A Family of MCMC Methods on Implicitly Defined Manifolds

Marcus A. Brubaker•,+, Mathieu Salzmann• and Raquel Urtasun•

• Toyota Technological Institute at Chicago

+ University of Toronto, Canada

Introduction:

- Traditional MCMC methods (e.g., Gauss-Metropolis, HMC) assume the target distribution is over a Euclidean space
- However, many problems exist which are most naturally characterized over a non-linear manifold
- Sampling from posteriors that arise in such problems has typically required the derivation of posterior-specific sampling schemes

Contributions:

- Here we derive an MCMC scheme based on Hamiltonian dynamics on an implicitly defined manifold $\mathcal{M} = \{q \in \mathbb{R}^n | c(q) = 0\}$
- We prove that, subject to suitable conditions, the Markov Chain converges to the target posterior $\pi(q)$
- We present constrained variants of several MCMC methods including: Gauss-Metropolis, Hamiltonian (and Langevin) Monte Carlo and Riemann Manifold HMC [6]
- These algorithms are demonstrated on a range of problems including:
  - Sampling from a linearly constrained Gaussian distribution
  - Sampling from the Bingham-von Mises-Fisher distribution over $\mathcal{S}^n$
  - Bayesian matrix factorization for collaborative filtering
  - Human pose estimation
- Matlab code available from: http://www.cs.toronto.edu/~mbrubake/

Previous Work:

- Similar methods are commonly used in molecular dynamics to compute the free energy of a constrained system (eg, [1-3])
- Gibbs samplers have been derived for some distributions (eg, [4]) but even those specialized methods are outperformed by methods presented here

Theoretical Result:

- Assume that $\mathcal{M} = \{q \in \mathbb{R}^n | c(q) = 0\}$ is connected, smooth and differentiable with $C(q) = \frac{\partial c}{\partial q}$ full-rank everywhere and the target posterior $\pi(q)$ is strictly positive on $\mathcal{M}$
- Given:
  - ✓ a mass matrix $M(q)$ which is positive definite on $\mathcal{M}$
  - ✓ a simulation potential energy function $\hat{U}(q)$ which is $\mathcal{C}^2$ continuous
  - ✓ a numerical integration method $\Phi_h^{\hat{\mathcal{H}}} : \mathcal{T}^*\mathcal{M} \to \mathcal{T}^*\mathcal{M}$ which is *symmetric*, *locally accessible*, *consistent* with the Simulation Hamiltonian $\hat{\mathcal{H}}$, and *symplectic* on the co-tangent bundle

$$\mathcal{T}^*\mathcal{M} = \left\{(p,q) | c(q) = 0 \text{ and } C(q)\frac{\partial \hat{\mathcal{H}}}{\partial p}(p,q) = 0\right\}$$

- *Theorem*: For all $q_0 \in \mathcal{M}$

$$\lim_{n\to\infty} \|T^n(q_0 \to \cdot) - \pi(\cdot)\| = 0$$

where $T^n(q_0 \to \cdot)$ denotes n steps of the Markov transition kernel of the Constrained Hamiltonian Monte Carlo algorithm

Constrained Hamiltonian Monte Carlo:

- *Input:* $q_0,\ M(q),\ h,\ L,\ \pi(q),\ \hat{U}(q)$
- *Define:*
  - *Co-tangent Projection:*

$$\mathbf{P}(q) = I - M(q)^{-T}C(q)^T\left(C(q)M(q)^{-1}M(q)^{-T}C(q)^T\right)^{-1}C(q)M(q)^{-1}$$

  - *Acceptance Hamiltonian:*

$$\mathcal{H}(p,q) = \tfrac{1}{2}p^TM(q)^{-1}p + \tfrac{1}{2}\log|2\pi\mathbf{P}(q)^TM(q)\mathbf{P}(q)| - \log\pi(q)$$

  - *Simulation Hamiltonian:*

$$\hat{\mathcal{H}}(p,q) = \tfrac{1}{2}p^TM(q)^{-1}p + \hat{U}(q)$$

1. $p_0' \sim \mathcal{N}(0, M(q_0))$, $p_0 \leftarrow \mathbf{P}(q_0)p_0'$
2. For $i = 1, \ldots, L$, $(p_i, q_i) \leftarrow \Phi_h^{\hat{\mathcal{H}}}(p_{i-1}, q_{i-1})$
3. With probability $\min\{1, \exp(\mathcal{H}(p_0,q_0) - \mathcal{H}(p_L,q_L))\}$
   - Return q_L
4. Else
   - Return q_0

Simulation of constrained Hamiltonian systems

- Need a symplectic, consistent and symmetric integration method on $\mathcal{M}$
- Generalized RATTLE Algorithm (see [5] for details and other options)

$$\begin{aligned} p_{1/2} &= p_0 - \frac{h}{2}\left(\frac{\partial\hat{\mathcal{H}}(p_{1/2},q_0)}{\partial q} + C(q_0)^T\lambda\right) & p_1 &= p_{1/2} - \frac{h}{2}\left(\frac{\partial\hat{\mathcal{H}}(p_{1/2},q_1)}{\partial q} + C(q_1)^T\mu\right) \\ q_1 &= q_0 + \frac{h}{2}\left(\frac{\partial\hat{\mathcal{H}}(p_{1/2},q_0)}{\partial p} + \frac{\partial\hat{\mathcal{H}}(p_{1/2},q_1)}{\partial p}\right) & 0 &= C(q_1)\frac{\partial\hat{\mathcal{H}}(p_1,q_1)}{\partial p} \\ 0 &= c(q_1) \end{aligned}$$

