The Annals of Applied Probability
2021, Vol. 31, No. 6, 2612–2662
https://doi.org/10.1214/20-AAP1659

RANDOMIZED HAMILTONIAN MONTE CARLO AS SCALING LIMIT OF THE BOUNCY PARTICLE SAMPLER AND DIMENSION-FREE CONVERGENCE RATES

BY GEORGE DELIGIANNIDIS[1,*], DANIEL PAULIN[2], ALEXANDRE BOUCHARD-CÔTÉ[3] AND ARNAUD DOUCET[1,†]

[1]*Department of Statistics, University of Oxford,* [*]*deligian@stats.ox.ac.uk;* [†]*doucet@stats.ox.ac.uk*
[2]*School of Mathematics, University of Edinburgh, dpaulin@ed.ac.uk*
[3]*Department of Statistics, University of British Columbia, bouchard@stat.ubc.ca*

The bouncy particle sampler is a Markov chain Monte Carlo method based on a nonreversible piecewise deterministic Markov process. In this scheme, a particle explores the state space of interest by evolving according to a linear dynamics which is altered by bouncing on the hyperplane perpendicular to the gradient of the negative log-target density at the arrival times of an inhomogeneous poisson process (PP) and by randomly perturbing its velocity at the arrival times of a homogeneous PP. Under regularity conditions, we show here that the process corresponding to the first component of the particle and its corresponding velocity converges weakly towards a randomized Hamiltonian Monte Carlo (RHMC) process as the dimension of the ambient space goes to infinity. RHMC is another piecewise deterministic nonreversible Markov process where a Hamiltonian dynamics is altered at the arrival times of a homogeneous PP by randomly perturbing the momentum component. We then establish dimension-free convergence rates for RHMC for strongly log-concave targets with bounded Hessians using coupling ideas and hypocoercivity techniques. We use our understanding of the mixing properties of the limiting RHMC process to choose the refreshment rate parameter of BPS. This results in significantly better performance in our simulation study than previously suggested guidelines.

CONTENTS

Received March 2020; revised November 2020.
MSC2020 subject classifications. Primary 65C05, 60F17; secondary 60J25.
Key words and phrases. Bouncy particle sampler, coupling, randomized Hamiltonian Monte Carlo, weak convergence, hypocoercivity.

1. Introduction. Assume one is interested in sampling from a target probability density on $\mathbb{R}^d$ which can be evaluated pointwise up to an intractable normalizing constant. In this context one can use Markov chain Monte Carlo (MCMC) algorithms to sample from, and compute expectations with respect to the target measure. Despite their great success, standard MCMC methods, such as the ubiquitous Metropolis–Hastings algorithm, tend to perform poorly on high-dimensional targets. To address this issue, several new methods have been proposed over the past few decades. Popular alternatives include the Metropolis-adjusted Langevin algorithm (MALA) [53, 54], Hamiltonian, or Hybrid, Monte Carlo (HMC) [25] and slice sampling [44].

Recently, a novel class of nonreversible, continuous-time MCMC algorithms based on piecewise-deterministic Markov processes (PDMP) has appeared in applied probability [10, 43], automatic control [39], physics [41, 45, 49] statistics and machine learning [6, 7, 14, 15, 47, 57, 61]. Most of the current literature revolves around two piecewise-deterministic MCMC (PDMCMC) schemes: the bouncy particle sampler (BPS) [15, 49] and the zig-zag sampler [7]. A practical advantage of the BPS and zig-zag algorithms is that in many models it is possible to simulate their piecewise linear paths without time-discretization [15]. In contrast, methods based on either diffusions or Hamiltonian paths require time discretization and moreover their performance is known to collapse if the discretization is too coarse. Despite the increasing interest in these piecewise linear PDMCMC algorithms, our theoretical understanding of their properties remains limited, although a fair amount of progress has been achieved recently in establishing geometric ergodicity; see [23, 26] for BPS and [11, 30] for zig-zag. However, all of these results tend to provide convergence rates that deteriorate with the dimension and thus fail to capture the empirical performance of these PDMCMC algorithms on high-dimensional targets.

Scaling limits have become a very popular tool for analysing and comparing MCMC algorithms in high-dimensional scenarios since their introduction in the seminal paper [51]; see, for example, [5, 52]. They have been used to establish the computational complexity of the most popular MCMC algorithms, which is $O(d^2)$ for random walk Metropolis (RWM), $O(d^{4/3})$ for MALA and $O(d^{5/4})$ for HMC; here computational complexity is defined in terms of the expected squared jump distance. In this direction, the recent work of Bierkens et al. [8] has established scaling limits for both zig-zag and global BPS for high-dimensional standard Gaussian targets. They obtain the scaling limits of several finite dimensional statistics, namely the angular velocity, the log-density and the first coordinate. In this context, it is shown that zig-zag has algorithmic complexity $O(d)$ for all three types of statistics, whereas global BPS has complexity $O(d)$ for angular momentum and $O(d^2)$ for the other two types of statistics. Benefits of zig-zag over global BPS are to be expected in this scenario. Indeed, when applied to a product target, the zig-zag sampler factorises into independent components and is closely related to Local-BPS (LBPS); see [15, 49]. The standard (global) BPS studied herein and in Bierkens et al. [8], just like RWM, MALA and HMC, is an algorithm whose dynamics do not distinguish between product and nonproduct targets.

In the present paper, we also study scaling limits for BPS on a very general class of targets that greatly extends the i.i.d. scenario, and its variants, often considered in the literature; see, for example, [5, 8, 51, 52]. We concentrate on the first coordinate and its corresponding velocity in a regime which differs from the one considered in [8] in the following three ways: (a) [8] considers BPS with the location evolving at unit speed, whereas in our scenario the velocity is Gaussian, therefore with speed scaling like $\sqrt{d}$ in the dimension; (b) [8] considers scaling limits for the first coordinate of the location process only, whereas we look at both location and velocity; and finally (c) [8] rescales time with a factor d, whereas we obtain our limiting process on the natural time scale. As a result we obtain a different scaling limit which suggests that BPS has algorithmic complexity $O(d^{3/2})$ if one is interested on low-dimensional projections, at least on weakly dependent targets. This is in agreement with the empirical results reported in [15]. Given the different regimes and different objects studied in [8] and the present paper, it is not surprising that the two scaling limits differ significantly, with our bound being tighter and seemingly better at capturing the empirical behaviour of the process. In [8] the first location coordinate converges to a Langevin diffusion, whereas in the present paper the process tracking the first location and velocity components converges to a piecewise deterministic Markov process known as randomized Hamiltonian Monte Carlo (RHMC). Although the corresponding Fokker–Planck equation was studied in Dolbeault et al. [24], using a related approach to ours, RHMC was first studied in a Monte Carlo context in [14].

To the best of our knowledge, our result is the first in the literature establishing a direct link between BPS and Hamiltonian dynamics. It is our understanding that the Langevin diffusion obtained in [8] can be obtained from RHMC by a further limiting procedure similar to the *overdamped* regime of the Langevin equation. In addition, the assumptions under which our scaling limit is obtained allow much more complex dependence structures than those considered in the literature; see, for example, [3–5, 17, 51, 52, 62], where the target is assumed to factorise or to possess a hierarchical structure. In addition, in the scenario we consider all dimensions have an impact, in contrast with the Hilbert-space setting; see, for example, [38], where only a fixed, finite number of dimensions is significant.

The second part of the paper is concerned with the convergence properties of RHMC. This process was studied in [14] where it was established that it is *geometrically ergodic*. However, it is not clear whether such an approach can provide dimension independent convergence rates. The earlier work of [24] studies the corresponding Fokker–Planck equation, tracking the evolution of densities rather than conditional expectations. In recent years, there has been great success in obtaining dimension-free convergence rates of MCMC schemes for strongly log-concave targets with bounded Hessians; see, for example, [13, 20, 27, 28, 37]. In particular, in relation to HMC, the papers [13, 37] use coupling techniques to obtain convergence rates in terms of Wasserstein or total variation distances, but these usually leverage independent momentum refreshment to obtain a Markov process in the location components only. We establish here these convergence rates in weighted Wasserstein distance using coupling ideas, and also in L^2 using *hypocoercivity*; see, for example, [48, 58]. The rates we provide may generally not be the optimal ones for specific scenarios. However, the optimal rates for a specific scenario can be obtained by solving a multivariate optimisation problem. Dolbeault et al. [24] also uses hypocoercivity, albeit with a much different flavour, and does not seem to provide explicit rates. After the first version of the present paper appeared online, the approach of [24] was extended in Andrieu et al. [1] to cover several PDMPs, including BPS, zig-zag and RHMC. Even more recently, the paper [36] appeared online, proving L^2 rates for three PDMPs (BPS, zig-zag and RHMC).

The approach in [24] and [1] is quite distinct to ours. In particular [1] also obtain dimension-free bounds for RHMC under similar assumptions; their explicit rates have a complex dependency on various parameters of the problem and therefore a detailed comparison

with the explicit rates in our Theorem 5 was not performed in [1]. In Remark 11 we perform a comparison, and find that in the strongly convex and smooth setting, neither of these two approaches outperforms the other in all cases, sometimes the bound of [1] is sharper, while in other scenarios our bound is sharper. Their approach is quite general but much less direct for RHMC than ours, as they rely on generic results by Dolbeault, Mouhot and Schmeiser. The approach in [36] is entirely different from [1] and ours, using sophisticated PDE methods to analyse the Fokker–Planck equations of the PDMP directly. In Remark 11, we include a detailed comparison with our results. In general, we find that the bounds in [36] for RHMC are sharper than ours in the condition number M/m, but the constant of proportionality is not explicitly stated, and might be nontrivial to obtain reasonably small constants.

In addition the bounds of [1] and [36] for BPS suggest that its computational cost scales like $O(d^2)$. This seems to capture the worst case scenario and agrees, for example, with results [8] for the log-density of the target, which recommends scaling the refreshment rate with the dimension. Our results suggest that when one is interested in low-dimensional projections, then it is computationally more efficient to not scale the refreshment rate with the dimension, achieving computational cost of order $O(d^{3/2})$. Empirical results in Section 2.5 seem to suggest that this may also be the case for certain classes of functions depending on all the coordinates, such as the sum of all coordinates. A common scenario where this type of scaling limit is extremely relevant is, for example, that of Bayesian inference where typically one may only be interested in estimating the posterior means, variances and covariances of the high-dimensional state components (this is a set of one- and two- dimensional marginals). Finally, it is intuitively clear that the log-density will not mix well in a high-dimensional target for the global BPS; see [8] for a detailed study. We conjecture that the functions that exhibit this type of behaviour form a low-dimensional subspace of $L^2(\pi)$. Recently [9] has obtained very detailed results on the whole spectrum of the one-dimensional zig-zag process, it would be interesting if similar results could be obtained for BPS in high-dimensional scenarios.

Apart from the intrinsic interest of the RHMC process, our motivation for studying its convergence rates is as follows. In the scaling literature for MCMC the limiting processes are usually Langevin diffusions. These have very well understood convergence rates which, at least under additional assumptions, are dimension-free. Therefore, in high-dimensions the cost of running the (time-rescaled) algorithm serves as a proxy for its computational complexity. In our case, the algorithm ran on its natural time scale converges to RHMC, which as we establish here, also enjoys dimension-free convergence rates under appropriate assumptions. Therefore the cost of running BPS for a unit of process time serves as a proxy for its algorithmic complexity.

The next section contains the statements of the main results of the paper along with necessary notation and definitions. The remaining sections contain the proofs of the main results.

2. Main results.

2.1. *Notation.* For $x \in \mathbb{R}$, let $x_+ = \max\{x, 0\}$. Let $k \geq 1$. For vectors $u, v \in \mathbb{R}^k$ we write $|v|$ and (u, v) for the Euclidean norm and inner product respectively. For matrices $A, B \in \mathbb{R}^{k\times k}$ we write $A \preceq B$ if $B - A$ is positive-definite. For a function $f : \mathbb{R}^k \mapsto \mathbb{R}$ we write ∇f, $\nabla^2 f$ for its (weak) gradient and Hessian respectively. When considering functions $f = f(a, b)$, where $a, b \in \mathbb{R}^k$, that is, $f : \mathbb{R}^{2k} \mapsto \mathbb{R}$, we will write $\nabla_a f$, $\nabla_b f$ to denote the gradient with respect to the variables $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^k$ respectively. Allowing a slight abuse of notation, for vector valued functions $f : \mathbb{R}^d \to \mathbb{R}^k$, we will also write ∇f for the Jacobian matrix of derivatives.

For $\mathcal{Z} = \mathbb{R}^k$, with $k \in \mathbb{N}$, let $C_0(\mathcal{Z})$ denote the space of continuous functions $f : \mathcal{Z} \mapsto \mathbb{R}$ that vanish at infinity. Recall that $C_0(\mathcal{Z})$ is a Banach space with respect to the $\|\cdot\|_\infty$ norm,

which is defined as usual through $\|f\|_\infty = \sup|f|$. Also let $C_c^\infty(\mathcal{Z})$ be the space of infinitely differentiable functions $f : \mathcal{Z} \mapsto \mathbb{R}$ with compact support.

For a measure π on $\mathcal{Z}$, we will write $L^2(\pi)$ for the usual, real Hilbert space, and $\langle\cdot,\cdot\rangle$, $\|\cdot\|$ to denote the inner product and norm in $L^2(\pi)$ respectively, whereas $L_0^2(\pi)$ will denote the orthogonal complement of the constant functions, that is, functions with mean zero under the distribution π. Finally for $f : \mathcal{Z} \to \mathbb{R}^d$ and $g : \mathcal{Z} \to \mathbb{R}^d$, with $d \geq 1$, we also write

$$\langle f, g\rangle = \int \pi(\mathrm{d}z)\big(f(z), g(z)\big).$$

It will be clear from the context whether $\langle\cdot,\cdot\rangle$ is applied to $\mathbb{R}-$ or $\mathbb{R}^d$-valued functions. We also define

$$H^1 := H^1(\pi) := \big\{h \in L_0^2(\pi) : \nabla_x h, \nabla_v h \in L^2(\pi)\big\},$$

the Sobolev space of centred functions in $L^2(\pi)$ with weak derivatives in $L^2(\pi)$ and for $f, g \in H^1(\pi)$ we will denote the inner product and norm on $H^1(\pi)$ with $\langle\!\langle\cdot,\cdot\rangle\!\rangle_{H^1(\pi)}$ and $\|\cdot\|_{H^1(\pi)}$ respectively, where

$$\langle\!\langle f, g\rangle\!\rangle_{H^1(\pi)} = \langle\nabla_x f, \nabla_x g\rangle + \langle\nabla_v f, \nabla_v g\rangle.$$

2.2. *The bouncy particle sampler.* Let $\mathcal{Z} := \mathbb{R} \times \mathbb{R}$ and for $n \geq 1$, define the Borel probability measure $\pi_n(\mathrm{d}z)$ on $\mathcal{Z}^n$ with density w.r.t. Lebesgue measure given by

$$\pi_n(z) = \pi_n(\boldsymbol{x}, \boldsymbol{v}) \propto \exp\big\{-U_n(\boldsymbol{x}) - |\boldsymbol{v}|^2/2\big\}, \quad (\boldsymbol{x}, \boldsymbol{v}) \in \mathcal{Z}^n,$$

where $U_n : \mathbb{R}^n \mapsto \mathbb{R}_+$ is a potential.

For $(\boldsymbol{x}, \boldsymbol{v}) \in \mathcal{Z}^n$, define

$$R_n(\boldsymbol{x})\boldsymbol{v} := \boldsymbol{v} - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\nabla U_n(\boldsymbol{x}). \tag{2.1}$$

The vector $R_n(\boldsymbol{x})\boldsymbol{v}$ can be interpreted as a Newtonian collision on the hyperplane orthogonal to the gradient of the potential U_n, hence the interpretation of $\boldsymbol{x}$ as a position, and $\boldsymbol{v}$, as a velocity.

The bouncy particle sampler (BPS), first introduced in [49] and in a statistical context in [15], defines a π_n-invariant, nonreversible, piecewise deterministic Markov process (PDMP) $\{\boldsymbol{Z}_n(t) : t \geq 0\} = \{(\boldsymbol{X}(t), \boldsymbol{V}(t)) : t \geq 0\}$ taking values in $\mathcal{Z}^n$ whose generator $\mathcal{A}_n$, for smooth enough functions $f : \mathcal{Z}^n \mapsto \mathbb{R}$, is given by

$$\begin{aligned}\mathcal{A}_n f(\boldsymbol{x}, \boldsymbol{v}) = {}& \big(\nabla f(\boldsymbol{x}, \boldsymbol{v}), \boldsymbol{v}\big) + \max\big\{0, \big(\nabla U_n(\boldsymbol{x}), \boldsymbol{v}\big)\big\}\big[\mathfrak{R}_n f(\boldsymbol{x}, \boldsymbol{v}) - f(\boldsymbol{x}, \boldsymbol{v})\big] \\ & + \lambda_{\mathrm{ref}}\big[Q_{\alpha,n} f(\boldsymbol{x}, \boldsymbol{v}) - f(\boldsymbol{x}, \boldsymbol{v})\big],\end{aligned}$$

where

$$\begin{aligned}\mathfrak{R}_n f(\boldsymbol{x}, \boldsymbol{v}) &:= f\big(\boldsymbol{x}, R_n(\boldsymbol{x})\boldsymbol{v}\big), \\ Q_{\alpha,n} f(\boldsymbol{x}, \boldsymbol{v}) &:= \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \mathrm{e}^{-|\boldsymbol{\xi}|^2/2} f\big(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\,\mathrm{d}\boldsymbol{\xi},\end{aligned}$$

for $0 \leq \alpha < 1$ and a positive refreshment rate $\lambda_{\mathrm{ref}} > 0$. We also write $\boldsymbol{Z}_n(t) = (Z_n^{(1)}(t), \ldots, Z_n^{(n)}(t))$ where $Z_n^{(k)}(t) = (X_n^{(k)}(t), V_n^{(k)}(t)) \in \mathcal{Z}$ is the kth component. The original formulation of the BPS algorithm corresponds to $\alpha = 0$, that is, refreshment occurs independently. The generalization $\alpha > 0$ [57] consists in refreshments that are performed according to an auto-regressive process.

2.3. *Randomized Hamiltonian Monte Carlo.* We define here RHMC as this is the process we will obtain as the weak limit of $Z_n^{(1)}(t) = (X_n^{(1)}(t), V_n^{(1)}(t)) \in \mathcal{Z}$ as $n \to \infty$. Define the Hamiltonian

$$H(x, v) = W(x) + |v|^2/2, \tag{2.2}$$

for $(x, v) \in \mathcal{Z}$ and the corresponding probability density on $\mathcal{Z}$

$$\pi(x, v) = \bar{\pi}(x) \cdot \psi(v) \propto \exp\{-W(x) - |v|^2/2\}. \tag{2.3}$$

The *Hamiltonian dynamics* associated to (2.2) is an ordinary differential equation in $\mathcal{Z}$ of drift $(\nabla_v H, -\nabla_x H) = (v, -\nabla W)$. The RHMC process, denoted $\{Z_t : t \geq 0\}$, can then be defined following Davis ([22], Section 24) as a PDMP with deterministic dynamics given by Hamiltonian dynamics with respect to H, fixed jump rate $\lambda_{\text{ref}} > 0$ and jump kernel

$$Q_\alpha f(x, v) := \frac{1}{(2\pi)^{n/2}} \int \mathrm{e}^{-|\xi|^2/2} f\big(x, \alpha v + \sqrt{1-\alpha^2}\xi\big)\, \mathrm{d}\xi, \tag{2.4}$$

for some $0 \leq \alpha < 1$. We will write $\{P^t : t \geq 0\}$ for the semigroup corresponding to $\{Z_t : t \geq 0\}$, that is,

$$P^t f(z) = \mathbb{E}\big[f(Z_t) | Z_0 = z\big].$$

It has been shown, [14], that RHMC admits π as an invariant distribution.

It can also be shown that for $f \in C_c^\infty(\mathcal{Z})$, the generator of the semigroup $\{P^t : t \geq 0\}$ is given by

$$\mathcal{A} f(x, v) = (\nabla_x f, v) - (\nabla_v f, \nabla W) + \lambda_{\text{ref}}\big[Q_\alpha f(x, v) - f(x, v)\big]. \tag{2.5}$$

The refreshment is done in an auto-regressive manner. From now on, we will restrict ourselves for BPS and RHMC to $0 < \alpha < 1$. The reason for using $\alpha > 0$ is that it allows us to establish the Feller property which greatly simplifies the rest of the proofs. Since the autoregressive process mixes exponentially fast there is no loss in terms of mixing potentially at the cost of more frequent refreshments, something which has also been observed empirically.

REMARK 1. As one of the referees kindly suggested, one may attempt to couple the process with $\alpha = 0$ with the process at $\alpha_n = o(1)$ in order to extend the result to the case $\alpha = 0$. Unfortunately, the obvious line of attack requires one to couple the full n-dimensional velocity vector at refreshments, so the maximal coupling deteriorates with the growing dimension; this approach would require a quantitative version of Theorem 1. It is possible that a different coupling can be used, but we did not pursue this issue further.

2.4. *Main results.*

2.4.1. *RHMC as scaling limit of BPS.* Before stating our weak convergence result, we will make some assumptions. We consider a sequence of targets π_n on $\mathbb{R}^n \times \mathbb{R}^n$ where $\pi_n(\boldsymbol{x}, \boldsymbol{v}) = \bar{\pi}_n(\boldsymbol{x})\psi_n(\boldsymbol{v})$, with ψ_n a standard n-dimensional Gaussian and $\bar{\pi}_n(\boldsymbol{x}) = \exp[-U_n(\boldsymbol{x})]$ for a sequence of potentials $U_n : \mathbb{R}^n \to [0, \infty)$ satisfying the following assumptions.

ASSUMPTION 1. The potential $U_n \in C^2(\mathbb{R}^n)$ is m-strongly convex with M-Lipschitz gradient

$$mI \preceq \nabla^2 U_n(\boldsymbol{x}) \preceq MI, \quad x \in \mathbb{R}^n, \text{ with } 0 < m \leq M < \infty, \tag{2.6}$$

and U_n achieves its minimum at 0, that is, $U_n(0) = 0$ and $\nabla U_n(0) = 0$.

ASSUMPTION 2. The marginal density of the first component of $\bar{\pi}_n$ is fixed and is given by

$$f(x) := \int \bar{\pi}_n(x, \boldsymbol{x}_{2:n})\, \mathrm{d}\boldsymbol{x}_{2:n}.$$

We assume that $f(x) = \exp[-W(x)]$ for a potential $W \in C^\infty(\mathbb{R}; [0, \infty))$ such that $\lim_{|x|\to\infty} W(x) = \infty$ and

$$\int \mathrm{e}^{-W(x)}\big(|W''(x)| + |W'(x)|^2\big)\, \mathrm{d}x < \infty.$$

Let $\{Z_t : t \geq 0\}$ be the RHMC process with potential W and write $\mathcal{A}$ for its generator given in (2.5). The following theorem is our first main result.

THEOREM 1. *Suppose Assumptions 1 and 2 hold, $0 < \alpha < 1$, $\lambda_{\mathrm{ref}} > 0$ and that the BPS process $\{\boldsymbol{Z}_n(t) : t \geq 0\}$ is initialized at stationarity, that is, $\boldsymbol{Z}_n(0) \sim \pi_n$. Then the process $\{Z_n^{(1)}(t) : t \geq 0\}$ corresponding to the first location and velocity components of the BPS process converges weakly to the RHMC process $\{Z_t : t \geq 0\}$ as $n \to \infty$.*

We would like to stress that there is no time-rescaling in the above result, and that the sequence of targets is not assumed to factorise into independent components, or to converge towards an infinite dimensional measure as the dimension $n \to \infty$.

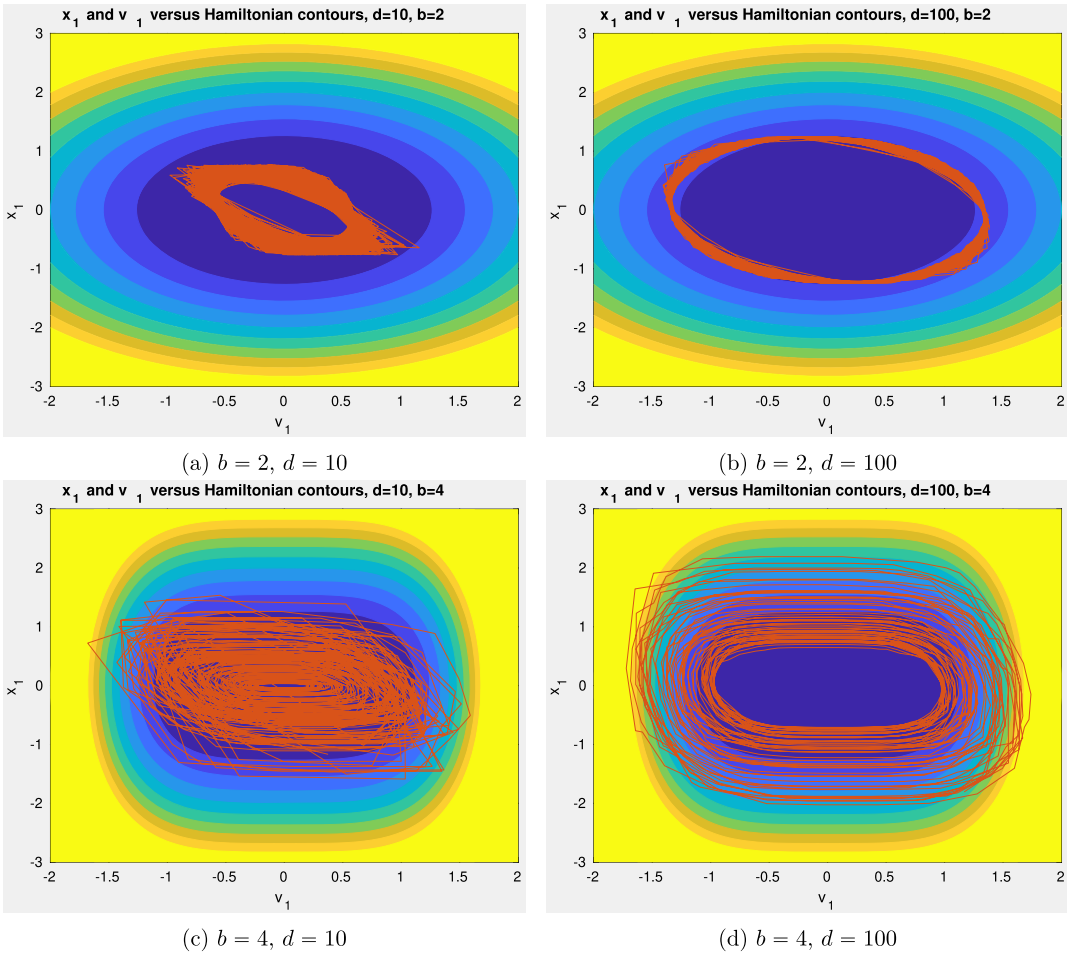
REMARK 2. Notice that Assumption 1 allows for the standard scenario where the target factorises in n i.i.d. copies which corresponds to $U_n(\boldsymbol{x}) = \sum_{i=1}^n U(x_i)$, for an m-strongly convex potential $U \in C^2(\mathbb{R})$ with $U'' \leq M$. Indeed in this case the Hessian matrix is diagonal and given by $(\nabla^2 U_n(\boldsymbol{x}))_{i,j} = U''(x_i)\delta_{i,j} \geq 0$. This was the scenario considered in an earlier version of the present paper. In fact, in this i.i.d. scenario the convexity assumption can be removed and the upper bound on U'' can be replaced by an upper bound on $U^{(k)}$ for any k, at the expense of additional technical complexity.

REMARK 3. From the proof (in particular, the bounds (3.14), (3.15), (3.16), (3.17), (3.18)) it is clear that the result remains true when m, M in Assumption 1 are allowed to depend on n, if in addition we assume that

$$\begin{aligned} &m_n n \to \infty, \qquad \frac{M_n}{m_n} = o\big(n^{1/4}\big), \qquad \frac{M_n^3}{m_n^{3/2}} = o\big(n^{1/2}\big), \\ &\frac{M_n^3}{m_n^2} = o\big(n^{1/2}\big), \qquad \frac{M_n^2}{m_n} = o\bigg(\frac{n^{1/2}}{(\log(n))^{1/2}}\bigg). \end{aligned} \tag{2.7}$$

REMARK 4. Scaling limits for non i.i.d. targets have appeared in the past. Bédard [3] studied targets that factorise into independent, but not identically distributed components; results on hierarchical targets can be found in [4, 62] and references therein. The case of Gibbs measures with finite range interactions was studied in [17]. Mattingly et al. [38] proved that a sequence of algorithms targeting finite dimensional projections of a measure admitting a density with respect to a reference Gaussian measure on a Hilbert space, converge to a Hilbert space-valued stochastic differential equation.

