Weird Rejection Sampler

$$\theta \sim \pi(\theta)$$

1) $X \sim \pi(\theta)$ 3) $u \sim \text{uni}(0,1)$

2) $\theta^* = f(X)$

$$q(\theta) = \pi(f^{-1}(\theta^*))\,|\nabla f^{-1}|$$

If

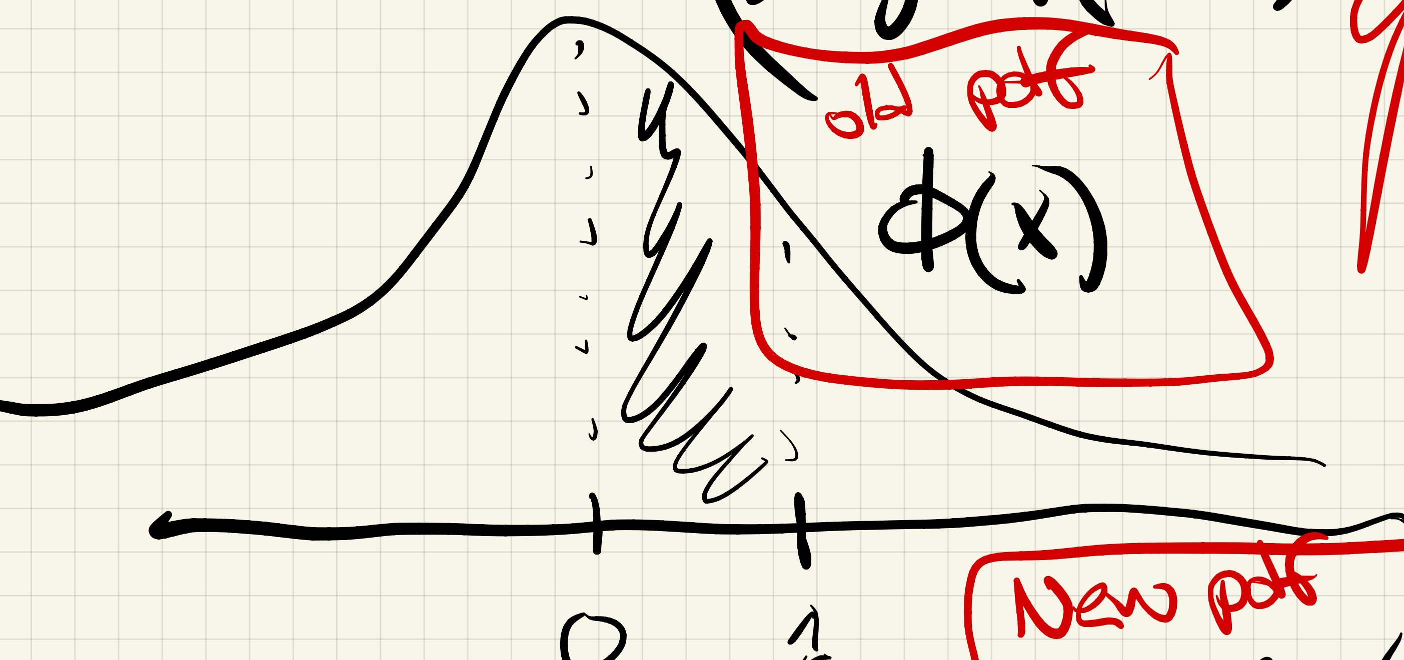
$$u < \frac{\pi(\theta^*)\,|\nabla f|}{\pi(X)\,M}$$

