

Hamiltonian Monte Carlo with Reduced Momentum Flips

Jascha Sohl-Dickstein
 Redwood Center for Theoretical Neuroscience
 University of California at Berkeley

May 10, 2012

Hamiltonian dynamics with partial momentum refreshment, in the style of [Horowitz, 1991], explore the state space more slowly than they otherwise would due to the momentum reversals which occur on proposal rejection. These cause trajectories to double back on themselves, leading to random walk behavior on timescales longer than the typical rejection time, and leading to slower mixing. I present a technique by which the number of momentum reversals can be reduced. This is accomplished by maintaining the net exchange of probability between states with opposite momenta, but reducing the rate of exchange in both directions such that it is 0 in one direction. An experiment illustrates these reduced momentum flips accelerating mixing for a particular distribution.

1 Formalism

A state $\zeta \in \mathcal{R}^{N \times 2}$ consists of a position $\mathbf{x} \in \mathcal{R}^N$ and an auxiliary momentum $\mathbf{v} \in \mathcal{R}^N$, $\zeta = \{\mathbf{x}, \mathbf{v}\}$. The state space has an associated Hamiltonian

$$H(\zeta) = E(\mathbf{x}) + \frac{1}{2} \mathbf{v}^T \mathbf{v}, \quad (1)$$

and a joint probability distribution

$$p(\mathbf{x}, \mathbf{v}) = p(\zeta) = \frac{1}{Z} \exp(-H(\zeta)), \quad (2)$$

where the normalization constant Z is the partition function.

The momentum flip operator $F : \mathcal{R}^{N \times 2} \rightarrow \mathcal{R}^{N \times 2}$ negates the momentum. It has the properties:

- F negates the momentum, $F\zeta = F\{\mathbf{x}, \mathbf{v}\} = \{\mathbf{x}, -\mathbf{v}\}$
- F is its own inverse, $F^{-1} = F$, $FF\zeta = \zeta$.
- F is volume preserving, $\det \left(\frac{\partial(F\zeta)}{\partial\zeta^T} \right) = 1$

- F doesn't change the probability of a state, $p(\zeta) = p(F\zeta)$

The leapfrog integrator $L(n, \epsilon) : \mathcal{R}^{N \times 2} \rightarrow \mathcal{R}^{N \times 2}$ integrates Hamiltonian dynamics for the Hamiltonian $H(\zeta)$, using leapfrog integration, for $n \in \mathcal{Z}^+$ integration steps with stepsize $\epsilon \in \mathcal{R}^+$. We assume that n and ϵ are constants, and write this operator simply as L . The leapfrog integrator L has the following relevant properties:

- L is volume preserving, $\det\left(\frac{\partial(L\zeta)}{\partial\zeta^T}\right) = 1$
- L is exactly reversible using momentum flips, $L^{-1} = FLF$, $\zeta = FLFL\zeta$

During sampling, state updates are performed using a transition operator $T(r) : \mathcal{R}^{N \times 2} \rightarrow \mathcal{R}^{N \times 2}$, where $r \sim U([0, 1])$ is drawn from the uniform distribution between 0 and 1,

$$T(r)\zeta = \begin{cases} L\zeta & r < P_{leap}(\zeta) \\ F\zeta & P_{leap} \leq r < P_{leap}(\zeta) + P_{flip}(\zeta) \\ \zeta & P_{leap} + P_{flip}(\zeta) \leq r \end{cases} . \quad (3)$$

$T(r)$ additionally depends on an acceptance probability for the leapfrog dynamics, $P_{leap}(\zeta) \in [0, 1]$, and a probability of negating the momentum, $P_{flip}(\zeta) \in [0, 1 - P_{leap}(\zeta)]$. These must be chosen to guarantee that $p(\zeta)$ is a fixed point of T .¹

