GEOMETRIC ERGODICITY OF THE BOUNCY PARTICLE SAMPLER

Alain Durmus, Arnaud Guillin, Pierre Monmarché

GEOMETRIC ERGODICITY OF THE BOUNCY PARTICLE SAMPLER

ALAIN DURMUS, ARNAUD GUILLIN, PIERRE MONMARCHÉ

Abstract. 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 [10] 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 is a core requirement in many applications, e.g. in computational statistics [20], machine learning [1], molecular dynamics [6]. 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

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

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 [44].

While the first and best-known MCMC methods are based on reversible chains, such as many Metropolis-Hastings type algorithms [30], there has been since the last decade an increasing interest in non-reversible discrete-time processes [11, 3, 38, 34]. 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 and therefore the process shows 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 [36], 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 non-reversible but still leaves μ invariant. By definition of $\tilde{\mathbf{M}}$, the probability of backtracking is smaller than for $\mathbf{M}$, *i.e.* $\tilde{\mathbf{M}}^2_{i,i} \leqslant \mathbf{M}^2_{i,i}$ for any $i \in \{1, \dots, n\}$. In addition, [36] 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 non-reversible Markov chain is based on lifting, see [11] 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 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 [11] converges to the integrated telegraph on the continuous torus [34], while the lifted chain defined in [45] for spin models converges to the Zig-zag process [4] (see also the event-chain MC with infinitesimal steps in the physics

literature [33, 38]). In these cases, the continuous-time limits belong to the class of velocity jump processes $(X_t, Y_t)_{t\geqslant 0}$ 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 $t \geqslant 0$ with $(Y_t)_{t\geqslant 0}$ is piecewise-constant on random time intervals. The velocity $(Y_t)_{t\geqslant 0}$ acts as an instantaneous memory, or inertia, so that $(X_t)_{t\geqslant 0}$ 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 [25, 19], 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 [16, 8, 17] 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 [28, 7, 27]. 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 [43, 13, 37, 41] and the reference therein.

This work deals with the velocity jump process introduced in [38, 35]. Following [7], 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, relaxing the conditions given in [10]. 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 [35].

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.

Notations. 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 $\mathsf{S}^d = \left\{v \in \mathbb{R}^d : \|v\| = 1\right\}$, the d-dimensional sphere with radius 1 and for all $x \in \mathbb{R}^d$, $r > 0$, by $\mathrm{B}(x, r) = \left\{w \in \mathbb{R}^d : \|w - x\| \leqslant r\right\}$ 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 $k \in \mathbb{N}$, $\mathrm{C}^k_c(\mathbb{R}^d)$ and $\mathrm{C}^k_b(\mathbb{R}^d)$ the set of 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

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

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

$$\|\mu-\nu\|_V = \sup\left\{\left|\int_{\mathbb{R}^d} f\mathrm{d}\mu - \int_{\mathbb{R}^d} f\mathrm{d}\nu\right|\ :\ f:\mathbb{R}^d\to\mathbb{R} \text{ is Borel measurable and } |f|<V\right\}\ .$$

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 $\inf\emptyset=+\infty$.

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, *i.e.* for any rotation $O\in\mathbb{R}^{d\times d}$, $O\mathsf{Y}=\mathsf{Y}$. The BPS process $(X_t,Y_t)_{t\geqslant 0}$ 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 $(E_i^1,E_i^2,G_i)_{i\in\mathbb{N}^*}$ on the same probability space $(\Omega,\mathcal{F},\mathbb{P})$, where for all $i\in\mathbb{N}^*$, E_i^1,E_i^2 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^1, E_i^2 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 $(X_t,Y_t)_{t\leqslant S_n}$ have been defined for $n\geqslant 0$. Consider

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

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

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

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

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

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 $T_{n+1}=T_{n+1}^1$, we say that, at time T_{n+1}, the velocity has been refreshed, and we call T_{n+1} a refreshment time. If $T_{n+1}=T_{n+1}^2$, 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 $t\geqslant\sup_{n\in\mathbb{N}}S_n$, $(X_t,Y_t)=\infty$, where ∞ is a cemetery point.

In fact, it is proven in [14, Proposition 10] that almost surely, $\sup_{n\in\mathbb{N}} S_n = +\infty$. Therefore almost surely $(X_t, Y_t)_{t\geqslant 0}$ is a $(\mathbb{R}^d \times \mathsf{Y})$-valued càdlàg process. By [9, Theorem 25.5], the BPS process $(X_t, Y_t)_{t\geqslant 0}$ defines a strong Markov semi-group $(P_t)_{t\geqslant 0}$ 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

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

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

Consider the following basic assumption.

A1. *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 [14, Proposition 24] that, under **A** 1, the probability measure $\tilde{\pi}$ defined by (2), is invariant for $(P_t)_{t\geqslant 0}$, *i.e.* $\tilde{\pi} P_t = \tilde{\pi}$ for all $t \geqslant 0$.

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

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

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

Our basic assumptions to prove geometric ergodicity are the following.

A 2. *(i) The potential U satisfies $\lim_{\|x\|\to+\infty} U(x) = +\infty$ and $\int_{\mathbb{R}^d} \exp\left(-U(x)/2\right) \mathrm{d}x < +\infty$. Moreover, and without loss of generality, for all $x \in \mathbb{R}^d$, $U(x) > 0$.*

(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 $(P_t)_{t\geqslant 0}$ is V-uniformly geometrically ergodicity. In fact, these conditions are derived from a more general result. However, since the assumptions and statement of the latter can seem very intricate, for sake of clarity we have decided to give this result next to its application.

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

A 3. *The potential U satisfies*

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

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

$$\liminf_{\|x\|\to+\infty} \left\{\|\nabla U(x)\| \, / \, U^{1-\varsigma}(x)\right\} > 0 , \ \limsup_{\|x\|\to+\infty} \left\{\|\nabla U(x)\| \, / \, U^{1-\varsigma/2}(x)\right\} < +\infty ,$$

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

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

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

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

$$\nabla U(x) = [\alpha x_1(1+x_1^2)^{\alpha/2-1}, \beta x_2(1+x_2^2)^{\beta/2-1}]^{\mathrm{T}}] \ , \ \nabla^2 U(x) = \begin{pmatrix} F(\alpha,x_1) & 0 \\ 0 & F(\beta,x_2) \end{pmatrix}$$

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

In that case **A**4 is satisfied if and only if $[(\alpha\vee\beta)/2, \alpha\wedge\beta]\neq\emptyset$, while **A**5 is satisfied if and only if $[2(\alpha\vee\beta)/(1+\alpha\vee\beta), \alpha\wedge\beta]\neq\emptyset$, chosing in both cases $\varsigma^{-1}>1$ in the corresponding interval. In particular, if both $\alpha,\beta\geqslant 2$, 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* $(P_t)_{t\geqslant 0}$ *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* $V(x,y)=\exp\left(\kappa U^{\varsigma}(x)\right)$.

Proof. The proof is postponed to Section 3.5. □

The geometric ergodicity of BPS was also shown in [10, Theorem 3.1] under **A**3 and with the condition that λ_{r} is sufficiently large. Note that we do not impose this last condition in Theorem 1.

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

The following result is a generalization of [10, Theorem 3.1] which consider $\mathsf{Y}=\mathsf{S}^d$ and μ_{v} is the uniform distribution on S^d.

Theorem 2. *Assume* ***A*** *1,* ***A*** *2,* ***A*** *6 and* Y *is bounded. Then, there exists* $\lambda_0>0$ *such that, if* $\lambda_{\mathrm{r}}\in(0,\lambda_0]$, $(P_t)_{t\geqslant 0}$ *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* $V(x,y)=\exp(U(x)/2)$.

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.

In the case where Y is unbounded, **A**4 must be strengthen as follow.

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

$$0 \ < \ \liminf_{\|x\|\to+\infty}\left\{\|\nabla U(x)\| \, / \, U^{1-\varsigma}(x)\right\} \ \leqslant \ \limsup_{\|x\|\to+\infty}\left\{\|\nabla U(x)\| \, / \, U^{1-\varsigma}(x)\right\} \ < \ +\infty \ ,$$

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

A7 (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* $t\geqslant 1$ *and* $x\in\mathbb{R}^d$ *with* $\|x\|\geqslant 1$,

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

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

Then ***A*** *7 holds with* $\varsigma = 1/\alpha$.

Proof. The proof is postponed to Appendix A. □

This class of potentials is considered in [24, 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* ***A*** *1,* ***A*** *2 ,* ***A*** *7 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* $(P_t)_{t\geqslant 0}$ *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* $V(x,y) = \exp\left(\kappa U^{\varsigma}(x)\right) + \exp(\eta\|y\|^2)$.

Proof. The proof is postponed to Section 3.7. □

As noticed before, Theorem 1, Theorem 2 and Theorem 4 ensue from a 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* $i = 1, \ldots, 4$ *satisfying the following conditions.*

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

$$\lim_{\|x\|\to+\infty} \bar{U}(x) = +\infty \, , \qquad \int_{\mathbb{R}^d} \exp\left(\bar{U}(x) - U(x)\right) \mathrm{d}x < +\infty \tag{6}$$

$$\sup_{x\in\mathbb{R}^d} \left\{ \exp\left(-\bar{U}(x)/4\right) \left(\left\|\nabla \bar{U}(x)\right\| + \left\|\nabla^2 \bar{U}(x)\right\| \right) \right\} < +\infty \, , \tag{7}$$

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

$$\left\|\nabla \bar{U}(x)\right\| \ell(x) \geqslant c_1 \, , \qquad \ell(x) \leqslant c_2 \, , \qquad \|\nabla U(x)\| \, \ell(x) / \left\|\nabla \bar{U}(x)\right\| \geqslant c_3 \, . \tag{8}$$

(ii) ***Conditions on*** μ_{v}.

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

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

$$\mathsf{A}_x = \left\{ y \in \mathsf{Y} \, : \, H(\|y\|) \leqslant 3\bar{U}(x) \right\} \, . \tag{9}$$

Assume that

$$\lim_{\|x\|\to+\infty} \left[\|\nabla \ell(x)\| \left\{ 1 \vee \sup_{y\in\mathsf{A}_x} \|y\| \right\} \right] = 0 \, , \tag{10}$$

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

$$\left\|\nabla^2 \bar{U}(x)\right\| \ell(x) \left\{ \sup_{y\in\mathsf{A}_x} \|y\|^2 \right\} \leqslant c_4 \, . \tag{11}$$

Theorem 5. *Assume* ***A*** *1-****A*** *2-****A*** *8. Assume in addition that the following inequalities hold*

$$\begin{aligned} &[16\lambda_{\mathrm{r}} c_2/(rc_1)] \wedge [64 c_4 c_2/(rc_1)^2] \\ &\qquad \leqslant [(1/3) \wedge \{\lambda_{\mathrm{r}} \delta r c_1/(16 c_4)\}] \left[\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}} \delta c_3/(100 r c_1)\}^{1/2} \right] \, . \end{aligned} \tag{12}$$

Then there exists $\kappa \in (0,1]$ *given below by* (32)*, such that* $(P_t)_{t\geqslant 0}$ *is* V*-uniformly geometrically ergodic with* V *given for all* $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$ *by* $V(x,y) = \exp\left(\kappa \bar{U}(x)\right) + \exp(H(\|y\|))$.

Proof. The proof is postponed to Section 3.4. □

Remark 6. *Note that, under* ***A*** *8,* (12) *is implied by either one of the two following additional assumptions:*

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

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* (12) *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 $(P_t)_{t\geqslant 0}$ is V-uniformly geometrically ergodic then by [18, Theorem 4.4], a functional Central Limit Theorem (FCLT) holds. Let $g : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}$ satisfying for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $|g|^2 \leqslant CV$ for some $C > 0$. Let $(X_t, Y_t)_{t\geqslant 0}$ be a BPS process with initial distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^d \times \mathsf{Y})$, satisfying $\mu_0(V) < +\infty$. For $t \geqslant 0$ and $n \in \mathbb{N}_*$, define

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

Then, there exists $\sigma_g \geqslant 0$ such that the sequence of processes $\{(G_t^n)_{t\geqslant 0}, n \in \mathbb{N}\}$ converges as $n \to \infty$ toward $(\sigma_g B_t)_{t\geqslant 0}$ in the Skorokhod space, where $(B_t)_{t\geqslant 0}$ is a standard Brownian motion. It is also possible to consider moderate deviation [21, 12] or large deviation principle [46, 26]

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 [35, 10, 5], 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 Determistic Markov Processes (PDMP). Indeed, consider the ordinary differential equation on $\mathbb{R}^{2d}$

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} y_t \\ 0 \end{pmatrix}, \tag{13}$$

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

$$\phi_t(x,y) = (x + ty, y)\,. \tag{14}$$

The family $(\phi_t)_{t\in\mathbb{R}_+}$ is referred to as the flow of diffeomorphisms associated with (13) *i.e.* for all $(x,y) \in \mathbb{R}^{2d}$, $t \mapsto \phi_t(x,y)$ is solution of (13) started at (x,y) and for all $t \geqslant 0$, $(x,y) \mapsto \phi_t(x,y)$ 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

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

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 [9]) $(\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0}$ which has the same distribution as $(X_t, Y_t)_{t \geqslant 0}$ 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 [23, Chapter 6].

Consider some initial data $(x,y) \in \mathbb{R}^{2d}$, a family of i.i.d. random variables $(\tilde{E}_i, \tilde{G}_i, \tilde{W}_i)_{i \geqslant 1}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ introduced in Section 2.1, where for all $i \geqslant 1$, $\tilde{E}_i$ 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 $n \geqslant 0$, let

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

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

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

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 $t \geqslant \sup_{n \in \mathbb{N}} \tilde{S}_n$, $(\tilde{X}_t, \tilde{Y}_t) = \infty$, 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 [14, Lemma 7], $(\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0}$ and $(X_t, Y_t)_{t \geqslant 0}$ have the same distribution (in particular, almost surely $\sup_{n \in \mathbb{N}} \tilde{S}_n = \infty$ and $(\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0}$ is a $(\mathbb{R}^d \times \mathsf{Y})$-valued càdlàg process).

Consider the canonical process associated with the BPS process $(X_t, Y_t)_{t \geqslant 0}$, still denoted by $(X_t, Y_t)_{t \geqslant 0}$ on the Skorokhod space $(\mathrm{D}(\mathbb{R}_+, \mathbb{R}^d \times \mathsf{Y}), \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}, (\mathbb{P}_{x,y})_{(x,y) \in \mathbb{R}^d \times \mathsf{Y}})$, where $\mathcal{F}$ is the Borel σ-field associated with the Skorokhod topology, $(\mathcal{F}_t)_{t \geqslant 0}$ 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 $t \geqslant 0$ 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

$$M_t^{f,g} = f(X_t, Y_t) - f(X_0, Y_0) - \int_0^t g(X_s, Y_s) \mathrm{d}s \, . \tag{15}$$

The (extended) generator and its domain $(\mathcal{A}, \mathrm{D}(\mathcal{A}))$ associated with the semi-group $(P_t)_{t \geqslant 0}$ 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 $(M_t^{f,g})_{t \geqslant 0}$ 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 $(P_t)_{t \geqslant 0}$ can be easily described. Indeed, [9, Theorem 26.14] shows that $\mathrm{D}(\mathcal{A}) = \mathsf{E}_1 \cap \mathsf{E}_2$ where

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

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 $(\mathcal{F}_t)_{t \geqslant 0}$-stopping time $(\sigma_n)_{n \geqslant 0}$, 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 $n\in\mathbb{N}^*$,

$$\mathbb{E}_{(x,y)}\left[\sum_{k=1}^{+\infty}\mathbb{1}_{\{S_k\leqslant\sigma_n\}}\left|f(X_{S_k},Y_{S_k})-f(X_{S_k-},Y_{S_k-})\right|\right]<+\infty\ . \tag{16}$$

Taking for all $n\in\mathbb{N}^*$, $\sigma_n = S_n\wedge n\wedge\upsilon_n$, where $\upsilon_n=\inf\{t\geqslant 0\,:\,\|X_t\|\geqslant n\}$, (16) is satisfied for any continuous f. As a consequence, $\mathrm{C}(\mathbb{R}^d\times\mathsf{Y})\subset\mathrm{D}(\mathcal{A})$.

