High-dimensional scaling limits of piecewise deterministic sampling algorithms

Bierkens, Joris; Kamatani, Kengo; Roberts, Gareth O.

DOI
10.1214/21-AAP1762

Publication date
2022

Document Version
Final published version

Published in
Annals of Applied Probability

The Annals of Applied Probability
2022, Vol. 32, No. 5, 3361–3407
https://doi.org/10.1214/21-AAP1762

HIGH-DIMENSIONAL SCALING LIMITS OF PIECEWISE DETERMINISTIC SAMPLING ALGORITHMS

BY JORIS BIERKENS[1,a], KENGO KAMATANI[2,b] AND GARETH O. ROBERTS[3,c]

[1]*Delft Institute of Applied Mathematics, Delft University of Technology,* [a]*joris.bierkens@tudelft.nl*
[2]*Graduate School of Engineering Science, Osaka University,* [b]*kamatani@sigmath.es.osaka-u.ac.jp*
[3]*Department of Statistics, University of Warwick,* [c]*Gareth.O.Roberts@warwick.ac.uk*

Piecewise deterministic Markov processes are an important new tool in the design of Markov chain Monte Carlo algorithms. Two examples of fundamental importance are the bouncy particle sampler (BPS) and the zig–zag process (ZZ). In this paper scaling limits for both algorithms are determined. Here the dimensionality of the space tends towards infinity and the target distribution is the multivariate standard normal distribution. For several quantities of interest (angular momentum, first coordinate and negative log-density) the scaling limits show qualitatively very different and rich behaviour. Based on these scaling limits the performance of the two algorithms in high dimensions can be compared. Although for angular momentum both processes require only a computational effort of $O(d)$ to obtain approximately independent samples, the computational effort for negative log-density and first coordinate differ: for these BPS requires $O(d^2)$ computational effort whereas ZZ requires $O(d)$. Finally we provide a criterion for the choice of the refreshment rate of BPS.

1. Introduction. Piecewise deterministic Markov processes (PDMPs, [13]) have turned out to be of substantial interest for Monte Carlo analysis, see, for example, [3, 8, 31, 37], which have particularly focused on potential for applications in Bayesian statistics, although their uses are far wider, see, for example, [29, 32] for applications in physics. However, there are still substantial gaps in our understanding of their theoretical properties. Even results about the ergodicity of these methods (including irreducibility and exponential ergodicity problems) often involve intricate and complex problems [5, 11, 14].

The two main PDMP methodologies for Monte Carlo algorithms are the zig–zag [3] and the bouncy particle sampler (BPS) [8], and we refer to these papers for applications of these methods. Interesting hybrid strategies are certainly possible but are currently under-explored. The important practical question for Monte Carlo practitioners concerns which methodology should be chosen, with currently available empirical comparisons giving mixed results.

The focus of the present paper is on shedding some light on these questions by providing a high-dimensional analysis of these two classes of PDMPs. Our approach will identify weak limits of PDMP chains (suitably speeded up) as dimension goes to infinity. Such analyses are of interest in connection with computational cost estimation of Monte Carlo methods (see, e.g., [34, 36]).

Since we focus on the theoretical properties, in this article we do not give the full description of the implementation of PDMPs. Monte Carlo methods based on PDMPs are new techniques and their implementation is not straightforward. The main difficulty in implementing PDMPs is the generation of nonhomogeneous Poisson processes corresponding to the jump components of PDMPs. This is an active area of research and progress has been made in this regard. See, for example, Section 3 of [3], Section 2.3 of [8] and [12].

Received July 2019; revised May 2021.
MSC2020 subject classifications. Primary 60F05; secondary 65C05.
Key words and phrases. Piecewise deterministic Markov processes, weak convergence, Markov chain Monte Carlo, exponential ergodicity, Gaussian process.

1.1. *Piecewise deterministic Markov processes.* We shall consider two particular classes of PDMPs (zig–zag and BPS) which have proved to be valuable for Monte Carlo sampling. Their constructions begin in the same way. We are interested in sampling from a target distribution Π^d which has density $Z_d^{-1}\exp(-\Psi^d(\xi))$ with respect to d-dimensional Lebesgue measure with normalising constant

$$Z_d = \int_{\mathbb{R}^d} \exp\bigl(-\Psi^d(\xi)\bigr)\,\mathrm{d}\xi < \infty. \tag{1.1}$$

Zig-zag and BPS proceed by augmenting this space to include an independent velocity variable taking values uniformly on a prescribed space $\Theta \subset \mathbb{R}^d$. Both algorithms define Piecewise deterministic Markov dynamics which preserve this extended target distribution on the augmented state space $E^d = \mathbb{R}^d \times \Theta$. The difference between zig–zag and BPS lies in the choice of Θ and the dynamics for moving between velocities.

For both algorithms we shall make use of independent standardised homogeneous Poisson measures, N say, on $\mathbb{R}_+ \times \mathbb{R}_+$, so that $\mathbb{E}[N(\mathrm{d}t, \mathrm{d}x)] = \mathrm{d}t\,\mathrm{d}x$. In our notation we will use a superscript Z to indicate the zig–zag process, and a superscript B to refer to the bouncy particle sampler.

1.1.1. *Zig–zag sampler.* For the zig–zag sampler the set of possible directions is given by

$$\Theta = \mathfrak{C}^{d-1} := \{-1, +1\}^d,$$

with χ_d denoting the uniform distribution on $\mathfrak{C}^{d-1}$, and constructs a Markov chain on the state space $E^{\mathrm{Z},d} = \mathbb{R}^d \times \mathfrak{C}^{d-1}$. Let $\lambda^{\mathrm{Z},d} = (\lambda_1^{\mathrm{Z},d}, \dots, \lambda_d^{\mathrm{Z},d}) : E^{\mathrm{Z},d} \to \mathbb{R}_+^d$. The zig–zag sampler with the jump rate $\lambda^{\mathrm{Z},d}$ generates a Markov process $\{x_t^{\mathrm{Z},d} = (\xi_t^{\mathrm{Z},d}, v_t^{\mathrm{Z},d})\}_{t\ge 0}$ on $E^{\mathrm{Z},d}$ such that

$$\xi_t^{\mathrm{Z},d} = \xi_0^{\mathrm{Z},d} + \int_0^t v_s^{\mathrm{Z},d}\,\mathrm{d}s, \quad (t \ge 0),$$

and $v_t^{\mathrm{Z},d} = (v_{1,t}^{\mathrm{Z},d}, \dots, v_{d,t}^{\mathrm{Z},d})$ is defined by

$$v_{i,t}^{\mathrm{Z},d} = v_{i,0}^{\mathrm{Z},d} - 2\int_{(0,t]\times\mathbb{R}_+} v_{i,s-}^{\mathrm{Z},d} 1_{\{z \le \lambda_i^{\mathrm{Z},d}(x_{s-}^{\mathrm{Z},d})\}} N^i(\mathrm{d}s, \mathrm{d}z) \quad (t \ge 0, i = 1, \dots, d)$$

for independent Poisson random measures $N^1, \dots N^d$, where $x_0^{\mathrm{Z},d} = (\xi_0^{\mathrm{Z},d}, v_0^{\mathrm{Z},d})$ is an $E^{\mathrm{Z},d}$-valued random variable.

1.1.2. *Bouncy particle sampler.* For the bouncy particle sampler the set of possible directions is given by

$$\Theta := \mathfrak{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|^2 = 1\}$$

with ψ_d denoting the uniform distribution on $\mathfrak{S}^{d-1}$, and constructs a Markov chain on the state space $E^{\mathrm{B},d} = \mathbb{R}^d \times \mathfrak{S}^{d-1}$. Let $\kappa^d : E^{\mathrm{B},d} \to \mathfrak{S}^{d-1}$ be a function and let $\lambda^{\mathrm{B},d} : E^{\mathrm{B},d} \to \mathbb{R}_+$. Then BPS with the jump rate $\lambda^{\mathrm{B},d}$ and the refreshment rate $\rho^d > 0$ generates a Markov process $\{x_t^{\mathrm{B},d} = (\xi_t^{\mathrm{B},d}, v_t^{\mathrm{B},d})\}_{t\ge 0}$ defined by

$$\xi_t^{\mathrm{B},d} = \xi_0^{\mathrm{B},d} + \int_0^t v_s^{\mathrm{B},d}\,\mathrm{d}s \quad (t \ge 0),$$

and $v_t^{\mathrm{B},d}$ is defined by

$$v_t^{\mathrm{B},d} = v_0^{\mathrm{B},d} + \int_{(0,t]\times\mathbb{R}_+} \left(\kappa^d\left(x_{s-}^{\mathrm{B},d}\right) - v_{s-}^{\mathrm{B},d}\right) 1_{\{z\le\lambda^{\mathrm{B},d}(x_{s-}^{\mathrm{B},d})\}} N(\mathrm{d}s,\mathrm{d}z) + \int_{(0,t]\times\mathfrak{S}^{d-1}} \left(u - v_{s-}^{\mathrm{B},d}\right) R_d(\mathrm{d}s,\mathrm{d}u),$$

where R_d is a homogeneous random measure which is independent from N with intensity measure

$$\mathbb{E}\left[R_d(\mathrm{d}s,\mathrm{d}u)\right] = \rho^d\,\mathrm{d}s\,\psi_d(\mathrm{d}u).$$

Without refreshment the bouncy particle sampler may not be ergodic in general [8]. The refreshment rate using the random measure R_d was referred to as restricted refreshment in [8], and other choices were also considered in that paper.

In this work, we describe the zig–zag sampler and the bouncy particle sampler as jump processes, since it is straightforward to derive semimartingale properties with these forms. In contrast, in the Monte Carlo community, these Markov processes are usually described with stopping times as in [3, 8]. It looks different, but the associated jump process can be recovered from stopping times as described in Section 4 of [13]. Using stopping times, the process $v_{i,t}^{\mathrm{Z},d}$ has the survival function

$$F_u = \exp\left(-\int_t^{t+u} \lambda_i^{\mathrm{Z},d}\left(x_s^{\mathrm{Z},d}\right)\mathrm{d}s\right) \quad (u>0)$$

for the next jump time when the current time is t. This gives the hazard rate $-F_t'/F_t = \lambda_i^{\mathrm{Z},d}(x_t^{\mathrm{Z},d})$. On the other hand, the jump size is $-2v_{i,t}^{\mathrm{Z},d}$. The compensator of the random measure associated with the jump of $z_{i,t}^{\mathrm{Z},d}$ is the product of the Dirac measure at the jump size $-2v_{i,t}^{\mathrm{Z},d}$ and the hazard rate distribution, that is,

$$\begin{aligned}\mu_i^P(\mathrm{d}t,\mathrm{d}x) &= \delta_{\{-2v_{i,t}^{\mathrm{Z},d}\}}(\mathrm{d}x)\lambda_i^{\mathrm{Z},d}\left(x_t^{\mathrm{Z},d}\right)\mathrm{d}t \\ &= \int_{z\in\mathbb{R}_+} \delta_{\{-2v_{i,t}^{\mathrm{Z},d}\}}(\mathrm{d}x) 1_{\{z\le\lambda_i^{\mathrm{Z},d}(x_t^{\mathrm{Z},d})\}}\,\mathrm{d}t\,\mathrm{d}z.\end{aligned}$$

Therefore, the random measure associated with the jump of $z_{i,t}^{\mathrm{Z},d}$ can be denoted as

$$\mu_i(\mathrm{d}t,\mathrm{d}x) = \int_{z\in\mathbb{R}_+} \delta_{\{-2v_{i,t}^{\mathrm{Z},d}\}}(\mathrm{d}x) 1_{\{z\le\lambda_i^{\mathrm{Z},d}(x_t^{\mathrm{Z},d})\}} N^i(\mathrm{d}t,\mathrm{d}z)$$

for a Poisson random measure N^i. The process $v_{i,t}^{\mathrm{Z},d}$ is a pure jump process, and so we have

$$v_{i,t}^{\mathrm{Z},d} = v_{i,0}^{\mathrm{Z},d} + \int_{(0,t]\times\mathbb{R}} x\,\mu_i(\mathrm{d}s,\mathrm{d}x),$$

which yield the form presented in Section 1.1.1. A similar derivation yields the form of the bouncy particle sampler.

1.2. *Finite dimensional properties.* In this section we briefly review finite dimensional properties of the piecewise deterministic processes. Here and elsewhere, we denote the d-dimensional Euclidean inner product by $\langle x,y\rangle = \sum_{i=1}^d x_i y_i$ and the norm by $\|x\| = (\langle x,x\rangle)^{1/2}$.

Let $F_i(v)$ be the function that switches the sign of the ith element of $v \in \mathfrak{C}^{d-1}$. By Theorem II.2.42 of [23] and Proposition VII.1.7 of [33], the infinitesimal generator $L^{\mathrm{Z},d}$ of the Markov process corresponding to the zig–zag sampler is defined by

$$\big(L^{\mathrm{Z},d}\varphi\big)(\xi, v) = \big\langle \nabla_\xi \varphi(\xi, v), v\big\rangle + \sum_{i=1}^{d} \lambda_i^{\mathrm{Z},d}(\xi, v)\big(\varphi\big(\xi, F_i(v)\big) - \varphi(\xi, v)\big)$$

for $\varphi : E^{\mathrm{Z},d} \to \mathbb{R}$ such that $\varphi(\cdot, v) \in C_0^1(\mathbb{R}^d)$ $(v \in \mathfrak{C}^{d-1})$ where $C_0^1(\mathbb{R}^d)$ is the set of differentiable functions with compact support. Here, $\nabla_\xi = (\partial/\partial \xi_i)_{i=1,\dots,d}$ is the derivative operator and we will denote it by ∇ when there is no ambiguity. Let $\Psi^d : \mathbb{R}^d \to \mathbb{R}_+$ be a smooth function with (1.1). Set $\lambda^{\mathrm{Z},d}(x)$ so that $\lambda_i^{\mathrm{Z},d}(\xi, v) - \lambda_i^{\mathrm{Z},d}(\xi, F_i(v)) = \partial_i \Psi^d(\xi) v_i$. As discussed in, for example, [3, 5], the Markov process corresponding to the zig–zag sampler is $\Pi^{\mathrm{Z},d}$-invariant where $\Pi^{\mathrm{Z},d} = \Pi^d \otimes \chi_d$.

The infinitesimal generator $L^{\mathrm{B},d}$ of the Markov process corresponding to the bouncy particle sampler is defined by

$$\begin{aligned}\big(L^{\mathrm{B},d}\varphi\big)(\xi, v) &= \big\langle \nabla_\xi \varphi(\xi, v), v\big\rangle + \lambda^{\mathrm{B},d}(\xi, v)\big(\varphi\big(\xi, \kappa^d(\xi, v)\big) - \varphi(\xi, v)\big) \\ &\quad + \rho^d\bigg(\int \varphi(\xi, u)\psi_d(\mathrm{d}u) - \varphi(\xi, v)\bigg)\end{aligned}$$

for continuous functions $\varphi : E^{\mathrm{B},d} \to \mathbb{R}$ satisfying $\varphi(\cdot, v) \in C_0^1(\mathbb{R}^d)$ $(v \in \mathfrak{S}^{d-1})$. We assume a constant refreshment rate, that is, $\rho^d \equiv \rho > 0$, and κ^d is a reflection function defined by

$$\kappa^d(\xi, v) = v - 2\frac{\langle \nabla \Psi^d(\xi), v\rangle}{\|\nabla \Psi^d(\xi)\|^2} \nabla \Psi^d(\xi) \tag{1.2}$$

and finally $\lambda^{\mathrm{B},d}(\xi, v) = \max\{\langle \nabla \Psi^d(\xi), v\rangle, 0\}$. As discussed in, for example, [8, 14] the Markov process corresponding to the bouncy particle sampler is $\Pi^{\mathrm{B},d}$ invariant, where $\Pi^{\mathrm{B},d} = \Pi^d \otimes \psi_d$.

1.3. *Summary of the main results.* In Section 2, we study the asymptotic properties of piecewise deterministic processes. This section summarises the main results in that section. For simplicity, all results in Section 2 assume that the initial value of ξ is generated from the target distribution, and the initial value of v is generated from the uniform distribution on the direction space. We only consider the standard normal case, that is,

$$\Psi^d(\xi) = \frac{\|\xi\|^2}{2}.$$

In agreement with this assumption, the jump rate of the zig–zag sampler is

$$\lambda_i^{\mathrm{Z},d}(\xi, v) = \max\{\xi_i v_i, 0\} = (\xi_i v_i)^+, \quad i = 1, \dots, d, (\xi, v) \in E^{\mathrm{Z},d},$$

and the jump rate and the refreshment rate of the bouncy particle sampler are

$$\lambda^{\mathrm{B},d}(\xi, v) = \max\big\{\langle \xi, v\rangle, 0\big\} = \langle \xi, v\rangle^+, \qquad \rho^d(\xi, v) = \rho > 0, \quad (\xi, v) \in E^{\mathrm{B},d},$$

and the reflection function satisfies (1.2). Analogous to [34], we focus on relevant finite-dimensional summary statistics. The *angular momentum* process, the *negative log-target density process* and the *first coordinate process* are defined by

$$t \mapsto \bigg\langle \xi_t, \frac{v_t}{\|v_t\|}\bigg\rangle, \qquad t \mapsto d^{1/2}\big(d^{-1}\|\xi_t\|^2 - 1\big), \quad t \mapsto \xi_{1,t},$$

TABLE 1
Size of continuous time intervals required to obtain approximately independent samples for the piecewise deterministic processes

Method	Angular momentum	Negative log-density	First Coordinate
ZZ	$O(1)$ (Thm. 2.2)	$O(1)$ (Thm. 2.5)	$O(1)$ (Thm. 2.6)
BPS	$O(1)$ (Thm. 2.8)	$O(d)$ (Thm. 2.10)	$O(d)$ (Thm. 2.13)

respectively, for both the zig–zag sampler and the bouncy particle sampler. As $d \to \infty$, the stationary distributions of these statistics converge to centred normal distributions (with variances 1, 2 and 1, respectively). We compare the convergence rates of the zig–zag sampler (ZZ) and the bouncy particle sampler (BPS) for these summary statistics. Table 1 summarises the results.

The computational effort per unit time of the processes is proportional to the number of switches per unit time interval, multiplied by the computational effort per switch. In complete generality, computational effort per switch of both zig–zag and BPS are $O(d)$. However zig–zag has the ability to exploit an available conditional independence structure to offer improved computational efficiency. For the sake of this discussion, we shall assume a particularly strong form of conditional independence, though weaker versions of this exist giving smaller computational advantages. We shall say that the target density has a *sparse conditional independence structure* if, for all i the derivative $\partial\Psi^d(\xi))$ depends only on an $O(1)$ number of components of the vector ξ. Note that this assumption is natural in statistics where models are constructed explicitly from such conditional independence relationships. In the case of sparse conditional independence structure, zig–zag achieves $O(1)$ computational effort per event. See, for example, [4] for a detailed consideration on how to benefit from sparse conditional independence.

However experiments and theory suggest that zig–zag may perform poorly in the case of highly anisotropic targets. See, for example, [1, 27].

For zig–zag and BPS, these are as given in Table 2. In particular for the case of product target distributions as studied theoretically in most of this paper, the zig–zag can be implemented with the higher efficiency described in the top row of Table 2. On the other hand we do not see a way in which the generic BPS as described in this paper can utilise conditional independence. However it is worth noting that generalisations of zig–zag termed *local BPS* by [8] and other variants as discussed in [32] also share computational advantages from sparse conditional independence. However in the context of a general partial correlation structure, implementation costs are an order of magnitude greater for the zig–zag (as is the case for relevant competitor algorithms such as MALA and HMC). Thus we give two complexities for zig–zag in Table 2 which can be thought of as best and worst cases according to the above discussion.

In order to obtain the algorithmic complexity required to draw approximately independent samples, we should multiply the required continuous time scaling with the computational

TABLE 2
Computational effort of the piecewise deterministic processes

Method	♯ events/unit time	Comp. effort/event	Combined effort/unit time
ZZ (with independence)	$O(d)$ (Cor. 2.4)	$O(1)$	$O(d)$
ZZ (general case)	$O(d)$ (Cor. 2.4)	$O(d)$	$O(d^2)$
BPS	$O(1)$ (Cor. 2.9)	$O(d)$	$O(d)$

TABLE 3
Algorithmic complexity to obtain approximately independent samples

Method	Angular momentum	Negative log-density	First Coordinate
ZZ (with independence)	$O(d)$	$O(d)$	$O(d)$
ZZ (general case)	$O(d^2)$	$O(d^2)$	$O(d^2)$
BPS	$O(d)$	$O(d^2)$	$O(d^2)$

complexity per continuous time unit. By doing so, we obtain the algorithmic complexities of the ZZ and BPS as listed in Table 3.

In terms of which algorithm, BPS or zig–zag should be implemented in any specific situation, the conclusions to the findings of Table 3 tentatively suggest that in the context of sparse conditional independence structure the zig–zag seems to have better complexity properties, but that for general target densities the methods have the same complexity. Of course these conclusions need to be treated with caution given the relatively specialised nature of the theory which underpins Table 1.

Analogous to [34], we also study the choice of the refreshment jump rate ρ that maximizes the speed of the limiting process. The limiting process of the negative log-target density of the BPS sampler is the Ornstein–Uhlenbeck process. The process attains its maximal speed when the ratio of the expected number of refreshment jumps to that of all jumps is approximately 0.7812 (see Figure 1). This result provides a practical criterion for selecting the refreshment rate; see Remark 2.12. In Section 2.3, we analyse this criterion for more general target probability distributions.

Asymptotic limit results illustrate some similarities and differences with the Metropolis–Hastings (MH) algorithm. Typically, high-dimensional limiting processes of MH algorithms are diffusions [34, 35]. In contrast, the first two summary statistics processes of ZZ converge to non-Markovian Gaussian processes and the first coordinate process of ZZ and the angular momentum process of BPS have pure jump process limits. At the same time, like MH algorithms, our results show that the piecewise deterministic processes can exhibit diffusive behaviour. In particular the latter two summary statistics processes for BPS have diffusion limits. Diffusion limits are known for PDMPs [9, 20], but have to our knowledge have not been established for dimension tending to infinity.

In this paper, we mainly consider the case of a standard normal stationary distribution. Experimental results of Section 3 suggest that the obtained results remain valid for general distributions of product form. For nonproduct strongly correlated distributions such as in [24] the convergence rates could be different. This remains a topic of active research; see also Section 4. Also, throughout this paper, we assume stationarity of the process. The behaviour

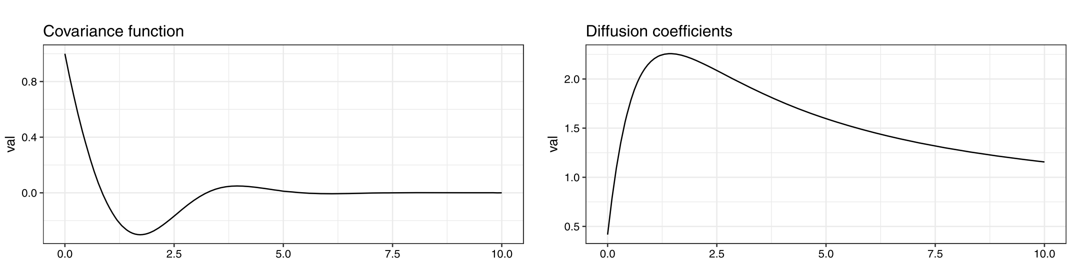


FIG. 1. *Monte Carlo estimated values of the covariance function $K(t,0)$ (left) and the diffusion coefficient $\sigma(\rho)^2$ (right).*

maybe different from the current study if the initial distribution is far from the centre region of the target distribution. This remains a topic of active research.

2. High-dimensional properties. We analyse high-dimension properties of the zig–zag and BPS samplers. Throughout in this paper, we assume strong stationarity of the Markov processes. Our first main objective is the analysis of the angular momentum processes

$$S_t^{\mathrm{Z},d} = \left\langle \xi_t^{\mathrm{Z},d}, \frac{v_t^{\mathrm{Z},d}}{\|v_t^{\mathrm{Z},d}\|} \right\rangle = d^{-1/2}\langle \xi_t^{\mathrm{Z},d}, v_t^{\mathrm{Z},d}\rangle, \tag{2.1}$$

$$S_t^{\mathrm{B},d} = \left\langle \xi_t^{\mathrm{B},d}, \frac{v_t^{\mathrm{B},d}}{\|v_t^{\mathrm{B},d}\|} \right\rangle = \langle \xi_t^{\mathrm{B},d}, v_t^{\mathrm{B},d}\rangle. \tag{2.2}$$

The behaviour of the angular momentum processes illustrates the dissimilarity of the zig–zag and BPS samplers.

