h \to 0$.

4.2 Local Discretization Error and Coupling Analysis

Lemma 4.1 (Local Error Estimate)**.** *Let $\{x_t\}_{t\in[0,h]}$ be the solution to*

$$\mathrm{d}x_t = -\nabla f(x_0)\, \mathrm{d}t + \sqrt{2}\, \mathrm{d}B_t, \quad x_0 = X_k,$$

and denote $X_{k+1}^c = x_h$ (the continuous evolution over time h with frozen drift). If $h \leq 1/(3L)$, then

$$\mathbb{E} \left\| X_{k+1}^c - X_k \right\|^2 \leq 6h^2\, \mathbb{E} \|\nabla f(X_k)\|^2 + 12h\, n.$$

Proof. A direct application of Itô's isometry and the Cauchy–Schwarz inequality yields the result; see also the detailed derivation in [Che23]. □

Lemma 4.2 (One-Step Contraction)**.** *Assume f is L-smooth and μ-strongly convex. Denote by μ_k the law of X_k generated by* (5)*. Then for $h \le \frac{\mu}{10L^2}$,*

$$W_2^2(\mu_{k+1},\mu) \le \Big(1-\frac{\mu h}{2}\Big)W_2^2(\mu_k,\mu) + \frac{32h^2L^2n}{\mu}.$$

Proof. The proof proceeds by coupling the discrete process and the continuous process over one step, controlling the contraction due to the strong convexity (via Theorem 3.1) and then bounding the additional error arising from discretization (using Lemma 4.1). An application of Grönwall's inequality finalizes the bound. □

4.3 Mixing Time of ULA

We now state a mixing time result for ULA.

Theorem 4.3 (Mixing Time for ULA)**.** *Assume that $\mu \propto e^{-f}$ satisfies $mI_n \preceq \nabla^2 f \preceq LI_n$, and define the condition number $\kappa = L/m$. Let the step size h satisfy*

$$h \lesssim \frac{\varepsilon^2}{L\kappa n},$$

and suppose that the initial distribution μ_0 is a point mass at the mode x^* of μ so that $W_2^2(\mu_0,\mu) \le \frac{n}{m}$. Then, after

$$K \gtrsim \frac{\kappa^2 n}{\varepsilon^2}\log\Big(\frac{\sqrt{m\,W_2^2(\mu_0,\mu)}}{\varepsilon}\Big)$$

iterations, one has

$$\sqrt{m\,W_2^2(\mu_K,\mu)} \le \varepsilon.$$

Proof Sketch. The proof involves two steps:

1. *One-step analysis:* Use Lemma 4.2 to relate $W_2^2(\mu_{k+1},\mu)$ to $W_2^2(\mu_k,\mu)$ plus a discretization error term.
2. *Iteration and Contraction:* Combine the one-step contraction with a recursive application and use the initial bound $W_2^2(\mu_0,\mu) \le \frac{n}{m}$ to deduce the required number of steps.

A careful balancing of the contraction term and the discretization error leads to the stated mixing time. □

5 Schrödinger Bridge Formulation for Exact Finite-Time Sampling

5.1 Problem Formulation and Change-of-Measure

Given a target measure μ on $\mathbb{R}^n$ with density f with respect to the canonical Gaussian measure

$$\gamma(\,\mathrm{d}x) = \frac{1}{(2\pi)^{n/2}}e^{-\|x\|^2/2}\,\mathrm{d}x,$$

define the set of path measures

$$\mathcal{M}^\mu = \Big\{ Q \text{ on } C([0,1],\mathbb{R}^n) \ : \ Q_0 = \delta_0,\ Q_1 = \mu \Big\}.$$

The Schrödinger bridge problem is to solve

$$Q^\mu = \arg\min_{Q \in \mathcal{M}^\mu} D_{\mathrm{KL}}(Q\|P),$$

where P is the Wiener measure on path space. By the chain rule for relative entropy, one can show that the unique minimizer satisfies

$$\frac{\mathrm{d}Q^\mu}{\mathrm{d}P}(W) = f(W_1),$$

which suggests that the optimal Q^μ is the law of Brownian motion conditioned on its terminal distribution being μ.

