INFERENCE AND REPRESENTATION: HOMEWORK 5

STUDENT: MARGARITA BOYARSKAYA

Problem 1. *Convexity of* $L(v, x)$.
Let us prove that the Hessian H is positive semidefinite.

$$H_{ij} = \frac{\partial^2 L}{\partial v_i \partial v_j} = \frac{\partial}{\partial v_i}\left(\frac{\partial}{\partial v_j}\frac{1}{2}\sum_k m_k v_k\right) = \frac{\partial}{\partial v_i} m_j v_j = \delta_{ij} m_i$$

where δ_{ij} is Kronecker's delta function.

It is evident that the positivity criterion is satisfied by H, establishing the convexity of L by v. $\square$

Problem 2. *Convexity of* $f^*(p)$.

In order to be convinced in the convexity of the Legendre-Fenchel conjugate f^* it suffices to note that f^* is the supremum of a family of affine continuous functions $(\langle y, \cdot\rangle - f(y)), y \in \Omega$.

Indeed, take $\{\phi_n\}_{n\in N}$ – a family of convex functions defined on a convex set $\Omega \subset R^k$, and denote with ϕ the supremum of $\{\phi_n\}$. Using the convexity of ϕ_n for any chosen index $n \in N$, we can see that, for any two points $\omega_1, \omega_2 \in \Omega$:

$$\phi_n(a\omega_1 + (1-a)\omega_2) \leq a\phi_n(\omega_1) + (1-a)\phi_n(\omega_2) \leq a\phi(\omega_1) + (1-a)\phi(\omega_2),$$

for any $a \in [0, 1]$. Transitioning to supremum on the left hand side, we obtain:

$$\phi(a\omega_1 + (1-a)\omega_2) \leq a\phi(\omega_1) + (1-a)\phi(\omega_2).$$

This shows that ϕ is convex. $\square$

Problem 3. *Derivation of the Legendre-Fenchel conjugate of* $L(v, x)$.

For a differentiable function f, the Legendre-Fenchel transform can be expressed as:

$$f^*(p) = max_{v\in\Omega}(\langle v, p\rangle - f(v)) = \langle v_p, p\rangle - f(v_p),$$

where v_p solves $p = f'(v_p)$.

Coming back to the Lagrangian, the expression $L'(v_p) = p$ immediately yields $v_p = M^{-1}p$. We thus obtain:

$$L^*(p) = \langle M^{-1}p, p\rangle - L(v_p) = \langle M^{-1}p, p\rangle - \frac{1}{2}\langle v_p, Mv_p\rangle + U(x) = \frac{1}{2}\langle M^{-1}p, p\rangle + U(x). \ \square$$

Problem 4. *Derivation of the Hamiltonian equations.*

$$H(p, x) = \langle v_p, p\rangle - L(v_p, x).$$

The total differential is:

$$\frac{\partial H}{\partial t} = \langle \frac{\partial H}{\partial p}, \dot{p}\rangle + \langle \frac{\partial H}{\partial x}, \dot{x}\rangle. \tag{1}$$

On the other hand, it holds that:

$$\frac{\partial H}{\partial t} = \langle p, \dot{v}_p\rangle + \langle v_p, \dot{p}\rangle - \langle \frac{\partial L}{\partial v_p}, \dot{v}_p\rangle - \langle \frac{\partial L}{\partial x}, \dot{x}\rangle,$$

which, considering the Euler-Lagrange equation and the definition $p = \frac{\partial L}{\partial v}\Big|_{v_p}$, becomes:

$$\frac{\partial H}{\partial t} = \langle p, \dot{v}_p\rangle + \langle v_p, \dot{p}\rangle - \langle p, \dot{v}_p\rangle - \langle \dot{p}, \dot{x}\rangle = \langle v_p, \dot{p}\rangle - \langle \dot{p}, \dot{x}\rangle. \tag{2}$$

Equating (1) and (2), we arrive at the sought identities:

$$\langle v_p, \dot{p}\rangle = \langle \frac{\partial H}{\partial p}, \dot{p}\rangle,$$

$$\langle \dot{p}, \dot{x}\rangle = \langle \frac{\partial H}{\partial x}, \dot{x}\rangle. \ \square$$

Problem 5. *Derivation of the canonical distribution.*

Using $U(x) = -\log \tilde{P}(x)$, we have $P(x) = \frac{1}{Z}e^{-U(x)}$. Let us denote $A(p) := \frac{1}{2}\langle p, M^{-1}p\rangle$ to obtain:

$$H(p, x) = A(p) + U(x).$$

Writing $P(x, p) = \frac{1}{Z}e^{-H(x,p)}$, the joint density is, finally,:

$$P(x, p) = \frac{1}{Z}e^{-U(x)}e^{-A(p)}, \tag{3}$$

and we see that x and p are independent, and each has canonical distributions with energey functions $U(x)$ and $A(p)$. $\square$

Since the joint probability factors into independent $x-$ and $p-$parts, we can obtain samples from $P(x)$ by sampling from $P(x, p)$ and ignoring the momentum variable p.

