Notes on Hamiltonian Monte Carlo

Junsong Tang

July 9, 2025

1 Fundamentals of Hamiltonian Dynamics

1.1 Notations

- $\boldsymbol{q} = (q_1, \dots, q_n)$: generalized coordinates
- $\dot{\boldsymbol{q}} = (\dot{q_1}, \dots, \dot{q_n})$: generalized velocity
- $\mathcal{L}(t, q, \dot{q}) = T - V = \frac{1}{2} M \|\dot{q}\|^2 - V(q)$: Lagrangian, where M is the mass matrix, T is kinetic energy and V is potential energy
- $\boldsymbol{p}$: generalized momentum, i.e., $\boldsymbol{p} = M \cdot \dot{\boldsymbol{q}}$

1.2 Euler–Lagrange Equation

Generalized version: given a funtion $\boldsymbol{y}(x)$ and its derivative: $\boldsymbol{y'}(x)$ and $f(x, \boldsymbol{y}, \boldsymbol{y'})$. Define the functional:

$$F[\boldsymbol{y}] = \int_a^b f(x, \boldsymbol{y}, \boldsymbol{y'}) dx \tag{1}$$

As the topic in calculus of variation, we want the functional $F[\boldsymbol{y}]$ to obtain its local min for some function $\boldsymbol{y}$. Perturbate $\boldsymbol{y}(x)$ with $\boldsymbol{y}(x) + \varepsilon \boldsymbol{u}(x)$ for any function $\boldsymbol{u}$ and $\varepsilon \in \mathbb{R}$ with small enough $|\varepsilon|$. and put

$$g(\varepsilon) = \int_a^b f(x, \boldsymbol{y} + \boldsymbol{\varepsilon u}, (\boldsymbol{y} + \boldsymbol{\varepsilon u})') dx$$

To make $F[\boldsymbol{y}]$ obtaining its local minimum for some $\boldsymbol{y}$ satisfying the initial conditions: $\boldsymbol{y}(a)$ and $\boldsymbol{y}(b)$ being fixed, i.e., $\boldsymbol{u}(a) = \boldsymbol{u}(b) = 0$, it is equivalent to require $g'(0) = 0$, hence:

$$\begin{aligned}
g'(0) &= \frac{d}{d\varepsilon}\Big(\int_a^b f(x, \boldsymbol{y} + \boldsymbol{\varepsilon u}, (\boldsymbol{y} + \boldsymbol{\varepsilon u})') dx\Big) \\
&= \int_a^b \Big(\frac{\partial f}{\partial \boldsymbol{y}} \boldsymbol{u} + \frac{\partial f}{\partial \boldsymbol{y'}} \boldsymbol{u'}\Big) dx \quad \text{By total derivative} \\
&= \int_a^b \boldsymbol{u} \frac{\partial f}{\partial \boldsymbol{y}} dx + \frac{\partial f}{\partial \boldsymbol{y'}} \boldsymbol{u}|_a^b - \int_a^b \boldsymbol{u} \frac{d}{dx}\big(\frac{\partial f}{\partial \boldsymbol{y'}}\big) dx \quad \text{integration by parts} \\
&= \int_a^b \boldsymbol{u}\Big(\frac{\partial f}{\partial \boldsymbol{y}} - \frac{d}{dx}\frac{\partial f}{\partial \boldsymbol{y'}}\Big) dx = 0 \quad \text{By initial condition: } \boldsymbol{u}(a) = \boldsymbol{u}(b) = 0
\end{aligned}$$

By the fundamental lemma of calculus of variation, since $\boldsymbol{u}(x)$ is arbitrary, hence we must have:

$$\frac{\partial f}{\partial \boldsymbol{y}} - \frac{d}{dx}\frac{\partial f}{\partial \boldsymbol{y'}} = 0 \tag{2}$$

Equation (2) is called the **Euler–Lagrangian Equation** in calculus of variation, and for $\boldsymbol{y}$ satisfying equation (2) is a sufficient condition for $F[\boldsymbol{y}]$ to have its local minimum.

