Submitted to the Annals of Statistics

EXPONENTIAL ERGODICITY OF THE BOUNCY PARTICLE SAMPLER

By George Deligiannidis[‡], Alexandre Bouchard-Côté[§,†] and Arnaud Doucet[‡,*]

University of Oxford[‡] *and University of British Columbia*[§]

Non-reversible Markov chain Monte Carlo schemes based on piecewise deterministic Markov processes have been recently introduced in applied probability, automatic control, physics and statistics. Although these algorithms demonstrate experimentally good performance and are accordingly increasingly used in a wide range of applications, geometric ergodicity results for such schemes have only been established so far under very restrictive assumptions. We give here verifiable conditions on the target distribution under which the Bouncy Particle Sampler algorithm introduced in [37] is geometrically ergodic and we provide a central limit theorem for the associated ergodic averages. This holds essentially whenever the target satisfies a curvature condition and the growth of the negative logarithm of the target is at least linear and at most quadratic. For target distributions with thinner tails, we propose an original modification of this scheme that is geometrically ergodic. For targets with thicker tails, we extend the idea pioneered in [26] in a random walk Metropolis context. We establish geometric ergodicity of the Bouncy Particle Sampler with respect to an appropriate transformation of the target. Mapping the resulting process back to the original parameterization, we obtain a geometrically ergodic piecewise deterministic Markov process.

1. Introduction.

Let $\bar{\pi}(\mathrm{d}x)$ be a Borel probability measure on $\mathbb{R}^d$ admitting a density $\bar{\pi}(x) = \exp\{-U(x)\}/\zeta$ with respect to the Lebesgue measure $\mathrm{d}x$ where $U : \mathbb{R}^d \mapsto [0, \infty)$ is a potential function with locally Lipschitz second derivatives. We assume that this potential function can be evaluated pointwise while ζ is intractable. In this context, one can sample approximately from $\bar{\pi}(\mathrm{d}x)$ and estimate expectations with respect to this measure using Markov chain Monte Carlo (MCMC) algorithms. A wide range of MCMC schemes have been proposed over the past 60 years since the introduction of the Metropolis algorithm.

*A. Doucet was partially supported by UK Engineering and Physical Sciences Research Council grant EP/K000276/1.

†A. Bouchard-Côté was partially supported by a Discovery Grant from the National Science and Engineering Research Council.

Keywords and phrases: piecewise deterministic Markov process, Markov chain Monte Carlo, geometric ergodicity, change of variable, central limit theorem

imsart-aos ver. 2014/10/16 file: finalversion_main_submitted.tex date: May 10, 2018

In particular, non-reversible MCMC algorithms based on piecewise deterministic Markov processes [10, 11] have recently emerged in applied probability [4, 18, 33], [13, Chapter 13], automatic control [30, 31], physics [27, 29, 37] and statistics [3, 6, 16, 36, 41, 42, 43]. These algorithms perform well empirically so they have already found many applications; see, e.g., [13, 22, 27, 35]. However, to the best of our knowledge, geometric convergence rates for this class of MCMC algorithms have only been established under stringent assumptions: [30] establishes geometric ergodicity of such a scheme but only for targets with exponentially decaying tails, [33] obtains sharp results but requires the state-space to be compact, while [2, 4, 18] consider targets on the real line. Similar restrictions apply to limit theorems for ergodic averages, where for example in [2], a Central Limit Theorem (CLT) has been obtained but this result is restricted to targets on the real line. Establishing exponential ergodicity and a CLT under weaker conditions is of interest theoretically but also practically as it lays the theoretical foundations justifying calibrated confidence intervals around Monte Carlo estimates (for a review, see e.g. [25]).

We focus here on the Bouncy Particle Sampler algorithm (BPS), a piecewise deterministic MCMC scheme proposed in [37] and subsequently studied in [6] and [33], as it has been observed to perform empirically very well when compared to other state-of-the-art MCMC algorithms. In addition it has recently been shown in [42] that BPS is the scaling limit of the (discrete-time) reflective slice sampling algorithm introduced in [34]. In this paper we give conditions on the target distribution $\bar{\pi}$ under which BPS is geometrically ergodic. These conditions hold whenever the target satisfies a curvature condition and has *"regular"* tails, in the sense that the potential U grows at least linearly and at most quadratically.

When the target has thin tails, that is U grows faster than a quadratic, we show that a simple modification of the original BPS algorithm provides a geometrically ergodic scheme. This modified BPS algorithm uses a position-dependent rate of refreshment and is easy to implement.

In the presence of thick tails, that is U grows sub-linearly, we follow the approach adopted in [26] for the random walk Metropolis algorithm. We change variables to obtain a transformed target satisfying our conditions and use BPS to sample this transformed target. Mapping this process back to the original parameterization, we obtain a geometrically ergodic algorithm.

All results in the present paper are of a qualitative nature. It would be of interest from a practitioner's point of view to obtain explicit convergence rates to guide the design of efficient algorithms. This is possible by keeping track of the constants in our proofs and applying for example [38, Corollary 4]. However, we expect any rates thus obtained to not be sharp.

We henceforth restrict our attention to dimensions $d \geq 2$; for $d = 1$ BPS coincides with the Zig-Zag process and this one-dimensional process has been shown to be geometrically ergodic under reasonable assumptions in [4]. After submission of this manuscript, two preprints have appeared establishing geometric ergodicity, when $d \geq 2$, of the Zig-Zag process [5] and of a related process modelling the motion of a bacterium [17].

The rest of the paper is structured as follows. Section 2 contains background information on continuous-time Markov processes, exponential ergodicity and BPS. The main results are stated in Section 3. Section 4 establishes several useful ergodic properties of BPS and of its novel variants proposed here. The proofs of the main results can be found in Section 5, whereas lengthy and technical proofs of auxiliary results are provided in the Supplementary Material [12].

2. Background and notation. Let $\{Z_t : t \geq 0\}$ denote a time-homogeneous, continuous-time Markov process on a topological space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$, where $\mathcal{B}(\mathcal{Z})$ is the Borel σ-field of $\mathcal{Z}$, and denote its transition semigroup by $\{P^t : t \geq 0\}$. For every initial condition $Z_0 := z \in \mathcal{Z}$, the process $\{Z_t : t \geq 0\}$ is defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}^z)$, with $\{\mathcal{F}_t\}$ the natural filtration, such that for any $n > 0$, times $0 < t_1 < t_2 < \cdots < t_n$ and $B_1, \ldots, B_n \in \mathcal{B}(\mathcal{Z})$ we have

$$\begin{aligned}
\mathbb{P}^z\{Z_{t_1} \in B_1\} &= \int_{B_1} P^{t_1}(z, \mathrm{d}z_1), \\
\mathbb{P}^z\{Z_{t_1} \in B_1, Z_{t_2} \in B_2\} &= \int_{B_1}\int_{B_2} P^{t_2-t_1}(z_1, \mathrm{d}z_2) P^{t_1}(z, \mathrm{d}z_1), \\
\mathbb{P}^z\left\{Z_{t_1} \in B_1, \ldots, Z_{t_n} \in B_n\right\} &= \int_{B_1}\cdots\int_{B_n} P^{t_n-t_{n-1}}(z_{n-1}, \mathrm{d}z_n) \\
&\qquad \times \cdots \times P^{t_2-t_1}(z_1, \mathrm{d}z_2) P^{t_1}(z, \mathrm{d}z_1).
\end{aligned}$$

We write $\mathbb{E}^z$ to denote expectation with respect to $\mathbb{P}^z$.

Let $\mathfrak{B}(\mathcal{Z})$ denote the space of bounded measurable functions on $\mathcal{Z}$, which is a Banach space with respect to the norm $\|f\|_\infty := \sup_{z\in\mathcal{Z}} |f(z)|$. We also write $\mathcal{M}(\mathcal{Z})$ for the space of σ-finite, signed measures on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$. Given a measurable function $V : \mathcal{Z} \to [1, \infty)$, we define a norm on $\mathcal{M}(\mathcal{Z})$ through

$$\|\mu\|_V := \sup_{|f|\leq V} |\mu(f)|.$$

For any transition kernel $K : \mathcal{Z} \times \mathfrak{B}(\mathcal{Z}) \to [0, 1]$, we define an operator $K : \mathfrak{B}(\mathcal{Z}) \to \mathfrak{B}(\mathcal{Z})$ through $Kf(z) = \int K(z, \mathrm{d}w) f(w)$. We will slightly abuse notation by letting K also denote the dual operator acting on $\mathcal{M}(\mathcal{Z})$ through $\mu K(A) = \int_{\mathcal{Z}} \mu(\mathrm{d}z)K(z,A)$ for $A \in \mathcal{B}(\mathcal{Z})$. With this notation, a σ-finite measure π on $\mathcal{B}(\mathcal{Z})$ is called *invariant* for $\{P^t : t \geq 0\}$ if $\pi P^t = \pi$ for all $t \geq 0$.

2.1. *Exponential ergodicity of continuous-time processes.* Suppose that a Borel probability measure π is invariant for $\{P^t : t \geq 0\}$. We are interested in the exponential convergence of the process in the sense of *V-uniform ergodicity*: that is there exists a measurable function $V : \mathcal{Z} \to [1,\infty)$ and constants $D < \infty$ and $\rho < 1$ such that

$$\|P^t(z,\cdot) - \pi(\cdot)\|_V \leq V(z)D\rho^t, \qquad t \geq 0. \tag{2.1}$$

The proof of V-uniform ergodicity usually proceeds through the verification of an appropriate *drift condition* which is often expressed in terms of the *strong generator* (see for example [11, pg. 28]). However, in this paper, it will prove useful to focus on the *extended generator* of the Markov process $\{Z_t : t \geq 0\}$ which is defined as follows. Let $\mathcal{D}(\widetilde{\mathcal{L}})$ denote the set of measurable functions $f : \mathcal{Z} \to \mathbb{R}$ for which there exists a measurable function $h : \mathcal{Z} \to \mathbb{R}$ such that $t \mapsto h(Z_t)$ is integrable $\mathbb{P}^z$-almost surely for each $z \in \mathcal{Z}$ and the process

$$f(Z_t) - f(z) - \int_0^t h(Z_s)\,\mathrm{d}s, \qquad t \geq 0,$$

is a local $\mathcal{F}_t$-martingale. Then we write $h = \widetilde{\mathcal{L}}f$ and we say that $(\widetilde{\mathcal{L}}, \mathcal{D}(\widetilde{\mathcal{L}}))$ is the *extended generator* of the process $\{Z_t : t \geq 0\}$. This is an extension of the usual *strong generator* associated with a Markov process; for more details see [11, Sections 14 and 26] and references therein. We will also need the concepts of *aperiodicity* (see [15, p. 1675]), *irreducibility*, *small sets* and *petite sets* ([15, p. 1674]).

2.2. *The Bouncy Particle Sampler.* We begin with some additional notation. We will consider $x \in \mathbb{R}^d$ as a column vector and we will write $|\cdot|$ and $\langle \cdot, \cdot \rangle$ to denote the Euclidean norm and scalar product in $\mathbb{R}^d$ respectively, whereas $\|A\| = \sup\{|Ax| : |x| = 1\}$ will denote the operator norm of the matrix $A \in \mathbb{R}^{d\times d}$. Let $B(x,\delta) := \{y \in \mathbb{R}^d : |x-y| < \delta\}$. For a function $U : \mathbb{R}^d \to [0,\infty)$ we write $\nabla U(x)$ and $\Delta U(x)$ for the gradient and the Hessian of $U(\cdot)$ evaluated at x and we adopt the convention of treating $\nabla U(x)$ as a column vector. For a differentiable map $h : \mathbb{R}^d \to \mathbb{R}^d$ we will write ∇h for the Jacobian of h; that is, letting $h = (h_1, \ldots, h_d)^T$, we have $(\nabla h)_{i,j} = \partial_{x_i} h_j$. Let us write ψ for the uniform measure on $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d : |v| = 1\}$, $p_\vartheta(\cdot)$

for the density of the angle between a fixed unit length vector and a random vector sampled from $\psi(\cdot)$, which is given by

$$p_{\vartheta}(\theta) := \kappa_d (\sin\theta)^{d-2}, \quad \kappa_d = \left(\int_0^{\pi} (\sin\theta)^{d-2}\, \mathrm{d}\theta\right)^{-1}, \quad \theta \in [0,\pi], \tag{2.2}$$

and let $\mathcal{Z} := \mathbb{R}^d \times \mathbb{S}^{d-1}$ and $\pi(\mathrm{d}x, \mathrm{d}v) := \bar{\pi}(\mathrm{d}x)\psi(\mathrm{d}v)$. For $(x,v) \in \mathcal{Z}$, we also define

$$R(x)v := v - 2\frac{\langle \nabla U(x), v\rangle}{|\nabla U(x)|^2}\nabla U(x). \tag{2.3}$$

The vector $R(x)v$ can be interpreted as a Newtonian collision on the hyperplane orthogonal to the gradient of the potential U, hence the interpretation of x as a position, and v, as a velocity.

BPS defines a π-invariant, non-reversible, piecewise deterministic Markov process $\{Z_t : t \geq 0\} = \{(X_t, V_t) : t \geq 0\}$ taking values in $\mathcal{Z}$. Since π admits $\bar{\pi}$ as a marginal, we can use this scheme to approximate expectations with respect to $\bar{\pi}$. We introduce here a slightly more general version of BPS than the one discussed in [1, 6, 33, 37]. Let

$$\begin{aligned} \bar{\lambda}(x,v) &:= \Lambda_{\mathrm{ref}}(x) + \lambda(x,v), \\ \lambda(x,v) &:= \max\{0, \langle \nabla U(x), v\rangle\} =: \langle \nabla U(x), v\rangle_+, \end{aligned} \tag{2.4}$$

where the *refreshment rate* $\Lambda_{\mathrm{ref}}(\cdot) : \mathbb{R}^d \mapsto (0,\infty)$ is allowed to depend on the location x. Previous versions of BPS restrict attention to the case $\Lambda_{\mathrm{ref}}(x) = \lambda_{\mathrm{ref}}$; the generalisation considered here will prove useful in establishing the geometric ergodicity of this scheme for thin-tailed targets.

Given any initial condition $z \in \mathcal{Z}$, a construction of a path of BPS is given in Algorithm 1. Step 4 of this algorithm corresponds to the simulation of the first arrival time of an inhomogeneous Poisson process. Simulating such arrival times is a well-studied problem and various exact simulation techniques can be found in [14, Chapter 6]. In the specific BPS context, these techniques have been detailed in [6, 37]. Equivalently, BPS can be defined as the Markov process on $\mathcal{Z}$ with extended generator given by

$$\widetilde{\mathcal{L}} f(x,v) = \mathfrak{V} f(x,v) + \bar{\lambda}(x,v)\left[K f(x,v) - f(x,v)\right], \tag{2.5}$$

for $f \in \mathcal{D}(\widetilde{\mathcal{L}})$, the domain of $\widetilde{\mathcal{L}}$ (see Section 5.1), where

$$\mathfrak{V} f(x,v) := \frac{\mathrm{d}}{\mathrm{d}t} f(x+tv, v)\Big|_{t=0+}, \tag{2.6}$$

Algorithm 1 : Bouncy Particle Sampler algorithm

$(X_0, V_0) \leftarrow (x, v)$
$t_0 \leftarrow 0$
for $k = 1, 2, 3, \dots$ **do**
  sample inter-event time τ_k, where τ_k is a positive random variable such that
$$\mathbb{P}[\tau_k \geq t] = \exp\left\{ - \int_0^t \bar{\lambda}(X_{t_{k-1}} + rV_{t_{k-1}}, V_{t_{k-1}})\, \mathrm{d}r \right\}$$
  for $r \in (0, \tau_k)$ set $(X_{t_{k-1}+r}, V_{t_{k-1}+r}) \leftarrow (X_{t_{k-1}} + rV_{t_{k-1}}, V_{t_{k-1}})$
  $t_k \leftarrow t_{k-1} + \tau_k$ ▷ Time of k-th event
  $X_{t_k} \leftarrow X_{t_{k-1}} + \tau_k V_{t_{k-1}}$
  if $U_k < \lambda(X_{t_k}, V_{t_{k-1}})/\bar{\lambda}(X_{t_k}, V_{t_{k-1}})$, where $U_k \sim \mathrm{Uniform}(0, 1)$ **then**
    $V_{t_k} \leftarrow R(X_{t_k})V_{t_{k-1}}$ ▷ Newtonian collision on the gradient ("bounce")
  else
    $V_{t_k} \sim \psi$ ▷ Refreshment of the velocity
  end if
end for

and the transition kernel $K : \mathcal{Z} \times \mathcal{B}(\mathcal{Z}) \mapsto [0, 1]$ is defined through
(2.7)
$$K\left((x, v), (\mathrm{d}y, \mathrm{d}w)\right) = \frac{\Lambda_{\mathrm{ref}}(x)}{\bar{\lambda}(x, v)} \delta_x(\mathrm{d}y)\psi(\mathrm{d}w) + \frac{\lambda(x, v)}{\bar{\lambda}(x, v)} \delta_x(\mathrm{d}y)\delta_{R(x)v}(\mathrm{d}w).$$

For a continuously differentiable $f \in \mathcal{D}(\widetilde{\mathcal{L}})$ the expression (2.5) reduces to

$$\widetilde{\mathcal{L}} f(x, v) = \langle \nabla_x f(x, v), v \rangle + \bar{\lambda}(x, v) \left[Kf(x, v) - f(x, v) \right]. \tag{2.8}$$

For $\Lambda_{\mathrm{ref}}(x) = \lambda_{\mathrm{ref}} > 0$, it has been shown in [6] that BPS is ergodic, provided U is continuously differentiable, when the velocities are distributed according to a normal distribution rather than uniformly on the sphere $\mathbb{S}^{d-1}$ as assumed here. Restricting velocities to $\mathbb{S}^{d-1}$ makes our calculations more tractable without significantly altering the properties of the process. In this context, [33] considers only compact state spaces but the arguments therein can be adapted to prove ergodicity in the general case.

3. Main results. In this paper, we provide sufficient conditions on the target measure $\bar{\pi}$ and the refreshment rate for BPS to be V-uniformly ergodic for the following Lyapunov function[1]

$$V(x, v) := \frac{\mathrm{e}^{U(x)/2}}{\bar{\lambda}(x, -v)^{1/2}}. \tag{3.1}$$

[1] In [30], the Lyapunov function $\mathrm{e}^{U(x)/2}\bar{\lambda}(x, v)^{1/2}$ is used to establish the geometric ergodicity of a different piecewise deterministic MCMC scheme for targets with exponential tails but we found this function did not apply to BPS.

	Target distributions					
		$\pi(x) \propto \exp\left\{-(1+\|x\|^2)^{\beta/2}\right\}$				
Sampling methods	t-distributions	$\beta \in (0,1)$ Thick tails	$\beta = 1$ Exponential	$\beta \in (1,2)$	$\beta = 2$ Gaussian	$\beta > 2$ Thin tails
BPS			**Yes** Thm. 3.1(B)	**Yes** Thm. 3.1(A)	**Yes** Thm. 3.1(A)	
BPS with position-dependent refreshment						**Yes** Thm. 3.2
BPS with transformation of the target	**Yes** Thm. 3.3(A)	**Yes** Thm. 3.3(B)				
Random walk Metropolis	**No** [24]	**No** [24]			**Yes** [40]	**Yes** [40]
Random walk Metropolis with transformation of the target	**Yes** [26] Sec. 3.3		**Yes** [26] Thm. 2 and 4	**Yes** [26] Cor. 1 and 2		
Metropolis Adjusted Langevin Algorithm (1D)	**No** [39] Thm. 4.3	**No** [39] Thm. 4.3	**Yes** [32] Sec. 16.1.3	**Yes** [39] Thm. 4.1	**Yes** [39] Thm. 4.1	**No** [39] Thm. 4.2
Hamiltonian Monte Carlo		**No** [28] Cor. 2.3(ii)	**Yes** [28] Cor. 2.3(i)	**Yes** [28] Cor. 2.3(i)	**Yes** [28] Cor. 2.3(i)	**No** [28] Cor. 2.3(ii)

TABLE 1. *Summary of geometric ergodicity (or proven lack thereof) for various sampling methods on targets with tails decreasing as* $\exp\left\{-|x|^\beta\right\}$ *for* $\beta \in (0,\infty)$ *and t-distributions. These models cover two important challenging situations: essentially, cases where the gradient of the potential becomes negligible in the tails (two leftmost columns) and cases where both the gradient and the Hessian are unbounded (rightmost column). See references for precise conditions.*

Throughout this section, we refer the reader to Table 1 for examples of target distributions with various tail behaviours where each of our Theorems is used to establish exponential ergodicity. All proofs are given in Section 5 and the Supplementary Material [12]. Before stating our results we make a few working assumptions.

