CPSC 661: Sampling Algorithms in ML

Andre Wibisono

April 28, 2021

Yale University

Last time

- Langevin dynamics
- Unadjusted Langevin Algorithm
- Variants of Langevin dynamics
  1. {Weighted, Mirror, Newton, Interacting} Langevin Dynamics
  2. Underdamped Langevin Dynamics

Today: Hamiltonian Monte Carlo

Hamiltonian Flow

Given $H: \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ differentiable (we call H the Hamiltonian)

$$(x, v) \longmapsto H(x, v)$$

The Hamiltonian flow generated by H is

$$(HF) \quad \begin{cases} \dot{x}_t = \dfrac{\partial H}{\partial v}(x_t, v_t) \\ \dot{v}_t = -\dfrac{\partial H}{\partial x}(x_t, v_t) \end{cases}$$

Fact: Hamiltonian flow preserves Hamiltonian H

$$H(x_t, v_t) = H(x_0, v_0)$$

Proof: $\dfrac{d}{dt} H(x_t, v_t) = \left\langle \dfrac{\partial H}{\partial x}(x_t, v_t), \dot{x}_t \right\rangle + \left\langle \dfrac{\partial H}{\partial v}(x_t, v_t), \dot{v}_t \right\rangle$

$$= \left\langle \frac{\partial H}{\partial x}, \frac{\partial H}{\partial v} \right\rangle + \left\langle \frac{\partial H}{\partial v}, -\frac{\partial H}{\partial x} \right\rangle$$

$$= 0.$$

- Eg. Newton's Law: $\ddot{X}_t = -\nabla f(X_t)$

$$\Leftrightarrow \begin{cases} \dot{X}_t = V_t = \frac{\partial}{\partial v}\left(\frac{1}{2}\|V_t\|^2\right) = \frac{\partial H}{\partial v}(X_t, V_t) \\ \dot{V}_t = -\nabla f(X_t) = -\frac{\partial}{\partial x}(f(X_t)) = -\frac{\partial H}{\partial x}(X_t, V_t) \end{cases}$$

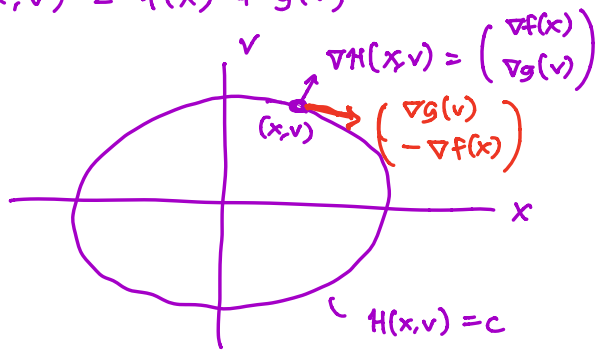
this is the Hamiltonian flow of

$$H(x,v) = \underbrace{f(x)}_{\text{potential energy}} + \underbrace{\tfrac{1}{2}\|v\|^2}_{\text{kinetic energy}}$$

- Hamiltonian flow of $H(x,v) = f(x) + g(v)$

is $\begin{cases} \dot{x} = \nabla g(v) \\ \dot{v} = -\nabla f(x) \end{cases}$

$$\Leftrightarrow \begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}}_{\text{Rotation by } 90^\circ} \begin{pmatrix} \nabla f(x) \\ \nabla g(v) \end{pmatrix}$$



- eg. $f(x) = \frac{\alpha}{2}x^2$ on $\mathbb{R}$

Hamiltonian flow $\begin{cases} \dot{x} = v \\ \dot{v} = -\alpha x \end{cases} \Leftrightarrow \ddot{x} = -\alpha x$

Solution is $\boxed{X_t = \cos(\sqrt{\alpha}t)\, x_0 + \frac{1}{\sqrt{\alpha}} \sin(\sqrt{\alpha}t)\, \dot{x}_0}$

$$\dot{X}_t = -\sqrt{\alpha}\sin(\sqrt{\alpha}t)\, x_0 + \cos(\sqrt{\alpha}t)\, \dot{x}_0$$

$$\ddot{X}_t = -\alpha\cos(\sqrt{\alpha}t)\, x_0 - \sqrt{\alpha}\sin(\sqrt{\alpha}t)\, \dot{x}_0$$

$$= -\alpha X_t$$

Hamiltonian Monte Carlo

To sample from $\nu(x) \propto e^{-f(x)}$ on $\mathbb{R}^n$

HMC: (Hybrid Monte Carlo [Duane et al. 1987]
Hamiltonian Monte Carlo)

* From $x_k \sim g_k$:
  - Run Hamiltonian flow for some integration time η:

$$\dot{X}_t = V_t$$
$$\dot{V}_t = -\nabla f(X_t)$$

this is Hamiltonian flow of $H(x, v) = f(x) + \frac{1}{2}\|v\|^2$

from $(X_0, V_0) = (x_k, z_k)$, where $z_k \sim \mathcal{N}(0, I)$ is independent

