Geodesic Monte Carlo

Simon Byrne and Mark Girolami
Department of Statistical Science, University of College London

1 Introduction

Many problems in directional statistics utilise probability distributions defined over manifolds embedded in Euclidean space, such as circles, spheres, tori and Stiefel manifolds. Unfortunately, it is often difficult to construct methods for independent sampling from such distributions, as the normalisation constants are often intractable, which means that standard approaches such as rejection sampling cannot be easily implemented.

As a result, Markov chain Monte Carlo (MCMC) methods are often used to obtain a sequence of dependent samples. However such an approach also faces challenges: the usual MCMC problem of ensuring adequate mixing so as to obtain a representative sample, but also the challenge of constructing a proposal mechanism that will provide samples from the manifold. Recent developments, such as Hoff (2009), utilise sophisticated reparametrisations and proposal distributions, often specifically adapted to each target distribution family.

In this talk, I will discuss our recent work in developing and applying Hamiltonian Monte Carlo (HMC) techniques for sampling from arbitrary distributions on embedded manifolds (Byrne and Girolami, 2013).

2 Hamiltonian Monte Carlo

The key challenge in any Metropolis–Hastings-based MCMC scheme is specifying a proposal distribution that is capable of making long-distance moves that also maintain a high probability of being accepted. Hamiltonian Monte Carlo (also known as Hybrid Monte Carlo, or HMC) are a class of general purpose MCMC techniques, that make such proposals by introducing an auxiliary "momentum" variable and simulating the dynamics of a hypothetical particle in a potential field determined by the target density.

For a target density $\pi(x)$ over $\mathbb{R}^n$, a simple choice for the auxiliary variable is $v \sim \mathsf{N}_n(0, I)$ which is updated at the start of each iteration. The *Hamiltonian* is defined to be the negative log joint density (up to proportionality):

$$H(x, v) = -\log \pi(x) + \tfrac{1}{2} v^\top v.$$

The dynamics of x and v over time is determined by Hamilton's equations

$$\frac{dx}{dt} = \frac{\partial H}{\partial v} = v, \qquad \frac{dv}{dt} = -\frac{\partial H}{\partial x} = \nabla_x \log \pi(x), \tag{1}$$

Of course, these equations cannot typically be solved exactly, and so we instead approximate the dynamics by alternately updating the momentum and velocity via what is known as a *leapfrog integrator*:

$$x \leftarrow x + \epsilon v, \qquad v \leftarrow v + \epsilon \nabla_x \log \pi(x) \tag{2}$$

By running the integrator for a large number of steps, we can obtain very large movements through the sample space. The exact Hamiltonian dynamics will preserve the Hamiltonian, giving a Metropolis acceptance probability of 1; the discretisation error means that this will not be the case for the approximation (2), but by choosing ϵ appropriately, a reasonably high acceptance ratio can still be maintained.

3 Geodesic Monte Carlo

Unfortunately, the leapfrog scheme cannot be applied directly to distributions on a manifold $\mathcal{M}$, as a linear jump $x + \epsilon v$ will typically not be in $\mathcal{M}$.

The first step is to constrain v to be in the tangent space T_x of $x \in \mathcal{M}$. If P_x is an orthogonal projection onto T_x, then we assume

$$v \mid x \sim \mathsf{N}(0, P_x^\top P_x)$$

A suitable integrator can be constructed via a procedure known as "Hamiltonian splitting", which alternates between two steps:

1. The first step involves updating the velocity according to Hamilton's equations, but constrained to stay within the tangent space. This is equivalent to simply applying an orthogonal projection
$$v \leftarrow P_x[v + \epsilon \nabla_x \pi(x)] \tag{3}$$

2. The second step involves transporting both x and v along the geodesic flow of $\mathcal{M}$ for length $\alpha = \|v\|$: this requires solving the geodesic equations, but for many manifolds that arise in directional problems, explicit forms for this are known. For instance on spheres, the geodesic flows are simply the "great circle" rotations
$$\begin{bmatrix} x & v \end{bmatrix} \leftarrow \begin{bmatrix} x & v \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \alpha^{-1} \end{bmatrix} \begin{bmatrix} \cos(\alpha\epsilon) & -\sin(\alpha\epsilon) \\ \sin(\alpha\epsilon) & \cos(\alpha\epsilon) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}$$

These methods are very flexible and straightforward to implement, requiring only the ability to evaluate the unnormalised log-density and its gradients. As with standard HMC, by adjusting the number of steps and step size, we are able to make large proposals yet also maintain high acceptance probabilities.

We demonstrate the utility of these methods on several problems in directional statistics, such as sampling from Bingham–von Mises–Fisher distributions, and distributions Stiefel manifolds.

References

Byrne, S. and Girolami, M. (2013) Geodesic Monte Carlo on embedded manifolds. *Scandinavian Journal of Statistics*, in press. ARXIV: 1301.6064.

Hoff, P. (2009) 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.