ASSUMPTIONS. *Let $U : \mathbb{R}^d \to [0, \infty)$ be such that*

$$\text{(A0)} \qquad \frac{\partial^2 U(x)}{\partial x_i \partial x_j} \text{ is locally Lipschitz continuous for all } i, j,$$

$$\text{(A1)} \qquad \int_{\mathbb{R}^d} \bar{\pi}(\,\mathrm{d}x)|\nabla U(x)| < \infty,$$

$$\text{(A2)} \qquad \varliminf_{|x|\to\infty} \frac{\mathrm{e}^{U(x)/2}}{\sqrt{|\nabla U(x)|}} > 0,$$

$$\text{(A3)} \qquad V \geq c, \qquad \text{for some } c > 0.$$

REMARK 1. *Assumption (A3) is not restrictive as in view of Assumption (A2), $V \geq c$ may only fail locally near the origin. Therefore if $V \geq c$ fails inside the ball $B(0, M)$, we can always replace V with $\widetilde{V} = V + \mathbb{1}_{B(0,M)} \geq 1$. This modified Lyapunov function still belongs to the domain $\mathcal{D}(\widetilde{\mathcal{L}})$ of the extended generator $\widetilde{\mathcal{L}}$ as explained in Section 5.1.*

REMARK 2. *From the proofs, it is clear that Theorems 3.1 and 3.2 detailed below remain true if we replace Assumption (A0) by the following slightly weaker assumption*

$$\text{(A0')} \qquad \begin{array}{l} t \mapsto \langle \nabla U(x + tv), v \rangle \text{ is locally Lipshitz for all } (x, v) \in \mathcal{Z}, \text{ and} \\ \text{(A0) holds for all } |x| > R, \text{ for some } R > 0. \end{array}$$

Although cumbersome, this alternative formulation is useful in the proof of Theorem 3.3 which relies on Theorems 3.1 and 3.2.

Rather than requiring that $U \geq 0$, we could equivalently require that U is bounded below. This guarantees that the density $\bar{\pi}$ is bounded above. Assumptions (A0) to (A3) are technical conditions and it may be possible to relax them. For example, Assumption (A1) allows us to use the approach of [8] to establish the invariance of the process. Other approaches exist, for example [6]. When the refreshment rate depends on the location we assume in addition that $\int \Lambda_{\text{ref}}(x)\bar{\pi}(\,\mathrm{d}x) < \infty$. Under these conditions the embedded discrete-time Markov chain $\{\Theta_k : k \geq 0\} := \{(X_{\tau_k}, V_{\tau_k}) : k \geq 0\}$ admits an invariant probability measure; see [8] and Lemma 1. The Lyapunov function (3.1) is proportional to the inverse of the square root of the invariant distribution of this embedded discrete-time Markov chain. There may exist other Lyapunov functions allowing different set of conditions compared to (A2) and (A3).

3.1. *"Regular" tails.* We now state our first main result.

THEOREM 3.1. *Suppose that Assumptions (A0)-(A3) hold. Let $\Lambda_{\text{ref}}(\cdot) = \lambda_{\text{ref}} > 0$ and suppose that one of the following sets of conditions holds:*

(A) $\underline{\lim}_{|x|\to\infty} |\nabla U(x)| = \infty$, $\overline{\lim}_{|x|\to\infty} \|\Delta U(x)\| \leq \alpha_1 < \infty$ *and* $\lambda_{\text{ref}} > (2\alpha_1 + 1)^2$,

(B) $\underline{\lim}_{|x|\to\infty} |\nabla U(x)| = 2\alpha_2 > 0$, $\overline{\lim}_{|x|\to\infty} \|\Delta U(x)\| \leq C < \infty$ *and* $\lambda_{\text{ref}} \leq \alpha_2/c_d$, *with* $c_d := 16\sqrt{d}$.

Then BPS is V-uniformly ergodic.

In summary, BPS with an appropriately chosen constant refreshment rate $\Lambda_{\text{ref}}(\cdot) = \lambda_{\text{ref}} > 0$ is exponentially ergodic for targets with tails that decay at least as fast as an exponential and at most as fast as a Gaussian. In addition the uniform bound on the Hessian imposes some regularity on the curvature of the target. The proof of Theorem 3.1 is provided in Section 5, building on the auxiliary results of Section 4. The conditions imposed on the refreshment rate are sufficient but not sharp.

We provide here an example of a common Bayesian model which yields a posterior density satisfying the assumptions of Theorem 3.1.

EXAMPLE 1. **Bayesian logistic regression.** *Consider binary observations $(y_1, ..., y_n) \in \{0, 1\}^n$ and associated $\mathbb{R}^d$-valued predictors $c_1, ..., c_n$. We assume the observations are conditionally independent given the predictors and regression coefficients $x \in \mathbb{R}^d$ and satisfy*

$$\mathbb{P}\left(Y_i = 1|x, c_i\right) = 1/\left(1 + e^{-\langle x,c_i\rangle}\right) = \rho_i\left(x\right).$$

We assign a prior distribution to x of negative log density $\sum_{k=1}^{d} g(x_k)$ where g is twice differentiable. Hence the potential associated to the posterior density of x given the observations satisfies

$$U\left(x\right) = \sum_{i=1}^{d} g(x_k) + \sum_{i=1}^{n} \left\{-y_i\langle c_i, x\rangle + \log\left(1 + e^{\langle x,c_i\rangle}\right)\right\}. \tag{3.2}$$

Its gradient with respect to x is given by

$$\nabla U\left(x\right)=\nabla g(x)+\sum_{i=1}^{n}\left\{-y_{i}+\rho_{i}\left(x\right)\right\}c_{i}, \tag{3.3}$$

where $\nabla g(x):=(g'(x_1),...,g'(x_d))^T$ while its Hessian satisfies

$$\Delta U\left(x\right):=\Delta g(x)+\sum_{i=1}^{n}\rho_{i}\left(x\right)\left\{1-\rho_{i}\left(x\right)\right\}c_{i}c_{i}^{T}, \tag{3.4}$$

with $\Delta g(x):=\mathrm{diag}(g''(x_1),...,g''(x_d))$. Hence for an isotropic Gaussian prior of covariance $\sigma^2 I_d$, we have $g(v)=v^2/(2\sigma^2)$ and U satisfies condition 3.1 (A). For a smoothed Laplace prior, i.e. $g(v)=(1+v^2/\sigma^2)^{1/2}$, U satisfies condition 3.1 (B).

Theorem 3.1 does not apply to targets with tails thinner than Gaussian or thicker than exponential distributions. As summarised in Table 1, it is also known that Metropolis adjusted Langevin algorithm (MALA), see [39, Theorems 4.2 and 4.3], and Hamiltonian Monte Carlo (HMC), see [28, Theorems 5.13 and 5.17], are not geometrically ergodic for such targets. We now turn our attention to these cases.

3.2. *Thin-tailed targets.* When the target is *thin-tailed*, in the sense that the gradient of its potential U grows super-linearly in the tails, any constant refreshment rate will eventually be negligible. It has been shown in [6] that BPS without refreshment is not ergodic as the process can remain forever outside a ball of positive radius. In our case the refreshment rate does not vanish, but refreshment in the tails will be extremely rare. This will result in long excursions during which the process will not explore the centre of the space.

The above discussion suggests that, for *thin-tailed* targets, we need to scale the refreshment rate accordingly in order for it to remain non-negligible in the tails. The next result makes this intuition more precise.

THEOREM 3.2. *Suppose that Assumptions (A0)-(A3) hold. Let $\lambda_{\mathrm{ref}}>0$ and define for some $\epsilon>0$*

$$\Lambda_{\mathrm{ref}}(x):=\lambda_{\mathrm{ref}}+\frac{|\nabla U(x)|}{\max\{1,|x|^{\epsilon}\}}. \tag{3.5}$$

Suppose that

$$\varliminf_{|x|\to\infty}\frac{|\nabla U(x)|}{|x|}=\infty,\qquad \lim_{|x|\to\infty}\frac{\|\Delta U(x)\|}{|\nabla U(x)|}|x|^{\epsilon}=0.$$

Then BPS is V-uniformly ergodic.

The proof of Theorem 3.2 is given in Section 5. It is worth noting that although Langevin diffusions can be geometrically ergodic for thin-tailed targets, they typically cannot be simulated exactly. When they are discretised, an additional Metropolis–Hastings step is needed to sample from the correct target distribution and the resulting MALA algorithm is not geometrically ergodic [39, Theorem 4.2].

We next provide an example of a common Bayesian model which yields a posterior density satisfying the assumptions of Theorem 3.2.

EXAMPLE 2. **Bayesian logistic regression (continued).** *In the context of the logistic regression model of Example 1, although priors whose tails decrease like a Gaussian or an exponential are very popular in the literature, alternatives have also been proposed, e.g., [21]. In particular, if we select $g(u) = (1 + u^2/\sigma^2)^{\beta/2}$ with $\beta > 2$ then the potential U given in (3.2) satisfies the conditions of Theorem 3.2.*

3.3. *Thick-tailed targets.* For targets with tails thicker than an exponential, that is when the gradient of the potential U vanishes in the tails, the lack of exponential ergodicity of gradient-based methods such as MALA and HMC, is natural—the vanishing gradient induces random-walk like behaviour in the tails. This seems to be the main obstruction preventing extension of Theorem 3.1 to thick-tailed distributions.

However, following the approach of [26], we can address this by transforming the target to one satisfying the assumptions of either Theorem 3.1, or Theorem 3.2. This guarantees that BPS with respect to the transformed target will be geometrically ergodic. By mapping back this BPS process to the original parameterization, we obtain a geometrically ergodic piecewise deterministic Markov process with non-linear dynamics.

As in [26] we define the following functions $f^{(1)}, f^{(2)} : [0, \infty) \to [0, \infty)$:

$$f^{(1)}(r) = \begin{cases} \mathrm{e}^{br} - \frac{\mathrm{e}}{3}, & r > \frac{1}{b}, \\ r^3 \frac{b^3 \mathrm{e}}{6} + r \frac{b\mathrm{e}}{2}, & r \le \frac{1}{b}, \end{cases} \tag{3.6}$$

and

$$f^{(2)}(r) = \begin{cases} r, & r \le R, \\ r + (r - R)^p, & r > R, \end{cases} \tag{3.7}$$

where $R, b > 0$ are arbitrary constants. We also define the *isotropic transformations* $h^{(i)} : \mathbb{R}^d \to \mathbb{R}^d$, given by

$$h^{(i)}(x) := \begin{cases} \dfrac{f^{(i)}(|x|)x}{|x|}, & \text{for } x \neq 0, \\ 0, & \text{for } x = 0. \end{cases} \tag{3.8}$$

From [26, Lemma 1] it follows that for $i = 1, 2$, $h = h^{(i)} : \mathbb{R}^d \mapsto \mathbb{R}^d$ defines a C^1-*diffeomorphism*, that is h is bijective with $h, h^{-1} \in C^1(\mathbb{R})$.

Let $h = h^{(i)}$ for some $i \in \{1, 2\}$, $X \sim \bar{\pi}$ and $Y = h^{-1}(X)$. Then $Y \in \mathbb{R}^d$ is distributed according to the Borel probability measure $\bar{\pi}_h$, with density given by $\bar{\pi}_h(y) = \exp\{-U_h(y)\}/\zeta_h$, where by [26, Eqns. (6) and (7)] we have that

$$U_h(y) = U(h(y)) - \log\det(\nabla h(y)), \tag{3.9}$$

$$\nabla U_h(y) = \nabla h(y) \nabla U(h(y)) - \nabla \log\det(\nabla h(y)). \tag{3.10}$$

Let $\{(Y_t, V_t); t \geq 0\}$ denote the trajectory produced by the BPS algorithm targeting $\pi_h(y, v) := \bar{\pi}_h(y)\psi(v)$ and let

$$V_h(x, v) := \frac{\mathrm{e}^{U_h(x)/2}}{[\Lambda_{\mathrm{ref}}(x) + \langle \nabla U_h(x), -v\rangle_+]^{1/2}}.$$

THEOREM 3.3. *Let U satisfy Assumption (A0).*

(A) *If for some $\mathfrak{d} > d$*

(i) $\overline{\lim}_{|x|\to\infty} |x||\nabla U(x)| < \infty$,

(ii) $\overline{\lim}_{|x|\to\infty} |x|^2 \|\Delta U(x)\| < \infty$, *and*

(iii) $\underline{\lim}_{|x|\to\infty} \langle x, \nabla U(x)\rangle = \mathfrak{d}$,

then $U_{h^{(1)}}$, with $h^{(1)}$ defined via (3.6), satisfies the assumptions of Theorem 3.1 (B). If in addition $\Lambda_{\mathrm{ref}}(\cdot) = \lambda_{\mathrm{ref}} \leq b(\mathfrak{d} - d)/32\sqrt{d}$, with b as in (3.6), then the process $\{(X_t, V_t) : t \geq 0\}$, where $X_t = h^{(1)}(Y_t)$, is π-invariant and $\widetilde{V}$-uniformly ergodic, where $\widetilde{V} = V_{h^{(1)}} \circ H^{(1)}$ with $H^{(1)}(x, v) := (h^{(1)}(x), v)$.

(B) *If for some $\beta \in (0, 1)$ we have*

(i) $\overline{\lim}_{|x|\to\infty} |x|^{1-\beta} |\nabla U(x)| < \infty$,

(ii) $\underline{\lim}_{|x|\to\infty} |x|^{-\beta} \langle x, \nabla U(x)\rangle > 0$, *and*

(iii) $\overline{\lim}_{|x|\to\infty} |x|^{2-\beta} \|\Delta U(x)\| < \infty$,

then $U_{h^{(2)}}$, with $h^{(2)}$ defined via (3.7) and p such that $\beta p > 2$, satisfies the assumptions of Theorem 3.2. If in addition $\Lambda_{\mathrm{ref}}(\cdot)$ is given by (3.5) with $U := U_{h^{(2)}}$, then the process $\{(X_t, V_t) : t \geq 0\}$, where $X_t = h^{(2)}(Y_t)$, is π-invariant and $\widetilde{V}$-uniformly ergodic, where $\widetilde{V} = V_{h^{(2)}} \circ H^{(2)}$ with $H^{(2)}(x, v) := (h^{(2)}(x), v)$.

The proof of this theorem is given in Section 5.3.

EXAMPLE 3. **Multivariate t-distribution.** *Suppose that $x \in \mathbb{R}^d$, for $d \geq 2$, $k > 1$, and let*

$$\bar{\pi}(x) \propto \mathrm{e}^{-U(x)} = \left[1 + \frac{|x|^2}{k}\right]^{-\frac{k+d}{2}}.$$

It follows that

$$\nabla U(x) = \frac{(k+d)}{(k+|x|^2)} x, \quad \Delta U(x) = \frac{k+d}{k+|x|^2}\mathbb{1}_d - 2\frac{(k+d)xx^T}{(k+|x|^2)^2},$$

where $\mathbb{1}_d$ is the $d \times d$ identity matrix. Then U satisfies the conditions of Theorem 3.3(A). We refer the reader to [26], Section 3.4, for a related example arising from Bayesian inference.

Generalised Gaussian distribution. *Let $U(x) = (1 + |x|^2)^{\beta/2}$ for some $\beta \in (0, 1)$. Then U satisfies the conditions of Theorem 3.3(B).*

REMARK 3. *In the context of Theorem 3.3(A), while geometric ergodicity holds for all positive fixed b, tuning this parameter may be useful in practice as pointed out by [26].*

3.4. *A Central Limit Theorem.* From the above results we obtain the following CLT, proven in Section 5.4, for the estimator $T^{-1}\int_0^T g(Z_s)\,\mathrm{d}s$ of $\pi(g)$. This estimator can be computed exactly when g is a multivariate polynomial of the components of z; see, e.g., [6, Section 2.4].

THEOREM 3.4. *Suppose that any of the conditions of Theorems 3.1 or 3.2 hold. Let $\varepsilon > 0$ such that $W := V^{1-\varepsilon}$, satisfies $\pi(W^2) < \infty$. Then for any $g : \mathcal{Z} \to \mathbb{R}$ such that $g^2 \leq W$ and for any initial distribution, we have that*

$$\frac{1}{\sqrt{T}} S_T\left[g - \pi(g)\right] \Rightarrow \mathcal{N}(0, \sigma_g^2),$$

with

$$S_T[g] := \int_0^T g(Z_s)\,\mathrm{d}s, \qquad \sigma_g^2 := 2\int \hat{g}(z)\left[g(z) - \pi(g)\right]\pi(\,\mathrm{d}z),$$

where $\hat{g}$ is the solution of the Poisson equation $g - \pi(g) = -\mathcal{L}\hat{g}$, and satisfies $|\hat{g}| \leq c_0(1+W)$ for some constant c_0.

COROLLARY 1. *Suppose that the conditions of Theorem 3.3(A) or Theorem 3.3(B) hold, let $h = h^{(1)}, h^{(2)}$ respectively, define $H(x,v) = (h(x), v)$, and let $\widetilde{V} = V_h \circ H$ denote the corresponding Lyapunov function. Let $\varepsilon > 0$ such that $W := \widetilde{V}^{1-\varepsilon}$, satisfies $\pi_h(W^2) < \infty$. Then for any $g : \mathcal{Z} \to \mathbb{R}$ such that $g^2 \leq W$ and for any initial distribution, we have that*

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \int_0^T [g(X_t, V_t) - \pi(g)] \,\mathrm{d}t \\
&= \frac{1}{\sqrt{T}} \int_0^T [g \circ H(Y_t, V_t) - \pi_h(g \circ H)] \,\mathrm{d}t \Rightarrow \mathcal{N}(0, \widetilde{\sigma}_g^2),
\end{aligned}$$

with

$$\widetilde{\sigma}_g^2 := 2 \int \widehat{g \circ H}(z) \, [g \circ H(z) - \pi_h(g)] \, \pi_h(\mathrm{d}z),$$

where $\widehat{g \circ H}$ is the solution of the Poisson equation $g \circ H - \pi(g \circ H) = -\mathcal{L}_h \widehat{g \circ H}$, and $\mathcal{L}_h$ is given in (2.5) with $\bar{\lambda}$ defined in (2.4) with U replaced by U_h and K defined in (2.7) using $R(x)v$ defined in (2.3) with ∇U_h replacing ∇U.

4. Auxiliary results. To prove V-uniform ergodicity we will use the following result.

THEOREM A. *[15, Theorem 5.2] Let $\{Z_t : t \geq 0\}$ be a Borel right Markov process taking values in a locally compact, separable metric space $\mathcal{Z}$ and assume it is non-explosive, irreducible and aperiodic. Let $(\widetilde{\mathcal{L}}, \mathcal{D}(\widetilde{\mathcal{L}}))$ be its extended generator. Suppose that there exists a measurable function $V : \mathcal{Z} \to [1, \infty)$ such that $V \in \mathcal{D}(\widetilde{\mathcal{L}})$, and that for a petite set $C \in \mathcal{B}(\mathcal{Z})$ and constants $b, c > 0$ we have*

$$\widetilde{\mathcal{L}} V \leq -cV + b\mathbb{1}_C. \tag{$\mathfrak{D}$}$$

Then $\{Z_t : t \geq 0\}$ is V-uniformly ergodic.

