

# STAT 238 - Bayesian Statistics

## Lecture Thirty Five

Spring 2026, UC Berkeley

Aditya Guntuboyina

24 April 2026

### 1 HMC and Leapfrog Discretization

There is a given target probability density  $\pi$  that we need to sample from. In HMC, given a current state  $x$ , one attempts to solve the dynamics:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{v}(t) \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial v}(x(t), v(t)) \\ -\frac{\partial H}{\partial x}(x(t), v(t)) \end{pmatrix} \text{ with initialization } \begin{pmatrix} x(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} x \\ v \end{pmatrix} \quad (1)$$

with  $v \sim N(0, I_d)$ . Here  $H(\cdot, \cdot)$  denotes the Hamiltonian:

$$H(x, v) = H(x_1, \dots, x_d, v_1, \dots, v_d) := -\log \pi(x) + \frac{1}{2} \|v\|^2. \quad (2)$$

Let  $T_\sigma$  denote the function which maps  $(x, v)$  to  $(x(\sigma), v(\sigma))$  obtained by solving (1) for  $t \in [0, \sigma]$ . We saw in the last lecture that  $T_\sigma$  is invertible (specifically,  $T_\sigma S = (T_\sigma S)^{-1}$  where  $S$  is the velocity sign-flip operator), Hamiltonian preserving (i.e.,  $H(T_\sigma(x, v)) = H(x, v)$ ) and Volume preserving (i.e., the determinant of the Jacobian of  $T_\sigma$  always equals 1).

The HMC ODE (1) cannot be solved exactly in practice, so one attempts to solve it approximately via discretization by fixing a step size  $\epsilon$  and solving (1) approximately for  $t = 0, \epsilon, 2\epsilon, \dots$ . To go from  $(x(t), v(t))$  to  $(x(t + \epsilon), v(t + \epsilon))$ , one uses the leapfrog discretization:

$$\begin{aligned} v\left(t + \frac{\epsilon}{2}\right) &= v(t) + \frac{\epsilon}{2} \nabla \log \pi(x(t)) \\ x(t + \epsilon) &= x(t) + \epsilon v\left(t + \frac{\epsilon}{2}\right) \\ v(t + \epsilon) &= v\left(t + \frac{\epsilon}{2}\right) + \frac{\epsilon}{2} \nabla \log \pi(x(t + \epsilon)) \end{aligned} \quad (3)$$

If we denote this map from  $(x(t), y(t))$  to  $(x(t + \epsilon), y(t + \epsilon))$  by  $T_\epsilon^{\text{disc}}(x, v)$  and then apply this mapping  $N$  times in succession starting from  $(x, v)$ , then the overall mapping is given by

$$T_\epsilon^{\text{disc}, N} = T_\epsilon^{\text{disc}} \circ \dots \circ T_\epsilon^{\text{disc}}.$$

The above mapping is an approximation to the continuous ODE flow  $T_\sigma$ :

$$T_\epsilon^{\text{disc}, N}(x, v) \approx T_\sigma(x, v) \quad \text{provided } \epsilon = \frac{\sigma}{N}.$$

In the last lecture, we observed that  $T_\epsilon^{\text{disc},N}$  is also invertible (with  $ST_\epsilon^{\text{disc},N} = (ST_\epsilon^{\text{disc},N})^{-1}$ ) and volume preserving. But it does not preserve the Hamiltonian i.e.,  $H(T_\epsilon^{\text{disc},N}(x, v))$  can be different from  $H(x, v)$ .

