Problem 1 (1 point)

We are interested in sampling from a target distribution $p(\theta)$ which is an equal mixture of two univariate normals: $N(2, 1^2)$ and $N(-2, 1^2)$.

We are interested in sampling using Hamiltonian Monte Carlo with a distribution for the auxiliary momentum variable given by $p(m) = N(0, 1^2)$.

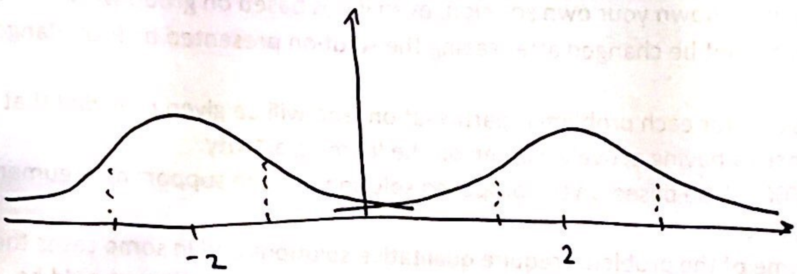
QUESTION: Sketch a plot of the density of the target distribution $p(\theta)$.

QUESTION: Derive an expression for the Hamiltonian $H(\theta, m)$ associated with this sampler.

QUESTION: Provide a qualitative sketch of the level sets associated with the Hamiltonian in (θ, m)-space.

$p(\theta)$:

1



2

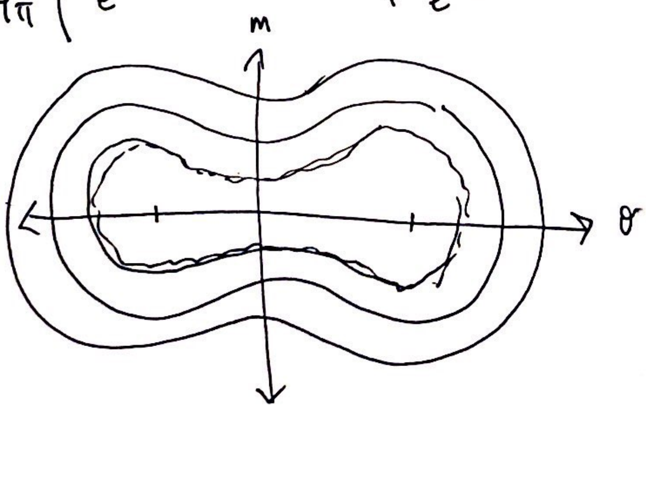
$$H = \log(p(\theta)) + \log(p(m))$$

$$= \log\left[\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-2)^2} + \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}(\theta+2)^2}\right] + \log\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}m^2}\right]$$

$$= \log\left[\frac{1}{4\pi} e^{-\frac{1}{2}(\theta-2)^2 - \frac{1}{2}m^2} + \frac{1}{4\pi} e^{-\frac{1}{2}(\theta+2)^2 - \frac{1}{2}m^2}\right]$$

$$= \log\left[\frac{1}{4\pi}\left(e^{-\frac{1}{2}(x^2 - 4x + 4) - \frac{1}{2}m^2} + e^{-\frac{1}{2}(x^2 + 4x + 4) - \frac{1}{2}m^2}\right)\right]$$

3



Problem 2 (1 point)

Consider further the setting of Problem 1. Assume that $\varepsilon = 0.1$ and $L = 10$.

QUESTION: Derive the formula for $\bigtriangledown \log p(\theta)$.

QUESTION: Derive the expressions for the updating rules for the θ and m parameters in each leap-frogging step in the Hamiltonian MC algorithm.

QUESTION: Starting from $\theta^* = 0$ and $m^* = 1$, use these expressions to sketch a trajectory for the proposal generation step of the sampler in the Hamiltonian level sets (which you sketch in Problem 1)?

(1) $\nabla \log(p(\theta)) = \frac{d}{d\theta} \log\left[\frac{1}{4\pi}\left(e^{-\frac{1}{2}(\theta-2)^2} + e^{-\frac{1}{2}(\theta+2)^2}\right)\right]$

$= 2\tanh(2\theta) - \theta$ (using mathematica)

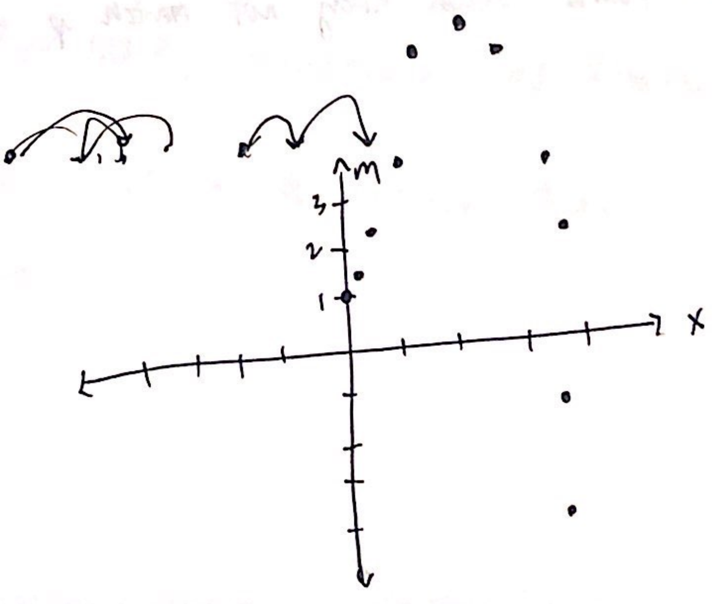
(2) $M = 1 \quad \epsilon = 0.1$

repeat 10 times: $m^* \leftarrow m^* + \frac{1}{2}(0.1) \times 2\tanh(2\theta^*) - \theta^*$

$\theta^* \leftarrow \theta^* + (0.1)\, m^*$

$m^* \leftarrow m^* + \frac{1}{2}(0.1) \times 2\tanh(2\theta^*) - \theta^*$

(3) $\theta_0 = 0 \quad m_0 = 1$



Problem 3 (1 point)

Consider further the setting of Problem 1.

QUESTION: Can you provide an intuition for why the proposals generated by the Hamiltonian MC sampler are similar to the density of the target distribution $p(\theta)$?

By adding $\nabla \log(p(\theta))$ to m^*, we always "pull" m^* down toward the minimum at $\log(p(\theta))$.

This ensures that m^* stays in regions where $p(\theta)$ has most of its mass and therefore that ~~most of them~~ the proposal points are accepted because they come from the distribution $p(\theta)$.

If we were plotting the proposal ~~th~~ trajectory from a worse algorithm with a lower acceptance rate, the proposal points would likely not match up well with $p(\theta)$.

Problem 4 (1 point)

QUESTION: Why does the time that it takes to characterize the posterior distribution increase exponentially with the number of unknown parameters in the grid method, but not in Hamiltonian Monte Carlo?

In the grid method, you need to construct an array with size ~~[illegible]~~ $n_1 \times n_2 \times \ldots n_d$ for d = # of unknown parameters.

$\therefore$ Size $\sim n^d$ grows exponentially with d. In the grid method, you have to do all the calculations $\forall$ grid points.

However, with Hamiltonian MC, the proposal trajectory is "smarter" and it stays in regions of positive / high mass for $p(\theta)$. Thus, it doesn't equally traverse the whole parameter space (which would cause time to go exponentially with d). Instead, it only stays in relevant parts of the parameter space, regardless of its dimension.

↳ this means HMC is much faster than the grid method.