1 Coding (To Metropolize or not?).

Sample from a $\mathcal{N}(0,1)$ distribution using a 1-d Langevin sampler

$$X_{t+1} = X_t + \delta \frac{d}{dx}(\log \pi(X_t)) + \sqrt{2\delta}\xi_t, \qquad \xi_t \in \mathcal{N}(0,1).$$

(a) When you change the step size δ, how does the histogram of samples change?

(b) Now add an acceptance-rejection step. How does the acceptance probability change depending on δ?

2 Coding (Sampling from a double-well potential).

Apply Langevin dynamics or Hamiltonian Monte Carlo to sample from a double-well potential

$$\pi(x) \propto \exp\bigl(-(x^2-4)^2\bigr).$$

What can you do to optimize? How often does the sampler transition from the right well (centered at $x=+2$) into the left well (centered at $x=-2$)?

3 Coding (Sampling from a doughnut).

Apply Hamiltonian Monte Carlo (HMC) to sample from a doughnut density

$$\pi(x,y) \propto \exp(-(x^2+y^2-25)^2).$$

What can you do to optimize? Is HMC better than Langevin dynamics for this problem?

4 Choose-your-own-adventure (Underdamped Langevin).

The underdamped Langevin equation is the stochastic differential equation

$$d\boldsymbol{x} = \boldsymbol{v}\,dt, \qquad d\boldsymbol{v} = \nabla \log \pi(\boldsymbol{x})\,dt - \gamma \boldsymbol{v}\,dt + \sqrt{2\gamma}\,dW,$$

where $\boldsymbol{v}$ is the velocity and $\gamma > 0$ is the fraction parameter.

(a) Write down the underdamped Langevin sampling algorithm based on an Euler discretization for the SDE above.

(b) If we want to correct the bias, how do we define acceptance probabilities?

(c) Apply underdamped Langevin dynamics to a two-dimensional Gaussian $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_{\max}, \lambda_{\min})$ (using math or experiments) and show it can outperform MALA.

(d) Show that the SDE has a unique stationary measure $(2\pi)^{d/2}\exp(-\|\boldsymbol{v}\|^2/2)\pi(\boldsymbol{x})$.

5 Math (No acceleration with Langevin dynamics).

Suppose the target distribution is a two-dimensional Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = \text{diag}(\lambda_{\max}, \lambda_{\min})$ and we sample using the Langevin sampler. We draw a random velocity according to $\boldsymbol{v} \in \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and apply the update

$$\boldsymbol{x} \leftarrow \boldsymbol{x} + \delta \nabla \log \pi(\boldsymbol{x}) + \sqrt{2\delta}\boldsymbol{v}.$$

(a) How is the position $\boldsymbol{x} = (x_1, x_2)$ updated at each iteration?

(b) If we want to correct the bias, how do we define acceptance probabilities?

(c) If we draw samples $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \ldots$, with $\boldsymbol{x}^{(0)}$ initialized according to the stationary measure, what is the smallest number of update steps s so that

$$|\rho_1(s)| = |\text{corr}[\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_1^{(s)}]| \leq 1/2, \qquad |\rho_2(s)| = |\text{corr}[\boldsymbol{x}_2^{(0)}, \boldsymbol{x}_2^{(s)}]| \leq 1/2.$$

6 Math (acceleration with Hamiltonian Monte Carlo).

Suppose the target distribution is a two-dimensional Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = \text{diag}(\lambda_{\max}, \lambda_{\min})$ and we sample using Hamiltonian Monte Carlo (HMC). We draw a random velocity according to $\boldsymbol{v} \in \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and apply L rounds of leapfrog updates:

$$\boldsymbol{v} \leftarrow \boldsymbol{v} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}), \qquad \boldsymbol{x} \leftarrow \boldsymbol{x} + \delta \boldsymbol{v}, \qquad \boldsymbol{v} \leftarrow \boldsymbol{v} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}).$$

(a) Show that the position $\boldsymbol{x} = (x_1, x_2)$ and velocity $\boldsymbol{v} = (v_1, v_2)$ are updated according to

$$\begin{pmatrix} x_i \\ v_i \end{pmatrix} \leftarrow \begin{pmatrix} 1 - \alpha_i & \delta \\ -\frac{1}{\delta}(2\alpha_i - \alpha_i^2) & 1 - \alpha_i \end{pmatrix}^L \begin{pmatrix} x_i \\ v_i \end{pmatrix},$$

where $\alpha_i = \delta^2/(2\lambda_i)$ for $i = 1, 2$.

(b) How large can we take δ to ensure a stable algorithm?

(c) If we want to correct the bias, how do we define acceptance probabilities?

(d) Show that the samples $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \ldots$ satisfy

$$\mathbb{E}[\boldsymbol{x}^{(t)} | \boldsymbol{x}^{(t-1)}] = \begin{pmatrix} \cos\big(L \arccos(1 - \alpha_1)\big) & \\ & \cos\big(L \arccos(1 - \alpha_2)\big) \end{pmatrix} \boldsymbol{x}^{(t-1)}.$$

(e) Set $\delta = \sqrt{2\lambda_{\min}}$ and set L to be the smallest odd number such that $\cos(\pi/(2L)) \geq 1 - 1/\kappa$, where $\kappa = \lambda_{\max}/\lambda_{\min}$. Then, show the number of leapfrog steps is bounded by

$$L \leq \pi\sqrt{\kappa/8} + 2$$

and the decay of correlations (under the stationary measure) is bounded by

$$0 \leq \rho_1(s) = \text{corr}[\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_1^{(s)}] \leq (\sqrt{3}/2)^s, \qquad \rho_2(s) = \text{corr}[\boldsymbol{x}_2^{(0)}, \boldsymbol{x}_2^{(s)}] = \delta(s).$$

This is an example of *acceleration*, where the amount of work to sample from a poorly scaled distribution only depends on $\sqrt{\kappa}$, not κ.