REMARK 5. To illustrate Theorem 1 in Figure 1 we have plotted the paths of the BPS process and the equi-energy contours of the Hamiltonian corresponding to the deterministic dynamics of RHMC. The target distribution has potential $U(\boldsymbol{x}) = \sum_{i=1}^n |x_i|^b/2$ and we have



(a) $b = 2$, $d = 10$ (b) $b = 2$, $d = 100$

(c) $b = 4$, $d = 10$ (d) $b = 4$, $d = 100$

FIG. 1. *Convergence of the BPS process to RHMC in high dimensions for* $U(x) = |x|^b/2$.

tested two values of b, $b = 2$ (Gaussian) and $b = 4$. These figures show the first coordinate of the position and velocity vectors. As we can see, as the dimension increases, the paths of BPS indeed appear more and more similar to the contours of the Hamiltonian.

REMARK 6. Theorem 1 can be straightforwardly extended to any fixed, finite number of coordinates $d > 1$. In this case the limiting process will be RHMC in $\mathbb{R}^d \times \mathbb{R}^d$ with respect to the potential $W : \mathbb{R}^d \to \mathbb{R}$ given by

$$W(\boldsymbol{x}) = -\log \int \bar{\pi}_n(\boldsymbol{x}, \boldsymbol{x}_{d+1:n})\, \mathrm{d}\boldsymbol{x}_{d+1:n}, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

with W satisfying a d-dimensional version of Assumption 2.

Sketch of proof. The full proof of this result is quite lengthy and will be given in Section 3. However, we now give the key idea without going into technical details, for the simpler i.i.d. scenario where $U_n(\boldsymbol{x}) = \sum_{i=1}^n U(x_i)$, for $U : \mathbb{R} \mapsto [0, \infty)$. In this case the limiting process has potential $W \equiv U$. Under the assumptions of Theorem 1 let $\boldsymbol{Z}_n = (\boldsymbol{X}_n, \boldsymbol{V}_n) \sim \pi_n$ and let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be smooth. We now consider the generator $\mathcal{A}_n$ of BPS targeting π_n and the generator $\mathcal{A}$ of RHMC targeting π, the marginal of the first location momentum pair under π_n, applied to the function f. By inspecting $\mathcal{A}_n(f)$, $\mathcal{A}(f)$ we find that the terms corresponding to the deterministic flow of BPS and the refreshment events coincide exactly with

corresponding terms in $\mathcal{A}(f)$. We therefore only have to consider the term corresponding to the "bounce events", that is,

$$\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\}\Big[f\Big(X_1, V_1 - 2\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}U'(X_1)\Big) - f(X_1, V_1)\Big],$$

and show that on average it is close to $-(\nabla_v f, \nabla U) = -U'(X_1)\partial_v f(X_1, V_1)$.

To see why this is true, after a Taylor expansion we can see that the bounce part of the BPS generator is close to

$$-2\max\Big\{0, \frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|}\Big\}\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|}\partial_v f(X_1, V_1)U'(X_1).$$

Looking closer one can see that

$$\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|} = \frac{\sum_{i=1}^n U'(X_i)V_i}{\sqrt{\sum_{i=1}^n U'(X_i)^2}},$$

and since the $(V_i)_i$ are i.i.d. standard Gaussians it is easily seen that

$$\frac{\sum_{i=1}^n U'(X_i)V_i}{\sqrt{\sum_{i=1}^n U'(X_i)^2}}\Big|(X_i)_{i=1}^n \sim \mathcal{N}(0, 1).$$

It now seems plausible that, letting $\xi \sim \mathcal{N}(0, 1)$, we have

$$\begin{aligned}&\mathbb{E}\Big\{\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\}\Big[f\Big(X_1, V_1 - 2\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}U'(X_1)\Big) - f(X_1, V_1)\Big]\Big|X_1, V_1\Big\}\\ &\quad\approx -2\mathbb{E}\big[\max\{0, \xi\}\xi\big]\partial_v f(X_1, V_1)U'(X_1) = -\partial_v f(X_1, V_1)U'(X_1) = -(\nabla_v f, \nabla U).\end{aligned}$$

2.4.2. *Dimension-free convergence rates for RHMC.* We consider the RHMC process on the target

$$\pi(\boldsymbol{x}, \boldsymbol{v}) = \bar{\pi}(\boldsymbol{x}) \cdot \psi(\boldsymbol{v}) \propto \exp\big\{-U(\boldsymbol{x}) - |\boldsymbol{v}|^2/2\big\},$$

defined on $\mathcal{Z} := \mathbb{R}^d \times \mathbb{R}^d$ for $\bar{\pi}(\cdot)$ a strongly log-concave target distribution on $\mathbb{R}^d$ having a potential with bounded Hessian. This is a standard assumption adopted in [13, 20, 27, 28, 37].

ASSUMPTION 3. Assume that $U \in C^2(\mathbb{R}^d)$ and that for some $0 < m < M$, and all $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^d$

$$m(\boldsymbol{v}, \boldsymbol{v}) \le \big(\boldsymbol{v}, \nabla^2 U(\boldsymbol{x})\boldsymbol{v}\big) \le M(\boldsymbol{v}, \boldsymbol{v}). \tag{2.8}$$

The following proposition, whose proof is given in the Appendix, shows that the expected number of bounces per unit time for BPS in stationary distribution is $O(\sqrt{d})$.

PROPOSITION 2. *Suppose that $\bar{\pi}(\boldsymbol{x}) \propto \exp(-U(\boldsymbol{x}))$ is a probability density on $\mathbb{R}^d$. Then the BPS process on $\mathcal{Z}$ targeting $\bar{\pi} \otimes \psi$ and initialized at stationarity, has the following expected number of bounces per unit time*:

$$\Lambda_b := \mathbb{E}_{X\sim\pi, V\sim N(0,\mathbf{I}_d)}\big[\big(\nabla U(\boldsymbol{X}), \boldsymbol{V}\big)_+\big],$$

for any choice of refreshment rate λ_{ref} and auto-regressive parameter α. Moreover, if $\bar{\pi}$ satisfies Assumption 3, then we have

$$\frac{\sqrt{m(d - 1/2)}}{\sqrt{2\pi}} \le \Lambda_b \le \frac{\sqrt{Md}}{\sqrt{2\pi}}. \tag{2.9}$$

Wasserstein distance. For $t \geq 0$, let $Z^{(1)}(t) = (X^{(1)}(t), V^{(1)}(t))$ denote a path of the RHMC process. We couple this with another path $Z^{(1)}(t) = (X^{(2)}(t), V^{(2)}(t))$ such that their refreshments happen simultaneously and the same multivariate normal random variables are used for updating their velocities. Therefore the difference between the paths $Z^{(1)}(\cdot)$ and $Z^{(2)}(\cdot)$ stems only from the different initialisations. Then the coupled process $(Z^{(1)}(t), Z^{(2)}(t))$ is Markov and we write $L_{1,2}$ for the corresponding generator. Notice that the 2×2 real valued matrix

$$A := \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \tag{2.10}$$

is positive definite, denoted $A \succeq 0$, if and only if $a > 0$, $c > 0$ and $b^2 < ac$. For such a matrix, let

$$\begin{aligned} &d_A^2\big(Z_1(t), Z_2(t)\big) \\ &\quad := a\big\|X^{(2)}(t) - X^{(1)}(t)\big\|^2 + 2b\big\langle X^{(2)}(t) - X^{(1)}(t), V^{(2)}(t) - V^{(1)}(t)\big\rangle \\ &\qquad + c\big\|V^{(2)}(t) - V^{(1)}(t)\big\|^2 \end{aligned}$$

denote a distance function called weighted distance. It is equivalent up to constant multiplicative factors to the standard Euclidean distance on $\mathbb{R}^{2d}$ and the standard Euclidean distance corresponds to the special case $a = 1$, $b = 0$, $c = 1$. However, due to the effect of the generator $L_{1,2}$ on $d_A^2(Z_1(t), Z_2(t))$, it will never be a contraction when $b = 0$, and thus weighting this distance is essential for obtaining convergence rates. Note that for every $p \geq 1$, the W_p-Wasserstein distance of two distributions ν_1, ν_2 on $\mathbb{R}^{2d}$ is defined as $W_p(\nu_1, \nu_2) = (\inf_{X_1 \sim \nu_1, X_2 \sim \nu_2} \mathbb{E}(|X_1 - X_2|^p))^{1/p}$, where the infimum is taken over all couplings with marginals ν_1 and ν_2.

Our main result in this section is the following.

THEOREM 3. *Suppose that* $0 \leq \alpha < 1$, *Assumption 3 holds and let*

$$\lambda_{\mathrm{ref}} = \frac{1}{1-\alpha^2}\left(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}}\right), \qquad \mu = \frac{(1+\alpha)m}{\sqrt{M+m}} - \frac{\alpha m^{3/2}}{2(M+m)}.$$

Then there exist constants a, b *and* c *depending on* m, M *and* α, *stated explicitly in* (4.8), *such that the corresponding matrix* A *is positive definite, and for any* $t \geq 0$ *we have*

$$L_{1,2} d_A^2\big(Z_1(t), Z_2(t)\big) \leq -\mu \cdot d_A^2\big(Z_1(t), Z_2(t)\big). \tag{2.11}$$

This directly implies that for any initial distribution ν *on* $\mathbb{R}^{2n}$, *for all* $t \geq 0$, *we have the following bounds on the* 2*-Wasserstein distance to the stationary distribution*:

$$W_2\big(P^t \nu, \pi\big)^2 \leq C_2 e^{-\mu t} W_2(\nu, \pi)^2, \tag{2.12}$$

for $C_2 = \frac{a+c+\sqrt{(a+c)^2-4(ac-b^2)}}{a+c-\sqrt{(a+c)^2-4(ac-b^2)}}$. *Moreover, for every* $f \in L_0^2(\pi)$, *for all* $t \geq 0$

$$\big\|P^t f\big\|^2 \leq \min\big(Ce^{-\mu t}, 1\big)\|f\|^2, \tag{2.13}$$

where $C = \frac{ac+b^2+2\sqrt{acb^2}}{ac-b^2}$.

REMARK 7. Due to the nonreversibility of RHMC, the convergence rates in Wasserstein distance do not directly imply bounds on the asymptotic variance for every function in $L^2(\pi)$, but only for Lipschitz functions. The argument for extending this contraction rate to all of $L^2(\pi)$, can be found in the second half of the proof of Theorem 5. This is based on the fact that Lipschitz functions are dense in $L^2(\pi)$.

REMARK 8. These results seem to suggest that choosing α close to 1 increases the convergence rate μ approximately by a factor of 2, at the expense of a higher refreshment rate. Hence in practice some tradeoff needs to be made between additional computational cost and the increased convergence rate. By Proposition 2, we know that the rate of bounces according to the stationary distribution is at least $\frac{\sqrt{m(d-1/2)}}{\sqrt{2\pi}}$, which will be significantly higher than the rate $\frac{1}{1-\alpha^2}(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}})$ in high dimensions, provided that $\sqrt{\frac{M+m}{m} \cdot \frac{1}{d-1/2}} \cdot \frac{1}{1-\alpha^2} \ll 1$. The choice $\alpha = 0.9$ is reasonable in most scenarios.

REMARK 9. We have been able to verify using Mathematica that if $M/m \geq 5$, and we choose $\lambda_{\text{ref}} \leq \frac{1}{2} \cdot \frac{1}{1-\alpha^2}(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}})$ (half the value recommended in Theorem 3), then the contraction (2.11) cannot hold for any choice of a, b and c. In general, if we choose $\lambda_{\text{ref}} = \frac{r}{1-\alpha^2}(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}})$ for some $r > 1$ (i.e., r times the refreshment rate recommended in Theorem 3), then it seems based on extensive experiments that the rate $\mu = \frac{1}{r}(\frac{(1+\alpha)m}{\sqrt{M+m}} - \frac{\alpha m^{3/2}}{2(M+m)})$ is attained (i.e., μ drops by a factor r); no values of a, b and c result in double the same rate. Obtaining a formula that describes sharp rates μ for a general choice of λ_{ref} seems difficult with our method of proof, as the inequalities that need to be checked in this case depend on many variables, and the calculations become intractable. We include in the electronic supplementary material Mathematica code that checks, for given values of m, M, α, λ_{ref}, μ, whether there exist a, b and c such that (2.13) holds, and returns a possible choice of these parameters if they exist.

As we shall see in the next proposition, it is possible to obtain faster convergence rates, that is, larger μ, for Gaussian target distributions. For this result, we consider a *weighted* distance of the form

$$d_D^2\big(Z_1(t), Z_2(t)\big) := \big\langle Z_2(t) - Z_1(t), D\big(Z_2(t) - Z_1(t)\big)\big\rangle, \tag{2.14}$$

where D is a real valued $2d \times 2d$ positive definite matrix.

PROPOSITION 4. *Suppose that $\bar{\pi}$ is Gaussian and its inverse covariance matrix H satisfies $mI \preceq H \preceq MI$. Let*

$$\lambda_{\text{ref}} = \frac{2\sqrt{m}}{1-\alpha}, \qquad \mu = \frac{\sqrt{m}}{3}.$$

Then there exists a $2d \times 2d$ real valued matrix D such that for any $t \geq 0$ we have

$$L_{1,2} d_D^2\big(Z_1(t), Z_2(t)\big) \leq -\mu \cdot d_D^2\big(Z_1(t), Z_2(t)\big). \tag{2.15}$$

Moreover, for every $f \in L_0^2(\pi)$, we have

$$\big\| P^t f \big\|^2 \leq \min\big(Ce^{-\mu t}, 1\big) \| f \|^2, \tag{2.16}$$

where $C = \frac{ac+b^2+2\sqrt{acb^2}}{ac-b^2}$.

Hypocoercivity. Our next convergence result is based on the hypocoercivity approach; see, for example, [24, 32, 42, 55, 58]. Our result will be stated in terms of the modified Sobolev norm $\langle\langle h, h\rangle\rangle^{1/2}$, where

$$\langle\langle h, h\rangle\rangle := a\|\nabla_v h\|^2 - 2b\langle \nabla_x h, \nabla_v h\rangle + c\|\nabla_x h\|^2, \tag{2.17}$$

which again for $a, c > 0$ and $b^2 < ac$ defines a norm equivalent to the H^1 norm. In particular following the calculations in [58], by Young's inequality we get

$$\left(1 + \frac{|b|}{\sqrt{ac}}\right)\big[a\|\nabla_v h\|^2 + c\|\nabla_x h\|^2\big] \geq \langle\langle h, h\rangle\rangle \geq \left(1 - \frac{|b|}{\sqrt{ac}}\right)\big[a\|\nabla_v h\|^2 + c\|\nabla_x h\|^2\big].$$

By the Efron–Stein–Steele inequality [56] and the fact that $\pi(x, v) = \overline{\pi}(x)\psi(v)$ is the product of two independent distributions, we have

$$\|h\|^2 = \mathrm{Var}_\pi(h) \le \mathrm{Var}_\psi\big(\mathbb{E}_{\overline{\pi}}(h)\big) + \mathrm{Var}_{\overline{\pi}}\big(\mathbb{E}_\psi(h)\big),$$

for any $h \in L_0^2(\pi)$. Now by using the Poincaré inequality [16] and the strong log-concavity of the distributions $\overline{\pi}$ and ψ, it is not difficult to show that

$$a\|\nabla_v h\|^2 + c\|\nabla_x h\|^2 \ge a \cdot 1 \cdot \mathrm{Var}_\psi\big(\mathbb{E}_{\overline{\pi}}(h)\big) + c \cdot m \cdot \mathrm{Var}_{\overline{\pi}}\big(\mathbb{E}_\psi(h)\big) \ge \min(a, cm)\|h\|^2.$$

Therefore convergence in the $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ norm implies convergence in $L_0^2(\pi)$.

THEOREM 5. *Suppose that Assumption 3 holds and let $\alpha \in [0, 1)$ and*

$$\lambda_{\mathrm{ref}} = \frac{1}{1-\alpha^2}\left(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}}\right), \qquad \mu = \frac{(1+\alpha)m}{\sqrt{M+m}} - \frac{\alpha m^{3/2}}{2(M+m)}.$$

Then there exist constants a, b, c depending on m, M and α such that $a > 0$, $c > 0$, $b^2 < ac$, and for every $f \in \mathcal{D}(B) \subset H^1(\pi) \subset L_0^2(\pi)$, with $B, \mathcal{D}(B)$ as defined in (5.1),

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\!\langle P^t f, P^t f \rangle\!\rangle \le -\mu \langle\!\langle P^t f, P^t f \rangle\!\rangle. \tag{2.18}$$

Moreover, for every $f \in L_0^2(\pi)$ and $t \ge 0$, we have

$$\|P^t f\|^2 \le \min\big(Ce^{-\mu t}, 1\big)\|f\|^2, \tag{2.19}$$

where $C = \frac{ac+b^2+2\sqrt{acb^2}}{ac-b^2}$.

REMARK 10. Although (2.18) only implies variance bounds for functions in H^1, we are able to extend this to functions in $L^2(\pi)$ in the second half of the proof of Theorem 5, given in Section 5.2. As our rates are the same as in Theorem 3, the optimal choice of α can be done as discussed in Remark 8.

Since the first-coordinate process of BPS converges to RHMC, whose mixing we established above, in the natural time-scale the computational cost of running BPS for one time unit serves as a proxy for its algorithmic complexity. This cost is proportional to the number of total events per time unit, including bounces and refreshments. Proposition 2 shows that the expected number of bounces per unit time under Assumption 3 is at least $\frac{\sqrt{m(d-1/2)}}{2\sqrt{\pi}}$, which is much larger than the expected number of refreshments (λ_{ref}) if the refreshment rate is chosen as recommended by Theorems 3 and 5 (as long as $M/m \ll d$ and α is not too close to 1). Therefore in these cases it is justified to choose λ_{ref} in order to maximize the contraction rate μ of the limiting RHMC process.

Since each bounce has a computational cost of order $O(1)$ in terms of gradient evaluations, our results suggests that BPS scales like $O(d^{1/2})$ in gradient evaluations under our assumptions. This is the scaling observed in the simulations presented in the next section.

REMARK 11. We state here the rates for RHMC obtained by [1] and [36] under the same set of assumptions on the potential, that is, $mI_d \preceq \nabla^2 U(\mathbf{x}) \preceq MI_d$. Both papers show L^2 bounds of the form

$$\|P^t f\| \le Ce^{-\mu t}\|f\| \quad \text{for every } f \in L_0^2(\pi).$$

The convergence rate μ in [1] in this setting is shown to satisfy the inequality $\alpha(\epsilon_0) \le \mu \le 3\alpha(\epsilon_0)$. After some calculations with Mathematica, we were able to show that

$$\frac{m^2}{30} \le \alpha(\epsilon_0) \le \frac{m^2}{5} \quad \text{for } 0 < m < 1 \quad \text{and} \quad 0.03 \le \alpha(\epsilon_0) \le 0.11 \quad \text{for } m > 1,$$

when the optimal choice of refreshment rate is chosen as

$$\lambda_{\mathrm{ref}}^{\mathrm{opt}} = \frac{8 - 2\sqrt{2} + 4\sqrt{3}}{\sqrt{2}} \approx 8.5583.$$

Assuming $\alpha = 0$ (no autoregressive part in the velocity refreshments), our results yield

$$\mu = \frac{m}{\sqrt{M+m}} \quad \text{for the choice } \lambda_{\mathrm{ref}} = 2\sqrt{M+m} - \frac{m}{\sqrt{M+m}}.$$

We can see that for large values of M/m, the convergence rate of [1] is sharper, while for smaller values, our rates are sharper. We note that the conditions in [1] are quite general, and only require a Poincaré inequality, hence they are applicable even without strong convexity. [36] shows that for RHMC, the convergence rate is $\mu = \Theta(\frac{m\lambda_{\mathrm{ref}}}{(\sqrt{(m)}+\lambda_{\mathrm{ref}})^2})$, which is maximized when $\lambda_{\mathrm{ref}} = \Theta(\sqrt{m})$, yielding $\mu = \Theta(\sqrt{m})$. The dependence of these results on the parameters m, M improves upon [1] and our paper, but the constant of proportionality is not known.

In the case of BPS, both [1] and [36] shows rates of the form $\mu = \Theta(\sqrt{d})$. The dependence on the parameters m and M is sharper in [36] compared to [1], but the constant of proportionality is unknown. In contrast with these results, our high-dimensional limit argument (Theorem 1) shows that for functions that only depend on a single coordinate (or on a fixed number of coordinates), in high dimensions, the convergence occurs according to a dimension independent rate μ as long as we choose the refreshment rate appropriately, at $\lambda_{\mathrm{ref}} = \Theta(1)$. This is useful in particular for situations where we are interested in estimating the posterior mean.

2.5. *Empirical results for different functions.* In this section, we show some simulation results about the computational cost of the BPS for a d-dimensional standard normal target, and seven different test functions defined as follows:

$$f_1(x) = x_1 \quad \text{(first coordinate)},$$

$$f_2(x) = \sum_{i=1}^{d} x_i \quad \text{(sum of all coordinates)},$$

$$f_3(x) = \sum_{i=1}^{d-1} \sin(x_i + x_{i+1}) \quad \text{(a sum of sines depending on two component each)},$$

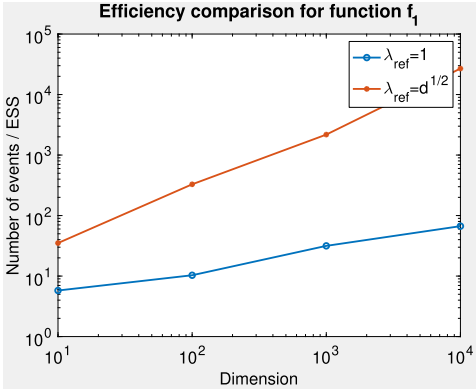
$$f_4(x) = |x| \quad \text{(radius)},$$

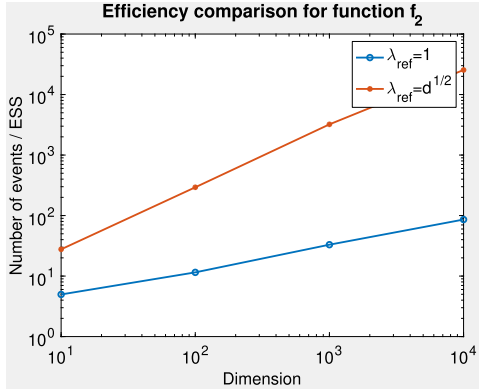
$$f_5(x) = \frac{|x|^2}{2} = \sum_{i=1}^{d} \frac{x_i^2}{2} \quad \text{(log-density)},$$

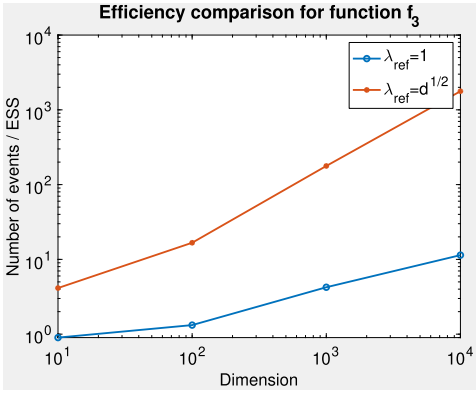
$$f_6(x) = x_1^2 \quad \text{(square of first coordinate)},$$

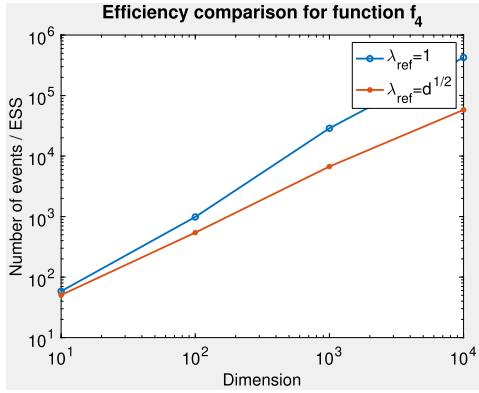
$$f_7(x) = x_1 x_2 \quad \text{(product of first and second coordinates)}.$$

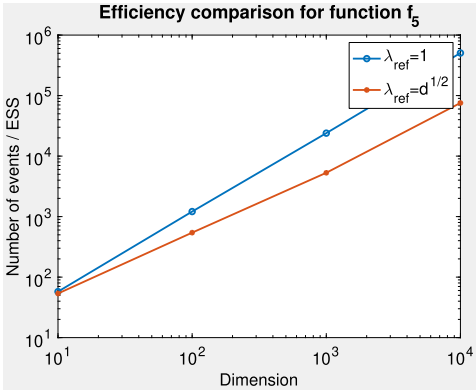
In order to estimate the effective sample sizes, we have run 100 parallel BPS simulations with 10^6 events per simulation, starting from the Gaussian target distribution. The autoregressive parameter α was set as $\alpha = 0$. Figure 2 shows the number of events required for one effective sample for dimensions $d = 10$, 100, 1000 and 10,000 for these seven functions, with refreshment parameter choices $\lambda_{\mathrm{ref}} = 1$ (as suggested by Theorems 3 and 5) and $\lambda_{\mathrm{ref}} = \sqrt{d}$ (as suggested by [8] and Table 1 of [1]). The number of events is a correct proxy for the computational cost as each event requires one gradient evaluation (see Section 2.3 of [15]

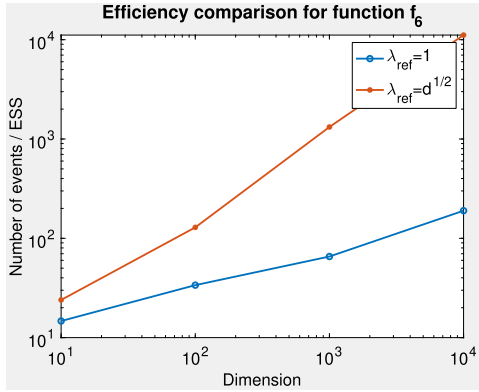












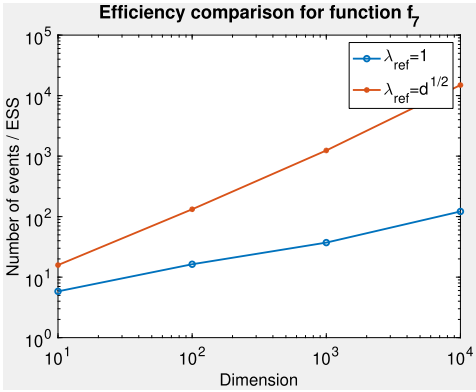


FIG. 2. *Number of BPS events per effective sample for* 7 *different functions for standard Gaussian target as a function of the dimension, with two different scalings of the refreshment rate* λ_{ref} *in terms of the dimension.*

for the description of the implementation of BPS for Gaussian targets). As we can see, if the refreshment rate is chosen as $\lambda_{\text{ref}} = 1$, these simulation results show $O(\sqrt{d})$ scaling in the number of events required for an effective sample for all of the functions except the radius and the log-density (f_4 and f_5). In contrast, the choice $\lambda_{\text{ref}} = \sqrt{d}$ seems to require significantly more events per effective sample, with $O(d)$ scaling observed empirically. In the cases of the radius and the log-density, the choice $\lambda_{\text{ref}} = \sqrt{d}$ still seems to require $O(d)$ events per effective sample, while $\lambda_{\text{ref}} = 1$ is doing worse, approximately $O(d^{4/3})$ events per effective sample is required. The scaling limits for this function were studied in [8], who has recommended choosing $\lambda_{\text{ref}} = O(\sqrt{d})$ to obtain the best mixing for the log-density, consistently with our empirical results.

To sum up, we can see that if the goal of the simulation is to estimate the posterior mean or posterior covariance matrix, or other quantities only depending a small subset of the coordinates, then choosing λ_{ref} as recommended by Theorems 3 and 5 yield good empirical performance ($O(\sqrt{d})$ scaling in the number of events required for an effective sample). For functions depending on all of the coordinates the situation is more complicated, and the best choice of λ_{ref} is strongly function dependent in this case.

3. Proof of weak convergence result—Theorem 1. The proof will be based on a sequence of auxiliary results. First we will show that the RHMC semigroup $\{P^t : t \geq 0\}$, acting on the Banach space $C_0(\mathcal{Z})$ with the sup-norm is Feller, and that the space $C_c^\infty(\mathcal{Z})$ is a *core* for its generator given in (2.5), in the sense that C_c^∞ is dense in $\mathcal{D}(\mathcal{A})$ with respect to the norm $|||\cdot||| := \|f\|_\infty + \|\mathcal{A}f\|_\infty$. This, and a sequence of auxiliary results, will allow us to apply [29], Corollary 8.6, to prove Theorem 1.

3.1. *Feller property.* Recall that in the context of Theorem 1, we have $d = 1$ and $\mathcal{Z} = \mathbb{R}^2$. A Markov process taking values in $\mathcal{Z}$, with transition semigroup $\{P^t : t \geq 0\}$, is called a Feller process and $\{P^t : t \geq 0\}$ a Feller semigroup, if it satisfies the following two properties:

Feller property: for all $t \geq 0$ and $f \in C_0(\mathcal{Z})$ we have $P^t f \in C_0(\mathcal{Z})$, and
Strong continuity: $\|P^t f - f\|_\infty \to 0$ as $t \to 0$ for $f \in C_0(\mathcal{Z})$.

