Efficient Sampling on Riemannian Manifolds via Langevin MCMC

Xiang Cheng, Jingzhao Zhang, Suvrit Sra

Sampling over Riemannian Manifolds

Given manifold (M, g), sample from $\boldsymbol{dp(x) = e^{-U(x)} dvol_g(x)}$

- $U(x): M \to \mathbb{R}$ is a potential function (e.g. negative-log-posterior)
- $dvol_g(x)$ is manifold volume, in coordinates, it is $\sqrt{\det(g)}$.

The Riemannian Langevin Diffusion (RLD):

$$dx(t) = -\mathrm{grad}\, U\big(x(t)\big)dt + dB_t^g$$

- $\mathrm{grad}\, U$ denotes the manifold gradient, and dB_t^g denotes the manifold Brownian motion
- Has invariant distribution $\boldsymbol{e^{-U(x)} dvol_g(x)}$

Sampling over Riemannian Manifolds

Given manifold (M, g), sample from $\boldsymbol{dp(x) = e^{-U(x)} dvol_g(x)}$

- $U(x): M \to \mathbb{R}$ is a potential function (e.g. negative-log-posterior)
- $dvol_g(x)$ is manifold volume, in coordinates, it is $\sqrt{\det(g)}$.

The Riemannian Langevin Diffusion (RLD):

$$dx(t) = -\mathrm{grad}\, U\big(x(t)\big)dt + dB_t^g$$

- $\mathrm{grad}\, U$ denotes the manifold gradient, and dB_t^g denotes the manifold Brownian motion
- Has invariant distribution $\boldsymbol{e^{-U(x)} dvol_g(x)}$

Riemannian Langevin MCMC

Based on the geometric Euler Murayama Discretization of RLD:

$$x_{(k+1)\delta} = Exp_{x_{k\delta}}\left(-\delta \, \mathrm{grad}\, U(x_{k\delta}) + \sqrt{\delta}\xi_k\right)$$

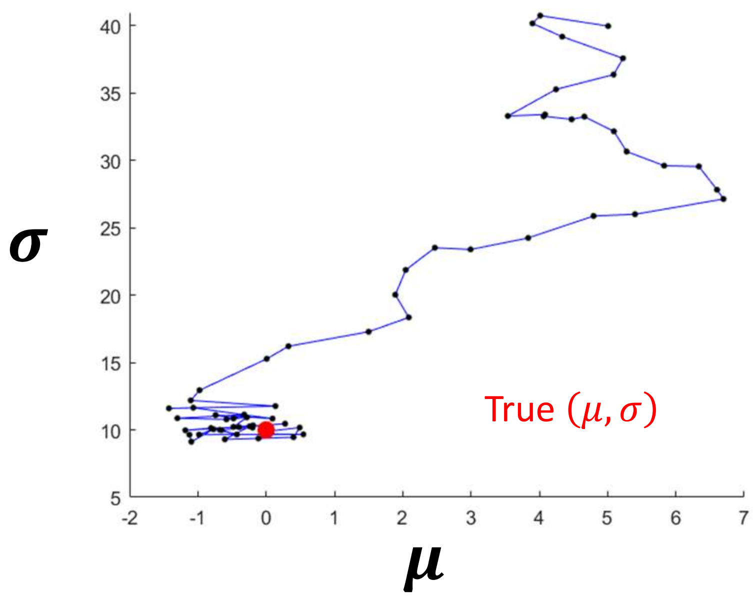
where ξ_k is "standard Gaussian" wrt an orthonormal basis at $T_{x_k}M$