The BPS processes considered in this paper can be easily seen to satisfy the *standard conditions* in [11, Section 24.8], and thus by [11, Theorem 27.8] it follows that they are Borel right Markov processes. In addition since the process moves at unit speed, for any $z = (x, v) \in \mathcal{Z}$ the first exit time from $B(0, |x| + M) \times \mathbb{S}^{d-1}$ is at least M, and thus, BPS is non-explosive.

We will next show that BPS remains π-invariant when the refreshment rate is allowed to vary with x, and that it is irreducible and aperiodic. Finally we will show that all compact sets are *small*, hence *petite*. To complete the proofs of Theorems 3.1 and 3.2 it remains to establish that V satisfies ($\mathfrak{D}$) which is done in Section 5.

LEMMA 1. *Suppose that the map $t \mapsto U(x+tv)$ is absolutely continuous for all $(x,v) \in \mathcal{Z}$, that Assumption (A1) holds and that $\int \Lambda_{\text{ref}}(x)\bar{\pi}(\,\mathrm{d}x) < \infty$. Then BPS with refreshment rate $\Lambda_{\text{ref}}(\cdot)$ is invariant with respect to π.*

The proof of Lemma 1 is based on [8], see also [9], where the authors provide a link between the invariant measures of $\{Z_t : t \geq 0\}$ and those of the embedded discrete-time Markov chain $\{\Theta_k : k \geq 0\} := \{(X_{\tau_k}, V_{\tau_k}) : k \geq 0\}$ which tracks the process just after events. The details are given in the Supplementary Material [12].

Notice that when $\Lambda_{\text{ref}}(\cdot)$ is given by (3.5), the condition $\int \Lambda_{\text{ref}}(x)\bar{\pi}(\,\mathrm{d}x) < \infty$ is implied by (A1).

REMARK 4. *The Markov chain $\{\Theta_k : k \geq 0\}$ admits an invariant probability measure proportional to $\bar{\lambda}(x,-v)\pi(\,\mathrm{d}x,\,\mathrm{d}v)$. It follows from a simple change of measure argument that under ergodicity and integrability conditions one has*

$$\frac{\sum_{k=1}^{n} g\left(X_{\tau_k}, V_{\tau_k}\right)/\bar{\lambda}\left(X_{\tau_k}, -V_{\tau_k}\right)}{\sum_{k=1}^{n} 1/\bar{\lambda}\left(X_{\tau_k}, -V_{\tau_k}\right)} \to \pi(g) \qquad \text{a.s. as } n \to \infty. \tag{4.1}$$

This is an alternative estimator of $\pi(g)$ compared to $T^{-1}\int_0^T g(Z_s)\,\mathrm{d}s$.

The next result establishes the existence of small sets as well as the irreducibility of the process.

LEMMA 2. *Suppose that $\Lambda_{\text{ref}}(\cdot) > \lambda_{\text{ref}} > 0$. For all $T > 0$, $z := (x_0, v_0) \in B(0, T/6) \times \mathbb{S}^{d-1}$, and Borel set $A \subseteq B(0, \frac{T}{6}) \times \mathbb{S}^{d-1}$,*

$$\mathbb{P}^z(Z_T \in A) \geq C(T, d, \lambda_{\text{ref}}) \iint_A \psi(\,\mathrm{d}v)\,\mathrm{d}x,$$

for some constant $C(T, d, \lambda_{\text{ref}}) > 0$ depending only on $T, d, \lambda_{\text{ref}}$. Hence, all compact sets are small. Moreover, the process $\{Z_t : t \geq 0\}$ is irreducible.

The proof of Lemma 2 leverages the refreshment events to construct paths connecting arbitrary points. The details are provided in the Supplementary Material [12].

Lemma 3. *The process $\{Z_t : t \geq 0\}$ is aperiodic.*

Proof of Lemma 3. We show that for some small set A', there exists a T such that $P^t(z, A') > 0$ for all $t \geq T$ and $z \in A'$.

Let $A' := B(0,1) \times \mathbb{S}^{d-1}$, $T = 6$, and suppose that $t > T$. By Lemma 2, for all $z \in B(0, t/6) \times \mathbb{S}^{d-1}$ and Borel set $A \subset B(0, t/6) \times \mathbb{S}^{d-1}$, we have

$$\mathbb{P}^z(Z_t \in A) \geq C(t, d, \lambda_{\text{ref}}) \iint_A \psi(\mathrm{d}v)\, \mathrm{d}x,$$

for some $C(t, d, \lambda_{\text{ref}}) > 0$. Hence, by picking $A = A'$, we have, since $B(0,1) \subset B(0, t/6)$, that for all $z \in A'$,

$$\mathbb{P}^z(Z_t \in A') \geq C(t, d, \lambda_{\text{ref}}) \iint_{A'} \psi(\mathrm{d}v)\, \mathrm{d}x > 0. \qquad \square$$

5. Proofs of main results. To complete the proofs of Theorems 3.1 and 3.2 it remains to show that $V : \mathcal{Z} \to [0, \infty)$, defined in (3.1), satisfies the *drift condition* ($\mathfrak{D}$).

5.1. *Extended Generator of BPS.* We first need show that V belongs to $\mathcal{D}(\widetilde{\mathcal{L}})$, the domain of the *extended generator* $\widetilde{\mathcal{L}}$ (see [11, Section 26]), which suffices for Theorem A to apply. By Assumption (A0'), or the stronger Assumption (A0), it easily follows that for all (x, v) the function $t \mapsto V(x + tv, v)$ is locally Lipschitz and thus absolutely continuous [11, Proposition 11.8]. Therefore by [11, Theorem 26.14], since there is no *boundary* (see [11, Section 24]), V is bounded as a function of v and the jump rate $\bar{\lambda}$ is locally bounded, it follows that $V \in \mathcal{D}(\widetilde{\mathcal{L}})$.

The fact that $\widetilde{\mathcal{L}}$ is given by (2.5) follows from the proof of [11, Theorem 26.14, bottom of page 71]. Indeed, for any fixed $z = (x, v) \in \mathcal{Z}$, let $\{T_i\}_{i \geq 1}$ denote the event times of BPS started from (x, v), the paths of which we denote with $\{Z_t : t \geq 0\}$, where $Z_t = (X_t, V_t)$. Since $t \mapsto V(x + tv, v)$ is absolutely continuous, its left and right derivatives of $V(x + tv, v)$ coincide almost everywhere and thus we can write

$$\begin{aligned} V\left(Z_{T_i^-}\right) - V\left(Z_{T_{i-1}}\right) &= \int_0^{T_i - T_{i-1}} \frac{\mathrm{d}}{\mathrm{d}s} V\left(X_{T_{i-1}} + sV_{T_{i-1}}, V_{T_{i-1}}\right) \mathrm{d}s \\ &= \int_0^{T_i - T_{i-1}} \mathfrak{V} V\left(X_{T_{i-1}} + sV_{T_{i-1}}, V_{T_{i-1}}\right) \mathrm{d}s, \end{aligned}$$

where $\mathfrak{V}$ is defined in (2.6). From this and the proof of the first part of [11, Theorem 26.14] it follows that

$$V(Z_t) - V(z) - \int_0^t \mathfrak{V} V(Z_s)\, \mathrm{d}s$$

is a local martingale and therefore $\widetilde{\mathcal{L}}$ defined in (2.5) coincides with the extended generator given in [11, Eq. (26.15)].

From the discussion in [11, p. 32], it is also clear that for $f \in \mathcal{D}(\widetilde{\mathcal{L}})$, the function $\widetilde{\mathcal{L}}f : \mathcal{Z} \to \mathbb{R}$ is uniquely defined everywhere except possibly on a set A of *zero potential*, that is

$$\int_0^\infty \mathbb{1}_A(Z_s)\,\mathrm{d}s = 0, \quad \mathbb{P}^z \text{ a.s.}, \quad \text{for all } z \in \mathcal{Z}.$$

For the proof of Theorem 3.2, $\nabla_x V(x, v)$ will not be well defined for the set $A := \{(x, v) \in \mathcal{Z} : |x| = 1\}$ which has zero potential, since the linear trajectories of BPS and the countable number of jumps imply it can intersect this set at most a countable number of times. The same argument also justifies Remark 1.

Finally, at points (x, v) where $\langle \nabla U(x), v\rangle \neq 0$, the gradient $\nabla_x V(x, v)$ exists and therefore we can use the more convenient expression (2.8), whereas we will use the original expression (2.5) whenever $\langle \nabla U(x), v\rangle = 0$.

5.2. *Proof of Theorem 3.1 and Theorem 3.2.* We have established that BPS satisfies all conditions of Theorem A and that the Lyapunov function defined in (3.1) belongs to the domain of the extended generator $\mathcal{D}(\widetilde{\mathcal{L}})$. The next result establishes the drift condition ($\mathfrak{D}$) for a constant refreshment rate and thus completes the proof of Theorem 3.1.

LEMMA 4 (Lyapunov function—Constant refreshment). *Let the refreshment rate be constant, i.e.,* $\Lambda_{\text{ref}}(\cdot) := \lambda_{\text{ref}}$. *The function V defined in (3.1) belongs to $\mathcal{D}(\widetilde{\mathcal{L}})$. If either of the conditions of Theorem 3.1 holds, V is a Lyapunov function as it satisfies ($\mathfrak{D}$).*

Next we establish the drift condition ($\mathfrak{D}$) for a location-dependent refreshment rate completing the proof of Theorem 3.2.

LEMMA 5 (Lyapunov function—Varying refreshment). *Let the refreshment rate $\Lambda_{\text{ref}}(\cdot)$ be given by* (3.5). *Then the function V defined in (3.1) belongs to $\mathcal{D}(\widetilde{\mathcal{L}})$. If in addition the assumptions of Theorem 3.2 hold, V is a Lyapunov function as it satisfies ($\mathfrak{D}$).*

The proofs are quite lengthy and technical and are thus given in the Supplementary Material [12].

5.3. *Proof of Theorem 3.3.* Next, we set the stage for the proof of Theorem 3.3. We will frequently use [26, Equations (11),(13)] which we state for the reader's convenience,

$$\nabla h(x) = \begin{cases} \dfrac{f(|x|)\mathbb{1}_d}{|x|} + \left[f'(|x|) - \dfrac{f(|x|)}{|x|}\right]\dfrac{xx^T}{|x|^2}, & x \neq 0, \\ f'(0)\mathbb{1}_d, & x = 0, \end{cases} \tag{5.1}$$

and

$$\det\left(\nabla h(x)\right) = \begin{cases} f'(|x|)\left(\dfrac{f(|x|)}{|x|}\right)^{d-1}, & x \neq 0, \\ f'(0)^d, & x = 0. \end{cases} \tag{5.2}$$

Let $\{Z_{h,t} = (Y_t, V_t); t \geq 0\}$ be a Markov process whose generator is given by (2.5) with U replaced by U_h, and write $\{P_h^t : t \geq 0\}$ for its transition kernels. Then letting $X_t := h(Y_t)$ for $t \geq 0$, from [7, Corollary 3], it follows that $\{Z_t = (X_t, V_t) : t \geq 0\}$ is also a Markov process with transition kernel given by $P^t(z, A) = P_h^t(H^{-1}(z), H^{-1}(A))$ for all $A \in \mathcal{B}(\mathcal{Z})$ where $H(x, v) = (h(x), v)$. It is also easy to see that if $Z_{h,t}$ is π_h-invariant, then Z_t will be π-invariant – see also the discussion in [26, Theorem 6].

Suppose now that $\{Z_{h,t} : t \geq 0\}$ is V_h-uniformly ergodic for some function V_h, that is

$$\|P_h^t(z, \cdot) - \pi_h\|_{V_h} \leq C_h V_h(z)\rho_h^t,$$

for some $C_h > 0$ and $\rho_h \in (0, 1)$ with π_h admitting the density $\bar{\pi}_h(y)\psi(v)$. Then we can see that

$$\int f \, \mathrm{d}P^t(z, \cdot) - \int f \, \mathrm{d}\pi = \int f \circ H \, \mathrm{d}P_h^t\left(H^{-1}(z), \cdot\right) - \int f \circ H \, \mathrm{d}\pi_h.$$

Therefore it follows that

$$\begin{aligned} &\sup_{|f| \leq V_h \circ H^{-1}} \left| \int f \, \mathrm{d}P^t(z, \cdot) - \int f \, \mathrm{d}\pi \right| \\ &= \sup_{|f| \leq V_h \circ H^{-1}} \left| \int f \circ H \, \mathrm{d}P_h^t\left(H^{-1}(z), \cdot\right) - \int f \circ H \, \mathrm{d}\pi_h \right| \\ &\leq \sup_{|g| \leq V_h} \left| \int g \, \mathrm{d}P_h^t\left(H^{-1}(z), \cdot\right) - \int g \, \mathrm{d}\pi_h \right| \\ &= \|P_h^t(H^{-1}(z), \cdot) - \pi_h\|_{V_h} \leq C_h V_h \circ H^{-1}(z)\rho_h^t, \end{aligned}$$

whence $Z_t = H(Z_{h,t})$ is $V_h \circ H^{-1}$-uniformly ergodic.

The proof of Theorem 3.3 then follows from the following two Lemmas the proofs of which are given in the Supplementary Material [12].

LEMMA 6. *Under the assumptions of Theorem 3.3, the potentials $U_h : \mathbb{R}^d \to [0,\infty)$ defined in (3.9) satisfy Assumptions (A0)-(A2), when $h = h^{(1)}$ or $h = h^{(2)}$.*

LEMMA 7. *The following results hold.*

(A) Under the assumptions of Theorem 3.3(A), the function $U_{h^{(1)}}$, defined through equations (3.6), (3.8) and (3.9), satisfies the conditions of Theorem 3.1(B) with $\alpha_2 := b(\mathfrak{d} - d)/2$.
(B) Under the assumptions of Theorem 3.3(B), the function $U_{h^{(2)}}$, defined through equations (3.7), (3.8) and (3.9), satisfies the conditions of Theorem 3.2.

5.4. *Proof of Theorem 3.4.* The proof of the CLT now follows from a standard result [20, Theorem 4.3].

PROOF OF THEOREM 3.4. Notice that if V satisfies ($\mathfrak{D}$) then for any $\varepsilon \in (0,1)$, by Jensen's inequality it follows that $\mathbb{E}^z\left[V^{1-\varepsilon}(Z_t)\right] \leq \mathbb{E}^z\left[V(Z_t)\right]^{1-\varepsilon}$. Since $\mathbb{E}^z\left[V^{\varepsilon}(Z_0)\right] = V(z)^{\varepsilon}$, it follows that

$$\begin{aligned}
\mathcal{L}V^{1-\varepsilon}(z) &= \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}^z\left[V^{1-\varepsilon}(Z_t)\right]\Big|_{t=0} \leq \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}^z\left[V(Z_t)\right]^{1-\varepsilon}\Big|_{t=0} \\
&= (1-\varepsilon)\frac{1}{\mathbb{E}^z[V(Z_t)]^{\varepsilon}}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}^z\left[V(Z_t)\right]\Big|_{t=0} \\
&= (1-\varepsilon)\frac{\mathcal{L}V(z)}{V(z)^{\varepsilon}} \leq -(1-\varepsilon)\delta\frac{V(z)}{V(z)^{\varepsilon}} + \frac{b\mathbb{1}_C(z)}{V(z)^{\varepsilon}},
\end{aligned}$$

and thus $W(z) := V(z)^{1-\varepsilon}$ also satisfies ($\mathfrak{D}$). An application of [20, Theorem 4.3] completes the proof. □

Acknowledgements. The authors are grateful to François Dufour for useful discussions and pointing them towards [8], to Pierre Del Moral for having brought their attention to reference [31] and the referees for helpful comments which have improved the presentation of the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Exponential Ergodicity of the Bouncy Particle Sampler"
(doi: COMPLETED BY THE TYPESETTER; .pdf). We provide detailed proofs of all results given in the main paper.

REFERENCES

[1] Bierkens, J., Bouchard-Côté, A., Doucet, A., Duncan, A.B., Fearnhead, P., Lienart, T., Roberts, G.O., and Vollmer, S.J. (2016). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *To appear in Statist. Probab. Lett.*, arXiv preprint arXiv:1701.04244.

[2] Bierkens, J., and Duncan, A. (2017). Limit theorems for the Zig-Zag process. *Adv. Appl. Probab.*, **49**:3, 791–825.

[3] Bierkens, J., Fearnhead, P., and Roberts, G.O. (2016). The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. arXiv preprint arXiv:1607.03188.

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

[5] Bierkens, J., Roberts, G.O., and Zitt, P.A. (2017). Ergodicity of the Zig-Zag process. arXiv preprint arXiv:1712.09875.

[6] Bouchard-Côté, A., Vollmer, S.J., and Doucet, A. (2015). The Bouncy Particle Sampler: a non-reversible rejection-free Markov chain Monte Carlo method. *To appear in J. Amer. Statist. Assoc.*, arXiv preprint arXiv:1510.02451.

[7] Burke, C.J. and Rosenblatt, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Statist.*, **29**:4, 1112–1122.

[8] Costa, O.L.V. (1990). Stationary distributions for piecewise-deterministic Markov processes. *J. Appl. Probab.*, **27**:1, 60–73.

[9] Costa, O.L.V., and Dufour, F. (2008). Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, **47**:2, 1053–1077.

[10] Davis, M.H.A. (1984). Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. Roy. Stat. Soc.* B, **46**:3, 353–388.

[11] Davis, M.H.A. (1993). *Markov Models and Optimization.* CRC Press.

[12] Deligiannidis, G., Bouchard-Côté, A. and Doucet, A. (2018). Supplement to "Exponential Ergodicity of the Bouncy Particle Sampler."

[13] Del Moral, P., and Penev, S. (2016). *Stochastic Processes: From Applications to Theory.* CRC Press.

[14] Devroye, L. (1986). *Non-uniform Random Variate Generation.* Springer-Verlag: New York.

[15] Down, D., Meyn, S.P., and Tweedie, R.L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23**:4, 1671–1691.

[16] Fearnhead, P., Bierkens, J., Pollock, M., and Roberts, G. O. (2016). Piecewise deterministic Markov processes for continuous-time Monte Carlo. arXiv preprint arXiv:1611.07873.

[17] Fétique, N. (2017). Long-time behaviour of generalised Zig-Zag process. arXiv preprint arXiv:1710.01087.

[18] Fontbona, J., Guérin, H. and Malrieu, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stoch. Proc. Appl.* **126**:10, 3077–3101.

[19] Gelman, A., Jakulin, A., Pitau, M.G. and Su, Y-S. (2008) A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Statist.*, **2**:4, 1360–1383.

[20] Glynn, P.W. and Meyn, S.P. (1996) A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, **24**:2, 916–931.

[21] Gramacy, R.B. and Polson, N.G. (2012) Simulation-based regularized logistic regression. *Bayesian Anal.*, **7**:3, 567–590.

[22] Isobe, M. and Krauth. W. (2015). Hard-sphere melting and crystallization with event-chain Monte Carlo. *J. Chem. Phys.*, **143**:8, 084509.

[23] Jarner, S.F., and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stoch. Proc. Appl.*, **85**:2, 341–361.

[24] Jarner, S. F., and Tweedie, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type. *Bernoulli*, **9**:4, 559–578.

[25] Jones, G.L., and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Stat. Science*, **16**:4, 312–334.

[26] Johnson, L.T., and Geyer, C.J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40**:6, 3050–3076.

[27] Kapfer, S.C. and Krauth, W. (2016). Cell-veto Monte Carlo algorithms for long-range systems. *Phys. Rev.* E, **94**:3 031302(R).

[28] Livingstone, S., Betancourt, M., Byrne, S. and Girolami, M. (2016). On the geometric ergodicity of Hamiltonian Monte Carlo. arXiv preprint arXiv:1601.08057.

[29] Michel, M., Kapfer, S.C. and Krauth, W. (2014). Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J. Chem. Phys.* **140**:5, 054116.

