STAT 238 - Bayesian Statistics
Lecture Thirty Three

Spring 2026, UC Berkeley

Aditya Guntuboyina

20 April 2026

1 Hamiltonian Monte Carlo

In HMC, the proposal y is generated by solving the following ODE:

$$x(0) = x \quad \dot{x}(0) = z \quad \ddot{x}(t) = \nabla \log \pi(x(t)) \quad y = x(\sigma). \tag{1}$$

Note that if $\nabla \log \pi(x(t))$ is replaced by $\nabla \log \pi(x)$, then it is easy to check that y equals the MALA proposal $y = x + (\sigma^2/2)\nabla \log \pi(x) + \sigma z$.

The ODE (1) can be written in the following alternative first order form:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{v}(t) \end{pmatrix} = \begin{pmatrix} v(t) \\ \nabla \log \pi(x(t)) \end{pmatrix} \text{ with initialization } \begin{pmatrix} x(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix} \tag{2}$$

It is easy to see that (1) and (2) are equivalent. Here is another common way of writing (2). Let

$$H(x, v) = H(x_1, \dots, x_d, v_1, \dots, v_d) := -\log \pi(x) + \frac{1}{2}\|v\|^2 \tag{3}$$

This quantity $H(x, v)$ is called the Hamiltonian associated with the ODE (2). It is then easy to see that

$$\frac{\partial H}{\partial x} = \left(\frac{\partial H}{\partial x_1}, \dots, \frac{\partial H}{\partial x_d}\right)^T = -\nabla \log \pi(x)$$

and

$$\frac{\partial H}{\partial v} = \left(\frac{\partial H}{\partial v_1}, \dots, \frac{\partial H}{\partial v_d}\right)^T = v.$$

As a result, (2) is equivalent to:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{v}(t) \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial v} \\ -\frac{\partial H}{\partial x} \end{pmatrix} \text{ with initialization } \begin{pmatrix} x(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix} \tag{4}$$

One important property of this ODE is that the Hamiltonian $H(x(t), v(t))$ remains constant in t for all times t:

$$H(x(t), v(t)) = \text{constant} = H(x(0), v(0)) = H(x, z) = -\log \pi(x) + \frac{1}{2}\|z\|^2. \tag{5}$$

To see (5), just note that

$$\begin{aligned}\frac{d}{dt}H(x(t),v(t)) &= \sum_{i=1}^{d}\left(\frac{\partial H}{\partial x_i}\dot{x}_i(t)+\frac{\partial H}{\partial v_i}\dot{v}_i(t)\right)\\ &= \sum_{i=1}^{d}\left(\frac{\partial H}{\partial x_i}\frac{\partial H}{\partial v_i}+\frac{\partial H}{\partial v_i}\left(-\frac{\partial H}{\partial x_i}\right)\right)=\sum_{i=1}^{d}\left(\frac{\partial H}{\partial x_i}\frac{\partial H}{\partial v_i}-\frac{\partial H}{\partial v_i}\frac{\partial H}{\partial x_i}\right)=0\end{aligned}$$

2 Stationarity of π for the HMC

The transition kernel given by (1) satisfies detailed balance with respect to π for every fixed $\sigma > 0$. A consequence of this is that π is invariant (stationary) for this Markov Chain (for every σ). We will sketch a proof of this invariance (and skip the proof of detailed balance).

Stationarity of π means that if $x \sim \pi$, then y given by (1) is also distributed as π. In order to verify this, we shall use the following result on density evolutions of ODEs. This result has been popular in the recent literature on generative modeling (see e.g., Lai et al. [1, Equation (5.2.8), Theorem 5.2.2, Section B.1.2]).