  - set $x_{k+1} = X_\eta$

Example: Gaussian

$\nu = \mathcal{N}(0, \frac{1}{\alpha})$ on $\mathbb{R}^1$, $\quad f(x) = \frac{\alpha}{2}x^2 + \text{constant}$

Hamiltonian flow $\quad \dot{X}_t = V_t$

$\dot{V}_t = -\alpha X_t$

$\Rightarrow X_t = \cos(\sqrt{\alpha}t)\, X_0 + \frac{1}{\sqrt{\alpha}} \sin(\sqrt{\alpha}t)\, V_0$

HMC: from $X_n \sim \delta_n$:

$e^{i\theta} = \cos\theta + i \sin\theta$

* draw $Z_n \sim \mathcal{N}(0, I)$
* set $\boxed{X_{n+1} = \cos(\sqrt{\alpha}\eta)\, X_n + \frac{1}{\sqrt{\alpha}} \sin(\sqrt{\alpha}\eta)\, Z_n}$

Consider for integration time η:

$\cos x \approx 1 - x^2$

$\sin x \approx x$

1) $0 < \eta \ll 1$: $\quad \cos(\sqrt{\alpha}\eta) \approx 1 - \alpha\eta^2$

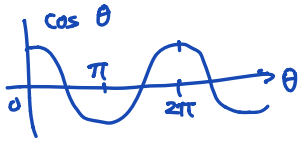
$\sin(\sqrt{\alpha}\eta) \approx \sqrt{\alpha}\eta$

then $\quad X_{n+1} \approx (1 - \alpha\eta^2) X_n + \eta Z_n$

c.f. OU (Langevin dynamics for $\nu = \mathcal{N}(0, \frac{1}{\alpha})$) is:

$$x_{n+1} = e^{-\alpha\eta} x_n + \sqrt{\frac{1 - e^{-2\alpha\eta}}{\alpha}}\, z_n, \quad z_n \sim \mathcal{N}(0, I)$$

2) Large η?



* if $\sqrt{\alpha}\eta = 2m\pi$ for some $m \in \mathbb{N}$

$\cos(\sqrt{\alpha}\eta) = 1$

$\sin(\sqrt{\alpha}\eta) = 0$

then $\boxed{x_{n+1} = x_n}$ in HMC

→ this is bad, no mixing to $\nu = \mathcal{N}(0, \frac{1}{\alpha})$

* if $\sqrt{\alpha}\eta = (2m-1)\pi$

$\cos(\sqrt{\alpha}\eta) = -1$

$\sin(\sqrt{\alpha}\eta) = 0$

then $\boxed{x_{n+1} = -x_n}$

→ this is also bad, no mixing to $\nu = \mathcal{N}(0, \frac{1}{\alpha})$

* if $\boxed{\sqrt{\alpha}\eta = \frac{\pi}{2} + m\pi}$

$\cos(\sqrt{\alpha}\eta) = 0$

$\sin(\sqrt{\alpha}\eta) = \pm 1$

then HMC becomes $x_{n+1} = 0 \cdot x_n + \frac{1}{\sqrt{\alpha}} z_n = \frac{1}{\sqrt{\alpha}} z_n \sim \mathcal{N}(0, \frac{1}{\alpha}) = \nu$

∴ optimal integration time is

$$\boxed{\eta = \frac{\pi}{2\sqrt{\alpha}}}$$

then HMC converges in 1 step to $\nu = \mathcal{N}(0, \frac{1}{\alpha})$

Why does HMC work?

Lemma: In one step of HMC, for any integration time $\eta > 0$, $x_k \sim \rho_k$ satisfies

$$\mathrm{KL}(\rho_{k+1} \,\|\, \nu) \le \mathrm{KL}(\rho_k \,\|\, \nu)$$

$$\mathrm{KL}(\rho \,\|\, \nu) = H_\nu(\rho) = \int \rho(x) \log \frac{\rho(x)}{\nu(x)} \, dx$$

Proof: From $x_k \sim \rho_k$, let $v_k \sim \mathcal{N}(0, I)$ independent $\Big\}$ set $(X_0, V_0) = (x_k, v_k) \sim \tilde{\rho}_0(x, v)$

In the joint space (x, v), the target distribution is

$$\tilde{\nu}(x, v) = \nu(x) \cdot \mathcal{N}(v; 0, I)$$

$$\propto e^{-f(x)} \cdot e^{-\frac{1}{2}\|v\|^2} = e^{-H(x,v)}$$

$$H(x, v) = f(x) + \tfrac{1}{2}\|v\|^2$$

In the initialization, $KL(\tilde{\rho}_0 \| \tilde{\nu}) = KL(\rho_n \| \nu)$ (*)

then run Hamiltonian flow

$$\left.\begin{aligned} \dot{X}_t &= V_t \\ \dot{V}_t &= -\nabla f(X_t) \end{aligned}\right\} \quad (X_t, V_t) \sim \tilde{\rho}_t(x,v) \text{ satisfies continuity equation}$$

