Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 125 (2015) 3522–3540
stochastic processes and their applications
www.elsevier.com/locate/spa

Variance reduction for diffusions

Chii-Ruey Hwang[a], Raoul Normand[a,*], Sheng-Jhih Wu[b]

[a] *Institute of Mathematics - Academia Sinica - 6F, Astronomy-Mathematics Building - No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan*

[b] *Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, China*

Received 13 January 2015; received in revised form 18 March 2015; accepted 19 March 2015
Available online 26 March 2015

Abstract

The most common way to sample from a probability distribution is to use Markov Chain Monte Carlo methods. One can find many diffusions with the target distribution as equilibrium measure, so that the state of the diffusion after a long time provides a good sample from that distribution. One naturally wants to choose the best algorithm. One way to do this is to consider a reversible diffusion, and add to it an antisymmetric drift which preserves the invariant measure. We prove that, in general, adding an antisymmetric drift reduces the asymptotic variance, and provide some extensions of this result.

MSC: 60J60; 65B99; 60J35; 47D07

Keywords: Asymptotic variance; Rate of convergence; Diffusion; Acceleration; Markov Chain Monte Carlo

1. Introduction

Sampling directly from a probability distribution is often infeasible in practice, in particular when the distribution is only known up to a multiplicative constant. Markov Chain Monte Carlo (MCMC) methods are popular for obtaining samples from a target probability π, see e.g. [3,9].

* Corresponding author.
E-mail addresses: crhwang@sinica.edu.tw (C.-R. Hwang), rnormand@math.sinica.edu.tw (R. Normand), szwu@suda.edu.cn (S.-J. Wu).

http://dx.doi.org/10.1016/j.spa.2015.03.006

The idea is to construct a Markov chain or Markov process (X_t), whose invariant distribution is π. Under reasonable assumptions ("ergodicity"), the distribution of (X_t) converges to π, so that, for t large enough, X_t provides a good sample of π. It is natural that one would prefer the dynamics that attain the desired equilibrium distribution as fast as possible. There are several ways to measure how good an algorithm is: mixing time, spectral gap, asymptotic variance.... Hence, comparing the performance of different algorithms also depends on the criterion used.

The discrete case has been extensively studied in the literature, see e.g. [4,7,16–18,20]. In general, one can choose reversible or non-reversible algorithms. The latter usually enjoy a faster convergence to equilibrium, but are more challenging to study. More specifically, of special interest to us is the scheme of modifying a given reversible process by a suitable antisymmetric perturbation so that the equilibrium is preserved, but the convergence is accelerated. All evidence points to this strategy always providing a better algorithm, be it in the discrete case [5] or in the continuous one [10,11,19]. In the continuous setting, the performance is understood in terms of the spectral gap of the generator of the diffusion. Our goal in this paper is to show that perturbing a diffusion by an antisymmetric drift also reduces the asymptotic variance. Let us note that the recent paper [24] provides a similar result in terms of large deviations, with a corollary for the asymptotic variance. However, the setting is limited to the d-dimensional torus, and some details seem to be lacking in the proof.

Let us informally introduce our setting. Let U be a given energy function and π the probability with density proportional to $e^{-U(x)}$. If one is interested in sampling from π, then a usually used diffusion is the time reversible Langevin equation with π as equilibrium measure

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

where (W_t) is a Brownian motion. Perturbing the reversible diffusion (1) by adding an antisymmetric drift term results in

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

where the vector field C is weighted divergence-free with respect to π, i.e., $\mathrm{div}(Ce^{-U}) = 0$. This ensures that the non-reversible diffusion (2) also has equilibrium π. That there are many ways to choose such a perturbation C, such as taking $C = Q\nabla U$ for an antisymmetric matrix Q. Note that, in any case, it is unnecessary to know the normalization constant for π.

Let $-L$ (we use the sign convention to make L positive) denote the infinitesimal generator of (1). More precisely,

$$L = -\Delta + \nabla U \cdot \nabla$$

on, say, C_c^∞, the space of infinitely differentiable functions with compact support. Informally, one may recover that π is invariant by noticing that the adjoint operator is $L^* f = -\Delta f - \mathrm{div}(f\nabla U)$, so that $L^*\pi = 0$. This process is reversible, which amounts to saying that L is symmetric in $\mathcal{L}^2(\pi)$, the space of square-integrable complex functions with respect to π. On the other hand, the generator of the modified equation is given by $-L_C$, where

$$L_C = L - C \cdot \nabla.$$

It has the adjoint $L_C^* f = L^* f + \mathrm{div}(fC)$. Hence, to ensure that the diffusion (2) also has π as its invariant measure, it is necessary to assume that $L_C^*\pi = 0$, i.e., $\mathrm{div}(Ce^{-U}) = 0$. In [11] (see also [10]), the spectral gap of these operators is employed as a measurement of the rate of