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

$$\begin{aligned}\mathcal{A}f(x,y) = D_yf(x,y) + (\langle y,\nabla U(x)\rangle)_+\left\{f(x,\mathrm{R}(x,y))-f(x,y)\right\}\\ +\lambda_{\mathrm{r}}\left\{\int_{\mathsf{Y}}f(x,w)\mathrm{d}\mu_{\mathrm{v}}(w)-f(x,y)\right\}\ ,\end{aligned} \tag{17}$$

where

$$D_yf(x,y)=\begin{cases}\lim_{t\to0}\frac{f(\varphi_t(x,y))-f(x,y)}{t}\ , & \text{if this limit exists}\\ 0 & \text{otherwise}\ .\end{cases}$$

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

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

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

$$\begin{aligned}\varphi(s)&=1 && \text{if } s\in(-\infty,-2]\\ 1+a(s+2)-\varepsilon\leqslant\varphi(s)&\leqslant 1+a(s+2)+\varepsilon && \text{if } s\in(-2,-1)\\ \varphi(s)&=1+b+s(b-a) && \text{if } s\in[-1,0]\\ 1+b+s(c-b)-\varepsilon\leqslant\varphi(s)&\leqslant 1+b+s(c-b)+\varepsilon && \text{if } s\in(0,1)\\ \varphi(s)&=1+c && \text{if } s\in[1,+\infty]\ ,\end{aligned} \tag{19}$$

and

$$\sup_{s\in[-2,-1]}\varphi'(s)\leqslant a+\varepsilon\ ,\ \sup_{s\in[0,1]}\varphi'(s)\leqslant c-b+\varepsilon\ . \tag{20}$$

In addition for $\kappa\in(0,1]$, under **A**8, define the Lyapunov function $V:\mathbb{R}^d\times\mathsf{Y}\to[1,+\infty)$ by

$$V(x,y)=\exp(\kappa\bar{U}(x))\varphi\left\{(2\ell(x)/(rc_1))\left\langle y,\nabla\bar{U}(x)\right\rangle\right\}+\exp(H(\|y\|))\ . \tag{21}$$

This section is devoted to the proof of a Foster-Lyapunov drift condition for the generator $\mathcal{A}$ given by (18) and the function V defined in (21).

Lemma 7. *Assume* ***A*** *1-****A*** *2-****A*** *8 and* (12) *hold. There exist* $a,b,c\in\mathbb{R}_+$, $a\leqslant b\leqslant c$, $c-b\leqslant b-a\leqslant a$, $\varepsilon\in(0,1]$ *and* $\kappa\in(0,1]$ *such that* $\mathcal{A}$ *given by* (18) *satisfies a Foster-Lyapunov drift condition with the Lyapunov function* V*, i.e. there exist* $A_1,A_2>0$ *such that, for all* $(x,y)\in\mathbb{R}^d\times\mathsf{Y}$,

$$\mathcal{A}V(x,y)\leqslant A_1\left(A_2-V(x,y)\right)\ . \tag{22}$$

Proof. For ease of notation, we denote in the following $\theta(x,y) = \left\langle \nabla \bar{U}(x), y \right\rangle$ for any $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$. From (18) and the facts that $\nabla \bar{U}(x) = \psi'(U(x)) \nabla U(x)$ and $\|\mathrm{R}(x,y)\| = \|y\|$, for any $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$,

$$\mathcal{A}V(x,y) = \exp\left(\kappa \bar{U}(x)\right) J(x,y) + \lambda_{\mathrm{r}} \left\{ \int_{\mathsf{Y}} \exp\left(H(\|w\|)\right) \mu_{\mathrm{v}}(\mathrm{d}w) - \exp\left(H(\|y\|)\right) \right\} , \tag{23}$$

where

$$\begin{aligned} J(x,y) &= \kappa \theta(x,y) \varphi\left\{2\ell(x)\theta(x,y)/(rc_1)\right\} \\ &+ (2/(rc_1)) \varphi'\left\{2\ell(x)\theta(x,y)/(rc_1)\right\} \left[\ell(x) \left\langle y, \nabla^2 \bar{U}(x) y \right\rangle + \theta(x,y) \left\langle \nabla \ell(x), y \right\rangle\right] \\ &+ \left\{\|\nabla U(x)\| / \left\|\nabla \bar{U}(x)\right\|\right\} \{\theta(x,y)\}_+ \left[\varphi\left\{-2\ell(x)\theta(x,y)/(rc_1)\right\} - \varphi\left\{2\ell(x)\theta(x,y)/(rc_1)\right\}\right] \\ &+ \lambda_{\mathrm{r}} \left\{ \int_{\mathsf{V}} \varphi\left\{(2\ell(x)/(rc_1)) \left\langle \nabla \bar{U}(x), w \right\rangle\right\} \mathrm{d}\mu_{\mathrm{v}}(w) - \varphi\left\{2\ell(x)\theta(x,y)/(rc_1)\right\} \right\} . \end{aligned} \tag{24}$$

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

$$\mathcal{A}V(x,y) \leqslant -A_{1,1} V(x,y) + A_{1,2} \text{ for any } (x,y) \in \mathbb{R}^d \times \mathsf{Y},\ y \notin \mathsf{A}_x , \tag{25}$$

where $\mathsf{A}_x \subset \mathsf{Y}$ is defined by (9). In a second step, we show that there exist $A_{2,1}, A_{2,2} > 0$ such that

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

Note that if (25) and (26) hold, then the proof is concluded.

Proof of (25). Let $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $y \notin \mathsf{A}_x$. From (24) and the facts that φ is bounded by $1+c$, that $\varphi(-s) - \varphi(s) \leqslant 0$ for any $s \in \mathbb{R}_+$ since φ is non-decreasing, and that $\sup_{s \in \mathbb{R}} \varphi'(s) \leqslant (a+\varepsilon) \vee b \vee ((c-b)+\varepsilon) \leqslant 1 + c$ since $\varepsilon \leqslant 1$, we have

$$\begin{aligned} &J(x,y) \\ &\leqslant (1+c) \left[\kappa \left\|\nabla \bar{U}(x)\right\| \|y\| + (2/(rc_1)) \left\{ \|y\| \|\nabla \ell(x)\| + \ell(x) \|y\|^2 \left\|\nabla^2 U(x)\right\| \right\} + \lambda_{\mathrm{r}} \right] . \end{aligned} \tag{27}$$

By (8) and (10) and since $\ell \in \mathrm{C}^1(\mathbb{R}^d)$, $\|\nabla \ell\|_\infty + \|\ell\|_\infty < \infty$. Therefore plugging (27) in (24) and using (7) and **A**8-(ii), we get

$$\begin{aligned} &\mathcal{A}V(x,y) \leqslant C_1 (1 \vee \|y\|^2) \exp(5\bar{U}(x)/4) + C_2 - \lambda_{\mathrm{r}} \exp(H(\|y\|)) , \\ &C_1 = (1+c) \left\{ (\kappa \|\nabla \bar{U} \mathrm{e}^{-\bar{U}/4}\|_\infty) \vee \frac{2\|\nabla \ell\|_\infty}{rc_1} \vee \lambda_{\mathrm{r}} \vee \frac{2\|\nabla^2 \bar{U} \mathrm{e}^{\bar{U}/4}\|_\infty \|\ell\|_\infty}{rc_1} \right\} < +\infty , \\ &C_2 = \lambda_{\mathrm{r}} \int_{\mathsf{Y}} \exp(H(\|y\|)) \mathrm{d}\mu_{\mathrm{v}}(w) < +\infty . \end{aligned} \tag{28}$$

Using now **A**8-(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 $y \notin \mathsf{A}_x$, $3\bar{U}(x) \leqslant H(\|y\|)$ and we obtain

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

The proof of (25) follows upon noting that $\kappa \leqslant 1$ and that φ is bounded by $1+c$, so that $V(x,y) \leqslant (2+c) \exp(H(\|y\|))$ if $y \notin \mathsf{A}_x$.

Proof of (26). We show in Lemma 8 that there exist $a, b, c \in \mathbb{R}_+$, $a \leqslant b \leqslant c$, $\varepsilon \in (0,1]$, $\kappa \in (0,1)$, $R_1 \in \mathbb{R}_+$ and $\eta \in \mathbb{R}_+^*$ such that for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $y \in \mathsf{A}_x$ and $\|x\| \geqslant R_1$, $J(x,y) < -\eta$. Note that if this result holds, then for all $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $y \in \mathsf{A}_x$ and $\|x\| \geqslant R_1$ by (23),

$$\mathcal{A}V(x,y) \leqslant -\eta \exp(\kappa \bar{U}(x)) + C_2 - \lambda_{\mathrm{r}} \exp(H(\|y\|)) \leqslant -((\eta/(1+c)) \wedge \lambda_{\mathrm{r}}) V(x,y) + C_2 , \tag{29}$$

where C_2 is given by (28) and we have used for the last inequality that φ is bounded by $1+c$. This result concludes the proof of (26) for $\|x\| \geqslant R_1$. It remains to consider the case $\|x\| \leqslant R_1$.

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 $y \in \mathsf{A}_x$, $H(\|y\|) \leqslant M_1$. Since $\sup_{w\in\mathsf{Y}} \|w\|^2 \, \mathrm{e}^{-H(\|w\|)} < +\infty$ by **A** 8-(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 $y \in \mathsf{A}_x$, $\mathcal{A}V(x,y) \leqslant C_5$ and $V(x,y) \leqslant C_6$. Combining this result and (29) concludes the proof of (26). □

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

$$a = 1 \wedge \left([(1/3) \wedge \{\lambda_{\mathrm{r}} \delta r c_1/(16 c_4)\}] \left[\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}} \delta c_3/(100 r c_1)\}^{1/2} \right] \right)^{-1} \tag{30}$$

$$b - a = a \, [(1/3) \wedge \{\lambda_{\mathrm{r}} \delta r c_1/(16 c_4)\}] \tag{31}$$

$$\begin{aligned} \kappa &= (b-a) \left[\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}} \delta c_3/(100 r c_1)\}^{1/2} \right] \\ &= a \, [(1/3) \wedge \{\lambda_{\mathrm{r}} \delta r c_1/(16 c_4)\}] \left[\{c_3/(4c_2)\} \wedge \{\lambda_{\mathrm{r}} \delta c_3/(100 r c_1)\}^{1/2} \right] \end{aligned} \tag{32}$$

$$c - b = [\delta \lambda_{\mathrm{r}} a/(4(4c_4/(rc_2) + 2\lambda_{\mathrm{r}}))] \wedge (b-a) \wedge [(b-a)c_3/(4\kappa c_2)] \wedge (\delta b/4) \tag{33}$$

$$\varepsilon = (1/2) \wedge (c-b) \wedge (\kappa r c_1/4) \wedge (\lambda_{\mathrm{r}} c_2) \, . \tag{34}$$

Note that $\kappa \leqslant 1$ and

$$0 \leqslant c - b \leqslant b - a \leqslant a \leqslant 1 \, . \tag{35}$$

Lemma 8. *Assume* ***A*** *1-****A*** *2-****A*** *8 and* (12) *hold. Then for* $a, b, c, \kappa, \varepsilon \in (0,1]$, *given in* (30)-(31)-(33)-(32)-(34) *respectively, there exist* $\tilde{R}, \eta > 0$ *such that for all* $x \in \mathbb{R}^d$ *with* $\|x\| \geqslant \tilde{R}$ *and all* $y \in \mathsf{A}_x$, $J(x,y) < -\eta$, *where* J *and* φ *are defined by* (24) *and* (19) *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, 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. For ease of notation, we denote for any $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $\theta(x,y) = \langle \nabla \bar{U}(x), y \rangle$ again.

By (10), there exists $R_1 \in \mathbb{R}_+$ such that for any $(x,y) \in \mathbb{R}^d$, $y \in \mathsf{A}_x$, $\|x\| \geqslant R_1$,

$$\|\nabla \ell(x)\| \, \|y\| \leqslant \varepsilon \, . \tag{36}$$

From (8), $\left\|\nabla \bar{U}(x)\right\| \ell(x) \geqslant c_1$ for all $x \in \mathbb{R}^d$ with $\|x\| \geqslant R$. Using **A** 8-(ii) and the facts that μ_{v} is rotation invariant and that φ is non-decreasing, bounded by $1+c$ and equal to 1 on $(-\infty, 2]$, we then have for any $x \in \mathbb{R}^d$ with $\|x\| \geqslant R$

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

Therefore, combining this result, (36), (11) and the fact that φ is non-decreasing so that $\varphi'(s) \geqslant 0$ for any $s \in \mathbb{R}$, we get, for any $x \in \mathbb{R}^d$ with $\|x\| \geqslant R_2 = R \vee R_1$ and all $y \in \mathsf{A}_x$,

$$\begin{aligned} J(x,y) &\leqslant \kappa \theta(x,y) \varphi\left\{2\ell(x)\theta(x,y)/(rc_1)\right\} + (2/(rc_1)) \varphi'\left\{2\ell(x)\theta(x,y)/(rc_1)\right\} \left[c_4 + |\theta|\,(x,y)\varepsilon\right] \\ &+ \left\{\|\nabla U(x)\| \,/\, \left\|\nabla \bar{U}(x)\right\|\right\} \{\theta(x,y)\}_+ \left[\varphi\left\{-2\ell(x)\theta(x,y)/(rc_1)\right\} - \varphi\left\{2\ell(x)\theta(x,y)/(rc_1)\right\}\right] \\ &+ \lambda_{\mathrm{r}} \left\{1 + (1-\delta/2)c - \varphi\left\{2\ell(x)\theta(x,y)/(rc_1)\right\}\right\} \, . \end{aligned} \tag{37}$$

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

Case 1 : $2\ell(x)\theta(x,y)/(rc_1) \in (-\infty,-2]$. Since for $s \in (-2,-\infty]$, $\varphi(s)=1$, (37) reads

$$J(x,y) \leqslant \kappa\theta(x,y) + (1-\delta/2)\lambda_{\mathrm{r}} c \,. \tag{38}$$

Using the facts that $2\ell(x)\theta(x,y)/(rc_1) \in (-\infty,-2]$, that $\ell(z) \leqslant c_2$ for all $z \in \mathbb{R}^d$ by (8), that $(b-a)\vee(c-b) \leqslant a$ by (35), that $a \leqslant rc_1\kappa/(6\lambda_{\mathrm{r}}c_2)$ by (32) and that (12) holds, we get

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

By this result and (38), we obtain

$$J(x,y) \leqslant -rc_1\kappa/(2c_2) \,. \tag{39}$$

Case 2 : $2\ell(x)\theta(x,y)/(rc_1) \in (-2,-1)$. By (19)-(20), $1+2a+sa-\varepsilon \leqslant \varphi(s) \leqslant 1+2a+sa+\varepsilon$ and $\varphi'(s) \leqslant a+\varepsilon$ for $s \in (-2,-1)$, so that (37) reads

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

where we have used that $2\ell(x)\theta(x,y)/(rc_1) \in [-2,-1]$ and that $\ell(x) \leqslant c_2$ by (8), and defined

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

First, (34) and (35) ensures that $\varepsilon \leqslant (1/2)\wedge a \wedge (\lambda_{\mathrm{r}}c_2)$, and therefore

$$B_1 - 2\,B_2 \geqslant \kappa/2 - 4\lambda_{\mathrm{r}} a c_2/(rc_1) \geqslant \kappa/4 \,,$$

where we have used that $a \leqslant rc_1\kappa/(16\lambda_{\mathrm{r}}c_2)$ for the last inequality, which is a consequence of (32) and (12). In particular, $B_1 \geqslant 2B_2$ and using again that $2\ell(x)\theta(x,y)/(rc_1) \in (-2,-1)$ and $\ell(x) \leqslant c_2$ from (8), then

$$J(x,y) \leqslant B_0 + (rc_1/(2c_2))(2B_2 - B_1) \leqslant B_0 - rc_1\kappa/(8c_2) \,. \tag{40}$$

