

UC Irvine

UC Irvine Previously Published Works

Title

Markov Chain Monte Carlo from Lagrangian Dynamics

Permalink

<https://escholarship.org/uc/item/23t5w14s>

Journal

Journal of Computational and Graphical Statistics, 24(2)

ISSN

1061-8600

Authors

Lan, Shiwei
Stathopoulos, Vasileios
Shahbaba, Babak
[et al.](#)

Publication Date

2015-04-03

DOI

10.1080/10618600.2014.902764

Peer reviewed

Lagrangian Dynamical Monte Carlo

Shiwei Lan*, Vassilios Stathopoulos†, Babak Shahbaba*, Mark Girolami†

November 22, 2012

Abstract

Hamiltonian Monte Carlo (HMC) improves the computational efficiency of the Metropolis algorithm by reducing its random walk behavior. Riemannian Manifold HMC (RMHMC) further improves HMC’s performance by exploiting the geometric properties of the parameter space. However, the geometric integrator used for RMHMC involves implicit equations that require costly numerical analysis (e.g., fixed-point iteration). In some cases, the computational overhead for solving implicit equations undermines RMHMC’s benefits. To avoid this problem, we propose an explicit geometric integrator that replaces the momentum variable in RMHMC by velocity. We show that the resulting transformation is equivalent to transforming Riemannian Hamilton dynamics to Lagrangian dynamics. Experimental results show that our method improves RMHMC’s overall computational efficiency. All computer programs and data sets are available online (<http://www.ics.uci.edu/~babaks/Site/Codes.html>) in order to allow replications of the results reported in this paper.

1 Introduction

Hamiltonian Monte Carlo (HMC) (Duane et al., 1987) reduces the random walk behavior of Metropolis by proposing samples that are distant from the current state, but nevertheless have a high probability of acceptance. These distant proposals are found by numerically simulating Hamiltonian dynamics for some specified amount of fictitious time (Neal, 2010). Hamiltonian dynamics can be represented by a function, known as the *Hamiltonian* function, of model parameters $\boldsymbol{\theta}$ and fictitious momentum parameters $\mathbf{p} \sim N(\mathbf{0}, \mathbf{M})$ (with the same dimension as $\boldsymbol{\theta}$) as follows:

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\log p(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^\top \mathbf{M}^{-1}\mathbf{p} \quad (1)$$

where \mathbf{M} is a symmetric, positive-definite *mass matrix*.

Hamilton’s equations, which involve differential equations of H , determine how $\boldsymbol{\theta}$ and \mathbf{p} change over time. In practice, however, solving these equations exactly is too hard, so we need to approximate them by discretizing time, using some small step size ε . For this purpose, the *leapfrog* method is commonly used.

*Department of Statistics, University of California, Irvine, USA.

†Department of Statistical Science, University College London, UK.

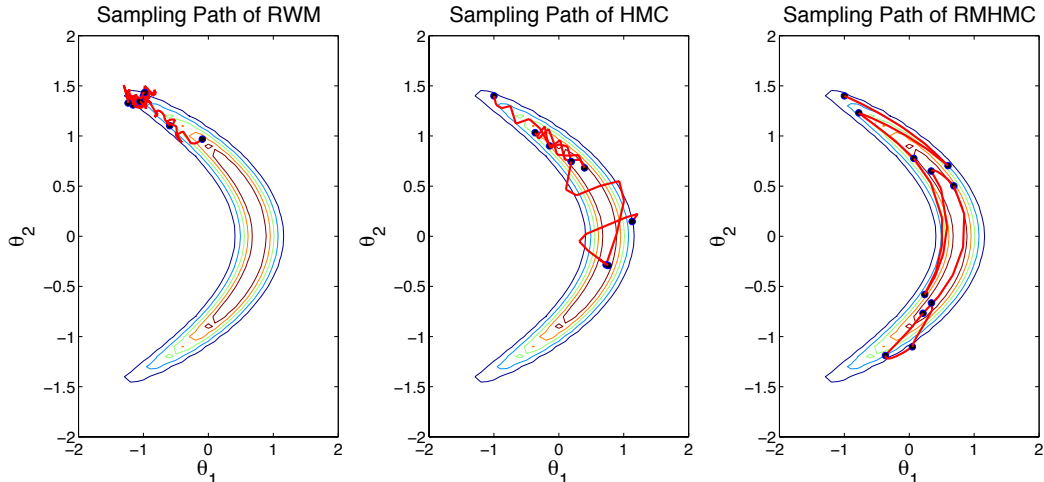


Figure 1: The first 10 iterations in sampling from a banana shaped distribution with random walk Metropolis (RWM), Hamiltonian Monte Carlo (HMC), and Riemannian Manifold HMC (RMHMC). For all three methods, the trajectory length (i.e., step size times number of integration steps) is set to 1. Solid red lines are the sampling path, and black circles are the accepted proposals.

As the dimension grows, the system becomes increasingly restricted by its smallest eigen-direction, requiring smaller step sizes to maintain the stability of numerical discretization. Girolami and Calderhead (2011) proposed a new method, called Riemannian Manifold HMC (RMHMC), that exploits the geometric properties of the parameter space to improve the efficiency of standard HMC. Simulating from the resulting dynamic, however, is computationally intensive since it involves solving two implicit equations, which require additional iterative numerical analysis (e.g., fixed-point iteration).

To increase RMHMC’s speed, we propose a new integrator that is completely explicit: we propose to replace momentum with velocity in Riemannian Manifold Hamilton dynamics. As we will see, this is equivalent to using Lagrangian dynamics as opposed to Hamiltonian dynamics. By doing so, we eliminate one of the implicit steps in RMHMC. Next, we construct a time symmetric integrator to remove the remaining implicit step in RHHMC. This leads to a sampling scheme, called e-RMHMC, that involves explicit equations only.

In what follows, we start with a brief review of RMHMC and its geometric integrator. Section 3 introduces our proposed semi-explicit integrator based on defining Hamiltonian dynamics in terms of velocity as opposed to momentum. Next, in Section 4, we eliminate the remaining implicit equation and propose a fully explicit integrator. In Section 5, we use simulated and real data to evaluate our methods’ performance. Finally, in Section 6, we discuss some possible future research directions.

2 Riemannian Manifold Hamiltonian Monte Carlo

As discussed above, although HMC explores the parameter space more efficiently than random walk Metropolis does, it does not fully exploits the geometric properties of parameter space defined by the density $p(\boldsymbol{\theta})$. Indeed, Girolami and Calderhead (2011) argue that dynamics over Euclidean space may not be appropriate to guide the exploration of parameter space. To address this issue, they propose a new method, called Riemannian Manifold HMC (RMHMC), that exploits the Riemannian geometry of the parameter space (Amari and Nagaoka, 2006) to improve standard HMC's efficiency by automatically adapting to the local structure. They do this by using a position-specific mass matrix $\mathbf{M} = \mathbf{G}(\boldsymbol{\theta})$. More specifically, they set $\mathbf{G}(\boldsymbol{\theta})$ to the Fisher information matrix. As a result, $\mathbf{p} = \mathbf{G}(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} \sim \mathcal{N}(0, \mathbf{G}(\boldsymbol{\theta}))$, and Hamiltonian is defined as follows:

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\log p(\boldsymbol{\theta}) + \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{p}^\top \mathbf{G}(\boldsymbol{\theta})^{-1} \mathbf{p} = \phi(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{p}^\top \mathbf{G}(\boldsymbol{\theta})^{-1} \mathbf{p} \quad (2)$$

where $\phi(\boldsymbol{\theta}) := -\log p(\boldsymbol{\theta}) + \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta})$. Based on this dynamic, Girolami and Calderhead (2011) propose the following HMC on Riemmanian manifold:

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \nabla_{\mathbf{p}} H(\boldsymbol{\theta}, \mathbf{p}) = \mathbf{G}(\boldsymbol{\theta})^{-1} \mathbf{p} \\ \dot{\mathbf{p}} &= -\nabla_{\boldsymbol{\theta}} H(\boldsymbol{\theta}, \mathbf{p}) = -\nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) + \frac{1}{2} \boldsymbol{\nu}(\boldsymbol{\theta}, \mathbf{p}) \end{aligned} \quad (3)$$

Using the shorthand notation $\partial_i = \partial/\partial\theta_i$ for partial derivative, the i th element of the vector $\boldsymbol{\nu}(\boldsymbol{\theta}, \mathbf{p})$ is

$$(\boldsymbol{\nu}(\boldsymbol{\theta}, \mathbf{p}))_i = -\mathbf{p}^\top \partial_i (\mathbf{G}(\boldsymbol{\theta})^{-1}) \mathbf{p} = (\mathbf{G}(\boldsymbol{\theta})^{-1} \mathbf{p})^\top \partial_i \mathbf{G}(\boldsymbol{\theta}) \mathbf{G}(\boldsymbol{\theta})^{-1} \mathbf{p}$$

The above dynamic is non-separable (it contains products of $\boldsymbol{\theta}$ and p), and the resulting map $(\boldsymbol{\theta}, \mathbf{p}) \rightarrow (\boldsymbol{\theta}^*, \mathbf{p}^*)$ based on the standard leapfrog method is neither time-reversible nor symplectic. Therefore, the standard leapfrog algorithm cannot be used for the above dynamic (Girolami and Calderhead, 2011). Instead, we can use the Stömer-Verlet (Verlet, 1967) method as follows:

$$\mathbf{p}^{(n+1/2)} = \mathbf{p}^{(n)} - \frac{\varepsilon}{2} \left[\nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n)}) - \frac{1}{2} \boldsymbol{\nu}(\boldsymbol{\theta}^{(n)}, \mathbf{p}^{(n+1/2)}) \right] \quad (4)$$

