Exponential Integration for Hamiltonian Monte Carlo

Wei-Lun Chao[1], Justin Solomon[2,3], Dominik L. Michels[3], and Fei Sha[1]

[1]University of Southern California, [2]Princeton University, [3]Stanford University

This work is supported by ARO Award # W911NF-12-1-0241, ONR # N000141210066, Alfred P. Sloan Fellowship, and the Max Planck Center for Visual Computing and Communication.

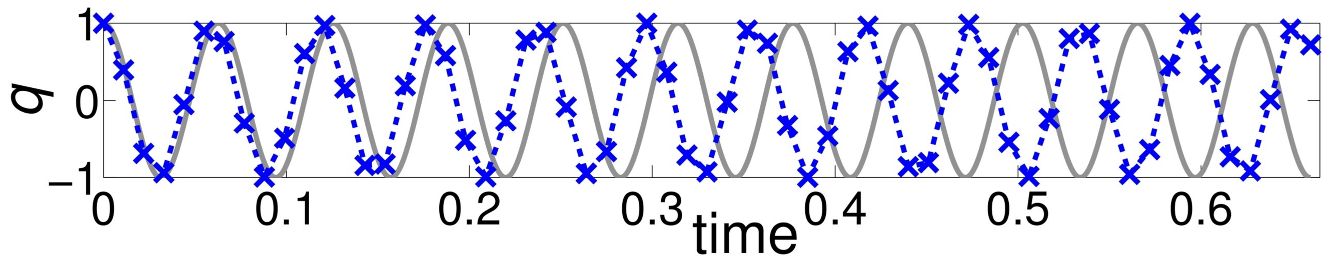
Highlights

- **Investigate numerical integration of ODEs for Hamiltonian Monte Carlo (HMC)**
- **Propose the use of a *exponential integrator*, which is more robust to stiff ODEs than the standard leapfrog integrator**
- **Generate effective samples with high acceptance in short time**

Goal

- Improve the technique of integration
  - ➢ Fast step, large h w/o tuning M
- Why does leapfrog need small h?
  - ➢ High-frequency (stiff) components of ODE

$$\ddot{q} + 10^4 \times q = 0 \;\Rightarrow\; q(t) = \cos(10^2 \times t)$$



  - ➢ Stiffness: distributions peak around modes

Algorithm

Leapfrog HMC (previous)	Exponential HMC (proposed)
$(\boldsymbol{q}_0, \boldsymbol{p}_0) \leftarrow (\boldsymbol{q}^{(i)}, \boldsymbol{p}^{(i)})$	$(\boldsymbol{r}_0, \dot{\boldsymbol{r}}_0) \leftarrow (\boldsymbol{M}^{1/2}\boldsymbol{q}^{(i)}, \boldsymbol{M}^{-1/2}\boldsymbol{p}^{(i)})$
for $j \leftarrow 0,1,\ldots,L-1$	**for** $j \leftarrow 0,1,\ldots,L-1$
$\boldsymbol{p}_{j+1/2} = \boldsymbol{p}_j - 0.5h\nabla_q U(\boldsymbol{q}_j)$	$\boldsymbol{r}_{j+1} = \cos(h\Omega)\boldsymbol{r}_j + \Omega^{-1}\sin(h\Omega)\dot{\boldsymbol{r}}_j - 0.5h^2\psi(h\Omega)\boldsymbol{F}(\phi(h\Omega)\boldsymbol{r}_j)$
$\boldsymbol{q}_{j+1} = \boldsymbol{q}_j + h\boldsymbol{M}^{-1}\boldsymbol{p}_{j+1/2}$	$\dot{\boldsymbol{r}}_{j+1} = -\Omega\sin(h\Omega)\boldsymbol{r}_j + \cos(h\Omega)\dot{\boldsymbol{r}}_j$
$\boldsymbol{p}_{j+1} = \boldsymbol{p}_{j+1/2} - 0.5h\nabla_q U(\boldsymbol{q}_{j+1})$	$-0.5h\left(\psi_0(h\Omega)\boldsymbol{F}(\phi(h\Omega)\boldsymbol{r}_j) + \psi_1(h\Omega)\boldsymbol{F}(\phi(h\Omega)\boldsymbol{r}_{j+1})\right)$
$(\boldsymbol{q}^*, \boldsymbol{p}^*) \leftarrow (\boldsymbol{q}_L, \boldsymbol{p}_L)$	$(\boldsymbol{q}^*, \boldsymbol{p}^*) \leftarrow (\boldsymbol{M}^{-1/2}\boldsymbol{r}_L, \boldsymbol{M}^{1/2}\dot{\boldsymbol{r}}_L)$