PROPOSITION 6. *Suppose that $W : \mathbb{R} \mapsto [0, \infty)$ is continuously differentiable and $\lim_{|x|\to\infty} W(x) = \infty$. Then the RHMC process $\{Z_t\}_{t\geq 0}$ with generator $\mathcal{A}$ given by (2.5) with Hamiltonian $H(x, v) = W(x) + |v|^2/2$, $\alpha \in (0, 1)$ and $\lambda_{\text{ref}} > 0$ is a Feller process. If in addition $W \in \mathcal{C}^\infty(\mathbb{R})$, then $C_c^\infty(\mathbb{R})$ is a core for its generator.*

Note that a more technical approach proposed recently in Holderrieth [33] requires weaker assumptions.

3.1.1. *Proof of Proposition 6.* Before we proceed let us first define the *resolvent operator* for $\lambda > 0$

$$\mathcal{R}_\lambda f(z) := \int_0^\infty \mathrm{e}^{-\lambda s} P^s f(z)\,\mathrm{d}s = \int_0^\infty \mathrm{e}^{-\lambda s} \mathbb{E}^z\big[f(Z_s)\big]\,\mathrm{d}s.$$

The proof will proceed as follows. First we will first show that $\mathcal{R}_\lambda : C_0(\mathcal{Z}) \to C_0(\mathcal{Z})$, and then use [12], Corollary 1.23, to establish that $\{P^t : t \geq 0\}$ has the Feller property, that is, for all $t \geq 0$ $P^t : C_0(\mathcal{Z}) \to C_0(\mathcal{Z})$. Once the Feller property is established by [12], Lemma 1.4, to prove strong continuity, it suffices to prove the weaker statement $P^t f(z) \to f(z)$, for all $f \in C_0(\mathcal{Z})$ and $z \in \mathcal{Z}$. We now establish this property. Let $T_1, T_2, \ldots$ be the arrival times of

the jumps. Then we have for $h > 0$

$$\begin{aligned} P^h f(z) - f(z) &= \mathbb{E}^z\big[f(Z_h)\big] - f(z) \\ &= \mathbb{E}^z\big[f(Z_h)\mathbb{1}\{T_1 \geq h\}\big] - f(z) + \mathbb{E}^z\big[f(Z_h)\mathbb{1}\{T_1 < h\}\big] \\ &= f\big(\Xi(h,z)\big)\mathrm{e}^{-\lambda_{\mathrm{ref}}h} - f(z) + \mathcal{E}, \end{aligned}$$

where we write $\Xi(z,t)$ for the solution of the Hamiltonian dynamics at time t initialized at $z_0 = z$. It is well known that if $H : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is continuously differentiable everywhere then $\Xi(z,s)$ is well defined for all $s > 0$ (see, e.g., [18], Theorem 1.186), $H(\Xi(z,s)) = H(z)$ for all $s > 0$ and $\Xi(z,h) \to z$ as $h \to 0$. Since f is bounded it easily follows that as $h \to 0$

$$|\mathcal{E}| \leq \|f\|_\infty\big(1 - \mathrm{e}^{-\lambda_{\mathrm{ref}}h}\big) \to 0.$$

Since $\Xi(z,h) \to z$ as $h \to 0$, the result follows.

Proof of the Feller property. From [19], equation (2.6), we know that we can express the resolvent kernel as follows for a measurable set A:

$$\mathcal{R}_\lambda(z,A) = \sum_{j=0}^{\infty} J_\lambda^j K_\lambda(z,A), \tag{3.1}$$

where

$$K_\lambda(z,A) := \int_0^\infty \mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}}s}\mathbb{1}_A\big(\Xi(z,s)\big)\,\mathrm{d}s, \tag{3.2}$$

$$J_\lambda(z,A) := \int_0^\infty \lambda_{\mathrm{ref}}\mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}}s} Q_\alpha\big(\Xi(z,s),A\big)\,\mathrm{d}s, \tag{3.3}$$

with $\Xi(z,s) = \Xi((x,v),s)$ as defined above.

We will now show that $\mathcal{R}_\lambda f \in C_0(\mathcal{Z})$ for any $f \in C_0(\mathcal{Z})$. This follows from the next result.

LEMMA 1. *$W \in C^1(\mathbb{R};[0,\infty))$, $W(x) \to \infty$ as $|x| \to \infty$ and let $f \in C_0(\mathcal{Z})$. Then, for any $\lambda > 0$, we have $J_\lambda f, K_\lambda f \in C_0(\mathcal{Z})$ and $\|J_\lambda f\|_\infty \leq \lambda_{\mathrm{ref}}/(\lambda + \lambda_{\mathrm{ref}})\|f\|_\infty$. In particular*

$$\mathcal{R}_\lambda f = \sum_{j=0}^{\infty} J_\lambda^j K_\lambda f \in C_0(\mathcal{Z}).$$

PROOF OF LEMMA 1. Let $\lambda > 0$ and let us first look at K_λ. Suppose now that $f \in C_0(\mathcal{Z})$ and that $z_n \to z$. Then

$$\big|K_\lambda f(z) - K_\lambda f(z_n)\big| \leq \int_0^\infty \lambda_{\mathrm{ref}}\mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}}s}\big|f\big(\Xi(z,s)\big) - f\big(\Xi(z_n,s)\big)\big|\,\mathrm{d}s \to 0,$$

by the bounded convergence theorem, since f is bounded and the functions $s \mapsto |f(\Xi(z,s)) - f(\Xi(z_n,s))|$ vanish pointwise by the continuity of f and the continuous dependence of the solution $\{\Xi(z,s) : s \geq 0\}$ on the initial condition; see, for example, [18], Theorem 1.3. This establishes that $K_\lambda f$ is continuous.

Next we prove that $K_\lambda f$ vanishes at infinity. Let $\epsilon > 0$ be arbitrary. Since $W(x) \to \infty$ as $|x| \to \infty$, the level sets $\mathcal{H}_L := \{z : H(z) \leq L\}$ are compact and $\mathcal{Z} = \bigcup_{L>0}\{z : H(z) \leq L\}$. Therefore we can find $L = L(\epsilon)$ such that $|f(z)| < \epsilon(\lambda + \lambda_{\mathrm{ref}})$ for $z \notin \mathcal{H}_L$. For all such z, since $H(\Xi(z,s)) = H(z)$ for all $s > 0$, we have that

$$\begin{aligned} \big|K_\lambda f(z)\big| &\leq \int_0^\infty \mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}}s}\big|f\big(\Xi(z,s)\big)\big|\,\mathrm{d}s \\ &< \epsilon(\lambda + \lambda_{\mathrm{ref}})\int_0^\infty \mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}}s}\,\mathrm{d}s = \epsilon. \end{aligned}$$

Thus we conclude that for all $\lambda > 0$ we have $K_\lambda : C_0(\mathcal{Z}) \to C_0(\mathcal{Z})$.

Now we move on to J_λ. First notice that for any $f \in C_0(\mathcal{Z})$ we have $Q_\alpha f$ is also continuous. To see why let $z_n = (x_n, v_n) \to z = (x, v)$ and notice that as $d = 1$

$$\begin{aligned}&\big|Q_\alpha f(z_n) - Q_\alpha f(z)\big| \\ &\quad\le \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \big|f\big(x_n, \alpha v_n + \sqrt{1-\alpha^2}\xi\big) - f\big(x, \alpha v + \sqrt{1-\alpha^2}\xi\big)\big| \mathrm{e}^{-\xi^2/2}\,\mathrm{d}\xi \to 0,\end{aligned}$$

by the bounded convergence theorem, since f is continuous and bounded, and therefore $Q_\alpha f$ is continuous. Next, for any $\delta > 0$ we can choose a compact set K_δ such that $|f(z)| < \delta$ for $z \notin K_\delta$. In particular, since K_δ is compact, for any $\delta > 0$ we can also find $M_\delta > 0$ such that

$$K_\delta \subset \big\{(x, v) : |x|, |v| \le M_\delta\big\}.$$

Fix $\epsilon \in (0, 1/2)$ and choose z_ϵ such that $\Phi(z_\epsilon) = 1 - \epsilon/2$, where Φ is the cumulative distribution function of the standard normal distribution. Then

$$\big|Q_\alpha f(z)\big| \le \epsilon \|f\|_\infty + \frac{1}{\sqrt{2\pi}} \int_{\xi=-z_\epsilon}^{z_\epsilon} \big|f\big(x, \alpha v + \sqrt{1-\alpha^2}\xi\big)\big| \mathrm{e}^{-\xi^2/2}\,\mathrm{d}\xi.$$

Then for all $z = (x, v)$ and ξ such that $|x| > M_\epsilon$, $|v| > (M_\epsilon + z_\epsilon)/\alpha$ and $|\xi| < z_\epsilon$ we have

$$\big|\alpha v + \sqrt{1-\alpha^2}\xi\big| \ge \alpha|v| - \sqrt{1-\alpha^2}|\xi| \ge \alpha|v| - |\xi| \ge M_\epsilon + z_\epsilon - z_\epsilon > M_\epsilon.$$

Therefore for such z we have that

$$\begin{aligned}\big|Q_\alpha f(z)\big| &\le \epsilon \|f\|_\infty + \frac{1}{\sqrt{2\pi}} \int_{\xi=-z_\epsilon}^{z_\epsilon} \big|f\big(x, \alpha v + \sqrt{1-\alpha^2}\xi\big)\big| \mathrm{e}^{-\xi^2/2}\,\mathrm{d}\xi \\ &< \epsilon \|f\|_\infty + \frac{\epsilon}{\sqrt{2\pi}} \int_{\xi=-z_\epsilon}^{z_\epsilon} \mathrm{e}^{-\xi^2/2}\,\mathrm{d}\xi,\end{aligned}$$

and since $\epsilon > 0$ is arbitrary it follows that $Q_\alpha f \in C_0(\mathcal{Z})$.

Observe that $J_\lambda f(z) = \lambda_{\mathrm{ref}} K_\lambda Q_\alpha f(z)$. Therefore if $f \in C_0(\mathcal{Z})$, since we have already shown that $Q_\alpha : C_0(\mathcal{Z}) \to C_0(\mathcal{Z})$ and $K_\lambda : C_0(\mathcal{Z}) \to C_0(\mathcal{Z})$, it follows that $J_\lambda f \in C_0(\mathcal{Z})$.

Finally, since clearly $\|Q_\alpha f(\Xi(z, s))\|_\infty \le \|f\|_\infty$

$$\begin{aligned}\|J_\lambda f\|_\infty &= \sup_z \left| \int_0^\infty \lambda_{\mathrm{ref}} \mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}} s} Q_\alpha f\big(\Xi(z, s)\big)\,\mathrm{d}s \right| \\ &\le \int_0^\infty \lambda_{\mathrm{ref}} \mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}} s} \big\|Q_\alpha f\big(\Xi(z, s)\big)\big\|_\infty \,\mathrm{d}s \\ &\le \int_0^\infty \lambda_{\mathrm{ref}} \mathrm{e}^{-\lambda s - \lambda_{\mathrm{ref}} s} \|f\|_\infty \,\mathrm{d}s = \frac{\lambda_{\mathrm{ref}}}{\lambda + \lambda_{\mathrm{ref}}} \|f\|_\infty,\end{aligned}$$

and since $\lambda > 0$ we can see that this is a strict contraction. From this, it follows that the sequence

$$\sum_{j=0}^{n} J_\lambda^j K_\lambda f,$$

is Cauchy in the Banach space $(C_0(\mathcal{Z}), \|\cdot\|_\infty)$, whence the conclusion follows. □

C_c^∞ is a core. Define the semigroup $\{Q^t : t \ge 0\}$, where for each $t \ge 0$ $Q^t : C_0(\mathcal{Z}) \mapsto C_0(\mathcal{Z})$ is defined through $Q^t f(z) = f(\Xi(z, t))$, with $\Xi(z, t)$ denoting as before the solution of the Hamiltonian dynamics started from z at time t. It can be easily shown that the generator of Q^t is given for $f \in C_c^\infty(\mathcal{Z})$ by

$$Bf(x, v) = (\nabla_x f, v) - \big(\nabla_v f, \nabla U(x)\big),$$

that is the first two terms of the generator $\mathcal{A}$ of RHMC.

Let f be supported on a compact set K. By our assumptions on the Hamiltonian H, there exists $L > 0$ such that $K \subseteq \mathcal{H}_L := \{(x, v) : H(x, v) \leq L\}$. Letting $z \notin \mathcal{H}_L$, for all $t \geq 0$, we have by definition $H(\Xi(z, t)) = H(z)$ and thus $\Xi(z, t) \notin \mathcal{H}_L$. Therefore $Q^t f$ will have compact support.

Notice next, since $W \in C^\infty(\mathbb{R})$, that for any $t \geq 0$ the mapping $z \mapsto \Xi(z, t)$ is infinitely differentiable; see, for example, [18], Exercise 1.185. From this and the above discussion we conclude that for any $f \in C_c^\infty(\mathcal{Z})$ and $t \geq 0$ we have $Q^t f \in C_c^\infty$. Therefore from Davies [21], Theorem 1.9, and since $C_c^\infty(\mathcal{Z}) \subset C_0(\mathcal{Z})$ is dense, we conclude that C_c^∞ is a core for B, and in particular that for any $f \in \mathcal{D}(B)$, there exists a sequence $\{f_n : n \geq 0\} \subset C_c^\infty(\mathcal{Z})$ such that

$$\|f_n - f\|_\infty + \|Bf_n - Bf\|_\infty \to 0.$$

Since the operator $\lambda_{\text{ref}}[Q_\alpha - I]$ is clearly bounded on $C_0(\mathcal{Z})$ for any α, it follows that $\mathcal{D}(\mathcal{A}) = \mathcal{D}(B)$, and that for the sequence $\{f_n\}$ above we also have

$$\|f_n - f\|_\infty + \|\mathcal{A}f_n - \mathcal{A}f\|_\infty \to 0,$$

proving that $C_c^\infty(\mathcal{Z})$ is a core for $\mathcal{A}$.

3.2. *Proof of Theorem 1.* Recall that we write $\{\mathbf{Z}_n(s) : s \geq 0\}$ for BPS initialized from π_n, the generator of which we denote with $\mathcal{A}_n$, and write $\{Z_n^{(1)}(s) : s \geq 0\}$ for its first component. In addition let

$$\mathcal{F}_t^n := \sigma\{\mathbf{Z}_n(s) : s \leq t\} \quad \text{and} \quad \mathcal{G}_t^n := \sigma\{Z_n^{(1)}(s) : s \leq t\}.$$

Let $\epsilon_n \to 0$ be monotone and to be specified later on. All expectations will be with respect to the path measure of BPS started from π_n. We proceed with the usual construction. For some function $f : \mathcal{Z} \to \mathbb{R}$, that is f is a function only of $Z_n^{(1)}$, such that $f \in C_c^\infty$, smooth with compact support, we define

$$\xi_n(t) := \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[f\big(Z_n^{(1)}(t+s)\big)|\mathcal{G}_t^n\big]\,\mathrm{d}s, \tag{3.4}$$

$$\phi_n(t) := \epsilon_n^{-1}\mathbb{E}\big[f\big(Z_n^{(1)}(t+\epsilon_n)\big) - f\big(Z_n^{(1)}(t)\big)|\mathcal{G}_t^n\big]. \tag{3.5}$$

Abusing notation, we will also write f for the mapping $\mathcal{Z}^n \mapsto \mathbb{R}$ given by $f(z_1, \ldots, z_n) = f(z_1)$. We have already established that $(\mathcal{A}, C_c^\infty)$ generates the strongly continuous semigroup $\{P^t : t \geq 0\}$ corresponding to RHMC. To apply [29], Corollary 8.6 of Chapter 4, we need to check the following:

- *Strongly separating algebra*: the closure of the linear span of C_c^∞ contains an algebra that strongly separates points; see [29], Section 3.4, for the definition. This is obvious since $C_c^\infty(\mathcal{Z})$ strongly separates points and is dense in the algebra $C_0(\mathcal{Z})$, since any function in $C_0(\mathcal{Z})$ can be approximated arbitrarily well by functions in $C_c(\mathcal{Z})$ by multiplying with, and then convolving with appropriate mollifiers.
- *Generator convergence*: for each $f \in C_c^\infty(\mathcal{Z})$ and $T > 0$, for ξ_n, ϕ_n as defined in (3.4), (3.5)

$$\sup_n \sup_{t \leq T} \mathbb{E}\big[|\xi_n(t)|\big] < \infty, \tag{3.6}$$

$$\sup_n \sup_{t \leq T} \mathbb{E}\big[|\phi_n(t)|\big] < \infty, \tag{3.7}$$

$$\lim_{n\to\infty} \mathbb{E}\big[|\xi_n(t) - f\big(Z_n^{(1)}(t)\big)|\big] = 0, \tag{3.8}$$

$$\lim_{n\to\infty} \mathbb{E}\big[|\phi_n(t) - \mathcal{A}f\big(Z_n^{(1)}(t)\big)|\big] = 0, \tag{3.9}$$

and in addition

$$\lim_{n\to\infty} \mathbb{E}\Big\{ \sup_{t\in\mathbb{Q}\cap[0,T]} \big|\xi_n(t) - f\big(Z_n^{(1)}(t)\big)\big| \Big\} = 0, \tag{3.10}$$

and for some $p > 1$

$$\sup_{n\to\infty} \mathbb{E}\Big[\Big(\int_0^T |\phi_n(s)|^p \mathrm{d}s \Big)^{1/p} \Big] < \infty. \tag{3.11}$$

3.2.1. *Proofs of equations* (3.10) *and* (3.8). Since condition (3.8) is implied by (3.10), we will establish (3.10).

First recall that for each n, BPS is nonexplosive. To see why, for each $\boldsymbol{x}$, $\boldsymbol{v}$, let $L > |v| > 0$ and consider

$$\tau_{n,L} := \inf\big\{ t \geq 0 : \boldsymbol{Z}_n(t) \notin B\big(x, L^2\big) \times B(0, L) \big\}.$$

Letting

$$\sigma^x_{n,L^2} := \inf\big\{ t \geq 0 : \boldsymbol{X}_n(t) \notin B\big(x, L^2\big) \big\}, \qquad \sigma^v_{n,L} := \inf\big\{ t \geq 0 : \boldsymbol{V}_n(t) \notin B(0, L) \big\},$$

we have

$$\begin{aligned} \tau_{n,L} &= \sigma^x_{n,L^2} \mathbb{1}\big\{ \sigma^x_{n,L^2} < \sigma^v_{n,L} \big\} + \sigma^v_{n,L} \mathbb{1}\big\{ \sigma^x_{n,L^2} \geq \sigma^v_{n,L} \big\} \\ &\geq L \mathbb{1}\big\{ \sigma^x_{n,L^2} < \sigma^v_{n,L} \big\} + \sigma^v_{n,L} \mathbb{1}\big\{ \sigma^x_{n,L^2} \geq \sigma^v_{n,L} \big\} \geq L \vee \sigma^v_{n,L}, \end{aligned}$$

where the first inequality follows, since on the event $\{\sigma^x_{L^2} \geq \sigma^v_L\}$ the maximum speed up to $\sigma^x_{L^2}$ is less than L. Since $|\boldsymbol{V}_n(t)|$ only changes at the arrivals of a homogeneous Poisson process with rate $\lambda_{\mathrm{ref}} > 0$, it is clear that as $L \to \infty$, $\sigma^v_{n,L} \to \infty$ and therefore $\tau_{n,L} \to \infty$.

Fix $T > 0$. Since BPS is nonexplosive for every n and $\delta > 0$ we can find a $K_{n,\delta} > 0$ such that

$$\mathbb{P}\Big[\sup_{t\leq T+1} \big|\boldsymbol{Z}_n(t)\big| \geq K_{n,\delta} \Big] \leq \delta.$$

For $\delta_n \to 0$ and by a diagonal argument, we can find a sequence K_{n,δ_n} such that

$$\mathbb{P}\Big[\sup_{t\leq T+1} \big|\boldsymbol{Z}_n(t)\big| \geq K_{n,\delta_n} \Big] \leq \delta_n \to 0.$$

We will write G_n for the event

$$G_n := \Big\{ \sup_{t\leq T+1} \big|\boldsymbol{Z}_n(t)\big| \leq K_{n,\delta_n} \Big\}.$$

Then we have for $\epsilon_n \to 0$, to be specified later on,

$$\begin{aligned} &\mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}} \big|\xi_n(t) - f\big(Z_n^{(1)}(t)\big)\big| \Big] \\ &\quad = \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}} \Big| \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[f\big(Z_n^{(1)}(t+r)\big) - f\big(Z_n^{(1)}(t)\big) | \mathcal{G}^n_t \big] \mathrm{d}r \Big| \Big] \\ &\quad = \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}} \Big| \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{ f\big(Z_n^{(1)}(t+r)\big) - f\big(Z_n^{(1)}(t)\big) | \mathcal{F}^n_t \big\} | \mathcal{G}^n_t \big] \mathrm{d}r \Big| \Big] \\ &\quad \leq \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}} \Big| \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{ \big(f\big(Z_n^{(1)}(t+r)\big) - f\big(Z_n^{(1)}(t)\big) \big) \mathbb{1}_{G_n} | \mathcal{F}^n_t \big\} | \mathcal{G}^n_t \big] \mathrm{d}r \Big| \Big] \\ &\qquad + \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}} \Big| \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{ \big(f\big(Z_n^{(1)}(t+r)\big) - f\big(Z_n^{(1)}(t)\big) \big) \mathbb{1}_{G_n^c} | \mathcal{F}^n_t \big\} | \mathcal{G}^n_t \big] \mathrm{d}r \Big| \Big] \\ &\quad := J_1 + J_2. \end{aligned}$$

For the term J_2 we have for $p > 1$

$$
\begin{aligned}
J_2 &\le 2\|f\|_\infty \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}} \mathbb{E}\big[\mathbb{1}_{G_n^c}|\mathcal{G}_t^n\big]\Big] \\
&\le 2\|f\|_\infty \mathbb{E}\Big[\Big(\sup_{t\in[0,T]\cap\mathbb{Q}} \mathbb{E}\big[\mathbb{1}_{G_n^c}|\mathcal{G}_t^n\big]\Big)^p\Big]^{1/p} \\
&\le 2\|f\|_\infty \frac{p}{p-1}\mathbb{E}\big[\mathbb{E}\big[\mathbb{1}_{G_n^c}|\mathcal{G}_T^n\big]^p\big]^{1/p} \le 2\|f\|_\infty \frac{p}{p-1}\mathbb{E}\big[\mathbb{1}_{G_n^c}^p\big]^{1/p} \\
&= 2\|f\|_\infty \frac{p}{p-1}\delta_n^{1/p},
\end{aligned}
\tag{3.12}
$$

where we used Jensen's inequality, the fact that for each n, $\{\mathbb{E}[\mathbb{1}_{G_n^c}|\mathcal{G}_t^n] : t \ge 0\}$ is a $\mathcal{G}_t^n$-martingale and Doob's martingale inequality.

We proceed with the term J_1 as follows:

$$
\begin{aligned}
J_1 \le\ & \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{\big[f\big(Z_n^{(1)}(t+r)\big) - f\big(Z_n^{(1)}(t)\big)\big] \\
&\times \mathbb{1}_{G_n}\mathbb{1}\{\tau_1^{\mathrm{ref}}(t) > \epsilon_n\}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Big|\Big] \\
&+ \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{\big[f\big(Z_n^{(1)}(t+r)\big) - f\big(Z_n^{(1)}(t)\big)\big] \\
&\times \mathbb{1}_{G_n}\mathbb{1}\{\tau_1^{\mathrm{ref}}(t) \le \epsilon_n\}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Big|\Big] \\
=:\ & J_{1,1} + J_{1,2},
\end{aligned}
$$

where we denote by $\tau_1^{\mathrm{ref}}(t)$ the first refreshment time after time t. Since refreshment happens independently we can bound $J_{1,2}$

$$
J_{1,2} \le 2\|f\|_\infty \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n}\big(1-\mathrm{e}^{-\lambda_{\mathrm{ref}}\epsilon_n}\big)\,\mathrm{d}r\Big|\Big] \le 2\|f\|_\infty\lambda_{\mathrm{ref}}\epsilon_n \to 0.
$$

We control the term $J_{1,1}$ in two steps. To keep notation short we introduce the notation $G_n'(t) := \{\tau_1^{\mathrm{ref}}(t) > \epsilon_n\}$. Then

$$
\begin{aligned}
J_{1,1} \le\ & \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{\big[f\big(X_n^{(1)}(t+r), V_n^{(1)}(t+r)\big) \\
&- f\big(X_n^{(1)}(t), V_n^{(1)}(t+r)\big)\big]\mathbb{1}_{G_n}\mathbb{1}_{G_n'(t)}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Big|\Big] \\
&+ \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{\big[f\big(X_n^{(1)}(t), V_n^{(1)}(t+r)\big) \\
&- f\big(X_n^{(1)}(t), V_n^{(1)}(t)\big)\big]\mathbb{1}_{G_n}\mathbb{1}_{G_n'(t)}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Big|\Big] =: J_{1,1,1} + J_{1,1,2}.
\end{aligned}
$$

For the first term, since only the location component changes we have

$$
\begin{aligned}
J_{1,1,1} &\le \|\partial_x f\|_\infty \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\big\{\big|X_n^{(1)}(t+r) - X_n^{(1)}(t)\big| \\
&\qquad \times \mathbb{1}_{G_n}\mathbb{1}_{G_n'(t)}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Big|\Big] \\
&\le \|\partial_x f\|_\infty \mathbb{E}\Big[\sup_{t\in[0,T]\cap\mathbb{Q}}\Big|\epsilon_n^{-1}\int_0^{\epsilon_n} \mathbb{E}\big[\epsilon_n\big|V_n^{(1)}(t)\big| \times \mathbb{1}_{G_n}\mathbb{1}_{G_n'(t)}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Big|\Big],
\end{aligned}
$$

where the second inequality follows from the linear dynamics of BPS, since on the event $G_n'(t)$ there is no refreshment event and therefore the norm of the velocity component does not change. Finally, recalling the definition of the event G_n we obtain

$$J_{1,1,1} \leq \|\partial_x f\|_\infty \epsilon_n K_{n,\delta_n}.$$

Next we have to control the term $J_{1,1,2}$ for which we point out that, since there is no refreshment event, the velocity will remain constant on the interval $[t, t+\epsilon_n]$ unless there is a bounce. Writing $\sigma_1(t)$ for the arrival time of the first bounce after time t we thus have

$$\begin{aligned}
&J_{1,1,2}\\
&\quad= \mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\big[\mathbb{E}\big\{\big[f\big(X_n^{(1)}(t), V_n^{(1)}(t+r)\big)\\
&\qquad - f\big(X_n^{(1)}(t), V_n^{(1)}(t)\big)\big]\mathbb{1}\{\sigma_1(t)<\epsilon_n\}\mathbb{1}_{G_n}\mathbb{1}_{G_n'(t)}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Bigg|\Bigg]\\
&\quad\leq 2\|f\|_\infty\mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\big[\mathbb{E}\big\{\mathbb{1}\{\sigma_1(t)<\epsilon_n\}\mathbb{1}_{G_n}\mathbb{1}_{G_n'(t)}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Bigg|\Bigg]\\
&\quad\leq 2\|f\|_\infty\mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\big[\mathbb{E}\big\{\mathbb{1}\{\sigma_1(t)<\epsilon_n\}|\mathcal{F}_t^n\big\}|\mathcal{G}_t^n\big]\,\mathrm{d}r\Bigg|\Bigg]\\
&\quad\leq 2\|f\|_\infty\mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\Bigg[\mathbb{E}\Bigg\{\Bigg[1\\
&\qquad - \exp\Bigg(-\int_0^{\epsilon_n}\big(\nabla U_n\big(\boldsymbol{X}_n(t+s)\big), \boldsymbol{V}_n(t+s)\big)_+\,\mathrm{d}s\Bigg)\Bigg]\Bigg|\mathcal{F}_t^n\Bigg\}\Bigg|\mathcal{G}_t^n\Bigg]\,\mathrm{d}r\Bigg|\Bigg],
\end{aligned}$$

where we dropped the indicators in order to be able to compute the probability of no bounce. We again decompose according to the event G_n in order to proceed

$$\begin{aligned}
&J_{1,1,2}\\
&\quad\leq 2\|f\|_\infty\mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\Bigg[\mathbb{E}\Bigg\{\Bigg[1\\
&\qquad - \exp\Bigg(-\int_0^{\epsilon_n}\big(\nabla U_n\big(\boldsymbol{X}_n(t+s)\big), \boldsymbol{V}_n(t+s)\big)_+\,\mathrm{d}s\Bigg)\Bigg]\mathbb{1}_{G_n}\Bigg|\mathcal{F}_t^n\Bigg\}\Bigg|\mathcal{G}_t^n\Bigg]\,\mathrm{d}r\Bigg|\Bigg]\\
&\qquad + 2\|f\|_\infty\mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\Bigg[\mathbb{E}\Bigg\{\Bigg[1\\
&\qquad - \exp\Bigg(-\int_0^{\epsilon_n}\big(\nabla U_n\big(\boldsymbol{X}_n(t+s)\big), \boldsymbol{V}_n(t+s)\big)_+\,\mathrm{d}s\Bigg)\Bigg]\mathbb{1}_{G_n^c}\Bigg|\mathcal{F}_t^n\Bigg\}\Bigg|\mathcal{G}_t^n\Bigg]\,\mathrm{d}r\Bigg|\Bigg].
\end{aligned}$$

Since the integrand is bounded above by 1, a calculation similar to the one for the term J_2 in (3.12) shows that the second term above vanishes as $n\to\infty$, and therefore using the inequality $1-\exp(-x)\leq x$ for $x>0$ we have for $p>1$

$$\begin{aligned}
&J_{1,1,2}\\
&\quad\leq C\delta_n^{1/p} + 2\|f\|_\infty\mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\\
&\qquad\times\int_0^{\epsilon_n}\mathbb{E}\Bigg[\mathbb{E}\Bigg\{\int_0^{\epsilon_n}\big|\nabla U_n\big(\boldsymbol{X}_n(t+s)\big)\big|\big|\boldsymbol{V}_n(t+s)\big|\,\mathrm{d}s\mathbb{1}_{G_n}\Bigg|\mathcal{F}_t^n\Bigg\}\Bigg|\mathcal{G}_t^n\Bigg]\,\mathrm{d}r\Bigg|\Bigg]\\
&\quad\leq C\delta_n^{1/p}
\end{aligned}$$

$$
\begin{aligned}
&+ 2\|f\|_\infty \mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\\
&\times \int_0^{\epsilon_n} \mathbb{E}\Bigg[\mathbb{E}\Bigg\{\int_0^{\epsilon_n}\Big(\frac{1}{2}\big|\nabla U_n\big(\boldsymbol{X}_n(t+s)\big)\big|^2 + \frac{1}{2}\big|\boldsymbol{V}_n(t+s)\big|^2\Big)\,\mathrm{d}s \times \mathbb{1}_{G_n}\Big|\mathcal{F}_t^n\Bigg\}\Big|\mathcal{G}_t^n\Bigg]\,\mathrm{d}r\Bigg|\Bigg]\\
&\leq C\delta_n^{1/p}\\
&+ 2C\|f\|_\infty \mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\\
&\times \int_0^{\epsilon_n} \mathbb{E}\Bigg[\mathbb{E}\Bigg\{\int_0^{\epsilon_n}\big(M\big|\boldsymbol{X}_n(t+s)\big|^2 + \big|\boldsymbol{V}_n(t+s)\big|^2\big)\,\mathrm{d}s \times \mathbb{1}_{G_n}\Big|\mathcal{F}_t^n\Bigg\}\Big|\mathcal{G}_t^n\Bigg]\,\mathrm{d}r\Bigg|\Bigg]
\end{aligned}
$$

since $|\nabla U_n(\boldsymbol{x})| = |\nabla U_n(\boldsymbol{x}) - \nabla U_n(0)| \leq M|\boldsymbol{x}|$ by Assumption 1

$$
\begin{aligned}
&\leq C\delta_n^{1/p}\\
&\quad + 2CM\|f\|_\infty \mathbb{E}\Bigg[\sup_{t\in[0,T]\cap\mathbb{Q}}\Bigg|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\big[C\epsilon_n\big|\boldsymbol{Z}_n(t+s)\big|^2\mathbb{1}_{G_n}\big|\mathcal{F}_t^n\big\}\big|\mathcal{G}_t^n\big]\,\mathrm{d}r\Bigg|\Bigg]\\
&\leq C\delta_n^{1/p} + 2C\|f\|_\infty\epsilon_n K_{n,\delta_n}^2.
\end{aligned}
$$

We choose ϵ_n such that $\epsilon_n K_{n,\delta_n}^2 \to 0$.