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \frac{\varepsilon}{2} \left[\mathbf{G}^{-1}(\boldsymbol{\theta}^{(n)}) + \mathbf{G}^{-1}(\boldsymbol{\theta}^{(n+1)}) \right] \mathbf{p}^{(n+1/2)} \quad (5)$$

$$\mathbf{p}^{(n+1)} = \mathbf{p}^{(n+1/2)} - \frac{\varepsilon}{2} \left[\nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n+1)}) - \frac{1}{2} \boldsymbol{\nu}(\boldsymbol{\theta}^{(n+1)}, \mathbf{p}^{(n+1/2)}) \right] \quad (6)$$

This is also known as generalized leapfrog (Leimkuhler and Reich, 2004). The above series of transformations are (i) deterministic (ii) reversible and (iii) volume-preserving. Therefore, the effective proposal distribution is a delta function $\delta((\boldsymbol{\theta}^{(1)}, \mathbf{p}^{(1)}), (\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}))$ and the acceptance probability is as follows:

$$\frac{\exp(-H(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}))}{\exp(-H(\boldsymbol{\theta}^{(1)}, \mathbf{p}^{(1)}))} \times \frac{\delta((\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}), (\boldsymbol{\theta}^{(1)}, \mathbf{p}^{(1)}))}{\delta((\boldsymbol{\theta}^{(1)}, \mathbf{p}^{(1)}), (\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}))} = \exp(H(\boldsymbol{\theta}^{(1)}, \mathbf{p}^{(1)}) - H(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)})) \quad (7)$$

Here, $(\boldsymbol{\theta}^{(1)}, \mathbf{p}^{(1)})$ is the current state, and $(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)})$ is the proposal after L leapfrog steps.

As an illustrative example, Figure 1 shows the sampling paths of random walk Metropolis (RWM), HMC, and RMHMC for an artificially created banana-shaped distribution (See Girolami and Calderhead, 2011, discussion by Luke Bornn and Julien Cornebise). For this example, we fixed the trajectory and chose the step sizes such that the acceptance probability for all three methods remains around 0.7. RWMH moves slowly and spends most of iterations at the distribution’s low-density tail, and HMC explores the parameter space in a tortuous way, while RMHMC moves directly to the high-density region and explores the distribution more efficiently.

One major drawback of this geometric integrator, which is both time-reversible and volume-preserving, is that it involves two implicit functions: Equations (4) and (5). These functions require extra numerical analysis (e.g. fixed-point iteration), which results in higher computational cost and simulation error. To address this problem, we propose an alternative approach that uses velocity instead of momentum.

3 Moving from Momentum to Velocity

In the Hamiltonian dynamic (3), the product of $\mathbf{G}(\boldsymbol{\theta})^{-1}$ and \mathbf{p} is in fact velocity, $\mathbf{v} = \mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p}$. This motivates us to define the dynamic in terms of \mathbf{v} instead of \mathbf{p} . The transformation $\mathbf{p} \mapsto \mathbf{v}$ changes the Hamiltonian dynamics (3) to the following form (derivation in Appendix A):

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{v}) - \mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta})\end{aligned}\tag{8}$$

where $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{v})$ is a vector whose k th element is $\sum_{i,j} \Gamma_{ij}^k(\boldsymbol{\theta})\mathbf{v}^i\mathbf{v}^j$. Here, $\Gamma_{ij}^k(\boldsymbol{\theta}) := \frac{1}{2} \sum_l \mathbf{g}^{kl}(\partial_i \mathbf{g}_{lj} + \partial_j \mathbf{g}_{il} - \partial_l \mathbf{g}_{ij})$ is Christoffel symbol whose (i, j) th element is $\mathbf{G}(\boldsymbol{\theta}) = (\mathbf{g}_{ij})$. Further, $\mathbf{G}(\boldsymbol{\theta})^{-1} = (\mathbf{g}^{ij})$.

This transformation moves the Hamiltonian dynamic’s complexity from its first equation for $\boldsymbol{\theta}$ to its second equation. In this way, we resolve one implicit function in the generalized leapfrog method and develop a semi-explicit integrator as follows:

$$\mathbf{v}^{(n+1/2)} = \mathbf{v}^{(n)} - \frac{\varepsilon}{2} [(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)}) \mathbf{v}^{(n+1/2)} + \mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n)})]\tag{9}$$

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \varepsilon \mathbf{v}^{(n+1/2)}\tag{10}$$

$$\mathbf{v}^{(n+1)} = \mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2} [(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)}) \mathbf{v}^{(n+1/2)} + \mathbf{G}(\boldsymbol{\theta}^{(n+1)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n+1)})]\tag{11}$$

Note that updating \mathbf{v} remains implicit (more details are available in Appendix B).

In general, the new dynamic (8) cannot be recognized as a Hamiltonian dynamic of $(\boldsymbol{\theta}, \mathbf{v})$. Nevertheless, it remains a valid proposal-generating mechanism, which preserves the original Hamiltonian $H(\boldsymbol{\theta}, \mathbf{p} = \mathbf{G}(\boldsymbol{\theta})\mathbf{v})$ (proof in Appendix A); thus, the acceptance probability is only determined by a discretization error from the numerical integration, as before.

Because $\mathbf{p} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}))$, the distribution of $\mathbf{v} = \mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p}$ is $\mathcal{N}(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta})^{-1})$. Therefore, we have

$$p(\mathbf{v}) = \frac{1}{(\sqrt{2\pi})^D \sqrt{\det(\mathbf{G}(\boldsymbol{\theta})^{-1})}} \exp \left\{ -\frac{1}{2} \mathbf{v}^\top \mathbf{G}(\boldsymbol{\theta}) \mathbf{v} \right\} \propto (\det \mathbf{G}(\boldsymbol{\theta}))^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{v}^\top \mathbf{G}(\boldsymbol{\theta}) \mathbf{v} \right\}$$

Algorithm 1 Semi-explicit Riemannian Manifold Lagrangian Monte Carlo (RMLMC)

```

Initialize  $\boldsymbol{\theta}^{(1)} = \text{current } \boldsymbol{\theta}$ 
Sample new velocity  $\mathbf{v}^{(1)} \sim \mathcal{N}(0, \mathbf{G}^{-1}(\boldsymbol{\theta}^{(1)}))$ 
Calculate current  $\mathbf{E}(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})$  according to equation (23)
for  $n = 1$  to  $L$  (leapfrog steps) do
  % Update the velocity with fixed point iterations
   $\hat{\mathbf{v}}^{(0)} = \mathbf{v}^{(n)}$ 
  for  $i = 1$  to NumOfFixedPointSteps do
     $\hat{\mathbf{v}}^{(i)} = \mathbf{v}^{(n)} - \frac{\varepsilon}{2} \mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1} [(\hat{\mathbf{v}}^{(i-1)})^\top \tilde{\Gamma}(\boldsymbol{\theta}^{(n)}) \hat{\mathbf{v}}^{(i-1)} + \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n)})]$ 
  end for
   $\mathbf{v}^{(n+1/2)} = \hat{\mathbf{v}}^{(\text{last } i)}$ 
  % Update the position only with simple one step
   $\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \varepsilon \mathbf{v}^{(n+1/2)}$ 
   $\Delta \log \det_n = \log \det(\mathbf{I} - \varepsilon (\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})) - \log \det(\mathbf{I} + \varepsilon (\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)}))$ 
  Update the velocity exactly
   $\mathbf{v}^{(n+1)} = \mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2} \mathbf{G}(\boldsymbol{\theta}^{(n+1)})^{-1} [(\mathbf{v}^{(n+1/2)})^\top \tilde{\Gamma}(\boldsymbol{\theta}^{(n+1)}) \mathbf{v}^{(n+1/2)} + \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n+1)})]$ 
end for
Calculate proposed  $\mathbf{E}(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})$  according to equation (23)
logRatio = -ProposedH + CurrentH +  $\sum_{n=1}^N \Delta \log \det_n$ 
Accept or reject according to Metropolis ratio

```

We define the *energy* function $\mathbf{E}(\boldsymbol{\theta}, \mathbf{v})$ as the sum of the potential energy, $U(\boldsymbol{\theta})$ and $K(\boldsymbol{\theta}, \mathbf{v})$, where $K(\boldsymbol{\theta}, \mathbf{v}) = -\log(p(\mathbf{v}))$.

Analogous to RMHMC, the acceptance probability is calculated based on $\mathbf{E}(\boldsymbol{\theta}, \mathbf{v})$, which is the negative log of the joint density of parameter $\boldsymbol{\theta}$ and the new auxiliary variable \mathbf{v} . Therefore, we can apply the generalized leapfrog scheme to this new dynamic (8) and derive a semi-implicit method. (See more details in Appendix B.) Although the resulting integrator is not symplectic, we nonetheless have detailed balance with volume correction. Algorithm 1 shows the corresponding steps for implementing this method.

In Appendix C, we show that the new dynamic (8) is essentially a *Lagrangian* dynamic. Therefore, we refer to the derived algorithm (Algorithm 1) as Riemannian Manifold Lagrangian Monte Carlo (RMLMC), which explores the parameter space along the path on a Riemannian manifold that minimizes the total Lagrangian. We can use this new proposal-generating mechanism, RMLMC, which is based on an Euler-Lagrange system (8) of $(\boldsymbol{\theta}, \mathbf{v})$ instead of the original Hamiltonian system (3) defined in terms of $(\boldsymbol{\theta}, \mathbf{p})$. The two methods use equivalent dynamics but differ numerically in the following way: RMHMC augments parameter space with momentum, while RMLMC augments parameter space with velocity. Later, we will show that switching to velocity leads to substantial improvement in computational efficiency.