Can show: Hamiltonian flow preserves KL divergence to $\tilde{\nu}$

(*) $KL(\tilde{\rho}_t \| \tilde{\nu}) = KL(\tilde{\rho}_0 \| \tilde{\nu}) \overset{(*)}{=} KL(\rho_n \| \nu)$

then at time $t=\eta$, want to set $x_{out} = X_\eta$

$(X_\eta, V_\eta) \sim \tilde{\rho}_\eta(x,v)$

Let X_η have marginal $\tilde{\rho}_{\eta,x}(x) = \int_{\mathbb{R}^n} \tilde{\rho}_\eta(x,v)\, dv$

V_η have marginal $\tilde{\rho}_{\eta,v}(v) = \int_{\mathbb{R}^n} \tilde{\rho}_\eta(x,v)\, dx$

then can show: since $\tilde{\nu}(x,v) = \nu(x) \cdot \mathcal{N}(v; 0, I)$

(*) $\Rightarrow KL(\tilde{\rho}_\eta \| \tilde{\nu}) = KL(\tilde{\rho}_{\eta,x} \| \nu) + KL(\tilde{\rho}_{\eta,v} \| \mathcal{N}(0,I)) + I(\tilde{\rho}_\eta)$

↖ mutual information

$$I(\tilde{\rho}_\eta) = \int \tilde{\rho}_\eta(x,v) \log \frac{\tilde{\rho}_\eta(x,v)\, dx\, dv}{\tilde{\rho}_{\eta,x}(x) \cdot \tilde{\rho}_{\eta,v}(v)}$$

Since each term is ≥ 0, when we set

$x_{out} = X_\eta$

$\rho_{out} = \tilde{\rho}_{\eta,x}$

$\Rightarrow KL(\rho_{out} \| \nu) \overset{(*)}{\leq} KL(\tilde{\rho}_\eta \| \tilde{\nu}) = KL(\tilde{\rho}_0 \| \tilde{\nu}) = KL(\rho_n \| \nu)$.

□

Choosing integration time

How to choose integration time η?

- Too small $\rightarrow$ not much progress
- Too large $\rightarrow$ come back to beginning

Choosing integration time

How to choose integration time η?

- Too small $\rightarrow$ not much progress
- Too large $\rightarrow$ come back to beginning

Will see:

1. Random integration time [Bou-Rabee & Sanz-Serna 2017]
2. Ideal HMC with short integration time [Chen & Vempala 2019]
3. Metropolized HMC with leapfrog integrator [Chen et al. 2020]

Other ways, e.g. No-U-Turn Sampler [Hoffman & Gelman 2014]

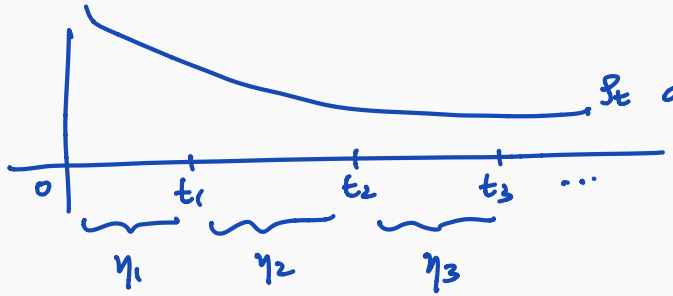
1. Randomized Hamiltonian Monte Carlo

HMC with integration time at step k:

[RHMC] $\eta_k \sim \mathrm{Exp}(\gamma)$ indpt. exp. R.V. with mean $\frac{1}{\gamma}$

$$\mathbb{P}(\eta_k \le t) = 1 - e^{-\frac{t}{\gamma}}$$

Let $t_k = \eta_1 + \cdots + \eta_k$



[Bou-Rabee & Sanz-Serna 2018]

Thm: [Lu & Wang 2021]:

Assume $\nu \propto e^{-f}$ satisfies α-Poincaré ineq.

