Adv. Appl. Prob. **49,** 791–825 (2017)
*doi:*10.1017/apr.2017.22
© *Applied Probability Trust* 2017

LIMIT THEOREMS FOR THE ZIG-ZAG PROCESS

JORIS BIERKENS,* *University of Warwick*
ANDREW DUNCAN,** *Imperial College London*

Abstract

Markov chain Monte Carlo (MCMC) methods provide an essential tool in statistics for sampling from complex probability distributions. While the standard approach to MCMC involves constructing discrete-time reversible Markov chains whose transition kernel is obtained via the Metropolis–Hastings algorithm, there has been recent interest in alternative schemes based on piecewise deterministic Markov processes (PDMPs). One such approach is based on the zig-zag process, introduced in Bierkens and Roberts (2016), which proved to provide a highly scalable sampling scheme for sampling in the big data regime; see Bierkens *et al.* (2016). In this paper we study the performance of the zig-zag sampler, focusing on the one-dimensional case. In particular, we identify conditions under which a central limit theorem holds and characterise the asymptotic variance. Moreover, we study the influence of the switching rate on the diffusivity of the zig-zag process by identifying a diffusion limit as the switching rate tends to ∞. Based on our results we compare the performance of the zig-zag sampler to existing Monte Carlo methods, both analytically and through simulations.

Keywords: MCMC; nonreversible Markov process; piecewise deterministic Markov process; continuous-time Markov process; central limit theorem; functional central limit theorem

2010 Mathematics Subject Classification: Primary 65C05
Secondary 60J25; 60F05; 60F17

1. Introduction

Markov chain Monte Carlo (MCMC) methods remain an essential computational tool in statistics and among other things have made it possible for Bayesian inference techniques to be applied to increasingly complex models. Due to its simplicity and wide applicability, the Metropolis–Hastings (MH) algorithm [15], [24] and its numerous variants remain the most widely used MCMC method for sampling from a general target probability distribution, despite having been introduced over sixty years ago. Given a target distribution π, the MH scheme defines a discrete-time Markov chain which will be both ergodic and reversible with respect to π. The fact that the Markov chain is reversible is a serious limitation. Indeed, it is now well known that nonreversible chains can significantly outperform reversible chains, in terms of rate of convergence to equilibrium [16], [22], asymptotic variance [6], [9], [34] as well as large deviation functionals [31]–[33]. One particular approach to improving performance is to introduce a velocity/momentum variable and construct Markovian dynamics which are able to mix more rapidly in the augmented state space. Such methods include hybrid Monte Carlo (HMC) methods, inspired by Hamiltonian dynamics, and numerous generalisations. While the

Received 9 September 2016; revision received 1 June 2017.
* Current address: Delft Institute of Applied Mathematics, Mekelweg 4, 2628 CD, Delft, The Netherlands.
Email address: joris.bierkens@tudelft.nl
** Current address: School of Mathematical and Physical Sciences, University of Sussex, Brighton BN1 9QH, UK.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

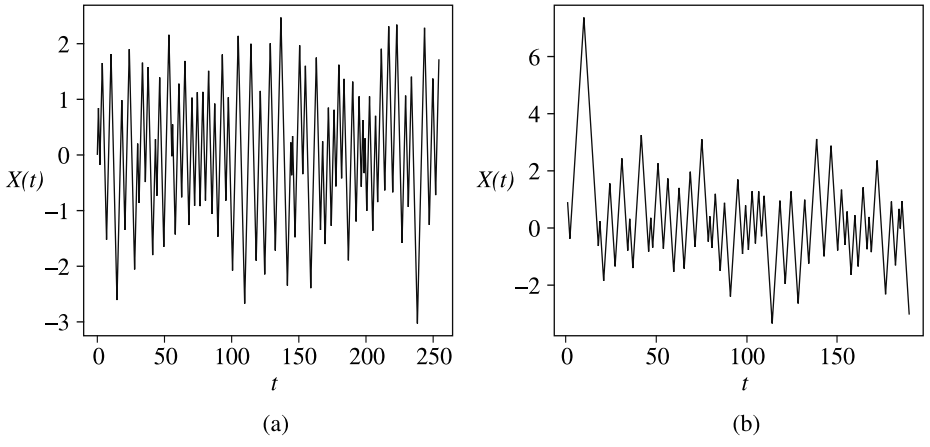


FIGURE 1: Example trajectories of the zig-zag process with the specified invariant distributions. (a) One-dimensional Gaussian and (b) one-dimensional Cauchy.

standard construction of HMC [8], [28] is reversible, it is straightforward to alter the scheme such that the resulting process is nonreversible [29].

In [2] the zig-zag process was introduced, a continuous-time piecewise deterministic process (PDMP) which provides a practical sampling scheme applicable for a wide class of probability distributions. Given a target density π, known up to a multiplicative constant, the one-dimensional zig-zag process is a continuous-time Markov process $(X(t), \Theta(t))_{t\geq 0}$ on $E = \mathbb{R} \times \{-1, +1\}$, such that $X(t)$ moves with constant velocity $\Theta(t)$. The velocity process $\Theta(t)$ switches its values between -1 and $+1$ at random times obtained from an inhomogeneous Poisson process with switching rate $\lambda(X(t), \Theta(t))$. If the switching rate is chosen to agree with the target distribution π in a certain way, this guarantees that the zig-zag process has stationary distribution μ on $\mathbb{R} \times \{-1, +1\}$, whose marginal distribution on $\mathbb{R}$ is proportional to π. As a consequence, the law of large numbers,

$$\mathbb{E}_\pi[f] = \int_{\mathbb{R}} f(x)\pi(x)\,\mathrm{d}x = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(X(s))\,\mathrm{d}s, \tag{1.1}$$

is satisfied, so that the zig-zag process can be used to approximate expectations with respect to π. Two one-dimensional examples of the zig-zag process are displayed in Figure 1.

While the construction and finite-time behaviour of PDMPs is well understood [7], their use within the context of sampling has only recently been considered and is mostly unexplored. The first such occurrence of a MCMC scheme based on PDMP appeared in the computational physics literature [30] and in one dimension coincides with the zig-zag sampler. This scheme was extended and analysed carefully in [4], where it was rechristened the bouncy particle sampler. In one dimension, the quantitative long-time behaviour of related PDMP schemes have been analysed in detail; see, for example, [1], [12], [13], [26], and [27]. More recently, in [3] the application of the zig-zag sampler to big data settings was investigated. It was found that the zig-zag sampler lends itself very well to such problems since subsampling can be introduced without affecting the stationary distribution, as opposed to standard subsampling techniques, such as stochastic gradient Langev in dynamics [35] which are inherently biased. By introducing appropriate control variates a 'super-efficient' sampling scheme for big data problems was produced, in the sense that it is able to generate independent samples from the

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

target distribution at a higher efficiency than directly generating independent and identically distributed (i.i.d.) samples using the entire data set for each sample.

In this paper we seek to better understand the qualitative performance of the zig-zag sampler. Focusing on the one-dimensional case, we study the important practical question of whether a central limit theorem (CLT) holds for the zig-zag process, i.e. whether for a given observable f,

$$\sqrt{t}\left(\frac{1}{t}\int_0^t f(X(s))\,\mathrm{d}s - \mathbb{E}_\pi[f]\right) \xrightarrow{\mathrm{D}} \mathcal{N}(0, \sigma_f^2) \quad \text{as } t \to \infty, \tag{1.2}$$

where σ_f^2 is the asymptotic variance and where '$\xrightarrow{\mathrm{D}}$' denotes convergence in distribution. Heuristically, once a CLT is known to hold, we know that the ergodic average in (1.1) converges at rate $\sigma_f/\sqrt{t}$, which is the best convergence to be expected in a Monte Carlo simulation. It is also clear that a smaller value of $\sigma_f > 0$ implies a faster convergence of the ergodic averages. Without a CLT, convergence may be arbitrarily slow. Starting from the case of a unimodal target distribution and extending to more general cases, we obtain sufficient conditions for (1.2) to hold. Moreover, we identify conditions under which the CLT can be strengthened to an invariance principle or functional central limit theorem (FCLT) [21]. For the one-dimensional zig-zag process, we obtain explicit expressions for the asymptotic variance, which we illustrate for various examples.

Given a target distribution π, there is some freedom in choosing the switching rate λ in such a way that π is invariant for the zig-zag process. This freedom is crucial for the ability of the subsampling zig-zag scheme of [3] to sample without bias. In Section 4 we study the influence of the particular choice of switching rate on the behaviour of the process. We show that as the switching rate is increased the zig-zag sampler will exhibit random walk behaviour. In particular, over an appropriate timescale the zig-zag sampler will behave asymptotically, as the excess switching rate tends to ∞, as an overdamped Langevin diffusion which is ergodic with respect to π.

As the zig-zag sampler is based upon a continuous-time process, it is not immediately clear how its performance can be compared to existing discrete-time sampling schemes. With this aim in mind, we derive approximations for the average switching rate of the process per unit time, and apply this to construct an *effective sample size* (ESS) for the zig-zag sampler which quantifies the number of independent samples generated in terms of the number of evaluations of the gradient of the log density. A suitable definition of effective sample size depends in an essential way on the asymptotic variance of the corresponding CLT, which further illustrates the importance of establishing a CLT from an applied viewpoint. Comparing to i.i.d. samples in some cases, we observe a remarkable feature: the effective sample size of the zig-zag sampler will be larger than that of i.i.d. samples, behaviour which is strongly tied to the nonreversibility of the scheme.

We structure the paper as follows. In Section 2 we review the construction of the zig-zag sampler in the one-dimensional case and explore its basic properties. Section 3 describes conditions for a CLT to hold for the one-dimensional zig-zag sampler and characterises the asymptotic variance. These results are demonstrated numerically for some standard probability distributions. In Section 4 the diffusive regime is investigated where the switching rate λ goes to ∞. Finally, in Section 5 an appropriate measure of effective sample size is introduced for the zig-zag sampler, and is used to compare the performance of the zig-zag sampler with other sampling techniques for some standard probability distributions. The proofs of most of the results can be found in Appendix A. In Appendix B we discuss the simulation of the zig-zag process, which provides the necessary background for Section 5.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

1.1. Notation

For E a topological space, the space of continuous functions $f\colon E \to \mathbb{R}$ is denoted by $C(E)$, and $\mathcal{M}(E)$ denotes the set of Borel measurable functions on E. The Borel sets in E are denoted by $\mathcal{B}(E)$. On a measurable space E, the measure δ_x for $x \in E$, is defined as the probability measure assigning mass 1 to x. Lebesgue measure on $\mathbb{R}^d$ is denoted by Leb. The Skorokhod space of càdlàg paths from an interval $I \subset \mathbb{R}$ into E is denoted by $D(I; E)$; see [11] for details. The Skorokhod space of càdlàg paths from I into $\mathbb{R}$ is also denoted by $D(I)$. We use the symbol '$\xrightarrow{\text{w}}$' to indicate weak convergence of probability distributions, where the relevant topology (either the natural topology on $\mathbb{R}$ or the Skorokhod topology on the space of càdlàg paths) can be deduced from the context. We write $\mathcal{L}(X)$ for the law of a random variable X. The pushforward $\mu_\star f$ of a measure μ on E by a measurable function $f\colon E \to F$, with E and F measurable spaces, is defined as $\mu_\star f(A) := \mu(f^{-1}(A))$ for measurable sets A in F. We write Φ for the cumulative distribution function of the standard normal distribution. We will use the notation π for a probability density function $\pi\colon \mathbb{R} \to [0, \infty)$, as well as for the associated probability measure, so, for example, $\pi(f) = \int_{\mathbb{R}} f(x)\pi(x)\,\mathrm{d}x$. For $a \in \mathbb{R}$, we will write $(a)^+$ and $(a)^-$ for the positive and negative parts of a, respectively, i.e. $(a)^+ = \max(0, a)$ and $(a)^- = \max(0, -a)$.

2. The zig-zag process

In this section we review some earlier established results on the zig-zag process. Let $E = \mathbb{R} \times \{-1, +1\}$ and equip E with the product topology of open sets in $\mathbb{R}$ and the discrete topology on $\{-1, +1\}$. The following assumption will be sufficient to define the zig-zag process, and ensure it has a unique invariant distribution.

Assumption 2.1. *We assume that $\lambda\colon E \to \mathbb{R}_+$ is continuous and the function*

$$U(x) := \int_0^x \{\lambda(\xi, +1) - \lambda(\xi, -1)\}\,\mathrm{d}\xi \tag{2.1}$$

satisfies

$$\int_{-\infty}^{\infty} \exp(-U(x))\,\mathrm{d}x < \infty.$$

Furthermore, for some $x_0 > 0$, we have $\lambda(x, \theta) > 0$ if $\theta x \geq x_0$.

An alternative and convenient way of writing (2.1) is $\lambda(x, \theta) - \lambda(x, -\theta) = \theta U'(x)$ for all $(x, \theta) \in E$. It is easy to check that (2.1) holds if and only if there exists a continuously differentiable function U and a continuous nonnegative function γ such that

$$\lambda(x, \theta) = \max(0, \theta U'(x)) + \gamma(x). \tag{2.2}$$

The switching rates λ for which $\gamma \equiv 0$ are called *canonical switching rates* and the corresponding zig-zag process is called the *canonical zig-zag process*.

Let ν denote a reference measure on E given by $\nu := \mathrm{Leb} \otimes (\delta_{-1} + \delta_{+1})$. We use ν to define the probability measure μ by

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x, \theta) = \frac{\exp(-U(x))}{2k}, \qquad (x, \theta) \in E,$$

where $k := \int_{\mathbb{R}} \exp(-U(x))\,\mathrm{d}x$. The marginal distribution of μ with respect to x has Lebesgue density proportional to $\exp(-U(x))$, denoted by π, i.e. $\pi(x) = \exp(-U(x))/k$.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Define an operator L with domain

$$\mathcal{D}(L) = \{f \in C(E)\colon f(\cdot,\theta) \text{ is absolutely continuous for } \theta = \pm 1\}$$

by

$$Lf(x,\theta) = \theta \partial_x f(x,\theta) + \lambda(x,\theta)(f(x,-\theta) - f(x,\theta)), \qquad (x,\theta) \in E, \tag{2.3}$$

which will service as the generator of the Markov semigroup of the zig-zag process, with dynamics as discussed in the introduction. In the following proposition, the notion of 'petite sets' can be found in [25].

Proposition 2.1. *Suppose that Assumption 2.1 holds. Then* $(L, \mathcal{D}(L))$ *is the extended generator of a piecewise deterministic Markov–Feller process* $(Z(t))_{t\geq 0} := (X(t), \Theta(t))_{t\geq 0}$ *in* E. *All compact sets are petite for* $(X(t), \Theta(t))$. *Finally,* μ *is the unique invariant probability distribution for* $(Z(t))_{t\geq 0}$.

Proof. The proof of this result can be found in Appendix A.1. □

The above setting can be used for Monte Carlo sampling as follows. Starting from a normalizable (but possibly unnormalized), strictly positive and continuously differentiable density $\widetilde{\pi}(x)$ on $\mathbb{R}$, we can define $U(x) := -\log \widetilde{\pi}(x)$, and define $\lambda(x,\theta)$ by (2.2) for some nonnegative function γ of our choice. Assuming that, for some $x_0 > 0$, either $\gamma(x) > 0$ for $|x| \geq x_0$, or that $\theta U'(x) > 0$ for $\theta x \geq x_0$, Assumption 2.1 is satisfied, and the process constructed in Proposition 2.1 has marginal stationary distribution π on $\mathbb{R}$, where π is the normalization of $\widetilde{\pi}$.

We call $(Z(t))_{t\geq 0} = (X(t), \Theta(t))_{t\geq 0}$ the *zig-zag process* with *switching intensity* $\lambda(x,\theta)$. Although the paths of the zig-zag process are continuous in E, in view of our goal of obtaining limit theorems for the zig-zag process we will consider its sample paths as elements in $D([0,\infty);\ E)$. For any probability distribution η on E, let $\mathbb{P}_\eta$ denote the probability measure on $D([0,\infty);\ E)$ for the zig-zag process with initial distribution η. In particular, under $\mathbb{P}_\mu$ the law of $(Z(t))_{t\geq 0}$ is stationary.

3. CLTs for the zig-zag process

First, in Section 3.1, we obtain a CLT for the zig-zag process in the simple and intuitive case in which the target distribution is unimodal and the excess switching rate $\gamma = 0$. Then we describe a general approach to the CLT in Section 3.2. We then illustrate the theory with several examples in Section 3.3.