[30] Mesquita, A.R. and Hespanha, J.P. (2010). Construction of Lyapunov functions for piecewise-deterministic Markov processes. *Proc. 49th IEEE Conf. Decision and Control*, 2408–2413.

[31] Mesquita, A.R. and Hespanha, J.P. (2012). Jump control of probability densities with applications to autonomous vehicle motion. *IEEE Trans. Auto. Control*, **57**:10, 2588–2598.

[32] Meyn, S.P. and Tweedie, R.L. (1993). Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.*, **25**:3, 518–548.

[33] Monmarché, P. (2016). Piecewise deterministic simulated annealing. *ALEA, Lat. Am. J. Proba. Math. Stat.*, **13**:1, 357–398.

[34] Neal, R.M. (2003). Slice sampling (with discussion). *Ann. Statist.*, **31**:3, 705–767.

[35] Nishikawa, Y. and Hukushima, K. (2016). Event-chain Monte Carlo algorithm for continuous spin systems and its application. *J. Phys: Conf. Ser.*, **750**:012014.

[36] Pakman, A., Gilboa, D., Carlson, D., and Paninski, L. (2017). Stochastic Bouncy Particle Sampler. *Proc. 34th Int. Conf. Machine Learning*, 2741–2750.

[37] Peters, E.A.J.F. and de With, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev.* E, **85**:026703, 1671–1691.

[38] Roberts, G. and Rosenthal, J. (1996). Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.*, **1**:9, 1–21.

[39] Roberts, G.O. and Tweedie, R.L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, **2**:4, 341–363.

[40] Roberts, G. O., and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, **83**:1, 95–110.

[41] Sherlock, C. and Thiery, A. (2017). A discrete Bouncy Particle Sampler. arXiv preprint arXiv:1707.05200.

[42] Vanetti, P., Bouchard-Côté, A., Deligiannidis, G. and Doucet, A. (2017). Piecewise deterministic Markov chain Monte Carlo. arXiv preprint arXiv:1707.05296.

[43] Wu, C. and Robert, C.P. (2017). Generalized Bouncy Particle Sampler. arXiv preprint arXiv:1706.04781.

Department of Statistics,
University of Oxford,
24-29 St. Giles,
OX1 3LB, Oxford, UK
E-mail: deligian@stats.ox.ac.uk

Department of Statistics,
University of British Columbia,
2207 Main Mall,
V6T 1Z4, Vancouver, Canada
E-mail: bouchard@stat.ubc.ca

Department of Statistics,
University of Oxford,
24-29 St. Giles,
OX1 3LB, Oxford, UK
E-mail: doucet@stats.ox.ac.uk

Submitted to the Annals of Statistics

SUPPLEMENT TO "EXPONENTIAL ERGODICITY OF THE BOUNCY PARTICLE SAMPLER"

By George Deligiannidis*, Alexandre Bouchard-Côté† and Arnaud Doucet*

University of Oxford and University of British Columbia†*

1. Proofs of Auxiliary results of Section 4.

Proof of Lemma 1. We prove invariance using the approach developed in [1], see also [2], where a link is provided between the invariant measures of $\{Z_t : t \geq 0\}$ and those of the embedded discrete-time Markov chain $\{\Theta_k : k \geq 0\} := \{(X_{\tau_k}, V_{\tau_k}) : k \geq 0\}$. The Markov transition kernel of this chain is given for $A \times B \in \mathcal{B}(\mathcal{Z})$ by

$$\begin{aligned}&\mathcal{Q}\left((x,v), A\times B\right)\\&= \int_0^\infty \exp\{-\int_0^s \bar{\lambda}(x+uv,v)\,\mathrm{d}u\}\bar{\lambda}(x+sv,v)K\left((x+sv,v),A\times B\right)\,\mathrm{d}s,\end{aligned}$$

where K is defined in equation (2.7) of the main manuscript. We also define for $A \times B \in \mathcal{B}(\mathcal{Z})$ the measure

$$\begin{aligned}\mu(A\times B) &:= \int \bar{\lambda}(x,v)\pi(\,\mathrm{d}x,\,\mathrm{d}v)K\left((x,v),A\times B\right)\\&= \int_{A\times B}\left[\Lambda_{\mathrm{ref}}(x)+\lambda(x,R(x)v)\right]\pi(\,\mathrm{d}x,\,\mathrm{d}v)\\&= \int_{A\times B}\bar{\lambda}(x,-v)\pi(\,\mathrm{d}x,\,\mathrm{d}v),\end{aligned}$$

as $\lambda(x,R(x)v) = \lambda(x,-v)$. This measure is finite by the integrability condition (A1). We set $\xi := (\mu(\mathcal{Z}))^{-1}$ and $\bar{\mu} := \xi\mu$. The measure $\bar{\mu}$ satisfies $\bar{\mu} = \mathcal{T}\pi$, where $\mathcal{T}$ is operator defined in [1, Section 3.3] mapping invariant measures of $\{Z_t : t \geq 0\}$ to invariant measures of $\{\Theta_k : k \geq 0\}$. By [1, Theorem 3], $\mathcal{T}$ is invertible. Therefore, from [1, Theorem 2], it suffices to prove the result to show that μ is invariant for $\{\Theta_k\}$ which we now establish.

For continuous, bounded $f : \mathcal{Z} \to \mathbb{R}$ we have

$$\begin{aligned}&\zeta \iiiint \mu(\,\mathrm{d}x,\,\mathrm{d}v)\mathcal{Q}\left((x,v),\,\mathrm{d}y,\,\mathrm{d}w\right)f(y,w)\\&= \iint \mathrm{e}^{-U(x)}\,\mathrm{d}x\psi(\,\mathrm{d}v)\bar{\lambda}(x,-v)\end{aligned}$$

imsart-aos ver. 2014/10/16 file: finalversion_supp_submitted.tex date: May 3, 2018

$$
\begin{aligned}
&\times \int_0^\infty \exp\{-\int_0^s \bar\lambda(x+uv,v)\,\mathrm{d}u\}\bar\lambda(x+sv,v)Kf(x+sv,v)\,\mathrm{d}s\\
&= \int_{s=0}^\infty \mathrm{d}s \iint \mathrm{e}^{-U(x)}\,\mathrm{d}x\psi(\,\mathrm{d}v)\bar\lambda(x,-v)\\
&\quad\times \exp\{-\int_0^s \bar\lambda(x+uv,v)\,\mathrm{d}u\}\bar\lambda(x+sv,v)Kf(x+sv,v)
\end{aligned}
$$

and letting $z = x+sv$

$$
\begin{aligned}
&= \int_{s=0}^\infty \mathrm{d}s \iint \,\mathrm{d}z\psi(\,\mathrm{d}v)\mathrm{e}^{-U(z-sv)}\bar\lambda(z-sv,-v)\\
&\quad\times \exp\{-\int_0^s \bar\lambda(z+(u-s)v,v)\,\mathrm{d}u\}\bar\lambda(z,v)Kf(z,v)\\
&= \int_{s=0}^\infty \mathrm{d}s \iint \,\mathrm{d}z\psi(\,\mathrm{d}v)\bar\lambda(z-sv,-v)\\
&\quad\times \exp\{-U(z-sv) - \int_0^s \bar\lambda(z-wv,v)\,\mathrm{d}w\}\bar\lambda(z,v)Kf(z,v).
\end{aligned}
$$

Since $t\mapsto U(x+tv)$ is absolutely continuous, we can write

$$
\begin{aligned}
U(z) &= U(z-sv) + \int_{w=0}^s \langle \nabla U(z-wv), v\rangle\,\mathrm{d}w\\
&= U(z-sv) + \int_{w=0}^s \big[\max\{\langle \nabla U(z-wv), v\rangle, 0\}\\
&\qquad + \min\{\langle \nabla U(z-wv), v\rangle, 0\}\big]\,\mathrm{d}w,
\end{aligned}
$$

and it follows that

$$
\begin{aligned}
&U(z) + \int_{w=0}^s \max\{\langle \nabla U(z-wv), -v\rangle, 0\}\,\mathrm{d}w\\
&\quad= U(z-sv) + \int_{w=0}^s \max\{\langle \nabla U(z-wv), v\rangle, 0\}\,\mathrm{d}w.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\zeta \iiiint \mu(\,\mathrm{d}x,\,\mathrm{d}v)\mathcal{Q}\left((x,v),\,\mathrm{d}y,\,\mathrm{d}w\right)f(y,w)\\
&= \int_{s=0}^\infty \mathrm{d}s \iint \,\mathrm{d}z\psi(\,\mathrm{d}v)\bar\lambda(z-sv,-v)\\
&\quad\times \exp\{-U(z) - \int_0^s \bar\lambda(z-wv,-v)\,\mathrm{d}w\}\bar\lambda(z,v)Kf(z,v)\\
&= \iint \mathrm{e}^{-U(z)}\,\mathrm{d}z\psi(\,\mathrm{d}v)\bar\lambda(z,v)Kf(z,v)\\
&\quad\times \int_{s=0}^\infty \mathrm{d}s\bar\lambda(z-sv,-v)\exp\{-\int_0^s \bar\lambda(z-wv,-v)\,\mathrm{d}w\}
\end{aligned}
$$

$$= \iint \mathrm{e}^{-U(z)} \,\mathrm{d}z \psi(\,\mathrm{d}v) \bar{\lambda}(z,v) K f(z,v)$$
$$= \zeta \iint \pi(\,\mathrm{d}z,\,\mathrm{d}v) \bar{\lambda}(z,v) K f(z,v) = \zeta \iint \mu(\,\mathrm{d}z,\,\mathrm{d}v) f(z,v),$$

proving that μ is invariant for $\mathcal{Q}$. □

Proof of Lemma 2. The proof is inspired by [5]. Let $f : B(0, T/6) \times \mathbb{S}^{d-1} \to [0,\infty)$ be a bounded non-negative function. Let E be the event that exactly two events have occurred up to time T, and both of them are refreshments. Then

$$\begin{aligned}
&\mathbb{E}^z[f(Z_T)] \\
&\geq \mathbb{E}^z[f(Z_T)\mathbf{1}_E] \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{t=0}^{T} \int_{s=0}^{T-t} \,\mathrm{d}s \,\mathrm{d}t \bar{\lambda}(x_0 + tv_0, v_0) \\
&\quad \times \exp\left\{ - \int_{u=0}^{t} \bar{\lambda}(x_0 + uv_0, v_0) \,\mathrm{d}u \right\} \frac{\Lambda_{\mathrm{ref}}(x + tv_0)}{\bar{\lambda}(x + tv_0, v_0)} \bar{\lambda}(x_0 + tv_0 + sv_1, v_1) \\
&\quad \times \exp\left\{ - \int_{w=0}^{s} \bar{\lambda}(x_0 + v_0 t + wv_1, v_1) \,\mathrm{d}w \right\} \frac{\Lambda_{\mathrm{ref}}(x + tv_0 + sv_1)}{\bar{\lambda}(x + tv_0 + sv_1, v_1)} \\
&\quad \times \exp\left\{ - \int_{r=0}^{T-s-t} \bar{\lambda}(x_0 + v_0 t + sv_1 + rv_2, v_2) \,\mathrm{d}r \right\} \\
&\quad \times f\left(x_0 + tv_0 + sv_1 + (T-s-t)v_2, v_2\right) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{t=0}^{T} \int_{s=0}^{T-t} \,\mathrm{d}s \,\mathrm{d}t \; \Lambda_{\mathrm{ref}}(x + tv_0) \\
&\quad \times \exp\left\{ - \int_{u=0}^{t} \bar{\lambda}(x_0 + uv_0, v_0) \,\mathrm{d}u \right\} \Lambda_{\mathrm{ref}}(x + tv_0 + sv_1) \\
&\quad \times \exp\left\{ - \int_{w=0}^{s} \bar{\lambda}(x_0 + v_0 t + wv_1, v_1) \,\mathrm{d}w \right\} \\
&\quad \times \exp\left\{ - \int_{r=0}^{T-s-t} \bar{\lambda}(x_0 + v_0 t + sv_1 + rv_2, v_2) \,\mathrm{d}r \right\} \\
&\quad \times f\left(x_0 + tv_0 + sv_1 + (T-s-t)v_2, v_2\right).
\end{aligned}$$

Since the process moves at unit speed and $|x_0| \leq T/6$, it follows that $\sup_{t \leq T} |X_t| \leq 7T/6$. Let

$$K := \sup \left\{ \bar{\lambda}(x,v) : |x| \leq 7T/6,\; v \in \mathbb{S}^{d-1} \right\} < \infty,$$

and recall that $\bar{\lambda}(x,v) \geq \lambda_{\mathrm{ref}} > 0$. Therefore

$$
\begin{aligned}
&\mathbb{E}^z[f(Z_T)] \\
&\geq \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{s=0}^{T} \int_{t=0}^{T-s} \mathrm{d}s\,\mathrm{d}t\ \lambda_{\mathrm{ref}}\lambda_{\mathrm{ref}} \\
&\qquad \times \exp\left\{-\int_{u=0}^{t} K\,\mathrm{d}u - \int_{w=0}^{s} K\,\mathrm{d}w - \int_{r=0}^{T-s-t} K\,\mathrm{d}r\right\} \\
&\qquad \times f\left(x_0 + sv_0 + tv_1 + (T-s-t)v_2, v_2\right) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{s=0}^{T} \int_{t=0}^{T-s} \mathrm{d}s\,\mathrm{d}t\lambda_{\mathrm{ref}}^2 \\
&\quad \times \exp\left\{-Kt - Ks - K(T-s-t)\right\} f\left(x_0 + sv_0 + tv_1 + (T-s-t)v_2, v_2\right) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{s=0}^{T} \mathrm{d}s \int_{t=s}^{T} \mathrm{d}t\lambda_{\mathrm{ref}}^2 \exp\left\{-KT\right\} \\
&\qquad \times f\left(x_0 + sv_0 + (t-s)v_1 + (T-t)v_2, v_2\right) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{t=0}^{T} \mathrm{d}t \\
&\qquad \times \int_{s=0}^{t} \mathrm{d}s\lambda_{\mathrm{ref}}^2 \exp\left\{-KT\right\} f\left(x_0 + sv_0 + (t-s)v_1 + (T-t)v_2, v_2\right), \\
&\geq \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{t=5T/6}^{T} \mathrm{d}t \\
&\qquad \times \int_{s=0}^{t} \mathrm{d}s\lambda_{\mathrm{ref}}^2 \exp\left\{-KT\right\} f\left(x_0 + sv_0 + (t-s)v_1 + (T-t)v_2, v_2\right) \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \int_{t=5T/6}^{T} \mathrm{d}t\ t \\
&\qquad \times \int_{r=0}^{1} \mathrm{d}r\lambda_{\mathrm{ref}}^2 \exp\left\{-KT\right\} f\left(x_0 + rtv_0 + (t-rt)v_1 + (T-t)v_2, v_2\right) \\
&\geq \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\psi(\,\mathrm{d}v_2) \frac{5T}{6} \int_{t=5T/6}^{T} \mathrm{d}t \\
&\qquad \times \int_{r=0}^{1} \mathrm{d}r\lambda_{\mathrm{ref}}^2 \exp\left\{-KT\right\} f\left(x_0 + trv_0 + t(1-r)v_1 + (T-t)v_2, v_2\right).
\end{aligned}
$$

Fix $t \in (5T/6, T]$ and $v_2 \in \mathbb{S}^{d-1}$ so that $x' := x_0 + (T-t)v_2$ is now fixed. Since $T \geq t > 5T/6$ it follows that $T - t < T/6$. Since also $|x_0| \leq T/6$ we must have that $|x'| \leq T/3$. Let $x'' \in B(0, T/6)$ be arbitrary. Then it follows that $|x' - x''| \leq T/2$.

We will now show that there exist $v_* \in \mathbb{S}^{d-1}$ and $r_* \in [0,1]$ such that

$$x' + tr_*v_0 + t(1-r_*)v_* = x''.$$

In the trivial case where $x'' = x' + trv_0$ for some $r \in [0,1]$, then since $t > |x'' - x'|$ it must hold that $r < 1$ and thus the representation is trivially satisfied with $v_* = -v_0$ and $r_* = r + (1-r)/2$.

If this is not the case, then for any $r \in [0,1]$ define $l(r) := rt + |x'' - x' - rtv_0| > rt$ and consider the path $f_r : [0, l(r)] \mapsto \mathbb{R}^d$ given by

$$f_r(s) = x' + sv_0, \text{ for } s \in [0, rt],$$
$$f_r(s) = x' + rtv_0 + (s - rt)\frac{x'' - x' - rtv_0}{|x'' - x' - rtv_0|}, \text{ for } s > rt.$$

For each $r \in [0,1]$ the path starts at x', ends at x'' and has length given by $l(r)$. By definition we have $l(0) = |x'' - x'|$ and since $t > |x'' - x'|$ it follows from the triangle inequality that $|x'' - x' - tv_0| \geq t - |x'' - x'| > 0$ and thus that $l(1) > t$. Since $l(\cdot)$ is continuous, by the intermediate value theorem there exists $r_* \in [0,1]$ such that $l(r_*) = t$, whence the definition of $l(\cdot)$ implies that

$$|x'' - x' - r_* tv_0| = (1 - r_*)t.$$

Letting

$$v_* := \frac{x'' - x' - r_* tv_0}{|x'' - x' - r_* tv_0|} \in \mathbb{S}^{d-1},$$

we have

$$x'' = x' + r_* tv_0 + t(1 - r_*)v_*.$$

To proceed we define the measure $\mu = \mu_{t,x_0,v_0}$ on $\mathbb{R}^d$ as

$$\mu(V) := \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1) \int_{r=0}^{1} \mathrm{d}r\, \mathbb{1}_V\left(x' + rtv_0 + t(1-r)v_1\right),$$

for Borel sets $V \subset \mathbb{R}^d$. It is obvious that μ is absolutely continuous with respect to d-dimensional Lebesgue measure. Letting $R \sim U[0,1]$ and $V \sim \psi$ be independent, we have for any $\delta > 0$ and $x'' \in B(0, T/6)$

$$\begin{aligned}
\mu\left(B(x'', \delta)\right) &= \mathbb{P}\left\{|x' + tRv_0 + t(1-R)V - x''| \leq \delta\right\} \\
&= \mathbb{P}\left\{|tRv_0 + t(1-R)V - tr_* v_0 - t(1-r_*)v_*| \leq \delta\right\} \\
&= \mathbb{P}\left\{|t(R - r_*)v_0 + t(V - v_*) - t(RV - r_* v_*)| \leq \delta\right\} \\
&= \mathbb{P}\left\{|t(R - r_*)v_0 + t(V - v_*) - t(RV - Rv_* + Rv_* - r_* v_*)| \leq \delta\right\} \\
&= \mathbb{P}\left\{|t(R - r_*)v_0 + t(V - v_*) - tR(V - v_*) - t(R - r_*)v_*)| \leq \delta\right\} \\
&\geq \mathbb{P}\left\{T|R - r_*| + T|V - v_*| + T|V - v_*| + T|R - r_*| \leq \delta\right\} \\
&= \mathbb{P}\left\{|R - r_*| + |V - v_*| \leq \frac{\delta}{2T}\right\}
\end{aligned}$$

$$\begin{aligned}
&\geq \mathbb{P}\left\{\max\{|R-r_*|,|V-v_*|\}\leq \frac{\delta}{4T}\right\}\\
&= \mathbb{P}\left\{|R-r_*|\leq \frac{\delta}{4T}\right\}\mathbb{P}\left\{|V-v_*|\leq \frac{\delta}{4T}\right\}\\
&\geq \frac{\delta}{4T}\times\Big(C_1\Big(\frac{\delta}{4T}\Big)\Big)^{d-1} = C_2(T)\delta^d,
\end{aligned}$$

where $C_1 > 0$ is a constant, and where $C_i(\cdot)$ denotes quantities depending only on the variables in the bracket. Therefore, for all $t > 5T/6$ and $v_2 \in \mathbb{S}^{d-1}$, we can bound the density of μ with respect to Lebesgue measure, on the set $B(0,T/6)$, from below by a constant $C_3(T,d) > 0$

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}v_1)\int_{r=0}^{1} \mathrm{d}r f\left(x_0 + rtv_0 + t(1-r)v_1 + (T-t)v_2, v_2\right)\\
&\geq \int \mu(\,\mathrm{d}x'')f(x'',v_2) \geq C_3(T,d)\int_{B(0,T/6)} f\left(x'',v_2\right)\,\mathrm{d}x'',
\end{aligned}$$

and thus

$$\begin{aligned}
\mathbb{E}^z[f(Z_T)] &\geq C_4(T,d,\lambda_{\mathrm{ref}})\int_{\mathbb{S}^{d-1}}\psi(\,\mathrm{d}v_2)\int_{t=5T/6}^{T}\mathrm{d}t\int_{x''\in B(0,T/6)} f\left(x'',v_2\right)\,\mathrm{d}x''\\
&\geq C_5(T,d,\lambda_{\mathrm{ref}})\int_{\mathbb{S}^{d-1}}\int_{x''\in B(0,T/6)} f\left(x'',v\right)\,\mathrm{d}x''\psi(\,\mathrm{d}v).
\end{aligned}$$

Since f is generic, we conclude that for all $z = (x'', v) \in B(0, T/6) \times \mathbb{S}^{d-1}$, and any Borel set $A \subseteq B(0,T/6)\times\mathbb{S}^{d-1}$

$$\mathbb{P}^z(Z_T \in A) \geq C_5(T,d,\lambda_{\mathrm{ref}})\iint_A \psi(\,\mathrm{d}v)\,\mathrm{d}x,$$

whence it follows that for any $R > 0$ the set $B(0,R)\times\mathbb{S}^{d-1}$ is petite.