Since $\varepsilon \leqslant a \wedge (c-b)$ by (34), $c-b \leqslant b-a$ by (33) and $b-a \leqslant a/3$ by (31), we have $B_0 \leqslant 4ac_4/(rc_1)$. Hence, (40) becomes

$$J(x,y) \leqslant 4ac_4/(rc_1) - rc_1\kappa/(8c_2) \leqslant -rc_1\kappa/(16c_2) \,, \tag{41}$$

where we have used (32) and (12) 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 (19), (37) reads

$$\begin{aligned} J(x,y) &\leqslant \kappa\theta(x,y)\{1+b+(b-a)2\ell(x)\theta(x,y)/(rc_1)\} + (2(b-a)/(rc_1))\{c_4 - \theta(x,y)\varepsilon\} \\ &\quad + \lambda_{\mathrm{r}}\{(1-\delta/2)c - b - 2\ell(x)\theta(x,y)(b-a)/(rc_1)\} \\ &\leqslant B_0 + B_1\theta(x,y) + B_2 2\ell(x)/(rc_1)\theta(x,y)^2 \leqslant B_0 + (B_1 - B_2)\theta(x,y) \,, \end{aligned} \tag{42}$$

where we have used that $2\ell(x)\theta(x,y)/(rc_1) \in [-1,0]$ and $\ell(x) \leqslant c_2$ by (8), 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 $c-b \leqslant \delta b/4 \leqslant \delta c/4$ and $a \leqslant c$ by (33) and (35), we have

$$B_0 \leqslant 2(b-a)c_4/(rc_1) - \lambda_{\mathrm{r}}\delta c/4 \leqslant 2(b-a)c_4/(rc_1) - \lambda_{\mathrm{r}}\delta a/4 \leqslant -a\lambda_{\mathrm{r}}\delta/8 \,, \tag{43}$$

where we have used that $b-a \leqslant \lambda_r \delta a r c_1/(16 c_4)$ by (31) for the last inequality. Second, from the facts that $\varepsilon \leqslant \lambda_r c_2$ by (34) and $(b-a) \leqslant a/3 \leqslant 1/3$ by (31)-(30), we have

$$B_2 - B_1 \leqslant \kappa(b-a) + 4\lambda_r c_2 (b-a)/(rc_1) - \kappa(1+b) \leqslant 4\lambda_r c_2 a/(rc_1) - \kappa \leqslant 0 \,, \tag{44}$$

where we used the definition of κ (32) and the condition (12) for the last inequality. Combining (43) and (44) in (42), we get

$$J(x,y) \leqslant -a\lambda_r \delta/8 \tag{45}$$

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 non-decreasing, we have for any $s \in [0,1]$,

$$\varphi(-s) - \varphi(s) \leqslant \varphi(-s) - \varphi(0) \leqslant -(b-a)s \,.$$

From this result and the fact by (19)-(20) that $1+b+s(c-b)-\varepsilon \leqslant \varphi(s) \leqslant 1+b+s(c-b)+\varepsilon$ and $\varphi'(s) \leqslant c-b+\varepsilon$ for $s \in (0,1)$ we get that (37) reads

$$\begin{aligned} J(x,y) &\leqslant \kappa\theta(x,y)\left\{1+b+2\ell(x)\theta(x,y)(c-b+\varepsilon)/(rc_1)+\varepsilon\right\} \\ &\quad + (2(c-b+\varepsilon)/(rc_1))\left\{c_4+\theta(x,y)\varepsilon\right\} - (\|\nabla U(x)\| / \left\|\nabla \bar{U}(x)\right\|) 2\ell(x)(b-a)\theta(x,y)^2/(rc_1) \\ &\quad + \lambda_r\{1+(1-\delta/2)c-1-b-2\ell(x)\theta(x,y)(c-b)/(rc_1)+\varepsilon\} \\ &\leqslant B_0 + B_1\theta(x,y) + 2\ell(x)B_2\theta(x,y)^2/(rc_1) \,, \end{aligned}$$

where we have used that $(\|\nabla U(x)\| / \left\|\nabla \bar{U}(x)\right\|)\ell(x) \geqslant c_3$ by (8), $\theta(x,y) \geqslant 0$ and defined

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

Since $\varepsilon \leqslant c-b$ by (34), $\ell(x) \leqslant c_2$ by (8) and $2\kappa c_2(c-b) \leqslant c_3(b-a)/2$ by (33), we get

$$B_2 \leqslant -\tilde{B}_2 = -c_3(b-a)/(2\ell(x)) \,, \tag{46}$$

and therefore

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

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

$$J(x,y) \leqslant B_0 + \theta(x,y) r c_1 B_1^2/(4\ell(x)\tilde{B}_2)$$

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

$$J(x,y) \leqslant -\lambda_r \delta c/16 \,, \tag{47}$$

it is sufficient to prove that

$$B_0 \leqslant -\lambda_r \delta c/4 \tag{48}$$

$$r c_1 B_1^2/(4\ell(x)\tilde{B}_2) \leqslant \lambda_r \delta c/8 \,. \tag{49}$$

First (48) holds since using that $\varepsilon \leqslant (c-b)$ by (34) and that $a \leqslant c$, we have

$$\begin{aligned} B_0 - \delta/4 &= 2c_4(c-b+\varepsilon)/(rc_1) + \lambda_r\{(1-\delta/4)c - b + \varepsilon\} \\ &\leqslant (4c_4/(rc_2) + 2\lambda_r)(c-b) - \delta a \lambda_r/4 \leqslant 0 \,, \end{aligned}$$

using $(c-b) \leqslant \delta a \lambda_r/(4(4c_4/(rc_2)+2\lambda_r))$ by (33) for the last inequality. It remains to establish (49) which is equivalent by definition of B_1 and $\tilde{B}_2$ (46) to

$$\kappa(1+b+\varepsilon) + 2\varepsilon(c-b+\varepsilon)/(rc_1) \leqslant \{\lambda_r c \delta c_3 (b-a)/(4rc_1)\}^{1/2} \,. \tag{50}$$

Since $\varepsilon \leqslant 1 \wedge (\kappa r c_1/4)$ by (34), $c-b \leqslant 1$ and $b \leqslant 2$ by (35) and (30), we get

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

This result, the inequality $b-a \leqslant c$ and the definition of κ (32) implies that (50) holds.

Case 5 : $2\ell(x)\theta(x,y)/(rc_1) \geqslant 1$. Since by (19), $\varphi(s) = 1+c$, $\varphi'(s) = 0$ and $\varphi(-s)-\varphi(s) \leqslant a-c$ for $s \geqslant 1$, (37) reads

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

where we used successively that by (8) $\ell(x) \leqslant c_2$ and $\|\nabla U(x)\|\, \ell(x)/\left\|\nabla \bar{U}(x)\right\| \geqslant c_3$. Since $c \leqslant 3$ by (35) we have

$$\begin{aligned} J(x,y) &\leqslant \{\kappa(1+c) - c_3(c-a)/c_2\}\theta(x,y) - \lambda_{\mathrm{r}}\delta c/2 \\ &\leqslant \{4\kappa - c_3(b-a)/c_2\}\theta(x,y) - \lambda_{\mathrm{r}}\delta c/2 \leqslant -\lambda_{\mathrm{r}}\delta c/2 \;, \end{aligned} \tag{51}$$

where we have used the definition of κ given by (32) and $\theta(x,y) \geqslant 0$ for the last inequality.

The proof follows from combining (39)-(41)-(45)-(47)-(51). □

Corollary 9. *Under* ***A*** *8, for all* $(x,y) \in \mathbb{R}\times\mathsf{Y}$ *and* $t \geqslant 0$,

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

where V *is given by* (21) *and* A_1, A_2 *are given by Lemma 7.*

Proof. By [9, Section 31.5], since $V \in \mathrm{D}(\mathcal{A})$, the process $(M_t)_{t\geqslant 0}$, defined for any $t \in \mathbb{R}_+$ by

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

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

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

Letting n go to infinity concludes the proof since it yields

$$\mathrm{e}^{A_1 t}\mathbb{E}\left[V(X_t,Y_t)\right] \;\leqslant\; V(x,y) + A_2\left(\mathrm{e}^{A_1 t}-1\right) \;.$$

□

3.3. **Mirror Coupling.** To obtain geometric ergodicity, the classical Meyn and Tweedie approach is, once a Lyapunov drift condition holds, to show that some sets $\mathsf{C} \subset \mathbb{R}^d \times \mathsf{Y}$ are small sets: there exist $t > 0$, $\varepsilon > 0$ and $\nu \in \mathcal{P}(\mathbb{R}^d \times \mathsf{Y})$, $\nu(\mathsf{C}) = 1$, such that

$$P_t((x,y),\mathsf{A}) \geqslant \varepsilon\nu(\mathsf{A}) \;, \text{ for all } \mathsf{A} \subset \mathsf{C} \;.$$

It is commonly known that this condition is equivalent to: there exist $t > 0$, $\varepsilon > 0$ such that for all $(x,y),(\tilde{x},\tilde{y}) \in \mathsf{C}$,

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

This section is devoted to the proof of the following result:

Lemma 10. *Assume* ***A*** *1 and* ***A*** *2-(ii). Then, any compact set* $\mathsf{K} \subset \mathbb{R}^d \times \mathsf{Y}$ *is a small set.*

Previous works [35, 10] show that Lemma 10 hold in case where $\mathsf{Y} = \mathsf{S}^d$. The proof consists in establishing that the occurrence of more than two refreshment events suffices for the distribution of X_t, $t \geqslant 0$, to have some density w.r.t. the Lebesgue density on a ball with a radius proportional to t. Nevertheless, the latter strategy usually yields a non-explicit 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 [35, Lemma 5.2] or [10, Lemma 2]. We will only treat the non-singular case, with a particular emphasis on the case where μ_{v} is a d-dimensional non-degenerate Gaussian distribution with zero-mean and covariance matrix Σ.

Before stating our main result, we need the following lemma concerning the reflexion coupling (see [29], [15] 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 be a positive definite matrix and $(W_t^{(1)})_{t\geqslant 0}$ be a standard one dimensional Brownian motion. Define $T_{\mathrm{c}} = \inf\{t \geqslant 0 : W_t^{(1)} \geqslant \|\Sigma^{-1/2}(x^{(2)} - x^{(1)})\|/2\}$, the stochastic process $(W_t^{(2)})_{t\geqslant 0}$ by*

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

and the d-dimensional random variables

$$G^{(1)} = W_1^{(1)} \mathrm{n}\left\{\Sigma^{-1/2}(x^{(2)} - x^{(1)})\right\} + \bar{G} , \quad G^{(2)} = W_1^{(2)} \mathrm{n}\left\{\Sigma^{-1/2}(x^{(2)} - x^{(1)})\right\} + \bar{G}$$
$$\bar{G} = \left(\mathrm{Id} - \mathrm{n}\left\{\Sigma^{-1/2}(x^{(2)} - x^{(1)})\right\} \mathrm{n}\left\{\Sigma^{-1/2}(x^{(2)} - x^{(1)})\right\}^{\mathrm{T}}\right) G ,$$

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

$$\mathbb{P}\left(x^{(1)} + \Sigma^{1/2}G^{(1)} = x^{(2)} + \Sigma^{1/2}G^{(2)} , \left\|G^{(1)} - \Sigma^{-1/2}(x^{(2)} - x^{(1)})/2\right\| \leqslant M\right)$$
$$= \tilde{\alpha}(\|\Sigma^{-1/2}(x^{(2)} - x^{(1)})\|, M) ,$$

where for all $r \geqslant 0$,

$$\tilde{\alpha}(r, M) = \frac{r}{2(2\pi)^{(d+1)/2}} \int_0^1 \left\{ s^{-3/2} \exp\left(-r^2/(8s)\right) \right.$$
$$\left. \int_{\mathbb{R}^d} \mathbb{1}_{[0,M]}\left(\left((1-s)w_1^2 + \cdots + w_d^2\right)^{1/2}\right) \mathrm{e}^{-\|x\|^2/2} \mathrm{d}w \right\} \mathrm{d}s . \tag{52}$$

Proof. Without loss of generalities, we can assume that $\Sigma = \mathrm{Id}$. By the Markov property of the Brownian motion $(W_t^{(1)})_{t\geqslant 0}$, since T_c is a $(\mathcal{F}_t^W)_{t\geqslant 0}$-stopping time, where $\mathcal{F}_t^W = \sigma(W_s^{(1)}, s \leqslant t)$, $W_t^{(2)}$ is a Brownian motion. Therefore, $G^{(1)}$ and $G^{(2)}$ are d-dimensional standard Gaussian random variables.

Using again the Markov property of $(W_t^{(1)})_{t\geqslant 0}$, given $T_{\mathrm{c}} < 1$, $W_1^{(1)} - W_{T_{\mathrm{c}}}^{(1)}$ is independent of $\mathcal{F}_{T_{\mathrm{c}}}^W$. Therefore, since $\{x^{(1)} + G^{(1)} = x^{(2)} + G^{(2)}\} = \{T_{\mathrm{c}} \leqslant 1\}$ and G is independent of

$(W_t^{(1)})_{t\geqslant 0}$, we get for all $M \geqslant 0$,

$$\begin{aligned}
&\mathbb{P}\left(x^{(1)} + \Sigma^{1/2} G^{(1)} = x^{(2)} + \Sigma^{1/2} G^{(2)}\,, \left\| G^{(1)} - \Sigma^{-1/2}(x^{(2)} - x^{(1)})/2 \right\| \leqslant M\right) \\
&\quad = \mathbb{E}\left[\mathbb{1}_{[0,1]}(T_{\mathrm{c}})\mathbb{P}\left(\left((W_1^{(1)} - W_{T_{\mathrm{c}}}^{(1)})^2 + \|\bar{G}\|^2\right)^{1/2} \leqslant M \middle| \mathcal{F}_{T_{\mathrm{c}}}^W\right)\right] \\
&\quad = (2\pi)^{-d/2}\mathbb{E}\left[\mathbb{1}_{[0,1]}(T_{\mathrm{c}}) \int_{\mathbb{R}^d} \mathbb{1}_{[0,M]}\left\{\left((1-T_{\mathrm{c}})w_1^2 + \cdots + w_d^2\right)^{1/2}\right\} \mathrm{e}^{-\|x\|^2/2} \mathrm{d}w\right] .
\end{aligned}$$

The proof then follows from the explicit expression of the density of T_{c} w.r.t. the Lebesgue measure (see e.g. [40, p. 107]). □

Lemma 12. *Assume* ***A*** *1,* $\mathsf{Y} = \mathbb{R}^d$ *and* μ_{v} *is the Gaussian measure with zero-mean and covariance matrix* Σ*. Then, for all* $t > 0$ *and all compact set* $\mathsf{K} \subset \mathrm{B}(0, R_{\mathsf{K}})$ *of* $\mathbb{R}^d \times \mathsf{Y}$*,* $R_{\mathsf{K}} \geqslant 0$*, for all* $(x,y),(\tilde{x},\tilde{y}) \in \mathsf{K}$ *and for all* $M \geqslant 0$*,*

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

where $\tilde{\alpha}$ *is given by* (52)*, for all* $r \geqslant 0$*,*

$$\begin{aligned}
g(r) &= \mathbb{P}\left(rM \sup_{z \in \mathrm{B}(0,(1+E_1/\lambda_{\mathrm{r}})R_{\mathsf{K}} + (r/\lambda_{\mathrm{r}})\tilde{M})} \|\nabla U(z)\| \geqslant E_3\right) , \\
\tilde{M} &= M + \|\Sigma^{1/2}\|(1 + E_1/\lambda_{\mathrm{r}})R_{\mathsf{K}} , \qquad (53)
\end{aligned}$$

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 $P_t((x,y),\cdot)$ 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 caracterization 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 Fig. 1, 2 and 3). We couple both processes to have the same two first refreshment times S_1 and S_2. At time S_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 S_2. At time S_2, both velocities are refreshed with the same Gaussian variable. Hence, with positive probability, at time S_2, the processes have the same position and same velocity, in which case we can keep them equal for all times $t \geqslant S_2$.

More precisely, the coupling we consider is defined as follows. Let $(\bar{E}_i^1, \bar{E}_i^2, \bar{E}_i^3, \bar{G}_i)_{i\in\mathbb{N}^*}$ be i.i.d. random variables, where for all $i \in \mathbb{N}^*$, $\bar{E}_i^1, \bar{E}_i^2, \bar{E}_i^3$ are independent exponential random variables with parameter 1 and $\bar{G}_i$ has distribution μ_{v} and is independent of $\bar{E}_i^j$, $j \in \{1,2,3\}$.

