TU Delft

Delft University of Technology

A piecewise deterministic Monte Carlo method for diffusion bridges

Bierkens, Joris; Grazzi, Sebastiano; van der Meulen, Frank; Schauer, Moritz

DOI
10.1007/s11222-021-10008-8

Publication date
2021

Document Version
Final published version

Published in
Statistics and Computing

Citation (APA)
Bierkens, J., Grazzi, S., van der Meulen, F., & Schauer, M. (2021). A piecewise deterministic Monte Carlo method for diffusion bridges. *Statistics and Computing*, *31*(3), Article 37. https://doi.org/10.1007/s11222-021-10008-8

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Statistics and Computing (2021) 31:37
https://doi.org/10.1007/s11222-021-10008-8

A piecewise deterministic Monte Carlo method for diffusion bridges

Joris Bierkens[1] · Sebastiano Grazzi[1] · Frank van der Meulen[1] · Moritz Schauer[2,3]

Received: 7 September 2020 / Accepted: 23 March 2021
© The Author(s) 2021

Abstract
We introduce the use of the Zig-Zag sampler to the problem of sampling conditional diffusion processes (diffusion bridges). The Zig-Zag sampler is a rejection-free sampling scheme based on a non-reversible continuous piecewise deterministic Markov process. Similar to the Lévy–Ciesielski construction of a Brownian motion, we expand the diffusion path in a truncated Faber–Schauder basis. The coefficients within the basis are sampled using a Zig-Zag sampler. A key innovation is the use of the *fully local* algorithm for the Zig-Zag sampler that allows to exploit the sparsity structure implied by the dependency graph of the coefficients and by the *subsampling* technique to reduce the complexity of the algorithm. We illustrate the performance of the proposed methods in a number of examples.

Keywords Diffusion bridge · Conditional diffusion · Diffusion process · Faber–Schauder basis · Intractable target density · Local Zig-Zag sampler · Piecewise deterministic Monte Carlo · High-dimensional simulation

1 Introduction

Diffusion processes are an important class of continuous-time probability models which find applications in many fields such as finance, physics and engineering. They naturally arise by adding Gaussian random perturbations (white noise) to deterministic systems. We consider diffusions described by a one-dimensional stochastic differential equation of the form

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = u, \tag{1}$$

where $(W_t)_{t\geq 0}$ is a driving scalar Wiener process defined in some probability space and b is the *drift* of the process. The solution of Eq. (1), assuming it exists, is an instance of one-dimensional time-homogeneous diffusion. We aim to sample X on $[0, T]$ conditional on $\{X_T = v\}$, also known as a *diffusion bridge*.

One driving motivation for studying this problem is estimation for discretely observed diffusions. Here, one assumes observations $\mathcal{D} = \{x_{t_1}, \ldots, x_{t_N}\}$ at observations times $t_1 < \ldots < t_N$ are given and interest lies in estimation of a parameter θ appearing in the drift b. It is well known that this problem can be viewed as a missing data problem as in Peters and With (2012), where one iteratively imputes the missing paths conditional on the parameter and the observations, and then the parameter conditional on the "full" continuous path. Due to the Markov property, the missing paths in between subsequent observations can be sampled independently and each of such segments constitutes a diffusion bridge. As this application requires sampling iteratively many diffusion bridges, it is crucial to have a fast algorithm for this step. We achieve this by adapting the Zig-Zag sampler for the simulation of diffusion bridges. The Zig-Zag sampler is an innovative non-reversible and rejection-free Markov process Monte Carlo algorithm which can exploit the structure present in this high-dimensional sampling problem. It is based on simulating a piecewise deterministic Markov process (PDMP). To the best of our knowledge, this is the first application of PDMPs for diffusion bridge simulation. This method also illustrates the use of a local version of the Zig-Zag sampler in a genuinely high-dimensional setting (arguably even an infinite-dimensional setting).

The problem of diffusion bridge simulation has received considerable attention over the past two decades, see, for

✉ Sebastiano Grazzi
s.grazzi@tudelft.nl

1 Delft Institute of Applied Mathematics (DIAM), Delft University of Technology, Delft, The Netherlands

2 Department of Mathematical Sciences, Chalmers University of Technology, Göteborg, Sweden

3 Department of Mathematical Sciences, University of Gothenburg, Göteborg, Sweden

Published online: 20 April 2021

example, Bladt and Sørensen (2014), Beskos et al. (2006), Roberts and Tweedie (1996), van der Meulen et al. (2018), and Bierkens (2020) and references therein. This far from exhaustive list of references includes methods that apply to a more general setting than considered here, such as multivariate diffusions, conditioning on partial observations and hypo-elliptic diffusions. Among the methods that can be applied, most of the methodologies available are of the acceptance–rejection type and scale poorly with respect to some parameters of the diffusion bridge. For example, if the proposed path is not informed by the target distribution, the probability of accepting the path depends strongly on the discrepancy between the proposed path and the target diffusion bridge measure and usually scales poorly as the time horizon of the diffusion bridge T grows. In contrast, gradient-based techniques which compute informed proposals (e.g. Metropolis-adjusted Langevin algorithm) require the evaluation of the gradient of the target distribution, which, in this case, is a path integral that has to be generally computed numerically and its computational cost is of order T, leading to computational limitations. The present work aims to alleviate such restrictions through the use of a rejection-free method and an exact subsampling technique which reduces the cost of evaluating the gradient. On a more abstract level, our method can be viewed as targeting a probability distribution which is obtained by a push-forward of Wiener measure through a change of measure. It then becomes apparent that the studied problem of diffusion bridge simulation is a nicely formulated non-trivial example problem within this setting to study the potential of simulation based on PDMPs. Our results open new paths towards applications of the Zig-Zag for high-dimensional problems.

1.1 Approach

In this section, we present the main ideas used in this paper.

1.1.1 Brownian motion expanded in the Faber–Schauder basis

Our starting point is the Lévy–Ciesielski construction of Brownian Motion. Define $\bar{\phi}(t) = \sqrt{t}$, $\phi_{0,0}(t) = \sqrt{T}\left((t/T)\mathbf{1}_{[0,T/2]}(t) + (1-t/T)\mathbf{1}_{(1/2,1]}(t)\right)$ and set

$$\phi_{i,j}(t) = 2^{-i/2}\phi_{0,0}(2^i t - jT), \quad \text{for} \quad i = 0, 1, \ldots, \quad j = 0, 1, \ldots 2^i - 1.$$

If $\bar{\xi}$ is standard normal and $\{\xi_{i,j}\}$ is a sequence of independent standard normal random variables (independent of $\bar{\xi}$), then

$$X^N(t) = \bar{\phi}(t)\bar{\xi} + \sum_{i=0}^{N}\sum_{j=0}^{2^i-1}\xi_{i,j}\phi_{i,j}(t) \tag{2}$$

converges almost surely on $[0, T]$ (uniformly in t) to a Brownian motion as $N \to \infty$ [see, for example, Sect. 1.2 of McKean (1969)]. The basis formed by $\bar{\phi}$ and $\{\phi_{i,j}\}$ is known as the Faber–Schauder basis (see Fig. 1). The larger the i, the smaller the support of $\phi_{i,j}$, reflecting that higher-order coefficients represent the fine details of the process. A Brownian bridge starting in u and ending in v can be obtained by fixing $\bar{\xi} = v/\sqrt{T}$ and adding the function $\bar{\bar{\phi}}(t)u = (1-t/T)u \; t \mapsto u(1-t/T)$ to (2). By sampling $\xi^N := (\xi_{0,0}, \xi_{1,0}, \ldots, \xi_{N,2^N-1})$ (which in this case are standard normal), approximate realisations of a Brownian bridge can be obtained.

1.1.2 Zig-Zag sampler for diffusion bridges

Let $\mathbb{Q}^u$ denote the Wiener measure on $C[0, T]$ with initial value $X_0 = u$ [cf. Sect. 2.4 of Karatzas and Shreve (1991)], and let $\mathbb{P}^u$ denote the law on $C[0, T]$ of the diffusion in (1). Under mild conditions on b, the two measures are absolutely continuous and their Radon–Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^u}{\mathrm{d}\mathbb{Q}^u}$ is given by the Girsanov formula. Denote by $\mathbb{P}^{u,v_T}$ and $\mathbb{Q}^{u,v_T}$ the measures of the diffusion bridge and the Wiener bridge, respectively, both starting at u and conditioned to hit a point v at time T. Applying the Bayes' law for conditional expectations (Klebaner 2005, Chapt. 10), we obtain:

$$\frac{\mathrm{d}\mathbb{P}^{u,v_T}}{\mathrm{d}\mathbb{Q}^{u,v_T}}(X) = \frac{q(0,u,T,v)}{p(0,u,T,v)}\frac{\mathrm{d}\mathbb{P}^u}{\mathrm{d}\mathbb{Q}^u}(X), \tag{3}$$

where p and q are the transition densities of X under $\mathbb{P}$, $\mathbb{Q}$, respectively, so that for $s < t$, $p(s,x,t,y)\mathrm{d}y = P(X_t \in \mathrm{d}y \mid X_s = x)$. As p is intractable, the Radon–Nikodym derivative for the diffusion bridge is only known up to proportionality constant. The main idea now consists of rewriting the Radon–Nikodym derivative in (3), evaluating it in X^N and running the Zig-Zag sampler for ξ^N targeting this density. Technicalities to actually get this to work are detailed in Sect. 3. A novelty is the introduction of a *local* version of the Zig-Zag sampler, analogously to the *local bouncy particle sampler* (Bouchard-Côté 2015). This allows for exploiting the sparsity in the dependence structure of the coefficients of the Faber–Schauder expansion efficiently, resulting in a reduction of the complexity of the algorithm. The methodology we propose is derived for one-dimensional diffusion processes with unit diffusivity. However, diffusions with state-dependent diffusivity can be transformed to this setting using the Lamperti transform. (An example is given in Sect. 5.3.) In Sect. 6.1, we generalise the method to multivariate diffusion processes with unit diffusivity, assuming the drift to be a conservative vector field.

1.2 Contributions of the paper

The Faber–Schauder basis offers a number of attractive properties:

(a) The coefficients of a diffusions have a structural conditional independence property (see Sect. 4 and Appendix A) which can be exploited in numerical algorithms to improve their efficiency.
(b) A diffusion bridge is obtained from the unconditioned process by simply fixing the coefficient $\bar{\xi}$.
(c) It will be shown (see, for example, Fig. 8) that the nonlinear component of the diffusion process is typically captured by coefficients ξ_{ij} in equation (2) for which i is small. This allows for a low-dimensional representation of the process and yet a good approximation. Therefore, the approximation error caused by leaving out fine details is equally divided over $[0, T]$, contrary to approaches where a proxy for the diffusion bridge is simulated by Euler discretisation of an SDE governing its dynamics. In the latter case, the discretisation error accumulates over the interval on which the bridge is simulated.
(d) It is very convenient from a computational point of view as each function is piecewise linear with compact support.

We adopt the Zig-Zag sampler (Bierkens et al. 2019) which is a sampler based on the theory of piecewise deterministic Markov processes (see Fearnhead 2018; Bouchard-Côté 2015; Andrieu 2018; Andrieu and Livingstone 2019). The main reasons motivating this choice are:

a. The partial derivatives of the log-likelihood of a diffusion bridge measure usually appear as a path integral that has to be computed numerically (introducing consequently computational burden derived by this step and its bias). The Zig-Zag sampler allows us to replace the gradient of the log-likelihood with an unbiased estimate of it without introducing bias in the target measure. This is done in Sect. 4.4 with the subsampling technique which was presented in Bierkens et al. (2019) for applications for which the evaluation of the log-likelihood is expensive due to the size of the dataset.
b. In the same spirit as the local Bouncy Particle Sampler of Bouchard-Côté (2015) and Mider (2019), the *local* and the *fully local* Zig-Zag sampler introduced in Sect. 4 reduces the complexity of the algorithm improving its efficiency with respect to the standard Zig-Zag algorithm as the dimensionality of the target distribution increases (see Sect. 6.2). This opens the way to high-dimensional applications of the Zig-Zag sampler when the dependency graph of the target distribution is not fully connected and when using subsampling. The factorisation of the log-likelihood and the local method we proposed is reminiscent of other work such as Faulkner (2018), Meulen and Schauer (2017) and Mider et al. (2020).
c. The method is a rejection-free sampler, differing from most of the methodologies available for simulating diffusion bridges.
d. The Zig-Zag sampler is defined and implemented in continuous time, eliminating the choice of tuning parameters appearing, for example, in the proposal density of the Metropolis–Hastings algorithm. This advantage comes at the cost of a more complicated method which relies upon bounding from above rates which are model specific and often difficult to derive (see Sect. 5 for our specific applications).
(e) The process is non-reversible: As shown, for example, in Diaconis (2000), non-reversibility generally enhances the speed of convergence to the invariant measure and mixing properties of the sampler. For an advanced analysis on convergences results for this class of non-reversible processes, we refer to the articles Andrieu (2018) and Andrieu and Livingstone (2019).

The local Zig-Zag sampler relies on the conditional independence structure of the coefficients only. This translates to other settings than diffusion bridge sampling, or other choices of basis functions. For this reason, Sect. 4 describes the algorithms of the sampler in their full generality, without referring to our particular application. A documented implementation of the algorithms used in this manuscript can be found in Roberts and Stramer (2001).

1.3 Outline

In Sect. 2, we set some notation and recap the Zig-Zag sampler. In Sect. 3, we expand a diffusion process in the Faber–Schauder basis and prove the aforementioned conditional dependence. The simulation of the coefficients ξ^N presents some challenges as it is high dimensional and its density is expressed by an integral over the path. We give two variants of the Zig-Zag algorithm which enables sampling in a high-dimensional setting. In particular, in Sect. 4 we present the local and fully local Zig-Zag algorithms which exploit a factorisation of the joint density (Appendix A) and a subsampling technique which, in this setting, is used to avoid the evaluation of the path integral appearing in the density (which otherwise would severely complicate the implementation of the sampler). In Sect. 5, we illustrate our methodology using a variety of examples, validate our approach and compare the Zig-Zag sampler with other benchmark MCMC algorithms. We conclude by sketching the extension of our method to multi-dimensional diffusion bridges, carrying out an informal scaling analysis and providing several remarks for future research (Sects. 6 and 7).

2 Preliminaries

Throughout, we denote by ∂_i the partial derivative with respect to the coefficient ξ_i, the positive part of a function f by $(f)^+$, the ith element and the Euclidean norm of a vector x, respectively, by $[x]_i$ and $\|x\|$. The cardinality of a countable set A is denoted by $|A|$.