3.2.2. *Proof of* (3.9). Next we prove (3.9). First, by stationarity notice that we can equivalently check

$$\mathbb{E}\big[\big|\phi_n(0) - \mathcal{A}f\big(Z_n^{(1)}(0)\big)\big|\big] \to 0.$$

Notice first that $f \in \mathcal{D}(\widetilde{\mathcal{A}}_n)$, the domain of the *extended generator*, since f is smooth and bounded (see [22], Theorem 26.14)

$$
\begin{aligned}
\phi_n(0) &= \epsilon_n^{-1}\mathbb{E}\big[f\big(Z_n^{(1)}(\epsilon_n)\big) - f\big(Z_n^{(1)}(0)\big)\big|\mathcal{G}_0^n\big]\\
&= \epsilon_n^{-1}\mathbb{E}\Big[\int_0^{\epsilon_n}\widetilde{\mathcal{A}}_n f\big(\boldsymbol{Z}_n(s)\big)\,\mathrm{d}s + \mathcal{R}_n(s)\Big|\mathcal{G}_0^n\Big]\\
&= \epsilon_n^{-1}\mathbb{E}\Big[\int_0^{\epsilon_n}\widetilde{\mathcal{A}}_n f\big(\boldsymbol{Z}_n(s)\big)\,\mathrm{d}s\Big|\mathcal{G}_0^n\Big],
\end{aligned}
$$

where we used the facts that $\mathcal{R}_n(t)$ is an $\mathcal{F}_t^n$-martingale and $\mathcal{F}_t^n \subseteq \mathcal{G}_t^n$, whence

$$\mathbb{E}\big[\mathcal{R}_n(s)|\mathcal{G}_0^n\big] = \mathbb{E}\big\{\mathbb{E}\big[\mathcal{R}_n(s)|\mathcal{F}_0^n\big]|\mathcal{G}_0^n\big\} = 0.$$

We also notice that $g_n := \widetilde{\mathcal{A}}_n f \in \mathrm{Dom}(\widetilde{\mathcal{A}}_n)$ the domain of the extended generator. Therefore

$$
\begin{aligned}
\phi_n(0) &= \epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\big[g_n\big(\boldsymbol{Z}_n(s)\big)|\mathcal{G}_0^n\big]\,\mathrm{d}s\\
&= \epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\Big[\widetilde{\mathcal{A}}_n f\big(\boldsymbol{Z}_n(0)\big) + \int_0^s \widetilde{\mathcal{A}}_n g_n\big(\boldsymbol{Z}_n(r)\big) + \mathcal{R}_n'(s)\,\mathrm{d}r\Big|\mathcal{G}_0^n\Big]\,\mathrm{d}s\\
&= \epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\Big[\widetilde{\mathcal{A}}_n f\big(\boldsymbol{Z}_n(0)\big) + \int_0^s \widetilde{\mathcal{A}}_n g_n\big(\boldsymbol{Z}_n(r)\big)\,\mathrm{d}r\Big|\mathcal{G}_0^n\Big]\,\mathrm{d}s,
\end{aligned}
$$

where, from [22], Theorem 26.12, it follows that the local martingale $\{\mathcal{R}_n'(s) : s \geq 0\}$ is actually a proper martingale, and therefore using the same arguments as before, for $s > 0$,

$$\mathbb{E}\big[\mathcal{R}_n'(s)|\mathcal{G}_0^n\big] = 0.$$

Then we have

$$\begin{aligned}\mathbb{E}\big[|\phi_n(0) - \mathcal{A}f(Z_n^{(1)}(0))|\big] &\le \mathbb{E}\big[|\mathbb{E}[\widetilde{\mathcal{A}}_n f(\mathbf{Z}_n(0))|\mathcal{G}_0^n] - \mathcal{A}f(Z_n^{(1)}(0))|\big] \\ &\quad + \mathbb{E}\left\{\left|\epsilon_n^{-1}\int_0^{\epsilon_n}\mathbb{E}\left[\int_0^s \widetilde{\mathcal{A}}_n g_n(\mathbf{Z}_n(r))\,\mathrm{d}r\middle|\mathcal{G}_0^n\right]\mathrm{d}s\right|\right\} \\ &\le \mathbb{E}\big[|\mathbb{E}[\widetilde{\mathcal{A}}_n f(\mathbf{Z}_n(0))|\mathcal{G}_0^n] - \mathcal{A}f(Z_n^{(1)}(0))|\big] \\ &\quad + \epsilon_n^{-1}\int_0^{\epsilon_n}\int_0^s \mathbb{E}\{\mathbb{E}[|\widetilde{\mathcal{A}}_n g_n(\mathbf{Z}_n(r))||\mathcal{G}_0^n]\}\,\mathrm{d}r\,\mathrm{d}s \\ &:= \mathbb{E}\big[|\mathbb{E}[\widetilde{\mathcal{A}}_n f(\mathbf{Z}_n(0))|\mathcal{G}_0^n] - \mathcal{A}f(Z_n^{(1)}(0))|\big] + \mathcal{R}_n,\end{aligned} \tag{3.13}$$

applying Jensen's inequality conditionally. Finally by the tower law and by stationarity of $\{\mathbf{Z}_n(t) : t \ge 0\}$ when initialized from π_n

$$\begin{aligned}\mathcal{R}_n &= \epsilon_n^{-1}\int_0^{\epsilon_n}\int_0^s \mathbb{E}\{\mathbb{E}[|\widetilde{\mathcal{A}}_n g_n(\mathbf{Z}_n(r))||\mathcal{G}_0^n]\}\,\mathrm{d}r\,\mathrm{d}s \\ &= \epsilon_n^{-1}\int_0^{\epsilon_n}\int_0^s \mathbb{E}\{|\widetilde{\mathcal{A}}_n g_n(\mathbf{Z}_n(r))|\}\,\mathrm{d}r\,\mathrm{d}s = \frac{\epsilon_n}{2}\mathbb{E}\{|\widetilde{\mathcal{A}}_n g_n(\mathbf{Z}_n(0))|\}.\end{aligned}$$

Error term. We will now control this error term. Recall first that for $f \in C_c^\infty(\mathcal{Z}) \subset \mathcal{D}(\mathcal{A}_n)$ we have

$$\begin{aligned}\mathcal{A}_n f(\boldsymbol{x},\boldsymbol{v}) &= (\nabla f(\boldsymbol{x}),\boldsymbol{v})) + \max\{0,(\nabla U_n(\boldsymbol{x}),\boldsymbol{v})\}[\mathfrak{R}_n f(\boldsymbol{x},\boldsymbol{v}) - f(\boldsymbol{x},\boldsymbol{v})] \\ &\quad + \lambda_{\mathrm{ref}}[Qf(\boldsymbol{x},\boldsymbol{v}) - f(\boldsymbol{x},\boldsymbol{v})], \\ \mathfrak{R}_n f(\boldsymbol{x},\boldsymbol{v}) &:= f\left(\boldsymbol{x},\boldsymbol{v} - 2\frac{(\nabla U_n(\boldsymbol{x}),\boldsymbol{v}))}{|\nabla U_n(\boldsymbol{x})|^2}\nabla U_n(\boldsymbol{x})\right), \\ Q_{\alpha,n} f(\boldsymbol{x},\boldsymbol{v}) &:= \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \mathrm{e}^{-|\boldsymbol{\xi}|^2/2} f\big(\boldsymbol{x},\alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\,\mathrm{d}\boldsymbol{\xi}.\end{aligned}$$

Potentially abusing notation, for $n \ge 1$ and $\boldsymbol{x} \in \mathbb{R}^n$ we define a mapping $R_n(\boldsymbol{x}) : \mathbb{R}^n \mapsto \mathbb{R}^n$ through

$$R_n(\boldsymbol{x})\boldsymbol{v} := \boldsymbol{v} - 2\frac{(\nabla U_n(\boldsymbol{x}),\boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\nabla U_n(\boldsymbol{x}),$$

with the convention that $R_n(x)\boldsymbol{v} = 0$, when $\nabla U_n(\boldsymbol{x}) = 0$.

We decompose the generator $\mathcal{A}_n$ into three parts

$$\mathcal{A}_n = \mathcal{A}_n^{(1)} + \mathcal{A}_n^{(2)} + \mathcal{A}_n^{(3)},$$

where

$$\begin{aligned}\mathcal{A}_n^{(1)} f(\boldsymbol{x},\boldsymbol{v}) &= \frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{x}+t\boldsymbol{v},\boldsymbol{v})\Big|_{t=0}, \\ \mathcal{A}_n^{(2)} f(\boldsymbol{x},\boldsymbol{v}) &= \max\{0,(\nabla U_n(\boldsymbol{x}),\boldsymbol{v})\}[\mathfrak{R}_n f(\boldsymbol{x},\boldsymbol{v}) - f(\boldsymbol{x},\boldsymbol{v})], \\ \mathcal{A}_n^{(3)} f(\boldsymbol{x},\boldsymbol{v}) &= \lambda_{\mathrm{ref}}[Qf(\boldsymbol{x},\boldsymbol{v}) - f(\boldsymbol{x},\boldsymbol{v})].\end{aligned}$$

REMARK 12. Notice that when f is differentiable we have

$$\mathcal{A}_n^{(1)} f(\boldsymbol{x},\boldsymbol{v}) = \langle\nabla f(\boldsymbol{x}),\boldsymbol{v}\rangle,$$

however, for $\mathcal{A}_n^{(1)} f(x,v)$ to be well defined we only need that $t \mapsto f(\boldsymbol{x}+t\boldsymbol{v},\boldsymbol{v})$ is absolutely continuous; see Davis [22], Chapter 2.22.

Therefore when considering $\mathcal{A}_n g_n = \mathcal{A}_n\mathcal{A}_n f_n$ we will need to consider all possible combinations $\mathcal{A}_n^{(i)}\mathcal{A}_n^{(j)}$ since the operators do not necessarily commute.

Case $i = 1$. Using the fact that $f(\boldsymbol{x}, \boldsymbol{v}) = f(x_1, v_1)$, where we write (x_1, v_1) for the first location and velocity components of $(\boldsymbol{x}, \boldsymbol{v})$, the first term reduces to

$$\begin{aligned}\mathcal{A}_n^{(1)}\mathcal{A}_n^{(1)} f(\boldsymbol{x}, \boldsymbol{v}) &= \frac{\mathrm{d}}{\mathrm{d}t}\big(\nabla f(\boldsymbol{x}), \boldsymbol{v}\big)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial x} f(x_1 + tv_1, v_1)v_1\Big|_{t=0} \\ &= \frac{\partial^2 f}{\partial x^2}(x_1, v_1)v_1^2.\end{aligned}$$

Since $f \in C_c^\infty(\mathbb{R} \times \mathbb{R})$, it follows that $\partial_x^2 f(x, v)$ is also continuous and compactly supported and therefore bounded. Thus

$$\mathbb{E}\left|\frac{\partial^2 f}{\partial x^2}\big(X^{(1)}, V^{(1)}\big)\big(V^{(1)}\big)^2\right| \le \left\|\frac{\partial^2 f}{\partial x^2}\right\|_\infty \mathbb{E}\big[\big(V^{(1)}\big)^2\big] \le \left\|\frac{\partial^2 f}{\partial x^2}\right\|_\infty = O(1),$$

since under π_n, $V^{(1)}$ is centered Gaussian with unit variance.

The second term (see Remark 12) takes the form

$$\begin{aligned}\mathcal{A}_n^{(1)}\mathcal{A}_n^{(2)} f(\boldsymbol{x}, \boldsymbol{v}) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}_n^{(2)} f(\boldsymbol{x} + t\boldsymbol{v}, \boldsymbol{v})\Big|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\max\big\{0, \big(\nabla U_n(\boldsymbol{x} + t\boldsymbol{v}), \boldsymbol{v}\big)\big\}\Big|_{t=0}\big[\Re f(\boldsymbol{x}, \boldsymbol{v}) - f(\boldsymbol{x}, \boldsymbol{v})\big] \\ &\quad + \frac{\mathrm{d}}{\mathrm{d}t}\big[\Re f(\boldsymbol{x} + t\boldsymbol{v}, \boldsymbol{v}) - f(\boldsymbol{x} + t\boldsymbol{v}, \boldsymbol{v})\big]\Big|_{t=0}\max\big\{0, \big(\nabla U_n(\boldsymbol{x}), \boldsymbol{v}\big)\big\} \\ &=: J_1 + J_2.\end{aligned}$$

For J_1, since by Assumption 1 ∇U_n is M-Lipschitz

$$\begin{aligned}&\big|\max\big\{0, \big(\nabla U_n(\boldsymbol{x} + h\boldsymbol{v}), \boldsymbol{v}\big)\big\} - \max\big\{0, \big(\nabla U_n(\boldsymbol{x}), \boldsymbol{v}\big)\big\}\big| \\ &\quad \le \big|\big(\nabla U_n(\boldsymbol{x} + h\boldsymbol{v}), \boldsymbol{v}\big) - \big(\nabla U_n(\boldsymbol{x}), \boldsymbol{v}\big)\big| \le Mh|\boldsymbol{v}|^2.\end{aligned}$$

Therefore we have that, for $h \in (0, 1)$

$$h^{-1}\big|\max\big\{0, \big(\nabla U_n(\boldsymbol{x} + h\boldsymbol{v}), \boldsymbol{v}\big)\big\} - \max\big\{0, \big(\nabla U_n(\boldsymbol{x}), \boldsymbol{v}\big)\big\}\big| \le M|\boldsymbol{v}|^2 \in L^1(\pi),$$

since the V_i are standard normal random variables. In addition since f is bounded it follows that $\Re f(\boldsymbol{x}, \boldsymbol{v}) \le \|f\|_\infty$. Therefore by the dominated convergence theorem, we can exchange the $h \to 0$ limit and expectation to obtain

$$\begin{aligned}\pi[J_1] &\le 2\|f\|_\infty \mathbb{E}\left[\left|\frac{\mathrm{d}}{\mathrm{d}t}\max\big\{0, \big(\nabla U_n(\boldsymbol{X} + t\boldsymbol{V}), \boldsymbol{V}\big)\big\}\Big|_{t=0}\right|\right] \\ &\le 2\|f\|_\infty \mathbb{E}\Big[\lim_{h\to 0} h^{-1}\big|\big(\nabla U_n(\boldsymbol{X} + h\boldsymbol{V}), \boldsymbol{V}\big) - \big(\nabla U_n(\boldsymbol{X}), \boldsymbol{V}\big)\big|\Big] \\ &\le 2\|f\|_\infty M\mathbb{E}\big[|\boldsymbol{V}|^2\big] = O(Mn).\end{aligned}$$

For J_2 a lengthy but straightforward calculation shows that

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\Re f(\boldsymbol{x} + t\boldsymbol{v}, \boldsymbol{v})\Big|_{t=0} \\ &\quad = \frac{\mathrm{d}}{\mathrm{d}t} f\left(x_1 + tv_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x} + t\boldsymbol{v}), \boldsymbol{v}))}{|\nabla U_n(\boldsymbol{x} + t\boldsymbol{v})|^2}\partial_1 U_n(\boldsymbol{x} + t\boldsymbol{v})\right)\Big|_{t=0} \\ &\quad = \partial_x f\left(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{x})\right)v_1\end{aligned}$$

$$- 2\partial_v f\left(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{x})\right)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{(\nabla U_n(\boldsymbol{x}+t\boldsymbol{v}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x}+t\boldsymbol{v})|^2}\partial_1 U_n(\boldsymbol{x}+t\boldsymbol{v})\right)\Big|_{t=0}$$
$$= (\mathcal{R}\partial_x f)(\boldsymbol{x}, \boldsymbol{v})v_1 - (\mathcal{R}\partial_v f)(\boldsymbol{x}, \boldsymbol{v}) \times \mathfrak{U}(\boldsymbol{x}, \boldsymbol{v}),$$

where

$$\mathfrak{U}(\boldsymbol{x}, \boldsymbol{v}) := \frac{\mathrm{d}}{\mathrm{d}t}\left(2\frac{(\nabla U_n(\boldsymbol{x}+t\boldsymbol{v}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x}+t\boldsymbol{v})|^2}\partial_1 U_n(\boldsymbol{x}+t\boldsymbol{v})\right)\Big|_{t=0}$$
$$= \frac{2}{|\nabla U_n(\boldsymbol{x})|^2}\left\{(\boldsymbol{v}, \nabla U_n^2(\boldsymbol{x})\boldsymbol{v})\partial_1 U_n(\boldsymbol{x}) + (\nabla U_n(\boldsymbol{x}), \boldsymbol{v})\sum_{j=1}^{n}\partial_{j,1}^2 U_n(\boldsymbol{x})v_j\right\}$$
$$- \frac{1}{|\nabla U_n(\boldsymbol{x})|^4}\{2\partial_1 U_n(\boldsymbol{x})(\nabla U_n(\boldsymbol{x}), \nabla^2 U_n(\boldsymbol{x})\boldsymbol{v})(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})\},$$

and thus by Assumption 1

$$|\mathfrak{U}(\boldsymbol{x}, \boldsymbol{v})| \le \frac{2}{|\nabla U_n(\boldsymbol{x})|^2}\{M|\nabla U_n(\boldsymbol{x})||\boldsymbol{v}|^2 + M|\boldsymbol{v}|^2|\nabla U_n(\boldsymbol{x})|\}$$
$$+ \frac{1}{|\nabla U_n(\boldsymbol{x})|^4}\{2M|\nabla U_n(\boldsymbol{x})|^3|\boldsymbol{v}|^2\},$$

whence

$$|\mathfrak{U}(\boldsymbol{x}, \boldsymbol{v})\max\{0, (\nabla U_n(\boldsymbol{x}), \boldsymbol{v}))\}| \le \frac{C|\boldsymbol{v}|}{|\nabla U_n(\boldsymbol{x})|}\{M|\nabla U_n(\boldsymbol{x})||\boldsymbol{v}|^2 + M|\boldsymbol{v}|^2|\nabla U_n(\boldsymbol{x})|\}$$
$$+ \frac{C|\boldsymbol{v}|}{|\nabla U_n(\boldsymbol{x})|^3}\{M|\nabla U_n(\boldsymbol{x})|^3|\boldsymbol{v}|^2\} \le CM|\boldsymbol{v}|^3.$$

Thus overall,

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\Re f(\boldsymbol{x}+t\boldsymbol{v}, \boldsymbol{v})\Big|_{t=0}\max\{0, (\nabla U_n(\boldsymbol{x}), \boldsymbol{v})\}\right| \le \|\partial_x f\|_\infty|\boldsymbol{v}|^2|\nabla U_n(\boldsymbol{x})| + CM\|\partial_v f\|_\infty|\boldsymbol{v}|^3.$$

On the other hand

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{x}+t\boldsymbol{v}, \boldsymbol{v})\Big|_{t=0}\max\{0, (\nabla U_n(\boldsymbol{x}), \boldsymbol{v}))\}\right| \le \|\partial_x f\|_\infty|\nabla U(\boldsymbol{x})||\boldsymbol{v}|^2.$$

Thus overall we have that, using the fact that $(V_1, \ldots, V_n)$ are i.i.d. standard Gaussians and Lemma A.3 in the Appendix

$$\pi[|J_2|] \le C\mathbb{E}[|\nabla U_n(\boldsymbol{X})|]\mathbb{E}[|\boldsymbol{V}^2|] + CM\mathbb{E}[|\boldsymbol{V}|^3]$$
$$\le CM^{1/2}n^{3/2} + CMn^{3/2} = O(Mn^{3/2})$$

and thus we have that $\pi[|\mathcal{A}_n^{(1)}\mathcal{A}_n^{(2)} f|] = O(Mn^{3/2})$.

For the final term, since $Qf(\boldsymbol{x}, \boldsymbol{v}) = Qf(x_1, v_1)$ we have

$$\mathcal{A}_n^{(1)}\mathcal{A}_n^{(3)} f(\boldsymbol{x}, \boldsymbol{v}) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}_n^{(3)} f(\boldsymbol{x}+t\boldsymbol{v}, \boldsymbol{v})\Big|_{t=0}$$
$$= \lambda_{\mathrm{ref}}\frac{\mathrm{d}}{\mathrm{d}t}\big[Qf(x_1 + tv_1, v_1) - f(x_1 + tv_1, v_1)\big]\Big|_{t=0}$$
$$= \lambda_{\mathrm{ref}}\big[Q(\partial_x f)(x_1 + tv_1, v_1) - \partial_x f(x_1, v_1)\big]v_1,$$

by an application of dominated convergence. We can easily see from the above that $\pi[\mathcal{A}_n^{(1)}\mathcal{A}_n^{(3)} f] = O(1)$ as $n \to \infty$.

Case $i = 2$. For the first term $\mathcal{A}_n^{(2)}\mathcal{A}_n^{(1)}f$, notice first that since $f(\boldsymbol{x}, \boldsymbol{v}) = f(x_1, v_1)$ we have

$$\mathcal{A}_n^{(1)}f(\boldsymbol{x}, \boldsymbol{v}) = \partial_x f(x_1, v_1)v_1 =: h(x_1, v_1).$$

Therefore

$$\mathfrak{R}_n h(\boldsymbol{x}, \boldsymbol{v}) = \partial_x f\biggl(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{x})\biggr)\biggl(v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{x})\biggr),$$

whence

$$\begin{aligned}\mathfrak{R}_n h(\boldsymbol{x}, \boldsymbol{v}) - h(\boldsymbol{x}, \boldsymbol{v}) &= v_1\bigl[\mathfrak{R}_n\partial_x f(\boldsymbol{x}, \boldsymbol{v}) - \partial_x f(\boldsymbol{x}, \boldsymbol{v})\bigr] \\ &\quad - 2\mathfrak{R}_n\partial_x f(\boldsymbol{x}, \boldsymbol{v})\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{x}),\end{aligned}$$

and thus

$$\begin{aligned}\mathbb{E}\bigl|\mathcal{A}_n^{(2)}\mathcal{A}_n^{(1)}f(\boldsymbol{X}, \boldsymbol{V})\bigr| &\le \mathbb{E}\bigl[\bigl|(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\bigr| \times |V_1| \times \bigl|\mathfrak{R}_n\partial_x f(\boldsymbol{X}, \boldsymbol{V}) - \partial_x f(\boldsymbol{X}, \boldsymbol{V})\bigr|\bigr] \\ &\quad + 2\mathbb{E}\biggl[\bigl|(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\bigr| \times \biggl|\mathfrak{R}_n\partial_x f(\boldsymbol{X}, \boldsymbol{V})\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}\partial_1 U_n(\boldsymbol{x})\biggr|\biggr] \\ &\le \bigl(\|\mathfrak{R}_n\partial_x f\|_\infty + \|\partial_x f\|_\infty\bigr)\mathbb{E}\bigl[|\boldsymbol{V}| \times |V_1|\bigr]\mathbb{E}\bigl[|\nabla U_n(\boldsymbol{X})|\bigr] \\ &\quad + 2\|\mathfrak{R}_n\partial_x f\|_\infty\mathbb{E}\biggl[\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})^2}{|\nabla U_n(\boldsymbol{X})|^2}\bigl|\partial_1 U_n(\boldsymbol{x})\bigr|\biggr] \\ &\le \bigl(\|\mathfrak{R}_n\partial_x f\|_\infty + \|\partial_x f\|_\infty\bigr)\mathbb{E}\bigl[|\boldsymbol{V}| \times |V_1|\bigr]\mathbb{E}\bigl[|\nabla U_n(\boldsymbol{X})|\bigr] \\ &\quad + 2\|\mathfrak{R}_n\partial_x f\|_\infty\mathbb{E}\bigl[\bigl|\partial_1 U_n(\boldsymbol{X})\bigr|\bigr],\end{aligned}$$

where for the second term we used the tower law and the fact that conditionally on $\boldsymbol{X}$, $(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})$ is Gaussian with mean 0 and variance $|\nabla U_n(\boldsymbol{X})|^2$. Using the Cauchy–Schwarz inequality and Lemma A.3 from the Appendix we have

$$\begin{aligned}\mathbb{E}\bigl|\mathcal{A}_n^{(2)}\mathcal{A}_n^{(1)}f(\boldsymbol{X}, \boldsymbol{V})\bigr| &\le \bigl(\|\mathfrak{R}_n\partial_x f\|_\infty + \|\partial_x f\|_\infty\bigr)CM\sqrt{n}\mathbb{E}\bigl[|V_1|^2\bigr]^{1/2}\mathbb{E}\bigl[|\boldsymbol{V}|^2\bigr]^{1/2} \\ &\quad + 2\|\mathfrak{R}_n\partial_x f\|_\infty\mathbb{E}\bigl[\bigl|\nabla U_n(\boldsymbol{X})\bigr|\bigr] = O(Mn).\end{aligned}$$

For the next term $\mathcal{A}_n^{(2)}\mathcal{A}_n^{(2)}f$ first we write

$$\mathcal{A}_n^{(2)}\mathcal{A}_n^{(2)}f(\boldsymbol{x}, \boldsymbol{v}) = \max\bigl\{0, \bigl(\nabla U_n(\boldsymbol{x}), \boldsymbol{v}\bigr)\bigr\}\bigl[\mathfrak{R}_n\mathcal{A}_n^{(2)}f(\boldsymbol{x}, \boldsymbol{v}) - \mathcal{A}_n^{(2)}f(\boldsymbol{x}, \boldsymbol{v})\bigr].$$

Then notice that

$$\begin{aligned}\mathfrak{R}_n\mathcal{A}_n^{(2)}f(\boldsymbol{x}, \boldsymbol{v}) &= \max\biggl\{0, \biggl(\nabla U_n(\boldsymbol{x}), \boldsymbol{v} - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\nabla U_n(\boldsymbol{x})\biggr)\biggr\} \\ &\quad \times \biggl[\mathfrak{R}_n f\biggl(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{X})\biggr) \\ &\qquad - f\biggl(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{X})\biggr)\biggr] \\ &= \max\bigl\{0, \bigl(\nabla U_n(\boldsymbol{x}), -\boldsymbol{v}\bigr)\bigr\} \\ &\quad \times \biggl[\mathfrak{R}_n f\biggl(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{X})\biggr) \\ &\qquad - f\biggl(x_1, v_1 - 2\frac{(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})}{|\nabla U_n(\boldsymbol{x})|^2}\partial_1 U_n(\boldsymbol{X})\biggr)\biggr],\end{aligned}$$

and therefore that

$$|\mathfrak{R}_n \mathcal{A}_n^{(2)} f(\boldsymbol{x}, \boldsymbol{v}) - \mathcal{A}_n^{(2)} f(\boldsymbol{x}, \boldsymbol{v})| \le 2\|f\|_\infty |(\nabla U_n(\boldsymbol{x}), \boldsymbol{v})|,$$
$$|\mathcal{A}_n^{(2)} \mathcal{A}_n^{(2)} f(\boldsymbol{x}, \boldsymbol{v})| \le 2\|f\|_\infty (\nabla U_n(\boldsymbol{x}), \boldsymbol{v})^2.$$

Thus

$$\mathbb{E}|\mathcal{A}_n^{(2)} \mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})| \le C\|f\|_\infty \mathbb{E}\big[(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})^2\big]$$
$$\le C\|f\|_\infty \mathbb{E}\big\{\mathbb{E}\big[(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})^2 | \boldsymbol{X}\big]\big\}$$

using the fact that conditionally on $\boldsymbol{X}$, $(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})$ is Gaussian

$$= C\|f\|_\infty \mathbb{E}\big\{|\nabla U_n(\boldsymbol{X})|^2\big\} = O(Mn)$$

from Lemma A.3 in the Appendix.