Theorem 2.1. *Consider the ODE*

$$\dot{S}(x,t) = V(t, S(x,t)) \quad \textit{with initial condition } S(x,0) = x. \tag{6}$$

Here $\dot{S}(x,t) = \frac{d}{dt}S(x,t)$, $x \in \mathbb{R}^d$ *and* $S(t,\cdot), V(t,\cdot)$ *are functions from* $\mathbb{R}^d$ *to* $\mathbb{R}^d$*. Suppose* ρ_b *is a density on* $\mathbb{R}^d$ *and let* $\rho(t,\cdot)$ *be the density of* $S(X,t)$ *with* $X \sim \rho_b$*. Then* $\rho(t,x)$ *satisfies the following PDE:*

$$\partial_t \rho(t,x) = -\nabla\cdot(V(t,x)\rho(t,x)) \quad \textit{with } \rho(0,\cdot) = \rho_b. \tag{7}$$

Here $\nabla\cdot$*, also known as the divergence, is defined as follows.* $\nabla\cdot G(x) = \sum_{i=1}^{d}\frac{\partial}{\partial x_i}G_i(x)$ *for a function* $G(x) = (G_1(x),\dots,G_d(x))$*. So* $\nabla\cdot(V\rho) = \operatorname{div}(V\rho) = \sum_{i=1}^{d}\frac{\partial}{\partial x_i}(V_i(x,t)\rho(x,t))$.

Proof of Theorem 2.1. $S(X,t)$ can be interpreted as the position at time t of a particle starting at x in a velocity vector field $V(t,x)$, formally given by the ODE (6). $\rho(t,\cdot)$ is the pdf of $S(X,t)$ with $X \sim \rho_b$.

We need to prove that $\rho(t,x)$ satisfies the PDE (7). Suppose f is a smooth function with compact support (i.e., $f \in \mathcal{C}_c^{\infty}$). Then $\mathbb{E}f(S(X,t))$ can be written in the following two equivalent ways:

$$\mathbb{E}f(S(X,t)) = \int f(x)\rho(t,x)dx = \int f(S(x,t))\rho_b(x)dx.$$

Differentiating both sides of this equality with respect to t, we obtain

$$\int f(x)\frac{\partial}{\partial t}\rho(t,x)dx = \int \left\langle \dot{S}(x,t), \nabla f(S(x,t))\right\rangle \rho_b(x)dx \tag{8}$$

The right hand side above can be simplified using $\dot{S}(x,t) = V(t,S(x,t))$ (because $S(x,t)$

satisfies (6)) to get

$$
\begin{aligned}
\int \left\langle \dot{S}(x,t), \nabla f(S(x,t)) \right\rangle \rho_b(x)dx &= \int \left\langle V(t, S(x,t)), \nabla f(S(x,t)) \right\rangle \rho_b(x)dx \\
&= \int \left\langle V(t, nx), \nabla f(x) \right\rangle \rho(t,x)dx \\
&= \sum_{i=1}^{d} \int V_i(t,x) \frac{\partial f}{\partial x_i}(x)\rho(t,x)dx \\
&= \sum_{i=1}^{d} \int V_i(t,x)\rho(t,x) \frac{\partial f}{\partial x_i}(x)dx.
\end{aligned}
$$

Integration by parts (note that we assumed that f has compact support) and combining with (8) allows us to derive

$$
\begin{aligned}
\int f(x) \frac{\partial}{\partial t}\rho(t,x)dx &= -\sum_{i=1}^{d} \int f(x) \frac{\partial}{\partial x_i} \left(V_i(t,x)\rho(t,x)\right) \\
&= -\int f(x) \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \left(V_i(t,x)\rho(t,x)\right) dx = -\int f(x) \mathrm{div}_x \left(V(t,x)\rho(t,x)\right) dx
\end{aligned}
$$

Because this holds for every f (with compact support), we deduce the pde (7). □

In the next lecture, we shall see how this result implies that π is invariant for the HMC chain. We will basically apply Theorem 2.1 with V given by:

$$
V(t,x,v) = \begin{pmatrix} v \\ \nabla \log \pi(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial v} \\ -\frac{\partial H}{\partial x} \end{pmatrix}.
$$

We shall also discuss the leapfrog discretization of the Hamiltonian ODE (2), and its Metropolization.

References

[1] Lai, C.-H., Song, Y., Kim, D., Mitsufuji, Y., and Ermon, S. (2025). The principles of diffusion models. *arXiv preprint arXiv:2510.21890.*