The Annals of Applied Probability
2020, Vol. 30, No. 5, 2069–2098
https://doi.org/10.1214/19-AAP1552

GEOMETRIC ERGODICITY OF THE BOUNCY PARTICLE SAMPLER

BY ALAIN DURMUS[1], ARNAUD GUILLIN[2] AND PIERRE MONMARCHÉ[3]

[1] *CMLA, ENS Cachan, alain.durmus@cmla.ens-cachan.fr*
[2] *Laboratoire de Mathématiques Blaise Pascal, Université Clermont-Auvergne, guillin@math.univ-bpclermont.fr*
[3] *LJLL, Sorbonne Université, pierre.monmarche@sorbonne-universite.fr*

The Bouncy Particle Sampler (BPS) is a Monte Carlo Markov chain algorithm to sample from a target density known up to a multiplicative constant. This method is based on a kinetic piecewise deterministic Markov process for which the target measure is invariant. This paper deals with theoretical properties of BPS. First, we establish geometric ergodicity of the associated semi-group under weaker conditions than in (*Ann. Statist.* **47** (2019) 1268–1287) both on the target distribution and the velocity probability distribution. This result is based on a new coupling of the process which gives a quantitative minorization condition and yields more insights on the convergence. In addition, we study on a toy model the dependency of the convergence rates on the dimension of the state space. Finally, we apply our results to the analysis of simulated annealing algorithms based on BPS.

1. Introduction. Markov chain Monte Carlo methods are a core requirement in many applications, for example, in computational statistics [22], machine learning [1], molecular dynamics [7]. These methods are used to get approximate samples from a target distribution denoted π, with density w.r.t.the Lebesgue measure given for all $x \in \mathbb{R}^d$ by

$$\pi(x) = \exp\big(-U(x)\big), \tag{1}$$

for a potential $U : \mathbb{R}^d \to \mathbb{R}$, known up to an additive constant. They rely on the construction of Markov chains which are ergodic with respect to π, see [48].

Whle the first and best-known MCMC methods are based on reversible chains, such as many Metropolis–Hastings type algorithms [34], there has been since the last decade an increasing interest in nonreversible discrete-time processes [4, 12, 38, 42]. Indeed, consider a Markov chains $(X_k)_{k\in\mathbb{N}}$ on the state space $\{1,\dots,n\}$. If $(X_k)_{k\in\mathbb{N}}$ is reversible, for any $n \in \mathbb{N}$, the event $\{X_{n+2} = X_n\}$ has a positive probability, which explains why reversible processes typically used in MCMC show a diffusive behaviour, covering a distance $\sqrt{K}$ after K iterations. This makes the exploration of the space slow and affects the efficiency of the algorithm. One of the first attempt to avoid this diffusive behaviour has been proposed in [40], where the author suggests to modify the transition matrix $\mathbf{M}$ of $(X_k)_{k\in\mathbb{N}}$, reversible with respect to μ, in such way that the obtained transition matrix is nonreversible but still leaves μ invariant. By definition of $\tilde{\mathbf{M}}$, the probability of backtracking is smaller than for $\mathbf{M}$, that is, $\tilde{\mathbf{M}}^2_{i,i} \le \mathbf{M}^2_{i,i}$ for any $i \in \{1,\dots,n\}$. In addition, [40] shows that the asymptotic variance of $\tilde{\mathbf{M}}$ is always smaller than the one of $\mathbf{M}$.

For general state space and in particular in order to sample from π defined by (1), a now popular idea to construct nonreversible Markov chain is based on lifting, see [12] and the references therein. The idea is to extend the state space $\mathbb{R}^d$ and consider a Markov chain $(X_k, Y_k)_{k\in\mathbb{N}}$ on $\mathbb{R}^d \times \mathsf{Y}$, $\mathsf{Y} \subset \mathbb{R}^d$, which admits an invariant distribution for which the first

Received July 2018; revised November 2019.
MSC2020 subject classifications. 60J22, 65C40.
Key words and phrases. Bouncy particle sampler, MCMC, geometric ergodicity, coupling, simulated annealing.

marginal is the probability measure of interest. It turns out that, appropriately scaled, some of these lifted chains converge to continuous-time Markov processes. For instance, the persistent walk on the discrete torus introduced in [12] converges to the integrated telegraph on the continuous torus [38], while the lifted chain defined in [49] for spin models converges to the Zig-zag process [5] (see also the event-chain MC with infinitesimal steps in the physics literature [37, 42]). In these cases, the continuous-time limits belong to the class of velocity jump processes $(X_t, Y_t)_{t\geq 0}$ on $\mathbb{R}^d \times \mathsf{Y}$, $\mathsf{Y} \subset \mathbb{R}^d$, satisfying $X_t = X_0 + \int_0^t Y_s \, \mathrm{d}s$ for all $t \geq 0$ with $(Y_t)_{t\geq 0}$ piecewise-constant on random time intervals. The velocity $(Y_t)_{t\geq 0}$ acts as an instantaneous memory, or inertia, so that $(X_t)_{t\geq 0}$ tends to continue in the same direction for some time instead of backtracking. In addition, these processes may be designed to target a given probability measure defined on $(\mathbb{R}^d \times \mathsf{Y}, \mathcal{B}(\mathbb{R}^d \times \mathsf{Y}))$ of the form

$$\tilde{\pi} = \pi \otimes \mu_{\mathrm{v}}, \tag{2}$$

where μ_{v} is a probability measure on Y, and therefore can be used as MCMC samplers. This kind of dynamics, which are not new [21, 29], have regained a particular interest in the last decade, in two separate fields: stochastic algorithms, as we presented, but also biological modelling, where they model the motion of a bacterium [9, 18, 19] and are sometimes called run-&-tumble processes.

From a numerical point of view, an advantage of these continuous-time processes is that, under appropriate conditions on the potential U, an exact simulation is possible, following a thinning strategy [8, 31, 32]. Therefore, no discretization schemes are needed to approximate the continuous time trajectory, contrary to Langevin diffusions or Hamiltonian dynamics. As a consequence, no Metropolis filter is necessary to preserve the invariance of π, see [14, 41, 45, 47] and the reference therein.

This work deals with the velocity jump process introduced in [39, 42]. Following [8], we refer to it as the Bouncy Particle Sampler (BPS). The aim of this paper is to establish geometric convergence to equilibrium for the BPS in dimension larger than 1. As detailed below, we relax the conditions of [11], in particular we show that any constant refreshment rate is sufficient for thin tail target distributions. The paper is organized as follows. Section 2.2 presents the BPS process and our main results, which are proven in Section 3. Finally, Section 4 is devoted to a discussion on our result and approach. First, in Section 4.1, we give explicit bound for a toy model, paying a particular attention to the dependency on the dimension of the state space in the constants we get. Second, in Section 4.2, we apply our results to study the annealing algorithm based on the BPS, extending the results of [39]. Some technical proofs are postponed to the Supplementary Material [16].

Although the work is restricted to the BPS, our arguments can easily be adapted to other velocity jump processes, such as randomized variants of the BPS. In particular, the coupling argument in Section 3.3 applies as soon as the process admits a refreshment mechanism.

Notation. For all $a, b \in \mathbb{R}$, we denote $a_+ = \max(0, a)$, $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$. Id stands for the identity matrix on $\mathbb{R}^d$.

For all $x, y \in \mathbb{R}^d$, the scalar product between x and y is denoted by $\langle x, y\rangle$ and the Euclidean norm of x by $\|x\|$. We denote by $\mathsf{S}^d = \{v \in \mathbb{R}^d : \|v\| = 1\}$, the d-dimensional sphere with radius 1 and for all $x \in \mathbb{R}^d$, $r > 0$, by $\mathrm{B}(x, r) = \{w \in \mathbb{R}^d : \|w - x\| \leq r\}$ the ball centered in x with radius r. For any d-dimensional matrix M, define by $\|M\| = \sup_{w\in \mathrm{B}(0,1)} \|Mw\|$ the operator norm associated with M.

Denote by $\mathrm{C}(\mathbb{R}^d)$ the set of continuous function from $\mathbb{R}^d$ to $\mathbb{R}$ and for all $k \in \mathbb{N}^*$, $\mathrm{C}^k(\mathbb{R}^d)$ the set of k-times continuously differentiable function from $\mathbb{R}^d \to \mathbb{R}$. Denote for all $k \in \mathbb{N}$, $\mathrm{C}_c^k(\mathbb{R}^d)$ and $\mathrm{C}_b^k(\mathbb{R}^d)$ the set of functions belonging to $\mathrm{C}^k(\mathbb{R}^d)$ with compact support and the set of bounded functions belonging to $\mathrm{C}^k(\mathbb{R}^d)$ respectively. For all function $f : \mathbb{R}^d \to \mathbb{R}$, we denote by ∇f and $\nabla^2 f$, the gradient and the Hessian of f respectively, if they exist. For all function $F : \mathbb{R}^d \to \mathbb{R}^m$ and compact set $\mathsf{K} \subset \mathbb{R}^d$, denote $\|F\|_\infty = \sup_{x\in\mathbb{R}^d} \|F(x)\|$, $\|F\|_{\infty,\mathsf{K}} = \sup_{x\in\mathsf{K}} \|F(x)\|$. We denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-field of and $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $\xi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is called a transference plan between μ and ν if for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$, $\xi(\mathsf{A} \times \mathbb{R}^d) = \mu(\mathsf{A})$ and $\xi(\mathbb{R}^d \times \mathsf{A}) = \nu(\mathsf{A})$. The set of transference plan between μ and ν is denoted $\Gamma(\mu, \nu)$. The random variables X and Y on $\mathbb{R}^d$ are a coupling between μ and ν if the distribution of (X, Y) belongs to $\Gamma(\mu, \nu)$. The total variation norm between μ and ν is defined by

$$\|\mu - \nu\|_{\mathrm{TV}} = 2 \inf_{\xi\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{1}_{\Delta^d_{\mathbb{R}}}(x, y)\, \mathrm{d}\xi(x, y),$$

where $\Delta_{\mathbb{R}^d} = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : x = y\}$. For $V : \mathbb{R}^d \to [1, +\infty)$, define the V-norm between μ and ν by

$$\|\mu - \nu\|_V = \sup\left\{\left|\int_{\mathbb{R}^d} f\, \mathrm{d}\mu - \int_{\mathbb{R}^d} f\, \mathrm{d}\nu\right| : f : \mathbb{R}^d \to \mathbb{R}, \|f/V\|_\infty < 1\right\}.$$

When $V(x) = 1$ for all $x \in \mathbb{R}^d$, the V-norm is simply the total variation norm. For all $\mu \in \mathcal{P}(\mathbb{R}^d)$, define the support of μ by

$$\operatorname{supp}\mu = \overline{\{x \in \mathbb{R}^d : \text{for all open set } \mathsf{U} \ni x, \mu(\mathsf{U}) > 0\}}.$$

In the sequel, we take the convention that $\inf \varnothing = +\infty$.

2. Geometric convergence of the BPS.

2.1. *Presentation of the BPS.* In all this work, we assume that the potential U, given by (1), is continuously differentiable on $\mathbb{R}^d$. Let $\mathsf{Y} \subset \mathbb{R}^d$ be a closed C^∞-submanifold $\mathsf{Y} \subset \mathbb{R}^d$, which is rotation invariant, that is, for any rotation $O \in \mathbb{R}^{d\times d}$, $O\mathsf{Y} = \mathsf{Y}$. The BPS process $(X_t, Y_t)_{t\geq 0}$ associated with U evolves on $(\mathbb{R}^d \times \mathsf{Y}, \mathcal{B}(\mathbb{R}^d \times \mathsf{Y}))$ and is defined as follows.

Consider some initial point $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, and a family of i.i.d. random variables $(E_i, F_i, G_i)_{i\in\mathbb{N}^*}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where for all $i \in \mathbb{N}^*$, E_i, F_i are exponential random variables with parameter 1, G_i is a random variable with a given distribution μ_{v} on $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$, referred to as the refreshment distribution. In addition, for all $i \in \mathbb{N}^*$, E_i, F_i and G_i are independent. Let $\lambda_{\mathrm{r}} > 0$, referred to as the refreshment rate, $(X_0, Y_0) = (x, y)$ and $S_0 = 0$. We define by recursion the jump times of the process and the process itself. Assume that S_n and $(X_t, Y_t)_{t\leq S_n}$ have been defined for $n \geq 0$. Consider

$$\begin{aligned} T^{(1)}_{n+1} &= E_{n+1}/\lambda_{\mathrm{r}}, \\ T^{(2)}_{n+1} &= \inf\left\{t \geq 0 : \int_0^t \langle Y_{S_n}, \nabla U(X_{S_n} + sY_{S_n})\rangle_+ \,\mathrm{d}s \geq F_{n+1}\right\}, \\ T_{n+1} &= T^{(1)}_{n+1} \wedge T^{(2)}_{n+1}. \end{aligned} \tag{3}$$

Set $S_{n+1} = S_n + T_{n+1}$, $(X_t, Y_t) = (X_{S_n} + tY_{S_n}, Y_{S_n})$, for all $t \in [S_n, S_{n+1})$, $X_{S_{n+1}} = X_{S_n} + T_{n+1}Y_{S_n}$ and

$$Y_{S_{n+1}} = \begin{cases} G_{n+1} & \text{if } T_{n+1} = T^{(1)}_{n+1}, \\ \mathrm{R}(X_{S_{n+1}}, Y_{S_n}) & \text{otherwise}, \end{cases}$$

where $\mathrm{R} : \mathbb{R}^{2d} \to \mathbb{R}^d$ is the function given for all $x, y \in \mathbb{R}^d$ by

$$\begin{aligned} \mathrm{R}(x, y) &= y - 2\langle y, \mathrm{n}(\nabla U(x))\rangle \mathrm{n}(\nabla U(x)), \\ \text{where for all } z \in \mathbb{R}^d, \quad \mathrm{n}(z) &= \begin{cases} z/\|z\| & \text{if } z \neq 0, \\ 0 & \text{otherwise}. \end{cases} \end{aligned} \tag{4}$$

Note that for all $(x,y)\in\mathbb{R}^{2d}$ with $\nabla U(x)\neq 0$, $\mathrm{R}(x,y)$ is the reflection of y orthogonal to $\nabla U(x)$ and therefore for all $(x,y)\in\mathbb{R}^{2d}$, $\|\mathrm{R}(x,y)\|=\|y\|$.

If $T_{n+1}=T^{(1)}_{n+1}$, we say that, at time T_{n+1}, the velocity has been refreshed, and we call T_{n+1} a refreshment time. If $T_{n+1}=T^{(2)}_{n+1}$, we say that, at time T_{n+1}, the process has bounced, and we call T_{n+1} a bounce time.

Then, (X_t,Y_t) is defined for all $t<\sup_{n\in\mathbb{N}} S_n$ and we set for all $t\geq \sup_{n\in\mathbb{N}} S_n$, $(X_t,Y_t)=\infty$, where ∞ is a cemetery point.

In fact, it is proven in [15], Proposition 10, that almost surely, $\sup_{n\in\mathbb{N}} S_n=+\infty$. Therefore, almost surely, $(X_t,Y_t)_{t\geq 0}$ is a $(\mathbb{R}^d\times\mathsf{Y})$-valued càdlàgprocess. By [10], Theorem 25.5, the BPS process $(X_t,Y_t)_{t\geq 0}$ defines a strong Markov semi-group $(P_t)_{t\geq 0}$ given for all $(x,y)\in\mathbb{R}^d\times\mathsf{Y}$ and $\mathsf{A}\in\mathcal{B}(\mathbb{R}^d\times\mathsf{Y})$ by

$$P_t\big((x,y),\mathsf{A}\big)=\mathbb{P}\big((X_t,Y_t)\in\mathsf{A}\big),$$

where $(X_t,Y_t)_{t\in\mathbb{R}_+}$ is the BPS process started from (x,y).

Consider the following basic assumption.

*A*1. The potential U is twice continuously differentiable, μ_{v} is rotation invariant and $(x,y)\mapsto\|y\|\|\nabla U(x)\|$ is integrable with respect to $\tilde{\pi}$ defined by (2).

It is shown in [15], Corollary 24, and contrary to the popular belief it is quite technical and difficult, that under *A*1, the probability measure $\tilde{\pi}$ defined by (2) is invariant for $(P_t)_{t\geq 0}$, that is, $\tilde{\pi}P_t=\tilde{\pi}$ for all $t\geq 0$.

2.2. *Main results.* For $V:\mathbb{R}^d\times\mathsf{Y}\to[1,+\infty)$, the semi-group $(P_t)_{t\geq 0}$ with invariant measure $\tilde{\pi}$ is said to be V-uniformly geometrically ergodic if there exist $C,\rho>0$ such that for all $t\geq 0$ and all $\mu\in\mathcal{P}(\mathbb{R}^d\times\mathsf{Y})$ with $\mu(V)<+\infty$, it holds

$$\|\mu P_t-\tilde{\pi}\|_V\leq C\mathrm{e}^{-\rho t}\mu(V). \tag{5}$$

We state in this section our main results regarding the V-uniform geometric ergodicity of the BPS.

Our basic assumptions to prove geometric ergodicity are the following.

*A*2.

(i) The potential U is positive and satisfies $\int_{\mathbb{R}^d}\exp(-U(x)/2)\,\mathrm{d}x<+\infty$ and $\lim_{\|x\|\to+\infty}U(x)=+\infty$.

(ii) μ_{v} admits a density w.r.t.the Lebesgue measure on $\mathbb{R}^d$ or there exists $r_0>0$ such that $\mu_{\mathrm{v}}(r_0\mathsf{S}^d)>0$.

Here, we establish practical conditions on the potential U, μ_{v} and Y implying that $(P_t)_{t\geq 0}$ is V-uniformly geometrically ergodicity. In fact, these conditions are derived from a more general result. However, since its assumptions and statement may seem very intricate, for the sake of clarity we have decided to give this result after its corollaries.

Consider the following alternative conditions, which will be used in the case where Y is bounded.