Set $(X_0,Y_0) = (x,y)$, $(\tilde{X}_0,\tilde{Y}_0) = (\tilde{x},\tilde{y})$, $S_0^{\mathrm{a}} = S_0 = \tilde{S}_0 = 0$, $N_0 = 0$, $H_1 = \bar{E}_1^1/\lambda_{\mathrm{r}}$ and $N_1 = 1$. We define then by recursion the jump times of the process and the process between these jump times. For $n \geqslant 0$, we consider $\mathsf{A}_n = \{(X_{S_n^{\mathrm{a}}}, Y_{S_n^{\mathrm{a}}}) = (\tilde{X}_{S_n^{\mathrm{a}}}, \tilde{Y}_{S_n^{\mathrm{a}}})\}$ and distinguish two cases.

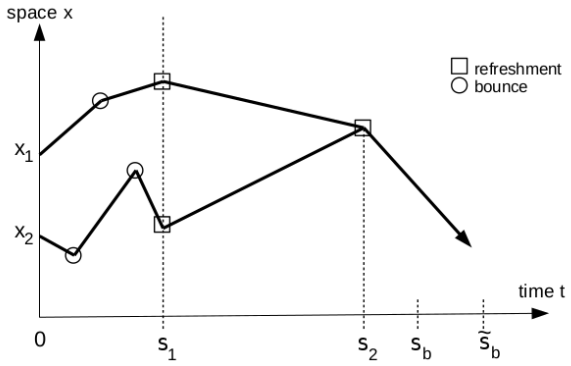

FIGURE 1. Before the first refreshment at time S_1, both processes may bounce freely. At time S_1, the Gaussian velocities are coupled so that, at time S_2 (which is the next refreshment time), provided this Gaussian coupling of the velocities succeeds, and provided they haven't bounced in the meanwhile (i.e. $S_2 < S_b \wedge \widetilde{S}_b$), both processes reach the same position. At time S_2, both processes take the same velocity: they have merged, the coupling is a success.

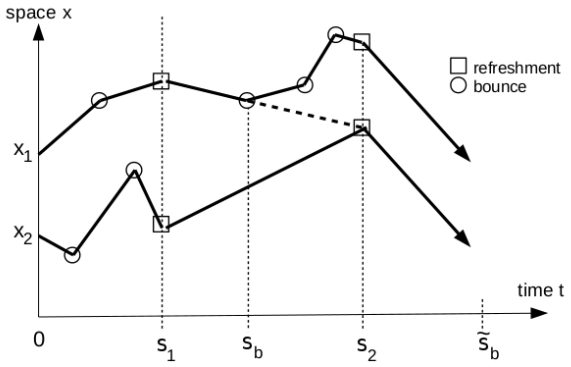

FIGURE 2. If one (at least) of the processes bounces between times S_1 and S_2, then the coupling fails. There may be other bounces after the first one.

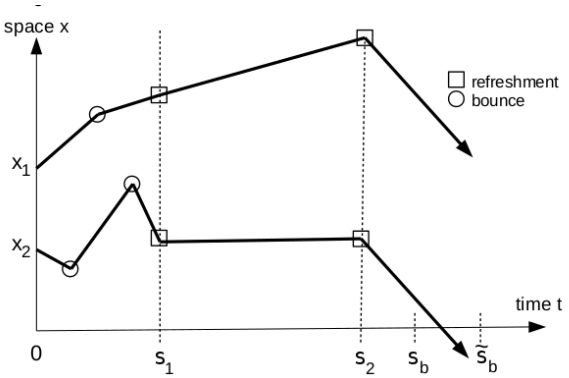

FIGURE 3. Even if none of the process bounces between time S_1 and S_2, the coupling may also fail if the Gaussian coupling of the velocities at time S_1 fails.

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

$$T^1_{n+1} = \inf\left\{ t \geqslant 0 \, : \, \int_0^t \left\{ \left\langle Y_{S^{\mathrm{a}}_n}, \nabla U(X_{S^{\mathrm{a}}_n} + sY_{S^{\mathrm{a}}_n}) \right\rangle_+ \right\} \mathrm{d}s \geqslant \bar{E}^1_{n+1} \right\} ,$$

$$\tilde{T}^1_{n+1} = \inf\left\{ t \geqslant 0 \, : \, \int_0^t \left\{ \left\langle \tilde{Y}_{S^{\mathrm{a}}_n}, \nabla U(\tilde{X}_{S^{\mathrm{a}}_n} + s\tilde{Y}_{\bar{S}^{\mathrm{a}}_n}) \right\rangle_+ \right\} \mathrm{d}s \geqslant \bar{E}^2_{n+1} \mathbb{1}_{\mathsf{A}^{\mathrm{c}}_n} + \bar{E}^1_{n+1} \mathbb{1}_{\mathsf{A}_n} \right\} ,$$

$$T_{n+1} = \bar{H}_{n+1} \wedge T^1_{n+1} \wedge \tilde{T}^1_{n+1} .$$

Set $S^{\mathrm{a}}_{n+1} = S^{\mathrm{a}}_n + T_{n+1}$, for all $t \in \left[\bar{S}_n, \bar{S}_{n+1}\right)$, $(X_t, Y_t) = \phi_t(X_{\bar{S}_n}, Y_{\bar{S}_n})$, $X_{S^{\mathrm{a}}_{n+1}} = X_{S^{\mathrm{a}}_n} + T_{n+1} Y_{S^{\mathrm{a}}_n}$, $(\tilde{X}_t, \tilde{Y}_t) = \phi_t(\tilde{X}_{S^{\mathrm{a}}_n}, \tilde{Y}_{S^{\mathrm{a}}_n})$, $\tilde{X}_{S^{\mathrm{a}}_{n+1}} = \tilde{X}_{S^{\mathrm{a}}_n} + T_{n+1} \tilde{Y}_{S^{\mathrm{a}}_n}$. If $\tilde{T}_{n+1} = \bar{H}_{n+1}$, consider the two random variables $G^{(1)}, G^{(2)}$ defined by Lemma 11, associated with a Brownian motion independent of $(\bar{E}^1_i, \bar{E}^2_i, \bar{E}^3_i, (\bar{G}_{i,j})_{j \in \mathbb{N}^*})_{i \in \mathbb{N}^*}$, and for $x^{(1)} = X_{S^{\mathrm{a}}_{n+1}}$, $x^{(2)} = \tilde{X}_{S^{\mathrm{a}}_{n+1}}$, Σ the co-variance matrix associated with μ_{v} multiplied by $\bar{E}^3_2 / \bar{\lambda}$, and $M \geqslant 0$.

$$\text{Still if } \tilde{T}_{n+1} = \bar{H}_{n+1}, \quad \text{set} \quad \left\{ \begin{array}{l} Y_{S^{\mathrm{a}}_{n+1}} = \Sigma^{1/2} G^{(1)} \, , \, \tilde{Y}_{S^{\mathrm{a}}_{n+1}} = \Sigma^{1/2} G^{(2)} \\ N_{n+2} = 2 \, , \, H_{n+2} = \bar{E}^3_{\bar{N}_{n+2}} / \bar{\lambda} \, , \end{array} \right.$$

Otherwise set $\bar{N}_{n+2} = \bar{N}_{n+1}$, $\bar{H}_{n+2} = \bar{H}_{n+1} - T_{n+1}$ and

$$\text{if } T_{n+1} = T^1_{n+1} = \tilde{T}^1_{n+1}, \; Y_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(\bar{X}_{\bar{S}_n} + T_{n+1} \bar{Y}_{\bar{S}_n}, \bar{Y}_{\bar{S}_n}) \, , \, Y^R_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(\tilde{X}_{S^{\mathrm{a}}_n} + T_{n+1} \bar{Y}^R_{\bar{S}_n}, \bar{Y}^R_{\bar{S}_n}) \, ,$$

$$\text{if } T_{n+1} = T^1_{n+1} < \tilde{T}^1_{n+1}, \; Y_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(X_{S^{\mathrm{a}}_n} + T_{n+1} Y_{S^{\mathrm{a}}_n}, Y_{S^{\mathrm{a}}_n}) \, , \, \tilde{Y}_{S^{\mathrm{a}}_{n+1}} = \tilde{Y}_{S^{\mathrm{a}}_n} \, ,$$

$$\text{if } T_{n+1} = \tilde{T}^1_{n+1} < T^1_{n+1}, \; \tilde{Y}_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(\tilde{X}_{S^{\mathrm{a}}_n} + T_{n+1} \tilde{Y}_{S^{\mathrm{a}}_n}, \tilde{Y}_{S^{\mathrm{a}}_n}) \, , \, Y_{S^{\mathrm{a}}_{n+1}} = Y_{S^{\mathrm{a}}_n} \, .$$

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

$$T^1_{n+1} = \inf\left\{ t \geqslant 0 \, : \, \int_0^t \left\{ \left\langle Y_{S^{\mathrm{a}}_n}, \nabla U(X_{S^{\mathrm{a}}_n} + sY_{S^{\mathrm{a}}_n}) \right\rangle_+ \right\} \mathrm{d}s \geqslant \bar{E}^1_{n+1} \right\} ,$$

$$\tilde{T}^1_{n+1} = \inf\left\{ t \geqslant 0 \, : \, \int_0^t \left\{ \left\langle \tilde{Y}_{S^{\mathrm{a}}_n}, \nabla U(\tilde{X}_{S^{\mathrm{a}}_n} + s\tilde{Y}_{\bar{S}^{\mathrm{a}}_n}) \right\rangle_+ \right\} \mathrm{d}s \geqslant \bar{E}^2_{n+1} \mathbb{1}_{\mathsf{A}^{\mathrm{c}}_n} + \bar{E}^1_{n+1} \mathbb{1}_{\mathsf{A}_n} \right\} ,$$

$$T_{n+1} = \bar{H}_{n+1} \wedge T^1_{n+1} \wedge \tilde{T}^1_{n+1} .$$

Set $S^{\mathrm{a}}_{n+1} = S^{\mathrm{a}}_n + T_{n+1}$, for all $t \in \left[\bar{S}_n, \bar{S}_{n+1}\right)$, $(X_t, Y_t) = \phi_t(X_{\bar{S}_n}, Y_{\bar{S}_n})$, $X_{S^{\mathrm{a}}_{n+1}} = X_{S^{\mathrm{a}}_n} + T_{n+1} Y_{S^{\mathrm{a}}_n}$, $(\tilde{X}_t, \tilde{Y}_t) = \phi_t(\tilde{X}_{S^{\mathrm{a}}_n}, \tilde{Y}_{S^{\mathrm{a}}_n})$, $\tilde{X}_{S^{\mathrm{a}}_{n+1}} = \tilde{X}_{S^{\mathrm{a}}_n} + T_{n+1} \tilde{Y}_{S^{\mathrm{a}}_n}$ and

$$\text{if } \tilde{T}_{n+1} = \bar{H}_{n+1}, \quad \left\{ \begin{array}{l} Y_{S^{\mathrm{a}}_{n+1}} = \bar{G}_{n+1} \, , \, \tilde{Y}_{S^{\mathrm{a}}_{n+1}} = \bar{G}_{n+1} \\ N_{n+2} = N_{n+1} + 1 \, , \, H_{n+2} = \bar{E}^3_{\bar{N}_{n+2}} / \bar{\lambda} \, , \end{array} \right.$$

Otherwise set $\bar{N}_{n+2} = \bar{N}_{n+1}$, $\bar{H}_{n+2} = \bar{H}_{n+1} - T_{n+1}$ and

$$\text{if } T_{n+1} = T^1_{n+1} = \tilde{T}^1_{n+1}, \; Y_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(\bar{X}_{\bar{S}_n} + T_{n+1} \bar{Y}_{\bar{S}_n}, \bar{Y}_{\bar{S}_n}) \, , \, Y^R_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(\tilde{X}_{S^{\mathrm{a}}_n} + T_{n+1} \bar{Y}^R_{\bar{S}_n}, \bar{Y}^R_{\bar{S}_n}) \, ,$$

$$\text{if } T_{n+1} = T^1_{n+1} < \tilde{T}^1_{n+1}, \; Y_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(X_{S^{\mathrm{a}}_n} + T_{n+1} Y_{S^{\mathrm{a}}_n}, Y_{S^{\mathrm{a}}_n}) \, , \, \tilde{Y}_{S^{\mathrm{a}}_{n+1}} = \tilde{Y}_{S^{\mathrm{a}}_n} \, ,$$

$$\text{if } T_{n+1} = \tilde{T}^1_{n+1} < T^1_{n+1}, \; \tilde{Y}_{S^{\mathrm{a}}_{n+1}} = \mathrm{R}(\tilde{X}_{S^{\mathrm{a}}_n} + T_{n+1} \tilde{Y}_{S^{\mathrm{a}}_n}, \tilde{Y}_{S^{\mathrm{a}}_n}) \, , \, Y_{S^{\mathrm{a}}_{n+1}} = Y_{S^{\mathrm{a}}_n} \, .$$

For $t \geqslant \sup_{n \in \mathbb{N}^*} S^{\mathrm{a}}_n$, set $(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0} = \infty$. Denote by $(\bar{\mathcal{F}}_n)_{n \geqslant 1}$, the filtration associated with $(\bar{E}^1_i, \bar{E}^2_i, \bar{E}^3_i, (\bar{G}_{i,j})_{j \in \mathbb{N}^*})_{i \in \mathbb{N}^*}$. In the construction of the two processes $(X_t, Y_t)_{t \geqslant 0}$, $(\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0}$, note that $(W_t)_{t \geqslant 0}$ is independent of $(\bar{E}^1_i, \bar{E}^2_i, \bar{E}^3_i, (\bar{G}_{i,j})_{j \in \mathbb{N}^*})_{i \in \mathbb{N}^*}$. By Lemma 11, we have that conditionally to $(\bar{E}^1_i, \bar{E}^2_i, \bar{E}^3_i, (\bar{G}_{i,j})_{j \in \mathbb{N}^*})_{i \in \mathbb{N}^*}$, $G^{(2)}$ is a standard d-dimensional Gaussian random variable. Therefore from their definitions and [14, Proposition 5], marginally, $(X_t, Y_t)_{t \geqslant 0}$ and $(\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0}$ are two BPS processes starting from (x, y) and $(\tilde{x}, \tilde{y})$. However,

since the conditional distribution of $(G^{(1)}, G^{(2)})$ given $(\bar{E}_i^1, \bar{E}_i^2, \bar{E}_i^3, (\bar{G}_{i,j})_{j\in\mathbb{N}^*})_{i\in\mathbb{N}^*}$ depends on $\bar{E}_2^3$, $(X_t, Y_t, \tilde{X}_t, \tilde{Y}_t)_{t\geqslant 0}$ is not Markovian.