2.1 Notation for the Faber–Schauder basis

To graphically illustrate the Faber–Schauder basis, a construction of a Brownian motion with the representation of the basis functions is given in Fig. 1. The Faber–Schauder functions are piecewise linear with compact support. The length of the support and the height of the function are determined by the first index, while the second index determines the location. All basis functions with first index i are referred to as *level* i basis functions. For convenience, we often swap between double and single indexing of Faber–Schauder functions. Denote the double indexing with (i, j) and the single indexing with n. We go from one to the other through the transformations

$$i = \lfloor \log_2(n) \rfloor, \qquad j = n - 2^i, \qquad n = 2^i + j;$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The basis with truncation level N has $M := 2^{N+1} - 1$ coefficients. Let ξ^N denote the vector of coefficients up to level N, i.e.

$$\xi^N := (\xi_{0,0}, \xi_{1,0}, \ldots, \xi_{N,2^N-1}) \in \mathbb{R}^M, \tag{4}$$

and let $X^{\xi^N} := X^N$ when we want to stress the dependencies of X^N on the coefficients ξ^N. Using double indexing, we denote by $S_{i,j} = \operatorname{supp} \phi_{i,j}$.

2.2 The Zig-Zag sampler

A *piecewise deterministic Markov process* (Davis 1993) is a continuous-time process with behaviour governed by random jumps at points in time, but deterministic evolution governed by an ordinary differential equation in between those times (yielding piecewise-continuous realisations). If the differential equation can be solved in closed form and the random event times can be sampled exactly, then the process can be simulated in continuous time without introducing any discretisation error (up to floating number precision) making it attractive from a computational point of view.

By a careful choice of the event times and deterministic evolution, it is possible to create and simulate an ergodic and non-reversible process with a desired unique invariant distribution (Fearnhead 2018). The Zig-Zag sampler (Bierkens et al. 2019) is a successful construction of such a processes. We now recap the intuition and the main steps behind the Zig-Zag sampler.

The *one-dimensional* Zig-Zag sampler is defined in the *augmented space* $(\xi, \theta) \in \mathbb{R} \times \{+1, -1\}$, where the first coordinate is viewed as the position of a moving particle and the second coordinate as its velocity. The dynamics of the process $t \mapsto (\xi(t), \theta(t))$ (not to be confused with the time indexing the diffusion process) are as follows: starting from $(\xi(0), \theta(0))$,

(a) its flow is deterministic and linear in its first component with direction $\theta(0)$ and constant in its second component until an event at time τ occurs. That is, $(\xi(t), \theta(t)) = (\xi(0) + t\theta(0), \theta(0))$, $0 \le t \le \tau$.
(b) At an event time τ, the process changes the sign of its velocity, i.e. $(\xi(\tau), \theta(\tau)) = (\xi(\tau-), -\theta(\tau-))$.

The event times are simulated from an inhomogeneous Poisson process with specified rate $\lambda : (\mathbb{R} \times \{1, -1\}) \to \mathbb{R}^+$ such that $P(\tau \in [t, t+\epsilon]) = \lambda(\xi(t), \theta(t))\epsilon + o(\epsilon), \epsilon \downarrow 0$.

The *d-dimensional* Zig-Zag sampler is conceived as the combination of d one-dimensional Zig-Zag samplers with rates $\lambda_i(\xi, \theta)$, $i = 1, \ldots, d$, where the rates create a coupling of the independent coordinate processes. The following result provides a sufficient condition for the d-dimensional Zig-Zag sampler to have a particular d-dimensional target density π as invariant distribution. Assume that the target d-dimensional distribution has strictly positive density with respect to the Lebesgue measure, i.e.

$$\pi(\mathrm{d}\xi) \propto \exp(-\psi(\xi))\mathrm{d}\xi, \qquad \xi \in \mathbb{R}^d.$$

Define the *flipping function* as $F_i(\theta) = (\theta_1, \ldots, -\theta_i, \ldots, \theta_d)$, for $\theta \in \{-1, +1\}^d$. For any $i = 1, \ldots, d$ and $(\xi, \theta) \in \mathbb{R}^d \times \{1, -1\}^d$, the Zig-Zag process with Poisson rates satisfying

$$\lambda_i(\xi, \theta) - \lambda_i(\xi, F_i(\theta)) = \theta_i \partial_i \psi(\xi), \tag{5}$$

has π as invariant density. Condition (5) is derived in the supplementary material of Bierkens et al. (2019). Condition (5) is equivalent to

$$\lambda_i(\xi, \theta) = (\theta_i \partial_i \psi(\xi))^+ + \gamma_i(\xi) \tag{6}$$

for some $\gamma_i(\xi) \ge 0$. Throughout, we set $\gamma_i(\xi) = 0$ because generally the algorithm is more efficient for lower Poisson event intensity (see, for example, Andrieu 2018, Sect. 5.4).

Assume the target density is $\pi(\xi) = c\tilde{\pi}(\xi)$. The process targets the specific distribution function through the Poisson rate λ which is a function of the gradient of $\xi \mapsto \psi(\xi) = -\log(\tilde{\pi}(\xi))$, so that any proportionality factor of the density disappears. Throughout we refer to the function ψ as the

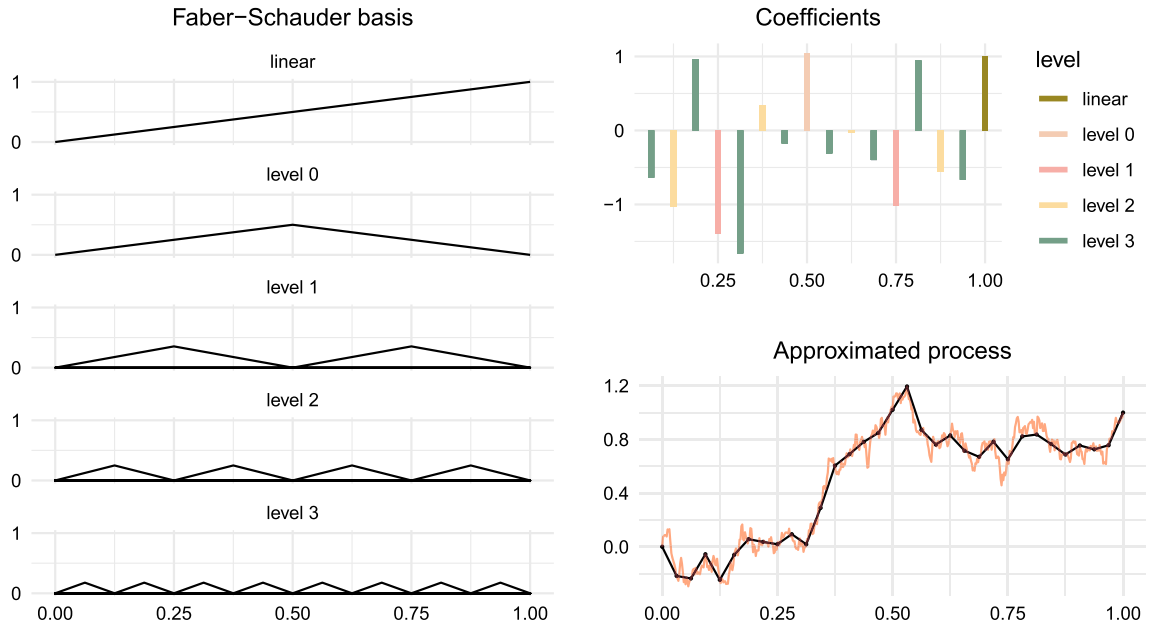

Fig. 1 Lévy–Ciesielski construction of a Brownian motion on $(0, 1)$. On the left the Faber–Schauder basis functions up to level $N = 3$, on the top right the values of the corresponding coefficients located at the peak of their relative FS basis function and on the bottom right the resulting approximated Brownian path X^N (black line) compared with a finer approximation (red line). The truncated sum defines the process in $2^{N+1} + 1$ finite dyadic points (black dots) with linear interpolation in between points. A finer approximation corresponds to Brownian fill-in noise between any two neighbouring dyadic points. (Color figure online)

energy function. As opposed to standard Markov chain Monte Carlo methods, the process is not reversible and it is defined in continuous time.

Example 2.1 Consider a d-dimensional Gaussian random variable with mean $\mu \in \mathbb{R}^d$ and positive-definite covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. Then,

- $\pi(\xi) \propto \exp\left(-(\xi-\mu)'\Sigma^{-1}(\xi-\mu)/2\right)$,
- $\partial_k \psi(\xi) = \left[\Sigma^{-1}(\xi-\mu)\right]_k$,
- $\lambda_k(\xi,\theta) = \left(\theta_k[\Sigma^{-1}(\xi-\mu)]_k\right)^+$.

Notice that if Σ is diagonal, then $\lambda_k(\xi,\theta) = 0$ whenever the process is directed towards the mean so that no jump occurs in the kth component when one of the following conditions is satisfied: $(\theta_k = -1, \xi_k - \mu_k \geq 0)$ or $(\theta_k = 1, \xi_k - \mu_k \leq 0)$. In Fig. 2, we simulate a realisation of the Zig-Zag sampler targeting a univariate standard normal random distribution.

Algorithm 1 shows the standard implementation of the Zig-Zag sampler. After initialisation, the first event time τ^* is determined by taking the minimum of event times $\tau_1, \tau_2, \ldots, \tau_d$ simulated according to the Poisson rates $\lambda_i, i = 1, 2, \ldots, d$. At event time τ^*, the velocity vector becomes $\theta(\tau^*) = F_{i^*}(\theta)$, with $i^* = \arg\min(\tau_1, \ldots, \tau_d)$. The algorithm iterates this step moving forward each time until the next simulated event time exceeds the final clock τ_{final}.

Although we consider the velocities for each dimension of a d-dimensional Zig-Zag process to be either 1 or -1, these can be taken to be any nonzero values $(\theta_i, -\theta_i)$ for $i = 1, \ldots, d$. A fine-tuning of $\theta_1, \ldots, \theta_N$ can improve the performance of the sampler. Note that the only challenge in implementing Algorithm 1 lies on the simulation of the waiting times which correspond to the simulation of the first event time of d inhomogeneous Poisson processes (IPPs) with rates $\lambda_1, \lambda_2, \ldots, \lambda_d$ which are functions of the state space (ξ, θ) of the process. Since the flow of the process is linear and deterministic, the Poisson rates are known at each time and are equal to

$$\lambda_i(t; \xi, \theta) = \lambda_i(\xi + t\theta, \theta), \qquad i = 1, 2, \ldots, d.$$

To lighten the notation, we write $\lambda_i(t) := \lambda_i(t; \xi, \theta)$ when ξ, θ are fixed. Given an initial position ξ and velocity θ, the waiting times $\tau_1, \ldots, \tau_d$ are computed by finding the roots for x of the equations

$$\int_0^x \lambda_i(s)\mathrm{d}s + \log(u_i) = 0, \qquad i = 1, 2, \ldots, d, \tag{7}$$

where $(u_i)_{i=1,2,\ldots,d}$ are independent realisations from the uniform distribution on $(0, 1)$. When it is not possible to find roots of equation (7) efficiently; for example, in closed form, it suffices to find upper bounds for the rate functions for which this is possible, Sect. 4.4 treats this problem for

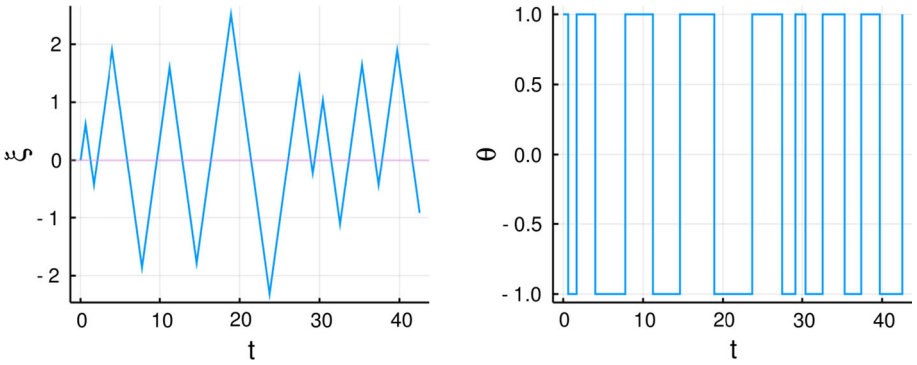

Fig. 2 One-dimensional Zig-Zag targeting a Gaussian random variable $\mathcal{N}(0, 1)$. Left: $t \mapsto \xi(t)$, right: $t \mapsto \theta(t)$

our particular setting. The linear evolution of the process and the jumps of the velocities are always trivially computed and implemented.

Algorithm 1 returns a *skeleton* of values corresponding to the position of the process at the event times. From these values, it is straightforward to reconstruct the continuous path of the Zig-Zag sampler. Given a sample path of the Zig-Zag sampler from 0 to τ_{final}, we can obtain a sample from the target distribution in the following way:

- Denote by $\xi(\tau)$ the value of the vector ξ at the Zig-Zag clock $\tau < \tau_{\text{final}}$. Fixing a sample frequency $\Delta\tau$, we can produce a sample from the density π by taking the values of the random vector ξ at time $\tau_{\text{burn-in}} + \Delta\tau, \tau_{\text{burn-in}} + 2\Delta\tau, \ldots, \tau_{\text{final}}$ where $\tau_{\text{burn-in}}$ is the initial burn-in time taken to ensure that the process has reached its stationary regime. Throughout the paper, we create samples using this approach.

Algorithm 1 Standard d-dimensional Zig-Zag sampler (Bierkens et al. 2019)

procedure ZIGZAG($\tau_{\text{final}}, \xi, \theta$)
  Initialise $k = 1$, $t = 0$
  $\tau_j \sim \text{IPP}\,(\lambda_j(\cdot; \xi, \theta))$, $j = 1, \ldots, d$ ▷ Draw from Inhomogeneous Poisson process (IPP)
  while $t \leq \tau_{\text{final}}$ **do**
    $\tau^*, i^* \leftarrow \text{findmin}(\tau_1, \ldots, \tau_d)$
    Update: $\xi \leftarrow \xi + \theta(\tau^* - t)$
    Update: $\theta_{i^*} \leftarrow -\theta_{i^*}$; $t \leftarrow \tau^*$
    Save $\xi^{(k)} \leftarrow \xi$; $t^{(k)} \leftarrow t$
    for $j = 1, \ldots, d$ **do**
      $\tau_j \sim t + \text{IPP}(\lambda_j(\cdot; \xi, \theta))$
    end for
    $k \leftarrow k + 1$
  end while
  return Skeletons $(\xi^{(l)}, t^{(l)})_{l=1,\ldots,k-1}$
end procedure

2.3 Zig-Zag sampler for Brownian bridges

The previous subsections contain all ingredients necessary to run the Zig-Zag sampler in a finite-dimensional projection of the Brownian bridge measure $\mathbb{Q}^{0,v}$ on the interval $[0, T]$. We fix $\bar{\xi}$ to v and run the Zig-Zag sampler for ξ^N as defined in (4) targeting a multivariate normal distribution. Figure 3 shows 100 samples obtained from one sample run of the Zig-Zag sampler where the coefficients are mapped to samples paths using (2). The final clock of the Zig-Zag is set to $\tau_{\text{final}} = 500$ with initial burning $\tau_{\text{burn-in}} = 10$.

