Inference and Representation, Fall 2016

Problem Set 6: MCMC.

Due: Monday, November 21, 2016 at 11:59pm (as a PDF document uploaded in Gradescope.)

Important: *See problem set policy on the course web site.*

Hamiltonian Monte-Carlo

This problem will explore Hamiltonian Dynamics as a tool to enhance classic MCMC methods, in the so-called *Hamiltonian Monte-Carlo (HMC)*.

The classical (non-relativistic) Lagrangian Mechanics describing the dynamics of N particles in $\mathbb{R}^3$ with positions $x_1, \ldots x_N$ and velocities $v_1, \ldots v_N$, $v_i = \dot{x}_i$ is given, in absence of external forces, by

$$\mathcal{L}(\mathbf{v}, \mathbf{x}) = \frac{1}{2}\langle \mathbf{v}, M\mathbf{v}\rangle - U(\mathbf{x}) \ , \ \mathbf{v} = (v_1, \ldots, v_N) \in \mathbb{R}^{3N}, \mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^{3N} \ , \tag{1}$$

where M is a $3N \times 3N$ diagonal, positive matrix of the form $M = \mathrm{diag}(m_1, m_1, m_1, \ldots, m_N, m_N, m_N)$ describing the masses of the particles, and $U(\mathbf{x})$ is a potential energy term, that only depends upon position variables $\mathbf{x}$.

1. Show that, for each fixed $\mathbf{x}$, $\mathcal{L}(\mathbf{v}, \mathbf{x})$ is convex with respect to $\mathbf{v}$.

By using the fact that Lagrangian solutions are stationary points and the fact that $\dot{\mathbf{x}} = \mathbf{v}$, we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = \frac{\partial \mathcal{L}}{\partial \mathbf{x}} \ . \tag{2}$$

This is called the *Euler-Lagrange* equation.

The first step to understand HMC is to derive Hamiltonian mechanics from Lagrangian mechanics. For that purpose, we recall the notion of convex or Legendre-Fenchel conjugate: given a convex function $f : \Omega \to \mathbb{R}$ defined on a convex set Ω, it is defined as

$$f^*(p) = \sup_{y \in \Omega} \left(\langle y, p\rangle - f(y) \right) \ . \tag{3}$$

2. Show that f^* is convex. *Hint: think about what happens if you take the maximum of two convex functions.*

3. Using the fact that $\mathcal{L}$ is differentiable, show that the Legendre-Fenchel conjugate of $\mathcal{L}(\mathbf{v}, \mathbf{x})$, for fixed $\mathbf{x}$, has the form

$$\mathcal{H}(\mathbf{p}, \mathbf{x}) = \frac{1}{2}\langle \mathbf{p}, M^{-1}\mathbf{p}\rangle + U(\mathbf{x}) \ , \text{ with } \mathbf{p} = \frac{\partial \mathcal{L}}{\partial \mathbf{v}} \ . \tag{4}$$

   This is the *Hamiltonian*, and is interpreted as the energy of the system in terms of position $\mathbf{x}$ and momentum $\mathbf{p}$.

The Legendre duality also gives the crucial interpretation of momentum variables $\mathbf{p}$ as the partial derivatives of $\mathcal{L}$ with respect to velocity, and

$$\mathcal{H}(\mathbf{p}, \mathbf{x}) = \langle \mathbf{v}, \mathbf{p} \rangle - \mathcal{L}(\mathbf{v}, \mathbf{x}) \ . \tag{5}$$

4. Using the previous results, take the differential of $\mathcal{H}(\mathbf{p}, \mathbf{x})$ with respect to time

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \langle \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \dot{\mathbf{p}} \rangle + \langle \frac{\partial \mathcal{H}}{\partial \mathbf{x}}, \dot{\mathbf{x}} \rangle$$

and use the Euler-Lagrange equation to derive the *Hamiltonian equations*:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{p}} = \dot{\mathbf{x}} = \mathbf{v} \ , \ \frac{\partial \mathcal{H}}{\partial \mathbf{x}} = -\dot{\mathbf{p}} \ . \tag{6}$$

Now that we have derived the Hamiltonian Dynamics, we need an algorithm to implement them in a computer. A popular strategy is the so-called *Leap Frog* method. Given a stepsize $\delta > 0$, it consists in the following steps:

- Take a half-step to update the momentum variable:

$$\mathbf{p}(t + \delta/2) = \mathbf{p}(t) - \frac{\delta}{2} \nabla_{\mathbf{x}} \mathcal{H}(\mathbf{p}(t), \mathbf{x}(t)) \ .$$

- Take a full-step to update the position variable:

$$\mathbf{x}(t + \delta) = \mathbf{x}(t) + \delta \nabla_{\mathbf{p}} \mathcal{H}(\mathbf{p}(t + \delta/2), \mathbf{x}(t)) \ .$$

- Take the remaining half-step to update momentum:

$$\mathbf{p}(t + \delta) = \mathbf{p}(t + \delta/2) - \frac{\delta}{2} \nabla_{\mathbf{x}} \mathcal{H}(\mathbf{p}(t + \delta/2), \mathbf{x}(t + \delta)) \ .$$

The goal of Hamiltonian Monte-Carlo is to use Hamiltonian Dynamics to approximate expectations on a given model of the form

$$p(\mathbf{x}) = \frac{\tilde{p}(\mathbf{x})}{Z} \ , \tag{7}$$

where Z is the partition function. By denoting $U(\mathbf{x}) = -\log p(\mathbf{x})$, this is interpreted as the *canonical* distribution of the system with energy $U(\mathbf{x})$.

5. If $U(\mathbf{x})$ in (7) denotes the potential energy of the system, derive the canonical distribution in terms of $\mathbf{x}$ and $\mathbf{p}$ using the Hamiltonian, and conclude that the joint canonical distribution $p(\mathbf{x}, \mathbf{p})$ is separable in $\mathbf{x}$ and $\mathbf{p}$, that is, $\mathbf{x}$ and $\mathbf{p}$ are independent: $p(\mathbf{x}, \mathbf{p}) = p(\mathbf{x})p(\mathbf{p})$. Explain how this property justifies using $\mathbf{p}$ as auxiliary variables to sample from $p(\mathbf{x})$.
6. Using (4), show that the marginal $p(\mathbf{p})$ is a Normal distribution with zero mean.
7. Show that the leap-frog algorithm does not require knowledge of Z.
8. Finally, let us consider the following proposal distribution. Given $\mathbf{x}^{(0)}$, we draw $\mathbf{p}^{(0)} \sim p(\mathbf{p})$ and run 1 step of the Leapfrog algorithm with step δ, to obtain $(\mathbf{x}^*, \mathbf{p}^*)$. Denote by

$$q(\mathbf{x}^*, \mathbf{p}^* \mid \mathbf{x}^{(0)}, \mathbf{p}^{(0)}) \tag{8}$$

the resulting distribution. Show that the Metropolis-Hastings algorithm using this proposal distribution produces samples from $p(\mathbf{x})$ when marginalizing over position variables.