Furthermore, from the construction of the two processes, for all $n \in \mathbb{N}$ if $(X_{S_n^{\mathrm{a}}}, Y_{S_n^{\mathrm{a}}}) = (\tilde{X}_{S_n^{\mathrm{a}}}, \tilde{Y}_{S_n^{\mathrm{a}}})$, then $(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)$ for all $t > S_n^{\mathrm{a}}$. Besides, consider $\tau = \inf\{n \in \mathbb{N} : N_{n+2} = 2\}$. Then by definition, if $T_{\tau+2} = \bar{H}_{\tau+2}$ and $X_{S_{\tau+1}^{\mathrm{a}}} + \bar{E}_2^3 G^{(1)}/\lambda_{\mathrm{r}} = \tilde{X}_{S_{\tau+1}^{\mathrm{a}}} + \bar{E}_2^3 G^{(2)}/\lambda_{\mathrm{r}}$, then $(X_{S_{\tau+2}^{\mathrm{a}}}, Y_{S_{\tau+2}^{\mathrm{a}}}) = (\tilde{X}_{S_{\tau+2}^{\mathrm{a}}}, \tilde{Y}_{S_{\tau+2}^{\mathrm{a}}})$. Finally, by definition of τ, $T_{\tau+1} = \bar{H}_{\tau+1}$ implies $S_{\tau+1} = \bar{E}_1^3/\lambda_{\mathrm{r}}$ and if $T_{\tau+2} = \bar{H}_{\tau+2}$, $S_{\tau+2} = S$, $S = (\bar{E}_1^3 + \bar{E}_2^3)/\lambda_{\mathrm{r}}$. Based on these three observations, we get for all $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\right) \\
&\geqslant \mathbb{P}\left(t \geqslant S, T_{\tau+2} = \bar{H}_{\tau+2}, X_{S_{\tau+1}^{\mathrm{a}}} + \bar{E}_2^3 \Sigma^{1/2} G^{(1)}/\lambda_{\mathrm{r}} = \tilde{X}_{S_{\tau+1}^{\mathrm{a}}} + \bar{E}_2^3 \Sigma^{1/2} G^{(2)}/\lambda_{\mathrm{r}}\right) \\
&\geqslant \mathbb{P}\left(\mathsf{A} \cap \{t \geqslant S\} \cap \{X_{\bar{E}_1^3/\lambda_{\mathrm{r}}} + \bar{E}_2^3 \Sigma^{1/2} G^{(1)}/\lambda_{\mathrm{r}} = \tilde{X}_{\bar{E}_1^3/\lambda_{\mathrm{r}}} + \bar{E}_2^3 \Sigma^{1/2} G^{(2)}/\lambda_{\mathrm{r}}\}\right) , \qquad (54)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathsf{A}_1 &= \left\{ \int_0^{\bar{E}_2^3/\lambda_{\mathrm{r}}} \left\{ \left\langle Y_{\bar{E}_1^3/\lambda_{\mathrm{r}}}, \nabla U(X_{\bar{E}_1^3/\lambda_{\mathrm{r}}} + s Y_{\bar{E}_1^3/\lambda_{\mathrm{r}}}) \right\rangle_+ \right\} \mathrm{d}s \geqslant \bar{E}_{\tau+2}^1 \right\} , \\
\mathsf{A}_2 &= \left\{ \int_0^{\bar{E}_2^3/\lambda_{\mathrm{r}}} \left\{ \left\langle \tilde{Y}_{\bar{E}_1^3/\lambda_{\mathrm{r}}}, \nabla U(\tilde{X}_{\bar{E}_1^3/\lambda_{\mathrm{r}}} + s \tilde{Y}_{\bar{E}_1^3/\lambda_{\mathrm{r}}}) \right\rangle_+ \right\} \mathrm{d}s \geqslant \bar{E}_{\tau+2}^2 \right\} .
\end{aligned}
$$

Since for all $n \in \{1, \ldots, \tau\}$, $T_{n+1} = T_{n+1}^1 \wedge \tilde{T}_{n+1}^1$, $\left\|Y_{S_n^{\mathrm{a}}}\right\| = \|y\|$, $\|\tilde{Y}_{S_n^{\mathrm{a}}}\| = \|\tilde{y}\|$, so for all $s \in \left[0, \bar{E}_1^3/\lambda_{\mathrm{r}}\right]$,

$$
\|X_s\| \leqslant \|x\| + (\bar{E}_1^3/\lambda_{\mathrm{r}})\,\|y\| \leqslant (1 + \bar{E}_1^3/\lambda_{\mathrm{r}}) R_{\mathsf{K}}, \|\tilde{X}_s\| \leqslant (1 + \bar{E}_1^3/\lambda_{\mathrm{r}}) R_{\mathsf{K}} \qquad (55)
$$

Therefore, for $i = 1, 2$, by the definition (53) of $\tilde{M}$,

$$
\mathsf{B} = \{\|G^{(i)} - (\Sigma^{1/2}/2)(X_{\bar{E}_1^3/\lambda_{\mathrm{r}}} - \tilde{X}_{\bar{E}_1^3/\lambda_{\mathrm{r}}})\| \leqslant M\} \subset \{\|G^{(i)}\| \leqslant \tilde{M}\} ,
$$

and

$$
\begin{aligned}
\mathsf{A}_1 \cap \mathsf{B} \subset \tilde{\mathsf{A}}_1 &= \left\{ (\bar{E}_2^3/\lambda_{\mathrm{r}}) M \sup_{z \in \mathrm{B}(0, (1 + \bar{E}_1^3/\lambda_{\mathrm{r}}) R_{\mathsf{K}} + (\bar{E}_2^3/\lambda_{\mathrm{r}}) M)} \|\nabla U(z)\| \geqslant \bar{E}_{\tau+2}^1 \right\} , \\
\mathsf{A}_2 \cap \mathsf{B} \subset \tilde{\mathsf{A}}_2 &= \left\{ (\bar{E}_2^3/\lambda_{\mathrm{r}}) \tilde{M} \sup_{z \in \mathrm{B}(0, (1 + \bar{E}_1^3/\lambda_{\mathrm{r}}) R_{\mathsf{K}} + (\bar{E}_2^3/\lambda_{\mathrm{r}}) \tilde{M})} \|\nabla U(z)\| \geqslant \bar{E}_{\tau+2}^2 \right\} .
\end{aligned}
$$

Then, we get by (54) setting $\tilde{\mathsf{A}} = \tilde{\mathsf{A}}_1 \cap \tilde{\mathsf{A}}_2$,

$$
\begin{aligned}
&\mathbb{P}\left((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\right) \\
&\geqslant \mathbb{P}\left(\tilde{\mathsf{A}} \cap \mathsf{B} \cap \{t \geqslant S\} \cap \{X_{\bar{E}_1^3/\lambda_{\mathrm{r}}} + \bar{E}_2^3 \Sigma^{1/2} G^{(1)}/\lambda_{\mathrm{r}} = \tilde{X}_{\bar{E}_1^3/\lambda_{\mathrm{r}}} + \bar{E}_2^3 \Sigma^{1/2} G^{(2)}/\lambda_{\mathrm{r}}\}\right) .
\end{aligned}
$$

Conditionning on $\mathcal{F}_{\tau+1}$ and $\bar{E}_2^3$, using that $\bar{E}_{\tau+2}^1, \bar{E}_{\tau+2}^2$ are independent and independent of $G^{(1)}$, $G^{(2)}$ $\bar{E}_2^3$ and $\mathcal{F}_{\tau+1}$, the definition of $G^{(1)}, G^{(2)}$ conditionnaly to $\bar{E}_2^3$ and $\mathcal{F}_{\tau+1}$, and Lemma 11 we have

$$
\begin{aligned}
&\mathbb{P}\left((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\right) \\
&\geqslant \mathbb{E}\left[\mathbb{1}_{[0,t]}(\lambda_{\mathrm{r}}^{-1}\{E_1 + E_2\}) \tilde{\alpha}\left(M, \left\|\Sigma^{1/2}(X_{\bar{E}_1^3/\lambda_{\mathrm{r}}} - \tilde{X}_{\bar{E}_1^3/\lambda_{\mathrm{r}}})\right\| \sqrt{\lambda_{\mathrm{r}}/E_2}\right) g^2(E_2)\right]
\end{aligned}
$$

Combining this result with (55) 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* $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu_{\mathrm{v}}(\mathsf{A}) \geqslant c\nu_{r,\delta}(\mathsf{A}) \ , \tag{56}$$

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}, $t_0 \geqslant 0$, $\varepsilon > 0$ *such that for* $s \geqslant t_0$, *there exists* $M \geqslant 0$ *satisfying for all* $x, \tilde{x} \in \mathsf{K}$,

$$\mathbb{P}\left(x + sG^{(1)} = \tilde{x} + sG^{(2)}, \left\|G^{(1)} - (x-\tilde{x})/2\right\| \leqslant M\right) \geqslant 1 - \varepsilon \ .$$

Proof. Let $x, \tilde{x} \in \mathsf{K} \subset \mathrm{B}(0, R_{\mathsf{K}})$, $R_{\mathsf{K}} \geqslant 0$. If $s > \|x - \tilde{x}\| / (2(r+\delta))$ and $M \geqslant R_{\mathsf{K}} + s(r+\delta)$, then $\mathsf{I}(x, \tilde{x}, s) = \{w \in \mathbb{R}^d,\ \|w\| \leqslant M\} \cap \{w \in \mathbb{R}^d : sr < \|w - x\| < s(r+\delta)\} \cap \{w \in \mathbb{R}^d : sr < \|w - \tilde{x}\| < s(r+\delta)\} \neq \emptyset$. 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 (56), there exists $\tilde{c} > 0$ such that

$$\bar{\nu}_{x,s}(\mathsf{A}) \wedge \bar{\nu}_{\tilde{x},s}(\mathsf{A}) \ \geqslant \tilde{c}\, \mathrm{Leb}\left(\mathsf{A} \cap \mathsf{I}(x, \tilde{x}, s)\right) \ . \tag{57}$$

Besides, (see e.g. [39] or [42]), 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 $\mathbb{P}\left(x + sG = \tilde{x} + sG\right) = \bar{\nu}_{x,s}(\mathsf{A}) \wedge \bar{\nu}_{\tilde{x},s}(\mathsf{A})$. Combining this result with (57), the fact the function in the right hand side of (57) is positive and depends continuously of x and $\tilde{x}$, hence is lower bounded on K, concludes. □

Lemma 14. *Assume* ***A*** *1 and* (56) *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* $t \geqslant t_0$,

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

Proof. The proof is exactly similar to the proof of Lemma 12. Indeed it suffices to consider a coupling of two BPS $(X_t, Y_t)_{t \geqslant 0}$ and $(\tilde{X}_t, \tilde{Y}_t)_{t \geqslant 0}$ 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 precise Lemma 10, in prevision of the low-temperature study of Section 4.2.

Lemma 15. *Assume* ***A*** *1. 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* $t \geqslant t_0$,

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

Proof. In the Gaussian case, 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 [35, Theorem 5.1]. First, in the spirit of the proof of Lemma 12 or of [35, Lemma 5.2], we study a BPS with no potential, i.e. 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 [32, Theorem 6.1]. However, [32, Theorem 6.1] is non quantitative and for the proofs of Section 4.2 need explicit bounds for the convergence of $(P_t)_{t \geqslant 0}$ to π. To this end, we give a quantitative version of Theorem 5 in Appendix B based on [22, 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 (12) 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, (12) holds by Remark 6 since under **A**4

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

Under **A**5, set $\bar{U}(x) = U^{\varsigma}(x)$ and $\ell(x) = 1/(1+\left\|\nabla \bar{U}(x)\right\|)$ for all $x \in \mathbb{R}^d$. All the conditions of **A**8 are satisfied and (12) 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, (12) 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

$$\begin{aligned} \{\sup_{y \in \mathsf{A}_x} \|y\|^2\} \left\|\nabla^2 \bar{U}(x)\right\| &\leqslant 3\eta^{-1} \bar{U}(x) \left\|\nabla^2 \bar{U}(x)\right\| \\ &\leqslant C U^{\varsigma}(x) \left(\left\|\nabla^2 U(x)\right\| U^{\varsigma-1}(x) + \|\nabla U(x)\|^2 U^{\varsigma-2}(x) \right) \end{aligned}$$

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 precise 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 exponential in the dimension d. 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 $d \geqslant 2$ and $\eta > 0$, and let $\mathrm{proj}^{\mathsf{D}} : \mathbb{R}^d \to \mathsf{D}$ be the corresponding projection. 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 $\sigma^2 \, \mathrm{Id}$, $\sigma^2 > 0$. In this setting, U is simply the function which is identically equal to 0 on D. A BPS sampler $(X_t, Y_t)_{t \geqslant 0}$ is defined as in Section 2.1 to target $\pi \otimes \mu_{\mathrm{v}}$. The construction is in all the respect the same, just by replacing the state space $\mathbb{R}^d \times \mathsf{Y}$ by $\mathsf{D} \times \mathsf{Y}$. To show the convergence of the corresponding semi-group $(P_t^{\mathsf{D}})_{t \geqslant 0}$, we show a uniform Doeblin condition [31, Chapter 16] holds using a direct coupling argument.

Note 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]. This 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 orde 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 $t > 0$,*

$$\|\delta_{(x,y)} P_t^{\mathsf{D}} - \delta_{(\tilde{x},\tilde{y})} P_t^{\mathsf{D}}\|_{\mathrm{TV}} \leqslant 2 \left[\mathbb{P}\left(N_t \leqslant 1\right) + \mathbb{E}\left[\mathbb{1}_{[2,+\infty]}(N_t) \left\{ 1 - 2\mathbf{\Phi}\left(\frac{(1+\eta^2(d-1))^{1/2}}{2(S_{N_t} - S_1)} \right) \right\} \right] \right] .$$

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

Proof. Let $(N_t)_{t\geqslant 0}$ 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 [29, Section 2], given $(S_i)_{i\in\mathbb{N}}$, there exist two Brownian motions $(W_t)_{t\geqslant 0}$ and $(\tilde{W}_t)_{t\geqslant 0}$ on D such that for any $t > 0$,

$$\mathbb{P}\left(X_{S_1} + W_t = \tilde{X}_{S_1} + \tilde{W}_t \middle| (S_k)_{k\geqslant 0} \right) = \mathbb{P}\left(T_{\mathrm{c}} \leqslant t | (S_k)_{k\geqslant 0}\right) = 1 - 2\mathbf{\Phi}\left(- \left\| X_{S_1} - \tilde{X}_{S_1} \right\| \Big/ (2t^{1/2}) \right) , \tag{58}$$

and

$$T_{\mathrm{c}} = \inf\{ s \geqslant 0 \,:\, X_{S_1} + W_s = \tilde{X}_{S_1} + \tilde{W}_s \} . \tag{59}$$

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

$$\begin{aligned} G_i &= (W_{(S_{i+1}-S_1)^2} - W_{(S_i - S_1)^2}) / (S_{i+1} - S_i) , \\ \tilde{G}_i &= (\tilde{W}_{(S_{i+1}-S_1)^2} - \tilde{W}_{(S_i - S_1)^2}) / (S_{i+1} - S_i) . \end{aligned} \tag{60}$$

Note that by the Markov property of $(W_t)_{t\geqslant 0}$ and $(\tilde{W}_t)_{t\geqslant 0}$, $(G_i)_{i\in\mathbb{N}^*}$ 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 $t \in [0, S_k]$, $k \geqslant 1$. 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 $t \geqslant 0$, $(X_t, Y_t)_{t\geqslant 0}$ is distributed according to $P_t^{\mathsf{D}}((x,y),\cdot)$ and $(\tilde{X}_t, \tilde{Y}_t)_{t\geqslant 0}$ is distributed according to $P_t^{\mathsf{D}}((\tilde{x},\tilde{y}),\cdot)$. Then it remains to bound $\mathbb{P}\left((X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\right)$ by definition of the total variation norm.

Note that if $(S_{i+1} - S_1)^2 \geqslant (t - S_1)^2 \geqslant (S_i - S_1)^2 \geqslant T_{\mathrm{c}} \geqslant (S_{i-1} - S_1)^2$, $i \geqslant 2$, we have by (59)-(60) and construction $(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)$. Therefore, we get $\{(S_{N_t} - S_1)^2 \geqslant T_{\mathrm{c}}\} \cap \{N_t > 1\} \subset \{(X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\}$ and we obtain

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

The proof is then concluded by conditioning with respect to $(S_k)_{k\in\mathbb{N}}$ using (58) and for any $x \in \mathsf{D}$, $\|x\| \leqslant (1 + \eta^2(d-1))^{1/2}$.

□

Corollary 16. *There exist $C \geqslant 0$ 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 $y, \tilde{y} \in \mathbb{R}^d$,*

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

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

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

Since $\{S_{N_t} \leqslant 3t/4\} \subset \{N_t - N_{3t/4} = 0\}$, 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 $t > 0$

$$2^{-1}\|\delta_{(x,y)} P^{\mathsf{D}}_t - \delta_{(\tilde{x},\tilde{y})} P^{\mathsf{D}}_t\|_{\mathrm{TV}} \leqslant \mathbb{P}\left(S_2 \geqslant t/4\right) + \mathrm{e}^{-\lambda t/4} + \frac{\{2(1+\eta^2(d-1))\}^{1/2}}{t\pi^{1/2}} \, .$$

The proof then follows from a straightforward computation. □

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

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

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 algorithm is a variation of the MCMC algorithm which, rather than computing expectation with respect to the distribution $\pi = \exp(-U)$, aims to find a global minimum of U. We will study in this section a simulated annealing algorithm based on the BPS, extending the results of [35, Theorem 1.5]. For the sake of simplicity, the study is restricted to the following case:

A 9. *(i) The potential $U \in \mathrm{C}^2(\mathbb{R}^d)$ satisfies*

$$\int_{\mathbb{R}^d} \exp(-U(x)/2)\mathrm{d}x < \infty \, , \qquad \lim_{\|x\| \to +\infty} U(x) = +\infty,$$

$$\liminf_{\|x\| \to \infty} \|\nabla U(x)\| > 0 \, , \qquad \sup_{x \in \mathbb{R}^d} \left\|\nabla^2 U(x)\right\| < \infty \, .$$

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, **A** 9 is enforced . However, note the arguments also work under **A** 8 (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 **A** 9.

