~~Quiz 7: tonight~~

HW 4 → ~~2~~ 1 weeks

Quiz 8: ~~next Wed~~ tonight, conjugate Bayesian

Hamiltonian MCMC

default sampler for many PPLs

stan, numpyro

highly automated

Dan Sheldon

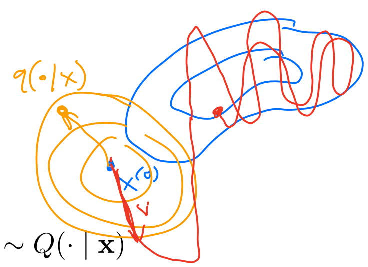
April 15, 2024

Review: Metropolis-Hastings

P

- Given: probability density $P(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$
- Goal: generate sample $\mathbf{x} \sim P$

Metropolis-Hastings

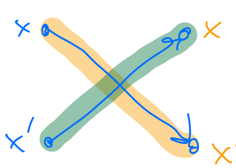


- Initialize $\mathbf{x}^{(0)}$ arbitrarily
- Given $\mathbf{x}^{(t)} = \mathbf{x}$, propose

$$\mathbf{x}' \sim Q(\cdot \mid \mathbf{x})$$

- Accept and set $\mathbf{x}^{(t+1)} = \mathbf{x}'$ with probability $\min(a, 1)$

$$a = \frac{P(\mathbf{x}')}{P(\mathbf{x})} \cdot \frac{Q(\mathbf{x} \mid \mathbf{x}')}{Q(\mathbf{x}' \mid \mathbf{x})}$$

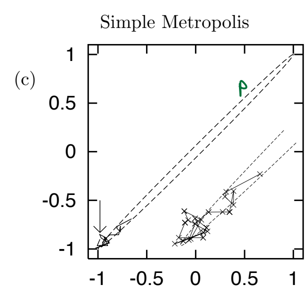


- Else reject and set $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}$

For large enough T, have $\mathbf{x}^{(T)} \sim P$

Problem: random walk

Slow mixing due to "random walk" behavior



Why? Typical proposal is a random *displacement*

- Spherical Gaussian → Brownian motion-like
- Ignores density surface

dist after n steps is order $\sqrt{n}$

Main Idea

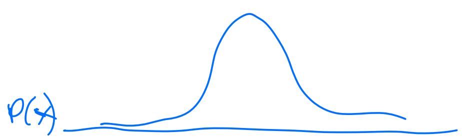
- Idea: use density to guide proposals
- Select random *velocity* $\mathbf{p}/m \in \mathbb{R}^d$
  - $\mathbf{p}$ = momentum, m = mass
- Simulate motion on energy surface

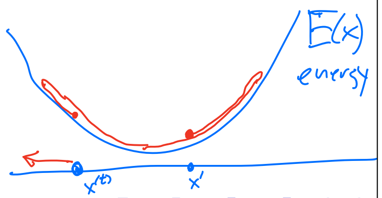
$$\{(\mathbf{x}, E(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^d\} \subseteq \mathbb{R}^{d+1}, \quad E(\mathbf{x}) = -\log P(\mathbf{x})$$

with initial velocity $\mathbf{p}/m$ for some amount of time to get proposal $\mathbf{x}'$.

$Q(x'|x)$

challenge: density



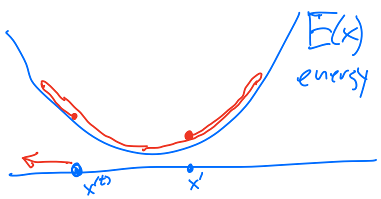


Main Idea

Demo: $P(x)$, $E(x)$, motion in 1D

Hamiltonian Mechanics

- Position $\mathbf{x} \in \mathbb{R}^d$
- Velocity $\mathbf{p}/m \in \mathbb{R}^d$
- Potential energy $E(\mathbf{x})$ (= height)



- Temporal dynamics

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{m} \leftarrow \text{vel.}$$
$$\frac{d\mathbf{p}}{dt} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} = -E'(x)$$

Puck of mass m sliding on frictionless surface with velocity $\mathbf{p}/m$, height at $\mathbf{x}$ equal to $E(\mathbf{x})$ (and thus "incline" $\partial E(\mathbf{x})/\partial \mathbf{x}$).