Algorithm 2 Explicit Riemannian Manifold Lagrangian Monte Carlo (e-RMLMC)

Initialize $\boldsymbol{\theta}^{(1)} = \text{current } \boldsymbol{\theta}$
Sample new velocity $\mathbf{v}^{(1)} \sim \mathcal{N}(0, \mathbf{G}(\boldsymbol{\theta}^{(1)})^{-1})$
Calculate current $\mathbf{E}(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})$ according to equation (23)
 $\Delta \log \det = 0$
for $n = 1$ to L **do**
 $\Delta \log \det = \Delta \log \det - \det(\mathbf{G}(\boldsymbol{\theta}^{(n)}) + \varepsilon/2\tilde{\boldsymbol{\Omega}}(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}))$
 Update the velocity **explicitly** with a half step:
 $\mathbf{v}^{(n+1/2)} = [\mathbf{G}(\boldsymbol{\theta}^{(n)}) + \frac{\varepsilon}{2}\tilde{\boldsymbol{\Omega}}(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})]^{-1}[\mathbf{G}(\boldsymbol{\theta}^{(n)})\mathbf{v}^{(n)} - \frac{\varepsilon}{2}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}^{(n)})]$
 $\Delta \log \det = \Delta \log \det + \det(\mathbf{G}(\boldsymbol{\theta}^{(n)}) - \varepsilon/2\tilde{\boldsymbol{\Omega}}(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n+1/2)}))$
 Update the position with a full step:
 $\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \varepsilon\mathbf{v}^{(n+1/2)}$
 $\Delta \log \det = \Delta \log \det - \det(\mathbf{G}(\boldsymbol{\theta}^{(n+1)}) + \varepsilon/2\tilde{\boldsymbol{\Omega}}(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n+1/2)}))$
 Update the velocity **explicitly** with a half step:
 $\mathbf{v}^{(n+1)} = [\mathbf{G}(\boldsymbol{\theta}^{(n+1)}) + \frac{\varepsilon}{2}\tilde{\boldsymbol{\Omega}}(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n+1/2)})]^{-1}[\mathbf{G}(\boldsymbol{\theta}^{(n+1)})\mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}^{(n+1)})]$
 $\Delta \log \det = \Delta \log \det + \det(\mathbf{G}(\boldsymbol{\theta}^{(n+1)}) - \varepsilon/2\tilde{\boldsymbol{\Omega}}(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n+1)}))$
end for
Calculate proposed $\mathbf{E}(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})$ according to equation (23)
 $\log \text{Ratio} = -\text{ProposedE} + \text{CurrentE} + \Delta \log \det$
Accept or reject the proposed state according to (15)

4 Explicit Riemannian Manifold Lagrangian Monte Carlo

We now propose a fully explicit integrator for Lagrangian dynamics (8) as follows:

$$\mathbf{v}^{(n+1/2)} = [\mathbf{I} + \frac{\varepsilon}{2}\boldsymbol{\Omega}(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})]^{-1}[\mathbf{v}^{(n)} - \frac{\varepsilon}{2}\mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}^{(n)})] \quad (12)$$

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \varepsilon\mathbf{v}^{(n+1/2)} \quad (13)$$

$$\mathbf{v}^{(n+1)} = [\mathbf{I} + \frac{\varepsilon}{2}\boldsymbol{\Omega}(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n+1/2)})]^{-1}[\mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2}\mathbf{G}(\boldsymbol{\theta}^{(n+1)})^{-1}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}^{(n+1)})] \quad (14)$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})$ is a matrix whose (i, j) th element is $\sum_k \mathbf{v}_k^{(n)} \Gamma_{kj}^i(\boldsymbol{\theta}^{(n)})$. This integrator is (i) reversible and (ii) energy-preserving up to order $\mathcal{O}(\varepsilon)$, where ε is the stepsize. The resulting map, however, is not volume-preserving and as such the effective proposal distribution will be the product of a delta function and the determinant of the transformation,

$$\frac{\exp(-\mathbf{E}(\boldsymbol{\theta}^{(L)}, \mathbf{v}^{(L)}))}{\exp(-\mathbf{E}(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}))} \times \frac{\delta((\boldsymbol{\theta}^{(L)}, \mathbf{v}^{(L)}), (\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}))}{\delta((\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}), (\boldsymbol{\theta}^{(L)}, \mathbf{v}^{(L)}))} \times \det \mathbf{J}$$

which simplifies to

$$\exp(\mathbf{E}(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}) - \mathbf{E}(\boldsymbol{\theta}^{(L)}, \mathbf{v}^{(L)})) \times \det \mathbf{J} \quad (15)$$

with \mathbf{J} the Jacobian matrix of $(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}) \rightarrow (\boldsymbol{\theta}^{(L)}, \mathbf{v}^{(L)})$. Detailed derivations and proofs are given in Appendix D.

We refer to this approach as explicit Riemannian Manifold Lagrangian Monte Carlo (e-RMLMC). Algorithm 2 shows the corresponding steps for this method. In this algorithm, we use $\tilde{\Omega}(\boldsymbol{\theta}, \mathbf{v})$ to denote $\mathbf{G}(\boldsymbol{\theta})\Omega(\boldsymbol{\theta}, \mathbf{v})$ whose (k, j) th element is equal to $\sum_i \mathbf{v}^i \tilde{\Gamma}_{ij}^k(\boldsymbol{\theta})$.

Our proposed e-RMLMC does not involve implicit functions for updating $(\boldsymbol{\theta}, \mathbf{v})$ in RMHMC. Because we remove multiple fixed-point iteration steps, we reduce the computation time by $\mathcal{O}(D^2)$ where D is the dimension of the parameters. Additionally, using this explicit updating, we resolve the convergence issue faced by fixed-point iterations. The connection terms $\tilde{\Gamma}(\boldsymbol{\theta})$ in $\tilde{\Omega}$ do not add substantial computational cost since they are obtained from permuting three dimensions of the array $\partial\mathbf{G}(\boldsymbol{\theta})$, which is computed in RMHMC. However, besides $\mathbf{G}(\boldsymbol{\theta})^{-1}$, which is required for $\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta})$, e-RMLMC has two extra matrix inversions to update \mathbf{v} , whose complexity in general is $\mathcal{O}(D^3)$. Therefore, as dimension grows, the efficiency gained by removing multiple fixed-point iterations may be overwhelmed by this additional overhead. This is evident from our experimental results presented in Section 5. Faster matrix inversion algorithms could be used to alleviate this issue.

5 Experimental Results

In this section, we use simulated and real data to evaluate our methods, RMLMC and e-RMLMC, compared to RMHMC. Following Girolami and Calderhead (2011), we use a time-normalized effective sample size (ESS) to compare these methods. For B posterior samples we calculate $\text{ESS} = B[1 + 2\Sigma_k\gamma(k)]^{-1}$ for each parameter and choose the minimum as the measure of sampling efficiency, where $\Sigma_k\gamma(k)$ is the sum of the K monotone sample autocorrelations estimated by the initial monotone sequence estimator (Geyer, 1992). All computer programs and data sets discussed in this paper are available online at <http://www.ics.uci.edu/~babaks/Site/Codes.html>.

5.1 Logistic Regression Models

We start by evaluating our methods based on five binary classification problems used in Girolami and Calderhead (2011). These are Australian Credit data, German Credit data, Heart data, Pima Indian data, and Ripley data. For each problem, we use a logistic regression model and run 20000 MCMC iterations. Results (after discarding the initial 5000 iterations) are summarized in Table 1, and show that in general our methods improve the sampling efficiency measured in terms of ESS per second compared to RMHMC.

5.2 Simulated Logistic Regression

Next, we construct some synthetic datasets with variable numbers of observations and increasing dimensionality for logistic regression. The simulated results are summarized in Table 2. In general, our methods RMLMC and e-RMLMC improve RMHMC in minimal ESS per second, but as expected, such an advantage gradually diminishes as the dimension increases.

Data	method	AP	s	ESS	min(ESS)/s
Australian D=14,N=690	RMHMC	0.74	2.40E-02	(8561,9595,10262)	23.77
	RMLMC	0.75	1.90E-02	(8038,10488,11468)	28.24
	e-RMLMC	0.75	1.55E-02	(9636,10443,11268)	41.34
German D=24,N=1000	RMHMC	0.77	5.63E-02	(15000,15000,15000)	17.76
	RMLMC	0.73	4.32E-02	(15000,15000,15000)	23.17
	e-RMLMC	0.70	3.59E-02	(13762,15000,15000)	25.57
Heart D=13,N=270	RMHMC	0.77	1.65E-02	(7050,8369,8905)	28.57
	RMLMC	0.78	1.09E-02	(10847,11704,12405)	66.25
	e-RMLMC	0.76	1.01E-02	(10347,10724,11773)	68.18
Pima D=7,N=532	RMHMC	0.81	1.25E-02	(4325,4622,4980)	23.10
	RMLMC	0.82	7.27E-03	(4713,5448,5576)	43.20
	e-RMLMC	0.82	7.04E-03	(4839,5193,5539)	45.85
Ripley D=2,N=250	RMHMC	0.78	8.39E-03	(15000,15000,15000)	119.20
	RMLMC	0.76	5.07E-03	(13498,15000,15000)	177.43
	e-RMLMC	0.79	4.77E-03	(12611,15000,15000)	176.37

Table 1: Comparing alternative methods using five binary classification problems discussed in Girolami and Calderhead (2011). For each dataset, the number of predictors, D , and the number of observations, N , are specified. For each method, we provide the acceptance probability (AP), the CPU time (s) for each iteration, and the time-normalized ESS.