The angular momentum processes do not completely capture the asymptotic properties of the Markov processes. For the understanding of long-time properties, it is more natural to consider the behavior of the negative log-target density. Observe that there is an interesting connection between the angular momentum process and the negative log-target density processes:

$$\mathrm{d}\|\xi_t^{\mathrm{Z},d}\|^2 = 2d^{1/2} S_t^{\mathrm{Z},d}\,\mathrm{d}t, \qquad \mathrm{d}\|\xi_t^{\mathrm{B},d}\|^2 = 2S_t^{\mathrm{B},d}\,\mathrm{d}t.$$

Additionally we will study the number of switches (jumps)

$$\sum_{0\le t\le T} 1_{\{\Delta S_t^{\mathrm{Z},d}\neq 0\}}, \qquad \sum_{0\le t\le T} 1_{\{\Delta S_t^{\mathrm{B},d}\neq 0\}}$$

up to $T > 0$, where $\Delta X_t = X_t - X_{t-}$. Finally, we will check the convergence rates for the coordinate processes.

REMARK 2.1 (Proof strategy). In the high-dimensional MCMC literature, as in [34], the Trotter–Kato-type approach is the most popular which uses convergence of generators to prove convergence of Markov processes. Classical literature is [19]. In this paper, we closely follow the semimartingale characteristics approach taken in [23], which is natural to the non-Markovian processes which arise in our analysis. See Section IX.2a of [23] for the connection between the two approaches.

2.1. *Asymptotic limit of the zig–zag sampler.* In this section, we study the asymptotic properties of the zig–zag sampler. All the proofs are postponed to Appendix A. To state the first results, we introduce a stationary piecewise deterministic jump process

$$\mathcal{T}_t = \mathcal{T}_0 + t - 2\int_{(0,t]\times\mathbb{R}_+} \mathcal{T}_{s-} 1_{\{z\le \mathcal{T}_{s-}\}} N(\mathrm{d}s, \mathrm{d}z) \tag{2.3}$$

with $\mathcal{T}_0 \sim \mathcal{N}(0, 1)$. We will show that the process has the same law as that of $(\xi_{i,t}^{\mathrm{Z},d} v_{i,t}^{\mathrm{Z},d})_{t\ge 0}$ for each $i = 1, \dots, d$ where $\xi_{i,t}^{\mathrm{Z},d}$ and $v_{i,t}^{\mathrm{Z},d}$ are ith components of $\xi_t^{\mathrm{Z},d}$ and $v_t^{\mathrm{Z},d}$ respectively. By Itô's formula, the process has the infinitesimal generator

$$Gf(x) = f'(x) + x^+\big(f(-x) - f(x)\big). \tag{2.4}$$

In Section C we show that there exists a unique solution of the martingale problem corresponding to G. By this expression, $\mathcal{N}(0, 1)$ is the invariant distribution of $\mathcal{T} = (\mathcal{T}_t)_{t\ge 0}$ by Proposition 4.9.2 of [19]. In particular, $\mathcal{T}$ is a stationary process. Set

$$K(s, t) = \mathbb{E}[\mathcal{T}_s \mathcal{T}_t]. \tag{2.5}$$

This covariance kernel will play an important role in this work. Some properties are collected in Proposition C.3.

Our first result is the asymptotic limit of the angular momentum process $S^{Z,d} = (S_t^{Z,d})_{t\geq 0}$. We discuss the continuity of the sample path of the limit process.

THEOREM 2.2. *The process $S^{Z,d}$, defined in (2.1), converges to $S^Z = (S_t^Z)_{t\geq 0}$ in distribution in Skorohod topology where S^Z is the non-Markovian stationary Gaussian process with mean 0 and covariance function $K(s,t)$. The Gaussian process is locally α-Hölder continuous for any $\alpha \in (0, 1/2)$, but it is not locally α-Hölder continuous for any $\alpha \geq 1/2$.*

REMARK 2.3 (Hölder continuity). See [25], eq. (2.2.8), or Section A.1 for the definition of local α-Hölder continuity.

The path continuity property explains why the path of S^Z resembles that of Markov diffusion processes since diffusion processes have the same local α-Hölder continuity property. Despite the similarity, the limit process S^Z is a non-Markovian Gaussian process, unlike most of the scaling limit results related to classical MCMC methods.

The second result concerns the number of switches for the zig–zag process, indicating the computational cost of the process. The following results show that the process $S^{Z,d}$, the number of switches per unit time is $O(d)$.

COROLLARY 2.4. *The number of switches of $S^{Z,d}$ over a time interval $(0, T]$ scaled by d^{-1} satisfies*

$$d^{-1} \sum_{0\leq t\leq T} 1_{\{\Delta S_t^{Z,d} \neq 0\}} \longrightarrow \frac{T}{\sqrt{2\pi}} \quad \text{in probability as } d \to \infty$$

for any $T > 0$.

The third result is the analysis of the negative log-target density process. As for the angular momentum process, the limiting process is a non-Markovian Gaussian process. We also discuss the sample path property. We call a process differentiable if there is a modification such that each path is differentiable almost surely. See Section A.3 for the definition.

THEOREM 2.5. *The negative log-density process*

$$Y_t^{Z,d} := \sqrt{d}\left(\frac{\|\xi_t^{Z,d}\|^2}{d} - 1\right)$$

converges to a non-Markovian stationary Gaussian process Y^Z with mean 0 and covariance function

$$L(s,t) = 2 - 2\int_s^t \int_s^t K(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

Moreover the Gaussian process Y^Z is differentiable with respect to the time index t.

Finally we consider the first coordinates of ξ. Let

$$\pi_k(\xi) = \pi_k^d(\xi) = (\xi_1, \ldots, \xi_k) \quad \text{for } \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d,$$

denote the operation of taking the first $k \in \{1, \ldots, d\}$ components of a d-dimensional vector. If $k > d$, then we set

$$\pi_k(\xi) = \pi_k^d(\xi) = (\xi_1, \ldots, \xi_d, \overbrace{0, \ldots, 0}^{k-d}).$$

Let $\phi_k(x)$ be the density of the k-dimensional standard normal distribution $\mathcal{N}(0, I_k)$.

THEOREM 2.6. *For any $k \in \mathbb{N}$ and $d \geq k$, the law of the process $Z_t^{Z,d,k} := \pi_k(\xi_t^{Z,d})$ does not depend on d, and $Z^{Z,d,k}$ is an ergodic process. In particular, for any $\mathcal{N}(0, I_k)$-integrable function $f : \mathbb{R}^k \to \mathbb{R}$, we have*

$$\frac{1}{T}\int_0^T f\left(Z_t^{Z,d,k}\right) \mathrm{d}t \longrightarrow \int_{\mathbb{R}^k} f(x)\phi_k(x)\,\mathrm{d}x \quad \textit{in probability as } T \to \infty.$$

REMARK 2.7 (Joint limiting processes of the angular momentum, the negative log-density and the first coordinate processes). The joint process of the first two processes has a Gaussian limit by the central limit theorem of the processes. The diagonal components of the corresponding covariance kernel are $K(s, t)$ and $L(s, t)$. The off-diagonal component is the covariance of the angular momentum and the log negative-density processes. Since $\mathrm{d}Y_t^{Z,d} = 2S_t^{Z,d}$, off-diagonal component of the corresponding covariance kernel is

$$M(s, t) := \mathbb{E}\left[Y_s^Z S_t^Z\right] = 2^{-1}\frac{\partial L(s, t)}{\partial t} = -2\int_s^t K(t, u)\,\mathrm{d}u.$$

The first coordinate process is asymptotically independent from the other processes.

2.2. *Asymptotic limit of the bouncy particle sampler.* In this section, we study the asymptotic properties of the bouncy particle sampler. All the proofs are postponed to Appendix B. The limiting process of the angular momentum is represented as

$$\begin{aligned} S_t^{\mathrm{B}} = S_0^{\mathrm{B}} + t - 2\int_{(0,t]\times\mathbb{R}_+} S_{s-}^{\mathrm{B}} 1_{\{z \leq S_{s-}^{\mathrm{B}}\}} N(\mathrm{d}s, \mathrm{d}z) \\ + \int_{(0,t]\times\mathbb{R}} \left(z - S_{s-}^{\mathrm{B}}\right) R(\mathrm{d}s, \mathrm{d}z), \end{aligned} \tag{2.6}$$

where R is the random measure with the intensity measure

$$\mathbb{E}\left[R(\mathrm{d}s, \mathrm{d}z)\right] = \rho\,\mathrm{d}s\phi(z)\,\mathrm{d}z,$$

where ϕ denotes the $\mathcal{N}(0, 1)$ density function. By Itô's formula, the process $S^{\mathrm{B}} = (S_t^{\mathrm{B}})_{t\geq 0}$ has the infinitesimal generator

$$Hf(x) = f'(x) + x^+\left(f(-x) - f(x)\right) + \rho\left(\int_{\mathbb{R}} \phi(y) f(y)\,\mathrm{d}y - f(x)\right). \tag{2.7}$$

In Section C we show that there exists a unique solution of the martingale problem corresponding to H.

The process is $\mathcal{N}(0, 1)$-invariant by Proposition 4.9.2 of [19]. Observe that the process S_t^{B} follow the same dynamics as that of $\mathcal{T}_t$ between the refreshment times.

THEOREM 2.8. *The process $S^{\mathrm{B},d}$, defined in (2.2), converges in law to S^{B}.*

In fact, if $\rho = 0$, then the law of $S^{\mathrm{B},d}$ is *identical* to the law of S^{B} for any $d \in \mathbb{N}$. Indeed, say $g(\xi, v) = f(\langle \xi, v\rangle)$ for $(\xi, v) \in E^{\mathrm{B},d}$. Then (without refreshment), for $s = \langle \xi, v\rangle$,

$$\begin{aligned} \left(L^{\mathrm{B},d} g\right)(\xi, v) &= \langle v, \nabla_\xi g(\xi, v)\rangle + \langle \xi, v\rangle^+\left(g(\xi, -v) - g(\xi, v)\right) \\ &= f'\left(\langle \xi, v\rangle\right)\langle v, v\rangle + \langle \xi, v\rangle^+\left(f\left(-\langle \xi, v\rangle\right) - f\left(\langle \xi, v\rangle\right)\right) \\ &= f'(s) + (s)^+\left(f(-s) - f(s)\right), \end{aligned}$$

which establishes that if $\rho = 0$, then $S^{\mathrm{B},d}$ is a Markov process with generator H.

COROLLARY 2.9. *The expected number of switches of $S^{\mathrm{B},d}$ over a time interval $(0,T]$ does not depend on d and is given by*

$$\mathbb{E}\Big[\sum_{0\le t\le T} 1_{\{\Delta S_t^{\mathrm{B},d}\neq 0\}}\Big] = T\Big(\frac{1}{\sqrt{2\pi}}+\rho\Big).$$

Observe that $T\rho$ is the expected number of refreshment jumps, and $T/\sqrt{2\pi}$ is the expected number of bounce jumps. Unlike the zig–zag sampler, the number of switches is random even in the limit $d\to\infty$. This is the reason why we consider expectation rather than the limit in Corollary 2.9. Note that each switch changes all components of the direction v. On the other hand, the zig–zag sampler only changes one component in each switch.

THEOREM 2.10. *The normalised negative log-target density process*

$$Y_t^{\mathrm{B},d} := \sqrt{d}\Big(\frac{\|\xi_{dt}^{\mathrm{B},d}\|^2}{d}-1\Big)$$

converges to the stationary Ornstein–Uhlenbeck process Y^{B} such that

$$\mathrm{d}Y_t^{\mathrm{B}} = -\frac{\sigma(\rho)^2}{4}Y_t^{\mathrm{B}}\,\mathrm{d}t + \sigma(\rho)\,\mathrm{d}W_t,$$

where

$$\sigma(\rho)^2 := 8\int_0^\infty e^{-\rho s}K(s,0)\,\mathrm{d}s$$

with $K(s,0)$ defined in (2.5), and where $(W_t)_{t\ge 0}$ is the one-dimensional standard Wiener process.

The speed of the negative log-target density process is determined by the diffusion coefficient $\sigma(\rho)^2$.

PROPOSITION 2.11. *The continuous function $\sigma(\rho)^2$ satisfies*

$$\lim_{\rho\to+0}\sigma(\rho)^2 = 8\int_0^\infty K(s,0)\,\mathrm{d}s = 0, \qquad \lim_{\rho\to+\infty}\sigma(\rho)^2 = 0.$$

In particular, there exists $\rho^*\in(0,\infty)$ such that $\sigma(\rho^*)^2 = \sup_{\rho\in(0,\infty)}\sigma(\rho)^2$.

The covariance function $K(t,0)$ and the diffusion coefficients $\sigma(\rho)^2$ do not admit simple expressions. These functions can be written as infinite sums of convolutions, and numerical evaluation is difficult. On the other hand, simple Monte Carlo calculations yield good estimates of these functions (Figure 1). The Monte Carlo estimates also provide that the value of ρ maximising $\sigma(\rho)^2$ is around $\rho^\star\approx 1.424$. The ratio of the expected number of the refreshment jumps to that of overall jumps is

$$\frac{\rho^*}{\frac{1}{\sqrt{2\pi}}+\rho^*}\approx 0.7812. \tag{2.8}$$

Note that the choice of ρ is not scale invariant, that is, if we apply the target distribution with the negative log-density $\Psi^d(\xi)=\|\xi\|^2/(2\gamma^2)$, the maximiser depends on $\gamma>0$. However the above jump ratio does not depend on the scale, making the 78.12% rule a possible criterion for the choice of the refreshment rate. In fact, as will be established in Section 2.3, the choice (2.8) maximizes $\sigma^2(\rho)$ for a general class of (non-i.i.d.) distributions.

REMARK 2.12 (Choice of the refreshment rate). In practice, by Corollary 2.9 and Theorem 2.10, a suitable optimization target would be

$$\frac{\text{comp. effort}}{\text{approx. indep. sample}} = \frac{\text{comp. effort}}{\text{time unit}} \times \frac{\text{\# time units}}{\text{indep. sample}} \propto \left(\frac{C_1}{\sqrt{2\pi}} + C_2\rho\right)\frac{1}{\sigma^2(\rho)}.$$

Here C_1 and C_2 represent the computational complexity of a bounce jump and of a refreshment, respectively. These constants depend strongly on implementational aspects, for example the Poisson thinning scheme used for simulating bounces. As a crude simplification, it may be argued that the computational cost of a bounce is significantly more expansive than that of a reflection, that is, $C_1 \gg C_2$, since bounce jumps involve a thinning procedure and the computation of a gradient of the target distribution. With this approximation in mind, it becomes reasonable to maximize $\sigma^2(\rho)$ with respect to ρ in order to obtain high computational efficiency.

Finally, we consider the coordinate process convergence for the bouncy particle sampler.

THEOREM 2.13. *For any $k \in \mathbb{N}$, the process $Z^{\mathrm{B},d,k} = (Z_t^{\mathrm{B},d,k})_{t\geq 0}$ defined by $Z_t^{\mathrm{B},d,k} := \pi_k(\xi_{dt}^{\mathrm{B},d})$ converges to the stationary Ornstein–Uhlenbeck process $Z^{\mathrm{B},k}$ satisfying the SDE*

$$\mathrm{d}Z_t^{\mathrm{B},k} = -\rho^{-1} Z_t^{\mathrm{B},k}\,\mathrm{d}t + \sqrt{2\rho^{-1}}\,\mathrm{d}W_t^k$$

for $k \in \mathbb{N}$ where W^k is the k-dimensional standard Wiener process. In particular, any bounded continuous function $f : \mathbb{R}^k \to \mathbb{R}$, we have

$$\frac{1}{T}\int_0^T f(Z_t^{\mathrm{B},d,k})\,\mathrm{d}t \underset{d,T\to\infty}{\longrightarrow} \int_{\mathbb{R}^k} f(x)\phi_k(x)\,\mathrm{d}x \quad \text{in probability.} \tag{2.9}$$

Note that (2.9) is a double limit. In other words, for all $\varepsilon > 0$ and $\gamma > 0$, there is a $K > 0$ such that for all $d > K$, $T > K$,

$$\mathbb{P}\left(\left|\frac{1}{T}\int_0^T f(Z_t^{\mathrm{B},d,k})\,\mathrm{d}t - \int_{\mathbb{R}^k} f(x)\phi_k(x)\,\mathrm{d}x\right| > \gamma\right) < \varepsilon.$$

This means that the limits with respect to d and T can be freely interchanged. The robustness of the result in terms of the choice of d and T is important for Markov chain Monte Carlo analysis since the practitioner may use $T = d^2$ or $T = d^{10}$, or even $T = \sqrt{d}$.

REMARK 2.14 (Joint convergence of the angular momentum, the negative log-density and the first component processes). Unlike the zig–zag sampler case, the angular momentum does not share its time scaling with the other observables for the bouncy particle sampler. Therefore, the only nontrivial joint process is the combination of the negative log-density and the first component processes, and these two processes are asymptotically independent.

2.3. *Choice of the refreshment ratio for the bouncy particle sampler in the general case.* In Section 2.2 we discussed the choice of the refreshment ratio which maximizes the diffusion coefficient. In this section we will show that it is also possible to estimate the coefficient directly, without resorting to scaling limits. This implies a more general validity of the rule (2.8) for maximizing the diffusive speed.

Let $0 = \sigma_0 < \sigma_1 < \cdots < \sigma_N$ be the refreshment times. The diffusion coefficient can be estimated by

$$\hat{\sigma}_N^2(\rho) := 4\rho N^{-1}\sum_{n=1}^N \left(\Psi^d(\xi_{\sigma_n}^{\mathrm{B},d}) - \Psi^d(\xi_{\sigma_{n-1}}^{\mathrm{B},d})\right)^2 \tag{2.10}$$

that is, an asymptotically unbiased estimator of the diffusion coefficient (see Proposition 2.16). Observe that $\rho^{-1}N$ is asymptotically equivalent to $T=\sigma_N$, and the sum in (2.10) is an estimator of the quadratic variation $\sigma^2(\rho)T$ of the process Y^{B} divided by 4. One could monitor the estimator for different values of ρ in order to select the choice of ρ which maximises $\hat{\sigma}_N^2(\rho)$.

For non-Gaussian, non-i.i.d. case, the relevance of the coefficient is not immediate. However, we may still treat it as a criterion since if the value is large, we expect that the process moves relatively well. So we want to understand the behaviour of the coefficient in situations that are different from the standard Gaussian case.

For the general case, we still assume stationarity of the process, and assume the following. Let Ψ^d be a thrice differentiable function, and assume the Lipschitz-type condition

$$\|\nabla\Psi^d(x)-\nabla\Psi^d(y)\|\le l(\|x-y\|), \tag{2.11}$$

where $l:\mathbb{R}_+\to\mathbb{R}_+$ is a nondecreasing function. We use notation

$$\nabla^2\Psi^d(\xi)[u,v]=\sum_{i=1}^d\frac{\partial^2\Psi^d(\xi)}{\partial\xi_i\,\partial\xi_j}u_iv_j,\qquad \nabla^3\Psi^d(\xi)[u,v,w]=\sum_{i=1}^d\frac{\partial^3\Psi^d(\xi)}{\partial\xi_i\,\partial\xi_j\partial\xi_k}u_iv_jw_k$$

and $\nabla^2\Psi^d(\xi)[u^{\otimes 2}]=\nabla^2\Psi^d(\xi)[u,u]$, and $\nabla^3\Psi^d(\xi)[u^{\otimes 3}]=\nabla^3\Psi^d(\xi)[u,u,u]$. We also assume consistency conditions

$$\mathbb{E}\left[\left|\frac{\|\nabla\Psi^d(\xi_0^{\mathrm{B},d})\|^2}{d}-H\right|\right]\xrightarrow[d\to\infty]{}0, \tag{2.12}$$

$$\mathbb{E}\left[\left|\nabla^2\Psi^d(\xi_0^{\mathrm{B},d})\left[(v_0^{\mathrm{B},d})^{\otimes 2}\right]-H\right|\right]\xrightarrow[d\to\infty]{}0 \tag{2.13}$$

for $H>0$. The following nonexplosive condition is also assumed:

$$\sup_{\xi\in\mathbb{R}^d}\sup_{u\in\mathfrak{S}_{d-1}}\left|\nabla^2\Psi^d(\xi)\left[u^{\otimes 2}\right]\right|<C,\qquad \sup_{\xi\in\mathbb{R}^d}\sup_{u\in\mathfrak{S}_{d-1}}\left|\nabla^3\Psi^d(\xi)\left[u^{\otimes 3}\right]\right|<C \tag{2.14}$$

for some $C>0$, that does not depend on d.

REMARK 2.15. Conditions (2.12) and (2.13) are related to the convergence of the trace of the Fisher information matrix when Ψ^d is the negative log-likelihood function.

In the following proposition, we denote $S_t^{\mathrm{B}}(\rho)$ for the process defined in (2.6) to specify the value of the refreshment rate.

PROPOSITION 2.16. *Under the assumption of* (2.11)–(2.14) *with* $\xi_0^{\mathrm{B},d}\sim\Pi^{\mathrm{B},d}$, *the stochastic process* $S^{\mathrm{B},d}=(S_t^{\mathrm{B},d})_{t\ge 0}$ *defined by* $S_t^{\mathrm{B},d}:=\langle\nabla\Psi^d(\xi_t^{\mathrm{B},d}),v_t^{\mathrm{B},d}\rangle$ *converges to another stochastic process* $(H^{1/2}S^{\mathrm{B},d}_{H^{1/2}t}(H^{-1/2}\rho))_{t\ge 0}$. *In particular,*

$$4\rho\mathbb{E}\left[\left(\Psi^d(\xi_{\sigma_n}^{\mathrm{B},d})-\Psi^d(\xi_{\sigma_{n-1}}^{\mathrm{B},d})\right)^2\right]\xrightarrow[d\to\infty]{}H^{1/2}\sigma^2\left(H^{-1/2}\rho\right).$$

Proof of this proposition is in Appendix D. Since the limit of $S_H^{\mathrm{B},d}$ is the time-scale change of S^{B}, we can still use the 78.12% rule (2.8) for the choice of ρ. More precisely, the expected number of all jumps and that of the refreshment jumps up to time T are

$$TH^{1/2}\left(\frac{1}{\sqrt{2\pi}}+H^{-1/2}\rho\right),\ T\rho$$

with respectively. Therefore the fraction of the number of refreshment jumps is

$$\frac{H^{-1/2}\rho}{\frac{1}{\sqrt{2\pi}} + H^{-1/2}\rho}.$$

On the other hand, $H^{1/2}\sigma^2(H^{-1/2}\rho)$ is maximised when $H^{-1/2}\rho = \rho^*$. Therefore we will have the same ratio 78.12% as before when $H^{1/2}\sigma^2(H^{-1/2}\rho)$ is maximised.

3. Experimental results.

3.1. *Validation of the scaling limits for target that are not standard normal.* In order to investigate the dependence of our results on the distributional assumptions we will carry out computer experiments with respect to four different d-dimensional target distributions:

(i) The standard normal distribution.

(ii) A correlated Gaussian distribution, for which $\mathrm{Var}(\xi_i) = 1$ and $\mathrm{Cov}(\xi_i, \xi_j) = \rho$ (for $i \neq j$) where we take $\rho = 0.9$.

(iii) $(\xi_1, \dots, \xi_d)$ are i.i.d. Student distributed with $\nu = 4$ degrees of freedom.

(iv) $(\xi_1, \dots, \xi_d)$ is a d-dimensional spherically symmetric Student distribution with $\nu = 4$ degrees of freedom (see [7]).

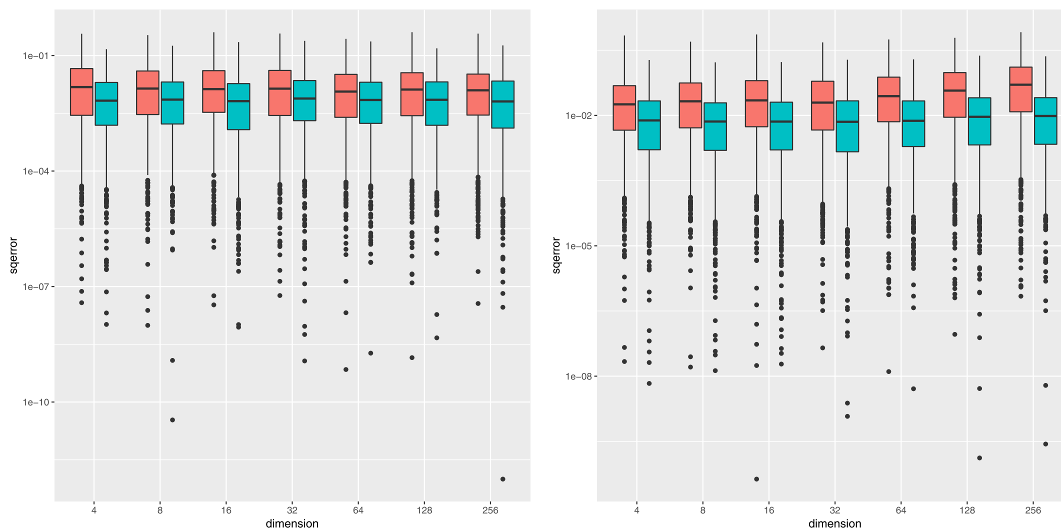
For these four distributions we run both the zig–zag sampler and the bouncy particle sampler with a refresh rate of 1.4. In all cases the zig–zag process with speeds $v^{\mathrm{Z}} \in \{-1, +1\}^d$ is run on a fixed continuous time interval $[0, T]$ where $T = 100$. The bouncy particle sampler with speeds $v^{\mathrm{B}} \in \mathfrak{S}^{d-1} = \{v \in \mathbb{R}^d : \|v\| = 1\}$ is run on a continuous time interval $[0, d \times T]$, which for the purpose of this section is equivalent to a BPS at speed increased by a factor d run on a time interval $[0, T]$. These combinations of velocities and interval length are such that the processes with respect to the observables "first coordinate" and "log density" converge in distribution to their limiting processes as specified in this paper, at least for the standard normal distribution. All processes are started from a random sample from their respective stationary distributions.