Both Brownian motion and the Brownian bridge are special in that all coefficients in the Faber–Schauder basis are independent. Of course, these processes can directly be simulated without need of a more advanced method like the Zig-Zag sampler. However, for a diffusion process with nonzero drift this property is lost. Nevertheless, we will see that when the process is expanded in the Faber–Schauder basis, many coefficients are still *conditionally* independent. This implies that the dependency graph of the joint density of the coefficients is sparse. We will show in Sect. 4 how this property can be exploited efficiently using the Zig-Zag sampler in its local version.

3 Faber–Schauder expansion of diffusion processes

We extend the results of Sect. 2 to one-dimensional diffusions governed by the SDE in (1). Although the density is defined in infinite-dimensional space, in this section we justify both intuitively and formally that the diffusion can be approximated to arbitrary precision by considering a finite-dimensional projection of it.

The intuition behind using the Faber–Schauder basis is that, under mild assumptions on the drift function b, any diffusion process behaves locally as a Brownian motion. Expanding the diffusion process with the Faber–Schauder functions, this notion translates to the existence of a level N such that the random coefficients at higher levels which

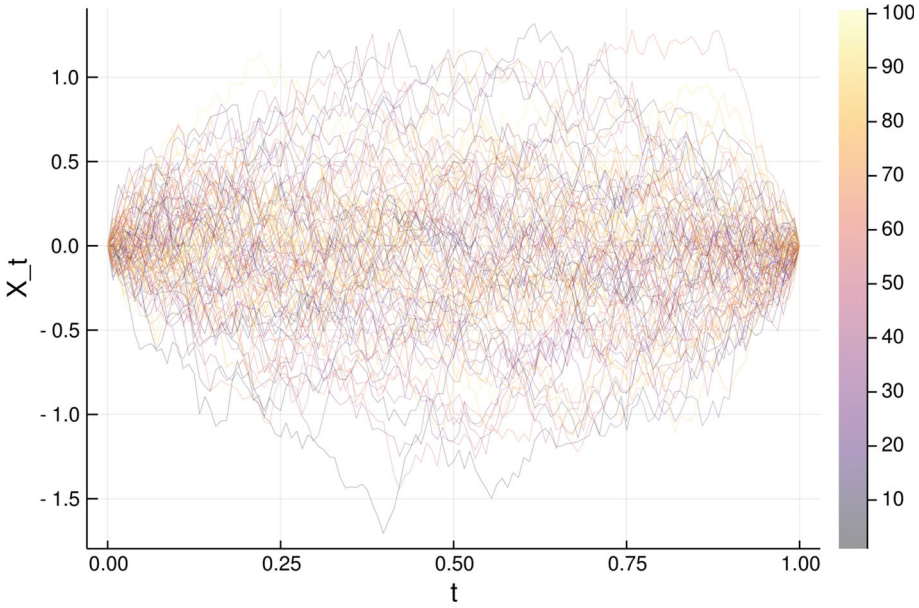

Fig. 3 100 samples from the Brownian bridge measure starting at 0 and hitting 0 at time 1 obtained by one run of the Zig-Zag sampler targeting the coefficients relative to the measure expanded with the Faber–Schauder basis. The resolution level is fixed to $N = 6$ and the Zig-Zag clock to $\tau_{\text{final}} = 500$ and initial burn in $\tau_{\text{burn-in}} = 10$

are associated with the Faber–Schauder basis are approximately independent standard normal and independent from ξ^N under the measure $\mathbb{P}$.

Define the function $Z_t : \mathbb{R}^+ \times C[0,T] \to \mathbb{R}^+$ given by

$$Z_t(X) = \exp\left(\int_0^t b(X_s)\mathrm{d}X_s - \frac{1}{2}\int_0^t b^2(X_s)\mathrm{d}s\right) \tag{8}$$

where the first integral is understood in the Itô sense and $X \equiv (X_s,\ s \in [0,T])$.

Assumption 3.1 Z_t is a $\mathbb{Q}$-martingale.

For sufficient conditions for verifying that this assumption applies, we refer to Remark 3.6, Remark 3.9 and Liptser et al. (2013), Chapter 6.

Theorem 3.2 (Girsanov's theorem) *If Assumption 3.1 is satisfied,*

$$\frac{\mathrm{d}\mathbb{P}^u}{\mathrm{d}\mathbb{Q}^u}(X) = Z_T(X). \tag{9}$$

Moreover, a weak solution of the stochastic differential equation exists which is unique in law.

Proof This is a standard result in stochastic calculus (see Liptser et al. 2013, Sect. 6). □

As we consider diffusions on $[0,T]$ with T fixed, we denote $Z(X) := Z_T(X)$. Due to the appearance of the stochastic Itô integral in $Z(X)$, we cannot substitute for X its truncated expansion in the Faber–Schauder basis. Clearly, whereas the approximation has finite quadratic variation, X has not. Assuming that b is differentiable and applying Itô's lemma to the function $B(x) = \int_0^x b(s)\mathrm{d}s$, the stochastic integral can be replaced and Eq. (8) is rewritten as

$$Z(X) = \exp\left(B(X_T) - B(X_0) - \frac{1}{2}\int_0^T \left(b^2(X_s) + b'(X_s)\right)\mathrm{d}s\right), \tag{10}$$

where b' is the derivative of b.

Definition 3.3 Let X be a diffusion governed by (1). Let X^N be the process derived from X by setting to zero all coefficients of level exceeding N in its Faber–Schauder expansion [see Eq. (2)]. Set

$$Z^N(X) = \exp\left(B\left(X_T^N\right) - B\left(X_0^N\right) - \frac{1}{2}\int_0^T \left[b^2\left(X_s^N\right) + b'\left(X_s^N\right)\right]\mathrm{d}s\right).$$

We define the approximating measure $\mathbb{P}_N$ by the change of measure

$$\frac{\mathrm{d}\mathbb{P}_N^u}{\mathrm{d}\mathbb{Q}^u}(X) = \frac{Z^N(X)}{c_N}, \tag{11}$$

where $c_N = \mathbb{E}_{\mathbb{Q}}\left(Z^N(X)\right)$.

Note that the measure $\mathbb{P}_N^u$ associated with the approximated stochastic process is still on an infinite-dimensional space and such that the joint measure of random coefficients ξ^N is different from the one under $\mathbb{Q}^u$, while the remaining coefficients stay independent standard normal and independent from ξ^N. This is equivalent to approximating the diffusion process at finite dyadic points with Brownian noise fill-in in between every two points. We now fix the final point v_T by setting $\bar{\xi} = v_T$. Define the *approximated stochastic*

bridge with measure $\mathbb{P}_N^{u,v_T}$ in an analogous way of equation (11), so that

$$\frac{\mathrm{d}\mathbb{P}_N^{u,v_T}}{\mathrm{d}\mathbb{Q}^{u,v_T}}(X) = \frac{Z^N(X)}{c_N^{v_T}}. \tag{12}$$

where $c_N^{v_T} = \mathbb{E}_{\mathbb{Q}^{u,v_T}}\left(Z^N(X)\right)$. The following is the main assumption made.

Assumption 3.4 The drift b is continuously differentiable, and $b^2 + b'$ is bounded from below.

Theorem 3.5 *If Assumptions 3.1 and 3.4 are satisfied, then $\mathbb{P}_N^{u,v_T}$ converges weakly to $\mathbb{P}^{u,v_T}$ as $N \to \infty$.*

Proof In the following, we lighten the notation by omitting the initial point u from the notation, which will be assumed fixed to $u = x_0$. We wish to show that $\mathbb{P}_N^{v_T}$ converges weakly to $\mathbb{P}^{v_T}$ as $N \to \infty$. This is equivalent to showing that $\int f \mathrm{d}\mathbb{P}_N^{v_T} \to \int f \mathrm{d}\mathbb{P}^{v_T}$ for all bounded and continuous functions f. Write $c_\infty^{v_T} = p(0, x_0, T, v_T)/q(0, x_0, T, v_T)$. By equation (3) and (9),

$$\mathbb{E}_{\mathbb{Q}^{v_T}} Z(X) = \mathbb{E}_{\mathbb{Q}^{v_T}} \frac{d\mathbb{P}^{x_0}}{d\mathbb{Q}^{x_0}} = c_\infty^{v_T} \mathbb{E}_{\mathbb{Q}^{v_T}} \left[\frac{d\mathbb{P}^{v_T}}{d\mathbb{Q}^{v_T}}\right] = c_\infty^{v_T}$$

and we have that

$$\begin{aligned}
&\left|\int f \mathrm{d}\mathbb{P}_N^{v_T} - \int f \mathrm{d}\mathbb{P}^{v_T}\right| \\
&= \left|\int f \left(\frac{Z^N}{c_N^{v_T}} - \frac{Z}{c_\infty^{v_T}}\right) \mathrm{d}\mathbb{Q}^{v_T}\right| \\
&\le \|f\|_\infty \int \left|\frac{Z^N(X)}{c_N^{v_T}} - \frac{Z(X)}{c_\infty^{v_T}}\right| \mathrm{d}\mathbb{Q}^{v_T}(X) \\
&\le \|f\|_\infty \left(\frac{1}{c_N^{v_T}} \int \left|Z^N(X) - Z(X)\right| \mathrm{d}\mathbb{Q}^{v_T}(X)\right. \\
&\quad \left. + \int Z(X) \left|\frac{1}{c_N^{v_T}} - \frac{1}{c_\infty^{v_T}}\right| \mathrm{d}\mathbb{Q}^{v_T}(X)\right) \\
&\le \|f\|_\infty \left(\frac{1}{c_N^{v_T}} \int \left|Z^N(X) - Z(X)\right| \mathrm{d}\mathbb{Q}^{v_T}(X) + \left|\frac{c_\infty^{v_T}}{c_N^{v_T}} - 1\right|\right)
\end{aligned} \tag{13}$$

where we used Assumption 3.1 for applying the change of measure between the conditional measures. Notice that $Z^N(X) = Z(X^N)$. The mapping $X \mapsto Z(X)$, as a function acting on $C(0, T)$ with uniform norm, is continuous, since B, b and b' are continuous. Therefore, it follows from the Lévy–Ciesielski construction of Brownian motion (see Sect. 1.1.1) and the continuous mapping theorem that

$$Z^N(X) \to Z(X) \qquad \mathbb{Q}^{v_T} - a.s.$$

Now, notice that, under conditional measures $\mathbb{Q}^{v_T}$ and $\mathbb{P}^{v_T}$, the term $B(X_T) - B(X_0)$ is fixed. By the assumptions on b and b', Z is a bounded function and by dominated convergence, we get that

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{Q}}^{v_T} |Z^N(X) - Z(X)| = 0$$

giving convergence to zero of the first term in (13). This implies that also the constant $c_N := \mathbb{E}_{\mathbb{Q}}^{v_T} |Z^N(X)|$ converges to $\mathbb{E}_{\mathbb{Q}}^{v_T} |Z(X)| = c_\infty^{v_T}$ so that all the terms in (13) converge to 0. □

We now list some technical conditions for the process to satisfy Assumptions 3.1 and 3.4.

Remark 3.6 If $|b(x)| \le c(1+|x|)$, for some positive constant c, then Assumption 3.1 is satisfied.

Proof See Liptser et al. (2013), Sect. 6, Example 3 (b). □

Remark 3.7 If b is globally Lipschitz and continuously differentiable, then Assumptions 3.1 and 3.4 are satisfied.

Proof Assumption 3.4 is trivially satisfied. By Remark 3.6, also Assumption 3.1 is satisfied. □

In Sect. 5.3, we will present an example where the drift b is not globally Lipschitz, yet Assumption 3.4 is satisfied.

Assumption 3.8 There exists a non-decreasing function $h : [0, \infty) \to [0, \infty)$ such that $B(x) \le h(|x|)$ and

$$\int_0^\infty \exp(h(x) - x^2/(2T))\, dx < \infty.$$

The above integrability condition is, for example, satisfied if $h(|x|) = c(1 + |x|)$ for some $c > 0$.

Remark 3.9 If Assumptions 3.4 and 3.8 hold, then Assumption 3.1 is satisfied.

Proof By Sect. 3.5 in Karatzas and Shreve (1991), (Z_t) is a local martingale. Say $b'(x) + b^2(x) \ge -2C$, where $C \ge 0$. Using the assumptions, we have

$$\begin{aligned}
Z_t &= \exp\left(B(X_t) - B(X_0) - \tfrac{1}{2}\int_0^t \{b'(X_s) + b^2(X_s)\}\, ds\right) \\
&\le A \exp(Ct) \exp(h(|X_t|)),
\end{aligned}$$

with constant $A = \exp(-B(X_0))$. Then,

$$\sup_{t\in[0,T]} Z_t \le A \sup_{t\in[0,T]} \exp(Ct) \exp(h(|X_t|)) \le A \exp(CT) \exp\left(h\left(\max_{t\in[0,T]} |X_t|\right)\right).$$

By Lemma 3.10,

$$\mathbb{E} \sup_{t\in[0,T]} Z_t \le A \exp(CT)\, \mathbb{E} \exp(h(\max_{t\in[0,T]} |X_t|)) < \infty.$$

Then, for a sequence of stopping times (τ_k) diverging to infinity such that $(Z_t^{\tau_k})_{0\le t\le T}$ is a martingale for all k, we have

$$\mathbb{E}Z_0 = \mathbb{E}Z_0^{\tau_k} = \mathbb{E}Z_t^{\tau_k} \to \mathbb{E}Z_t$$

as $k \to \infty$ by dominated convergence. □

Lemma 3.10 *Suppose* $h\colon [0,\infty) \to [0,\infty)$ *is non-decreasing. Let* $N_T = \max_{0\le t\le T} |X_t|$ *where* (X_t) *is a Brownian motion. Then,*

$$\mathbb{E}\exp h(N_T) \le 4\int_0^\infty \frac{1}{\sqrt{2\pi T}} \exp(h(x) - x^2/(2T))\,dx.$$

Proof The maximum $M_T = \max_{0\le t\le T} X_t$ of a Brownian motion is distributed as the absolute value of a Brownian motion and thus has density function $\frac{2}{\sqrt{2\pi T}}\exp(-x^2/(2T))$, see Karatzas and Shreve (1991), Sect. 2.8. We have $\mathbb{P}(N_T \ge y) \le 2\mathbb{P}(M_T \ge y)$ from which the result follows. □

Finally, we mention that Theorem 3.5 can be generalised in the following way to diffusions without a fixed end point.