3.1. The CLT for the special case of a unimodal invariant distribution

If the potential $U(x)$ is continuously differentiable and is monotonically nondecreasing (nonincreasing) for $x \geq 0$ ($x \leq 0$) then the canonical switching rates associated with U satisfy $\lambda(x,+1) = 0$ for $x \leq 0$, and $\lambda(x,-1) = 0$ for $x \geq 0$. In this situation trajectories of the canonical zig-zag process will always pass through the origin $x = 0$ between switches. This regular behaviour makes it possible to obtain a CLT in a very straightforward way: by inspecting the contributions towards the total variance of trajectory segments between crossings of the origin.

Assumption 3.1. (i) *We assume that* $U\colon \mathbb{R} \to [0,\infty)$ *is continuously differentiable and is monotonically nondecreasing (nonincreasing) for* $x \geq 0$ ($x \leq 0$). *Furthermore,* $k := \int_{\mathbb{R}} \exp(-U(x))\,\mathrm{d}x < \infty$.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

(ii) *Assume that* $g\colon \mathbb{R} \to \mathbb{R}$ *is integrable with respect to* π *and satisfies* $\int_{\mathbb{R}} g(x)\pi(x)\,\mathrm{d}x = 0$, *where* $\pi(x) := \exp(-U(x))/k$.

(iii) *We have*

$$\int_{\mathbb{R}} |U'(t)| \exp(-U(t)) \left(\int_0^t g(s)\,\mathrm{d}s \right)^2 \mathrm{d}t < \infty.$$

(iv) *Assume that* $\lambda(x,\theta)$ *are the canonical switching rates defined by* $\lambda(x,\theta) = (\theta U'(x))^+$.

Note that the definition of π agrees with the definition of π below Assumption 2.1. Furthermore, the fact that $\exp(-U(x))$ is integrable, combined with the monotonicity assumption, implies that the switching rates $\lambda(x,\theta)$ are positive for $\theta x \geq x_0$, for some fixed $x_0 > 0$, so that Assumption 3.1 implies Assumption 2.1.

Theorem 3.1. *Suppose that Assumption 3.1 holds. Let* $(X(t), \Theta(t))$ *denote the zig-zag process with switching rates* $\lambda(x,\theta)$. *Then*

$$\frac{1}{\sqrt{t}} \int_0^t g(X(s))\,\mathrm{d}s \xrightarrow{\mathrm{w}} \mathcal{N}(0, \sigma_g^2),$$

where

$$\sigma_g^2 := \frac{2\int_{\mathbb{R}} |U'(t)| \exp(-U(t)) \left(\int_0^t g(s)\,\mathrm{d}s \right)^2 \mathrm{d}t - 4\left(\int_0^\infty \exp(-U(t)) g(t)\,\mathrm{d}t \right)^2}{\int_{-\infty}^{\infty} \exp(-U(t))\,\mathrm{d}t}. \tag{3.1}$$

Proof. Iteratively define random times $(T_i^{\pm})_{i\in\mathbb{N}}$ and $(S_i^{\pm})_{i\in\mathbb{N}}$ as follows:

$$\begin{aligned}
T_0^+ &= \inf\{t \geq 0 \colon X(t) = 0 \text{ and } \Theta(t) = +1\},\\
T_i^- &= \inf\{t > T_{i-1}^+ \colon X(t) = 0 \text{ and } \Theta(t) = -1\}, \qquad i = 1, 2, 3, \ldots,\\
T_i^+ &= \inf\{t > T_i^- \colon X(t) = 0 \text{ and } \Theta(t) = +1\}, \qquad i = 1, 2, 3, \ldots,\\
S_i^+ &= \inf\{t > T_{i-1}^+ \colon \Theta(t) = -1\}, \qquad i = 1, 2, 3, \ldots,\\
S_i^- &= \inf\{t > T_i^- \colon \Theta(t) = +1\}, \qquad i = 1, 2, 3, \ldots.
\end{aligned}$$

See Figure 2 for a graphical illustration of these times.

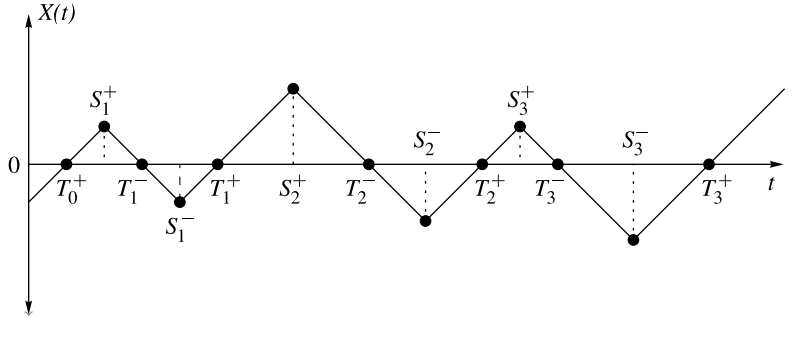


FIGURE 2: Graphical illustration of the random times $S_i^{\pm}$, $T_i^{\pm}$ introduced in the proof of Theorem 3.1.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Now for $i = 1, 2, \dots,$ define the random variables

$$Y_i^+ := \int_{T_{i-1}^+}^{T_i^-} g(s)\,\mathrm{d}s = 2\int_{T_{i-1}^+}^{S_i^+} g(s)\,\mathrm{d}s,$$
$$Y_i^- := 2\int_{T_i^-}^{T_i^+} g(s)\,\mathrm{d}s = 2\int_{T_i^-}^{S_i^-} g(s)\,\mathrm{d}s, \qquad Y_i := Y_i^+ + Y_i^-.$$

Let $N(t) := \sup\{i : T_i^+ \le t\}$. Then

$$\frac{1}{\sqrt{t}}\int_0^t g(X(s))\,\mathrm{d}s = \frac{1}{\sqrt{t}}\left(\int_0^{T_0^+} g(X(s))\,\mathrm{d}s + \sum_{i=1}^{N(t)} Y_i + \int_{T_{N(t)}^+}^t g(s)\,\mathrm{d}s\right).$$

Note that (Y_i) are i.i.d. with distribution identical to that of the random variable $Y := Y^+ + Y^-$, where Y^+ and Y^- are independent random variables defined by

$$Y^+ := 2\int_0^{\tau^+} g(s)\,\mathrm{d}s, \qquad Y^- := 2\int_0^{\tau^-} g(-s)\,\mathrm{d}s,$$

where $\mathbb{P}(\tau^\pm \ge t) = \exp(-\int_0^t \lambda(\pm s, \pm 1)\,\mathrm{d}s)$. We compute

$$\begin{aligned}\mathbb{E}[Y^+] &= \int_0^\infty \lambda(t,+1)\exp\left(-\int_0^t \lambda(s,+1)\,\mathrm{d}s\right)\left(2\int_0^t g(s)\,\mathrm{d}s\right)\mathrm{d}t\\ &= 2\int_0^\infty U'(t)\exp(-U(t))\left(\int_0^t g(s)\,\mathrm{d}s\right)\mathrm{d}t,\end{aligned}$$

and, using Assumption 3.1(ii),

$$\begin{aligned}\mathbb{E}[Y^-] &= \int_0^\infty \lambda(-t,-1)\exp\left(-\int_0^t \lambda(-s,-1)\,\mathrm{d}s\right)\left(2\int_0^t g(-s)\,\mathrm{d}s\right)\mathrm{d}t\\ &= -2\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\exp\left(-\int_0^t \lambda(-s,-1)\,\mathrm{d}s\right)\left(\int_0^t g(-s)\,\mathrm{d}s\right)\mathrm{d}t\\ &= 2\int_0^\infty \exp\left(-\int_0^t \lambda(-s,-1)\,\mathrm{d}s\right)g(-t)\mathrm{d}t\\ &= 2\exp(U(0))\int_{-\infty}^\infty \exp(-U(t))g(t)\,\mathrm{d}t - 2\exp(U(0))\int_0^\infty \exp(-U(t))g(t)\,\mathrm{d}t\\ &= -\mathbb{E}[Y^+].\end{aligned}$$

Next,

$$\begin{aligned}\mathbb{E}[(Y^+)^2] &= 4\int_0^\infty \lambda(t,+1)\exp\left(-\int_0^t \lambda(s,+1)\,\mathrm{d}s\right)\left(\int_0^t g(s)\,\mathrm{d}s\right)^2\mathrm{d}t\\ &= 4\int_0^\infty U'(t)\exp(-U(t))\left(\int_0^t g(s)\,\mathrm{d}s\right)^2\mathrm{d}t\end{aligned}$$

and, similarly,

$$\mathbb{E}[(Y^-)^2] = 4\int_{-\infty}^0 (-U'(t))\exp(-U(t))\left(\int_t^0 g(s)\,\mathrm{d}s\right)^2\mathrm{d}t.$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

By Assumption 3.1(iii),

$$\mathbb{E}[Y^2] = \mathbb{E}[(Y^+ + Y^-)^2] \le 2\mathbb{E}[(Y^+)^2] + 2\mathbb{E}[(Y^-)^2] < \infty.$$

Also by this assumption, $\int_0^{T_0^+} g(X(s))\,\mathrm{d}s$ and $\int_{T_{N(t)}^+}^t g(X(s))\,\mathrm{d}s$ are bounded in probability. Furthermore,

$$\mathbb{E}[\tau^+ + \tau^-] = \int_{-\infty}^{\infty} \exp(-U(t))\,\mathrm{d}t < \infty$$

since $\pi(t) \propto \exp(-U(t))$ is a probability measure. By the strong law for renewal processes, [10, Theorem 1.7.3], $N(t)/t \to 1/\mathbb{E}[2\tau^+ + 2\tau^-]$ almost surely. It follows from Lemma A.1 (located in the appendix) that

$$\frac{1}{\sqrt{t}} \sum_{i=1}^{N(t)} Y_i \xrightarrow{\mathrm{w}} \mathcal{N}\left(0, \frac{\mathbb{E}[Y^2]}{\mathbb{E}[2\tau^+ + 2\tau^-]}\right) \quad \text{as } t \to \infty,$$

where

$$\mathbb{E}[Y^2] = \mathbb{E}[(Y^+)^2] + \mathbb{E}[(Y^-)^2] - 2\mathbb{E}[Y^+]^2.$$

Combining all terms gives the stated expression for the asymptotic variance. □

3.2. General approach to the CLT

The approach of Section 3.1 is intuitively appealing. However, the required assumptions are very restrictive. In this section we will employ a far more general approach to obtaining a CLT. In particular, this approach allows us to include nonunimodal cases, as well as situations in which the excess switching rate γ in (2.2) is nonzero.

First we recall two key results from the literature which will be helpful for our purposes. Recall the definition of a *petite set* from, for example, [25].

Assumption 3.2. *We assume that $(Z(t))_{t\ge 0}$ is a φ-irreducible continuous-time Markov process in a Borel space E with extended generator L. For a function $f : E \to [1, \infty)$, a petite set $C \in \mathcal{B}(E)$, a constant $b < \infty$ and a function $V : E \to [0, \infty)$, $V \in \mathcal{D}(L)$,*

$$LV(z) \le -f(z) + b\,\mathbf{1}_C(z), \qquad z \in E,$$

where $\mathbf{1}$ is the indicator function.

Proposition 3.1. ([14, Theorem 3.2].) *Suppose that Assumption 3.2 is satisfied. Then $(Z(t))_{t\ge 0}$ is positive Harris recurrent with invariant probability distribution μ and $\mu(f) < \infty$. For some $c_0 < \infty$ and any $|g| \le f$, the Poisson equation*

$$L\phi = \mu(g) - g \tag{3.2}$$

admits a solution ϕ satisfying the bound $|\phi| \le c_0(V + 1)$.

Define a sequence of stochastic processes $(Y_n(t))_{t\ge 0}$, $n \in \mathbb{N}$, by

$$Y_n(t) = \frac{1}{\sqrt{n}}\left(\int_0^{nt} \{\pi(g) - g(Z(s))\}\,\mathrm{d}s\right), \qquad t \ge 0.$$

The following general result establishes sufficient conditions for a FCLT to hold. Part of the results in this section can be obtained simply by verifying the conditions of the following theorem, although, in particular, work needs to be done to find suitable functions f and V satisfying Assumption 3.2.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Proposition 3.2. ([14, Theorem 4.3].) *Suppose that Assumption 3.2 is satisfied. If* $\mu(V^2) < \infty$ *then, for any* $|g| \leq f$*, there exists a constant* $0 \leq \gamma_g < \infty$ *such that under* $\mathbb{P}_\eta$*,* $Y_n \overset{\mathrm{w}}{\to} \gamma_g B$*, with* B *a standard Brownian motion, as* $n \to \infty$ *in* $D[0,1]$ *for any initial distribution* η*. Furthermore, the constant* γ_g^2 *can be defined as* $\gamma_g^2 = 2\int_E \phi(x)\{\pi(g) - g(x)\}\pi(\mathrm{d}x)$*, where* ϕ *is the solution to the Poisson equation given in Proposition 3.1.*

In situations where $\mu(V^2) < \infty$ cannot be established, we will have to establish a weaker (nonfunctional) form of the CLT, which will depend on a CLT for martingales such as [21, Theorem 2.1]. We require the following lemmas, the proofs of which may be found in Appendix A.2.

Lemma 3.1. *Suppose that Assumption 3.2 is satisfied. Let* $g \in \mathcal{M}(E)$ *be measurable, and satisfy* $|g| \leq f$ *and* $\mu(g) = 0$*. Suppose that* ϕ *is a solution to the Poisson equation (3.2) for the generator* L *given by (2.3) and suppose* $\mu(|\phi|) < \infty$*. Define the process*

$$M(t) := \phi(Z(t)) - \phi(Z(0)) + \int_0^t g(Z(s))\,\mathrm{d}s, \quad t \geq 0, \tag{3.3}$$

where $(Z(t))_{t\geq 0}$ *denote trajectories of the zig-zag process. Then* M *is a martingale with respect to the stationary measure* $\mathbb{P}_\mu$*. Define* $\psi(x) := \frac{1}{2}(\phi(x,+1) - \phi(x,-1))$ *and for a given trajectory* $Z(t) = (X(t), \Theta(t))$ *of the zig-zag process, let* $N(t)$ *denote the process counting the switches in* Θ*, and let* $(T_i)_{i=1}^\infty$ *denote the random times at which these switches occur. The quadratic variation process* $[M]$ *and predictable quadratic variation process* $\langle M\rangle$ *admit the following expressions:*

$$[M](t) = 4\sum_{i=1}^{N(t)} \psi^2(X(T_i)), \qquad \langle M\rangle(t) = 4\int_0^t \lambda(X(s), \Theta(s))\psi^2(X(s))\,\mathrm{d}s.$$

Lemma 3.2. *Suppose that Assumption 3.2 holds and* $\pi(x)V(x,\pm 1) \to 0$ *as* $|x| \to \infty$*. Let* $g \in \mathcal{M}(E)$ *such that* $|g| \leq f$ *and* $\mu(g) = 0$*. Let* $\phi\colon E \to \mathbb{R}$ *be as in Proposition 3.1. Define* $\psi(x) := \frac{1}{2}(\phi(x,+1) - \phi(x,-1))$*. Then* ψ *admits the representation (3.5). Furthermore, if, for some* $\delta \in \mathbb{R}$*, we have* $\lim_{x\to\infty} |x|^\delta \pi(x) = 0$ *and*

$$\lim_{|x|\to\infty} \frac{(g(x,+1) + g(x,-1))\pi(x)}{|x|^\delta \pi'(x)} = 0, \tag{3.4}$$

then

$$\lim_{|x|\to\infty} \frac{\psi(x)}{|x|^\delta} = 0.$$

Theorem 3.2. (CLT for the zig-zag process.) *Suppose that Assumption 3.2 is satisfied for the zig-zag process with generator (2.3) and let* $g \in \mathcal{M}(E)$ *satisfy* $|g| \leq f$ *and* $\mu(g) = 0$*. Furthermore, suppose that* V *satisfies* $\mu(V) < \infty$*, or, alternatively,* $\mu(|\phi|) < \infty$*, where* ϕ *is the solution of the Poisson equation given by Proposition 3.1. Let* ψ *be given by*

$$\psi(x) = \frac{1}{2\pi(x)}\int_x^\infty \{g(\xi,+1) + g(\xi,-1)\}\pi(\xi)\,\mathrm{d}\xi \tag{3.5}$$

and define

$$\sigma_g^2 := 4\int_E \lambda(x,\theta)\psi^2(x)\,\mathrm{d}\mu(x,\theta). \tag{3.6}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

If $\sigma_g^2 < \infty$ then under the stationary distribution $\mathbb{P}_\mu$ over the trajectories of the zig-zag process,

$$\frac{1}{\sqrt{t}} \int_0^t g(Z(s))\, \mathrm{d}s \xrightarrow{\mathrm{w}} \mathcal{N}(0, \sigma_g^2) \quad \text{as } t \to \infty.$$

Proof. Let $(Z(t))_{t\geq 0} = (X(t), \Theta(t))_{t\geq 0}$ denote the stationary zig-zag process defined on an underlying probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}_\mu)$. Let ϕ denote the solution of the Poisson equation (3.2), and define the martingale M as in Lemma 3.1, using the fact that $\mu(|\phi|) < \infty$. Indeed, $|\phi| \leq c_0(V+1)$ by Proposition 3.1, and it is assumed that either $\mu(V) < \infty$ or otherwise $\mu(|\phi|) < \infty$. By Lemma 3.2, $\psi(x) := \frac{1}{2}(\phi(x,-1) - \phi(x,+1))$ admits the stated expression. Due to the stationarity of the zig-zag process, M is stationary, and $\sigma_g^2 := \mathbb{E}|M(1)|^2$. By [21, Theorem 2.1], it follows that $M(t)/\sqrt{t}$ converges in distribution to $\mathcal{N}(0, \sigma_g^2)$. Because $(Z(t))_{t\geq 0}$ is stationary under $\mathbb{P}_\mu$, it follows that $\mathcal{L}(\phi(Z(t))) = \mathcal{L}(\phi(Z(0))) = \mu_\star \phi$ (the pushforward of μ by ϕ). As a consequence,

$$\frac{1}{\sqrt{t}}(\phi(Z(t)) - \phi(Z(0))) \xrightarrow{\mathrm{w}} 0.$$

The stated result now follows by combining the obtained limits in (3.3). □

We have now obtained two different expressions for the asymptotic variance, namely (3.1) and (3.6). In cases where both Theorems 3.1 and 3.2 apply these expressions, of course, have the same value. In Appendix A.3 we show the equality of both expressions directly.