In the experiments, for a given trajectory $(\xi(t))_{t\geq 0}$, we define the standardised error with respect to an observable h as

$$E_h = \frac{\frac{1}{T}\int_0^T h(\xi(s))\,\mathrm{d}s - \pi^d(h)}{\sqrt{\mathrm{Var}_{\pi^d}(h)}},$$

where π^d represents the probability distribution with unnormalised negative log density Ψ^d. The "first coordinate" observable corresponds to $h(\xi) = \xi_1$ and the "log density" observable corresponds to $h(\xi) = \|\xi\|^2$. The continuous time integral representing the ergodic average (given the piecewise deterministic trajectory $(\xi(t))_{0\leq t\leq T}$) can be evaluated analytically. In the box plots below the standardised squared error is displayed for increasing dimension, based on 1000 experiments.

As to be expected from the theory developed in this paper the distribution of the standardized squared error for the standard normal distribution (Figure 2) is stable with respect to increase in dimension. BPS seems to be more robust in the presence of correlations (Figure 3), in particular with respect to the first coordinate. In the case of a factorized heavy tailed distribution (Figure 4) we see that the behaviour of both zig–zag and BPS is very robust. Finally in the case of a spherically symmetric example (Figure 5) we see similar behaviour for the different samplers with a nonconstant dependence on dimension.

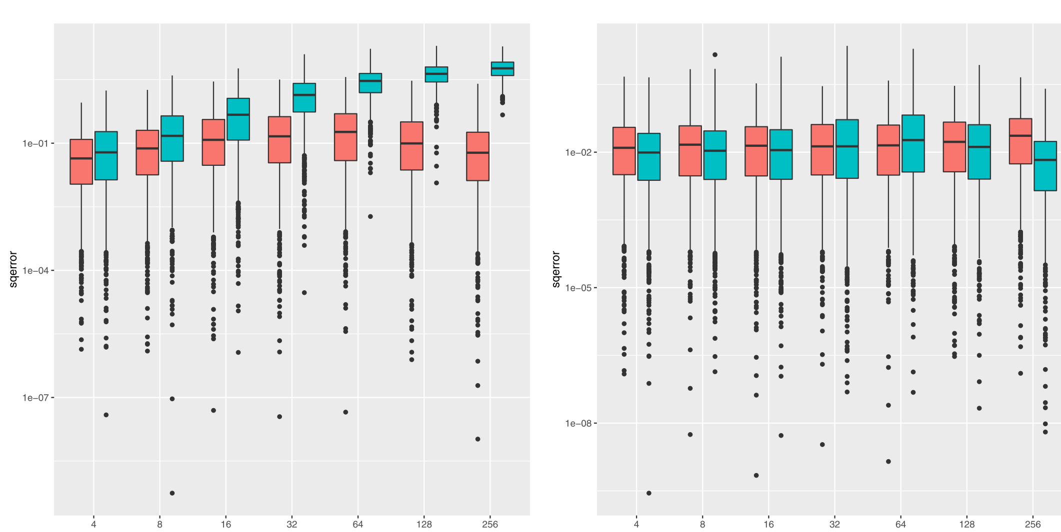


(a) First coordinate (b) Log density

FIG. 2. *Standardized squared errors standard normal distribution. ZZ is cyan, BPS is red.*

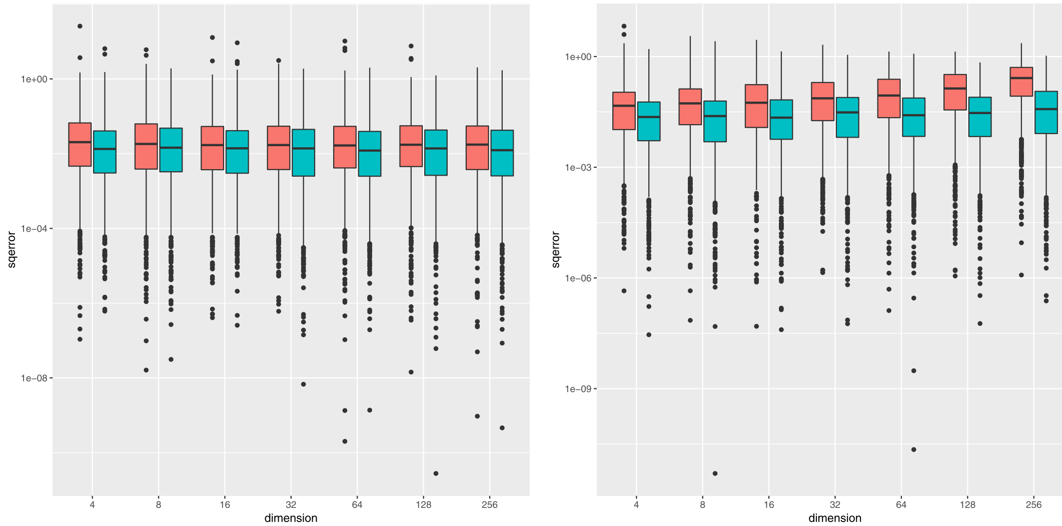
3.2. *Validation of the refreshment rate choice.* In this section we consider the optimality criterion (2.8) in the non-Gaussian, non-i.i.d. setting. Specifically, we consider a simulated logistic regression problem, constructed as follows. We randomly generate a d-dimensional multivariate standard-normal "true" parameter ξ and n covariates $x^{(i)} \in \mathbb{R}^d$, $(i = 1, \dots, n)$ with the first component fixed at one and the other components generated according to a $d-1$-dimensional standard normal distribution. Next independent Bernoulli observations $y^{(i)} \in \{0, 1\}$ are generated according to the logistic probability

$$\mathbb{P}(Y^{(i)} = 1 \mid x^{(i)}, \xi) = \frac{1}{1 + \exp(-\xi^\top x^{(i)})}.$$



(a) First coordinate (b) Log density

FIG. 3. *Standardized squared errors correlated Gaussian distribution*, $\mathrm{Var}(\xi_i) = 1$, $\mathrm{Cov}(\xi_i, \xi_j) = \rho = 0.9$, $i \neq j$. *ZZ is cyan, BPS is red.*



(a) First coordinate (b) Log density

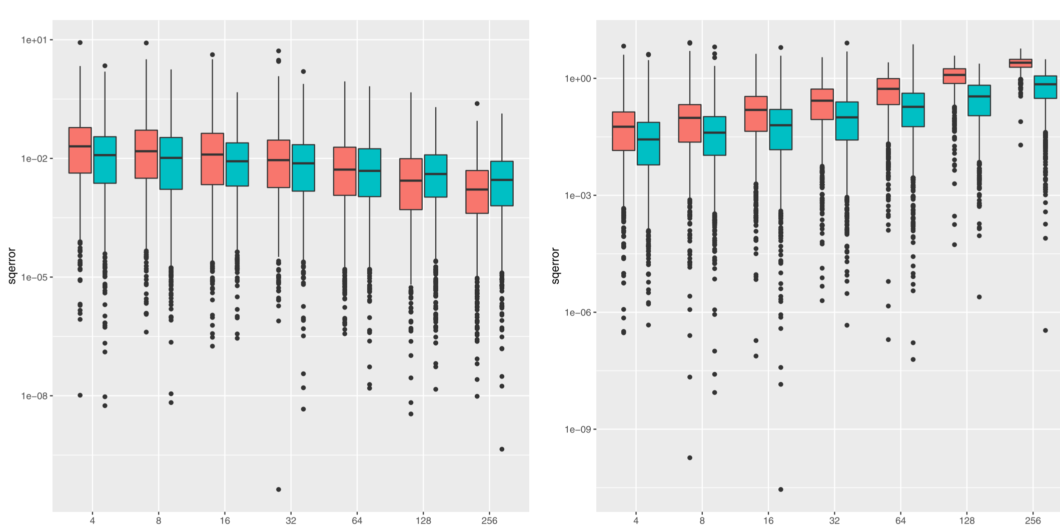
FIG. 4. *Standardized squared errors i.i.d. Student distribution with* $\nu = 4$ *degrees of freedom. ZZ is cyan, BPS is red.*

The same procedure that generates the data $(x^{(i)}, y^{(i)})_{i=1,\dots,n}$ is used to specify the prior and likelihood (conditional on $x^{(i)}$) for ξ, that is, the prior is a standard normal distribution and the likelihood is the product of logistic probabilities, as follows:

$$\pi_0(\xi) \sim \mathcal{N}(0, I_d), \qquad L(\xi \mid x, y) = \prod_{i=1}^{n} \mathbb{P}(Y^{(i)} = y^i \mid x^{(i)}, \xi).$$

This determines the posterior probability distribution

$$\pi(\xi \mid x^{(i)}, y^{(i)}) = L(\xi \mid x, y)\pi_0(\xi).$$



(a) First coordinate (b) Log density

FIG. 5. *Standardized squared errors spherically symmetric Student distribution with* $\nu = 4$ *degrees of freedom. ZZ is cyan, BPS is red.*

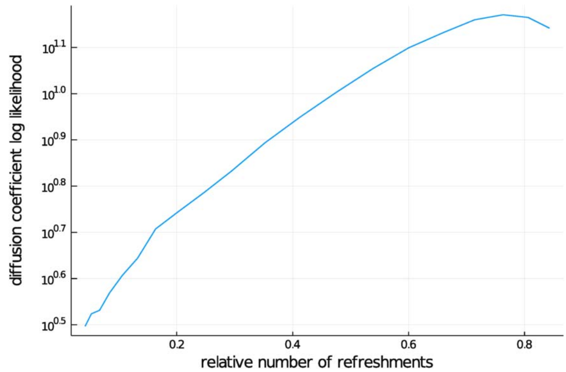


FIG. 6. *Experimental estimation of the average limiting diffusivity of the log density process for a logistic regression posterior distribution as a function of the relative amount of refreshments, as described in Section* 3.2. *We see that the diffusive speed is maximized around a relative number of refreshments of approximately* 0.75, *agreeing closely with* (2.8). *Here we have taken the dimensionality to be* $d = 50$ *and the number of covariates* $n = 1000$. *Experiments have been carried out for a discrete grid of* 20 *different values for* ρ, *ranging from* 0.125 *to* 16, *and for each value of* ρ *the relative amount of refreshments was obtained over* 100 *experiments over a continuous time horizon* $T = 10{,}000$, *along with the average estimated diffusive speed according to* (2.10).

For this target probability distribution, the bouncy particle sampler is run for different values of the refreshment rate ρ. For every experiment, the relative number of refreshments is recorded along with the estimated diffusive speed of the log density process, $-\Psi(\xi) = \log \pi(\xi \mid x^{(i)}, y^{(i)})$. To estimate the diffusive speed of the process we use the estimator (2.10). The experimental result is displayed in Figure 6, and shows that also for a high-dimensional, correlated, non-Gaussian target distribution the limiting diffusive speed of the log density processes is maximized at the refreshment rate ρ^* satisfying (2.8).

4. Discussion. In this paper we considered the high-dimensional asymptotic analysis of ZZ and BPS. The target probability distribution is assumed to be the standard normal distribution. This assumption is indeed restrictive, but the results can be extended to more general target distributions. For the ZZ sampler, it is straight forward to generalise it to a target distribution with a product from $\prod_{i=1}^d F(\mathrm{d}\xi_i)$ where F is a probability measure on $\mathbb{R}$. For the BPS sampler, we have proved convergence of the angular momentum for a general target distribution in Section 2.3.

One of the major computational advantages of PDMP methods is the easy applicability of principled subsampling methods giving substantial computational advantages for instance, in the setting of simulation of Bayesian posteriors with large data sets. However it has also been noted empirically that subsampling can slow down the convergence of PDMP samplers, see, for example, [3]. Therefore it would be very interesting to generalise our work in this paper to consider limits of subsampled PDMPs in order to quantify the effect of subsampling on algorithm performance. It would also be natural to generalise our results to cover the various generalisations and alternatives of BPS and zig–zag such as the coordinate sampler [39] and random velocity zig–zag [38].

Recently, the convergence rates of BPS and ZZ have also been studied by [1] and [15]. In the former article, they studied L^2-exponential convergence rates of Markov semigroups corresponding to the PDMPs under fairly general assumptions. The BPS convergence rate $O(d)$ considered here is in agreement with their results after noticing that they assumed $\sqrt{d}\mathfrak{S}_{d-1}$ as the direction space. On the other hand, the convergence rate obtained by [15] is $O(d^{1/2})$, which is different from ours since this work studied a different scaling limit regime. In Theorem 2.13, we obtained the Ornstein–Uhlenbeck process limit for the first coordinate process. This scaling limit regime does not describe the optimal choice of ρ for a single component since it can be accelerated arbitrarily by taking $\rho \downarrow 0$. The work [15] studied another scaling limit regime by taking $\rho = O(d^{-1/2})$, and proved that the first coordinate process, together with the velocity, converges to a randomised Hamiltonian Monte Carlo process. However, in their regime, the negative log density process will be degenerate by Theorem 2.5. Therefore, if one starts from an initial point $\xi_0^d \in \mathbb{R}^d$ with large log negative density, the process stays in that region in the limit. Therefore, the process does not reach to the smallest log negative density area in this regime. For this reason, we did not use a diminishing refreshment rate. See Section 2.5 of [15] for some empirical comparison between the two regimes. Similarly, our results agree with the very recent contribution [27].

APPENDIX A: THE CONVERGENCE OF THE ZIG–ZAG SAMPLER

A.1. Proof of Theorem 2.2. Let $S^{\mathrm{Z}} = (S_t^{\mathrm{Z}})_{t\ge 0}$ be a Gaussian process with mean 0 and covariance $K(s,t) = \mathbb{E}[\mathcal{T}_s\mathcal{T}_t]$ where $\mathcal{T}$ is defined in (2.3). First, we prove that the Gaussian process S^{Z} is not a Markov process, although $\mathcal{T}$ is a Markov process.

LEMMA A.1. *The stationary Gaussian process S^{Z} is not a Markov process.*

PROOF. By Theorem V.8.1 of [17] together with the continuity of $t \mapsto K(t,0)$, if S^{Z} is a Markov process, then

$$K(t,0) = e^{-ct}$$

for some $c \in \mathbb{R}$. Therefore, the first and the second derivatives of $K(t,0)$ at $t = 0$ are $-c$ and c^2 with respectively. However, this is impossible by derivatives calculated in Proposition C.3. Thus the process S^{Z} is not a Markov process. □

Next we prove convergence of $S^{\mathrm{Z},d}$. We denote the space of continuous and càdlàg functions on $[0,\infty)$ by $\mathbb{C}[0,\infty)$ and $\mathbb{D}[0,\infty)$, respectively. A sequence of $\mathbb{D}[0,\infty)$-valued processes $X^d = (X_t^d)_{t\ge 0}$ is called $\mathbb{C}$-tight if it is tight and any limit point is in $\mathbb{C}[0,\infty)$ with probability 1. By Corollary VI.3.33 of [23], if X^d and Y^d are $\mathbb{C}$-tight, then $(X_t^d + Y_t^d)_{t\ge 0}$ is $\mathbb{C}$-tight. On the other hand, the sum of tight sequence of processes is not tight in general.

LEMMA A.2. *The process $S^{\mathrm{Z},d}$ converges to S^{Z}.*

PROOF. Observe that $S_t^{\mathrm{Z},d} = d^{-1/2}\sum_{i=1}^d \mathcal{T}_{i,t}^d$ where

$$\mathcal{T}_{i,t}^d = \xi_{i,t}^{\mathrm{Z},d} v_{i,t}^{\mathrm{Z},d}. \tag{A.1}$$

By construction, $(\mathcal{T}_{i,t}^d)_{t\ge 0}$ $(i = 1,\dots,d)$ are independent processes and have the same law as that of $\mathcal{T}$. By using the fact, we prove tightness of the sequence of processes $(S_t^{\mathrm{Z},d})_{t\in(0,T]}$ for each $T > 0$ by the central limit theorem of stochastic processes. By (C.1), we have

$$\sup_{0\le t,u\le T} |\mathcal{T}_u - \mathcal{T}_t| \le 2 \sup_{0\le t\le T} |\mathcal{T}_t| \le 2|\mathcal{T}_0| + 2T.$$

Observe that any moments of the right-hand side of the above inequality exist since $\mathcal{T}_0 \sim \mathcal{N}(0,1)$. By using this bound, for the Poisson random measure $N(\mathrm{d}t,\mathrm{d}z)$, we have

$$\sum_{t<v\le u} 1_{\{\Delta\mathcal{T}_v \ne 0\}} = \int_{(t,u]\times\mathbb{R}_+} 1_{\{z\le\mathcal{T}_{s-}\}} N(\mathrm{d}s,\,\mathrm{d}z) \le N(A),$$

where $A = (t, u] \times (0, |\mathcal{T}_0| + T]$ since $|\mathcal{T}_t| \le |\mathcal{T}_0| + t$ by (C.1). Let $\lambda = |u - t|(|\mathcal{T}_0| + T)$. Then we have

$$\mathbb{P}(N(A) \ge 1) = 1 - e^{-\lambda} \le \lambda, \qquad \mathbb{P}(N(A) \ge 2) = 1 - e^{-\lambda} - \lambda e^{-\lambda} \le \frac{\lambda^2}{2}.$$

Since if there is no jump, $\mathcal{T}_t$ has the deterministic move, and we have

$$\sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 0 \quad \Longrightarrow \quad \mathcal{T}_u - \mathcal{T}_s = u - s.$$

Hence if $t \le u \le T$, we have

$$\begin{aligned}
\mathbb{E}[(\mathcal{T}_u - \mathcal{T}_t)^2] &= \mathbb{E}\Big[(\mathcal{T}_u - \mathcal{T}_t)^2, \sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 0\Big] \\
&\quad + \mathbb{E}\Big[(\mathcal{T}_u - \mathcal{T}_t)^2, \sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} \ge 1\Big] \\
&\le |u-t|^2 + \mathbb{E}\big[(2|\mathcal{T}_0| + 2T)^2 \times \lambda\big] \\
&\le C|u-t|
\end{aligned}$$

for some $C = C_T > 0$. On the other hand, if $s \le t \le u \le T$

$$\sum_{s<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 1 \quad \Longrightarrow \quad \sum_{s<v\le t} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 0 \quad \text{or} \quad \sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 0$$

and hence

$$\sum_{s<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 1 \quad \Longrightarrow \quad (\mathcal{T}_u - \mathcal{T}_t)^2(\mathcal{T}_t - \mathcal{T}_s)^2 \le |u-s|^2\big(2|\mathcal{T}_0| + 2T\big)^2.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[(\mathcal{T}_u - \mathcal{T}_t)^2(\mathcal{T}_t - \mathcal{T}_s)^2] &= \mathbb{E}\Big[(\mathcal{T}_u - \mathcal{T}_t)^2(\mathcal{T}_t - \mathcal{T}_s)^2, \sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 0\Big] \\
&\quad + \mathbb{E}\Big[(\mathcal{T}_u - \mathcal{T}_t)^2(\mathcal{T}_t - \mathcal{T}_s)^2, \sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} = 1\Big] \\
&\quad + \mathbb{E}\Big[(\mathcal{T}_u - \mathcal{T}_t)^2(\mathcal{T}_t - \mathcal{T}_s)^2, \sum_{t<v\le u} 1_{\{\Delta \mathcal{T}_v \neq 0\}} \ge 2\Big] \\
&\le |u-s|^4 + |u-s|^2 \mathbb{E}\big[(2|\mathcal{T}_0| + 2T)^2\big] \\
&\quad + \mathbb{E}\Big[(2|\mathcal{T}_0| + 2T)^4 \times \frac{\lambda^2}{2}\Big] \\
&\le C|u-s|^2
\end{aligned}$$

for some $C = C_T > 0$. These inequalities imply the conditions (i, ii) in Theorem 2 of [21]. Therefore, by Theorem 2 of [21], we have central limit theorems for the sum of the copies of $(\mathcal{T}_t)_{t\in(0,T]}$. In particular, $(S_t^{\mathrm{Z},d})_{t\in(0,T]}$ is tight.

On the other hand, for any $0 \le t_1 < t_2 < \cdots < t_k$, any k-dimensional random variable

$$\big(S_{t_1}^{\mathrm{Z},d}, \dots, S_{t_k}^{\mathrm{Z},d}\big) = d^{-1/2} \sum_{i=1}^{d} \big(\mathcal{T}_{i,t_1}^d, \dots, \mathcal{T}_{i,t_k}^d\big)$$

converges to a normal distribution by the finite-dimensional central limit theorem since the random variables $(\mathcal{T}_{i,t_1}^d, \dots, \mathcal{T}_{i,t_k}^d)$ $(i = 1, \dots, d)$ are independent and have the same law as that of $(\mathcal{T}_{t_1}, \dots, \mathcal{T}_{t_k})$. Hence $(S_t^{Z,d})_{t\in(0,T]}$ converges to $(S_t^{Z})_{t\in(0,T]}$ by Lemma VI.3.19 of [23]. Then the convergence of $(S_t^{Z,d})_{t\geq 0}$ to $(S_t^{Z})_{t\geq 0}$ also follows from Theorem 16.7 of [6]. □

We call a $\mathbb{D}[0,\infty)$-valued processes $X = (X_t)_{t\geq 0}$ *locally α-Hölder continuous* if there is a $\mathbb{C}[0,\infty)$-valued process $\tilde{X}$ with the same law as that of X such that there exists $\delta_T > 0$, $h_T(\omega) > 0$ and

$$\mathbb{P}\left(\omega \in \Omega : \sup_{|u-v|\leq h_T(\omega), 0\leq u,v\leq T} \frac{|\tilde{X}_u(\omega) - \tilde{X}_v(\omega)|}{|u-v|^\alpha} \leq \delta_T\right) = 1$$

for any $T > 0$.

LEMMA A.3. *S^Z is locally α-Hölder continuous for $\alpha \in (0, 1/2)$ but it is not locally α-Hölder continuous for any $\alpha \geq 1/2$.*

PROOF. The mean zero Gaussian process S^Z satisfies $S_t^Z \sim \mathcal{N}(0,1)$ and $S_t^Z - S_0^Z \sim \mathcal{N}(0, \sigma(t)^2)$ where

$$\sigma(t)^2 := \mathbb{E}[(S_t^Z - S_0^Z)^2] = \mathbb{E}[(S_t^Z)^2] + \mathbb{E}[(S_0^Z)^2] - 2\mathbb{E}[S_t^Z S_0^Z] = 2 - 2K(t,0).$$

Observe that $\sigma^2(0) = 0$. By Proposition C.3 we have

$$\begin{aligned} \frac{\sigma(t)^2}{t} &= \frac{\sigma(t)^2 - \sigma^2(0)}{t} \\ &= -2\frac{K(t,0) - K(0,0)}{t} \underset{t\to 0}{\longrightarrow} -2\partial_t K(t,0)|_{t=0} = 8\phi(0), \end{aligned}$$

and in particular, for sufficiently small $h > 0$, we have a local bound $ct \leq |\sigma(t)^2| \leq Ct$ $(0 \leq t \leq h)$ for some $c, C > 0$. On the other hand, since we have $\sigma(t)^2 \leq 2\mathbb{E}[(S_t^Z)^2 + (S_0^Z)^2] = 4$, there is a global bound $|\sigma(t)^2| \leq Ct$ $(t \geq 0)$ for some constant $C > 0$. Therefore, the $(2n)$th moment of $S_t^Z - S_0^Z \sim \mathcal{N}(0, \sigma(t)^2)$ is

$$\mathbb{E}[|S_t^Z - S_0^Z|^{2n}] = (2n-1)!!|\sigma(t)^2|^n \leq C|t|^n$$

for some $C > 0$ for any $n \in \mathbb{N}$. Thus, local α-Hölder continuity for any $\alpha \in (0, 1/2)$ follows from Kolmogorov–Čentsov's theorem (Theorem 2.2.8 of [25]).