*A*3. The potential U satisfies

$$\lim_{\|x\|\to+\infty}\|\nabla U(x)\|=\infty,\qquad \sup_{x\in\mathbb{R}^2}\|\nabla^2 U(x)\|<\infty.$$

*A*4. There exists $\varsigma \in (0,1)$ such that

$$\liminf_{\|x\|\to+\infty} \{\|\nabla U(x)\|/U^{1-\varsigma}(x)\} > 0,$$

$$\limsup_{\|x\|\to+\infty} \{\|\nabla U(x)\|/U^{1-\varsigma/2}(x)\} < +\infty,$$

$$\limsup_{\|x\|\to+\infty} \{\|\nabla^2 U(x)\|/U^{1-\varsigma}(x)\} < +\infty.$$

*A*5. The potential U satisfies $\lim_{\|x\|\to+\infty} \|\nabla^2 U(x)\|/\|\nabla U(x)\| = 0$ and there exists $\varsigma \in (0,1)$ such that

$$\liminf_{\|x\|\to+\infty} \|\nabla U(x)\|/U^{1-\varsigma}(x) > 0 \quad \text{and} \quad \lim_{\|x\|\to+\infty} \|\nabla U(x)\|/U^{2(1-\varsigma)}(x) = 0.$$

Note that *A*5 is similar to *A*4 but these two conditions are different: none of them implies the other. Indeed, on $\mathbb{R}^2$, consider $U(x_1, x_2) = (1 + |x_1|^2)^{\alpha/2} + (1 + |x_2|^2)^{\beta/2}$ for some $\alpha, \beta > 1$. Then for all $(x_1, x_2) \in \mathbb{R}^2$, we have

$$\nabla U(x) = \left[\alpha x_1 (1 + x_1^2)^{\alpha/2-1}, \beta x_2 (1 + x_2^2)^{\beta/2-1}\right]^{\mathrm{T}},$$

$$\nabla^2 U(x) = \begin{pmatrix} F(\alpha, x_1) & 0 \\ 0 & F(\beta, x_2) \end{pmatrix},$$

$$\text{where} \quad F(\alpha, x_1) = \alpha(1 + x_1^2)^{\alpha/2-1} + 2\alpha x_1^2 (\alpha/2 - 1)(1 + x_1^2)^{\alpha/2-2}.$$

In that case *A*4 is satisfied if and only if $[(\alpha \vee \beta)/2, \alpha \wedge \beta] \neq \varnothing$, while *A*5 is satisfied if and only if $[2(\alpha \vee \beta)/(1+\alpha \vee \beta), \alpha \wedge \beta] \neq \varnothing$, chosing in both cases $\varsigma^{-1} > 1$ in the corresponding interval. In particular, if both $\alpha, \beta \geq 2$, then *A*5 is satisfied, but *A*4 may not (if $\alpha > 2\beta$ for instance). On the contrary if, say, $\alpha = 4/3$ and $\beta \in (1, 8/7)$, then *A*4 holds while *A*5 does not.

THEOREM 1. *Assume A*1, *A*2, Y *is bounded and either A*3, *A*4 *or A*5. *In the case where A*3 *holds, set* $\varsigma = 1$. *Then, for any refreshment rate* $\lambda_{\mathrm{r}} > 0$, *there exists* $\kappa \in (0, 1]$ *such that* $(P_t)_{t\geq 0}$ *is V-uniformly geometrically ergodic with* $V : \mathbb{R}^d \times \mathsf{Y} \to [1, +\infty)$ *given for all* $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ *by* $V(x, y) = \exp(\kappa U^{\varsigma}(x))$.

PROOF. The proof is postponed to Section 3.5. □

Note that *A*3, *A*4 and *A*5 all require that $\lim_{\|x\|\to+\infty} \|\nabla U(x)\| = +\infty$. We consider now the case where $\liminf_{\|x\|\to+\infty} \|\nabla U(x)\| < +\infty$ possibly.

*A*6. The potential U satisfies

$$\liminf_{\|x\|\to+\infty} \|\nabla U(x)\| > 0 \quad \text{and} \quad \lim_{\|x\|\to+\infty} \|\nabla^2 U(x)\| = 0.$$

THEOREM 2. *Assume A*1, *A*2, *A*6 *and* Y *is bounded. Then, there exists* $\lambda_0 > 0$ *such that, if* $\lambda_{\mathrm{r}} \in (0, \lambda_0]$, $(P_t)_{t\geq 0}$ *is V-uniformly geometrically ergodic with* $V : \mathbb{R}^d \times \mathsf{Y} \to [1, +\infty)$ *given for all* $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ *by* $V(x, y) = \exp(\kappa U(x))$, *for* $\kappa \in (0, 1]$.

PROOF. The proof is postponed to Section 3.6. □

Note that contrary to the setting of Theorem 1, the result of Theorem 2 requires that the refreshment rate λ_{r} is sufficiently small for the BPS to be V-uniformly geometrically ergodic.

We now turn to the case where Y is unbounded. Indeed, this case is interesting from the numerical experiments conducted in [8], Section 4.3, which shows that the choice of $\mathsf{Y} = \mathbb{R}^d$ and μ_{v} being the d-dimensional Gaussian distribution appears to be better and less sensitive to the choice of the refreshment rate λ_{r} compared to $\mathsf{Y} = \mathsf{S}^d$ and the uniform distribution on this set.

In the case where Y is unbounded, *A*4 must be strengthened as follows.

*A*7. There exists $\varsigma \in (0, 1)$ such that

$$\liminf_{\|x\|\to+\infty} \{\|\nabla U(x)\| / U^{1-\varsigma}(x)\} > 0,$$

$$\limsup_{\|x\|\to+\infty} \{\|\nabla U(x)\| / U^{1-\varsigma}(x)\} < +\infty,$$

$$\limsup_{\|x\|\to+\infty} \{\|\nabla^2 U(x)\| / U^{1-2\varsigma}(x)\} < +\infty.$$

*A*7 (and therefore *A*4) holds when U is a perturbation of an α-homogeneous function:

PROPOSITION 3. *Let $\alpha \in (1, +\infty)$ and assume that $U = U_1 + U_2$ with $U_1, U_2 \in \mathrm{C}^2(\mathbb{R}^d)$ satisfying*:

- *U_1 is α-homogeneous: for all $t \geq 1$ and $x \in \mathbb{R}^d$ with $\|x\| \geq 1$,*

$$U_1(tx) = t^\alpha U_1(x) \quad \textit{and} \quad \lim_{\|x\|\to+\infty} U_1(x) = +\infty.$$

-

$$\limsup_{\|x\|\to+\infty} \{U_2(x)/\|x\|^\alpha + \|\nabla U_2(x)\|/\|x\|^{\alpha-1} + \|\nabla^2 U_2(x)\|/\|x\|^{\alpha-2}\} = 0.$$

Then A7 holds with $\varsigma = 1/\alpha$.

PROOF. The proof is postponed to Section S1.1 in the Supplementary Material [16]. □

This class of potentials is considered in [27], Theorem 4.6, which shows that the random walk metropolis algorithm is geometrically ergodic for target distributions π associated to a potential belonging to this class.

THEOREM 4. *Assume A1, A2, A7 and μ_{v} admits a Gaussian moment: there exists $\eta > 0$ such that $\int_{\mathsf{Y}} \mathrm{e}^{\eta\|y\|^2} \mu_{\mathrm{v}}(\mathrm{d}y) < +\infty$. Then, for any refreshment rate $\lambda_{\mathrm{r}} > 0$, there exists $\kappa \in (0, 1]$ such that $(P_t)_{t\geq 0}$ is V-uniformly geometrically ergodic with $V : \mathbb{R}^d \times \mathsf{Y} \to [1, +\infty)$ given for all $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ by $V(x, y) = \exp(\kappa U^\varsigma(x)) + \exp(\eta\|y\|^2)$.*

PROOF. The proof is postponed to Section 3.7. □

We now compare our results to the ones established by [11]. First, their results deal only with the case where $\mathsf{Y} = \mathsf{S}^d$ and μ_{v} is the uniform distribution on S^d, while our work can be applied to much broader cases. We discuss in the following our main contributions compared to [11] in the case where Y is bounded. The basic assumptions of [11] are the following: (i) $\nabla^2 U$ is locally Lipschitz; (ii) $\int_{\mathbb{R}^d} \|\nabla U(x)\| \, \mathrm{d}\pi(x) < +\infty$; (iii) $\liminf_{\|x\|\to+\infty}\{\mathrm{e}^{U(x)/2}/\|\nabla U(x)\|^{1/2}\} > 0$;

$$\text{(iv)} \quad \inf_{(x,v)\in\mathbb{R}^d\times\mathsf{S}^d} \frac{\mathrm{e}^{U(x)/2}}{\{\langle\nabla U(x), v\rangle_+ \Lambda_{\mathrm{ref}}\}^{1/2}} > 0,$$

where $\Lambda_{\mathrm{ref}} : \mathbb{R}^d \to \mathbb{R}_+$ is a function chosen in the results. These conditions are similar to $A1$ and $A2$ in our work. We now give the results obtained by [11] in detail in order to highlight the differences with the present work. Apart from the CLT which is a consequence of the others, there are three main results in [11] for the geometric ergodicity of the BPS. The first one, concerning regular tail distributions ([11], Theorem 3.1), establishes that the BPS process as defined at the beginning of Section 2.1 is V-geometrically ergodic if $\Lambda_{\mathrm{ref}} = \lambda_{\mathrm{r}}$ and one of the following conditions holds:

(A) $\liminf_{\|x\|\to+\infty} \|\nabla U(x)\| = +\infty$, $\limsup_{\|x\|\to+\infty} \|\nabla^2 U(x)\| < +\infty$ and $\lambda_{\mathrm{r}} > C_1$ for some constant $C_1 > 0$.

(B) $\liminf_{\|x\|\to+\infty} \|\nabla U(x)\| > 0$, $\lim_{\|x\|\to+\infty} \|\nabla^2 U(x)\| = 0$[1] and $\lambda_{\mathrm{r}} < C_2$ for some constant $C_2 > 0$.

Note that Theorem 1 applied with $A3$ generalizes [11], Theorem 3.1(A) since no condition on λ_{r} is required, which is nice in practice. In addition, Theorem 1 can be applied with other conditions than $A3$ that is, $A4$ and $A5$, which yields new results. Also, Theorem 2 is similar to [11], Theorem 3.1(B), except that, as stated before, it holds with more general choices for Y.

The second result of [11] studies, in the case of thin tail distributions, the BPS process where λ_{r} is replaced by $\Lambda_{\mathrm{ref}} : \mathbb{R}^d \to \mathbb{R}_+$ defined for any $x \in \mathbb{R}^d$ by $\lambda_{\mathrm{r}} + \|\nabla U(x)\| / \max(1, \|x\|^\epsilon)$ for some $\epsilon > 0$. Then, under the conditions that

$$\lim_{\|x\|\to+\infty} \|\nabla U(x)\| / \|x\| = +\infty, \qquad \lim_{\|x\|\to+\infty} \{\|\nabla^2 U(x)\| \|x\|^\epsilon / \|\nabla U(x)\|\} = 0,$$

[11], Theorem 3.2, shows that the BPS with refreshment rate Λ_{ref} is V-geometrically ergodic. The use of a nonconstant, unbounded refreshment rate is motivated in [11] by the fact that [11], Theorem 3.1 (the result with constant rate) does not apply to potentials equivalent at infinity to $\|x\|^\alpha$, $\alpha > 2$. For instance, the case of the Bayesian logistic regression presented in [11], Example 2, for which

$$U(x) = \sum_{i=1}^{d} g(x_k) + \sum_{i=1}^{n_l} \big(-b_i \langle c_i, x\rangle + \log\big(1 + \mathrm{e}^{\langle c_i, x\rangle}\big)\big), \tag{6}$$

with $y_i \in \{0, 1\}$ and $c_i \in \mathbb{R}^d$ for all $i \in \{1, \ldots, n_l\}$, $n_l \in \mathbb{N}^*$ is the number of data points, and $g(u) = (1 + u^2/\sigma^2)^{\beta/2}$ for some parameters $\sigma > 0$ and $\beta > 2$, is covered by [11], Theorem 3.2, but not [11], Theorem 3.1. Following the results of [11], one would use a nonconstant, unbounded refreshment rate in that practical case. However, first, from a computational point of view, this kind of refreshment rate function may be problematic when there is no simple thinning method to sample the refreshment times exactly. Even when a thinning method is available, the cost of each jump is increased since ∇U has to be computed when a refreshment is proposed. Moreover, at least for $d = 1$ (see [3]), increasing the refreshment rate—hence the amount of randomness in the system and its diffusive behaviour—increases the asymptotic variance. For these reasons, it was an important question to understand whether the use of a nonconstant, unbounded refreshment rate in [11] was a practical necessity or a technical restriction in the theoretical study. Although the assumptions of Theorem 1 are slightly more restrictive than the conditions of [11], Theorem 3.2, our results show that a constant refreshment (with any positive value) is in fact sufficient for a large class of thin tail distributions, including the logistic regression case (6) or more generally the cases where U behaves at infinity like $\|x\|^\alpha$ for any $\alpha > 1$ (from Theorem 1 with $A4$ thanks to Proposition 3).

[1] In the statement of the Theorem, the authors claim that $\limsup_{\|x\|\to+\infty} \|\nabla^2 U(x)\| < +\infty$ but a careful reading of the proof shows that $\lim_{\|x\|\to+\infty} \|\nabla^2 U(x)\| = 0$ is necessary.

Finally, [11], Theorem 3.3, deals with thick tail distributions. It consists in applying smooth bijective parametrizations of the space proposed by [28] to get geometric ergodicity of Metropolis–Hastings algorithms for thick tail distributions by transforming the target into a thin tail one. It is in fact a general trick that could also be applied in combination of our results.

As noticed before, Theorem 1, Theorem 2 and Theorem 4 ensue from more general results, which holds under the following assumption.

*A*8. There exist some positive functions $H \in \mathrm{C}(\mathbb{R}_+)$, $\psi \in \mathrm{C}^2(\mathbb{R})$, $\ell \in \mathrm{C}^1(\mathbb{R}^d)$, and some constants $R, r, \delta > 0$, $c_i > 0$ for $i = 1, \dots, 4$ satisfying the following conditions.

(i) *Conditions on U*. The function $\bar{U}$, defined by $\bar{U} = \psi \circ U$, satisfies

$$\lim_{\|x\| \to +\infty} \bar{U}(x) = +\infty, \tag{7}$$
$$\int_{\mathbb{R}^d} \exp(\bar{U}(x) - U(x))\,\mathrm{d}x < +\infty,$$

$$\sup_{x \in \mathbb{R}^d} \{\exp(-\bar{U}(x)/4)(\|\nabla \bar{U}(x)\| + \|\nabla^2 \bar{U}(x)\|)\} < +\infty, \tag{8}$$

and for all $x \in \mathbb{R}^d$ with $\|x\| > R$,

$$\|\nabla \bar{U}(x)\|\ell(x) \geq c_1, \qquad \ell(x) \leq c_2, \qquad \|\nabla U(x)\|\ell(x)/\|\nabla \bar{U}(x)\| \geq c_3. \tag{9}$$

(ii) *Conditions on μ_{v}*.

$$\int_{\mathsf{Y}} \mathrm{e}^{H(\|y\|)} \mu_{\mathrm{v}}(\mathrm{d}y) < \infty,$$
$$\sup_{y \in \mathsf{Y}} \{\mathrm{e}^{-H(\|y\|)/2}\|y\|^2\} < \infty,$$
$$\int_{\mathsf{Y}} \mathbb{1}_{[r,+\infty)}(y_1)\mu_{\mathrm{v}}(\mathrm{d}y) \geq \frac{\delta}{2}.$$

(iii) *Conditions on U and μ_{v}*. For $x \in \mathbb{R}^d$, define

$$\mathsf{A}_x = \{y \in \mathsf{Y} : H(\|y\|) \leq 3\bar{U}(x)\}. \tag{10}$$

Assume that

$$\lim_{\|x\| \to +\infty} \Big[\|\nabla \ell(x)\| \Big\{1 \vee \sup_{y \in \mathsf{A}_x} \|y\|\Big\}\Big] = 0, \tag{11}$$

and for all $x \in \mathbb{R}^d$ with $\|x\| > R$,

$$\|\nabla^2 \bar{U}(x)\|\ell(x)\Big\{\sup_{y \in \mathsf{A}_x} \|y\|^2\Big\} \leq c_4. \tag{12}$$

THEOREM 5. *Assume A1–A2–A8. Assume in addition that the following inequalities hold*:

$$\begin{aligned}&[16\lambda_{\mathrm{r}}c_2/(rc_1)] \vee [64c_4c_2/(rc_1)^2] \\ &\quad \leq [(1/3) \wedge \{\lambda_{\mathrm{r}}\delta r c_1/(16c_4)\}][\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}}\delta c_3/(100rc_1)\}^{1/2}].\end{aligned} \tag{13}$$

Then there exists $\kappa \in (0,1]$ given below by (33), such that $(P_t)_{t \geq 0}$ is V-uniformly geometrically ergodic with V given for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$ by $V(x,y) = \exp(\kappa \bar{U}(x)) + \exp(H(\|y\|))$.

PROOF. The proof is postponed to Section 3.4. □

REMARK 6. Note that, under $A8$, (13) is implied by either one of the two following additional assumptions:

(a) $\lim_{\|x\|\to+\infty} \|\nabla\bar{U}(x)\| = +\infty$;
(b) $\lim_{\|x\|\to+\infty} \ell(x) = 0$;
(c) $\lim_{\|x\|\to+\infty} \|\nabla U(x)\|\ell(x)/\|\nabla\bar{U}(x)\| = +\infty$.

Indeed, if a holds, then c_1 can be chosen as large as necessary while c_2, c_4, c_3 can be held fixed so that (13) is satisfied. If b holds, then c_2 can be chosen as small as necessary while c_1, c_3, c_4 can be held fixed. Finally if c holds, then c_3 can be chosen as large as necessary while c_1, c_2, c_4 can be held fixed.

Note that if $(P_t)_{t\geq 0}$ is V-uniformly geometrically ergodic then, by [20], Theorem 4.4, a functional central limit theorem (FCLT) holds. Let $g : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}$ satisfying for all $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, $|g|^2 \leq CV$ for some $C > 0$. Let $(X_t, Y_t)_{t\geq 0}$ be a BPS process with initial distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^d \times \mathsf{Y})$, satisfying $\mu_0(V) < +\infty$. For $t \geq 0$ and $n \in \mathbb{N}_*$, define

$$G_t^n = \frac{1}{\sqrt{n}} \int_0^{nt} \big(g(X_s, Y_s) - \tilde{\pi}(g)\big)\,\mathrm{d}s.$$

Then, there exists $\sigma_g \geq 0$ such that the sequence of processes $\{(G_t^n)_{t\geq 0}, n \in \mathbb{N}\}$ converges as $n \to \infty$ toward $(\sigma_g B_t)_{t\geq 0}$ in the Skorokhod space, where $(B_t)_{t\geq 0}$ is a standard Brownian motion. It is also possible to consider moderate deviation [13, 23] or large deviation principle [30, 50].

