

Chap. 4.2: Hamiltonian Monte Carlo

LIU, Ran

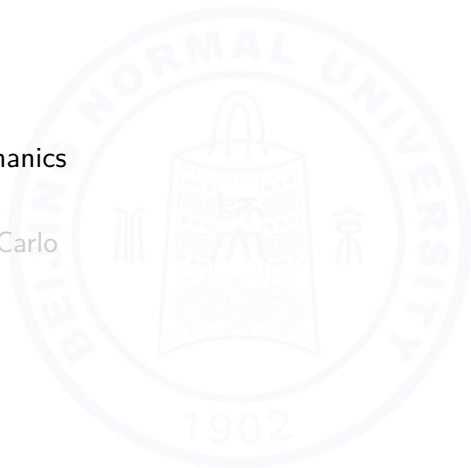
April 4, 2025

LIU, Ran - Department of Statistics @BNU

Summary

Hamiltonian mechanics

Hamilton Monte Carlo



LIU, Ran - Department of Statistics @BNU

势能与力

常见的力比如重力、万有引力等。力做功会导致能量的转化，而势能 (Potential Energy) 描述的是物体由于其位置或状态而具有的能量。

对于保守力 (做功只与初末状态有关，而与路径无关)，势能 $U(x)$ 与力 $F(x)$ 之间存在以下关系：

$$F = -\frac{dU}{dx}$$

这意味着力可以通过势能的负梯度 (即对势能的空间导数取负) 来计算。可以把势能理解成要克服这个力做的功。

Liu, Ran - Department of Statistics @BNU

具体来说：

- 重力势能：在近地表区域，重力势能由物体的高度 z 决定：

$$U(z) = mgz$$

其对应的重力为：

$$F = -\frac{dU}{dz} = -mg$$

这表明重力方向向下，与势能的变化趋势相反。

- 万有引力势能：对于质量 m_1 和 m_2 之间的万有引力势能：

$$U(r) = -G \frac{m_1 m_2}{r}$$

由此计算得到的引力大小为：

$$F = -\frac{dU}{dr} = -\left(-G \frac{m_1 m_2}{r^2}\right) = G \frac{m_1 m_2}{r^2}$$

这与万有引力公式一致，表明引力方向指向质量中心。

物体在重力作用下从高度 h 下降 z 到高度 $h - z$ ，其势能的变化为：

$$\Delta U = U_{\text{final}} - U_{\text{initial}} = mg(h - z) - mgh = -mgz$$

此时重力与位移同方向，我们可以发现做功为：

$$W = \int F \cdot dx = mgz = -\Delta U$$

在一维情况下，力与势能的关系可以写为：

$$F = -\frac{dU}{dx}$$

在三维情况下，力由势能的梯度（Gradient）决定：

$$\mathbf{F} = -\nabla U = \left(-\frac{\partial U}{\partial x}, -\frac{\partial U}{\partial y}, -\frac{\partial U}{\partial z} \right)$$

其中 ∇U 表示势能的梯度，是一个向量，指向势能增长最快的方向。由式可得，势能的零点可以根据实际情况自由选择，差一个常数情况下，而不影响力的计算结果。

哈密顿量的定义

在分析动力学系统时，我们可以引入哈密顿量 $H(x, p)$ ，它表示系统的总能量，由物体的位置和动量描述：

$$H(x, p) = \frac{p^2}{2m} + U(x).$$

其中：

- 动能 $\frac{p^2}{2m} = \frac{1}{2}mv^2$ ，描述物体因运动而具有的能量。
- 势能 $U(x)$ ，表示物体因位置而具有的能量，通常由外部力场（如引力或弹性力）决定。

哈密顿量在经典力学中扮演着重要角色，它不仅代表能量，还能用于推导系统的运动方程。

在推导正则方程之前，我们首先回顾牛顿第二定律，它描述了力与动量变化之间的关系：

$$F = ma = m \frac{d^2x}{dt^2} = \frac{dp}{dt},$$

其中，动量 p 定义为：

$$p = mv = m \frac{dx}{dt}.$$

动量是描述物体运动状态的重要物理量，它的变化率由外力决定。

LIU, Ran - Department of Statistics @BNU

哈密顿正则方程的推导

为了进一步理解哈密顿量的物理意义，我们观察它对 x 和 p 的偏导数：

$$\frac{\partial H}{\partial x} = \frac{\partial U}{\partial x} = -F = -\frac{dp}{dt},$$

$$\frac{\partial H}{\partial p} = \frac{p}{m} = \frac{dx}{dt}.$$

这两条公式可以整理为：

哈密顿正则方程

$$\frac{dp}{dt} = -\frac{\partial H}{\partial x}, \quad \frac{dx}{dt} = \frac{\partial H}{\partial p}.$$

