

Subgeometric hypocoercivity for piecewise-deterministic markov process monte carlo methods

Andrieu, Christophe ; Dobson, Paul; Wang, Andi Q.

DOI

[10.1214/21-EJP643](https://doi.org/10.1214/21-EJP643)

Publication date

2021

Document Version

Final published version

Published in

Electronic Journal of Probability

Citation (APA)

Andrieu, C., Dobson, P., & Wang, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic markov process monte carlo methods. *Electronic Journal of Probability*, 26, 1-26. Article 78. <https://doi.org/10.1214/21-EJP643>

Important note

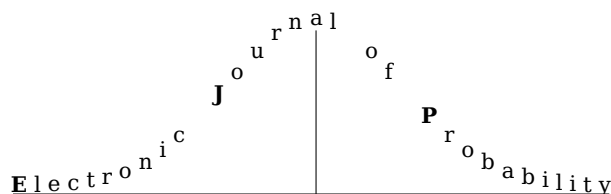
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Electron. J. Probab. **26** (2021), article no. 78, 1–26.
ISSN: 1083-6489 <https://doi.org/10.1214/21-EJP643>

Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods*

Christophe Andrieu[†] Paul Dobson[‡] Andi Q. Wang[§]

Abstract

We extend the hypocoercivity framework for piecewise-deterministic Markov process (PDMP) Monte Carlo established in [2] to heavy-tailed target distributions, which exhibit subgeometric rates of convergence to equilibrium. We make use of weak Poincaré inequalities, as developed in the work of [15], the ideas of which we adapt to the PDMPs of interest. On the way we report largely potential-independent approaches to bounding explicitly solutions of the Poisson equation of the Langevin diffusion and its first and second derivatives, required here to control various terms arising in the application of the hypocoercivity result.

Keywords: Piecewise-deterministic Markov process; hypocoercivity; Markov chain Monte Carlo; subgeometric convergence.

MSC2020 subject classifications: 60J25; 65C05.

Submitted to EJP on December 3, 2020, final version accepted on May 4, 2021.

Supersedes arXiv:2011.09341v2.

1 Introduction

In this work, we study piecewise-deterministic Markov processes (PDMPs) which are used in the context of Monte Carlo inference to draw samples from some given target density π on $\mathbb{R}^d$, for instance in Bayesian computation. Notable examples of such processes are the Zig-zag process of [5] and the Bouncy Particle Sampler of [9]. PDMPs have gained attention within the field of Markov chain Monte Carlo (MCMC) because such methods depart significantly from traditional reversible MCMC approaches

*Research of CA supported by EPSRC grants Bayes4Health, ‘New Approaches to Bayesian Data Science: Tackling Challenges from the Health Sciences’ (EP/R018561/1) and ‘CoSInES (COmputational Statistical Inference for Engineering and Security)’ (EP/R034710/1). Research of PD supported by the research programme ‘Zigzagging through computational barriers’ with project number 016.Vidi.189.043, financed by the Dutch Research Council (NWO). Research of AQW supported by EPSRC grant CoSInES (EP/R034710/1).

[†]University of Bristol, United Kingdom. E-mail: c.andrieu@bristol.ac.uk

[‡]Delft University of Technology, The Netherlands. E-mail: p.dobson@tudelft.nl

[§]University of Bristol, United Kingdom. E-mail: andi.wang@bristol.ac.uk

based on Metropolis–Hastings. These processes are constructed to be nonreversible, intuitively enabling persistent exploration within the state space, rather than the diffusive exploration characteristic of reversible schemes.

However, this nonreversibility also introduces mathematical difficulties when analysing the theoretical properties of the resulting algorithms, such as rates of convergence. Traditional methods based on spectral theory for self-adjoint operators in Hilbert spaces can no longer be applied, and furthermore the underlying operators which define the process tend to be non-coercive: the symmetric component of the operator has a nontrivial kernel. This implies that one cannot expect straightforward geometric convergence of the semigroup (P_t) , in the sense that there exists some $\rho > 0$, such that for appropriate functions f ,

$$\|P_t f\| \leq C \|f\| \xi(t), \quad \forall t \geq 0, \quad (1.1)$$

for some $C \leq 1$, $\xi(t) = \exp(-\rho t)$ for $t \geq 0$ and an appropriate norm $\|\cdot\|$. In order to understand degenerate dynamics, the hypocoercivity framework has been developed, following the approach of [11].

This framework was first applied to PDMPs in [2], where exponential convergence of the semigroups was proven as in (1.1) when the target density π satisfies a Poincaré inequality: for some constant $C_P > 0$,

$$\|\nabla f\|_2^2 \geq C_P \|f\|_2^2, \quad (1.2)$$

for suitably differentiable functions $f \in L^2(\pi)$, with $\int f \, d\pi = 0$, where $\|\cdot\|_2$ is the norm in $L^2(\pi)$. The authors were able to conclude that (1.1) holds for such targets, with an exponential rate function ξ , and for some constant $C > 1$.

The goal of this work is to extend the hypocoercivity results of [2] to targets π which do not possess a Poincaré inequality (1.2), but instead possess a weak Poincaré inequality of the form

$$\|f\|_2^2 \leq \alpha(r) \|\nabla f\|_2^2 + r \Psi(f), \quad \forall r > 0,$$

where $\alpha : (0, \infty) \rightarrow [1, \infty)$ is a decreasing function, typically divergent as $r \downarrow 0$, and Ψ is an appropriate functional. This encompasses target distributions which possess subgeometric tail decay, and are typically referred to as ‘heavy-tailed’. To do this, we will utilize the approach of [15], where such inequalities were studied for degenerate diffusions. Our main abstract result will be a convergence result of the form (1.1) where the rate function ξ is in fact subgeometric.

As a concrete application, our bounds on the semigroup of the form (1.1) will allow us to check conditions which ensure that a central limit theorem holds for (appropriately scaled) ergodic averages of the process.

1.1 Contribution

In this subsection, we carefully describe our contributions in relation to the literature, particularly references [2, 15] and [14]. Readers interested in our actual results are encouraged to move on to the following subsection where we define our notation, or to Section 2 for our assumptions, Section 3 for our abstract result, Section 4 for our result for PDMPs, or to Section 5 where we compare our bounds theoretically and empirically on illustrative examples.

In the present manuscript, we work in and extend the general framework laid out in [2] for PDMPs. The analysis carried out in [2] crucially relied on the existence of a (strong) Poincaré inequality (1.2), which enabled the authors to establish geometric convergence of the semigroup (1.1). These results rely themselves on the framework proposed by [11] for which the first rigorous proof was established in [14] and whose

results were adapted to take into account technical specificities of PDMPs in [2]. Our work aims to combine the framework recently proposed in [15] to tackle scenarios where application of the ideas of [11, 14] is sought, but only a weak form of the Poincaré inequality is satisfied, and [2] which takes into account PDMP idiosyncrasies.

More specifically, while our abstract assumptions and the resulting theorem and its proof (see Sections 2, 3) may superficially appear very similar to those of [15], we were not able to straightforwardly apply their results and have adapted them following [2]. This disparity fundamentally arises from the differences in the how the corresponding processes arise. For PDMPs, as in the present work and in [2], the initial point of departure is an explicit construction of the process, in terms of the deterministic dynamics and the switching mechanism, driven by an inhomogeneous Poisson process. The infinitesimal generator is a by-product, and its closure not sufficiently tractable to work with. On the other hand, for diffusion processes as in [15], one can begin with an appropriate differential operator, the putative infinitesimal generator, take the closure, and then via the standard operator-theoretic machinery define the ensuing semigroup and stochastic process. The key technical differences arise on the level of checking the closure of certain operators and we comment on this point in relevant places in the text.

In relation to the actual results of [15], by specialising to our particular PDMP setting we are also able to obtain slightly better constants in the decay of the semigroup. In relation to [2], by leveraging the powerful results of [16], we establish (quasi) potential-independent approaches to bounding the difficult cross terms arising in the application of the hypocoercivity result, which depend on smoothness estimates of the solution of the Poisson equation for the Langevin diffusion process.

The recent work [17] also studies geometric convergence of PDMP semigroups. Their work also crucially relies on a (strong) Poincaré inequality, and their Assumption 3 typically holds when the potential $U(x) = -\log \pi(x)$ grows at a superlinear rate in $|x|$. Thus the framework of [17] cannot currently be applied to heavy-tailed targets.

Finally, we briefly mention some other related work. For an accessible introduction to hypocoercivity, we recommend [1], which focusses on the finite-dimensional ODE setting. In [10], hypocoercivity techniques were used to study randomized Hamiltonian Monte Carlo (RHMC) and derive dimension-free exponential convergence rates. For connections between hypocoercivity and convergence proofs based on Lyapunov functions, see [18]. For a broad and recent review of the current literature on hypocoercivity, see [4].

1.2 Notation

- $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$ and for $v, w \in \mathbb{R}^d$, $\langle v, w \rangle = v^\top w$ is the associated inner product where $v^\top$ is the transpose of v .
- I_d denotes the $d \times d$ identity matrix.
- For a vector $w \in \mathbb{R}^d$ we will write w_i , $i = 1, \dots, d$, for its coordinates with respect to the standard basis.
- For $R > 0$, $B_R := \{x \in \mathbb{R}^d : |x| \leq R\}$.
- For any $s \in \mathbb{R}$, $(s)_+ := \max\{s, 0\}$ denotes the positive part.
- For a set, $\mathbb{I}A$ is the associated indicator or characteristic function.
- For a smooth manifold M and $k \in \mathbb{N} \cup \{\infty\}$, $C^k(M, \mathbb{R}^m)$ denotes the set of k -times continuously differentiable functions $f : M \rightarrow \mathbb{R}^m$. $C^k(M)$ denotes $C^k(M, \mathbb{R})$. $C_b^k(M)$ denotes functions in $C^k(M)$ which are in addition bounded and have bounded derivatives up to order k . A subscript $C_c^k(M)$ denotes functions in $C^k(M)$ which are compactly supported.
- For $\alpha \in (0, 1)$, $C^{k+\alpha}(M, \mathbb{R}^m)$ denotes the set of k -times continuously differentiable functions with locally α -Hölder continuous k -th derivative.

- For $f \in C^k(M)$, $i \in \{1, 2, \dots, d\}$, $x \mapsto \partial_i f(x)$ denotes the partial derivative of f with respect to the i th coordinate, for $k \geq 1$, and analogously for $i, j \in \{1, 2, \dots, d\}$, $\partial_{i,j} f$ denotes $\partial_i \partial_j f$, for $k \geq 2$.
- For a function $f \in C^2(X)$, $\nabla^2 f$ denotes the Hessian matrix of second-order partial derivatives. For $f \in C^2(X)$, $\Delta_x f := \sum_{i=1}^d \partial_{i,i} f$ denotes the Laplacian.
- For $f = (f_1, \dots, f_m) \in C^k(M, \mathbb{R}^m)$, $\nabla_x f$ is the gradient of f , defined at any $x \in M$ by $\nabla_x f(x) = (\partial_j f_i(x))_{i \in \{1, \dots, m\}, j \in \{1, \dots, d\}} \in \mathbb{R}^{d \times m}$.
- For any measurable space $(M, \mathcal{F})$ with probability measure m , we let $L^2(m)$ be the Hilbert space of (equivalence classes of) real measurable functions f with $\int_M |f|^2 dm < \infty$, with inner product $\langle f, g \rangle_2 = \int_M fg dm$ and corresponding norm $\|\cdot\|_2$. When there is ambiguity we may also write $\langle f, g \rangle_{L^2(m)}$, $\|\cdot\|_{L^2(m)}$ or $\langle f, g \rangle_m$, $\|\cdot\|_m$. We use the same notation for $\langle F, G \rangle_m = \int_M \text{Tr}(F^\top G) dm$ with $F, G: M \rightarrow \mathbb{R}^d$.
- $\text{Id}: L^2(m) \rightarrow L^2(m)$ denotes the identity mapping, $f \mapsto f$.
- We let $H^1(M, \mathbb{R}^m, m) = \{g \in C^1(M, \mathbb{R}^m) \cap L^2(m): \|\nabla_x g\|_2 < \infty\}$, $H^1(M) := H^1(M, \mathbb{R}, m)$ and $H^2(M) := \{g \in H^1(M) \cap C^2(M): \|\nabla_x^2 g\|_2 < \infty\}$. Note that we use the notation normally associated with Sobolev spaces, but our derivatives are not weak derivatives.
- For a measurable function $f: M \rightarrow \mathbb{R}$, let $\|f\|_{\text{osc}} := \text{ess}_m \sup f - \text{ess}_m \inf f$.
- Let $L_m^2(M; \mathbb{R}^d)$ be the space of functions $f: M \rightarrow \mathbb{R}^d$ with $\|h\|_{L_m^2(M; \mathbb{R}^d)}^2 < \infty$, endowed with the norm

$$\|h\|_{L_m^2(M; \mathbb{R}^d)}^2 = \int_M |h(x)|^2 dm(x).$$

- $L^\infty(m)$ will denote the Banach space of (equivalence classes of) measurable functions $f: M \rightarrow \mathbb{R}$ with $\text{ess}_m \sup |f| < \infty$.
- For an unbounded operator $(\mathcal{A}, D(\mathcal{A}))$, we let $\text{Ran}(\mathcal{A}) := \{\mathcal{A}f: f \in D(\mathcal{A})\}$ and $\text{Ker}(\mathcal{A}) := \{f \in D(\mathcal{A}): \mathcal{A}f = 0\}$. If $\mathcal{A}$ is closable, we let $\bar{\mathcal{A}}$ denote its closure.