3. Proofs of the main results. For the proof Theorem 5, we follow the Meyn and Tweedie approach, based upon two ingredients: a Foster–Lyapunov drift and a local Doeblin condition on compact sets. This section is organized as follows. Before showing the Foster–Lyapunov drift in Section 3.2, we introduce the generator of the BPS in Section 3.1. Then in Section 3.3, we show that under appropriate conditions, the BPS satisfies a local Doeblin condition on compact sets. Contrary to the previous works [6, 11, 39], this result is obtained in the case where μ_{v} has a density with respect to the Lebesgue measure by a direct coupling. With these two elements in hand, Theorem 5 is proven in 3.4. The proofs of Theorem 1, Theorem 2 and Theorem 4 are given in Section 3.5, Section 3.6 and Section 3.7.

3.1. *Generator of the BPS.* The BPS process belongs to the class of piecewise deterministic Markov processes (PDMP). Indeed, consider the ordinary differential equation on $\mathbb{R}^{2d}$

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} y_t \\ 0 \end{pmatrix}, \tag{14}$$

and define for all $t \geq 0$, the map $\phi_t : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ given for all $(x, y) \in \mathbb{R}^{2d}$ by

$$\phi_t(x, y) = (x + ty, y). \tag{15}$$

The family $(\phi_t)_{t\in\mathbb{R}_+}$ is referred to as the flow of diffeomorphisms associated with (14) that is, for all $(x, y) \in \mathbb{R}^{2d}$, $t \mapsto \phi_t(x, y)$ is solution of (14) started at (x, y) and for all $t \geq 0$, $(x, y) \mapsto \phi_t(x, y)$ is a C^∞-diffeomorphism. In addition to the deterministic flow $(\phi_t)_{t\in\mathbb{R}_+}$, the BPS, as a PDMP, is characterized by a function $\lambda : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}_+$, referred to as the jump rate, and a Markov kernel Q on $\mathbb{R}^d \times \mathsf{Y} \times \mathcal{B}(\mathbb{R}^d \times \mathsf{Y})$, defined for all $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d \times \mathsf{Y})$ by

$$\lambda(x, y) = \langle y, \nabla U(x)\rangle_+ + \bar{\lambda},$$

$$Q\big((x, y), \mathsf{A}\big) = \left[\delta_x \otimes \left\{\frac{\langle y, \nabla U(x)\rangle_+}{\lambda(x, y)}\delta_{\mathsf{R}(x,y)} + \frac{\bar{\lambda}}{\lambda(x, y)}\mu_{\mathrm{v}}\right\}\right](\mathsf{A}),$$

where δ_x is the Dirac measure at $x \in \mathbb{R}^d$. With these definitions in mind, we can define a PDMP (in the sense of [10]) $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ which has the same distribution as $(X_t, Y_t)_{t\geq 0}$ on the space $\mathrm{D}(\mathbb{R}_+, \mathbb{R}^d)$ of càdlàg functions $\omega : \mathbb{R}_+ \to \mathbb{R}^d$, endowed with the Skorokhod topology, see [26], Chapter 6.

Consider some initial condition $(x, y) \in \mathbb{R}^{2d}$, a family of i.i.d. random variables $(\tilde{E}_i, \tilde{G}_i, \tilde{W}_i)_{i\geq 1}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ introduced in Section 2.1, where for all $i \geq 1$, $\tilde{E}_i$ is an exponential random variable with parameter 1, $\tilde{G}_i$ is a random variable with distribution μ_{v}, $\tilde{W}_i$ is a uniform random variable and $\tilde{E}_i$, $\tilde{G}_i$ and $\tilde{W}_i$ are independent. Set $(\tilde{X}_0, \tilde{Y}_0) = (x, y)$ and $\tilde{S}_0 = 0$. We define by recursion the jump times of the process and the process itself. For all $n \geq 0$, let

$$\tilde{T}_{n+1} = \inf\left\{t \geq 0 : \int_0^t \lambda\{\phi_s(\tilde{X}_{\tilde{S}_n}, \tilde{Y}_{\tilde{S}_n})\}\,\mathrm{d}s \geq \tilde{E}_{n+1}\right\}.$$

Set $\tilde{S}_{n+1} = \tilde{S}_n + \tilde{T}_{n+1}$, $(\tilde{X}_t, \tilde{Y}_t) = \phi_t(\tilde{X}_{\tilde{S}_n}, \tilde{Y}_{\tilde{S}_n})$ for all $t \in [\tilde{S}_n, \tilde{S}_{n+1})$, $\tilde{X}_{\tilde{S}_{n+1}} = \tilde{X}_{\tilde{S}_n} + \tilde{T}_{n+1}\tilde{Y}_{\tilde{S}_n}$ and

$$\tilde{Y}_{\tilde{S}_{n+1}} = \begin{cases} \tilde{G}_{n+1} & \text{if } \tilde{W}_{n+1} \leq \bar{\lambda}/\lambda(\tilde{X}_{\tilde{S}_{n+1}}, \tilde{Y}_{\tilde{S}_n}), \\ \mathrm{R}(\tilde{X}_{\tilde{S}_{n+1}}, \tilde{Y}_{\tilde{S}_n}) & \text{otherwise,} \end{cases}$$

where R is defined by (4). Thus, $(\tilde{X}_t, \tilde{Y}_t)$ is defined for all $t < \sup_{n\in\mathbb{N}} \tilde{S}_n$ and we set for all $t \geq \sup_{n\in\mathbb{N}} \tilde{S}_n$, $(\tilde{X}_t, \tilde{Y}_t) = \infty$, where ∞ is a cemetery point. Note that for all $n \in \mathbb{N}^*$, $(\tilde{X}_{\tilde{S}_n}, \tilde{Y}_{\tilde{S}_n})$ is distributed according to $Q((\tilde{X}_{\tilde{S}_n}, \tilde{Y}_{\tilde{S}_{n-1}}), \cdot)$.

From [15], Lemma 7, $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ and $(X_t, Y_t)_{t\geq 0}$ have the same distribution (in particular, almost surely $\sup_{n\in\mathbb{N}} \tilde{S}_n = \infty$ and $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ is a $(\mathbb{R}^d \times \mathsf{Y})$-valued càdlàgprocess).

Consider the canonical process associated with the BPS process $(X_t, Y_t)_{t\geq 0}$, still denoted by $(X_t, Y_t)_{t\geq 0}$ on the Skorokhod space $(\mathrm{D}(\mathbb{R}_+, \mathbb{R}^d \times \mathsf{Y}), \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, (\mathbb{P}_{x,y})_{(x,y)\in\mathbb{R}^d\times\mathsf{Y}})$, where $\mathcal{F}$ is the Borel σ-field associated with the Skorokhod topology, $(\mathcal{F}_t)_{t\geq 0}$ is the completed natural filtration, and for all $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, $\mathbb{P}_{x,y}$ is the distribution of the BPS process starting from $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$. For all $t \geq 0$ and Borel measurable functions $f, g : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}$ such that, for all $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, $s \mapsto g((X_s, Y_s))$ is integrable $\mathbb{P}_{(x,y)}$-almost surely, denote

$$M_t^{f,g} = f(X_t, Y_t) - f(X_0, Y_0) - \int_0^t g(X_s, Y_s)\,\mathrm{d}s. \tag{16}$$

The (extended) generator and its domain $(\mathcal{A}, \mathrm{D}(\mathcal{A}))$ associated with the semi-group $(P_t)_{t\geq 0}$ are defined as follows: $f \in \mathrm{D}(\mathcal{A})$ if there exists a Borel measurable function $g : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}$ such that $(M_t^{f,g})_{t\geq 0}$ is a local martingale under $\mathbb{P}_{(x,y)}$ for all $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ and, for such a function, $\mathcal{A}f = g$. Despite its very formal definition, $(\mathcal{A}, \mathrm{D}(\mathcal{A}))$ associated with $(P_t)_{t\geq 0}$ can be easily described. Indeed, [10], Theorem 26.14, shows that $\mathrm{D}(\mathcal{A}) = \mathsf{E}_1 \cap \mathsf{E}_2$ where

$$\begin{aligned} \mathsf{E}_1 = \{f \in \mathbb{M}(\mathbb{R}^d \times \mathsf{Y}) : t \mapsto f(\phi_t(x, y)) \\ \text{is absolutely continuous on } \mathbb{R}_+ \text{ for all } (x, y) \in \mathbb{R}^{2d}\}, \end{aligned}$$

and E_2 is the set of Borel measurable functions $f : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}$ such that there exists an increasing sequence of $(\mathcal{F}_t)_{t\geq 0}$-stopping time $(\sigma_n)_{n\geq 0}$, such that for all $(x, y) \in \mathbb{R}^{2d}$, $\lim_{n\to+\infty} \sigma_n = +\infty \mathbb{P}_{(x,y)}$-almost surely, and for all $n \in \mathbb{N}^*$,

$$\mathbb{E}_{(x,y)}\left[\sum_{k=1}^{+\infty} \mathbb{1}_{\{S_k \leq \sigma_n\}} |f(X_{S_k}, Y_{S_k}) - f(X_{S_k-}, Y_{S_k-})|\right] < +\infty. \tag{17}$$

Taking for all $n \in \mathbb{N}^*$, $\sigma_n = S_n \wedge n \wedge \upsilon_n$, where $\upsilon_n = \inf\{t \geq 0 : \|X_t\| \geq n\}$, (17) is satisfied for any function $f \in \mathrm{C}(\mathbb{R}^d \times \mathsf{Y})$ such that for all $x \in \mathbb{R}^d$, $\int_{\mathsf{Y}} |f(x, w)|\,\mathrm{d}\mu_{\mathrm{v}}(w) < \infty$.

Then, for all $f \in \mathrm{D}(\mathcal{A})$ and $x, y \in \mathbb{R}^d \times \mathsf{Y}$,

$$\begin{aligned}\mathcal{A}f(x,y) = D_y f(x,y) + (\langle y, \nabla U(x)\rangle)_+ \{f(x, \mathrm{R}(x,y)) - f(x,y)\} \\ + \lambda_{\mathrm{r}}\left\{\int_{\mathsf{Y}} f(x,w)\,\mathrm{d}\mu_{\mathrm{v}}(w) - f(x,y)\right\},\end{aligned} \tag{18}$$

where

$$D_y f(x,y) = \begin{cases} \lim\limits_{t\to 0} \dfrac{f(\varphi_t(x,y)) - f(x,y)}{t} & \text{if this limit exists,} \\ 0 & \text{otherwise.}\end{cases}$$

In particular, if $x \mapsto f(x,y)$ is C^1 for all $y \in \mathsf{Y}$, then

$$\begin{aligned}\mathcal{A}f(x,y) = \langle y, \nabla f(x,y)\rangle + (\langle y, \nabla U(x)\rangle)_+ \{f(x, \mathrm{R}(x,y)) - f(x,y)\} \\ + \lambda_{\mathrm{r}}\left\{\int_{\mathsf{Y}} f(x,w)\,\mathrm{d}\mu_{\mathrm{v}}(w) - f(x,y)\right\}.\end{aligned} \tag{19}$$

3.2. *Foster–Lyapunov drift condition.* For $a, b, c \in \mathbb{R}_+$, $a \le b \le c$, $c - b \le b - a \le a$ and $\varepsilon \in (0,1]$ consider a nondecreasing continuously differentiable function $\varphi : \mathbb{R}_+ \to [1, +\infty)$ satisfying

$$\begin{aligned}\varphi(s) = 1 \quad \text{if } s \in (-\infty, -2], \\ 1 + a(s+2) - \varepsilon \le \varphi(s) \le 1 + a(s+2) + \varepsilon \quad \text{if } s \in (-2,-1), \\ \varphi(s) = 1 + b + s(b-a) \quad \text{if } s \in [-1, 0], \\ 1 + b + s(c-b) - \varepsilon \le \varphi(s) \le 1 + b + s(c-b) + \varepsilon \quad \text{if } s \in (0,1), \\ \varphi(s) = 1 + c \quad \text{if } s \in [1, +\infty]\end{aligned} \tag{20}$$

and

$$\sup_{s\in[-2,-1]} \varphi'(s) \le a + \varepsilon, \qquad \sup_{s\in[0,1]} \varphi'(s) \le c - b + \varepsilon. \tag{21}$$

In addition for $\kappa \in (0,1]$, under A8, define the Lyapunov function $V : \mathbb{R}^d \times \mathsf{Y} \to [1, +\infty)$ by

$$V(x,y) = \exp(\kappa \bar{U}(x))\varphi\{(2\ell(x)/(rc_1))\langle y, \nabla \bar{U}(x)\rangle\} + \exp(H(\|y\|)). \tag{22}$$

This section is devoted to the proof of a Foster–Lyapunov drift condition for the generator $\mathcal{A}$ given by (19) and the function V defined in (22).

LEMMA 7. *Assume A1–A2–A8 and (13) hold. There exist $a, b, c \in \mathbb{R}_+$, $a \le b \le c$, $c - b \le b - a \le a$, $\varepsilon \in (0,1]$ and $\kappa \in (0,1]$ such that $\mathcal{A}$ given by (19) satisfies a Foster–Lyapunov drift condition with the Lyapunov function V, that is, there exist $A_1, A_2 > 0$ such that, for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$,*

$$\mathcal{A}V(x,y) \le A_1(A_2 - V(x,y)). \tag{23}$$

Inequality (23) means that, away from a given compact set, in average, V tends to decay along a trajectory of the BPS. Before proceeding into the details, let us give a brief explanation on the roles of the different parts of V in this decay. When x has a large norm and $y \notin \mathsf{A}_x$, the leading term of both V and $\mathcal{A}V$ is $\exp(H(\|y\|))$, which appears in $\mathcal{A}V$, thanks to the refreshment operator, with the negative factor $-\lambda_{\mathrm{r}}$. In other words, when the scalar velocity is large, then it will typically decrease at the next refreshment time, so that V will decrease. The main difficulty appears as $y \in \mathsf{A}_x$. The reason why V should decrease in average depends on $\theta(x,y)=\langle y,\nabla\bar U(x)\rangle$: when this is large enough, the process is likely to bounce, which causes $\varphi(\theta)$ to change to $\varphi(-\theta)$, which is smaller, so that V decreases. When θ is negative enough, the deterministic transport leads $\exp(\kappa\bar U)$, hence V, to decrease. Finally, when $|\theta|$ is small, $\varphi(\theta)$ is close to 1, hence is larger than its mean with respect to μ_{v}, so that it can be expected to decrease at the next refreshment time.

Remark that, because of the operator $f\mapsto\int_{\mathsf Y}f(\cdot,w)\,\mathrm{d}\mu_{\mathrm{v}}(w)$, the construction of V at a point (x,y) influences the value of $\mathcal{A}V$ at all points $\{(x,v),\ v\in\mathsf Y\}$. Similarly, the term $f(x,\mathrm{R}(x,y))$ is nonlocal. This yields contradictory constraints: for instance, when θ is large, while the bounce mechanism typically makes $\varphi(\theta)$ decrease, the deterministic transport leads $\exp(\kappa\bar U)$ to increase. Thus, in order for V to decrease in average, we need κ to be small enough. On the contrary, when θ is negative enough, $\exp(\kappa\bar U)$ tends to decrease, but then $\varphi(\theta)$ is below its mean with respect to μ_{v}, so that it is expected to increase at the next refreshment time. Then we would like κ to be large enough. The condition (13) on the c_i's and on λ_{r} ensures that the different constraints are compatible.

PROOF. For ease of notation, we denote in the following for any $(x,y)\in\mathbb{R}^d\times\mathsf Y$ $\theta(x,y)=\langle\nabla\bar U(x),y\rangle$. From (19) and the facts that $\nabla\bar U(x)=\psi'(U(x))\nabla U(x)$ and $\|\mathrm{R}(x,y)\|=\|y\|$, for any $(x,y)\in\mathbb{R}^d\times\mathsf Y$,

$$\mathcal{A}V(x,y)=e^{\kappa\bar U(x)}J(x,y)+\lambda_{\mathrm{r}}\left\{\int_{\mathsf Y}e^{H(\|w\|)}\mu_{\mathrm{v}}(\mathrm{d}w)-e^{H(\|y\|)}\right\},\tag{24}$$

where

$$\begin{aligned}J(x,y)={}&\kappa\theta(x,y)\varphi\{2\ell(x)\theta(x,y)/(rc_1)\}\\&+\big(2/(rc_1)\big)\varphi'\{2\ell(x)\theta(x,y)/(rc_1)\}\\&\times\big[\ell(x)\langle y,\nabla^2\bar U(x)y\rangle+\theta(x,y)\langle\nabla\ell(x),y\rangle\big]\\&+\frac{\|\nabla U(x)\|}{\|\nabla\bar U(x)\|}\{\theta(x,y)\}_+\\&\times\big[\varphi\{-2\ell(x)\theta(x,y)/(rc_1)\}-\varphi\{2\ell(x)\theta(x,y)/(rc_1)\}\big]\\&+\lambda_{\mathrm{r}}\left\{\int_{\mathsf V}\varphi\{\big(2\ell(x)/(rc_1)\big)\langle\nabla\bar U(x),w\rangle\}\,\mathrm{d}\mu_{\mathrm{v}}(w)\right.\\&\left.-\varphi\{2\ell(x)\theta(x,y)/(rc_1)\}\right\}.\end{aligned}\tag{25}$$

The first step of the proof is to show that there exist $A_{1,1},A_{1,2}>0$ such that

$$\mathcal{A}V(x,y)\le -A_{1,1}V(x,y)+A_{1,2}\quad\text{for any }(x,y)\in\mathbb{R}^d\times\mathsf Y,\ y\notin\mathsf A_x,\tag{26}$$

where $\mathsf A_x\subset\mathsf Y$ is defined by (10). In a second step, we show that there exist $A_{2,1},A_{2,2}>0$ such that

$$\mathcal{A}V(x,y)\le -A_{2,1}V(x,y)+A_{2,2}\quad\text{for any }(x,y)\in\mathbb{R}^d\times\mathsf Y,\ y\in\mathsf A_x.\tag{27}$$

Note that if (26) and (27) hold, then the proof is concluded.