Given any compact set $U \subset \mathbb{R}^d \times \mathbb{S}^{d-1}$, we can find $R > 0$ such that $U \subset B(0,R)\times\mathbb{S}^{d-1}$, and we can easily conclude using the above that U must also be petite.

Finally let $A \subseteq \mathcal{Z}$ such that $\iint_A \psi(\,\mathrm{d}v)\,\mathrm{d}x > 0$. We can find $R > 0$ such that the set $A' := A \cap (B(0,R)\times\mathbb{S}^{d-1})$ satisfies $\iint_{A'}\psi(\,\mathrm{d}v)\,\mathrm{d}x > 0$. Let $z = (x,v) \in \mathcal{Z}$ be arbitrary and for some fixed $\epsilon > 0$ define $T := \max\{6|x|+\epsilon, 6R+\epsilon\}$. Then $z \in B(0,T/6)\times\mathbb{S}^{d-1}$, $A' \subseteq B(0,T/6)\times\mathbb{S}^{d-1}$ and thus by the first part of the lemma

$$\mathbb{P}^z(Z_T \in A') \geq C_5(T,d,\lambda_{\mathrm{ref}})\iint_A \psi(\,\mathrm{d}v)\,\mathrm{d}x > 0.$$

Therefore since $A' \subseteq A$, writing $\tau_B := \inf\{t \geq 0 : Z_t \in B\}$ for a measurable set $B \subseteq \mathcal{Z}$, we have

$$\mathbb{P}^z\left\{\tau_A < \infty\right\} \geq \mathbb{P}^z\left\{\tau_{A'} < \infty\right\} \geq \mathbb{P}^z\left\{Z_T \in A'\right\} > 0.$$

Irreducibility then follows from [4, Proposition 2.1]. □

2. Proofs of results of Section 5.

Proof of Lemma 4. That $V \in \mathcal{D}(\widetilde{\mathcal{L}})$ follows from the discussion in Section 5.1. We now establish that V is a Lyapunov function. First we compute $\widetilde{\mathcal{L}}V(x,v)$. Notice that if $\langle \nabla U(x), v\rangle \neq 0$, then by continuity there will be a neighbourhood of (x,v) on which $V(x,v)$ will be differentiable. Therefore at those points we will use equation (2.8) of the main manuscript.

Case $\langle \nabla U(x), v\rangle > 0$. We have

$$\langle \nabla V(x,v), v\rangle = \frac{1}{2}V(x,v)\langle \nabla U(x), v\rangle,$$

and adding the reflection part we obtain

$$\begin{aligned}
&\langle \nabla V(x,v), v\rangle + \langle \nabla U(x), v\rangle \left[V(x,R_x v) - V(x,v)\right]\\
&= \frac{1}{2}V(x,v)\langle \nabla U(x), v\rangle\\
&\qquad + \langle \nabla U(x), v\rangle \left[\frac{e^{U(x)/2}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle_+}}\frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle_+}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), v\rangle}} - V(x,v)\right]\\
&= -\frac{1}{2}V(x,v)\langle \nabla U(x), v\rangle + \langle \nabla U(x), v\rangle V(x,v)\frac{\sqrt{\lambda_{\mathrm{ref}}}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), v\rangle}}.
\end{aligned}$$

The refreshment term is given by

$$\begin{aligned}
&e^{U(x)/2}\lambda_{\mathrm{ref}}\int \psi(\,\mathrm{d}w)\left[\frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle_+}} - \frac{1}{\sqrt{\lambda_{\mathrm{ref}}}}\right]\\
&= e^{U(x)/2}\lambda_{\mathrm{ref}}\int \psi(\,\mathrm{d}w)\frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle_+}} - \lambda_{\mathrm{ref}}V(x,v)\\
&= e^{U(x)/2}\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x), w\rangle > 0} \psi(\,\mathrm{d}w)\frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle_+}}\\
&\qquad + e^{U(x)/2}\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x), w\rangle \leq 0} \psi(\,\mathrm{d}w)\frac{1}{\sqrt{\lambda_{\mathrm{ref}}}} - \lambda_{\mathrm{ref}}V(x,v)\\
&= e^{U(x)/2}\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x), w\rangle > 0} \psi(\,\mathrm{d}w)\frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle_+}} - \frac{1}{2}\lambda_{\mathrm{ref}}V(x,v),
\end{aligned}$$

since $\psi\{w : \langle \nabla U(x), w\rangle > 0\} = 1/2$. Thus overall when $\langle \nabla U(x), v\rangle > 0$ we have

$$\widetilde{\mathcal{L}}V(x,v) = \frac{1}{2}V(x,v)\langle \nabla U, v\rangle + \langle \nabla U(x), v\rangle \left[V(x,R_x v) - V(x,v)\right]$$

$$
\begin{aligned}
&+ \mathrm{e}^{U(x)/2}\lambda_{\mathrm{ref}} \int_{\mathbb{S}^{d-1}} \psi(\,\mathrm{d}w) \left[\frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle_+}} - \frac{1}{\sqrt{\lambda_{\mathrm{ref}}}}\right] \\
&= -\frac{1}{2}V(x,v)\Bigg[\langle \nabla U(x), v\rangle - 2\frac{\langle \nabla U(x), v\rangle\sqrt{\lambda_{\mathrm{ref}}}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), v\rangle}} + \lambda_{\mathrm{ref}} \\
&\qquad - 2\lambda_{\mathrm{ref}}^{3/2} \int_{\langle \nabla U(x), w\rangle \geq 0} \frac{\psi(\,\mathrm{d}w)}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle}}\Bigg] \\
&= -\frac{1}{2}V(x,v)\Bigg[\langle \nabla U(x), v\rangle - 2\frac{\langle \nabla U(x), v\rangle\sqrt{\lambda_{\mathrm{ref}}}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), v\rangle}} + \lambda_{\mathrm{ref}} \\
&\qquad - 2\lambda_{\mathrm{ref}} \int_{\langle \nabla U(x), w\rangle \geq 0} \frac{\psi(\,\mathrm{d}w)}{\sqrt{1 + \frac{\langle \nabla U(x), w\rangle}{\lambda_{\mathrm{ref}}}}}\Bigg].
\end{aligned}
$$

Case $\langle \nabla U(x), v\rangle < 0$. In this case

$$
\langle \nabla V(x,v), v\rangle = -\frac{1}{2}V(x,v)\left[\langle \nabla U, -v\rangle - \frac{1}{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle\right].
$$

Since $\langle \nabla U(x), v\rangle_+ = 0$ there is no reflection and thus overall

$$
\begin{aligned}
\widetilde{\mathcal{L}}V(x,v) &= -\frac{1}{2}V(x,v)\left[\langle \nabla U(x), -v\rangle - \frac{1}{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle\right] \\
&\quad - \lambda_{\mathrm{ref}} V(x,v) + \frac{1}{2}\lambda_{\mathrm{ref}}\frac{\mathrm{e}^{U(x)/2}}{\sqrt{\lambda_{\mathrm{ref}}}} \\
&\quad + \lambda_{\mathrm{ref}} V(x,v) \int_{\langle \nabla U(x), w\rangle \geq 0} \psi(\,\mathrm{d}w)\frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle}} \\
&= -\frac{1}{2}V(x,v)\left[\langle \nabla U(x), -v\rangle - \frac{1}{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle + 2\lambda_{\mathrm{ref}}\right] \\
&\quad + \frac{1}{2}\lambda_{\mathrm{ref}} V(x,v)\frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\mathrm{ref}}}} + \\
&\quad + \lambda_{\mathrm{ref}} V(x,v) \int_{\langle \nabla U(x), w\rangle \geq 0} \psi(\,\mathrm{d}w)\frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle}} \\
&= -\frac{1}{2}V(x,v)\Bigg[\langle \nabla U(x), -v\rangle + 2\lambda_{\mathrm{ref}} - \frac{1}{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle \\
&\quad - \lambda_{\mathrm{ref}}\frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\mathrm{ref}}}} \\
&\qquad - 2\lambda_{\mathrm{ref}} \int_{\langle \nabla U(x), w\rangle \geq 0} \psi(\,\mathrm{d}w)\frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle}}\Bigg].
\end{aligned}
$$

Case $\langle \nabla U(x), v\rangle = 0$. In this case we compute $\widetilde{\mathcal{L}}V(x,v)$ as

(2.1)
$$\widetilde{\mathcal{L}}V(x,v) = \frac{\mathrm{d}}{\mathrm{d}t}V(x+tv,v)\Big|_{t=0+} + \lambda_{\mathrm{ref}}\left[\int \psi(\,\mathrm{d}w)V(x,w) - V(x,v)\right],$$

since the reflection term vanishes.

We first compute the directional derivative for which we can distinguish two cases. Suppose first that $\langle \Delta U(x)v, -v\rangle > 0$. Then we have that for all $t > 0$ small enough

$$\langle \nabla U(x+tv), -v\rangle = 0 + t\langle \Delta U(x)v, -v\rangle + o(t) \geq 0.$$

Therefore, since $\langle \nabla U(x), v\rangle = 0$, in this case we can compute the first term of (2.1) as follows

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}V(x+tv,v)\Big|_{t=0+} \\
&= \lim_{t\to 0+}\frac{1}{t}\left[\frac{\exp(U(x+tv)/2)}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x+tv), -v\rangle_+}} - \frac{\exp(U(x)/2)}{\sqrt{\lambda_{\mathrm{ref}}}}\right] \\
&= \lim_{t\to 0+}\frac{1}{t}\left[\frac{\exp(U(x+tv)/2)}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x+tv), -v\rangle}} - \frac{\exp(U(x)/2)}{\sqrt{\lambda_{\mathrm{ref}}}}\right] \\
&= \lim_{t\to 0+}\frac{1}{t}\Big[\frac{\exp(U(x+tv)/2) - \exp(U(x)/2)}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x+tv), -v\rangle}} \\
&\qquad + \exp(U(x)/2)\left(\frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x+tv), -v\rangle}} - \frac{1}{\sqrt{\lambda_{\mathrm{ref}}}}\right)\Big] \\
&= 0 - \frac{1}{2}\frac{\exp(U(x)/2)}{\sqrt{\lambda_{\mathrm{ref}}}^3}\langle \Delta U(x)v, -v\rangle = -\frac{1}{2}V(x,v)\frac{\langle \Delta U(x)v, -v\rangle}{\lambda_{\mathrm{ref}}}
\end{aligned}$$

Now consider the case where $\langle \Delta U(x)v, -v\rangle \leq 0$, then for all $t > 0$ small enough

$$\langle \nabla U(x+tv), -v\rangle = 0 + t\langle \Delta U(x)v, -v\rangle + o(t) \leq 0,$$

and therefore

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}V(x+tv,v)\Big|_{t=0+} \\
&= \lim_{t\to 0+}\frac{1}{t}\left[\frac{\exp(U(x+tv)/2)}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x+tv), -v\rangle_+}} - \frac{\exp(U(x)/2)}{\sqrt{\lambda_{\mathrm{ref}}}}\right] \\
&= \lim_{t\to 0+}\frac{1}{t}\left[\frac{\exp(U(x+tv)/2)}{\sqrt{\lambda_{\mathrm{ref}} + 0}} - \frac{\exp(U(x)/2)}{\sqrt{\lambda_{\mathrm{ref}}}}\right] = 0.
\end{aligned}$$

Overall we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}V(x+tv,v)\Big|_{t=0+} = -\frac{1}{2}V(x,v)\langle \Delta U(x)v,-v\rangle_+.$$

Adding the refreshment term we find that in this case

$$\widetilde{\mathcal{L}}V(x,v) = -\frac{1}{2}V(x,v)\Bigg[\frac{\langle \Delta U(x)v,-v\rangle_+}{\lambda_{\mathrm{ref}}} + \lambda_{\mathrm{ref}} - 2\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x),w\rangle\geq 0}\frac{\psi(\,\mathrm{d}w)}{\sqrt{1+\frac{\langle \nabla U(x),w\rangle}{\lambda_{\mathrm{ref}}}}}\Bigg].$$

Combining the three cases we obtain

(2.2)

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} = \begin{cases} -\Big[\frac{\langle \Delta U(x)v,-v\rangle_+}{\lambda_{\mathrm{ref}}} + \lambda_{\mathrm{ref}} \\ \quad -2\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x),w\rangle\geq 0}\frac{\sqrt{\lambda_{\mathrm{ref}}}\psi(\,\mathrm{d}w)}{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),w\rangle}}\Big] & \langle \nabla U(x),v\rangle = 0, \\ -\Big[\langle \nabla U(x),v\rangle - 2\frac{\langle \nabla U(x),v\rangle\sqrt{\lambda_{\mathrm{ref}}}}{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),v\rangle}} + \lambda_{\mathrm{ref}} \\ \quad -2\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x),w\rangle\geq 0}\frac{\sqrt{\lambda_{\mathrm{ref}}}\psi(\,\mathrm{d}w)}{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),w\rangle}}\Big], & \langle \nabla U(x),v\rangle > 0, \\ -\Big[\langle \nabla U(x),-v\rangle + 2\lambda_{\mathrm{ref}} - \frac{1}{\lambda_{\mathrm{ref}}+\langle \nabla U(x),-v\rangle}\langle v,\Delta U(x)v\rangle \\ \quad -\lambda_{\mathrm{ref}}\frac{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),-v\rangle}}{\sqrt{\lambda_{\mathrm{ref}}}} \\ \quad -2\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x),w\rangle\geq 0}\psi(\,\mathrm{d}w)\frac{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),-v\rangle}}{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),w\rangle}}\Big], & \langle \nabla U(x),v\rangle < 0. \end{cases}$$

*Condition (*A*).* We have that $\overline{\lim}_{|x|\to\infty}\|\Delta U(x)\| \leq \alpha_1$ and $\underline{\lim}_{|x|\to\infty}|\nabla U(x)| = \infty$. Since $\lambda_{\mathrm{ref}} > 0$ and $1/\sqrt{\cos(\theta)} \in L^1([0,\pi/2],\,\mathrm{d}\theta)$, for any $\epsilon > 0$ we can find $K > 0$ such that for all $|x| > K$

$$\int_{\langle \nabla U(x),w\rangle\geq 0}\frac{\psi(\,\mathrm{d}w)}{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),w\rangle}} \leq \int_{\theta=0}^{\pi/2}\frac{p_\vartheta(\theta)\,\mathrm{d}\theta}{\sqrt{|\nabla U(x)|\cos(\theta)}} \leq \frac{\epsilon}{\sqrt{\lambda_{\mathrm{ref}}}}, \tag{2.3}$$

Case $\langle \nabla U(x),v\rangle = 0$. Suppose that $|x| > K$. Then from (2.2), by dropping the first term which is negative,

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \leq -\left[\lambda_{\mathrm{ref}} - 2\lambda_{\mathrm{ref}}\int_{\langle \nabla U(x),w\rangle\geq 0}\frac{\sqrt{\lambda_{\mathrm{ref}}}\psi(\,\mathrm{d}w)}{\sqrt{\lambda_{\mathrm{ref}}+\langle \nabla U(x),w\rangle}}\right] \leq -\lambda_{\mathrm{ref}}(1-2\epsilon).$$

Case $\langle \nabla U(x), v\rangle > 0$. Again let $|x| > K$. From (2.2)

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \leq -\left[\langle \nabla U(x), v\rangle - 2\frac{\langle \nabla U(x), v\rangle\sqrt{\lambda_{\text{ref}}}}{\sqrt{\lambda_{\text{ref}} + \langle \nabla U(x), v\rangle}} + \lambda_{\text{ref}}(1-2\epsilon)\right].$$

For $w > 0$ consider the function

$$w - 2\frac{w\sqrt{\lambda_{\text{ref}}}}{\sqrt{\lambda_{\text{ref}} + w}} + \lambda_{\text{ref}}(1-2\epsilon).$$

Since for $w > 0$, $\lambda_{\text{ref}} + w \geq 2\sqrt{w}\sqrt{\lambda_{\text{ref}}}$ we have that

$$w - 2\frac{w\sqrt{\lambda_{\text{ref}}}}{\sqrt{\lambda_{\text{ref}} + w}} + \lambda_{\text{ref}}(1-2\epsilon) \geq w - 2\frac{w\sqrt{\lambda_{\text{ref}}}}{\sqrt{2}\lambda_{\text{ref}}^{1/4}w^{1/4}} + \lambda_{\text{ref}}(1-2\epsilon)$$

$$= w - \sqrt{2}w^{3/4}\lambda_{\text{ref}}^{1/4} + \lambda_{\text{ref}}(1-2\epsilon) =: f_\epsilon(w) =: f(w). \tag{2.4}$$

Then

$$f'(w) = 1 - \frac{3\lambda_{\text{ref}}^{1/4}}{2\sqrt{2}w^{1/4}},$$

and thus f is minimised at $w_* = 81\lambda_{\text{ref}}/64$ and

$$f(w_*) = \left(\frac{37}{64} - 2\epsilon\right)\lambda_{\text{ref}}.$$

For any $\lambda_{\text{ref}} > 0$ we can choose $\epsilon > 0$ small enough so that $f(w_*) > 0$. From (2.3) we can choose K large enough, so that for all $|x| > K$ and all v such that $\langle \nabla U(x), v\rangle > 0$

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} < -\delta,$$

for some $\delta = f(w_*) > 0$.