While implementing leapfrog discretization to compute  $x(N\epsilon), v(N\epsilon)$ , it can be checked that the following iterations can be used which combine the two velocity updates into one:

$$\begin{aligned}x(j\epsilon) &= x((j-1)\epsilon) + \epsilon v((j-1/2)\epsilon) \\v((j+1/2)\epsilon) &= v((j-1/2)\epsilon) + \epsilon x(j\epsilon)\end{aligned}$$

for  $j = 1, \dots, N-1$ . This iteration needs to be initialized using  $x(0) = x$  and  $v(\epsilon/2) = v(0) + (\epsilon/2)\nabla \log \pi(x(0))$ . At iteration  $N-1$ , we obtain  $x((N-1)\epsilon)$  and  $v((N-1/2)\epsilon)$ . From here  $x(N\epsilon)$  and  $v(N\epsilon)$  can be computed by:

$$\begin{aligned}x(N\epsilon) &= x((N-1)\epsilon) + \epsilon v((N-1/2)\epsilon) \\v(N\epsilon) &= v((N-1/2)\epsilon) + (\epsilon/2)\nabla \log \pi(x(N\epsilon)).\end{aligned}$$

## 2 The HMC Algorithm

We are now ready to formally describe the HMC algorithm. The algorithm starts with an arbitrary initialization  $x^{(0)}$  and then repeats the following for  $t = 0, 1, 2, \dots$ :

1. Generate  $v^{(t)} \sim N(0, I_d)$ .
2. Run discretized Leapfrog Hamiltonian dynamics starting at  $(x^{(t)}, v^{(t)})$  for some fixed step size  $\epsilon$  and trajectory length  $\sigma$ . This gives  $(y, w) = T_\epsilon^{\text{disc},N}(x, v)$  (here  $N = \sigma/\epsilon$  is the number of leapfrog steps).
3. Compute the acceptance probability:

$$\min\left(1, \exp(H(x^{(t)}, v^{(t)}) - H(y, w))\right)$$

and update  $x^{(t+1)}$  as:

$$x^{(t+1)} = \begin{cases} y & \text{with probability } \min(1, \exp(H(x^{(t)}, v^{(t)}) - H(y, w))), \\ x^{(t)} & \text{with probability } 1 - \min(1, \exp(H(x^{(t)}, v^{(t)}) - H(y, w))). \end{cases}$$

We need to verify that the Markov Chain given by the above algorithm maintains detailed balance with respect to  $\pi$ . This is difficult to verify directly for the chain with transitions to  $x^{(t+1)}$  from  $x^{(t)}$ . Instead, we shall verify detailed balance for the chain which transitions to  $(x^{(t+1)}, v^{(t+1)})$  from  $(x^{(t)}, v^{(t)})$ . Note that  $v^{(t+1)}$  does not appear in the above description of the HMC algorithm. We rewrite the algorithm below using  $v^{(t+1)}$  explicitly.

In addition to using  $v^{(t+1)}$  explicitly, another change in the following algorithm is that the equation  $(y, w) = T_\epsilon^{\text{disc},N}(x, v)$  is now changed to  $(y, w) = ST_\epsilon^{\text{disc},N}(x, v)$ . In other words, we do a velocity flip after running the discretized Hamiltonian dynamics. This velocity flip makes the deterministic transition from  $(x, v)$  to  $(y, w)$  an involution (i.e., its inverse is itself:  $(ST_\epsilon^{\text{disc},N})^{-1} = ST_\epsilon^{\text{disc},N}$ ) which makes the argument for detailed balance very clean. Note also that the velocity flip does not change the Hamiltonian (because  $H(y, w) = H(y, -w)$  as the dependence on  $w$  is through  $\|w\|^2$ ) so the acceptance probability does not change. Algorithmically, both the versions of the algorithm lead to identical output.

1. Generate  $v^{(t)} \sim N(0, I_d)$ .
2. Run discretized Leapfrog Hamiltonian dynamics (followed by a velocity flip) starting at  $(x^{(t)}, v^{(t)})$  for some fixed step size  $\epsilon$  and trajectory length  $\sigma$ . This gives  $(y, w) = ST_\epsilon^{\text{disc}, N}(x, v)$  (here  $N = \sigma/\epsilon$  is the number of leapfrog steps).
3. Compute the acceptance probability:

$$\min\left(1, \exp(H(x^{(t)}, v^{(t)}) - H(y, w))\right)$$

and update  $(x^{(t+1)}, v^{(t+1)})$  as:

$$(x^{(t+1)}, v^{(t+1)}) = \begin{cases} (y, w) & \text{with probability } \min\left(1, \exp(H(x^{(t)}, v^{(t)}) - H(y, w))\right), \\ (x^{(t)}, v^{(t)}) & \text{with probability } 1 - \min\left(1, \exp(H(x^{(t)}, v^{(t)}) - H(y, w))\right). \end{cases}$$

Next we will verify detailed balance of the above Markov Chain with respect to the joint density:

$$\tilde{\pi}(x, v) \propto \exp(-H(x, v)).$$

Note that, under  $\tilde{\pi}$ , the variables  $x \sim \pi$  and  $v \sim N(0, I_d)$  are independent (because  $H(x, v) = -\log \pi(x) + \|v\|^2/2$ ).