1.3 PDMP notation

We summarize here our PDMP notation; for the underlying assumptions, see Section 2. Given potential $U: X \rightarrow \mathbb{R}$, we will denote the target distribution of interest by $\pi = e^{-U} / \int_X e^{-U(y)} dy$ on $X = \mathbb{R}^d$ equipped with its Borel σ -algebra. $V \subset \mathbb{R}^d$ is a closed subset and we have a probability measure ν defined on V equipped with its Borel σ -algebra $\mathcal{V}$. Then set $E = X \times V$ and define the augmented probability measure $\mu = \pi \otimes \nu$. We will be working with PDMPs whose generators are of the form, for $f \in C_b^2(E)$, $(x, v) \in E$,

$$\mathcal{L}f(x, v) = v^\top \nabla_x f(x, v) + \sum_{k=1}^K \lambda_k(x, v) [(\mathcal{B}_k - \text{Id})f](x, v) + m_2^{1/2} \lambda_{\text{ref}} \mathcal{R}_v f(x, v). \quad (1.3)$$

Here $\mathcal{R}_v$ is the refreshment operator, given for any $f \in L^2(\mu)$, by

$$\mathcal{R}_v f := \Pi_v f - f \text{ with } \Pi_v f(x, v) := \int_V f(x, w) d\nu(w).$$

For a sequence of continuous vector fields $F_k: X \rightarrow \mathbb{R}^d$, $k \in \{1, 2, \dots, K\}$, such that $\nabla_x U = \sum_{k=1}^K F_k$, we now define the corresponding bounce operators $\mathcal{B}_k$. For each $k \in \{1, 2, \dots, K\}$, $x \in X$ set

$$n_k(x) = \begin{cases} F_k(x)/|F_k(x)| & |F_k(x)| \neq 0, \\ 0 & \text{else,} \end{cases} \quad (1.4)$$

then set for any $f : E \rightarrow \mathbb{R}$, $(x, v) \in E$,

$$\mathcal{B}_k f(x, v) = f(x, v - 2(v^\top \mathbf{n}_k(x)) \mathbf{n}_k(x)).$$

The intensity $\lambda_k(x, v)$ has an explicit form, depending on the dynamics, which ensures invariance of μ . In particular, we require that $\lambda_k(x, v) - \lambda_k(x, -v) = v^\top F_k(x)$, which is a necessary condition for μ to be an invariant measure (see our Assumption 2.5). Finally, we set

$$m_2 := \int_{\mathcal{V}} v_1^2 d\nu(v),$$

which is assumed to be finite.

Remark 1.1. The PDMPs considered in [2] are slightly more general as they include the possibility of non-linear drift, that is the first order derivative term

$$v^\top \nabla_x f(x, v) - F_0(x)^\top \nabla_v f(x, v),$$

This includes important examples such as randomized HMC [8] and the Boomerang Sampler [6]; to simplify the expressions we have not included this term but the results hold in this more general setting also. More complex refreshment operators can also be considered; see [2].

2 PDMP assumptions

In what follows, ‘Conditions’ will refer to the conditions needed for the abstract hypocoercivity result to hold; these are inspired by [15]. ‘Assumptions’ will refer to the assumptions made on our PDMP process, which we will show imply that the Conditions hold. We first give a basic condition, following [2].

Condition 2.1. We have:

1. The operator $(\mathcal{L}, D(\mathcal{L}))$ is closed in $L^2(\mu)$ and generates a strongly continuous contraction semigroup $(P_t)_{t \geq 0}$ on $L^2(\mu)$.
2. μ is a stationary measure for $(P_t)_{t \geq 0}$.
3. There exists a core C for $\mathcal{L}$ such that C is dense in $L^2(\mu)$ and $C \subset D(\mathcal{L}) \cap D(\mathcal{L}^*)$, where $(\mathcal{L}^*, D(\mathcal{L}^*))$ is the adjoint of $\mathcal{L}$ on $L^2(\mu)$.

We now give the assumptions on the potential U .

Assumption 2.2. The potential U is such that,

1. $U \in C^{2+\alpha}(\mathbb{X})$ for some $\alpha \in (0, 1)$;
2. there exists a constant $c_U \geq 0$ such that for each $x \in \mathbb{X}$, $\nabla_x^2 U(x) \succeq -c_U I_d$ in the sense of definiteness of matrices;
3. either of the following holds:

- (a) $\nabla_x U$ is bounded,
- (b) $\|\nabla_x U\|_\pi < \infty$ and for some $C_U > 0$ and $\omega \geq 0$,

$$\Delta_x U(x) \leq C_U d^{1+\omega} + |\nabla_x U(x)|^2/2 \text{ for all } x \in \mathbb{X}. \quad (2.1)$$

We remark that this is where our present work diverges substantially from [2]. Since we are interested in studying subgeometric rates of convergence, we do not assume a Poincaré inequality here as in [2]: instead, we will later assume a weak Poincaré inequality.

Example 2.3. We shall consider two different heavy-tailed distributions:

1. Set $U(x) = \frac{1}{2}(d+p) \log(1+|x|^2)$, for some $p > 0$, in this case

$$\pi(dx) = \frac{Z_{d,p}}{(1+|x|)^{d+p}} dx$$

for some normalising constant $Z_{d,p}$;

2. Set $U(x) = \sigma|x|^\delta$, some $\sigma > 0$ and $0 < \delta < 1$, then

$$\pi(dx) = Z_{\sigma,\delta} e^{-\sigma|x|^\delta} dx$$

for some normalising constant $Z_{\sigma,\delta}$.

In both of these cases Assumption 2.2 holds since $\nabla_x U$ and $\nabla_x^2 U$ are both bounded. Both of these examples have subexponential decay but satisfy a weak Poincaré inequality as we show in Example 4.3 using the results of [21]. We assume the following assumptions on the vector fields, as in [2].

Assumption 2.4. The family of vector fields $\{F_k : X \rightarrow \mathbb{R}^d; k \in \{0, 1, \dots, K\}\}$ satisfies:

1. for $k \in \{0, 1, \dots, K\}$, $F_k \in C^2(X, \mathbb{R}^d)$;
2. for all $x \in X$, $\nabla_x U(x) = \sum_{k=1}^K F_k(x)$;
3. for all $k \in \{0, 1, \dots, K\}$, there exists $a_k \geq 0$ such that for all $x \in X$,

$$|F_k(x)| \leq a_k \{1 + |\nabla_x U(x)|\}.$$

Following [2], we make the following assumption on the event rate.

Assumption 2.5. The events rates are given by $\lambda_k(x, v) = \varphi(v^\top F_k(x))$, for each $(x, v) \in E$ and $k = 1, 2, \dots, K$, where $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ is a continuous function satisfying for any $s \in \mathbb{R}$,

$$\varphi(s) - \varphi(-s) = s, \quad |s| \leq \varphi(s) + \varphi(-s) \leq c_\varphi m_2^{1/2} + C_\varphi |s|,$$

for some constants $c_\varphi \geq 0, C_\varphi \geq 1$.

This assumption allows the canonical choices of rates, $\varphi(s) = (s)_+$ as well as smooth versions as in [3], such as $\varphi = -\log(\phi(\exp(-s)))$, for $\phi(r) = r/(1+r)$ or arbitrarily precise uniform approximation of canonical rates.

Example 2.6. Many standard PDMP algorithms satisfy these assumptions:

1. Let $K = d$ and for $k \in \{1, \dots, d\}$, $x \in X$, $F_k(x) = \partial_k U(x) e_k$, where e_k is the canonical basis, then we have the Zig-Zag process [5].
2. The choice $K = 1$ and $F_1 = \nabla_x U$ gives the Bouncy Particle Sampler [9].

Now we give assumptions on V and ν :

Assumption 2.7. We assume the following.

1. V is stable under bounces, i.e. for all $(x, v) \in E$ and $k \in \{1, \dots, K\}$, $v - 2(v^\top n_k(x))n_k(x) \in V$, where $n_k(x)$ is defined by (1.4);
2. For any $A \subseteq V$ Borel measurable, $x \in X$, we have $\nu((\text{Id} - 2n_k(x)n_k(x)^\top)A) = \nu(A)$, for any $k \in \{1, \dots, K\}$.
3. For any bounded and measurable function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, $i, j \in \{1, \dots, d\}$ such that $i \neq j$, $\int_V g(v_i, v_j) d\nu(v) = \int_V g(-v_i, v_j) d\nu(v)$;
4. ν has finite fourth order marginal moment,

$$m_4 := (1/3)\|v_1^2\|_2^2 = (1/3) \int_V v_1^4 d\nu(v) < \infty,$$

and for any $i, j, k, l \in \{1, 2, \dots, d\}$ such that $\int_V v_i v_j v_k v_l d\nu(v) = 0$ whenever $\text{card}(\{i, j, k, l\}) > 2$.

5. Assume that $m_2 \geq 1$.

The last condition is purely technical and allows for simpler expressions in Theorem 3.6.

We note that this assumption precludes the use of heavy-tailed distributions for the velocity component v . By the discussion after [2, H4] if ν is rotation invariant then Assumption 2.7-1-2-3 are satisfied. Assumption 2.7-2 above implies also that

$$m_{2,2} := \|v_1 v_2\|_2^2 = \int_{\mathcal{V}} v_1^2 v_2^2 d\nu(v) < \infty.$$

Assumption 2.8. The refreshment mechanism is given by

$$\mathcal{R}_v = \Pi_v - \text{Id}.$$

Assumption 2.9. The refreshment rate $\lambda_{\text{ref}} : X \rightarrow \mathbb{R}$ is bounded from below and above as follows: there exist $\underline{\lambda} > 0, c_\lambda \geq 0$ such that for each $x \in X$,

$$0 < \underline{\lambda} \leq \lambda_{\text{ref}}(x) \leq \underline{\lambda}(1 + c_\lambda |\nabla_x U(x)|).$$

3 Abstract result

We will decompose our operator $\mathcal{L}$ into symmetric and antisymmetric parts, $\mathcal{L} = \mathcal{S} + \mathcal{T}$ on $\mathcal{C}$, as in Condition 2.1, where

$$\mathcal{S} := (\mathcal{L} + \mathcal{L}^*)/2, \quad \mathcal{T} := (\mathcal{L} - \mathcal{L}^*)/2, \quad \mathcal{D}(\mathcal{S}) = \mathcal{D}(\mathcal{T}) = \mathcal{C}. \quad (3.1)$$

We remark that while our upcoming abstract result, Theorem 3.6, closely resembles Theorem 2.1 of [15], our definitions of the abstract operators $\mathcal{S}, \mathcal{T}$ are given by (3.1) above, which follows the approach of [2] instead. In our PDMP setting we require this approach in order to explicitly identify the operators $\mathcal{S}, \mathcal{T}$ in Section 4 and define intermediate quantities and their properties below. By contrast, in the diffusion setting of [15], the authors are able to employ Itô's Formula to identify their corresponding symmetric and antisymmetric operators. Thus we cannot simply use the approach and Theorem 2.1 of [15] directly, but we combine the two approaches of [2, 15].

Condition 3.1. $\Pi_v \mathcal{C} \subset \mathcal{C}$.

As in [2] we note that under Conditions 2.1 and 3.1, $\mathcal{T}\Pi_v$ is closable, with closure $(\overline{\mathcal{T}\Pi_v}, \mathcal{D}(\overline{\mathcal{T}\Pi_v}))$; this follows from the fact that $\mathcal{T}$ is antisymmetric and $\mathcal{C}$ is dense as shown in the proof of Lemma 3.11. This allows us to define, by [19, Theorem 5.1.9],

$$\mathcal{A} := (m_2 \text{Id} + (\mathcal{T}\Pi_v)^*(\overline{\mathcal{T}\Pi_v}))^{-1} (-\mathcal{T}\Pi_v)^*, \quad \mathcal{D}(\mathcal{A}) = \mathcal{D}((\mathcal{T}\Pi_v)^*).$$

As detailed in [2, Lemma 3] $\mathcal{A}$ is closable with bounded closure (on $L^2(\mu)$), and we denote its closure by $\overline{\mathcal{A}}$ hereafter. This is different from [14, 15] where instead it is assumed that $\mathcal{T}$ is closed, or closable, leading to a definition of $\mathcal{A}$ either involving $\mathcal{T}\Pi_v$, or $\overline{\mathcal{T}\Pi_v}$, and for which we could not establish key intermediate results, given the level of current understanding of PDMPs of the type considered in this manuscript.

Condition 3.2. We have:

1. $\text{Ran}(\Pi_v) \subset \text{Ker}(\mathcal{S}^*)$.
2. For any $f \in \mathcal{C}$, $\Pi_v \mathcal{T}\Pi_v f = 0$.

Condition 3.3. There exists some $R_0 \geq 1$ such that for any $f \in \mathcal{C}$,

$$|\langle \overline{\mathcal{A}} \mathcal{T}(\text{Id} - \Pi_v)f, f \rangle_2 + \langle \overline{\mathcal{A}} \mathcal{S}f, f \rangle_2| \leq R_0 \|(\text{Id} - \Pi_v)f\|_2 \|\Pi_v f\|_2.$$

We state the assumptions on the functional Ψ which will appear in our weak Poincaré inequalities.

