Stochastic Simulation

Autumn Semester 2024

Prof. Fabio Nobile *Assistant: Matteo Raviola*

Project - 2

Submission deadline: 16 January 2025

Hamiltonian Monte Carlo

Introduction and background

A common shortfall of the simple random walk Metropolis proposal in the Metropolis-Hastings algorithm is the slow exploration rate of the state space. Much effort has been devoted in recent years to devise proposals with more efficient exploration rates (i.e "distant" proposals). One idea borrowed from physics, which, can be applied to most problems with continuous state space, is to introduce a "fictitious" Hamiltonian dynamics and "fictitious" momentum variables.

We recall that a Hamiltonian dynamical system is characterized by a Hamiltonian function $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $H = H(q, p)$, that is conserved during dynamics. Here $q = (q_1, \dots, q_d)$ denotes the position vector and $p = (p_1, \dots, p_d)$ denotes the momentum vector. The Hamiltonian dynamics is governed by the equations

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \tag{1}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \tag{2}$$

for $i = 1, \dots, d$. In general, the above equation can be understood as a conservation of the total energy of a system in time.

Hamiltonian Monte Carlo takes inspiration from the previous physical system in order to construct a Markov Chain Monte Carlo algorithm with a given invariant density $\pi(q)$ on the position variables $(q_1, \dots, q_d)$. To do so, we introduce the *potential energy* $U(q) = -\log \pi(q)$, a *kinetic energy* $K(p) = \sum_{i=1}^{d} \frac{p_i^2}{2m_i}$, for some mass parameters m_i, $i = 1, \dots, d$, and the Hamiltonian $H(q, p) = U(q) + K(p)$. Having introduced these functions, we can then simulate a Markov chain in which each iteration re-samples the momentum, evolves the Hamiltonian system for a certain time, and then performs a Metropolis-type acceptance-rejection step on the new position vector. More concretely, we consider the so-called Gibbs distribution, given by

$$G(q, p) = \frac{1}{Z} \exp(-H(p, q)),$$

where Z is the (unknown) normalizing constant. Notice that such a Gibbs distribution nat-

urally factorizes as:

$$G(q,p) = \frac{1}{\tilde{Z}} \exp(-U(q)) \frac{1}{\prod_{i=1}^{d} \sqrt{2\pi m_i}} \exp(-K(p)),$$

where $\frac{1}{\tilde{Z}} \exp(-U(q))$ is the probability density we are interested in sampling from, whereas $\frac{\exp(-K(p))}{\prod_{i=1}^{d} \sqrt{2\pi m_i}}$ is a multivariate Gaussian distribution $N(0, M)$ with $M = \text{diag}(m_1, ..., m_d)$. Given the state q^n at iteration n, the idea of the algorithm is then to sample a momentum vector p^n from $N(0, M)$, and compute $H(q^n, p^n)$. The Hamiltonian system is then evolved starting from $q(0) = q^n, p(0) = p^n$, on a time interval $[0, T]$ using equations (1), and (2) for some arbitrary final time T, to obtain $(q(T), p(T))$, where, in general, $q(T) \neq q(0)$. This state is then taken as the proposal state in a Metropolis-Hastings step to generate the new state q^{n+1}. For many problems of modern relevance, it is not possible to compute the dynamics exactly and numerical discretization is needed. A convenient time discretization scheme is the *Verlet's method*: the time interval $[0, T]$ is divided into N_t intervals of size $\epsilon > 0$ and for each particle i the position q_i and momemtum p_i are updated as follows

$$\begin{aligned} p_i(t + \epsilon/2) &= p_i(t) - (\epsilon/2) \frac{\partial U(q(t))}{\partial q_i} \\ q_i(t + \epsilon) &= q_i(t) + \epsilon \frac{p_i(t + \epsilon/2)}{m_i} \\ p_i(t + \epsilon) &= p_i(t + \epsilon/2) - (\epsilon/2) \frac{\partial U(q(t + \epsilon))}{\partial q_i}. \end{aligned} \tag{3}$$

The main steps of the Hamiltonian Monte Carlo algorithm using Verlet's method are outlined in Algorithm 1. There, N is the length of the chain, ϵ the time step in Verlet's method, and T the final integration time.

Algorithm 1 Hamiltonian Monte Carlo

1: **For** $n = 1, \ldots, N$, **do**:

2: Sample new values for the momentum variables,

$$p_i^n \sim \mathcal{N}(0, m_i), \quad i = 1, \ldots, d$$

3: Given the current state (q^n, p^n), propose a new state (q^*, p^*) by evolving the Hamiltonian system (1), (2) using Verlet's method.

4: Set $q^{n+1} = q^*$ with probability α, where

$$\alpha = \min\left[1, \exp\left(-U(q^*) + U(q^n) - K(p^*) + K(p^n)\right)\right]$$

otherwise, set $q^{n+1} = q^n$.