Proposition 3.11 *If Assumption 3.4 is satisfied and B is bounded, then* $\mathbb{P}_N$ *converges weakly to* $\mathbb{P}$.

The proof follows the same steps of the one of Theorem 3.5. In this case, we need to pay attention on B, as for unconditioned process, the final point is not fixed. If B is bounded, then Assumption 3.8 is satisfied. By Remark 3.9, also Assumption 3.1 is satisfied so that we can apply Theorem 3.2 for the change of measure. Finally, by the assumptions on b and B, the function Z is bounded and by dominated convergence, we get that

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{Q}}|Z^N(X) - Z(X)| = 0.$$

4 A local Zig-Zag algorithm with subsampling for high-dimensional structured target densities

In Sect. 4.4, we will show that the task of sampling diffusion bridges boils down to the task of sampling a high-dimensional vector $\xi^N \in \mathbb{R}^M$ under the measure $\mathbb{P}_N^{\mu,v_T}$. Define by P_{ξ^N} the distribution of the vector ξ^N. Under the target measure,

$$P_{\xi^N}(\mathrm{d}\xi^N) = \pi(\xi^N)\mathrm{d}\xi^N.$$

We take the density π to be the M-dimensional invariant density (target density) for the Zig-Zag sampler. An efficient implementation of piecewise deterministic Monte Carlo methods including the local and fully local Zig-Zag sampler can be found in Roberts and Stramer (2001).

4.1 Subsampling technique

In our setting, the integral appearing in the Girsanov formula (10) poses difficulties when finding the root of equation (7) and would require numerical evaluation of the integral, hence also introducing a bias. By adapting the subsampling technique presented in Bierkens et al. (2019) (Sect. 4), we avoid this problem altogether (see Sect. 4.4). In general, this technique requires

(a) unbiased estimators for $\partial_i\psi$, i.e. random functions $\partial_i\tilde{\psi}_i(\xi, U_i)$ such that

$$E_{U_i}[\partial_i\tilde{\psi}_i(\xi, U_i)] = \partial_i\psi(\xi),$$

for all i and ξ. These random functions create new (random) Poisson rates given by

$$\tilde{\lambda}_i(t;\xi,\theta;U_i) = (\theta_i\partial_i\tilde{\psi}(\xi(t), U_i))^+, \qquad i = 1,2,\ldots,d, \tag{14}$$

whose evaluation becomes feasible and computationally more efficient compared to the original Poisson rates given by Eq. (6).

(b) upper bounds $\bar{\lambda}_i : (\mathbb{R}^+ \times \mathbb{R}^d \times \{-1,+1\}^d) \to \mathbb{R}^+$ for all $i = 1,\ldots,d$ such that for any point (ξ,θ) and $t \ge 0$, we have

$$P\left(\tilde{\lambda}_i(t;\xi,\theta;U_i) \le \bar{\lambda}_i(t;\xi,\theta)\right) = 1. \tag{15}$$

As we show in Algorithm 2 and in Sect. 5, these upper bounds are used for finding the roots of Eq. (7).

Algorithm 2 gives the algorithm for the Zig-Zag sampler with subsampling. It can be proved (see Bierkens et al. 2019) that the Zig-Zag sampler with subsampling has the same invariant distribution as its original and therefore does not introduce any bias. Note that we slightly modified the algorithm from Bierkens et al. (2019) in order to reduce its complexity. In particular, it is sufficient to draw new waiting times and to save the coordinates only when the *if* condition at the *subsampling step* of Algorithm 2 is true.

4.2 Local Zig-Zag sampler

Section 3.1 of Bouchard-Côté (2015) proposes a local algorithm for the *Bouncy Particle Sampler* which is a process belonging to the class of piecewise-deterministic Markov processes. Similar ideas apply to our setting.

Assumption 4.1 The Poisson rate λ_i for a d-dimensional target distribution is a function of the coordinates $N_i \subset \{1,\ldots,d\}$,

Algorithm 2 d-dimensional Zig-Zag sampler with subsampling

procedure ZIGZAG_WS($\tau_{\text{final}}, \xi, \theta$)
  Initialise $k = 1$, $t = 0$
  $\tau_j \sim \text{IPP}(\bar{\lambda}_j(\cdot; \xi, \theta))$, $j = 1, \dots, d$
  while $t \leq \tau_{\text{final}}$ **do**
    $\tau^*, i^* \leftarrow \text{findmin}(\tau_1, \dots, \tau_d)$
    $\xi^{old} \leftarrow \xi$
    Update: $\xi \leftarrow \xi + \theta(\tau^* - t)$
    Update: $\Delta t \leftarrow \tau^* - t$; $t \leftarrow \tau^*$
    $U_{i^*} \sim \text{Law}(U_{i^*})$, $V \sim \text{Unif}(0, 1)$
    if $V \leq \tilde{\lambda}_{i^*}(0, \xi, \theta, U_{i^*})/\bar{\lambda}_{i^*}(\Delta t; \xi^{old}, \theta)$ **then** ▷ Subsampling step
      Save $\xi^{(k)} \leftarrow \xi$, $t^{(k)} \leftarrow t$
      $k \leftarrow k + 1$
      $\theta_{i^*} \leftarrow -\theta_{i^*}$
      for $j \in \{1, \dots, d\} \setminus \{i^*\}$ **do**
        $\tau_j \sim t + \text{IPP}(\bar{\lambda}_j(\cdot; \xi, \theta))$
      end for
    else
      $\tau_{i^*} \sim t + \text{IPP}(\bar{\lambda}_{i^*}(\cdot; \xi, \theta))$
    end if
  end while
  return Skeletons $(\xi^{(l)}, t^{(l)})_{l=1,2,\dots,k-1}$
end procedure

$$\lambda_i(s; \xi, \theta) = \lambda_i(s; \xi_k, \theta_k : k \in N_i).$$

Recall that by the definition of λ_i (see equation (6)), the ith partial derivative of the negative log-likelihood determines the sets N_i. Now, let us suppose that the first event time τ is triggered by the coordinate i so that at event time, the velocity θ_i is flipped. For all λ_k which are not function of this coordinate ($k \notin N_i$), we have

$$\lambda_k^{old}(\tau + s) = \lambda_k^{new}(s),$$

which implies that the waiting times drawn before τ are still valid after switching the velocity i. This allows us to rescale the previous waiting time and reduce the number of computations at each step. The sets $N_1, \dots, N_d$ are connected to the factorisation of the target distribution and define its conditional dependence structure. Indeed, take a d-dimensional target distribution with the following decomposition

$$\pi(\xi) = \prod_{i=1}^{N} \pi_i(\xi^{(i)})$$

where $\xi^{(i)} := \{\xi_j : j \in \Gamma_i\}$ and $\Gamma_i \subset \{1, 2, \dots, N\}$ defines a subset of indices. We have that

$$-\partial_k \log(\pi(\xi)) = -\sum_{i=1}^{N} \partial_k \log \pi_i(\xi^{(i)}), \quad k = 1, \dots, d$$

where the ith term in the sum is equal to 0 if $k \notin \Gamma_i$. Since the Poisson rates (6) are defined through the partial derivatives, the factorisation defines the sets $N_1, \dots, N_d$ of Assumption 4.1.

Algorithm 3 shows the implementation of the local sampler which exploits any conditional independence structure so that the complexity of the algorithm scales well with the number of dimensions.

The local Zig-Zag sampler simplifies to independent one-dimensional Zig-Zag processes if the coefficients are pairwise-independent coefficients, as it was the case in the example of sampling a Brownian motion or Brownian bridge (see Sect. 2.3). On the other hand, it defaults to Algorithm 1 when the dependency graph is fully connected, that is if $N_i = \{1, \dots, d\}, \forall i$.

Algorithm 3 d-dimensional local Zig-Zag sampler

Input: The bounds $\bar{\lambda}_i$ depend only on ξ_k, θ_k, for $k \in N_i$
procedure ZIGZAG_LOCAL($\tau_{\text{final}}, \xi, \theta$)
  Initialise $k = 1$, $t = 0$
  $\tau_j \sim \text{IPP}(\lambda_j(\cdot; \xi, \theta))$, $j = 1, \dots, d$
  while $t \leq \tau_{\text{final}}$ **do**
    $\tau^*, i^* \leftarrow \text{findmin}(\tau_1, \dots, \tau_d)$
    Update: $\xi \leftarrow \xi + \theta(\tau^* - t)$
    Update: $\theta_{i^*} \leftarrow -\theta_{i^*}$; $t \leftarrow \tau^*$
    Save $\xi^{(k)} \leftarrow \xi$; $t^{(k)} \leftarrow t$
    $k \leftarrow k + 1$
    for j in N_{i^*} **do** ▷ Local step
      $\tau_j \sim t + \text{IPP}(\lambda_j(\cdot; \xi, \theta))$
    end for
  end while
  return Skeletons $(\xi^{(l)}, t^{(l)})_{l=1,\dots,k-1}$
end procedure

4.3 Fully local Zig-Zag sampler

Combining the subsampling technique and the local ZZ can lead to a further reduction of the complexity of the algorithm. Indeed, the bounds for the Poisson rates might induce sparsity as $\bar{\lambda}_i$ can be function of few coordinates (see, for example, Sect. 5.2). This means that, after flipping θ_i, $\bar{\lambda}_j^{old}(\tau + t) = \bar{\lambda}_j^{new}(t)$ for almost all $j \neq i$ making the if statement in the *local step* of Algorithm 3 almost always satisfied and improving the efficiency of the algorithm. This means that, after flipping θ_i, we have that $\bar{\lambda}_j^{old}(\tau + t) = \bar{\lambda}_j^{new}(t)$ for almost all $j \neq i$ or, in other words, the cardinality of the set N_i in the *local step* of Algorithm 3 is small. Furthermore, the evaluation of $\tilde{\lambda}_i(t, \xi, \theta)$ and $\bar{\lambda}_i(t, \xi, \theta)$ for $i = 1, 2, \dots, d$ does not necessarily require to access the location of all the coordinates ξ_j so that, by assigning an independent time for each coordinate and updating only the coordinates needed for the evaluation of $\tilde{\lambda}_i$ and $\bar{\lambda}_i$, the algorithm can be made more efficient. This is shown in the fully local ZZ sampler (Algorithm 4) where $\bar{N}_i$, $\tilde{N}_i(U_i)$ define, respectively, the sub-

Algorithm 4 Implementation of the d-dimensional fully local Zig-Zag sampler

Input: The bounds $\bar{\lambda}_i$ depend only on ξ_k, θ_k, for $k \in \bar{N}_i$ and the random Poisson rates $\tilde{\lambda}_i$ (eq. (14)) depends only on U_i (the randomizing argument of $\tilde{\partial}_i \psi$) and ξ_k, θ_k for $k \in \tilde{N}_i(U_i)$
procedure ZIGZAG_FULLY_LOCAL($\tau_{\text{final}}, \xi, \theta$)
  Initialise: $k = 1, t = \mathbf{0} \in \mathbb{R}^d, \tau^* = 0$
  $\tau_j \sim \text{IPP}(\bar{\lambda}_j(\cdot; \xi, \theta))$, $j = 1, \ldots, d$
  while $\max(t) \leq \tau_{\text{final}}$ **do**
    $\tau_{i^*}^{old} \leftarrow \tau^*, \quad \xi_{i^*}^{old} \leftarrow \xi_{i^*}$
    $\tau^*, i^* \leftarrow \text{findmin}(\tau_1, \ldots, \tau_d)$
    $U_{i^*} \sim \text{Law}(U_{i^*})$
    for j in $\bar{N}_{i^*} \cup \tilde{N}_{i^*}(U_{i^*})$ **do**
      Update: $\xi_j \leftarrow \xi_j + \theta_j(\tau^* - t_j)$
      Update: $t_j \leftarrow \tau^*$
    end for
    $V \sim \text{Unif}(0, 1)$
    if $V \leq \tilde{\lambda}_{i^*}(0; \xi, \theta; U_{i^*})/\bar{\lambda}_{i^*}(\tau^* - \tau_{i^*}^{old}; \xi^{old}, \theta)$ **then**
      Update: $\theta_{i^*} \leftarrow -\theta_{i^*}$
      Update: $k \leftarrow k + 1$
      Save: $i^{(k)} \leftarrow i^*, s^{(k)} \leftarrow \tau^*, \xi^{(k)} \leftarrow \xi_{i^*}$
      for n in $\left(\bigcup_{j \in \bar{N}_{i^*}} \bar{N}_j\right) \setminus \left(\bar{N}_{i^*} \cup \tilde{N}_{i^*}(U_{i^*})\right)$ **do**
        Update: $\xi_n \leftarrow \xi_n + \theta_n(\tau^* - t_n)$
        Update: $t_n \leftarrow \tau^*$
      end for
      for j in $\bar{N}_{i^*} \setminus \{i^*\}$ **do**
        $\tau_j \sim \tau^* + \text{IPP}(\bar{\lambda}_j(\cdot; \xi, \theta))$
        $\tau_j^{old} \leftarrow \tau^*, \quad \xi_j^{old} \leftarrow \xi_j$
      end for
    end if
    $\tau_{i^*} \sim \tau^* + \text{IPP}(\bar{\lambda}_{i^*}(\cdot; \xi, \theta))$
  end while
  return reflection tuples $((i^{(l)}, s^{(l)}, \xi^{(l)}))_{l=1,\ldots,k}$
end procedure

set and the random subset of the coordinates required for the evaluation of $\bar{\lambda}_i(\cdot; \xi, \theta)$ and $\tilde{\lambda}_i(\cdot; \xi, \theta; U_i)$.

4.4 Sampling diffusion bridges

In order to employ the Zig-Zag sampler to simulate from the bridge measure, we choose the truncation level N in Eq. (2). Then, under $\mathbb{P}_N^{u,v_T}$

$$\pi(\mathrm{d}\xi^N) \propto Z^N(X) \exp\left(\frac{-\|\xi^N\|^2}{2}\right) \mathrm{d}\xi^N.$$

This is a straightforward consequence of the change of measure in (12) and the Lévy–Ciesielski construction.

We need to make one further assumption:

Assumption 4.2 The drift b of the diffusion process is twice differentiable.

Assumption 4.2 is necessary in order to compute the ξ_k-partial derivative of the energy function, which becomes

$$\partial_k \psi(\xi^N) = \frac{1}{2} \int_{S_k} h_k(s; \xi^N) \mathrm{d}s + \xi_k, \tag{16}$$

where

$$h_k(s; \xi^N) = \phi_k(s) \left(2b(X_s^N) b'(X_s^N) + b''(X_s^N)\right).$$

As the index k in the Faber–Schauder basis function gets larger, both the magnitude of ϕ_k and the size of its support decrease so that typically $\int h_k(s; \xi^N) \mathrm{d}s$ gets smaller and $\partial_k \psi(\xi) \approx \xi_k$ which corresponds to the partial derivative of the energy function of a standardised Gaussian random variable with independent components. This justifies one more time the intuition that for high levels i, the random variables $\xi_{ij}, j = 1, \ldots, 2^i - 1$ are approximately normally distributed and almost independent from the other random coefficients.