Condition 3.4. We have a functional $\Psi : L^2(\mu) \rightarrow [0, \infty]$ such that the set $\{f \in L^2(\mu) : \Psi(f) < \infty\}$ is dense in $L^2(\mu)$. For any $f \in D(\mathcal{L})$ there exists a sequence $\{f_n\}_{n=1}^\infty \subset C$ such that $f_n \rightarrow f$ in $L^2(\mu)$ and

$$\limsup_{n \rightarrow \infty} \langle -\mathcal{L}f_n, f_n \rangle_2 \leq \langle -\mathcal{L}f, f \rangle_2, \quad \limsup_{n \rightarrow \infty} \Psi(f_n) \leq \Psi(f). \quad (3.2)$$

Further, setting $G := (\mathcal{T}\Pi_v)^* \overline{\mathcal{T}\Pi_v}$ with $D(G) = \{f \in D(\overline{\mathcal{T}\Pi_v}) : \overline{\mathcal{T}\Pi_v}f \in D((\mathcal{T}\Pi_v)^*)\}$, we also assume that Ψ satisfies for each $f \in L^2(\mu)$ and $t \geq 0$,

$$\Psi(P_t f) \leq \Psi(f), \quad \Psi(e^{-tG} f) \leq \Psi(f) \text{ and } \Psi(\Pi_v f) \leq \Psi(f). \quad (3.3)$$

We now state the required weak Poincaré inequalities.

Condition 3.5. Assume we have the following weak Poincaré inequalities: for some decreasing functions $\alpha_1, \alpha_2 : (0, \infty) \rightarrow [1, \infty)$,

$$\|\Pi_v f - \mu(f)\|_2^2 \leq \alpha_1(r) \|\mathcal{T}\Pi_v f\|_2^2 + r\Psi(\Pi_v f), \quad \forall f \in D(\mathcal{T}\Pi_v), r > 0, \quad (3.4)$$

$$\|(\text{Id} - \Pi_v)f\|_2^2 \leq \alpha_2(r) \langle -\mathcal{S}f, f \rangle_2 + r\Psi(f), \quad \forall f \in C, r > 0. \quad (3.5)$$

The following abstract result is inspired by Theorem 2.1 of [15] and Theorem 4 of [2].

Theorem 3.6. Assume Conditions 2.1, 3.1, 3.2, 3.3, 3.4, 3.5. Then there exist constants $c_1, c_2 > 0$ such that

$$\|P_t f - \mu(f)\|_2^2 \leq \xi(t) \left(\|f\|_2^2 + \Psi(f) \right), \quad \forall t \geq 0, f \in D(\mathcal{L}), \quad (3.6)$$

for

$$\xi(t) := c_1 \inf \left\{ r > 0 : c_2 t \geq \alpha_1(r)^2 \alpha_2 \left(\frac{r}{\alpha_1(r)^2} \right) \log \frac{1}{r} \right\}. \quad (3.7)$$

Expressions for c_1 and c_2 are given in (3.16).

Corollary 3.7. Assume the same conditions as Theorem 3.6, except that (3.5) is replaced by a strong Poincaré inequality,

$$\|(\text{Id} - \Pi_v)f\|_2^2 \leq C_P \langle -\mathcal{S}f, f \rangle_2, \quad \forall f \in C, \quad (3.8)$$

and assume furthermore that for each $f \in L^2(\mu)$ with $\Psi(f) < \infty$, we can find a sequence $(f_n) \subset D(\mathcal{L})$ with

$$f_n \rightarrow f \text{ in } L^2(\mu), \quad \limsup_{n \rightarrow \infty} \Psi(f_n) \leq \Psi(f). \quad (3.9)$$

Then we have that (3.6) holds, additionally, for any $f \in L^2(\mu)$, with

$$\xi(t) := c_1 \inf \left\{ r > 0 : c'_2 t \geq \alpha_1(r)^2 \log 1/r \right\},$$

for some $c_1, c'_2 > 0$.

As we shall see in Example 4.2, our application of Theorem 3.6 to PDMPs, given in Theorem 4.1, greatly broadens the class of PDMP Monte Carlo processes for which a central limit theorem holds. The following corollary will be applied to our examples in Section 5.

Corollary 3.8. Whenever

$$\int_0^\infty t^{-3/2} \xi^{1/2}(t) dt < \infty,$$

then for any $f \in L^2(\mu)$ such that $\mu(f) = 0$ and $\Psi(f) < \infty$, the finite-dimensional distributions of the rescaled process

$$t \mapsto \sigma^{-1} N^{-1/2} \int_0^{Nt} f(X_s, V_s) ds,$$

converges as $N \rightarrow \infty$ to those of a standard one-dimensional Brownian motion, where $\sigma > 0$ is an appropriately chosen constant defined in [22, Theorem MW].

Proof of Corollary 3.7. Fix some $f \in L^2(\mu)$. If $\Psi(f) = \infty$ then (3.6) vacuously holds. So assume $\Psi(f) < \infty$, and choose a sequence $(f_n) \subset D(\mathcal{L})$ satisfying (3.9). We can apply (3.6) to each f_n , and by taking the lim sup we conclude the inequality (3.6) is valid for f also. The alternative expression for ξ is immediate from the expression (3.7) since in this case, α_2 can be uniformly bounded from above. If c_2 denotes the constant in (3.7), then we can choose

$$c'_2 = \frac{c_2}{C_P}. \quad \square$$

Proof of Corollary 3.8. This is a direct application of [22, Theorem MW] which holds whenever, with $f \in L^2(\mu)$ such that $\int f d\mu = 0$ and $v_t := \int_0^t P_s f ds$ for each $t > 0$,

$$\int_0^\infty t^{-3/2} \|v_t\|_2 dt < \infty. \quad (3.10) \quad \square$$

Remark 3.9. The proof of Theorem 3.6 follows the proof of Theorem 2.1 of [15], however we have adapted the proof to take into account our differing assumptions and have been careful to track the constants involved. Due to the structure of the generators of the PDMPs we consider, under Assumptions 2.7 and 2.8, the operator $\mathcal{S}$ will satisfy (3.5) with $\alpha_2(r)$ constant and hence we may set $\Psi = 0$ in this inequality and obtain (3.8); see Corollary 3.7. This stems from the specific refreshment mechanism employed here – the measure ν may be required to satisfy a weak Poincaré inequality of the form of (4.2) when using a diffusion for this update; see Section 4.3 for a similar situation involving SDEs. Therefore we include the details of the proof of Theorem 3.6, making it straightforward to see how the constants simplify in Corollary 3.7.

Before we prove Theorem 3.6 we need the following two lemmas. The following is taken from [15, Lemma 2.3]:

Lemma 3.10 ([15, Lemma 2.3]). *Let $(A, D(A))$ be a densely defined closed linear operator on a separable Hilbert space $(H, \langle \cdot, \cdot \rangle, \|\cdot\|)$. Let $(T_t)_{t \geq 0}$ be the C_0 -contraction semigroup generated by the self-adjoint operator $-A^*A$ with domain $D(A^*A) := \{f \in D(A) : Af \in D(A^*)\}$. If the weak Poincaré inequality*

$$\|f\|^2 \leq \alpha(r) \|Af\|^2 + r\Psi(f), \quad r > 0, f \in D(A) \quad (3.11)$$

holds for some decreasing $\alpha : (0, \infty) \rightarrow (0, \infty)$, where $\Psi : H \rightarrow [0, \infty]$ satisfies

$$\Psi(T_t f) \leq \Psi(f), \quad t \geq 0, f \in D(A).$$

Then, for any $m_2 > 0$,

$$\|f\|^2 \leq (m_2 + \alpha(r)) \langle (m_2 \text{Id} + A^*A)^{-1} A^* A f, f \rangle + r\Psi(f), \quad r > 0, f \in D(A).$$

The following is a consequence of Lemma 3.10.

Lemma 3.11. *Assume Conditions 2.1, 3.1, 3.2, 3.3, 3.4 3.5, then for any $f \in C$*

$$\langle \overline{A\mathcal{T}} \Pi_v \Pi_v f, \Pi_v f \rangle_2 \leq -\frac{1}{m_2 + \alpha_1(r)} \|\Pi_v f\|_2^2 + \frac{r}{m_2 + \alpha_1(r)} \Psi(f).$$

Proof. We apply Lemma 3.10 with $H := L_0^2(\pi) = \{f \in L^2(\pi) : \pi(f) = 0\}$ and $(A, D(A)) = ((\mathcal{T}\Pi_v)^* \overline{\mathcal{T}\Pi_v})^{1/2}|_H, D(\overline{\mathcal{T}\Pi_v})$. The existence of A follows from: (a) $\mathcal{T}$ is anti-symmetric and $\mathbb{C}$ dense, implying that $\mathcal{T}\Pi_v$ is closable of closure we denote $(\overline{\mathcal{T}\Pi_v}, D(\overline{\mathcal{T}\Pi_v}))$, (b) $(\mathcal{T}\Pi_v)^* \overline{\mathcal{T}\Pi_v}$ is well defined and self-adjoint by [19, Theorem 5.1.9] and (c) $(A, D(A))$ exists as defined and is self-adjoint and positive [20, Theorem VIII.32]. Note that (3.11) holds by (3.4) and the fact that $D(\mathcal{T}\Pi_v)$ is a core for $(\overline{\mathcal{T}\Pi_v}, D(\overline{\mathcal{T}\Pi_v}))$. Now since $D(\mathcal{A}) = D((\mathcal{T}\Pi_v)^*)$, for $f \in D((\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v}))$,

$$\overline{\mathcal{A}\mathcal{T}\Pi_v}f = -(m_2 \text{Id} + (\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v}))^{-1} (\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v})f = -\phi((\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v}))f,$$

where $\phi(s) := s/(m_2 + s)$. Further from [19, Theorem 5.1.9] $D((\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v}))$ is a core for $D(\overline{\mathcal{T}\Pi_v})$ and $\overline{\mathcal{A}\mathcal{T}\Pi_v}$ is bounded, implying that on $D(\overline{\mathcal{T}\Pi_v})$, $\overline{\mathcal{A}\mathcal{T}\Pi_v} = -\phi((\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v}))$. This together with Condition 3.4 allows us to apply Lemma 3.10 to conclude that

$$\|f\|_2^2 \leq (m_2 + \alpha_1(r)) \langle \phi((\mathcal{T}\Pi_v)^* (\overline{\mathcal{T}\Pi_v}))f, f \rangle_2 + r\Psi(f), \quad \forall r > 0, f \in D(\overline{\mathcal{T}\Pi_v}).$$

Rearranging, setting $f = \Pi_v g$ for $g \in \mathbb{C} \subset D(\overline{\mathcal{T}\Pi_v})$ and using Condition 3.4 we obtain

$$\langle \Pi_v g, \Pi_v \mathcal{T}^* \mathcal{A}^* g \rangle_2 \leq -\frac{1}{m_2 + \alpha_1(r)} \|\Pi_v g\|_2^2 + \frac{r}{m_2 + \alpha_1(r)} \Psi(g). \quad \square$$

We now establish the following result which is based on [2, Lemma 5]. To prove the above theorem we need to define the closure $\bar{\mathcal{A}}$ of $\mathcal{A}$ which is defined on the whole space $L^2(\mu)$, this is possible since $\mathcal{A}$ is a bounded operator see [2, Lemma 3]. Define for any $g \in D(\mathcal{L})$

$$\mathcal{F}_1(g) = \langle \mathcal{L}g, g \rangle_2, \quad \mathcal{F}_2(g) = \langle \mathcal{L}g, \bar{\mathcal{A}}g \rangle_2, \quad \mathcal{F}_3(g) = \langle \bar{\mathcal{A}}\mathcal{L}g, g \rangle_2.$$

Lemma 3.12. Assume that Conditions 2.1, 3.1, 3.2, 3.3, 3.4, 3.5 hold. Then for any $g \in D(\mathcal{L})$ we have for any $r_1, r_2 > 0$,

$$\begin{aligned} \mathcal{F}_1(g) &\leq -\frac{1}{\alpha_2(r_2)} \|(\text{Id} - \Pi_v)g\|_2^2 + \frac{r_2}{\alpha_2(r_2)} \Psi(g), & \mathcal{F}_2(g) &\leq \|(\text{Id} - \Pi_v)g\|_2^2, \\ \mathcal{F}_3(g) &\leq -\frac{1}{m_2 + \alpha_1(r_1)} \|\Pi_v g\|_2^2 + \frac{r_1}{m_2 + \alpha_1(r_1)} \Psi(g) + R_0 \|(\text{Id} - \Pi_v)g\|_2 \|\Pi_v g\|_2. \end{aligned} \quad (3.12)$$

Proof. Note that the proof of the inequality for $\mathcal{F}_2$ in [2, Lemma 5] does not rely upon the Poincaré inequality so we may use the same proof to obtain for all $g \in D(\mathcal{L})$,

$$\mathcal{F}_2(g) \leq \|(\text{Id} - \Pi_v)g\|_2^2.$$

Fix $g \in \mathbb{C}$ and using that $\mathcal{T}$ is antisymmetric we have that $\langle \mathcal{L}g, g \rangle_2 = \langle \mathcal{S}g, g \rangle_2$ and by (3.5) we have for $r > 0$,

$$\mathcal{F}_1(g) \leq -\frac{1}{\alpha_2(r)} \|(\text{Id} - \Pi_v)g\|_2^2 + \frac{r}{\alpha_2(r)} \Psi(g).$$

