Langevin MC,
or the curse of dimensionality

B. Han, T.M. Hodgson, M. Holden & M. Puza

March 7, 2019

1 Motivation

2 Introduction

Markov chain Monte Carlo (MCMC) methods are a family of algorithms for calculating numerical approximations of integrals, with applications in computational physics, biology and statistics.

Monte Carlo Integration

Th

3 Langevin Monte Carlo Algorithms

The Langevin equation is a stochastic differential equation (SDE) originally developed to model the movement of a Brownian particle [1]. The form of interest here is the *overdamped* Langevin equation, in which the particle experiences no average acceleration. The equation is thus

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t. \tag{1}$$

Here, W_t is a d-dimensional Wiener process (Brownian motion) and $U : \mathbb{R}^d \to \mathbb{R}$ is the potential function. The equation can be thought of as modelling a particle in a potential well with shape U. As each particle moves randomly, it is natural to ask what is the average position of many particles in such a well? It can be shown that in fact the position of a particle moving according to the above dynamics is exactly π +++Reference to earlier section/first mention of distribution+++. For a diffusion process this is called the *stationary distribution*[a] To show that π is indeed the stationary distribution the following lemma

Lemma. *For a one-dimensional Itô diffusion,*

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma^2(X_t)dW_t,$$

the Fokker-Planck operator, $\mathcal{L}^*$, is

$$\mathcal{L}^* := -\partial_x(\mu(x)\cdot) + \frac{1}{2}\partial_x^2(\sigma^2(x)\cdot).$$

A measure π is invariant for the diffusion if and only if

$$\mathcal{L}^*\pi = 0$$

The proof of this is omitted however it can be seen by forming the Fokker-Planck equation for the probability density of the diffusion. The proof that π is the stationary measure of Equation (1) is given only in the one dimensional case, however it is extendable to higher dimensions. For the Langevin equation, the Fokker-Planck operator is

$$\mathcal{L}^* = \partial_x(U'(x)\cdot) + \partial_{xx}\cdot\,.$$

[a] Another common term is *invariant measure* +++

So it remains to calculate $\mathcal{L}^*\pi$.

$$\begin{aligned}\mathcal{L}^*\pi &= \frac{\partial}{\partial x}\left[U'(x)\pi(x) + \frac{\partial}{\partial x}\pi(x)\right] \\ &= \frac{\partial}{\partial x}\left[U'(x)\mathcal{Z}\mathrm{e}^{-U(x)} + \left(-U'(x)\mathcal{Z}\mathrm{e}^{-U(x)}\right)\right] \\ &= \frac{\partial}{\partial x}[0] \\ &= 0\end{aligned}$$

Hence π is indeed the invariant measure of (1). ■

Although this shows that the Langevin equation has an invariant measure, the question of convergence to this measure remains unanswered. Roberts and Tweedie give the following restriction [3].

Theorem 3.1 (Theorem 2.1, [3])**.** *Let $P_X^t(x, A) = \mathbb{P}(X_t \in A | X_0 = x_0)$ and suppose that $\nabla U(x)$ is continuously differentiable and that, for some $N, a, b < \infty$,*

$$\nabla U(x) \cdot x \leq a|x|^2 + b, \qquad |x| > N.$$

Then the measure π is invariant for the Langevin diffusion X. Moreover, for all $x \in \mathbb{R}^d$ and Borel sets A,

$$\|P_X^t(x, \cdot) - \pi\| = \frac{1}{2}\sup_A \left|P_X^t(x, A) - \pi(A)\right| \to 0$$

+++Should this norm be an integral? Figure of $U = x^2$? Very small step EM scheme +++
The problem of sampling from the high dimensional distribution has been reduced to being able to accurately simulate Langevin dynamics. However, this is not as simple as it sounds. To simulate the continuous process (1), it must first be discretised. However, doing so may not preserve the convergence to the invariant measure. The discretised process may not have the same stationary measure or the masure may not even exist. This means that the method used to discretise must be chosen carefully to ensure good convergence properties. The most natural way to discretise an SDE is to use the stochastic analogue of the Euler method used on ordinary differential equations, known as the Euler-Maruyama (EM) method. Doing so leads to the Unadjusted Langevin Algorithm (`ULA`).

3.1 The Unadjusted Langevin Algorithm

Applying the Euler-Maruyama method to Equation (1) gives the following iterative scheme.

$$X_{n+1} = X_n - \gamma\nabla U(X_n) + \sqrt{2\gamma}Z_{n+1}, \qquad X_0 = x_0$$

Here the Z_n are i.i.d. standard normal random variables and γ is the step size. This is equivalent to $X_{n+1} \sim N(X_n - \gamma\nabla U)X_n), 2\gamma I_d$.[b] A simple example shows that this discretisation does not converge to π. Let $U(x) = |x|^2/2$ and $\gamma = 1$.

However, it is well known that this method does not always give the correct answer, and diverges whenever $\nabla U(x)$ is superlinear (See Roberts & Tweedie 1996). How can we solve this problem? We can look at either discretising the SDE in a different way (HOLA, LM); or we can modify the Euler scheme to mitigate the issues. Here we will focus on the latter, although our code has algorithms from both approaches.