Generalization: Kinetic Energy

scalar $\frac{d}{dp}\frac{p^2}{2m} = \frac{p}{m}$

vector $\frac{d}{dp}\frac{p^Tp}{2m} = \frac{p}{m}$

Define $K(\mathbf{p}) = \dfrac{\mathbf{p}^T\mathbf{p}}{2m}$

kinetic energy

$$\boxed{\begin{aligned}\frac{d\mathbf{x}}{dt} &= \frac{\mathbf{p}}{m}\\ \frac{d\mathbf{p}}{dt} &= -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\end{aligned}} \Longrightarrow \boxed{\begin{aligned}\frac{d\mathbf{x}}{dt} &= \frac{\partial K(\mathbf{p})}{\partial \mathbf{p}}\\ \frac{d\mathbf{p}}{dt} &= -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\end{aligned}}$$

Generalization: The Hamiltonian

$\frac{\partial}{\partial p}\left(E(x)+K(p)\right) = \frac{\partial}{\partial p}K(p)$

Define $H(\mathbf{x},\mathbf{p}) = E(\mathbf{x}) + K(\mathbf{p})$ $\quad \frac{\partial}{\partial x}\left(E(x)+K(p)\right) = \frac{\partial}{\partial x}E(x)$

Hamiltonian or total energy

$$\boxed{\begin{aligned}\frac{d\mathbf{x}}{dt} &= \frac{\partial K(\mathbf{p})}{\partial \mathbf{p}} = \frac{p}{m}\\ \frac{d\mathbf{p}}{dt} &= -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\end{aligned}} \Longrightarrow \boxed{\begin{aligned}\frac{d\mathbf{x}}{dt} &= \frac{\partial H(\mathbf{x},\mathbf{p})}{\partial \mathbf{p}}\\ \frac{d\mathbf{p}}{dt} &= -\frac{\partial H(\mathbf{x},\mathbf{p})}{\partial \mathbf{x}}\end{aligned}}$$

Simluating Hamiltonian Mechanics

$$\begin{matrix} x(t) \\ p(t)\end{matrix} \longmapsto \begin{matrix} x(t+\epsilon) \\ p(t+\epsilon)\end{matrix}$$

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{m} \qquad \frac{d\mathbf{p}}{dt} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}$$

Euler's method

$$\mathbf{x}(t+\varepsilon) = \mathbf{x}(t) + \varepsilon\frac{\mathbf{p}(t)}{m}$$

$$\mathbf{p}(t+\varepsilon) = \mathbf{p}(t) - \varepsilon\frac{\partial E(\mathbf{x}(t))}{\partial \mathbf{x}}$$

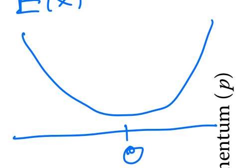
Problem: numerically unstable

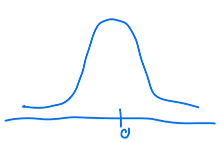
Euler's Method

$E(x) = \frac{x^2}{2}$ $\quad K(p) = \frac{p^2}{2}$

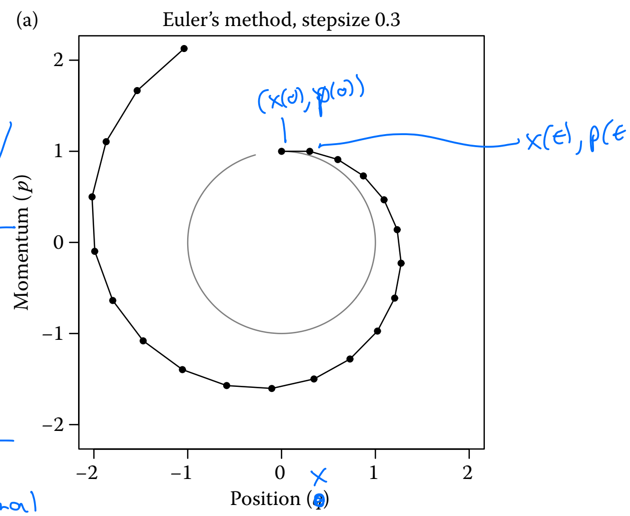
total energy $= H(x,p) = \frac{x^2}{2} + \frac{p^2}{2}$ = const