PROOF OF (26). Let $(x,y)\in\mathbb{R}^d\times\mathsf Y$, $y\notin\mathsf A_x$. From (25) and the facts that φ is bounded by $1+c$, that $\varphi(-s)-\varphi(s)\le 0$ for any $s\in\mathbb{R}_+$ since φ is nondecreasing, and that $\sup_{s\in\mathbb{R}}\varphi'(s)\le(a+\varepsilon)\vee b\vee((c-b)+\varepsilon)\le 1+c$ since $\varepsilon\le 1$, we have

$$\begin{aligned}J(x,y)\le{}&(1+c)\big[\kappa\|\nabla\bar U(x)\|\|y\|\\&+\big(2/(rc_1)\big)\{\|y\|\|\nabla\ell(x)\|+\ell(x)\|y\|^2\|\nabla^2U(x)\|\}+\lambda_{\mathrm{r}}\big].\end{aligned}\tag{28}$$

By (9) and (11) and since $\ell \in \mathrm{C}^1(\mathbb{R}^d)$, $\|\nabla \ell\|_\infty + \|\ell\|_\infty < \infty$. Therefore plugging (28) in (25) and using (8) and A8(ii), we get

$$
\begin{aligned}
\mathcal{A}V(x,y) &\le C_1(1 \vee \|y\|^2)\exp(5\bar{U}(x)/4) + C_2 - \lambda_{\mathrm{r}} \exp(H(\|y\|)),\\
C_1 &= (1+c)\{(\kappa \|\nabla \bar{U} \mathrm{e}^{-\bar{U}/4}\|_\infty) \vee (2\|\nabla \ell\|_\infty/(rc_1)) \\
&\qquad \vee \lambda_{\mathrm{r}} \vee (2\|\nabla^2 \bar{U}\mathrm{e}^{\bar{U}/4}\|_\infty \|\ell\|_\infty/(rc_1))\} < +\infty,\\
C_2 &= \lambda_{\mathrm{r}} \int_{\mathsf{Y}} \exp(H(\|y\|))\,\mathrm{d}\mu_{\mathrm{v}}(w) < +\infty.
\end{aligned} \tag{29}
$$

Using now A8(ii) and the continuity of H, we get that $C_3 = C_1 \sup_{y \in \mathsf{Y}}(1 \vee \|y\|^2)\mathrm{e}^{-H(\|y\|)/2}$ is finite. Since $y \notin \mathsf{A}_x$, $3\bar{U}(x) \le H(\|y\|)$ and we obtain

$$
\begin{aligned}
\mathcal{A}V(x,y) &\le C_3 \exp(11H(\|y\|)/12) + C_2 - \lambda_{\mathrm{r}} \exp(H(\|y\|))\\
&\le -(\lambda_{\mathrm{r}}/2)\exp(H(\|y\|)) + C_4,\\
C_4 &= C_2 + \sup_{s \in \mathbb{R}_+}\{C_3 \mathrm{e}^{11s/12} - \lambda_{\mathrm{r}} \mathrm{e}^s\}.
\end{aligned}
$$

The proof of (26) follows upon noting that $\kappa \le 1$ and that φ is bounded by $1+c$, so that $V(x,y) \le (2+c)\exp(H(\|y\|))$ if $y \notin \mathsf{A}_x$. $\square$

PROOF OF (27). We show in Lemma 8 below that there exist $a,b,c \in \mathbb{R}_+$, $a \le b \le c$, $\varepsilon \in (0,1]$, $\kappa \in (0,1)$, $R_1 \in \mathbb{R}_+$ and $\eta \in \mathbb{R}_+^*$ such that for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $y \in \mathsf{A}_x$ and $\|x\| \ge R_1$, $J(x,y) < -\eta$. Note that if this result holds, then for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $y \in \mathsf{A}_x$ and $\|x\| \ge R_1$, by (24),

$$
\begin{aligned}
\mathcal{A}V(x,y) &\le -\eta \exp(\kappa \bar{U}(x)) + C_2 - \lambda_{\mathrm{r}} \exp(H(\|y\|))\\
&\le -\{(\eta/(1+c)) \wedge \lambda_{\mathrm{r}}\} V(x,y) + C_2,
\end{aligned} \tag{30}
$$

where C_2 is given by (29) and we have used for the last inequality that φ is bounded by $1+c$. This result concludes the proof of (27) for $\|x\| \ge R_1$. It remains to consider the case $\|x\| \le R_1$.

Since ψ and U are continuous, so is $\bar{U}$, so that there exists M_1 such that for all $x \in \mathrm{B}(0,R_1)$ and $y \in \mathsf{A}_x$, $H(\|y\|) \le M_1$. Since $\sup_{w \in \mathsf{Y}} \|w\|^2 \mathrm{e}^{-H(\|w\|)} < +\infty$ by A8(ii), it follows that there exists M_2 such that for all $x \in \mathrm{B}(0,R_1)$, $\mathsf{A}_x \subset \mathrm{B}(0,M_2)$. Then, using that $\bar{U} \in \mathrm{C}^2(\mathbb{R}^d)$, $\ell \in \mathrm{C}^1(\mathbb{R}^d)$, $H \in \mathrm{C}(\mathbb{R}_+)$ and $\varphi \in \mathrm{C}^1(\mathbb{R})$ we get that there exists C_5, C_6 such that for all $x \in \mathrm{B}(0,R_1)$ and $y \in \mathsf{A}_x$, $\mathcal{A}V(x,y) \le C_5$ and $V(x,y) \le C_6$. Combining this result and (30) concludes the proof of (27). $\square$

Let us now precise the parameters we chose in the definition of V. Set

$$
a = 1 \wedge ([(1/3) \wedge \{\lambda_{\mathrm{r}}\delta r c_1/(16c_4)\}][\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}}\delta c_3/(100rc_1)\}^{1/2}])^{-1}, \tag{31}
$$

$$
b - a = a[(1/3) \wedge \{\lambda_{\mathrm{r}}\delta r c_1/(16c_4)\}], \tag{32}
$$

$$
\begin{aligned}
\kappa &= (b-a)[\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}}\delta c_3/(100rc_1)\}^{1/2}]\\
&= a[(1/3) \wedge \{\lambda_{\mathrm{r}}\delta r c_1/(16c_4)\}][\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}}\delta c_3/(100rc_1)\}^{1/2}],
\end{aligned} \tag{33}
$$

$$
c - b = [\delta \lambda_{\mathrm{r}} a/(4(4c_4/(rc_2) + 2\lambda_{\mathrm{r}}))] \wedge (b-a) \wedge [(b-a)c_3/(4\kappa c_2)] \wedge (\delta b/4), \tag{34}
$$

$$
\varepsilon = (1/2) \wedge (c-b) \wedge (\kappa r c_1/4) \wedge (\lambda_{\mathrm{r}} c_2). \tag{35}
$$

Note that $\kappa \le 1$ and

$$
0 \le c - b \le b - a \le a \le 1. \tag{36}
$$

LEMMA 8. *Assume A1–A2–A8 and (13) hold. Then for $a, b, c, \kappa, \varepsilon \in (0, 1]$, given in (31)–(32)–(34)–(33)–(35) respectively, there exist $\tilde{R}, \eta > 0$ such that for all $x \in \mathbb{R}^d$ with $\|x\| \geq \tilde{R}$ and all $y \in \mathsf{A}_x$, $J(x, y) < -\eta$, where J and φ are defined by (25) and (20) respectively.*

PROOF. In the proof, we first give a bound on J for any $(x, y) \in \mathbb{R}^d$, $y \in \mathsf{A}_x$. Second, denoting again $\theta(x, y) = \langle \nabla \bar{U}(x), y \rangle$ for $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, we distinguish five cases depending on the value of $2\ell(x)\theta(x, y)/(rc_1)$ which determines the contribution of φ and φ' in J.

By (11), there exists $R_1 \in \mathbb{R}_+$ such that for any $(x, y) \in \mathbb{R}^d$, $y \in \mathsf{A}_x$, $\|x\| \geq R_1$,

$$\|\nabla \ell(x)\| \|y\| \leq \varepsilon. \tag{37}$$

From (9), $\|\nabla \bar{U}(x)\| \ell(x) \geq c_1$ for all $x \in \mathbb{R}^d$ with $\|x\| \geq R$. Using A8(ii) and the facts that μ_{v} is rotation invariant and that φ is nondecreasing, bounded by $1 + c$ and equal to 1 on $(-\infty, 2]$, we then have for any $x \in \mathbb{R}^d$ with $\|x\| \geq R$

$$\begin{aligned}
&\int_{\mathsf{V}} \varphi\left\{\frac{2\ell(x)}{rc_1}\langle \nabla \bar{U}(x), w\rangle\right\} \mathrm{d}\mu_{\mathrm{v}}(w) \\
&\quad = \int_{\mathsf{V}} \varphi\left\{\frac{2\ell(x)|\nabla \bar{U}(x)|w_1}{rc_1}\right\} \mathrm{d}\mu_{\mathrm{v}}(w) \\
&\quad \leq \int_{\mathsf{V}} \mathbb{1}_{(-\infty,-r]}(w_1)\,\mathrm{d}\mu_{\mathrm{v}}(w) + (1 + c)\int_{\mathsf{V}} \mathbb{1}_{(-r,+\infty)}(w_1)\,\mathrm{d}\mu_{\mathrm{v}}(w) \\
&\quad \leq 1 + (1 - \delta/2)c.
\end{aligned}$$

Therefore, combining this result, (37), (12) and the fact that φ is nondecreasing so that $\varphi'(s) \geq 0$ for any $s \in \mathbb{R}$, we get, for any $x \in \mathbb{R}^d$ with $\|x\| \geq R_2 = R \vee R_1$ and all $y \in \mathsf{A}_x$,

$$\begin{aligned}
J(x, y) \leq{}& \kappa\theta(x, y)\varphi\{2\ell(x)\theta(x, y)/(rc_1)\} \\
&+ \big(2/(rc_1)\big)\varphi'\{2\ell(x)\theta(x, y)/(rc_1)\}\big[c_4 + |\theta|(x, y)\varepsilon\big] \\
&+ \frac{\|\nabla U(x)\|}{\|\nabla \bar{U}(x)\|}\{\theta(x, y)\}_+\big[\varphi\{-2\ell(x)\theta(x, y)/(rc_1)\} \\
&- \varphi\{2\ell(x)\theta(x, y)/(rc_1)\}\big] \\
&+ \lambda_{\mathrm{r}}\big\{1 + (1 - \delta/2)c - \varphi\{2\ell(x)\theta(x, y)/(rc_1)\}\big\}.
\end{aligned} \tag{38}$$

Let $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, $y \in \mathsf{Y}$, $\|x\| \geq R_2$. We consider now five cases.

Case 1: $2\ell(x)\theta(x, y)/(rc_1) \in (-\infty, -2]$. Since for $s \in (-2, -\infty]$, $\varphi(s) = 1$, (38) reads

$$J(x, y) \leq \kappa\theta(x, y) + (1 - \delta/2)\lambda_{\mathrm{r}}c. \tag{39}$$

Using the facts that $2\ell(x)\theta(x, y)/(rc_1) \in (-\infty, -2]$, that $\ell(z) \leq c_2$ for all $z \in \mathbb{R}^d$ by (9), that $(b - a) \vee (c - b) \leq a$ by (36), that $a \leq rc_1\kappa/(6\lambda_{\mathrm{r}}c_2)$ by (33) and that (13) holds, we get

$$rc_1\kappa/\big(2\ell(x)\big) \geq rc_1\kappa/(2c_2) \geq 3\lambda_{\mathrm{r}}a \geq (1 - \delta/2)\lambda_{\mathrm{r}}c.$$

By this result and (39), we obtain

$$J(x, y) \leq -rc_1\kappa/(2c_2). \tag{40}$$

Case 2: $2\ell(x)\theta(x, y)/(rc_1) \in (-2, -1)$. By (20)–(21), $1 + 2a + sa - \varepsilon \leq \varphi(s) \leq 1 + 2a + sa + \varepsilon$ and $\varphi'(s) \leq a + \varepsilon$ for $s \in (-2, -1)$, so that (38) reads

$$\begin{aligned}
J(x, y) \leq{}& \kappa\theta(x, y)\{1 + 2a + 2a\ell(x)\theta(x, y)/(rc_1) - \varepsilon\} \\
&+ \big(2(a + \varepsilon)/(rc_1)\big)\{c_4 - \varepsilon\theta(x, y)\}
\end{aligned}$$

$$
\begin{aligned}
&+ \lambda_{\mathrm{r}}\{(1-\delta/2)c - 2a - 2a\ell(x)\theta(x,y)/(rc_1) + \varepsilon\} \\
&\le B_0 + B_1\theta(x,y) + 2\ell(x)B_2\theta(x,y)^2/(rc_1) \\
&\le B_0 + (B_1 - 2B_2)\theta(x,y),
\end{aligned}
$$

where we have used that $2\ell(x)\theta(x,y)/(rc_1) \in (-2,-1)$ and that $\ell(x) \le c_2$ by (9), and defined

$$
\begin{aligned}
B_0 &= 2(a+\varepsilon)c_4/(rc_1) + \lambda_{\mathrm{r}}\{(1-\delta/2)c - 2a + \varepsilon\}, \\
B_1 &= \kappa(1+2a-\varepsilon) - 2\lambda_{\mathrm{r}}ac_2/(rc_1) - 2\varepsilon(a+\varepsilon)/(rc_1), \\
B_2 &= \kappa a.
\end{aligned}
$$

First, (35) and (36) ensures that $\varepsilon \le (1/2) \wedge a \wedge (\lambda_{\mathrm{r}}c_2)$, and therefore

$$B_1 - 2B_2 \ge \kappa/2 - 4\lambda_{\mathrm{r}}ac_2/(rc_1) \ge \kappa/4,$$

where we have used that $a \le rc_1\kappa/(16\lambda_{\mathrm{r}}c_2)$ for the last inequality, which is a consequence of (33) and (13). In particular, $B_1 \ge 2B_2$ and using again that $2\ell(x)\theta(x,y)/(rc_1) \in (-2,-1)$ and $\ell(x) \le c_2$ from (9), then

$$J(x,y) \le B_0 + \big(rc_1/(2c_2)\big)(2B_2 - B_1) \le B_0 - rc_1\kappa/(8c_2). \tag{41}$$

Since $\varepsilon \le a \wedge (c-b)$ by (35), $c - b \le b - a$ by (34) and $b - a \le a/3$ by (32), we have $B_0 \le 4ac_4/(rc_1)$. Hence, (41) reads

$$J(x,y) \le 4ac_4/(rc_1) - rc_1\kappa/(8c_2) \le -rc_1\kappa/(16c_2), \tag{42}$$

where we have used (33) and (13) for the last inequality.

Case 3: $2\ell(x)\theta(x,y)/(rc_1) \in [-1,0]$. Using the expression of φ on $[-1,0]$ given by (20), (38) reads

$$
\begin{aligned}
J(x,y) \le{}& \kappa\theta(x,y)\{1 + b + (b-a)2\ell(x)\theta(x,y)/(rc_1)\} \\
&+ \big(2(b-a)/(rc_1)\big)\{c_4 - \theta(x,y)\varepsilon\} \\
&+ \lambda_{\mathrm{r}}\{(1-\delta/2)c - b - 2\ell(x)\theta(x,y)(b-a)/(rc_1)\} \\
\le{}& B_0 + B_1\theta(x,y) + B_2 2\ell(x)/(rc_1)\theta(x,y)^2 \\
\le{}& B_0 + (B_1 - B_2)\theta(x,y),
\end{aligned}
\tag{43}
$$

where we have used that $2\ell(x)\theta(x,y)/(rc_1) \in [-1,0]$ and $\ell(x) \le c_2$ by (9), and defined

$$
\begin{aligned}
B_0 &= 2(b-a)c_4/(rc_1) + \lambda_{\mathrm{r}}\{(1-\delta/2)c - b\}, \\
B_1 &= \kappa(1+b) - 2(\varepsilon + \lambda_{\mathrm{r}}c_2)(b-a)/(rc_1), \\
B_2 &= \kappa(b-a).
\end{aligned}
$$

First, since $c - b \le \delta b/4 \le \delta c/4$ and $a \le c$ by (34) and (36), we have

$$
\begin{aligned}
B_0 &\le 2(b-a)c_4/(rc_1) - \lambda_{\mathrm{r}}\delta c/4 \\
&\le 2(b-a)c_4/(rc_1) - \lambda_{\mathrm{r}}\delta a/4 \\
&\le -a\lambda_{\mathrm{r}}\delta/8,
\end{aligned}
\tag{44}
$$

where we have used that $b - a \le \lambda_{\mathrm{r}}\delta arc_1/(16c_4)$ by (32) for the last inequality. Second, using $\varepsilon \le \lambda_{\mathrm{r}}c_2$ by (35), $(b-a) \le a/3 \le 1/3$ by (32)–(31), we have

$$
\begin{aligned}
B_2 - B_1 &\le \kappa(b-a) + 4\lambda_{\mathrm{r}}c_2(b-a)/(rc_1) - \kappa(1+b) \\
&\le 4\lambda_{\mathrm{r}}c_2 a/(rc_1) - \kappa \le 0,
\end{aligned}
\tag{45}
$$

where we used the definition of κ (33) and the condition (13) for the last inequality. Combining (44) and (45) in (43), we get

$$J(x,y) \le -a\lambda_{\mathrm{r}}\delta/8. \tag{46}$$

Case 4: $2\ell(x)\theta(x,y)/(rc_1) \in (0,1)$. First, note that since $\varphi(s) = 1+b+s(b-a)$ for $s \in [-1,0]$, and φ is nondecreasing, we have for any $s \in [0,1]$,

$$\varphi(-s) - \varphi(s) \le \varphi(-s) - \varphi(0) \le -(b-a)s.$$

From this result and the fact by (20)–(21) that $1+b+s(c-b)-\varepsilon \le \varphi(s) \le 1+b+s(c-b)+\varepsilon$ and $\varphi'(s) \le c-b+\varepsilon$ for $s \in (0,1)$ we get that (38) reads

$$\begin{aligned} J(x,y) &\le \kappa\theta(x,y)\{1+b+2\ell(x)\theta(x,y)(c-b+\varepsilon)/(rc_1)+\varepsilon\} \\ &\quad + \bigl(2(c-b+\varepsilon)/(rc_1)\bigr)\{c_4+\theta(x,y)\varepsilon\} \\ &\quad - \bigl(\|\nabla U(x)\|/\|\nabla \bar{U}(x)\|\bigr)2\ell(x)(b-a)\theta(x,y)^2/(rc_1) \\ &\quad + \lambda_{\mathrm{r}}\{1+(1-\delta/2)c-1-b-2\ell(x)\theta(x,y)(c-b-\varepsilon)/(rc_1)+\varepsilon\} \\ &\le B_0 + B_1\theta(x,y) + 2\ell(x)B_2\theta(x,y)^2/(rc_1), \end{aligned}$$

where we have used that $(\|\nabla U(x)\|/\|\nabla \bar{U}(x)\|)\ell(x) \ge c_3$ by (9), $\theta(x,y) \ge 0$ and defined

$$\begin{aligned} B_0 &= 2c_4(c-b+\varepsilon)/(rc_1) + \lambda_{\mathrm{r}}\{(1-\delta/2)c-b+\varepsilon\}, \\ B_1 &= \kappa(1+b+\varepsilon) + 2\varepsilon(c-b+\varepsilon)/(rc_1), \\ B_2 &= \{\kappa(c-b+\varepsilon) - c_3(b-a)/\ell(x)\}. \end{aligned}$$