We will now introduce some specific assumptions on the switching rates which will suffice to establish a CLT for the zig-zag process.

3.2.1. *The exponentially ergodic case.*

Assumption 3.3. *The switching rate $\lambda\colon E \to \mathbb{R}$ is continuous and there exists a constant $x_0 > 0$ such that*

(i) $\inf_{x\geq x_0} \lambda(x,+1) > \sup_{x\geq x_0} \lambda(x,-1)$, *and*

(ii) $\inf_{x\leq -x_0} \lambda(x,-1) > \sup_{x\leq -x_0} \lambda(x,+1)$.

In other words, there are constants $M^- > m^- \geq 0$ and $M^+ > m^+ \geq 0$ such that

$$\lambda(x,+1) \geq M^+ > m^+ \geq \lambda(x,-1) \quad \text{for all } x \geq x_0,$$
$$\lambda(x,-1) \geq M^- > m^- \geq \lambda(x,+1) \quad \text{for all } x \leq -x_0.$$

It was established in [2, Theorem 5] that under these conditions the zig-zag process is exponentially ergodic.

Theorem 3.3. (CLT and FCLT for the zig-zag process in the exponentially ergodic case.) *Suppose that Assumption 3.3 is satisfied. Let $(Z(t))_{t\geq 0}$ denote the zig-zag process with generator (2.3). Then there exists a unique invariant probability distribution μ on E for $(Z(t))_{t\geq 0}$. Furthermore, there are constants $0 < \alpha^+ \leq M^+ - m^+$ and $0 < \alpha^- \leq M^- m^-$, with $M^\pm$ and $m^\pm$ as above, such that for any function $g \in \mathcal{M}(E)$ satisfying $\mu(g) = 0$ and, for $\theta = \pm 1$,*

$$\limsup_{x\to+\infty} \frac{1}{|x|} \log|g(x,\theta)| < \alpha^+ \quad \text{and} \quad \limsup_{x\to-\infty} \frac{1}{|x|} \log|g(x,\theta)| < \alpha^-, \tag{3.7}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

and if σ_g^2 as given by (3.6) satisfies $\sigma_g^2 < \infty$, then

$$\frac{1}{\sqrt{t}} \int_0^t g(Z(s))\,\mathrm{d}s \xrightarrow{\mathrm{w}} \mathcal{N}(0, \sigma_g^2) \quad \text{as } t \to \infty.$$

If, in addition, $\mu(\mathbf{1}_{[0,\infty)}(x) \exp(2\alpha^+ x)) < \infty$ and $\mu(\mathbf{1}_{(-\infty,0]}(x) \exp(-2\alpha^- x)) < \infty$, then $\sigma_g^2 < \infty$ and, for any initial distribution η on E, under $\mathbb{P}_\eta$ the process $(Z(t))_{t\geq 0}$ satisfies a FCLT, in the sense that

$$\left(\frac{1}{\sqrt{n}} \int_0^{nt} g(Z(s))\,\mathrm{d}s \right)_{t\in[0,1]} \xrightarrow{\mathrm{w}} \sigma_g B \quad \text{as } n \to \infty,$$

where B denotes a standard Brownian motion and the weak convergence is with respect to the Skorokhod topology on $D([0, 1])$.

Although the constants $\alpha^\pm$ are not explicitly specified in the formulation of Theorem 3.3, their construction can be traced in the proof of [2, Theorem 5]. Note that, irrespective of the value of $\alpha^\pm$, (3.7) is satisfied for any subexponential function g.

Proof of Theorem 3.3. Assumption 3.3 implies Assumption 2.1. By Proposition 2.1, it follows that $(Z(t))_{t\geq 0}$ admits a unique invariant probability distribution μ. By tracing the proof of [2, Theorem 5], it follows that there exists a Lyapunov function $V : E \to [0, \infty)$ such that

$$V(x, \theta) = \begin{cases} c^+(\theta) \exp(\alpha^+ x), & x \geq x_0, \\ c^-(\theta) \exp(-\alpha^- x), & x \leq -x_0, \end{cases}$$

for some constants $c^\pm > 0$ and $\alpha^\pm$ as specified in the statement of the theorem, and such that Assumption 3.2 is satisfied with $f := V$. By the stated assumptions on g, possibly after a rescaling by a constant factor, it follows that $|g| \leq f$. By Proposition 3.1, $\mu(f) < \infty$ and there exists a solution ϕ for the Poisson equation (3.2) satisfying $\mu(\phi) = 0$ and $|\phi| \leq c_0(V + 1)$ for some constant $c_0 > 0$. In particular, $\mu(|\phi|) < \infty$. The CLT is therefore a result of Theorem 3.2. Under the stronger assumption, $\mu(V^2) < \infty$ and, therefore, the FCLT follows by Proposition 3.2. □

Remark 3.1. A sufficient condition for $\sigma_g^2 < \infty$ is that $g \in \mathcal{M}(E)$ and $\lambda : E \to [0, \infty)$ are of polynomial growth in x. Indeed, if $g(x, \theta) = O(|x|^\beta)$ then by Lemma 3.2, for any $\delta > \beta$, $\psi(x) = o(|x|^\delta)$. Then, since $\pi(x) = O(\exp(-(M^+ - m^+)x))$ for $x \geq x_0$ (and, similarly, for $x \leq -x_0$), it follows that $\psi^2(x)\lambda(x, \theta)\pi(x)$ has bounded integral.

3.2.2. *Heavy-tailed distributions.*

Assumption 3.4. *We assume that $\lambda : E \to [0, \infty)$ is continuous. There exist constants $\alpha > 0$ and $0 \leq \kappa \leq 1$ such that $\lambda(x, +1) \geq \alpha x^{-\kappa}$ for $x > x_0$ and $\lambda(x, -1) \geq \alpha(-x)^{-\kappa}$ for $x < -x_0$, with $\alpha > 2$ in the $\kappa = 1$ case. Furthermore, $\lambda(x, -1) = 0$ for $x > x_0$ and $\lambda(x, +1) = 0$ for $x < -x_0$.*

Lemma 3.3. *Suppose that Assumption 3.4 is satisfied. Let $1 \leq \beta < \alpha$ in the $\kappa = 1$ case, and $1 \leq \beta < \infty$ in the $\kappa < 1$ case. There exists a norm-like function $V : E \to [0, \infty)$, and a function f of the form $f(x, \theta) = c|x|^{\beta-1}$ for some $c > 0$, and $x_1 > 0$ such that*

$$LV(x, \theta) \leq -f(x, \theta), \qquad |x| > x_1, \quad \theta \in \{-1, +1\}.$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Proof. Let V be given, for $x > x_0$, by $V(x, +1) = kx^\beta$ and $V(x, -1) = (1/\beta)x^\beta$, with

$$k = \begin{cases} \dfrac{2\alpha}{\beta(\alpha - \beta)} & \text{if } \kappa = 1, \\ \dfrac{2}{\beta} & \text{if } 0 \leq \kappa < 1. \end{cases}$$

Then, for $x > x_0$, $LV(x, -1) = -x^{\beta-1}$ and

$$\begin{aligned} LV(x, +1) &= k\beta x^{\beta-1} + \lambda(x, +1)\left(\frac{1}{\beta} - k\right)x^\beta \\ &\leq k\beta x^{\beta-1} + \alpha\left(\frac{1}{\beta} - k\right)x^{\beta-\kappa} \\ &= \begin{cases} -\dfrac{\alpha}{\beta}x^{\beta-1} & \text{if } \kappa = 1, \\ 2x^{\beta-1} - \dfrac{\alpha}{\beta}x^{\beta-\kappa} & \text{if } 0 \leq \kappa < 1. \end{cases} \end{aligned}$$

In the $\kappa < 1$ case, the negative term will dominate for sufficiently large x. It follows in either case that for a suitable constant $c > 0$ and $x_1 \geq x_0$, $LV(x, \pm 1) \leq -cx^{\beta-1} \leq -1$ for all $x \geq x_1$. The situation for $x \leq -x_0$ is completely analogous, and within $[-x_0, x_0]$, the function V can be continuously and differentiably extended. □

Remark 3.2. In fact, for Lemma 3.3, we only require $\alpha > 1$ in the $\kappa = 1$ case, because this allows us to choose $\beta \in [1, \alpha)$. However, in order to obtain $\mu(V) < \infty$ as required for the proof of the following theorem, we need the stronger assumption $\alpha > 2$ in the $\kappa = 1$ case.

Theorem 3.4. (CLT and FCLT for the zig-zag process with a heavy-tailed stationary distribution.) *Suppose that Assumption 3.4 is satisfied. Let $(Z(t))_{t\geq 0}$ denote the zig-zag process with generator (2.3). Then there exists a unique invariant probability distribution μ on E for $(Z(t))_{t\geq 0}$. Suppose that $g \in \mathcal{M}(E)$ with $\mu(g) = 0$ and $g(x, \theta) = O(|x|^{\beta-1})$, where $1 \leq \beta < \alpha - 1$ in the $\kappa = 1$ case and $1 \leq \beta < \infty$ in the $\kappa < 1$ case. Furthermore, suppose that $\sigma_g^2 := 4\int_E \lambda(x, \theta)\psi^2(x)\,\mathrm{d}\mu(x, \theta) < \infty$, where ψ is given by (3.5).*

Then the stationary zig-zag process $(Z(t))_{t\geq 0}$ with switching rates λ satisfies a CLT with asymptotic variance σ_g^2, i.e. under the stationary measure $\mathbb{P}_\mu$ on the trajectories of the zig-zag process,

$$\frac{1}{\sqrt{t}}\int_0^t g(Z(s))\,\mathrm{d}s \xrightarrow{\mathrm{w}} \mathcal{N}(0, \sigma_g^2) \quad \text{as } t \to \infty.$$

If, furthermore, either

- $\kappa < 1$, *or*
- $\kappa = 1$, $\alpha > 3$, *and* $1 \leq \beta < (\alpha - 1)/2$,

then $\sigma_g^2 < \infty$ and, for any initial distribution η on E, under $\mathbb{P}_\eta$ the process $(Z(t))_{t\geq 0}$ satisfies a FCLT, in the sense that

$$\left(\frac{1}{\sqrt{n}}\int_0^{nt} g(Z(s))\,\mathrm{d}s\right)_{t\in[0,1]} \xrightarrow{\mathrm{w}} \sigma_g B \quad \text{as } n \to \infty,$$

where B denotes a standard Brownian motion and the weak convergence is with respect to the Skorokhod topology on $D([0, 1])$.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Proof. Assumption 3.4 implies Assumption 2.1 so that by Proposition 3.1 there is a unique invariant probability distribution μ. If $\kappa = 1$ in Assumption 3.4 then $\mathrm{d}\mu(x,\theta)/\mathrm{d}x = O(|x_0/x|^{\alpha})$. Because $\alpha > 2$ we can choose $1 \leq \beta < \alpha - 1$ in Lemma 3.3, and it follows that the Lyapunov function $V(x,\theta) = O(|x|^{\beta})$ satisfies $\mu(V) < \infty$. If $0 \leq \kappa < 1$ then $\mathrm{d}\mu(x,\theta)/\mathrm{d}x = O(\exp(-\alpha/(1-\kappa)|x|^{1-\kappa}))$ and again $\mu(V) < \infty$. The CLT now follows from Theorem 3.2. Under the stronger assumptions, $\mu(V^2) < \infty$ using the above asymptotic analysis, so that the FCLT follows from Proposition 3.2. □

Remark 3.3. A sufficient condition for $\sigma_g^2 < \infty$ in the $\kappa = 1$ case is that $\alpha > 2$, $1 \leq \beta < \min(\alpha - 1, \frac{1}{2}\alpha)$, and $\lambda(x,+1) = O(x^{-1})$. Indeed, in this case there exists a $\delta \in (\beta, \alpha/2)$. Since $\pi(x) = O(|x|^{-\alpha})$, and $\delta < \alpha$, we have $\pi(x)|x|^{\delta} \to 0$. Furthermore, (3.4) is satisfied as $g(x) = O(|x|^{\beta-1})$ and $\pi(x)/\pi'(x) = O(|x|^{-1})$, so we may deduce from Lemma 3.2 that $\psi(x) = o(|x|^{\delta})$. Hence, $\lambda(x)\psi^2(x)\pi(x) = o(|x|^{2\delta-1-\alpha}) = o(|x|^{-1})$ using the fact that $\delta < \alpha/2$.

3.2.3. *Comparison with Langevin diffusion.* Let A denote the generator of the Langevin diffusion with invariant density $\pi(x) = \exp(-U(x))/k$, i.e.

$$Af(x) = f''(x) - U'(x)f'(x),$$

with domain including at least all twice continuously differentiable functions f for which Af is a bounded continuous function.

Proposition 3.3. *Suppose that $g \in L^2(\pi)$ with $\pi(g) = 0$ and let ψ as in (3.5). If $\psi \in L^2(\pi)$ then under the stationary measure $\mathbb{P}_\pi$, the Langevin diffusion $(X(t))_{t\geq 0}$ with generator A satisfies the CLT with asymptotic variance is given by $\widetilde{\sigma}_g^2 := 2\int_{\mathbb{R}} |\psi(x)|^2\pi(x)\,\mathrm{d}x < \infty$, i.e.*

$$\lim_{t\to\infty} \frac{1}{\sqrt{t}} \int_0^t g(X_s)\,\mathrm{d}s \xrightarrow{\mathrm{w}} \mathcal{N}(0, \widetilde{\sigma}_g^2).$$

Conversely, if $\int_{\mathbb{R}} |\psi(x)|^2\pi(x)\,\mathrm{d}x = \infty$ then $\limsup_{t\to\infty}(1/t)\operatorname{var}_\pi(\int_0^t g(X_s)\,\mathrm{d}s) = \infty$.

Proof. The proof of this result may be found in Appendix A.4. □

In cases where both a CLT holds for the Langevin diffusion and the zig-zag process, and the function of interest g does not depend on θ, we can compare the asymptotic variances, given by

$$\begin{aligned}
\widetilde{\sigma}_g^2 &= 2\int_{\mathbb{R}} \psi^2(x)\pi(x)\,\mathrm{d}x \quad \text{(Langevin asymptotic variance)},\\
\sigma_g^2 &= 2\int_{\mathbb{R}} (\lambda(x,+1) + \lambda(x,-1))\psi^2(x)\pi(x)\,\mathrm{d}x\\
&= 2\int_{\mathbb{R}} (|U'(x)| + 2\gamma(x))\psi^2(x)\pi(x)\,\mathrm{d}x \quad \text{(zig-zag asymptotic variance)},
\end{aligned}$$

where we used (2.2) to obtain the last equality.