$E(x)$





$p(x)$ = std. normal



Leapfrog Method

Euler: $\begin{matrix} x(t) \\ p(t)\end{matrix} \longmapsto \begin{matrix} x(t+\epsilon) \\ p(t+\epsilon)\end{matrix}$

Leapfrog

1. $p(t) \longmapsto p(t+\frac{\epsilon}{2})$ "half step"
2. $x(t) \longmapsto x(t+\epsilon)$ "full step"
3. $p(t+\frac{\epsilon}{2}) \longmapsto p(t+\epsilon)$ "half step"

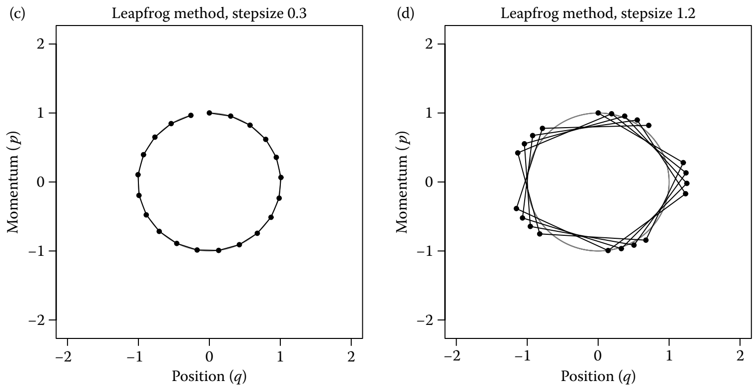
More accurate and stable method

$$\mathbf{p}(t+\varepsilon/2) = \mathbf{p}(t) - (\varepsilon/2)\frac{\partial E(\mathbf{x}(t))}{\partial \mathbf{x}}$$

$$\mathbf{x}(t+\varepsilon) = \mathbf{x}(t) + \varepsilon\frac{\mathbf{p}(t+\varepsilon/2)}{m} \qquad t \qquad t+\epsilon$$

$$\mathbf{p}(t+\varepsilon) = \mathbf{p}(t+\varepsilon/2) - (\varepsilon/2)\frac{\partial E(\mathbf{x}(t+\varepsilon))}{\partial \mathbf{x}}$$

Leapfrog Method



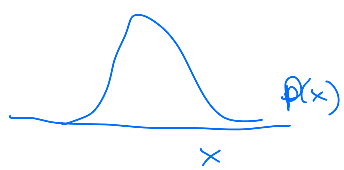
Hamiltonian MCMC

Random velocity/momentum instead of random displacement

- Start at $\mathbf{x}$
- Choose random momentum $\mathbf{p} \sim \exp(-\mathbf{p}^T\mathbf{p}/2m)$
- Simluate Hamiltonian mechanics for s time units $\longrightarrow$ end at $\mathbf{x}'$
- Propose $\mathbf{x}'$

Problem: how to compute $\boxed{Q(\mathbf{x}' \mid \mathbf{x})}$ for acceptance probability?

Auxilliary Variables



pos mom.

Sample both $\mathbf{x}$ and $\mathbf{p}$ from

$E(x) + K(p)$

$$P(\mathbf{x}, \mathbf{p}) \propto \exp(-H(\mathbf{x}, \mathbf{p}))$$
$$\propto \underbrace{\exp(-E(\mathbf{x}))}_{p(x)} \underbrace{\exp(-K(\mathbf{p}))}_{\text{"prior over } p\text{"}},$$

when done, discard $\mathbf{p}$ values

Note: $\mathbf{x}$ and $\mathbf{p}$ are independent

Given:

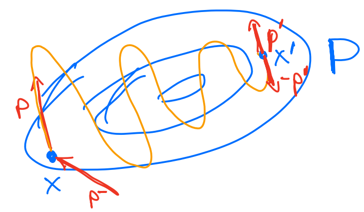
$P(x) \xrightarrow{\text{augment}} P(x, p)$