这组方程提供了一种描述运动的等价方法，与拉格朗日力学不同，它直接基于动量和坐标，而非仅基于坐标和速度。

哈密顿量的守恒性

最后，我们验证哈密顿量是否是守恒量，即：

$$\frac{dH}{dt} = \frac{\partial H}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial H}{\partial p} \cdot \frac{dp}{dt}.$$

将哈密顿正则方程代入：

$$\frac{dH}{dt} = \left(\frac{\partial H}{\partial x} \right) \left(\frac{\partial H}{\partial p} \right) + \left(\frac{\partial H}{\partial p} \right) \left(-\frac{\partial H}{\partial x} \right)$$

由于两项互相抵消，我们得到：

$$\frac{dH}{dt} = 0.$$

当系统包含多个粒子时，势能不仅依赖于单个粒子的位置，还可能涉及粒子之间的相互作用。因此，我们将势能推广为依赖于所有粒子坐标的函数，表示为 $U(x_1, x_2, \dots, x_N)$ 。例如，两粒子万有引力势能

$$U(x_1, x_2) = -G \frac{m_1 m_2}{|x_1 - x_2|}$$

在多粒子系统中，哈密顿量可以推广为：

$$H = \sum_i \frac{p_i^2}{2m_i} + U(x_1, x_2, \dots, x_N)$$

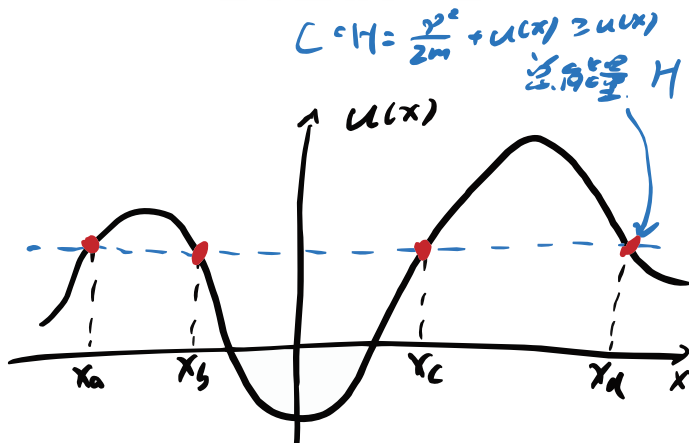
这时候 H 以每个粒子的动量和位置作为参数。

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial x_i} \quad \frac{dx_i}{dt} = \frac{\partial H}{\partial p_i}$$

$$\sum_i F_i dx_i = -dU \Leftrightarrow F_i = -\frac{\partial U}{\partial x_i}$$

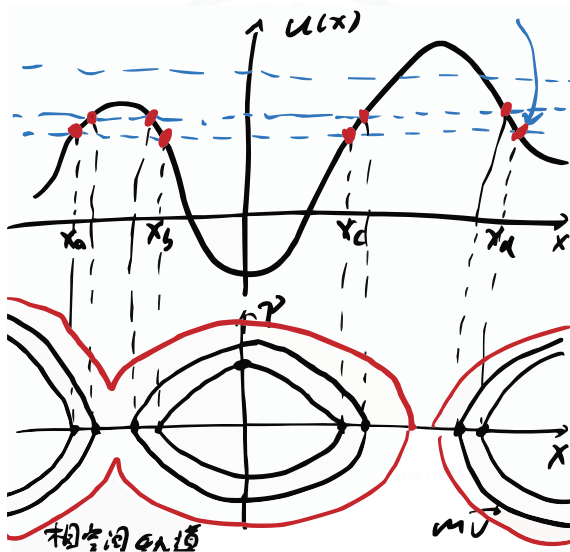
其中 dx_i 为粒子的虚位移矢量， F_i 是第 i 个粒子的受力。

看一个粒子的例子，在一个轨道能量守恒的运动：



Liou, Yan - Department of Statistics (UCSD)