Next we consider the term $\mathcal{A}_n^{(2)} \mathcal{A}_n^{(3)} f$. Since f is bounded, it easily follows that $\mathcal{A}_n^{(3)} f$ is also bounded and therefore that

$$|\mathcal{A}_n^{(2)} \mathcal{A}_n^{(3)} f(\boldsymbol{x}, \boldsymbol{v})| = \max\big\{0, (\nabla U_n(\boldsymbol{x}), \boldsymbol{v})\big\} |\mathfrak{R}_n \mathcal{A}_n^{(3)} f(\boldsymbol{x}, \boldsymbol{v}) - \mathcal{A}_n^{(3)} f(\boldsymbol{x}, \boldsymbol{v})|$$
$$\le 2\lambda_{\mathrm{ref}} \|f\|_\infty \max\big\{0, (\nabla U_n(\boldsymbol{x}), \boldsymbol{v})\big\}.$$

Therefore

$$\mathbb{E}|\mathcal{A}_n^{(2)} \mathcal{A}_n^{(3)} f(\boldsymbol{X}, \boldsymbol{V})| \le C\mathbb{E}|(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})| \le C\mathbb{E}\big[(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})^2\big]^{1/2} = O\big(M^{1/2} n^{1/2}\big),$$

from Lemma A.3 and calculations similar to the previous term.

Case $i = 3$. The first term to consider is

$$\mathcal{A}_n^{(3)} \mathcal{A}_n^{(1)} f(\boldsymbol{x}, \boldsymbol{v})$$
$$= \lambda_{\mathrm{ref}} \big[Q\mathcal{A}_n^{(1)} f(\boldsymbol{x}, \boldsymbol{v}) - \mathcal{A}_n^{(1)} f(\boldsymbol{x}, \boldsymbol{v})\big]$$
$$= \lambda_{\mathrm{ref}} \int \big[\mathcal{A}_n^{(1)} f\big(x_1, \alpha v_1 + \sqrt{1-\alpha^2}\xi\big) - \mathcal{A}_n^{(1)} f(\boldsymbol{x}, \boldsymbol{v})\big]\phi(\xi)\,\mathrm{d}\xi$$
$$= \lambda_{\mathrm{ref}} \int \big[\partial_x f\big(x_1, \alpha v_1 + \sqrt{1-\alpha^2}\xi\big)\big(\alpha v_1 + \sqrt{1-\alpha^2}\xi\big) - \partial_x f(x_1, v_1) v_1\big]\phi(\xi)\,\mathrm{d}\xi,$$

where ϕ denotes the standard normal density. Since $\|\partial_x f\|_\infty < \infty$ we have

$$\mathbb{E}|\mathcal{A}_n^{(3)} \mathcal{A}_n^{(1)} f(\boldsymbol{X}, \boldsymbol{V})| \le \lambda_{\mathrm{ref}} \|\partial_x f\|_\infty \mathbb{E}\big[|\alpha V_1 + \sqrt{1-\alpha^2}\xi| + |V_1|\big] = O(1),$$

as $n \to \infty$.

For the second term we have, using Jensen's inequality on the Markov kernel Q,

$$\mathbb{E}|\mathcal{A}_n^{(3)} \mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})| \le \lambda_{\mathrm{ref}} \mathbb{E}\big[|Q\mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})|\big] + \lambda_{\mathrm{ref}} \mathbb{E}\big[|\mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})|\big]$$
$$\le \lambda_{\mathrm{ref}} \mathbb{E}\big[Q\big(|\mathcal{A}_n^{(2)} f|\big)(\boldsymbol{X}, \boldsymbol{V})\big] + \lambda_{\mathrm{ref}} \mathbb{E}\big[|\mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})|\big].$$

At this point notice that Q is π_n-invariant and therefore

$$\mathbb{E}\big[Q\big(|\mathcal{A}_n^{(2)} f|\big)(\boldsymbol{X}, \boldsymbol{V})\big] = \mathbb{E}\big[|\mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})|\big],$$

whence we conclude that

$$\mathbb{E}|\mathcal{A}_n^{(3)} \mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})| \le 2\lambda_{\mathrm{ref}} \mathbb{E}\big[|\mathcal{A}_n^{(2)} f(\boldsymbol{X}, \boldsymbol{V})|\big]$$
$$\le 4\lambda_{\mathrm{ref}} \|f\|_\infty \mathbb{E}\big[|(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})|\big]$$
$$= 4\sqrt{\frac{2}{\pi}} \lambda_{\mathrm{ref}} \|f\|_\infty \mathbb{E}\big[|\nabla U_n(\boldsymbol{X})|\big] = O\big(M^{1/2} n^{1/2}\big),$$

using Lemma A.3 and the fact that conditionally on $\boldsymbol{X}$, $(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})$ is a mean zero Gaussian with variance $|\nabla U_n(\boldsymbol{X})|^2$.

Finally, by similar arguments as above the last term is given by

$$\begin{aligned}\mathbb{E}|\mathcal{A}_n^{(3)}\mathcal{A}_n^{(3)} f(\boldsymbol{X}, \boldsymbol{V})| &\le 2\lambda_{\mathrm{ref}}\mathbb{E}\big[|\mathcal{A}_n^{(3)} f(\boldsymbol{X}, \boldsymbol{V})|\big] \\ &\le 4\lambda_{\mathrm{ref}}^2 \|f\|_\infty = O(1).\end{aligned}$$

Overall we have shown that the error term defined in (3.13) satisfies

$$\mathcal{R}_n = \frac{\epsilon_n}{2}\mathbb{E}\big[|\mathcal{A}_n\mathcal{A}_n f(\boldsymbol{Z}_n(0))|\big] = O(Mn^{3/2}\epsilon_n) = o(1), \tag{3.14}$$

since we have chosen ϵ_n such that $\epsilon_n n^2 \to 0$, as $n \to \infty$.

Main term. Having controlled the error term, we now focus on the main term given by

$$\mathbb{E}\big[|\mathbb{E}[\widetilde{\mathcal{A}}_n f(\boldsymbol{Z}_n(0))|\mathcal{G}_0^n] - \mathcal{A}f(Z_n^{(1)}(0))|\big],$$

where we recall that $\widetilde{\mathcal{A}}_n$ is the extended generator. Notice that for $f(\boldsymbol{x}, \boldsymbol{v}) = f(x_1, v_1)$,

$$\begin{aligned}\mathcal{A}_n f(\boldsymbol{x}, \boldsymbol{v}) &= \partial_x f(x_1, v_1)v_1 + \max\{0, (\nabla U_n(\boldsymbol{x}), \boldsymbol{v})\}\big[\mathfrak{R}_n f(\boldsymbol{x}, \boldsymbol{v}) - f(\boldsymbol{x}, \boldsymbol{v})\big] \\ &\quad + \lambda_{\mathrm{ref}}\big[Qf(x_1, v_1) - f(x_1, v_1)\big], \\ \mathcal{A}f(x_1, v_1) &= \partial_x f(x_1, v_1)v_1 - \partial_v f(x_1, v_1)W'(x_1) + \lambda_{\mathrm{ref}}\big[Qf(x_1, v_1) - f(x_1, v_1)\big],\end{aligned}$$

and thus the first and third terms are in fact identical and will cancel out. We thus only have to consider the difference of the second terms. We apply a first order Taylor expansion

$$\begin{aligned}&\mathbb{E}\big[\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\}\big[\mathfrak{R}_n f(\boldsymbol{X}, \boldsymbol{V}) - f(\boldsymbol{X}, \boldsymbol{V})\big]|\mathcal{G}_0^n\big] \\ &\quad = \mathbb{E}\Bigg[\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\} \\ &\qquad \times \Bigg[f\bigg(X_1, V_1 - 2\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}\partial_1 U_n(\boldsymbol{X})\bigg) - f(X_1, V_1)\Bigg]\Big|\mathcal{G}_0^n\Bigg] \\ &\quad = \mathbb{E}\Bigg[\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\} \\ &\qquad \times \partial_v f(X_1, V_1)\bigg\{-2\frac{(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}\partial_1 U_n(\boldsymbol{X})\bigg\}\Big|\mathcal{G}_0^n\Bigg] + \mathcal{E}_1,\end{aligned}$$

where $\mathcal{E}_1$ is the remainder. At this point notice that, by the tower law and the fact that $(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})$ is Gaussian conditionally on $\boldsymbol{X}$,

$$\begin{aligned}\mathbb{E}|\mathcal{E}_1| &\le \|\partial_v f\|_\infty \mathbb{E}\bigg[\frac{|(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})|^3|\partial_1 U_n(\boldsymbol{X}|)}{|\nabla U_n(\boldsymbol{X})|^4}\bigg] \\ &= \|\partial_v f\|_\infty \mathbb{E}\bigg\{\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|^4}\mathbb{E}\big[|(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})|^3|\boldsymbol{X}\big]\bigg\} \\ &\le C\|\partial_v f\|_\infty \mathbb{E}\bigg\{\frac{|\partial_1 U_n(\boldsymbol{X})||\nabla U_n(\boldsymbol{X})|^{3/2}}{|\nabla U_n(\boldsymbol{X})|^4}\bigg\} \qquad (3.15) \\ &\le C\|\partial_v f\|_\infty \mathbb{E}\bigg\{\frac{|\nabla U_n(\boldsymbol{X})|^{5/2}}{|\nabla U_n(\boldsymbol{X})|^4}\bigg\} = C\|\partial_v f\|_\infty \mathbb{E}\bigg\{\frac{1}{|\nabla U_n(\boldsymbol{X})|^{3/2}}\bigg\} \\ &\le C\|\partial_v f\|_\infty \bigg[\frac{1}{(nm)^{3/2}} + \bigg(\frac{\sqrt{M}}{m}\bigg)^{3/2}\exp\bigg(-\frac{nm^2}{4M^2}\bigg) + \frac{m^{3/2}}{2^n M^{3/4} n}\bigg]\end{aligned}$$

by Lemma A.5 in the Appendix, which tends to 0 as $n \to \infty$.

Finally, having controlled the error terms, to complete the proof of (3.9), it remains to show that the following term vanishes:

$$\mathbb{E}_\pi\Big[\big|\partial_v f(X_1,V_1)\big| \times \Big|\mathbb{E}\Big[\max\{0,(\nabla U_n(\boldsymbol{X}),\boldsymbol{V})\}\Big(\frac{-2(\nabla U_n(\boldsymbol{X}),\boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}\Big)\partial_1 U_n(\boldsymbol{X})\Big|X_1,V_1\Big] - W'(X_1)\Big|\Big].$$

First notice that, since $V_2,\dots,V_n$ are independent of V_1 and $\boldsymbol{X}$, we can write

$$\begin{aligned}
&I(X_1,V_1)\\
&:=\mathbb{E}\Big[\max\{0,(\nabla U_n(\boldsymbol{X}),\boldsymbol{V})\}\Big(\frac{(\nabla U_n(\boldsymbol{X}),\boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}\Big)\partial_1 U_n(\boldsymbol{X})\Big|X_1,V_1\Big]\\
&=\mathbb{E}\Big\{\frac{\partial_1 U_n(\boldsymbol{X})}{|\nabla U_n(\boldsymbol{X})|^2}\mathbb{E}\big[\max\{0,(\nabla U_n(\boldsymbol{X}),\boldsymbol{V})\}^2|\boldsymbol{X},V_1\big]\Big|X_1,V_1\Big\}\\
&=\mathbb{E}\Big\{\frac{\partial_1 U_n(\boldsymbol{X})}{|\nabla U_n(\boldsymbol{X})|^2}\mathbb{E}\Big[\max\Big\{0,\partial_1 U_n(\boldsymbol{X})V_1+\sqrt{\sum_{j=2}^n[\partial_j U_n(\boldsymbol{X})]^2}\times\xi\Big\}^2\Big|\boldsymbol{X},V_1\Big]\Big|X_1,V_1\Big\}\\
&=\mathbb{E}\Big\{\frac{\partial_1 U_n(\boldsymbol{X})}{|\nabla U_n(\boldsymbol{X})|^2}\max\Big\{0,\partial_1 U_n(\boldsymbol{X})V_1+\sqrt{\sum_{j=2}^n[\partial_j U_n(\boldsymbol{X})]^2}\times\xi\Big\}^2\Big|X_1,V_1\Big\},
\end{aligned}$$

where ξ is a standard Gaussian random variable, independent from $\boldsymbol{X}$ and V_1. Continuing we have

$$I(X_1,V_1)=\mathbb{E}\Big\{\frac{\partial_1 U_n(\boldsymbol{X})}{|\nabla U_n(\boldsymbol{X})|^2}\max\Big\{0,\sqrt{\sum_{j=2}^n[\partial_j U_n(\boldsymbol{X})]^2}\times\xi\Big\}^2\Big|X_1,V_1\Big\}+\mathcal{E}_2(X_1,V_1)$$

where

$$\begin{aligned}
\mathcal{E}_2(X_1,V_1)&\le C\mathbb{E}\Big\{\frac{|\partial_1 U_n(\boldsymbol{X})|^3}{|\nabla U_n(\boldsymbol{X})|^2}\Big|X_1,V_1\Big\}+C\mathbb{E}\Big\{\frac{|\partial_1 U_n(\boldsymbol{X})|^2}{|\nabla U_n(\boldsymbol{X})|}\Big|X_1,V_1\Big\}\\
&=:\mathcal{E}_{2,1}(X_1,V_1)+\mathcal{E}_{2,2}(X_1,V_1).
\end{aligned}$$

We control the first term using the Cauchy–Schwarz inequality as follows:

$$\mathbb{E}[\mathcal{E}_{2,1}(X_1,V_1)]\le C\mathbb{E}\Big\{\frac{|\partial_1 U_n(\boldsymbol{X})|^4}{|\nabla U_n(\boldsymbol{X})|^4}\Big\}^{1/2}\mathbb{E}\{|\partial_1 U_n(\boldsymbol{X})|^2\}^{1/2}$$

and since $|\partial_1 U_n(\boldsymbol{x})|^2/|\nabla U_n(\boldsymbol{x})|^2\le 1$

$$\begin{aligned}
&\le C\mathbb{E}\Big\{\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|}\Big\}^{1/2}\mathbb{E}\{|\partial_1 U_n(\boldsymbol{X})|^2\}^{1/2}\\
&\le CM^{1/2}\Big[\frac{M^2}{m^2\sqrt{n}}+\frac{M^2\mathbb{E}|X_1|}{m^{3/2}n^{1/2}}+\frac{M}{m}\sqrt{\frac{\log n}{n}}+\frac{1}{n}\Big]^{1/2}
\end{aligned}\tag{3.16}$$

by Lemmas A.4, A.3 in the Appendix, where we used the fact that by Assumption 1 we have that $\mathbb{E}|X_1|<\infty$ (this follows, e.g., by the $\mathbb{L}^1$ Poincaré inequality applied on the function $f(X)=X_1$; see Corollary 1.9 of [2]).

For the second error term we have, again using the Cauchy–Schwarz inequality,

$$
\begin{aligned}
\mathbb{E}\big[\mathcal{E}_{2,2}(X_1, V_1)\big] &= C\mathbb{E}\left\{\frac{|\partial_1 U_n(\boldsymbol{X})|^2}{|\nabla U_n(\boldsymbol{X})|}\right\} \\
&\le \mathbb{E}\left\{\frac{|\partial_1 U_n(\boldsymbol{X})|^2}{|\nabla U_n(\boldsymbol{X})|^2}\right\}^{1/2} \mathbb{E}\big\{\big|\partial_1 U_n(\boldsymbol{X})\big|^2\big\}^{1/2} \\
&\le C\mathbb{E}\left\{\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|}\right\}^{1/2} \mathbb{E}\big\{\big|\partial_1 U_n(\boldsymbol{X})\big|^2\big\}^{1/2} \\
&\le CM^{1/2}\left[\frac{M^2}{m^2\sqrt{n}} + \frac{M^2\mathbb{E}|X_1|}{m^{3/2}n^{1/2}} + \frac{M}{m}\sqrt{\frac{\log n}{n}} + \frac{1}{n}\right]^{1/2}
\end{aligned} \tag{3.17}
$$

as before.

Finally notice that

$$
\begin{aligned}
&\mathbb{E}\left[\max\left\{0, \sum_{j=2}^n \partial_j U_n(\boldsymbol{X}) V_j\right\}\left(\frac{-2\sum_{j=2}^n \partial_j U_n(\boldsymbol{X}) V_j}{|\nabla U_n(\boldsymbol{X})|^2}\right)\partial_1 U_n(\boldsymbol{X})\Big| X_1, V_1\right] \\
&\quad = -2\mathbb{E}\left\{\mathbb{E}\left[\frac{\max\{0, \sum_{j=2}^n \partial_j U_n(\boldsymbol{X}) V_j\}^2}{|\nabla U_n(\boldsymbol{X})|^2}\partial_1 U_n(\boldsymbol{X})\Big|\boldsymbol{X}\right]\Big| X_1, V_1\right\} \\
&\quad = -2\mathbb{E}\left\{\mathbb{E}\left[\mathbb{1}\{\xi > 0\}\frac{(\sum_{j=2}^n [\partial_j U_n(\boldsymbol{X})]^2)\xi^2}{|\nabla U_n(\boldsymbol{X})|^2}\partial_1 U_n(\boldsymbol{X})\Big|\boldsymbol{X}\right]\Big| X_1, V_1\right\},
\end{aligned}
$$

where ξ is an independent standard Gaussian

$$
\begin{aligned}
&= -\mathbb{E}\left\{\partial_1 U_n(\boldsymbol{X})\frac{(\sum_{j=2}^n [\partial_j U_n(\boldsymbol{X})]^2)}{|\nabla U_n(\boldsymbol{X})|^2}\Big| X_1, V_1\right\} \\
&= -\mathbb{E}\big\{\partial_1 U_n(\boldsymbol{X}) | X_1, V_1\big\} + \mathcal{E}_3(X_1, V_1),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}\big[\big|\mathcal{E}_3(X_1, V_1)\big|\big] &\le \mathbb{E}\left[\frac{|\partial_1 U_n(\boldsymbol{X})|^3}{|\nabla U_n(\boldsymbol{X})|^2}\right] \\
&\le CM^{1/2}\left[\frac{M^2}{m^2\sqrt{n}} + \frac{M^2\mathbb{E}|X_1|}{m^{3/2}n^{1/2}} + \frac{M}{m}\sqrt{\frac{\log n}{n}} + \frac{1}{n}\right]^{1/2},
\end{aligned} \tag{3.18}
$$

by calculations similar to those for the error term $\mathcal{E}_{2,1}$. Finally

$$
\begin{aligned}
-\mathbb{E}\big\{\partial_1 U_n(\boldsymbol{X}) | X_1, V_1\big\} &= -\frac{\int \frac{\partial}{\partial x_1} U_n(x_1; x_{2:n}) \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n}}{\int \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n}} \\
&= -\frac{\frac{\partial}{\partial x_1}\int U_n(x_1; x_{2:n}) \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n}}{\int \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n}} \\
&= -\frac{\frac{\partial}{\partial x_1}\int \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n}}{\int \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n}} = -\frac{\partial}{\partial x_1}\log\int \mathrm{e}^{-U_n(x_1; x_{2:n})}\,\mathrm{d}x_{2:n} \\
&= -\frac{\partial}{\partial x_1}\log \mathrm{e}^{-W(x_1)} = W'(x_1).
\end{aligned}
$$

Overall we have shown that

$$
\mathbb{E}\left[\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\}\left(\frac{-2(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2}\right)\partial_1 U_n(\boldsymbol{X})\Big| X_1, V_1\right] = W'(X_1) + \mathcal{E}_2 + \mathcal{E}_3,
$$

where $\mathbb{E}[|\mathcal{E}_2|], \mathbb{E}[|\mathcal{E}_3|] \to 0$ as $n \to \infty$. Therefore we have

$$\mathbb{E}_\pi\Bigg[|\partial_v f(X_1, V_1)| \times \Bigg| \mathbb{E}\Bigg[\max\{0, (\nabla U_n(\boldsymbol{X}), \boldsymbol{V})\} \Bigg(\frac{-2(\nabla U_n(\boldsymbol{X}), \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2} \Bigg) \partial_1 U_n(\boldsymbol{X}) \Big| X_1, V_1 \Bigg] - W'(X_1) \Bigg| \Bigg] \to 0$$

as $n \to \infty$.

3.2.3. *Proof of* (3.11). Next we need to verify (3.11) for some $p > 1$ for which we proceed as follows:

$$\begin{aligned}
\mathbb{E}\Bigg[\Bigg(\int_0^T |\phi_n(t)|^p \, \mathrm{d}t \Bigg)^{1/p} \Bigg]^p &\le \mathbb{E}\Bigg[\int_0^T |\phi_n(t)|^p \, \mathrm{d}t \Bigg] = \int_0^T \mathbb{E}\big[|\phi_n(t)|^p \big] \mathrm{d}t \\
&= \int_0^T \mathbb{E}\big[|\epsilon_n^{-1} \mathbb{E}\{ f(Z_n^{(1)}(t+\epsilon_n)) - f(Z_n^{(1)}(t)) | \mathcal{G}_t^n \}|^p \big] \mathrm{d}t \\
&= \int_0^T \mathbb{E}\Bigg[\Bigg| \epsilon_n^{-1} \mathbb{E}\Bigg\{ \int_0^{\epsilon_n} (\widetilde{\mathcal{A}}_n f(Z_n^{(1)}(t+s)) + R_{t+s}) \, \mathrm{d}s \Big| \mathcal{G}_t^n \Bigg\} \Bigg|^p \Bigg] \mathrm{d}t
\end{aligned}$$

and using the fact that $\mathbb{E}[R_{t+s}|\mathcal{G}_t^n] = \mathbb{E}[\mathbb{E}[R_{t+s}|\mathcal{F}_t^n]|\mathcal{G}_t^n] = 0$,

$$= \int_0^T \mathbb{E}\Bigg[\Bigg| \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\{ \widetilde{\mathcal{A}}_n f(Z_n^{(1)}(t+s)) | \mathcal{G}_t^n \} \, \mathrm{d}s \Bigg|^p \Bigg] \mathrm{d}t$$

and by Jensen's inequality,

$$\begin{aligned}
&\le \int_0^T \mathbb{E}\Bigg[\epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\{ |\widetilde{\mathcal{A}}_n f(Z_n^{(1)}(t+s))|^p | \mathcal{G}_t^n \} \, \mathrm{d}s \Bigg] \mathrm{d}t \\
&= \int_0^T \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[\mathbb{E}\{ |\widetilde{\mathcal{A}}_n f(Z_n^{(1)}(t+s))|^p | \mathcal{G}_t^n \} \big] \mathrm{d}s \, \mathrm{d}t \\
&= \int_0^T \epsilon_n^{-1} \int_0^{\epsilon_n} \mathbb{E}\big[|\widetilde{\mathcal{A}}_n f(Z_n^{(1)}(t+s))|^p \big] \mathrm{d}s \, \mathrm{d}t \\
&= T \mathbb{E}\big[|\widetilde{\mathcal{A}}_n f(Z_n^{(1)}(0))|^p \big],
\end{aligned}$$

by stationarity. Next recalling the decomposition of $\widetilde{\mathcal{A}_n}$ into $\mathcal{A}_n^{(i)}$, $i = 1, 2, 3$ notice that

$$\sup_{x,v} |\mathcal{A}_n^{(1)} f(x, v)| = \sup_{x,v} |\partial_x f(x, v) v| < \infty,$$

since $(x, v) \mapsto \partial_x f(x, v) v$ is continuous and has compact support, since f has compact support. Similarly, it follows easily that $\|\mathcal{A}_n^{(3)} f\|_\infty < \infty$ and therefore the only term we have to control corresponds to $\mathcal{A}_n^{(2)}$. For this term notice that

$$\begin{aligned}
&\mathbb{E}\big[|\mathcal{A}_n^{(2)} f(Z_n^{(1)}(0))|^p \big] \\
&= \mathbb{E}\Bigg[\Bigg| \max\{0, (\nabla U_n(\boldsymbol{X}, \boldsymbol{V})\} \Bigg[f\Bigg(X_1, V_1 - 2 \frac{(\nabla U_n(\boldsymbol{X}, \boldsymbol{V})}{|\nabla U_n(\boldsymbol{X})|^2} \partial_1 U_n(\boldsymbol{X}) \Bigg) - f(X_1, V_1) \Bigg] \Bigg|^p \Bigg] \\
&\le 2^p \|\partial_v f\|_\infty \mathbb{E}\Bigg[\frac{|(\nabla U_n(\boldsymbol{X}, \boldsymbol{V})|^{2p} |\partial_1 U_n(\boldsymbol{X})|^p}{|\nabla U_n(\boldsymbol{X})|^{2p}} \Bigg] \\
&\le 2^p \|\partial_v f\|_\infty \mathbb{E}\Bigg\{ \mathbb{E}\Bigg[\frac{|(\nabla U_n(\boldsymbol{X}, \boldsymbol{V})|^{2p} |\partial_1 U_n(\boldsymbol{X})|^p}{|\nabla U_n(\boldsymbol{X})|^{2p}} \Big| \boldsymbol{X} \Bigg] \Bigg\} \\
&\le 2^p \|\partial_v f\|_\infty \mathbb{E}\Bigg\{ \mathbb{E}\Bigg[\frac{|\nabla U_n(\boldsymbol{X})|^p |\partial_1 U_n(\boldsymbol{X})|^p}{|\nabla U_n(\boldsymbol{X})|^{2p}} \Big| \boldsymbol{X} \Bigg] \Bigg\} \le 2^p \|\partial_v f\|_\infty = O(1).
\end{aligned}$$

3.2.4. *Proofs of* (3.6) *and* (3.7). Notice that (3.6) follows immediately since $\|f\|_\infty < \infty$, whereas (3.7) follows from calculations similar to the ones used to prove (3.11).