Trivially, if $\lambda(x,+1) + \lambda(x,-1) \leq 1$ for all $x \in \mathbb{R}$, the asymptotic variance of the zig-zag process is less than or equal to the asymptotic variance of the Langevin diffusion, but this is a very restrictive condition. More generally, the asymptotic variance of the zig-zag process is smaller than that of the Langevin if the switching rates are small where $\psi^2\pi$ has most of its mass. It is also clear from the above expression that having a nonzero excess switching rate γ increases the asymptotic variance of the zig-zag process.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

3.3. Examples

To illustrate the effectiveness of the developed theory, we consider several examples. We consider

- Gaussian distributions, which have light tails and for which the associated zig-zag process is exponentially ergodic, and
- Student t-distributions, which are heavy tailed so that the associated zig-zag process is not exponentially ergodic.

For both families of distributions we will consider two types of observables: moments and tail probabilities.

3.3.1. *Gaussian distribution.* The family of centered one-dimensional Gaussian distributions $\mathcal{N}(0, \nu^2)$ is described by the potential functions and canonical switching rates

$$U(x) = \frac{x^2}{2\nu^2} \quad \text{and} \quad \lambda(x, \theta) = \left(\frac{\theta x}{\nu^2}\right)^+. \tag{3.8}$$

Example 3.1. (*Moments of a Gaussian distribution.*) First we consider the asymptotic variance associated with the kth moment for positive integer values of k. This corresponds to the mean-zero functional $g(x) = x^k - m_k$, where

$$m_k = \frac{1}{\sqrt{2\pi\nu^2}} \int_{\mathbb{R}} x^k \exp\left(-\frac{x^2}{2\nu^2}\right) \mathrm{d}x = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \nu^k (k-1)!! = \dfrac{\nu^k k!}{2^{k/2}(k/2)!} & \text{if } k \text{ is even.} \end{cases}$$

Assumption 3.1 is satisfied for any $k \geq 0$ so that a CLT holds by Theorem 3.1. The asymptotic variance can be computed using (3.1) to be

$$\sigma_g^2 = \frac{\nu^{2k+1}}{\sqrt{2\pi}} \begin{cases} 2^{k+2}\left(\dfrac{2(k!)}{k+1} - \dfrac{(((k-1)/2)!)^2}{2}\right) & \text{for } k \text{ odd,} \\ \dfrac{8(k!)^2}{2^k((k/2)!)^2} + \dfrac{8(k!)(2^k - k - 2)}{k+1} & \text{for } k \text{ even.} \end{cases}$$

The variance of g under π is given by

$$\mathrm{var}_\pi(g) = \nu^{2k} \begin{cases} (2k-1)!! & \text{for } k \text{ odd,} \\ (2k-1)!! - ((k-1)!!)^2 & \text{for } k \text{ even.} \end{cases}$$

In order to compare the asymptotic variance of the Langevin diffusion, we compute

$$\psi(x) = \exp\left(\frac{x^2}{2\nu^2}\right) \int_x^\infty (\xi^k - m_k) \exp\left(-\frac{\xi^2}{2\nu^2}\right) \mathrm{d}\xi.$$

Expressions for $\psi(x)$ for different values of k are given, along with the computed asymptotic variance for the zig-zag process (σ_g^2) and Langevin diffusion ($\widetilde{\sigma}_g^2$) in Table 1.

For each of these moments, we note that $\sigma_g^2/\widetilde{\sigma}_g^2 \propto \nu^{-1}$, which suggests that for large variance distributions, the variance of an estimator for $\pi(g)$ using the zig-zag process will be considerably lower than that of an estimator generated from a Langevin trajectory.

The result of Theorem 3.1 can be strengthened since by Theorem 3.3 the FCLT holds for this entire family of examples.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

TABLE 1.

k	1	2	3	4
$\mathrm{var}_\pi(g)$	ν^2	$2\nu^4$	$15\nu^6$	$96\nu^8$
$\psi(x)$	ν^2	$\nu^2 x$	$\nu^2(x^2+2\nu^2)$	$\nu^2 x(x^2+3\nu^2)$
σ_g^2	$2\sqrt{\frac{2}{\pi}}\nu^3$	$4\sqrt{\frac{2}{\pi}}\nu^5$	$40\sqrt{\frac{2}{\pi}}\nu^7$	$228\sqrt{\frac{2}{\pi}}\nu^9$
$\widetilde{\sigma}_g^2$	ν^4	ν^6	$11\nu^8$	$42\nu^{10}$

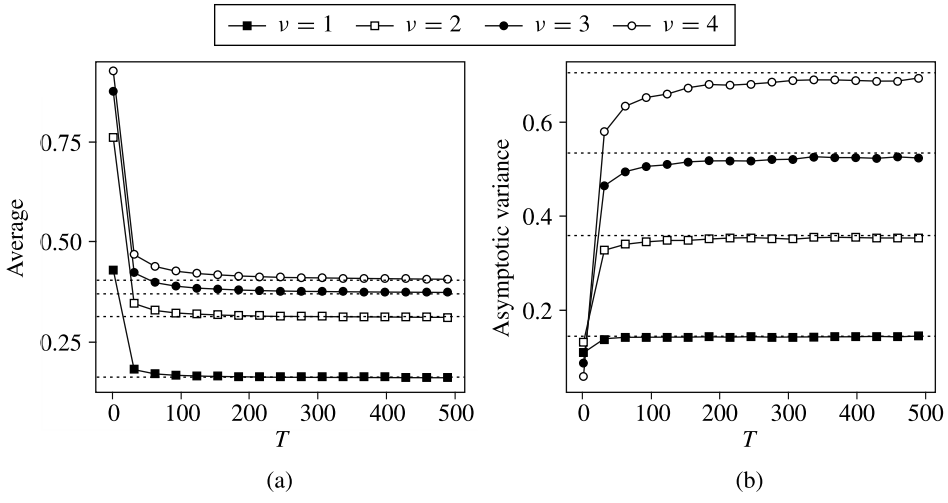


FIGURE 3: Mean and variance estimates for the tail probabilities of a Gaussian $\mathcal{N}(0, \nu^2)$ distribution obtained from simulations compared to predicted estimates. (a) Plot of ergodic average $\pi_T(f)$ of $f(x) = \mathbf{1}_{[1,\infty)}(x)$ as a function of time T, for the one-dimensional zig-zag process ergodic with respect to $\mathcal{N}(0, \nu^2)$ for different values of ν. (b) Corresponding values of $T\,\mathrm{var}[\pi_T(f)]$ as a function of time T, for different values of ν. The dashed lines denote the corresponding asymptotic variance obtained via (15).

Example 3.2. (*Tail probabilities for a Gaussian distribution.*) Next consider the tail probabilities $p_a := \pi(x \geq a)$ for a $\mathcal{N}(0, \nu^2)$-distribution. The potential and associated switching rates are given by (3.8). We have $p_a = 1 - \Phi(a/\nu)$ and $g(x) = \mathbf{1}_{[a,\infty)}(x) - p_a$. Assumption 3.1 is satisfied for any value of $\nu > 0$ so that Theorem 3.1 gives a CLT. Again, using Theorem 3.3 we obtain a FCLT in this family of examples. Computing the necessary integrals in (3.1) gives the asymptotic variance

$$\sigma_g^2 = \frac{-4a(1-p_a)p_a\nu\sqrt{2\pi} + 4(1-2p_a)\nu^2\exp(-a^2/(2\nu^2)) + (8-2\pi)p_a^2\nu^2}{\sqrt{2\pi\nu^2}}, \tag{3.9}$$

while the variance of g is given by $\mathrm{var}_\pi(g) = p_a(1-p_a)$.

In Figure 3 we compare the expression (3.9) with the variance estimated from 10^5 independent simulations of the zig-zag process for different values of ν^2.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

3.3.2. *Student t-distribution.* Consider the family of Student t-distributions with $\nu > 0$ degrees of freedom,

$$\pi(x) \propto \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2},$$

and let λ denote the canonical switching rates, given by

$$\lambda(x, \theta) = \begin{cases} \dfrac{(\nu+1)|x|}{\nu + x^2} & \text{if } \theta x > 0, \\ 0 & \text{if } \theta x \leq 0. \end{cases}$$

Example 3.3. (*Moments for the Student t-distribution.*) For integer values of k with $0 \leq k < \nu$, we can compute the values of the moments to be

$$m_k := \int_{\mathbb{R}} x^k \pi(x)\, \mathrm{d}x = \begin{cases} 0 & \text{if } k \text{ is odd}, \\ \dfrac{\nu^{(k+1)/2}\Gamma((k+1)/2)\Gamma((\nu-k)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} & \text{if } k \text{ is even}. \end{cases}$$

The mean-zero function representing the observable of interest is $g(x) = x^k - m_k$. Assumption 3.1 is satisfied if $k < (\nu - 1)/2$. Moreover, we may apply Theorem 3.4 with $\alpha < \nu + 1$, $\gamma = 1$, and $\beta = k + 1$ to see that in the above cases a FCLT is satisfied under the stated assumption that $k < (\nu - 1)/2$.

This may be compared to the random walk Metropolis (RWM) algorithm. In [17, p. 796] it was established that for a finite variance proposal distribution, the range of parameter values for which a CLT holds is $k < \nu/2 - 1$, which is slightly more restrictive. By tuning the proposal distribution in the RWM to have the same decay in the tails, this range can be improved to $k < \nu/2$.

Using (3.1) we obtain, for the zig-zag process,

$$\sigma_g^2 = \frac{2\nu^{k+1}\Gamma(k+1)\Gamma((\nu-2k-1)/2)}{(1+k)\sqrt{\pi\nu}} - \frac{\nu^{k+1}\Gamma((1+k)/2)^2\Gamma((\nu-k)/2)^2}{\sqrt{\pi\nu}\Gamma(\nu/2)\Gamma((\nu+1)/2)} \quad \text{if } k \text{ is odd}.$$

For k even an also explicit but more cumbersome expression can be obtained from (3.1).

It may be verified that $\psi(x) = O(|x|^{k+1})$, $\lambda(x, \theta) = O(|x|^{-1})$, and $\pi(x) = O(|x|^{-(\nu+1)})$ as $|x| \to \infty$. In particular, the Langevin asymptotic variance, $\widetilde{\sigma}_g^2 = 2\int_{\mathbb{R}} \psi^2(x)\pi(x)\, \mathrm{d}x$ is finite if and only if $k < (\nu - 2)/2$, so that the zig-zag process has finite asymptotic variance for a wider range of combinations of k and ν.

Example 3.4. (*Tail probabilities for the Student t-distribution.*) Suppose that now we wish to consider the behaviour of the zig-zag process with respect to the observable given by the tail probability $p_a = \int_a^\infty \pi(x)\, \mathrm{d}x$ for $a \in \mathbb{R}$. The associated functional of interest is $g(x) = \mathbf{1}_{[a,\infty)}(x) - p_a$. Assume that $a \geq 0$ for simplicity. Assumption 3.1 is satisfied if $\nu > 1$, so that for these values of ν a CLT holds. Using Theorem 3.4, a FCLT holds at least for those cases for which $\nu > 2$.

It may be verified that $\psi(x) = O(|x|)$, $\lambda(x, \theta) = O(|x|^{-1})$, and $\pi(x) = O(|x|^{-(\nu+1)})$ as $|x| \to \infty$. Using Proposition 3.3 the asymptotic variance $\widetilde{\sigma}_g^2 = 2\int_{\mathbb{R}} \psi^2(x)\pi(x)\, \mathrm{d}x$ of the associated Langevin diffusion is finite if and only if $\nu > 2$. So for heavy-tailed distributions the zig-zag process allows for a larger range of parameter values ν with finite asymptotic variance.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

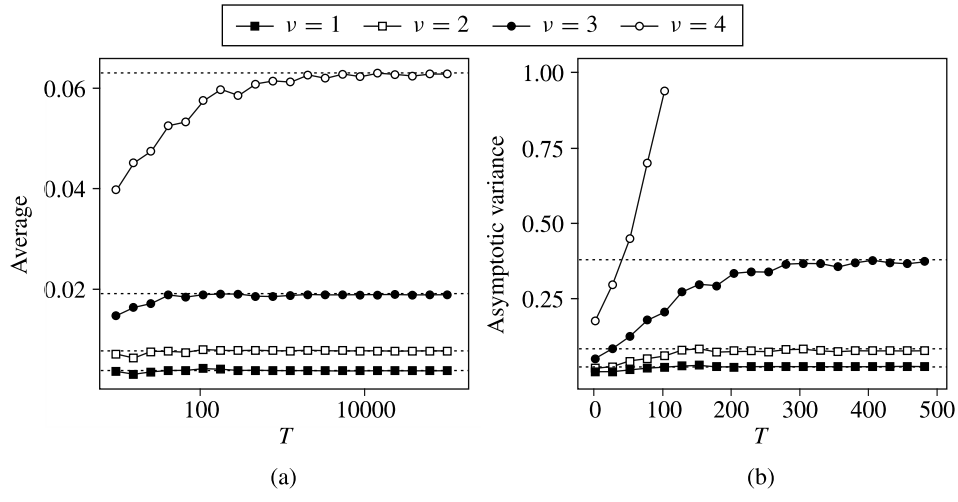


FIGURE 4: Convergence towards the ergodic average and asymptotic variance of the zig-zag process corresponding to the family of Student t-distributions with ν degrees of freedom and with the tail probabilities $\mathbf{1}_{[5,\infty)}$ as observable. The dashed lines denote the analytically derived values. (a) Plot of ergodic average $\pi_T(f)(1/T)\int_0^T f(X(s))\,\mathrm{d}s$ of $f(x) = \mathbf{1}_{[5,\infty)}(x)$ as a function of time T, for the one-dimensional zig-zig process ergodic with respect to (16) for different values of ν. (b) Corresponding values of $T\,\mathrm{var}[\pi_T(f)]$ as a function of time T, for different values of ν, approximated from 10^4 independent realisations of the zig-zig process.

After evaluating the necessary integrals in (3.1), we find the asymptotic variance of the zig-zag process to be

$$\sigma_g^2 = \frac{4(1-2p_a)\nu}{z(\nu-1)}\left(1+\frac{a^2}{\nu}\right)^{(1-\nu)/2} - 4a(1-p_a)p_a + \frac{8p_a^2\nu}{z(\nu-1)} - zp_a^2,$$

where

$$z = \int_{\mathbb{R}} \exp(-U(x))\,\mathrm{d}x = \frac{\sqrt{\nu\pi}\,\Gamma(\nu/2)}{\Gamma((\nu+1)/2)}$$

and, writing ${}_2F_1$ for the hypergeometric function,

$$\begin{aligned} p_a &= \frac{1}{z}\int_a^\infty \exp(-U(x))\,\mathrm{d}x \\ &= \frac{1}{2} - \frac{a\Gamma((\nu+1)/2)\,{}_2F_1(1/2,(\nu+1)/2,3/2,-a^2/\nu)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}. \end{aligned}$$

For $\nu = 2$, the above expressions simplify to

$$\sigma_g^2 = \frac{\sqrt{2}+2a+\sqrt{2}a^2 - a\sqrt{4+2a^2}}{2+a^2} \quad \text{and} \quad \mathrm{var}_\pi(g) = p_a(1-p_a) = \frac{1}{4+2a^2},$$

whereas for other values of ν the expression for the asymptotic variance can typically not be significantly simplified. See Figure 4 for an experimental verification of these results. We see good agreement with theoretical predictions. Also from Figure 4(b) the rescaled variance of the estimator for $\nu = 1$ appears to diverge to ∞ as $T \to \infty$, which suggests that no CLT holds in this case, and, thus, the condition $\nu > 1$ is indeed tight.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

4. Diffusion limit of the zig-zag process

In this section we will consider the one-dimensional zig-zag process with switching rates of the form

$$\lambda(x,\theta) = \max(0, \theta U'(x)) + \gamma(x)$$

for a general nonvanishing space-dependent switching rate γ. An example arising from applications where γ is positive is when zig-zag sampling is used in combination with subsampling, as discussed in [3]. It is observed in simulations that this gives rise to diffusive behaviour. In this section we show that under an appropriate time change the zig-zag process converges weakly to an Itô diffusion, ergodic with respect to π, with a space-dependent diffusion coefficient inversely proportional to the switching rate γ.

We shall focus on behaviour of the zig-zag process in the large $\|\gamma\|_\infty$ limit. To this end, we shall introduce the rescaling $\gamma^\epsilon = \epsilon^{-1}\gamma$, and denote by $Z^\epsilon(t) = (X^\epsilon(t), \Theta^\epsilon(t))$ the corresponding zig-zag process, with generator defined by

$$L^\epsilon f(x,\theta) = \theta \partial_x f(x,\theta) + (\lambda^0(x,\theta) + \gamma^\epsilon(x))[f(x,-\theta) - f(x,\theta)],$$

where $\lambda^0(x,\theta) = \max(0, \theta U'(x))$. Our objective is to prove the following result.