Assume $\|\nabla^2 f(x)\| \le M(1 + \|\nabla f(x)\|)$, f is superlinear.

then RHMC with any $\gamma > 0$ satisfies

$$\chi^2_\nu(\rho_t) \le C_0 \cdot e^{-2\lambda t}\, \chi^2_\nu(\rho_0)$$

where $\lambda = \dfrac{\alpha\gamma}{(\sqrt{\alpha}+\gamma+R)^2}$

where:

1) if f is convex, then $R = 0$

then with $\gamma = \sqrt{\alpha} \Rightarrow \boxed{\lambda = \dfrac{\sqrt{\alpha}}{4}}$ accelerated rate (similar to underdamped, c.f. Langevin dynamics)

2) if $\nabla^2 f(x) \gtrsim -K \cdot I$, then $R = \sqrt{K}$

then with $\gamma = \sqrt{\alpha} + \sqrt{K}$: $\lambda = \dfrac{\alpha}{4(\sqrt{\alpha}+\sqrt{K})}$

2. Ideal HMC with short integration time

Consider HMC with constant $\eta = \frac{1}{2\sqrt{L}}$

Assume f is α- strongly convex, L- smooth $\left(\alpha I \le \nabla^2 f(x) \le L I\right)$

$\kappa = \frac{L}{\alpha}$ is condition number

Lemma: Hamiltonian flow is a contraction

$x_0 \xrightarrow{v}$ ---- $\ddot{x}_t = -\nabla f(x_t)$

$\dot{x}_0 = \dot{y}_0 = v$

$\xrightarrow{v}$ ---- y_0 $\quad \ddot{y}_t = -\nabla f(y_t)$

for $0 \le t \le \frac{1}{2\sqrt{L}}$

$$\|x_t - y_t\|^2 \le \left(1 - \frac{\alpha t^2}{4}\right)\|x_0 - y_0\|^2$$

Theorem: [Chen & Vempala 2020]

Along HMC, $W_2^2(g_n, \nu) \le \left(1 - \frac{1}{16\kappa}\right)^k W_2^2(g_0, \nu)$

$\Rightarrow$ mixing time is $\tilde{O}(\kappa)$

* Matching lower bound:

Consider $\nu = \mathcal{N}\left(0, \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{L} \end{pmatrix}\right)$, so $\nabla^2 f(x) = \begin{pmatrix} \alpha & \\ & L \end{pmatrix}$

Theorem: HMC with $\eta = \frac{1}{c\sqrt{L}}$

has spectral gap $\gamma \leq \frac{1}{2c^2 K}$

$\Rightarrow$ so mixing time is $\Omega(K)$

3. Metropolis HMC with Leapfrog Integrator

To sample from $\nu(x) \propto e^{-f(x)}$

* From $x_u \sim g_u$:

  - Run Leapfrog integrator (to approx. Hamiltonian flow $\dot{X}_t = V_t$, $\dot{V}_t = -\nabla f(X_t)$)

    from $(\tilde{X}_m, \tilde{V}_m)$:

$$\begin{cases} \tilde{V}_{m+\frac{1}{2}} = \tilde{V}_m - \frac{\eta}{2} \nabla f(\tilde{X}_m) \\ \tilde{X}_{m+1} = \tilde{X}_m + \eta \tilde{V}_{m+\frac{1}{2}} \\ \tilde{V}_{m+1} = \tilde{V}_{m+\frac{1}{2}} - \frac{\eta}{2} \nabla f(\tilde{X}_{m+1}) \end{cases}$$

    for M times, from $(\tilde{X}_0, \tilde{V}_0) = (x_u, z_u)$, $z_u \sim N(0, I)$

  - proposed point $\tilde{X}_M$

  - (Metropolis-Hastings): accept / reject

$$\text{set } x_{k+1} = \begin{cases} \tilde{x}_M & \text{with prob. } \alpha_k = \min\left\{1, \dfrac{\tilde{\nu}(\tilde{x}_M, \tilde{v}_M)}{\tilde{\nu}(\tilde{x}_0, \tilde{v}_0)}\right\} \\ x_k & \text{with prob. } 1-\alpha_k. \end{cases}$$

Notes:

1) Acceptance prob. α_k is simple, because leapfrog is invertible

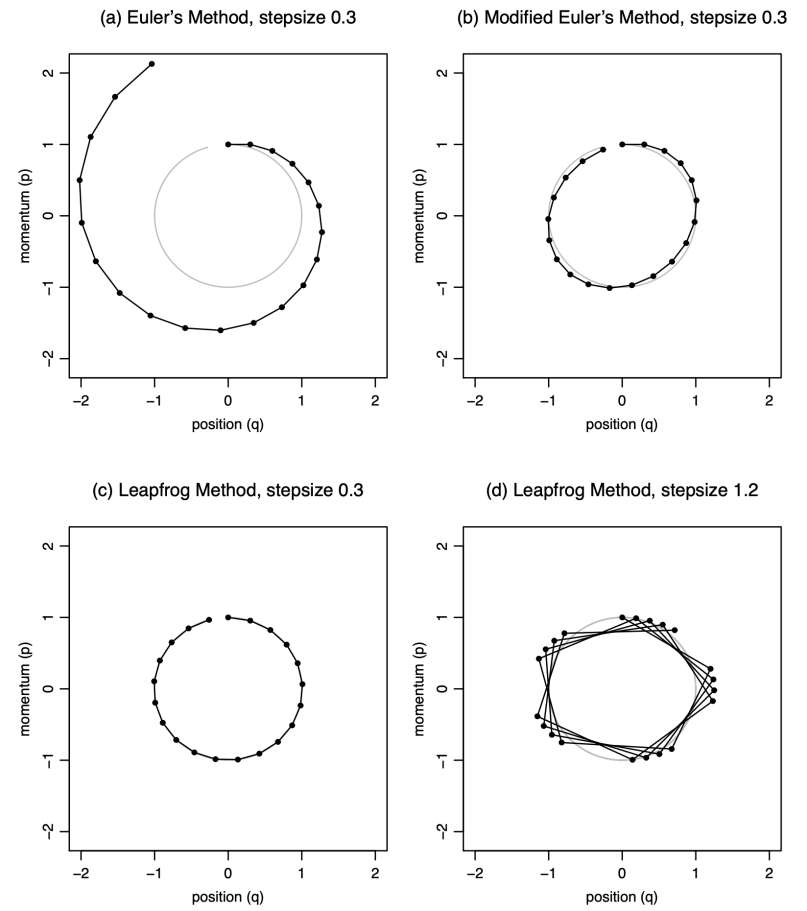
2) M = # of steps, each with step size η

$\Rightarrow$ $\sim$ approximating Hamiltonian flow for time $\eta \cdot M = T$

3) $M = 1$, then this recovers MALA

because $\tilde{x}_{k+1} = x_k - \frac{\eta^2}{2} \nabla f(x_k) + \eta z_k, \quad z_k \sim \mathcal{N}(0, I)$

Leapfrog Integrator



[Neal, *MCMC using Hamiltonian dynamics*, Handbook of MCMC, 2010]

Metropolis HMC with Leapfrog

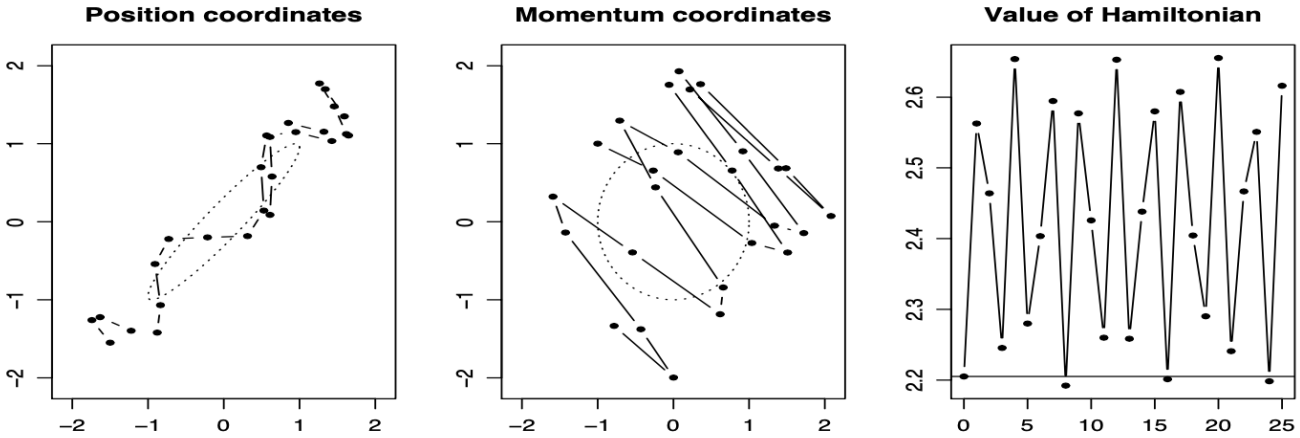


Figure 3: A trajectory for a 2D Gaussian distribution, simulated using 25 leapfrog steps with a stepsize of 0.25. The ellipses plotted are one standard deviation from the means. The initial state had $q = [-1.50, -1.55]^T$ and $p = [-1, 1]^T$.

[Neal, *MCMC using Hamiltonian dynamics*, Handbook of MCMC, 2010]