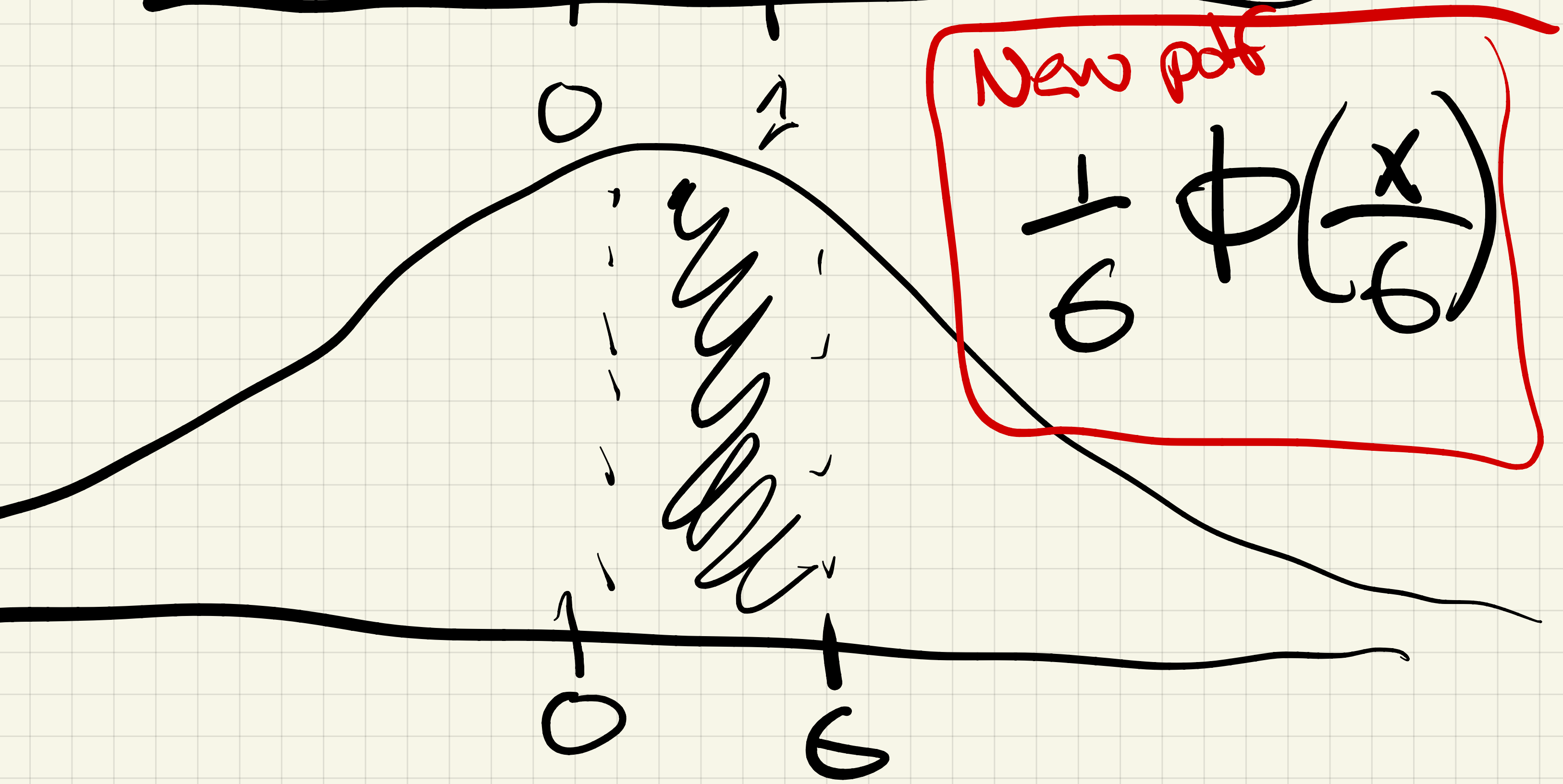
$$|\nabla f^{-1}| = \frac{1}{|\nabla f|}$$

Then $\theta \longmapsto \theta^*$ $\theta \sim \pi(\theta)$

Metropolis-Hastings-Green

$$a = \min\left(1, \frac{\pi(\theta^*)}{\pi(\theta^{(s-1)})} * |\nabla g|\right)$$





q position $\qquad d$

p momentum $\qquad 2d$

$H(q, p)$

$$\boxed{\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = \frac{-\partial H}{\partial q}}$$

1) $T_t(q_0, p_0) = (q_t, p_t)$ Reversible?

$T_t^{-1}(q_t, p_t) = T_{-t}(q_t, p_t) = (q_0, p_0)$

---

2) $$\frac{dH}{dt} = \sum_{i=1}^{d} \frac{\partial H}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial H}{\partial p_i}\frac{dp_i}{dt}$$

$$= \sum_{i=1}^{d} \frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_i}$$

$$= 0$$

Conservation of energy

3) $|\nabla T| = 1$ Volume preservation

If divergence $= 0$
then volume is preserved

$$\nabla \cdot F = \sum_{i=1}^{d} \nabla F_i$$

$$F = \cancel{\sum_{i=1}^{d} \frac{dq_i}{dt}} + \frac{dp_i}{dt}$$

$$F = \begin{pmatrix} \frac{dq_1}{dt} \\ \vdots \\ \frac{dq_d}{dt} \\ \frac{dp_1}{dt} \\ \vdots \\ \frac{dp_d}{dt} \end{pmatrix}$$

$$\sum_{i=1}^{d} \frac{\partial}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial}{\partial p_i}\frac{dp_i}{dt}$$

$$= \sum_{i=1}^{d} \frac{\partial}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial}{\partial p_i}\frac{\partial H}{\partial q_i}$$

$$= 0$$

$$q \sim \pi(q) \qquad q \perp p$$

$$p \sim \xi(p) \propto \exp\left(-\frac{p^T M^{-1} p}{2}\right)$$

joint distribution has pdf

$$\pi(q)\,\xi(p)$$

By fiat: $H(q,p) = -\log\left(\pi(q)\,\xi(p)\right)$

$$> 0$$

$$H(q,p) = -\log \pi(q)\,\xi(p)$$

$$= -\log \pi(q) = U(q)$$

$$- \log \xi(p) = K(p)$$

Algorithm for sampling $q \sim \pi(q)$

1) choose any t $\quad q_0, p_0$

2) Advance system according to

$$\frac{\partial q}{\partial t} = \frac{\partial H}{\partial p} = M^{-1} p$$

$$\frac{\partial p}{\partial t} = -\frac{\partial H}{\partial q} = \underline{\nabla \log \pi(q)}$$