To extend this to $D(\mathcal{L})$ fix $f \in D(\mathcal{L})$ and let $\{f_n\}_n \subseteq \mathbb{C}$ be as in Condition 3.4, then we have

$$\begin{aligned} \mathcal{F}_1(f) &\leq \limsup_{n \rightarrow \infty} \mathcal{F}_1(f_n) \\ &\leq \limsup_{n \rightarrow \infty} \left(-\frac{1}{\alpha_2(r)} \|(\text{Id} - \Pi_v)f_n\|_2^2 + \frac{r}{\alpha_2(r)} \Psi(f_n) \right). \end{aligned}$$

Now since $\text{Id} - \Pi_v$ is bounded and $\limsup_{n \rightarrow \infty} \Psi(f_n) \leq \Psi(f)$ we obtain

$$\mathcal{F}_1(f) \leq -\frac{1}{\alpha_2(r)} \|(\text{Id} - \Pi_v)f\|_2^2 + \frac{r}{\alpha_2(r)} \Psi(f).$$

Now consider $\mathcal{F}_3$, for any $g \in \mathcal{C} \subseteq D(\mathcal{L}) \cap D(\mathcal{L}^*) \cap D(\mathcal{T}\Pi_v)$ we have by Condition 3.3,

$$\begin{aligned}\mathcal{F}_3(g) &= \langle \bar{\mathcal{A}}\mathcal{T}\Pi_v g, g \rangle_2 + \langle \bar{\mathcal{A}}\mathcal{T}(\text{Id} - \Pi_v)g, g \rangle_2 + \langle \bar{\mathcal{A}}\mathcal{S}g, g \rangle_2 \\ &\leq \langle \Pi_v g, \Pi_v \mathcal{T}^* \mathcal{A}^* g \rangle_2 + R_0 \|(\text{Id} - \Pi_v)g\|_2 \|\Pi_v g\|_2.\end{aligned}$$

This inequality can be extended to $g \in D(\mathcal{L})$ since $\mathcal{C}$ is dense in $D(\mathcal{L})$ and $\bar{\mathcal{A}}\mathcal{T}\Pi_v$ is bounded, and Π_v and $\text{Id} - \Pi_v$ are bounded. From Lemma 3.11, we have

$$\langle \Pi_v g, \Pi_v \mathcal{T}^* \mathcal{A}^* g \rangle_2 \leq -\frac{1}{m_2 + \alpha_1(r)} \|\Pi_v g\|_2^2 + \frac{r}{m_2 + \alpha_1(r)} \Psi(g), \quad (3.13)$$

for $g \in \mathcal{C} \subseteq D(\bar{\mathcal{A}}\mathcal{T}\Pi_v)$. Note that (3.13) can be extended to $f \in D(\mathcal{L})$: fix $f \in D(\mathcal{L})$ and let $\{f_n\}_n \subseteq \mathcal{C}$ be as in Condition 3.4. Then (3.13) holds for each f_n and can be extended to f , since $\bar{\mathcal{A}}\mathcal{T}\Pi_v$ is bounded and by (3.2). Finally therefore, putting the pieces together we can conclude that for each $g \in D(\mathcal{L})$,

$$\mathcal{F}_3(g) \leq -\frac{1}{m_2 + \alpha_1(r)} \|\Pi_v g\|_2^2 + \frac{r}{m_2 + \alpha_1(r)} \Psi(g) + R_0 \|(\text{Id} - \Pi_v)g\|_2 \|\Pi_v g\|_2. \quad \square$$

Proof of Theorem 3.6. We combine approach of [2, Theorem 4] with that of [15, Theorem 2.1]. Without loss of generality we may assume $\mu(f) = 0$. Let us define for any $\epsilon > 0$ and $g \in L^2(\mu)$,

$$\mathcal{H}_\epsilon(g) := \frac{1}{2} \|g\|_2^2 + \epsilon \langle g, \bar{\mathcal{A}}g \rangle.$$

As in [2], we have the equivalence, for any $0 < \epsilon < (m_2/2)^{1/2}$ and $g \in L^2(\mu)$,

$$\frac{1 - (m_2/2)^{-1/2}\epsilon}{2} \|g\|_2^2 \leq \mathcal{H}_\epsilon(g) \leq \frac{1 + (m_2/2)^{-1/2}\epsilon}{2} \|g\|_2^2. \quad (3.14)$$

For $f \in L^2(\mu)$ let us write for convenience, $f_t := P_t f$ for each $t \geq 0$. Then from the Dynkin formula we know that $f_t \in D(\mathcal{L})$ and $df_t/dt = \mathcal{L}f_t$ for each $t > 0$. Then, we can use Lemma 3.12 to obtain

$$\begin{aligned}-\frac{d}{dt} \mathcal{H}_\epsilon(f_t) &= -[\mathcal{F}_1(f_t) + \epsilon\{\mathcal{F}_2(f_t) + \mathcal{F}_3(f_t)\}] \\ &\geq \left(\frac{1}{\alpha_2(r_2)} - \epsilon \right) \|(\text{Id} - \Pi_v)f_t\|_2^2 + \frac{\epsilon}{m_2 + \alpha_1(r_1)} \|\Pi_v f_t\|_2^2 \\ &\quad - \left(\frac{r_2}{\alpha_2(r_2)} + \frac{\epsilon r_1}{m_2 + \alpha_1(r_1)} \right) \Psi(f_t) - \epsilon R_0 \|(\text{Id} - \Pi_v)f_t\|_2 \|\Pi_v f_t\|_2.\end{aligned}$$

We now follow the calculations in the proof of [15, Theorem 2.1]. Our approach is very similar, however we obtain slightly better bounds (from our Lemma 3.12) which lead to slightly better constants in the end; hence we include a full proof. We use Young's inequality to bound the cross term

$$\epsilon R_0 \|(\text{Id} - \Pi_v)f_t\|_2 \|\Pi_v f_t\|_2 \leq \frac{\epsilon \|\Pi_v f_t\|_2^2}{2(\alpha_1(r_1) + m_2)} + \frac{\epsilon R_0^2(\alpha_1(r_1) + m_2)}{2} \|(\text{Id} - \Pi_v)f_t\|_2^2.$$

This gives us

$$\begin{aligned}-\frac{d}{dt} \mathcal{H}_\epsilon(f_t) &\geq \left(\frac{1}{\alpha_2(r_2)} - \frac{\epsilon R_0^2(\alpha_1(r_1) + m_2)}{2} - \epsilon \right) \|(\text{Id} - \Pi_v)f_t\|_2^2 \\ &\quad + \frac{\epsilon}{2(\alpha_1(r_1) + m_2)} \|\Pi_v f_t\|_2^2 - \left(\frac{r_2}{\alpha_2(r_2)} + \frac{\epsilon r_1}{m_2 + \alpha_1(r_1)} \right) \Psi(f_t).\end{aligned}$$

Now we take

$$\epsilon = \frac{1}{\alpha_2(r_2)[R_0^2(\alpha_1(r_1) + m_2) + 2]} < \frac{1}{2}, \quad (3.15)$$

and using the fact that $\|f_t\|_2^2 = \|\Pi_v f_t\|_2^2 + \|(\text{Id} - \Pi_v)f_t\|_2^2$ and (3.3) we get

$$\begin{aligned} -\frac{d}{dt} \mathcal{H}_\epsilon(f_t) &\geq \frac{1}{2\alpha_2(r_2)} \|(\text{Id} - \Pi_v)f_t\|_2^2 \\ &\quad + \frac{1}{2\alpha_2(r_2)(\alpha_1(r_1) + m_2)[R_0^2(\alpha_1(r_1) + m_2) + 2]} \|\Pi_v f_t\|_2^2 \\ &\quad - \left(\frac{r_2}{\alpha_2(r_2)} + \frac{r_1}{(m_2 + \alpha_1(r_1))\alpha_2(r_2)[R_0^2(\alpha_1(r_1) + m_2) + 2]} \right) \Psi(f_t) \\ &\geq \frac{1}{4R_0^2\alpha_2(r_2)(\alpha_1(r_1) + m_2)^2} \|f_t\|_2^2 \\ &\quad - \left(\frac{r_2}{\alpha_2(r_2)} + \frac{r_1}{R_0^2\alpha_2(r_2)(\alpha_1(r_1) + m_2)^2} \right) \Psi(f_t). \end{aligned}$$

Now since $\epsilon < 1/2$ we use (3.14) to see that for $g \in L^2(\mu)$

$$\|g\|_2^2 \geq \frac{4\mathcal{H}_\epsilon(g)}{2 + (m_2/2)^{-1/2}},$$

so

$$\begin{aligned} -\frac{d}{dt} \mathcal{H}_\epsilon(f_t) &\geq \frac{\mathcal{H}_\epsilon(f_t)}{(2 + (m_2/2)^{-1/2})R_0^2\alpha_2(r_2)(\alpha_1(r_1) + m_2)^2} \\ &\quad - \left(\frac{r_2}{\alpha_2(r_2)} + \frac{r_1}{R_0^2\alpha_2(r_2)(\alpha_1(r_1) + m_2)^2} \right) \Psi(f_t). \end{aligned}$$

So by Gronwall's lemma, for $t \geq 0$,

$$\begin{aligned} \mathcal{H}_\epsilon(f_t) &\leq \exp \left[-\frac{t}{(2 + (m_2/2)^{-1/2})R_0^2\alpha_2(r_2)(\alpha_1(r_1) + m_2)^2} \right] \mathcal{H}_\epsilon(f) \\ &\quad + \left[r_2(2 + (m_2/2)^{-1/2})R_0^2(\alpha_1(r_1) + m_2)^2 + r_1(2 + (m_2/2)^{-1/2}) \right] \Psi(f). \end{aligned}$$

Now we choose $r_1 = r$, $r_2 = r/\alpha_1(r_1)^2$, and then using that $m_2 \geq 1$ (Assumption 2.7) and (3.14),

$$\|f_t\|_2^2 \leq c_1 \exp \left[-\frac{c_2 t}{\alpha_1(r)^2 \alpha_2(r/\alpha_1(r)^2)} \right] \|f\|_2^2 + c_1 r \Psi(f),$$

for $r > 0$, $f \in D(\mathcal{L})$, $t \geq 0$. Here we can take for ϵ as defined in (3.15),

$$\begin{aligned} c_1 &= \frac{2 \max \left\{ \frac{1+\epsilon(m_2/2)^{-1/2}}{2}, (2 + (m_2/2)^{-1/2}) [R_0^2(1 + m_2)^2 + 1] \right\}}{1 - \epsilon(m_2/2)^{-1/2}}, \\ c_2 &= \frac{1}{R_0^2(2 + (m_2/2)^{-1/2})(1 + m_2)^2}. \end{aligned} \quad (3.16)$$

Thus we can conclude that (3.6) holds with ξ as in (3.7) for some $c_1, c_2 > 0$. $\square$

4 Application to PDMPs

In this case the operator $\mathcal{L}$ acts on smooth functions in $C = C_b^2(E)$ as follows,

$$\mathcal{L}f(x, v) = v^\top \nabla_x f(x, v) + \sum_{k=1}^K \lambda_k(x, v)(\mathcal{B}_k - \text{Id})f(x, v) + m_2^{1/2} \lambda_{\text{ref}}(x) \mathcal{R}_v f(x, v). \quad (4.1)$$

In which case we have that the operators $\mathcal{S}$ and $\mathcal{T}$ are given for functions $f \in \mathbb{C}$ by

$$\begin{aligned}\mathcal{S}f(x, v) &= \frac{1}{2} \sum_{k=1}^K \lambda_k^e(x, v) (\mathcal{B}_k - \text{Id}) f(x, v) + m_2^{1/2} \lambda_{\text{ref}}(x) \mathcal{R}_v f(x, v), \\ \mathcal{T}f(x, v) &= v^\top \nabla_x f(x, v) + \frac{1}{2} \sum_{k=1}^K v^\top F_k (\mathcal{B}_k - \text{Id}) f(x, v).\end{aligned}$$

Where, $\lambda_k^e(x, v) := \lambda_k(x, v) + \lambda_k(x, -v)$, for $k \in \{1, \dots, K\}$ and $(x, v) \in \mathbb{E}$. Recall that $G := (\mathcal{T}\Pi_v)^* \mathcal{T}\Pi_v$ with $D(G) = \{f \in D(\mathcal{T}\Pi_v) : \mathcal{T}\Pi_v f \in D((\mathcal{T}\Pi_v)^*)\}$. By [2, Lemma 9(b)], we have that $C_b^2(\mathbb{E}) \subset D(G)$ and for any $f \in C_b^2(\mathbb{E})$, $Gf = m_2 \nabla_x^* \nabla_x \Pi_v f$, here ∇^* is the adjoint of ∇ as an operator from $D(\nabla) \subseteq L^2(\pi) \rightarrow L^2_\pi(\mathbb{X}; \mathbb{R}^d)$. We take $\Psi = \|\cdot\|_{\text{osc}}^2$. Our goal in the section is the following theorem, which will follow from Corollary 3.7 once we have checked that the abstract conditions hold.