在以 x 为横坐标、 p 为纵坐标的坐标系中，我们绘制哈密顿量 $H(x, p) = \frac{p^2}{2m} + U(x)$ 的等高线图，这对应于系统在相空间中的等能面。



总而言之，哈密顿动力学描述了动量与位置的关系，我们将利用正则方程，根据目标分布的梯度，产生 MH 算法合适的候选值，再判断接受与否。

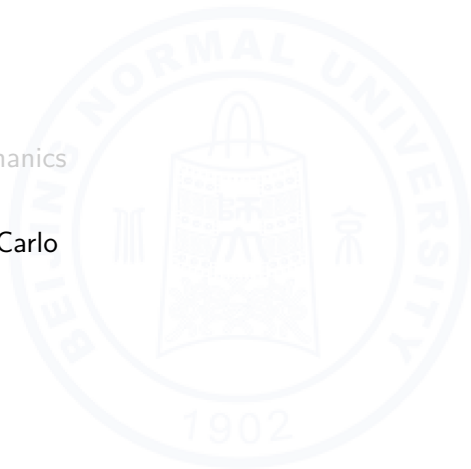
相关参考资料：经典力学新讲.

LIU, Ran - Department of Statistics @BNU

Summary

Hamiltonian mechanics

Hamilton Monte Carlo



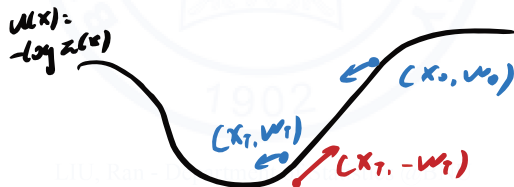
LIU, Ran - Department of Statistics @BNU

HMC 简述

利用梯度产生合理的候选值，

- 初始化：随机采样动量 W_0 (辅助变量，最终会舍弃)。
- 哈密顿动力学：演化 L 步，总共 T 时间，保持总能量守恒。
- 反转动量： $-W_T$ 保证可逆性：

$$(X_0, W_0) \rightarrow (X_T, W_T) \rightarrow (X_T, -W_T)$$



如何设置势能以及如何设置 W_0 的分布？

玻尔兹曼分布与能量函数建模的启发

- 玻尔兹曼分布 (或称 Gibbs 分布) 是统计物理中描述系统状态概率的模型:

$$p(x) = \frac{1}{Z} \exp\left(-\frac{E(x)}{kT}\right)$$

- 其中:
 - $E(x)$ 是状态 x 的能量;
 - T 是温度, k 是玻尔兹曼常数;
 - Z 是归一化常数 (配分函数)。
- 能量越低的状态, 出现概率越高。
- 启发: 可以用一个能量函数 $E(x)$ 对状态打分, 构造出概率分布。
- 这种思想影响了机器学习中的能量建模方法, 发展为能量基模型 (energy-based model, EBM)。

能量基模型

能量基模型通过能量函数定义概率密度：

$$p(x) = \frac{1}{Z} \exp(-E(x))$$

我们可以不直接建模 $p(x)$ ，而是定义能量函数 $E(x)$ ，使得：

- 真样本能量低；假样本能量高；
- 可用于判别或采样等任务。

LIU, Ran - Department of Statistics @BNU

HMC 中能量与分布的转换

- 为 $x \sim \pi(x)$ 构造能量函数，定义：

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

- 为动量 w 构造分布：

$$V(w) = \frac{1}{2}w^\top M^{-1}w \Rightarrow p(w) \propto \exp(-\frac{1}{2}w^\top M^{-1}w)$$

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

- 联合能量函数（哈密顿量）为：

$$H(x, w) = U(x) + V(w) = -\log \pi(x) + \frac{1}{2}w^\top M^{-1}w$$

- 检查一下此时的联合分布为：

$$p(x, w) \propto \exp(-H(x, w)) = \exp(-U(x)) \cdot \exp(-V(w)) \propto \pi(x)p(w)$$

HMC 的步骤为:

- 采样动量: $w_0 \sim p(w) \propto e^{-V(w)} \propto N(0, M^{-1})$.
- 势能设定为负的对数目标分布: $U(x) = -\log \pi(x)$, 则此时我们的哈密顿常量为:

$$H(x, w) = U(x) + V(w) = -\log \pi(x) + \frac{1}{2} w^\top M^{-1} w$$

根据动力学, 满足微分方程。往前走 T 时间

$$(x_0, w_0) \rightarrow (x_T, w_T) \xrightarrow{\text{为了可逆}} (x_T, -w_T)$$

不然 $P(x' \rightarrow x) = 0$.