4 Taming

Taming has been suggested for ULA by Brosse, Durmus, Moulines and Sabanis, as well as by Roberts & Tweedie for MALA (they called the result MALTA). The idea is to scale the gradient

5 ULA

$$X_{n+1} = X_n - \gamma\nabla U(X_n) + \sqrt{2\gamma}Z_{n+1}, \qquad X_0 = x_0$$

[b] I_d denotes the $d \times s$ identity matrix.

6 MALA

[3] Propose V_{n+1} using Langevin dynamics:

$$V_{n+1} = X_n - \gamma \nabla U(X_n) + \sqrt{2\gamma} Z_{n+1}$$

Calculate acceptance probability

$$\alpha(X_n, V_{n+1}) = 1 \wedge \frac{\pi(V_{n+1}) q(V_{n+1}, X_n)}{\pi(X_n) q(X_n, V_{n+1})}$$

Here $q(x, y)$ is the transition probability, $\mathbb{P}(X_{n+1} = y | X_n = x)$. If `rand`$\leq \alpha$,

$$X_{n+1} = V_{n+1}.$$

That is,

$$X_{n+1} = \mathbb{I}(u \leq \alpha) V_{n+1} + \mathbb{I}(u > \alpha) X_n$$

7 Taming the Gradient

7.1 tULA/c

$$X_{n+1} = X_n - \gamma T_\gamma(X_n) + \sqrt{2\gamma} Z_{n+1}, \qquad X_0 = x_0$$

where $T_\gamma(x) = \frac{\nabla U(x)}{1 + \|\nabla U(x)\|}$ or $T_\gamma(x) = \left(\frac{\nabla U(x)}{1 + |\partial_i U(x)|} \right)_{i = \{1, \ldots, d\}}$

7.2 tHOLA

[4] Use an Itô-Taylor expansion

$$X_{n+1} - X_n + \mu_\gamma(X_n)\gamma + \sigma_\gamma(X_n)\sqrt{\gamma} Z_{n+1}$$

where

$$\mu_\gamma(x) = -\nabla U_\gamma(x) + \frac{\gamma}{2} \left(\left(\nabla^2 U \nabla U \right)_\gamma (x) - \vec{\Delta}(\nabla U)_\gamma(x) \right),$$

and $\sigma_\gamma(x) = \operatorname{diag}\left(\left(\sigma_\gamma^{(k)}(x) \right)_{k \in \{1, \ldots, d\}} \right)$ with,

$$\sigma_\gamma^{(k)}(x) = \sqrt{2 + \frac{2\gamma^2}{3} \sum_{j=1}^{d} |\nabla^2 U_\gamma^{(k,j)}(x)|^2 - 2\gamma \nabla^2 U_\gamma^{(k,k)}(x)}$$

ALSO need to define the gamma subscript, i.e. the tamed variables. is gamma best subscript? Although fn depends on gamma it doesn't indicate that taming has occurred.

7.3 tMALA/c

Use the same taming T as in `tULA`. Is this sensible? Could compare with `MALTA`.

7.4 MALTA?

[**?**] Tame with

$$T = \frac{\nabla U(x)}{1 \vee \gamma |\nabla U(x)|}$$

for some constant $D > 0$

8 LM

[2] Non-Markovian scheme,

$$X_{n+1} = X_n + \gamma \nabla U(X_n) + \sqrt{\frac{\gamma}{2}} (Z_n + Z_{n+1})$$

9 RWM

Popular variant of the Metropolis-Hastings algorithm (CITE) with a normal proposal.

$$U_{n+1} = X_n + \sqrt{2\gamma} Z_{n+1}$$

Calculate acceptance probability

$$\alpha(X_n, U_{n+1}) = 1 \wedge \frac{\pi(U_{n+1}) q(U_{n+1}, X_n)}{\pi(X_n) q(X_n, U_{n+1})}$$

Here $q(x, y)$ is the transition probability, $\mathbb{P}(X_{n+1} = y | X_n = x)$. If `rand`$\leq \alpha$,

$$X_{n+1} = U_{n+1}.$$

That is,

$$X_{n+1} = \mathbb{I}(u \leq \alpha) U_{n+1} + \mathbb{I}(u > \alpha) X_n$$

10 Beyond Moments

References

[1] Don S. Lemons and Anthony Gythiel. Paul Langevins 1908 paper On the Theory of Brownian Motion [Sur la thorie du mouvement brownien, C. R. Acad. Sci. (Paris) 146, 530533 (1908)]. *American Journal of Physics*, 65(11):1079–1081, 1997.

[2] Charles Matthews and Benedict Leimkuhler. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Applied Mathematics Research eXpress*, 2013(1):34–56, 06 2012.

[3] Gareth O. Roberts and Richard L. Tweedie. Exponential Convergence of Langevin Distributions and Their Discrete Approximations. *Bernoulli*, 2(4):341–363, 1996.

[4] Sotirios Sabanis and Ying Zhang. Higher Order Langevin Monte Carlo Algorithm. Workingpaper, ArXiv, 8 2018.