Theorem 4.1. *Suppose that $\gamma \in C^1(\mathbb{R})$ is positive. Consider the process $Z^\epsilon(t) = (X^\epsilon(t), \Theta^\epsilon(t))$ with initial condition $(X^\epsilon(0), \Theta^\epsilon(0)) \sim \eta$ on E. Suppose that the Itô stochastic differential equation (SDE)*

$$\mathrm{d}\xi(t) = -\frac{1}{2}\left(\frac{U'(\xi(t))}{\gamma(\xi(t))} + \frac{\gamma'(\xi(t))}{\gamma(\xi(t))^2}\right)\mathrm{d}t + \sqrt{\frac{1}{\gamma(\xi(t))}}\mathrm{d}W(t), \tag{4.1}$$

where $\xi(0)$ is distributed according to the marginal distribution of η with respect to x, and where $W(t)$ is a standard Brownian motion independent from $\xi(0)$, has a unique weak solution for $t \geq 0$. Then as $\epsilon \to 0$, the process $X^\epsilon(t/\epsilon)$ converges weakly in $C([0,\infty),\mathbb{R})$ to the solution $\xi(t)$ of (4.1).

Remark 4.1. If the process $(\xi(t))_{t\geq 0}$ exists and is nonexplosive then it is ergodic with unique stationary distribution $\pi(x) \propto \exp(-U(x))$.

To prove this result, we will follow an approach similar to that of [13, Theorem 1.5]. The main distinction is that, in [13, Theorem 1.5] the authors introduced a random time-change for the PDMP which produces a limiting SDE with additive noise. On the other hand, the limiting SDE (4.1) is qualitatively different, in particular, it will have multiplicative noise dependent on the switching rate γ and, moreover, is ergodic with respect to the unique stationary distribution π. The proof of Theorem 4.1 will be deferred to Appendix A.5.

Example 4.1. We demonstrate the conclusions of Theorem 4.1 using a simple example. Given $U(x) = x^2/(2\sigma^2)$ consider the family of zig-zag processes $Z^\epsilon(t) = (X^\epsilon(t), \Theta^\epsilon(t))$ with switching rates

$$\lambda^\epsilon(x,\theta) = \max(0, \theta U'(x)) + \frac{1}{\epsilon}\gamma(x), \tag{4.2}$$

where we choose $\gamma(x) = (1 + x^2)$ for a positive parameter $\epsilon > 0$. The resulting process is ergodic, with unique invariant distribution $\pi \sim \mathcal{N}(0, \sigma^2)$. Applying Theorem 4.1 we know

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

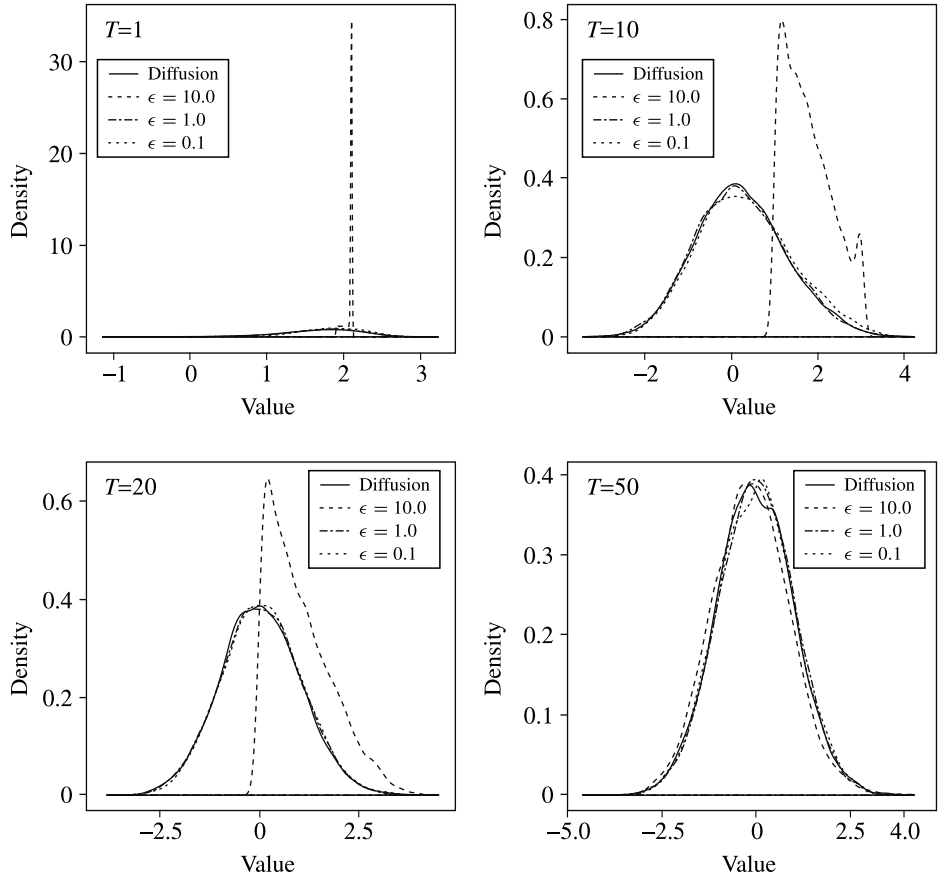


FIGURE 5: Comparison of distribution of zig-zag versus Itô diffusion processes. The dashed lines denote the zig-zag process $X^\epsilon(t)$ with switching rate (4.2) for different ϵ while the solid line denotes the density of the diffusion process given by (4.3).

that, in the limit $\epsilon \to 0$, the time-changed process $X^\epsilon(t/\epsilon)$ will converge weakly to an Itô diffusion process $\xi(t)$ given by the unique solution of

$$\mathrm{d}\xi(t) = -\left(\frac{1}{2\sigma^2}\frac{\xi(t)}{1+\xi^2(t)} + \frac{2\xi(t)}{(1+\xi^2(t))^2}\right)\mathrm{d}t + \sqrt{\frac{1}{1+\xi^2(t)}}\,\mathrm{d}W(t). \tag{4.3}$$

It is straightforward to show that $(\xi(t))_{t\geq 0}$ is an ergodic process with unique invariant distribution π. In Figure 5 we demonstrate this result numerically. Choosing $\sigma^2 = 1$ and $\epsilon = 10, 1, 0.1$, we plot a histogram of the values of $Z^\epsilon(t)$ at values $t/\epsilon = 1$, 10, 20, and 50 over 10^4 independent realisations starting from $X^\epsilon(0) = 2.0$. We compare the result with the corresponding distribution of the diffusion process (4.3) denoted by the solid line. While for larger values of ϵ there is a clear discrepancy between $X^\epsilon(t)$ and $\xi(t)$, as the speed of the switching rate increases, the zig-zag process displays increasing random walk behaviour and shows very good agreement with the diffusion process.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

5. Effective sample size for the zig-zag process

Provided that a CLT holds, for large T, the variance of the estimator $\pi_T(f)$ is given to leading order by $T^{-1}\sigma_f^2$, where σ_f^2 is the asymptotic variance for the observable f. Suppose that we wish to obtain an approximation of $\pi(f)$ within a given error tolerance δ^2 (in the sense of mean-square error), we can obtain an estimate of the amount of time T that the zig-zag process must be simulated, namely

$$T \approx \frac{\sigma_f^2}{\delta^2}. \tag{5.1}$$

In general, (5.1) does not reflect the true cost of simulating the zig-zag sampler. Indeed, as with all continuous-time processes, one can accelerate the mixing of a process simply by introducing a time change $Z^a(t) = Z(at)$ for $a > 0$. In reality, introducing such a time change will increase the number of switches which occur per unit time, thus, increasing the computational effort required to simulate the process up to a given final time T.

Assume that $Z(t)$ is simulated using the direct method (see Algorithm B.1 in Appendix B). The switching times are determined by a Poisson process with inhomogeneous rate $\int_0^t \lambda(X(s), \Theta(s))\,\mathrm{d}s$. Therefore, the average number of switches occurring in time $[0, T]$ is given by

$$N(T) := \mathbb{E}\left[\int_0^T \lambda(X(s), \Theta(s))\,\mathrm{d}s\right].$$

To quantify the average computational cost of simulating a zig-zag sampler, we introduce the average switching rate $N_{\mathrm{S}} = \lim_{t\to\infty} t^{-1}N(t)$, which measures the average number of switches occurring per unit time. Since $Z(t)$ is ergodic, we have

$$\begin{aligned} N_{\mathrm{S}} &= \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{E}[\lambda(X(s), \Theta(s))]\,\mathrm{d}s \\ &= \frac{1}{2}\sum_{\theta=\pm 1}\int_{\mathbb{R}} \lambda(x, \theta)\pi(x)\,\mathrm{d}x \\ &= \frac{1}{2}\int_{\mathbb{R}} (|U'(x)| + 2\gamma(x))\pi(x)\,\mathrm{d}x, \end{aligned} \tag{5.2}$$

where we used the explicit formula for $\lambda(x, \theta)$ given in (2.2). Thus, assuming that N_{S} is finite, after an initial transient period the number of switchings will increase linearly in time with rate N_{S}. In terms of computational cost per simulated unit time interval, it is clear that using canonical switching (i.e. $\gamma = 0$) is the cheapest option. In this case, the average switching rate will be determined entirely by the target distribution.

For the purpose of comparison with other sampling schemes, it would be ideal to obtain an expression for the variance of the estimator $(1/T)\int_0^T f(X_s)\,\mathrm{d}s$ as a function of the number of switches required to simulate the zig-zag process up to time T. For large T the average number of switches that occurred over $[0, T]$ is approximately TN_{S}, where N_{S} is given by (5.2). Over large time-scales the variance of the estimator $\pi_T(f)$ is thus given (for the canonical switching rates, $\gamma = 0$) by

$$\mathrm{var}[\pi_T(f)] \approx \frac{\sigma_f^2 N_{\mathrm{S}}}{N(T)} = \frac{1}{N(T)}\int_{\mathbb{R}} |U'(x)|\psi^2(x)\pi(x)\,\mathrm{d}x \int_{\mathbb{R}} |U'(x)|\pi(x)\,\mathrm{d}x,$$

where $N(T)$ is the number of switches that occurred up to time T and ψ is given by (3.5).

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

A useful measure of the effectiveness of a sampling scheme is the ESS, which provides a measurement of the equivalent number of i.i.d. draws from π which would be required to obtain an estimate for $\pi(f)$ with similar variance. For the zig-zag sampler, it is natural to define the ESS as follows:

$$\mathrm{ESS} := \frac{\mathrm{var}_\pi[f]}{\mathrm{var}[\pi_T(f)]} = \frac{\mathrm{var}_\pi[f]}{\sigma_f^2 N_S} N(T).$$

This expression provides a far more natural measure of the effectiveness of the zig-zag sampler than, for example, (5.1). In particular, it is trivial to check that $\mathrm{var}_\pi[f]/(\sigma_f^2 N_S)$ is invariant under time rescaling $t \to at$ for $a > 0$. The use of the number of switches $N(T)$ as a measure of computational cost is also well justified. One can see from Algorithm B.1 that this coincides with the number of evaluations of the gradient of the log target distribution $U(x)$, which in high dimensions, or in the large data regime for Bayesian inference problems (as considered in [3]) would be the most expensive operation required to compute the next term in the event chain. The ESS is linearly increasing with $N(T)$ by a factor equal to $\mathrm{var}_\pi[f]/(\sigma_f^2 N_S)$, which determines the efficiency of the zig-zag sampler.

Example 5.1. (*Moments of the Gaussian distribution.*) Consider the problem of computing moments x^k of the Gaussian distribution $\mathcal{N}(0, \nu^2)$, where k is a natural number. In this case, we can compute the effective switching rate to be $N_S = (2\pi\nu^2)^{-1/2}$, so that using the expression for the asymptotic variance obtained in Example 3.1, we have, for k odd,

$$\begin{aligned}\frac{\mathrm{ESS}}{N(T)} &= \nu^{2k}\sqrt{2\pi\nu^2}(2k-1)!!\left[\frac{\nu^{2k+1}}{\sqrt{2\pi}}2^{k+2}\left(\frac{2k!}{k+1} - \frac{1}{2}\left(\frac{k-1}{2}\right)!\right)^2\right]^{-1} \\ &= 2\pi(2k-1)!!\left[2^{k+2}\left(\frac{2k!}{k+1} - \frac{1}{2}\left(\frac{k-1}{2}\right)!\right)^2\right]^{-1}, \end{aligned} \tag{5.3}$$

which is independent of ν. A tedious calculation reveals that $\mathrm{ESS} > N(T)$ for all such k. A similar computation gives, for k even,

$$\frac{\mathrm{ESS}}{N(T)} = 2\pi\left\{((2k-1)!! - ((k-1)!!)^2)\left[\frac{8(k!)^2}{2^k((k/2)!)^2} + \frac{8k!(2^k - k - 2)}{k+1}\right]^{-1}\right\}. \tag{5.4}$$

Evaluating numerically the first few moments using (5.3) and (5.4) we obtain the results presented in Table 2. We see that the relation $\mathrm{ESS} > N(T)$ appears to hold for general k. This demonstrates a nonintuitive phenomenon: the effective sample size of the zig-zag process is higher than the number of i.i.d. samples. Thus, an ergodic average generated from a trajectory of the zig-zag process with N switches will tend to have lower variance than a Monte Carlo average of N i.i.d. samples of π.

To demonstrate the performance of the zig-zag sampler, we generate 10^5 independent realisations of the process ergodic with respect to $\mathcal{N}(0, 4)$, and in Figure 6 plot the variance for estimators of the first two moments as a function of N (the maximum number of switches). We also plot the variance for a Monte Carlo average generated from i.i.d. samples, as well as for

TABLE 2.

k	1	2	3	4	5	6
$\mathrm{ESS}/N(T)$	1.570 80	1.570 80	1.178 10	1.322 78	1.220 73	1.334 59

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

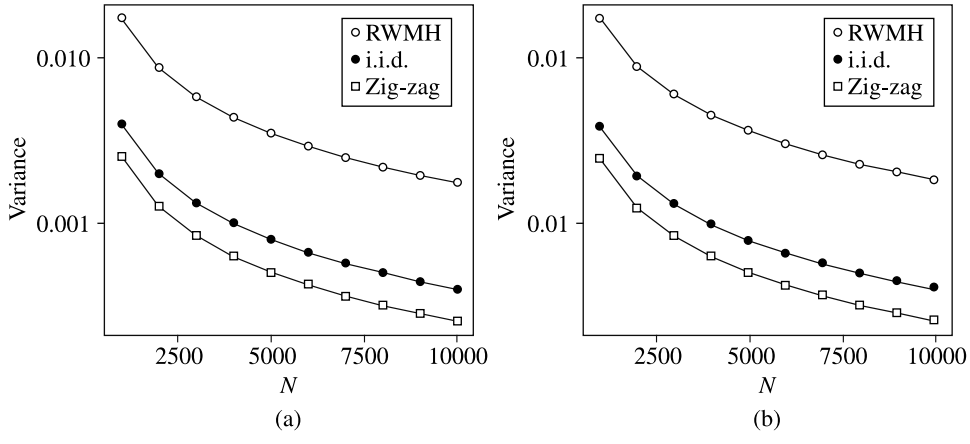


FIGURE 6: Variance of estimator $\pi_T(f)$ of (a) $f(x) = x$ and (b) $f(x) = x^2$ respectively, as a function of the number of switches. For comparison, the variances of the Monte Carlo estimator using i.i.d. samples and a tuned RWMH chain are also displayed.

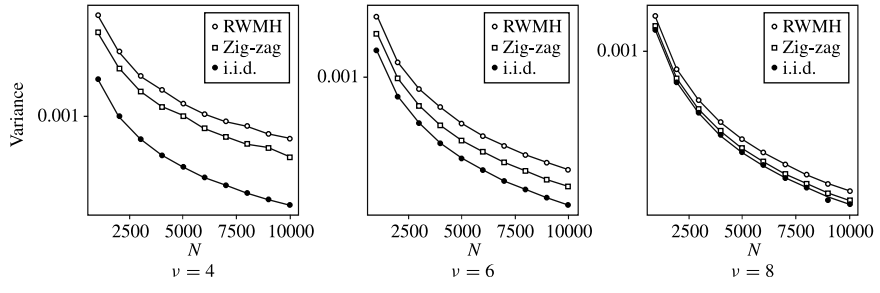


FIGURE 7: Variance of estimator $\pi_T(f)$ of $f(x) = x$ respectively, as a function of the number of switches for the student t-distribution. For comparison, the variance of Monte Carlo estimator using i.i.d. samples and a tuned random walk Metropolis–Hastings chain are also displayed.

a random walk Metropolis–Hastings (RWMH) scheme with manually tuned step-size. We see that even after manually tuning the step-size of the RWMH chain, the asymptotic variance of the corresponding estimator is still an order of magnitude higher than that of the i.i.d. chain and zig-zag sampler. In both cases, the ratio of variances for the zig-zag sampler and i.i.d. average is constant, independent of N, as predicted by (5.3) and (5.4).