4. Proofs of Wasserstein rates.

4.1. *Proof of Theorem 3.* Let $\widetilde{X}(t) := X^{(2)}(t) - X^{(1)}(t)$ and $\widetilde{V}(t) := V^{(2)}(t) - V^{(1)}(t)$ denote the differences between the two paths in position and momentum. Ignoring for the moment the refreshment events, $(\widetilde{X}(t), \widetilde{V}(t))$ will evolve according to the Hamiltonian dynamics, that is,

$$\begin{aligned}
&\widetilde{X}'(t) = \widetilde{V}(t),\\
&\widetilde{V}'(t) = -\big(\nabla U\big(X^{(2)}(t)\big) - \nabla U\big(X^{(1)}(t)\big)\big) = -H(t)\widetilde{X}(t) \quad \text{where}\\
&\mathcal{H}(t) := \int_{s=0}^{1} \nabla^2 U\big(sX^{(1)}(t) + (1-s)X^{(2)}(t)\big)\,\mathrm{d}s.
\end{aligned} \tag{4.1}$$

By convexity, we can see that $\mathcal{H}(t)$ satisfies that $mI \preceq \mathcal{H}(t) \preceq MI$ where I denotes the identity matrix, where we write $A \preceq B$ to denote that $B - A$ is positive definite. The effect of the generator $L_{1,2}$ on $|\widetilde{X}(t)|^2$, $\langle \widetilde{X}(t), \widetilde{V}(t)\rangle$ and $|\widetilde{V}(t)|^2$ is given by

$$\begin{aligned}
L_{1,2}\big|\widetilde{X}(t)\big|^2 &= 2\big\langle \widetilde{X}(t), \widetilde{V}(t)\big\rangle,\\
L_{1,2}\widetilde{X}(t)^T\widetilde{V}(t) &= \big|\widetilde{V}(t)\big|^2 - \widetilde{X}(t)^T\mathcal{H}(t)\widetilde{X}(t) - \lambda_{\mathrm{ref}}(1-\alpha)\widetilde{X}(t)^T\widetilde{V}(t),\\
L_{1,2}\big|\widetilde{V}(t)\big|^2 &= -2\widetilde{V}(t)^T\mathcal{H}(t)\widetilde{X}(t) - \lambda_{\mathrm{ref}}\big(1-\alpha^2\big)\big|\widetilde{V}(t)\big|^2.
\end{aligned} \tag{4.2}$$

The claim of Theorem 3 is equivalent to showing that $-\mu \cdot d_A^2(Z_1(t), Z_2(t)) - L_{1,2}d_A^2(Z_1(t), Z_2(t)) \geq 0$. This can be expressed as

$$\begin{aligned}
&-\mu \cdot d_A^2\big(Z_1(t), Z_2(t)\big) - L_{1,2}d_A^2\big(Z_1(t), Z_2(t)\big)\\
&\quad = -\mu a\big|\widetilde{X}(t)\big|^2 + 2\big[-\mu b + \lambda_{\mathrm{ref}}(1-\alpha)b - a\big]\widetilde{X}(t)^T\widetilde{V}(t)\\
&\qquad + \big[-c\mu + \lambda_{\mathrm{ref}}\big(1-\alpha^2\big)c - 2b\big]\big|\widetilde{V}(t)\big|^2\\
&\qquad + 2b\widetilde{X}(t)^T\mathcal{H}(t)\widetilde{X}(t) + 2c\widetilde{V}(t)^T\mathcal{H}(t)\widetilde{X}(t).
\end{aligned}$$

Let

$$X := \begin{pmatrix} |\widetilde{X}(t)|^2 & \widetilde{X}(t)^T\widetilde{V}(t) \\ \widetilde{X}(t)^T\widetilde{V}(t) & |\widetilde{V}(t)|^2 \end{pmatrix}, \qquad P := \begin{pmatrix} \widetilde{X}(t)^T\mathcal{H}(t)\widetilde{X}(t) & \widetilde{V}(t)^T\mathcal{H}(t)\widetilde{X}(t) \\ \widetilde{V}(t)^T\mathcal{H}(t)\widetilde{X}(t) & \widetilde{V}(t)^T\mathcal{H}(t)\widetilde{V}(t) \end{pmatrix},$$

$$V := \begin{pmatrix} -\mu a & -a + b\lambda_{\mathrm{ref}}(1-\alpha) - \mu b \\ -a + b\lambda_{\mathrm{ref}}(1-\alpha) - \mu b & -c\mu + c\lambda_{\mathrm{ref}}\big(1-\alpha^2\big) - 2b \end{pmatrix}, \qquad W := \begin{pmatrix} 2b & c \\ c & 0 \end{pmatrix}.$$

We have

$$-\mu \cdot d_A^2\big(Z_1(t), Z_2(t)\big) - L_{1,2}d_A^2\big(Z_1(t), Z_2(t)\big) = \mathrm{Tr}(VX + WP),$$

so our goal is to show that $\mathrm{Tr}(VX + WP) \geq 0$ for all the possible X, P. Using the fact that $mI \preceq \mathcal{H}(t) \preceq MI$, we have $0 \preceq mX \preceq P \preceq MX$. Let $Y := P - mX$, and $Z := MX - P$, then $Y \succeq 0$, $Z \succeq 0$, and for $M > m$, we have

$$X = \frac{Y+Z}{M-m}, \qquad P = \frac{MY + mZ}{M-m},$$

and hence

$$\operatorname{Tr}(VX + WP) = \frac{1}{M-m}\big(\operatorname{Tr}\big((V+MW)Y + (V+mW)Z\big)\big).$$

When $M = m$, we have $\mathcal{H}(t) = MI$ and $P = MX$, hence

$$\operatorname{Tr}(VX + WP) = \operatorname{Tr}\big((V+MW)X\big).$$

Note that in both cases, $\operatorname{Tr}(VX + WP) \geq 0$ if both $V + MW \succeq 0$ and $V + mW \succeq 0$. This can be equivalently written as the following set of inequalities:

$$-\mu a + 2Mb \geq 0, \tag{4.3}$$

$$-\mu a + 2mb \geq 0, \tag{4.4}$$

$$-c\mu + c\lambda_{\mathrm{ref}}\big(1-\alpha^2\big) - 2b \geq 0, \tag{4.5}$$

$$\big(-a + b\lambda_{\mathrm{ref}}(1-\alpha) - \mu b + Mc\big)^2 \leq (-\mu a + 2Mb)\big(-c\mu + c\lambda_{\mathrm{ref}}\big(1-\alpha^2\big) - 2b\big), \tag{4.6}$$

$$\big(-a + b\lambda_{\mathrm{ref}}(1-\alpha) - \mu b + mc\big)^2 \leq (-\mu a + 2mb)\big(-c\mu + c\lambda_{\mathrm{ref}}\big(1-\alpha^2\big) - 2b\big). \tag{4.7}$$

These inequalities correspond to the diagonal elements and the determinants of $V + mW$ and $V + MW$ being nonnegative. As we have stated, let $\lambda_{\mathrm{ref}} = \frac{1}{1-\alpha^2}(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}})$, $\mu = \frac{(1+\alpha)m}{\sqrt{M+m}} - \frac{\alpha m^{3/2}}{2(M+m)}$. Moreover, let

$$\begin{aligned} a &:= 1, \\ b &:= \frac{1+\alpha-\alpha(\frac{m}{M+m})^{3/4} + \frac{3}{4}\frac{\alpha m}{M+m}}{2\sqrt{M+m}}, \\ c &:= \frac{1+\alpha-\frac{\alpha}{2}(\frac{m}{M+m})^{1/2}}{M+m}. \end{aligned} \tag{4.8}$$

Notice that by the change of variables $m \to 1$, $M \to M/m$, and updating a, b, c and μ and λ with these new values, inequalities (4.3)–(4.7) are kept invariant (they have this homogeneity property). Hence, without loss of generality, we can assume that $m = 1$. For the choice of a, b, c as in (4.8), the five inequalities can be shown to hold for every possible $0 \leq \alpha < 1$ and M using, for example, Mathematica. Hence the bound (2.11) follows.

Now we are going to show the Wasserstein bounds. Note that the matrix A satisfies that $\lambda_{\min}(A) = \frac{a+c-\sqrt{(a+c)^2-4(ac-b^2)}}{2}$ and $\lambda_{\max}(A) = \frac{a+c+\sqrt{(a+c)^2-4(ac-b^2)}}{2}$, hence by defining

$$W_{2,d_A}(\nu_1,\nu_2) = \Big(\inf_{X_1\sim\nu_1, X_2\sim\nu_2} d_A(X_1,X_2)^2\Big)^{1/2},$$

then using the assumption $b^2 < ac$, we have $W_2(\nu P^t,\pi)^2 \leq \frac{1}{\lambda_{\min}(A)} W_{2,d_A}(\nu P^t,\pi)^2$. Let $Z_1(0)$, $Z_2(0)$ be coupled according to the optimal coupling of ν and π according to W_2 distance satisfying that $\mathbb{E}(|Z_1(0) - Z_2(0)|^2) = W_2(\nu,\pi)^2$ (existence is shown by Theorem 4.1 of [59]). Using (2.11) along with Grönwall's lemma, and the definition of the Wasserstein distance, it follows that

$$\begin{aligned} W_2\big(\nu P^t,\pi\big)^2 &\leq \frac{1}{\lambda_{\min}(A)} W_{2,d_A}\big(\nu P^t,\pi\big)^2 \leq \frac{1}{\lambda_{\min}(A)}\mathbb{E}\big(d_A^2\big(Z_1(t),Z_2(t)\big)\big) \\ &\leq \frac{e^{-\mu t}}{\lambda_{\min}(A)}\mathbb{E}\big(d_A^2\big(Z_1(0),Z_2(0)\big)\big) \leq \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} e^{-\mu t} W_2(\nu,\pi)^2, \end{aligned}$$

hence (2.12) follows.

To show our L^2 bounds, we are also going to study the adjoint process $(P^t)^*$. Using the exact same coupling as before, the dynamics (4.1) ran backwards in time becomes

$$\begin{aligned}\widetilde{X}'(t) &= -\widetilde{V}(t),\\ \widetilde{V}'(t) &= \mathcal{H}(t)\widetilde{X}(t),\end{aligned} \tag{4.9}$$

with $\mathcal{H}(t)$ defined as in (4.1). For the velocity updates, forward in time we had $v' = \alpha v + \sqrt{1-\alpha^2}Z$ where $Z \sim N(0, I_d)$. Since in stationary we have $v, v' \sim N(0, I_d)$ and $\mathbb{E}(v(v')^T) = \rho I_d$, one can see that the updates backward in time are still the same. Hence the effect of the adjoint becomes

$$\begin{aligned}L_{1,2}^*\big|\widetilde{X}(t)\big|^2 &= -2\widetilde{X}(t)^T\widetilde{V}(t),\\ L_{1,2}^*\widetilde{X}(t)^T\widetilde{V}(t) &= \big|\widetilde{V}(t)\big|^2 - \widetilde{X}(t)^T\mathcal{H}(t)\widetilde{X}(t) + \lambda_{\mathrm{ref}}(1-\alpha)\widetilde{X}(t)^T\widetilde{V}(t),\\ L_{1,2}^*\big|\widetilde{V}(t)\big|^2 &= 2\widetilde{V}(t)^T\mathcal{H}(t)\widetilde{X}(t) - \lambda_{\mathrm{ref}}\big(1-\alpha^2\big)\big|\widetilde{V}(t)\big|^2.\end{aligned} \tag{4.10}$$

Notice that this is very similar to the forward case (4.2), except that we need to replace $\widetilde{V}(t)$ by $-\widetilde{V}(t)$. Based on this, by repeating the previous argument for $A' := \left(\begin{smallmatrix} a & -b \\ -b & c\end{smallmatrix}\right)$, we have

$$L_{1,2}^* d_{A'}^2\big(Z_1(t), Z_2(t)\big) \le -\mu \cdot d_{A'}^2\big(Z_1(t), Z_2(t)\big), \tag{4.11}$$

where a, b and c are defined as in (4.8).

Hence we have shown that the adjoint process is also a contraction with the same rate μ, but with respect to a different metric $d_{A'}$ instead of d_A used for the forward process. Now we are going to show that d_A^2 and $d_{A'}^2$ are equivalent up to a constant factor $C := \frac{ac+b^2+2\sqrt{acb^2}}{ac-b^2}$. Notice that for any $z_1, z_2 \in \mathbb{R}^{2d}$,

$$d_A^2(z_1, z_2)/C \le d_{A'}^2(z_1, z_2) \le d_A^2(z_1, z_2)\cdot C, \tag{4.12}$$

as long as $A \preceq CA'$ and $A' \preceq CA$, and by rearrangement, this is equivalent to

$$\begin{pmatrix} a(C-1) & -b(1+C) \\ -b(1+C) & c(C-1)\end{pmatrix} \succeq 0 \quad \text{and} \quad \begin{pmatrix} a(C-1) & b(C+1) \\ b(C+1) & c(C-1)\end{pmatrix} \succeq 0,$$

which holds for C defined as above.

For $f : \mathbb{R}^{2d} \to \mathbb{R}$, let

$$\|f\|_{\mathrm{Lip},d_A} := \sup_{z_1,z_2\in\mathbb{R}^{2d}, z_1\neq z_2} \frac{|f(z_1) - f(z_2)|}{d_A(z_1, z_2)},$$

be its Lipschitz coefficient with respect to the d_A distance. Then based on (2.11), (4.11) and (4.12), for any $t \ge 0$, $f : \mathbb{R}^{2d} \to \mathbb{R}$, have

$$\begin{aligned}\big\|(P^t)^* P^t f\big\|_{\mathrm{Lip},d_A} &\le \sqrt{C}\big\|(P^t)^* P^t f\big\|_{\mathrm{Lip},d_{A'}} \le \sqrt{C}\exp\left(-\frac{\mu t}{2}\right)\big\|P^t f\big\|_{\mathrm{Lip},d_{A'}}\\ &\le C\exp\left(-\frac{\mu t}{2}\right)\big\|P^t f\big\|_{\mathrm{Lip},d_A} \le C\exp(-\mu t)\|f\|_{\mathrm{Lip},d_A}.\end{aligned}$$

Based on Propositions 29 and 30 of [46] with $\kappa = 1 - C\exp(-\mu t)$, it follows that for any $t > \frac{\log(C)}{\mu}$, the reversible kernel $(P^t)^* P^t$ has as spectral radius of at most $C\exp(-\mu t)$. Thus for every $f \in L_0^2(\pi)$, we have

$$\big\|P^t f\big\|^2 = \big\langle f, (P^t)^* P^t f\big\rangle \le \|f\|\big\|(P^t)^* P^t f\big\| \le Ce^{-\mu t}\|f\|^2, \tag{4.13}$$

and the claim of the theorem follows by noticing that $\|P^t f\|^2 \le \|f\|^2$ for every $t \ge 0$.

REMARK 13. We note that for any given $\lambda_{\mathrm{ref}} > 0$, $\mu > 0$, the contraction rate of $d_A^2(Z_1(t), Z_2(t))$ is at least μ as long as there are constants a, b, c such that $a > 0$, $c > 0$, $b^2 < ac$ and inequalities (4.3)–(4.7) hold. Unfortunately due to the nonlinearity of these inequalities we did not manage to find an analytical expression for the largest possible μ for a given λ_{ref} (and then the largest possible μ for any λ_{ref}). The reader can possibly slightly improve these rates by numerical optimization for a given α, m and M. Note, however, that in our numerical experiments, it seems that the choices of λ_{ref} as stated leads to μ that is close to optimal in most of the domain $0 \le \alpha < 1$, and $0 < m \le M$ (i.e., if we increase μ by a few percent, typically there is no longer a $\lambda_{\mathrm{ref}} > 0$ and parameters a, b, c satisfying all of the inequalities).

4.2. *Proof of Proposition 4.* Assume without loss of generality that $m = 1$ (the general case can be obtained from this by rescaling). Let $D := \left(\begin{smallmatrix} aH & bI \\ bI & cI \end{smallmatrix}\right)$ be a block matrix. Then

$$d_D^2\big(Z_1(t), Z_2(t)\big) = a\widetilde{X}(t)^T H\widetilde{X}(t) + 2b\widetilde{X}(t)^T\widetilde{V}(t) + c\big|\widetilde{V}(t)\big|^2,$$

and the effect of the generator on these terms equal

$$L_{1,2}\widetilde{X}(t)^T H\widetilde{X}(t) = 2\widetilde{X}(t)^T H\widetilde{V}(t), \tag{4.14}$$

$$L_{1,2}\widetilde{X}(t)^T\widetilde{V}(t) = \big|\widetilde{V}(t)\big|^2 - \widetilde{X}(t)^T H\widetilde{X}(t) - \lambda_{\mathrm{ref}}(1-\alpha)\widetilde{X}(t)^T\widetilde{V}(t), \tag{4.15}$$

$$L_{1,2}\big|\widetilde{V}(t)\big|^2 = -2\widetilde{V}(t)^T H\widetilde{X}(t) - \lambda_{\mathrm{ref}}\big(1-\alpha^2\big)\big|\widetilde{V}(t)\big|^2. \tag{4.16}$$

We have

$$\begin{aligned}
&-\mu\cdot d_D^2\big(Z_1(t), Z_2(t)\big) - L_{1,2}d_D^2\big(Z_1(t), Z_2(t)\big)\\
&\quad= 2\big[-\mu b + \lambda_{\mathrm{ref}}(1-\alpha)b\big]\widetilde{X}(t)^T\widetilde{V}(t) + \big[-c\mu + \lambda_{\mathrm{ref}}\big(1-\alpha^2\big)c - 2b\big]\big|\widetilde{V}(t)\big|^2\\
&\qquad+ (2b - \mu a)\widetilde{X}(t)^T H\widetilde{X}(t) + 2(c-a)\widetilde{V}(t)^T H\widetilde{X}(t).
\end{aligned}$$

Let X and P be defined as in the proof of Theorem 3, and let

$$V := \begin{pmatrix} 0 & b\lambda_{\mathrm{ref}}(1-\alpha) - \mu b \\ b\lambda_{\mathrm{ref}}(1-\alpha) - \mu b & -c\mu + c\lambda_{\mathrm{ref}}\big(1-\alpha^2\big) - 2b \end{pmatrix}, \qquad W := \begin{pmatrix} 2b - \mu a & c-a \\ c-a & 0 \end{pmatrix}.$$

Then we have $-\mu\cdot d_D^2(Z_1(t), Z_2(t)) - L_{1,2}d_D^2(Z_1(t), Z_2(t)) = \mathrm{Tr}(VX + WP)$, and using the same argument as in the proof of Theorem 3, it follows that $\mathrm{Tr}(VX + WP) \ge 0$ if both $V + MW \succeq 0$ and $V + mW \succeq 0$. This can be verified (e.g., by Mathematica) for the choices $\lambda_{\mathrm{ref}} = 2\sqrt{m}/(1-\alpha)$, $\mu = \frac{\sqrt{m}}{3}$, $a = 1$, $b = \frac{1}{4}$, $c = 1$. The proof of (2.16) is analogous to the proof of (2.13). First we show that for $D' := \left(\begin{smallmatrix} aH & -bI \\ -bI & cI \end{smallmatrix}\right)$,

$$L_{1,2}^* d_{D'}^2\big(Z_1(t), Z_2(t)\big) \le -\mu\cdot d_{D'}^2\big(Z_1(t), Z_2(t)\big), \tag{4.17}$$

then use the same argument as previously.

5. Proof of Theorem 5. The generator of the RHMC process will be denoted by $\mathcal{A}$ and it is given for smooth enough functions by

$$\mathcal{A}f(x,v) = \langle\nabla_x f, v\rangle - \langle\nabla U, \nabla_v f\rangle + \lambda_{\mathrm{ref}}\big[Q_\alpha f(x,v) - f(x,v)\big],$$

where recall that $\alpha \in (0,1)$ and

$$Q_\alpha f(x,v) := \frac{1}{\sqrt{2\pi}^d}\int \mathrm{e}^{-\xi'\xi/2} f\big(x, \alpha v + \sqrt{1-\alpha^2}\xi\big)\,\mathrm{d}\xi.$$

Hypo-coercivity, exponential convergence and asymptotic variance. In the context of MCMC one is interested in optimising the computational resources needed to produce an estimate of a certain precision. For this reason we are also interested in understanding the asymptotic variance. Geometric ergodicity is enough to show that a large class of functions, determined by the Lyapunov function, have finite asymptotic variance. However, since the convergence rates are not explicit in the parameters of the process, geometric ergodicity often does not allow one to optimise the asymptotic variance.

Usually controlling the asymptotic variance for a large enough class of functions is closely related to establishing a *spectral gap*, that is showing that the $L^2(\pi)$ spectrum of the generator $\mathcal{L}$ lies in $\{z \in \mathbb{C} : \Re z \leq -\mu\}$, for some $\mu > 0$. In the reversible case, it is well known that geometric ergodicity is equivalent to having a spectral gap, but in the nonreversible case this is no longer true; see [34] and references therein (although it may be equivalent to a spectral gap on a different Banach space). For reversible processes, an L^2-spectral gap is also equivalent to *coercivity* of the associated Dirichlet form, that is, $\langle -\mathcal{L}f, f\rangle \geq \mu\|f\|^2$, for all $f \in L_0^2(\pi)$. Moreover, coercivity is equivalent to $\|P^t f\| \leq \mathrm{e}^{-\mu t}\|f\|$, for all $f \in L_0^2(\pi)$, for *all* Markov processes, whether reversible or not. For this reason, and perhaps abusively, coercivity is sometimes in the literature referred to as a spectral gap, or a spectral gap inequality. Another reason is that, an inequality of the form $\langle -\mathcal{L}f, f\rangle \geq \mu\|f\|^2$ is often easy to prove, for example, for diffusions, by rewriting the Dirichlet form in a form involving the Sobolev norm and then applying a Poincaré inequality.

Interestingly enough, however, for nonreversible processes it is possible that coercivity fails to hold, although we still have $\|P^t f\| \leq C\mathrm{e}^{-\mu t}\|f\|$, for all $f \in L_0^2(\pi)$, for some $C > 1$. This is not possible for reversible processes, since one can use spectral calculus to show that $\|P^t f\| \leq C\mathrm{e}^{-\mu t}\|f\|$, for all $f \in L_0^2(\pi)$ also implies the same inequality with $C \equiv 1$. This fact is actually observed for piecewise deterministic Markov processes such as the BPS and zig-zag samplers; see [7, 15, 49]. This class of processes also includes RHMC. Although geometric ergodicity has been established for BPS [23, 26], zig-zag (see [11, 30]) and RHMC [14], an easy calculation shows that, writing $\mathcal{L}$ for the generator of any of the above processes, we have $\langle \mathcal{L}f, f\rangle = 0$ for any function $f \in L^2(\pi)$ such that $f(x, v) = f(x)$, that is functions of the location only. The reason for this is that the Dirichlet form $\mathcal{E}(f, f) := \langle \mathcal{L}f, f\rangle$ only captures the symmetric part of the generator $\mathcal{L}$, which in these processes only affects the velocity component, whereas the location component is only affected by the anti-symmetric part of the generator. This means that although BPS, zig-zag and RHMC are geometrically ergodic, we certainly cannot have an inequality of the form $\langle -\mathcal{L}f, f\rangle \geq \mu\|f\|^2$ for all $f \in L_0^2(\pi)$. However, it may still be true that these processes admit a spectral gap in the classical sense, and in fact this has been shown for one-dimensional zig-zag in Bierkens and Lunel [9]. Notice, however, that in the nonreversible case, a classical spectral gap requires additional work, and potentially assumptions, to guarantee exponential decay of the semigroup; see [9], Section 4.2.

In fact, this situation arises very often in so called kinetic equations which include, for example, the underdamped Langevin processes. For such processes a range of methods have been developed recently that are widely termed as *hypocoercivity*; see [24, 32, 58] and references therein. In fact, such methods have already been applied to piecewise deterministic Markov processes; see [42]. Although this approach is often quite deep and involved, the underlying principle is that of adjusting the norm, or metric, in which the convergence is studied. This principle has been extremely successful recently, for example, in the convergence of HMC when log-concavity fails locally in [13]. In the case of hypocoercive estimates, the principle is to move from the L^2 norm to a stronger norm, usually some form of Sobolev norm.

5.1. *Strong continuity in $H^1(\pi)$.* We will establish that the abstract Cauchy problem

$$\frac{\partial u(t,z)}{\partial t} = \mathcal{A}u,$$
$$u(0,z) = f,$$

where the class of initial conditions f will be specified in the sequel, admits a unique solution in $H^1(\pi)$ given by $u(t,z) := P^t f(z)$. This will justify computing the time derivatives of $\langle\langle P^t f, P^t f\rangle\rangle$.

Before we proceed we will need to introduce some additional notation. We decompose the generator $\mathcal{A}$ of RHMC into its symmetric and antisymmetric component as follows:

$$\mathcal{A}f(x,v) = Bf(x,v) + \lambda_{\mathrm{ref}}(-S)f,$$

where

$$Bf := \langle \nabla_x f, v\rangle - \langle \nabla_v f, \nabla U\rangle, \qquad Sf := [I - Q_\alpha]f. \tag{5.1}$$

As before we write $\{P^t : t \geq 0\}$ for the semigroup of transition kernels of RHMC, but in this section we slightly change our point of view and consider it as a semigroup on $L^2(\pi)$, that is, $P^t : L^2(\pi) \to L^2(\pi)$. Its generator will be given by $\mathcal{A}$ for smooth enough functions.

In fact, even more is true as we will next show that P^t is also strongly continuous as a semigroup on $H^1(\pi)$. To see why, first recall that the anti-symmetric operator B generates the Hamiltonian flow $z \mapsto \Xi(t,z)$ with respect to $H(\boldsymbol{x},\boldsymbol{v}) = U(\boldsymbol{x}) + |\boldsymbol{v}|^2/2$. Let us write $\{T^t : t \geq 0\}$ for the semigroup generated by B, that is, $T^t f(z) = f(\Xi(t,z))$ for $z \in \mathcal{Z}$. Then given a smooth function $f \in H^1(\pi)$, from the chain rule we have

$$\nabla T^t f(z) = \nabla f\big(\Xi(t,z)\big)\nabla_z \Xi(t,z).$$

From the variational equations of the Hamiltonian dynamics (see Section 6.1.2 of [40]) and the upper bounds M and 1 of the Hessians of $U(\boldsymbol{x})$ and $\frac{\|\boldsymbol{v}\|^2}{2}$ it follows that for $C = \max(1, M)$, we have $\|\nabla_z \Xi(t,z)\| \leq \mathrm{e}^{Ct}$ for every $t \geq 0$. Using this, we conclude that

$$\begin{aligned}\|\nabla_x T^t f\|^2 + \|\nabla_v T^t f\|^2 &\leq \mathrm{e}^{2Ct} \iint \pi(\mathrm{d}z)\big[\big|\nabla_x f\big(\Xi(t,z)\big)\big|^2 + \big|\nabla_v f\big(\Xi(t,z)\big)\big|^2\big] \\ &= \mathrm{e}^{2Ct} \iint \pi(\mathrm{d}z)\big[|\nabla_x f(z)|^2 + |\nabla_v f(z)|^2\big],\end{aligned}$$

by stationarity of the flow. By an approximation argument we can further show that $T^t : H^1(\pi) \to H^1(\pi)$ for all $t \geq 0$. Finally $\{T^t : t \geq 0\}$ is strongly continuous on $H^1(\pi)$, since

$$\begin{aligned}\|\nabla T^s f - \nabla f\|^2 &= \int \big|\nabla f\big(\Xi(s,z)\big)\nabla_z \Xi(s,z) - \nabla f(z)\big|^2 \pi(\mathrm{d}z) \\ &\leq \int \big|\nabla f\big(\Xi(s,z)\big)\big[\nabla_z \Xi(s,z) - I\big]\big|^2 \pi(\mathrm{d}z) \\ &\quad + \int \big|\nabla f\big(\Xi(s,z)\big) - \nabla f(z)\big|^2 \pi(\mathrm{d}z) \\ &\leq \int \big|\nabla f\big(\Xi(s,z)\big)\big|^2 \big|\nabla_z \Xi(s,z) - I\big|^2 \pi(\mathrm{d}z) \\ &\quad + \int \big|\nabla f\big(\Xi(s,z)\big) - \nabla f(z)\big|^2 \pi(\mathrm{d}z) \\ &\leq \int \big|\nabla f\big(\Xi(s,z)\big)\big|^2 \big|\nabla_z \Xi(s,z) - I\big|^2 \pi(\mathrm{d}z) \\ &\quad + 2\int \big|T^s \nabla f(z) - \nabla f(z)\big|^2 \pi(\mathrm{d}z).\end{aligned} \tag{5.2}$$

Since $g := \nabla f \in L^2(\pi)$, for every $\epsilon > 0$ there is a smooth, compactly supported function g_ϵ such that $\|g - g_\epsilon\|_{L^2(\pi)} < \epsilon$. Then

$$\begin{aligned}\int |T^s g(z) - g(z)|^2 \pi(\mathrm{d}z) &= \int |T^s g(z) - T^s g_\epsilon(z) + T^s g_\epsilon(z) - g_\epsilon(z) + g_\epsilon(z) - g(z)|^2 \pi(\mathrm{d}z) \\ &\le \int \pi(\mathrm{d}z) |g(\Xi(s,z)) - g_\epsilon(\Xi(s,z))|^2 + \int \pi(\mathrm{d}z) |g(z) - g_\epsilon(z)|^2 \\ &\quad + \int \pi(\mathrm{d}z) |g_\epsilon(\Xi(s,z)) - g_\epsilon(z)|^2 \\ &= 2\|g - g_\epsilon\| + \int \pi(\mathrm{d}z) |g_\epsilon(\Xi(s,z)) - g_\epsilon(z)|^2 \\ &\le 2\epsilon + \int \pi(\mathrm{d}z) |g_\epsilon(\Xi(s,z)) - g_\epsilon(z)|^2.\end{aligned}$$

For every fixed $\epsilon > 0$, the second term vanishes by bounded convergence. Since $\epsilon > 0$ is arbitrary this shows that $\|T^s \nabla f - \nabla f\|^2 \to 0$ as $s \to 0$.

Going back to (5.2), notice that the first term also vanishes by the dominated convergence theorem, since $|\nabla_z \Xi(s,z) - I| \le 2\mathrm{e}^{Cs}$ uniformly in z, $|\nabla_z \Xi(s,z) - I| \to 0$ pointwise. Thus T^t is strongly continuous and therefore it admits a densely defined generator, which we denote by B,

$$B : \mathcal{D}(B) \subseteq H^1(\pi) \to H^1(\pi).$$

Again it is straightforward to check that B has the expression given earlier.