Case $\langle \nabla U(x), v\rangle < 0$. Then from (2.2)

$$\begin{aligned} 2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} = -\Bigg[&\langle \nabla U(x), -v\rangle + 2\lambda_{\text{ref}} - \frac{1}{\lambda_{\text{ref}} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle \\ &- \lambda_{\text{ref}}\frac{\sqrt{\lambda_{\text{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\text{ref}}}} \\ &- 2\lambda_{\text{ref}}\int_{\langle \nabla U(x), w\rangle \geq 0}\psi(\,\mathrm{d}w)\frac{\sqrt{\lambda_{\text{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\text{ref}} + \langle \nabla U(x), w\rangle}}\Bigg], \end{aligned}$$

and arguing in the same way as in the previous case, given $\epsilon > 0$ we can choose $K > 0$ such that for all $|x| > K$ we have similarly to (2.3)

$$\int_{\langle \nabla U(x), w\rangle \geq 0} \psi(\,\mathrm{d}w) \frac{1}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle}} \leq \int_{\langle \nabla U(x), w\rangle \geq 0} \psi(\,\mathrm{d}w) \frac{1}{\sqrt{\langle \nabla U(x), w\rangle}}$$
$$= \int_{\theta=0}^{\pi/2} \frac{p_\vartheta\,(\theta)\,\mathrm{d}\theta}{\sqrt{|\nabla U(x)|\cos(\theta)}} \leq \frac{\epsilon}{\lambda_{\mathrm{ref}}}.$$

Since $\overline{\lim}_{|x|\to\infty} \|\Delta U(x)\| \leq \alpha_1$, for K large enough and $|x| > K$ we have $\|\Delta U(x)\| \leq 2\alpha_1$. Thus overall when $\langle \nabla U(x), v\rangle < 0$

$$\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \leq -\frac{1}{2}\Bigg[\langle \nabla U(x), -v\rangle + 2\lambda_{\mathrm{ref}} - \frac{2}{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}\alpha_1$$
$$- \lambda_{\mathrm{ref}} \frac{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\mathrm{ref}}}} - 2\epsilon\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), -v\rangle}\Bigg].$$

For $w = \langle \nabla U(x), -v\rangle > 0$ define

$$g(w) := w + 2\lambda_{\mathrm{ref}} - \frac{2\alpha_1}{\lambda_{\mathrm{ref}} + w} - \lambda_{\mathrm{ref}} \frac{\sqrt{\lambda_{\mathrm{ref}} + w}}{\sqrt{\lambda_{\mathrm{ref}}}} - 2\epsilon\sqrt{\lambda_{\mathrm{ref}} + w},$$
$$g'(w) = 1 + \frac{2\alpha_1}{(\lambda_{\mathrm{ref}} + w)^2} - \frac{\sqrt{\lambda_{\mathrm{ref}}}}{2\sqrt{\lambda_{\mathrm{ref}} + w}} - \frac{\epsilon}{\sqrt{\lambda_{\mathrm{ref}} + w}}$$
$$\geq 1 + \frac{2\alpha_1}{(w + \lambda_{\mathrm{ref}})^2} - \frac{1}{2} - \frac{\epsilon}{\sqrt{\lambda_{\mathrm{ref}}}} = \frac{1}{2} + \frac{2\alpha_1}{(w + \lambda_{\mathrm{ref}})^2} - \frac{\epsilon}{\sqrt{\lambda_{\mathrm{ref}}}},$$

and thus for all λ_{ref} we can choose ϵ small enough so that $g'(w) \geq 0$ for all $w \geq 0$. Therefore

$$g(w) \geq g(0) = \lambda_{\mathrm{ref}} - 2\epsilon\sqrt{\lambda_{\mathrm{ref}}} - \frac{2\alpha_1}{\lambda_{\mathrm{ref}}}.$$

If $\lambda_{\mathrm{ref}} \geq (2\alpha_1 + 1)^2$ then for ϵ small enough we have that $g(w) \geq \delta > 0$, for some δ.

Thus, there exists $K > 0$ large enough so that for all $|x| > K$ and v such that $\langle \nabla U(x), v\rangle < 0$ we have $\widetilde{\mathcal{L}}V(x,v)/V \leq -\delta$. Therefore ($\mathfrak{D}$) holds with $C = B(0,K) \times \mathbb{S}^{d-1}$.

Condition (B). Recall that $2\alpha_2 := \underline{\lim}_{|x|\to\infty} |\nabla U(x)|$, so that we can choose K large enough so that for all $|x| > K$ we have $|\nabla U(x)| \geq \alpha_2$. Thus when $|x| > K$

$$\int_{\langle \nabla U(x), w\rangle \geq 0} \frac{\psi(\,\mathrm{d}w)}{\sqrt{\lambda_{\mathrm{ref}} + \langle \nabla U(x), w\rangle}} = \int_{\theta=0}^{\pi/2} \frac{p_\vartheta(\,\mathrm{d}\theta)}{\sqrt{\lambda_{\mathrm{ref}} + |\nabla U(x)|\cos(\theta)}}$$

$$\leq \frac{1}{\sqrt{\lambda_{\text{ref}}}} \int_{\theta=0}^{\pi/2} \frac{p_\vartheta(\,\mathrm{d}\theta)}{\sqrt{1+\frac{\alpha_2}{\lambda_{\text{ref}}}\cos(\theta)}}$$
$$= \frac{1}{\sqrt{\lambda_{\text{ref}}}} F\left(\frac{\alpha_2}{\lambda_{\text{ref}}}, d\right),$$

where

$$F\,(u,d) := \mathbb{E}\left[\frac{\mathbb{I}\,(\vartheta \in [0,\pi/2])}{\sqrt{1+u\cos\vartheta}}\right], \quad \vartheta \sim p_\vartheta\,(\cdot)\,,$$

with p_ϑ as defined in equation (2.2) of the main manuscript. Clearly $F(u,d) \leq F(0,d) = 1/2$ for all u and $F(u,d) \to 0$ as $u \to \infty$ for all d. Thus we can choose a sequence $\{c_d\}_{d\geq 0}$ such that $F(c_d,d) \leq 1/4$. One such choice given in the statement of the Theorem is $c_d = 16\sqrt{d}$. Indeed as $d \to \infty$,

$$F(c_d,d) = \kappa_d \int_{\theta=0}^{\pi/2} \frac{(\sin\theta)^{d-2}\,\mathrm{d}\theta}{\sqrt{1+c_d\cos\theta}}$$
$$\leq \frac{1}{4d^{1/4}} \int_{\theta=0}^{\pi/2} \frac{\kappa_d(\sin\theta)^{d-2}\,\mathrm{d}\theta}{\sqrt{\cos\theta}}$$
$$\leq \frac{\kappa_d}{4d^{1/4}} \int_{\theta=0}^{\pi/2} (\sin\theta)^{d-2}(\cos\theta)^{-1/2}\,\mathrm{d}\theta = \frac{1}{8}\frac{1}{d^{1/4}}\frac{\mathrm{Beta}\left(\frac{d-1}{2},\frac{1}{4}\right)}{\mathrm{Beta}\left(\frac{d-1}{2},\frac{1}{2}\right)} < \frac{1}{4}.$$

Case $\langle \nabla U(x), v\rangle = 0$. For $|x| > K$ and (2.2) we have

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \leq -\left[\lambda_{\text{ref}} - 2\lambda_{\text{ref}} \int_{\theta=0}^{\pi/2} \frac{\sqrt{\lambda_{\text{ref}}}p_\vartheta\,(\theta)\,\mathrm{d}\theta}{\sqrt{\lambda_{\text{ref}}+|\nabla U(x)|\cos(\theta)}}\right]$$
$$\leq -\lambda_{\text{ref}}\,(1-2F\,(\alpha_2/\lambda_{\text{ref}},d))$$
$$\leq -\lambda_{\text{ref}}\left(1-2\times\frac{1}{4}\right) = -\frac{\lambda_{\text{ref}}}{2},$$

as long as $\alpha_2/\lambda_{\text{ref}} > c_d$, with c_d defined as in the statement of Theorem 3.1(B).

Case $\langle \nabla U(x), v\rangle > 0$. From (2.2) we have

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \leq -\left[\langle \nabla U(x), v\rangle - 2\frac{\langle \nabla U(x), v\rangle\sqrt{\lambda_{\text{ref}}}}{\sqrt{\lambda_{\text{ref}}+\langle \nabla U(x), v\rangle}} + \lambda_{\text{ref}} - 2\lambda_{\text{ref}}F\left(\frac{\alpha_2}{\lambda_{\text{ref}}},d\right)\right].$$

For $w \geq 0$, using again that $\lambda_{\text{ref}} + w \geq 2\sqrt{w}\sqrt{\lambda_{\text{ref}}}$, we have

$$w - 2\frac{w\sqrt{\lambda_{\text{ref}}}}{\sqrt{\lambda_{\text{ref}}+w}} + \lambda_{\text{ref}}(1-2\epsilon) \geq w - \sqrt{2}\lambda_{\text{ref}}^{1/4}w^{3/4} + \lambda_{\text{ref}}(1-2\epsilon) = f_\epsilon(w).$$

Recall from (2.4) that for all $w \geq 0$

$$f_\epsilon(w) \geq \lambda_{\rm ref}\left(\frac{37}{64} - 2\epsilon\right) > 0,$$

as long as $\epsilon < 37/128$. For each d and $\alpha_2 > 0$ we can choose $\lambda_{\rm ref}$ small enough so that $F(\alpha_2/\lambda_{\rm ref}, d) < 37/128$. Then following a similar reasoning as before it can be easily seen that, as long as $\lambda_{\rm ref}$ is small enough, then there exists a $\delta > 0$, such that for all $|x| > K$ and $\langle \nabla U(x), v\rangle > 0$ we have

$$2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \leq -\delta.$$

Case $\langle \nabla U(x), v\rangle < 0$. Suppose that $\lambda_{\rm ref} \leq \alpha_2/c_d$ or equivalently that $\alpha_2/\lambda_{\rm ref} \geq c_d$. Then since $\|\Delta U(x)\| \to 0$, for all $\epsilon_1 > 0$, there is a $K > 0$ such that for all $|x| > K$ and $\lambda_{\rm ref}$ small enough

$$\begin{aligned}
&2\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \\
&\leq -\Bigg[\langle \nabla U(x), -v\rangle + 2\lambda_{\rm ref} - \frac{1}{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle \\
&\qquad - \lambda_{\rm ref}\frac{\sqrt{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\rm ref}}} - 2\lambda_{\rm ref}\frac{\sqrt{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\rm ref}}}F(\alpha_2/\lambda_{\rm ref}, d)\Bigg] \\
&\leq -\Bigg[\langle \nabla U(x), -v\rangle + 2\lambda_{\rm ref} - \frac{1}{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}\langle v, \Delta U(x) v\rangle \\
&\qquad - \lambda_{\rm ref}\frac{\sqrt{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\rm ref}}} - 2\lambda_{\rm ref}\frac{\sqrt{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\rm ref}}}F(c_d, d)\Bigg] \\
&\leq -\Bigg[\langle \nabla U(x), -v\rangle + 2\lambda_{\rm ref} - \frac{\epsilon_1}{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle} \\
&\qquad - \lambda_{\rm ref}\frac{\sqrt{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}}{\sqrt{\lambda_{\rm ref}}} - \frac{1}{2}\sqrt{\lambda_{\rm ref}}\sqrt{\lambda_{\rm ref} + \langle \nabla U(x), -v\rangle}\Bigg].
\end{aligned}$$

Let $w = \langle \nabla U(x), -v\rangle > 0$ and consider

$$g(w) := w + 2\lambda_{\rm ref} - \frac{\epsilon_1}{\lambda_{\rm ref} + w} - \lambda_{\rm ref}\frac{\sqrt{\lambda_{\rm ref} + w}}{\sqrt{\lambda_{\rm ref}}} - \frac{1}{2}\sqrt{\lambda_{\rm ref}}\sqrt{\lambda_{\rm ref} + w}.$$

Then we obtain

$$g'(w) = 1 + \frac{\epsilon_1}{(\lambda_{\rm ref} + w)^2} - \frac{3\sqrt{\lambda_{\rm ref}}}{4\sqrt{\lambda_{\rm ref} + w}}. \geq 0.$$

Thus, we have

$$g(w) \geq g(0) = \frac{\lambda_{\text{ref}}}{2} - \frac{\epsilon_1}{\lambda_{\text{ref}}},$$

which is strictly positive as long as $\epsilon_1 < \lambda_{\text{ref}}^2/2$, and the result follows. □

PROOF OF LEMMA 5. First notice that V with $\lambda_{\text{ref}}(x)$ as defined in the statement of the Lemma also belongs to $\mathcal{D}(\widetilde{\mathcal{L}})$ from the same arguments as in the proof of Lemma 4. We now prove that V satisfies ($\mathfrak{D}$). From the form of ($\mathfrak{D}$) it follows that we can assume without loss of generality that $|x| > 1$, so that

$$\Lambda_{\text{ref}}(x) = \lambda_{\text{ref}} + \frac{|\nabla U(x)|}{|x|^{\epsilon}}.$$

First we restrict our attention to the case where $\langle \nabla U(x), v\rangle \neq 0$, for which we compute

$$\begin{aligned}
\frac{\partial V}{\partial x_i} &= \frac{1}{2} V(x,v) U_{x_i}(x) \\
&\quad - \frac{1}{2} \frac{\mathrm{e}^{U(x)/2}}{(\Lambda_{\text{ref}}(x) + \langle \nabla U(x), -v\rangle_+)^{3/2}} \left[\frac{\partial}{\partial x_i} \Lambda_{\text{ref}}(x) + \frac{\partial}{\partial x_i} \langle \nabla U(x), -v\rangle_+ \right], \\
\langle \nabla V, v\rangle &= \frac{1}{2} V(x,v) \langle \nabla U(x), v\rangle \\
&\quad - \frac{1}{2} \frac{\mathrm{e}^{U(x)/2}}{(\Lambda_{\text{ref}}(x) + \langle \nabla U(x), -v\rangle_+)^{3/2}} [\langle \nabla \Lambda_{\text{ref}}(x), v\rangle \\
&\quad - \langle v, \Delta U(x) v\rangle \mathbb{1}\{\langle \nabla U(x), -v\rangle > 0\}] \\
&= \frac{1}{2} V(x,v) \Big\{ \langle \nabla U(x), v\rangle - \frac{\langle \nabla \Lambda_{\text{ref}}(x), v\rangle}{\Lambda_{\text{ref}}(x) + \langle \nabla U(x), -v\rangle_+} \\
&\qquad + \frac{\langle v, \Delta U(x) v\rangle \mathbb{1}\{\langle \nabla U(x), -v\rangle > 0\}}{\Lambda_{\text{ref}}(x) + \langle \nabla U(x), -v\rangle_+} \Big\}.
\end{aligned}$$

After adding the reflection and refreshment terms we get

$$\begin{aligned}
\widetilde{\mathcal{L}} V(x,v) &= \frac{1}{2} V(x,v) \Big\{ \langle \nabla U(x), v\rangle - \frac{\langle \nabla \Lambda_{\text{ref}}(x), v\rangle}{\Lambda_{\text{ref}}(x) + \langle \nabla U(x), -v\rangle_+} \\
&\qquad + \frac{\langle v, \Delta U(x) v\rangle \mathbb{1}\{\langle \nabla U(x), -v\rangle > 0\}}{\Lambda_{\text{ref}}(x) + \langle \nabla U(x), -v\rangle_+} \Big\} \\
&\quad + \Lambda_{\text{ref}}(x) \int V(x,w) \psi(\mathrm{d}w) - \Lambda_{\text{ref}}(x) V(x,v) \\
&\quad + \langle \nabla U(x), v\rangle \mathbb{1}\{\langle \nabla U(x), v\rangle \geq 0\} [V(x, R(x)v) - V(x,v)],
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} &= \frac{1}{2}\Bigg\{\langle\nabla U(x),v\rangle - \frac{\langle\nabla\Lambda_{\mathrm{ref}}(x),v\rangle}{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),-v\rangle_+} \\
&\quad + \frac{\langle v,\Delta U(x)v\rangle\mathbb{1}\{\langle\nabla U(x),-v\rangle>0\}}{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),-v\rangle_+}\Bigg\} \\
&\quad + \Lambda_{\mathrm{ref}}(x)\int\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),-v\rangle_+}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),w\rangle_+}}-1\right]\psi(\mathrm{d}w) \\
&\quad + \langle\nabla U(x),v\rangle\mathbb{1}\{\langle\nabla U(x),v\rangle\geq 0\}\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),-v\rangle_+}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),v\rangle_+}}-1\right].
\end{aligned}$$

Thus when $\langle\nabla U(x),v\rangle>0$ we have

$$\begin{aligned}
\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} &= \frac{1}{2}\langle\nabla U(x),v\rangle - \frac{1}{2}\frac{\langle\nabla\Lambda_{\mathrm{ref}}(x),v\rangle}{\Lambda_{\mathrm{ref}}(x)} \qquad (2.5)\\
&\quad + \langle\nabla U(x),v\rangle\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),v\rangle}}-1\right] \\
&\quad + \Lambda_{\mathrm{ref}}(x)\int_{\langle\nabla U(x),w\rangle\geq 0}\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),w\rangle_+}}-1\right]\psi(\mathrm{d}w). \qquad (2.6)
\end{aligned}$$

When $\langle\nabla U(x),v\rangle<0$ then

$$\begin{aligned}
\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} &= \frac{1}{2}\Bigg\{\langle\nabla U(x),v\rangle - \frac{\langle\nabla\Lambda_{\mathrm{ref}}(x),v\rangle - \langle v,\Delta U(x)v\rangle}{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),-v\rangle}\Bigg\} \\
&\quad + \Lambda_{\mathrm{ref}}(x)\int\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),-v\rangle_+}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),w\rangle_+}}-1\right]\psi(\mathrm{d}w). \qquad (2.7)
\end{aligned}$$

When $\langle\nabla U(x),v\rangle=0$, similarly to the proof of Lemma 4, by considering separately the case where $\langle\Delta U(x),-v\rangle>0$ and $\langle\Delta U(x),-v\rangle\leq 0$ we find that

$$\begin{aligned}
&\lim_{t\to 0+}\frac{\mathrm{d}}{\mathrm{d}t}V(x+tv,v) \\
&= \lim_{t\to 0+}\frac{1}{t}\left\{V(x+tv,v)-V(x,v)\right\} \\
&= \lim_{t\to 0+}\frac{1}{t}\left\{\frac{\exp\left(U(x+tv)/2\right)}{\sqrt{\Lambda_{\mathrm{ref}}(x+tv)+\langle\nabla U(x+tv),-v\rangle_+}} - \frac{\exp\left(U(x)/2\right)}{\sqrt{\Lambda_{\mathrm{ref}}(x)}}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}\frac{\exp\left(U(x)/2\right)}{\sqrt{\Lambda_{\text{ref}}(x)}^3}\left[\langle \nabla \Lambda_{\text{ref}}(x), v\rangle + \langle \Delta U(x) v, -v\rangle_+\right] \\
&= -\frac{V(x,v)}{2}\frac{\left[\langle \nabla \Lambda_{\text{ref}}(x), v\rangle + \langle \Delta U(x) v, -v\rangle_+\right]}{\Lambda_{\text{ref}}(x)}.
\end{aligned}
$$

Thus for $\langle \nabla U(x), v\rangle = 0$, after adding the refreshment term we have

$$
\begin{aligned}
\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} &= -\frac{1}{2}\frac{\langle \nabla \Lambda_{\text{ref}}(x), v\rangle + \langle \Delta U(x) v, -v\rangle_+}{\Lambda_{\text{ref}}(x)} \\
&\quad + \Lambda_{\text{ref}}(x)\int\left[\frac{\sqrt{\Lambda_{\text{ref}}(x)}}{\sqrt{\Lambda_{\text{ref}}(x)+\langle \nabla U(x), w\rangle_+}} - 1\right]\psi(\,\mathrm{d}w) \qquad (2.8)\\
&\leq -\frac{1}{2}\frac{\langle \nabla \Lambda_{\text{ref}}(x), v\rangle}{\Lambda_{\text{ref}}(x)} \qquad (2.9)\\
&\quad + \Lambda_{\text{ref}}(x)\int\left[\frac{\sqrt{\Lambda_{\text{ref}}(x)}}{\sqrt{\Lambda_{\text{ref}}(x)+\langle \nabla U(x), w\rangle_+}} - 1\right]\psi(\,\mathrm{d}w).
\end{aligned}
$$

From the definition of $\Lambda_{\text{ref}}(x)$ and the chain rule

$$
\nabla \Lambda_{\text{ref}}(x) = |x|^{-\epsilon}\nabla|\nabla U(x)| + |\nabla U(x)|\nabla\left(|x|^{-\epsilon}\right).
$$