Theorem 4.1. Assume that our basic Condition 2.1 and our PDMP Assumptions 2.2, 2.4, 2.5, 2.7, 2.8, 2.9 hold. Then we have convergence of the semigroup,

$$\|P_t f\|_2^2 \leq \xi(t) \left(\|f\|_2^2 + \|f\|_{\text{osc}}^2 \right), \quad \forall t \geq 0, f \in L^2(\mu),$$

where

$$\xi(t) := c_1 \inf \{r > 0 : c_2 t \geq \alpha_1(r)^2 \log 1/r\},$$

for a decreasing function $\alpha_1 : (0, \infty) \rightarrow [1, \infty)$ and some constants $c_1, c_2 > 0$. In particular, c_2 may be taken to be

$$c_2 = \frac{\lambda m_2^{1/2}}{R_0^2(2 + (m_2/2)^{-1/2})(1 + m_2)^2}.$$

Example 4.2 (Central limit theorems.). From our results in Example 5.1, we can apply Corollary 3.8 to our running examples to see that a central limit theorem holds for $f \in L^2(\mu)$ such that $\|f\|_{\text{osc}} < \infty$,

1. for $U(x) = \frac{1}{2}(d+p) \log(1 + |x|^2)$ whenever p is large enough so that $\tau < 1/2$, where τ is defined later in (4.3),
2. for $U(x) = \sigma|x|^\delta$ for any $\sigma > 0$, $0 < \delta < 1$.

4.1 Checking condition 2.1

In [2] it is argued that the BPS and the ZZ processes are both well-defined Markov processes satisfying Condition 2.1 with $C = C_b^2(\mathbb{E})$ as a core (see their remarks after their Corollary 2). In order to help the reader we provide here a brief overview of existing theoretical results which have been used to establish a similar property, and can be adapted to establish Condition 2.1. For the BPS, it is shown in [12], in full detail, that $C_c^1(\mathbb{E})$ is core for its generator on $C_0(\mathbb{E})$, the set of continuous functions vanishing at infinity. This relies on a stability property of $C^1(\mathbb{E})$, the set of continuously differentiable functions, under the semigroup [12, Lemma 17], which by using [13, Proposition 3.3] implies the core property. As remarked in [12, Remark 18], [12, Lemma 17] can be extended to cover stability of $C_c^k(\mathbb{E})$ for $k \geq 2$ and an application of [13, Proposition 3.3] leads to the desired conclusion for the generator on $C_0(\mathbb{E})$. Crucially [12, Lemma 17] requires the intensity to belong to $C^1(\mathbb{E})$, a property not satisfied by the standard BPS or ZZ when using the “canonical” choice of intensity. This is however relaxed for BPS by utilising their [12, Theorem 21] in conjunction with [12, Propositions 9 and 23]. In [3, Theorem 5.11] these ideas are used in the context of the ZZ process to establish in detail that $C_c^1(\mathbb{E})$ is a core for the generator on $L^2(\mu)$, assuming that the intensities involved belong to $C^1(\mathbb{E})$ —this extends directly to the scenario where $C^1(\mathbb{E})$ is replaced

with $C^2(E)$. We note that [3, Proposition 5.17] defines a family of smooth intensities, $C^2(E)$ under Condition 2.2, uniformly converging to the popular canonical choice, so we may apply [12, Proposition 11 and 27] to get that $C_c^2(E)$ is a core for the generator in $C_0(E)$ of ZZ with the canonical intensity. Since the semigroup preserves the domain of the generator in $C_0(E)$, when we extend the semigroup to $L^2(\mu)$ we have that the domain of the generator is a core for the generator in $L^2(\mu)$ by using [13, Proposition 3.3]. Then since convergence in $C_0(E)$ implies convergence in $L^2(\mu)$ we have that $C_c^2(E)$ is a core for the generator of ZZ with the canonical intensity in $L^2(\mu)$. This immediately implies that $C_b^2(E)$ is a core in $L^2(\mu)$ as desired. See also [7] for a direct approach in the one-dimensional ZZ case.

4.2 Checking Condition 3.1 and 3.2

Condition 3.1 follows from Lemma 9 of [2]. Condition 3.2-1 is immediate from the definition of $\mathcal{S}$. Finally, Condition 3.2-2 follows from Lemma 9 of [2].

4.3 Checking Condition 3.5: weak Poincaré inequalities

To establish weak Poincaré inequalities, our starting point is [21], as in [15]. [21, Theorem 3.1] and the subsequent remark allow us to deduce that there exists decreasing functions $\alpha_1, \alpha_2 : (0, \infty) \rightarrow [1, \infty)$ such that the weak Poincaré inequalities hold

$$\pi(f^2) - \pi(f)^2 \leq \alpha_1(r)\pi(|\nabla_x f|^2) + r\|f\|_{\text{osc}}^2, \forall f \in C_b^1(X), r > 0. \quad (4.2)$$

We need to show conditions (3.4) and (3.8) hold. First by [2, Lemma 9 (b)] we have for any $f \in C$,

$$(\mathcal{T}\Pi_v)^*(\mathcal{T}\Pi_v)f = m_2\nabla_x^*\nabla_x\Pi_v f.$$

Multiplying by $\Pi_v f$ and integrating we obtain

$$\|(\mathcal{T}\Pi_v)f\|_2^2 = m_2\|\nabla\Pi_v f\|_2^2.$$

Therefore substituting the above expression into (4.2) we have

$$\|\Pi_v f - \mu(f)\|_2^2 \leq \frac{\alpha_1(r)}{m_2}\|\mathcal{T}\Pi_v f\|_2^2 + r\|f\|_{\text{osc}}^2, \forall f \in C_b^1(E), r > 0.$$

Thus we have (3.4). Note that (3.8) follows immediately from [2, Proposition 10], since we have the same refreshment mechanism. Indeed, we can obtain for $f \in C$,

$$\|(\text{Id} - \Pi_v)f\|_2^2 \leq \frac{1}{\lambda m_2^{1/2}} \langle -\mathcal{S}f, f \rangle.$$

Example 4.3 (Example 2.3 continued). By [21, Theorem 3.2] we have that (4.2) holds and

$$\alpha(r) = \frac{4R_r^2}{\pi^2} e^{\delta_{R_r}(U)}$$

where $R_r := \inf\{s > 0 : \pi(B_s^c) \leq r/(1+r)\}$, $\delta_R(U) = \sup_{\{x,y \in B_R\}} U(x) - U(y)$ and $B_R = \{x \in \mathbb{R}^d : |x| \leq R\}$. We shall exhibit α for the two cases considered in Example 2.3

1. If $U(x) = \frac{1}{2}(d+p)\log(1+|x|^2)$, for some $p > 0$ then by [21, Example 1.4(a)] we have that (4.2) holds with

$$\alpha_1(r) = c(1+r^{-\tau})$$

$$\tau = \min \left\{ \frac{d+p+2}{p}, \frac{4p+4+2d}{[p^2-4-2d-2p]^+} \right\}, \quad (4.3)$$

where c is a universal constant independent of dimension and r . Note that [21, Example 1.4(a)] considers a slightly different function, $V(x) = (d + p) \log(1 + |x|)$. However, the difference between these functions is bounded so if (4.2) holds for V , then it also holds for U .

2. If $U(x) = \sigma|x|^\delta$ then by [21, Example 1.4 (c)] we have that (4.2) holds with

$$\alpha_1(r) = c [1 + \log(1 + r^{-1})]^{4(1-\delta)/\delta}. \quad (4.4)$$

We detail convergence rates these potentials lead to in Section 5.

4.4 Checking Condition 3.3: finding R_0

The most difficult part of the proof is checking Condition 3.3 which will control the remainder terms. For light-tailed targets in [2] this is done by showing that solutions of the Poisson equation

$$m_2(\text{Id} + \nabla_x^* \nabla_x)u_f = \Pi_v f, \quad (4.5)$$

have polynomially growing derivatives. This implies in their setting that they are π -integrable, however for heavy-tailed measures π this is not sufficient. By using Schauder estimates we know that the solution u_f of the Poisson equation is twice differentiable, however we do not know in general that the derivatives are π -integrable. By multiplying the solution by smooth cut-off functions it is shown in [16] that under Assumption 2.2 the first and second derivatives are in $L^2(\pi)$. We can write down the solution as

$$u_f = m_2^{-1}(\text{Id} + \nabla_x^* \nabla_x)^{-1} \Pi_v f,$$

since ∇_x is a densely-defined closed operator on $L^2(\pi)$, [2, Proposition 26] shows that $(\text{Id} + \nabla_x^* \nabla_x)^{-1}$ is a positive self-adjoint bounded operator on $L^2(\pi)$, which furthermore is a bijection between $L^2(\pi)$ and $D(\nabla_x^* \nabla_x)$. In our case, we will utilize the powerful abstract result of [16].

We also remark here that our subsequent argument patches a minor omission in [15], in their proof of (H3) for degenerate diffusions. In [15], the authors reference [11] and [14, Section 5.1] for elliptic a priori estimates. However, the cited references assume the existence of a strong Poincaré inequality, which precisely falls outside the scope of the processes under consideration.

Lemma 4.4. *Under our Assumption 2.2 for any $f \in C_b^2(E)$,*

1. *the solution u_f of the Poisson equation (4.5) is uniquely defined and $u_f \in C^3(X)$;*
2. *there exist some $\kappa_1, \kappa_2 > 0$ such that:*

$$\max \{ \|u_f\|_2, \|\nabla_x u_f\|_2 \} \leq m_2^{-1} \|\Pi_v f\|_2, \quad (4.6)$$

$$\|\nabla_x^2 u_f\|_2 \leq m_2^{-1} \kappa_1 \|\Pi_v f\|_2, \quad (4.7)$$

$$\|(\nabla_x U)^\top \nabla_x u_f\|_2 \leq m_2^{-1} \kappa_2 \|\Pi_v f\|_2. \quad (4.8)$$

Here

$$\kappa_1 = \sqrt{2(2 + c_U)},$$

and if $|\nabla_x U|$ is bounded then we may take $\kappa_2 = \sup_{x \in X} |\nabla_x U(x)|$, otherwise $\kappa_2 = \sqrt{4(4\kappa_1 + C_U d^{1+\omega})}$, where ω, c_U and C_U are as in Assumption 2.2.

Proof. Note by rescaling f we may take $m_2 = 1$. The fact that $u_f \in L^2(\pi)$ follows from the fact that $(\text{Id} + \nabla_x^* \nabla_x)^{-1}$ is a positive self-adjoint bounded operator on $L^2(\pi)$, as detailed in [2, Proposition 26]. We now make use of [16, Theorem 3.3]. Since we are dealing with a simplified version of the Poisson equation (4.5), the Hypotheses 2.1(i)-(iii) of [16] are trivially satisfied. Hypothesis 2.1(iv) of [16] is equivalent in our setting to Assumption 2.2-2. Hence we have satisfied the hypotheses of [16, Theorem 3.3]. The bounds (4.6) then follow immediately from [16, Theorem 3.3]. The finiteness and upper bound of $\|\nabla_x^2 u_f\|_2$ follows from revisiting the proof of [16, Theorem 3.3], see Appendix A.2. If there exists κ_2 such that $\|\nabla_x U\|_\infty \leq \kappa_2$ then (4.8) follows from (4.6). Now consider the case where (2.1) holds.

In the following, for functions $\mathbb{R}^d \rightarrow \mathbb{R}^d$, we will use the bare notation $\|\cdot\|$ to denote the norm $\|\cdot\|_{L_\pi^2(\mathbb{R}^d; \mathbb{R}^d)}$. Following the proof of [2, Lemma 34] we have for any $\phi \in C_c^\infty(X)$ and $\varepsilon > 0$ that

$$\|\phi \nabla_x U\|^2 - \langle \phi^2, \Delta_x U \rangle_2 \leq \varepsilon^{-1} \|\nabla_x \phi\|^2 + \varepsilon \|\phi \nabla_x U\|^2.$$

Now using (2.1) we obtain

$$\frac{1}{2} \|\phi \nabla_x U\|^2 - C_U d^{1+\omega} \|\phi\|_2^2 \leq \varepsilon^{-1} \|\nabla_x \phi\|^2 + \varepsilon \|\phi \nabla_x U\|^2.$$

Rearranging and setting $\varepsilon = 1/4$ gives

$$\frac{1}{4} \|\phi \nabla_x U\|^2 \leq 4 \|\nabla_x \phi\|^2 + C_U d^{1+\omega} \|\phi\|^2. \quad (4.9)$$

As C_c^∞ is a core for $(\nabla_x, D(\nabla_x))$ we have the above inequality for any $\phi \in D(\nabla_x)$. In particular we shall set $\phi(x) = \partial_i u_f$ which gives

$$\frac{1}{4} \|\partial_i u_f \nabla_x U\|^2 \leq 4 \|\nabla_x \partial_i u_f\|^2 + C_U d^{1+\omega} \|\partial_i u_f\|^2.$$

Now summing over $i \in \{1, \dots, d\}$ we obtain

$$\begin{aligned} \|\nabla_x u_f\| \|\nabla_x U\|^2 &= \sum_{i=1}^d \|\partial_i u_f \nabla_x U\|^2 \\ &\leq 4(4 \|\nabla_x^2 u_f\|_{L_\pi^2(\mathbb{R}^d; \mathbb{R}^{d \times d})}^2 + C_U d^{1+\omega} \|\nabla_x u_f\|^2) \\ &\leq 4(4\kappa_1 + C_U d^{1+\omega}) \|\Pi_v f\|_2^2. \end{aligned}$$

Recall here that $\|\nabla_x u_f(x)\|$ denotes the Euclidean norm of $\nabla_x u_f(x)$. Finally by using Cauchy-Schwarz we obtain the desired result with $\kappa_2 = \sqrt{4(4\kappa_1 + C_U d^{1+\omega})}$. $\square$

A proof of the following result is given in [2, Lemma 9(c)]; however it relies on a density argument involving $C_{\text{poly}}^3(X)$, therefore requiring the existence of moments under π and hence stronger assumption on U . The below establishes that this assumption is not required.