In addition notice that S is a bounded operator on $H^1(\pi)$. To see why first notice that an easy calculation, which will be provided later on in Section 5.2 for completeness, shows that $\nabla_x Q_\alpha = Q_\alpha \nabla_x$ and $\nabla_v Q_\alpha = \alpha Q_\alpha \nabla_v$ whence

$$\|\nabla_x Q_\alpha f\|^2 + \|\nabla_v Q_\alpha f\|^2 \le \|Q_\alpha \nabla_x f\|^2 + \alpha \|Q_\alpha \nabla_v f\|^2 \le C\big(\|\nabla_x f\|^2 + \|\nabla_v f\|^2\big),$$

since Q_α is a contraction on $L^2(\pi)$. Therefore, applying [50], Theorem 3.2, the operator $\mathcal{A} := B + \lambda_{\mathrm{ref}}(-S)$ has domain $\mathcal{D}(B)$ and generates a strongly continuous on $H^1(\pi)$, which we will denote again by $\{P^t : t \ge 0\}$. This implies that for every $f \in \mathcal{D}(B)$, $P^t f \in \mathcal{D}(\mathcal{A})$ for all $t \ge 0$ and $\mathcal{A} P^t f = P^t \mathcal{A} f$. This essentially shows that given $f \in \mathcal{D}(B)$ the abstract Cauchy problem

$$\begin{aligned}\frac{\partial u(t,z)}{\partial t} &= \mathcal{A} u, \\ u(0,z) &= f,\end{aligned}$$

admits a unique solution in $H^1(\pi)$ given by $u(t,z) := P^t f(z)$.

5.2. *Proof of Theorem* 5. We introduce some additional notation to keep the presentation concise. First recall the decomposition $\mathcal{A} = B + \lambda_{\mathrm{ref}}(-S)$ where

$$Bf = \langle \nabla_x f, v\rangle - \langle \nabla_v f, \nabla U\rangle, \qquad Sf = [I - Q_\alpha] f,$$

and let us define the Dirichlet form $\mathcal{E}(f,g) := \langle f, Sg\rangle$. We will also write $A := \nabla_v$, $C := \nabla_x$. From [58], page 40, or an easy calculation, we have

$$[A,B] = AB - BA = \nabla_x \quad \text{and} \quad [B,C] = \nabla^2 U \cdot \nabla_v = \nabla^2 U \cdot A.$$

Since $P^t = \exp(t\mathcal{A})$, where $\mathcal{A}$ is the generator of the RHMC process, an easy calculation shows that for all $f, g \in \mathcal{D}(\mathcal{B})$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle P^t f, P^t g\rangle \Big|_{t=0} = \langle \mathcal{A} f, g\rangle + \langle f, \mathcal{A} g\rangle,$$

This also implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle P^t f, P^t f\rangle\Big|_{t=0} = 2\langle \mathcal{A}f, f\rangle = -2\lambda_{\mathrm{ref}}\mathcal{E}(f,f),$$

since B is antisymmetric, in the sense that $\langle Bf, g\rangle = -\langle f, Bg\rangle$.

We want to compute $\mathrm{d}\langle\!\langle P^t f, P^t f\rangle\!\rangle/\mathrm{d}t|_{t=0}$. To keep notation to a minimum we will write h rather than $P^t f$. We proceed by computing the derivative of each term individually,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\|Ah\|^2 &= 2\langle Ah, A\mathcal{A}h\rangle = -2\lambda_{\mathrm{ref}}\langle Ah, ASh\rangle + 2\langle Ah, ABh\rangle,\\
\frac{\mathrm{d}}{\mathrm{d}t}\langle Ch, Ah\rangle &= \big\langle Ch, A(-\lambda_{\mathrm{ref}}S + B)h\big\rangle + \big\langle C(-\lambda_{\mathrm{ref}}S + B)h, Ah\big\rangle,\\
\frac{\mathrm{d}}{\mathrm{d}t}\|Ch\|^2 &= 2\langle Ch, C\mathcal{A}h\rangle = -2\lambda_{\mathrm{ref}}\langle Ch, CSh\rangle + 2\langle Ch, CBh\rangle.
\end{aligned}$$

Term one. We now compute the first term which is given by

$$-2\lambda_{\mathrm{ref}}\langle Ah, ASh\rangle + 2\langle Ah, ABh\rangle.$$

Notice that

$$\begin{aligned}
\frac{\partial}{\partial v_i}Q_\alpha f(\boldsymbol{x},\boldsymbol{v}) &= \frac{\partial}{\partial v_i}\mathbb{E}\big[f\big(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\big]\\
&= \mathbb{E}\big[\alpha f_{v_i}\big(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\big]\\
&= \alpha\mathbb{E}\big[f_{v_i}\big(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\big],
\end{aligned}$$

where to keep notation clear we write $\partial G(\boldsymbol{x},\boldsymbol{v})/\partial v_i$ to denote the derivative of the expression $G(x,v)$ w.r.t. v_i, whereas we write $f_{v_i}(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi})$ to denote the derivative of f w.r.t. v_i evaluated at $\alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}$.

The above calculation shows that $AQ_\alpha = \alpha Q_\alpha A$ and therefore

$$\begin{aligned}
-\lambda_{\mathrm{ref}}\langle Ah, ASh\rangle &= \lambda_{\mathrm{ref}}\big\langle Ah, A(Q_\alpha - I)h\big\rangle\\
&= \lambda_{\mathrm{ref}}\langle Ah, AQ_\alpha h\rangle - \lambda_{\mathrm{ref}}\langle Ah, Ah\rangle\\
&= \lambda_{\mathrm{ref}}\alpha\langle Ah, Q_\alpha Ah\rangle - \lambda_{\mathrm{ref}}\langle Ah, Ah\rangle\\
&= \lambda_{\mathrm{ref}}\big\langle Ah, (\alpha Q_\alpha - I)Ah\big\rangle\\
&= \lambda_{\mathrm{ref}}\big\langle Ah, \alpha(Q_\alpha - I)Ah\big\rangle - (1-\alpha)\lambda_{\mathrm{ref}}\langle Ah, Ah\rangle\\
&= -\lambda_{\mathrm{ref}}\alpha\langle Ah, SAh\rangle - (1-\alpha)\lambda_{\mathrm{ref}}\langle Ah, Ah\rangle.
\end{aligned}$$

Continuing we have

$$\begin{aligned}
\langle Ah, ABh\rangle &= \big\langle Ah, (AB - BA)h\big\rangle + \langle Ah, BAh\rangle\\
&= \big\langle Ah, [A,B]h\big\rangle + 0 = \langle Ah, Ch\rangle,
\end{aligned}$$

since by the anti-symmetry of B, it follows that $\langle g, Bg\rangle = 0$ for any g.

Term two. We next compute the second term

$$\big\langle Cf, A(-\lambda_{\mathrm{ref}}S + B)f\big\rangle + \big\langle C(-\lambda_{\mathrm{ref}}S + B)f, Af\big\rangle.$$

First we compute the derivative along B

$$\langle ABh, Ch\rangle + \langle Ah, CBh\rangle = \langle ABh, Ch\rangle + \langle Ah, BCh\rangle + \big\langle Ah, [C,B]h\big\rangle$$

and using that $B^* = -B$

$$\begin{aligned}
&= \langle ABh, Ch\rangle - \langle BAh, Ch\rangle + \langle Ah, [C, B]h\rangle \\
&= \langle [A, B]h, Ch\rangle + \langle Ah, [C, B]h\rangle \\
&= \langle Ch, Ch\rangle + \langle Ah, [C, B]h\rangle \\
&= \|Ch\|^2 - \langle Ah, \nabla^2 U Ah\rangle.
\end{aligned}$$

To compute the derivative along S first notice that $CQ_\alpha = Q_\alpha C$, where in the r.h.s. we tensorise Q_α allowing it to act on each component separately, in the sense that

$$\frac{\partial}{\partial x_i}\mathbb{E}\big[f\big(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\big] = \mathbb{E}\bigg[\frac{\partial}{\partial x_i} f\big(\boldsymbol{x}, \alpha\boldsymbol{v} + \sqrt{1-\alpha^2}\boldsymbol{\xi}\big)\bigg].$$

Therefore

$$\begin{aligned}
&-\lambda_{\mathrm{ref}}\langle ASh, Ch\rangle - \lambda_{\mathrm{ref}}\langle Ah, CSh\rangle \\
&\quad = \lambda_{\mathrm{ref}}\langle A(Q_\alpha - I)h, Ch\rangle + \lambda_{\mathrm{ref}}\langle Ah, C(Q_\alpha - I)h\rangle \\
&\quad = \lambda_{\mathrm{ref}}\langle (\alpha Q_\alpha - I)Ah, Ch\rangle + \lambda_{\mathrm{ref}}\langle Ah, (Q_\alpha - I)Ch\rangle \\
&\quad = \alpha\lambda_{\mathrm{ref}}\langle (Q_\alpha - I)Ah, Ch\rangle + (\alpha - 1)\lambda_{\mathrm{ref}}\langle Ah, Ch\rangle + \lambda_{\mathrm{ref}}\langle Ah, (Q_\alpha - I)Ch\rangle \\
&\quad = -(1+\alpha)\lambda_{\mathrm{ref}}\langle SAh, Ch\rangle - (1-\alpha)\lambda_{\mathrm{ref}}\langle Ah, Ch\rangle,
\end{aligned}$$

where we used again the fact that Q_α is positive.

Term three. Using the same arguments as before we have

$$\begin{aligned}
\langle Ch, CQ_\alpha h\rangle &= \sum_{i=1}^d \bigg\langle \frac{\partial}{\partial x_i} h, \frac{\partial}{\partial x_i} Q_\alpha h\bigg\rangle \\
&= \sum_{i=1}^d \bigg\langle \frac{\partial}{\partial x_i} h, Q_\alpha \frac{\partial}{\partial x_i} h\bigg\rangle = \langle Ch, Q_\alpha Ch\rangle,
\end{aligned}$$

where we are overloading the inner product by allowing it to take both vectors and scalars as arguments, in the case of scalars it integrates the product, in the case of vectors the vector inner product. Therefore

$$-\lambda_{\mathrm{ref}}\langle Ch, CSh\rangle = \lambda_{\mathrm{ref}}\langle Ch, C(Q_\alpha - I)h\rangle = -\lambda_{\mathrm{ref}}\langle Ch, SCh\rangle.$$

The next one is

$$\begin{aligned}
\langle Ch, CBh\rangle &= \langle Ch, CBh\rangle \\
&= \langle Ch, BCh\rangle - \langle Ch, [B, C]h\rangle = 0 - \langle Ch, \nabla^2 U \cdot Ah\rangle.
\end{aligned}$$

Combining all the terms. We now have the tools to compute the derivative of

$$\langle\langle h, h\rangle\rangle := a\|Ah\|^2 - 2b\langle Ch, Ah\rangle + c\|Ch\|^2,$$

which, after multiplying by -1, is given by

$$\begin{aligned}
&-\frac{\mathrm{d}}{\mathrm{d}t}\langle\langle h, h\rangle\rangle \\
&\quad = -a\frac{\mathrm{d}}{\mathrm{d}t}\|Ah\|^2 + 2b\frac{\mathrm{d}}{\mathrm{d}t}\langle Ah, Ch\rangle - c\frac{\mathrm{d}}{\mathrm{d}t}\|Ch\|^2 \\
&\quad = 2a\big[\lambda_{\mathrm{ref}}(1-\alpha)\|Ah\|^2 + \lambda_{\mathrm{ref}}\alpha\langle SAh, Ah\rangle - \langle Ah, Ch\rangle\big]
\end{aligned}$$

$$
\begin{aligned}
&+ 2b\big[\|Ch\|^2 - \langle \nabla^2 U\, Ah, Ah\rangle - (1+\alpha)\lambda_{\mathrm{ref}}\langle S^{1/2}Ah, S^{1/2}Ch\rangle - (1-\alpha)\lambda_{\mathrm{ref}}\langle Ah, Ch\rangle\big] \\
&+ 2c\big[\lambda_{\mathrm{ref}}\langle SCh, Ch\rangle + \langle \nabla^2 U\, Ah, Ch\rangle\big] \\
&= 2a\lambda_{\mathrm{ref}}(1-\alpha)\|Ah\|^2 - 2\big(a + (1-\alpha)b\lambda_{\mathrm{ref}}\big)\langle Ah, Ch\rangle + 2b\|Ch\|^2 \\
&- 2b\langle \nabla^2 U\, Ah, Ah\rangle + 2c\langle \nabla^2 U\, Ah, Ch\rangle \\
&+ 2a\lambda_{\mathrm{ref}}\alpha\langle SAh, Ah\rangle + 2c\lambda_{\mathrm{ref}}\langle SCh, Ch\rangle - 2(1+\alpha)b\lambda_{\mathrm{ref}}\langle SAh, Ch\rangle.
\end{aligned}
$$

REMARK 14. At this stage we can rewrite the above inequality as

$$
\begin{aligned}
-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle\!\langle h, h\rangle\!\rangle \geq{}& \big[a(1-\alpha)\lambda_{\mathrm{ref}} - bM\big]\|Ah\|^2 + b\|Ch\|^2 - \|JAh\|\|Ch\| \\
&+ a\alpha\lambda_{\mathrm{ref}}\big\|S^{1/2}Ah\big\|^2 + c\lambda_{\mathrm{ref}}\big\|S^{1/2}Ch\big\|^2 \\
&- (1+\alpha)b\lambda_{\mathrm{ref}}\big\|S^{1/2}Ah\big\|\,\big\|S^{1/2}Ch\big\|,
\end{aligned}
\tag{5.3}
$$

where $S^{1/2}$ is the positive, self-adjoint square root of S, and

$$
Jf := \big(aI + (1-\alpha)b\lambda_{\mathrm{ref}}I - c\nabla^2 U\big)f,
$$

which is also self-adjoint, since $\nabla^2 U$ is symmetric, whence its norm is given by

$$
\begin{aligned}
\sup_{\|f\|=1}\big|\langle Jf, f\rangle\big| &= \sup_{\|f\|=1}\big|\big[a + b\lambda_{\mathrm{ref}}(1-\alpha)\big]\langle f, f\rangle - c\langle \nabla^2 Uf, f\rangle\big| \\
&= \sup_{\|f\|=1}\max\big\{\big[a + b\lambda_{\mathrm{ref}}(1-\alpha)\big]\langle f, f\rangle - c\langle \nabla^2 Uf, f\rangle, \\
&\quad c\langle \nabla^2 Uf, f\rangle - \big[a + b\lambda_{\mathrm{ref}}(1-\alpha)\big]\langle f, f\rangle\big\} \\
&\leq \sup_{\|f\|=1}\max\big\{\big(a + (1-\alpha)\lambda_{\mathrm{ref}}b\big)\|f\| - cm\|f\|, \\
&\quad cM\|f\| - \big(a + (1-\alpha)\lambda_{\mathrm{ref}}b\big)\|f\|\big\} \\
&= \max\big\{a + (1-\alpha)\lambda_{\mathrm{ref}}b - cm, cM - a - (1-\alpha)\lambda_{\mathrm{ref}}b\big\}.
\end{aligned}
$$

Therefore, if we can find $a, b, c > 0$, such that $b < \sqrt{4a\alpha c}/(1+\alpha)$ and

$$
4\big[a(1-\alpha)\lambda_{\mathrm{ref}} - bM\big]b > \max\big\{cM - a - (1-\alpha)\lambda_{\mathrm{ref}}b, a + (1-\alpha)b\lambda_{\mathrm{ref}} - cm\big\}^2,
$$

then the RHS of (5.3) is a positive definite quadratic form. In principle this can be used to optimise the convergence rates among norms of the form (2.17).

We take a slightly different approach. Our goal is to show that for every h, we have $\frac{\mathrm{d}}{\mathrm{d}t}\langle\!\langle h, h\rangle\!\rangle \leq -\mu\langle\!\langle h, h\rangle\!\rangle$, or equivalently

$$
-\frac{\mathrm{d}}{\mathrm{d}t}\langle\!\langle h, h\rangle\!\rangle - \mu\langle\!\langle h, h\rangle\!\rangle \geq 0.
$$

After rearrangement, we obtain that

$$
\begin{aligned}
&-\frac{\mathrm{d}}{\mathrm{d}t}\langle\!\langle h, h\rangle\!\rangle - \mu\langle\!\langle h, h\rangle\!\rangle \\
&\quad= a\big(2\lambda_{\mathrm{ref}}(1-\alpha) - \mu\big)\|Ah\|^2 - 2\big(a + (1-\alpha)b\lambda_{\mathrm{ref}} - \mu b\big)\langle Ah, Ch\rangle \\
&\qquad+ (2b - c\mu)\|Ch\|^2 - 2b\langle \nabla^2 U\, Ah, Ah\rangle + 2c\langle \nabla^2 U\, Ah, Ch\rangle \\
&\qquad+ 2a\lambda_{\mathrm{ref}}\alpha\langle SAh, Ah\rangle + 2c\lambda_{\mathrm{ref}}\langle SCh, Ch\rangle - 2(1+\alpha)b\lambda_{\mathrm{ref}}\langle SAh, Ch\rangle.
\end{aligned}
\tag{5.4}
$$

We will use the following two lemmas.

LEMMA 2. *If $V, W, Z, A \in \mathbb{R}^{2\times 2}$ are symmetric matrices such that $0 \preceq A$, $-Z \preceq A$, $A \preceq V + mW$ and $A \preceq V + MW$, then $\operatorname{Tr}(VX + WP + ZQ) \geq 0$ for all symmetric matrices X, P, Q such that $0 \preceq Q \preceq X$ and $mX \preceq P \preceq MX$.*

PROOF OF LEMMA 2. First, suppose that $M = m$. By the assumptions we have $P = mX$, $A \succeq 0$ and $Z + A \succeq 0$. Note that if S, T are symmetric positive semidefinite matrices, then $\operatorname{Tr}(ST) \geq 0$. Using this fact, it follows that

$$\begin{aligned}\operatorname{Tr}(VX + WP + ZQ) &= \operatorname{Tr}\big((V + mW)X + ZQ\big) \geq \operatorname{Tr}\big(AX + (Z + A)Q - AQ\big)\\ &\geq \operatorname{Tr}\big(A(X - Q)\big) \geq 0.\end{aligned}$$

Now suppose that $M > m$. Let

$$A_1 = Z + A, \qquad A_2 = A,$$
$$A_3 = \frac{1}{M-m}(V + MW - A), \qquad A_4 = \frac{1}{M-m}(V + mW - A).$$

Then $A_1, A_2, A_3, A_4 \succeq 0$, and

$$V = A_2 - mA_3 + MA_4, \qquad W = A_3 - A_4, \qquad Z = A_1 - A_2.$$

So

$$\begin{aligned}VX + WP + ZQ &= (A_2 - mA_3 + MA_4)X + (A_3 - A_4)P + (A_1 - A_2)Q\\ &= A_1Q + A_2(X - Q) + A_3(P - mX) + A_4(MX - P).\end{aligned}$$

Using positive definiteness of both terms in the matrix products, we have

$$\operatorname{Tr}(A_1Q), \operatorname{Tr}\big(A_2(X - Q)\big), \operatorname{Tr}\big(A_3(P - mX)\big), \operatorname{Tr}\big(A_4(MX - P)\big) \geq 0,$$

and therefore

$$\operatorname{Tr}(VX + WP + ZQ) \geq 0. \qquad \square$$

Now we are ready to complete the proof of Theorem 5.

PROOF OF THEOREM 5. Let $a := 1$, and

$$b := \frac{1 + \alpha - \alpha(\frac{m}{M+m})^{3/4} + \frac{3}{4}\frac{\alpha m}{M+m}}{2\sqrt{M+m}},$$
$$c := \frac{1 + \alpha - \frac{\alpha}{2}(\frac{m}{M+m})^{1/2}}{M+m},$$
$$X := \begin{pmatrix} \|Ah\|^2 & \langle Ah, Ch\rangle \\ \langle Ah, Ch\rangle & \|Ch\|^2 \end{pmatrix},$$
$$P := \begin{pmatrix} \langle \nabla^2 U(x)Ah, Ah\rangle & \langle \nabla^2 U(x)Ah, Ch\rangle \\ \langle \nabla^2 U(x)Ah, Ch\rangle & \langle \nabla^2 U(x)Ch, Ch\rangle \end{pmatrix},$$
$$Q := \begin{pmatrix} \langle SAh, Ah\rangle & \langle SAh, Ch\rangle \\ \langle SAh, Ch\rangle & \langle SCh, Ch\rangle \end{pmatrix},$$
$$V := \begin{pmatrix} 2a(1-\alpha)\lambda_{\text{ref}} - a\mu & -a - (1-\alpha)b\lambda_{\text{ref}} + b\mu \\ -a - (1-\alpha)b\lambda_{\text{ref}} + b\mu & 2b - c\mu \end{pmatrix},$$

$$W := \begin{pmatrix} -2b & c \\ c & 0 \end{pmatrix},$$

$$Z := \begin{pmatrix} 2a\alpha\lambda_{\mathrm{ref}} & -(1+\alpha)b\lambda_{\mathrm{ref}} \\ -(1+\alpha)b\lambda_{\mathrm{ref}} & 2c\lambda_{\mathrm{ref}} \end{pmatrix},$$

$$A := \begin{pmatrix} \dfrac{4(-3+2m-2M)(-1+\alpha)}{3\sqrt{m+M}(1+\alpha)} & -\dfrac{(-3+2m-2M)(-1+\alpha)}{3(m+M)} \\ -\dfrac{(-3+2m-2M)(-1+\alpha)}{3(m+M)} & -\dfrac{(-3+2m-2M)(-1+\alpha)(1+\alpha)}{3(m+M)^{3/2}} \end{pmatrix}.$$

Using the fact that $x^{3/4} - 3/4x \geq 0$ for $x \in [0, 1/2]$, it is easy to check that $b^2 < ac$. Using the assumption that $mI \preceq \nabla^2 U \preceq MI$, we have $mX \preceq P \preceq MX$. Moreover, using the fact that $0 \preceq S \preceq I$, we have $0 \preceq Q \preceq X$. Based on (5.4) and the above definitions it follows that

$$-\frac{\mathrm{d}}{\mathrm{d}t}\langle\langle h, h\rangle\rangle - \mu\langle\langle h, h\rangle\rangle = \mathrm{Tr}(VX + WP + ZQ). \tag{5.5}$$

One can check, for example, using Mathematica, that for every $M \geq 1$, $0 \leq \alpha < 1$, the inequalities $0 \preceq A$, $-Z \preceq A$, $A \preceq V + mW$ and $A \preceq V + MW$ hold for A defined as above. Therefore (5.7) follows from Lemma 2, and by Grönwall's lemma, this implies that $\langle\langle P^t f, P^t f\rangle\rangle \leq \exp(-\mu t)\langle\langle f, f\rangle\rangle$.

5.2.1. *From H^1 to L^2.* To show our L^2 bound, we study the reversed process. Denote the variant of the scalar product $\langle\langle\cdot,\cdot\rangle\rangle$ when b is replaced by $-b$ by $\langle\langle\cdot,\cdot\rangle\rangle'$, that is,

$$\langle\langle h, h\rangle\rangle' := a\|\nabla_v h\|^2 + 2b\langle\nabla_x h, \nabla_v h\rangle + c\|\nabla_x h\|^2. \tag{5.6}$$

Then by repeating the same arguments as above with v replaced by $-v$ everywhere, one can show that we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\big\langle\big\langle (P^*)^t f, (P^*)^t f\big\rangle\big\rangle' \leq -\mu\big\langle\big\langle (P^*)^t f, (P^*)^t f\big\rangle\big\rangle', \tag{5.7}$$

and hence $\langle\langle (P^*)^t f, (P^*)^t f\rangle\rangle' \leq \exp(-\mu t)\langle\langle f, f\rangle\rangle'$. Similar to the previous proofs, we can show that $\langle\langle\cdot,\cdot\rangle\rangle$ and $\langle\langle\cdot,\cdot\rangle\rangle'$ are equivalent up to the same constant factor C, and

$$\big\langle\big\langle (P^t)^* P^t f, (P^t)^* P^t f\big\rangle\big\rangle \leq C^2 \exp(-2\mu t)\langle\langle f, f\rangle\rangle.$$

In addition, there exist constants $C_1, C_2 > 0$ such that $\langle\langle f, f\rangle\rangle \leq C_1\|\nabla f\|^2$ and $\|f\|^2 \leq C_2\langle\langle f, f\rangle\rangle$. Thus, letting f be k-Lipschitz we have

$$\begin{aligned} \big\|(P^t)^* P^t f\big\|^2 &\leq C_2\big\langle\big\langle (P^t)^* P^t f, (P^t)^* P^t f\big\rangle\big\rangle' \\ &\leq C_2 \exp(-2\mu t)\langle\langle f, f\rangle\rangle' \\ &\leq C_1 C_2 \exp(-2\mu t)\|\nabla f\|^2 \leq C_1 C_2 k^2 \exp(-2\mu t). \end{aligned}$$

Choose t such that $C_1 C_2 k^2 \mathrm{e}^{-2\mu t} =: 1 - \kappa < 1$ and define the self-adjoint operator $Q = (P^t)^* P^t$. Iterating the above we have for $n \geq 1$ that

$$\big\|Q^n f\big\|^2 \leq C_1 C_2 (1-\kappa)^{2n} k^2 =: C(f)(1-\kappa)^{2n}.$$

The rest is similar to the proof of Proposition 2.8 from Hairer et al. [31]. Let f be k-Lipschitz, and without loss of generality also assume that $\|f\| = 1$. Let ν_f be the spectral measure

corresponding to the self-adjoint operator Q applied to the function f. In particular, since $\|f\| = 1$, ν_f is a probability measure. Then

$$\begin{aligned}
\|Q^n f\|^2 &= \int_{-1}^{1} t^{2n} \nu_f(\mathrm{d}t) \\
&= \int_{-1}^{1} t^{2n(2n+2m)/(2n+2m)} \nu_f(\mathrm{d}t) \\
&\le \left(\int_{-1}^{1} t^{2(n+m)} \nu_f(\mathrm{d}t) \right)^{\frac{2n}{2(n+m)}} \\
&= \left(\|Q^{n+m} f\|^2 \right)^{\frac{2n}{2(n+m)}} \\
&\le \left[C(f)(1-\kappa)^{2(n+m)} \right]^{\frac{2n}{2(n+m)}} \\
&\le C(f)^{\frac{2n}{2(n+m)}} (1-\kappa)^{2n},
\end{aligned}$$

and letting $m \to \infty$ we get for any k-Lipschitz f

$$\|Q^n f\|^2 \le \|f\|^2 (1-\kappa)^{2n},$$

noticing that the upper bound is independent of the Lipschitz constant. Since Lipschitz functions are dense we conclude. □

REMARK 15. Given any $\lambda_{\mathrm{ref}} > 0$, $\mu > 0$, the contraction $\frac{\mathrm{d}}{\mathrm{d}t}\langle\langle h, h\rangle\rangle \le -\mu\langle\langle h, h\rangle\rangle$ holds as long as there exists coefficients $a, b, c \in \mathbb{R}$ and a 2×2 real valued symmetric matrix A such that $a > 0$, $c > 0$, $b^2 < ac$ and $0 \preceq A$, $-Z \preceq A$, $A \preceq V + mW$ and $A \preceq V + MW$ (with V and W defined as above). Note that as in the proof of Theorem 3, due to the nonlinearity of the constraints we did not manage to find an analytical expression for the largest possible μ for a given λ_{ref}, and the largest possible μ for any λ_{ref}. However, we believe that the choice of λ_{ref} and μ as given here is close to optimal in most of the parameter range $0 < m \le M < \infty$, $0 \le \alpha < 1$.

APPENDIX: AUXILIARY RESULTS

Notice that using the independence of X and Z, and the fact that the standard normal distribution is isotropic, we have

$$\mathbb{E}_{X\sim\pi, Z\sim N(0,\mathbb{I}_d)}\left[(\nabla U(X), Z)_+\right] = \mathbb{E}\left(|\nabla U(X)|\right)\mathbb{E}\left[(w, Z)_+\right],$$

where w is an arbitrary fixed d-dimensional unit vector. Now noticing that (w, Z) is a one-dimensional standard normal random variable, it follows that $\mathbb{E}[(w, Z)_+] = \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} x \times \exp(-\frac{x^2}{2})\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}$. Hence the key part of the proof is to find lower and upper bounds on

$$\mathbb{E}\left(|\nabla U(X)|\right) = \frac{\int_{x\in\mathbb{R}^d} |\nabla U(x)| e^{-U(x)}\,\mathrm{d}x}{\int_{x\in\mathbb{R}^d} e^{-U(x)}\,\mathrm{d}x}.$$

By shifting U, we can assume without loss of generality that $U(0) = 0$ and $\nabla U(0) = 0$ (hence the minimum is taken in the origin 0). Let $\mathbb{S}_1^d$ denote the d-dimensional unit sphere, then by writing the above integrals along half-lines, we have

$$\begin{aligned}
\mathbb{E}\left(|\nabla U(X)|\right) &= \frac{\int_{u\in\mathbb{S}_1^d} \int_{r=0}^{\infty} |\nabla U(ru)| e^{-U(ru)} r^{d-1}\,\mathrm{d}r\,\mathrm{d}u}{\int_{u\in\mathbb{S}_1^d} \int_{r=0}^{\infty} e^{-U(ru)} r^{d-1}\,\mathrm{d}r\,\mathrm{d}u} \\
&\ge \frac{\int_{u\in\mathbb{S}_1^d} \int_{r=0}^{\infty} |\frac{\partial}{\partial r} U(ru)| e^{-U(ru)} r^{d-1}\,\mathrm{d}r\,\mathrm{d}u}{\int_{u\in\mathbb{S}_1^d} \int_{r=0}^{\infty} e^{-U(ru)} r^{d-1}\,\mathrm{d}r\,\mathrm{d}u}.
\end{aligned} \tag{A.1}$$

If we could lower bound the ratios of the one-dimensional integrals

$$\frac{\int_{r=0}^{\infty} |\frac{\partial}{\partial r} U(ru)| e^{-U(ru)} r^{d-1}\,\mathrm{d}r}{\int_{r=0}^{\infty} e^{-U(ru)} r^{d-1}\,\mathrm{d}r},$$

then a lower bound for $\mathbb{E}(|\nabla U(X)|)$ follows by rearrangement. This is shown in the following lemma.