In order to avoid the evaluation of the integral appearing in (16) and the difficulty of drawing a Poisson time from its corresponding rate (6), we employ the subsampling technique. Considering ξ^N non-random, we take as an unbiased estimator for $\partial_k \psi(\xi_N)$ the (random) function

$$\frac{1}{2} |S_k| h_k(U_k; \xi^N) + \xi_k, \tag{17}$$

where $U_k \sim \text{Unif}(S_k)$ and as the bounding intensity rate

$$\bar{\lambda}_k(t, \xi^N, \theta^N) = \frac{1}{2} |S_k| |\theta_k| \bar{\Phi}_k f(\xi^N(t)) + (\theta_k \xi_k(t))^+, \quad \xi^N \in \mathbb{R}^M, \tag{18}$$

where $\bar{\Phi}_k = \max_s(\phi_k(s))$ and $f(\xi^N) \geq \left|2b(X_s^{\xi^N}) b'(X_s^{\xi^N}) + b''(X_s^{\xi^N})\right|$, $\forall s \in [0, T]$, $\xi^N \in \mathbb{R}^M$. The subsampling technique avoids the numerical computation of the time integral (16), thus avoiding a numerical bias and reducing the computational effort from $\mathcal{O}(T)$ (for fixed discretisation size) to $\mathcal{O}(1)$. The variance of this unbiased estimator can be reduced by averaging over multiple independent uniform draws or similar strategies (see, for example, Sect. 5.4), albeit at the cost of additional computations. In Sect. 5, we show specifically for each numerical experiment how we derived the Poisson upper bounds $\bar{\lambda}_i$.

The compact support of the Faber–Schauder functions induces a sparse dependency structure on the target measure π. Indeed, X_t only depends on those values of $\xi_{l,k}$ for which $t \in S_{l,k}$. See Fig. 4 for an illustration. It is easy to see that this implies that $\frac{\partial \psi(\xi^N)}{\partial \xi_{(i,j)}}$ depends only on those $\xi_{(k,l)}$ for which the interior of $S_{i,j} \cap S_{k,l}$ is non-empty. In particular, define the relation $\xi_{i,j} \ll \xi_{k,l}$ to hold if $S_{k,l} \subset S_{i,j}$. If this happens, then we refer to $\xi_{i,j}$ as the *ancestor* of $\xi_{k,l}$ (and conversely $\xi_{k,l}$ as the *descendant*). Then, the sets in Assumption 4.1 (using double indexing) can be chosen as $N_{i,j} = \{\xi_{h,d} : \xi_{h,d} \ll \xi_{i,j} \vee \xi_{h,d} \gg \xi_{i,j}\}$ with cardinality $|N_{i,j}| = 2^{N-i+1} + i - 1$, where N is the truncation level. Formally, $N_{i,j}$ are the neighbourhoods of the interval graph induced

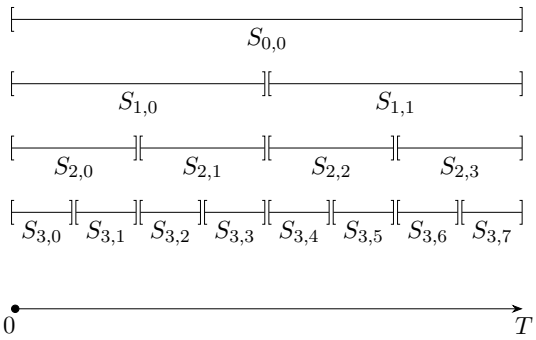

Fig. 4 Support of the Faber–Schauder functions $(\phi_{i,j} : i \in \{0, 1, \ldots, N\},\ j = \{0, 1, \ldots, 2^i - 1\}$ with $N = 3$. The coefficient $\xi_{i,j}$ is independent of the coefficient $\xi_{k,l}$ conditionally on the set of common ancestors $(\xi_{m,n} : S_{m,n} \cap S_{i,j} \neq \emptyset \wedge S_{m,n} \cap S_{k,l} \neq \emptyset)$ if $S_{i,j} \cap S_{k,l} = \emptyset$

by $((S_{i,j} : i \in \{1, 2, \ldots, N\},\ j \in \{0, 1, \ldots, 2^i - 1\}))$ with vertices $\{(i, j) : i \in \{1, 2, \ldots, N\},\ j \in \{0, 1, \ldots, 2^i - 1\}\}$, where there is an edge between (i, j) and (l, k) if the interior of $S_{i,j} \cap S_{k,l}$ is non-empty (see Fig. 11). The factorisation of the partial derivatives leads to a specific dependency structure of the coefficients under the target diffusion bridge measure: the coefficient $\xi_{i,j}$ is conditionally independent of the coefficient $\xi_{k,l}$ if $S_{i,j} \cap S_{k,l} = \emptyset$ conditionally on the set of common ancestors $(\xi_{m,n} : \xi_{m,n} \ll \xi_{i,j} \wedge \xi_{m,n} \ll \xi_{k,l})$. This argument is made more formal by decomposing the likelihood function in Appendix A.

5 Numerical results

We show numerical results for three representative examples. In general, when applying our method, we start from a model (1) and devise a representation of the approximate diffusion bridge (12) that we sample using generic implementations of algorithms 1-4 from our package, which are easily adapted to the task of sampling the coefficients of the Faber–Schauder expansion. To this end, we provide the k-th partial derivative of the energy function (16) or an upper bound to the Poisson rate (18) as argument for the sampler, as well as the sets $N_{i,j}$ as given in Sect. 4.4. The reader is referred to the file `faberschauder.jl` in the public repository https://github.com/SebaGraz/ZZDiffusionBridge/src for the implementation of the expansion and for the generic implementation of the different variants of the Zig-Zag sampler to our package (Roberts and Stramer 2001).

The first class of diffusion processes considered are diffusions with linear drift function (Sect. 5.1). This is a special case, where our method does not require the subsampling technique described in Sect. 4.1 and only Algorithm 3 has been employed. Notice that for this class, the transition kernel of the conditioned process is known. In Sect. 5.2, we apply our method for diffusions which substantially differ from Brownian motions, being highly nonlinear and multimodal and therefore creating challenging bridge distributions for standard MCMC. Here, we use the fully local algorithm (Algorithm 4). In the specific example considered, the implementation of the Zig-Zag sampler is facilitated by the drift function and its derivatives being bounded, and therefore, a bounded Poisson rate for the subsampling technique is available. In view of this, we choose for the third numerical experiment a diffusion with unbounded drift (Sect. 5.3). For all the models, Assumptions 3.1, 3.4 and 4.2 are immediate to verify and Assumption 4.1 is satisfied. For each experiment, the burn-in $\tau_{\text{burn-in}}$ and final clock τ_{final} are manually tuned by inspecting the trace of ξ^N and ensuring that the process reached stationarity before $\tau_{\text{burn-in}}$ and fully explore the state space before the final clock τ_{final}. The computations are performed with a conventional laptop with a 1.8GHz intel core i7-8550U processor and 8GB DDR4 RAM. We wrote the program in Julia 1.4.2 which allows profiling and optimizing the code for high performance. The program is publicly available on GitHub at https://github.com/SebaGraz/ZZDiffusionBridge where the reader can follow the documentation to reproduce the results.

5.1 Linear diffusions

A linear stochastic differential equation conditioned to hit a final point v_T has the form

$$\mathrm{d}X_t = (\alpha + \beta X_t)\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 = u, X_T = v_T \tag{19}$$

for some $(\alpha, \beta) \in \mathbb{R}^2$. Assumptions 3.1, 3.4 and 4.2 can be easily verified. In this case, the energy function of the target distribution is

$$\psi(\xi^N) = C_1 - \ln(Z^N(X)) + \frac{\|\xi^N\|^2}{2} = C_2 + \frac{1}{2}\int_0^T \left(\beta^2 \left(X_t^{\xi^N}\right)^2 + 2\alpha\beta X_t^{\xi^N}\right)\mathrm{d}t + \frac{\|\xi^N\|^2}{2},$$

for some constant C_1, C_2. Note that ψ is a quadratic function of ξ, which means that the target density is still Gaussian under $\mathbb{P}_N^{u,v_T}$. It follows that

$$\partial_{\xi_k}\psi(\xi^N) = \int_{t\in S_k} \phi_k(t)\left(\beta^2\left(\bar{\bar{\phi}}(t)u + \bar{\phi}(t)v_T/\sqrt{T} + \sum_{j\in N_k}\phi_j\xi_j\right) + \alpha\beta\right)\mathrm{d}t + \xi_k.$$

Interchanging the integral and the sum, this becomes

$$\partial_{\xi_k}\psi(\xi^N) = \beta^2\left(\bar{\bar{\Phi}}_k u + \bar{\Phi}_k v_T/\sqrt{T} + \sum_{j\in N_k}\Phi_{jk}\xi_j\right) + \alpha\beta\Phi_k + \xi_k,$$

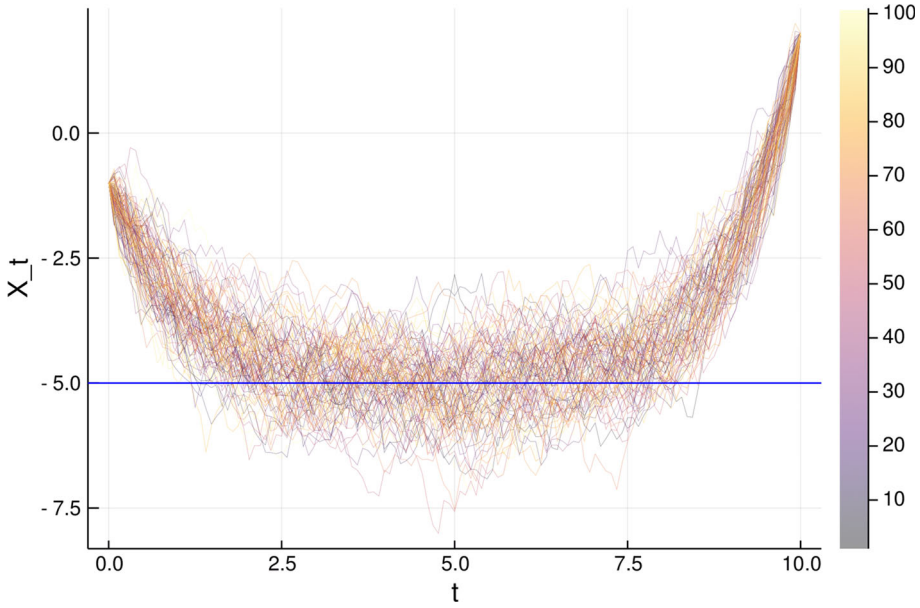

Fig. 5 Simulation of the diffusion bridge measure (100 samples) given by Eq. (19) starting at -1.0 and conditioned to hit 2.0 at $T = 10$. $\alpha = -5.0$, $\beta = -1.0$ which is equivalent to a mean reverting process with mean reversion at $x = -5$ (straight line). The truncation level is $N = 6$, final clock $\tau_{\text{final}} = 1000$ and burn-in $\tau_{\text{burn-in}} = 10$

where $\Phi_k = \int \phi_k \mathrm{d}t$, $\Phi_{jk} = \int \phi_k \phi_j \mathrm{d}t$, $\bar{\Phi}_k = \int \bar{\phi} \phi_k \mathrm{d}t$ and $\bar{\bar{\Phi}} = \int \bar{\bar{\phi}} \phi_k \mathrm{d}t$. This is a linear function of ξ^N, and for each i, the event times with rates λ_i, see (6), can be directly simulated without upper bounds. Figure 5 shows samples from the resulting diffusion bridge measure with $\alpha = -5$, $\beta = -1$ obtained with this method running the Zig-Zag sampler for $\tau_{\text{final}} = 1000$, with a burn-in time of $\tau_{\text{burn-in}} = 10$. The closed form of the expansion of linear processes, or more generally, reciprocal linear processes, with the Faber–Schauder basis was also found and used in Schauer and Grazzi (2020) for the problem of nonparametric drift estimation of diffusion processes. The results are validated by computing analytically the density of the random variable $X_{T/2}$ (which, for the linear case, is known in close form) and comparing this with its empirical density obtained from one sample of the Zig-Zag process (see Fig. 7, left panel).

5.2 Nonlinear multimodal diffusions

The stochastic differential equation considered here has the form

$$\mathrm{d}X_t = \alpha \sin(X_t)\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 = u, X_T = v_T \tag{20}$$

for some $\alpha \geq 0$. When $\alpha = 0$, the process is a standard Brownian motion, while for positive α, the process is attracted to its stable points $(2k-1)\pi$, $k \in \mathbb{N}$. Assumptions 3.1, 3.4, 4.2 follow from drift, its primitive and its derivative being globally bounded. Fixing N, the energy function is given by

$$\psi(\xi^N) = \frac{\alpha}{2}\int_0^T \left(\alpha \sin^2(X_t^{\xi^N}) + \cos(X_t^{\xi^N})\right)\mathrm{d}t + \frac{\|\xi^N\|^2}{2}.$$

Using trigonometric identities, we obtain that

$$\partial_{\xi_k}\psi(\xi^N) = \frac{1}{2}\int_{S_k} \phi_k(t)\left(\alpha^2 \sin\left(2X_t^{\xi^N,k}\right) - \alpha \sin\left(X_t^{\xi^N,k}\right)\right)\mathrm{d}t + \xi_k$$

where $X_t^{\xi^N,k} := \bar{\bar{\phi}}(t)u + \bar{\phi}(t)v_T/\sqrt{T} + \sum_{j\in N_k}\phi_j(t)\xi_j$. To avoid the need to find the roots of Eq. (7), we apply the subsampling technique described in Sect. 4.1. Since the drift and its derivatives are bounded, we can easily find the following upper bound for (14):

$$\bar{\lambda}_k(t) = |\theta_k|a_1 + (\theta_k \xi_k(t))^+, \tag{21}$$

with $a_1 = \bar{\Phi}_k S_k(\alpha^2+\alpha)/2$, $\bar{\Phi}_k = \max(\phi_k)$ and $\xi_k(t) = \xi_k + \theta_k t$. In this case, the upper bound $\bar{\lambda}_i$ is a function only of the coefficient ξ_i. Figure 6 shows the results obtained with this method setting $\alpha = 0.7$. For this diffusion, the nonlinearity and multiple modes make the mixing of the Zig-Zag sampler slower, so we set $\tau_{\text{final}} = 10{,}000$ and burn-in $\tau_{\text{burn-in}} = 10$.