Since $\varepsilon \le c-b$ by (35), $\ell(x) \le c_2$ by (9) and $2\kappa c_2(c-b) \le c_3(b-a)/2$ by (34), we get

$$B_2 \le -\tilde{B}_2 = -c_3(b-a)/\bigl(2\ell(x)\bigr), \tag{47}$$

and therefore

$$J(x,y) \le B_0 + B_1\theta(x,y) - 2\ell(x)\tilde{B}_2\theta(x,y)^2/(rc_1).$$

Then, using that $s \mapsto C_1 s - C_2 s^2$ is bounded by $C_1^2/(2C_2)$ on $\mathbb{R}$, we obtain

$$J(x,y) \le B_0 + \theta(x,y)rc_1 B_1^2/\bigl(4\ell(x)\tilde{B}_2\bigr).$$

Therefore, since $\theta(x,y) \in (0,1)$, to show that

$$J(x,y) \le -\lambda_{\mathrm{r}}\delta c/16, \tag{48}$$

it is sufficient to prove that

$$B_0 \le -\lambda_{\mathrm{r}}\delta c/4, \tag{49}$$

$$rc_1 B_1^2/\bigl(4\ell(x)\tilde{B}_2\bigr) \le \lambda_{\mathrm{r}}\delta c/8. \tag{50}$$

First (49) holds since using that $\varepsilon \le (c-b)$ by (35) and that $a \le c$, we have

$$\begin{aligned} B_0 - \delta/4 &= 2c_4(c-b+\varepsilon)/(rc_1) + \lambda_{\mathrm{r}}\{(1-\delta/4)c-b+\varepsilon\} \\ &\le \bigl(4c_4/(rc_2)+2\lambda_{\mathrm{r}}\bigr)(c-b) - \delta a\lambda_{\mathrm{r}}/4 \le 0, \end{aligned}$$

using $(c-b) \le \delta a\lambda_{\mathrm{r}}/(4(4c_4/(rc_2)+2\lambda_{\mathrm{r}}))$ by (34) for the last inequality. It remains to establish (50) which is equivalent by definition of B_1 and $\tilde{B}_2$ (47) to

$$\kappa(1+b+\varepsilon) + 2\varepsilon(c-b+\varepsilon)/(rc_1) \le \bigl\{\lambda_{\mathrm{r}}c\delta c_3(b-a)/(4rc_1)\bigr\}^{1/2}. \tag{51}$$

Since $\varepsilon \le 1 \wedge (\kappa r c_1/4)$ by (35), $c - b \le 1$ and $b \le 2$ by (36) and (31), we get

$$\kappa(1 + b + \varepsilon) + 2\varepsilon(c - b + \varepsilon)/(rc_1) \le 5\kappa.$$

This result, the inequality $b - a \le c$ and the definition of κ (33) implies that (51) holds.

Case 5: $2\ell(x)\theta(x, y)/(rc_1) \ge 1$. Since by (20), $\varphi(s) = 1 + c$, $\varphi'(s) = 0$ and $\varphi(-s) - \varphi(s) \le a - c$ for $s \ge 1$, (38) reads

$$\begin{aligned} J(x, y) &\le \kappa\theta(x, y)(1 + c) - \{\|\nabla U(x)\|/\|\nabla \bar{U}(x)\|\}\theta(x, y)(c - a) - \lambda_{\mathrm{r}}\delta c/2 \\ &\le \kappa\theta(x, y)(1 + c) - \{\|\nabla U(x)\|\ell(x)/(c_2\|\nabla \bar{U}(x)\|)\}\theta(x, y)(c - a) - \lambda_{\mathrm{r}}\delta c/2 \\ &\le \{\kappa(1 + c) - c_3(c - a)/c_2\}\theta(x, y) - \lambda_{\mathrm{r}}\delta c/2, \end{aligned}$$

where we have used by (9) that $\ell(x) \le c_2$ and $\|\nabla U(x)\|\ell(x)\|\nabla \bar{U}(x)\|^{-1} \ge c_3$. From $c \le 3$ by (36) we obtain

$$\begin{aligned} J(x, y) &\le \{\kappa(1 + c) - c_3(c - a)/c_2\}\theta(x, y) - \lambda_{\mathrm{r}}\delta c/2 \\ &\le \{4\kappa - c_3(b - a)/c_2\}\theta(x, y) - \lambda_{\mathrm{r}}\delta c/2 \le -\lambda_{\mathrm{r}}\delta c/2, \end{aligned} \tag{52}$$

where we have used the definition of κ given by (33) and $\theta(x, y) \ge 0$ for the last inequality.

The proof follows from combining (40)–(42)–(46)–(48)–(52). □

COROLLARY 9. *Under A8, for all $(x, y) \in \mathbb{R} \times \mathsf{Y}$ and $t \ge 0$,*

$$P_t V(x, y) \le V(x, y)\mathrm{e}^{-A_1 t} + A_2(1 - \mathrm{e}^{-A_1 t}),$$

where V is given by (22) and A_1, A_2 are given by Lemma 7.

PROOF. By [10], Section 31.5, since $V \in \mathrm{D}(\mathcal{A})$, the process $(M_t)_{t \ge 0}$, defined for any $t \in \mathbb{R}_+$ by

$$M_t = \mathrm{e}^{A_1 t} V(X_t, Y_t) - V(x, y) - \int_0^t \{A_1 \mathrm{e}^{A_1 s} V(X_s, Y_s) + \mathrm{e}^{A_1 s}\mathcal{A}V(X_s, Y_s)\}\,\mathrm{d}s,$$

is a local martingale. Therefore $(M_{t\wedge\tau_n})_{t\ge 0}$ is a martingale where for all $n \in \mathbb{N}^*$, $\tau_n = \inf\{t \ge 0 : \|X_t\| + \|Y_t\| \ge n\}$ and

$$\begin{aligned} &\mathbb{E}\big[\mathrm{e}^{A_1(t\wedge\tau_n)} V(X_{t\wedge\tau_n}, Y_{t\wedge\tau_n})\big] - V(x, y) \\ &\quad = \mathbb{E}\bigg[\int_0^{t\wedge\tau_n} \mathrm{e}^{A_1 s}\{A_1 V(X_s, Y_s) + \mathcal{A}V(X_s, Y_s)\}\,\mathrm{d}s\bigg] \\ &\quad \le \mathbb{E}\bigg[\int_0^{t\wedge\tau_n} \mathrm{e}^{A_1 s} A_1 A_2\,\mathrm{d}s\bigg] \le A_2(\mathrm{e}^{A_1 t} - 1). \end{aligned}$$

Letting n go to infinity concludes the proof since it yields

$$\mathrm{e}^{A_1 t}\mathbb{E}\big[V(X_t, Y_t)\big] \le V(x, y) + A_2(\mathrm{e}^{A_1 t} - 1).$$ □

3.3. *Mirror coupling*. To obtain geometric ergodicity, the classical Meyn and Tweedie approach is, once a Lyapunov drift condition holds, to show a Doeblin condition for some $\mathsf{C} \subset \mathbb{R}^d \times \mathsf{Y}$, that is, that the following holds: there exist $t > 0$, $\varepsilon > 0$ and $\nu \in \mathcal{P}(\mathbb{R}^d \times \mathsf{Y})$, such that

$$P_t\big((x, y), \mathsf{A}\big) \ge \varepsilon\nu(\mathsf{A}) \quad \text{for all } \mathsf{A} \in \mathcal{B}\big(\mathbb{R}^d \times \mathsf{Y}\big), (x, y) \in \mathsf{C}.$$

A set C that satisfies this is called a small set.

LEMMA 10. *Assume A1 and A2(ii). Then, any compact set $\mathsf{K} \subset \mathbb{R}^d \times \mathsf{Y}$ is a small set.*

Previous works [11, 39] establish Lemma 10 in the case where $\mathsf{Y} = \mathsf{S}^d$. The proof relies on the fact that after two refreshment events the distribution of X_t has some density w.r.t.the Lebesgue density on a ball with a radius proportional to t. Nevertheless, the latter strategy yields a nonexplicit rate of convergence. In particular the dependence of the obtained rate in the dimension of the space is either intractable or very rough.

For this reason, we will present a different argument, based on an explicit coupling of two BPS processes. However, this will only work under the assumption that μ_{v} is not singular with respect to the Lebesgue measure on $\mathbb{R}^d$, which rules out, for example, the case of the uniform measure on S^d. A general proof of Lemma 10, with no additional assumption on μ_{v}, may be obtained by a straightforward adaptation of [39] or [11], Lemma 2, Lemma 5.2. We will only treat the nonsingular case, with a particular emphasis on the case where μ_{v} is a d-dimensional nondegenerate Gaussian distribution with zero-mean and covariance matrix Σ.

The aim of the rest of this section is to establish the following coupling condition: for any compact set $\mathsf{C} \subset \mathbb{R}^d \times \mathsf{Y}$, there exist $t > 0$, $\varepsilon > 0$ such that for all $(x, y), (\tilde{x}, \tilde{y}) \in \mathsf{C}$,

$$\|P_t\big((x, y), \cdot\big) - P_t\big((\tilde{x}, \tilde{y}), \cdot\big)\|_{\mathrm{TV}} \leq 2(1 - \varepsilon).$$

This is clearly implied by Lemma 10. However, in order to get good explicit rates of convergence, it may be more efficient to establish directly a coupling condition, which can then be directly used to obtain quantitative estimates (see for instance, Theorem S7 in Appendix and the exemple in Section 4.1).

Before stating our main result, we need the following lemma concerning the reflexion coupling (see [17, 33] and references therein) between two d standard Gaussian random variables with different means.

LEMMA 11. *Let $x^{(1)}, x^{(2)} \in \mathbb{R}^d$, Σ_{R} be a positive definite matrix and $(W_t^{(1)})_{t\geq 0}$ be a standard one-dimensional Brownian motion. Define $T_{\mathrm{c}} = \inf\{t \geq 0 : W_t^{(1)} \geq \|\Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})\|/2\}$, the stochastic process $(W_t^{(2)})_{t\geq 0}$ by*

$$W_t^{(2)} = \begin{cases} -W_t^{(1)} & \text{if } t \leq T_{\mathrm{c}}, \\ -\|\Sigma_{\mathrm{R}}^{-1}(x^{(2)} - x^{(1)})\| + W_t^{(1)} & \text{otherwise}, \end{cases}$$

and the d-dimensional random variables

$$\begin{aligned} G^{(1)} &= W_1^{(1)} \mathrm{n}\{\Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})\} + G_{\mathrm{P}}, \\ G^{(2)} &= W_1^{(2)} \mathrm{n}\{\Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})\} + G_{\mathrm{P}}, \\ G_{\mathrm{P}} &= \big(\mathrm{Id} - \mathrm{n}\{\Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})\}\mathrm{n}\{\Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})\}^{\mathrm{T}}\big)G, \end{aligned}$$

where G is a standard d-dimensional Gaussian random variable independent of $(W_t^{(1)})_{t\geq 0}$ and n *is given by* (4). *Then $G^{(1)}$ and $G^{(2)}$ are d-dimensional standard Gaussian random variables and for all $M \geq 0$,*

$$\begin{aligned} &\mathbb{P}\big(x^{(1)} + \Sigma_{\mathrm{R}}^{1/2} G^{(1)} = x^{(2)} + \Sigma_{\mathrm{R}}^{1/2} G^{(2)}, \|G^{(1)} - \Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})/2\| \leq M\big) \\ &\qquad = \tilde{\alpha}\big(\|\Sigma_{\mathrm{R}}^{-1/2}(x^{(2)} - x^{(1)})\|, M\big), \end{aligned}$$

where for all $r \geq 0$,

$$\begin{aligned} \tilde{\alpha}(r, M) = {} & \frac{r}{2(2\pi)^{(d+1)/2}} \int_0^1 \bigg\{ s^{-3/2} \exp\big(-r^2/(8s)\big) \\ & \times \int_{\mathbb{R}^d} \mathbb{1}_{[0,M]}\big(\big((1-s)w_1^2 + \cdots + w_d^2\big)^{1/2}\big) \mathrm{e}^{-\|x\|^2/2} \, \mathrm{d}w \bigg\} \, \mathrm{d}s. \end{aligned} \tag{53}$$

PROOF. By the Markov property of the Brownian motion $(W_t^{(1)})_{t\geq 0}$, since T_c is a $(\mathcal{F}_t^W)_{t\geq 0}$-stopping time, where $\mathcal{F}_t^W = \sigma(W_s^{(1)}, s \leq t)$, $W_t^{(2)}$ is a Brownian motion. Therefore, $G^{(1)}$ and $G^{(2)}$ are d-dimensional standard Gaussian random variables.

Using again the Markov property of $(W_t^{(1)})_{t\geq 0}$, given $T_c < 1$, $W_1^{(1)} - W_{T_c}^{(1)}$ is independent of $\mathcal{F}_{T_c}^W$. Therefore, since $\{x^{(1)} + \Sigma_R^{1/2} G^{(1)} = x^{(2)} + \Sigma_R^{1/2} G^{(2)}\} = \{T_c \leq 1\}$ and G is independent of $(W_t^{(1)})_{t\geq 0}$, we get for all $M \geq 0$,

$$
\begin{aligned}
&\mathbb{P}\big(x^{(1)} + \Sigma_R^{1/2} G^{(1)} = x^{(2)} + \Sigma_R^{1/2} G^{(2)}, \|G^{(1)} - \Sigma_R^{-1/2}(x^{(2)} - x^{(1)})/2\| \leq M\big) \\
&\quad = \mathbb{E}\big[\mathbb{1}_{[0,1]}(T_c)\mathbb{P}\big(\big((W_1^{(1)} - W_{T_c}^{(1)})^2 + \|\bar{G}\|^2\big)^{1/2} \leq M | \mathcal{F}_{T_c}^W\big)\big] \\
&\quad = (2\pi)^{-d/2}\mathbb{E}\bigg[\mathbb{1}_{[0,1]}(T_c)\int_{\mathbb{R}^d} \mathbb{1}_{[0,M]}\big\{\big((1-T_c)w_1^2 + \cdots + w_d^2\big)^{1/2}\big\} \mathrm{e}^{-\|x\|^2/2}\,\mathrm{d}w\bigg].
\end{aligned}
$$

The proof then follows from the explicit expression of the density of T_c w.r.t.the Lebesgue measure (see, e.g., [44], p. 107). □

LEMMA 12. *Assume A1*, $\mathsf{Y} = \mathbb{R}^d$ *and* μ_v *is the Gaussian measure with zero-mean and covariance matrix* Σ. *Then, for all* $t > 0$ *and all compact set* $\mathsf{K} \subset \{(z, w) \in \mathbb{R}^d \times \mathsf{Y} : \|z\| + \|w\| \leq R_{\mathsf{K}}\}$ *of* $\mathbb{R}^d \times \mathsf{Y}$, $R_{\mathsf{K}} \geq 0$, *for all* $(x, y), (\tilde{x}, \tilde{y}) \in \mathsf{K}$ *and for all* $M \geq 0$,

$$
\begin{aligned}
&(1/2)\big\|P_t\big((x, y), \cdot\big) - P_t\big((\tilde{x}, \tilde{y}), \cdot\big)\big\|_{\mathrm{TV}} \\
&\quad \leq 1 - \mathbb{E}\big[\mathbb{1}_{[0,\lambda_r t]}(E_1 + E_2)\tilde{\alpha}\big(2(\lambda_r + E_1)R_{\mathsf{K}}\|\Sigma^{-1/2}\|E_2, M\big)g(E_2/\lambda_r)\big],
\end{aligned}
$$

where $\tilde{\alpha}$ *is given by* (53), *for all* $r \geq 0$,

$$
\begin{aligned}
g(r) &= \mathbb{P}\Big(r\tilde{M} \sup_{z \in \mathrm{B}(0,(1+E_1/\lambda_r)R_{\mathsf{K}} + (r/\lambda_r)\tilde{M})} \|\nabla U(z)\| \geq E_3\Big), \\
\tilde{M} &= M + \|\Sigma^{1/2}\|(1 + E_1/\lambda_r)R_{\mathsf{K}},
\end{aligned}
\tag{54}
$$

and E_1, E_2, E_3 *are three independent exponential random variables with parameter* 1.

PROOF. Let K be a compact set of $\mathbb{R}^{2d}$. Let $(x, y), (\tilde{x}, \tilde{y}) \in \mathsf{K}$, $(x, y) \neq (\tilde{x}, \tilde{y})$. We construct a non Markovian coupling $(X_t, Y_t, \tilde{X}_t, \tilde{Y}_t)$ between the two distributions $P_t((x, y), \cdot)$ and $P_t((\tilde{x}, \tilde{y}), \cdot)$ for all $t > 0$, and lower bound the quantity $\mathbb{P}((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t))$, which will conclude the proof using the characterization of the total variation distance by coupling.

