Inference and Representation, Fall 2017

Problem Set 5: Hamiltonian Monte-Carlo

Zhuoru Lin
zlin@nyu.edu

Disclaimer: I adhered to NYU honor code in this assignment.

1. To show $\mathcal{L}(\mathbf{v}, \mathbf{x})$ is convex given $\mathbf{v}$, we show the second derivative of $\mathcal{L}$ with respect to $\mathbf{v}$ is larger than zero for all $\mathbf{v}$. In fact:

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{v}^2} = M. \tag{1}$$

By definition M is a positive matrix. Therefore $\mathcal{L}$ is convex with respect to $\mathbf{v}$.

2. We use the fact that the pointwise suprenum of a set of convex functions is also convex. Consider any $y \in \Omega$, we must have $< y, p > +f(y)$ is convex with respect to p, therefore the convex conjugate:

$$f^*(p) = \sup_y (< y, p > +f(x)) \tag{2}$$

is convex.

3. By the definition of $\mathcal{L}$ and $\mathbf{p}$ we have:

$$\mathbf{p} = \frac{\partial \mathcal{L}}{\partial \mathbf{v}} = M\mathbf{v}. \tag{3}$$

The convex conjugate of $\mathcal{L}(\mathbf{v})$ is:

$$\begin{aligned}\mathcal{L}^*(v) &= \sup_v [< v, p > -\mathcal{L}(\mathbf{v})] \\ &= \sup_v [< v, Mv > -\frac{1}{2} < v, mv > +U(\mathbf{x})] \\ &= \sup_v [\frac{1}{2} < v, mv > +U(\mathbf{x})]\end{aligned}$$

There fore the Hamilton $\mathcal{H}$ must have the form:

$$\mathcal{H}(\mathbf{p}, \mathbf{x}) = \frac{1}{2} < \mathbf{p}, M^{-1}\mathbf{p} > +U(\mathbf{x}) \tag{4}$$

4. By the definition of $\mathcal{H}$ (equation 4), we have:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{p}} = M^{-1}\mathbf{p} = \mathbf{v}. \tag{5}$$

By the definition of $\mathcal{H}$ and $\mathcal{L}$ we also have:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = -\frac{\partial U}{\partial \mathbf{x}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \tag{6}$$

By Euler-Lagrange equation and the definition of $\mathbf{p}$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{x}} = -\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = -\frac{d\mathbf{p}}{dt}. \tag{7}$$

Equation 5 and 7 are *Hamiltonian Equations.*

5. Without loss of generality, let the temperature $T = 1$. The canonical distribution of hamiltonian $\mathcal{H}(\mathbf{x}, \mathbf{p})$ is:

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{p}) &= \frac{1}{Z} \exp[-\mathcal{H}(\mathbf{x}, \mathbf{p})] && (8) \\
&= \frac{1}{Z} \exp[-\mathbf{p}^T M^{-1} \mathbf{p} - U(\mathbf{x})] \\
&= \frac{1}{Z} \exp[-\mathbf{p}^T M^{-1} \mathbf{p}] \exp[-U(\mathbf{x})], && (9)
\end{aligned}
$$

where Z is the partition function. Observed that $p(\mathbf{x}, \mathbf{p})$ is seperable function, therefore one can factor $p(\mathbf{x}, \mathbf{p}) = p(\mathbf{x})p(\mathbf{p})$.

6. From equation 9, we can derive that the expression of marginal distribution:

$$p(\mathbf{p}) = \int p(\mathbf{x}, \mathbf{p}) d\mathbf{x} \propto \exp[-\mathbf{p}^T M^{-1} \mathbf{p}]. \tag{10}$$

   From equation 10, we know that $p(\mathbf{p})$ is a Gaussian distribution with mean $\mathbf{0}$ and Varaince M.

7. The update of $\mathbf{x}$ and $\mathbf{p}$ is based on *Hamiltonian equation*, which is depended only on the hamiltonian $\mathcal{H}(\mathbf{x}, \mathbf{p})$ but not the partition function Z. Therefore the leapfrog algorithm does not require the knowledge of the partition function.