Data	Method	AP	s	ESS	min(ESS)/s
N=200,D=10	RMHMC	0.84	1.85e+02	(4837, 4902, 4968)	26.20
	RMLMC	0.86	4.44e+01	(5000, 5000, 5000)	112.60
	e-RMLMC	0.83	7.42e+01	(3792, 4310, 4671)	51.11
N=400,D=20	RMHMC	0.82	9.56e+02	(4727, 4893, 5000)	4.95
	RMLMC	0.80	1.28e+02	(4680, 4819, 4857)	36.44
	e-RMLMC	0.81	2.59e+02	(2964, 3543, 3968)	11.46
N=800,D=40	RMHMC	0.82	3.15e+03	(4691, 4983, 5000)	1.49
	RMLMC	0.82	6.88e+02	(4749, 4836, 4960)	6.91
	e-RMLMC	0.81	1.09e+03	(2902, 3636, 4127)	2.65
N=1600,D=80	RMHMC	0.81	9.87e+03	(3712, 4515, 4950)	0.38
	e-RMLMC	0.83	4.64e+03	(4002, 4672, 4919)	0.86
	e-RMLMC	0.80	1.19e+04	(2565, 3415, 4081)	0.22
N=3200,D=160	RMLMC	0.79	1.63e+05	(3160, 3959, 4464)	0.02
	RMLMC	0.83	1.44e+05	(3458, 4221, 4676)	0.02
	e-RMLMC	0.80	1.20e+05	(2708, 3548, 4156)	0.02

Table 2: Time, ESS and time-normalized ESS for logistic regression with simulated datasets. Results are calculated on a 5,000 sample chain with a 5,000 sample burn-in session.

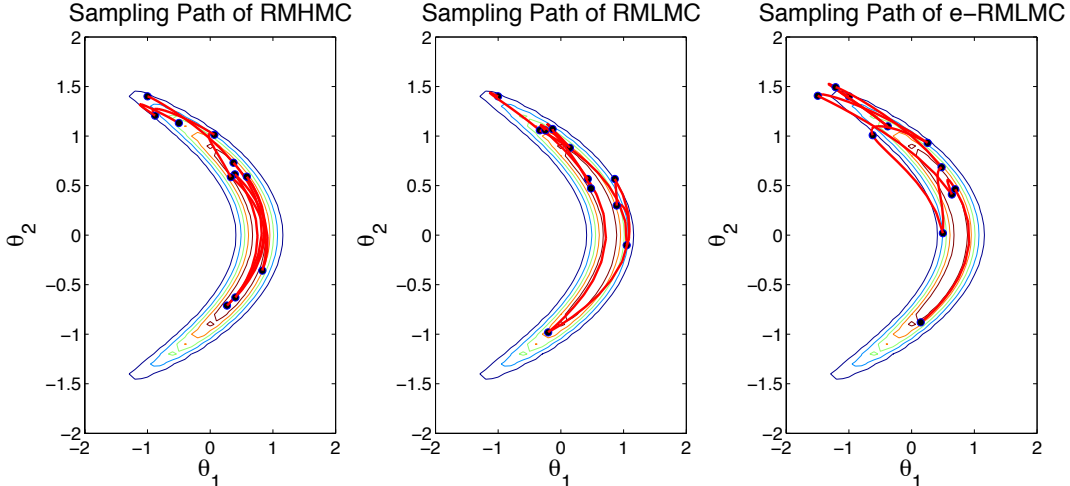


Figure 2: The first 10 iterations in sampling from the banana-shaped distribution with Riemannian Manifold HMC (RMHMC), Riemannian Manifold Lagrange Monte Carlo (RMLMC) and explicit RMLMC (e-RMLMC). For all three methods, the trajectory length (i.e., step size times number of integration steps) is set to 1.45. Solid red lines show the sampling path, and each point represents an accepted proposal.

5.3 Simulating a banana-shaped distribution

The banana-shaped distribution, which we used above for illustration, can be constructed as the posterior distribution of $\theta = (\theta_1, \theta_2)|y$ based on the following model:

$$\begin{aligned} y|\theta &\sim N(\theta_1 + \theta_2^2, \sigma_y^2) \\ \theta &\sim N(0, \sigma_\theta^2) \end{aligned}$$

The data $\{y_i\}_{i=1}^{100}$ are generated with $\theta_1 + \theta_2^2 = 1, \sigma_y = 2$. We set $\sigma_\theta = 1$.

We want to investigate how the three algorithms, RMHMC, RMLMC, e-RMLMC, explore the parameter space. Fig.2 shows the first 10 iterations for each algorithm using fixed trajectory length of 1.45. We can see that RMLMC and e-RMLMC explore the parameter space according to its curvature and mix quickly, similar to RMHMC.

Table 3 compares the performances of these algorithms based on 5000 MCMC iterations (after burning the initial 1000 iteration). For this example, RMLMC has the highest ESS/s. As discussed above, the additional overhead of matrix inversion for e-RMLMC occasionally overwhelms its gain in computational efficiency.

5.4 Finite Mixture of Gaussians

Finally we consider finite mixtures of univariate Gaussian components of the form

$$p(x_i|\boldsymbol{\theta}) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i|\mu_k, \sigma_k^2) \quad (16)$$

Method	AP	s	ESS	min(ESS)/s
RMHMC	0.73	8.21e-03	(729,1117,1506)	17.76
RMLMC	0.79	5.36e-03	(857,1317,1777)	31.99
e-RMLMC	0.78	5.61e-03	(585,1085,1585)	20.85

Table 3: Comparing alternative methods using a banana-shaped distribution. For each method, we provide the acceptance probability (AP), the CPU time (s) for each iteration, and the time-normalized ESS.

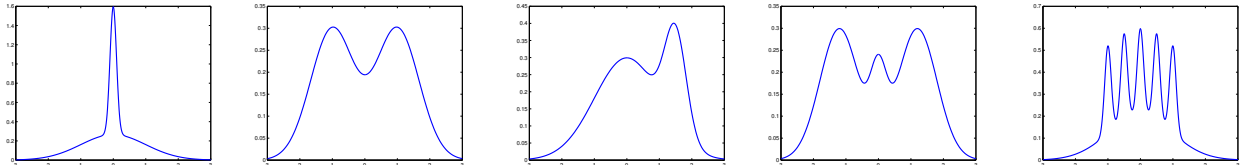


Figure 3: Densities used to generate synthetic datasets. From left to right the densities are in the same order as in Table 4. The densities are taken from McLachlan and Peel (2000)

where $\boldsymbol{\theta}$ is the vector of size $D = 3K$ of all the parameters π_k , μ_k and σ_k^2 and $\mathcal{N}(\cdot|\mu, \sigma^2)$ is a Gaussian density with mean μ and variance σ^2 . A common choice of prior takes the form

$$p(\boldsymbol{\theta}) = \mathcal{D}(\pi_1, \dots, \pi_K|\lambda) \prod_{k=1}^K \mathcal{N}(\mu_k|m, \beta^{-1}\sigma_k^2) \mathcal{IG}(\sigma_k^2|b, c) \quad (17)$$

where $\mathcal{D}(\cdot|\lambda)$ is the symmetric Dirichlet distribution with parameter λ and $\mathcal{IG}(\cdot|b, c)$ is the inverse Gamma distribution with shape parameter b and scale parameter c .

Although the posterior distribution associated with this model is formally explicit, it is computationally intractable, since it can be expressed as a sum of K^N terms corresponding to all possible allocations of observations x_i to mixture components (Marin et al, 2005, chap. 9). We want to use this model to test the efficiency of posterior sampling $\boldsymbol{\theta}$ using the three methods. A more extensive comparison of Riemannian Manifold MCMC and HMC, Gibbs sampling and standard Metropolis-Hastings for finite Gaussian mixture models can be found at Stathopoulos and Girolami (2011). Due to the non-analytic nature of the expected Fisher Information, $\mathbf{I}(\boldsymbol{\theta})$, we use the empirical Fisher information as metric tensor, defined in (McLachlan and Peel, 2000, chap. 2):

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{S}^T \mathbf{S} - \frac{1}{N} \mathbf{s} \mathbf{s}^T$$

where $N \times D$ score matrix \mathbf{S} has elements $S_{i,d} = \frac{\partial \log p(x_i|\boldsymbol{\theta})}{\partial \theta_d}$ and $\mathbf{s} = \sum_{i=1}^N \mathbf{S}_{i,\cdot}^T$.

Depending on allocations of π , we show several classical mixtures showing in the following Table 4 and Figure 3. Their sampling efficiency is compared in Table 5. As before, our two algorithms outperform RMHMC.