Before proceeding to its precise definition, let us give a brief and informal description of this coupling (see Figure 1, Figure 2 and Figure 3). We couple both processes to have the same two first refreshment times H_1 and H_2. At time H_1, the Gaussian velocities are chosen according to Lemma 11 so that, in the absence of bounces in the meanwhile, with positive probability, the processes will reach the same position at time H_2. At time H_2, both velocities are refreshed with the same Gaussian variable. Hence, with positive probability, at time H_2, the processes have the same position and same velocity, in which case we can keep them equal for all times $t \geq H_2$.

More precisely, the coupling we consider is defined as follows. Let $(E_i, F_i, \bar{G}_i)_{i\in\mathbb{N}^*}$ be i.i.d. random variables, where for all $i \in \mathbb{N}^*$, E_i, F_i are independent exponential random variables with parameter 1 and $\bar{G}_i$ has distribution μ_v and is independent from E_i, F_i. In addition, let G be a standard d-dimensional Gaussian random variable and $(W_t)_{t\geq 0}$ be a d-dimensional standard Brownian motion such that G, $(W_t)_{t\geq 0}$ and $(E_i, F_i, \bar{G}_i)_{i\in\mathbb{N}^*}$ are independent.

Set $(X_0, Y_0) = (x, y)$, $(\tilde{X}_0, \tilde{Y}_0) = (\tilde{x}, \tilde{y})$, $S_0 = 0$, $H_0 = 0$, $N_0 = 0$, $H_1 = E_1/\lambda_r$ and $N_1 = 1$. The process and its jump times are defined by recursion. Assume that S_n, N_{n+1}, H_{n+1} and $(X_t, Y_t, \tilde{X}_t, \tilde{Y}_t)_{t\in[0,S_n]}$ have been defined for some $n \in \mathbb{N}$. We distinguish two cases.

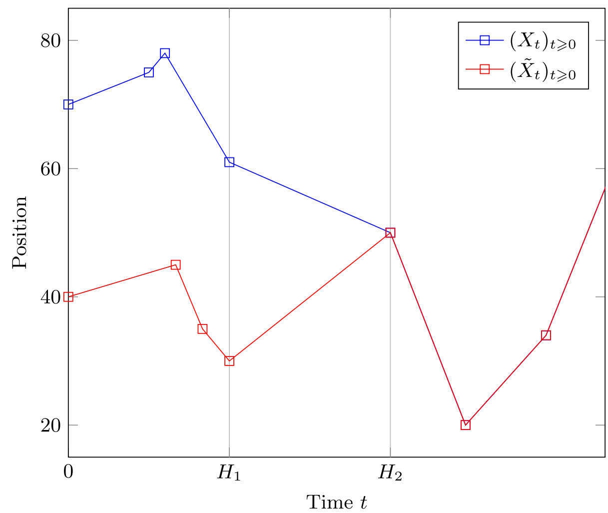

FIG. 1. *Before the first refreshment at time H_1, both processes may bounce freely. At time H_1, the Gaussian velocities are coupled so that, at time H_2 (which is the next refreshment time), provided this Gaussian coupling of the velocities succeeds, and provided they have not bounced in the meanwhile, both processes reach the same position. At time H_2, both processes take the same velocity: they have merged, the coupling is a success.*

(A) If $N_{n+1} = 1$. Define

$$T_{n+1}^{(1)} = \inf\left\{t \geq 0 : \int_0^t \{\langle Y_{S_n}, \nabla U(X_{S_n} + sY_{S_n})\rangle_+\}\,\mathrm{d}s \geq F_{n+1}\right\},$$

$$\tilde{T}_{n+1}^{(1)} = \inf\left\{t \geq 0 : \int_0^t \{\langle \tilde{Y}_{S_n}, \nabla U(\tilde{X}_{S_n} + s\tilde{Y}_{\tilde{S}_n})\rangle_+\}\,\mathrm{d}s \geq F_{n+1}\right\},$$

$$T_{n+1} = H_{n+1} \wedge T_{n+1}^{(1)} \wedge \tilde{T}_{n+1}^{(1)}.$$

Set $S_{n+1} = S_n + T_{n+1}$, for all $t \in [S_n, S_{n+1})$, $(X_t, Y_t) = \phi_t(X_{S_n}, Y_{S_n})$, $X_{S_{n+1}} = X_{S_n} + T_{n+1}Y_{S_n}$, $(\tilde{X}_t, \tilde{Y}_t) = \phi_t(\tilde{X}_{S_n}, \tilde{Y}_{S_n})$, $\tilde{X}_{S_{n+1}} = \tilde{X}_{S_n} + T_{n+1}\tilde{Y}_{S_n}$. If $T_{n+1} = \bar{H}_{n+1}$, consider the two random variables $G^{(1)}, G^{(2)}$ defined by Lemma 11, associated with $(W_t)_{t\geq 0}$ and G, and

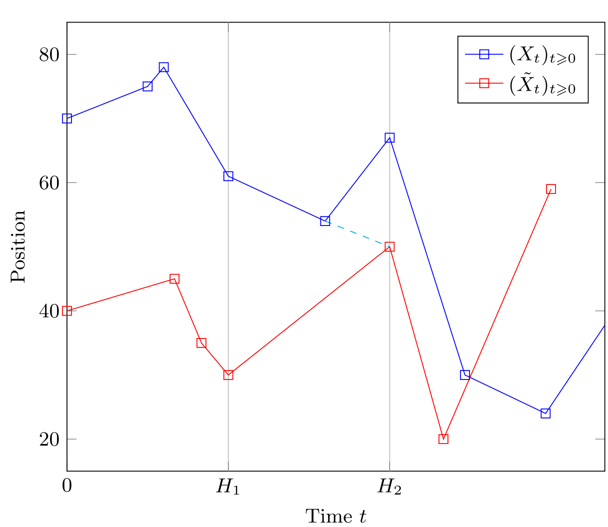

FIG. 2. *If one (at least) of the processes bounces between times H_1 and H_2, then the coupling fails. There may be other bounces after the first one.*

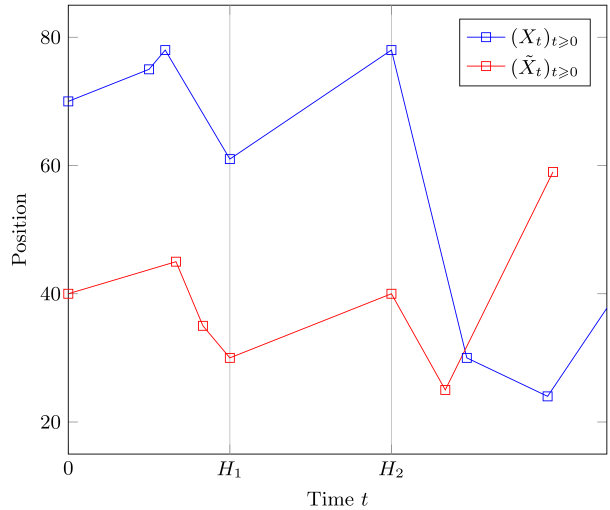

FIG. 3. *Even if none of the process bounces between time H_1 and H_2, the coupling may also fail if the Gaussian coupling of the velocities at time H_1 fails.*

for $x^{(1)} = X_{S_{n+1}}$, $x^{(2)} = \tilde{X}_{S_{n+1}}$, $\Sigma_{\mathrm{R}} = E_2\Sigma/\lambda_{\mathrm{r}}$, and $M \geq 0$.

$$\text{Still if } T_{n+1} = H_{n+1}, \quad \text{set } \begin{cases} Y_{S_{n+1}} = \Sigma^{1/2}G^{(1)} & \tilde{Y}_{S_{n+1}} = \Sigma^{1/2}G^{(2)}, \\ N_{n+2} = 2 & H_{n+2} = E_{N_{n+2}}/\lambda_{\mathrm{r}}. \end{cases}$$

Otherwise set $N_{n+2} = N_{n+1}$, $H_{n+2} = H_{n+1} - T_{n+1}$ and

$$\begin{aligned} &\text{if } T_{n+1} = T^{(1)}_{n+1} = \tilde{T}^{(1)}_{n+1}, && Y_{S_{n+1}} = \mathrm{R}(X_{\tilde{S}_n} + T_{n+1}Y_{S_n}, Y_{S_n}), \\ & && \tilde{Y}_{S_{n+1}} = \mathrm{R}(\tilde{X}_{S_n} + T_{n+1}\tilde{Y}_{S_n}, \tilde{Y}_{S_n}), \\ &\text{if } T_{n+1} = T^{(1)}_{n+1} < \tilde{T}^{(1)}_{n+1}, && Y_{S_{n+1}} = \mathrm{R}(X_{S_n} + T_{n+1}Y_{S_n}, Y_{S_n}), \qquad \tilde{Y}_{S_{n+1}} = \tilde{Y}_{S_n}, \\ &\text{if } T_{n+1} = \tilde{T}^{(1)}_{n+1} < T^{(1)}_{n+1}, && \tilde{Y}_{S_{n+1}} = \mathrm{R}(\tilde{X}_{S_n} + T_{n+1}\tilde{Y}_{S_n}, \tilde{Y}_{S_n}), \qquad Y_{S_{n+1}} = Y_{S_n}, \end{aligned}$$

where R is defined by (4).

(B) If $N_{n+1} \geq 2$. Define

$$T^{(1)}_{n+1} = \inf\left\{t \geq 0 : \int_0^t \{\langle Y_{S_n}, \nabla U(X_{S_n} + sY_{S_n})\rangle_+\}\,\mathrm{d}s \geq F_{n+1}\right\},$$

$$\tilde{T}^{(1)}_{n+1} = \inf\left\{t \geq 0 : \int_0^t \{\langle \tilde{Y}_{S_n}, \nabla U(\tilde{X}_{S_n} + s\tilde{Y}_{\tilde{S}_n})\rangle_+\}\,\mathrm{d}s \geq F_{n+1}\right\},$$

$$T_{n+1} = H_{n+1} \wedge T^{(1)}_{n+1} \wedge \tilde{T}^{(1)}_{n+1}.$$

Set $S_{n+1} = S_n + T_{n+1}$, for all $t \in [S_n, S_{n+1})$, $(X_t, Y_t) = \phi_t(X_{S_n}, Y_{S_n})$, $X_{S_{n+1}} = X_{S_n} + T_{n+1}Y_{S_n}$, $(\tilde{X}_t, \tilde{Y}_t) = \phi_t(\tilde{X}_{S_n}, \tilde{Y}_{S_n})$, $\tilde{X}_{S_{n+1}} = \tilde{X}_{S_n} + T_{n+1}\tilde{Y}_{S_n}$ and

$$\text{if } \tilde{T}_{n+1} = H_{n+1}, \quad \begin{cases} Y_{S_{n+1}} = \bar{G}_{n+1} & \tilde{Y}_{S_{n+1}} = \bar{G}_{n+1}, \\ N_{n+2} = N_{n+1} + 1 & H_{n+2} = E_{N_{n+2}}/\lambda_{\mathrm{r}}. \end{cases}$$

Otherwise set $N_{n+2} = N_{n+1}$, $H_{n+2} = H_{n+1} - T_{n+1}$ and

$$\begin{aligned} &\text{if } T_{n+1} = T^{(1)}_{n+1} = \tilde{T}^{(1)}_{n+1}, && Y_{S_{n+1}} = \mathrm{R}(X_{\tilde{S}_n} + T_{n+1}Y_{S_n}, Y_{S_n}), \\ & && \tilde{Y}_{S_{n+1}} = \mathrm{R}(\tilde{X}_{S_n} + T_{n+1}\tilde{Y}_{S_n}, \tilde{Y}_{S_n}), \end{aligned}$$

$$\text{if } T_{n+1} = T^{(1)}_{n+1} < \tilde{T}^{(1)}_{n+1}, \quad Y_{S_{n+1}} = \mathrm{R}(X_{S_n} + T_{n+1}Y_{S_n}, Y_{S_n}), \qquad \tilde{Y}_{S_{n+1}} = \tilde{Y}_{S_n},$$
$$\text{if } T_{n+1} = \tilde{T}^{(1)}_{n+1} < T^{(1)}_{n+1}, \quad \tilde{Y}_{S_{n+1}} = \mathrm{R}(\tilde{X}_{S_n} + T_{n+1}\tilde{Y}_{S_n}, \tilde{Y}_{S_n}), \qquad Y_{S_{n+1}} = Y_{S_n}.$$

For $t \geq \sup_{n\in\mathbb{N}^*} S_n$, set $(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t) = \infty$. Remark that, since the conditional distribution of $(G^{(1)}, G^{(2)})$ given $(E_i, F_i, \bar{G}_i)_{i\in\mathbb{N}^*}$ depends on E_2, $(X_t, Y_t, \tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ is not Markovian. However, according to Lemma 11, conditionally to $(E_i, (F_{i,j})_{j\in\mathbb{N}^*}, G_i)_{i\in\mathbb{N}^*}$, $G^{(1)}$ and $G^{(2)}$ are both d-dimensional standard Gaussian random variables. As a consequence, from [15], Proposition 5, marginally, $(X_t, Y_t)_{t\geq 0}$ and $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ are two BPS processes starting from (x, y) and $(\tilde{x}, \tilde{y})$.

Further, from the construction of the two processes, for all $n \in \mathbb{N}$ if $(X_{S_n}, Y_{S_n}) = (\tilde{X}_{S_n}, \tilde{Y}_{S_n})$, then $(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)$ for all $t > S_n$. Besides, consider $\tau = \inf\{n \in \mathbb{N} : N_{n+2} = 2\}$. Then by definition, if $T_{\tau+2} = H_{\tau+2}$ and $X_{S_{\tau+1}} + E_2 G^{(1)}/\lambda_{\mathrm{r}} = \tilde{X}_{S_{\tau+1}} + E_2 G^{(2)}/\lambda_{\mathrm{r}}$, then $(X_{S_{\tau+2}}, Y_{S_{\tau+2}}) = (\tilde{X}_{S_{\tau+2}}, \tilde{Y}_{S_{\tau+2}})$. Finally, by definition of τ, $T_{\tau+1} = H_{\tau+1}$ implies $S_{\tau+1} = E_1/\lambda_{\mathrm{r}}$ and if in addition $T_{\tau+2} = H_{\tau+2}$, we have that $S_{\tau+2} = S$ with $S = (E_1 + E_2)/\lambda_{\mathrm{r}}$. Based on these three observations, we get for all $t > 0$,

$$\begin{aligned}
&\mathbb{P}\big((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\big) \\
&\quad \geq \mathbb{P}\left(t \geq S_{\tau+2}, T_{\tau+2} = H_{\tau+2}, X_{S_{\tau+1}} + \frac{E_2\Sigma^{1/2}G^{(1)}}{\lambda_{\mathrm{r}}} = \tilde{X}_{S_{\tau+1}} + \frac{E_2\Sigma^{1/2}G^{(2)}}{\lambda_{\mathrm{r}}}\right) \qquad (55)\\
&\quad \geq \mathbb{P}\left(\mathsf{A} \cap \{t \geq S\} \cap \left\{X_{E_1/\lambda_{\mathrm{r}}} + \frac{E_2\Sigma^{1/2}G^{(1)}}{\lambda_{\mathrm{r}}} = \tilde{X}_{E_1/\lambda_{\mathrm{r}}} + \frac{E_2\Sigma^{1/2}G^{(2)}}{\lambda_{\mathrm{r}}}\right\}\right),
\end{aligned}$$

where $\mathsf{A} = \mathsf{A}_1 \cap \mathsf{A}_2$,

$$\mathsf{A}_1 = \left\{\int_0^{E_2/\lambda_{\mathrm{r}}} \{\langle Y_{E_1/\lambda_{\mathrm{r}}}, \nabla U(X_{E_1/\lambda_{\mathrm{r}}} + sY_{E_1/\lambda_{\mathrm{r}}})\rangle_+\}\,\mathrm{d}s \geq F_{\tau+2}\right\},$$
$$\mathsf{A}_2 = \left\{\int_0^{E_2/\lambda_{\mathrm{r}}} \{\langle \tilde{Y}_{E_1/\lambda_{\mathrm{r}}}, \nabla U(\tilde{X}_{E_1/\lambda_{\mathrm{r}}} + s\tilde{Y}_{E_1/\lambda_{\mathrm{r}}})\rangle_+\}\,\mathrm{d}s \geq F_{\tau+2}\right\}.$$

Since for all $n \in \{1, \dots, \tau\}$, $T_{n+1} = T^{(1)}_{n+1} \wedge \tilde{T}(1)_{n+1}$, $\|Y_{S_n}\| = \|y\|$, $\|\tilde{Y}_{S_n}\| = \|\tilde{y}\|$, so for all $s \in [0, E_1/\lambda_{\mathrm{r}}]$,

$$\begin{aligned}
&\|X_s\| \leq \|x\| + (E_1/\lambda_{\mathrm{r}})\|y\| \leq (1 + E_1/\lambda_{\mathrm{r}})R_{\mathsf{K}}, \\
&\|\tilde{X}_s\| \leq (1 + E_1/\lambda_{\mathrm{r}})R_{\mathsf{K}}.
\end{aligned} \qquad (56)$$

For $i = 1, 2$, by the definition (54) of $\tilde{M}$, we obtain that

$$\mathsf{B} = \bigcap_{i=1}^{2}\{\|G^{(i)} - (\Sigma^{1/2}/2)(X_{E_1/\lambda_{\mathrm{r}}} - \tilde{X}_{E_1/\lambda_{\mathrm{r}}})\| \leq M\} \subset \bigcap_{i=1}^{2}\{\|G^{(i)}\| \leq \tilde{M}\}.$$

Using that by definition, $S_{\tau+1} = E_1/\lambda_{\mathrm{r}}$, $N_{\tau+1} = 1$, so $Y_{S_{\tau+1}} = \Sigma^{1/2}G^{(1)}$ and $Y_{S_{\tau+1}} = \Sigma^{1/2}G^{(1)}$, we get that $\mathsf{A}_1 \cap \mathsf{A}_2 \cap \mathsf{B} \subset \tilde{\mathsf{A}}$ where

$$\tilde{\mathsf{A}} = \left\{(E_2/\lambda_{\mathrm{r}})\tilde{M} \sup_{z\in \mathrm{B}(0,(1+E_1/\lambda_{\mathrm{r}})R_{\mathsf{K}}+(E_2/\lambda_{\mathrm{r}})M)} \|\nabla U(z)\| \geq F_{\tau+2}\right\}.$$

Then, we get by (55)

$$\begin{aligned}
&\mathbb{P}\big((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\big) \\
&\quad \geq \mathbb{P}\left(\tilde{\mathsf{A}} \cap \{t \geq S\} \cap \left\{X_{E_1/\lambda_{\mathrm{r}}} + \frac{E_2\Sigma^{1/2}G^{(1)}}{\lambda_{\mathrm{r}}} = \tilde{X}_{E_1/\lambda_{\mathrm{r}}} + \frac{E_2\Sigma^{1/2}G^{(2)}}{\lambda_{\mathrm{r}}}\right\}\right).
\end{aligned}$$

Denoting by $(\bar{\mathcal{F}}_n)_{n\geq 1}$ the filtration associated with $(E_i, F_i, \bar{G}_i)_{i\in\mathbb{N}^*}$, conditioning on $\bar{\mathcal{F}}_{\tau+1}$ and E_2 and using that $F_{\tau+2}$ is independent from $G^{(1)}$, $G^{(2)}$ E_2 and $\bar{\mathcal{F}}_{\tau+1}$, the definition of $G^{(1)}, G^{(2)}$ conditionally to E_2 and $\bar{\mathcal{F}}_{\tau+1}$, Lemma 11 and since $S = (E_1 + E_2)/\lambda_\mathrm{r}$ by definition, we have

$$\mathbb{P}\big((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\big) \geq \mathbb{E}\bigg[\mathbb{1}_{[0,t]}\bigg\{\frac{E_1 + E_2}{\lambda_\mathrm{r}}\bigg\}\tilde{\alpha}\bigg(\frac{\|\Sigma^{-1/2}(X_{E_1/\lambda_\mathrm{r}} - \tilde{X}_{E_1/\lambda_\mathrm{r}})\|\lambda_\mathrm{r}}{E_2}, M\bigg)g\bigg(\frac{E_2}{\lambda_\mathrm{r}}\bigg)\bigg].$$

Combining this result with (56) concludes the proof. □

Consider the more general case where μ_v is rotation invariant and not singular with respect to the Lebesgue measure on $\mathbb{R}^d$. The previous proof may be adapted to this case but the result is less explicit.