Similar complexity: Matrix operations (e.g., filter functions in blue) are pre-computable

Introduction

- HMC belongs to Metropolis-Hastings MCMC: proposal + acceptance test
- Proposal based on classical mechanics

$$\boldsymbol{q} \sim p(\boldsymbol{q}) \Rightarrow (\boldsymbol{q}, \boldsymbol{p}) \sim p(\boldsymbol{q}, \boldsymbol{p}) = p(\boldsymbol{q})p(\boldsymbol{p})$$

$\boldsymbol{q}$: position $\quad \boldsymbol{p}$: momentum

$$H(\boldsymbol{q}, \boldsymbol{p}) = -\log p(\boldsymbol{q}) - \log p(\boldsymbol{p}) = \underbrace{U(\boldsymbol{q})}_{\text{potential}} + \underbrace{\frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p} + \text{const.}}_{\text{kinetic}}$$

- Final positions of motion (after time t) as proposals, governed by ODEs

$$\begin{cases} \dot{\boldsymbol{q}} = \nabla_p H = \boldsymbol{M}^{-1}\boldsymbol{p} \\ \dot{\boldsymbol{p}} = -\nabla_q H = -\nabla_q U(\boldsymbol{q}) \end{cases} \equiv \ddot{\boldsymbol{q}} + \boldsymbol{M}^{-1}\nabla_q U(\boldsymbol{q}) = 0$$

- In practice: divide t into L steps h
  - ➢ Acceptance: $H(\boldsymbol{q}^*, \boldsymbol{p}^*) \approx H(\boldsymbol{q}^{(i)}, \boldsymbol{p}^{(i)})$

$h \downarrow \Rightarrow$ acceptance $\uparrow$ VS. $L \uparrow \Rightarrow$ time $\uparrow$

- Challenges: numerical integrators
  - ➢ Leapfrog: fast step, limited h
  - ➢ Preconditioning: suitable $\boldsymbol{M}$
  - ➢ RMHMC: adaptive $\boldsymbol{M}$, slow step

Approach

Exponential Integrator:

- Explicitly consider different frequencies
- Decompose the ODE into two terms
  1) Linear: high frequency (unstable)
  2) Nonlinear: low frequency (stable)
- Integrate the linear term analytically
- Integrate the nonlinear term numerically using *variation of constants*

Exponential HMC:

- Decomposition

$$\nabla_q U(\boldsymbol{q}) = \Sigma^{-1}\boldsymbol{q} + \boldsymbol{f}(\boldsymbol{q}) = \Sigma^{-1}\boldsymbol{q} + \left(\nabla_q U(\boldsymbol{q}) - \Sigma^{-1}\boldsymbol{q}\right)$$

$$-\nabla_q \log$$

$$p(\boldsymbol{q}) = \mathcal{N}(\boldsymbol{q}; \mu, \Sigma) \times \frac{p(\boldsymbol{q})}{\mathcal{N}(\boldsymbol{q}; \mu, \Sigma)}$$

- Gaussian approximation:
  1) Laplace approximation (expHMC)
  2) Empirical statistics ($\widehat{\text{exp}}$HMC)
- Integration: ensure convergence of HMC
  - ➢ Explicit integrators [Hairer & Lubich, 2000]
  - ➢ Filter function [Garcia-Archilla et al., 1998]