5.2 Gradient Ansatz and Girsanov Transformation

Assume we seek an Itô diffusion process $\{X_t\}_{t\in[0,1]}$ satisfying

$$\mathrm{d}X_t = -\nabla v(X_t,t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0, \tag{6}$$

for some function $v : \mathbb{R}^n \times [0,1] \to \mathbb{R}$. By Girsanov's theorem, the Radon–Nikodym derivative of the law Q of (6) with respect to P is

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(W) = \exp\left\{ -\int_0^1 \langle \nabla v(W_t,t),\, \mathrm{d}W_t\rangle - \frac{1}{2}\int_0^1 \|\nabla v(W_t,t)\|^2 \,\mathrm{d}t \right\}.$$

Applying Itô's formula to $v(W_t,t)$ and rearranging terms, one obtains

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(W) = \exp\Big\{ v(0,0) - v(W_1,1) + \int_0^1 \Big[\partial_t v(W_t,t) + \tfrac{1}{2}\Delta v(W_t,t) - \tfrac{1}{2}\|\nabla v(W_t,t)\|^2 \Big]\,\mathrm{d}t \Big\}.$$

To force the integrand to vanish, we require that

$$\partial_t v(x,t) + \frac{1}{2}\Delta v(x,t) = \frac{1}{2}\|\nabla v(x,t)\|^2, \quad (x,t) \in \mathbb{R}^n \times [0,1], \tag{7}$$

with terminal condition $v(x,1) = -\log\psi(x)$ for a strictly positive function ψ. Defining the Cole–Hopf transform

$$h(x,t) = e^{-v(x,t)},$$

one shows that h satisfies the linear heat equation

$$\partial_t h(x,t) + \frac{1}{2}\Delta h(x,t) = 0, \quad h(x,1) = \psi(x),$$

and by the Feynman–Kac formula,

$$h(x,t) = \mathbb{E}_P\Big[\psi(W_1) \mid W_t = x\Big].$$

Choosing $\psi(x) = f(x)$ and noting that $v(0,0) = -\log h(0,0)$, we deduce

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(W) = \frac{f(W_1)}{\mathbb{E}_\gamma[f(W_1)]},$$

which is consistent with the desired change of measure.

5.3 Optimal Control Perspective

An equivalent formulation is to consider the controlled SDE

$$\mathrm{d}X_t^u = u_t \,\mathrm{d}t + \mathrm{d}W_t, \quad X_0^u = 0, \tag{8}$$

with cost functional

$$J(u) = \mathbb{E}^u \left[\frac{1}{2} \int_0^1 \|u_t\|^2 \, \mathrm{d}t - \log f(X_1^u) \right].$$

The corresponding Hamilton–Jacobi–Bellman equation for the value function

$$v(x,t) = \min_u \mathbb{E}^u \left[\frac{1}{2} \int_t^1 \|u_s\|^2 \, \mathrm{d}s - \log f(X_1^u) \,\middle|\, X_t^u = x \right]$$

is equivalent to (7), and the optimal control is given by $u^*(x,t) = -\nabla v(x,t)$ (the Föllmer drift). This confirms that the optimal process exactly transports the initial measure to μ in finite time.

6 Comparative Insights and Discussion

The two paradigms we have developed differ fundamentally in their methodology and error analysis:

- **Langevin/ULA Approach:** The continuous-time Langevin diffusion converges exponentially fast to μ, as measured by both W_2 and χ^2 distances. ULA approximates these dynamics via an Euler discretization. Its convergence is established by careful control of the one-step discretization error (via coupling arguments and local error bounds) and the contraction properties inherent in strong convexity. The mixing time of ULA is explicitly determined as a function of the step size h, the dimension n, and the condition number κ.
- **Schrödinger Bridge Approach:** In contrast, the Schrödinger bridge formulation constructs an Itô diffusion process with a time-dependent drift that is exactly optimal in the sense of relative entropy. By solving a nonlinear PDE (which is linearized via the Cole–Hopf transformation), one obtains an explicit formula for the drift (the Föllmer drift) that guarantees the terminal distribution is exactly μ. This method does not suffer from discretization error, as it achieves exact finite-time sampling.

These methods exemplify two philosophies: the ULA method relies on iterative, asymptotic convergence with quantifiable discretization error, while the Schrödinger bridge method achieves exact matching by solving an optimal transport problem in path space.