We first compute

$$
\frac{\partial}{\partial x_i}|\nabla U(x)| = \frac{\partial}{\partial x_i}\sqrt{\sum_j\left(\frac{\partial}{\partial x_j}U\right)^2} = |\nabla U(x)|^{-1}\sum_j \frac{\partial}{\partial x_j}U(x)\frac{\partial^2}{\partial x_i \partial x_j}U(x),
$$

whence it follows that

$$
\begin{aligned}
|\nabla\, |\nabla U|| &= |\nabla U|^{-1}|\left\{\sum_{i=1}^d\left[\sum_{j=1}^d \frac{\partial}{\partial x_j}U(x)\frac{\partial^2}{\partial x_i \partial x_j}U(x)\right]^2\right\}^{1/2} \\
&\leq |\nabla U(x)|^{-1}|\nabla U|\|\Delta U\| = \|\Delta U\|
\end{aligned}
$$

Thus we have that

$$
\begin{aligned}
\frac{|\nabla \Lambda_{\text{ref}}(x)|}{|\Lambda_{\text{ref}}(x)|} &\leq \frac{\|\Delta U(x)\|/|x|^{\epsilon}}{|\nabla U(x)|/|x|^{\epsilon}} + \frac{|\nabla U(x)|}{|x|^{1+\epsilon}\times|\nabla U(x)|/|x|^{\epsilon}} \\
&= \frac{\|\Delta U(x)\|}{|\nabla U(x)|} + \frac{1}{|x|} \to 0,
\end{aligned}
$$

where we also used the fact that $|\nabla(|x|^{-\epsilon})| = \epsilon|x|^{-1-\epsilon}$. It therefore follows that

$$
\overline{\lim_{|x|\to\infty}}\frac{|\langle \nabla \Lambda_{\text{ref}}(x), v\rangle|}{\Lambda_{\text{ref}}(x)} = 0, \qquad (2.10)
$$

so that this term can be ignored for large $|x|$. Also notice that

$$\begin{aligned}
&\int_{w:\langle\nabla U(x),w\rangle\geq 0}\psi(\,\mathrm{d}w)\frac{1}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),w\rangle_+}}\\
&=\frac{1}{|\nabla U(x)|^{1/2}}\int_{w:\langle\nabla U(x),w\rangle\geq 0}\frac{\psi(\,\mathrm{d}w)}{\sqrt{\frac{\Lambda_{\mathrm{ref}}(x)}{|\nabla U(x)|}+\left\langle\frac{\nabla U(x)}{|\nabla U(x)|},w\right\rangle}}\\
&=\frac{1}{|\nabla U(x)|^{1/2}}\int_{\theta=0}^{\pi/2}\frac{p_\vartheta(\,\mathrm{d}\theta)}{\sqrt{\frac{\Lambda_{\mathrm{ref}}(x)}{|\nabla U(x)|}+\cos(\theta)}},
\end{aligned}$$

where d is the dimension. As $|x|\to\infty$, our definition of $\Lambda_{\mathrm{ref}}(x)$ ensures that

$$\int_{\theta=0}^{\pi/2}\frac{p_\vartheta(\,\mathrm{d}\theta)}{\sqrt{\frac{\Lambda_{\mathrm{ref}}(x)}{|\nabla U(x)|}+\cos(\theta)}}\to\int_{\theta=0}^{\pi/2}\frac{p_\vartheta(\,\mathrm{d}\theta)}{\sqrt{\cos(\theta)}}=-\frac{3\Gamma\left(-\frac{3}{4}\right)\Gamma\left(\frac{d}{2}\right)}{8\sqrt{\pi}\Gamma\left(\frac{d}{2}-\frac{1}{4}\right)}=:\gamma_d>0.$$

Case $\langle\nabla U(x),v\rangle=0$. Thus when $\langle\nabla U(x),v\rangle=0$, for $|x|$ large we have

$$\begin{aligned}
\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)}&\sim\Lambda_{\mathrm{ref}}(x)\int\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),w\rangle_+}}-1\right]\psi(\,\mathrm{d}w)\\
&\sim\Lambda_{\mathrm{ref}}(x)\int_{\langle\nabla U(x),w\rangle<0}[1-1]\,\psi(\,\mathrm{d}w)+\Lambda_{\mathrm{ref}}(x)\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{|\nabla U(x)|}}\gamma_d-\frac{1}{2}\right]\\
&\sim\Lambda_{\mathrm{ref}}(x)\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{|\nabla U(x)|}}\gamma_d-\frac{1}{2}\right],
\end{aligned}$$

and from the definition of $\Lambda_{\mathrm{ref}}(x)$ it easily follows that

$$\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)}\sim-\frac{1}{2}\Lambda_{\mathrm{ref}}(x)\to-\infty.$$

Case $\langle\nabla U(x),v\rangle>0$. For $|x|$ large we have

$$\begin{aligned}
\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)}\sim\frac{1}{2}\langle\nabla U(x),v\rangle+\langle\nabla U(x),v\rangle\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{\Lambda_{\mathrm{ref}}(x)+\langle\nabla U(x),v\rangle}}-1\right]\\
+\Lambda_{\mathrm{ref}}(x)\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x)}}{\sqrt{|\nabla U(x)|}}\gamma_d-\frac{1}{2}\right].
\end{aligned}$$

Using the definition of $\Lambda_{\mathrm{ref}}(x)$, and letting $\langle\nabla U(x),v\rangle=|\nabla U(x)|\cos(\theta)$ for $\theta\in[0,\pi/2)$ we have for $\langle\nabla U(x),v\rangle>0$ as $|x|\to\infty$

$$\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)}\sim\frac{1}{2}|\nabla U(x)|\cos(\theta)$$

$$
\begin{aligned}
&+ |\nabla U(x)|\cos(\theta)\left[\frac{\sqrt{|\nabla U(x)|/|x|^\epsilon}}{\sqrt{|\nabla U(x)|/|x|^\epsilon + |\nabla U(x)|\cos(\theta)}} - 1\right]\\
&+ \frac{|\nabla U(x)|}{|x|^\epsilon}\left[\frac{\sqrt{|\nabla U(x)|/|x|^\epsilon}}{\sqrt{|\nabla U(x)|}}\gamma_d - \frac{1}{2}\right]\\
&= |\nabla U(x)|\left[\cos(\theta)\left(\frac{1}{2} + \frac{1}{\sqrt{1+|x|^\epsilon\cos(\theta)}} - 1\right) + \frac{1}{|x|^\epsilon}\left[\frac{\gamma_d}{|x|^{\epsilon/2}} - \frac{1}{2}\right]\right]\\
&= \frac{|\nabla U(x)|}{|x|}|x|\left[\cos(\theta)\left(-\frac{1}{2} + \frac{1}{\sqrt{1+|x|^\epsilon\cos(\theta)}}\right) + \frac{1}{|x|^\epsilon}\left[\frac{\gamma_d}{|x|^{\epsilon/2}} - \frac{1}{2}\right]\right]\\
&\leq |x|\left[\cos(\theta)\left(-\frac{1}{2} + \frac{1}{\sqrt{1+|x|^\epsilon\cos(\theta)}}\right) + \frac{1}{|x|^\epsilon}\left[\frac{\gamma_d}{|x|^{\epsilon/2}} - \frac{1}{2}\right]\right],
\end{aligned}
$$

since $|\nabla U(x)|/|x| \to \infty$ and the quantity in brackets is clearly negative for large enough $|x|$. Let $u = \cos(\theta)$ and $r = |x|$. Then observe that we can rewrite the right hand side as

$$
\begin{aligned}
&r^{1-\epsilon}r^\epsilon u\left(\frac{1}{\sqrt{1+r^\epsilon u}} - \frac{1}{2}\right) - \frac{r^{1-\epsilon}}{2} + O(r^{1/4})\\
&\qquad\qquad \leq \frac{r^{1-\epsilon}}{4} - \frac{r^{1-\epsilon}}{2} + O(r^{1-3\epsilon/2}) = -\frac{\sqrt{r}}{4} + O(r^{1-3\epsilon/2}),
\end{aligned}
$$

since for $w = r^\epsilon u > 0$ it can be shown that

$$
w\left(\frac{1}{\sqrt{1+w}} - \frac{1}{2}\right) \leq \frac{1}{4}.
$$

Thus it follows that for $\langle \nabla U(x), v\rangle > 0$ we have that $\overline{\lim}_{|x|\to\infty} \widetilde{\mathcal{L}}V/V = -\infty$.

Case $\langle \nabla U(x), v\rangle < 0$. From (2.7) and (2.10) we have as $|x| \to \infty$

$$
\begin{aligned}
&2\,\overline{\lim_{|x|\to\infty}}\,\frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)}\\
&= \overline{\lim_{|x|\to\infty}}\left\{\langle \nabla U(x), v\rangle - \frac{\langle \nabla\Lambda_{\mathrm{ref}}(x), v\rangle - \langle v, \Delta U(x) v\rangle}{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle}\right.\\
&\qquad\left. + 2\Lambda_{\mathrm{ref}}(x)\int\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle_+}}{\sqrt{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), w\rangle_+}} - 1\right]\psi(\,\mathrm{d}w)\right\}\\
&= \overline{\lim_{|x|\to\infty}}\left\{\langle \nabla U(x), v\rangle + \frac{\langle v, \Delta U(x) v\rangle}{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle}\right.
\end{aligned}
$$

$$
\begin{aligned}
&+ \Lambda_{\mathrm{ref}}(x)\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle}}{\sqrt{\Lambda_{\mathrm{ref}}(x)}} - 1\right] \\
&+ 2\Lambda_{\mathrm{ref}}(x)\frac{\sqrt{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle_+}}{\sqrt{|\nabla U(x)|}}\gamma_d - \Lambda_{\mathrm{ref}}(x)\Bigg\} \\
= \overline{\lim_{|x|\to\infty}} &\Bigg\{\langle \nabla U(x), v\rangle + \Lambda_{\mathrm{ref}}(x)\left[\frac{\sqrt{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle}}{\sqrt{\Lambda_{\mathrm{ref}}(x)}} - 1\right] \\
&+ 2\Lambda_{\mathrm{ref}}(x)\frac{\sqrt{\Lambda_{\mathrm{ref}}(x) + \langle \nabla U(x), -v\rangle_+}}{\sqrt{|\nabla U(x)|}}\gamma_d - \Lambda_{\mathrm{ref}}(x)\Bigg\},
\end{aligned}
$$

since $\overline{\lim}_{|x|\to\infty} \|\Delta U(x)\|/\Lambda_{\mathrm{ref}}(x) \to 0$. Thus letting θ be the angle between $U(x)$ and $-v$, we have

$$
\begin{aligned}
&2\,\overline{\lim_{|x|\to\infty}} \frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} \\
&= \overline{\lim_{|x|\to\infty}} \Bigg\{ -|\nabla U(x)|\cos(\theta) + \frac{|\nabla U(x)|}{|x|^\epsilon}\left[\frac{\sqrt{|\nabla U(x)|/|x|^\epsilon + |\nabla U(x)|\cos(\theta)}}{\sqrt{|\nabla U(x)|/|x|^\epsilon}} - 1\right] \\
&\quad + 2\frac{|\nabla U(x)|}{|x|^\epsilon}\left(\frac{\sqrt{|\nabla U(x)|/|x|^\epsilon + |\nabla U(x)|\cos(\theta)}}{\sqrt{|\nabla U(x)|}}\gamma_d - \frac{1}{2}\right)\Bigg\} \\
&= \overline{\lim_{|x|\to\infty}} \Bigg\{ -|\nabla U(x)|\cos(\theta) + \frac{|\nabla U(x)|}{|x|^\epsilon}\left[\sqrt{1 + |x|^\epsilon\cos(\theta)} - 1\right] \\
&\quad + 2\frac{|\nabla U(x)|}{|x|^\epsilon}\left(\sqrt{\frac{1}{|x|^\epsilon} + \cos(\theta)}\gamma_d - \frac{1}{2}\right)\Bigg\} \\
&= \overline{\lim_{|x|\to\infty}} |\nabla U(x)|\Bigg\{ -\cos(\theta) + \frac{1}{|x|^\epsilon}\left[\sqrt{1 + |x|^\epsilon\cos(\theta)} - 1\right] + \\
&\quad \frac{2}{|x|^\epsilon}\left(\sqrt{\frac{1}{|x|^\epsilon} + \cos(\theta)}\gamma_d - \frac{1}{2}\right)\Bigg\} \\
&= \overline{\lim_{|x|\to\infty}} \frac{|\nabla U(x)|}{|x|}|x|\Bigg\{ -\cos(\theta) + \frac{1}{|x|^\epsilon}\left[\sqrt{1 + |x|^\epsilon\cos(\theta)} - 1\right] \\
&\quad + \frac{2}{|x|^\epsilon}\left(\sqrt{\frac{1}{|x|^\epsilon} + \cos(\theta)}\gamma_d - \frac{1}{2}\right)\Bigg\} \\
&\le \overline{\lim_{|x|\to\infty}} |x|\Bigg\{ -\cos(\theta) + \frac{1}{|x|^\epsilon}\left[\sqrt{1 + |x|^\epsilon\cos(\theta)} - 1\right]
\end{aligned}
$$

$$+ \frac{2}{|x|^\epsilon}\left(\sqrt{\frac{1}{|x|^\epsilon} + \cos(\theta)}\gamma_d - \frac{1}{2}\right)\Bigg\},$$

since the right hand side is clearly negative for $|x|$ large enough.

For $u = \cos(\theta) \in [0,1]$ define the function

$$f(u) := -u + \frac{1}{|x|^\epsilon}\left[\sqrt{1+|x|^\epsilon u} - 1\right] + \frac{2}{|x|^\epsilon}\left(\sqrt{\frac{1}{|x|^\epsilon} + u}\gamma_d - \frac{1}{2}\right).$$

Then

$$f'(u) = -1 + \frac{|x|^\epsilon}{2|x|^\epsilon\sqrt{1+|x|^\epsilon u}} + \frac{1}{|x|^\epsilon\sqrt{\frac{1}{|x|^\epsilon}+u}}\gamma_d.$$

This is negative for all $u \geq 0$ for $|x|$ large enough. Therefore

$$f(u) \leq f(0) = \frac{2}{|x|^\epsilon}\left(\sqrt{\frac{1}{|x|^\epsilon}}\gamma_d - \frac{1}{2}\right) \sim -\frac{1}{|x|^\epsilon},$$

as $|x| \to \infty$. Hence

$$\overline{\lim_{|x|\to\infty}} \frac{\widetilde{\mathcal{L}}V(x,v)}{V(x,v)} = -\infty,$$

and the result follows. □

Proof of Lemma 6. *Checking Assumption* (A0'). Notice that from equations (3.6), (3.7) and (3.8) of the main manuscript, the functions $h^{(i)}$, are infinitely differentiable except perhaps for $x = 0$ and $|x| = 1/b$ for $i = 1$, or $|x| = R$ for $i = 2$. Thus U_h will satisfy Assumption (A0) for $|x|$ large enough and in fact everywhere except for $|x| = 0, 1/b$ for $i = 1$, and $|x| = 0, R$ for $i = 2$. It remains to show that the mapping $t \mapsto \langle \nabla U_h(x+tv), v\rangle$ is locally Lipschitz at these points. First, from the definition of $f = f^{(i)}$, it follows easily that the mapping $t \mapsto \langle \nabla U_h(x+tv), v\rangle$ will be continuous and piecewise smooth, and thus locally Lipschitz, at $|x| = 1/b$ and $|x| = R$ for $i = 1, 2$ respectively. To deal with the remaining case $x = 0$, we next show that $t \mapsto \langle \nabla U_h(tv), v\rangle$ is in fact differentiable at $t = 0$.

Recall the decomposition of ∇U_h given in (3.10) of the main manuscript. The first term of (3.10) is given by

$$\nabla \log\det(\nabla h(x)) = \begin{cases} \left[\frac{f''(|x|)}{f'(|x|)} + (d-1)\left(\frac{f'(|x|)}{f(|x|)} - \frac{1}{|x|}\right)\right]\frac{x}{|x|}, & x \neq 0, \\ 0, & x = 0, \end{cases} \tag{2.11}$$

whence we can compute

$$\frac{1}{t}\left[\langle \nabla \log\det(\nabla h(tv)), v\rangle - \langle \nabla \log\det(\nabla h(0)), v\rangle\right]$$

$$= \frac{1}{t}\left[\frac{f''(t)}{f'(t)} + (d-1)\left(\frac{f'(t)}{f(t)} - \frac{1}{t}\right)\right]\left\langle \frac{tv}{|tv|}, v\right\rangle$$
$$= \frac{1}{t}\left[\frac{f''(t)}{f'(t)} + (d-1)\left(\frac{f'(t)}{f(t)} - \frac{1}{t}\right)\right].$$

In the case $f = f^{(1)}$ we have

$$f(0) = 0, \quad f'(0) = \frac{b\mathrm{e}}{2}, \quad f''(0) = 0, \quad f'''(0) = b^3\mathrm{e},$$

and thus using Taylor expansions

$$\begin{aligned}
&\lim_{t\to 0}\frac{1}{t}\left[\frac{f''(t)}{f'(t)} + (d-1)\left(\frac{f'(t)}{f(t)} - \frac{1}{t}\right)\right]\\
&= 0 + (d-1)\lim_{t\to 0}\frac{1}{t^2 f(t)}[tf'(t) - f(t)]\\
&= (d-1)\lim_{t\to 0}\frac{1}{t^3 f'(0)}\Big(tf'(0) + t^2 f''(0) + \frac{t^3}{2}f'''(0) - f(0)\\
&\qquad - tf'(0) - \frac{t^2}{2}f''(0) - \frac{t^3}{6}f'''(0) + o(t^3)\Big)\\
&= (d-1)\lim_{t\to 0}\frac{1}{t^3 f'(0)}\left(\frac{t^3}{2}f'''(0) - \frac{t^3}{6}f'''(0) + o(t^3)\right) = \frac{(d-1)}{3}\frac{f'''(0)}{f'(0)}.
\end{aligned}$$

In the case $f = f^{(2)}$, we have for $t > 0$ small enough

$$\frac{1}{t}\left[\frac{f''(t)}{f'(t)} + (d-1)\left(\frac{f'(t)}{f(t)} - \frac{1}{t}\right)\right] = 0.$$

Thus overall $t \mapsto \langle \nabla \log\det(\nabla h(tv)), v\rangle$ is differentiable at $t = 0$ and thus locally Lipschitz.

We now deal with the second term of (3.10). From (5.1) we have

$$\begin{aligned}
&\frac{1}{t}\left[\langle \nabla h(tv)\nabla U(h(tv)), v\rangle - \langle \nabla h(0)\nabla U(0), v\rangle\right]\\
&= \frac{1}{t}\left[\langle \nabla h(tv)\nabla U(h(tv)), v\rangle - f'(0)\langle \nabla U(0), v\rangle\right]\\
&= \frac{1}{t}\left[\frac{f(t)}{t}\langle \nabla U(h(tv)), v\rangle - f'(0)\langle \nabla U(0), v\rangle\right] + \frac{1}{t}\left[f'(t) - \frac{f(t)}{t}\right]\frac{\langle tv, v\rangle\langle \nabla U(h(tv)), tv\rangle}{t^2}\\
&= I_1 + I_2.
\end{aligned}$$

For the first term we have

$$I_1 = \frac{1}{t}\frac{f(t) - tf'(0)}{t}\langle \nabla U(h(tv)), v\rangle + \frac{1}{t}f'(0)\left[\langle \nabla U(h(tv)), v\rangle - \nabla U(0), v\rangle\right].$$

Since U satisfies (A0) and h is differentiable, the second term of I_1 clearly converges. The first term also converges since ∇U is continuous and

$$\frac{f(t) - tf'(0)}{t^2} = \frac{f(0) + tf'(0) + t^2 f''(0) + o(t^2) - tf'(0)}{t^2}.$$

For the second term we have

$$\begin{aligned} I_2 &= \frac{1}{t}\left[f'(t) - \frac{f(t)}{t}\right]\langle \nabla U(h(tv)), v\rangle \\ &= \frac{f'(t)t - f(t)}{t^2}\langle \nabla U(h(tv)), v\rangle \\ &= \frac{f'(0)t + f''(0)t^2 + o(t^2) - f(0) - tf'(0) - \frac{t^2}{2}f''(0)}{t^2}\langle v, v\rangle\langle \nabla U(h(tv)), v\rangle \\ &\to 0. \end{aligned}$$

It follows that $t \mapsto \langle \nabla h(x + tv)\nabla U(h(x + tv)), v\rangle$ is differentiable at $t = 0$.