5: **end For**

Notice that, similar to the random-walk Metropolis, this algorithm depends on few parameters, namely, ϵ, T , and m_i, which should be properly tuned.

Goals of the project

(a) Show that the Hamiltonian is preserved by the dynamics (1)-(2), i.e, $\frac{dH}{dt} = 0$ along a trajectory.

(b) Suppose you could evolve the Hamiltonian system exactly without a need for a numerical discretization. How does the acceptance rate in Algorithm 1 look like in this case? Now, suppose we can't evolve the Hamiltonian exactly, and we have to use Verlet's method to numerically approximate the time integration. Does your answer to the previous question change?

(c) Suppose the time integration is being carried-out exactly, i.e., there is no numerical error. Prove that the HMC algorithm leaves the Gibbs distribution invariant. Conclude that if $q^0 \sim \pi(q)$, then, $q^k \sim \pi(q)$ for all k. **Hint:** See Theorems 2.3 and 5.1 in [1]. Can you show the same when the Verlet's method is used to discretize the dynamics?

(d) Consider the problem of sampling from a distribution π with unnormalized density

$$f_1(q_1, q_2) = e^{-\alpha(q_1^2+q_2^2-\frac{1}{4})^2}.$$

Test both cases $\alpha = 10$ and $\alpha = 10^3$. Implement a random walk Metropolis algorithm to sample from said distribution. Implement a HMC sampling scheme using Verlet's method for different values of ϵ, m_1, m_2, T. Discuss your results. Moreover, compare your results in terms of autocorrelation of the chains obtained with each method and in terms of effective sample size vs number of evaluations of f and $\partial f/\partial q_i, i \in \{1, 2\}$.

(e) We consider a logistic regression model $\mathbb{P}(y_i = 1|x_i^T q) = \left[1 + \exp(-x_i^T q)\right]^{-1}$, where y_i is the binary response for the ith subject $i = 1, \ldots, n$ and $y = (y_1, \ldots, y_n)$ is the vector of responses for all subjects. The co-variate values for the ith subject are $x_i^T = (x_{i,0}, \ldots, x_{i,p})$. The full design matrix is written as $X = (x_1^T; \ldots; x_n^T) \in \mathbb{R}^{n\times(p+1)}$. The regression coefficients for the $p+1$ covariates are given by a vector $q = (\beta_0, \ldots, \beta_p)^T$. The un-normalized log-posterior for this model, with Gaussian prior $N(0, \text{diag}(\sigma_0^2, \ldots, \sigma_p^2))$ on q can be written as

$$\log f(q|y, X, \sigma_q^2) = q^T X^T (y - \mathbf{1}_n) - \mathbf{1}_n^T \left[\log(1 + \exp(-x_i^T q))\right]_{n\times 1} - \frac{1}{2}\sum_{i=0}^{p} \frac{q_i^2}{\sigma_i^2}, \qquad (4)$$

where $\mathbf{1}_n$ indicates an n-dimensional vector of ones, $\left[\log(1 + \exp(-x_i^T q))\right]_{n\times 1}$ indicates an $n \times 1$ vector $\forall i = 1, \ldots, n$, and I_{p+1} indicates the identity matrix in $\mathbb{R}^{p+1}$. We analyze data of 189 births at a U.S. hospital to examine the risk factors of low birth weight ($y = 1$ if low weight, $y = 0$ otherwise). The data is available on the course website and is taken from [4] (Table 1.6). We construct a model by setting $x_{i,0} = 1$, and letting the remaining covariates x_i to indicate mother's age in years, $x_{i,1}$, mother's weight in pounds at last menstrual period $x_{i,2}$, race $x_{i,3}$ (African-American) $x_{i,4}$ (other race), smoking during pregnancy, $x_{i,5}$, premature birth, $x_{i,6}$, hypertension, $x_{i,7}$, presence of uterine irritability, $x_{i,8}$, one physician visit during the first trimester, $x_{i,9}$, and two or more physician visits during the first trimester, $x_{i,10}$. Implement an HMC sampler to perform the logistic regression on q. Choose suitable values for the prior covariance. Monitor the convergence of your algorithm. Present histograms of your coefficients.

(f) Consider now the task of computing the posterior expectation of q in the logistic regression problem of the previous point. Compare the performance of HMC with that of another method of your choice.

References

[1] Bou-Rabee, Nawaf, and Jesús María Sanz-Serna. "Geometric integrators and the Hamiltonian Monte Carlo method." *Acta Numerica 27* (2018): 113-206.

[2] Duane, Simon, et al. "Hybrid Monte Carlo." *Physics letters B 195.2* (1987): 216-222.

[3] Thomas, Samuel, and Wanzhu Tu. "Learning Hamiltonian Monte Carlo in R." *arXiv preprint arXiv:2006.16194 (2020).*

[4] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X.. "Applied logistic regression". John Wiley & Sons (2013).