Dataset name	Density function	Num. of parameters
Kurtotic	$\frac{2}{3}\mathcal{N}(x 0, 1) + \frac{1}{3}\mathcal{N}\left(x 0, \left(\frac{1}{10}\right)^2\right)$	6
Bimodal	$\frac{1}{2}\mathcal{N}\left(x -1, \left(\frac{2}{3}\right)^2\right) + \frac{1}{2}\mathcal{N}\left(x 1, \left(\frac{2}{3}\right)^2\right)$	6
Skewed	$\frac{3}{4}\mathcal{N}(x 0, 1) + \frac{1}{4}\mathcal{N}\left(x \frac{3}{2}, \left(\frac{1}{3}\right)^2\right)$	6
Trimodal	$\frac{9}{20}\mathcal{N}\left(x -\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{9}{20}\mathcal{N}\left(x \frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{1}{10}\mathcal{N}\left(x 0, \left(\frac{1}{4}\right)^2\right)$	9
Claw	$\frac{1}{2}\mathcal{N}(x 0, 1) + \sum_{i=0}^4 \frac{1}{10}\mathcal{N}\left(x \frac{i}{2} - 1, \left(\frac{1}{10}\right)^2\right)$	18

Table 4: Densities used for the generation of synthetic Mixture of Gaussian data sets.

Data	Method	AP	s	ESS	min(ESS)/s
claw	RMHMC	0.80	2.54e+03	(1524, 3474, 4586)	0.60
	RMLMC	0.86	1.88e+03	(2531, 4332, 5000)	1.35
	e-RMLMC	0.82	1.46e+03	(2436, 3455, 4608)	1.67
trimodal	RMHMC	0.79	4.97e+02	(4701, 4928, 5000)	9.46
	RMLMC	0.82	2.01e+02	(4978, 5000, 5000)	24.77
	e-RMLMC	0.80	2.42e+02	(4899, 4982, 5000)	20.21
skewed	RMHMC	0.85	2.55e+02	(5000, 5000, 5000)	19.63
	RMLMC	0.82	1.13e+02	(4698, 4940, 5000)	41.68
	e-RMLMC	0.84	1.26e+02	(4935, 5000, 5000)	39.09
kurtotic	RMHMC	0.82	2.36e+02	(5000, 5000, 5000)	21.20
	RMLMC	0.85	1.27e+02	(5000, 5000, 5000)	39.34
	e-RMLMC	0.81	1.35e+02	(5000, 5000, 5000)	36.90
bimodal	RMHMC	0.86	2.69e+02	(5000, 5000, 5000)	18.56
	RMLMC	0.81	1.03e+02	(4935, 4996, 5000)	48.00
	e-RMLMC	0.85	1.08e+02	(5000, 5000, 5000)	46.43

Table 5: Time, ESS and time-normalized ESS for Gaussian mixture models. Results are calculated on a 5,000 sample chain with a 5,000 sample burn-in session.

6 Conclusions and Discussion

Following the method of Girolami and Calderhead (2011) for more efficient exploration of parameter space, we have proposed new sampling schemes to reduce the computational cost associated with using a position-specific mass matrix. To this end, we have developed a semi-explicit (RMLMC) integrator and a fully explicit (e-RMLMC) integrator for RMHMC and demonstrated their advantage in improving computational efficiency over the generalized leapfrog (RMHMC) method used by Girolami and Calderhead (2011). It is easy to show that for $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{I}$, our method degenerates to standard HMC.

Future directions could involve splitting Hamiltonian (Dullweber et al., 1987; Sexton and Wein-

garten, 1992; Neal, 2010; Shahbaba et al., 2011) to develop explicit geometric integrators. For example, one could split a non-separable Hamiltonian dynamics into several smaller dynamics some of which can be analytically solved. A similar idea has been explored by (Chin, 2009), where the Hamiltonian, instead of the dynamic, is split.

Another possible research direction could be to approximate the mass matrix (Christoffel Symbols). For many large-dimensional problems, the mass matrix could be appropriately approximated by a highly sparse matrix. This could further improve our method's computational efficiency.

References

- Amari, S. and H. Nagaoka (2000). *Methods of Information Geometry*. Oxford University Press.
- Bishop, Richard L. and Samuel I. Goldberg (2000). *Tensor Analysis on Manifolds*. Dover Publications, Inc.
- Chin, S. A. (2009). Explicit symplectic integrators for solving nonseparable Hamiltonians. *Physical Review E* 80, 037701.
- Christensen, O.F., G.O. Roberts, J.S. Rosenthal (2005, April). Scaling Limits for the Transient Phase of Local Metropolis-Hastings Algorithms. *Journal of the Royal Statistical Society: Series B* 67(2), 253–268.
- Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth (1987). Hybrid monte carlo. *Physics Letters B* 195(2), 216 – 222.
- Dullweber, A., B. Leimkuhler, R. McLachlan (1997). Split-Hamiltonian Methods for Rigid Body Molecular Dynamics. *Journal of Chemical Physics* 107(15), 5840 – 5852.
- Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical Science* 7(4), 473–483.
- Girolami, M. and Calderhead, B. (2011), Riemann manifold Langevin and Hamiltonian Monte Carlo methods, *Journal of the Royal Statistical Society, Series B*, (with discussion) 73, 123–214.
- Hoffman, M. D. and A. Gelman (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. arXiv:1111.4246v1.
- Leimkuhler, B. and S. Reich (2004). *Simulating Hamiltonian Dynamics*. Cambridge University Press.
- Liu, J. S. (2001). Molecular Dynamics and Hybrid Monte Carlo. In Liu, Jun S., *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag.
- Marin, J.M., K.L. Mengersen and C. Robert (2005). Bayesian modeling and inference on mixtures of distributions. In D. Dey and C.R. Rao, *Handbook of Statistics Volume 25*. Elsevier.
- McLachlan, G.J. and D. Peel (2000). *Finite Mixture Models*. John Wiley & Sons, Inc., New York.

- Neal, R. M. (2010). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, and X. L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.
- Sexton, J. C. and D.H. Weingarten (1992). Hamiltonian evolution for the hybrid Monte Carlo algorithm. *Nuclear Physics B* 380(3), 665–677.
- Shahbaba, B., S. Lan, W. Johnson, and R. Neal (2011). Split hamiltonian monte carlo. arXiv:1106.5941.
- Stathopoulos, V. and Girolami, M. (2011) Manifold MCMC for Mixtures. in *K. Mengersen, C. P. Robert and M. D. Titteringhton, Mixture Estimation and Applications* John Wiley & Sons
- Verlet, Loup (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* 159, 98–103.

Appendix: Derivations and Proofs

In what follows, we show the detailed derivations of our methods. We adopt Einstein notation (summation convention), so whenever the index appears twice in a mathematical expression, we sum over it: e.g., $a_i b^i := \sum_i a_i b^i$, $\Gamma_{ij}^k v^i v^j := \sum_{i,j} \Gamma_{ij}^k v^i v^j$. A lower index is used for the covariant tensor, whose components vary by the same transformation as the change of basis, whereas the upper index is reserved for the contravariant tensor, whose components vary in the opposite way as the change of basis in order to compensate: e.g. velocity vector. Interested readers should refer to Bishop and Goldberg (1980).

A Transformation of Hamiltonian Dynamics

To derive the dynamic (8) from the Hamiltonian dynamic (3), the first equation in (8) is directly obtained from the assumed transformation: $\dot{\theta}^k = \mathbf{g}^{kl} \mathbf{p}_l = \mathbf{v}^k$. For the second equation in (8), we have

$$\dot{\mathbf{p}}_l = \frac{d(\mathbf{g}_{lj}(\boldsymbol{\theta}) \mathbf{v}^j)}{dt} = \frac{\partial \mathbf{g}_{lj}}{\partial \theta_i} \dot{\theta}^i \mathbf{v}^j + \mathbf{g}_{lj} \dot{\mathbf{v}}^j = \partial_i \mathbf{g}_{lj} \mathbf{v}^i \mathbf{v}^j + \mathbf{g}_{lj} \dot{\mathbf{v}}^j$$

Further, from Equation (3) we have

$$\begin{aligned} \dot{\mathbf{p}}_l &= -\partial_l \phi(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{v}^\top \partial_l \mathbf{G}(\boldsymbol{\theta}) \mathbf{v} = -\partial_l \phi + \frac{1}{2} \mathbf{g}_{ij,l} \mathbf{v}^i \mathbf{v}^j \\ &= \partial_i \mathbf{g}_{lj} \mathbf{v}^i \mathbf{v}^j + \mathbf{g}_{lj} \dot{\mathbf{v}}^j \end{aligned}$$

which means

$$\mathbf{g}_{lj} \dot{\mathbf{v}}^j = -(\partial_i \mathbf{g}_{lj} - \frac{1}{2} \partial_l \mathbf{g}_{ij}) \mathbf{v}^i \mathbf{v}^j - \partial_l \phi$$

By multiplying $\mathbf{G}^{-1} = (\mathbf{g}^{kl})$ on both sides, we have

$$\dot{\mathbf{v}}^k = \delta_j^k \dot{\mathbf{v}}^j = -\mathbf{g}^{kl} (\partial_i \mathbf{g}_{lj} - \frac{1}{2} \partial_l \mathbf{g}_{ij}) \mathbf{v}^i \mathbf{v}^j - \mathbf{g}^{kl} \partial_l \phi \quad (18)$$

Since i, j are symmetric in the first summand, switching them gives the following equations:

$$\dot{\mathbf{v}}^k = -\mathbf{g}^{kl} (\partial_j \mathbf{g}_{li} - \frac{1}{2} \partial_l \mathbf{g}_{ji}) \mathbf{v}^i \mathbf{v}^j - \mathbf{g}^{kl} \partial_l \phi \quad (19)$$

which in turn gives the final form of Equation (8) after adding equations (18) and (19) and dividing the results by two:

$$\dot{\mathbf{v}}^k = -\Gamma_{ij}^k(\boldsymbol{\theta}) \mathbf{v}^i \mathbf{v}^j - \mathbf{g}^{kl}(\boldsymbol{\theta}) \partial_l \phi(\boldsymbol{\theta})$$

Here, $\Gamma_{ij}^k(\boldsymbol{\theta}) := \frac{1}{2} \mathbf{g}^{kl} (\partial_i \mathbf{g}_{lj} + \partial_j \mathbf{g}_{il} - \partial_l \mathbf{g}_{ij})$ is Christoffel Symbol of second kind.