Checking Assumption (A1). For both $h = h^{(1)}$ and $h = h^{(2)}$, a change of variable leads to

(2.12)
$$\begin{aligned} &\int \pi_h(y)|\nabla U_h(y)|\,\mathrm{d}y \\ &= \int \pi_h(h^{-1}(x))|\nabla h^{-1}(x)||\nabla U_h(h^{-1}(x))|\,\mathrm{d}x \\ &= \int \pi(h(h^{-1}(x)))|\nabla U_h(h^{-1}(x))||\nabla h(h^{-1}(x))||\nabla h^{-1}(x)|\,\mathrm{d}x \end{aligned}$$

(2.13)
$$\begin{aligned} &= \int \pi(x)|\nabla U_h(h^{-1}(x))|\,\mathrm{d}x \\ &\le \int \pi(x)[|\nabla\{h\}(h^{-1}(x))\nabla U(x)| + |\nabla \log\det(\nabla\{h\}(h^{-1}(x)))|]\,\mathrm{d}x \end{aligned}$$

(2.14)
$$\le \int \pi(x)[\|\nabla\{h\}(h^{-1}(x))\||\nabla U(x)| + |\nabla \log\det(\nabla\{h\}(h^{-1}(x)))|]\,\mathrm{d}x.$$

Here for clarity we use the notation $\nabla\{\cdot\}(x)$ for the gradient of the function in the bracket evaluated at x and we will similarly use $\Delta\{\cdot\}(x)$ for its Hessian. We begin with the first term in (2.14). Under the assumptions of Theorem 3.3(A) we have, for $|x| > R$ and some constant $C > 0$, that $|\nabla U(x)| \le C|x|^{-1}$ and thus

$$\int \pi(x)|\nabla U(x)|\|\nabla\{h\}(h^{-1}(x))\|\,\mathrm{d}x \le C + C\int_{|x|>R}\pi(x)\frac{1}{|x|}f(|h^{-1}(x)|)\,\mathrm{d}x$$

$$\leq C + C \int_{|x|>R} \pi(x) \frac{1}{|x|} |x| \, \mathrm{d}x \leq 2C,$$

since clearly $f(|h^{-1}(x)|) = |x|$.

Under the assumptions of Theorem 3.3(B), by Assumption (B)-(ii), we can assume that there exists $K > 0$ such that if $|x| > K$ then $\langle x, \nabla U(x) \rangle \geq C|x|^\beta$ for some $C > 0$. Thus for $|x|$ large enough, say $K/|x| < 1/2$, we have

$$\begin{aligned} U(x) &= U\left(K\frac{x}{|x|}\right) + \int_{t=K/|x|}^{1} \frac{\mathrm{d}U(tx)}{\mathrm{d}t} \, \mathrm{d}t \\ &\geq U\left(K\frac{x}{|x|}\right) + \int_{t=1/2}^{1} \frac{1}{t} \langle \nabla U(tx), tx \rangle \, \mathrm{d}t \\ &\geq U\left(K\frac{x}{|x|}\right) + C\int_{t=1/2}^{1} \frac{1}{t} t^\beta |x|^\beta \, \mathrm{d}t \geq C|x|^\beta, \end{aligned} \tag{2.15}$$

since $U \geq 0$. Therefore

$$\begin{aligned} \int \pi(x) |\nabla U(x)| \|\nabla\{h\}(h^{-1}(x))\| \, \mathrm{d}x &\leq C\Big[1 + \int_{|x|>R} \pi(x) |x|^{\beta-1} f(|h^{-1}(x)|) \, \mathrm{d}x\Big] \\ &\leq C\Big[1 + \int_{|x|>R} \mathrm{e}^{-C|x|^\beta} |x|^\beta \, \mathrm{d}x\Big] < \infty. \end{aligned}$$

For the second term of (2.14), let

$$L'(x) := |\nabla \log \det(\nabla h(x))| \, .$$

From equation (5.2) of the main manuscript it follows easily that L' is bounded for both $h = h^{(1)}$ and $h = h^{(2)}$, and thus

$$\int \pi(x) |\nabla \log \det(\nabla\{h\}(h^{-1}(x)))| \, \mathrm{d}x < \infty.$$

Checking Assumption (A2). For $h = h^{(1)}$, notice that by [3, Lemma 4], and the fact that $h(\cdot)$ is isotropic in the sense of [3], it follows that

$$\overline{\lim_{|y|\to\infty}} |\nabla \log \det(\nabla h(y))| < C,$$

for some $C > 0$. Therefore

$$\begin{aligned} |\nabla U_h(y)| &\leq |\nabla h(y) \nabla U(h(y))| + |\nabla \log \det(\nabla h(y))| \\ &\leq \|\nabla h(y)\| \, |\nabla U(h(y))| + C \end{aligned}$$

and using Assumption (A)-(i) and equation (5.1) of the main manuscript

$$\leq C \, \|\nabla h(y)\| \, / |h(y)| + C \leq C,$$

since $\|\nabla h(y)\| \leq C|h(y)|$. Thus it follows that

$$\frac{\mathrm{e}^{U_h(y)/2}}{\sqrt{|\nabla U_h(y)|}} \geq C\mathrm{e}^{U_h(y)/2} \to \infty,$$

as $|y| \to \infty$ since $\mathrm{e}^{-U_h(y)}$ is integrable.

On the other for $h = h^{(2)}$ notice that by [3, Lemma 2], and the fact that $h(\cdot)$ is isotropic in the sense of [3], we obtain

$$\overline{\lim_{|y|\to\infty}} |\nabla \log \det(\nabla h(y))| = 0. \tag{2.16}$$

Therefore

$$\begin{aligned} |\nabla U_h(y)| &\leq |\nabla h(y) \nabla U(h(y))| + |\nabla \log \det(\nabla h(y))| \\ &\leq \|\nabla h(y)\| \, |\nabla U(h(y))| + C. \end{aligned}$$

From equation (5.1) of the main manuscript it follows that $\|\nabla h(y)\| \leq C|y|^{p-1}$. Therefore, using Assumption (B)-(i)

$$\begin{aligned} |\nabla U_h(y)| &\leq C \, \|\nabla h(y)\| \, |h(y)|^{\beta-1} + C \\ &\leq C|y|^{p-1+p\beta-p} = C|y|^{p\beta-1}. \end{aligned}$$

Thus

$$\frac{\mathrm{e}^{U_h(y)/2}}{\sqrt{|\nabla U_h(y)|}} \geq C \frac{\mathrm{e}^{U(h(y))/2}}{\det\left(\nabla h(y)\right) \sqrt{|y|^{p\beta-1}}}.$$

Finally, recalling (2.15), for $|x|$ large enough, say $K/|x| < 1/2$, we have $U(x) \geq C|x|^{\beta}$. Since by definition $h(y) \sim |y|^p$, and from equation (5.2) of the main manuscript $\det(\nabla h(y))$ grows at most polynomially, we obtain

$$\frac{\mathrm{e}^{U(h(y))/2}}{\det\left(\nabla h(y)\right) \sqrt{|y|^{p\beta-1}}} \geq \frac{\mathrm{e}^{C|h(y)|^{\beta}/2}}{\det\left(\nabla h(y)\right) \sqrt{|y|^{p\beta-1}}} \geq \frac{\mathrm{e}^{C|y|^{p\beta}/2}}{\det\left(\nabla h(y)\right) \sqrt{|y|^{p\beta-1}}} \to \infty.$$

□

PROOF OF LEMMA 7(A). For notational simplicity, we assume $b = 1$ but the argument can be generalized to other values. We start by establishing the first condition of Theorem 3.1(B), i.e. that U_h satisfies our definition of exponential tail behaviour. In the remaining, assume $|x| > b^{-1} = 1$.

By Assumption (A)-(i) and Cauchy-Schwartz, we have for $|x|$ large enough

$$\left| \frac{\langle x, \nabla U(x) \rangle}{|x|} \right| \leq |\nabla U(x)| \leq \frac{c_1}{|x|},$$

hence π is a *sub-exponentially light density* as defined in [3, p. 3052]. This combined with Assumption (A)-(iii), which is equivalent to [3, Eq. (17)], means that we can apply [3, Theorem 3] to obtain that π_h is an *exponentially light density* as defined in [3, p. 3052]. More specifically, from the proof of [3, Theorem 3] it follows that

$$\overline{\lim_{|x|\to\infty}} \frac{-\langle x, \nabla U_h(x)\rangle}{|x|} = -b(\mathfrak{d} - d) < 0.$$

Applying Cauchy-Schwartz again, we obtain

$$0 < b(\mathfrak{d} - d) = \underline{\lim_{|x|\to\infty}} \frac{\langle x, \nabla U_h(x)\rangle}{|x|} \leq \underline{\lim_{|x|\to\infty}} |\nabla U_h(x)|,$$

which establishes the first condition of Theorem 3.1(B) with $\alpha_2 = b(\mathfrak{d} - d)/2$.

We now turn our attention to the Hessian condition of Theorem 3.1(B). We first decompose the norm of the Hessian as follows:

$$\|\Delta U_h(x)\| \leq \|\Delta\{U \circ h\}(x)\| + \|\Delta\{\log\det(\nabla h(x))\}(x)\|. \tag{2.17}$$

From [3, Lemma 1], we have for $|x| \geq 1$

$$\log\det(\nabla h(x)) = |x| + (d-1)[\log(e^{|x|} - e/3) - \log(|x|)] =: L(|x|),$$

hence

$$\Delta\{\log\det(\nabla h(x))\}(x) = \frac{L''(|x|)}{|x|^2} x x^T + \frac{L'(|x|)}{|x|} I_d - \frac{L'(|x|)}{|x|^3} x x^T.$$

We have

$$L'(r) = 1 + (d-1)\left[\frac{e^r}{e^r - e/3} - \frac{1}{r}\right],$$
$$L''(r) = (1-d)\left[\frac{e^{r+1}/3}{(e^r - e/3)^2} + \frac{1}{r^2}\right],$$

so $|L'(r)| \to d$, $|L''(r)| \to 0$ and therefore

$$\overline{\lim_{|x|\to\infty}} \|\Delta U_h(x)\| \leq \overline{\lim_{|x|\to\infty}} \|\Delta\{U \circ h\}(x)\| + \overline{\lim_{|x|\to\infty}} \|\Delta\{\log\det(\nabla h(x))\}(x)\|$$
$$= \overline{\lim_{|x|\to\infty}} \|\Delta\{U \circ h\}(x)\|.$$

To control this remaining term, we bound the operator norm with the Frobenius norm and write

$$\begin{aligned}\overline{\lim}_{|x|\to\infty} \|\Delta\{U\circ h\}(x)\|^2 &\leq \overline{\lim}_{|x|\to\infty} \sum_{i=1}^d \sum_{j=1}^d |\partial_i\partial_j\{U\circ h\}(x)|^2 \\ &= \overline{\lim}_{|x|\to\infty} \sum_{i=1}^d \sum_{j=1}^d \left|\partial_i \left\{\sum_{k=1}^d \partial_k\{U\}(h(x))\partial_j\{h_k\}(x)\right\}\right|^2,\end{aligned}$$

where we write $\partial_i\{\cdot\}(x)$ as a shorthand for the i-th partial derivative, $(\nabla\{\cdot\}(x))_i$.

It is enough to bound the d^2 expressions of the form

$$\begin{aligned}\left|\partial_i \left\{\sum_{k=1}^d \partial_k\{U\}(h(x))\partial_j\{h_k\}(x)\right\}\right| \leq \sum_{k=1}^d \Big[&|\partial_i\{\partial_k\{U\}\circ h\}(x)\}\partial_j\{h_k\}(x)| \\ &+ |\partial_k\{U\}(h(x))\partial_i\{\partial_j\{h_k\}(x)\}|\}\Big].\end{aligned} \tag{2.18}$$

The first term in Equation (2.18) is controlled as follows:

$$|\partial_i\{\partial_k\{U\}\circ h\}(x)\partial_j\{h_k\}(x)| \tag{2.19}$$

$$\leq |\partial_j\{h_k\}(x)| \sum_{m=1}^d |\partial_m\partial_k\{U\}(h(x))||\partial_i\{h_m\}(x)|. \tag{2.20}$$

Using again [3, Lemma 1], and the fact that $|h(x)| = f(|x|) \leq f'(|x|)$, for $|x|$ large enough,

$$|\partial_i\{h_j\}(x)| = \left|\frac{f(|x|)}{|x|}\mathbf{1}[i=j] + \left[f'(|x|) - \frac{f(|x|)}{|x|}\right]\frac{x_i x_j}{|x|^2}\right| \leq 3f(|x|), \tag{2.21}$$

hence using Assumption (A)-(ii), for $|x|$ large enough,

$$|\partial_i\{\partial_k\{U\}\circ h\}(x)\partial_j\{h_k\}(x)| \leq d\frac{c_2}{(h(|x|))^2}(3f(|x|))^2.$$

The second term in Equation (2.18) is controlled similarly, this time using Assumption (A)-(i), for $|x|$ large enough,

$$|\partial_k\{U\}(h(x))\partial_i\{\partial_j\{h_k\}(x)\}| \leq \frac{c_1}{h(|x|)}(8f(|x|)), \tag{2.22}$$

since it follows from [3, Lemma 1] that

$$|\partial_i\partial_j\{h_k\}(x)| \leq 8f(|x|). \qquad \square$$

Proof of Lemma 7(B). Let $f := f^{(2)}$, $h := h^{(2)}$ given in (3.7) and (3.8) of the main manuscript respectively. We need to check that the assumptions of Theorem 3.2 are satisfied. First we check that

$$\lim_{|x|\to\infty} \frac{|\nabla U_h(x)|}{|x|} = \infty.$$

From (2.16) and (3.10) of the main manuscript it follows that

$$\lim_{|x|\to\infty} \frac{|\nabla U_h(x)|}{|x|} = \lim_{|x|\to\infty} \frac{|\nabla h(x) \nabla U\left(h(x)\right)|}{|x|}.$$

Recall from [3, Lemma 1] that for $x \neq 0$

$$\nabla h(x) = \frac{f(|x|)}{|x|}\mathbb{1}_d + \left[f'(|x|) - \frac{f(|x|)}{|x|}\right]\frac{xx^T}{|x|^2}, \tag{2.23}$$

where $\mathbb{1}_d$ is the $d \times d$-identity matrix. Therefore we have

$$\begin{aligned}\nabla h(x)\nabla U\left(h(x)\right) &= \frac{f(|x|)}{|x|}\nabla U\left(h(x)\right) + \left[f'(|x|) - \frac{f(|x|)}{|x|}\right]\left\langle \nabla U\left(h(x)\right), \frac{x}{|x|}\right\rangle \frac{x}{|x|} \\ &= f'(|x|)\left\langle \nabla U\left(h(x)\right), \frac{x}{|x|}\right\rangle \frac{x}{|x|} + \frac{f(|x|)}{|x|}P_x^{\perp}\nabla U\left(h(x)\right),\end{aligned}$$

where $P_x^{\perp}$ denotes the orthogonal projection on the plane normal to x. Therefore, since by definition $h(x) := f(|x|)x/|x|$, we have that

$$\begin{aligned}\left|\nabla h(x)\nabla U\left(h(x)\right)\right| &\geq f'(|x|)\left|\left\langle \nabla U\left(h(x)\right), \frac{x}{|x|}\right\rangle\right| \\ &= \frac{f'(|x|)}{f(|x|)}\left|\left\langle \nabla U\left(h(x)\right), h(x)\right\rangle\right| \\ &= \frac{f'(|x|)}{f(|x|)}|h(x)|^{\beta}\left[|h(x)|^{-\beta}\left|\left\langle \nabla U\left(h(x)\right), h(x)\right\rangle\right|\right].\end{aligned}$$

Since $|h(x)| \to \infty$ as $|x| \to \infty$, Assumption (B)-(ii) and the definitions of f and h yield

$$\begin{aligned}\lim_{|x|\to\infty} \frac{|\nabla U_h(x)|}{|x|} &\geq \lim_{|x|\to\infty} |x|^{-1}\left\{\frac{f'(|x|)}{f(|x|)}|h(x)|^{\beta}\left[|h(x)|^{-\beta}\left|\left\langle \nabla U\left(h(x)\right), h(x)\right\rangle\right|\right]\right\} \\ &\geq C \lim_{|x|\to\infty} |x|^{-1-1+\beta p} = C|x|^{\beta p - 2} = \infty, \end{aligned} \tag{2.24}$$

since $\beta p > 2$.

Finally we need to check, that for some $\epsilon > 0$ we have

$$\lim_{|x|\to\infty} \frac{\|\Delta U_h(x)\|}{|\nabla U_h(x)|}|x|^{\epsilon} = 0.$$

Recall the expression (2.17). It follows easily from the definitions of h, f and [3, Lemma 1, Eq.(13)] that

$$\lim_{|x|\to\infty} \|\Delta \log\det(\nabla h)(x)\| = 0.$$

Therefore we focus on the first term of (2.17). As in the proof of the first part of the Theorem, we need essentially to control terms of the form (2.19) and terms of the form (2.22). To this end, using Assumption (B)-(iii), we estimate

$$\begin{aligned}
&|\partial_i\{\partial_k\{U\}\circ h\}(x)\partial_j\{h_k\}(x)| \\
&\le |\partial_j\{h_k\}(x)| \sum_{m=1}^{d} |\partial_m\partial_k\{U\}(h(x))||\partial_i\{h_m\}(x)| \\
&\le C|x|^{2p-2}|h(x)|^{\beta-2} \le C|x|^{2p-2+p\beta-2p} = C|x|^{p\beta-2},
\end{aligned}$$

since from equation (5.1) of the main manuscript and the definitions of f and h one can easily show that $|\partial_i h_k(x)| \le |x|^{p-1}$. On the other hand, from Assumption (B)-(i) and the fact that $|\partial_i\partial_j\{h_k\}(x)| \le C|x|^{p-2}$, which follows again from (5.1), the remaining terms can be estimated through

$$|\partial_k\{U\}(h(x))\partial_i\{\partial_j\{h_k\}\}(x)| \le |h(x)|^{\beta-1}|x|^{p-2} \le C|x|^{p\beta-2}.$$

Therefore combining the above with the arguments leading to (2.24) we have that as $|x|\to\infty$

$$\frac{\|\Delta U_h(x)\|}{|\nabla U_h(x)|}|x|^{\epsilon} \le C\frac{|x|^{\beta p-2}}{|x|^{\beta p-1}}|x|^{\epsilon} \to 0, \qquad \square$$

REFERENCES

[1] Costa, O.L.V. (1990). Stationary distributions for piecewise-deterministic Markov processes. *J. Appl. Probab.*, **27**:1, 60–73.

[2] Costa, O.L.V., and Dufour, F. (2008). Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, **47**:2, 1053–1077.

[3] Johnson, L.T., and Geyer, C.J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40**:6, 3050–3076.

[4] Meyn, S.P. and Tweedie, R.L. (1993). Stability of Markovian processes II: Continuous-time processes and sampled chains. *Adv. Appl. Probab.*, **25**:3, 487–517.

[5] Monmarché, P. (2016). Piecewise deterministic simulated annealing. *ALEA, Lat. Am. J. Proba. Math. Stat.*, **13**:1, 357–398.

Department of Statistics,
University of Oxford,
24-29 St. Giles,
OX1 3LB, Oxford, UK
E-mail: deligian@stats.ox.ac.uk

Department of Statistics,
University of British Columbia,
2207 Main Mall,
V6T 1Z4, Vancouver, Canada
E-mail: bouchard@stat.ubc.ca

Department of Statistics,
University of Oxford,
24-29 St. Giles,
OX1 3LB, Oxford, UK
E-mail: doucet@stats.ox.ac.uk