For a measurable function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$, referred to in the following as cooling schedule, we consider in this section the simulated annealing BPS process (X^β_t, Y^β_t) 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 $(E^1_i, E^2_i, G_i)_{i \in \mathbb{N}^*}$ introduced in Section 2.1. Let $\lambda_{\mathrm{r}} > 0$, $(X^\beta_0, Y^\beta_0) = (x, y)$ and $S^\beta_0 = 0$. We

define by recursion the jump times of the process and the process itself. For all $n \geqslant 0$, consider

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

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

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

where R is defined by (4). Note that under **A**9, Y is bounded and therefore by [14, Proposition 10], $\sup_{n\in\mathbb{N}} S_n^\beta = +\infty$.

Therefore almost surely $(X_t^\beta, Y_t^\beta)_{t\geqslant 0}$ is a $(\mathbb{R}^d \times \mathsf{Y})$-valued càdlàg process. By [9, Theorem 25.5], the BPS process $(X_t^\beta, Y_t^\beta)_{t\geqslant 0}$ defines a non-homogeneous strong Markov semi-group $(P_t)_{t\geqslant 0}$ 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

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

where $(X_u^\beta, Y_u^\beta)_{u\in\mathbb{R}_+}$ is the annealed BPS process started from (x, y) and cooling schedule $s \mapsto \beta(t + s)$. $(P_{s,t})_{t\geqslant s\geqslant 0}$ is associated with the family of generator $(\mathcal{A}_{\beta(t)})_{t\geqslant 0}$ where for any $\beta > 0$, $\mathcal{A}_\beta$ is defined for any $f \in \mathrm{C}^1(\mathbb{R}^d \times \mathsf{Y})$ by

$$\begin{aligned} \mathcal{A}_\beta f(x,y) &= \langle y, \nabla f(x,y)\rangle + \beta(\langle y, \nabla U(x)\rangle)_+ \{f(x, \mathrm{R}(x,y)) - f(x,y)\} \\ &\quad + \lambda_{\mathrm{r}}\left\{\int_{\mathsf{Y}} f(x,w)\mathrm{d}\mu_{\mathrm{v}}(w) - f(x,y)\right\}\,. \end{aligned} \tag{61}$$

As it is usual in simulated annealing if $t \mapsto \beta(t)$ goes to infinity sufficiently slowly for the process (X_t^β, Y_t^β) to approach its instantaneous equilibrium $\exp(-\beta(t)U) \otimes \mu_{\mathrm{v}}$, then X_t^β should be close to a global minimum of U with high probability.

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

Theorem 17. *Assume **A**9. There exists $\theta > 0$ such that if **A**10 holds with $D_1 \leqslant \theta^{-1}$, 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 $t > 0$,*

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

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

First, we establish a Foster-Lyapunov drift condition for $\mathcal{A}_\beta$ uniformly on $\beta \geqslant 1$.

Lemma 18. *Assume **A**9. There exist $A_1, A_2, A_3 > 0$, $\beta_* \geqslant 1$ and $V_1, V_2 \in \mathrm{C}^1(\mathbb{R}^d \times \mathsf{Y})$, with $V_i \exp(-U/2)$ bounded above and below by positive constants for $i = 1, 2$, such that for all $\beta \geqslant \beta_*$,*

$$\mathcal{A}_\beta V_1 \leqslant A_1(A_2 - V_1)\,,$$

and for all $\beta \geqslant 1$,

$$\mathcal{A}_\beta V_2 \leqslant A_3 V_2 \ .$$

Proof. We check that **A** 8 holds for β large enough, with $\bar{U} = U/2$ and the potential $x \mapsto U_\beta(x)$. Indeed, set $\ell(x) = 1$ for all $x \in \mathbb{R}^d$ and $H(t) = t^2$ for $t \in \mathbb{R}$. Then all the conditions of **A** 8 are clearly satisfied, with c_1, c_2 and c_4 which does not depend on β, and $c_3 = \beta$. Let β_* be large enough so that (12) holds for $\beta \geqslant \beta_*$ and $\kappa = 1$ defined in (32).

Let V_1 be the function defined by (21). According to Lemma 7, there exist $A_1, A_2 > 0$ such that

$$\mathcal{A}_{\beta_*} V_1 \leqslant A_1 (A_2 - V_1).$$

Now, for $\beta \geqslant \beta_*$, keeping the notations of Section 3.2,

$$(\mathcal{A}_\beta - \mathcal{A}_{\beta_*}) V_1(x,y) \ = \ (\beta - \beta_*) e^{U(x)/2} \left\langle y, \nabla U(x) \right\rangle_+ (\varphi(-\theta) - \varphi(\theta)) \ \leqslant \ 0 \ .$$

Second, set for any $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$, $V_2(x,y) = \exp(U(x)/2)\varphi_2(\langle y, \nabla U(x) \rangle)$, where $\varphi_2 \in \mathrm{C}^1(\mathbb{R})$ is an increasing function such that $\varphi(s) = 1$ for $s \leqslant -1$ and $\varphi(s) = 3$ for $s \geqslant 1$. Then, for all $\beta \geqslant 1$,

$$\begin{aligned} e^{-U(x)/2} \mathcal{A}_\beta V_2(x,y) &\leqslant \langle y, \nabla U(x) \rangle \, \varphi_2(\langle y, \nabla U(x) \rangle) + M^2 \left\| \nabla^2 U \right\|_\infty \left\| \varphi_2' \right\|_\infty + 2\lambda_{\mathrm{r}} \\ &\quad + \beta \left\langle y, \nabla U(x) \right\rangle_+ \{ \varphi(- \langle y, \nabla U(x) \rangle) - \varphi(\langle y, \nabla U(x) \rangle) \} \\ &\leqslant 3 + M^2 \left\| \nabla^2 U \right\|_\infty \left\| \varphi_2' \right\|_\infty + 2\lambda_{\mathrm{r}} \ , \end{aligned}$$

and we conclude by noting that $\exp(U(x)/2) \leqslant V_2(x,y)$ for any $(x,y) \in \mathbb{R}^d \times \mathsf{Y}$. □

Corollary 19. *Assume that the assumptions of Theorem 17 hold. Then there exists* $A_4 > 0$ *such that for all* $t, s \geqslant 0$ *and* $(x,y) \in \mathbb{R} \times \mathsf{Y}$,

$$P_{t,t+s} V_1(x,y) \leqslant A_4 e^{A_3 s} V_1(x,y) \ ,$$

and for all $t \geqslant 0$ *such that* $\beta(t) \geqslant \beta_*$,

$$P_{t,t+s} V_1(x,y) \leqslant e^{-A_1 s} V_1(x,y) + (1 - e^{-A_1 s}) A_2 \ ,$$

where $V_1 \in \mathrm{C}^1(\mathbb{R}^d \times \mathsf{Y})$, A_1, A_2, A_3 *are given by Lemma 18.*

Proof. The proof follows the same line as the proof of Corollary 9, using Lemma 18 and V_1/V_2 is bounded above and below by positive constants. □

Lemma 20. *Assume that the assumptions of Theorem 17 hold. Then, for all compact set* K *of* $\mathbb{R}^d \times \mathsf{Y}$, *there exist* $s_1, \chi, A_5 > 0$ *which depend on* K, μ_{v}, λ_{r} *and* U *but not on* $t \mapsto \beta(t)$, *such that for all* $(x,y), (\tilde{x}, \tilde{y}) \in \mathsf{K}$, *all* $t \geqslant 0$ *and all* $s \geqslant s_1$,

$$\| P_{t,t+s}((x,y), \cdot) - P_{t,t+s}((\tilde{x}, \tilde{y}), \cdot) \|_{\mathrm{TV}} \leqslant 2 \left[1 - \chi \exp\left(-A_5 \int_t^{t+s} \beta(u) \mathrm{d}u \right) \right] \ .$$

Proof. The arguments are exactly those of the proof of Lemma 15, hence of [35, Theorem 5.1], so that we only give a sketch of proof. First, considering the case $\beta = 0$, we have already shown in Section 3.3 that, starting from two different points in a given compact K, it is possible to merge two processes in a time $s_1 > 0$ while staying in a compact K', with some probability $\chi > 0$. Call E this event. Then, considering the case $\beta > 0$, we follow the same coupling up to the first bounce time. The processes have merged if this first bounce happens after time s_1, which occurs with probability

$$\mathbb{P}\left(\int_t^{t+s_1} \beta(u) \left\langle Y_u^\beta, \nabla U(X_u^\beta) \right\rangle_+ \mathrm{d}u \geqslant E_1^2 \middle| \mathsf{E} \right) \ \geqslant \ \exp\left(-M \left\| \nabla U \right\|_{\infty, K'} \int_t^{t+s_1} \beta(u) \mathrm{d}u \right) \ ,$$

where $M = \sup_{(w,z) \in \mathsf{K}'} \|z\|$. □

Let V_1 and A_i, $i \in \{1,\dots,4\}$, be given by Corollary 19. Then, let $s_1, \chi, A_5 > 0$ be given by Lemma 20, with $\mathsf{K} = \{(x,y) \in \mathbb{R}^d \times \mathsf{Y},\ V_1(x) \leqslant 2A_2\}$. Let

$$t_0 = \inf\{t \geqslant 0,\ \beta(t) \geqslant \beta_*\} , \tag{62}$$

and for $t \geqslant t_0$, define

$$\mathbf{n}(t) = \lfloor (t - t_0)/s_1 \rfloor . \tag{63}$$

Consider the following decomposition,

$$P_{0,t} = P_{0,t-\mathbf{n}(t)s_1} Q_0 Q_1 \cdots Q_{\mathbf{n}(t)-1} Q_{\mathbf{n}(t)} ,$$

where Q_0 is the identity kernel and for $k \in \{1,\dots,n(t)\}$, we set

$$Q_k = P_{t-(\mathbf{n}(t)-k+1)s_1, t-(\mathbf{n}(t)-k)s_1} . \tag{64}$$

For any measurable function $\varphi : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}$ and $\zeta \geqslant 0$, we set

$$\|\varphi\|_{\zeta,V_1} = \sup_{(x,y)\in\mathbb{R}^d\times\mathsf{Y}} \left\{ \frac{|\varphi(x,y)|}{1+\zeta V_1(x)} \right\} ,$$

and consider the weighted V_1-norm on $\mathcal{P}_{V_1}(\mathbb{R}^d \times \mathsf{Y}) = \{\mu \in \mathcal{P}(\mathbb{R}^d \times \mathsf{Y}) : \mu(V_1) < \infty\}$, defined for $\mu_1, \mu_2 \in \mathcal{P}_{V_1}(\mathbb{R}^d \times \mathsf{Y})$ by

$$\rho_\zeta(\mu_1,\mu_2) = \sup\{\mu_1(\varphi) - \mu_2(\varphi) : \|\varphi\|_{\zeta,V_1} \leqslant 1\} . \tag{65}$$

Note that $\rho_\zeta(\mu_1,\mu_2)$ increases with ζ and that $\rho_0 = \|\cdot\|_{\mathrm{TV}}$. In addition, for any $\mu_1, \mu_2 \in \mathcal{P}_{V_1}(\mathbb{R}^d \times \mathsf{Y})$,

$$\rho_\zeta(\mu_1,\mu_2) \leqslant \|\mu_1 - \mu_2\|_{V_1} \leqslant (1+\zeta)^{-1} \rho_\zeta(\mu_1,\mu_2) .$$

Lemma 21. *Assume that the conditions of Theorem 17 hold. Then for all $\nu_1, \nu_2 \in \mathcal{P}_{V_1}(\mathbb{R}^d \times \mathsf{Y})$, $t \geqslant t_0$ and all $k \in \{1,\dots,\mathbf{n}(t)\}$,*

$$\rho_{\epsilon_k}(\nu_1 Q_k, \nu_2 Q_k) \leqslant \kappa_k \rho_{\epsilon_k}(\nu_1,\nu_2) , \tag{66}$$

where

$$\epsilon_k = \frac{\chi}{(1-\gamma)A_2} \exp\left(-A_5 \int_{t-(n-k+1)s_1}^{t-(n-k)s_1} \beta_u \mathrm{d}u \right) ,$$

$$\kappa_k = 1 - \left(\frac{\chi}{2} \wedge \frac{1-\gamma}{4} \right) \exp\left(-A_5 \int_{t-(n-k+1)s_1}^{t-(n-k)s_1} \beta_u \mathrm{d}u \right) , \gamma = \exp(-s_1 A_1) .$$

Proof. It is a direct application to Q_k for all k of Theorem 24 based on Lemma 20 and Corollary 19. □

For a fixed $\beta \geqslant 0$, let $(P_t^{(\beta)})_{t\geqslant 0}$ be the semi-group of the BPS sampler associated with the potential $x \mapsto \beta U(x)$ and, for $t \geqslant t_0$ and $k \in \{0,\dots,\mathbf{n}(t)\}$, let

$$Q'_k = P_{s_1}^{(\beta_k)} , \tag{67}$$

where for ease of notation simplicity we denote

$$\beta_k = \beta_{t-(\mathbf{n}(t)-k)s_1} . \tag{68}$$

In other words, Q'_k is similar to Q_k except that the inverse temperature is frozen. Let $\tilde{\pi}_k$ be the invariant measure of Q'_k, namely

$$\tilde{\pi}_k = \pi_k \otimes \mu_{\mathrm{v}} ,$$

where π_k admits a density with respect to the Lebesgue measure given for any $x \in \mathbb{R}^d$ by

$$\pi_k(x) = \mathrm{Z}_k^{-1} \exp(-\beta_k U(x)) \mathrm{d}x , \qquad \mathrm{Z}_k = \int_{\mathbb{R}^d} \exp(-\beta_k U(\tilde{x})) \mathrm{d}\tilde{x} .$$

We know that the mass of π_k concentrates, as $k \to \infty$, around the vicinity of the global minima of U. To get the same with $P_{0,t}((x,y),\cdot)$, we need to show that $\|\tilde{\pi}_{\mathbf{n}(t)} - P_{0,t}((x,y),\cdot)\|_{\mathrm{TV}}$ vanishes as $t \to \infty$. Denoting, for $t \geqslant t_0$ and $k \in \{0,\dots,\mathbf{n}(t)\}$, $\nu_k = \delta_{(x,y)} P_{0,t-\mathbf{n}(t)s_1} Q_0 Q_1 \cdots Q_{k-1} Q_k$, where Q_k is defined in (64), it is then natural to study

$$u_k = \rho_{\epsilon_k}(\nu_k, \tilde{\pi}_k) \; . \tag{69}$$

From (66), for any $t \geqslant t_0$, $k \in \{1,\dots,\mathbf{n}(t)\}$

$$u_k \leqslant \kappa_k \rho_{\epsilon_k}(\nu_{k-1}, \tilde{\pi}_{k-1}) + \rho_{\epsilon_k}(\tilde{\pi}_{k-1}Q_k, \tilde{\pi}_k) \;\leqslant\; \kappa_k u_{k-1} + e_k \tag{70}$$

where we defined

$$e_k = \rho_{\epsilon_k}(\tilde{\pi}_{k-1}Q_k, \tilde{\pi}_k)$$

and used that $\rho_{\epsilon_k}(\nu_{k-1}, \tilde{\pi}_{k-1}) \leqslant \rho_{\epsilon_{k-1}}(\nu_{k-1}, \tilde{\pi}_{k-1})$ since $(\epsilon_k)_{k\geqslant 0}$ is non-increasing.

Lemma 22. *Assume that the conditions of Theorem 17 hold. Then, there exists $A_6 > 0$ such that for all $t \geqslant t_0$, all $k \in \{1,\dots,\mathbf{n}(t)\}$ and $l \geqslant 1$, there exists $A_l > 0$ such that*

$$e_k \leqslant A_l(\sqrt{\beta_k - \beta_{k-1}} + \beta_k - \beta_{k-1}) + A_6 \mathrm{e}^{-\frac{1}{2}(\beta_{k-1}-1)l} \; ,$$

where β_k, $\mathbf{n}$ and t_0 are defined by (68), (63) and (62) respectively.