Exponential maps can be approximated to high accuracy efficiently

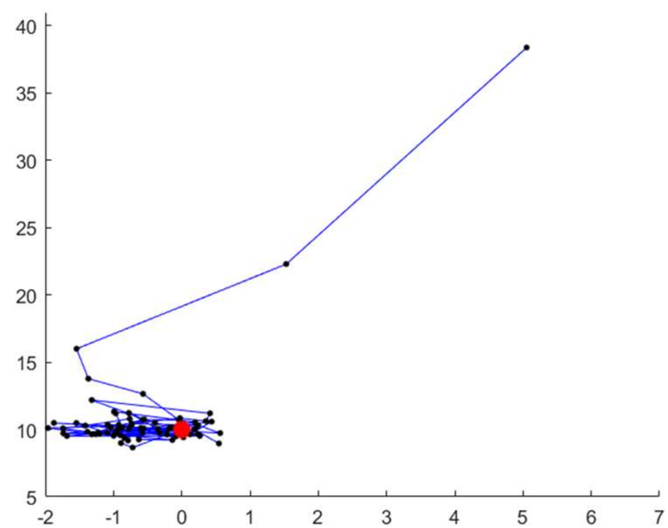
RLMCMC can be much faster than Euclidean Langevin MCMC

- Given: unobserved $(\mu = 0, \sigma = 10)$, observe samples $x_1 \dots x_{100} \sim \mathcal{N}(\mu, \sigma^2)$.
- Task: sample from the posterior distribution $p(\mu, \sigma | x_1 \dots x_{100}) \propto \exp\left(\sum_i \frac{\|x_i - \mu\|^2}{2\sigma^2} - N \log \sigma\right)$
- Fisher-Rao manifold: $\left(M = \mathbb{R} \times \mathbb{R}^+,\ g(\mu, \sigma) = \begin{bmatrix} N/\sigma^2 & 0 \\ 0 & 2N/\sigma^2 \end{bmatrix}\right)$

Euclidean Langevin MCMC

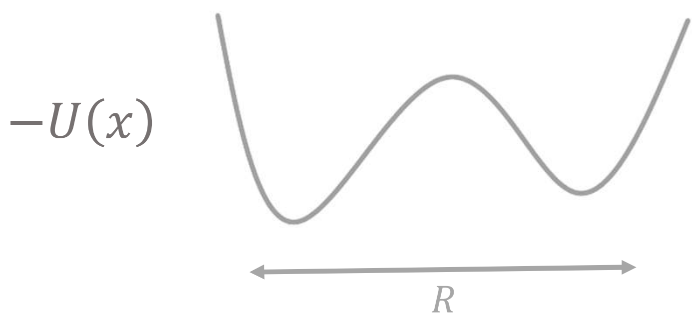


RLMCMC



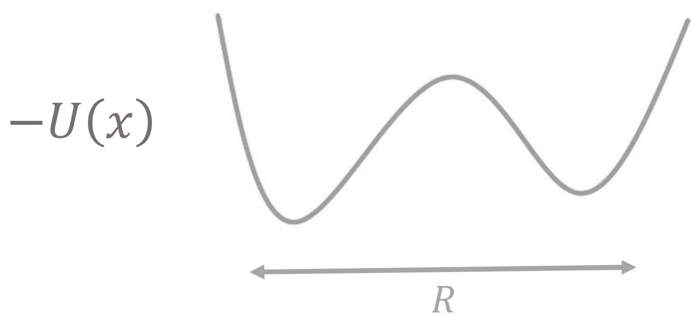
Key Assumptions

- Assume (M, g) satisfies
  - Ricci curvature lower bounded by $-L_{ric}$
  - Absolute value of sectional curvature upper bounded by L_{sec}
- Assume $-U$ satisfies
  - (gradient Lipschitz) $\mathbf{Hess}(\boldsymbol{U})[\boldsymbol{v}, \boldsymbol{v}] \leq \boldsymbol{L_U} \|\boldsymbol{v}\|^2,$ for all $x \in M, v \in T_x M$
  - (distant dissipativity) $\left\langle \Gamma_x^y \operatorname{grad} U(y) - \operatorname{grad} U(x), \operatorname{Exp}_x^{-1}(y) \right\rangle \geq m \operatorname{dist}(x, y)^2,$ for all $\operatorname{dist}(x, y) > R$ and some $m > -L_{Ric}$