On the other hand, by Proposition C.3, the second derivative of $K(t,0)$ around $t = 0$ is positive and hence $\sigma(t)^2$ is concave around $t = 0$. Therefore, by Slepian's lemma (Theorem 7.2.10 of [28]), we have

$$\lim_{t\to 0} \sup_{|u-v|\leq t, 0\leq u,v\leq 1} \frac{|S_u^Z(\omega) - S_v^Z(\omega)|}{\sqrt{2\sigma^2(u-v)\log(1/|u-v|)}} \geq 1 \tag{A.2}$$

almost surely. If S^Z is locally 1/2-Hölder continuous, then there exists a process $\tilde{S}^Z$, with the same law as S^Z, such that for $t \geq 0$, and for some $\delta > 0$,

$$\frac{|\tilde{S}_{t+h}^Z(\omega) - \tilde{S}_t^Z(\omega)|}{\sqrt{2\sigma^2(h)\log(1/|h|)}} \leq \delta \frac{|h|^{1/2}}{\sqrt{2c|h|\log(1/|h|)}}$$

for sufficiently small h. The right hand side converges to 0 which contradicts (A.2). Thus $\tilde{S}^Z$ and S^Z cannot be locally 1/2-Hölder continuous with probability 1. □

PROOF OF THEOREM 2.2. The claim follows by Lemmas A.1–A.3. □

A.2. Proof of Corollary 2.4. PROOF OF COROLLARY 2.4. The convergence of the switching rate comes from the law of large numbers. Observe that

$$d^{-1}\sum_{0\le t\le T} 1_{\{\Delta S_t^{Z,d}\neq 0\}} = d^{-1}\sum_{i=1}^{d}\sum_{0\le t\le T} 1_{\{\Delta \mathcal{T}_{i,t}^{d}\neq 0\}},$$

where $(\mathcal{T}_{i,t}^d)_{t\ge 0}$ $(i=1,\dots,d)$ are independent copies of (2.3). See the proof of Lemma A.2. Therefore, by the law of large numbers, we have

$$\begin{aligned} d^{-1}\sum_{0\le t\le T} 1_{\{\Delta S_t^{d}\neq 0\}} \underset{d\to\infty}{\longrightarrow} \mathbb{E}\Big[\sum_{0\le t\le T} 1_{\{\Delta \mathcal{T}_t\neq 0\}}\Big] &= \mathbb{E}\Big[\int_{(0,T]\times\mathbb{R}_+} 1_{\{z\le \mathcal{T}_t\}}\,\mathrm{d}z\,\mathrm{d}t\Big] \\ &= \int_0^T \mathbb{E}[\mathcal{T}_t^+]\,\mathrm{d}t = \frac{T}{\sqrt{2\pi}} \end{aligned}$$

by $\mathcal{T}_t\sim\mathcal{N}(0,1)$. □

A.3. Proof of Theorem 2.5. We call a $\mathbb{D}[0,\infty)$-valued processes X differentiable with respect to the time index t if there is a $\mathbb{C}[0,\infty)$-valued process $\tilde{X}$ with the same law as that of X and another $\mathbb{C}[0,\infty)$-valued process $(\partial\tilde{X}_t(\omega))_{t\ge 0}$ on the same probability space as that of $\tilde{X}$ such that

$$\mathbb{P}\Big(\omega\in\Omega: \lim_{h\to 0}\frac{\tilde{X}_{t+h}(\omega)-\tilde{X}_t(\omega)}{h} = \partial\tilde{X}_t(\omega), \forall t\in(0,T)\Big) = 1$$

for any $T>0$.

PROOF OF THEOREM 2.5. The map $(\alpha_t)_{t\ge 0}\mapsto(\int_0^t \alpha_s\,\mathrm{d}s)_{t\ge 0}$ from $\mathbb{D}[0,\infty)$ to $\mathbb{C}[0,\infty)$ is continuous. Also, by Theorem 2.2, the sequence $S^{Z,d}$ converges in law to S^{Z}. Therefore, the sequence of processes $(Y_t^{Z,d}-Y_0^{Z,d})_{t\ge 0}$ $(d\in\mathbb{N})$ is $\mathbb{C}$-tight since

$$\begin{aligned} Y_t^{Z,d}-Y_0^{Z,d} &= d^{-1/2}(\|\xi_t^{Z,d}\|^2-\|\xi_0^{Z,d}\|^2) \\ &= 2\int_0^t S_u^{Z,d}\,\mathrm{d}u \underset{d\to\infty}{\longrightarrow} 2\int_0^t S_u^{Z}\,\mathrm{d}u \end{aligned} \tag{A.3}$$

in distribution in Skorohod topology. Also, $\xi_0^{Z,d}\sim\mathcal{N}_d(0,I_d)$ and we have

$$Y_0^{Z,d} = \sqrt{d}\Big(\frac{\|\xi_0^{Z,d}\|^2}{d}-1\Big) \underset{d\to\infty}{\Longrightarrow} \mathcal{N}(0,2).$$

Thus $(Y_t^{Z,d})_{t\ge 0} = ((Y_t^{Z,d}-Y_0^{Z,d})+Y_0^{Z,d})_{t\ge 0}$ is $\mathbb{C}$-tight. On the other hand, by the finite-dimensional central limit theorem, $(Y_{t_1}^{Z,d}, Y_{t_2}^{Z,d},\dots,Y_{t_k}^{Z,d})$ converges in distribution to some normal distribution for any $k\in\mathbb{N}$ and any $t_1<\cdots<t_k$, since

$$\begin{aligned} (Y_{t_1}^{Z,d}, Y_{t_2}^{Z,d},\dots,Y_{t_k}^{Z,d}) &= \sqrt{d}\Big(\frac{\|\xi_{t_1}^{Z,d}\|^2}{d}-1,\dots,\frac{\|\xi_{t_k}^{Z,d}\|^2}{d}-1\Big) \\ &= \sqrt{d}^{-1}\sum_{i=1}^{d}(\|\xi_{i,t_1}^{Z,d}\|^2-1,\dots,\|\xi_{i,t_k}^{Z,d}\|^2-1) \\ &=: \sqrt{d}^{-1}\sum_{i=1}^{d} U_i^d \end{aligned}$$

and U_i^d $(i=1,\dots,d, d\in\mathbb{N})$ are mean 0 and independent and identically distributed since every component of $\xi^{Z,d}$ is an independent zig–zag process due to the decoupling of the switching rate. Thus by Lemma VI.3.19 of [23], $Y^{Z,d}$ converges to a Gaussian process, which will be denoted by Y^{Z} with a covariance function denoted by $L(s,t)$. Since the covariance function of $Y^{\mathrm{Z},d}$ and Y^{Z} are the same, and $\mathrm{d}Y_t^{\mathrm{Z},d} = 2S_t^{\mathrm{Z},d}\,\mathrm{d}t$, we have

$$\begin{aligned}
L(s,t) &= \mathbb{E}[Y_s^{\mathrm{Z},d} Y_t^{\mathrm{Z},d}] \\
&= \frac{1}{2}(\mathbb{E}[(Y_s^{\mathrm{Z},d})^2] + \mathbb{E}[(Y_t^{\mathrm{Z},d})^2] - \mathbb{E}[(Y_s^{\mathrm{Z},d} - Y_t^{\mathrm{Z},d})^2]) \\
&= \frac{1}{2}\left(4 - 4\mathbb{E}\left[\left\{\int_s^t S_u^{\mathrm{Z},d}\,\mathrm{d}u\right\}^2\right]\right) \\
&= 2 - 2\int_s^t \int_s^t \mathbb{E}[S_u^{\mathrm{Z},d} S_v^{\mathrm{Z},d}]\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

Furthermore, since the covariance function of $S^{\mathrm{Z},d}$ and $\mathcal{T}$ are the same, we have

$$L(s,t) = 2 - 2\int_s^t \int_s^t \mathbb{E}[\mathcal{T}_u \mathcal{T}_v]\,\mathrm{d}u\,\mathrm{d}v = 2 - 2\int_s^t \int_s^t K(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

From this expression, we can conclude that the limiting process Y^{Z} is non-Markovian as in Lemma A.1. Because if it is a Gaussian process, the second derivative of the convariance function $L(t,0)$ at $t=0$ should be negative that is impossible by the expression of $L(s,t)$.

Finally, since $(Y_t^{\mathrm{Z}} - Y_0^{\mathrm{Z}})_{t\geq 0}$ and $(2\int_0^t S_u^{\mathrm{Z}}\,\mathrm{d}u)_{t\geq 0}$ have the same law by (A.3) and the latter process is differentiable, the process Y^{Z} has a differentiable version. □

A.4. Proof of Theorem 2.6. PROOF OF THEOREM 2.6. Let $(\xi_t^{\mathrm{Z}})_{t\geq 0}$ be the process such that $\xi_0^{\mathrm{Z}} \sim \mathcal{N}(0,1)$ and v_0^{Z} are independent and $\mathbb{P}(v_0^{\mathrm{Z}} = +1) = \mathbb{P}(v_0^{\mathrm{Z}} = -1) = 1/2$ and

$$\xi_t^{\mathrm{Z}} = \xi_0^{\mathrm{Z}} + \int_0^t v_s^{\mathrm{Z}}\,\mathrm{d}s \quad (t \geq 0),$$

and

$$v_t^{\mathrm{Z}} = v_0^{\mathrm{Z}} - 2\int_{(0,t]\times\mathbb{R}_+} v_{s-}^{\mathrm{Z}} 1_{\{z \leq \xi_{s-}^{\mathrm{Z}}\}} N(\mathrm{d}s, \mathrm{d}z) \quad (t \geq 0),$$

where $N(\mathrm{d}t, \mathrm{d}x)$ is the homogeneous Poisson measure with the intensity measure $\mathrm{d}t\,\mathrm{d}x$. The process $(\xi_t^{\mathrm{Z}})_{t\geq 0}$ was studied extensively by [2]. In particular, it is ergodic by Proposition 2.2 of [2]. Therefore, for $k \in \mathbb{N}$, if $(\xi_{i,t}^{\mathrm{Z}})_{t\geq 0}$ $(i = 1, \ldots, k)$ are independent copies of $(\xi_t^{\mathrm{Z}})_{t\geq 0}$, we have

$$\frac{1}{T}\int_0^T f(\xi_{1,t}^{\mathrm{Z}}, \ldots, \xi_{k,t}^{\mathrm{Z}})\,\mathrm{d}t \underset{T\to\infty}{\longrightarrow} \int_{\mathbb{R}^k} f(x)\phi_k(x)\,\mathrm{d}x \tag{A.4}$$

almost surely, where $f : \mathbb{R}^k \to \mathbb{R}$ is a $N_k(0, I_k)$-integrable function.

On the other hand, the processes $(\xi_{k,t}^{\mathrm{Z},d})_{t\geq 0}$ $(k \in \{1, \ldots, d\}, d \in \mathbb{N})$ are independent and identically distributed with the same law as that of $(\xi_t^{\mathrm{Z}})_{t\geq 0}$. Since

$$\frac{1}{T}\int_0^T f(\pi_k(\xi_t^{\mathrm{Z},d}))\,\mathrm{d}t = \frac{1}{T}\int_0^T f(\xi_{1,t}^{\mathrm{Z},d}, \ldots, \xi_{k,t}^{\mathrm{Z},d})\,\mathrm{d}t$$

has the same law as that of the left-hand side of (A.4), the claim follows. □

APPENDIX B: THE CONVERGENCE OF THE BOUNCY PARTICLE SAMPLER

B.1. Some preliminary results.

B.1.1. *Some remarks on semimartingale characteristics and majoration hypothesis.* As commented at the end of Section C, we use the martingale problem approach to show scaling limit results instead of the classical Trotter–Kato approach. For this approach, we need some knowledge on semimartingale theory. A nice introduction to semimartingale theory can be found in Chapters I and II of [23]. Our notation will generally follow this reference. A semimartingale $X = (X_t)_{t\geq 0}$, is called locally square-integrable if it has the canonical decomposition

$$X_t = X_0 + M_t + B'_t, \quad t \geq 0,$$

such that $M = (M_t)_{t\geq 0}$ is locally square-integrable local martingale, and $B' = (B'_t)_{t\geq 0}$ is predictable process with finite variation (see Definition II.2.27). We consider the convergence of a sequence of semimartingales. We prove the convergence by using the so-called characteristics (B', C, ν) and the modified second characteristic $\tilde{C}'$. We briefly explain these characteristics for locally square-integrable semimartingale. Note that as in Section IX.3b.2, for a locally square-integrable semimartingale, we can treat the characteristics without truncation function $h(x)$ in Definition II.2.16.

The first characteristic B' was already introduced as above. We denote μ^X for the random measure associated to the jumps of X, that is,

$$\mu^X(\omega;\, \mathrm{d}t, \mathrm{d}x) = \sum_{s>0} 1_{\{\Delta X_s(\omega)\neq 0\}}\delta_{(s,\Delta X_s(\omega))}(\mathrm{d}t, \mathrm{d}x).$$

The third characteristic $\nu(\omega;\, \mathrm{d}t, \mathrm{d}x)$ is the intensity measure of the random measure μ^X, and $\tilde{C}' = (\tilde{C}'_t)_{t\geq 0}$ is the predictable quadratic variation of M. The second characteristic C is the predictable quadratic variation of the continuous part of X, but in this section, $C \equiv 0$ since the processes S^{B} and $S^{\mathrm{B},d}$ do not have continuous martingale parts.

For example, the Markov process S^{B} defined in (2.6) has the following decomposition where we use the single random measure form implicitly (see Remark B.1 below). By the definition for the stochastic integral with respect to random measures (Section II.1d), the square integrable martingale part is

$$M_t = M_t(S^{\mathrm{B}}) = -2\int_{(0,t]\times\mathbb{R}_+} S^{\mathrm{B}}_{s-} 1_{\{z\leq S^{\mathrm{B}}_{s-}\}}\{N(\mathrm{d}s, \mathrm{d}z) - \mathrm{d}s\,\mathrm{d}z\}$$
$$+ \int_{(0,t]\times\mathbb{R}} (z - S^{\mathrm{B}}_{s-})\{R(\mathrm{d}s, \mathrm{d}z) - \rho\, \mathrm{d}s\phi(z)\,\mathrm{d}z\}.$$

The predictable process part is

$$B'_t = B'_t(S^{\mathrm{B}}) = t - 2\int_0^t \{(S^{\mathrm{B}}_s)^+\}^2\,\mathrm{d}s - \rho\int_0^t S^{\mathrm{B}}_s\,\mathrm{d}s, \tag{B.1}$$

which is the sum of the deterministic part t and the intensity measure of the random measure part. By Theorem II.1.33, the predictable quadratic variation of M is

$$\tilde{C}'_t(S^{\mathrm{B}}) := 4\int_0^t \{(S^{\mathrm{B}}_s)^+\}^3\,\mathrm{d}s + \rho\int_0^t (1 + (S^{\mathrm{B}}_s)^2)\,\mathrm{d}s.$$

The random measure $\mu = \mu^{S^{\mathrm{B}}}$ is defined by the integral form

$$g * \mu_t := \int g(x)\mu_t(\mathrm{d}x)$$

$$
\begin{aligned}
&:= \int_{(0,t]\times\mathbb{R}_+} g(-2S^{\mathrm{B}}_{s-})1_{\{z\le S^{\mathrm{B}}_{s-}\}}N(\mathrm{d}s,\mathrm{d}z)\\
&\quad + \int_{(0,t]\times\mathbb{R}} \big(g(z-S^{\mathrm{B}}_{s-})\big)R(\mathrm{d}s,\mathrm{d}z),
\end{aligned}
$$

where $g:\mathbb{R}\to[0,\infty)$ is a continuous bounded function. The random measure $\nu(\omega;\,\mathrm{d}t,\mathrm{d}x)$ is its compensator which is defined by

$$
\begin{aligned}
g*\nu_t &:= \int g(x)\nu_t(\mathrm{d}x)\\
&:= \int_0^t g(-2S^{\mathrm{B}}_s)(S^{\mathrm{B}}_s)^+\,\mathrm{d}s + \rho\int_0^t\int_{\mathbb{R}}\big(g(z-S^{\mathrm{B}}_s)\big)\,\mathrm{d}s\phi(z)\,\mathrm{d}z.
\end{aligned}
\tag{B.2}
$$

By this decomposition S^{B} is also a homogeneous jump process in the sense of Section III.2c, where $b(x)=1-2(x^+)^2-\rho x$, $c(x)\equiv 0$ and $K(x,\mathrm{d}y)=(x^+)\delta_{\{-2x\}}(\mathrm{d}y)+\rho\phi(y-x)\,\mathrm{d}y$.

On the other hand, the process $S^{\mathrm{B},d}$ is not a Markov process, and has the expression

$$
\begin{aligned}
S^{\mathrm{B},d}_t &= S^{\mathrm{B},d}_0 + t - 2\int_{(0,t]\times\mathbb{R}_+} S^{\mathrm{B},d}_{s-}1_{\{z\le S^{\mathrm{B},d}_{s-}\}}N(\mathrm{d}s,\mathrm{d}z)\\
&\quad + \int_{(0,t]\times\mathfrak{S}^{d-1}}\big(\langle\xi^{\mathrm{B},d}_s,u\rangle - S^{\mathrm{B},d}_{s-}\big)R_d(\mathrm{d}s,\mathrm{d}u),
\end{aligned}
\tag{B.3}
$$

by Itô's formula. We denote (B'^d, C^d, ν^d) and $\tilde{C}'^d$ for the characteristics and modified second characteristic of $S^{\mathrm{B},d}$. As in the above example, we have

$$
\begin{aligned}
B'^d_t &:= t - 2\int_0^t \{(S^{\mathrm{B},d}_s)^+\}^2\,\mathrm{d}s - \rho\int_0^t S^{\mathrm{B},d}_s\,\mathrm{d}s,\\
\tilde{C}'^d_t &:= 4\int_0^t \{(S^{\mathrm{B},d}_s)^+\}^3\,\mathrm{d}s + \rho\int_0^t\left(\frac{\|\xi^{\mathrm{B},d}_s\|^2}{d} + (S^{\mathrm{B},d}_s)^2\right)\mathrm{d}s,
\end{aligned}
$$

and

$$
\begin{aligned}
g*\nu^d_t := \int g(x)\nu^d_t(\mathrm{d}x) &:= \int_0^t g(-2S^{\mathrm{B},d}_s)(S^{\mathrm{B},d}_s)^+\,\mathrm{d}s\\
&\quad + \rho\int_0^t\int_{\mathfrak{S}^{d-1}} g\big(\langle\xi^{\mathrm{B},d}_s,u\rangle - S^{\mathrm{B},d}_s\big)\,\mathrm{d}s\psi_d(\mathrm{d}u)
\end{aligned}
$$

for a continuous bounded function g.

Finally, we introduce strong majorisation property which is important to prove tightness of the sequence of processes. For two increasing processes $X=(X_t)_{t\ge0}$, $Y=(Y_t)_{t\ge0}$, X *strongly majorises* Y if $X-Y=(X_t-Y_t)_{t\ge0}$ is an increasing process, that is, almost all paths of $X_t(\omega)-Y_t(\omega)$ are increasing; see [23], Definition VI.3.34. We denote $Y\prec X$ if X strongly majorises Y.

REMARK B.1. Piecewise deterministic Markov processes in this paper are naturally described by stochastic integrals with respect to several independent random measures. However, it can also be possible to express these integrals by using single random measures. At the same time, we can always recover the separate random measure expression from a single random measure representation. We do not use the single random measure representation explicitly in this paper, but we implicitly use the form when we apply theorems in literature that use the single random measure form.

B.1.2. *Some remark on spherically symmetric distribution.* Some of the characteristics of semimartingales $S^{\mathrm{B},d}$ and $Y^{\mathrm{B},d}$ are written by the expectation of U^d which will be defined in (B.4), and U^d will be approximated by a Gaussian random variable. We will quantify this approximation error by the result in [16].

As mentioned above, we need to show that

$$U^d := d^{1/2}\langle e, v\rangle, \tag{B.4}$$

where $v \sim \psi_d$ and e is a unit vector, converges to the standard normal distribution and we need to quantify the approximation error. The distribution is extensively studied by [16]. For example, since $|\langle e, v\rangle|^2$ follows the Beta distribution with parameters $1/2$ and $(d-1)/2$, we have

$$\mathbb{E}\big[|U^d|^\alpha\big] = \frac{d^{\alpha/2} B(\frac{\alpha+1}{2}, \frac{d-1}{2})}{B(\frac{1}{2}, \frac{d-1}{2})} \underset{d\to\infty}{\longrightarrow} \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\frac{1}{2})} 2^{\alpha/2} \tag{B.5}$$

for $\alpha > -1$, where we used Stirling's approximation. Moreover,

$$\|\mathcal{L}(U^d) - \mathcal{N}(0,1)\|_{\mathrm{TV}} = O(1/d) \tag{B.6}$$

for $\|\nu\|_{\mathrm{TV}} = \sup |\int h(x)\nu(\mathrm{d}x)|$ where the supremum is evaluated over those measurable function $h(x)$ bounded above by 1. Since the expectations in the semimartingale characteristics are not bounded functions, we need the following proposition to quantify the approximation error.

PROPOSITION B.2. *For any* $\epsilon > 0, k \in \mathbb{N}$ *and* $W \sim \mathcal{N}(0,1)$,

$$\sup_{|h(x)| \le (1+|x|)^k} \big|\mathbb{E}\big[h(U^d)\big] - \mathbb{E}\big[h(W)\big]\big| = O\big(d^{\epsilon-1}\big).$$

PROOF. Without loss of generality, we can assume $\epsilon \in (0, 1/2)$. Let $|h(x)| \le (1+|x|)^k$. To apply (B.6), we consider a bounded modification

$$h_a = h(x)1_{\{|h(x)|\le a\}}$$

for $a > 0$. Then

$$\big|\mathbb{E}\big[h_{d^\epsilon}(U^d)\big] - \mathbb{E}\big[h_{d^\epsilon}(W)\big]\big| \le d^\epsilon \|\mathcal{L}(U^d) - \mathcal{N}(0,1)\|_{\mathrm{TV}} = O\big(d^{\epsilon-1}\big).$$

By Markov's inequality, the error due to the modification of $h(U^d)$ is

$$\begin{aligned}
\big|\mathbb{E}\big[h_{d^\epsilon}(U^d)\big] - \mathbb{E}\big[h(U^d)\big]\big| &\le \mathbb{E}\big[\big|h(U^d)\big|, \big|h(U^d)\big| > d^\epsilon\big] \\
&\le \mathbb{E}\bigg[\big|h(U^d)\big|\bigg\{\frac{|h(U^d)|}{d^\epsilon}\bigg\}^{(1-\epsilon)/\epsilon}\bigg] \\
&\le d^{\epsilon-1}\mathbb{E}\big[\big(1+\big|U^d\big|\big)^{k(1+(1-\epsilon)/\epsilon)}\big] = O\big(d^{\epsilon-1}\big)
\end{aligned}$$

by (B.5). Similarly, the error due to the modification of $h(W)$ is dominated by

$$\begin{aligned}
\big|\mathbb{E}\big[h_{d^\epsilon}(W)\big] - \mathbb{E}\big[h(W)\big]\big| &\le \mathbb{E}\big[\big|h(W)\big|, \big|h(W)\big| > d^\epsilon\big] \\
&\le \mathbb{E}\bigg[\big|h(W)\big|\bigg\{\frac{|h(W)|}{d^\epsilon}\bigg\}^{(1-\epsilon)/\epsilon}\bigg] \\
&\le d^{\epsilon-1}\mathbb{E}_y\big[\big(1+|W|\big)^{k(1+(1-\epsilon)/\epsilon)}\big] = O\big(d^{\epsilon-1}\big).
\end{aligned}$$

Hence the claim follows by the triangle inequality. □

B.1.3. *Remark on Stein's method.* We will use a martingale problem approach for the convergence of stochastic processes and hence we will show the convergence of characteristics of semimartingales. In order to prove the convergence of characteristics, we will use Stein's identity and Stein's method.

Thanks to the results in Section B.1.2, the semimartingale characteristics are, essentially, written by expectations with respect to normal distributions. For calculation involving Gaussian random variables, Stein's identity is useful:

$$\mathbb{E}[Wf(W)] = \mathbb{E}[f'(W)], \tag{B.7}$$

where $W \sim \mathcal{N}(0,1)$ and f is sufficiently smooth.

Stein identity (B.7) characterises the standard normal distribution: $W \sim \mathcal{N}(0,1)$ if and only if (B.7) is satisfied for every differentiable function f with $\mathbb{E}|f'(W)| < \infty$. Moreover, by using *Stein's method*, the deviation from $\mathcal{N}(0,1)$ is bounded by the deviation from Stein's identity. The usefulness of Stein's method is illustrated in the monographs Chen et al. [10] and Nourdin and Peccati [30]. In this paper, we will use the following result due to Proposition 3.2.2 of [30].

LEMMA B.3. *For any $h : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[|h(W)|] < \infty$ for $W \sim \mathcal{N}(0,1)$, there is the unique solution $f : \mathbb{R} \to \mathbb{R}$ of the ordinary differential equation (called Stein's equation)*

$$Lf(s) := f'(s) - sf(s) = h(s) - \mathbb{E}[h(W)] \tag{B.8}$$

such that $\lim_{x\to\pm\infty} \phi(x) f(x) = 0$.