The fact that the zig-zag sampler is able to achieve effective sample sizes which beat i.i.d. is a property which is closely tied to the nonreversible nature of the zig-zag process. While we have demonstrated this property for the Gaussian case, one should not interpret this as a general result. Indeed, in the following example we repeat the above experiment for the Student t-distribution, and we show that although the zig-zag sampler outperforms the corresponding RWMH chain, it will not have ESS higher than that of an i.i.d. chain.

Example 5.2. (*Moments of the Student t-distribution.*) Following Example 3.3, we consider once again the problem of the first moment of the Student t-distribution with ν degrees of freedom. In Figure 7 we plot the variance of estimates for the first moment obtained from the zig-zag process using a canonical switching rate (B.3) for $\nu = 4$, 6, and 8. Each point is

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

generated from $M = 10^5$ independent realisations of the process. Note that for the observable $f(x) = x$, Assumption 3.1 holds for each value of ν. As in the previous example, we also plot the variance of a Monte Carlo estimator generated from M i.i.d. samples, as well as from a manually tuned RWMH chain.

In this case the effective sample size of the zig-zag sampler will not be higher than that of the i.i.d. estimator, in general. However, as the degrees of freedom ν goes to ∞, the target distribution becomes increasingly Gaussian, and, for sufficiently large ν, the zig-zag sampler will exhibit lower variance than the corresponding i.i.d. scheme.

Appendix A

A.1. Proof of Proposition 2.1

Because λ is locally bounded, [7, Assumption 3.1] is satisfied, and a piecewise deterministic Markov process can be constructed as described in [7]. Then, by [7, Theorem 5.5], L is the extended generator. The Feller property is established by tracing the proof of [2, Proposition 4], for which only continuity of λ is required. Since λ is continuous and because $\lambda(x, \theta) > 0$ for $\theta x \geq x_0$, we have, in fact, that, for any $x_1 > x_0$, there exists a $c > 0$ such that

$$\lambda(x, \theta) \geq c \quad \text{for all } (x, \theta) \text{ satisfying } x_0 \leq \theta x \leq x_1.$$

The proof that compact sets are petite is now a straightforward adaptation of the proof of [2, Lemma 15], and a Markov process with this property is φ-irreducible; in particular, there exists at most a single invariant measure. The stationarity of μ was established in [2, Proposition 5].

A.2. Technical results towards the CLT

The following lemma is a continuous-time variant of [10, Exercise 2.4.6].

Lemma A.1. *Let (Y_i) be sequence of i.i.d. mean-zero random variables with $\mathbb{E}[Y_i^2] = \sigma^2 < \infty$. Suppose that $a\colon [0, \infty) \to \mathbb{N}$ such that $\lim_{t\to\infty} a(t) = \infty$ and $(N(t))_{t\geq 0}$ is a random process such that $\lim_{t\to\infty} N(t)/a(t) = 1$ in probability. Then*

$$\frac{1}{\sqrt{a(t)}} \sum_{i=1}^{N(t)} Y_i \xrightarrow{\text{w}} \mathcal{N}(0, \sigma^2) \quad \text{as } t \to \infty.$$

Proof. Let $\varepsilon > 0$ and $\gamma > 0$. Let $\beta = \varepsilon\gamma^2/(2\sigma^2)$. Pick $T > 0$ such that, for all $t \geq T$, $|N(t)/a(t) - 1| > \beta$ with probability at most $\varepsilon/2$. For fixed $t \geq T$, let $\Omega(t)$ denote the event in which $|N(t)/a(t) - 1| \leq \beta$. On $\Omega(t)$, $|N(t) - a(t)| \leq \lfloor \beta a(t) \rfloor \leq \beta a(t)$. By Kolmogorov's maximal inequality,

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{\sqrt{a(t)}} \left| \sum_{i=1}^{N(t)} Y_i - \sum_{i=1}^{a(t)} Y_i \right| \geq \gamma \right) &\leq \frac{\varepsilon}{2} + \mathbb{P}\left(\mathbf{1}_{\Omega(t)} \frac{1}{\sqrt{a(t)}} \left| \sum_{i=1}^{N(t)} Y_i - \sum_{i=1}^{a(t)} Y_i \right| \geq \gamma \right) \\
&\leq \frac{\varepsilon}{2} + \mathbb{P}\left(\sup_{m \in \{1, \ldots, \lfloor \beta a(t) \rfloor\}} \frac{1}{\sqrt{a(t)}} \left| \sum_{i=1}^{m} Y_i \right| \geq \gamma \right) \\
&\leq \frac{\varepsilon}{2} + \frac{\beta\sigma^2 a(t)}{\gamma^2 a(t)} \\
&\leq \varepsilon.
\end{aligned}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

This establishes that $(1/\sqrt{a(t)})(\sum_{i=1}^{N}(t)Y_i - \sum_{i=1}^{a(t)} Y_i)$ converges in probability to 0. The stated result now follows from the classical CLT applied to $(1/\sqrt{a(t)})\sum_{i=1}^{a(t)} Y_i$. □

Proof of Lemma 3.1. Since $\phi \in \mathcal{D}(L)$, it follows that M is a local martingale. Due to stationarity, we have

$$\mathbb{E}_\mu|\phi(Z(t))| = \mathbb{E}_\mu|\phi(Z(0)| = \mu(|\varphi|) < \infty$$

and

$$\mathbb{E}_\mu\left|\int_0^t g(Z(s))\,\mathrm{d}s\right| \le \int_0^t \mathbb{E}_\mu|g(Z(s))|\,\mathrm{d}s = t\mu(|g|) < \infty,$$

where we used the fact that $|g| \le f$ and $\mu(f) < \infty$ by Proposition 3.1. It follows that M is a martingale. We have

$$\begin{aligned}
M(t) &= \phi(Z(t)) - \phi(Z(0)) - \int_0^t L\phi(Z(s))\,\mathrm{d}s \\
&= \int_0^t \Theta(s)\phi'(Z(s))\,\mathrm{d}s + \sum_{i=1}^{N(t)}\{\phi(Z(T_i)) - \phi(Z(T_i-))\} \\
&\quad - \int_0^t \{\Theta(s)\phi'(Z(s)) + \lambda(Z(s))(\phi(X(s),\Theta(s)) - \phi(X(s),-\Theta(s)))\}\,\mathrm{d}s \\
&= \sum_{i=1}^{N(t)}\{\phi(X(T_i),\Theta(T_i)) - \phi(X(T_i),-\Theta(T_i))\} \\
&\quad - \int_0^t \lambda(Z(s))(\phi(X(s),\Theta(s)) - \phi(X(s),-\Theta(s)))\,\mathrm{d}s \\
&= -2\sum_{i=1}^{N(t)}\psi(Z(T_i)) + 2\int_0^t \lambda(Z(s))\psi(Z(s))\,\mathrm{d}s,
\end{aligned}$$

where $\psi(x) = \frac{1}{2}(\phi(x,+1) - \phi(x,-1))$. Using [18, Theorems 26.6(vii) and 2.6(viii)] the quadratic variation of M and predictable quadratic variation are given by the stated expressions. This completes the proof. □

In Lemma 3.1 we introduced the function $\psi : \mathbb{R} \to \mathbb{R}$. In the following lemma we collect some useful properties of this function.

Proof of Lemma 3.2. Assume without loss of generality that $\mu(g) = 0$. Writing out the relation $L\phi(x,\theta) = -g(x,\theta)$ for $\theta = \pm 1$ and adding the two equations yields

$$\begin{aligned}
&\frac{\mathrm{d}\phi(x,+1)}{\mathrm{d}x} - \frac{\mathrm{d}\phi(x,-1)}{\mathrm{d}x} - (\lambda(x,+1) - \lambda(x,-1))(\phi(x,+1) - \phi(x,-1)) \\
&\quad = -\frac{(g(x,+1) + g(x,-1))}{2},
\end{aligned}$$

i.e.

$$\psi'(x) - U'(x)\psi(x) = -\frac{(g(x,+1) + g(x,-1))}{2}.$$

This equation may be solved to give

$$\psi(x) = \frac{c}{\pi(x)} + \frac{1}{2\pi(x)}\int_x^\infty \{g(\xi,+1) + g(\xi,-1)\}\pi(\xi)\,\mathrm{d}\xi, \qquad x \in \mathbb{R}. \tag{A.1}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

It remains to verify that the constant c vanishes. By Proposition 3.1, we have $|\phi| \le c_0(V+1)$ and, hence,

$$|\psi(x)| = |\phi(x,+1) - \phi(x,-1)| \le c_0(V(x,+1) + V(x,-1) + 2).$$

By the assumption that $\pi(x)V(x,\pm 1) \to 0$, it therefore follows that $\pi(x)\psi(x) \to 0$ as $|x| \to \infty$. Multiplying (A.1) by π, we have

$$\pi(x)\psi(x) = c + \frac{1}{2}\int_x^\infty \{g(\xi,+1) + g(\xi,-1)\}\pi(\xi)\,\mathrm{d}\xi \to c \quad \text{as } |x| \to \infty,$$

so that necessarily $c = 0$.

Now suppose, for some $\delta \in \mathbb{R}$, that $|x|^\delta \pi(x) \to 0$ as $|x| \to \infty$ and (3.4) holds. Then, since $h(x) := \int_{-\infty}^x \{g(\xi,+1) + g(\xi,-1)\}\pi(\xi)\,\mathrm{d}\xi \to 0$ as $x \to \pm\infty$, using l'Hôpital's rule gives

$$\begin{aligned}
\frac{\psi(x)}{x^\delta} &= \frac{h(x)}{x^\delta \pi(x)} \\
&\sim \frac{h'(x)}{\delta x^{\delta-1}\pi(x) + x^\delta \pi'(x)} \\
&= \frac{(g(x,+1) + g(x,-1))\pi(x)}{\underbrace{\delta x^{\delta-1}\pi(x)}_{\to 0} + x^\delta \pi'(x)} \\
&\to 0 \quad \text{as } |x| \to \infty.
\end{aligned}$$

□

A.3. Equivalence of expressions for asymptotic variance

A natural question to ask is whether the two expressions for asymptotic variance, given by (3.1) and (3.6), are equivalent in cases where both expressions are valid. Suppose that, for an observable g such that $\pi(g) = 0$,

$$\lim_{x \to \pm\infty} \mathrm{e}^{-U(x)} \left(\int_0^x g(y)\,\mathrm{d}y \right)^2 = 0 \tag{A.2}$$

and

$$\lim_{x \to \pm\infty} \mathrm{e}^{U(x)} \left(\int_{-\infty}^x g(y)\,\mathrm{e}^{-U(y)}\mathrm{d}y \right)^2 = 0. \tag{A.3}$$

Assuming that (A.2) and (A.3) hold, and that the potential U satisfies $U(0) = 0$, then we can show that both expressions are equal. Considering the term

$$\begin{aligned}
&\int_0^\infty U'(x)\,\mathrm{e}^{U(x)} \left(\int_{-\infty}^x g(y)\,\mathrm{e}^{-U(y)}\,\mathrm{d}y \right)^2 \mathrm{d}x \\
&\quad = -2\int_0^\infty g(x) \left(\int_{-\infty}^x g(y)\,\mathrm{e}^{-U(y)}\,\mathrm{d}y \right) \mathrm{d}x - \left(\int_{-\infty}^0 \mathrm{e}^{-U(y)} g(y)\,\mathrm{d}y \right)^2,
\end{aligned}$$

where we use (A.3) to eliminate the contribution due to the upper integration limit. Similarly, we have

$$\begin{aligned}
&\int_0^\infty U'(x)\,\mathrm{e}^{-U(x)} \left(\int_0^x g(y)\,\mathrm{d}y \right)^2 \mathrm{d}x \\
&\quad = 2\int_0^\infty \mathrm{e}^{-U(x)} g(x) \left(\int_0^x g(y)\,\mathrm{d}y \right) \mathrm{d}x - \lim_{x \to \infty} \mathrm{e}^{-U(x)} \left(\int_0^x g(y)\,\mathrm{d}y \right)^2,
\end{aligned}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

for which the second term is 0, by (A.2). Exchanging the integrals, we obtain

$$2\int_0^\infty \mathrm{e}^{-U(x)}g(x)\int_0^x g(y)\,\mathrm{d}y\,\mathrm{d}x = 2\int_0^\infty g(x)\int_x^\infty \mathrm{e}^{-U(y)}g(y)\,\mathrm{d}y\,\mathrm{d}x.$$

Since $\pi(g) = 0$, it follows that

$$\int_x^\infty \mathrm{e}^{-U(y)}g(y)\,\mathrm{d}y = -\int_{-\infty}^x \mathrm{e}^{-U(y)}g(y)\,\mathrm{d}y,$$

and so

$$2\int_0^\infty \mathrm{e}^{-U(x)}g(x)\int_0^x g(y)\,\mathrm{d}y\,\mathrm{d}x = -2\int_0^\infty g(x)\int_{-\infty}^x \mathrm{e}^{-U(y)}g(y)\,\mathrm{d}y\,\mathrm{d}x$$

so that

$$\begin{aligned}&\int_0^\infty U'(x)\,\mathrm{e}^{U(x)}\left(\int_{-\infty}^x g(y)\,\mathrm{e}^{-U(y)}\,\mathrm{d}y\right)^2\mathrm{d}x\\ &\quad= \int_0^\infty U'(x)\,\mathrm{e}^{-U(x)}\left(\int_0^x g(y)\,\mathrm{d}y\right)^2\mathrm{d}x - \left(\int_{-\infty}^0 \mathrm{e}^{-U(y)}g(y)\,\mathrm{d}y\right)^2. \qquad \text{(A.4)}\end{aligned}$$

Arguing similarly, if follows that

$$\begin{aligned}&\int_{-\infty}^0 U'(x)\,\mathrm{e}^{U(x)}\left(\int_{-\infty}^x g(y)\,\mathrm{e}^{-U(y)}\,\mathrm{d}y\right)^2\mathrm{d}x\\ &\quad= \int_{-\infty}^0 U'(x)\,\mathrm{e}^{-U(x)}\left(\int_0^x g(y)\,\mathrm{d}y\right)^2\mathrm{d}x + \left(\int_{-\infty}^0 \mathrm{e}^{-U(y)}g(y)\,\mathrm{d}y\right)^2. \qquad \text{(A.5)}\end{aligned}$$

Combining (A.4) and (A.5), it follows immediately that the expressions for asymptotic variance respectively given by (3.1) and (3.6) are equal.

A.4. Proof of Proposition 3.3

Write $(P^s)_{s\geq 0}$ for the Markov semigroup corresponding to the Langevin diffusion, with generator A. By [19, Corollary 1.9], a CLT is satisfied if there exists a constant $c > 0$ such that

$$\langle g, f\rangle_{L^2(\pi)} \leq c\langle -Af, f\rangle_{L^2(\pi)}^{1/2} \quad \text{for all } f \in \mathcal{D}(A),$$

where the domain of A is interpreted as corresponding to the domain of the semigroup generator in $L^2(\pi)$. It is sufficient to check this condition for f in the space $C_c^\infty(\mathbb{R})$ of infinitely differentiable functions with compact support, as this is a core for A. By partial integration on both sides, the above condition then becomes

$$\langle -\psi, f'\rangle_{L^2(\pi)} \leq c\|f'\|_{L^2(\pi)} \quad \text{for all } f \in C_c^\infty(\mathbb{R}),$$

which is satisfied for $c = \|\psi\|_{L^2(\pi)}$. In this case, by [19, Corollary 1.9], the asymptotic variance admits the expression

$$\begin{aligned}\widetilde{\sigma}_g^2 &= 2\langle \varphi, g\rangle_{L^2(\pi)}\\ &= -2\int_{-\infty}^\infty \varphi(x)\left(\frac{1}{\pi(x)}\frac{\mathrm{d}}{\mathrm{d}x}(\pi(x)\varphi'(x))\right)\pi(x)\,\mathrm{d}x\\ &= 2\int_{-\infty}^\infty (\varphi'(x))^2\pi(x)\,\mathrm{d}x,\end{aligned}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

where φ satisfies the Poisson equation $A\varphi = -g$. By the Poisson equation for φ,

$$\pi(x)\varphi'(x) = \int_x^\infty \pi(\xi)g(\xi)\,\mathrm{d}\xi + c.$$

By a similar argument as in the proof of Lemma 3.2, using the fact that $\varphi \in \mathcal{D}(A)$ and, hence, $\varphi' \in L^2(\pi)$, it follows that $c = 0$ and, hence, $\varphi'(x) = -\psi(x)$.