The transitions of the HMC Markov Chain from  $(x, v)$  to  $(y, w)$  are deterministic (given by the involutive map  $ST_\epsilon^{\text{disc}, N}$ ) so we need to be careful about the form of detailed balance.

### 3 Recap: detailed balance

So far in this course (e.g., see Lectures 25 and 31), we have used the following definition of detailed balance. The Markov Chain with transition densities given by  $Q(x, y)$  satisfies detailed balance with respect to the density  $\pi(x)$  provided:

$$\pi(x)Q(x, y) = \pi(y)Q(y, x). \tag{4}$$

This formulation is only true if the transitions have a density given by  $Q(x, \cdot)$ . For HMC (viewed as a chain over  $(x^{(t)}, v^{(t)})$ ), the transitions are deterministic so we cannot use the detailed balance characterization in (4). The more general way of writing the detailed balance condition is:

$$\mathbb{P}\{\Theta^{(t)} \in A, \Theta^{(t+1)} \in B\} = \mathbb{P}\{\Theta^{(t)} \in B, \Theta^{(t+1)} \in A\} \tag{5}$$

assuming that  $\Theta^{(t)} \in A$  and the conditional distribution of  $\Theta^{(t+1)}$  given  $\Theta^{(t)}$  is given by the Markov chain. Also (5) is required to hold for all subsets  $A$  and  $B$  of the state space.

In the case when the Markov chain transitions are given by densities  $Q(x, \cdot)$ , then it is easy to show that (5) is equivalent to (4). For example, if we restrict  $A$  to be a small region around a point  $x$ , and  $B$  to be a small region around a point  $y$ , then the left hand side of (5) is approximately  $\pi(x)Q(x, y)\text{vol}(A)\text{vol}(B)$  and the right hand side is  $\pi(y)Q(y, x)\text{vol}(A)\text{vol}(B)$ . One can then cancel the product of the volume terms to obtain (4) from (5).

On the other hand, when the Markov chain transitions do not have a density, we need to work with the general condition (5).

## 4 Detailed Balance for HMC

We shall now derive the detailed balance condition (5) for HMC. We need to prove that:

$$\mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in A, (x^{(t+1)}, v^{(t+1)}) \in B \right\} = \mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in B, (x^{(t+1)}, v^{(t+1)}) \in A \right\} \quad (6)$$

where  $(x^{(t)}, v^{(t)})$  has density proportional to  $\exp(-H(x, v))$ . We compute the left hand side as

$$\begin{aligned} & \mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in A, (x^{(t+1)}, v^{(t+1)}) \in B \right\} \\ &= \int_A \exp(-H(x, v)) \mathbb{P} \left\{ (x^{(t+1)}, v^{(t+1)}) \in B \mid x^{(t)} = x, v^{(t)} = v \right\} dx dv \end{aligned}$$

To write the conditional probability  $\mathbb{P} \left\{ (x^{(t+1)}, v^{(t+1)}) \in B \mid x^{(t)} = x, v^{(t)} = v \right\}$ , note that  $(x^{(t+1)}, v^{(t+1)})$  either equals  $ST_\epsilon^{\text{disc}, N}(x, v)$  with probability  $\min(1, \exp(H(x^{(t)}, v^{(t)}) - H(ST_\epsilon^{\text{disc}, N}(x, v))))$  or it equals  $(x, v)$  with the opposite probability. For simplicity of notation, let us denote the function  $ST_\epsilon^{\text{disc}, N}$  by  $\Upsilon$ . Thus

$$\begin{aligned} & \mathbb{P} \left\{ (x^{(t+1)}, v^{(t+1)}) \in B \mid x^{(t)} = x, v^{(t)} = v \right\} \\ &= \min \left( 1, \frac{\exp(-H(\Upsilon(x, v)))}{\exp(-H(x, v))} \right) I\{\Upsilon(x, v) \in B\} + \left[ 1 - \min \left( 1, \frac{\exp(-H(\Upsilon(x, v)))}{\exp(-H(x, v))} \right) \right] I\{(x, v) \in B\}. \end{aligned}$$