Lemma 4.5. *Under Assumption 2.2, then for any $f \in L^2(\pi)$*

$$\{m_2 \text{Id} + (\mathcal{T} \Pi_v)^* \overline{\mathcal{T} \Pi_v}\}^{-1} f = m_2 \{\text{Id} + \nabla_x^* \nabla_x \Pi_v\}^{-1} f. \quad (4.10)$$

Proof. The proof is along the same lines as that of [2, Lemma 9(c)], replacing $C_{\text{poly}}^3(X)$ with $H^2(X)$, thanks to the results of [16]. More precisely from [16, Lemma 3.1] we have that for any $g \in H^2(X)$ we can find $\{g_n \in C_b^2(X)\}_{n \in \mathbb{N}}$ such that $\nabla_x g_n \rightarrow \nabla_x g$ and $\nabla_x^2 g_n \rightarrow \nabla_x^2 g$ in $L^2(\pi)$, therefore implying $\{m_2 \text{Id} + (\mathcal{T} \Pi_v)^* \overline{\mathcal{T} \Pi_v}\} g_n \rightarrow m_2 \{\text{Id} + \nabla_x^* \nabla_x\} g$ and $\{m_2 \text{Id} + (\mathcal{T} \Pi_v)^* \overline{\mathcal{T} \Pi_v}\} g_n \rightarrow m_2 \{\text{Id} + \nabla_x^* \nabla_x\}^{-1} g$ since $\nabla_x^* \nabla_x h = -\Delta_x h + \nabla_x U^\top \nabla_x h$ for $h \in H^2(X)$ (see Lemma A.3) and the two operators are closed. Therefore the two

operators coincide on $H^2(X) \subset D(m_2 \text{Id} + (\mathcal{T}\Pi_v)^* \overline{\mathcal{T}\Pi_v}) \cap D(\text{Id} + \nabla_x^* \nabla_x)$. Again from [16] we have $C_b^2(X) \subset m_2(\text{Id} + \nabla_x^* \nabla_x)(H^2(X))$ (see Lemma 4.4), which is dense in $L^2(\pi)$ and we deduce (4.10) by boundedness of the two inverses. $\square$

In order to show that Condition 3.3 follows we use [2, Lemma 12 & 13] which states the following.

Lemma 4.6 ([2, Lemma 12 & 13]). *Assume that $\mathcal{L}$ is given by (4.1). Assume in addition that Assumptions 2.2 - 2.9 hold.*

1. For any $f \in C_b^2(E)$,

$$|\langle \mathcal{AS}(\text{Id} - \Pi_v)f, f \rangle_2| \leq \|(\text{Id} - \Pi_v)f\|_2 \|(\text{Id} - \Pi_v)\mathcal{SA}^*f\|_2.$$

2. For any $f \in C_b^2(E)$,

$$\begin{aligned} \|(\text{Id} - \Pi_v)\mathcal{SA}^*f\|_2 &\leq m_2(\|\lambda_{\text{ref}} \nabla_x u_f\|_2 + c_\varphi K \|\nabla_x u_f\|_2) \\ &\quad + C_\varphi \sqrt{2m_{2,2} + 3(m_4 - m_{2,2})_+} \sum_{k=1}^K \|F_k^\top \nabla_x u_f\|_2. \end{aligned}$$

3. For any $f \in C_b^2(E)$,

$$|\langle \mathcal{AT}(\text{Id} - \Pi_v)f, f \rangle_2| \leq \|(\text{Id} - \Pi_v)f\|_2 \|(\text{Id} - \Pi_v)\mathcal{TA}^*f\|_2.$$

4. For any $f \in C_b^2(E)$,

$$\|(\text{Id} - \Pi_v)\mathcal{TA}^*f\|_2 \leq \sqrt{3(m_4 - m_{2,2})_+ + 2m_{2,2}(\kappa_1 + \kappa_2)} \|\Pi_v f\|_2,$$

with κ_1 and κ_2 positive constants as defined in Lemma 4.4.

Corollary 4.7. *Assume that $\mathcal{L}$ is given by (4.1). Assume in addition that Assumptions 2.2 - 2.9 hold. Then Condition 3.3 holds with*

$$\begin{aligned} R_0 &= \sqrt{3(m_4 - m_{2,2})_+ + 2m_{2,2}(\kappa_1 + \kappa_2)} \\ &\quad + m_2^{-1} \left(m_2(\lambda(1 + c_\lambda \kappa_2) + c_\varphi K) + C_\varphi(1 + \kappa_2) \sqrt{2m_{2,2} + 3(m_4 - m_{2,2})_+} \sum_{k=1}^K a_k \right). \end{aligned}$$

Proof of Corollary 4.7. By Lemma 4.6, we have

$$\begin{aligned} &|\langle \mathcal{AT}(\text{Id} - \Pi_v)f, f \rangle_2 + \langle \mathcal{AS}f, f \rangle_2| \leq \\ &\quad \sqrt{3(m_4 - m_{2,2})_+ + 2m_{2,2}(\kappa_1 + \kappa_2)} \|\Pi_v f\|_2 \|(1 - \Pi_v)f\|_2 \\ &\quad + \left(C_\varphi \sqrt{2m_{2,2} + 3(m_4 - m_{2,2})_+} \sum_{k=1}^K \|F_k^\top \nabla_x u_f\|_2 \right. \\ &\quad \left. + m_2(\|\lambda_{\text{ref}} \nabla_x u_f\|_2 + c_\varphi K \|\nabla_x u_f\|_2) \right) \|(1 - \Pi_v)f\|_2. \end{aligned}$$

Now by Assumption 2.4-3 and Lemma 4.4 we have

$$\|F_k^\top \nabla_x u_f\|_2 \leq a_k(\|\nabla_x u_f\|_2 + \|\nabla_x U \nabla_x u_f\|_2) \leq m_2^{-1} a_k(1 + \kappa_2) \|\Pi_v f\|_2.$$

Similarly using Assumption 2.9 we have

$$\|\lambda_{\text{ref}} \nabla_x u_f\|_2 \leq \lambda \|(1 + c_\lambda |\nabla_x U(x)|) \nabla_x u_f\|_2 \leq m_2^{-1} \lambda(1 + c_\lambda \kappa_2) \|\Pi_v f\|_2.$$

Therefore Condition 3.3 follows with

$$\begin{aligned} R_0 &= \sqrt{3(m_4 - m_{2,2})_+ + 2m_{2,2}(\kappa_1 + \kappa_2)} \\ &\quad + m_2^{-1} \left(m_2(\lambda(1 + c_\lambda \kappa_2) + c_\varphi K) + C_\varphi(1 + \kappa_2) \sqrt{2m_{2,2} + 3(m_4 - m_{2,2})_+} \sum_{k=1}^K a_k \right). \quad \square \end{aligned}$$

4.5 Checking Condition 3.4

Recall that here $\Psi(\cdot) = \|\cdot\|_{\text{osc}}^2$. We combine the approaches of [15] and [2]. Note then that the conditions in (3.3) follow immediately from the definition of Ψ and the contractivity in $L^\infty(\mu)$ of the Markov semigroups P_t and e^{-tG} ; the latter corresponds to a diffusion semigroup. Now we check (3.2). Fix some $f \in D(\mathcal{L})$ with $\Psi(f) = \|f\|_{\text{osc}}^2 < \infty$. Without loss of generality, by translating f , we can assume that $\mu(f) = 0$. Hence we have $\gamma_1 := \text{ess}_\mu \inf f \leq 0$ and $\gamma_2 := \text{ess}_\mu \sup f \geq 0$. Since $C = C_b^2(E)$ is a core, we can choose a sequence $\{g_n\}_{n=1}^\infty \subset C$ such that $g_n \rightarrow f$ and $\mathcal{L}g_n \rightarrow \mathcal{L}f$ in $L^2(\mu)$. We take, as in [15], for each $n \in \mathbb{N}$ a monotone increasing function $h_n \in C^\infty(\mathbb{R})$ which satisfies $0 \leq h'_n \leq 1$ and

$$h_n(r) = \begin{cases} r & r \in [\gamma_1, \gamma_2], \\ \gamma_1 - \frac{1}{2n} & r \leq \gamma_1 - \frac{1}{n}, \\ \gamma_2 + \frac{1}{2n} & r \geq \gamma_2 + \frac{1}{n}. \end{cases}$$

Now similarly we set $f_n := h_n(g_n) \in C$ and we have $f_n \rightarrow f$ in $L^2(\mu)$. By construction $\|f_n\|_{\text{osc}} \leq \gamma_2 - \gamma_1 + \frac{1}{n}$ so we have

$$\limsup_{n \rightarrow \infty} \Psi(f_n) = \limsup_{n \rightarrow \infty} \|f_n\|_{\text{osc}}^2 \leq (\gamma_2 - \gamma_1)^2 = \|f\|_{\text{osc}}^2.$$

Additionally,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \langle -\mathcal{L}f_n, f_n \rangle &= \limsup_{n \rightarrow \infty} \langle -\mathcal{S}f_n, f_n \rangle \\ &= \limsup_{n \rightarrow \infty} \sum_{k=0}^K \frac{1}{2} \langle \lambda_k^e f_n, f_n - \mathcal{B}_k f_n \rangle - \langle m_2^{1/2} \lambda_{\text{ref}} \mathcal{R}_v f_n, f_n \rangle \\ &= \limsup_{n \rightarrow \infty} \frac{1}{4} \sum_{k=0}^K \int \lambda_k^e(x, v) \{f_n(x, v) - \mathcal{B}_k f_n(x, v)\}^2 d\mu(x, v) \\ &\quad + m_2^{1/2} \int \lambda_{\text{ref}}(x) \{(\text{Id} - \Pi_v) f_n(x, v)\}^2 d\mu(x, v) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{4} \sum_{k=0}^K \int \lambda_k^e(x, v) \{g_n(x, v) - \mathcal{B}_k g_n(x, v)\}^2 d\mu(x, v) \\ &\quad + m_2^{1/2} \int \lambda_{\text{ref}}(x) \{(\text{Id} - \Pi_v) g_n(x, v)\}^2 d\mu(x, v) \\ &= \limsup_{n \rightarrow \infty} \langle -\mathcal{S}g_n, g_n \rangle \\ &= \langle -\mathcal{L}f, f \rangle. \end{aligned}$$

The third equality follows from the fact that $\mathcal{B}_k$ is symmetric on $L^2(\mu)$, and $\mathcal{B}_k \lambda_k^e(x, v) = \lambda_k^e(x, v)$, as in the proof of Proposition 7 of [3]. The inequality follows from the fact that h_n is 1-Lipschitz; we have that $(h_n(x) - h_n(y))^2 \leq (x - y)^2$ for any $x, y \in \mathbb{R}$. Thus we have verified Condition 3.4.

Since C is a core for $\mathcal{L}$, and $\mathcal{L}$ is densely defined, C is also dense in $L^2(\mu)$. So given some $f \in L^2(\mu)$ with $\Psi(f) < \infty$, this argument also allows us to conclude that there exists a sequence $(f_n) \subset D(\mathcal{L})$ satisfying (3.9) as required in the assumptions of Corollary 3.7.

5 Examples

In this section we apply Theorem A.3 to our running examples and obtain explicit bounds on convergence rate. We further explore the tightness of such bounds on various examples, both theoretically and empirically. Our main finding is that although our

bounds are useful (e.g. we establish the existence of a central limit theorem for a large class of problems; see Example 4.2) and widely applicable, they are not sharp and rather pessimistic. In particular we find that the bounds we obtain for PDMPs do not compare favourably with the corresponding bounds for (reversible) Langevin diffusions for a particular heavy-tailed target density. Informally, this should not be surprising since for PDMPs, condition (3.4) is precisely that required of a Langevin diffusion (there is equivalence in the reversible case) to achieve a particular subgeometric rate of convergence. This condition drives all subsequent developments where the nonreversible nature of the initial process does not seem to play a rôle anymore.

Example 5.1 (Example 2.3 and 4.3 continued). In Example 4.3 we showed the weak Poincaré Inequality holds for the two examples considered and we now show what rate we obtain by applying Theorem 4.1. These obtained rates will immediately allow us to check condition (3.10) to ensure central limit theorems, as discussed in Example 4.2.