LEMMA 13. *Assume for all* $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu_\mathrm{v}(\mathsf{A}) \geq c\nu_{r,\delta}(\mathsf{A}), \tag{57}$$

for some $r, \delta, c > 0$, *where* $\nu_{r,\delta}$ *the uniform law on* $\{y \in \mathbb{R}^d, r < \|y\| < r + \delta\}$. *Let* $\mathsf{K} \subset \mathbb{R}^d$, *be a compact set. Then there exists two random variables* $G^{(1)}, G^{(2)}$ *with distribution* μ_v, $t_0 \geq 0$, $\varepsilon > 0$ *such that for* $s \geq t_0$, *there exists* $M \geq 0$ *satisfying for all* $x, \tilde{x} \in \mathsf{K}$,

$$\mathbb{P}\big(x + sG^{(1)} = \tilde{x} + sG^{(2)}, \|G^{(1)} - (x - \tilde{x})/2\| \leq M\big) \geq \varepsilon.$$

PROOF. Let $x, \tilde{x} \in \mathsf{K} \subset \mathrm{B}(0, R_\mathsf{K})$, $R_\mathsf{K} \geq 0$. If $s > \|x - \tilde{x}\|/(2(r + \delta))$ and $M \geq R_\mathsf{K} + s(r + \delta)$, then $\mathsf{I}(x, \tilde{x}, s) = \{w \in \mathbb{R}^d, \|w\| \leq M\} \cap \{w \in \mathbb{R}^d : sr < \|w - x\| < s(r + \delta)\} \cap \{w \in \mathbb{R}^d : sr < \|w - \tilde{x}\| < s(r + \delta)\} \neq \varnothing$. Writing $\bar{\nu}_{x,s}$ the law of $x + sG$ where G has law μ_v, then for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$, by (57), there exists $\tilde{c} > 0$ such that

$$\bar{\nu}_{x,s}(\mathsf{A}) \wedge \bar{\nu}_{\tilde{x},s}(\mathsf{A}) \geq \tilde{c}\,\mathrm{Leb}\big(\mathsf{A} \cap \mathsf{I}(x, \tilde{x}, s)\big). \tag{58}$$

Besides (see, e.g., [43] or [46]), we can construct a pair (G_1, G_2) of random variables with both G_1 and G_2 distributed according to μ_v, and such that $\mathbb{P}(x + sG = \tilde{x} + sG) = \bar{\nu}_{x,s}(\mathsf{A}) \wedge \bar{\nu}_{\tilde{x},s}(\mathsf{A})$. Combining this result with (58), the fact the function in the right-hand side of (58) is positive and depends continuously of x and $\tilde{x}$, hence is lower bounded on K, concludes. □

LEMMA 14. *Assume A1 and (57) for some* $r, \delta, c > 0$, *where* $\nu_{r,\delta}$ *the uniform law on* $\{y \in \mathbb{R}^d, r < \|y\| < r + \delta\}$. *Then, for all compact set* K *of* $\mathbb{R}^d \times \mathsf{Y}$, *there exists* $t_0, \alpha > 0$ *such that for all* $(x, y), (\tilde{x}, \tilde{y}) \in \mathsf{K}$ *and all* $t \geq t_0$,

$$\|P_t\big((x, y), \cdot\big) - P_t\big((\tilde{x}, \tilde{y}), \cdot\big)\|_{\mathrm{TV}} \leq 2(1 - \alpha).$$

PROOF. The proof is exactly similar to the proof of Lemma 12. Indeed it suffices to consider a coupling of two BPS $(X_t, Y_t)_{t\geq 0}$ and $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ defined similarly to the processes defined in the proof of Lemma 12 but $G^{(1)}, G^{(2)}$ are chosen according to Lemma 13 in place of Lemma 11. □

Finally, let us detail Lemma 10, in prevision of the low-temperature study of Section 4.2.

LEMMA 15. *Assume A1. Then, for all compact set* $\mathsf{K} \subset \mathbb{R}^d \times \mathsf{Y}$, *there exist* $t_0, \varepsilon, C, R > 0$, *which depend on* K, μ_v *and* λ_r *but not on* U, *such that for all* $(x, y), (\tilde{x}, \tilde{y}) \in \mathsf{K}$ *and all* $t \geq t_0$,

$$\|P_t\big((x, y), \cdot\big) - P_t\big((\tilde{x}, \tilde{y}), \cdot\big)\|_{\mathrm{TV}} \leq 2\big[1 - \varepsilon\exp\big(-C\|\nabla U\|_{\infty,\mathrm{B}(0,R)}\big)\big].$$

PROOF. In the case where μ_{v} is a Gaussian distribution, the proof follows from the statement of Lemma 12. In the general case, we only give a sketch of proof, since this is a direct adaptation of [39], Theorem 5.1. First, in the spirit of the proof of Lemma 12 or of [39], Lemma 5.2, we study a BPS with no potential, that is, with $U=0$, and we show that we may couple them so that, with some probability $\alpha>0$, they merge in a given time t_0, without leaving a given compact set. Then we add independent bounces, and say that the coupling is still a success if no bounce happens before time t_0, which gives the desired dependency with respect to U. □

3.4. *Proof of Theorem* 5. The proof follows from Lemma 7 and Lemma 10, and an application of [36], Theorem 6.1. However, [36], Theorem 6.1, is nonquantitative and for the proofs of Section 4.2 need explicit bounds for the convergence of $(P_t)_{t\geq 0}$ to π. To this end, we give a quantitative version of Theorem 5 in Section S2 based on [25], Theorem 1.2.

3.5. *Proofs of Theorem* 1. In each case, we apply Theorem 5. Set $H(t)=t^2$ for $t\in\mathbb{R}$. Consider $r>0$ such that $\delta=\mathbb{P}(|Y_1|>r)>0$ where $Y=(Y_1,\dots,Y_d)\in\mathsf{Y}$ is distributed according to μ_{v}. Note that *A*8(ii) is automatically satisfied in all the cases.

Under *A*3, set $\bar U(x)=U(x)$ and $\ell(x)=1$ for all $x\in\mathbb{R}^d$. All the conditions of *A*8 are sastisfied and so is (13) by Remark 6 since $\lim_{\|x\|\to+\infty}\|\nabla U(x)\|=+\infty$.

Under *A*4, set $\bar U(x)=U^{\varsigma}(x)$ and $\ell(x)=1$ for any $x\in\mathbb{R}^d$. Then *A*8 is satisfied. In addition, (13) holds by Remark 6 since under *A*4

$$\lim_{\|x\|\to+\infty}\{\ell(x)\|\nabla U(x)\|/\|\nabla\bar U(x)\|\}=+\infty.$$

Under *A*5, set $\bar U(x)=U^{\varsigma}(x)$ and $\ell(x)=1/(1+\|\nabla\bar U(x)\|)$ for all $x\in\mathbb{R}^d$. All the conditions of *A*8 are satisfied and (13) holds by Remark 6 since $\lim_{\|x\|\to+\infty}\ell(x)=0$.

3.6. *Proof of Theorem* 2. We apply Theorem 5 again. Set $H(t)=t^2$ for $t\in\mathbb{R}$. Consider $r>0$ such that $\delta=\mathbb{P}(|Y_1|>r)>0$ where $Y=(Y_1,\dots,Y_d)\in\mathsf{Y}$ is distributed according to μ_{v}. Note that *A*8(ii) is automatically satisfied. Set $\bar U(x)=U(x)$ and $\ell(x)=1$ for any $x\in\mathbb{R}^d$. Then, the conditions of *A*8 hold with c_4 arbitrarily small. Therefore, (13) is satisfied if λ_{r} is small enough.

3.7. *Proof of Theorem* 4. We apply Theorem 5. Set $H(t)=\eta t^2$ for η small enough such that *A*8(ii) is satisfied. Set $\bar U(x)=U^{\varsigma}(x)$ for any $x\in\mathbb{R}^d$. Note that

$$\begin{aligned}&\Big\{\sup_{y\in\mathsf{A}_x}\|y\|^2\Big\}\|\nabla^2\bar U(x)\|\\&\quad\leq 3\eta^{-1}\bar U(x)\|\nabla^2\bar U(x)\|\\&\quad\leq CU^{\varsigma}(x)\big(\|\nabla^2U(x)\|U^{\varsigma-1}(x)+\|\nabla U(x)\|^2U^{\varsigma-2}(x)\big)\end{aligned}$$

for some $C>0$, hence is bounded. Then, the proof follows the same lines as the proof of Theorem 1 under *A*4, and is omitted.

4. Miscellaneous.

4.1. *A specific and explicit bound for a toy model.* Following carefully the proofs of Theorem 5, it is possible to get explicit bounds on the values of $C,\rho>0$ such that (5) holds. Nevertheless, the obtained bounds are not sharp. In particular, in Section 3.3, when we try to couple two processes, we do not make any use of the potential U. In fact, at this step, U only plays the role of an hindrance in the minorization condition given by Lemma 10 based on

Lemma 12–Lemma 15. We try to couple the processes using only the refreshment jumps, and hope that, during this attempt, no bounce occurs. We now illustrate on a toy model how an analysis which is model specific can circumvent this flaw. It shows that the explicit bounds we obtain in Lemma 12 may be far from optimality for some problems.

Consider the smooth manifold $\mathsf{D} = (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\eta\mathbb{Z})^{d-1}$ for $d \geq 2$ and $\eta > 0$, and let $\mathrm{proj}^{\mathsf{D}} : \mathbb{R}^d \to \mathsf{D}$ be the corresponding projection (also referred to as quotient map). We set in this section π to be the uniform distribution on D, $\mathsf{Y} = \mathbb{R}^d$ and μ_{v} to be the zero-mean d-dimensional Gaussian distribution on $\mathbb{R}^d$ with covariance matrix $\sigma^2 \operatorname{Id}$, $\sigma^2 > 0$. In this setting, U is simply the function which is identically equal to 0 on D. A BPS sampler $(X_t, Y_t)_{t\geq 0}$ is defined as in Section 2.1 to target $\pi \otimes \mu_{\mathrm{v}}$. The construction is in all respects the same, just by replacing the state space $\mathbb{R}^d \times \mathsf{Y}$ by $\mathsf{D} \times \mathsf{Y}$ and setting $X_t = \mathrm{proj}^{\mathsf{D}}(X_{S_n} + tY_{S_n})$ for $t \in [S_n, S_{n+1})$ in place of $X_t = X_{S_n} + tY_{S_n}$. To show the convergence of the corresponding semi-group $(P_t^{\mathsf{D}})_{t\geq 0}$, we show a uniform Doeblin condition [35], Chapter 16, holds using a direct coupling argument.

Note that D has no boundary and therefore no reflexion has to be take care of but it is worthwhile to mention that by a deterministic transformation of this process from D to $[0, 1] \times [0, \eta]^{d-1}$, we end up with the reflected PDMP process targeting the uniform distribution on $[0, 1] \times [0, \eta]^{d-1}$ described in [2].

The process that we consider in this section can be seen as a toy model for convex potentials. If η is small, which is the analogous of multi-scales problems, then the proof of Theorem 5 would yield a mixing time of order η^d. Indeed, in Section 3.3, the coupling is considered a failure as soon as one of the processes bounce (or, here, is reflected at the boundary). Hence, a successful coupling would need that, at the first refreshment time, the new Gaussian velocity is directed mainly according to the first dimension, which is unlikely. As we will see, this is a too pessimistic bound.

PROPOSITION 1. *For all $x, \tilde{x} \in \mathsf{D}$, $y, \tilde{y} \in \mathbb{R}^d$ and $t > 0$,*

$$\begin{aligned}
&\|\delta_{(x,y)} P_t^{\mathsf{D}} - \delta_{(\tilde{x},\tilde{y})} P_t^{\mathsf{D}}\|_{\mathrm{TV}} \\
&\quad \leq 2\left[\mathbb{P}(N_t \leq 1) + \mathbb{E}\left[\mathbb{1}_{[2,+\infty]}(N_t)\left\{1 - 2\boldsymbol{\Phi}\left(\frac{(1+\eta^2(d-1))^{1/2}}{2(S_{N_t} - S_1)}\right)\right\}\right]\right],
\end{aligned}$$

where $\boldsymbol{\Phi}$ is the cumulative distribution function of the standard Gaussian distribution on $\mathbb{R}$, $(N_t)_{t\geq 0}$ is a Poisson process with rate λ_{r} and jump times $(S_i)_{i\in\mathbb{N}}$, with $S_0 = 0$.

PROOF. Let $(N_t)_{t\geq 0}$ be a Poisson process with rate λ_{r} and jump times $(S_i)_{i\in\mathbb{N}}$, with $S_0 = 0$. Set first for $t \in [0, S_1)$, $X_t = \mathrm{proj}^{\mathsf{D}}(x + ty)$, $Y_t = y$, $X_{S_1} = \mathrm{proj}^{\mathsf{D}}(x + S_1 y)$, $\tilde{X}_t = \mathrm{proj}^{\mathsf{D}}(\tilde{x} + t\tilde{y})$, $\tilde{Y}_t = \tilde{y}$, $\tilde{X}_{S_1} = \mathrm{proj}^{\mathsf{D}}(\tilde{x} + S_1\tilde{y})$. By [33], Section 2, given $(S_i)_{i\in\mathbb{N}}$, there exist two Brownian motions $(W_t)_{t\geq 0}$ and $(\tilde{W}_t)_{t\geq 0}$ on D such that for any $t > 0$,

$$\begin{aligned}
\mathbb{P}\big(X_{S_1} + W_t = \tilde{X}_{S_1} + \tilde{W}_t \,|\, (S_k)_{k\geq 0}\big) &= \mathbb{P}\big(T_{\mathrm{c}} \leq t \,|\, (S_k)_{k\geq 0}\big) \\
&= 1 - 2\boldsymbol{\Phi}\big(-\|X_{S_1 - \tilde{X}_{S_1}}\|/(2t^{1/2})\big)
\end{aligned} \tag{59}$$

and

$$T_{\mathrm{c}} = \inf\{s \geq 0 : X_{S_1} + W_s = \tilde{X}_{S_1} + \tilde{W}_s\}. \tag{60}$$

We can define then, for any $i \in \mathbb{N}^*$,

$$\begin{aligned}
G_i &= (W_{(S_{i+1} - S_1)^2} - W_{(S_i - S_1)^2})/(S_{i+1} - S_i), \\
\tilde{G}_i &= (\tilde{W}_{(S_{i+1} - S_1)^2} - \tilde{W}_{(S_i - S_1)^2})/(S_{i+1} - S_i).
\end{aligned} \tag{61}$$

Note that by the Markov property of $(W_t)_{t\geq 0}$ and $(\tilde{W}_t)_{t\geq 0}$, $(G_i)_{i\in\mathbb{N}^*}$ and $(\tilde{G}_i)_{i\in\mathbb{N}^*}$ are sequences of i.i.d. d-dimensional standard Gaussian random variables.

Define $Y_{S_i} = G_1$, $\tilde{Y}_{S_1} = \tilde{G}_1$ and now assume that (X_t, Y_t), $(\tilde{X}, \tilde{Y}_t)$ are defined for $t \in [0, S_k]$, $k \geq 1$. Set for $t \in [S_k, S_{k+1}]$, $X_t = \mathrm{proj}^{\mathsf{D}}(X_{S_k} + (t - S_k)G_{k+1})$, $\tilde{X}_t = \mathrm{proj}^{\mathsf{D}}(\tilde{X}_{S_k} + (t - S_k)\tilde{G}_{k+1})$, for $t \in [S_k, S_{k+1})$, $Y_t = Y_{S_k}$, $\tilde{Y}_t = \tilde{Y}_{S_k}$ and $Y_{S_{k+1}} = G_{k+1}$, $\tilde{Y}_{S_{k+1}} = \tilde{G}_{k+1}$. It follows then by construction that for any $t \geq 0$, $(X_t, Y_t)_{t\geq 0}$ is distributed according to $P_t^{\mathsf{D}}((x, y), \cdot)$ and $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ is distributed according to $P_t^{\mathsf{D}}((\tilde{x}, \tilde{y}), \cdot)$. Then it remains to bound $\mathbb{P}((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t))$ by definition of the total variation norm.