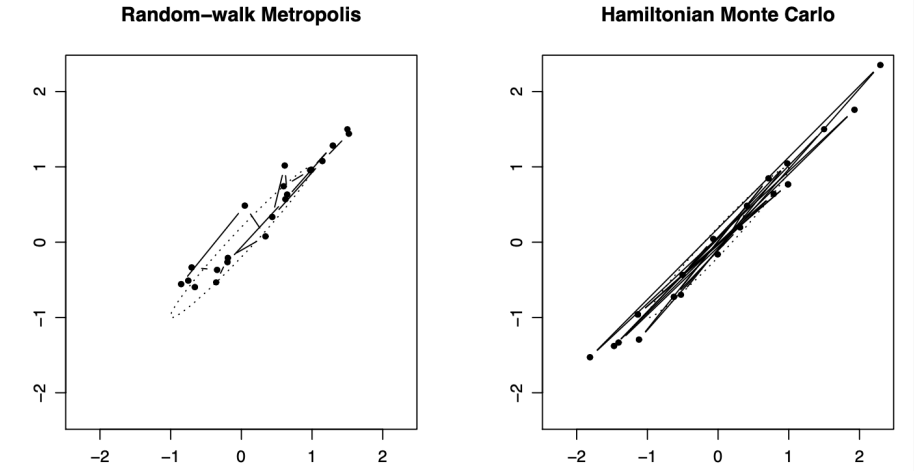


Figure 4: Twenty iterations of the random-walk Metropolis method (with 20 updates per iteration) and of the Hamiltonian Monte Carlo method (with 20 leapfrog steps per trajectory) for a 2D Gaussian distribution with marginal standard deviations of one and correlation 0.98. Only the two position coordinates are plotted, with ellipses drawn one standard deviation away from the mean.

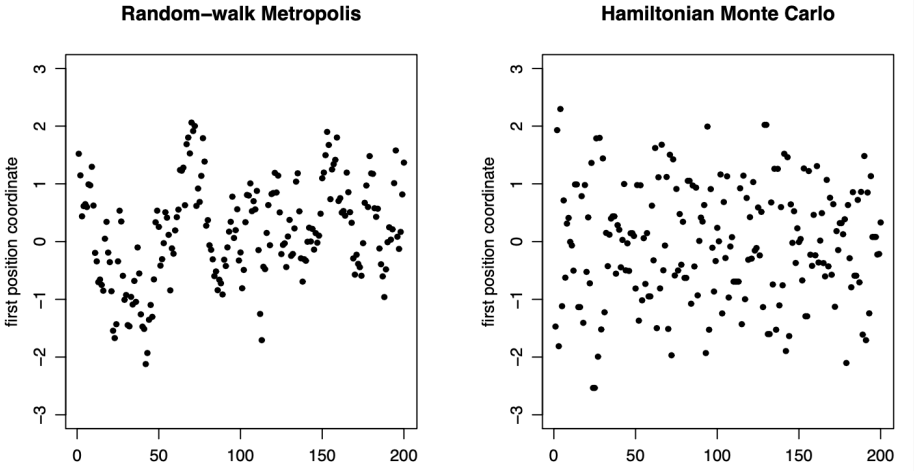


Figure 5: Two hundred iterations, starting with the twenty iterations shown above, with only the first position coordinate plotted.

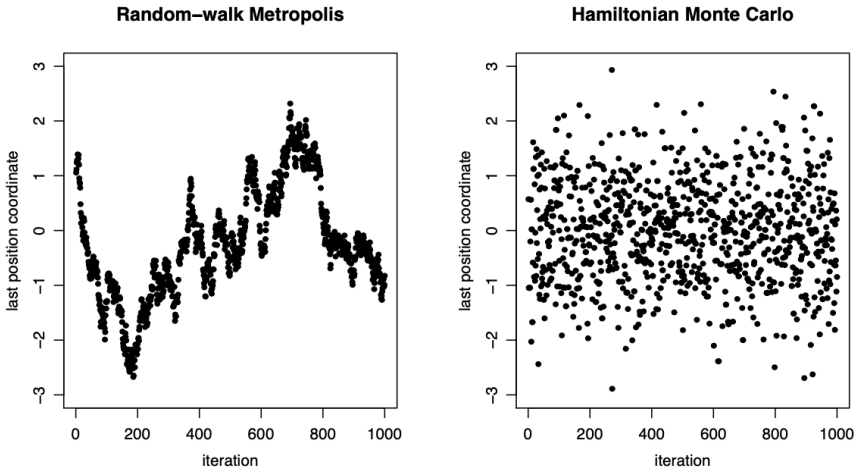


Figure 6: Values for the variable with largest standard deviation for the 100-dimensional example, from a random-walk Metropolis run and an HMC run with $L = 150$. To match computation time, 150 updates were counted as one iteration for random-walk Metropolis.