- (a) For the case $U(x) = \frac{1}{2}(d+p) \log(1+|x|^2)$ for some $p > 0$, we have from Example 4.3 (a) that α_1 is given by (4.3). Hence by Theorem 4.1, we have the bound, for some $c > 0$,

$$\|P_t f\|_2^2 \leq c t^{-\frac{1}{2\tau}} \left(\|f\|_2^2 + \|f\|_{\text{osc}}^2 \right), \quad \forall t \geq 0, f \in \mathcal{D}(\mathcal{L}).$$

- (b) Consider the case $U(x) = \sigma|x|^\delta$, for $x \in \mathbb{R}^d$ with $|x| \geq M$, for some $\delta, \sigma, M > 0$, by Example 4.3 (b) we have that α_1 is given by (4.4). By Theorem 4.1 we have that (3.6) holds with $\xi(t) = \inf \{r > 0 : c_2 t \geq \alpha_1(r)^2 \log(r^{-1})\}$. Setting $r = \exp(-kt^{\frac{\delta}{8-\delta}})$ we have

$$\begin{aligned} \alpha_1(r)^2 \log(r^{-1}) &= c^2 [1 + \log(1 + r^{-1})]^{8(1-\delta)/\delta} \log(r^{-1}) \\ &= c^2 k \left[1 + \log(1 + \exp(kt^{\frac{\delta}{8-\delta}})) \right]^{8(1-\delta)/\delta} t^{\frac{\delta}{8-\delta}} \\ &\leq c^2 k \left[1 + \exp(-kt^{\frac{\delta}{8-\delta}}) + \log(\exp(kt^{\frac{\delta}{8-\delta}})) \right]^{8(1-\delta)/\delta} t^{\frac{\delta}{8-\delta}} \end{aligned}$$

to obtain the inequality in the last line we use that $\log(1+x) - \log(x) = \log(1+x^{-1}) \leq x^{-1}$ for $x \geq 1$. Now for $t \leq 1$ the required bound is immediate so we shall assume $t \geq 1$ in which case $1 + \exp(-kt^{\frac{\delta}{8-\delta}}) \leq 2 \leq 2kt^{\frac{\delta}{8-\delta}}$ so there exists $C(k, \delta) > 0$ such that

$$\begin{aligned} \alpha_1(r)^2 \log(r^{-1}) &\leq C(k, \delta) \left[t^{\frac{\delta}{8-\delta}} \right]^{8(1-\delta)/\delta} t^{\frac{\delta}{8-\delta}} \\ &= C(k, \delta) t^{\frac{8(1-\delta)}{8-\delta}} t^{\frac{\delta}{8-\delta}} = C(k, \delta) t. \end{aligned}$$

Therefore we have, for some $c > 0$,

$$\|P_t f\|_2^2 \leq \exp(-kt^{\frac{\delta}{8-\delta}}) \left(\|f\|_2^2 + \|f\|_{\text{osc}}^2 \right), \quad \forall t \geq 0, f \in \mathcal{D}(\mathcal{L}).$$

Let us compare the rates we obtain with those found for the reversible and non-reversible Langevin diffusion. In [21] the authors consider the reversible (over-damped) Langevin diffusion,

$$dX_t = -\nabla_x U(X_t) dt + \sqrt{2} dB_t,$$

for heavy-tailed target distributions and prove convergence to equilibrium by using the weak Poincaré Inequality and standard techniques, whereas in [15]

Table 1: Comparison of rates of convergence for scenarios in Example 2.3-4.3.

U in scenario:	(a)	(b)
rate of PDMP	$t^{-\frac{1}{2\tau}}$	$\exp(-kt^{\frac{\delta}{8-7\delta}})$
rate of reversible Langevin	$t^{-\frac{1}{\tau}}$	$\exp(-kt^{\frac{\delta}{4-3\delta}})$
nonreversible Langevin with $V_1 = U$, $V_2 = v^2/2$	$t^{-\frac{1}{2\tau}}$	$\exp(-kt^{\frac{\delta}{8-7\delta}})$
nonreversible Langevin with $V_1 = v^2/2$, $V_2 = U$	$t^{-\frac{1}{\tau}}$	$\exp(-kt^{\frac{\delta}{4-3\delta}})$

they use hypocoercivity to prove convergence to equilibria for the nonreversible (underdamped) Langevin diffusion,

$$\begin{aligned} dX_t &= \nabla_v V_2(V_t) dt, \\ dV_t &= -[\nabla_x V_1(X_t) + \nabla_v V_2(V_t)] dt + \sqrt{2} dB_t. \end{aligned}$$

In this case the diffusion has a unique invariant measure with density which is proportional to $e^{-V_1(x)-V_2(v)}$. In [21, Example 1.4] and [15, Example 1.3] they find the rates of convergence to equilibria which we summarise in the table below. We can see in Table 1 that for the examples we have considered we obtain the same rate of convergence for PDMP to those obtained in [15] for non-reversible Langevin dynamics which have Gaussian velocity. This is a demonstration of the limits of hypocoercivity theory for subgeometric target distributions, since the rate $\xi(t)$ given by Theorem 3.6 depends only on α_1 , α_2 and c_2 but α_1 and α_2 are given by the target measure so are the same for each algorithm.

Example 5.2. Let $d = 1$ and $U(x) = \log(1 + x^2)$, for this choice we can give the rate explicitly. By [21, Theorem 3.2] we have that

$$\alpha_1(r) = \frac{4R_r^2}{\pi^2} \exp(\delta_{R_r}(U)).$$

Here $R_r := \inf \{U(x) - U(y) : x, y \in B_r\}$ and $\delta_R(U) = \sup \{U(x) - U(y) : x, y \in B_R\}$. Now,

$$\pi(B_s^c) = 2 \int_s^\infty \frac{1}{1+x^2} dx = \pi - 2 \arctan(s).$$

Rearranging, we find

$$\begin{aligned} R_r &= \tan\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right), \\ \delta_R(U) &= \log(1 + R^2). \end{aligned}$$

Substituting these expressions into the definition of $\alpha_1(r)$ we have

$$\alpha_1(r) = \frac{4}{\pi^2} \frac{\tan\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right)}{\cos\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right)^2} \sim \frac{32}{\pi^2 r^3}.$$

Now we will run the Zig-Zag sampler with this potential, in which case one finds that

$$c'_2 = \frac{\lambda_{\text{ref}}}{4R_0^2(2 + \sqrt{2})}, \quad R_0 = (1 + \sqrt{1 + c_U/2}) + \lambda_{\text{ref}} + 2\sqrt{2}, \quad c_1 = 2 + \sqrt{2}.$$

For this choice of U we find $c_U = 0.3$. Then

$$\xi(t) := \inf \left\{ r > 0 : c'_2 t \geq \frac{16}{\pi^4} \frac{\tan\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right)^2}{\cos\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right)^4} \log\left(\frac{1}{r}\right) \right\}.$$

To leading order¹ $\xi(t) \sim W(\frac{3c'_2\pi^4}{512}t)^{-1/6}$, here $W(x)$ is the Lambert function defined as the inverse of xe^x . We can also compare this with the numerical performance for the Zig-Zag sampler with canonical switching rates, below is a plot of $\mathbb{E}[f(X_t)]^2$ started with initial condition $X_0 = -5$ and with the velocity V_0 drawn uniformly from $\{1, -1\}$. To run the simulation we generated $N = 10^7$ Zig-Zag samplers and then calculated $N^{-1} \sum_{n=1}^N f(X_t^n)$ to estimate $\mu(f)$. In the figure below we have used $f(x) = \mathbb{I}\{x \geq 5\}$. As we can see in the plot the process appears to converge much faster than the theoretical bound of $\xi(t)$ which is included as a reference. Note that in some of the plots the error is converging to a constant value (around 10^{-4}) this is due to error in running a finite number of particles to estimate the expectation. We have also included on the plot a simulation for the non-reversible Langevin SDE

¹This is found by first approximating to leading order

$$\frac{16 \tan\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right)^2}{\pi^4 \cos\left(\frac{1}{2}\left(\pi - \frac{r}{r+1}\right)\right)^4} \sim \frac{1024}{\pi^4 r^6}$$

then we can solve

$$c'_2 t = \frac{1024}{\pi^4 r^6} \log\left(\frac{1}{r}\right) = \frac{512}{3\pi^4 r^6} \log\left(\frac{1}{r^6}\right).$$

In which case $r^{-6} = W(\frac{3c'_2\pi^4}{512}t)$, therefore $\xi(t) \sim W(\frac{3c'_2\pi^4}{512}t)^{-1/6}$.

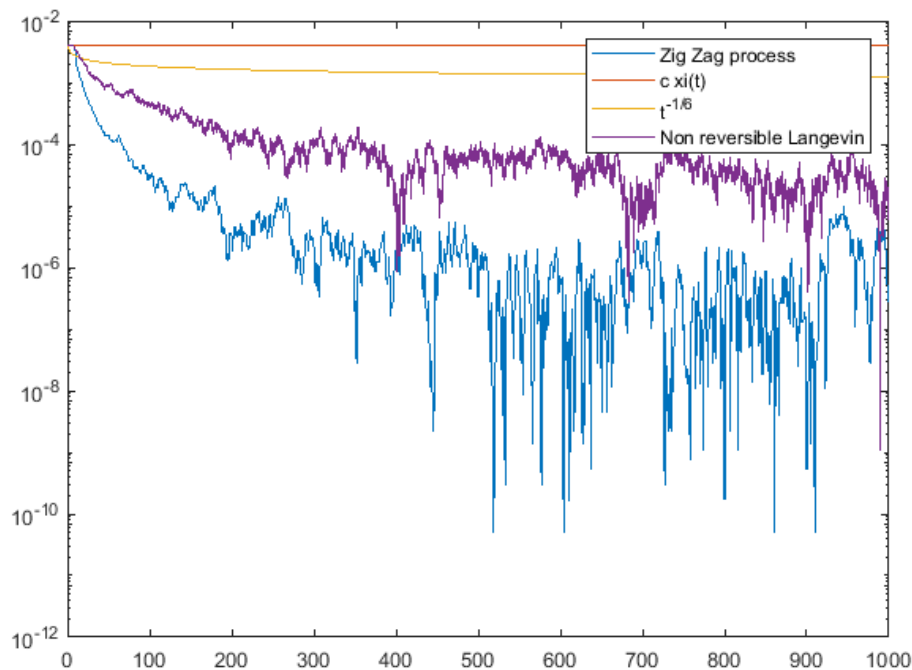


Figure 1: A plot of the error $|P_t f(-5, v) - \mu(f)|^2$ for $f(x, v) = \mathbb{I}\{x \geq 5\}$, in red is a plot of $c\xi(t)$ where c is chosen so that curves are initially equal. The line in blue is the error from the Zig-Zag process whereas the purple is an Euler approximation to the non-reversible Langevin process using a step size of 0.01.

$$dX_t = V_t dt$$

$$dV_t = -\nabla_x U(X_t) dt - V_t dt + \sqrt{2} dB_t,$$

where $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion. To simulate the non-reversible Langevin process we used the Euler-Maruyama scheme with step size 0.01. We see for this example that the Zig-Zag process is converging to zero faster than the non-reversible Langevin SDE.

A Proof of Lemma 4.4

Definition A.1. A (smooth) cut-off function $\vartheta: X \rightarrow [0, 1]$ is such that

1. $\vartheta \in C_c^\infty(X)$,
2. $\vartheta \equiv 0$ on the complement of the unit ball B_1^c ,
3. $\vartheta \equiv 1$ on the ball $B_{1/2}$.

Remark A.2. Define the mollifier $\omega \in C_c^\infty(\mathbb{R}^d)$,

$$\omega(x) = \begin{cases} \exp(-1/(1-|x|^2)) & \text{if } |x| < 1 \\ 0 & \text{otherwise,} \end{cases}$$

then $\vartheta(x) = 2^{-d} \int \mathbb{I}\{|y-x| \leq 1\} \omega(2y) dy$ is a cut-off function. We now show that if $|x_2| \geq |x_1|$ then $\vartheta(x_2) \leq \vartheta(x_1)$. First notice that since ω is invariant by rotation, so is ϑ , and we can focus on the following scenario. Let $x \in B_{1/2}^c$ and $\lambda > 1$ then $\{y \in B_{1/2}: |y-\lambda x| \leq 1\} \subset \{y \in B_{1/2}: |y-x| \leq 1\}$ and therefore $\vartheta(\lambda x) \leq \vartheta(x)$, from which we conclude. The inclusion above is justified by the fact that for $(x, y) \in B_{1/2}^c \times B_{1/2}$ $\lambda \mapsto |\lambda x - y|^2$ is non-decreasing since, with $n(x) = x/|2x|$,

$$\begin{aligned} |\lambda x - y|^2 &= |\lambda x - n(x)|^2 + |n(x) - y|^2 + 2\langle \lambda x - n(x), n(x) - y \rangle \\ &= (\lambda|x| - 1/2)^2 + |n(x) - y|^2 + 2(\lambda|x| - 1/2)\langle n(x), n(x) - y \rangle, \end{aligned}$$

and $\langle n(x), n(x) - y \rangle = \langle n(x), n(x) \rangle - \langle n(x), y \rangle \geq 0$ as $\langle n(x), y \rangle \leq |n(x)| |y| \leq |n(x)|^2$.