- 接受率为 (后面会证明):

$$\alpha(x_0, x_T) = \min \left(1, \frac{\exp(-H(x_T, w_T))}{\exp(-H(x_0, w_0))} \right) = 1$$

具体动力学实现

$$\frac{dx_j(t)}{dt} = \frac{\partial H}{\partial w_j}, \quad \frac{dw_j}{dt} = -\frac{\partial H}{\partial x_j}$$

此时 $H(x, w)$ 及其偏导数均是已知的，我们通过欧拉公式或者 leap frog 往前走。

LIU, Ran - Department of Statistics @BNU

为什么是 $U(x) = -\log \pi(x)$ 以及 HMC 为什么是生成了一个好的 proposal ?

- 类似能量基模型的定义。
- 动量变化方向是 $\pi(x)$ 的梯度方向

$$\frac{dW_j}{dt} = -\frac{\partial H}{\partial x_j} = \frac{\partial \log \pi(x)}{\partial x_j}$$

类似 $F = ma_j = \frac{dw_j}{dt}$ 。动力学有让系统有往势能低的地方的趋势，而我们的目的是找到 $\log \pi(x)$ 大的地方，所以符号相反。

- 计算接受率的时候，如果要凑成哈密顿常量，需要 $U(x) = -\log \pi(x)$ 。（见后面详细的计算过程）

Liu, Ran - Department of Statistics @BNU

计算接受率

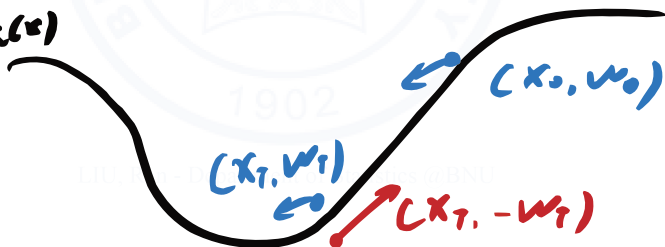
为了保证细致平衡，我们需要

$$(x_0, w_0) \rightarrow (x_T, -w_T) \rightarrow (x_0, w_0)$$

$$x_0 \xrightarrow{\text{采样 } w_0} (x_0, w_0) \xrightarrow{\text{动}} (x_T, w_T) \xrightarrow{\text{负}} (x_T, -w_T) \xrightarrow{\text{舍弃 } -w_T} x_T$$

$$x_T \xrightarrow{\text{采样 } -w_T} (x_T, -w_T) \xrightarrow{\text{动}} (x_0, -w_0) \xrightarrow{\text{负}} (x_0, w_0) \xrightarrow{\text{舍弃 } -w_0} x_0$$

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



计算 $\alpha(x_0, x_T)$ 只需考虑随机的部分，因为决定性部分的概率为 1（动力学能抵达是 1，不能抵达是 0）。根据细致平衡条件，

$$\pi(x)q(x' | x)\alpha(x, x') = \pi(x')q(x | x')\alpha(x', x)$$

我们取接受率为

$$\alpha(x, x') = \min \left(1, \frac{\pi(x')q(x | x')}{\pi(x)q(x' | x)} \right)$$

不失一般性，我们假设 $M = I$, $w_0 \sim N(0, M^{-1})$, 则有

$$\begin{aligned} \frac{\pi(x')q(x | x')}{\pi(x)q(x' | x)} &= \frac{\frac{\pi(x_T)}{\sqrt{2\pi}} \exp\left(-\frac{w_T^2}{2}\right) \cdot 1}{\frac{\pi(x_0)}{\sqrt{2\pi}} \exp\left(-\frac{w_0^2}{2}\right) \cdot 1} = \frac{\pi(x_T) \exp\left(-\frac{w_T^2}{2}\right)}{\pi(x_0) \exp\left(-\frac{w_0^2}{2}\right)} \\ &= \frac{\exp\left(-(-\log \pi(x_T) + \frac{1}{2}w_T^2)\right)}{\exp\left(-(-\log \pi(x_0) + \frac{1}{2}w_0^2)\right)} = \frac{\exp(-H(x_T, w_T))}{\exp(-H(x_0, w_0))} \end{aligned}$$