$= P(x) \cdot \exp(-K(p))$

$x^{(1)}, \cancel{p^{(1)}}$
$x^{(2)}, \cancel{p^{(2)}}$
$x^{(3)}, \cancel{p^{(3)}}$
$\vdots$

Hamiltonian MCMC

$$a = \frac{P(x', p')\,Q(x, p \mid x', p')}{P(x, p)\,Q(x', p' \mid x, p)}$$



Gibbs step

- Start at $(\mathbf{x}, \mathbf{p}^-)$
- Choose random momentum $\mathbf{p} \sim \exp(-\mathbf{p}^T\mathbf{p}/2m)$
- End at $(\mathbf{x}, \mathbf{p})$

$E(x) = -\log P(x)$

Metropolis-Hastings step

- Start at $(\mathbf{x}, \mathbf{p})$
- Simluate Hamiltonian mechanics $\rightarrow$ end at $(\mathbf{x}', \mathbf{p}')$
- Propose $(\mathbf{x}', -\mathbf{p}')$

Acceptance Probability?

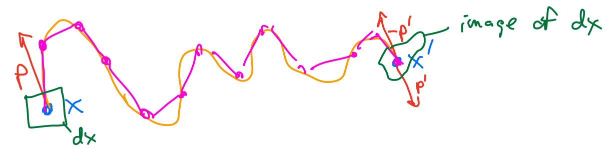
$$P(x,p) = \exp(-E(x))\exp(-K(p))$$
$$= \exp(-E(x) - K(p))$$

$$
\begin{aligned}
a &= \boxed{\frac{P(\mathbf{x}', -\mathbf{p}')}{P(\mathbf{x}, \mathbf{p})}} \cdot \frac{Q(\mathbf{x}, \mathbf{p} \mid \mathbf{x}', -\mathbf{p}')}{Q(\mathbf{x}', -\mathbf{p}' \mid \mathbf{x}, \mathbf{p})} \\
&= \frac{P(\mathbf{x}', -\mathbf{p}')}{P(\mathbf{x}, \mathbf{p})} \qquad \text{(reversibility, volume preservation)} \\
&= \exp(E(\mathbf{x}) - E(\mathbf{x}') + K(\mathbf{p}) - K(\mathbf{p}')) \qquad (K(\mathbf{p}') = K(-\mathbf{p}')) \\
&\approx 1 \qquad \text{(conservation of energy)}
\end{aligned}
$$

Reversibility

Let $T_{L,\varepsilon} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ be simulation mapping with L steps at time increment ε

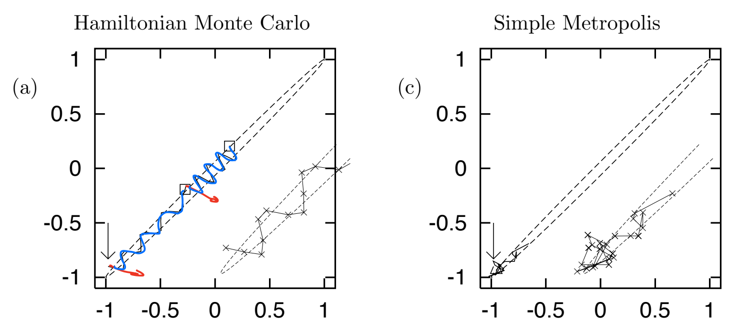
$$T_{L,\varepsilon}(\mathbf{x}, \mathbf{p}) = (\mathbf{x}', \mathbf{p}') \quad \Longrightarrow \quad T_{L,\varepsilon}(\mathbf{x}', -\mathbf{p}') = (\mathbf{x}, -\mathbf{p})$$



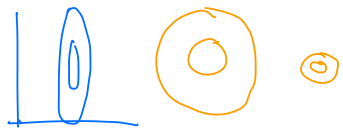
Demo

Demo with sampling

Example



Example



Setup: 100D Gaussian, standard deviations in different dimensions are $0.01, 0.02, \ldots, \boxed{1.00}$