[Neal, *MCMC using Hamiltonian dynamics*, Handbook of MCMC, 2010]

HMC Mixing Time

$$r(s) := 1 + \max\left\{\left(\frac{\log(1/s)}{d}\right)^{1/4}, \left(\frac{\log(1/s)}{d}\right)^{1/2}\right\}, \tag{11a}$$

for $s > 0$, and involves the step-size choices

$$\eta_{\text{warm}} = \sqrt{\frac{1}{cL \cdot r\left(\frac{\epsilon^2}{2\beta}\right) d^{\frac{7}{6}}}}, \quad \text{and} \quad \eta_{\text{feas}} = \sqrt{\frac{1}{cL \cdot r\left(\frac{\epsilon^2}{2\kappa^d}\right)} \min\left\{\frac{1}{d\kappa^{\frac{1}{2}}}, \frac{1}{d^{\frac{2}{3}}\kappa^{\frac{5}{6}}}, \frac{1}{d^{\frac{1}{2}}\kappa^{\frac{3}{2}}}\right\}}. \tag{11b}$$

With these definitions, we have the following:

Corollary 1. *Consider an (L, L_H, m)-strongly log-concave target distribution Π^* (cf. Assumption (B)) such that $L_H^{2/3} = O(L)$, and any error tolerance $\epsilon \in (0, 1)$.*

(a) Suppose that $\kappa = O(d^{\frac{2}{3}})$ and $\beta = O\left(\exp\left(d^{\frac{2}{3}}\right)\right)$. Then with any β-warm initial distribution μ_0, hyper-parameters $K = d^{\frac{1}{4}}$ and $\eta = \eta_{warm}$, the HMC-(K, η) chain satisfies

$$\tau_2^{HMC}(\epsilon; \mu_0) \leq c\, d^{\frac{2}{3}}\, \kappa\, r\left(\frac{\epsilon^2}{2\beta}\right) \log\left(\frac{\log \beta}{\epsilon}\right). \tag{12a}$$

(b) With the initial distribution $\mu_\dagger = \mathcal{N}(x^\star, \frac{1}{L}\mathbb{I}_d)$, hyper-parameters $K = \kappa^{\frac{3}{4}}$ and $\eta = \eta_{feas}$, the HMC-(K, η) chain satisfies

$$\tau_2^{HMC}(\epsilon; \mu_\dagger) \leq c\, r\left(\frac{\epsilon^2}{2\kappa^d}\right) \max\left\{d \log \kappa, \max\left[d, d^{\frac{2}{3}}\kappa^{\frac{1}{3}}, d^{\frac{1}{2}}\kappa\right] \log\left(\frac{d \log \kappa}{\epsilon}\right)\right\}. \tag{12b}$$

[Chen, Dwivedi, Wainwright, & Yu, *Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients*, JMLR 2020]

HMC Mixing Time: Warm Start

Assume ρ_0 is β-warm with respect to ν

Sampling algorithm	Mixing time	#Gradient evaluations
MRW [24, Theorem 2]	$d\kappa^2 \cdot \log \frac{1}{\epsilon}$	NA
MALA [24, Theorem 1]	$d\kappa \cdot \log \frac{1}{\epsilon}$	$d\kappa \cdot \log \frac{1}{\epsilon}$
HMC-(K, η) [ours, Corollary 1]	$d^{\frac{2}{3}}\kappa \cdot \log \frac{1}{\epsilon}$	$d^{\frac{11}{12}}\kappa \cdot \log \frac{1}{\epsilon}$

Table 2. Summary of the ϵ-mixing time and the corresponding number of gradient evaluations for MRW, MALA and HMC from a *warm start* with an (L, L_{H}, m)-strongly-log-concave target. These statements hold under the assumption $L_{\mathrm{H}}^{2/3} = O(L)$, $\kappa = \frac{L}{m} \ll d$, and omit logarithmic terms in dimension.

[Chen, Dwivedi, Wainwright, & Yu, *Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients*, JMLR 2020]

HMC Mixing Time: Feasible Start

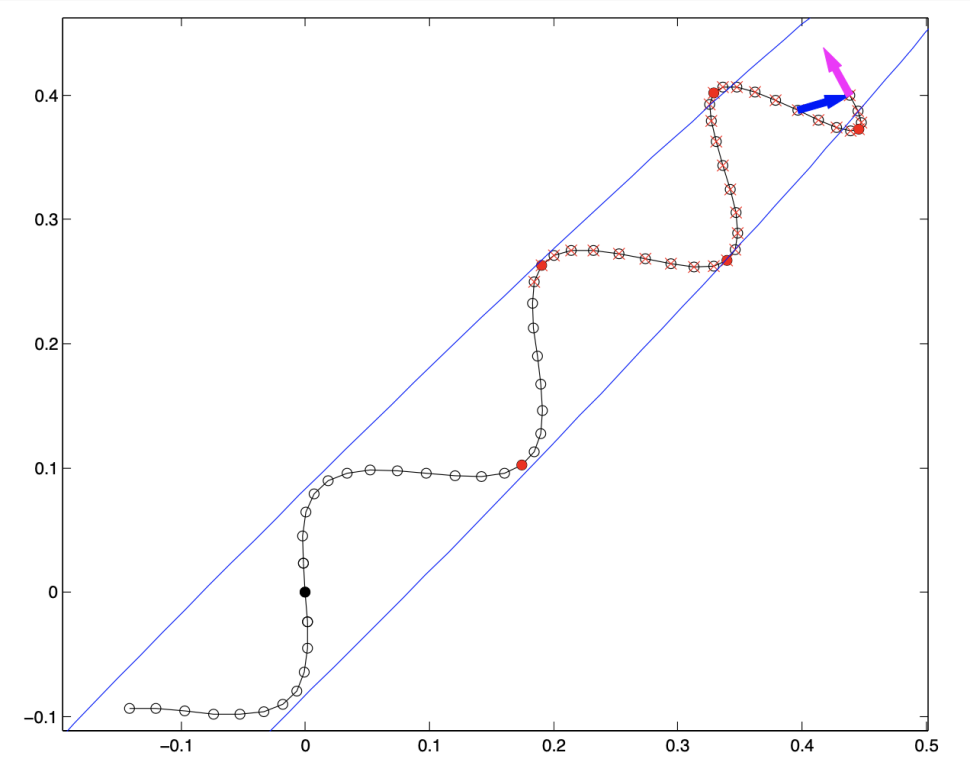
Assume $\rho_0 = \mathcal{N}(x^*, \frac{1}{L} I)$, which is warm with $\beta = \kappa^{d/2}$