Note that if $(S_{i+1} - S_1)^2 \geq (t - S_1)^2 \geq (S_i - S_1)^2 \geq T_{\mathrm{c}} \geq (S_{i-1} - S_1)^2$, $i \geq 2$, we have by (60)–(61) and construction $(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)$. Therefore, we get $\{(S_{N_t} - S_1)^2 \geq T_{\mathrm{c}}\} \cap \{N_t > 1\} \subset \{(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\}$ and we obtain

$$\begin{aligned}\mathbb{P}\big((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\big) &\leq \mathbb{P}\big(\{S_{N_t} \leq S_1 + T_{\mathrm{c}}\} \cap \{N_t \leq 1\}\big) \\ &\leq \mathbb{P}(N_t \leq 1) + \mathbb{P}\big(\{N_t \geq 2\} \cap \big\{(S_{N_t} - S_1)^2 \geq T_{\mathrm{c}}\big\}\big).\end{aligned}$$

The proof is then concluded by conditioning with respect to $(S_k)_{k\in\mathbb{N}}$ using (59) and for any $x \in \mathsf{D}$, $\|x\| \leq (1 + \eta^2(d-1))^{1/2}$. $\square$

COROLLARY 16. *There exist $C \geq 0$ and $\varepsilon \in (0, 1]$ independent of d such that setting $t_{\mathrm{c}} = Cd^{1/2}$, for all $x, \tilde{x} \in \mathsf{D}$ and $y, \tilde{y} \in \mathbb{R}^d$,*

$$\big\|\delta_{(x,y)} P_{t_{\mathrm{c}}}^{\mathsf{D}} - \delta_{(\tilde{x},\tilde{y})} P_{t_{\mathrm{c}}}^{\mathsf{D}}\big\|_{\mathrm{TV}} \leq (1 - \varepsilon).$$

PROOF. By Proposition 1 and using the same notation, for all $x, \tilde{x} \in \mathsf{D}$, $y, \tilde{y} \in \mathbb{R}^d$ and $t > 0$, we have since for any $s \geq 0$, $1/2 - \mathbf{\Phi}(-s) \leq 1 \wedge \{s/(2\pi)^{1/2}\}$,

$$\begin{aligned}&2^{-1}\big\|\delta_{(x,y)} P_t^{\mathsf{D}} - \delta_{(\tilde{x},\tilde{y})} P_t^{\mathsf{D}}\big\|_{\mathrm{TV}} \\ &\quad\leq \mathbb{P}(S_2 \geq t/4) + \mathbb{P}(S_2 \leq t/4, S_{N_t} - S_2 \leq t/2) \\ &\qquad + \mathbb{E}\bigg[\mathbb{1}_{[0,t/4]}(S_2)\mathbb{1}_{[t/2,+\infty)}(S_{N_t} - S_2)\bigg\{1 - 2\mathbf{\Phi}\bigg(\frac{(1 + \eta^2(d-1))^{1/2}}{2(S_{N_t} - S_1)}\bigg)\bigg\}\bigg] \\ &\quad\leq \mathbb{P}(S_2 \geq t/4) + \mathbb{P}(S_{N_t} \leq 3t/4) + \frac{\{2(1 + \eta^2(d-1))\}^{1/2}}{t\pi^{1/2}}.\end{aligned}$$

Since $\{S_{N_t} \leq 3t/4\} \subset \{N_t - N_{3t/4} = 0\}$, and $N_t - N_{3t/4}$ follows a Poisson distribution with parameter $t\lambda/4$, we get for all $x, \tilde{x} \in \mathsf{D}$, $y, \tilde{y} \in \mathbb{R}^d$ and $t > 0$

$$2^{-1}\big\|\delta_{(x,y)} P_t^{\mathsf{D}} - \delta_{(\tilde{x},\tilde{y})} P_t^{\mathsf{D}}\big\|_{\mathrm{TV}} \leq \mathbb{P}(S_2 \geq t/4) + \mathrm{e}^{-\lambda t/4} + \frac{\{2(1 + \eta^2(d-1))\}^{1/2}}{t\pi^{1/2}}.$$

The proof then follows from a straightforward computation. $\square$

A direct consequence of Corollary 16 is that, with the same notation, for all $\nu \in \mathcal{P}(\mathsf{D} \times \mathbb{R}^d)$ and $t \geq 0$,

$$\big\|\nu P_t^{\mathsf{D}} - \pi \otimes \mu_{\mathrm{v}}\big\|_{\mathrm{TV}} \leq (1 - \varepsilon)^{\lfloor t/t_{\mathrm{c}}\rfloor}.$$

As a conclusion, for the considered toy model, we get that the rate of convergence scales only as $d^{1/2}$. Note that this result is optimal since the process has unit constant speed and the diameter of D is $d^{1/2}$.

4.2. *The metastable regime and annealing.* The simulated annealing methodology (see [24] and references therein) aims at finding a global minimum of a function U and not sampling to the target distribution π given by (1). However, roughly, these methods need to approximately sample from the family of distributions $\{\pi_\beta : \beta > 0\}$, where π_β is the distribution on $\mathbb{R}^d$ associated with the potential $x \mapsto \beta U(x)$, for $\beta > 0$. To do so, we will study in this section a simulated annealing algorithm based on the BPS, extending the results of [39], Theorem 1.5, on the torus $(\mathbb{R}/\mathbb{Z})^d$. For the sake of simplicity, the study is restricted to the following case:

A9.

(i) The potential $U \in \mathrm{C}^2(\mathbb{R}^d)$ satisfies

$$\int_{\mathbb{R}^d} \exp\bigl(-U(x)/2\bigr)\,\mathrm{d}x < \infty, \qquad \lim_{\|x\|\to+\infty} U(x) = +\infty,$$
$$\liminf_{\|x\|\to\infty} \|\nabla U(x)\| > 0, \qquad \sup_{x\in\mathbb{R}^d} \|\nabla^2 U(x)\| < \infty.$$

Moreover, without loss of generality, $U(0) = \min_{\mathbb{R}^d} U = 0$.

(ii) $\mathsf{Y} = \mathrm{B}(0, M)$ for $M > 0$ and the distribution μ_{v} on Y is rotation invariant.

In the rest of this section, *A9* is enforced. However, note the arguments also work under *A8* (in particular when $Y = \mathbb{R}^d$, μ_{v} has a Gaussian moment and U is a perturbation of an χ-homogeneous potential with $\chi > 1$, as in Proposition 3), which is not implied by *A9*.

For a measurable function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$, referred to in the following as the cooling schedule, we consider in this section the simulated annealing BPS process $(X_t^{(\beta)}, Y_t^{(\beta)})$ defined as follows. Consider some initial point $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$, and the family of i.i.d. random variables $(E_i, F_i, G_i)_{i\in\mathbb{N}^*}$ introduced in Section 2.1. Let $\lambda_{\mathrm{r}} > 0$, $(X_0^{(\beta)}, Y_0^{(\beta)}) = (x, y)$ and $S_0^{(\beta)} = 0$. We define by recursion the jump times of the process and the process itself. For all $n \geq 0$, consider

$$\begin{aligned}
T_{n+1}^{(1,\beta)} &= E_{n+1}/\lambda_{\mathrm{r}},\\
T_{n+1}^{(2,\beta)} &= \inf\Bigl\{t \geq 0 : \int_0^t \{\beta(s)\langle Y_{S_n^{(\beta)}}^{(\beta)}, \nabla U\bigl(X_{S_n^{(\beta)}}^{(\beta)} + sY_{S_n^{(\beta)}}^{(\beta)}\bigr)\rangle_+\}\,\mathrm{d}s \geq E_{n+1}^2\Bigr\},\\
T_{n+1}^{(\beta)} &= T_{n+1}^{(1,\beta)} \wedge T_{n+1}^{(2,\beta)}.
\end{aligned}$$

Set $S_{n+1}^{(\beta)} = S_n^{(\beta)} + T_{n+1}^{(\beta)}$, $(X_t^{(\beta)}, Y_t^{(\beta)}) = (X_{S_n^{(\beta)}}^{(\beta)} + tY_{S_n^{(\beta)}}^{(\beta)}, Y_{S_n^{(\beta)}}^{(\beta)})$, for all $t \in [S_n^{(\beta)}, S_{n+1}^{(\beta)})$, $X_{S_{n+1}^{(\beta)}}^{(\beta)} = X_{S_n^{(\beta)}}^{(\beta)} + T_{n+1}^{(\beta)} Y_{S_n^{(\beta)}}^{(\beta)}$ and

$$Y_{S_{n+1}^{(\beta)}}^{(\beta)} = \begin{cases} G_{n+1} & \text{if } T_{n+1}^{(\beta)} = T_{n+1}^{(1,\beta)},\\ \mathrm{R}\bigl(X_{S_{n+1}^{(\beta)}}^{(\beta)}, Y_{S_n^{(\beta)}}^{(\beta)}\bigr) & \text{otherwise,}\end{cases}$$

where R is defined by (4). Note that under *A9*, Y is bounded and therefore by [15], Proposition 10, $\sup_{n\in\mathbb{N}} S_n^{(\beta)} = +\infty$.

Therefore almost surely $(X_t^{(\beta)}, Y_t^{(\beta)})_{t\geq 0}$ is a $(\mathbb{R}^d \times \mathsf{Y})$-valued càdlàgprocess. By [10], Theorem 25.5, the BPS process $(X_t^{(\beta)}, Y_t^{(\beta)})_{t\geq 0}$ defines a nonhomogeneous strong Markov semigroup $(P_t)_{t\geq 0}$ given for all $s, t \in \mathbb{R}_+$, $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d \times \mathsf{Y})$ by

$$P_{t,t+s}^{(\beta)}\bigl((x, y), \mathsf{A}\bigr) = \mathbb{P}\bigl((X_s^{(\beta)}, Y_s^{(\beta)}) \in \mathsf{A}\bigr),$$

where $(X_u^{(\beta)}, Y_u^{(\beta)})_{u\in\mathbb{R}_+}$ is the annealed BPS process started from (x, y) and cooling schedule $s \mapsto \beta(t+s)$. As it is usual in simulated annealing if $t \mapsto \beta(t)$ goes to infinity sufficiently slowly for the process $(X_t^{(\beta)}, Y_t^{(\beta)})$ to approach its instantaneous equilibrium $\exp(-\beta(t)U) \otimes \mu_{\mathrm{v}}$, then $X_t^{(\beta)}$ should be close to a global minimum of U with high probability.

*A*10. The function $t \mapsto \beta(t)$ is increasing, satisfies $\lim_{t\to+\infty} \beta(t) = +\infty$, $\beta(0) \geq 1$ and there exist $s_0, D_1, D_2 > 0$ with $D_1 \geq D_2$ such that for all t large enough, $\beta(t) \geq D_2 \ln t$ and $\beta(t+s_0) - \beta(t) \leq D_1/t$.

We can then adapt well-known techniques from the simulated annealing literature to extend the result of [39] which restricts its study to the torus $(\mathbb{R}/\mathbb{Z})^d$. A crucial step is to show that for fixed $s, t \geq 0$, $s \leq t$, the Markov kernel $P_{s,t}^{\beta}$ is a contraction in an appropriate metric with constants which have to be explicit in s, t and the cooling schedule β. However, using our approach for the proof of the geometric ergodicity of BPS, we were able to complete such a task.

THEOREM 17. *Assume A9. There exists $\theta > 0$ such that if A10 holds with $D_1 \leq \theta^{-1}$, then for any $(x, y) \in \mathbb{R}^d \times \mathsf{Y}$ and any levels $\eta > \eta' > 0$, there exists $A > 0$ such that, for all $t > 0$,*

$$\mathbb{P}\Big(U\big(X_t^{(\beta)}\big) > \eta + \min_{\mathbb{R}^d} U\Big) \leq A \exp\big(U(x)/2\big)/t^p,$$

where $p = (1 - \theta D_1) \wedge (D_2 \eta') > 0$ and $(X_t^{(\beta)}, Y_t^{(\beta)})$ is the annealed BPS process starting from (x, y).

PROOF. The proof is postponed to Section S1.2 in the Supplementary Material [16]. □

Acknowledgements. Alain Durmus acknowledges support from Chaire BayeScale "P. Laffitte". Pierre Monmarché acknowledges support from the French ANR project ANR-12-JS01-0006-PIECE and the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 614492. Arnaud Guillin and Pierre Monmarché acknowledge support from the French ANR-17-CE40-0030-EFI-Entropy, flows, inequalities.

SUPPLEMENTARY MATERIAL

Supplement to "Geometric ergodicity of the Bouncy Particle Sampler" (DOI: 10.1214/19-AAP1552SUPP; .pdf). Supplementary information.

REFERENCES

[1] ANDRIEU, C., DE FREITAS, N., DOUCET, A. and JORDAN, M. I. (2003). An introduction to MCMC for machine learning. *Mach. Learn.* **50** 5–43.

[2] BIERKENS, J., BOUCHARD-CÔTÉ, A., DOUCET, A., DUNCAN, A. B., FEARNHEAD, P., LIENART, T., ROBERTS, G. and VOLLMER, S. J. (2018). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *Statist. Probab. Lett.* **136** 148–154. MR3806858 https://doi.org/10.1016/j.spl.2018.02.021

[3] 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

[4] 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

[5] BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. MR3655855 https://doi.org/10.1214/16-AAP1217

[6] 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

[7] BINDER, K., HEERMANN, D., ROELOFS, L., MALLINCKRODT, A. J. and MCKAY, S. (1993). Monte Carlo simulation in statistical physics. *Comput. Phys.* **7** 156–157.

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

[9] CALVEZ, V., RAOUL, G. and SCHMEISER, C. (2015). Confinement by biased velocity jumps: Aggregation of *Escherichia coli*. *Kinet. Relat. Models* **8** 651–666. MR3375486 https://doi.org/10.3934/krm.2015.8.651

[10] DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. CRC Press, London. MR1283589 https://doi.org/10.1007/978-1-4899-4483-2

[11] 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

[12] DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978 https://doi.org/10.1214/aoap/1019487508

[13] DOUC, R., GUILLIN, A. and MOULINES, E. (2008). Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 239–257. MR2446322 https://doi.org/10.1214/07-AIHP109

[14] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

[15] DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2018). Piecewise deterministic markov processes and their invariant measures.

[16] DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2020). Supplement to "Geometric ergodicity of the Bouncy Particle Sampler." https://doi.org/10.1214/19-AAP1552SUPP.

[17] EBERLE, A. (2016). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** 851–886. MR3568041 https://doi.org/10.1007/s00440-015-0673-1

[18] ERBAN, R. and OTHMER, H. G. (2004/05). From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.* **65** 361–391. MR2123062 https://doi.org/10.1137/S0036139903433232

[19] 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

[20] GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

[21] GOLDSTEIN, S. (1951). On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.* **4** 129–156. MR0047963 https://doi.org/10.1093/qjmam/4.2.129

[22] GREEN, P. J., ŁATUSZYŃSKI, K., PEREYRA, M. and ROBERT, C. P. (2015). Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput.* **25** 835–862. MR3360496 https://doi.org/10.1007/s11222-015-9574-5

[23] GUILLIN, A. (2000). Uniform moderate deviations of functional empirical processes of Markov chains. *Probab. Math. Statist.* **20** 237–260. MR1825641

[24] HAARIO, H. and SAKSMAN, E. (1991). Simulated annealing process in general state space. *Adv. in Appl. Probab.* **23** 866–893. MR1133733 https://doi.org/10.2307/1427681

[25] HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Birkhäuser/Springer Basel AG, Basel. MR2857021 https://doi.org/10.1007/978-3-0348-0021-1_7

[26] 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

[27] JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

[28] JOHNSON, L. T. and GEYER, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076. MR3097969 https://doi.org/10.1214/12-AOS1048

[29] KAC, M. (1974). A stochastic model related to the telegrapher's equation. *Rocky Mountain J. Math.* **4** 497–509. Reprinting of an article published in 1956, papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972). MR0510166 https://doi.org/10.1216/RMJ-1974-4-3-497

[30] KONTOYIANNIS, I. and MEYN, S. P. (2005). Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.* **10** 61–123. MR2120240 https://doi.org/10.1214/EJP.v10-231

[31] LEMAIRE, V., THIEULLEN, M. and THOMAS, N. (2018). Exact simulation of the jump times of a class of piecewise deterministic Markov processes. *J. Sci. Comput.* **75** 1776–1807. MR3798122 https://doi.org/10.1007/s10915-017-0607-4

[32] LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* **26** 403–413. MR0546120 https://doi.org/10.1002/nav.3800260304

[33] LINDVALL, T. and ROGERS, L. C. G. (1986). Coupling of multidimensional diffusions by reflection. *Ann. Probab.* **14** 860–872. MR0841588

[34] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

[35] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. With a prologue by Peter W. Glynn. MR2509253 https://doi.org/10.1017/CBO9780511626630

[36] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295 https://doi.org/10.2307/1427522

[37] 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.

[38] MICLO, L. and MONMARCHÉ, P. (2013). Étude spectrale minutieuse de processus moins indécis que les autres. In *Séminaire de Probabilités XLV. Lecture Notes in Math.* **2078** 459–481. Springer, Cham. MR3185926 https://doi.org/10.1007/978-3-319-00321-4_18

[39] MONMARCHÉ, P. (2016). Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 357–398. MR3487077

[40] NEAL, R. M. (2004). Improving asymptotic variance of MCMC estimators: Non-reversible chains are better Technical Report Technical Report No. 0406, Dept. of Statistics, University of Toronto.

[41] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, 113–162. CRC Press, Boca Raton, FL. MR2858447

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

[43] PITMAN, J. W. (1976). On coupling of Markov chains. *Z. Wahrsch. Verw. Gebiete* **35** 315–322. MR0415775 https://doi.org/10.1007/BF00532957

[44] REVUZ, D. and YOR, M. (1994). *Continuous Martingales and Brownian Motion*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1303781

[45] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

[46] ROSENTHAL, J. S. (1995). Rates of convergence for Gibbs sampling for variance component models. *Ann. Statist.* **23** 740–761. MR1345197 https://doi.org/10.1214/aos/1176324619

[47] ROSSKY, P. J., DOLL, J. D. and FRIEDMAN, H. L. (1978). Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.* **69** 4628–4633.

[48] TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. MR1329166 https://doi.org/10.1214/aos/1176325750

[49] TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D, Nonlinear Phenom.* **240** 410–414.

[50] WU, L. (2000). Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal.* **172** 301–376. MR1753178 https://doi.org/10.1006/jfan.1999.3544