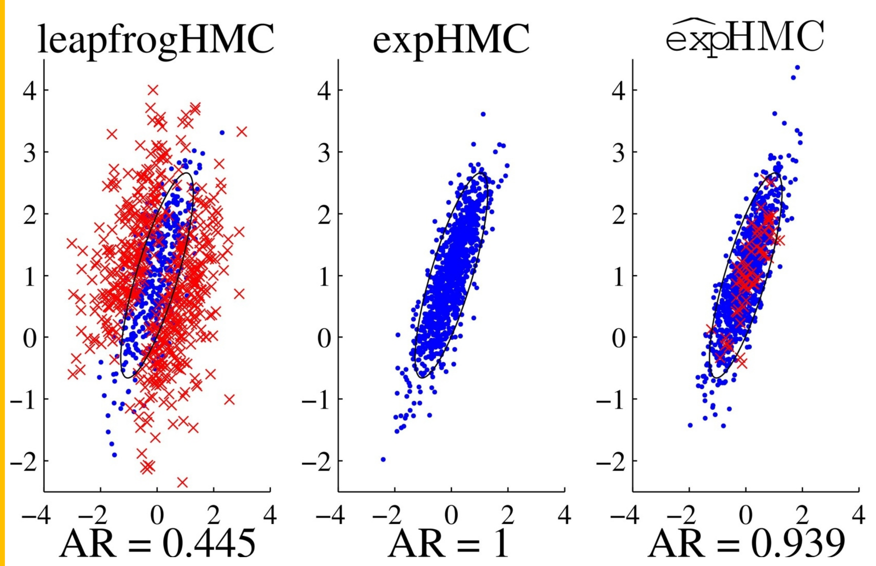
$$\ddot{\boldsymbol{q}} + \boldsymbol{M}^{-1}\left(\Sigma^{-1}\boldsymbol{q} + \boldsymbol{f}(\boldsymbol{q})\right) = 0 \Rightarrow \ddot{\boldsymbol{r}} + \Omega^2\boldsymbol{r} + \boldsymbol{F}(\boldsymbol{r}) = 0$$

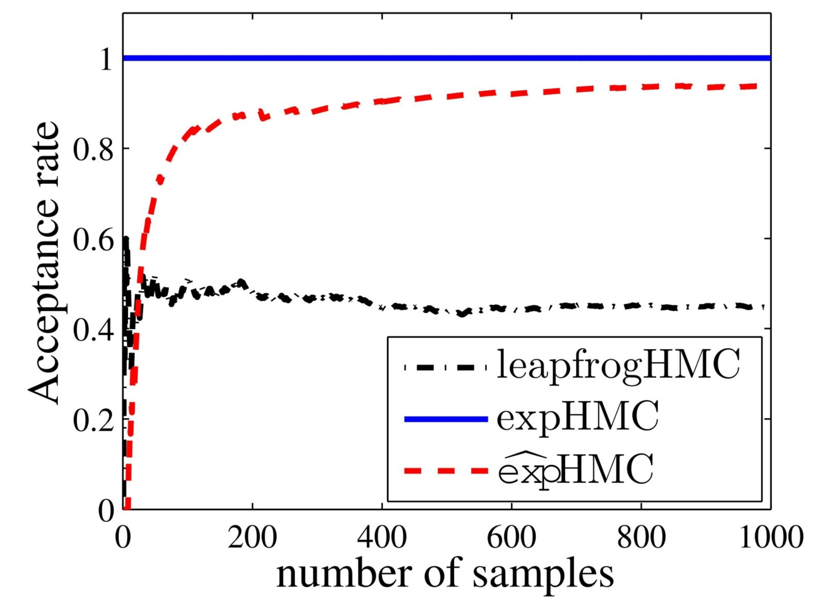
Experiments

- Evaluation:

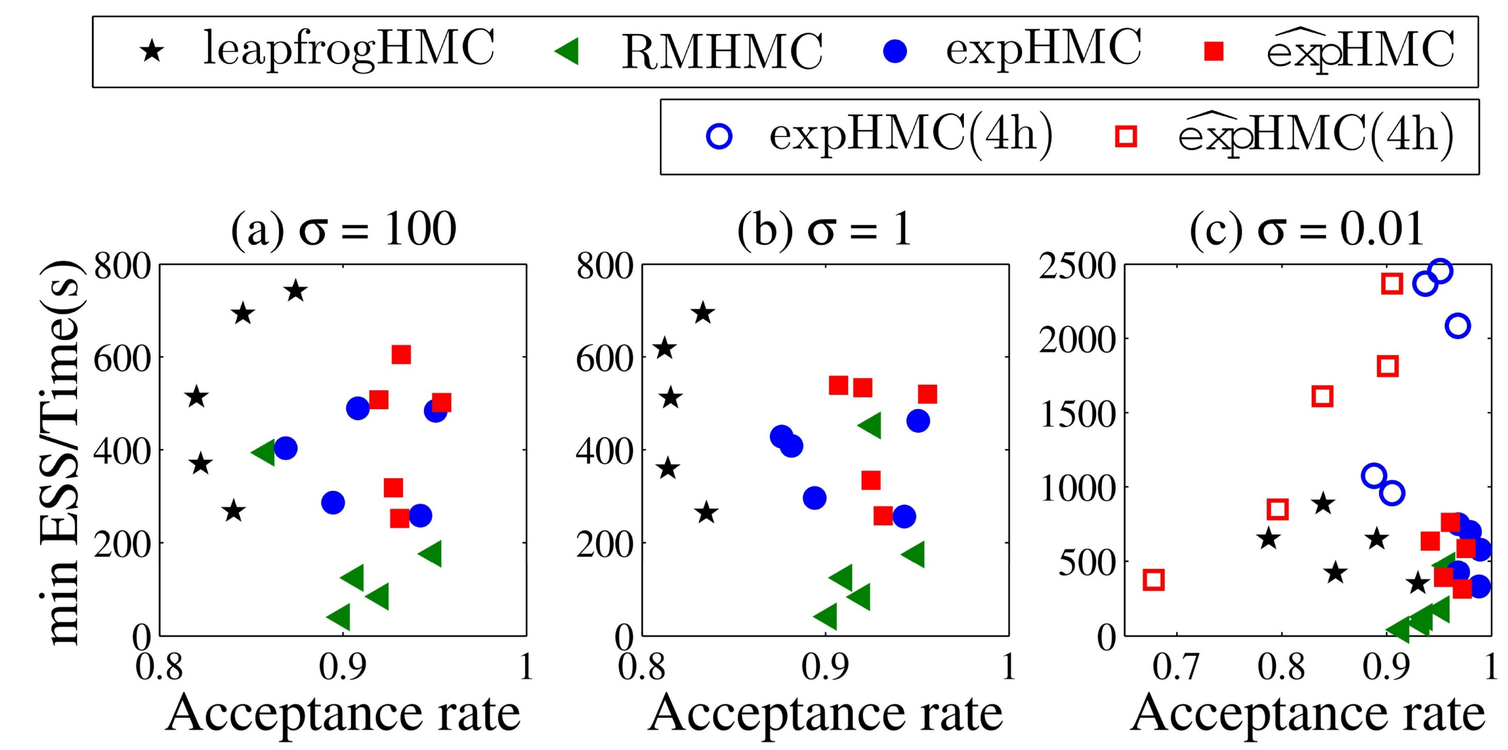
(1) Acceptance rate (AR) (2) min effective sample size (ESS) (3) min ESS/Time (s)

- Gaussian distributions
  - ➢ $(h, L) = (0.6, 8)$





- Posterior of Bayesian logistic regression (BLR)
  - ➢ 5 datasets from UCI repository
  - ➢ Prior: $\mathcal{N}(0, \sigma \boldsymbol{I})$ with $\sigma \in \{0.01, 1, 100\}$
  - ➢ L = 100, h selected to ensure suitable AR



- More details in the paper and suppl.
  - ➢ BLR: scalability & preconditioning & splitting
  - ➢ Posterior of Independent Component Analysis

[1] R. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2011

[2] M. Girolami and B. Calderhead. Riemann Manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society*, 2011