LEMMA A.1. *Let $d \in \mathbb{Z}_{\geq 1}$, $m \in \mathbb{R}_{>0}$, and let $V : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a differentiable function such that $x \mapsto V(x) - m\frac{x^2}{2}$ is convex, and $V'(0) = 0$. Let $A = \int_0^\infty x^{d-1} e^{-V(x)}\,\mathrm{d}x$ and $B = \int_0^\infty V'(x) x^{d-1} e^{-V(x)}\,\mathrm{d}x$. Then $B \geq \sqrt{2m}\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} A$.*

PROOF. First let $d = 1$. Then $B = \int_0^\infty (-e^{-V(x)})'\,\mathrm{d}x = e^{-V(0)}$. We have $V(x) \geq V(0) + m\frac{x^2}{2}$ for $x \geq 0$, so $A \leq \int_0^\infty e^{-V(0) - m\frac{x^2}{2}}\,\mathrm{d}x = e^{-V(0)}\sqrt{\frac{\pi}{2m}} = \sqrt{\frac{\pi}{2m}} B$, so $B \geq \sqrt{2m}\frac{\Gamma(1)}{\Gamma(\frac{1}{2})} A$.

Now let $d \geq 2$. Then

$$B = \int_0^\infty \big((-x^{d-1} e^{-V(x)})' + (d-1) x^{d-2} e^{-V(x)}\big)\,\mathrm{d}x = (d-1)\int_0^\infty x^{d-2} e^{-V(x)}\,\mathrm{d}x,$$

so the claim is equivalent to $\int_0^\infty (c - x) x^{d-2} e^{-V(x)}\,\mathrm{d}x \geq 0$, where $c = \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \cdot \frac{\sqrt{2}}{\sqrt{m}}$ (here we have used $\Gamma(\frac{d+1}{2}) = \frac{d-1}{2}\Gamma(\frac{d-1}{2})$). The function $x \mapsto V(x) - m\frac{x^2}{2}$ is convex, and its derivative at $x = 0$ is 0, so this function is monotone increasing on $\mathbb{R}_{\geq 0}$. Hence $V(x) \geq V(c) + \frac{m}{2}(x^2 - c^2)$ if $x \geq c$, and $V(x) \leq V(c) + \frac{m}{2}(x^2 - c^2)$ if $x \leq c$. Thus

$$\begin{aligned}\int_0^\infty (c - x) x^{d-2} e^{-V(x)}\,\mathrm{d}x &\geq \int_0^\infty (c - x) x^{d-2} e^{-V(c) - \frac{m}{2}(x^2 - c^2)}\,\mathrm{d}x \\ &= e^{\frac{m}{2}c^2 - V(c)} \int_0^\infty (c - x) x^{d-2} e^{-\frac{m}{2}x^2}\,\mathrm{d}x.\end{aligned}$$

We have $\int_0^\infty x^\alpha e^{-\frac{m}{2}x^2}\,\mathrm{d}x = \frac{1}{2}(\frac{m}{2})^{-\frac{\alpha+1}{2}}\Gamma(\frac{\alpha+1}{2})$ for every $m > 0$ and $\alpha > -1$. So

$$\begin{aligned}&\int_0^\infty (c - x) x^{d-2} e^{-V(x)}\,\mathrm{d}x \\ &\quad \geq e^{\frac{m}{2}c^2 - V(c)} \frac{1}{2}\left(c\left(\frac{m}{2}\right)^{-\frac{d-1}{2}}\Gamma\left(\frac{d-1}{2}\right) - \left(\frac{m}{2}\right)^{-\frac{d}{2}}\Gamma\left(\frac{d}{2}\right)\right) = 0.\end{aligned}$$

□

The following lemma will be used to find a simpler lower bound for the ratio $\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}$.

LEMMA A.2. *If $s > 0$, then $\frac{\Gamma(s+\frac{3}{4})}{\Gamma(s+\frac{1}{4})} > \sqrt{s}$.*

PROOF. Let $\phi(s) := \frac{1}{\sqrt{s}}\frac{\Gamma(s+\frac{3}{4})}{\Gamma(s+\frac{1}{4})}$ for $s > 0$. Stirling's formula implies that $\lim_{s\to\infty}\phi(s) = 1$. For $s > 0$ we have $\phi(s) > 0$ and $(\frac{\phi(s+1)}{\phi(s)})^2 = 1 - \frac{1}{(1+s)(1+4s)^2} < 1$, so $\phi(s) > \phi(s+1)$. Thus $\phi(s) > \phi(s+1) \geq \phi(s+n)$ for every $n \in \mathbb{Z}_{\geq 1}$, and taking $n \to \infty$ we get $\phi(s) > 1$. □

Taking $s = \frac{d-\frac{1}{2}}{2}$ for $d \in \mathbb{Z}_{\geq 1}$, we get

$$\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} > \sqrt{\frac{d-\frac{1}{2}}{2}}. \tag{A.2}$$

The next lemma will show the upper bound.

LEMMA A.3. *Suppose that the potential $U : \mathbb{R}^n \to \mathbb{R}$ satisfies Assumption 1. Then for every $1 \le i \le n$, we have $\mathbb{E}((\partial_i U(X))^2) \le M$, implying that $\mathbb{E}(|\nabla U_n(\boldsymbol{X})|^2) \le nM$ and $\mathbb{E}(|\nabla U_n(\boldsymbol{X})|) \le \sqrt{nM}$.*

PROOF. By Jensen's inequality, we have

$$\mathbb{E}\big(|\nabla U(X)|\big) \le \big[\mathbb{E}\big(|\nabla U(X)|^2\big)\big]^{1/2} = \bigg[\mathbb{E}\Big(\sum_{i=1}^{d}\big(\partial_i U(X)\big)^2\Big]^{1/2}.$$

Here

$$\mathbb{E}\big(\big(\partial_i U(X)\big)^2\big) = \frac{\int_{x\in\mathbb{R}^d}(\partial_i U(x))^2 \exp(-U(x))\,\mathrm{d}x}{\int_{x\in\mathbb{R}^d}\exp(-U(x))\,\mathrm{d}x},$$

and from integration by parts, it follows that for every $1 \le i \le d$, we have

$$\begin{aligned}
&\int_{x\in\mathbb{R}^d}\big(\partial_i U(x)\big)^2\exp\big(-U(x)\big)\,\mathrm{d}x\\
&\quad= \int_{x_{-i}\in\mathbb{R}^{d-1}}\int_{x_i\in\mathbb{R}}\big(\partial_i U(x)\big)^2\exp\big(-U(x)\big)\,\mathrm{d}x_i\,\mathrm{d}x_{-i}\\
&\quad= \int_{x_{-i}\in\mathbb{R}^{d-1}}\bigg\{\big[-\partial_i U(x)\exp\big(-U(x)\big)\big]_{x_i=-\infty}^{\infty} + \int_{x_i\in\mathbb{R}}\partial_i^2 U(x)\exp\big(-U(x)\big)\,\mathrm{d}x_i\bigg\}\,\mathrm{d}x_{-i}\\
&\quad= \int_{x_{-i}\in\mathbb{R}^{d-1}}\int_{x_i\in\mathbb{R}}\partial_i^2 U(x)\exp\big(-U(x)\big)\,\mathrm{d}x_i\,\mathrm{d}x_{-i} \le M\int_{x\in\mathbb{R}^d}\exp\big(-U(x)\big)\,\mathrm{d}x.
\end{aligned}$$

The second and third claims now follow by summing up in i, and using Jensen's inequality. □

PROOF OF PROPOSITION 2. The result follows from Lemmas A.1, A.2 and A.3. □

LEMMA A.4. *Suppose that $U_n(\boldsymbol{X}) : \mathbb{R}^n \to \mathbb{R}$ with $m\boldsymbol{I}_d \preceq \nabla^2 U_n(\boldsymbol{X}) \preceq M\boldsymbol{I}_d$. Then*

$$\mathbb{E}\bigg[\frac{\partial_1 U_n(\boldsymbol{X})V_1}{(\sum_{j=1}^n[\partial_j U_n(\boldsymbol{X})]^2)^{1/2}}\Big|X_1, V_1\bigg] \to 0 \quad \text{as } n\to\infty.$$

PROOF.

$$\mathbb{E}\bigg[\frac{\partial_1 U_n(\boldsymbol{X})V_1}{(\sum_{j=1}^n[\partial_j U_n(\boldsymbol{X})]^2)^{1/2}}\Big|X_1, V_1\bigg] \le \mathbb{E}\bigg[\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|}\Big|X_1\bigg]\cdot|V_1|.$$

Let us denote $X_{-1} := (X_2,\dots,X_n)$, then X_{-1} given X_1 has a conditional distribution with density that is proportional to $\exp(-U_n(X_{-1}, X_1))$, which is a log-concave function of X_{-1}, with Hessian bounded between m and M. By Theorem 5.2 of [35], $\mathcal{L}(X_{-1}|X_1)$ satisfies a log-Sobolev inequality with constant $C := m^{-1}$. The functions $|\nabla U_n(\boldsymbol{X})|$ and $|\partial_1 U_n(\boldsymbol{X})|$ are M-Lipschitz in X_{-1} given a fixed X_1, and hence by Herbst's argument (see equation (5.8) on page page 95 of [35]),

$$\mathbb{P}\big(|\partial_1 U_n(\boldsymbol{X})| - \mathbb{E}\big(|\partial_1 U_n(\boldsymbol{X})||X_1\big) \ge t|X_1\big) \le \exp\bigg(-t^2\cdot\frac{2m}{M^2}\bigg),$$

$$\mathbb{P}\big(|\nabla U_n(\boldsymbol{X})| - \mathbb{E}\big(|\nabla U_n(\boldsymbol{X})||X_1\big) \le -t|X_1\big) \le \exp\bigg(-t^2\cdot\frac{2m}{M^2}\bigg).$$

Conditionally on X_1, define the event G_t as

$$G_t := \big\{|\partial_1 U_n(\boldsymbol{X})| - \mathbb{E}\big(|\partial_1 U_n(\boldsymbol{X})||X_1\big) < t \text{ and } |\nabla U_n(\boldsymbol{X})| - \mathbb{E}\big(|\nabla U_n(\boldsymbol{X})||X_1\big) > -t\big\},$$

then by the above bounds, we have $\mathbb{P}(G_t|X_1) \geq 1 - 2\exp(-t^2 \cdot \frac{2m}{M^2})$ for every $t \geq 0$. Let G_t^c denote the complement of G_t. Assuming that $0 < t < \mathbb{E}(|\nabla U_n(\boldsymbol{X})||X_1)$, the quantity of interest can be bounded as

$$\begin{aligned}\mathbb{E}\left[\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|}\Big|X_1\right] &= \mathbb{E}\left[\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|} \cdot 1_{G_t}\Big|X_1\right] + \mathbb{E}\left[\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|} \cdot 1_{G_t^c}\Big|X_1\right] \\ &\leq \frac{\mathbb{E}(|\partial_1 U_n(\boldsymbol{X})||X_1) + t}{\mathbb{E}(|\nabla U_n(\boldsymbol{X})||X_1) - t} + 2\exp\left(-t^2 \cdot \frac{2m}{M^2}\right),\end{aligned} \tag{A.3}$$

where we have used the fact that $\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|} \leq 1$. By Lemma 9 and equation (A.2), it follows that for any $n \geq 2$,

$$\mathbb{E}\big(|\nabla U_n(\boldsymbol{X})||X_1\big) \geq \mathbb{E}\big(|\partial_{-1} U_n(\boldsymbol{X})||X_1\big) \geq \sqrt{m(n-3/2)},$$

where $\partial_{-1} U_n(\boldsymbol{X})$ denotes the gradient vector without the first component. By Lemma 11,

$$\mathbb{E}\big(|\partial_1 U_n(\boldsymbol{X})|\big) \leq \sqrt{M}.$$

Note that $|\partial_{-1} U_n(\boldsymbol{X}_{-1}, X_1) - \partial_{-1} U_n(\boldsymbol{X}_{-1}, X_1')| \leq M|X_1 - X_1'|$, and by Proposition 19 of [60], it follows that

$$W_1\big(\mathcal{L}(X_{-1}|X_1), \mathcal{L}\big(X_{-1}|X_1'\big)\big) \leq \frac{M}{m}|X_1 - X_1'|,$$

therefore $g(X_1) := \mathbb{E}(|\partial_1 U_n(\boldsymbol{X})||X_1)$ is $\frac{M^2}{m}$-Lipschitz in X_1. By log-Sobolev inequality and Herbst's argument, for any $s \geq 0$, we have

$$\mathbb{P}\big(\big|X_1 - \mathbb{E}(X_1)\big| \geq s\big) \leq 2\exp\big(-s^2 \cdot 2m\big).$$

Therefore, it follows that

$$\begin{aligned}\sqrt{M} \geq \mathbb{E}[g] &= \mathbb{E}\big[g(X_1) - g\big(\mathbb{E}(X_1)\big)\big] + g\big(\mathbb{E}(X_1)\big) \\ &\geq -\int_{r=0}^{\infty} \mathbb{P}\big[g(X_1) - g\big(\mathbb{E}(X_1)\big) \leq -r\big]\mathrm{d}r + g\big(\mathbb{E}(X_1)\big) \\ &\geq -\int_{r=0}^{\infty} \mathbb{P}\left[\big|X_1 - \mathbb{E}(X_1)\big| \geq r\frac{m}{M^2}\right]\mathrm{d}r + g\big(\mathbb{E}(X_1)\big) \geq -\int_{r=0}^{\infty} 2\exp\left(-r^2\frac{2m^3}{M^4}\right)\mathrm{d}r \\ &\geq g\big(\mathbb{E}(X_1)\big) - 2\sqrt{(\pi/2)M^4/m^3} \geq g\big(\mathbb{E}(X_1)\big) - 3\frac{M^2}{m^{3/2}}.\end{aligned}$$

Thus $g(\mathbb{E}(X_1)) \leq 4\frac{M^2}{m^{3/2}}$, which implies by the Lipschitz property that implying that

$$\mathbb{E}\big(|\partial_1 U_n(\boldsymbol{X})||X_1\big) = g(X_1) \leq 4\frac{M^2}{m^{3/2}} + \frac{M^2}{m}|X_1 - \mathbb{E}(X_1)|.$$

By simple algebra, $t = \sqrt{\log(n)M^2/(2m)}$ satisfies that for $n \geq 3/2 + 2\log(n)\frac{M^2}{m^2}$, we have $t \leq \frac{1}{2}\sqrt{m(n-3/2)}$. By combining the above bound with (A.3) and using this t, we have

$$\mathbb{E}\left[\frac{|\partial_1 U_n(\boldsymbol{X})|}{|\nabla U_n(\boldsymbol{X})|}\Big|X_1\right] \leq \frac{4\frac{M^2}{m^{3/2}} + \frac{M^2}{m}|X_1 - \mathbb{E}(X_1)| + \sqrt{\log(n)M^2/(2m)}}{\frac{1}{2}\sqrt{m(n-3/2)}} + \frac{2}{n},$$

as long as $n \geq 3/2 + 2\log(n)\frac{M^2}{m^2}$. This tends to 0 as $n \to \infty$. $\square$

LEMMA A.5. *Suppose that U_n satisfies Assumption 1 and let $\boldsymbol{X} \sim \bar{\pi}_n$. Then for any $\alpha > 0$*

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{1}{|\nabla U_n(\boldsymbol{X})|^\alpha}\right] = 0.$$

PROOF OF LEMMA A.5. We have

$$\mathbb{E}\left[\frac{1}{|\nabla U_n(\boldsymbol{X})|^\alpha}\right] = \int_{t=0}^{\infty} \mathbb{P}\left[\frac{1}{|\nabla U_n(\boldsymbol{X})|^\alpha} \geq t\right] \mathrm{d}t = \mathbb{P}\big[|\nabla U_n(\boldsymbol{X})| \leq t^{-1/\alpha}\big]\,\mathrm{d}t. \tag{A.4}$$

The function $|\nabla U_n(\boldsymbol{x})|$ is M-Lipschitz in $\boldsymbol{x}$, so by the log-Sobolev inequality and Herbst's argument (see [35]), for any $s \geq 0$, we have

$$\mathbb{P}\big(|\nabla U_n(\boldsymbol{x})| \leq \mathbb{E}\big(|\nabla U_n(\boldsymbol{x})|\big) - s\big) \leq \exp\left(-s^2 \cdot \frac{2m}{M^2}\right).$$

In the proof of Proposition 2, we have shown that $\mathbb{E}(|\nabla U_n(\boldsymbol{x})|) \geq \sqrt{n - \frac{1}{2}}\sqrt{m}$, hence for any $s \geq 0$,

$$\mathbb{P}\left(|\nabla U_n(\boldsymbol{x})| \leq \sqrt{n - \frac{1}{2}}\sqrt{m} - s\right) \leq \exp\left(-s^2 \cdot \frac{2m}{M^2}\right). \tag{A.5}$$

This bound will be used to control $\mathbb{P}[|\nabla U_n(\boldsymbol{X})| \leq t^{-1/\alpha}]$ for small and intermediate values of t. However, for large t, the above concentration bound is not sufficiently sharp, as it does not tends to zero as $t \to \infty$. Hence we will use a different argument, that upper bounds the density of $\bar{\pi}_n$ and the volume of the space where $|\nabla U_n(\boldsymbol{X})| \leq r$.

First, note that by Assumption 1, we have $U_n(0) = 0$ and U_n is minimized in 0. Using the lower and upper bounds on the Hessian of U_n, it follows that $\frac{m}{2}|x|^2 \leq U_n(\boldsymbol{x}) \leq \frac{M}{2}|\boldsymbol{x}|^2$. These bounds correspond to the log-likelihoods of Gaussian densities, so the normalising constant of U_n can be bounded as

$$\frac{(2\pi)^{n/2}}{M^{n/2}} \leq \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp\big(-U_n(\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{x} \leq \frac{(2\pi)^{n/2}}{m^{n/2}}. \tag{A.6}$$

Moreover, using the bounds on the Hessian of U_n, it follows that $|\nabla U_n(\boldsymbol{X})| \leq r$ implies that $|\boldsymbol{X}| \leq \frac{r}{m}$. Since the volume of a ball of radius $\frac{r}{m}$ in $\mathbb{R}^n$ is

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\left(\frac{r}{m}\right)^n \leq 6\left(\frac{r}{m}\right)^n,$$

it follows that

$$\mathbb{P}\big(|\nabla U_n(\boldsymbol{X})| \leq r\big) \leq \mathbb{P}\left(|\boldsymbol{X}| \leq \frac{r}{m}\right) \leq 6\frac{M^{n/2}}{(2\pi)^{n/2}}\left(\frac{r}{m}\right)^n. \tag{A.7}$$

Let $a := (\sqrt{n - \frac{1}{2}}\sqrt{m}/2)^{-\alpha}$, and $b = (\frac{m\sqrt{2\pi}}{2\sqrt{M}})^{-\alpha}$. By upper bounding $\mathbb{P}[|\nabla U_n(\boldsymbol{X})| \leq t^{-1/\alpha}]$ by 1 for $0 \leq t \leq a$, by $\exp(-\frac{(n-\frac{1}{2})m^2}{2M^2})$ for $a < t \leq b$ (using (A.5)), and by $6t^{-n/\alpha}(\frac{\sqrt{M}}{m\sqrt{2\pi}})^n$ for $t > b$, by (A.4), for $n > \alpha$, we have

$$\begin{aligned}\mathbb{E}\left[\frac{1}{|\nabla U_n(\boldsymbol{X})|^\alpha}\right] &\leq \left(\sqrt{n - \frac{1}{2}}\sqrt{m}/2\right)^{-\alpha} + \left(\frac{m\sqrt{2\pi}}{2\sqrt{M}}\right)^{-\alpha} \cdot \exp\left(-\frac{(n-\frac{1}{2})m^2}{2M^2}\right) \\ &\quad + 6\left(\frac{\sqrt{M}}{m\sqrt{2\pi}}\right)^n \frac{b^{-\frac{n}{\alpha}+1}}{\frac{n}{\alpha} - 1}\end{aligned}$$

$$\leq \left(\sqrt{n-\frac{1}{2}}\sqrt{m}/2\right)^{-\alpha} + \left(\frac{m\sqrt{2\pi}}{2\sqrt{M}}\right)^{-\alpha} \cdot \exp\left(-\frac{(n-\frac{1}{2})m^2}{2M^2}\right) + 6\left(\frac{\sqrt{M}}{m}\right)^{\alpha} \frac{2^{-n}}{\frac{n}{\alpha}-1}$$

which tends to 0 as $n \to \infty$. □

Acknowledgements. The authors would like to thank Peter Holderrieth for a careful reading of the manuscript and his invaluable suggestions and Philippe Gagnon for his insightful comments on the manuscript. G.D. would like to thank Gabriel Stoltz for many useful discussions. The authors would also like to thank the anonymous referees for numerous suggestions that have greatly improved the content and the presentation of the paper. A part of this research was done while A. Doucet, G. Deligiannidis and D. Paulin were hosted by the Institute for Mathematical Sciences in Singapore.

Funding. This material is based upon work supported in part by the U.S. Army Research Laboratory and the U. S. Army Research Office, and by the U.K. Ministry of Defence (MoD) and the U.K. Engineering and Physical Research Council (EPSRC) under grant number EP/R013616/1 and by the EPSRC EP/R034710/1.

REFERENCES

[1] ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2018). Hypocoercivity of piecewise deterministic Markov process–Monte Carlo. Preprint. Available at arXiv:1808.08592.

[2] BAKRY, D., BARTHE, F., CATTIAUX, P. and GUILLIN, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** 60–66. MR2386063 https://doi.org/10.1214/ECP.v13-1352

[3] BÉDARD, M. (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.* **17** 1222–1244. MR2344305 https://doi.org/10.1214/105051607000000096

[4] BÉDARD, M. (2019). Hierarchical models and tuning of random walk Metropolis algorithms. *J. Probab. Stat.* **2019** Art. ID 8740426, 24. MR4002236 https://doi.org/10.1155/2019/8740426

[5] BESKOS, A., PILLAI, N., ROBERTS, G., SANZ-SERNA, J.-M. and STUART, A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19** 1501–1534. MR3129023 https://doi.org/10.3150/12-BEJ414

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

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

[8] BIERKENS, J., KAMATANI, K. and ROBERTS, G. O. (2018). High-dimensional scaling limits of piecewise deterministic sampling algorithms. Preprint. Available at arXiv:1807.11358.

[9] BIERKENS, J. and LUNEL, S. M. V. (2019). Spectral analysis of the zigzag process. Preprint. Available at arXiv:1905.01691.

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

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

[12] BÖTTCHER, B., SCHILLING, R. and WANG, J. (2013). *Lévy Matters. III: Lévy-Type Processes: Construction, Approximation and Sample Path Properties. Lecture Notes in Math.* **2099**. Springer, Cham. MR3156646 https://doi.org/10.1007/978-3-319-02684-8

[13] BOU-RABEE, N., EBERLE, A. and ZIMMER, R. (2020). Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **30** 1209–1250. MR4133372 https://doi.org/10.1214/19-AAP1528

[14] Bou-Rabee, N. and Sanz-Serna, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. MR3693523 https://doi.org/10.1214/16-AAP1255

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

[16] Brascamp, H. J. and Lieb, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480 https://doi.org/10.1016/0022-1236(76)90004-5

[17] Breyer, L. A. and Roberts, G. O. (2000). From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Process. Appl.* **90** 181–206. MR1794535 https://doi.org/10.1016/S0304-4149(00)00041-7

[18] Chicone, C. (1999). *Ordinary Differential Equations with Applications*. *Texts in Applied Mathematics* **34**. Springer, New York. MR1707333

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

[20] Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

[21] Davies, E. B. (1980). *One-Parameter Semigroups*. *London Mathematical Society Monographs* **15**. Academic Press,New York. MR0591851

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

[23] Deligiannidis, G., Bouchard-Côté, A. and Doucet, A. (2019). Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.* **47** 1268–1287. MR3911112 https://doi.org/10.1214/18-AOS1714

[24] Dolbeault, J., Mouhot, C. and Schmeiser, C. (2015). Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.* **367** 3807–3828. MR3324910 https://doi.org/10.1090/S0002-9947-2015-06012-7

[25] Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

[26] Durmus, A., Guillin, A. and Monmarché, P. (2020). Geometric ergodicity of the bouncy particle sampler. *Ann. Appl. Probab.* **30** 2069–2098. MR4149523 https://doi.org/10.1214/19-AAP1552

[27] Durmus, A. and Moulines, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

[28] Dwivedi, R., Chen, Y., Wainwright, M. J. and Yu, B. (2019). Log-concave sampling: Metropolis–Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** Paper No. 183, 42. MR4048994

[29] Ethier, S. N. and Kurtz, T. G. (2009). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0838085 https://doi.org/10.1002/9780470316658

[30] Fétique, N. (2017). Long-time behaviour of generalised zig-zag process. Preprint. Available at arXiv:1710.01087.

[31] Hairer, M., Stuart, A. M. and Vollmer, S. J. (2014). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. MR3262508 https://doi.org/10.1214/13-AAP982

[32] Helffer, B. and Nier, F. (2005). *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians*. *Lecture Notes in Math.* **1862**. Springer, Berlin. MR2130405 https://doi.org/10.1007/b104762

[33] Holderrieth, P. (2019). Cores for piecewise deterministic Markov processes. Preprint. Available at arXiv:1910.11429.

[34] Kontoyiannis, I. and Meyn, S. P. (2012). Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probab. Theory Related Fields* **154** 327–339. MR2981426 https://doi.org/10.1007/s00440-011-0373-4

[35] Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347 https://doi.org/10.1090/surv/089

[36] Lu, J. and Wang, L. (2020). On explicit L^2-convergence rate estimate for piecewise deterministic Markov processes. Preprint. Available at arXiv:2007.14927.

[37] Mangoubi, O. and Smith, A. (2017). Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. Preprint. Available at arXiv:1708.07114.

[38] MATTINGLY, J. C., PILLAI, N. S. and STUART, A. M. (2012). Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.* **22** 881–930. MR2977981 https://doi.org/10.1214/10-AAP754

[39] MESQUITA, A. R. and HESPANHA, J. P. (2012). Jump control of probability densities with applications to autonomous vehicle motion. *IEEE Trans. Automat. Control* **57** 2588–2598. MR2991659 https://doi.org/10.1109/TAC.2012.2192356

[40] MEYER, K. R., HALL, G. R. and OFFIN, D. (2009). *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, 2nd ed. *Applied Mathematical Sciences* **90**. Springer, New York. MR2468466

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

[42] MONMARCHÉ, P. (2014). Hypocoercive relaxation to equilibrium for some kinetic models. *Kinet. Relat. Models* **7** 341–360. MR3195078 https://doi.org/10.3934/krm.2014.7.341

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

[44] NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

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

[46] OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937 https://doi.org/10.1016/j.jfa.2008.11.001

[47] PAKMAN, A., GILBOA, D., CARLSON, D. and PANINSKI, L. (2017). Stochastic bouncy particle sampler. In *International Conference on Machine Learning* 2741–2750.

[48] PAVLIOTIS, G. A. (2014). *Stochastic Processes and Applications*: *Diffusion Processes*, *the Fokker–Planck and Langevin Equations*. *Texts in Applied Mathematics* **60**. Springer, New York. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

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

[50] PHILLIPS, R. S. (1953). Perturbation theory for semi-groups of linear operators. *Trans. Amer. Math. Soc.* **74** 199–221. MR0054167 https://doi.org/10.2307/1990879

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

[52] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

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

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

[55] ROUSSEL, J. and STOLTZ, G. (2018). Spectral methods for Langevin dynamics and associated error estimates. *ESAIM Math. Model. Numer. Anal.* **52** 1051–1083. MR3865558 https://doi.org/10.1051/m2an/2017044

[56] STEELE, J. M. (1986). An Efron–Stein inequality for nonsymmetric statistics. *Ann. Statist.* **14** 753–758. MR0840528 https://doi.org/10.1214/aos/1176349952

[57] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2017). Piecewise deterministic Markov chain Monte Carlo. Preprint. Available at arXiv:1707.05296.

[58] VILLANI, C. (2009). Hypocoercivity. *Mem. Amer. Math. Soc.* **202** iv+141. MR2562709 https://doi.org/10.1090/S0065-9266-09-00567-5

[59] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[60] VONO, M., PAULIN, D. and DOUCET, A. (2020). Efficient MCMC sampling with dimension-free convergence rate using ADMM-type splitting. Preprint. Available at arXiv:1905.11937v5.

[61] WU, C. and ROBERT, C. P. (2017). Generalized bouncy particle sampler. Preprint. Available at arXiv:1706.04781.

[62] YANG, J., ROBERTS, G. O. and ROSENTHAL, J. S. (2020). Optimal scaling of random-walk Metropolis algorithms on general target distributions. *Stochastic Process. Appl.* **130** 6094–6132. MR4140028 https://doi.org/10.1016/j.spa.2020.05.004