There are many important properties of the solution of Stein's equation. We remark here the integration-by-parts formula

$$\int (Lf)(x) g(x)\phi(x)\,\mathrm{d}x = -\int f(x) g'(x)\phi(x)\,\mathrm{d}x \tag{B.9}$$

for smooth functions f, g. Also, we would like to remark the following lemma which provides a sufficient condition for $\mathcal{N}(0,1)$-integrability of Stein's solution. For $\beta > 0$, let

$$|||f|||_\beta = \sup_{x\in\mathbb{R}} e^{-\beta|x|} |f(x)|.$$

If $|||f|||_\beta < \infty$, f is $\mathcal{N}(0,1)$-integrable.

LEMMA B.4. *For $\beta > 0$, there exists $C_\beta < \infty$ such that for any $h : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{E}[h(W)] = 0$ for $W \sim \mathcal{N}(0,1)$, we have*

$$|||f|||_\beta \le C_\beta |||h|||_\beta, \tag{B.10}$$

where f is the solution to (B.8) such that $\lim_{x\to\pm\infty} \phi(x) f(x) = 0$.

PROOF. Without loss of generality, we can assume $|||h|||_\beta < \infty$. By equation (3.23) of [30], Stein's solution is given by

$$f(x) = \phi(x)^{-1} \int_{-\infty}^x h(y)\phi(y)\,\mathrm{d}y = -\phi(x)^{-1} \int_x^\infty h(y)\phi(y)\,\mathrm{d}y.$$

Therefore, if $x \ge 0$, we have

$$\begin{aligned} e^{-\beta x}|f(x)| &= \left(e^{\beta x}\phi(x)\right)^{-1} \left| \int_x^\infty h(y)\phi(y)\,\mathrm{d}y \right| \\ &\le \left(e^{\beta x}\phi(x)\right)^{-1} \int_x^\infty |h(y)|\phi(y)\,\mathrm{d}y \qquad \text{(B.11)} \\ &\le |||h|||_\beta \left(e^{\beta x}\phi(x)\right)^{-1} \int_x^\infty e^{\beta y}\phi(y)\,\mathrm{d}y. \end{aligned}$$

With a similar calculation for $x \le 0$, we obtain the inequality (B.10) with the constant

$$C_\beta = \sup_{x \ge 0} c_\beta(x), \qquad c_\beta(x) := \left(e^{\beta x}\phi(x)\right)^{-1} \int_x^\infty e^{\beta y}\phi(y)\,\mathrm{d}y.$$

Observe that $e^{\beta x}\phi(x) = e^{\beta^2/2}\phi(x-\beta)$. Also, if $y \ge 1$, we have $\phi(y) \le y\phi(y)$ and hence $\Phi(-x) \le \phi(x)$ by integrating $y \in [x, \infty)$. Therefore, if $x \ge \beta + 1$,

$$\begin{aligned} c_\beta(x) &= \phi(x-\beta)^{-1} \int_x^\infty \phi(y-\beta)\,\mathrm{d}y \\ &= \phi(x-\beta)^{-1}\Phi\left(-(x-\beta)\right) \le 1. \end{aligned}$$

Also, $x \mapsto c_\beta(x)$ is continuous, and hence bounded on $[0, \beta+1]$. Hence $C_\beta < \infty$ and the claim follows. □

B.2. Proof of Theorem 2.8. PROOF OF THEOREM 2.8. We apply [23], Theorem IX.3.48, to $S^{\mathrm{B},d}$ with stopping time

$$\tau_a(S^{\mathrm{B}}) = \inf\{t > 0 : |S_t^{\mathrm{B}}| \ge a \text{ or } |S_{t-}^{\mathrm{B}}| \ge a\}$$

for $a > 0$. Let $\tau_a^d = \tau_a(S^{\mathrm{B},d})$. First we prove the local strong majoration hypothesis (i) of Theorem IX.3.48. By the expression of the predictable process B' in (B.1), the total variation process (see Section I.3a) of B' up to the stopping time τ_a is

$$\mathrm{Var}(B')_t^{\tau_a} = \int_0^{t\wedge\tau_a} \left|1 - 2\{(S_s^{\mathrm{B}})^+\}^2 - \rho S_s^{\mathrm{B}}\right| \mathrm{d}s.$$

By construction of ν in (B.2), we have

$$\{|x|^2 * \nu\}_t^{\tau_a} = \int |x|^2 \nu_{t\wedge\tau_a}(\mathrm{d}x) = \int_0^{t\wedge\tau_a} \{4\{(S_s^{\mathrm{B}})^+\}^3 + \rho(1 + (S_s^{\mathrm{B}})^2)\}\,\mathrm{d}s.$$

Hence

$$\mathrm{Var}(B')^{\tau_a} \prec F_1(a), \qquad \{(|x|^2) * \nu\}^{\tau_a} \prec F_2(a),$$

where

$$F_1(a)_t = t\left(1 + 2a^2 + \rho a\right), \qquad F_2(a)_t = t\left(4a^3 + \rho\left(1 + a^2\right)\right).$$

Note that $C \equiv 0$. Thus (i) of Theorem IX.3.48 follows, since $\mathrm{Var}(B')^{\tau_a}$ and $\{(|x|^2) * \nu\}^{\tau_a}$ are strongly majorised by $F(a) = F_1(a) + F_2(a)$.

Second we prove (ii)–(v) of Theorem IX.3.48. If we take $b > 2a$, then

$$\begin{aligned} \{|x|^2 1_{\{|x|>b\}} * \nu\}^{t\wedge\tau_a} &= \rho \int_0^{t\wedge\tau_a} \int_{\mathbb{R}} |z - S_s^{\mathrm{B}}|^2 1_{\{|z - S_s^{\mathrm{B}}| > b\}} \phi(z)\,\mathrm{d}z\,\mathrm{d}s \\ &\le \rho \int_0^t \int_{\mathbb{R}} (|z| + a)^2 1_{\{|z| + a > b\}} \phi(z)\,\mathrm{d}z\,\mathrm{d}s \underset{b\to+\infty}{\longrightarrow} 0, \end{aligned}$$

which proves (ii) of Theorem IX.3.48. The existence and uniqueness of the martingale problem of (2.7) is proved in Section C. Thus local uniqueness condition (iii) of Theorem IX.3.48 comes from Lemma IX 4.4. Continuity condition (iv) is obvious. Since we assume stationarity, both $S_0^{\mathrm{B},d}$ and S_0^{B} follow the standard normal distribution. Thus (v) of Theorem IX.3.48 follows.

Finally we check the condition (vi) of Theorem IX.3.48. Recall that, by construction,

$$\|\xi_t^{\mathrm{B},d} - \xi_0^{\mathrm{B},d}\| \le t \quad \Longrightarrow \quad \sup_{0\le t\le T}\left|\frac{\|\xi_t^{\mathrm{B},d}\|^2}{d} - 1\right| = o_{\mathbb{P}}(1) \tag{B.12}$$

for any $0 \le t \le T$ since $\|v_t^{\mathrm{B},d}\| = 1$ and $\xi_0^{\mathrm{B},d}$ follows the standard normal distribution. Thus for any $0 \le s \le t$,

$$B_s'^d - B_s'(S^{\mathrm{B},d}) = 0, \qquad \left|\tilde{C}_s'^d - \tilde{C}_s'(S^{\mathrm{B},d})\right| \le \rho \int_0^t \left|\frac{\|\xi_s^{\mathrm{B},d}\|^2}{d} - 1\right| \mathrm{d}s = o_{\mathbb{P}}(1),$$

and hence the conditions [Sup-β_{loc}'] and [γ_{loc}'-D] of (vi) are satisfied. For Condition IX.3.49 of (vi), let $g_b(x) = x^2 1_{\{|x|>b\}}$ for $b > 2a$. Then

$$\begin{aligned} g_b * \nu_{t\wedge\tau_a^d}^d &= \rho \int_0^{t\wedge\tau_a^d} \int_{\mathfrak{S}^{d-1}} g_b\big(\langle\xi_s^{\mathrm{B},d}, u\rangle - S_s^{\mathrm{B},d}\big)\psi_d(\mathrm{d}u)\,\mathrm{d}s \\ &\le \rho \int_0^t \int_{\mathfrak{S}^{d-1}} g_b\big(|\langle\xi_s^{\mathrm{B},d}, u\rangle| + a\big)\psi_d(\mathrm{d}u)\,\mathrm{d}s. \end{aligned}$$

By stationarity together with the fact that $\mathcal{L}(\langle\xi_0^{\mathrm{B},d}, u\rangle) = \mathcal{L}(S_0^{\mathrm{B},d})$, we have

$$\begin{aligned} \mathbb{P}\big(g_b * \nu_{t\wedge\tau_a^d}^d > \epsilon\big) &\le \epsilon^{-1}\mathbb{E}\Big[\rho \int_0^t \int_{\mathfrak{S}^{d-1}} g_b\big(|\langle\xi_s^{\mathrm{B},d}, u\rangle| + a\big)\psi_d(\mathrm{d}u)\,\mathrm{d}s\Big] \\ &= \epsilon^{-1} t\rho\mathbb{E}\big[g_b\big(|S_0^{\mathrm{B},d}| + a\big)\big]. \end{aligned}$$

Therefore, by taking the lim sup as $d \to \infty$ of the expectation on the right-hand side of the above inequality gives

$$\limsup_{d\to\infty} \mathbb{P}\big(g_b * \nu_{t\wedge\tau_a^d}^d > \epsilon\big) \le \epsilon^{-1} t\rho\mathbb{E}\big[g_b\big(|S_0^{\mathrm{B}}| + a\big)\big] \underset{b\to\infty}{\longrightarrow} 0$$

by $S_0^{\mathrm{B}} \sim \mathcal{N}(0, 1)$ which establishes Condition IX.3.49 of (vi). Finally, we check [δ_{loc}-D] of (iv). By construction for any bounded, continuous function g, we have

$$\begin{aligned} \epsilon_t^d &:= g * \nu_t^d - (g * \nu_t) \circ S^{\mathrm{B},d} \\ &= \rho \int_0^t \int_{\mathfrak{S}^{d-1}} g\big(\langle\xi_s^{\mathrm{B},d}, u\rangle - S_s^{\mathrm{B},d}\big)\psi_d(\mathrm{d}u)\,\mathrm{d}s - \rho \int_0^t \int_{\mathbb{R}} g\big(z - S_s^{\mathrm{B},d}\big)\phi(z)\,\mathrm{d}z\,\mathrm{d}s. \end{aligned}$$

Therefore, by stationarity of the process, we have

$$\begin{aligned} &\mathbb{E}\Big[\sup_{0\le s\le t} |\epsilon_s^d|\Big] \\ &\quad\le t\rho\mathbb{E}\Big[\Big|\int_{\mathfrak{S}^{d-1}} g\big(\langle\xi_0^{\mathrm{B},d}, u\rangle - S_0^{\mathrm{B},d}\big)\psi_d(\mathrm{d}u) - \int_{\mathbb{R}} g\big(z - S_0^{\mathrm{B},d}\big)\phi(z)\,\mathrm{d}z\Big|\Big] \\ &\quad\le t\rho\|g\|_\infty \mathbb{E}\big[\big\|\mathcal{L}_0\big(\langle\xi_0^{\mathrm{B},d}, u\rangle\big) - \mathcal{N}(0, 1)\big\|_{\mathrm{TV}}\big], \end{aligned}$$

where $u \sim \psi_d$ and $\mathcal{L}_0(X)$ is the conditional distribution of X given $\xi_0^{\mathrm{B},d}$ and $v_0^{\mathrm{B},d}$, and $\|g\|_\infty = \sup_{x\in\mathbb{R}} |g(x)|$. By the property of the spherically symmetric distribution ψ_d, we have

$$\begin{aligned} \mathcal{L}_0\big(\langle\xi_0^{\mathrm{B},d}, u\rangle\big) &= \mathcal{L}_0\bigg(\frac{\|\xi_0^{\mathrm{B},d}\|}{d^{1/2}} d^{-1/2}\bigg\langle \frac{\xi_0^{\mathrm{B},d}}{\|\xi_0^{\mathrm{B},d}\|}, u\bigg\rangle\bigg) = \mathcal{L}_0\big(\alpha^d U^d\big), \\ (\alpha^d)^2 &:= \frac{\|\xi_0^{\mathrm{B},d}\|^2}{d}. \end{aligned} \tag{B.13}$$

Therefore the total variation distance in the above expectation is

$$\|\mathcal{L}_0(\alpha^d U^d) - \mathcal{N}(0,1)\|_{\mathrm{TV}} \le \|\mathcal{L}_0(\alpha^d U^d) - \mathcal{N}(0,(\alpha^d)^2)\|_{\mathrm{TV}} + \|\mathcal{N}(0,(\alpha^d)^2) - \mathcal{N}(0,1)\|_{\mathrm{TV}}.$$

The first term in the right-hand side equals to (B.6) which converges to 0, and the second term is dominated by

$$2|1-(\alpha^d)^2| \underset{d\to\infty}{\longrightarrow} 0 \quad \text{in } \mathbb{P}$$

by Proposition 3.6.1 of [30]. This proves [δ_{loc}-D]. Thus, the condition (iv) of Theorem IX.3.48 of [23] is proved. Hence the claim follows. □

B.3. Proof for Corollary 2.9. PROOF FOR COROLLARY 2.9. By the expression (B.3), the expected number of switches of $S^{\mathrm{B},d}$ per unit time is

$$\begin{aligned}\mathbb{E}\Big[\sum_{0\le t\le T} 1_{\{\Delta S_t^{\mathrm{B},d}\neq 0\}}\Big] &= \mathbb{E}\Big[\int_{(0,T]\times\mathbb{R}_+} 1_{\{z\le S_{s-}^{\mathrm{B},d}\}} N(\mathrm{d}s,\mathrm{d}z) + R_d\big((0,T]\times\mathbb{R}\big)\Big] \\ &= \mathbb{E}\Big[\int_0^T (S_s^{\mathrm{B},d})^+\,\mathrm{d}s + \rho T\Big] \\ &= T\mathbb{E}[(S_0^{\mathrm{B},d})^+ + \rho] \\ &= T\Big\{\int_{\mathbb{R}} x^+\phi(x)\,\mathrm{d}x + \rho\Big\} = T\Big(\frac{1}{\sqrt{2\pi}} + \rho\Big).\end{aligned}$$

□

B.4. Proof for Theorem 2.10. Thanks to the memoryless property of the exponential distribution, we can assume that a refreshment jump occurs at $t=0$ since it does not affect the law of $(\xi_t^{\mathrm{B},d}, v_t^{\mathrm{B},d})_{t\ge 0}$. By Proposition II.1.14 of [23], we can construct a probability space so that there are stopping times $0=\sigma_0<\sigma_1<\sigma_2<\cdots$ with $\mathcal{F}_{\sigma_n}$-measurable random variables W_n^d $(n\ge 1)$ such that

$$R_d(\mathrm{d}t,\mathrm{d}x) = \sum_{n\ge 1} 1_{\{\sigma_n<\infty\}}\delta_{(\sigma_n, W_n^d)}(\mathrm{d}t,\mathrm{d}x), \tag{B.14}$$

where $\mathbb{P}(W_n^d\in A|\mathcal{F}_{\sigma_n -}) = \psi_d(A)$.

The proof strategy of Theorem 2.10 is as follows. The first step is to show the convergence of $Y^{\mathrm{B},d}$ at refreshment times $(\sigma_n)_{n\ge 0}$. For that purpose, we consider a pure step Markov process $\overline{Y}^{\mathrm{B},d}$ defined by

$$\overline{Y}_t^{\mathrm{B},d} := \sum_{n\ge 0} Y_{\sigma_n/d}^{\mathrm{B},d} 1_{[\frac{\sigma_n}{d},\frac{\sigma_{n+1}}{d})}(t) = \sum_{n\ge 0} d^{1/2}\Big(\frac{\|\xi_{\sigma_n}^{\mathrm{B},d}\|^2}{d} - 1\Big) 1_{[\frac{\sigma_n}{d},\frac{\sigma_{n+1}}{d})}(t).$$

The pure step Markov process has a simpler structure which is characterised by the so-called finite transition measure. Since $\sigma_j/d - \sigma_{j-1}/d$ follows the exponential distribution with mean $1/\rho d$, its finite transition measure $K^d(x,\mathrm{d}y)$ is

$$\int_{\mathbb{R}} f(y) K^d(x,\mathrm{d}y) = \rho d\,\mathbb{E}[f(Y_{\sigma_1/d}^{\mathrm{B},d} - Y_0^{\mathrm{B},d})|Y_0^{\mathrm{B},d} = x]$$

in the sense of IX.4.19 of [23]. Then we will apply Theorem IX.4.21 of [23] to the Markov process $\overline{Y}^{\mathrm{B},d}$ in Lemma B.5. To apply the theorem, the key step is the proof for the convergence of the semimartingale characteristics. For this step, Stein's techniques work efficiently. After the proof of Lemma B.5, finally we will show that the difference between $\overline{Y}^{\mathrm{B},d}$ and $Y^{\mathrm{B},d}$ is ignorable.

LEMMA B.5. *The process $\overline{Y}^{B,d}$ converges in law to Y^{B}.*

PROOF. We can construct $(S_t^{B,d})_{t\in[0,\sigma_1)}$ so that

$$S_t^{B,d} = \mathcal{T}_t \quad (0 \le t < \sigma_1), \tag{B.15}$$

where $\mathcal{T}$ follows (2.3) with $\mathcal{T}_0 = S_0^{B,d} = x$, and independent from the refreshment times $(\sigma_n)_{n\ge 0}$. We apply Theorem IX.4.21 of [23]. Since the limiting process is the Ornstein–Uhlenbeck process, hypothesis [23], IX.4.3, is satisfied. By the central limit theorem, $\mathcal{L}(\overline{Y}_0^{B,d})$ converges to $\mathcal{N}(0,2) = \mathcal{L}(Y_0^{B})$, and hence condition (iii) is also satisfied. Therefore, it is sufficient to prove conditions (i) and (ii).

The condition (i) corresponds to the (locally uniformly in y) convergence of

$$b'^d(y) := \rho d\mathbb{E}\big[Y_{\sigma_1/d}^{B,d} - Y_0^d \,|\, Y_0^{B,d} = y\big] \quad \text{and}$$
$$\tilde{c}'^d(y) := \rho d\mathbb{E}\big[\big(Y_{\sigma_1/d}^{B,d} - Y_0^d\big)^2 \,|\, Y_0^{B,d} = y\big].$$

For simplicity, we will denote $\mathbb{E}[\cdot|Y_0^{B,d} = y]$ by $\mathbb{E}_y[\cdot]$. First, we check the convergence of the drift coefficient b'^d. Since $\mathrm{d}\|\xi_t^{B,d}\|^2 = 2S_t^{B,d}\,\mathrm{d}t$, we have

$$\|\xi_{\sigma_1}^{B,d}\|^2 - \|\xi_0^{B,d}\|^2 = 2\int_0^{\sigma_1} S_t^{B,d}\,\mathrm{d}t = 2\int_0^{\sigma_1} \mathcal{T}_t\,\mathrm{d}t = 2\int_0^{\infty} 1_{\{t\le\sigma_1\}}\mathcal{T}_t\,\mathrm{d}t.$$

Since σ_1 and $\mathcal{T}$ are independent, we can rewrite $b'^d(y)$ as

$$\begin{aligned} b'^d(y) &= \rho d^{1/2}\mathbb{E}_y\big[\|\xi_{\sigma_1}^{B,d}\|^2 - \|\xi_0^{B,d}\|^2\big] \\ &= 2\rho d^{1/2}\int_0^{\infty} \mathbb{P}_y(t\le\sigma_1)\mathbb{E}_y\big[\mathbb{E}[\mathcal{T}_t|\mathcal{T}_0 = S_0^{B,d}]\big]\,\mathrm{d}t \\ &= 2\rho d^{1/2}\int_0^{\infty} e^{-\rho t}\mathbb{E}_y\big[h_t(S_0^{B,d})\big]\,\mathrm{d}t, \end{aligned}$$

where $h_t(x) := \mathbb{E}[\mathcal{T}_t|\mathcal{T}_0 = x]$. Now we are going to approximate $S_0^{B,d}$ by a Gaussian random variable. For $\alpha > 0$, by (C.1), we have

$$\big|h_t(\alpha x)\big| = \big|\mathbb{E}[\mathcal{T}_t|\mathcal{T}_0 = \alpha x]\big| \le \mathbb{E}\big[|\mathcal{T}_t|\,|\mathcal{T}_0 = \alpha x\big] \le |\alpha x| + t \le \big(|\alpha| + t\big)\big(1+|x|\big).$$

Conditioned on y, we show that the difference of the law of $S_0^{B,d}$ and the normal distribution $\mathcal{N}(0,(\alpha^d)^2)$ is small, where

$$(\alpha^d)^2 := \frac{\|\xi_0^{B,d}\|^2}{d} = 1 + d^{-1/2}y.$$

By the property of ψ_d, we can rewrite the expectation of $S_0^{B,d}$ in terms of U^d (see (B.4)) since $\mathcal{L}_y(S_0^{B,d}) = \mathcal{L}_y(\alpha^d U^d)$ as in (B.13) where $\mathcal{L}_y$ is the conditional distribution given $Y_0^{B,d} = y$. Therefore, we can apply Proposition B.2 with $k = 1$ and $\epsilon \in (0,1/2)$ to $S_0^{B,d}$. We have

$$\big|\mathbb{E}_y\big[h_t(S_0^{B,d})\big] - \mathbb{E}_y\big[h_t(\alpha^d W)\big]\big| \le \big(|\alpha^d| + t\big)O\big(d^{\epsilon-1}\big), \tag{B.16}$$

where $W \sim \mathcal{N}(0,1)$. Since $\alpha^d \to 1$ locally uniformly in y, we obtain that the drift coefficient is an expectation of the Gaussian random variable with ignorable approximation error:

$$b'^d(y) = 2\rho d^{1/2}\int_0^{\infty} e^{-\rho t}\mathbb{E}_y\big[h_t(\alpha^d W)\big]\,\mathrm{d}t + O\big(d^{\epsilon-1/2}\big).$$

We are in a position to apply Stein's method. Let f_t be Stein's solution for $Lf_t = h_t$. Observe that $\mathbb{E}[h_t(W)] = \mathbb{E}[h_t(\mathcal{T}_0)] = \mathbb{E}[\mathcal{T}_t] = 0$. By Lemma B.4, f_t and $f_t' = xf_t + h_t$ are $\mathcal{N}(0,1)$-integrable. Therefore,

$$\begin{aligned}\mathbb{E}_y\big[h_t(\alpha^d W)\big] &= \mathbb{E}_y\big[f_t'(\alpha^d W) - \alpha^d W f_t(\alpha^d W)\big]\\ &= \mathbb{E}_y\big[f_t'(\alpha^d W) - (\alpha^d)^2 f_t'(\alpha^d W)\big]\\ &= (1-(\alpha^d)^2)\mathbb{E}_y\big[f_t'(\alpha^d W)\big]\\ &= -d^{-1/2} y \mathbb{E}_y\big[f_t'(\alpha^d W)\big],\end{aligned}$$

where we used Stein's identity in the second line. Since $\alpha^d \longrightarrow_{d\to\infty} 1$ locally uniformly in y, by the dominated convergence theorem, we have

$$b'^d(y) \underset{d\to\infty}{\longrightarrow} b'(y) := -2\rho y \int_0^\infty \int_{\mathbb{R}} e^{-\rho t} f_t'(x)\phi(x)\,\mathrm{d}x\,\mathrm{d}t.$$

To finish the calculation of the drift coefficient, we rewrite the expectation in the right hand side without using Stein's solution. By Stein's identity together with (B.9),

$$\begin{aligned}\int_{\mathbb{R}} f_t'(x)\phi(x)\,\mathrm{d}x &= \int_{\mathbb{R}} x f_t(x)\phi(x)\,\mathrm{d}x\\ &= \int_{\mathbb{R}} \left(\frac{x^2}{2}\right)' f_t(x)\phi(x)\,\mathrm{d}x\\ &= -\int_{\mathbb{R}} \frac{x^2}{2} h_t(x)\phi(x)\,\mathrm{d}x\\ &= -\mathbb{E}\left[\left(\frac{\mathcal{T}_0^2}{2}\right)\mathbb{E}[\mathcal{T}_t|\mathcal{T}_0]\right] = -2^{-1}\mathbb{E}[\mathcal{T}_0^2\mathcal{T}_t].\end{aligned}$$

We used Stein's identity in the first line, and the integration by parts formula (B.9) with $f_t = Lh_t$ and $g(x) = x^2/2$ in the third line. We can rewrite this expectation as an integration with respect to the covariance function $K(s,t)$. By the mixing property (C.6) with $k=2$, the right-hand side of the above equation equals

$$2^{-1}\lim_{s\to\infty}\mathbb{E}[(\mathcal{T}_s^2 - \mathcal{T}_0^2)\mathcal{T}_t] = \mathbb{E}\left[\int_0^\infty \mathcal{T}_s\mathcal{T}_t\,\mathrm{d}s\right] = \int_0^\infty K(s,t)\,\mathrm{d}s = \int_0^t K(s,0)\,\mathrm{d}s,$$

where we used (C.1) in the first equation, and (C.2) with $K(s,t) = K(t-s,0)$ for the last equation. Therefore we obtain the expression of the drift coefficient:

$$b'(y) = -2\rho y \int_0^\infty e^{-\rho t}\int_0^t K(s,0)\,\mathrm{d}s\,\mathrm{d}t = -2y\int_0^\infty e^{-\rho s} K(s,0)\,\mathrm{d}s.$$