Sampling algorithm	Mixing time	# Gradient Evaluations	
		general κ	$\kappa \ll d^{\frac{1}{2}}$
MRW [ours, Theorem 2]	$d\kappa^2$	NA	NA
MALA [ours, Theorem 2]	$\max\left\{d\kappa, d^{\frac{1}{2}}\kappa^{\frac{3}{2}}\right\}$	$\max\left\{d\kappa, d^{\frac{1}{2}}\kappa^{\frac{3}{2}}\right\}$	$d\kappa$
HMC-(K,η) [ours, Corollary 1]	$\max\left\{d, d^{\frac{2}{3}}\kappa^{\frac{1}{3}}, d^{\frac{1}{2}}\kappa\right\}$	$\max\left\{d\kappa^{\frac{3}{4}}, d^{\frac{2}{3}}\kappa^{\frac{13}{12}}, d^{\frac{1}{2}}\kappa^{\frac{7}{4}}\right\}$	$d\kappa^{\frac{3}{4}}$

Table 3. Summary of the ϵ-mixing time and the corresponding number of gradient evaluations for MRW, MALA and HMC from the *feasible start* $\mu_\dagger = \mathcal{N}(x^\star, \frac{1}{L}\mathbb{I}_d)$ for an (L, L_{H}, m)-strongly-log-concave target. Here $x^\star$ denotes the unique mode of the target distribution. These statements hold uner the assumption $L_{\mathrm{H}} = O(L^{\frac{3}{2}})$, and hide the logarithmic factors in ϵ, d and $\kappa = L/m$.

[Chen, Dwivedi, Wainwright, & Yu, *Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients*, JMLR 2020]

No-U-Turn Sampler



[Hoffman & Gelman, *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, JMLR 2014]

References

- Duane, Kennedy, Pendleton, & Roweth, *Hybrid Monte Carlo*, Physics Letters B, 1987
- Neal, *MCMC using Hamiltonian dynamics*, Handbook of MCMC, 2010
- Girolami & Calderhead, *Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods*, Journal of the Royal Statistical Society: Series B, 2011
- Hoffman & Gelman, *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, JMLR 2014
- Bou-Rabee & Sanz-Serna, *Randomized Hamiltonian Monte Carlo*, The Annals of Applied Probability, 2017
- Lee & Vempala, *Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope computation*, STOC 2018
- Chen & Vempala, *Optimal Convergence Rate of Hamiltonian Monte Carlo for Strongly Logconcave Distributions*, APPROX/RANDOM 2019
- Chen, Dwivedi, Wainwright, & Yu, *Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients*, JMLR 2020
- Lu & Wang, *On Explicit L^2-Convergence Rate Estimate for Piecewise Deterministic Markov Processes in MCMC Algorithms*, arXiv 2021

Recap

I. Classical Sampling

1. Introduction
2. Markov Chain
3. Spectral Theory
4. Conductance
5. s-Conductance
6. Metropolis-Hastings
7. Isoperimetry
8. From Isoperimetry to Conductance
9. MRW
10. MALA

II. Sampling as Optimization

11. Optimization and Dynamics
12. Optimization Algorithms
13. Optimization on Manifold
14. Wasserstein Metric
15. Otto Calculus
16. Potential Energy
17. Entropy
18. Langevin Dynamics
19. Langevin Dynamics under Isoperimetry
20. ULA
21. ULA Analysis

III. Variations

22. Variations of Langevin Dynamics
23. Underdamped Langevin Dynamics
24. Hamiltonian Monte Carlo
25. Student presentation I
26. Student presentation II

Thank you!