2 Making the distribution of interest a fixed point

In order to make $p(\zeta)$ a fixed point, we will choose the Markov dynamics T so that on average as many transitions enter as leave state ζ at equilibrium. This is *not* pairwise detailed balance — instead we are directly enforcing zero net change in the probability of each state by summing over all allowed transitions into or out of the state. This constraint is analogous to Kirchhoff's current law, where the total current entering a node is set to 0. As can be seen from Equation 3 and the definitions in Section 1, and as is illustrated in Figure 1, a state ζ can only lose probability to the two states $L\zeta$ and $F\zeta$, and gain probability from the two states $L^{-1}\zeta$ and $F^{-1}\zeta$. Equating the rates of probability inflow and outflow, we find

$$p(\zeta) P_{leap}(\zeta) + p(\zeta) P_{flip}(\zeta) = p(L^{-1}\zeta) P_{leap}(L^{-1}\zeta) + p(F^{-1}\zeta) P_{flip}(F^{-1}\zeta) \quad (4)$$

$$= p(L^{-1}\zeta) P_{leap}(L^{-1}\zeta) + p(\zeta) P_{flip}(F\zeta) \quad (5)$$

$$P_{flip}(\zeta) - P_{flip}(F\zeta) = \frac{p(L^{-1}\zeta)}{p(\zeta)} P_{leap}(L^{-1}\zeta) - P_{leap}(\zeta). \quad (6)$$

¹This fixed point requirement can be written as $p(\zeta) = \int d\zeta' p(\zeta') \int_0^1 dr \delta(\zeta - T(r)\zeta')$.

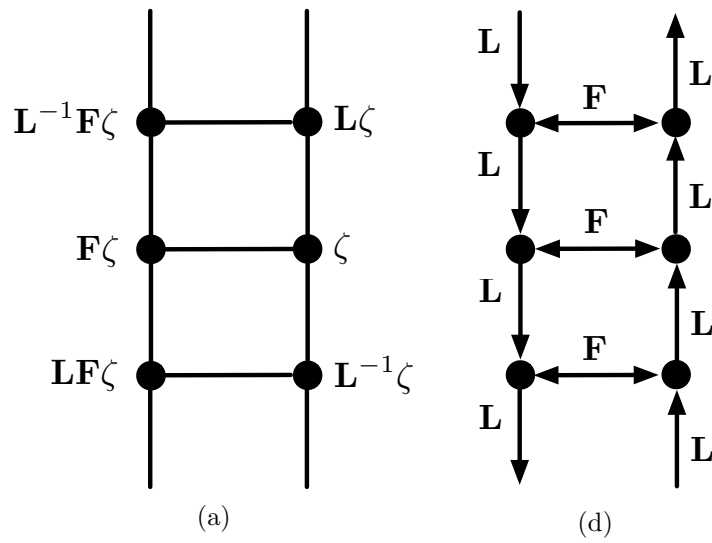


Figure 1: This diagram illustrates the possible transitions between states using the Markov transition operator from Equation 3. In (a) the relevant states, represented by the nodes, are labeled. In (b) the possible transitions, represented by the arrows, are labeled. In Section 2, the net probability flow into and out of the state ζ is set to 0.

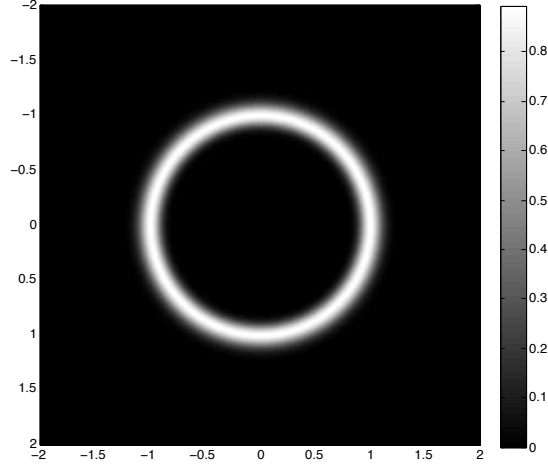


Figure 2: A two dimensional image of the distribution used in Section 3. Pixel intensity corresponds to the probability density function at that location.