Second, we check convergence of the diffusion coefficient. By $\mathrm{d}\|\xi_t^{\mathrm{B},d}\|^2 = 2S_t^{\mathrm{B},d}\,\mathrm{d}t$,

$$\begin{aligned}\tilde{c}'^d(y) &= \rho\mathbb{E}_y\big[(\|\xi_{\sigma_1}^{\mathrm{B},d}\|^2 - \|\xi_0^{\mathrm{B},d}\|^2)^2\big]\\ &= 4\rho\mathbb{E}_y\left[\left\{\int_0^{\sigma_1} S_t^{\mathrm{B},d}\,\mathrm{d}t\right\}^2\right] = 4\rho\mathbb{E}_y\left[\left\{\int_0^{\sigma_1}\mathcal{T}_t\,\mathrm{d}t\right\}^2\right].\end{aligned}$$

As in the drift coefficient case, since σ_1 and $\mathcal{T}_t$ are independent, we have

$$\begin{aligned}\mathbb{E}\left[\left\{\int_0^{\sigma_1}\mathcal{T}_t\,\mathrm{d}t\right\}^2\middle|\mathcal{T}_0 = S_0^{\mathrm{B},d}\right] &= \int_0^\infty\int_0^\infty \mathbb{E}[1_{\{s,t\le\sigma_1\}}\mathcal{T}_t\mathcal{T}_s|\mathcal{T}_0 = S_0^{\mathrm{B},d}]\,\mathrm{d}t\,\mathrm{d}s\\ &= \int_0^\infty\int_0^\infty e^{-\rho\max\{s,t\}} h_{s,t}(S_0^{\mathrm{B},d})\,\mathrm{d}t\,\mathrm{d}s,\end{aligned}$$

where $h_{s,t}(x) = \mathbb{E}[\mathcal{T}_t\mathcal{T}_s|\mathcal{T}_0 = x]$. Observe that if $t \geq s \geq 0$, by (C.1), we have

$$|h_{s,t}(\alpha x)| = |\mathbb{E}[\mathcal{T}_t\mathcal{T}_s|\mathcal{T}_0 = \alpha x]| \leq (|\alpha x| + t)(|\alpha x| + s) \leq (|\alpha| + t)^2(1 + |x|)^2.$$

Therefore by Proposition B.2 with $k = 2$, we can approximate the expectation of $S_0^{\mathrm{B},d}$ by that of a Gaussian random variable:

$$|\mathbb{E}[h_{s,t}(S_0^{\mathrm{B},d})] - \mathbb{E}[h_{s,t}(\alpha^d\mathcal{T}_0)]| \leq (|\alpha^d| + \max\{s,t\})^2 O(d^{\epsilon-1})$$

for any $\epsilon \in (0, 1)$. Therefore, we can conclude

$$\tilde{c}'^d(y) = 4\rho \int_0^\infty \int_0^\infty e^{-\rho\max\{s,t\}}\mathbb{E}_y[h_{s,t}(\alpha^d W)]\,\mathrm{d}s\,\mathrm{d}t + O(d^{\epsilon-1}).$$

Hence by the dominated convergence theorem, we have

$$\tilde{c}'^d(y) \underset{d\to\infty}{\longrightarrow} 4\rho \int_0^\infty \int_0^\infty e^{-\rho\max\{s,t\}}K(s,t)\,\mathrm{d}s\,\mathrm{d}t =: \tilde{c}'(y),$$

since $\mathbb{E}[h_{s,t}(W)] = \mathbb{E}[h_{s,t}(\mathcal{T}_0)] = \mathbb{E}[\mathcal{T}_t\mathcal{T}_s] = K(s,t)$. By change of variable $(s,t) \mapsto (t - s, t) =: (u, t)$, we have

$$\begin{aligned}\tilde{c}'(y) &= 8\rho \int_{0<s\leq t<\infty} e^{-\rho t}K(s,t)\,\mathrm{d}s\,\mathrm{d}t \\ &= 8\rho \int_0^\infty \int_u^\infty e^{-\rho t}K(u,0)\,\mathrm{d}t\,\mathrm{d}u \\ &= 8\int_0^\infty e^{-\rho u}K(u,0)\,\mathrm{d}u.\end{aligned} \tag{B.17}$$

Therefore, the condition (i) of Theorem IX.4.21 of [23] follows.

Finally, we check condition (ii). By Markov property, for any $\epsilon > 0$, we have

$$\int_{\mathbb{R}} K^d(x,\mathrm{d}y)|y|^2 1_{\{|y|>\epsilon\}} \leq \epsilon^{-2}\int_{\mathbb{R}} K^d(x,\mathrm{d}y)|y|^4 =: \epsilon^{-2}\delta^d(y).$$

By construction of K^d, we can rewrite $\delta^d(y)$ as

$$\delta^d(y) = \rho d\mathbb{E}[(Y_{\sigma_1/d}^{\mathrm{B},d} - Y_0^{\mathrm{B},d})^4|Y_0^{\mathrm{B},d} = y].$$

By Hölder's inequality,

$$\begin{aligned}\delta^d(y) &= \rho d^{-1}\mathbb{E}_y[(\|\xi_{\sigma_1}^{\mathrm{B},d}\|^2 - \|\xi_0^{\mathrm{B},d}\|^2)^4] \\ &= 16\rho d^{-1}\mathbb{E}_y\left[\left\{\int_0^{\sigma_1} S_t^{\mathrm{B},d}\,\mathrm{d}t\right\}^4\right] \\ &= 16\rho d^{-1}\mathbb{E}_y\left[\left\{\int_0^{\sigma_1} \mathcal{T}_t\,\mathrm{d}t\right\}^4\right] \\ &\leq 16\rho d^{-1}\mathbb{E}_y[\sigma_1^4(|S_0^{\mathrm{B},d}| + \sigma_1)^4] \\ &= O(d^{-1})\end{aligned}$$

locally uniformly in y where we used (C.1) in the inequality. Therefore, the condition (ii) follows. Thus, the claim follows by Theorem IX.4.21 of [23]. □

PROOF OF THEOREM 2.10. We showed that the process $\overline{Y}^{\mathrm{B},d}$ converges in law to Y^{B}. Therefore, by Lemma VI.3.31 of [23], it is sufficient to show

$$\epsilon_T^d := \sup_{0\leq t\leq T}|Y_t^{\mathrm{B},d} - \overline{Y}_t^{\mathrm{B},d}| \underset{d\to\infty}{\longrightarrow} 0$$

in probability for any $T > 0$. Let

$$A_T = (0, T] \times \mathbb{R} \quad \text{and} \quad \lambda_T = \rho T.$$

Then $R_d(A_T)$ follows the Poisson distribution with mean λ_T. In particular, $R_d(A_{dT})/d$ is tight. Since $R_d(A_T)$ is the number of the refreshment jumps until $T > 0$, we have

$$\epsilon_T^d \le \sup_{j \le R_d(A_{dT})} \sup_{\sigma_j \le dt < \sigma_{j+1}} \left|Y_t^{\mathrm{B},d} - Y_{\sigma_j/d}^{\mathrm{B},d}\right|.$$

On the other hand, for $\sigma_j \le dt < \sigma_{j+1}$, we have

$$\begin{aligned} \left|Y_t^{\mathrm{B},d} - Y_{\sigma_j/d}^{\mathrm{B},d}\right| &= d^{-1/2}\left|\|\xi_{td}^{\mathrm{B},d}\|^2 - \|\xi_{\sigma_j}^{\mathrm{B},d}\|^2\right| \\ &\le 2d^{-1/2}\int_{\sigma_j}^{\sigma_{j+1}} \left|S_t^{\mathrm{B},d}\right| \mathrm{d}t \\ &\le 2d^{-1/2}\int_{\sigma_j}^{\sigma_{j+1}} \left(\left|S_{\sigma_j}^{\mathrm{B},d}\right| + t\right) \mathrm{d}t \\ &= 2d^{-1/2}\left(\left|S_{\sigma_j}^{\mathrm{B},d}\right|(\sigma_{j+1} - \sigma_j) + \frac{1}{2}(\sigma_{j+1} - \sigma_j)^2\right), \end{aligned}$$

where we used (C.1) in the third line. Therefore, for any $J \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(\epsilon_T^d > \epsilon) &\le \mathbb{P}(R_d(A_{dT}) > dJ) \\ &\quad + \mathbb{P}\left(2d^{-1/2}\sup_{j \le dJ}\left(\left|S_{\sigma_j}^{\mathrm{B},d}\right|(\sigma_{j+1} - \sigma_j) + \frac{1}{2}(\sigma_{j+1} - \sigma_j)^2\right) > \epsilon\right) \\ &\le \mathbb{P}(R_d(A_{dT}) > dJ) + dJ\,\mathbb{P}\left(2d^{-1/2}\left(\left|S_0^{\mathrm{B},d}\right|\sigma_1 + \frac{1}{2}\sigma_1^2\right) > \epsilon\right). \end{aligned}$$

If we take $J \in \mathbb{N}$ large enough, the first probability in the right-hand side of the above inequality can be small. The second term converges to 0 by Markov's inequality together with the fact that $S_0^{\mathrm{B},d} \sim \mathcal{N}(0, 1)$ and σ_1 follows the exponential distribution with mean $1/\rho$. Hence the claim follows. □

B.5. Proof for Proposition 2.11.

PROOF FOR PROPOSITION 2.11. By Proposition C.3 together with Lebesgue's dominated convergence theorem, the claim is obvious. □

B.6. Proof for Theorem 2.13. First, we prove that the process $Z^{\mathrm{B},d,k}$ can be approximated by a pure step Markov process. Second, we show that this approximated process converges to an Ornstein–Uhlenbeck process which completes the proof of Theorem 2.13.

B.6.1. *Approximation of the process*. Let

$$\overline{Z}_t^{\mathrm{B},d,k} := \sum_{n \ge 0} Z_{\sigma_n/d}^{\mathrm{B},d,k} 1_{[\frac{\sigma_n}{d}, \frac{\sigma_{n+1}}{d})}(t),$$

be the pure step version of $Z^{\mathrm{B},d,k}$. By construction, we have the following decomposition imitating the Doob–Meyer decomposition

$$\Delta\overline{Z}_{\sigma_{n+1}/d}^{\mathrm{B},d,k} := \overline{Z}_{\sigma_{n+1}/d}^{\mathrm{B},d,k} - \overline{Z}_{\sigma_n/d}^{\mathrm{B},d,k} = \int_{\sigma_n}^{\sigma_{n+1}} \pi_k\left(v_t^{\mathrm{B},d}\right) \mathrm{d}t = M_{n+1}^d + A_{n+1}^d, \tag{B.18}$$

where

$$M_{n+1}^d = \int_{\sigma_n}^{\sigma_{n+1}} \pi_k\big(v_{\sigma_n}^{\mathrm{B},d}\big)\,\mathrm{d}t, \qquad A_{n+1}^d = \int_{\sigma_n}^{\sigma_{n+1}} \pi_k\big(v_t^{\mathrm{B},d} - v_{\sigma_n}^{\mathrm{B},d}\big)\,\mathrm{d}t.$$

Now we want to extract a predictable component from A_{n+1}^d. Let $(\mathcal{F}_t^d)_{t\geq 0}$ be the underlying filtration. For $N \in \mathbb{N}$, we show the following.

LEMMA B.6.

$$\mathbb{E}\left[\sum_{i=0}^{dN-1} \|A_{i+1}^d\|^2\right] \longrightarrow 0.$$

PROOF. By stationarity assumption, each A_n^d has the same law. Therefore it is sufficient to show that $d\mathbb{E}[\|A_1^d\|^2] \longrightarrow 0$. For the spherical symmetricity of the process $v^{\mathrm{B},d}$, we have

$$d\mathbb{E}[\|A_1^d\|^2] = d\mathbb{E}\left[\left\|\int_0^{\sigma_1} \pi_k\big(v_t^{\mathrm{B},d} - v_0^{\mathrm{B},d}\big)\,\mathrm{d}t\right\|^2\right] = k\mathbb{E}\left[\left\|\int_0^{\sigma_1} v_t^{\mathrm{B},d} - v_0^{\mathrm{B},d}\,\mathrm{d}t\right\|^2\right].$$

Since the stopping time σ_1 is independent from $\mathcal{F}_{\sigma_1-}^d$, by the dominated convergence theorem, it is sufficient to show that $\mathbb{E}[\|v_t^{\mathrm{B},d} - v_0^{\mathrm{B},d}\|^2] \longrightarrow 0$ for any $t > 0$, where $v_t^{\mathrm{B},d}$ follows the stochastic differential equation defined in Section 1.1.2 without refreshment jumps. We have

$$v_t^{\mathrm{B},d} - v_0^{\mathrm{B},d} = \int_{(0,t]\times\mathbb{R}_+} \psi(u,z) N(\mathrm{d}u,\mathrm{d}z), \qquad \psi(t,z) := -2S_{t-}^{\mathrm{B},d} \frac{\xi_{t-}^{\mathrm{B},d}}{\|\xi_{t-}^{\mathrm{B},d}\|^2} 1_{\{z\leq S_{t-}^{\mathrm{B},d}\}}.$$

Therefore by Theorem II.1.33 of [23],

$$\mathbb{E}[\|v_t^{\mathrm{B},d} - v_0^{\mathrm{B},d}\|^2] = \mathbb{E}\left[\int_{(0,t]\times\mathbb{R}_+} \|\psi(u,z)\|^2\,\mathrm{d}u\,\mathrm{d}z + \left\|\int_{(0,t]\times\mathbb{R}_+} \psi(u,z)\,\mathrm{d}u\,\mathrm{d}z\right\|^2\right].$$

By stationarity of the process together with Fubini's theorem, we have a bound

$$\mathbb{E}[\|v_t^{\mathrm{B},d} - v_0^{\mathrm{B},d}\|^2] \leq 4\mathbb{E}\left[t\{(S_0^{\mathrm{B},d})^+\}^3 \frac{1}{\|\xi_0^{\mathrm{B},d}\|^2} + t^2\{(S_0^{\mathrm{B},d})^+\}^4 \frac{1}{\|\xi_0^{\mathrm{B},d}\|^2}\right].$$

The variable $S_0^{\mathrm{B},d}$ follows the standard normal distribution, and $\|\xi_0^{\mathrm{B},d}\|^{-2}$ follows the inverse of the chi-squared distribution with d degrees of freedom which is on the order of d^{-1} by Lemma 4.1 of [24]. Thus, the expectation in the above has on the order of d^{-1} by the Cauchy–Schwarz inequality. Thus $\mathbb{E}[\|A_{i+1}^d\|^2]$ is on the order of d^{-2} which proves the claim. □

COROLLARY B.7.

$$\sup_{n=1,\dots,dN}\left\|\sum_{i=0}^{n-1} A_{i+1}^d - \mathbb{E}[A_{i+1}^d|\mathcal{F}_{\sigma_i-}^d]\right\| \underset{d\to\infty}{\longrightarrow} 0$$

in probability.

PROOF. Consider a filtration $(\mathcal{F}_{\sigma_n-}^d)_n$. A discrete process $(X_n)_{n=0,1,\dots}$ is L-dominated by $(Y_n)_{n=0,1,\dots}$ in the sense of I.3.29 of [23], that is, $\mathbb{E}[|X_\tau|] \leq \mathbb{E}[|Y_\tau|]$ for any bounded $(\mathcal{F}_{\sigma_n-}^d)_n$-stopping time τ where

$$X_n := \left\|\sum_{i=0}^{n-1} A_{i+1}^d - \mathbb{E}[A_{i+1}^d|\mathcal{F}_{\sigma_i-}^d]\right\|^2, \qquad Y_n := \sum_{i=0}^{n-1}\|A_{i+1}^d\|^2.$$

Then, by Lenglart's inequality (I.3.30 of [23]), we have

$$\mathbb{P}\Big(\sup_{n\leq dN} X_n \geq \epsilon\Big) \leq \frac{\eta}{\epsilon} + \mathbb{P}(Y_{dN} \geq \eta)$$

for $\epsilon, \eta > 0$. Therefore, the convergence of $\sup_{n\leq dN} X_n$ comes from Lemma B.6. □

Now we show that a predictable component $\mathbb{E}[A^d_{n+1}|\mathcal{F}^d_{\sigma_n-}]$ has a simpler expression $\mathbb{E}[B^d_{n+1}|\mathcal{F}^d_{\sigma_n-}]$ where

$$B^d_{n+1} = -\int_{\sigma_n}^{\sigma_{n+1}} \int_0^t \frac{1}{d}\pi_k(\xi^{\mathrm{B},d}_{\sigma_n})\,\mathrm{d}s\,\mathrm{d}t$$

up to negligible term. Note that

$$\mathbb{E}[B^d_{n+1}|\mathcal{F}^d_{\sigma_n-}] = -\frac{\rho^{-2}}{d}\pi_k(\xi^{\mathrm{B},d}_{\sigma_n}).$$

LEMMA B.8.

$$\sup_{n=1,\dots,dN}\Bigg\|\sum_{i=0}^{n-1}\mathbb{E}[A^d_{i+1}|\mathcal{F}^d_{\sigma_i-}] - \mathbb{E}[B^d_{i+1}|\mathcal{F}^d_{\sigma_i-}]\Bigg\| \underset{d\to\infty}{\longrightarrow} 0$$

in probability.

PROOF. By stationarity of the process together with the Cauchy–Schwarz inequality, it is sufficient to show $d\mathbb{E}[\|\mathbb{E}[A^d_1 - B^d_1|\mathcal{F}^d_{0-}]\|^2]^{1/2} \longrightarrow 0$. By spherical symmetricity of the processes, we have

$$\begin{aligned}\mathbb{E}[\|\mathbb{E}[A^d_1 - B^d_1|\mathcal{F}^d_{0-}]\|^2] &= \frac{k}{d}\mathbb{E}\bigg[\bigg\|\mathbb{E}\bigg[\int_0^{\sigma_1}\int_{(0,t]\times\mathbb{R}_+}\psi(s,z)N(\mathrm{d}s,\,\mathrm{d}z)\,\mathrm{d}t\,|\mathcal{F}^d_{0-}\bigg]\bigg\|^2\bigg]\\ &= \frac{k}{d}\mathbb{E}\bigg[\bigg\|\mathbb{E}\bigg[\int_0^{\sigma_1}\int_0^t \psi(s)\,\mathrm{d}s\,\mathrm{d}t\,|\mathcal{F}^d_{0-}\bigg]\bigg\|^2\bigg],\end{aligned}$$

where

$$\begin{aligned}\psi(t,z) &= -2S^{\mathrm{B},d}_{t-}\frac{\xi^{\mathrm{B},d}_{t-}}{\|\xi^{\mathrm{B},d}_{t-}\|^2}1_{\{z\leq S^{\mathrm{B},d}_{t-}\}} + \frac{\xi^{\mathrm{B},d}_{t-}}{d},\\ \psi(t) &= -2\{(S^{\mathrm{B},d}_t)^+\}^2\frac{\xi^{\mathrm{B},d}_t}{\|\xi^{\mathrm{B},d}_t\|^2} + \frac{\xi^{\mathrm{B},d}_t}{d}.\end{aligned}$$

By the Cauchy–Schwarz inequality together with the dominated convergence theorem, it is sufficient to prove $d\mathbb{E}[\|\mathbb{E}[\int_0^t \psi(s)\,\mathrm{d}s|\mathcal{F}^{\mathrm{B},d}_{0-}]\|^2] \longrightarrow 0$ where $\xi^{\mathrm{B},d}_t$ and $v^{\mathrm{B},d}_t$ follow the stochastic differential equation defined in Section 1.1.2 without refreshment jumps. Let

$$\begin{aligned}\psi_1(t) &= -2\{(S^{\mathrm{B},d}_t)^+\}^2\bigg\{\frac{\xi^{\mathrm{B},d}_t}{\|\xi^{\mathrm{B},d}_t\|^2} - \frac{\xi^{\mathrm{B},d}_t}{d}\bigg\},\\ \psi_2(t) &= -\{2\{(S^{\mathrm{B},d}_t)^+\}^2 - 1\}\frac{\xi^{\mathrm{B},d}_t - \xi^{\mathrm{B},d}_0}{d},\\ \psi_3(t) &= -\{2\{(S^{\mathrm{B},d}_t)^+\}^2 - 1\}\frac{\xi^{\mathrm{B},d}_0}{d},\end{aligned}$$

so that $\psi(t) = \psi_1(t) + \psi_2(t) + \psi_3(t)$. Convergence of $d\mathbb{E}[\|\psi_1(t)\|^2] = d\mathbb{E}[\|\psi_1(0)\|^2]$ follows from the Cauchy–Schwarz inequality as in the proof of Lemma B.6. Convergence of $d\mathbb{E}[\|\psi_1(t)\|^2]$ also follows by the Cauchy–Schwarz inequality together with the uniform bound $\|\xi_t^{\mathrm{B},d} - \xi_0^{\mathrm{B},d}\| \le t$. Therefore the proof will be completed if we show $d\mathbb{E}[\|\mathbb{E}[\int_0^t \psi_3(s)\,\mathrm{d}s|\mathcal{F}_{0-}^{\mathrm{B},d}]\|^2] \longrightarrow 0$.

By (B.3), up to the refreshment time, we have

$$S_t^{\mathrm{B},d} = S_0^{\mathrm{B},d} + t - 2\int_{(0,t]\times\mathbb{R}} S_{s-}^{\mathrm{B},d} 1_{\{z\le S_{s-}^{\mathrm{B},d}\}} N(\mathrm{d}s, \mathrm{d}z).$$

By this fact,

$$\begin{aligned}
\mathbb{E}\left[\int_0^t \psi_3(s)\,\mathrm{d}s|\mathcal{F}_{0-}^{\mathrm{B},d}\right] &= \mathbb{E}\left[t - 2\int_{(0,t]\times\mathbb{R}} S_{s-}^{\mathrm{B},d} 1_{\{z\le S_{s-}^{\mathrm{B},d}\}} N(\mathrm{d}s, \mathrm{d}z)|\mathcal{F}_{0-}^{d}\right]\frac{\xi_0^{\mathrm{B},d}}{d} \\
&= \mathbb{E}[S_t^{\mathrm{B},d} - S_0^{\mathrm{B},d}|\mathcal{F}_{0-}^{d}]\frac{\xi_0^{\mathrm{B},d}}{d} \\
&= \mathbb{E}[h_t(S_0^{\mathrm{B},d}) - S_0^{\mathrm{B},d}|\mathcal{F}_{0-}^{d}]\frac{\xi_0^{\mathrm{B},d}}{d},
\end{aligned}$$

where $h_t(x) = \mathbb{E}[\mathcal{T}_t|\mathcal{T}_0 = x]$. Let $\mathcal{L}_{0-}(X)$ be the distribution of X conditioned on $\mathcal{F}_{0-}^d$. Since the initial velocity is independent from the initial state, we have $\mathcal{L}_{0-}(S_0^{\mathrm{B},d}) = \mathcal{L}_{0-}(\alpha^d U^d)$ as in (B.13) where U^d is defined in (B.4) and $(\alpha^d)^2 = \|\xi_0^{\mathrm{B},d}\|^2/d$. In particular, $\mathbb{E}[S_0^{\mathrm{B},d}|\mathcal{F}_{0-}^d] = 0$. Moreover, by (B.16), we can substitute $h_t(S_0^{\mathrm{B},d})$ by $h_t(\alpha^d W)$ where W follows the standard normal distribution. Finally the claim follows by the dominated convergence theorem since $\alpha^d \to 1$ and $\mathbb{E}[h_t(W)|\mathcal{F}_{0-}^d] = 0$. Therefore, $d\mathbb{E}[\|\mathbb{E}[\int_0^t \psi_3(s)\,\mathrm{d}s|\mathcal{F}_{0-}^{\mathrm{B},d}]\|^2] \longrightarrow 0$ which proves the claim. □

LEMMA B.9.

$$\sup_{n=1,\dots,dN}\left\|\int_0^{\sigma_n/d} b'(\overline{Z}_t^{\mathrm{B},d,k})\,\mathrm{d}t - \sum_{i=0}^{n-1}\mathbb{E}[B_{i+1}^d|\mathcal{F}_{\sigma_i-}^d]\right\| \xrightarrow[d\to\infty]{} 0$$

in probability, where $b'(x) = -\rho^{-1}x$.