Proof. Let $t \geqslant t_0$, $k \in \{1,\dots,\mathbf{n}(t)\}$ and $l \geqslant 1$. In the proof, C stands for a constant which may change from line to line but does not depend on k, l and β. We bound

$$e_k \leqslant \rho_{\epsilon_k}(\tilde{\pi}_{k-1}, \tilde{\pi}_k) + \rho_{\epsilon_k}(\tilde{\pi}_{k-1}Q_k, \tilde{\pi}_{k-1}) \tag{71}$$

and deal with each terms of the right hand side apart. Indeed, for the first one, the first marginal of $\tilde{\pi}_{k-1}$ and $\tilde{\pi}_k$ having an explicit density, and their second marginal being equal, we bound

$$\begin{aligned}
\rho_{\epsilon_k}(\tilde{\pi}_{k-1}, \tilde{\pi}_k) &= \int_{\mathbb{R}^d \times \mathsf{Y}} (1 + \epsilon_k V_1(x,y)) |\pi_k(x) - \pi_{k-1}(x)| \mathrm{d}x \mu_{\mathrm{v}}(\mathrm{d}y) \\
&\leqslant C \int_{\mathbb{R}^d} \mathrm{e}^{U(x)/2} \left| \frac{\mathrm{e}^{-\beta_k U(x)}}{\mathrm{Z}_k} - \frac{\mathrm{e}^{-\beta_{k-1} U(x)}}{\mathrm{Z}_{k-1}} \right| \mathrm{d}x \\
&\leqslant C \mathrm{e}^{l/2} \int_{\mathbb{R}^d} \left| \frac{\mathrm{e}^{-\beta_k U(x)}}{\mathrm{Z}_k} - \frac{\mathrm{e}^{-\beta_{k-1} U(x)}}{\mathrm{Z}_{k-1}} \right| \mathrm{d}x \\
&\quad + C \int_{\{U>l\}} \frac{\mathrm{e}^{-(\beta_k - \frac{1}{2})U(x)}}{\mathrm{Z}_k} + \frac{\mathrm{e}^{-(\beta_{k-1} - \frac{1}{2})U(x)}}{\mathrm{Z}_{k-1}} \mathrm{d}x \; .
\end{aligned} \tag{72}$$

We treat the two terms in the right-hand-side apart. The first term is the total variation distance between π_k and π_{k-1}. Since $\beta_{k-1} \leqslant \beta_k$ since β is non-decreasing, $\mathrm{Z}_{k-1} \geqslant \mathrm{Z}_k$. Using Pinsker's inequality and this result, we get

$$\begin{aligned}
\left(\int_{\mathbb{R}^d} \left| \frac{\mathrm{e}^{-\beta_k U(x)}}{\mathrm{Z}_k} - \frac{\mathrm{e}^{-\beta_{k-1} U(x)}}{\mathrm{Z}_{k-1}} \right| \mathrm{d}x \right)^2 &\leqslant 2 \int_{\mathbb{R}^d} \ln \left(\frac{\mathrm{e}^{-\beta_{k-1} U(x)} \mathrm{Z}_k}{\mathrm{e}^{-\beta_k U(x)} \mathrm{Z}_{k-1}} \right) \pi_{k-1}(x) \mathrm{d}x \\
&\leqslant 2(\beta_k - \beta_{k-1}) \int_{\mathbb{R}^d} U(x) \frac{\mathrm{e}^{-\beta_{k-1} U(x)}}{\mathrm{Z}_{k-1}} \mathrm{d}x \\
&\leqslant 2(\beta_k - \beta_{k-1})(1 + C\sqrt{\beta_{k-1}} \mathrm{e}^{-\beta_{k-1}+1}) \\
&\leqslant C(\beta_k - \beta_{k-1}) \; ,
\end{aligned} \tag{73}$$

where we used for the two last inequalities that

$$\int_{\{U>1\}} U(x) \mathrm{e}^{-U(x)} \mathrm{d}x \;\leqslant\; 2 \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)/2} \mathrm{d}x \;<\; \infty$$

and since $U(0)=0$, $U(x) \leqslant \left\|\nabla^2 U\right\|_\infty \|x\|^2$, for any $x \in \mathbb{R}^d$ by **A**9,

$$\mathrm{Z}_{k-1} \geqslant \int_{\mathbb{R}^d} \mathrm{e}^{-\beta_{k-1}\|\nabla^2 U\|_\infty \|x\|^2} \mathrm{d}x \geqslant C\beta_{k-1}^{-d/2} > 0 \,.$$

Similarly, for the second term of (72) we obtain

$$\int_{\{U>l\}} \frac{\mathrm{e}^{-(\beta_k-\frac{1}{2})U(x)}}{\mathrm{Z}_k} \mathrm{d}x \leqslant C\beta_k^{d/2} \mathrm{e}^{-(\beta_k-1)l} \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)/2} \mathrm{d}x \,.$$

Using that for any $t \geqslant 1$, $t^{d/2}\exp(-l(t-1)/2) \leqslant (d/l)^{d/2}\exp(-(d-l)/2)$ if $d \geqslant \ell$ and $t^{d/2}\exp(-l(t-1)/2) \leqslant 1$ otherwise, there exists $A_{6,1}$ which does not depend on l such that

$$\int_{\{U>l\}} \frac{\mathrm{e}^{-(\beta_k-\frac{1}{2})U(x)}}{\mathrm{Z}_k} \mathrm{d}x \leqslant A_{6,1}\mathrm{e}^{-(\beta_k-1)l/2} \,. \tag{74}$$

Combining this bound and (73) in (72), we get that there exists $A_{l,1} \geqslant 0$ such that

$$\rho_{\epsilon_k}(\tilde{\pi}_{k-1}, \tilde{\pi}_k) \leqslant A_{l,1}\sqrt{\beta_k - \beta_{k-1}} + A_{6,1}\mathrm{e}^{-(\beta_k-1)l/2} \,. \tag{75}$$

The second term of (71) is treated through a synchronous coupling similar to [14, Section 6]. Indeed, $\tilde{\pi}_{k-1}$ being invariant for Q'_{k-1} defined in (67) and by (65),

$$\begin{aligned} \rho_{\epsilon_k}(\tilde{\pi}_{k-1}Q_k, \tilde{\pi}_{k-1}) &= \rho_{\epsilon_k}(\tilde{\pi}_{k-1}Q_k, \tilde{\pi}_{k-1}Q'_{k-1}) \\ &= \sup_{\|\varphi\|_{\epsilon_k,V_1}\leqslant 1} \left\{\mathbb{E}[\varphi(X_{s_1}, Y_{s_1}) - \varphi(X'_{s_1}, Y'_{s_1})]\right\} \,, \end{aligned} \tag{76}$$

where $(X_t, Y_t)_{t\geqslant 0}$ (resp. $(X'_t, Y'_t)_{t\geqslant 0}$) is a BPS process with a fixed temperature β_{k-1} (resp. a annealed BPS process with cooling schedule $s \mapsto \beta(t-(\mathbf{n}(t)-k+1)s_1+s)$) and $(X_0, Y_0) = (X'_0, Y'_0)$ is distributed according to $\tilde{\pi}_{k-1}$. Following [14, Section 6], we construct such processes in such a way $(X_t, Y_t) = (X'_t, Y'_t)$ up to time T'_{b}, where T'_{b} is the first time $(X'_t, Y'_t)_{t\geqslant 0}$ bounces while $(X_t, Y_t)_{t\geqslant 0}$ does not, defined by

$$T'_{\mathrm{b}} = \inf\left\{\tau \geqslant 0,\ E < \int_0^\tau (\beta_k - \beta_{t-(\mathbf{n}(t)-n)s_1+s}) \left\langle Y^\beta_s, \nabla U(X^\beta_s)\right\rangle_+ \mathrm{d}s\right\} \,,$$

where E is a standard exponential random variable independent of Z.

Consider the compact sets $\mathsf{K} = \{(x,y) \in \mathbb{R}^d \times \mathsf{Y} : U(x) < l\}$ and $\tilde{\mathsf{K}} = \{(x,y) \in \mathbb{R}^d \times \mathsf{Y} : \mathrm{dist}((x,y), \mathsf{K}) \leqslant Ms_1\}$, where $\mathrm{dist}(\cdot, \mathsf{K})$ is the distance from K and $M = \sup_{z\in\mathsf{Y}} \|z\|$. That way, if a BPS with refreshment law μ_{y} over Y have an initial condition in K, then on the time interval $[0, s_1]$ it necessarily stays in $\tilde{\mathsf{K}}$.

Consider $\varphi : \mathbb{R}^d \times \mathsf{Y} :\to \mathbb{R}$ with $\|\varphi\|_{\epsilon_{n+1},V_1} \leqslant 1$ and the following decomposition

$$\begin{aligned} \mathbb{E}[\varphi(X_{s_1}, Y_{s_1}) - \varphi(X'_{s_1}, Y'_{s_1})] &= \mathbb{E}[\mathbb{1}_{\mathsf{K}}(X_0, Y_0)\{\varphi(X_{s_1}, Y_{s_1}) - \varphi(X'_{s_1}, Y'_{s_1})\}] \\ &\quad + \mathbb{E}[\mathbb{1}_{\mathbb{R}^d\times\mathsf{Y}\setminus\mathsf{K}}(X_0, Y_0)\{\varphi(X_{s_1}, Y_{s_1}) - \varphi(X'_{s_1}, Y'_{s_1})\}] \,. \end{aligned} \tag{77}$$

We bound the two terms separately. First, using that if $(X_0, Y_0) \in \mathsf{K}$, then for any $t \in [0, s_1]$, $(X'_t, Y'_t) \in \tilde{\mathsf{K}}$, we have

$$\begin{aligned} &\mathbb{E}[\mathbb{1}_{\mathsf{K}}(X_0, Y_0)\{\varphi(X_{s_1}, Y_{s_1}) - \varphi(X'_{s_1}, Y'_{s_1})\}] \\ &\quad = 2(1+\epsilon_1)\|V_1\|_{\infty,\tilde{\mathsf{K}}} \mathbb{P}\left((X_0, Y_0) \in \mathsf{K}, (X_{s_1}, Y_{s_1}) \neq (X'_{s_1}, Y'_{s_1})\right) \\ &\quad = 2(1+\epsilon_1)\|V_1\|_{\infty,\tilde{\mathsf{K}}} \mathbb{P}\left((X_0, Y_0) \in \mathsf{K}, T'_{\mathrm{b}} < s_1\right) \\ &\quad \leqslant 2(1+\epsilon_1)\|V_1\|_{\infty,\tilde{\mathsf{K}}} \mathbb{P}\left(E < M\|\nabla U\|_{\infty,\tilde{\mathsf{K}}} \int_0^{s_1} (\beta_k - \beta_{t-(\mathbf{n}(t)-n)s_1+s}) \mathrm{d}s\right) \\ &\quad \leqslant 2(1+\epsilon_1)\|V_1\|_{\infty,\tilde{\mathsf{K}}} M\|\nabla U\|_{\infty,\tilde{\mathsf{K}}}\, s_1(\beta_k - \beta_{k-1}) \,, \end{aligned} \tag{78}$$

where $M = \sup_{z\in\mathsf{Y}} \|z\|$. Note that $\|\nabla U\|_{\infty,\tilde{\mathsf{K}}}$ depends on $\tilde{\mathsf{K}}$, hence on l.

Next using Lemma 18 and the Markov property, we get

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{\mathbb{R}^d\times\mathsf{Y}\setminus\mathsf{K}}(X_0,Y_0)\{\varphi(X_{s_1},Y_{s_1})-\varphi(X'_{s_1},Y'_{s_1})\}] & \\
&\leqslant (1+\epsilon_1)\mathbb{E}[\mathbb{1}_{\mathbb{R}^d\times\mathsf{Y}\setminus\mathsf{K}}(X_0,Y_0)\{V_1(X_{s_1})+V_1(X'_{s_1})\}] \\
&\leqslant 2(1+\epsilon_1)\left(\mathbb{E}\left[\mathbb{1}_{\mathbb{R}^d\times\mathsf{Y}\setminus\mathsf{K}}(X_0,Y_0)V_1(X_0)\right]+A_2\right) \\
&\leqslant 2(1+\epsilon_1)\left(C\int_{U\geqslant l}\mathrm{e}^{U(x)/2}\tilde{\pi}_{k-1}(\mathrm{d}x)+A_2\right) \\
&\leqslant 2(1+\epsilon_1)\left(C\mathrm{e}^{-(\beta_k-1)l/2}+A_2\right) .
\end{aligned}$$

where we used for the penultimate inequality that (X_0,Y_0) is distributed according to $\tilde{\pi}_{k-1}$, Combining this result and (78) in (77) and (76), we get there exist $A_{6,2}\geqslant 0$ independent of l and $A_{l,2}\geqslant 0$ satisfying

$$\rho_{\epsilon_k}(\tilde{\pi}_{k-1}Q_k,\tilde{\pi}_{k-1}) \leqslant A_{l,2}(\beta_k-\beta_{k-1})+A_{6,2}\mathrm{e}^{-(\beta_k-1)l/2} .$$

The proof is concluded combining this result and (75) in (71). □

Lemma 23. *Assume* ***A*** *9. There exists $\theta>0$ such that if* ***A*** *10 holds with $D_1\leqslant\theta^{-1}$, then there exists $A_7>0$ satisfying for all $t\geqslant t_0$, $k\leqslant\mathbf{n}(t)$ and $(x,y)\in\mathbb{R}^d\times\mathsf{Y}$, $u_k\leqslant A_7V_1(x,y)/k^{q_1}$ where u_k is given in (69) and $q_1=(1/2)(1-\theta D_1)$.*

Proof. Let $l\geqslant 1$, $t\geqslant t_0$ and $k\in\{1,\dots\mathbf{n}(t)\}$. In the proof, C stands for a constant which may change from line to line but does not depend on k, l and β. Denoting $d_0=0$ and $d_k=\kappa_k d_{k-1}+e_k$, (70) reads

$$u_k-d_k\leqslant\kappa_k(u_{k-1}-d_{k-1})$$

and yields

$$u_k\leqslant u_0\prod_{j=1}^k\kappa_j+\sum_{i=1}^k\left\{e_i\prod_{j=i}^{k-1}\kappa_j\right\} , \tag{79}$$

with the convention that $\prod_{j=k}^{k-1}\kappa_j=1$. From Lemma 22 applied with $l=1/D_2$, and bounding

$$\beta_k-\beta_{k-1} \leqslant \frac{D_1\lceil s_1/s_0\rceil}{t-(\mathbf{n}(t)-k+1)s_1}$$

for k large enough, we get

$$e_k\leqslant C/\sqrt{k} \tag{80}$$

Let $\theta=2A_5s_1$, so that, by definition of κ_k given in Lemma 21 the condition β, and using $1-s\leqslant\mathrm{e}^{-s}$, we have

$$\kappa_k \leqslant 1-\left(\frac{\chi}{2}\wedge\frac{1-\gamma}{4}\right)\exp(-\theta\beta_k/2) \leqslant \exp(-Cn^{-\theta D_1/2})$$

Hence, for $i\in\{1,\dots,k\}$,

$$\prod_{j=i}^k\kappa_j \leqslant\exp\left(-C\sum_{j=i}^n j^{-\theta D_1/2}\right)\leqslant\exp\left(-(C/q)\{(n+1)^q-i^q\}\right) ,$$

with $q=1-\theta D_1/2\in(1/2,1)$ by assumption. Thus, combining this result and (80) in (79), we get

$$u_k\leqslant u_0\mathrm{e}^{-(C/q)(n^q-1)}+C(k^{-1/2}+I(k))$$

with

$$\begin{aligned} I(k) &= \mathrm{e}^{-(C/q)k^q} \int_1^k \frac{1}{\sqrt{s}} \mathrm{e}^{(C/q)s^q} \mathrm{d}s \\ &= \frac{1}{C}\left(\frac{1}{k^{q-\frac{1}{2}}} - \mathrm{e}^{-(C/q)(k^q-1)}\right) + \frac{\mathrm{e}^{-(C/q)k^q}}{C}(q-1/2)\int_1^k s^{q-\frac{3}{2}}\mathrm{e}^{(C/q)s^q}\mathrm{d}s \\ &\leqslant \frac{1}{Ck^{q-\frac{1}{2}}} + \frac{\mathrm{e}^{-(C/q)k^q}}{C}\int_1^{k_0} s^{q-\frac{3}{2}}\mathrm{e}^{(C/q)s^q}\mathrm{d}s + \frac{I(k)}{Ck_0^{1-q}} \end{aligned}$$

for all $k_0 \geqslant 1$. In particular, for $k_0 \geqslant (2/C)^{1/(1-q)}$, this means $I(k) \leqslant Ck^{1/2-q}$.