Analysing the goodness of the empirical diffusion bridge distribution obtained is a difficult task since the true conditional distribution is not known in a tractable form. We start by checking if some geometrical properties of the diffusion bridge distributions are preserved in the simulations. For example, in Fig. 6, it can be noticed that the diffusion is attracted to the stable points $\pm\pi$, $\pm 3\pi$, ..., and symmetric (geometrically speaking, after rotation) around the vertical axes $t = T/2$. We furthermore validate our method by simulating forward diffusion processes, using Euler discretisation in a fine grid and retaining only the paths which end in a ϵ-ball of a certain point at time T (ϵ-ball forward simulation). If the final point is such that the probability of ending in this ϵ-ball is high enough, we can create in this way a sample

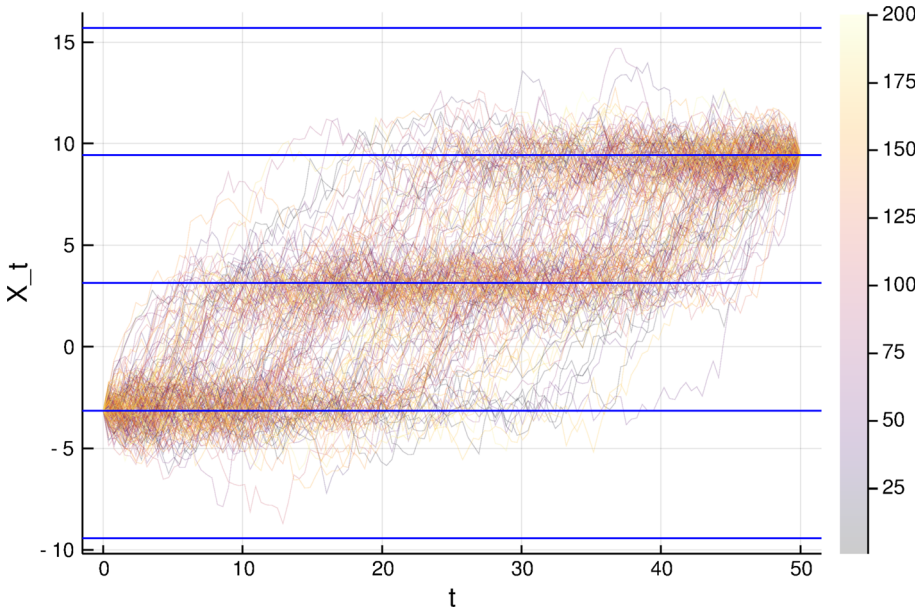

Fig. 6 Simulation of the diffusion bridge measure (200 samples) given by Eq. (20) with $\alpha = 0.7$ starting at $-\pi$ at time 0 and hitting 3π at $T = 50$. Truncation level $N = 6$, final clock $\tau_{\text{final}} = 10{,}000$ and burn-in 10. The straight horizontal lines are the attraction points of the process

from the approximated bridge and compare it to the samples obtained from the Zig-Zag. The right panel of Fig. 7 shows the joint empirical distribution with the two methods of the first quarter and third quarter random variables. Finally, Fig. 8 illustrates that the marginal distribution of the coefficients in higher levels is approximately Gaussian and the nonlinearity of the process is absorbed by the coefficients in low levels.

5.3 Diffusions with unbounded drift

Here, we consider *stochastic exponential logistic models*. For this class, the process grows exponentially with rate r until it reaches its saturation point K. Its dynamics are perturbed by noise which grows as the population grows. The resulting stochastic differential equation takes the form

$$\begin{aligned} &\mathrm{d}Y_t = rY_t(1 - Y_t/K)\mathrm{d}t + \beta Y_t \mathrm{d}W_t, \\ &X_0 = u > 0, \quad X_T = v_T > 0. \end{aligned} \tag{22}$$

We can transform the process in order to get a new process with unitary diffusivity $\sigma = 1$ (Lamperti transform with $X_t = -\log(Y_t)/\beta$). The transformed differential equation becomes

$$\begin{aligned} &\mathrm{d}X_t = (c_1 + c_2 e^{-\beta X_t})\mathrm{d}t + \mathrm{d}W_t, \\ &X_0 = -\log(u)/\beta, \quad X_T = -\log(v)/\beta. \end{aligned}$$

with $c_1 = \beta/2 - r/\beta$ and $c_2 = r/(\beta K)$. Note that the drift function b of the transformed process is not global Lipschitz continuous. Nevertheless, Assumptions 3.4 and 4.2 are satisfied and by Remark 3.9, also Assumption 3.1 is verified. In this case, the partial derivative of the energy function is given by

$$\partial_k \psi(\xi^N) = \frac{1}{2}\int_{S_k} \phi_k(s)\left(a_1 e^{-\beta X_s^{\xi^N}} - a_2 e^{-2\beta X_s^{\xi^N}}\right)\mathrm{d}s + \xi_k,$$

where $a_1 = 2r^2/(\beta K)$, $a_2 = a_1/K$. As before, it is not possible to simulate directly the first event time using the Poisson rates given by Eq. (6). The subsampling technique requires an upper bound for the unbiased estimator (14). Define the following quantities

$$b_k^{(1)} := \inf_{s\in S_k}\left\{\bar{\bar{\phi}}(s)u_0 + \bar{\phi}(s)v_T/\sqrt{T} + \sum_{i\in N_k}\phi_i(s)\xi_i\right\},$$

$$b_k^{(2)} := \inf_{s\in S_k}\left\{\sum_{i\in N_k}\phi_i(s)\theta_i\right\}.$$

For any $a, b, c \in \mathbb{R}$, $(a+b+c)^+ \le (a)^+ + (b)^+ + (c)^+$, and hence, a valid upper bound for the Poisson rate (14) is given by

$$\bar{\lambda}_k(t) = \lambda_k^{(1)}(t) + \lambda_k^{(2)}(t) + \lambda_k^{(3)}(t) \tag{23}$$

with

$$\begin{aligned} \lambda_k^{(1)}(t) &= \max\left(0, \theta_k \xi_k(t)\right), \\ \lambda_k^{(2)}(t) &= \max\left(0, \frac{1}{2}\theta_k \bar{\phi}_k S_k z_k^{(1)} e^{-\beta_k^\star t}\right), \\ \lambda_k^{(3)}(t) &= \max\left(0, -\frac{1}{2}\theta_k \bar{\phi}_k S_k z_k^{(2)} e^{2\beta_k^\star t}\right) \end{aligned}$$

and

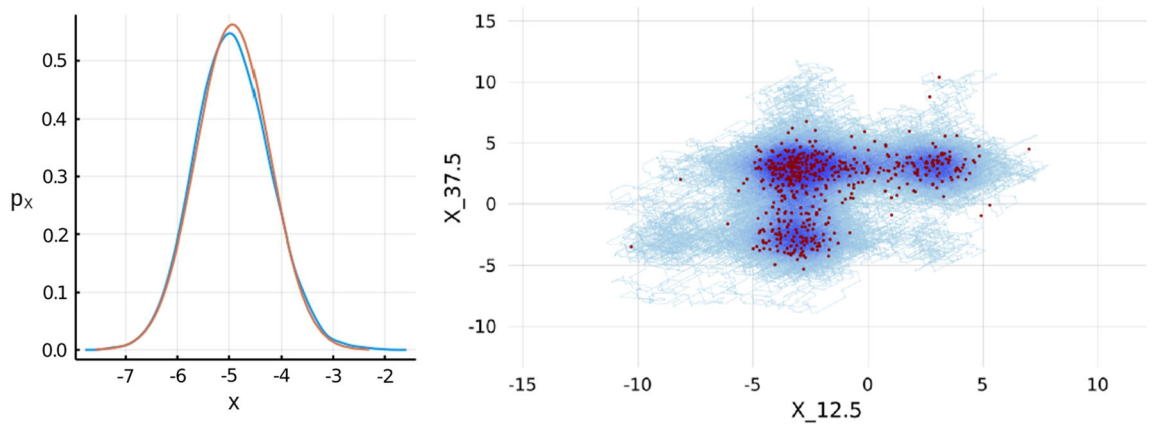

Fig. 7 On the left panel: comparison between empirical distribution (blue line, computed with a kernel estimator) and the exact distribution (red line) of the mid-point random variable X_5 for the linear diffusion (Eq. 19) with $a = -5$ and $b = -1$. The empirical distribution has been extracted from the same experiment shown in Fig. 5. On the right panel: comparison between the joint distribution of the variables $X_{T/4}$ and $X_{3T/4}$ of the process given in Eq. (20) starting at $-\pi$ and hitting π at $T = 50$. The scatter plot with red dots is obtained with ϵ-ball Euler simulation with $\epsilon = 0.1$ and discretisation $\Delta t = 0.0005$, while the blue continuous path is the Zig-Zag path. (Color figure online)

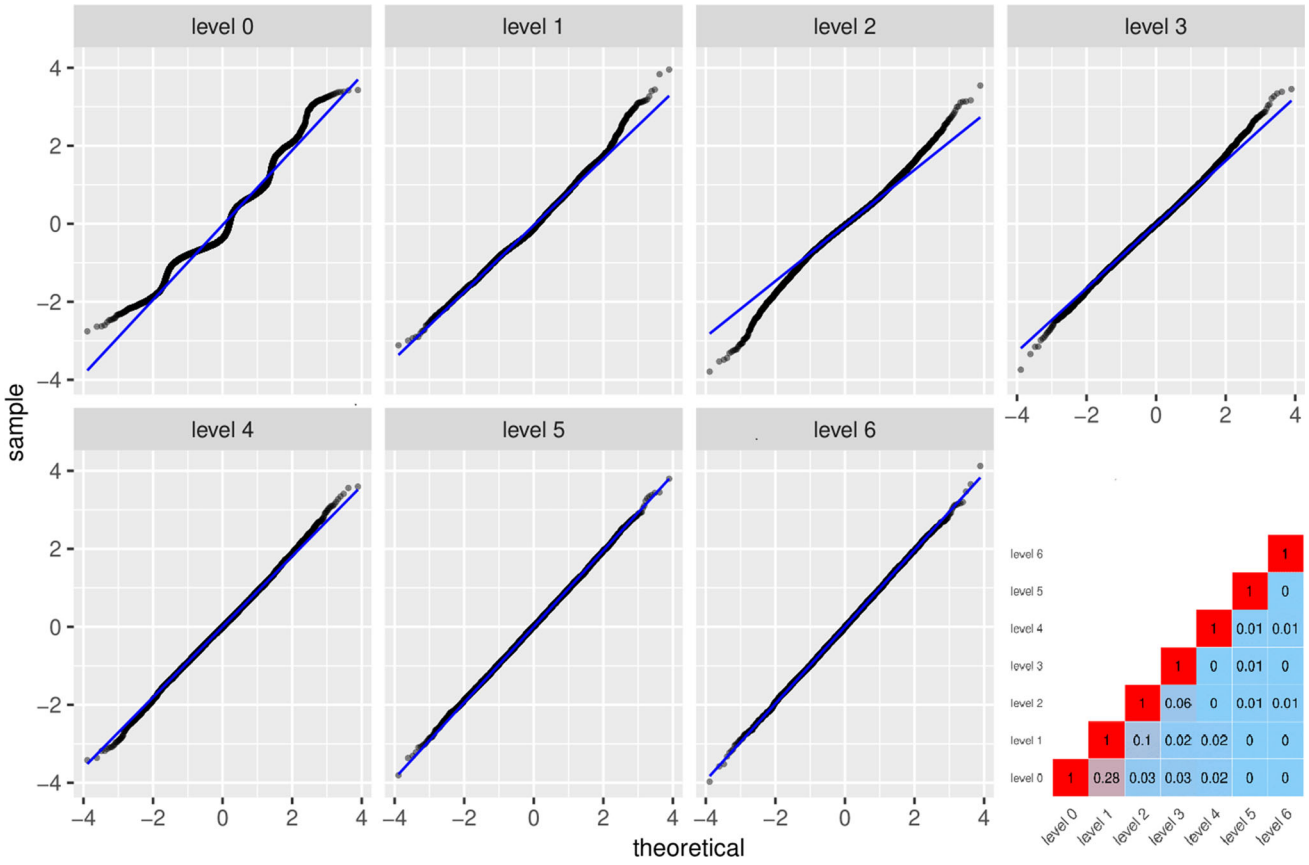

Fig. 8 Q–Q (quantile–quantile) plot against standard normal distributions of the sample path of 7 coefficients, respectively, at level 0, 1, 2, 3, 4, 5, 6 targeting the conditional bridge measure given by Eq. (20) with $\alpha = 0.7$ and initial point $u = 0$ and final point $v = 0$ at $T = 100$. On the bottom right panel, the heatmap of the absolute value of the sample correlation between the coefficients at different levels. The blue straight lines correspond to the marginal measures of the coefficients relatively to a Brownian bridge. (Color figure online)

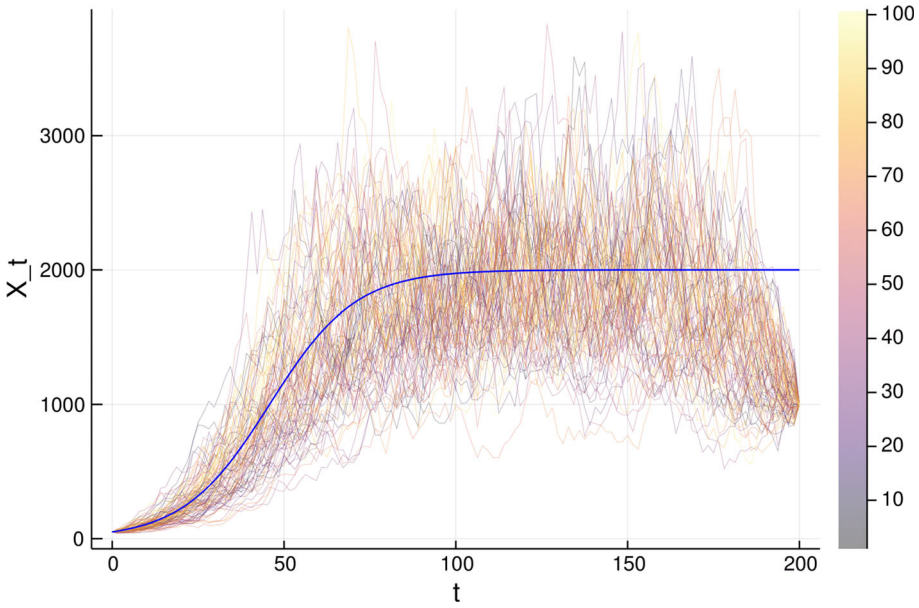

Fig. 9 Simulation of the diffusion bridge measure (100 samples) given by the logistic growth model [Eq. (22)] with parameters $K = 2000, r = 0.08, \beta = 0.1$, starting at the value 50 and hitting 1000 at time 200. Truncation level $N = 6$, final clock $\tau_{\text{final}} = 1000$ and burn-in $\tau_{\text{burn-in}} = 10$. The blue smooth line is the solution of the deterministic logistic model without final condition. (Color figure online)

$$z_k^{(1)} = a_1 \exp(-\beta b_k^{(1)}),\ z_k^{(2)} = z_k^{(1)} \exp(-\beta b_k^{(1)}),$$
$$\beta_k^\star = -\beta b_k^{(2)},\ \bar{\phi}_i = \max_s \phi_i(s).$$

Using the *superposition theorem* (see, for example, Grimmett and Stirzaker 2001), we can simulate a waiting time with Poisson rate (23) by means of simulating three waiting times according to the Poisson rates $\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)}$ and then take the minimum of the three realisations. Since at any time $t > 0$, either $\lambda_k^{(2)}(t)$ or $\lambda_k^{(3)}(t)$ is 0, we just need to evaluate two waiting times. Figure 9 shows the results obtained with our method for this process. The final clock of the Zig-Zag sampler is set to $T^\star = 1000$ and initial burn-in time $\tau_{\text{burn-in}} = 10$.