3) stop at time t

4)* MHG

$$\alpha = 1 \wedge \frac{\pi(q_+)\,\xi(p_+)}{\pi(q_0)\,\xi(p_0)} \quad \cancel{\nabla \pi_+}$$

$$H(q,p) = U(q) + K(p)$$

potential   kinetic

$$K(p) = \frac{p^T M^{-1} p}{2}$$ Gaussian

$$K(p) = |p|_1$$ Laplace

$$a = 1 \wedge \frac{\pi(q_t)\xi(p_t)}{\pi(q_0)\xi(p_0)} \cancel{\wedge} \ \xrightarrow{1}$$

$$= 1 \wedge \frac{\pi(q_t)\xi(p_t)}{\pi(q_0)\xi(p_0)}$$

$$\log(a) = 0 \wedge$$

$$-\log(\pi_t \xi_t) + \log(\pi_0 \xi_0)$$

$$= 0 \wedge -H(q_t, p_t) + H(p_0, \xi_0)$$

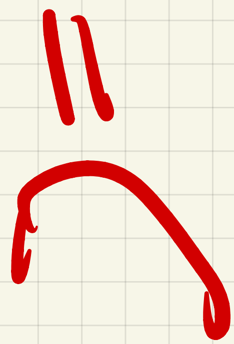
Conservation of Energy $\frac{\partial H}{\partial t} = 0 \Rightarrow \log(a) = 0$

$$\Rightarrow a = 1$$

M?

Think Newton's Method:

$$M = \nabla^2 \log \pi(q)$$

$M(q) =$ 

If $q \sim$ Gaussian

$$\mathcal{N}(0, M^{-1})$$

Harmonic oscillator

$$\Rightarrow \frac{\partial q}{dt} = \frac{\partial H}{\partial p} = M^{-1} p$$

$$\frac{\partial p}{dt} = -\frac{\partial H}{\partial q} = M q$$

Euler's method

$$q_t, p_t, \varepsilon > 0$$

$$q(t+\varepsilon) = q(t) + \varepsilon M^{-1} p(t)$$

$$p(t+\varepsilon) = p(t) - \varepsilon \nabla U(q(t))$$

$$= p(t) + \varepsilon \nabla \log \pi(q(t))$$

Modified Euler

$$q(t+\varepsilon) = q(t) + \varepsilon M^{-1} p(t)$$

$$p(t+\varepsilon) = p(t) - \varepsilon \nabla U(q(t+\varepsilon))$$

Leapfrog (Störmer-Verlet)

$$p(t+\tfrac{\varepsilon}{2}) = p(t) - \tfrac{\varepsilon}{2} \nabla U(q(t)) \quad \checkmark$$

$$q(t+\varepsilon) = q(t) + \varepsilon M^{-1} p(t+\varepsilon/2) \quad \checkmark$$

$$p(t+\varepsilon) = p(t+\tfrac{\varepsilon}{2}) - \tfrac{\varepsilon}{2} \nabla U(q(t+\varepsilon))$$

$$p(t+\varepsilon+\tfrac{1}{2}\varepsilon) = p(t+\varepsilon) - \tfrac{\varepsilon}{2} \nabla U(q(t+\varepsilon))$$

$q \sim \pi(q) \qquad p \sim N(0, M)$

$q \sim N(0, \Sigma) \qquad p \sim N(0, \Sigma^{-1})$

$$p(t + \varepsilon/2) = p(t) - \frac{\varepsilon}{2} \nabla U(q(t))$$

Ask: $p, \nabla U(q) \sim N(0, \Sigma^{-1})$

$$U(q) \propto \frac{1}{2} q^T \Sigma^{-1} q$$

$$\nabla U(q) = \Sigma^{-1} q \sim N(0, \Sigma^{-1} \Sigma \Sigma^{-1})$$

$$\equiv N(0, \Sigma^{-1})$$

$$q(t + \varepsilon) = q(t) + \varepsilon M^{-1} p(t + \varepsilon/2)$$

$$q, M^{-1} p \sim N(0, \Sigma)$$

if $p \sim N(0, \Sigma^{-1})$

$$\Rightarrow M = \Sigma^{-1}$$

$$M^{-1} p = \Sigma p \sim N(0, \cancel{\Sigma} \cancel{\Sigma^{-1}} \Sigma)$$

Connection to 1st Year Asymptotics

$$\ell(\theta)$$

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta) \sim N(\theta, I^{-1}(\theta))$$

$$\nabla \ell(\hat{\theta}) \sim N(0, I(\theta))$$

<u>Assume</u>: $\theta \sim N(0, \Sigma)$

$$I(\theta) = -\nabla^2 \log(\pi(\theta))$$

$$= \nabla^2 \theta^T \Sigma^{-1} \theta / 2$$

$$= \nabla(\nabla \underbrace{\theta^T \Sigma^{-1} \theta}_{2}) = \nabla(\Sigma^{-1}\theta) = \Sigma^{-1}$$

See "Riemannian manifold HMC"

HMC :(

1) ill-conditioning $\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_D \end{pmatrix}$

λ_1 / λ_D BIG

2) Tuning (L, ε, M)

3) $\nabla U(q)$ evals