Note that the new dynamic (8) still preserves the original Hamiltonian $H(\boldsymbol{\theta}, \mathbf{p} = \mathbf{G}(\boldsymbol{\theta})\mathbf{v})$. This is of course intuitive, but it also can be proven as follows:

$$\begin{aligned}
\frac{d}{dt}H(\boldsymbol{\theta}, \mathbf{G}(\boldsymbol{\theta})\mathbf{v}) &= \dot{\boldsymbol{\theta}}^\top \frac{\partial}{\partial \boldsymbol{\theta}} H(\boldsymbol{\theta}, \mathbf{G}(\boldsymbol{\theta})\mathbf{v}) + \dot{\mathbf{v}}^\top \frac{\partial}{\partial \mathbf{v}} H(\boldsymbol{\theta}, \mathbf{G}(\boldsymbol{\theta})\mathbf{v}) \\
&= \mathbf{v}^\top \left[\nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{v}^\top \partial \mathbf{G}(\boldsymbol{\theta}) \mathbf{v} \right] + [-\mathbf{v}^\top \Gamma(\boldsymbol{\theta}) \mathbf{v} - \mathbf{G}(\boldsymbol{\theta})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta})]^\top \mathbf{G}(\boldsymbol{\theta}) \mathbf{v} \\
&= \mathbf{v}^\top \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) - (\nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}))^\top \mathbf{v} + \frac{1}{2} \mathbf{v}^\top (\mathbf{v}^\top \partial \mathbf{G}(\boldsymbol{\theta}) \mathbf{v}) - (\mathbf{v}^\top \tilde{\Gamma}(\boldsymbol{\theta}) \mathbf{v})^\top \mathbf{v} \\
&= 0 + 0 = 0
\end{aligned}$$

where $\mathbf{v}^\top \Gamma(\boldsymbol{\theta}) \mathbf{v}$ is a vector whose k th element is $\Gamma_{ij}^k(\boldsymbol{\theta}) \mathbf{v}^i \mathbf{v}^j$. The second 0 is due to the triple form $(\mathbf{v}^\top \tilde{\Gamma}(\boldsymbol{\theta}) \mathbf{v})^\top \mathbf{v} = \tilde{\Gamma}_{ijk} \mathbf{v}^i \mathbf{v}^j \mathbf{v}^k = \frac{1}{2} \partial_k \mathbf{g}_{ij} \mathbf{v}^i \mathbf{v}^j \mathbf{v}^k$, where $\tilde{\Gamma}$ is Christoffel Symbol of first kind with elements $\tilde{\Gamma}_{ijk}(\boldsymbol{\theta}) := \mathbf{g}_{kl} \Gamma_{ij}^l(\boldsymbol{\theta}) = \frac{1}{2} (\partial_i \mathbf{g}_{kj} + \partial_j \mathbf{g}_{ik} - \partial_k \mathbf{g}_{ij})$.

B Derivation of semi-explicit Riemannian Manifold Lagrangian Monte Carlo (RMLMC)

Consider the following generalized leapfrog integration scheme:

$$\begin{aligned}
\mathbf{p}^{(n+1/2)} &= \mathbf{p}^{(n)} - \frac{\varepsilon}{2} \frac{\partial H}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^{(n)}, \mathbf{p}^{(n+1/2)}) \\
\boldsymbol{\theta}^{(n+1)} &= \boldsymbol{\theta}^{(n)} + \frac{\varepsilon}{2} \left[\frac{\partial H}{\partial \mathbf{p}}(\boldsymbol{\theta}^{(n)}, \mathbf{p}^{(n+1/2)}) + \frac{\partial H}{\partial \mathbf{p}}(\boldsymbol{\theta}^{(n+1)}, \mathbf{p}^{(n+1/2)}) \right] \\
\mathbf{p}^{(n+1)} &= \mathbf{p}^{(n+1/2)} - \frac{\varepsilon}{2} \frac{\partial H}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^{(n)}, \mathbf{p}^{(n+1/2)})
\end{aligned}$$

We composite implicit steps for velocity and explicit step for position within a leapfrog step to integrate dynamic (8) and derive the following semi-explicit integrator:

$$\mathbf{v}^{(n+1/2)} = \mathbf{v}^{(n)} - \frac{\varepsilon}{2} [(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)}) \mathbf{v}^{(n+1/2)} + \mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n)})] \quad (20)$$

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \varepsilon \mathbf{v}^{(n+1/2)} \quad (21)$$

$$\mathbf{v}^{(n+1)} = \mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2} [(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)}) \mathbf{v}^{(n+1/2)} + \mathbf{G}(\boldsymbol{\theta}^{(n+1)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n+1)})] \quad (22)$$

The time-reversibility of this integrator can be shown by switching $(\boldsymbol{\theta}, \mathbf{v})^{(n+1)}$ and $(\boldsymbol{\theta}, \mathbf{v})^{(n)}$ and negating velocity. The resulting integrator, however, is no longer volume-preserving (see subsection B.1). Nevertheless, based on proposition 1, we can still have detailed balance after determinant adjustment. (See Liu, 2001, for more details.)

Proposition 1 (Detailed Balance Condition with determinant adjustment). *Denote $\mathbf{z} = (\boldsymbol{\theta}, \mathbf{p})$, $\mathbf{z}' = \hat{T}_L(\mathbf{z})$ for some time reversible integrator \hat{T}_L to the Lagrangian dynamic. If the acceptance probability is adjusted in the following way:*

$$\tilde{\alpha}(\mathbf{z}, \mathbf{z}') = \min \left\{ 1, \frac{\exp(-H(\mathbf{z}'))}{\exp(-H(\mathbf{z}))} |\det \hat{T}_L| \right\}$$

then the detailed balance condition still holds

$$\tilde{\alpha}(\mathbf{z}, \mathbf{z}')\mathbb{P}(d\mathbf{z}) = \tilde{\alpha}(\mathbf{z}', \mathbf{z})\mathbb{P}(d\mathbf{z}')$$

Proof.

$$\begin{aligned} \tilde{\alpha}(\mathbf{z}, \mathbf{z}')\mathbb{P}(d\mathbf{z}) &= \min \left\{ 1, \frac{\exp(-H(\mathbf{z}'))}{\exp(-H(\mathbf{z}))} \left| \frac{d\mathbf{z}'}{d\mathbf{z}} \right| \right\} \exp(-H(\mathbf{z}))d\mathbf{z} \\ &\stackrel{\mathbf{z}=\hat{T}_L^{-1}(\mathbf{z}')}{=} \min \left\{ \exp(-H(\mathbf{z})), \exp(-H(\mathbf{z}')) \left| \frac{d\mathbf{z}'}{d\mathbf{z}} \right| \right\} \left| \frac{d\mathbf{z}}{d\mathbf{z}'} \right| d\mathbf{z}' \\ &= \min \left\{ 1, \frac{\exp(-H(\mathbf{z}))}{\exp(-H(\mathbf{z}'))} \left| \frac{d\mathbf{z}}{d\mathbf{z}'} \right| \right\} \exp(-H(\mathbf{z}'))d\mathbf{z}' = \tilde{\alpha}(\mathbf{z}', \mathbf{z})\mathbb{P}(d\mathbf{z}') \end{aligned}$$

□

□

Therefore, the acceptance probability could be calculated based on $H(\boldsymbol{\theta}, \mathbf{G}(\boldsymbol{\theta})\mathbf{v})$. However, it also could be also calculated as follows based on the energy function $\mathbf{E}(\boldsymbol{\theta}, \mathbf{v})$ defined in section 3,

$$\mathbf{E}(\boldsymbol{\theta}, \mathbf{v}) = U(\boldsymbol{\theta}) + K(\boldsymbol{\theta}, \mathbf{v}) = -\log p(\boldsymbol{\theta}) - \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{v}^\top \mathbf{G}(\boldsymbol{\theta})\mathbf{v} \quad (23)$$