5.4 Numerical comparisons

In this section, we benchmark the fully local Zig-Zag sampler against the Metropolis-adjusted Langevin algorithm (MALA) (Roberts and Rosenthal 1998), Hamiltonian Monte Carlo (HMC) (Duane 1987) and another well-known PDMP, the Bouncy particle sampler (Bouchard-Côté 2015). The Bouncy Particle sampler can use the exact subsampling technique in a very similar way as explained in Sect. 4.1. According to the scaling limit results obtained in Bierkens et al. (2020), the Zig-Zag is more efficient compared to the Bouncy Particle sampler in a high-dimensional setting when the conditional dependency graph corresponding to the target measure exhibits sparsity (which clearly is the case here). The MALA sampler is a well-known discrete-time Markov chain Monte Carlo method which performs informed updates through the gradient of the target distribution. HMC is considered a state-of-the-art algorithm. In contrast to PDMPs, for HMC and MALA the gradient needs to be fully evaluated and no subsampling methods can be exploited. Thus, the integral in (16) needs to be computed numerically, introducing bias. Furthermore, contrary to PDMPs, the resulting Markov chain is reversible. We study the performance of the samplers for the stochastic differential equation (20) with $u, v = 0$ and the time horizon $T = 100$ and we let α vary. As α increases, the target distribution on the coefficients presents higher peaks and valleys and is therefore a challenging distribution for general Markov chain Monte Carlo methods. We fix the refreshment rate of the Bouncy Particle sampler to 1 to avoid a degenerate behaviour and implement the MALA algorithm with adaptive step size over 250,000 iterations. We used the automatically tuned dynamic integration time HMC Algorithm (Betancourt 2018) with 3000 iterations and with diagonal mass matrix and integrator step size both adaptively tuned in a warm-up phase of 2000 iterations, with the latter adapted using a dual-averaging algorithm (Hoffman and Gelman 2014) with target acceptance statistic of 0.8. The algorithm is provided in the package `AdvancedHMC.jl` (see Ge et al. 2018) with 3000 iterations. The integral appearing in the gradient of the energy function is computed for the MALA sampler and for the HMC sampler numerically with a simple Euler integration scheme over 2^{N+1} points, where N is the truncation level which is fixed to 6 for all the experiments. The final clock for the PDMPs is $T' = 25,000$. We also include the numerical results of two variants of the Zig-Zag sampler:

(ZZv1) where the partial derivative in (16) is estimated by averaging over multiple independent realisations of (17), with the number of realisations proportional to the length of the range of the integral in (16);

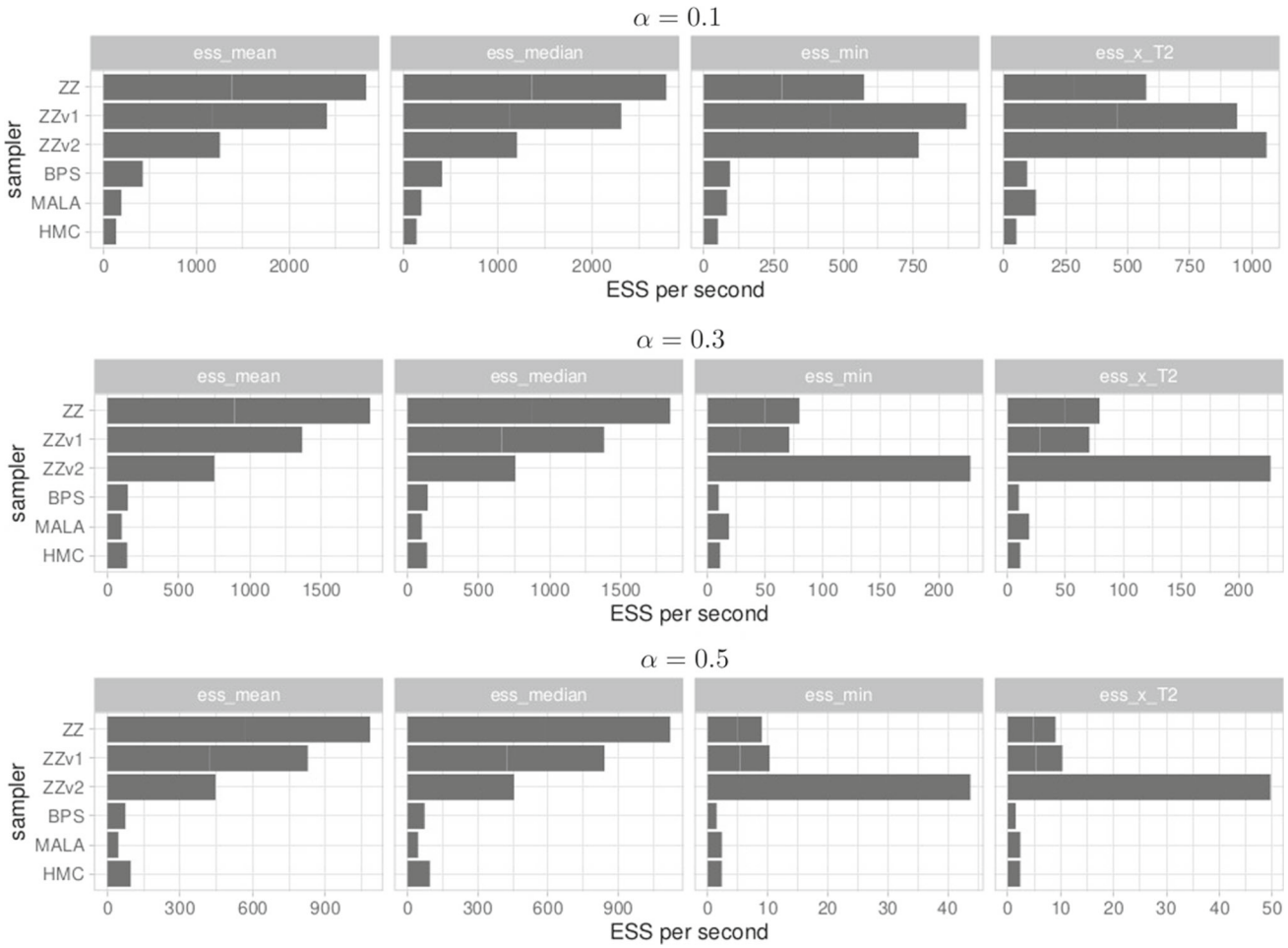

Fig. 10 Performance comparison of the fully local Zig-Zag (ZZ), its variants (ZZv1 and ZZv2), the Bouncy Particle sampler with subsampling (refreshment rate set to 1), MALA and HMC sampler. The performance measure considered here is, respectively, the effective sample size (ESS) of the middle point $X_{T/2}$, the median and the minimum of the ESS over the dimension of the coefficients of the expansion. The target diffusion bridge with drift $b(x) = \alpha \sin(x)$ with $u, v = 0$ and $T = 50$ and truncation level $N = 6$. The final clock for the PDMPs is set to $T' = 25{,}000$, the number of iterations for the MALA is set to be 250,000 with adaptive time step targeting the acceptance rate 0.6 (Pierre 2020), and the number of iteration for the HMC is 3000 with the algorithm fine-tuned by the package `AdvancedHMC.jl`. All the quantities are normalised by the runtime of execution. The asymptotic variance estimate used for computing the ESS is obtained using batch means. Notice that while the subsampling technique adopted for the piecewise deterministic Monte Carlo methods does not introduce bias on the target distribution, the numerical integration adopted for the MALA and HMC samplers introduces bias on the target distribution

(ZZv2) where the partial derivative in (16) is estimated by decomposing the range of the integral into N subintervals (with N proportional to the length of the range of the integral) and evaluating the integrand at a random point drawn inside each subinterval.

These variants of the Zig-Zag have been proposed after noticing that the coefficients at low levels are the ones deviating the most from normality and the partial derivative with respect to those coefficients have larger support. This suggests that refining the estimates of the partial derivative of the energy function only with respect to those coefficients can be beneficial and improve the performances of the PDMPs. Figure 10 shows the results obtained. The fully local Zig-Zag and its variants always outperform the Bouncy Particle sampler, the MALA and the HMC with respect to the statistics considered, namely the mean, median and minimum of the effective sample size computed for each coefficient of the Faber–Schauder expansion and the effective sample size of the coefficient $\xi_{0,0}$, which gives the middle point $X_{T/2}$ and, as shown in Fig. 10, is one of the most difficult coefficients to sample.

6 Extensions

In this section, we briefly sketch the extension of the approach presented in Sect. 3 to a class of multi-dimensional diffusion bridges. Then, we study the scaling properties of the algorithm with respect to three quantities: the time horizon of the

diffusion bridge T, the truncation level N and the dimensionality of the diffusion bridge d.

6.1 Multivariate diffusion bridge

Consider a d-dimensional diffusion bridge given the stochastic differential equation

$$\mathrm{d}X_t = \nabla B(X_t)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = u, X_T = v_T, \quad u, v_T \in \mathbb{R}^d,$$

where $(W_t)_{t\geq 0}$ is a d-dimensional Wiener process and $\nabla B : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a conservative vector field, i.e. the gradient of some scalar-valued function B. Denote its law by $\mathbb{P}^{u,v_T}$. Similarly to Eq. (10), under mild assumption on $\nabla B(X_t)$, we can write the change of measure between $\mathbb{P}^{u,v_T}$ and the standard d-dimensional Wiener bridge measure $\mathbb{Q}^{u,v_T}$ as

$$\frac{\mathrm{d}\mathbb{P}^{u,v_T}}{\mathrm{d}\mathbb{Q}^{u,v_T}}(X) = C \exp\left\{B(X_T) - B(X_0) - \frac{1}{2}\int_0^T \|b(X_t)\|^2 + \Delta B(X_t)\,\mathrm{d}t\right\},$$

where $b = \nabla B$, ΔB is the Laplacian of B, and C is a normalisation constant which depends on u, v_T and T. It is straightforward to derive an equivalent approximated measure as done in equation (12) and prove Theorem 3.11 in the multi-dimensional setting. In this case, the d-dimensional diffusion bridge measure is approximated by the same truncated expansion of equation (2) with coefficients $\xi_{i,j}, i = 0, \ldots, N;\ j = 0, \ldots, 2^N$ which now are d-dimensional random vectors. The total dimensionality of the target density for diffusion bridges becomes $d(2^{N+1} - 1)$. Similarly to the one-dimensional case, Proposition 7.1 holds. (The proof follows in a similar fashion of the proof of Proposition 7.1 and is omitted for brevity.) The Poisson rates $\lambda^k_{i,j}$ (where, $k \in \{1, \ldots, d\}$ defines the coordinate of the d-dimensional process) are functions of the sets $N^k_{i,j}$ which have maximum admissible size $|N^k_{i,j}| = d(2^{N-i+1} + i - 1) \leq d(2^{N+1} - 1)$ so that Assumption 4.1 holds.

6.2 Scaling for large *T*, *N*, *d*

The following scaling analysis serves as preliminary work for future explorations. The expected run time of the fully local Zig-Zag sampler (Algorithm 4) is intimately related to the number of Poisson event times for a fixed final clock τ_{final} and the conditional independence structure appearing in the target measure. The former is determined by the size of the Poisson bounding rates $\bar{\lambda}_1, \ldots, \bar{\lambda}_M$, while the latter is defined by the sets $N_1, \ldots, N_M$ and determines the complexity of the *local step* of Algorithm 3.

Remark 6.1 For a fixed position and velocity, the Poisson bounding rates used in the Zig-Zag sampler with subsampling (Algorithm 2) for diffusion bridges are of the form $\bar{\lambda}_{i,j} = C_1 T^{3/2} 2^{-3i/2} + C_2$, $i = 0, 1, \ldots, N;\ j = 0, 1, \ldots, 2^i - 1$, for some terms C_1 and C_2 which do not depend on i and T.

Proof For every $i = 0, 1, \ldots, N;\ j = 0, 1, \ldots, 2^i - 1$, the time horizon T and scaling index i enter in the bounding rates of (18) through the terms $S_{i,j}$ and $\bar{\phi}_{i,j}$. The first term is of $\mathcal{O}(T2^{-i})$, and the second one is of $\mathcal{O}(\sqrt{T}2^{-i/2})$. □

Proposition 6.1 helps in understanding how the complexity of the algorithm scales as T grows and as the truncation level N grows. As T grows, the Poisson rates increase with order $T^{3/2}$ so that the total number of Poisson events for a fixed Zig-Zag clock increases with the same order.

Furthermore, as the truncation level N grows, the change of measure affects less and less the coefficients in high levels and the partial derivative of the energy function goes to zero with rate $2^{-3N/2}$) implying that the for large N, $\bar{\lambda}_{N,j} \approx C_2 = (\xi_{N,j}\theta_{N,j})^+$ (which is the Poisson rate for the Brownian bridge). As a consequence, the Poisson processes of the coefficients in high levels (i large) will be approximately independent with all the other coefficients and not function of the level i so that the complexity of Algorithm 4 scales approximately linearly with the number of mesh points. This is opposed to the standard Zig-Zag algorithm (Algorithm 1) which does not take advantage of the approximate independence of the coefficients in high levels so that the $2^{N+1} - 1$ waiting times have to be renovated at every reflection of each coefficient.

The scaling result under mesh refinement (when N grows) is unsatisfactory as the algorithm deteriorates when the resolution of the path increases. A partial solution can be obtained by letting the absolute value of the marginal velocities $|\theta_{N,j}|$ to decrease as N increases. This would enhance the scaling property of the algorithm under mesh refinement at the cost of a slow mixing of high-level components. An alternative solution is considered in Bierkens et al. (2018) where the authors enhance the scaling property of the algorithm by replacing the Zig-Zag algorithm with the *Factorised Boomerang sampler*. The Factorised Boomerang sampler differs from the Zig-Zag by having curved trajectories which are invariant to a prescribed Gaussian measure. This allows the process to sample from the Gaussian measure (Brownian bridge measure) at barely no cost. However, the main drawback of the factorised Boomerang sampler is the current limiting techniques for simulating Poisson times given the curved trajectories which lead to Poisson upper bounds which are not tight.