We now prove the converse. To this end, suppose that

$$V := \limsup_{t\to\infty} \frac{1}{t}\operatorname{var}_\pi\left(\int_0^t g(X_s)\,\mathrm{d}s\right) = \int_0^\infty \int_{\mathbb{R}} (P^s g(x))^2 \pi(x)\,\mathrm{d}x\,\mathrm{d}s < \infty,$$

where the equality holds due to [5, Lemma 2.3]. For any $t > 0$, define $g_t := -\int_0^t P^s g\,\mathrm{d}s$. Note that $g_t \in \mathcal{D}(A)$ and satisfies

$$Ag_t = (I - P^t)g. \tag{A.6}$$

We follow the approach of [5, Theorem 3.3]. Below, let f' denote $\mathrm{d}f/\mathrm{d}x$. Given $s \le t$,

$$\begin{aligned}\int (g_t' - g_s')^2\pi\,\mathrm{d}x &= \int_{\mathbb{R}} -A(g_t - g_s)(g_t - g_s)\,\pi(x)\,\mathrm{d}x \\ &= \int_s^t \int_{\mathbb{R}} (P^s g - P^t g)(P^r g)\pi\,\mathrm{d}x\,\mathrm{d}r \\ &= \int_s^t \int_{\mathbb{R}} \{(P^{(r+s)/2}g)^2 - (P^{(r+t)/2}g)^2\}\,\pi\,\mathrm{d}x\,\mathrm{d}r \\ &\le 2\int_s^\infty \int_{\mathbb{R}} (P^r g)^2\pi\,\mathrm{d}x\,\mathrm{d}r.\end{aligned}$$

It follows that the family $(g_t')_{t>0}$ is Cauchy in $L^2(\pi)$, so that it strongly converges to a limit $-\eta \in L^2(\pi)$. The weak formulation of (A.6) is given by

$$-\langle g_t', v'\rangle_{L^2(\pi)} = \langle g, v\rangle_{L^2(\pi)} - \langle P^t g, v\rangle_{L^2(\pi)}, \qquad v \in C_c^\infty(\mathbb{R}). \tag{A.7}$$

We have $\lim_{t\to\infty} P^t g = \pi(g) = 0$, so that by dominated convergence $\langle P^t g, v\rangle_{L^2(\pi)} \to 0$ as $t \to \infty$, and, thus, taking the $t \to \infty$ limit in (A.7) yields

$$\langle \eta, v'\rangle_{L^2(\pi)} = \langle g, v\rangle_{L^2(\pi)}, \qquad v \in C_c^\infty(\mathbb{R}).$$

By the definition of ψ, we also have, for all $v \in C_c^\infty(\mathbb{R})$, $\langle \psi, v'\rangle_{L^2(\pi)} = \langle g, v\rangle_{L^2(\pi)}$, so that $\langle(\psi - \eta), v'\rangle_{L^2(\pi)} = 0$. Hence, in the sense of distributions, $(\psi - \eta)' = 0$, from which it follows (see, for example, [20, Section 21.4]) that $\eta = \psi +$ constant. In order for η to belong to $L^2(\pi)$, by a similar argument as in the proof of Lemma 3.2, the constant should be equal to 0 and, hence, $\psi = \eta \in L^2(\pi)$.

A.5. Proof of Theorem 4.1

In this section we prove Theorem 4.1, following the approach of [12]. To this end, consider the function

$$f(x,\theta) := x + \frac{\epsilon}{2}\frac{\theta}{\gamma(x)} - \frac{\epsilon^2}{2}\frac{\theta}{\gamma^2(x)}\lambda^0(x,\theta), \qquad (x,\theta) \in E.$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Using the fact that

$$\tfrac{1}{2}\theta(\lambda^0(x,+1)+\lambda^0(x,-1)) = \theta\lambda^0(x,\theta) - \tfrac{1}{2}(\lambda^0(x,1)-\lambda^0(x,-1)) = \theta\lambda^0(x,\theta) - \tfrac{1}{2}U'(x),$$

we obtain

$$\begin{aligned} L^\epsilon f(x,\theta) &= \theta - \frac{\epsilon}{2}\frac{\gamma'(x)}{\gamma^2(x)} - \epsilon\frac{\lambda^0(x,\theta)}{\gamma(x)}\theta - \theta + \frac{\epsilon}{2}\frac{1}{\gamma(x)}\theta[\lambda^0(x,-1)+\lambda^0(x,+1)] \\ &\quad + \epsilon^2 R_1(x,\theta), \\ &= -\frac{\epsilon}{2}\frac{\gamma'(x)}{\gamma^2(x)} - \epsilon\frac{\lambda^0(x,\theta)}{\gamma(x)}\theta + \epsilon\frac{1}{\gamma(x)}\left[\theta\lambda^0(x,\theta) - \frac{1}{2}(\lambda^0(x,+1)-\lambda^0(x,-1))\right] \\ &\quad + \epsilon^2 R_1(x,\theta) \\ &= \epsilon b(x) + \epsilon^2 R_1(x,\theta), \end{aligned}$$

where

$$b(x) = -\frac{1}{2}\left(\frac{\gamma'(x)}{\gamma^2(x)} + \frac{U'(x)}{\gamma(x)}\right)$$

and

$$R_1(x,\theta) = -\frac{\lambda^0(x,\theta)}{2\gamma(x)^2}U'(x) + \frac{\lambda^0(x,\theta)\gamma'(x)}{\gamma(x)^3} - \frac{\partial_x\lambda^0(x,\theta)}{2\gamma(x)^2}$$

is a remainder term which is measurable and independent of ϵ. Defining

$$Y^\epsilon(t) := f(X^\epsilon(t),\Theta^\epsilon(t)) \quad \text{and} \quad j^\epsilon(t) := \epsilon b(X^\epsilon(t)) + \epsilon^2 R_1(X^\epsilon(t),\Theta^\epsilon(t)),$$

it follows (using the fact that f is in the domain of the extended generator; see [7, Theorem 5.5]) that

$$M^\epsilon(t) := Y^\epsilon(t) - \int_0^t j^\epsilon(s)\,\mathrm{d}s$$

is a local martingale with respect to the filtration $\mathcal{F}_t^\epsilon$ generated by $\{Z^\epsilon(t) : t \in [0,T]\}$. Similarly, applying the generator to $g(x,\theta) := f^2(x,\theta)$, we obtain

$$\begin{aligned} L^\epsilon g(x,\theta) &= 2\theta x - 2\frac{\gamma(x)}{\gamma(x)}\theta x \\ &\quad + 2\epsilon\left(\frac{1}{2}\partial_x\left(\frac{x}{\gamma(x)}\right) - \frac{\theta x}{\gamma(x)}\left(\lambda^0(x,\theta) - \frac{1}{2}(\lambda^0(x,+1)+\lambda^0(x,-1))\right)\right) \\ &\quad + \epsilon^2 R_2(x,\theta) \\ &= \epsilon(a(x)+2xb(x)) + \epsilon^2 R_2(x,\theta), \end{aligned}$$

where $b(x)$ is as above, $a(x) = 1/\gamma(x)$, and $R_2(x,\theta)$ can be written as $R_2 = R_2^{(1)} + \epsilon R_2^{(2)} + \epsilon^2 R_2^{(3)}$, where the terms

$$\begin{aligned} R_2^{(1)}(x,\theta) &= -\frac{|U'(x)|}{2\gamma(x)^2} - \frac{\theta\gamma'(x)}{2\gamma(x)^3} \\ &\quad + x\left(\frac{2\theta\lambda^0(x,\theta)^2}{\gamma(x)^2} - \frac{U'(x)\lambda^0(x,\theta)}{\gamma(x)^2} + 2\frac{\lambda^0(x,\theta)\gamma'(x)}{\gamma(x)^3} - \frac{\partial_x\lambda^0(x,\theta)}{\gamma(x)^2}\right), \end{aligned}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

$$R_2^{(2)}(x,\theta) = \frac{3}{2}\frac{\lambda^0(x,\theta)\gamma'(x)}{\gamma(x)^4} + \frac{|U'(x)|^2}{4\gamma(x)^3} - \frac{\theta\partial_x\lambda^0(x,\theta)}{2\gamma(x)^4},$$
$$R_2^{(3)}(x,\theta) = \frac{\lambda^0(x,-\theta)^2\lambda^0(x,\theta)}{4\gamma(x)^4} - \frac{\lambda^0(x,\theta)^3}{4\gamma(x)^4} - \frac{\theta\lambda^0(x,\theta)^2\gamma'(x)}{\gamma(x)^5} + \frac{\theta\lambda^0(x,\theta)\partial_x\lambda^0(x,\theta)}{2\gamma(x)^4}$$

are measurable and independent of ϵ. Thus, it follows that

$$N^\epsilon(t) := (Y^\epsilon(t))^2 - \epsilon\int_0^t \{a(X^\epsilon(s)) + 2X^\epsilon(s)b(X^\epsilon(s)) - \epsilon R_2(X^\epsilon(s),\Theta^\epsilon(t))\}\,\mathrm{d}s,$$

is a local martingale with respect to the filtration $\mathcal{F}_t^\epsilon$. We now decompose the square local martingale $(M^\epsilon(t))^2$ into a local martingale term and a remainder. To this end, defining $J^\epsilon(t) = \int_0^t j^\epsilon(s)\,\mathrm{d}s$ and using integration by parts, we obtain

$$\begin{aligned}
(M^\epsilon(t))^2 &= (Y^\epsilon(t))^2 - 2J^\epsilon(t)Y^\epsilon(t) + (J^\epsilon(t))^2\\
&= (Y^\epsilon(t))^2 - 2\int_0^t Y^\epsilon(s)j^\epsilon(s)\,\mathrm{d}s - 2\int_0^t J^\epsilon(s)\,\mathrm{d}Y^\epsilon(s) + 2\int_0^t J^\epsilon(s)j^\epsilon(s)\,\mathrm{d}s\\
&= (Y^\epsilon(t))^2 - 2\int_0^t Y^\epsilon(s)j^\epsilon(s)\,\mathrm{d}s - 2\int_0^t J^\epsilon(s)\,\mathrm{d}M^\epsilon(s)\\
&= N^\epsilon(t) - 2\int_0^t J^\epsilon(s)\,\mathrm{d}M^\epsilon(s)\\
&\quad + \epsilon\int_0^t \{a(X^\epsilon(s)) + 2X^\epsilon(s)b(X^\epsilon(s)) - \epsilon R_2(X^\epsilon(s),\Theta^\epsilon(s))\}\mathrm{d}s\\
&\quad - 2\int_0^t \left(X^\epsilon(s) + \frac{\epsilon}{2}\frac{\Theta^\epsilon(s)}{\gamma(X^\epsilon(s))} - \epsilon^2\frac{\Theta^\epsilon(s)}{2\gamma(X^\epsilon(s))^2}\lambda^0(X^\epsilon(s),\Theta^\epsilon(s))\right)\\
&\quad \times (\epsilon b(X^\epsilon(s)) + \epsilon^2 R_1(X^\epsilon(s),\Theta^\epsilon(s)))\,\mathrm{d}s\\
&= N^\epsilon(t) - 2\int_0^t J^\epsilon(s)\,\mathrm{d}M^\epsilon(s) + \epsilon\int_0^t \{a(X^\epsilon(s)) + \epsilon R_3(X^\epsilon(s),\Theta^\epsilon(s))\}\,\mathrm{d}s,
\end{aligned}$$

where the terms of order ϵ^2 or higher are collected in the remainder term $R_3(x,\theta)$. It follows that

$$H^\epsilon(t) := (M^\epsilon(t))^2 - \epsilon\int_0^t \{a(X^\epsilon(s)) + \epsilon R_3(X^\epsilon(s),\theta^\epsilon(s))\}\,\mathrm{d}s$$

is a local martingale with respect to $\mathcal{F}_t^\epsilon$. Applying the time change $t \to t/\epsilon$, we see that

$$M^\epsilon\left(\frac{t}{\epsilon}\right) = f\left(X^\epsilon\left(\frac{t}{\epsilon}\right),\Theta^\epsilon\left(\frac{t}{\epsilon}\right)\right) - \int_0^t \left\{b\left(X^\epsilon\left(\frac{s}{\epsilon}\right)\right) + \epsilon R_1\left(X^\epsilon\left(\frac{s}{\epsilon}\right),\Theta^\epsilon\left(\frac{s}{\epsilon}\right)\right)\right\}\mathrm{d}s$$

and

$$H^\epsilon\left(\frac{t}{\epsilon}\right) = \left(M^\epsilon\left(\frac{t}{\epsilon}\right)\right)^2 - \int_0^t \left\{a\left(X^\epsilon\left(\frac{s}{\epsilon}\right)\right) + \epsilon R_3\left(X^\epsilon\left(\frac{s}{\epsilon}\right),\Theta^\epsilon\left(\frac{s}{\epsilon}\right)\right)\right\}\mathrm{d}s$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

are local martingales with respect to the filtration $\mathcal{F}_t^\epsilon := \mathcal{F}_{t/\epsilon}$, $t \geq 0$. We now verify the conditions of [11, Theorem VII.4.1] to derive the diffusive limit. To this end, define

$$A^\epsilon(t) := \int_0^t a\left(X^\epsilon\left(\frac{s}{\epsilon}\right)\right) + \epsilon R_3\left(X^\epsilon\left(\frac{s}{\epsilon}\right), \Theta^\epsilon\left(\frac{s}{\epsilon}\right)\right) \mathrm{d}s,$$

$$B^\epsilon(t) := -\frac{\epsilon}{2}\frac{\Theta^\epsilon(t/\epsilon)}{\gamma(X^\epsilon(t/\epsilon))} + \frac{\epsilon^2}{2}\frac{\Theta^\epsilon(t/\epsilon)}{\gamma^2(X^\epsilon(t/\epsilon))}\lambda^0\left(Z^\epsilon\left(\frac{t}{\epsilon}\right)\right) + \epsilon\int_0^t \left\{b\left(X^\epsilon\left(\frac{s}{\epsilon}\right)\right) + \epsilon R_1\left(X^\epsilon\left(\frac{s}{\epsilon}\right), \Theta^\epsilon\left(\frac{s}{\epsilon}\right)\right)\right\} \mathrm{d}s,$$

as well as the stopping time $\tau_R^\epsilon := \inf\{t \geq 0 \colon |X^\epsilon(t/\epsilon)| \geq R \text{ or } |X^\epsilon(t/\epsilon-)| \geq R\}$. From our assumptions, it follows that, for each $R \geq 0$,

$$\sup_{t \leq T \wedge \tau_R^\epsilon} \left| A^\epsilon(t) - \int_0^t a\left(X^\epsilon\left(\frac{s}{\epsilon}\right)\right) \mathrm{d}s \right| \leq \epsilon \sup_{t \leq T \wedge \tau_R^\epsilon} \left| \int_0^t R_3\left(X^\epsilon\left(\frac{s}{\epsilon}\right), \Theta^\epsilon\left(\frac{s}{\epsilon}\right)\right) \mathrm{d}s \right| \leq \epsilon T K_R,$$

and, thus, converges to 0 almost surely as $\epsilon \to 0$. Similarly,

$$\sup_{t \leq T \wedge \tau_R^\epsilon} \left| B^\epsilon(t) - \int_0^t b\left(X^\epsilon\left(\frac{s}{\epsilon}\right)\right) \mathrm{d}s \right| \to 0$$

almost surely as $\epsilon \to 0$. Finally, noting that $X^\epsilon(t)$ is continuous, we have

$$\lim_{\epsilon \to 0} \mathbb{E}\left[\sup_{t \leq T \wedge \tau_R^\epsilon} \left| X^\epsilon\left(\frac{t}{\epsilon}\right) - X^\epsilon\left(\frac{t}{\epsilon-}\right) \right|^2 \right] = 0$$

and, similarly, for every $R \geq 0$,

$$\lim_{\epsilon \to 0} \mathbb{E}\left[\sup_{t \leq T \wedge \tau_R^\epsilon} |A^\epsilon(t) - A^\epsilon(t-)|^2 \right] = \lim_{\epsilon \to 0} \mathbb{E}\left[\sup_{t \leq T \wedge \tau_R^\epsilon} |B^\epsilon(t) - B^\epsilon(t-)|^2 \right] = 0.$$