We then get

$$\begin{aligned} & \mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in A, (x^{(t+1)}, v^{(t+1)}) \in B \right\} \\ &= \int_A \exp(-H(x, v)) \mathbb{P} \left\{ (x^{(t+1)}, v^{(t+1)}) \in B \mid x^{(t)} = x, v^{(t)} = v \right\} dx dv \\ &= \int_A \exp(-H(x, v)) \min \left( 1, \frac{\exp(-H(\Upsilon(x, v)))}{\exp(-H(x, v))} \right) I\{\Upsilon(x, v) \in B\} dx dv \\ &+ \int_A \exp(-H(x, v)) \left[ 1 - \min \left( 1, \frac{\exp(-H(\Upsilon(x, v)))}{\exp(-H(x, v))} \right) \right] I\{(x, v) \in B\} dx dv \\ &= \int_A \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in B\} dx dv \\ &+ \int_{A \cap B} [\exp(-H(x, v)) - \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v))))] dx dv. \end{aligned}$$

By the same argument (but with  $A$  and  $B$  switched), the right hand side of (6) satisfies:

$$\begin{aligned} & \mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in B, (x^{(t+1)}, v^{(t+1)}) \in A \right\} \\ &= \int_B \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in A\} dx dv \\ &+ \int_{B \cap A} [\exp(-H(x, v)) - \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v))))] dx dv. \end{aligned}$$

It is easy to see that the second terms in the expressions for  $\mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in A, (x^{(t+1)}, v^{(t+1)}) \in B \right\}$  and  $\mathbb{P} \left\{ (x^{(t)}, v^{(t)}) \in B, (x^{(t+1)}, v^{(t+1)}) \in A \right\}$  above are identical. So we only need to prove that the first terms also coincide i.e., we need to show

$$\begin{aligned} & \int_A \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in B\} dx dv \\ &= \int_B \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in A\} dx dv \end{aligned} \quad (7)$$

for all  $A$  and  $B$ . For this, let us take the right hand side integral and use the change of variable  $(y, w) = \Upsilon(x, v)$ , or equivalently  $(x, v) = \Upsilon^{-1}(y, w)$ . Because  $\Upsilon = ST_\epsilon^{\text{disc}, N}$  is an involution, we have  $\Upsilon^{-1} = \Upsilon$ . Thus

$$\begin{aligned} & \int_B \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in A\} dx dv \\ &= \int \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in A, (x, v) \in B\} dx dv \\ &= \int \min(\exp(-H(y, w)), \exp(-H(\Upsilon(y, w)))) I\{\Upsilon(y, w) \in A, (y, w) \in B\} dy dw |\det J\Upsilon(y, w)|. \end{aligned}$$

Because  $\Upsilon$  is volume-preserving (we proved, in the last lecture, that leapfrog discretization preserves volumes), the determinant term above equals one. We thus have

$$\begin{aligned} & \int_B \min(\exp(-H(x, v)), \exp(-H(\Upsilon(x, v)))) I\{\Upsilon(x, v) \in A\} dx dv \\ &= \int \min(\exp(-H(y, w)), \exp(-H(\Upsilon(y, w)))) I\{\Upsilon(y, w) \in A, (y, w) \in B\} dy dw \end{aligned}$$

Similarly, one can also prove that the left hand side of (7) also equals the above. This completes the proof of (7) establishing detailed balance of the HMC markov chain.

For more details behind HMC, the paper Neal [1] is highly recommended.

## References

- [1] Radford M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, pages 47–95. Chapman and Hall/CRC, 2011.