PROOF. Since the difference in the norm is

$$-\sum_{i=0}^{N-1}(\sigma_{i+1} - \sigma_i - \rho^{-1})\frac{\rho^{-1}}{d}\pi_k(\xi_{\sigma_i}^{\mathrm{B},d})$$

and it is a martingale. Therefore the claim follows from Doob's inequality (I.1.43 of [23]). □

Since $\overline{Z}^{\mathrm{B},d,k}$ is a pure step process, the semimartingale characteristics are entirely described by a random measure as described in Theorem II.3.11(b) of [23] (See also Proposition II.2.17). Therefore, we have the first and modified second characteristics as follows:

$$\begin{aligned}
B_T'^d &= \sum_{n:\sigma_n\le T}\mathbb{E}[\Delta\overline{Z}_{\sigma_n/d}^{\mathrm{B},d,k}|\mathcal{F}_{\sigma_{n-1}-}], \\
\tilde{C}_T'^d &= \sum_{n:\sigma_n\le T}\mathbb{E}[(\Delta\overline{Z}_{\sigma_n/d}^{\mathrm{B},d,k})^{\otimes 2}|\mathcal{F}_{\sigma_{n-1}-}] - \mathbb{E}[(\Delta\overline{Z}_{\sigma_n/d}^{\mathrm{B},d,k})|\mathcal{F}_{\sigma_{n-1}-}]^{\otimes 2}.
\end{aligned}$$

Also the corresponding random measure is

$$g * \nu_T^d = \sum_{n:\sigma_n \le T} \mathbb{E}[g(\Delta \overline{Z}^{\mathrm{B},d,k}_{\sigma_n/d})|\mathcal{F}_{\sigma_{n-1}-}]$$

for a bounded smooth function $g(x)$. Here, for a vector $v = (v_1, \dots, v_k) \in \mathbb{R}^d$, $v^{\otimes 2}$ is a $k \times k$ matrix with (i, j)th element $v_i v_j$.

LEMMA B.10. *The process $\overline{Z}^{\mathrm{B},d,k}$ converges in law to $Z^{\mathrm{B},k}$.*

PROOF. The first and the modified second characteristics of $Z^{\mathrm{B},k}$ are

$$B'_T = \int_0^T b'(Z_t^{\mathrm{B},k})\,\mathrm{d}t, \qquad \tilde{C}'_T = 2T\rho^{-1}.$$

We apply Theorem IX.3.48 of [23]. Conditions (i)–(iv) are obvious since the limit is the Ornstein–Uhlenbeck process. The condition (v) is also clear since in this case, both η^d and η are the k-dimensional standard normal distribution. Therefore we only need to check four conditions in (vi).

First we can assume that the number of refreshment jumps until time T, $R_d((0, T] \times \mathbb{R})$ is smaller than dN for some $N \in \mathbb{N}$ by the argument of the proof of Theorem 2.10. Let ν^d be the random measure corresponding to $\overline{Z}^{\mathrm{B},d,k}$. For $g \in C_1(\mathbb{R})$ (See VII.2.7 of [23]), we can assume that $|g(x)| \le 1$ for any x and $g(x) = 0$ for $|x| < b$ for some $b > 0$. Then

$$g * \nu_T^d \le \sum_{n:\sigma_n \le T} \mathbb{P}(\|\Delta \overline{Z}^{\mathrm{B},d,k}_{\sigma_n/d}\| > b|\mathcal{F}_{\sigma_{n-1}-}).$$

Therefore, it is sufficient to prove

$$\sum_{i=1}^{dN} \mathbb{P}(\|\Delta \overline{Z}^{\mathrm{B},d,k}_{\sigma_n/d}\| > b) \xrightarrow[d\to\infty]{} 0$$

for [δ_{loc}-D]. This is also a sufficient condition for 3.49 of Theorem IX.3.48. By equation (B.18), we have

$$\|\Delta \overline{Z}^{\mathrm{B},d,k}_{\sigma_n/d}\| \le \|M_n^d\| + \|A_n^d\|.$$

The convergence of A_n^d part directly follows from Lemma B.6 with Chevyshev's inequality, and the convergence of M_n^d part follows from Markov's inequality together with the fact that the square of each component of $v^{\mathrm{B},d}_{\sigma_n}$ follows the Beta distribution with parameter $1/2$ and $(d-1)/2$. Condition [Sup-β'_{loc}] follows by Corollary B.7–B.9. Finally we check [γ'_{loc}-D].

By the decomposition of $\Delta\overline{Z}^{\mathrm{B},d,k}$, we have

$$\begin{aligned}\tilde{C}'^d_T = \sum_{n;\sigma_n \le T} \mathbb{E}[&(M_n^d)^{\otimes 2} + M_n^d \otimes (A_n^d - \mathbb{E}[A_n^d|\mathcal{F}_{\sigma_{n-1}-}]) \\ &+ (A_n^d - \mathbb{E}[A_n^d|\mathcal{F}_{\sigma_{n-1}-}]) \otimes M_n^d + (A_n^d - \mathbb{E}[A_n^d|\mathcal{F}_{\sigma_{n-1}-}])^{\otimes 2}|\mathcal{F}_{\sigma_{n-1}-}].\end{aligned}$$

The first term is

$$\mathbb{E}[(M_n^d)^{\otimes 2}|\mathcal{F}_{\sigma_{n-1}-}] = \mathbb{E}[(\sigma_n - \sigma_{n-1})^2 \pi_k(v^{\mathrm{B},d}_{\sigma_{n-1}})^{\otimes 2}|\mathcal{F}_{\sigma_{n-1}-}] = 2\rho^{-2}d^{-1}I_k.$$

From this fact together with Lemma B.6, the other term converges to 0. By the same argument as Lemma B.9, the claim follows. □

LEMMA B.11. *The process $Z^{\mathrm{B},d,k}$ converges in law to $Z^{\mathrm{B},k}$.*

PROOF. By Lemma VI.3.31 of [23], it is sufficient to show

$$\epsilon_T^d := \sup_{0\le t\le T} \|Z_t^{\mathrm{B},d,k} - \overline{Z}_t^{\mathrm{B},d,k}\| \underset{d\to\infty}{\longrightarrow} 0$$

in probability. Let A_t be as in the proof of Theorem 2.10. Then we have

$$\epsilon_T^d \le \sup_{0\le j\le R_d(A_{dT})} \sup_{\sigma_j\le t<\sigma_{j+1}} \|\pi_k(\xi_t^{\mathrm{B},d}) - \pi_k(\xi_{\sigma_j}^{\mathrm{B},d})\|.$$

Therefore, for $J\in\mathbb{N}$,

$$\begin{aligned}\mathbb{P}(\epsilon_t^d > \epsilon) &\le \mathbb{P}(R_d(A_{dT}) > dJ)\\ &\quad + \mathbb{P}\Big(\sup_{0\le j\le dJ} \sup_{\sigma_j\le t<\sigma_{j+1}} \|\pi_k(\xi_t^{\mathrm{B},d}) - \pi_k(\xi_{\sigma_j}^{\mathrm{B},d})\| > \epsilon\Big)\\ &\le \mathbb{P}(R_d(A_{dT}) > dJ) + dJ\,\mathbb{P}\Big(\sup_{0\le t<\sigma_1} \|\pi_k(\xi_t^{\mathrm{B},d}) - \pi_k(\xi_0^{\mathrm{B},d})\| > \epsilon\Big).\end{aligned}$$

On the other hand,

$$\|\pi_k(\xi_s^{\mathrm{B},d}) - \pi_k(\xi_0^{\mathrm{B},d})\| \le \int_0^s \|\pi_k(v_u^{\mathrm{B},d})\|\,\mathrm{d}u$$

and the forth moment of the norm is on the order of d^{-2}. Thus by Markov's inequality, ϵ_T^d is negligible. □

B.6.2. *Proof of Theorem* 2.13.

PROOF OF THEOREM 2.13. Weak convergence of $Z^{\mathrm{B},d,k}$ has been proved. Therefore, the proof of Theorem 2.13 will be completed if we can show the law of large numbers (2.9). The proof is essentially the same as that of Lemma B.4 of [24].

Let $\|f\|_\infty = \sup_{x\in\mathbb{R}^k} |f(x)|$. Without loss of generality, we can assume $\int f(x)\phi_k(x)\,\mathrm{d}x = 0$. It is sufficient to show that

$$I_{d,T} := \mathbb{E}\bigg[\bigg|\frac{1}{T}\int_0^T f(Z_t^{\mathrm{B},d,k})\,\mathrm{d}t\bigg|\bigg] \underset{d,T\to\infty}{\longrightarrow} 0.$$

Since the limiting process is the ergodic Ornstein–Uhlenbeck process, for any $\epsilon > 0$ we can find $T_0 > 0$ so that

$$I_{T_0} = \mathbb{E}\bigg[\bigg|\frac{1}{T_0}\int_0^{T_0} f(Z_t^{\mathrm{B},k})\,\mathrm{d}t\bigg|\bigg] < \epsilon$$

by the law of large numbers. By dividing the interval $[0, T]$ into shorter intervals with length T_0, we have

$$\begin{aligned}I_{d,T} &= \mathbb{E}\bigg[\bigg|\frac{1}{T}\sum_{k=0}^{[T/T_0]-1}\int_{kT_0}^{(k+1)T_0} f(Z_t^{\mathrm{B},d,k})\,\mathrm{d}t + \frac{1}{T}\int_{T_0[T/T_0]}^T f(Z_t^{\mathrm{B},d,k})\,\mathrm{d}t\bigg|\bigg]\\ &\le \frac{T_0}{T}\sum_{k=0}^{[T/T_0]-1}\mathbb{E}\bigg[\bigg|\frac{1}{T_0}\int_{kT_0}^{(k+1)T_0} f(Z_t^{\mathrm{B},d,k})\,\mathrm{d}t\bigg|\bigg]\\ &\quad + \frac{1}{T}\int_{T_0[T/T_0]}^T \mathbb{E}[|f(Z_t^{\mathrm{B},d,k})|]\,\mathrm{d}t.\end{aligned}$$

Then by stationarity of the process $Z^{\mathrm{B},d,k}$ together with the weak convergence of $Z^{\mathrm{B},d,k}$, we have

$$I_{d,T} \le \frac{T_0}{T}\bigg[\frac{T}{T_0}\bigg] I_{d,T_0} + \frac{T - T_0[T/T_0]}{T}\|f\|_\infty \underset{d,T\to\infty}{\longrightarrow} I_{T_0} \le \epsilon,$$

which completes the proof. □

APPENDIX C: ERGODIC PROPERTIES OF THE LIMITING PROCESSES

In this section, we study the ergodic properties of the limiting processes S^{B} and $\mathcal{T}$ together with existence and uniqueness of the solution. First we show the existence and uniqueness of the strong solutions of (2.3) and (2.6). The existence of a strong solution and pathwise uniqueness of the processes directly comes from Theorem IV.9.1 of [22]. Here we show a basic idea of the proof because of its importance as well as the fact that we are using an explicit form of pathwise construction in this paper. By Proposition II.1.14 of [23] (see also III.1.24), there are stopping times $0 < \tau_1 < \tau_2 < \cdots$ with $\mathcal{F}_{\tau_n}$-measurable random variables Z_n $(n \geq 1)$ such that

$$N(\mathrm{d}t, \mathrm{d}z) = \sum_{n\geq 1} 1_{\{\tau_n < \infty\}} \delta_{(\tau_n, Z_n)}(\mathrm{d}t, \mathrm{d}z).$$

By this expression, we can construct the process $\mathcal{T}$ uniquely in the time interval $[0, \tau_1]$ by

$$\mathcal{T}_t = \begin{cases} x + t, & 0 \leq t < \tau_1, \\ \mathcal{T}_{\tau_1 -} + 1_{\{Z_n \leq \mathcal{T}_{\tau_1 -}\}}(-2\mathcal{T}_{\tau_1 -}), & t = \tau_1. \end{cases}$$

Similarly, we can construct a pathwise unique solution in the time interval $[0, \tau_n]$ for any $n \in \mathbb{N}$, and hence $\mathcal{T}_t$ is determined globally. It is easy to see that the process is nonexplosive since $|\mathcal{T}_t| \leq |\mathcal{T}_0| + t$.

For (2.6), in the same way, there are stopping times $0 < \sigma_1 < \sigma_2 < \cdots$ with $\mathcal{F}_{\sigma_n}$-measurable random variables W_n $(n \geq 1)$ such that $\mathcal{L}(W_n | \mathcal{F}_{\sigma_n -}) = \mathcal{N}(0, 1)$ and

$$R(\mathrm{d}t, \mathrm{d}x) = \sum_{n\geq 1} 1_{\{\sigma_n < \infty\}} \delta_{(\sigma_n, W_n)}(\mathrm{d}t, \mathrm{d}x).$$

Then we can construct the unique solution in time interval $[0, \sigma_1]$ by

$$S_t^{\mathrm{B}} = \begin{cases} \mathcal{T}_t, & 0 \leq t < \sigma_1, \\ W_1, & t = \sigma_1. \end{cases}$$

Then, the process $(S_t^{\mathrm{B}})_{t\in[\sigma_1, \sigma_2]}$ proceeds according to (2.3) up to time σ_2- starting from $S_{\sigma_1}^{\mathrm{B}} = W_1$ in the same way as above. By iterating this procedure, we can construct a unique solution in time interval $[0, \sigma_n]$ for any $n \in \mathbb{N}$, and hence S_t^{B} is determined globally. Since $\sigma_n - \sigma_{n-1}$ $(n = 1, 2, \ldots)$ are identically distributed, the process is nonexplosive. Therefore, the existence and uniqueness of the strong solutions of (2.3) and (2.6) follows.

Next we consider existence and uniqueness of solution of martingale problems corresponding to (2.4) and (2.7). For this purpose, it might be natural to use a combined representation for (2.6) introduced in Remark B.1. Then by Itô's formula, weak solutions of (2.3) and (2.6) solve martingale problems corresponding to (2.4) and (2.7). Since weak solutions of (2.3) and (2.6) are unique by pathwise uniqueness, we also have uniqueness of martingale problems by Theorem 2.3 of [26].

Let ψ be a σ-finite measure on a measurable space $(E, \mathcal{E})$. Then a continuous time Markov process X_t is said to be (ψ-)irreducible if

$$\psi(A) > 0 \quad \Longrightarrow \quad \mathbb{E}_x[\eta_A] > 0 \quad (\forall x \in E),$$

where η_A is the occupation time defined by

$$\eta_A = \int_0^\infty 1_{\{X_t \in A\}} \, \mathrm{d}t.$$

A simple sufficient condition for ψ-irreducibility is

$$\psi(A) > 0 \quad \Longrightarrow \quad P_t(x, A) =: \mathbb{P}_x(X_t \in A) > 0 \quad (\forall x \in E, t \geq T)$$

for some $T > 0$ which is also a sufficient condition for aperiodicity of the Markov process. A measurable set $C \in \mathcal{E}$ is said to be small if there exists $t > 0$, $\epsilon > 0$ and a probability measure ν such that

$$P_t(x, A) \geq \epsilon \nu(A) \quad (\forall x \in C, \forall A \in \mathcal{E}).$$

This Markov process is said to be V-uniformly ergodic if there exists a probability measure Π, a constant $\gamma \in (0, 1)$, $C > 0$ and $V : E \to [1, \infty)$ such that

$$\| P_t(x, \cdot) - \Pi \|_V \leq C V(x) \gamma^t,$$

where

$$\|\nu\|_V := \sup_{\substack{f : E \to \mathbb{R} \\ |f(x)| \leq V(x)}} \left| \int_E f(x) \nu(\mathrm{d}x) \right|.$$

A simple Foster–Lyapunov-type drift condition was established by [18]. By using their results the following can be proved.

THEOREM C.1. *The Markov process S^{B} is irreducible, aperiodic and any compact set is a small set. Moreover, it is V-uniformly ergodic for $V(x) = 1 + x^2$.*

THEOREM C.2. *The Markov process $\mathcal{T}$ is irreducible, aperiodic and any compact set is a small set. Moreover, it is V-uniformly ergodic for some $e^{|x|} \leq V(x) \leq 2e^{|x|}$.*

Since the process $\mathcal{T}$ only changes the sign of the process in each jump time, by Itô's formula, it satisfies that

$$\mathcal{T}_t^2 - \mathcal{T}_0^2 = 2 \int_0^t \mathcal{T}_s \, \mathrm{d}s, \quad \text{and} \quad |\mathcal{T}_t| - |\mathcal{T}_0| = \int_0^t \operatorname{sgn}(\mathcal{T}_s) \, \mathrm{d}s, \tag{C.1}$$

where $\operatorname{sgn}(x)$ is the sign of $x \in \mathbb{R}$. Here, in order to apply Itô's formula for the latter, first apply the formula to $f_\epsilon(x) = (\epsilon + x^2)^{1/2}$ and then take the limit $\epsilon \to 0$. Moreover, the following result summarizes some properties of the covariance kernel of $\mathcal{T}$. See also Figure 1.

PROPOSITION C.3. *The covariance function $K(s, t) = \mathbb{E}[\mathcal{T}_s \mathcal{T}_t]$ of $\mathcal{T}$ satisfies*

$$\int_0^\infty K(s, 0) \, \mathrm{d}s = 0 \tag{C.2}$$

and

$$\partial_t K(t, 0)|_{t=0} = -4\phi(0) = -2\sqrt{\frac{2}{\pi}}, \qquad \partial_t^2 K(t, 0)|_{t=0} = 1. \tag{C.3}$$

C.1. Proof of Theorem C.1. Construct $\mathcal{T}$ and S^{B} as in the above. We also set $\sigma_0 = 0$ and $W_0 = S_0^{\mathrm{B}}$. First, we prove irreducibility and aperiodicity of the Markov process. For $K > 0$, let ν_K be the Lebesgue measure restricted to$[-K, K]$. Consider an event

$$B_T = \{\omega \in \Omega : R((0, T] \times \mathbb{R}) = 1, N(C_T) = 0\},$$

where $C_T = (0, T] \times [0, |x| + |W_1| + T]$. On the event, since $R((0, T] \times \mathbb{R}) = 1$ there is a single refreshment jump σ_1 until $T > 0$. Recall that in each interval $[\sigma_i, \sigma_{i+1})$, the process S^{B} has the same behavior as that of $\mathcal{T}$ with $\mathcal{T}_{\sigma_i} = W_i$. Therefore, by (C.1), we have

$$\omega \in B_T \implies |S_t^{\mathrm{B}}| \leq \begin{cases} |x| + t & \text{if } t < \sigma_1, \\ W_1 + t & \text{if } \sigma_1 \leq t \leq T, \end{cases} \implies \sup_{t \leq T} |S_t^{\mathrm{B}}| \leq |x| + |W_1| + T.$$

Therefore, on the event B_T, the number of jumps due to N up to time T is

$$\int_{(0,T]\times\mathbb{R}_+} 1_{\{z\le S^{\mathrm{B}}_{s-}\}} N(\mathrm{d}s,\mathrm{d}z) \le N(C_T) = 0.$$

Therefore, except for the refreshment jump time σ_1, S^{B} moves deterministically, and hence

$$\omega \in B_T \quad \Longrightarrow \quad S^{\mathrm{B}}_t = \begin{cases} x+t & \text{if } t<\sigma_1, \\ W_1 + (t-\sigma_1) & \text{if } \sigma_1 \le t \le T. \end{cases}$$

Now we calculate the probability of the event B_T. Since R_d, N and W are independent, if $|x|\le K$, then

$$\begin{aligned}\mathbb{P}_x(B_T) &= \mathbb{P}_x\big(R\big((0,T]\times\mathbb{R}\big)=1\big)\times\mathbb{P}\big(N(C_T)=0\big) \\ &= \{\rho T e^{-\rho T}\}\times\left\{\int_{\mathbb{R}} e^{-(|x|+|y|+T)T}\phi(y)\,\mathrm{d}y\right\} \\ &\ge \{\rho T e^{-\rho T}\}\times\{c_T e^{-(K+T)T}\},\end{aligned}$$

where $c_T = \int \exp(-T|y|)\phi(y)\,\mathrm{d}y$. On the other hand, for the Markov semigroup $(P_t)_{t\ge 0}$ of S^{B}, we have

$$\begin{aligned}P_T(x,A) &= \mathbb{P}_x\big(S^{\mathrm{B}}_T\in A\big) \ge \mathbb{P}_x\big(S^{\mathrm{B}}_T\in A, B_T\big) \\ &= \mathbb{P}_x\big(W_1+(T-\sigma_1)\in A, B_T\big) \\ &= \mathbb{E}_x\left[\int_A \phi\big(y-(T-\sigma_1)\big)\,\mathrm{d}y, B_T\right] \\ &\ge \inf_{0\le s\le T}\int_{A\cap K}\phi\big(y-(T-s)\big)\,\mathrm{d}y\,\mathbb{P}_x(B_T) \\ &\ge \kappa_T\nu_K(A)\mathbb{P}_x(B_T),\end{aligned}$$

where $\kappa_T = \inf_{0\le s\le T}\inf_{y\in K}\phi(y-(T-s))$. By these estimates, we obtain

$$P_T(x,A)\ge\kappa_T\nu_K(A)\{\rho T e^{-\rho T}\}\times\{c_T e^{-(K+T)T}\}\quad\text{for } x\in[-K,K].$$

Thus, the Markov process is ν_K-irreducible and aperiodic, and any compact set is a small set.

Second, we prove V-uniform ergodicity. We need to check

$$HV(x)\le -\gamma V(x)+b1_C \tag{C.4}$$

for some $\gamma, b>0$, a small set C and a drift function $V:\mathbb{R}\to[1,\infty)$ where H is defined in (2.7). However, by taking $V(x)=1+x^2$, we have

$$\frac{HV(x)}{V(x)} = \frac{2x+\rho(1-x^2)}{1+x^2}\xrightarrow[|x|\to\infty]{}-\rho.$$

Thus, the drift condition is satisfied for $C=[-R,R]$ and $\gamma=\rho/2$ when R is sufficiently large. Thus V-uniform ergodicity follows by Theorem 5.2 of [18].

C.2. Proof of Theorem C.2. Let $K>0$ and consider $x\in[-K,K]$. Let $T=2K+1$, and define

$$B_T = \{\omega\in\Omega : N(\omega;C_T)=N(\omega;D_T)=1\},$$

where $D_T\subset C_T$ are subsets of $\mathbb{R}_+\times\mathbb{R}_+$ such that

$$C_T=(0,T]\times\big[0,|x|+T\big],\qquad D_T=\big[(1-x)^+,T\big]\times[0,1].$$

On the event B_T, the number of jumps until time T is

$$\int_{(0,T]\times\mathbb{R}_+} 1_{\{z\le\mathcal{T}_{s-}\}} N(\mathrm{d}s,\mathrm{d}z) \le N(C_T) = 1$$

since $|\mathcal{T}_t| \le |x| + T$ $(0 \le t \le T)$ by (C.1). Thus the number of jumps is at most 1. On the other hand, if there is no jump, then $\mathcal{T}_t = x + t$ $(0 \le t \le T)$. However, since $(1-x)^+ \le t \Longrightarrow 1 \le x + t = \mathcal{T}_{t-}$ we have

$$\int_{(0,T]\times\mathbb{R}_+} 1_{\{z\le\mathcal{T}_{s-}\}} N(\mathrm{d}s,\mathrm{d}z) \ge N(D_T) = 1.$$

Therefore, there is a single jump until time T. Then, on the event B_T, we have

$$\mathcal{T}_t = \begin{cases} x+t & \text{if } t < \tau_1, \\ -(x+\tau_1) + (t-\tau_1) & \text{if } \tau_1 \le t \le T, \end{cases}$$

and hence, for a Markov semigroup $(P_t)_{t\ge0}$ of $\mathcal{T}$, we have

$$\begin{aligned} P_T(x,A) &= \mathbb{P}_x(\mathcal{T}_T \in A) \ge \mathbb{P}_x(\mathcal{T}_T \in A, B_T) \\ &= \mathbb{P}_x\bigl(-(x+\tau_1) + (T-\tau_1) \in A, B_T\bigr) \\ &= \mathbb{P}_x\bigl(-(x+\tau_1) + (T-\tau_1) \in A | B_T\bigr) \times \mathbb{P}_x(B_T). \end{aligned}$$

We have

$$\begin{aligned} &\mathbb{P}_x\bigl(-(x+\tau_1) + (T-\tau_1) \in A | B_T\bigr) \\ &\quad = \int_{(1-x)^+}^{T} 1_A\bigl(-(x+s) + (T-s)\bigr) \frac{\mathrm{d}s}{T-(1-x)^+} \\ &\quad \ge T^{-1}\,\mathrm{Leb}\bigl(A \cap \bigl[-x-T, T-x-2(1-x)^+\bigr]\bigr), \end{aligned}$$

where Leb is the Lebesgue measure. On the other hand,

$$\begin{aligned} \mathbb{P}_x(B_T) &= \mathbb{P}_x\bigl(N(D_T) = 1\bigr) \times \mathbb{P}\bigl(N\bigl(C_T \cap D_T^c\bigr) = 0\bigr) \\ &= \bigl(T-(1-x)^+\bigr)e^{-(T-(1-x)^+)} \times e^{-(T(|x|+T)-(T-(1-x)^+))} \\ &=: c(T,x). \end{aligned}$$

Since $c(T,x) > 0$ $(x \in \mathbb{R})$, the Markov process is Leb-irreducible and aperiodic since we have $P_T(x,A) > 0$ by taking $T > 0$ sufficiently large. Also, by $c_T := \inf_{x\in[-K,K]} c(T,x) > 0$ we have

$$P_T(x,A) \ge c_T T^{-1}\,\mathrm{Leb}\bigl(A \cap \bigl[K-T, T-K-2(1+K)^+\bigr]\bigr) \quad \bigl(x \in [-K,K]\bigr).$$

Thus any compact set is a small set.