The conditions of [11, Theorem VII.4.1] are satisfied and, thus, it follows that $(X^\epsilon(t/\epsilon))_{t \geq 0}$ converges in distribution to the solution of the martingale problem for the operator $(G, \mathcal{D}(G))$, where $\mathcal{D}(G) = C_c^2(\mathbb{R})$ and, for $h \in \mathcal{D}(G)$,

$$Gh(x) = b(x)\partial_x h(x) + \frac{1}{2}a(x)\partial_x^2 h(x) = -\frac{1}{2}\left(\frac{\gamma'(x)}{\gamma^2(x)} + \frac{U'(x)}{\gamma(x)}\right)\partial_x h(x) + \frac{1}{2\gamma(x)}\partial_x^2 h(x).$$

Since the well-posedness of this martingale problem is equivalent to the existence and uniqueness of a weak solution $(\xi(t))_{t \geq 0}$ for (4.1), the proof is complete. □

Appendix B. Simulation of the zig-zag process

In this section we describe some computational methods for simulating the process $Z(t) = (X(t), \Theta(t))$ and use results from previous sections in analyzing these methods. As with the rest of this paper, we shall focus, in particular, on the one-dimensional case, referring the reader to [3] for specifics of the general case.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

B.1. Direct simulation of the switching times

Clearly, it is sufficient to be able to simulate the random switching times $(T_i)_{i\in\mathbb{N}}$. Indeed, given initial conditions $(x,\theta)\in E$ and switching times $(T_i)_{i\in\mathbb{N}}$, the process $Z(t)=(X(t),\Theta(t))$ is defined for all $t\geq 0$ as

$$\Theta(t)=(-1)^k\theta,\quad t\in[T_k,T_{k+1}),\qquad X(t)=X(T_k)+(t-T_k)\Theta(T_k),\quad t\in[T_k,T_{k+1}].$$

Given the state $(X(T_0),\Theta(T_0))=(x_0,\theta_0)$ at switching time T_0, the next random switching time is given by $T_1=T_0+\tau$, where τ satisfies

$$\mathbb{P}[\tau>t]=\exp\biggl(-\int_0^t\lambda(x_0+s\theta_0,\theta_0)\,\mathrm{d}s\biggr). \tag{B.1}$$

In the case where $G(t)=\int_0^t\lambda(x_0+s\theta_0,\theta_0)\,\mathrm{d}s$ has an explicitly computable generalised inverse

$$H(y)=\inf\{t\geq 0\colon G(t)\geq y\},\qquad y\in[0,1],$$

then applying an inverse transformation, the random variable $\tau=H(-\log u)$, $u\sim U[0,1]$, satisfies (B.1). An algorithm for simulating $Z(t)$ based on this approach is detailed next.

Algorithm B.1. (*Direct zig-zag sampling.*) *Input:* initial condition $(x,\theta)\in E$.
Output: the event chain $(T_k,X(T_k),\Theta(T_k))_{k=0}^{\infty}$.

1: Set $(T_0,X(T_0),\Theta(T_0))=(0,x,\theta)$.

2: **for** $k=0,1,2,\ldots$ **do**

3: Draw $u\sim U[0,1]$.

4: Set $\tau=H(-\log u)$.

5: Set

$$\begin{aligned}T_{k+1}&=T_k+\tau,\\ X(T_{k+1})&=X(T_k)+\tau\Theta(T_k),\\ \Theta(T_{k+1})&=-\Theta(T_k).\end{aligned}$$

6: **end for**

The computational cost of Algorithm B.1 clearly depends on the switching intensity, i.e. a zig-zag sampler with a higher switching intensity will require more computational cost to be simulated up to a fixed time T. Indeed, while the zig-zag sampler does not reject samples such as a MH scheme, frequent switching will cause the process $Z(t)$ to mix slowly.

Example B.1. (*Sampling from a Gaussian distribution* $\mathcal{N}(0,\nu^2)$.) A straightforward calculation shows that, given $(x,\theta)\in E$, the generalised inverse of $G(t)=\nu^{-2}\int_0^t\max(0,x_0+s\theta)\,\mathrm{d}s$ can be written explicitly as

$$H(z)=\begin{cases}-\theta x+\sqrt{2\nu^2z} & \text{if } \theta x\leq 0,\\ -\theta x+\sqrt{x^2+2z\nu^2} & \text{otherwise,}\end{cases}\qquad \text{for } z>0.$$

In this case, the average switching rate is then given by $N_{\mathrm{S}}=(2\pi\nu^2)^{-1/2}$.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

Example B.2. (*Sampling from a Student t-distribution.*) It is also possible to sample from a Student t-distribution with ν degrees of freedom, i.e.

$$\pi(x) \propto \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \tag{B.2}$$

using the direct zig-zag sampling approach. For this distribution, the canonical switching function is given by

$$\lambda(x, \theta) = \begin{cases} \dfrac{(\nu+1)\theta x}{\nu + x^2} & \text{if } \theta x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{B.3}$$

Given $(x, \theta) \in E$, the generalised inverse of $G(t) = \int_0^t \lambda(x + \theta s, \theta)\, \mathrm{d}s$ can be written as

$$H(z) = \begin{cases} -\theta x + \sqrt{\left(-1 + \exp\left(\dfrac{2z}{1+\nu}\right)\right)\nu} & \text{if } \theta x \leq 0, \\ -\theta x + \exp\left(\dfrac{z}{1+\nu}\right)\left(x^2 + \nu - \nu \exp\left(-\dfrac{2z}{1+\nu}\right)\right)^{1/2} & \text{otherwise.} \end{cases}$$

The average switching rate is equal to the normalization constant for (B.2), i.e. $N_S = \Gamma((\nu + 1)/2)/\sqrt{\nu\pi}\Gamma(\nu/2)$. The resulting process will be ergodic with respect to the target distribution π for all $\nu > 0$. Conditions under which a CLT holds were studied in Section 3.3.

B.2. Sampling with Poisson thinning

In general, we will not be able to compute the generalized inverse of G explicitly. In many cases, however, it is possible to obtain an upper bound $\Lambda(t; x, \theta_0)$ such that $m(t) := \lambda(x_0 + \theta_0 t, \theta_0) \leq \Lambda(t; x_0, \theta_0)$ for all $t \geq 0$, $(x_0, \theta_0) \in E$, and where $\Lambda(t)$ has an explicitly computable inverse $\widetilde{H}$. In this case, one can simulate the random switching times using a standard Poisson thinning approach [23]. Using the upper bound $\Lambda(t; x_0, \theta_0)$, a candidate switching time $T_1 = T_0 + \tau$ is generated, such that

$$\mathbb{P}[\tau > t] = \exp\left(-\int_0^t \Lambda(s; x_0, \theta_0)\, \mathrm{d}s\right).$$

A switch (i.e. $\Theta_1 = -\Theta_0$) will occur at T_1 with probability $m(t_1)/\Lambda(t_1; x_0, \theta_0)$. An algorithm for sampling $Z(t)$ based on this approach is detailed as follows.

Algorithm B.2. (*Zig-zag sampling with thinning.*) *Input:* initial condition $(x, \theta) \in E$. *Output:* the event chain $(T_k, X(T_k), \Theta(T_k))_{k=0}^{\infty}$.

1: Set $(T_0, X(T_0), \Theta(T_0)) = (0, x, \theta)$.

2: **for** $k = 0, 1, 2, \ldots$ **do**

3: Draw $u \sim U[0, 1]$.

4: Set $\tau = \widetilde{H}(-\log u)$.

5: Set

$$T_{k+1} = T_k + \tau,$$
$$X(T_{k+1}) = X(T_k) + \tau\Theta_k.$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

6: With probability
$$\frac{\lambda(X(T_{k+1}), \Theta_k)}{\Lambda(T_{k+1}; X(T_k), \Theta(T_k))},$$
set $\Theta_{k+1} = -\Theta_k$ otherwise $\Theta_{k+1} = \Theta_k$.

7: **end for**

Identifying such a computable upper bound is highly problem specific, however we can highlight two frequently arising scenarios where upper bounds can be easily constructed.

(i) Suppose that the log density is globally bounded, i.e. $|U'(x)| \leq K$ for all $x \in \mathbb{R}$. In this case, we can simply choose $\Lambda(t; x, \theta) = K$. This case arises, in particular, for heavy-tailed distributions, for example the Cauchy distribution with $\pi \propto (1 + x^2)^{-1}$.

(ii) Suppose instead that the second derivative of the log density is absolutely bounded, i.e. $|U''(x)| \leq L$ for all $x \in \mathbb{R}$. In this case, we have
$$\theta U'(x + \theta t) = \theta U'(x) + \int_0^t U''(x + \theta s)\, \mathrm{d}s,$$
so that
$$\lambda(x + \theta t, \theta) \leq \max(0, \theta U'(x) + Lt) := \Lambda(t; x, \theta).$$
For fixed $(x, \theta) \in E$, the integrated intensity function $G(t) = \int_0^t \Lambda(s; x, \theta)\, \mathrm{d}s$ has generalised inverse
$$\widetilde{H}(z) = \begin{cases} \dfrac{-\theta U'(x) + \sqrt{2Lz}}{L} & \text{if } \theta U'(x) \leq 0, \\ \dfrac{-\theta U'(x) + \sqrt{2Lz + U'(x)^2}}{L} & \text{otherwise.} \end{cases}$$
This case arises naturally in various Bayesian inference problems, in particular, logistic regression; see [3, Section 6.5].

The number of switches that occur in a given time interval will depend on the intensity function $\Lambda(t; x, \theta)$, and clearly, a poor choice of this upper bound will cause Algorithm B.2 to undergo many potential switch events which are rejected. In particular, if the process $Z(t)$ is in stationarity then the average switching rate will always be higher or equal to that of the direct scheme described in Algorithm B.1.

B.3. Computing ergodic averages

While the event chain $(X(T_k), \Theta(T_k))_{k=0}^{\infty}$ defines a Markov chain, it will not be ergodic with respect to the target distribution π. To compute an ergodic average for a given observable f, the entire continuous-time realisation must be used as follows:
$$\pi_T(f) = \frac{1}{T} \int_0^T f(X_s)\, \mathrm{d}s.$$
Since the zig-zag process moves linearly between switches, this can be decomposed into a sum of integrals over straight lines. Indeed, for $T = T_K$, for some K, we have
$$\pi_{T_K}(f) = \frac{1}{T_K} \sum_{k=0}^{K-1} \int_{X(T_k)}^{X(T_{k+1})} f(x)\, \mathrm{d}x = \frac{1}{\sum_{k=0}^{K-1} \tau_k} \sum_{k=0}^{K-1} \int_0^{\tau_k} f(X(T_k) + \Theta(T_k)s)\, \mathrm{d}s, \quad \text{(B.4)}$$

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

where $\tau_k = T_{k+1} - T_k$. In many cases, the integral in (B.4) can be computed exactly. For example, the first and pth moment can be computed ergodically via the expressions

$$\pi_{T_K}(x) = \frac{1}{\sum_{k=0}^{K-1} \tau_k} \sum_{k=0}^{K-1} \tau_k X(T_k) + \frac{1}{2}\Theta(T_k)\tau_k^2$$

and

$$\pi_{T_K}(x^p) = \frac{1}{\sum_{k=0}^{K-1} \tau_k} \sum_{k=0}^{K-1} \Theta(T_k) \frac{-X(T_k)^{1+p} + (\tau_k \Theta(T_k) + X(T_k))^{1+p}}{(1+p)},$$

respectively. For more complicated observables it will not be possible to evaluate (B.4) analytically, and one must resort to some form of quadrature scheme, for example, Euler or other higher-order methods.

Acknowledgements

The authors acknowledge the EPSRC for support under grants EP/D002060/1, EP/K014463/1 (Joris Bierkens), EP/J009636/1, EP/L020564/1 (Andrew Duncan), as well as EP/K009788/2 (both authors). Furthermore, we acknowledge support of the Lloyds Registry Foundation through the Alan Turing Institute. We are grateful to the anonymous referee and associate editor for useful suggestions with regards to the mathematical exposition, which have certainly helped to improve this paper.

References

[1] Azaïs, R., Bardet, J.-B., Génadot, A., Krell, N. and Zitt, P.-A. (2014). Piecewise deterministic Markov process—recent results. *ESAIM: Proc.* **44,** 276–290.

[2] Bierkens, J. and Roberts, G. (2016). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Prob.* **27,** 846–882.

[3] Bierkens, J., Fearnhead, P. and Roberts, G. (2016). The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. Preprint. Available at https://arxiv.org/abs/1607.03188.

[4] Bouchard-Côté, A., Vollmer, S. J. and Doucet, A. (2017). The bouncy particle sampler: a non-reversible rejection-free Markov chain Monte Carlo method. To appear in *J. Amer. Statist. Assoc.* Available at http://dx.doi.org/10.1080/01621459.2017.1294075.

[5] Cattiaux, P., Chafaï, D. and Guillin, A. (2012). Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA Lat. Am. J. Prob. Math. Statist.* **9,** 1–39.

[6] Chen, T.-L. and Hwang, C.-R. (2013). Accelerating reversible Markov chains. *Statist. Prob. Lett.* **83,** 1956–1962.

[7] Davis, M. H. A. (1984). Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. With discussion. *J. Roy. Statist. Soc. B* **46,** 353–388.

[8] Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195,** 216–222.

[9] Duncan, A. B., Lelièvre, T. and Pavliotis, G. A. (2016). Variance reduction using nonreversible Langevin samplers. *J. Statist. Phys.* **163,** 457–491.

[10] Durrett, R. (1996). *Probability: Theory and Examples*, 2nd edn. Duxbury Press, Belmont, CA.

[11] Ethier, S. N. and Kurtz, T. G. (2005). *Markov Processes*. John Wiley, New York.

[12] Fontbona, J., Guérin, H. and Malrieu, F. (2012). Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. Appl. Prob.* **44,** 977–994.

[13] Fontbona, J., Guérin, H. and Malrieu, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stoch. Process. Appl.* **126,** 3077–3101.

[14] Glynn, P. W. and Meyn, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Prob.* **24,** 916–931.

[15] Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika* **57,** 97–109.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press

[16] Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S. J. (1993). Accelerating gaussian diffusions. *Ann. Appl. Prob.* **3,** 897–913.
[17] Jarner, S. F. and Roberts, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.* **34,** 781–815.
[18] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
[19] Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104,** 1–19.
[20] Kolmogorov, A. N. and Fomin, S. V. (1975). *Introductory Real Analysis*. Dover, New York.
[21] Komorowski, T., Landim, C. and Olla, S. (2012). *Fluctuations in Markov Processes* (Fund. Principles Math. Sci. **345**). Springer, Heidelberg. 491pp.
[22] Lelièvre, T., Nier, F. and Pavliotis, G. A. (2013). Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Statist. Phys.* **152,** 237–274.
[23] Lewis, P. A. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logist. Quart.* **26,** 403–413.
[24] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1092.
[25] Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markovian processes II: Continuous-time processes and sampled chains. *Adv. Appl. Prob.* **25,** 487–517.
[26] Monmarché, P. (2015). On $\mathcal{H}^1$ and entropic convergence for contractive PDMP. *Electron. J. Prob.* **20,** 1–30.
[27] Monmarché, P. (2016). Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Prob. Math. Statist.* **13,** 357–398.
[28] Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo.* Chapman and Hall, Boca Raton, FL, pp. 113–162.
[29] Ottobre, M., Pillai, N. S., Pinski, F. J. and Stuart, A. M. (2016). A function space HMC algorithm with second order Langevin diffusion limit. *Bernoulli* **22,** 60–106.
[30] Peters, E. A. J. F. and de With, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85,** 026703.
[31] Rey-Bellet, L. and Spiliopoulos, K. (2015). Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity* **28,** 2081–2103.
[32] Rey-Bellet, L., Spiliopoulos, K. (2015). Variance reduction for irreversible Langevin samplers and diffusion on graphs. *Electron. Commun. Prob.* **20,** 16pp.
[33] Rey-Bellet, L. and Spiliopoulos, K. (2016). Improving the convergence of reversible samplers. *J. Statist. Phys.* **164,** 472–494.
[34] Sun, Y., Schmidhuber, J. and Gomez, F. J. (2010). Improving the asymptotic performance of Markov chain Monte–Carlo by inserting vortices. *Adv. Neural Inf. Process. Syst.* **23,** 2235–2243.
[35] Welling, M. and Teh, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. *Proc. 28th Int. Conf. Machine Learning, ICML-11*, pp. 681–688.

https://doi.org/10.1017/apr.2017.22 Published online by Cambridge University Press