Finally, from the first part Corollary 19, $u_0 \leqslant CV_1(x,y)$. □

Proof of Theorem 17. Let $t > t_0$, $n = \mathbf{n}(t)$, $\eta > \eta' > 0$. In the proof, C stands for a constant which may change from line to line but does not depend on n, η, η', t and β. First,

$$\mathbb{P}(U(X_t^\beta) > \eta + \min U) \leqslant \int_{\{U \geqslant \eta\}} \tilde{\pi}_k(\mathrm{d}x, \mathrm{d}y) + (1/2)\|P_{0,t}((x,y),\cdot) - \tilde{\pi}_k\|_{\mathrm{TV}} \ .$$

Similarly to (74),

$$\int_{\{U \geqslant \eta\}} \tilde{\pi}_k(\mathrm{d}x, \mathrm{d}y) \leqslant C\mathrm{e}^{-\beta_k \eta'} \ \leqslant \ Ct^{-D_2\eta'} \ .$$

We conclude, with Lemma 23 and the first part of Corollary 19, by

$$\|\nu P_{0,t} - \tilde{\pi}_k\|_{\mathrm{TV}} \leqslant u_k \ \leqslant CV_1(x,y)/t^q \ .$$

□

Acknowledgements

Alain Durmus acknowledges support from Chaire BayeScale "P. Laffitte". Pierre Monmarché acknowledges support from the French ANR project ANR-12-JS01-0006 - PIECE. Arnaud Guillin and Pierre Monmarché acknowledge support from the French ANR-17-CE40-0030 - EFI - Entropy, flows, inequalities.

References

[1] C. Andrieu, N. De Freitas, A. Doucet, and M. I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[2] J. Bierkens, A. Bouchard-Côté, A. Doucet, A. B. Duncan, P. Fearnhead, G. Roberts, and S. J. Vollmer. Piecewise Deterministic Markov Processes for Scalable Monte Carlo on Restricted Domains. *ArXiv e-prints*, January 2017.

[3] J. Bierkens, P. Fearnhead, and G. Roberts. The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data. *ArXiv e-prints*, July 2016.

[4] J. Bierkens and G. Roberts. A piecewise deterministic scaling limit of lifted Metropolis-Hastings in the Curie-Weiss model. *Ann. Appl. Probab.*, 27(2):846–882, 2017.

[5] J. Bierkens, G. Roberts, and P.-A. Zitt. Ergodicity of the zigzag process. *arXiv preprint arXiv:1712.09875*, 2017.

[6] K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, and S. McKay. Monte carlo simulation in statistical physics. *Computers in Physics*, 7(2):156–157, 1993.

[7] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The Bouncy Particle Sampler: A Non-Reversible Rejection-Free Markov Chain Monte Carlo Method. *ArXiv e-prints*, October 2015.

[8] V. Calvez, G. Raoul, and C. Schmeiser. Confinement by biased velocity jumps: aggregation of *escherichia coli*. *Kinet. Relat. Models*, 8(4):651–666, 2015.

[9] M. Davis. *Markov Models & Optimization*, volume 49. CRC Press, 1993.

[10] G. Deligiannidis, A. Bouchard-Côté, and A. Doucet. Exponential Ergodicity of the Bouncy Particle Sampler. *ArXiv e-prints*, May 2017. Accepted for publication in Annals of Statistics.

[11] P. Diaconis, S. Holmes, and R. M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3):726–752, 2000.
[12] R. Douc, A. Guillin, and E. Moulines. Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):239–257, 2008.
[13] S. Duane, A.D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222, 1987.
[14] A. Durmus, A. Guillin, and P. Monmarché. Piecewise deterministic markov processes and their invariant measures. 2018.
[15] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, pages 1–36, 2015.
[16] R. Erban and H.G. Othmer. From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.*, 65(2):361–391, 2004/05.
[17] J. Fontbona, H. Guérin, and F. Malrieu. Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.*, 126(10):3077–3101, 2016.
[18] P. W. Glynn and S. P. Meyn. A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, 24(2):916–931, 1996.
[19] S. Goldstein. On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.*, 4:129–156, 1951.
[20] P. J. Green, K. Łatuszyński, M. Pereyra, and C. P. Robert. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25(4):835–862, 2015.
[21] A. Guillin. Uniform moderate deviations of functional empirical processes of Markov chains. *Probab. Math. Statist.*, 20(2, Acta Univ. Wratislav. No. 2256):237–260, 2000.
[22] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.
[23] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
[24] S. F. Jarner and E. Hansen. Geometric ergodicity of metropolis algorithms. *Stochastic Processes and their Applications*, 85(2):341 – 361, 2000.
[25] M. Kac. A stochastic model related to the telegrapher's equation. *Rocky Mountain J. Math.*, 4:497–509, 1974. Reprinting of an article published in 1956, Papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972).
[26] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10:no. 3, 61–123, 2005.
[27] V. Lemaire, M. Thieullen, and N. Thomas. Exact simulation of the jump times of a class of Piecewise Deterministic Markov Processes. *ArXiv e-prints*, February 2016.
[28] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logist. Quart.*, 26(3):403–413, 1979.
[29] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.
[30] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
[31] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.
[32] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.
[33] M. Michel, S. C. Kapfer, and W. Krauth. Generalized event-chain monte carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *The Journal of chemical physics*, 140(5):054116, 2014.
[34] L. Miclo and P. Monmarché. étude spectrale minutieuse de processus moins indécis que les autres. In *Séminaire de Probabilités XLV*, volume 2078 of *Lecture Notes in Math.*, pages 459–481. Springer, Cham, 2013.
[35] P. Monmarché. Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.*, 13(1):357–398, 2016.
[36] R. M. Neal. Improving asymptotic variance of MCMC estimators: Non-reversible chains are better. Technical Report Technical Report No. 0406, Dept. of Statistics, University of Toronto, 2004.
[37] R. M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162, 2011.

[38] E. A. J. F. Peters and G. de With. Rejection-free monte carlo sampling for general potentials. *Phys. Rev. E 85, 026703*, 2012.
[39] J. W. Pitman. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(4):315–322, 1976.
[40] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1994.
[41] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
[42] J. S. Rosenthal. Rates of convergence for Gibbs sampling for variance component models. *Ann. Statist.*, 23(3):740–761, 1995.
[43] P. J. Rossky, J. D. Doll, and H. L. Friedman. Brownian dynamics as smart Monte Carlo simulation. *The Journal of Chemical Physics*, 69(10):4628–4633, 1978.
[44] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1728, 12 1994.
[45] K. S. Turitsyn, M. Chertkov, and M. Vucelja. Irreversible Monte Carlo algorithms for efficient sampling. *Physica D Nonlinear Phenomena*, 240:410–414, February 2011.
[46] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal.*, 172(2):301–376, 2000.

Appendix A. Postponed proof

Proof of Proposition 3. Note that since for all $x \in \mathbb{R}^d$, $\|x\| \geqslant 1$,

$$U_1(x) = \|x\|^\alpha U_1(x/\|x\|) \,, \tag{81}$$

that it is sufficient to show that there exists $C_1, C_2 > 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geqslant 1$ such that

$$C_1 \|x\|^{\alpha-1} \leqslant \|\nabla U_1(x)\| \leqslant C_2 \|x\|^{\alpha-1} \tag{82}$$

$$\left\|\nabla^2 U_1(x)\right\| \leqslant C_2 \|x\|^{\alpha-2} \,. \tag{83}$$

(82) is just a consequence of [24, Lemma 4.5]. As for (83), we have by (81) for all $x \in \mathbb{R}^d$, $\|x\| \geqslant 1$,

$$\begin{aligned}
\nabla U_1(x) &= \alpha \|x\|^{\alpha-2} x U_1(x/\|x\|) + \|x\|^\alpha \left\{\mathrm{Id} - xx^{\mathrm{T}}/\|x\|^2\right\} \nabla U_1(x/\|x\|) \\
\nabla U_2(x) &= \alpha \left\{\|x\|^{\alpha-2} + (\alpha-2)\|x\|^{\alpha-4} x^{\mathrm{T}}\right\} U_1(x/\|x\|) \\
&+ \alpha \|x\|^{\alpha-2} \left[\left\{\mathrm{Id} - xx^{\mathrm{T}}/\|x\|^2\right\} \nabla U_1(x/\|x\|) x^{\mathrm{T}} + x \nabla U_1(x/\|x\|)^{\mathrm{T}}(x) \left\{\mathrm{Id} - xx^{\mathrm{T}}/\|x\|^2\right\}\right] \\
&\quad + \|x\|^{\alpha-2} \left\{\nabla U_1(x/\|x\|) x^{\mathrm{T}} + x \nabla U_1(x/\|x\|)^{\mathrm{T}} + 2 \nabla U_1(x/\|x\|)^{\mathrm{T}} x x x^{\mathrm{T}}\right\} \\
&\quad + \|x\|^\alpha \left\{\mathrm{Id} - xx^{\mathrm{T}}/\|x\|^2\right\} \nabla^2 U_1(x/\|x\|) \,.
\end{aligned}$$

Since the U_1 is assumed to be twice continuously differentiable, the proof is finished. □

Appendix B. Quantitative contraction rates for Markov chains

In this section, we give for completeness a quantitative version of [22, Theorem 1.2] which is used in Section 4.2. Let Q be a Markov operator on a smooth finite dimension manifold M (in our applications $Q = P_{t_0}$ for some $t_0 > 0$, with $\mathsf{M} = \mathbb{R}^d \times \mathsf{Y}$) and $V : \mathsf{M} \to [1, +\infty)$ (which can be thought as the one given by (21)).

For any measurable function $\varphi : \mathsf{M} \to \mathbb{R}$ and $\zeta \geqslant 0$, we set

$$\|\varphi\|_{\zeta,V} = \sup_{x \in \mathsf{M}} \left\{\frac{|\varphi(x)|}{1 + \zeta V(x)}\right\} \,,$$

and consider the weighted V-norm on $\mathcal{P}_V(\mathsf{M}) = \{\mu \in \mathcal{P}(\mathsf{M}) : \mu(V) < \infty\}$, defined for $\mu_1, \mu_2 \in \mathcal{P}_V(\mathsf{M})$ by

$$\rho_\zeta(\mu_1, \mu_2) = \sup\{\mu_1(\varphi) - \mu_2(\varphi) : \|\varphi\|_{\zeta,V} \leqslant 1\}. \tag{84}$$

Theorem 24. *Suppose that there exist $\alpha, \gamma \in (0,1)$, $C_1 > 0$ and $C_2 > 2C_1$ such that for all $x, y, z \in \mathsf{M}$, $V(x) + V(y) \leqslant C_2$,*

$$\|Q(x,\cdot) - Q(y,\cdot)\|_{\mathrm{TV}} \leqslant 2(1-\alpha)\,, \qquad QV(z) \leqslant \gamma V(z) + C_1(1-\gamma)\,.$$

Then there exists $\zeta > 0$ and $\kappa \in (0,1)$ such that for all $\mu_1, \mu_2 \in \mathcal{P}_V(\mathsf{M})$,

$$\rho_\zeta(\mu_1 Q, \mu_2 Q) \leqslant \kappa \rho_\zeta(\mu_1, \mu_2)\,,$$

where ρ_ζ is defined by (84). More precisely, if $C_2 = 4C_1$, then this holds with

$$\zeta = \alpha((1-\gamma)C_1)^{-1}\,, \qquad \kappa = (1-\alpha/2) \vee ((3+\gamma)/4)\,.$$

Proof. [22, Lemma 2.1] shows that

$$\rho_\zeta(\mu_1, \mu_2) = \sup\{\mu_1(\varphi) - \mu_2(\varphi) : |\varphi|_\zeta \leqslant 1\}\,,$$

where ρ_ζ is defined by (65) and

$$|\varphi|_\zeta = \sup_{x \neq y} \left\{ \frac{|\varphi(x) - \varphi(y)|}{2 + \zeta V(x) + \zeta V(y)} \right\} = \inf_{c \in \mathbb{R}} \|\varphi + c\|_{\zeta,V}.$$

Let φ be a measurable function such that $|\varphi|_\zeta = \|\varphi\|_{\zeta,V} = 1$. We aim to show that $|Q\varphi|_\zeta \leqslant \kappa$ or, in other words, that

$$|Q\varphi(x) - Q\varphi(y)| \leqslant \kappa(2 + \zeta V(x) + \zeta V(y))$$

for all $x, y \in \mathsf{M}$.

First, consider the case where $V(x) + V(y) \geqslant C_2$. For $\zeta > 0$, set $\kappa_1 = \gamma + (1-\gamma)\frac{1+\zeta C_1}{1+\zeta C_2/2}$. Note that $\gamma < \kappa_1 < 1$, and

$$2(1-\kappa_1) + (\gamma - \kappa_1)\zeta C_2 + 2\zeta C_1(1-\gamma) \leqslant 0\,.$$

Hence,

$$\begin{aligned}
|Q\varphi(x) - Q\varphi(y)| &\leqslant 2 + \zeta QV(x) + \zeta QV(y) \leqslant 2 + \zeta\gamma V(x) + \zeta\gamma V(y) + 2\zeta C_1 \\
&\leqslant \kappa_1(2 + \zeta V(x) + \zeta V(y)) + 2(1-\kappa_1) + (\gamma - \kappa_1)\zeta(V(x) + V(y)) + 2\zeta C_1 \\
&\leqslant \kappa_1(2 + \zeta V(x) + \zeta V(y))\,.
\end{aligned}$$

Second, consider the case where $V(x) + V(y) \leqslant C_2$. Let (Z_x, Z_y) be an optimal coupling of $Q(x,\cdot)$ and $Q(y,\cdot)$. Then, writing $\kappa_2 = (1 - \alpha + \zeta C_1(1-\gamma)/2) \vee \gamma$ (which is smaller than 1 for ζ small enough),

$$\begin{aligned}
|Q\varphi(x) - Q\varphi(y)| &\leqslant \mathbb{P}(Z_x \neq Z_y)\mathbb{E}\left[|\varphi(Z_x) - \varphi(Z_y)|\,|\, Z_x \neq Z_y\right] \\
&\leqslant \mathbb{P}(Z_x \neq Z_y)(2 + \zeta\,\mathbb{E}\left[V(Z_x) + V(Z_y)\right]) \\
&\leqslant 2(1-\alpha) + \zeta\gamma(V(x) + V(y)) + \zeta C_1(1-\gamma) \\
&\leqslant \kappa_2(2 + \zeta V(x) + \zeta V(y))\,,
\end{aligned}$$

which concludes the general proof.

For $C_2 = 4C_1$, we chose $\zeta = \frac{\alpha}{(1-\gamma)C_1}$, so that $\kappa_2 = 1 - \alpha/2$ and

$$\kappa_1 \;=\; \gamma + (1-\gamma)\left(1 - \frac{\alpha}{1-\gamma+2\alpha}\right) \;=\; 1 - \frac{\alpha(1-\gamma)}{1-\gamma+2\alpha}\,.$$

Using that, for $a, b > 0$,

$$\frac{ab}{a+b} \;=\; \frac{a \wedge b}{1 + \frac{a \wedge b}{a \vee b}} \;>\; \frac{a \wedge b}{2}\,,$$

we get

$$\kappa_1 \leqslant 1 - \frac{(2\alpha) \wedge (1-\gamma)}{4} = (1-\alpha/2) \vee ((3+\gamma)/4) \, .$$

□

Remark that, under the same assumptions that Theorem 24 but with $\alpha = 0$, the same proof yields, for all $\zeta > 0$ and all $\mu_1, \mu_2 \in \mathcal{P}_V(\mathsf{M})$,

$$\rho_\zeta(\mu_1 Q, \mu_2 Q) \leqslant (1 + \zeta C_1) \rho_\zeta(\mu_1, \mu_2) \, .$$

CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France
E-mail address: alain.durmus@cmla.ens-cachan.fr

Laboratoire Jacques-Louis Lions and Laboratoire de Chimie Théorique, Sorbonne Université
E-mail address: pierre.monmarche@sorbonne-universite.fr
URL: https://www.ljll.math.upmc.fr/monmarche/

Laboratoire de Mathématiques Blaise Pascal, CNRS - UMR 6620, Université Clermont-Auvergne, Avenue des landais,, 63177 Aubiere cedex, France
E-mail address: guillin@math.univ-bpclermont.fr
URL: http://math.univ-bpclermont.fr/∼guillin