We choose the standard Metropolis-Hastings acceptance rules for $P_{leap}(\zeta)$,

$$P_{leap}(\zeta) = \min\left(1, \frac{p(L\zeta)}{p(\zeta)}\right). \quad (7)$$

Substituting this in to Equation 6, we find

$$P_{flip}(\zeta) - P_{flip}(F\zeta) = \frac{p(L^{-1}\zeta)}{p(\zeta)} \min\left(1, \frac{p(LL^{-1}\zeta)}{p(L^{-1}\zeta)}\right) - \min\left(1, \frac{p(L\zeta)}{p(\zeta)}\right) \quad (8)$$

$$= \min\left(1, \frac{p(L^{-1}\zeta)}{p(\zeta)}\right) - \min\left(1, \frac{p(L\zeta)}{p(\zeta)}\right) \quad (9)$$

$$= \min\left(1, \frac{p(LF\zeta)}{p(\zeta)}\right) - \min\left(1, \frac{p(L\zeta)}{p(\zeta)}\right). \quad (10)$$

Satisfying Equation 10 we choose² the following form for $P_{flip}(\zeta)$,

$$P_{flip}(\zeta) = \max\left(0, \min\left(1, \frac{p(LF\zeta)}{p(\zeta)}\right) - \min\left(1, \frac{p(L\zeta)}{p(\zeta)}\right)\right). \quad (11)$$

Note that $P_{flip}(\zeta) \leq 1 - P_{leap}(\zeta)$, where $1 - P_{leap}(\zeta)$ is the rejection rate, and thus the momentum flip rate, in standard HMC. Using this form for $P_{flip}(\zeta)$ will generally reduce the number of momentum flips required.

²To recover standard HMC, instead set $P_{flip}(\zeta) = 1 - P_{leap}(\zeta)$. One can verify by substitution that this satisfies Equation 10.

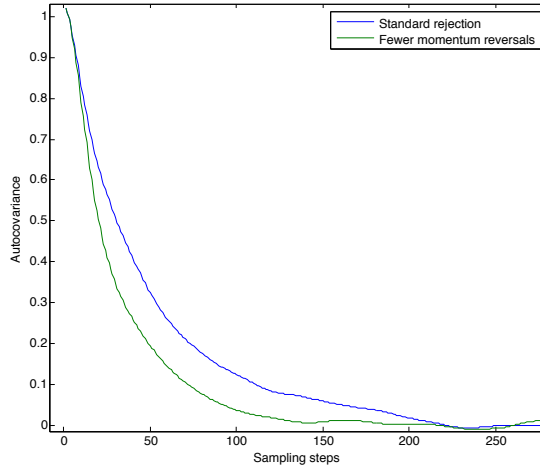


Figure 3: The covariance between samples as a function of the number of intervening sampling steps for HMC with standard rejection and rejection with fewer momentum reversals. Reducing the number of momentum reversals causes faster mixing, as evidenced by the faster falloff of the autocovariance.

3 Example

In order to demonstrate the accelerated mixing provided by this technique, samples were drawn from a simple distribution with standard rejection, and with separate rejection and momentum flipping rates as described above. In both cases, the leapfrog step length ϵ was set to 0.1, the number of integration steps n was set to 1, and the momentum corruption rate β was set so as to corrupt half the momentum per unit stimulation time. Both samplers were run for 100,000 sampling steps. The distribution used was described by the energy function

$$E = 100 \log^2 \left(\sqrt{x_1^2 + x_2^2} \right). \tag{12}$$

A 2 dimensional image of this distribution can be seen in Figure 2. The autocovariance of the returned samples can be seen, as a function of the number of intervening sampling steps, in Figure 3. Sampling using the technique presented here led to more rapid decay of the autocovariance, consistent with faster mixing.

References

[Horowitz, 1991] Horowitz, A. (1991). A generalized guided Monte Carlo algorithm. *Physics Letters B*, 268(2):247–252.