Finally, we prove V-uniform ergodicity. We need to check the drift criterion (C.4) for $\gamma > 0$, a small set C and $V : E \to [1,\infty)$ and G defined in (2.4) in place of H. Construct a continuously differentiable function $V : E \to [1,\infty)$ so that

$$V(x) = \begin{cases} 2\exp(x) - 1, & x > 4, \\ \exp(-x), & x \le 0. \end{cases} \tag{C.5}$$

Then $GV(x) = (2-x)e^x + xe^{-x} \le -V(x)$ for $x > 4$, since $xe^{-x} \le (e^x - 1)e^{-x} \le 1$ for $x > 0$. Also, $GV(x) = -V(x)$ for $x < 0$. Thus the drift condition holds with $V(x)$, $C = [0,4]$ and $\gamma = 1$. Thus the claim follows by Theorem 5.2 of [18].

C.3. Proof of Proposition C.3. By V-uniform ergodicity of the Markov process $\mathcal{T}$, for $s \le t$ and $k \in \mathbb{N}$, we have

$$\left|\mathbb{E}[\mathcal{T}_t^k | \mathcal{T}_s = x] - \int y^k \phi(y)\,\mathrm{d}y\right| \le C_k \gamma^{t-s} V(x) \tag{C.6}$$

for some $C_k > 0$, $\gamma \in (0,1)$ and hence the covariance function has exponential decay property

$$|K(s,t)| = |\mathbb{E}[\mathcal{T}_s \mathbb{E}[\mathcal{T}_t|\mathcal{T}_s]]| \le C_1 \gamma^{t-s} \mathbb{E}[|\mathcal{T}_s| V(\mathcal{T}_s)] = C\gamma^{t-s}$$

for some $C > 0$ since the marginal distribution of $\mathcal{T}$ is the standard normal distribution and using the explicit form of V given by (C.5).

PROOF OF PROPOSITION C.3. By (C.6) with $k = 2$, we have

$$\begin{aligned} 0 &= \lim_{t\to\infty} \mathbb{E}[(\mathcal{T}_t^2 - \mathcal{T}_0^2)\mathcal{T}_0] \\ &= \lim_{t\to\infty} \mathbb{E}\left[\left(\int_0^t 2\mathcal{T}_s\,\mathrm{d}s\right)\mathcal{T}_0\right] = 2\lim_{t\to\infty}\int_0^t K(s,0)\,\mathrm{d}s = 2\int_0^\infty K(s,0)\,\mathrm{d}s. \end{aligned}$$

Hence we have (C.2).

Next we calculate the derivatives of $K(t) := K(t,0)$. By Itô's formula together with the Lebesgue convergence theorem, we have

$$\begin{aligned} h^{-1}(K(t+h) - K(t)) &= h^{-1}\mathbb{E}[(\mathcal{T}_{t+h} - \mathcal{T}_t)\mathcal{T}_0] \\ &= h^{-1}\mathbb{E}\left[\int_0^h (1 - 2(\mathcal{T}_{t+s}^+)^2)\,\mathrm{d}s\mathcal{T}_0\right] \\ &\xrightarrow[h\to 0]{} \mathbb{E}[(1 - 2(\mathcal{T}_t^+)^2)\mathcal{T}_0]. \end{aligned}$$

The first derivative at $t = 0$ is

$$K'(0) = \mathbb{E}[(1 - 2(\mathcal{T}_0^+)^2)\mathcal{T}_0] = -2\int_0^\infty x^3\phi(x)\,\mathrm{d}x = -2\sqrt{\frac{2}{\pi}}.$$

Similarly, the second derivative at $t = 0$ is

$$\begin{aligned} h^{-1}(K'(h) - K'(0)) &= h^{-1}\mathbb{E}[\{(1 - 2(T_h^+)^2)\mathcal{T}_0\} - \{(1 - 2(T_0^+)^2)\mathcal{T}_0\}] \\ &= -h^{-1}\mathbb{E}\left[\int_0^h (4\mathcal{T}_t^+ - 2(\mathcal{T}_t^+)^3)\,\mathrm{d}s\mathcal{T}_0\right] \\ &\xrightarrow[h\to 0]{} -\mathbb{E}[4(\mathcal{T}_0^+)^2 - 2(\mathcal{T}_0^+)^4] \\ &= -\mathbb{E}[2T_0^2 - T_0^4] = 1. \end{aligned}$$

□

APPENDIX D: NON-GAUSSIAN RESULTS

First we show that the process $S^{\mathrm{B},d}$ converges to $S^{\mathrm{B}}_{H,t} := H^{1/2} S^{\mathrm{B}}_{H^{1/2}t}(H^{-1/2}\rho)$. Let $B'^0(\rho)$, $\tilde{C}'^0(\rho)$ and $\nu^0(\rho)$ be the first, modified second and third characteristics of $S^{\mathrm{B}}(\rho)$ (See Section B.1.1). Then the first and modified second characteristics of the process S^{B}_H are given by

$$B'_T = H^{1/2} B'^0_{H^{1/2}T}(H^{-1/2}\rho), \qquad \tilde{C}'_T = H\tilde{C}'^0_{H^{1/2}T}(H^{-1/2}\rho)$$

and the third characteristic is given by

$$g * \nu_T = g(H^{1/2}\cdot) * \nu^0_{H^{1/2}T}(H^{-1/2}\rho).$$

Therefore, by the change of variable formula, we have

$$B'_T = HT - 2\int_0^T \{(S^{\mathrm{B}}_{H,t})^+\}^2\,\mathrm{d}t - \rho\int_0^T S^{\mathrm{B}}_{H,t}\,\mathrm{d}t,$$

$$\tilde{C}'_T = 4\int_0^T \{(S^{\mathrm{B}}_{H,t})^+\}^3\,\mathrm{d}t + \rho\int_0^T \big(H + (S^{\mathrm{B}}_{H,t})^2\big)\,\mathrm{d}t$$

and

$$g * \nu_T = \int_0^T g(-2S^{\mathrm{B}}_{H,t})(S^{\mathrm{B}}_{H,t})^+\,\mathrm{d}t + \rho\int_0^T \int_{\mathbb{R}} \big(g(u - S^{\mathrm{B}}_{H,t})\big)\phi_H(u)\,\mathrm{d}u\,\mathrm{d}t,$$

where ϕ_H is the probability density function of $\mathcal{N}(0, H)$. On the other hand, the process $S^{\mathrm{B},d}$ satisfies

$$S^{\mathrm{B},d}_T = S^{\mathrm{B},d}_0 + \int_0^T \nabla^2\Psi^d(\xi^{\mathrm{B},d}_t)\big[(v^{\mathrm{B},d}_t)^{\otimes 2}\big]\,\mathrm{d}t - 2\int_{(0,T]\times\mathbb{R}_+} S^{\mathrm{B},d}_{t-} 1_{\{z\le S^{\mathrm{B},d}_{t-}\}} N(\mathrm{d}t, \mathrm{d}z)$$
$$+ \rho\int_{(0,T]\times\mathfrak{S}^{d-1}} \big(\langle\nabla\Psi^d(\xi^{\mathrm{B},d}_{t-}), u\rangle - S^{\mathrm{B},d}_{t-}\big) R_d(\mathrm{d}t, \mathrm{d}u),$$

by Itô's formula. The first and modified second characteristics are

$$B'^d_T := \int_0^T \nabla^2\Psi^d(\xi^{\mathrm{B},d}_t)\big[(v^{\mathrm{B},d}_t)^{\otimes 2}\big]\,\mathrm{d}t - 2\int_0^T \{(S^{\mathrm{B},d}_t)^+\}^2\,\mathrm{d}t - \rho\int_0^T S^{\mathrm{B},d}_t\,\mathrm{d}t,$$

$$\tilde{C}'^d_T := 4\int_0^T \{(S^{\mathrm{B},d}_t)^+\}^3\,\mathrm{d}t + \rho\int_0^T \left(\frac{\|\nabla\Psi^d(\xi^{\mathrm{B},d}_t)\|^2}{d} + (S^{\mathrm{B},d}_t)^2\right)\mathrm{d}t,$$

and the third characteristic is

$$g * \nu^d_T := \int g(x)\nu^d_T(\mathrm{d}x) := \int_0^T g(-2S^{\mathrm{B},d}_t)(S^{\mathrm{B},d}_t)^+\,\mathrm{d}t$$
$$+ \rho\int_0^T \int_{\mathfrak{S}^{d-1}} g\big(\langle\nabla\Psi^d(\xi^{\mathrm{B},d}_t), u\rangle - S^{\mathrm{B},d}_t\big)\,\mathrm{d}t\,\psi_d(\mathrm{d}u)$$

for a continuous bounded function g. For the proof of Proposition 2.16, we will apply Theorem IX.3.48 [23] by showing convergences of the characteristics. To show the convergence of $\tilde{C}'^d$, we need the next lemma.

LEMMA D.1. *For $T > 0$, we have*

$$\sup_{0\le t\le T}\left|\frac{\|\nabla\Psi^d(\xi^{\mathrm{B},d}_t)\|^2}{d} - H\right| \xrightarrow[d\to\infty]{} 0. \tag{D.1}$$

PROOF. Let X^d_T be the left-hand side of (D.1). By Itô's formula,

$$\frac{\|\nabla\Psi^d(\xi^{\mathrm{B},d}_T)\|^2}{d} - \frac{\|\nabla\Psi^d(\xi^{\mathrm{B},d}_0)\|^2}{d} = 2d^{-1}\int_0^T \nabla^2\Psi^d(\xi^{\mathrm{B},d}_t)\big[\nabla\Psi^d(\xi^{\mathrm{B},d}_t), v^{\mathrm{B},d}_t\big]\,\mathrm{d}t.$$

Let $M_t := \nabla^2\Psi^d(\xi^{\mathrm{B},d}_t)$, $a_t := \nabla\Psi^d(\xi^{\mathrm{B},d}_t)/\|\nabla\Psi^d(\xi^{\mathrm{B},d}_t)\|$ and $b_t := v^{\mathrm{B},d}_t$. By (2.14), we have

$$\big|M_t[a_t, b_t]\big| = \frac{1}{2}\big|M_t[a_t^{\otimes 2}] + M_t[b_t^{\otimes 2}] - M_t\big[(a_t - b_t)^{\otimes 2}\big]\big| \le 2C.$$

Also, we have a bound

$$\frac{\|\nabla\Psi^d(\xi^{\mathrm{B},d}_t)\|^2}{d} \le X^d_t + H$$

by the triangle inequality. Therefore we have

$$\begin{aligned}X_T^d &\le X_0^d + \sup_{0\le t\le T}\left|\frac{\|\nabla\Psi^d(\xi_t^{\mathrm{B},d})\|^2}{d} - \frac{\|\nabla\Psi^d(\xi_0^{\mathrm{B},d})\|^2}{d}\right| \\ &\le X_0^d + 2d^{-1}\sup_{0\le t\le T}\left|\int_0^t M_s[a_s,b_s]\times\|\nabla\Psi^d(\xi_t^{\mathrm{B},d})\|\,\mathrm{d}s\right| \\ &\le X_0^d + 4CTd^{-1/2}(X_T^d+H)^{1/2} \le X_0^d + 4CTd^{-1/2}(1+X_T^d+H),\end{aligned}$$

where we used $a^{1/2}\le 1+a$ for $a>0$. Hence

$$X_T^d \le (1-4CTd^{-1/2})^{-1}\big(X_0^d + 4CTd^{-1/2}(1+H)\big) \underset{d\to\infty}{\longrightarrow} 0$$

in probability since $X_0^d \longrightarrow 0$ in probability by (2.12). □

Next we show the following lemma to prove the convergence of B'^d.

LEMMA D.2. *For $T>0$, we have*

$$\sup_{0\le t\le T}\big|[\nabla^2\Psi^d(\xi_t^{\mathrm{B},d})][(v_t^{\mathrm{B},d})^{\otimes 2}]\,\mathrm{d}s - H\big| \underset{d\to\infty}{\longrightarrow} 0. \tag{D.2}$$

PROOF. Let $N_R(t)$ and $N_B(t)$ be the number of refreshment jumps and that of bouncy jumps respectively. Since $N_R(T)=R_d((0,T]\times\mathbb{R})$ follows the Poisson distribution with intensity ρT, it is $\mathbb{P}$-tight. Suppose that the interval $[s,t)$ does not include refreshment jump times. Then, by Itô's formula, we have

$$\big||S_t^{\mathrm{B},d}| - |S_s^{\mathrm{B},d}|\big| = \left|\int_s^t \nabla^2\Psi^d(\xi_u^{\mathrm{B},d})[(v_u^{\mathrm{B},d})^{\otimes 2}]\,\mathrm{sgn}(S_u^{\mathrm{B},d})\,\mathrm{d}u\right| \le CT. \tag{D.3}$$

Therefore, if $0=\sigma_0<\sigma_1<\cdots$ are the refreshment jump times, we have a bound

$$\sup_{t\in[0,T]}|S_t^{\mathrm{B},d}| \le CT + \sup_{n=0,\dots,N_R(T)}|S_{\sigma_n}^{\mathrm{B},d}|.$$

The right-hand side is $\mathbb{P}$-tight since $S_{\sigma_n}^{\mathrm{B},d}$ $(n=1,2,\dots)$ has the same law as that of $S_0^{\mathrm{B},d}$, and $N_R(T)$ is $\mathbb{P}$-tight. Thus $B_T^d := \sup_{t\in[0,T]}|S_t^{\mathrm{B},d}|$ is $\mathbb{P}$-tight. By this fact,

$$N_B(T) = \int_{(0,T]\times\mathbb{R}} 1_{\{z\le S_{t-}^{\mathrm{B},d}\}} N(\mathrm{d}s,\,\mathrm{d}z) \le N\big((0,T]\times[0,B_T^d]\big)$$

is also $\mathbb{P}$-tight.

Let X_t^d be the random variable in the absolute value in the left-hand side of (D.2). For $\epsilon>0$, let $D_\epsilon=\{0=t_0<\cdots<t_N\}\subset[0,T]$ be a finite set that includes all refreshment jump times and $\max|t_i-t_{i-1}|<\epsilon$. If the interval $[s,t)$ does not include refreshment jump times, then

$$\begin{aligned}X_t^d - X_s^d &= \int_s^t \nabla^3\Psi^d(\xi_u^{\mathrm{B},d})[(v_u^{\mathrm{B},d})^{\otimes 3}]\,\mathrm{d}u \\ &\quad - \int_{(s,t]\times\mathbb{R}} \nabla^2\Psi(\xi_{u-}^{\mathrm{B},d})[(\kappa^d(x_{u-}^{\mathrm{B},d}))^{\otimes 2} - (v_{u-}^{\mathrm{B},d})^{\otimes 2}]1_{\{z\le S_{u-}^{\mathrm{B},d}\}}N(\mathrm{d}u,\mathrm{d}z).\end{aligned}$$

By (2.14), we have

$$|X_t^d - X_s^d| \le |t-s|\big(C+2CN_B(T)\big).$$

Then we have

$$\sup_{0\le t\le T}|X_t^d| \le \sup_{t\in D}|X_t^d| + \epsilon(C + 2CN_B(T))$$

and the first term in the right-hand side converges to 0 by (2.13) which proves the claim. □

LEMMA D.3. *$S^{B,d}$ converges to S_H^B.*

PROOF. We apply Theorem IX.3.48 [23]. The proof follows the same line as that of Theorem 2.8 and conditions (i)–(iv) of Theorem IX.3.48 directly follow from the argument in the proof of Theorem 2.8. The condition (v) follows from (B.6) with condition (2.12). Conditions [δ_{loc}-D] and 3.49 can be proved in the same line as that of Theorem 2.8. Finally, we need to check conditions [Sup-β'_{loc}], [γ'_{loc}-D] of (vi) which follow from Lemmas D.1 and D.2. □

PROOF OF PROPOSITION 2.16. By stationarity,

$$\begin{aligned}4\rho\mathbb{E}\big[\big(\Psi^d(\xi_{\sigma_1}^{B,d}) - \Psi^d(\xi_{\sigma_0}^{B,d})\big)^2\big] &= 4\rho\mathbb{E}\left[\left\{\int_0^{\sigma_1} S_t^{B,d}\,\mathrm{d}t\right\}^2\right] \\ &= 4\rho\int_0^\infty\int_0^\infty \mathbb{E}\big[1_{\{s,t\le\sigma_1\}}S_s^{B,d}S_t^{B,d}\big]\,\mathrm{d}s\,\mathrm{d}t.\end{aligned}$$

By (2.14) together with Itô's formula for $S^{B,d}$, we have a uniform bound

$$|S_t^{B,d}| \le |S_0^{B,d}| + Ct$$

by (D.3). Thus, for $s \le t$,

$$\begin{aligned}\big|\mathbb{E}\big[1_{\{s,t\le\sigma_1\}}S_s^{B,d}S_t^{B,d}\big]\big| &\le \mathbb{E}\big[1_{\{t\le\sigma_1\}}\big(|S_0^{B,d}| + Ct\big)^2\big] \\ &= \mathbb{P}(t\le\sigma_1)\mathbb{E}\big[\big(|S_0^{B,d}| + Ct\big)^2\big] \\ &\le \mathbb{P}(t\le\sigma_1)2\mathbb{E}\big[|S_0^{B,d}|^2 + (Ct)^2\big] \\ &= e^{-\rho t}2\mathbb{E}\left[\frac{\|\nabla\Psi(\xi_0^{B,d})\|^2}{d} + (Ct)^2\right].\end{aligned}$$

Therefore, by (2.14), this value is bounded above by $\exp(-\rho t)$ times a polynomial of t. Thus by the dominated convergence theorem,

$$4\rho\mathbb{E}\big[\big(\Psi^d(\xi_{\sigma_1}^{B,d}) - \Psi^d(\xi_{\sigma_0}^{B,d})\big)^2\big] \underset{d\to\infty}{\longrightarrow} 4\rho\int_0^\infty\int_0^\infty \mathbb{E}\big[1_{\{s,t\le\sigma_1\}}S_{H,s}^B S_{H,t}^B\big]\,\mathrm{d}s\,\mathrm{d}t.$$

Now we are going to substitute $S_{H,t}^B$ in the right hand side by $H^{1/2}S_{H^{1/2}t}^B(H^{-1/2}\rho)$. For this substitution, the refreshment jump time σ_1 is also changed to $H^{-1/2}\sigma_1$. Therefore, the right-hand side of the above equation equals to

$$\begin{aligned}&4\rho\int_0^\infty\int_0^\infty \mathbb{E}\big[1_{\{s,t\le H^{-1/2}\sigma_1\}}\big(H^{1/2}S_{H^{1/2}s}^B(H^{-1/2}\rho)\big)\big(H^{1/2}S_{H^{1/2}t}^B(H^{-1/2}\rho)\big)\big]\,\mathrm{d}s\,\mathrm{d}t \\ &= 4\rho\int_0^\infty\int_0^\infty \mathbb{E}\big[1_{\{s,t\le\sigma_1\}}S_s^B(H^{-1/2}\rho)S_t^B(H^{-1/2}\rho)\big]\,\mathrm{d}s\,\mathrm{d}t \\ &= 4\rho\int_0^\infty\int_0^\infty e^{-H^{-1/2}\rho\max\{s,t\}}K(s-t,0)\,\mathrm{d}s\,\mathrm{d}t = H^{1/2}\sigma^2\big(H^{-1/2}\rho\big),\end{aligned}$$

where the last equation follows from the change-of-variable formula (B.17). □

Funding. JB is supported by the research programme "Zig-zagging through computational barriers" with project number 016.Vidi.189.043, financed by the Netherlands Organisation for Scientific Research (NWO). KK is supported by JST CREST Grant Number JPMJCR14D7 and JSPS KAKENHI Grant Number JP16K00046. GOR is supported by the EPSRC grants: *ilike* (EP/K014463/1), CoSInES (EP/R034710/1) and Bayes for Health (EP/R018561/1).

REFERENCES

[1] ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2021). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* **31** 2478–2517. MR4332703 https://doi.org/10.1214/20-aap1653

[2] BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the zig–zag process. *Adv. in Appl. Probab.* **49** 791–825. MR3694318 https://doi.org/10.1017/apr.2017.22

[3] BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig–zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

[4] BIERKENS, J., GRAZZI, S., VAN DER MEULEN, F. and SCHAUER, M. (2021). A piecewise deterministic Monte Carlo method for diffusion bridges. *Stat. Comput.* **31** Paper No. 37. MR4249043 https://doi.org/10.1007/s11222-021-10008-8

[5] BIERKENS, J., ROBERTS, G. O. and ZITT, P.-A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** 2266–2301. MR3983339 https://doi.org/10.1214/18-AAP1453

[6] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

[7] BOISBUNON, A. (2012). The class of multivariate spherically symmetric distributions. Technical Report,# 2012-005, Université de Rouen.

[8] BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

[9] BOUGUET, F. and CLOEZ, B. (2018). Fluctuations of the empirical measure of freezing Markov chains. *Electron. J. Probab.* **23** Paper No. 2. MR3751077 https://doi.org/10.1214/17-EJP130

[10] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications*. Springer, Berlin.

[11] COSTA, O. L. V. and DUFOUR, F. (2008). Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **47** 1053–1077. MR2385873 https://doi.org/10.1137/060670109

[12] COTTER, S. L., HOUSE, T. and PAGANI, F. (2020). The nuzz: Numerical zigzag sampling for general models. Available at arXiv:2003.03636.

[13] DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models (with discussion). *J. Roy. Statist. Soc. Ser. B* **46** 353–388. MR0790622

[14] DELIGIANNIDIS, G., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2019). Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.* **47** 1268–1287. MR3911112 https://doi.org/10.1214/18-AOS1714

[15] DELIGIANNIDIS, G., PAULIN, D., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2018). Randomized Hamiltonian Monte Carlo as Scaling Limit of the Bouncy Particle Sampler and Dimension-Free Convergence Rates. ArXiv e-prints. Available at arXiv:1808.04299.

[16] DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 397–423. MR0898502

[17] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York. MR0058896

[18] DOWN, D., MEYN, S. P. and TWEEDIE, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** 1671–1691. MR1379163

[19] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. MR0838085 https://doi.org/10.1002/9780470316658

[20] FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.* **126** 3077–3101. MR3542627 https://doi.org/10.1016/j.spa.2016.04.002

[21] HAHN, M. G. (1978). Central limit theorems in $d[0, 1]$. *Z. Wahrsch. Verw. Gebiete* **44** 89–101. https://doi.org/10.1007/BF00533047

[22] IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. *North-Holland Mathematical Library* **24**. North-Hollan, Amsterdam. MR1011252

[23] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR1943877 https://doi.org/10.1007/978-3-662-05265-5

[24] KAMATANI, K. (2018). Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. *Bernoulli* **24** 3711–3750. MR3788187 https://doi.org/10.3150/17-BEJ976

[25] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2

[26] KURTZ, T. G. (2011). Equivalence of stochastic equations and martingale problems. In *Stochastic Analysis* 2010 113–130. Springer, Heidelberg. MR2789081 https://doi.org/10.1007/978-3-642-15358-7_6

[27] LU, J. and WANG, L. (2020). On explicit L^2-convergence rate estimate for piecewise deterministic Markov processes. ArXiv.

[28] MARCUS, M. B. and ROSEN, J. (2006). *Markov Processes*, *Gaussian Processes*, *and Local Times*. *Cambridge Studies in Advanced Mathematics* **100**. Cambridge Univ. Press, Cambridge. MR2250510 https://doi.org/10.1017/CBO9780511617997

[29] MICHEL, M., KAPFER, S. C. and KRAUTH, W. (2014). Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J*. *Chem*. *Phys*. **140** 054116. https://doi.org/10.1063/1.4863991

[30] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659

[31] PAKMAN, A., GILBOA, D., CARLSON, D. and PANINSKI, L. (2016). Stochastic bouncy particle sampler. ArXiv preprint. Available at arXiv:1609.00770.

[32] PETERS, E. A. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85** 026703.

[33] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

[34] ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann*. *Appl*. *Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

[35] ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist*. *Sci*. **16** 351–367. MR1888450 https://doi.org/10.1214/ss/1015346320

[36] ROBERTS, G. O. and ROSENTHAL, J. S. (2016). Complexity bounds for Markov chain Monte Carlo algorithms via diffusion limits. *J*. *Appl*. *Probab.* **53** 410–420. MR3514287 https://doi.org/10.1017/jpr.2016.9

[37] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2017). Piecewise deterministic markov chain monte carlo. ArXiv preprint. Available at arXiv:1707.05296.

[38] VASDEKIS, G. (2020). On Zig–Zag Extensions and Related Ergodicity Properties. Ph.D. thesis, Univ. Warwick.

[39] WU, C. and ROBERT, C. P. (2020). Coordinate sampler: A non-reversible Gibbs-like MCMC sampler. *Stat*. *Comput*. **30** 721–730. MR4065228 https://doi.org/10.1007/s11222-019-09913-w