所以 $\alpha = 1$ 。

Liu, Ran - Department of Statistics @BNU

我们用 Leapfrog 实现 HMC 中动力学的离散化步骤方法。假设 x_0 是输入位置，我们只能计算 $r(x) \propto \pi(x)$ 。另外，设 ϵ 为离散化的步长， L 为动力学更新次数，即 $\epsilon \cdot L = T$ 为动力学应用的总时间。

① 生成初始动量 $\omega_0 \sim N(0, M^{-1})$ 。(不能为 0)

② 设 $x^{(0)} = x_0$ 。

③ 对动量进行半步更新：

$$\omega^{(0)} = \omega_0 - \frac{\epsilon}{2} \nabla \log r(x^{(0)})$$

④ 对于 $\ell = 1, \dots, L-1$ ，执行以下步骤：

① 位置更新：

$$x^{(\ell)} = x^{(\ell-1)} + \epsilon \cdot M^{-1} \omega^{(\ell-1)}$$

② 动量更新：

$$\omega^{(\ell)} = \omega^{(\ell-1)} - \epsilon \cdot \nabla \log r(x^{(\ell)})$$

⑤ 最后进行一次位置更新：

$$x^{(L)} = x^{(L-1)} + \epsilon \cdot M^{-1} \omega^{(L-1)}$$

⑤ 进行另一次动量的半步更新：

$$\omega^{(L)} = \omega^{(L-1)} - \frac{\epsilon}{2} \nabla \log r(x^{(L)})$$

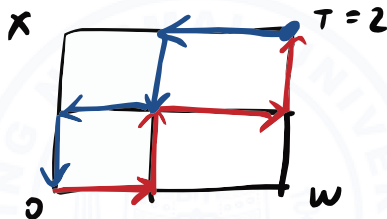
⑤ 计算接受概率：

$$\alpha(x_0, \omega_0, x^{(L)}, \omega^{(L)}) = \min \left\{ 1, \frac{\exp(-H(x^{(L)}, \omega^{(L)}))}{\exp(-H(x_0, \omega_0))} \right\}$$

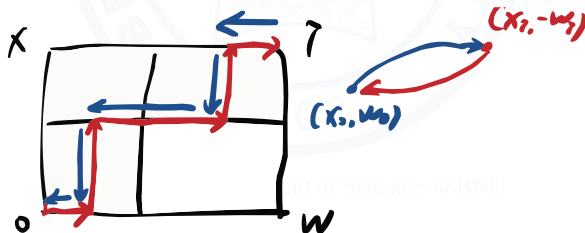
⑤ 以概率 $\alpha(x_0, \omega_0, x^{(L)}, \omega^{(L)})$ 接受 $x_{\text{new}} = x^{(L)}$ 。如果拒绝，则 $x_{\text{new}} = x_0$ 。

⑤ 返回 x_{new} 。

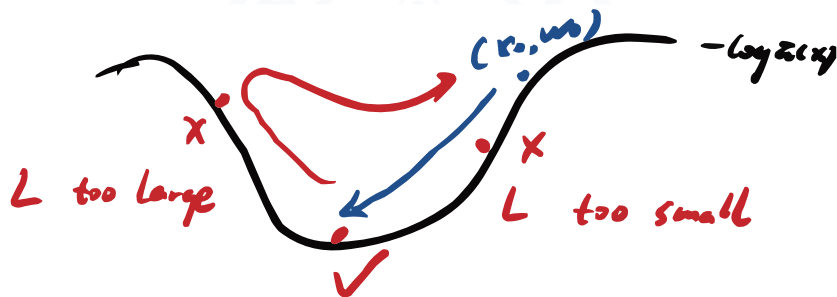
若不 half-step update, 路径不一样, 可能回不到起始状态:



half-step update, 保证动量更新开始, 动量更新结束 (对称性):



No U-turn Sampler(NUTS)



LIU, Ran - Department of Statistics @BNU