To show their equivalence, we note that $\left| \frac{\partial(\boldsymbol{\theta}', \mathbf{p}')}{\partial(\boldsymbol{\theta}, \mathbf{p})} \right| = \frac{\det(\mathbf{G}(\boldsymbol{\theta}'))}{\det(\mathbf{G}(\boldsymbol{\theta}))} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{v}')}{\partial(\boldsymbol{\theta}, \mathbf{v})} \right|$ and proved our claim as follows:

$$\begin{aligned} \tilde{\alpha} &= \min \left\{ 1, \frac{\exp(-H(\boldsymbol{\theta}', \mathbf{p}'))}{\exp(-H(\boldsymbol{\theta}, \mathbf{p}))} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{p}')}{\partial(\boldsymbol{\theta}, \mathbf{p})} \right| \right\} = \min \left\{ 1, \frac{\exp(-H(\boldsymbol{\theta}', \mathbf{G}(\boldsymbol{\theta}')\mathbf{v}'))}{\exp(-H(\boldsymbol{\theta}, \mathbf{G}(\boldsymbol{\theta})\mathbf{v}))} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{p}')}{\partial(\boldsymbol{\theta}, \mathbf{p})} \right| \right\} \\ &= \min \left\{ 1, \frac{\exp\{-\log p(\boldsymbol{\theta}') + \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta}') + \frac{1}{2} \mathbf{v}'^\top \mathbf{G}(\boldsymbol{\theta}')\mathbf{v}'\}}{\exp\{-\log p(\boldsymbol{\theta}) + \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{v}^\top \mathbf{G}(\boldsymbol{\theta})\mathbf{v}\}} \frac{\det(\mathbf{G}(\boldsymbol{\theta}'))}{\det(\mathbf{G}(\boldsymbol{\theta}))} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{v}')}{\partial(\boldsymbol{\theta}, \mathbf{v})} \right| \right\} \\ &= \min \left\{ 1, \frac{\exp\{-\log p(\boldsymbol{\theta}') - \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta}') + \frac{1}{2} \mathbf{v}'^\top \mathbf{G}(\boldsymbol{\theta}')\mathbf{v}'\}}{\exp\{-\log p(\boldsymbol{\theta}) - \frac{1}{2} \log \det \mathbf{G}(\boldsymbol{\theta}) + \frac{1}{2} \mathbf{v}^\top \mathbf{G}(\boldsymbol{\theta})\mathbf{v}\}} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{v}')}{\partial(\boldsymbol{\theta}, \mathbf{v})} \right| \right\} \\ &= \min \left\{ 1, \frac{\exp(-\mathbf{E}(\boldsymbol{\theta}', \mathbf{v}'))}{\exp(-\mathbf{E}(\boldsymbol{\theta}, \mathbf{v}))} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{v}')}{\partial(\boldsymbol{\theta}, \mathbf{v})} \right| \right\} \end{aligned}$$

B.1 Volume Correction

To adjust volume, we must derive the Jacobian determinant, $\det \mathbf{J} := \left| \frac{\partial(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})}{\partial(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})} \right|$ using wedge products.

Definition 1 (Differential Forms, Wedge Product). *Differential one-form* $\alpha : TM^D \rightarrow \mathbb{R}$ on a differential manifold M^D is a smooth mapping from tangent space TM^D to \mathbb{R} , which can be expressed as linear combination of differentials of local coordinates: $\alpha = f_i dx^i =: f \cdot dx$.

For example, if $f : \mathbb{R}^D \rightarrow \mathbb{R}$ is a smooth function, then its directional derivative along a vector $v \in \mathbb{R}^D$, denoted by $df(v)$ is given by

$$df(v) = \frac{\partial f}{\partial z_i} v^i$$

then $df(\cdot)$ is a linear functional of v , called the differential of f at z and is an example of a differential one-form. In particular, $dz^i(v) = v^i$, thus

$$df(v) = \frac{\partial f}{\partial z_i} dz^i(v), \quad \text{then } df = \frac{\partial f}{\partial z_i} dz^i$$

Wedge Product of two one-form α, β is a 2-form $\alpha \wedge \beta$ anti-symmetric bilinear function on tangent space which has the following properties (α, β, γ one-forms, A be a square matrix of same dimension D):

- $\alpha \wedge \alpha = 0$
- $\alpha \wedge (\beta + \gamma) = \alpha \wedge \beta + \alpha \wedge \gamma$ (thus $\alpha \wedge \beta = -\beta \wedge \alpha$)
- $\alpha \wedge A\beta = A^T \alpha \wedge \beta$

The following proposition enables us to calculate the Jacobian determinant denoted as $\det \mathbf{J}$.

Proposition 2. Let $T_L : (\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}) \rightarrow (\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})$ be evolution of a smooth flow, then

$$d\boldsymbol{\theta}^{(L+1)} \wedge d\mathbf{v}^{(L+1)} = \frac{\partial(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})}{\partial(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})} d\boldsymbol{\theta}^{(1)} \wedge d\mathbf{v}^{(1)}$$

Note that the Jacobian determinant $\det \mathbf{J}$ can also be regarded as Radon-Nikodym derivative of two probability measures: $\det \mathbf{J} = \frac{\mathbb{P}(d\boldsymbol{\theta}^{(L+1)}, d\mathbf{v}^{(L+1)})}{\mathbb{P}(d\boldsymbol{\theta}^{(1)}, d\mathbf{v}^{(1)})}$, where $\mathbb{P}(d\boldsymbol{\theta}, d\mathbf{v}) = p(\boldsymbol{\theta}, \mathbf{v}) d\boldsymbol{\theta} d\mathbf{v}$. We have

$$\begin{aligned} d\mathbf{v}^{(n+1/2)} &= d\mathbf{v}^{(n)} - \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)}) d\mathbf{v}^{(n+1/2)} + (**) d\boldsymbol{\theta}^{(n)} \\ d\boldsymbol{\theta}^{(n+1)} &= d\boldsymbol{\theta}^{(n)} + \varepsilon d\mathbf{v}^{(n+1/2)} \\ d\mathbf{v}^{(n+1)} &= d\mathbf{v}^{(n+1/2)} - \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)}) d\mathbf{v}^{(n+1/2)} + (**) d\boldsymbol{\theta}^{(n+1)} \end{aligned}$$

where $\mathbf{v}^\top \Gamma(\boldsymbol{\theta})$ is a matrix whose (k, j) th element is $\mathbf{v}^i \Gamma_{ij}^k(\boldsymbol{\theta})$. Therefore,

$$\begin{aligned} d\boldsymbol{\theta}^{(n+1)} \wedge d\mathbf{v}^{(n+1)} &= [\mathbf{I} - \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})]^\top d\boldsymbol{\theta}^{(n+1)} \wedge d\mathbf{v}^{(n+1/2)} \\ &= [\mathbf{I} - \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})]^\top d\boldsymbol{\theta}^{(n)} \wedge d\mathbf{v}^{(n+1/2)} \\ &= [\mathbf{I} - \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})]^\top [\mathbf{I} + \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-T} d\boldsymbol{\theta}^{(n)} \wedge d\mathbf{v}^{(n)} \end{aligned}$$

For volume adjustment, we must use the following Jacobian determinant accumulated along leap frog steps:

$$\det \mathbf{J}_{LMC} := \left| \frac{\partial(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})}{\partial(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})} \right| = \prod_{n=1}^L \frac{\det(\mathbf{I} - \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)}))}{\det(\mathbf{I} + \varepsilon(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)}))} \quad (24)$$

As a result, the acceptance probability becomes

$$\alpha_{LMC} = \min\{1, \exp(-\mathbf{E}(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)}) + \mathbf{E}(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})) |\det \mathbf{J}_{LMC}|\}$$

Using this acceptance probability, we are able to derive a semi-explicit integrator for RMLMC as shown in Algorithm 1. In this approach, the updates for $\boldsymbol{\theta}$ are explicit, while updating \mathbf{v} remains implicit.

C Connection to Lagrangian Dynamics

We now show that the above dynamic (8) is indeed *Lagrangian* dynamic. We define Lagrangian as follows:

$$\mathbf{L} = \frac{1}{2} \mathbf{v}^\top \mathbf{G}(\boldsymbol{\theta}) \mathbf{v} - \phi(\boldsymbol{\theta})$$

Using variation calculus to minimize the Lagrangian, we obtain a Euler-Lagrange equation of the second kind,

$$\frac{\partial \mathbf{L}}{\partial \boldsymbol{\theta}} = \frac{d}{dt} \frac{\partial \mathbf{L}}{\partial \dot{\boldsymbol{\theta}}}$$

which is

$$\ddot{\boldsymbol{\theta}} = -\dot{\boldsymbol{\theta}}^\top \Gamma(\boldsymbol{\theta}) \mathbf{v} - \mathbf{G}(\boldsymbol{\theta})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) \quad (25)$$

This is equivalent to the new dynamic (8) by taking the time derivative on the first equation and equating it with the time derivative on the second.

D Derivation of explicit Riemannian Manifold Lagrangian Monte Carlo (e-RMLMC)

We now propose an additional modification of Algorithm 1 to resolve the remaining implicit equation (20), while keeping time-reversibility to ensure the ergodicity of the induced Markov chain. We do this by modifying the symmetric quadratic form in equations (20),

$$\mathbf{v}^{(n+1/2)} = \mathbf{v}^{(n)} - \frac{\varepsilon}{2} [(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)}) \mathbf{v}^{(n+1/2)} + \mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n)})] \quad (26)$$

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \varepsilon \mathbf{v}^{(n+1/2)} \quad (27)$$

$$\mathbf{v}^{(n+1)} = \mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2} [(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)}) \mathbf{v}^{(n+1)} + \mathbf{G}(\boldsymbol{\theta}^{(n+1)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n+1)})] \quad (28)$$

The resulting integrator is completely explicit since both updates of velocity (26) and (28) can be solved as follows:

$$\mathbf{v}^{(n+1/2)} = [\mathbf{I} + \frac{\varepsilon}{2} (\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-1} [\mathbf{v}^{(n)} - \frac{\varepsilon}{2} \mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n)})] \quad (29)$$

$$\mathbf{v}^{(n+1)} = [\mathbf{I} + \frac{\varepsilon}{2} (\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})]^{-1} [\mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2} \mathbf{G}(\boldsymbol{\theta}^{(n+1)})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}^{(n+1)})] \quad (30)$$