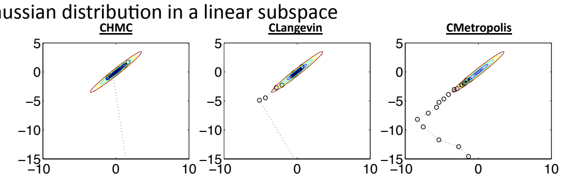
- If $\mathcal{M} = \mathbb{R}^n$ and the mass matrix is constant, RATTLE reduces to Leapfrog

Instances of Constrained HMC:

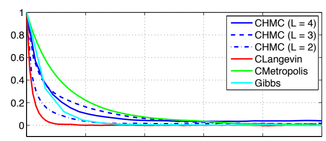
- Gauss-Metropolis with covariance Σ can expressed as HMC with $\hat{U}(q) = 0$ and $M(q) = \Sigma^{-1}$. Constrained Gauss-Metropolis is thus similarly defined.
- Constrained Langevin Monte Carlo arises with $L = 1$
- Constrained Riemann Manifold HMC [6] arises for suitable choices of $M(q)$

Experimental Results:

- Gaussian distribution in a linear subspace



- Bingham-von Mises-Fisher $\mathcal{M} = \mathcal{S}^n \quad \pi(q) \propto \exp(d^Tq + q^TAq)$

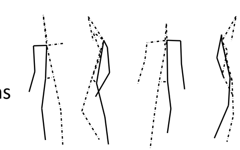


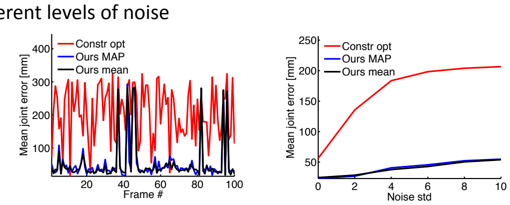
Method	$E[-\log\pi(q)]$	ESS %	ESS/second
CHMC (L = 4)	-999.021	27.3	183.756
CHMC (L = 3)	-998.759	25.4	217.427
CHMC (L = 2)	-999.121	37.9	440.898
CLangevin	-998.757	33.0	619.339
CMetropolis	-998.82	3.8	90.1513
Gibbs [4]	-998.742	50.8	160.722

- Collaborative filtering $\mathcal{M} = V_r(\mathbb{R}^N) \times V_r(\mathbb{R}^M) \times \mathbb{R}^r \quad \pi(\mathbf{U},\mathbf{S},\mathbf{V}) \propto \prod_{(i,j)\in\mathcal{E}} \exp\left(-\frac{(f(\mathbf{U}_i\mathbf{S}\mathbf{V}_j) - \mathbf{Y}_{i,j})^2}{2\sigma_p^2}\right)$

	1M Movie Lens (RMSE)			EachMovie (RMSE)		
r	5	10	15	5	10	15
HMC	1.577 ± 0.39	2.001 ± 0.66	2.306 ± 0.25	1.153 ± 0.002	1.161 ± 0.002	1.204 ± 0.018
HMC-l	0.909 ± 0.008	0.949 ± 0.01	0.99 ± 0.01	1.155 ± 0.007	1.164 ± 0.001	1.184 ± 0.004
CHMC	0.893 ± 0.01	0.888 ± 0.01	0.889 ± 0.01	1.144 ± 0.002	1.121 ± 0.001	1.116 ± 0.001
CHMC-l	**0.888 ± 0.01**	**0.881 ± 0.01**	**0.881 ± 0.01**	**1.137 ± 0.003**	**1.115 ± 0.002**	**1.11 ± 0.002**

- Human pose estimation
  - Pose is a set of 3D joint positions
  - Manifold is induced by the limb length constraints of the skeleton
  - Posterior combines noisy 2D joint projections with a PCA based prior model of pose
  - Compared with direct optimization for different levels of noise





References:

1. G. Ciccotti and J. P. Ryckaert. Molecular dynamics simulation of rigid molecules. Computer Physics Report, 4(6):346–392, 1986
2. C. Hartmann. An ergodic sampling scheme for constrained Hamiltonian systems with applications to molecular dynamics. Journal of Statistical Physics, 130:687–711, 2008
3. T. Lelièvre, M. Rousset, and G. Stoltz. Free energy computations: A Mathematical Perspective. Imperial College Press, 2010
4. P. D. Hoff. Simulation of the matrix Bingham-von Mises-FIsher distribution, with applications to multivariate and relational data. Journal of Computational and Graphical Statistics, 18:438–456, 2009
5. E. Hairer, C. Lubich, and G. Wanner. Geometric Numerical Integration. Springer, 2nd edition, 2006
6. M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society: Series B, 73:123–214, 2011