In a dynamical system, with Lagrangian $\mathcal{L}(t, \boldsymbol{q}, \dot{\boldsymbol{q}})$, we replace f with $\mathcal{L}$, $\boldsymbol{y}$ with generalized coordiantes $\boldsymbol{q}$, and x with time t, we get the **Lagrangian Equation** for dynamical system:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} = 0 \tag{3}$$

1.3 Legendre Transform

If $f : \mathcal{X} \to \mathbb{R}$ is convex, then $\forall t \in X$, then $f(t) \geq pt + b$, where $p = f'(x)$ and $b = f(x) - px$. This implies that:

$$pt - f(t) \leq px - f(x), \forall t \in \mathcal{X}$$

Then the one-dimensional Legendre transform is defined as:

$$f^*(p) = px - f(x) = \sup\{pt - f(t) : t \in \mathcal{X}\}$$

If $\mathcal{X} \subset \mathbb{R}^n$, and $\boldsymbol{p} = \nabla f = \frac{\partial f}{\partial \boldsymbol{x}}$, then the generalized Legendre tranform is defined as:

$$f^*(\boldsymbol{p}) = \boldsymbol{p} \cdot \boldsymbol{x} - f(\boldsymbol{x}) = \sup\{\boldsymbol{p} \cdot \boldsymbol{t} - f(\boldsymbol{t}) : \boldsymbol{t} \in \mathcal{X}\}$$

where $\boldsymbol{p} \cdot \boldsymbol{x} = \boldsymbol{p}^\top \boldsymbol{x}$ representing the inner product.

Note that f^* is also a convex function, since given a $t \in \mathcal{X}$, the map $p \mapsto pt - f(t)$ is linear, hence the supremum of the family of linear maps: $\{p \mapsto pt - f(t)\}_{t \in \mathcal{X}}$ gives an envolope, which is f^*. On the other hand, one can show:

$$f(x) = \sup\{px - f^*(p) : p \in \mathcal{X}^*\}$$

which means that $f(x)$ is the envolope of the family of linear maps: $\{x \mapsto px - f^*(p)\}_{p \in \mathcal{X}^*}$.

1.4 Hamiltonian Equation

Hamiltonian is the Legendre transform of Lagrangian, hence we have:

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p} \cdot \dot{\boldsymbol{q}} - \mathcal{L}(t, \boldsymbol{q}, \dot{\boldsymbol{q}}) \tag{4}$$

Since $\boldsymbol{p} \cdot \dot{\boldsymbol{q}} = 2T$, so $\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = 2T - (T - V) = T + V$ and hence Hamiltonian can be interpreted as the total energy of a dynamical system.

We can derive Hamiltonian equations from Lagrange eqaution. Note that:

$$\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} = M \cdot \dot{\boldsymbol{q}} = \boldsymbol{p} \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} = \dot{\boldsymbol{p}}$$

so with (4), we must have:

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}} \tag{5}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} = -\frac{\partial \mathcal{L}}{\partial q} = -\dot{\boldsymbol{p}} \tag{6}$$

(5) and (6) are called **Hamiltonian equations**

Or equivalently, we can take the total differentiation on both sides of (4):

$$\begin{aligned}
d\mathcal{H} &= d(\boldsymbol{p} \cdot \dot{\boldsymbol{q}}) - d\mathcal{L}(t, \boldsymbol{q}, \dot{\boldsymbol{q}}) \\
&= \dot{\boldsymbol{q}} \cdot d\boldsymbol{p} + \boldsymbol{p} \cdot d\dot{\boldsymbol{q}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} d\boldsymbol{q} - \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} d\dot{\boldsymbol{q}} - \frac{\partial \mathcal{L}}{\partial t} dt \\
&= \dot{\boldsymbol{q}} \cdot d\boldsymbol{p} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} d\boldsymbol{q} - \frac{\partial \mathcal{L}}{\partial t} dt \\
&= \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} d\boldsymbol{p} + \frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} d\boldsymbol{q} + \frac{\partial \mathcal{H}}{\partial t} dt
\end{aligned}$$

Hence we can correspond the coefficients to get (5) and (6).

2 Hamiltonian Monte Carlo