Finally, when the dimensionality of the diffusion bridge is $d \gg 1$, both the dimensionality of the target density of the Zig-Zag sampler and the sets $N^k_{i,j}$ for $i = 0, \ldots, N;\ j = 0, \ldots, 2^i - 1;\ k = 1, \ldots, d$ grow linearly with d so that, in general, we expect the computational time to grow with rate d^2. When the drift of the multi-dimensional bridge presents a sparse structure, i.e. not all coordinates of the differential equation interact directly with each other, as common in the high-dimensional case arising from discretised stochastic partial differential equations (e.g. Michel et al. 2019, Sect. 6),

the size of those sets reduces considerably until the extreme case of d independent diffusion bridges where the sets $N^k_{i,j}$ are not anymore a function of d and clearly the complexity grows linearly with the dimensionality d.

7 Conclusions

In this paper, we have introduced a new method for the simulation of diffusion bridges which substantially differs from existing methods by using the Zig-Zag sampler and the basis of representation adopted. We motivated both choices and presented the method and its implementation. The resulting simulated bridge measures are shown to be close to the real measures, even for low-dimensional approximations and bridges which are highly nonlinear. We took advantage of the subsampling technique and a local version of the Zig-Zag to sample high-dimensional approximation to conditional measures of diffusions with intractable transition densities. The subsampling technique is a key property in favour of using piecewise deterministic Monte Carlo methods for diffusion bridges (and whenever the target measure is expressed as an integral that requires numerical evaluation). However, the main limitation found for these methods is that they rely on upper bounds of the Poisson rates which are model-specific. Upper bounds for PDMC are easily derived in situations where the log-likelihood has a bounded Hessian. In our setting, this means that we wish for the function $b^2(x) - b'(x)$ to have bounded second derivative. In other cases, tailor-made bounds need to be derived which can be substantially more complicated. Furthermore, the performance of these samplers can be affected if the upper bounds are too large.

In conclusion, this is the first time (to our knowledge) the Zig-Zag has been employed in a high-dimensional practical setting. We claim that the promising results will open research towards applications of the Zig-Zag for high-dimensional problems. We mention below some possible extensions of the methodology proposed which are left for future research:

a. The hierarchical structure of the Faber–Schauder basis suggests that the Zig-Zag should explore the space at different velocities to achieve optimal performance. Unfortunately, it is not immediately clear how to tune the velocity vector;
b. In Sect. 6, we anticipated the possibility to simulate multi-dimensional diffusion bridges. In order to generalise the results presented in this paper, we assumed the drift being a conservative vector field. In order to relax this limiting assumption, new convergence results have to be derived which deal explicitly with the stochastic integral appearing in equation (8).
c. The driving motivation for proposing this methodology is to perform parameter estimation of a discretely observed diffusion model. For this purpose, the Zig-Zag sampler runs jointly on the augmented path space given by the coefficients ξ and the parameter space Θ.

Acknowledgements This work is part of the research programme *Bayesian inference for high-dimensional processes* with project number 613.009.034c, which is (partly) financed by the Dutch Research Council (NWO) under the *Stochastics – Theoretical and Applied Research* (STAR) grant. J. Bierkens acknowledges support by the NWO for the research project *Zig-zagging through computational barriers* with project number 016.Vidi.189.043. The authors are thankful to Gareth Roberts and Marcin Mider for fruitful discussion and grateful to the reviewers for valuable input.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A Factorisation of the diffusion bridge measure

Here, we derive rigorously the conditional independence structure of the coefficients which arise from the compact support of the Faber–Schauder functions as shown in Fig. 4. Recall that the relation $\xi_{i,j} \ll \xi_{k,l}$ holds if $S_{k,l} \subset S_{i,j}$ and in that case we refer to $\xi_{i,j}$ as the *ancestor* of $\xi_{k,l}$ (and conversely $\xi_{k,l}$ as the *descendant*). Notice that each coefficient is both descendant and ancestor of itself.

Proposition 7.1 (Conditional independence structure) *Denote the set of common ancestors of $\xi_{i,j}$ and $\xi_{k,l}$ by $A_{(i,j;k,l)} := \{\xi_{h,d} : \xi_{h,d} \ll \xi_{k,l} \wedge \xi_{h,d} \ll \xi_{i,j}\}$. Under $\mathbb{P}_N^{v_T}$, $\xi_{i,j}$ is conditionally independent from $\xi_{k,l}$, given the set $A_{(i,j;k,l)}$, whenever the interior of the supports of their basis function is disjoint that is neither $\xi_{i,j} \ll \xi_{k,l}$ nor $\xi_{k,l} \ll \xi_{i,j}$ is satisfied.*

Proof For $i = 1, \ldots, N$; $j = 1, \ldots, 2^i - 1$, define the vectors of ancestors and descendants of $\xi_{i,j}$ as $\xi^{(i,j)} := \{\xi_{m,n} : \xi_{m,n} \ll \xi_{i,j} \vee \xi_{m,n} \gg \xi_{i,j}\}$. Assume, without loss of generality, that $i \leq k$ and consider two coefficients $\xi_{i,j}, \xi_{k,l}$. We factorise $Z^N(X)$ by partitioning the integration interval $[0, T]$ in a sequence of sub-intervals $S_{k,0}, S_{k,1}, \ldots, S_{k,2^k-1}$ so that

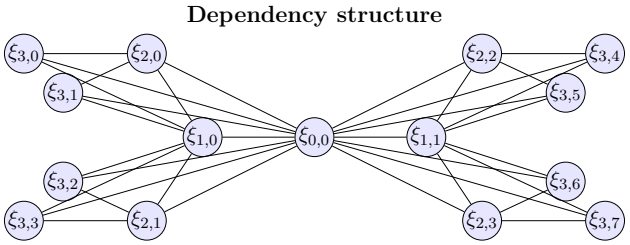

Fig. 11 Graphical representation of the dependency structure of the random vector of the coefficients under $\mathbb{P}_N^{u,v_T}$. $\xi_{i,j} \perp\!\!\!\perp \xi_{k,l}$ conditionally on the vertices which have a direct hedge to both $\xi_{i,j}$ and $\xi_{k,l}$ if $\xi_{i,j}$ does not have a direct edge to $\xi_{k,l}$. The dependency graph is a *chordal graph*

$$Z^N(X) = \prod_{p=1}^{2^k-1} f_{k,p}(\xi^{(k,p)}). \tag{24}$$

Here,

$$f_{k,p}(\xi^{(k,p)}) = \exp\left(B(X^N_{\max S_{k,p}}) - B(X^N_{\min S_{k,p}}) \times -\frac{1}{2}\int_{S_{k,p}} b^2\left(X_s^{N;k,p}\right) + b'\left(X_s^{N;k,p}\right) \mathrm{d}s\right).$$

with

$$X_s^{N;k,p} = \bar{\bar{\phi}}(s)u + \bar{\phi}(s)v_T/\sqrt{T} + \sum_{(i,j):\, \xi_{i,j} \ll \xi_{k,p}} \phi_{i,j}(s)\xi_{i,j}$$

and we used that $X_s^N = X_s^{N;i,j}$ when $s \in S_{i,j}$, $X_T^N = \bar{\phi}(T)v_T/\sqrt{T}$ and $X_0^N = \bar{\bar{\phi}}(0)u$. Now, just notice that, under this factorisation, the only factor which is a function of $\xi_{k,l}$ is $f_{k,l}(\xi^{(k,l)})$. Here, if $\xi_{i,j} \not\ll \xi_{k,l}$, then $\xi^{(k,l)}$ does not contain $\xi_{i,j}$. Conversely, the factors containing $\xi_{i,j}$ are those $f_{k,p}(\xi^{(k,p)})$ such that $\xi_{i,j} \ll \xi_{k,p}$ with $p = 0, 1, \ldots, 2^k - 1$. If $\xi_{i,j} \not\ll \xi_{k,l}$, none of the vectors $\xi^{(k,p)}$ contains $\xi_{k,l}$. Since, under the measure $\mathbb{Q}^{u,v_T}$, the random variables in the vector ξ^N are pairwise independent, the factorisation on $Z^N(X)$ defines the dependency structure of the vector ξ^N under $\mathbb{P}_N^{v_T}$ so that $\xi_{i,j}$ and $\xi_{k,l}$ are independent conditionally on their common coefficients given by the set $A_{(i,j;k,l)}$. □

More intuitively, the factorisation of $Z(X)$ gives rise to the dependency graph displayed in Fig. 11 which shows that the coefficients in high *levels* (i large) are coupled with just few other coefficients and conditionally independent from all the remaining. The conditional independence of the coefficients implies that the partial derivatives of the energy function (and consequently the Poisson rates given by equation (6)) are functions of only few coefficients in the sense of Assumption 4.1. In particular, the sets in Assumption 4.1 (using double indexing) can be chosen as $N_{i,j} = \{\xi_{h,d} : \xi_{h,d} \ll \xi_{i,j} \vee \xi_{h,d} \gg \xi_{i,j}\}$ with size $|N_{i,j}| = 2^{N-i+1} + i - 1$, where N is the truncation level.

References

Andrieu, C., Livingstone, S.: Peskun-Tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario (2019). arXiv:1906.06197

Andrieu, C. et al.: Hypocoercivity of Piecewise Deterministic Markov Process-Monte Carlo. (2018). arXiv:1808.08592

Beskos, A., Papaspiliopoulos, O., Roberts, G.O. et al.: Retrospective exact simulation of diffusion sample paths with applications. Bernoulli **12**(6), pp. 1077–1098 (2006)

Betancourt M. : A Conceptual Introduction to Hamiltonian Monte Carlo (2018). arXiv:1701.02434

Bierkens, J., Fearnhead, P., Roberts, G.: The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. Ann. Stat. **47**(3) , pp. 1288–1320 (2019). https://doi.org/10.1214/18-AOS1715

Bierkens, J., Kamatani, K., Roberts, G.O.: High-dimensional scaling limits of piecewise deterministic sampling algorithms (2018). arXiv:1807.11358

Bierkens, J. et al.: The Boomerang Sampler. (2020). arXiv:2006.13777

Bierkens, J., van der Meulen, F., Schauer, M.: Simulation of elliptic and hypo-elliptic conditional diffusions. Adv. Appl. Probab. **52**(1), 173–212 (2020). https://doi.org/10.1017/apr.2019.54

Bladt, M., Sørensen, M., et al.: Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Bernoulli **20**(2), 645–675 (2014)

Bouchard-Côté, A., Vollmer, S.J., Doucet, A.: The Bouncy particle sampler: a non-reversible rejection-free markov chain monte carlo method. In: (2015). arXiv:1510.02451

Davis, M.H.A.: Markov Models & Optimization. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis (1993). ISBN: 9780412314100

Diaconis, P., Holmes, S., Neal, R.M .: Analysis of a nonreversible Markov chain sampler. In: Annals of Applied Probability, pp. 726–752 (2000)

Duane, S. et al.: Hybrid Monte Carlo. Phys. Lett. B **195**(2), 216–222 (1987). ISSN: 0370-2693. https://doi.org/10.1016/0370-2693(87)91197-X

Faulkner, M.F. et al.: All-atom computations with irreversible Markov chains. J. Chem. Phys. **149**(6), 064113 (2018). ISSN: 1089-7690. https://doi.org/10.1063/1.5036638

Fearnhead, P. et al.: Piecewise deterministic markov processes for continuous-time monte carlo. Stat. Sci. **33**(3), 386–412 (2018). https://doi.org/10.1214/18-STS648

Ge, H., Xu, K., Ghahramani, Z.: Turing: a language for exible probabilistic inference. In: International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain, pp. 1682–1690 (2018). http://proceedings.mlr.press/v84/ge18b.html

Grimmett, G., Stirzaker, D.: Probability and Random Processes. Oxford University Press, Oxford (2001)

Hoffman M.D., Gelman A : The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res.15(1), 1593–1623 (2014)

Karatzas, I., Shreve, S.E .: Brownian motion and stochastic calculus. In: Graduate texts in Mathematics 113 (1991)

Klebaner, F.C.: Introduction to Stochastic Calculus with Applications. World Scientific Publishing Company (2005)

Liptser, R.S., Aries, B., Shiryaev, A.N.: Statistics of Random Processes: I. General Theory. Stochastic Modelling and Applied Probability. Springer, Berlin (2013). ISBN: 9783662130438

McKean, H.P.: Stochastic Integrals, vol. 353. American Mathematical Society (1969)

Michel, M., Tan, X., Deng, Y.: Clock Monte Carlo methods. Phys. Rev. E **99**(1) (2019). ISSN: 2470-0053. https://doi.org/10.1103/physreve.99.010105

Mider, M., Schauer, M., van der Meulen, F.: Continuous-discrete smoothing of diffusions. (2020). arXiv: 1712.03807

Mider, M. et al.: Simulating bridges using con uent diffusions (2019). arXiv: 1903.10184

Peters, E.A.J.F.: Rejection-free Monte Carlo sampling for general potentials. Phys. Rev. E **85**(2) (2012). ISSN: 1550-2376. https://doi.org/10.1103/PhysRevE.85.026703

Pierre, M. et al.: Velocity jump processes: an alternative to multi-time step methods for faster and accurate molecular dynamics simulations. J. Chem. Phys. **153**(2), 024101 (2020). ISSN: 1089-7690. https://doi.org/10.1063/5.0005060

Roberts, G.O., Rosenthal, J.S.: Optimal scaling of discrete approximations to Langevin diffusions. J. R. Stat. Soc.: Ser. B (Statistical Methodology) **60**(1), 255–268 (1998)

Roberts, G.O., Stramer, O.: On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika **88**(3), 603–621 (2001)

Roberts, G.O., Tweedie, R.L.: Exponential convergence of Langevin distributions and their discrete approximations. In: Bernoulli 2.4 , pp. 341–363 (1996). ISSN: 13507265. http://www.jstor.org/stable/3318418

Schauer, M., Grazzi, S.: ZigZagBoomerang: v0.5.3. https://www.github.com/mschauer/ZigZagBoomerang.jl. 2020. https://doi.org/10.5281/zenodo.3931118

van der Meulen, F., Schauer, M., van Waaij, J.: Adaptive nonparametric drift estimation for diffusion processes using Faber–Schauder expansions. Stat. Inference Stoch. Process. **21**(3), 603–628 (2018)

van der Meulen, F., Schauer, M.: Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. Electron. J. Stat. **11**(1), 2358–2396 (2017). https://doi.org/10.1214/17-EJS1290

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.