D.1 Convergence of Numerical Solution

We now show that the discretization error $e_n = \|\mathbf{z}(t_n) - \mathbf{z}^{(n)}\| = \|(\boldsymbol{\theta}(t_n), \mathbf{v}(t_n)) - (\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})\|$ (i.e. the difference between the true solution and the numerical solution) accumulated over final time interval $[0, T]$, is bounded and goes to zeros as the stepsize ε goes to zero. (See Leimkuhler and Reich (2004) for a similar proof for the generalized leapfrog method.) Here, we assume that $F(\boldsymbol{\theta}, \mathbf{v}) := \mathbf{v}^\top \Gamma(\boldsymbol{\theta}) \mathbf{v} + \mathbf{G}(\boldsymbol{\theta})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta})$ is smooth; hence, F and its derivatives are uniformly

bounded as $(\boldsymbol{\theta}, \mathbf{v})$ evolves within finite time duration T . We expand the true solution $\mathbf{z}(t_{n+1})$ at t_n :

$$\begin{aligned}\mathbf{z}(t_{n+1}) &= \mathbf{z}(t_n) + \dot{\mathbf{z}}(t_n)\varepsilon + \frac{1}{2}\ddot{\mathbf{z}}(t_n)\varepsilon^2 + o(\varepsilon) \\ &= \begin{bmatrix} \boldsymbol{\theta}(t_n) \\ \mathbf{v}(t_n) \end{bmatrix} + \begin{bmatrix} \mathbf{v}(t_n) \\ -F(\boldsymbol{\theta}(t_n), \mathbf{v}(t_n)) \end{bmatrix} \varepsilon + \frac{1}{2} \begin{bmatrix} -F(\boldsymbol{\theta}(t_n), \mathbf{v}(t_n)) \\ -\frac{\partial F}{\partial \boldsymbol{\theta}} \mathbf{v}(t_n) + \frac{\partial F}{\partial \mathbf{v}} F(\boldsymbol{\theta}(t_n), \mathbf{v}(t_n)) \end{bmatrix} \varepsilon^2 + o(\varepsilon) \\ &= \begin{bmatrix} \boldsymbol{\theta}(t_n) \\ \mathbf{v}(t_n) \end{bmatrix} + \begin{bmatrix} \mathbf{v}(t_n) \\ -F(\boldsymbol{\theta}(t_n), \mathbf{v}(t_n)) \end{bmatrix} \varepsilon + O(\varepsilon^2)\end{aligned}$$

Next, we simplify the expression of the numerical solutions $\mathbf{z}^{(n+1)} = \begin{bmatrix} \boldsymbol{\theta}^{(n+1)} \\ \mathbf{v}^{(n+1)} \end{bmatrix}$ for the fully explicit integrator and compare it to the above true solutions. To this end, we rewrite equation (29) as follows:

$$\begin{aligned}\mathbf{v}^{(n+1/2)} &= [\mathbf{I} + \frac{\varepsilon}{2}(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-1}[\mathbf{v}^{(n)} - \frac{\varepsilon}{2}\mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}^{(n)})] \\ &= \mathbf{v}^{(n)} - [\mathbf{I} + \frac{\varepsilon}{2}(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-1}[(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})\mathbf{v}^{(n)} + \frac{\varepsilon}{2}\mathbf{G}(\boldsymbol{\theta}^{(n)})^{-1}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}^{(n)})] \\ &= \mathbf{v}^{(n)} - [\mathbf{I} + \frac{\varepsilon}{2}(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-1}\frac{\varepsilon}{2}F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) \\ &= \mathbf{v}^{(n)} - \frac{\varepsilon}{2}F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) + \frac{\varepsilon^2}{4}[\mathbf{I} + \frac{\varepsilon}{2}(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-1}[(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) \\ &= \mathbf{v}^{(n)} - \frac{\varepsilon}{2}F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) + O(\varepsilon^2)\end{aligned}$$

Similarly, from equation (30) we have

$$\mathbf{v}^{(n+1)} = \mathbf{v}^{(n+1/2)} - \frac{\varepsilon}{2}F(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n+1/2)}) + O(\varepsilon^2)$$

Substituting $\mathbf{v}^{(n+1/2)}$ in the above equation, we obtain $\mathbf{v}^{(n+1)}$ as follows:

$$\begin{aligned}\mathbf{v}^{(n+1)} &= \mathbf{v}^{(n)} - \frac{\varepsilon}{2}F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) - \frac{\varepsilon}{2}F(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n)}) + O(\varepsilon^2) \\ &= \mathbf{v}^{(n)} - F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})\varepsilon + \frac{\varepsilon}{2}[F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) - F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})] + O(\varepsilon^2) \\ &= \mathbf{v}^{(n)} - F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})\varepsilon + O(\varepsilon^2)\end{aligned}$$

From (29), (27), and (30), we have the following numerical solution:

$$\mathbf{z}^{(n+1)} = \begin{bmatrix} \boldsymbol{\theta}^{(n+1)} \\ \mathbf{v}^{(n+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(n)} \\ \mathbf{v}^{(n)} \end{bmatrix} + \begin{bmatrix} \mathbf{v}^{(n)} \\ -F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)}) \end{bmatrix} \varepsilon + O(\varepsilon^2)$$

Therefore, the local error is

$$\begin{aligned}e_{n+1} &= \|\mathbf{z}(t_{n+1}) - \mathbf{z}^{(n+1)}\| = \left\| \begin{bmatrix} \boldsymbol{\theta}(t_n) - \boldsymbol{\theta}^{(n)} \\ \mathbf{v}(t_n) - \mathbf{v}^{(n)} \end{bmatrix} + \begin{bmatrix} \mathbf{v}(t_n) - \mathbf{v}^{(n)} \\ -[F(\boldsymbol{\theta}(t_n), \mathbf{v}(t_n)) - F(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})] \end{bmatrix} \varepsilon + O(\varepsilon^2) \right\| \\ &\leq (1 + M\varepsilon)e_n + O(\varepsilon^2)\end{aligned}$$

where $M = c \sup_{t \in [0, T]} \|\nabla F(\boldsymbol{\theta}(t), \mathbf{v}(t))\|$ for some constant $c > 0$. Accumulating the local errors by iterating the above inequality for $L = T/\varepsilon$ steps provides the following global error:

$$\begin{aligned} e_{n+1} &\leq (1 + M\varepsilon)e_n + O(\varepsilon^2) \leq (1 + M\varepsilon)^2 e_{n-1} + 2O(\varepsilon^2) \leq \dots \leq (1 + M\varepsilon)^n e_1 + nO(\varepsilon^2) \\ &\leq (1 + M\varepsilon)^L \varepsilon + LO(\varepsilon^2) \leq (e^{MT} + T)\varepsilon \rightarrow 0, \quad \text{as } \varepsilon \rightarrow 0 \end{aligned}$$

D.2 Volume Correction

As before, using wedge product calculation on the system (29), (27), and (30), the Jacobian matrix is

$$\begin{aligned} \frac{\partial(\boldsymbol{\theta}^{(n+1)}, \mathbf{v}^{(n+1)})}{\partial(\boldsymbol{\theta}^{(n)}, \mathbf{v}^{(n)})} &= [\mathbf{I} + \frac{\varepsilon}{2}(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})]^{-T} [\mathbf{I} - \frac{\varepsilon}{2}(\mathbf{v}^{(n+1)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})]^\top \\ &\quad [\mathbf{I} + \frac{\varepsilon}{2}(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^{-T} [\mathbf{I} - \frac{\varepsilon}{2}(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)})]^\top \end{aligned}$$

As these new equations show, our derived integrator is not symplectic so the acceptance probability needs to be adjusted by the following Jacobian determinant, $\det \mathbf{J}$, in order to preserve the detailed balance condition:

$$\begin{aligned} \det \mathbf{J}_{e-LMC} &:= \left| \frac{\partial(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)})}{\partial(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})} \right| \\ &= \prod_{n=1}^L \frac{\det(I - \varepsilon/2(\mathbf{v}^{(n+1)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})) \det(I - \varepsilon/2(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n)}))}{\det(I + \varepsilon/2(\mathbf{v}^{(n+1/2)})^\top \Gamma(\boldsymbol{\theta}^{(n+1)})) \det(I + \varepsilon/2(\mathbf{v}^{(n)})^\top \Gamma(\boldsymbol{\theta}^{(n)}))} \\ &= \prod_{n=1}^L \frac{\det(\mathbf{G}(\boldsymbol{\theta}^{(n+1)}) - \varepsilon/2(\mathbf{v}^{(n+1)})^\top \tilde{\Gamma}(\boldsymbol{\theta}^{(n+1)})) \det(\mathbf{G}(\boldsymbol{\theta}^{(n)}) - \varepsilon/2(\mathbf{v}^{(n+1/2)})^\top \tilde{\Gamma}(\boldsymbol{\theta}^{(n)}))}{\det(\mathbf{G}(\boldsymbol{\theta}^{(n+1)}) + \varepsilon/2(\mathbf{v}^{(n+1/2)})^\top \tilde{\Gamma}(\boldsymbol{\theta}^{(n+1)})) \det(\mathbf{G}(\boldsymbol{\theta}^{(n)}) + \varepsilon/2(\mathbf{v}^{(n)})^\top \tilde{\Gamma}(\boldsymbol{\theta}^{(n)}))} \end{aligned} \quad (31)$$

As a result, the acceptance probability is

$$\alpha_{e-LMC} = \min\{1, \exp(-\mathbf{E}(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)}) + \mathbf{E}(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})) |\det \mathbf{J}_{e-LMC}|\}$$

We can now derive a completely explicit integrator for RMLMC defined in terms of $(\boldsymbol{\theta}, \mathbf{v})$. We refer to this integrator as e-RMLMC for which the corresponding steps are presented in Algorithm 2. In both algorithms 1 and 2, the position update is relatively simple while the computational time is dominated by choosing the “right” direction (velocity) using the geometry of parameter space. Finally, it is easy to show that for $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{I}$, our method degenerates to standard HMC.