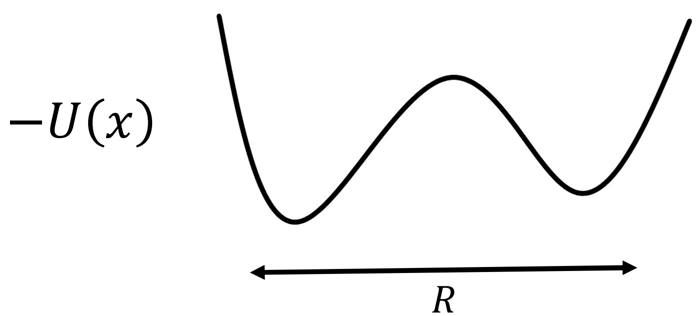
Key Assumptions

- Assume (M, g) satisfies
  - Ricci curvature lower bounded by $-L_{ric}$
  - Absolute value of sectional curvature upper bounded by L_{sec}
- Assume $-U$ satisfies
  - (gradient Lipschitz) $\mathbf{Hess}(\boldsymbol{U})[\boldsymbol{v}, \boldsymbol{v}] \leq \boldsymbol{L_U} \|\boldsymbol{v}\|^{\mathbf{2}},$ for all $x \in M, v \in T_x M$
  - (distant dissipativity) $\left\langle \Gamma_x^y \mathrm{grad}\, U(y) - \mathrm{grad}\, U(x), \mathrm{Exp}_x^{-1}(y) \right\rangle \geq m\, \mathrm{dist}(x, y)^2$, for all $\mathrm{dist}(x, y) > R$ and some $m > -L_{Ric}$



Key Assumptions

- Assume (M, g) satisfies
  - Ricci curvature lower bounded by $-L_{ric}$
  - Absolute value of sectional curvature upper bounded by L_{sec}
- Assume $-U$ satisfies
  - (gradient Lipschitz) $\mathbf{Hess}(\boldsymbol{U})[\boldsymbol{v}, \boldsymbol{v}] \leq \boldsymbol{L_U} \|\boldsymbol{v}\|^2$, for all $x \in M, v \in T_x M$
  - (distant dissipativity) $\langle \Gamma_x^y \operatorname{grad} U(y) - \operatorname{grad} U(x), \operatorname{Exp}_x^{-1}(y) \rangle \geq m \operatorname{dist}(x, y)^2$, for all $\operatorname{dist}(x, y) > R$ and some $m > -L_{Ric}$



Main Theoretical Result

- Assume (M, g) satisfies
  - Ricci curvature lower bounded by $-L_{ric}$, for some $L_{Ric} > 0$
  - Absolute value of sectional curvature upper bounded by L_{sec}
- Assume $-U$ satisfies
  - (gradient Lipschitz) $\mathbf{Hess}(\boldsymbol{U})[\boldsymbol{v}, \boldsymbol{v}] \leq \boldsymbol{L_U} \|\boldsymbol{v}\|^{\mathbf{2}}$, for all $x \in M, v \in T_x M$
  - (distant dissipativity) $\langle \Gamma_x^y \text{grad } U(y) - \text{grad } U(x), \text{Exp}_x^{-1}(y) \rangle \geq m \text{ dist}(x, y)^2$, for all $\text{dist}(x, y) > R$ and some $m > L_{Ric}$

Theorem 1

Let $x_{k\delta}$ be iterates of RLMCMC, and let $y(t)$ denote RLD, then

$$\mathbb{E}\left[\text{dist}\left(x_{K\delta}, y(K\delta)\right)\right] \leq \epsilon$$

for $K = \text{poly}\left(e^{(L_U + L_{ric})R^2}, L_{sec}, L_U, d, \frac{1}{m - L_{Ric}}\right) \cdot 1/\epsilon^2$