Indeed, we are free to draw values of p independently of the current values of x from its correct conditional distribution given x (which is precisely the marginal $A(p)$, due to independence). The canonical joint distribution remains invariant.

The subsequent step changes both x and p with a Metropolis update, new values suggested by the Hamiltonian dynamics. The canonical distribution $P(x, p)$ still remains invariant. From the current state simulation continues for some number of steps using the Leapfrog algorithm.

Problem 6. *The marginal $P(p)$.*

$P(p) = e^{-A(x)} = e^{-\frac{1}{2}\langle p, M^{-1}p\rangle}$. Since the normalizing constant ensures that the total area under the curve is equal to one, it suffices to take $\sigma = \sqrt{M}$ to see that $P(p) \sim \mathcal{N}(0, M)$. $\square$

Problem 7. *The Leapfrog algorithm does not require knowledge of Z.*

The variable Z is a normalizing constant that functions to scale the canonical distribution to make it valid a probability distribution that sums to one. This variable is inconsequential, because it cancels out in the acceptance rule. We accept the proposed (x^*, p^*) with probability:

$$\min(1, exp(H(x^*, p^*) - H(x, p)) = \min[1, exp(-U(x^*) + U(x) - K(p^*) + K(p))].$$

Problem 8. *The Metropolis-Hastings algorithm produces samples from $P(x)$ when marginalizing over position variables.*

In order to prove that the Metropolis update (and hence, following from discussion in the solution to problem 5, the HMC method overall) leaves the canonical distribution invariant, we will explore the idea of volume preservation [1].

Note that the Metropolis update is reversible with respect to the canonical distribution for x and p. Suppose that we partition the (x, p) space into regions C_k, each with the same small volume V. Denote $B_k = Leapfrog_L(C_k)$, that is, the image of C_k after performing the Leapfrog algorithm for L steps (negation the momenta). Using the reversibility, we see that B_k also partition the space, and since the leapfrog steps preserve volume (as does negation), each B_k will also have volume V.

Notice that for any the following identity for the conditional probabilities holds: $i, j, i = j$, $P(C_i|B_j) = P(B_j|C_i) = 0$. It follows, then, that:

$$P(C_i)P(B_j|C_i) = P(B_j)P(C_i|B_j).$$

Transitioning to the limit, the regions C_k and B_k become smaller, and the Hamiltonian can be treated as constant within each region, together with the canonical probability density. We thus get:

$$\frac{V}{Z}e^{-H_{C_i}} \min[1, exp(-H_{B_i} + H_{C_i})] = \frac{V}{Z}e^{-H_{B_i}} \min[1, exp(-H_{C_i} + H_{B_i})],$$

where H_X is the value of the Hamiltonian in region X.

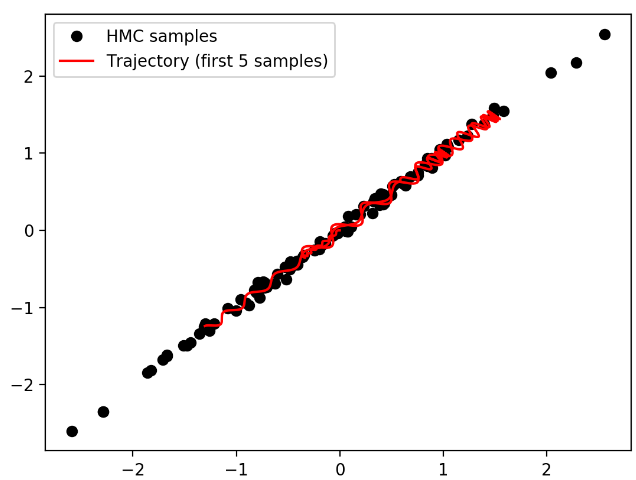
This implies that Metropolis update leaves the canonical distribution for x and p invariant. Denote $R(X)$ the probability that the Metropolis update for a state in small region X leads to a rejection. The probability of the next state belonging to a small region B_k is the sum of the probability that the current state is in B_k and the update leads to rejection, and the probability that the current state is in some region from which a move to B_k is proposed and accepted. The probability of the next state being in B_k can therefore be written as:

$$P(B_i)R(B_i) + \sum_k P(C_k)P(B_i|C_k) = P(B_i)R(B_i) + P(B_i)(1 - R(B_i)) = P(B_i).$$

We have prove that the Metropolis update within HMC therefore leaves the canonical distribution invariant. $\square$

Problem 9. *Python implementation.*

The code that provides the solution to the problem is attached in a file titled *problem9.py*. Below is the image illustrating the trajectories and the samples that result.



References

[1] http://www.tjsullivan.org.uk/pdf/2016-05-31_UQ_seminar_GradientMC.pdf

[2] https://theclevermachine.wordpress.com/2012/11/18/mcmc-hamiltonian-monte-carlo-a-k-a-hybrid-monte-carlo