Discussion of the Read Paper by Girolami and Calderhead "Riemann manifold Langevin and Hamiltonian Monte Carlo methods" read to the Society on October 13th, 2010

Christian P. Robert

Université Paris–Dauphine, CEREMADE, and CREST, Paris

This paper is an interesting addition to recent MCMC literature and I am eager to see how the community is going to react to this potential addition to the MCMC toolbox. I am however wondering about the impact of the paper on MCMC practice. Indeed, while the dynamic on the level sets of

$$\mathscr{H}(\theta,\mathbf{p}) = -\mathcal{L}(\theta) + \frac{1}{2}\log\{(2\pi)^D|\mathbf{G}(\theta)|\} + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{G}(\theta)^{-1}\mathbf{p}\,,$$

is associated with Hamilton's equations, in that those moves preserve the potential $\mathscr{H}(\theta,\mathbf{p})$ and hence the target distribution at all times t, I argue that the transfer to the simulation side, i.e. the discretisation part, is not necessarily useful, or at least that it does not need to be so painstakingly reproducing the continuous phenomenon.

In a continuous time-frame, the purpose of the auxiliary vector $\mathbf{p}$ is clearly to speed up the exploration of the posterior surface by taking advantage of the additional energy it provides. In the discrete-time universe of simulation, on the one hand, the fact that the discretised (Euler) approximation to Hamilton's equations are not exact nor available in closed form does not present such a challenge in that approximations can be corrected by a Metropolis-Hastings step, provided of course all terms in the Metropolis-Hastings ratio are available. On the other hand, the continuous time (physical or geometric) analogy at the core of the Hamiltonian may be unnecessary costly when trying to carry a physical pattern in a discrete (algorithmic) time. MCMC algorithms are not set to work in continuous time and therefore the invariance and stability properties of the continuous time process that motivates the method do not carry to the discretised version of the process. For one thing, the (continuous) time unit has no equivalent in discrete time. Therefore, the dynamics of the Hamiltonian do not tell us how long the discretised version should run, as illustrated on Figure 1. As a result, convergence issues (of the MCMC algorithm) should not be impacted by inexact renderings of the continuous time process in discrete time. For instance, when considering the Langevin diffusion, the corresponding Langevin algorithm could as well use another scale η for the gradient than the one τ used for the noise, i.e.

$$y = x^t + \eta \nabla \pi(x) + \tau \epsilon_t$$

rather than a strict Euler discretisation where $\eta = \tau^2/2$. A few experiments run in Robert and Casella (1999, Chapter 6, Section 6.5) showed that using a different scale η could actually lead to improvements, even though we never pursued the matter any further.

References

Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. 1st ed. Springer-Verlag, New York.

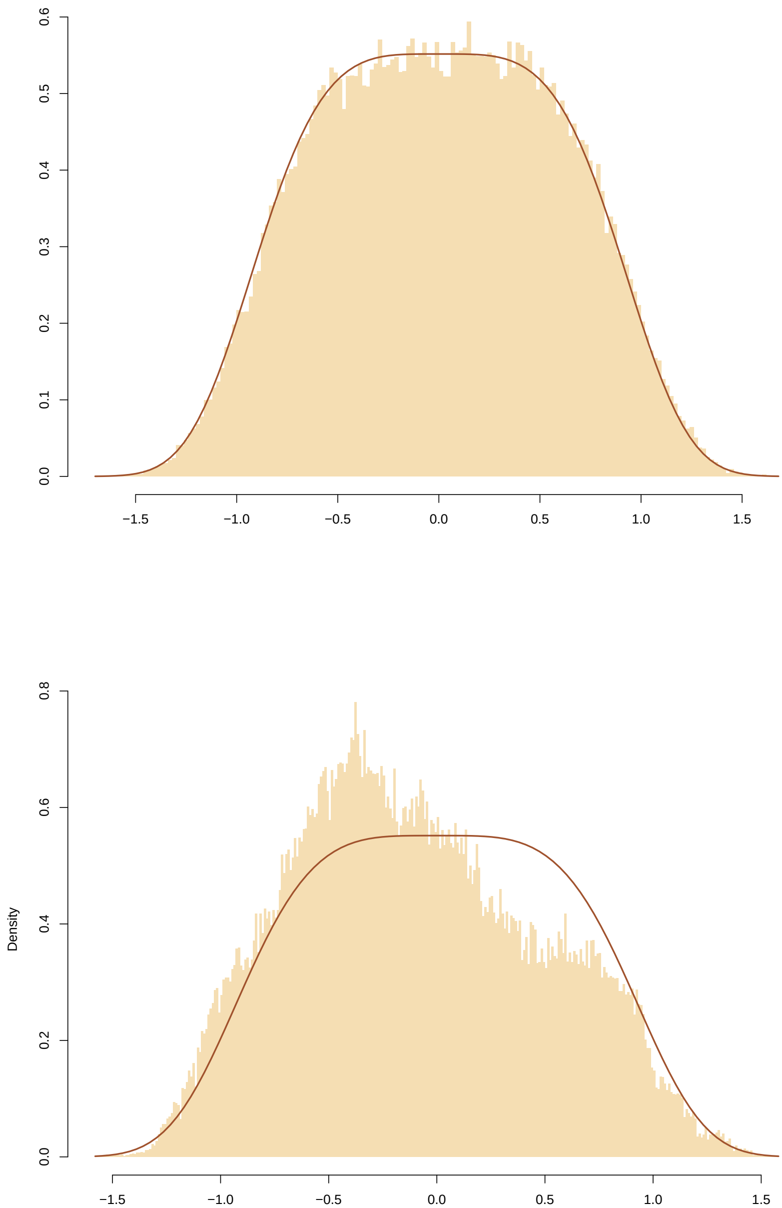

Fig. 1. Comparison of the fits of discretised Langevin diffusions to the target $f(x) \propto \exp(-x^4)$ when using a discretisation step $\sigma^2 = .01$ *(left)* and $\sigma^2 = .0001$ *(right)*, after $T = 10^7$ steps. This comparison illustrates the need for more time steps when using a smaller discretisation step.