

Langevin Sampling - Setting

Suppose we have access to (some version of) the unnormalized density, specifically, we are given query access to: $\nabla \ln \pi^*(x)$.

Like the M-H algorithm, we can construct a Markov Process whose invariant dist. is $\pi^*(x)$.

The Markov Process evolves in continuous state space and cont. time:

Langevin
Equation

$$dx_t = \nabla \ln \pi^*(x_t) dt + \sqrt{2} dB_t,$$

where B_t is Brownian motion.

This is an SDE (Stochastic Diff. Eqn), which is a cont. time, cont. state space Markov Process, and as we shall see, this has the invariant dist of $\pi^*(x)$.

.....

Variants of this equation have been studied by different communities. Two directions that

we highlight:

② **Optimization**: Goal is to $\min_{x \in \mathbb{R}^d} U(x)$. As previously discussed, one approach is to instead sample from:

$$dX_t = -\nabla U(X_t) dt + \sigma(t) dB_t.$$

for suitably chosen $\sigma(t)$. If $\sigma(t) = \sqrt{2T}$, where T is the fixed "temperature" (constant), then we can show that:

$$X_{\infty} \sim \frac{e^{-U(x)/T}}{Z}, \quad \text{where } Z \text{ is the normalizing constant.}$$

Observe that as $T \rightarrow 0$, the limiting samples will concentrate on $\{y : y \in \operatorname{argmin}_x U(x)\}$.

A **temperature-annealed** variant of the basic Langevin equation is:

$$dX_t = -\nabla U(X_t) dt + \sqrt{\frac{2c}{\ln(2+t)}} dB_t,$$

time-varying diffusion coefficient

then (under additional technical conditions, e.g., $U(\cdot)$ is twice differentiable, etc.; see ref. below):

$$\mu_{x_t} \xrightarrow{\text{dist.}} \pi_0, \quad \text{where}$$

density of x_t

$$\pi_0(\{y : U(y) = \inf_x U(x)\}) = 1$$

Ref: S. Geman and C. Hwang, Diffusions for global optimization, SIAM J. of Opt & Control, Sep. 1986.

(b) **Generative Modeling**: Goal is to generate samples from $\pi^*(x)$. However, we do not have access to $\nabla \ln \pi^*(x)$. Instead, we can learn (future notes) good approximations to

$$\nabla \ln \pi_t^*(x), \quad \text{where} \quad \pi_t^*(x) = \pi^*(x) * N(0, \sigma_t^2)$$

where σ_t^2 is an appropriate variance term, with $\sigma_t^2 \rightarrow 0$ as $t \rightarrow \infty$.

The associated **Gaussian-annealed** Langevin dynamics

is given by: $X_0 \sim N(0, I)$, and for $t \in [0, T]$,

$$dX_t = \nabla \ln \pi_{T-t}^*(t) dt + \sqrt{2} dB_t$$

(annealed Langevin dynamics)

Another dynamic is the DDPM dynamic, where $X_0 \sim N(0, I)$, and the following is run for finite interval of time $t \in [0, T]$:

$$dX_t = (X_t + 2 \nabla \ln \pi_{T-t}^*(t)) dt + \sqrt{2} dB_t$$

(DDPM dynamics)

Motivated by the above applications, we will study the "basic" Langevin dynamic and develop the theoretical tools to show the following two properties:

① The invariant distribution of the Langevin dynamic: $dX_t = \nabla \ln \pi^*(X_t) dt + \sqrt{2} dB_t$

is $X_\infty \sim \pi^*(x)$, $x \in \mathbb{R}^d$

(b) The rate of convergence of the density of X_t (μ_{X_t}) to the stationary dist. $\pi^*(x)$ is exponentially fast in time under the condition that:

$\pi^*(x)$ is a (strongly) log-concave distribution.

i.e., $\pi^*(x) \propto e^{-V(x)}$, $V(x)$ is β -strongly convex.

To show (a), we will develop some theory for diffusion processes and the Kolmogorov forward equation (aka Fokker-Planck equation) to study density evolution.

To show (b), we will develop some theory on Poincaré inequalities, the Bakry-Emery criterion, and log-Sobolev inequalities for Markov processes.