A.1 Expression for ∇^*

Lemma A.3. Let Assumption 2.2 hold. Then with $(\nabla_x, D(\nabla_x))$ with $\nabla_x: D(\nabla_x) \rightarrow L^2(\pi)^d$ where $D(\nabla_x) \subset L^2(\pi)$,

1. for $F \in H^1(X, \mathbb{R}^d)$,

$$\nabla_x^* F = -\operatorname{div}_x F + \nabla_x U^\top F,$$

2. for $f \in H^2(X)$,

$$\nabla_x^* \nabla_x f = -\Delta_x f + \nabla_x U^\top \nabla_x f.$$

Proof. We follow the ideas of [16]. Consider first the case where $d = 1$. Let $f \in H^1(X)$ and $\vartheta: X \rightarrow [0, 1]$ a cut-off function, then we define for $n \in \mathbb{N}_*$ and $x \in X$ $f_n(x) := f(x)\vartheta(x/n)$. Note that

$$\partial f_n(x) = \vartheta(x/n) \partial f(x) + \frac{1}{n} f(x) \partial \vartheta(x/n).$$

Let $f, g \in H^1(X)$. Then for any $n \in \mathbb{N}_*$, using integration by parts and noting that $\vartheta(\cdot/n) \equiv 0$ on B_n^c , we have

$$\begin{aligned} & \int f_n(x) \pi(x) \partial g_n(x) dx \\ &= -Z^{-1} \int g_n(x) [\partial f_n(x) - f_n(x) \partial U(x)] \exp(-U(x)) dx \\ &= - \int \vartheta^2(x/n) g(x) [\partial f(x) - f(x) \partial U(x) + f(x) \frac{\partial \vartheta(x/n)}{n}] \pi(dx). \end{aligned}$$

Letting $n \rightarrow \infty$ and applying the dominated convergence theorem leads to the desired result; the required integrability follows from applying the Cauchy–Schwarz inequality several times, Assumption 2.2–3 and that $f(x)\partial U(x) \in L^2(\pi)$. Note if ∂U is bounded then it is immediate to see $f(x)\partial U(x) \in L^2(\pi)$ on the other hand if ∂U is not bounded then Assumption 2.2- 3b holds and we use (4.9) with $\varphi = f$ to show $f(x)\partial U(x) \in L^2(\pi)$, indeed this gives

$$\|f\partial U\|_2^2 \leq 16\|\partial f\|_2^2 + 4C_U d^{1+\omega} \|f\|_2^2.$$

When $d > 1$, simply notice that for $f \in H^1(X)$ and $G = (g_1, \dots, g_d) \in H^1(X, \mathbb{R}^d)$, $\langle \nabla_x f, G \rangle_\pi = \sum_{i=1}^d \langle \partial_i f, g_i \rangle_\pi$ and apply the result above. The second statement is immediate. $\square$

Corollary A.4. *Let $\mathcal{L}$ be as in (1.3) and assume Assumption 2.2 holds. Then for $f \in C_b^2(E)$ and $(x, v) \in E$,*

$$\mathcal{L}^* f(x, v) = -v^\top \nabla_x f(x, v) + \sum_{k=1}^K \lambda(x, -v) (\mathcal{B}_k - \text{Id}) f(x, v) + m_2^{1/2} \lambda_{\text{ref}}(x) \mathcal{R}_v f(x).$$

Proof. From Lemma A.3, for $F \in H^1(X, \mathbb{R}^d)$ and $g \in H^1(\pi)$ we have $\langle \nabla_x g, F \rangle_\pi = \langle g, -\text{div}_x F + \nabla_x U^\top F \rangle_\pi$. We need to check that this is applicable in the case $x \mapsto F_v(x) := v \cdot f(x, v)$ for fixed $v \in V$, provided $F_v \in H^1(E, \mathbb{R}^d)$, which is clearly true here. This is true for $v = 0$. For $v \neq 0$, noting that $\nabla_x(v \cdot f) = v(\nabla_x f)^\top$ we deduce that $F_v \in H^1(X, \mathbb{R}^d)$. The rest then follows from a calculation identical to that in the proof of [2, Proposition 7]. $\square$

A.2 Bound on $\|\nabla_x^2 u_f\|_2$ in Lemma 4.4

Fix $f \in C_b^2(E)$, and let $u = u_f$ be the solution of the Poisson equation (4.5). Since $C_c^\infty(X)$ is dense in $H^2(X)$ (see [16, Lemma 3.1]) we may assume that f is smooth with compact support. By rescaling f we may assume that $m_2 = 1$. Differentiating (4.5) we have for $i \in \{1, \dots, d\}$,

$$\partial_i u + \nabla_x^* \nabla_x \partial_i u + (\nabla_x \partial_i U)^\top \nabla_x u = \partial_i f, \quad (\text{A.1})$$

since from Lemma 4.4 $u \in C^3(X)$ as $U \in C^{2+\alpha}(X)$. Let $\vartheta: X \rightarrow [0, 1]$ be a cut-off function. Define $\vartheta_n(x) := \vartheta(x/n)$ for $(n, x) \in \mathbb{N}_* \times X$ and note that $\sup_{x \in B_n} |\nabla_x \vartheta_n(x)| \leq C_1 n^{-1}$ with $C_1 := \sup_{x \in B_1} |\nabla_x \vartheta(x)|$. Now let $n \geq n_0$ for $n_0 \in \mathbb{N}_*$ such that $\vartheta_{n_0} \equiv 1$ on the support of f . Throughout this section, we will write $\langle \cdot, \cdot \rangle_\pi$ as shorthand for $\langle \cdot, \cdot \rangle_{L^2(\pi)}$. Now multiply (A.1) by $\vartheta_n^2 \partial_i u$, sum over $i \in \{1, \dots, d\}$ and integrate to get

$$\begin{aligned} \langle \vartheta_n^2 \nabla_x u, \nabla_x u \rangle_\pi + \sum_{i=1}^d \left(\langle \vartheta_n^2 \partial_i u, \nabla_x^* \nabla_x \partial_i u \rangle_\pi + \langle \vartheta_n^2 \partial_i u, (\nabla_x \partial_i U)^\top \nabla_x u \rangle_\pi \right) \\ = \langle \vartheta_n^2 \nabla_x u, \nabla_x f \rangle_\pi. \end{aligned} \quad (\text{A.2})$$

Now since $\vartheta_n \equiv 1$ on the support of f and from Lemma A.3, we deduce $\langle \vartheta_n^2 \nabla_x u, \nabla_x f \rangle_\pi = \langle \nabla_x^* \nabla_x u, f \rangle_\pi$ and therefore

$$\begin{aligned} \|\vartheta_n \nabla_x u\|_2^2 + \sum_{i=1}^d \left(\langle \vartheta_n^2 \partial_i u, \nabla_x^* \nabla_x \partial_i u \rangle_\pi + \langle \vartheta_n^2 \partial_i u, (\nabla_x \partial_i U)^\top \nabla_x u \rangle_\pi \right) \\ = \langle \nabla_x^* \nabla_x u, f \rangle_\pi. \end{aligned} \quad (\text{A.3})$$

Now consider the second term on the left hand side of (A.3),

$$\begin{aligned} \sum_{i=1}^d \langle \vartheta_n^2 \partial_i u, \nabla_x^* \nabla_x \partial_i u \rangle_\pi &= \sum_{i=1}^d \langle \nabla_x (\vartheta_n^2 \partial_i u), \nabla_x \partial_i u \rangle_\pi \\ &= \|\vartheta_n \nabla_x^2 u\|_2^2 + 2 \sum_{i,j=1}^d \int \vartheta_n \partial_j \vartheta_n \partial_i u \partial_{ij} u \, d\pi. \end{aligned}$$

Therefore (A.3) becomes

$$\|\vartheta_n \nabla_x u\|_2^2 + \|\vartheta_n \nabla_x^2 u\|_2^2 = I_1 + I_2 + I_3,$$

with

$$\begin{aligned} I_1 &:= \langle \nabla_x^* \nabla_x u, f \rangle_\pi, \quad I_2 := -2 \langle \vartheta_n \nabla_x \vartheta_n, \nabla_x^2 u \nabla_x u \rangle_\pi \\ I_3 &:= -\langle \vartheta_n \nabla_x u, \nabla_x^2 U (\vartheta_n \nabla_x u) \rangle_\pi. \end{aligned} \quad (\text{A.4})$$

It remains to estimate each of the terms I_1, I_2, I_3 . For the first term I_1 , recall that $\nabla_x^* \nabla_x u = f - u$ and $\|u\|_2 \leq \|f\|_2$ by (4.6) so we have

$$|\langle \nabla_x^* \nabla_x u, f \rangle_\pi| \leq \|\nabla_x^* \nabla_x u\|_2 \|f\|_2 \leq \|f - u\|_2 \|f\|_2 \leq 2\|f\|_2^2. \quad (\text{A.5})$$

For the third term I_3 we use that $\nabla_x^2 U \succeq -c_U \text{Id}$ by Assumption 2.2 for some $c_U \geq 0$, $\vartheta_n \leq 1$ and that $\|\nabla_x u\|_2 \leq \|f\|_2$ by (4.6),

$$|\langle \vartheta_n \nabla_x u, \nabla_x^2 U (\vartheta_n \nabla_x u) \rangle_\pi| \leq c_U \|\vartheta_n \nabla_x u\|_2^2 \leq c_U \|f\|_2^2. \quad (\text{A.6})$$

For the second term I_2 , we use that $|\nabla_x \vartheta_n(x)| \leq C_1 n^{-1}$ then

$$|\langle \vartheta_n \nabla_x \vartheta_n, \nabla_x^2 u \nabla_x u \rangle_\pi| \leq C_1 n^{-1} \int \vartheta_n |\nabla_x^2 u| |\nabla_x u| \, d\pi.$$

Now Young's inequality for any $\varepsilon > 0$ and (4.6) yield

$$|\langle \vartheta_n \nabla_x \vartheta_n, \nabla_x^2 u \nabla_x u \rangle_\pi| \leq \varepsilon \int \vartheta_n^2 |\nabla_x^2 u|^2 \, d\pi + \frac{C_1^2 n^{-2}}{4\varepsilon} \int |\nabla_x u|^2 \, d\pi \quad (\text{A.7})$$

$$\leq \varepsilon \|\vartheta_n \nabla_x^2 u\|_2^2 + \frac{C_1^2 n^{-2}}{4\varepsilon} \|f\|_2^2. \quad (\text{A.8})$$

Combining (A.4), (A.5), (A.6), (A.7) we have

$$\|\vartheta_n \nabla_x u\|_\pi^2 + (1 - \varepsilon) \|\vartheta_n \nabla_x^2 u\|_2^2 \leq \left(2 + c_U + \frac{C_1^2 n^{-2}}{4\varepsilon} \right) \|f\|_2^2.$$

Finally taking $\varepsilon = 1/4$ we obtain

$$\|\vartheta_n \nabla_x^2 u\|_2^2 \leq 2(2 + c_U + C_1^2 n^{-2}) \|f\|_2^2.$$

The result follows by choosing ϑ as in Remark A.2, letting $n \rightarrow \infty$ and invoking the monotone convergence theorem.

References

- [1] Franz Achleitner, Anton Arnold, and Beatrice Signorello. On Optimal Decay Estimates for ODEs and PDEs with Modal Decomposition. In Giambattista Giacomini, Stefano Olla, Ellen Saada, Herbert Spohn, and Gabriel Stoltz, editors, *Stochastic Dynamics Out of Equilibrium*, volume 282 of *PROMS*, pages 241–264. Springer New York LLC, 2019. MR-3986066

- [2] Christophe Andrieu, Alain Durmus, Nikolas Nüsken, and Julien Roussel. Hypocoercivity of Piecewise Deterministic Markov Process-Monte Carlo. *To appear in Annals of Applied Probability* <https://arxiv.org/abs/1808.08592>, 2018.
- [3] Christophe Andrieu and Samuel Livingstone. Peskun-Tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario. *To appear in Annals of Statistics* <https://arxiv.org/abs/1906.06197>, 2019.
- [4] E. Bernard, M. Fathi, A. Levitt, and G. Stoltz. Hypocoercivity with Schur complements. *arXiv preprint* <https://arxiv.org/abs/2003.00726>, 2020.
- [5] Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. *Annals of Statistics*, 47(3):1288–1320, 2019. MR-3911113
- [6] Joris Bierkens, Sebastiano Grazi, Kengo Kamatani, and Gareth Roberts. The Boomerang Sampler. *Proceedings of the 37th International Conference on Machine Learning*, 119:908–918 2020.
- [7] Joris Bierkens and Sjoerd M. Verduyn Lunel. Spectral analysis of the zigzag process. *arXiv preprint* <https://arxiv.org/abs/1905.01691>, 2019.
- [8] Nawaf Bou-Rabee and Jesús María Sanz-Serna. Randomized Hamiltonian Monte Carlo. *Annals of Applied Probability*, 27(4):2159–2194, 2017. MR-3693523
- [9] Alexandre Bouchard-Côté, Sebastian J. Vollmer, and Arnaud Doucet. The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method. *Journal of the American Statistical Association*, 113(522):855–867, 2018. MR-3832232
- [10] George Deligiannidis, Daniel Paulin, Alexandre Bouchard-Côté, and Arnaud Doucet. Randomized Hamiltonian Monte Carlo as Scaling Limit of the Bouncy Particle Sampler and Dimension-Free Convergence Rates. *To appear in Annals of Applied Probability* <https://arxiv.org/abs/1808.04299>, 2018. MR-3911112
- [11] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6):6012–6019, 2015. MR-3324910
- [12] Alain Durmus, Arnaud Guillin, and Pierre Monmarché. Piecewise Deterministic Markov Processes and their invariant measure. *arXiv preprint* <https://arxiv.org/abs/1807.05421>, 2018.
- [13] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009. MR-0838085
- [14] Martin Grothaus and Patrik Stilgenbauer. Hypocoercivity for Kolmogorov backward evolution equations and applications. *Journal of Functional Analysis*, 267(10):3515–3556, 2014. MR-3266239
- [15] Martin Grothaus and Feng-Yu Wang. Weak Poincaré inequalities for convergence rate of degenerate diffusion processes. *Annals of Probability*, 47(5):2930–2952, 2019. MR-4021241
- [16] Luca Lorenzi and Alessandra Lunardi. Elliptic operators with unbounded diffusion coefficients in L^2 spaces with respect to invariant measures. *Journal of Evolution Equations*, 6:691–709, 2006. MR-2267704
- [17] Jianfeng Lu and Lihan Wang. On explicit L^2 -convergence rate estimate for piecewise deterministic Markov processes. *arXiv preprint*, <https://arxiv.org/abs/2007.14927> 2020.
- [18] Pierre Monmarché. L^2 hypocoercivity, deviation bounds, hitting times and Lyapunov functions. *arXiv preprint* <https://arxiv.org/abs/1911.01748>, 2019.
- [19] Gert K Pedersen. *Analysis Now*, volume 118 of *Graduate texts in mathematics*. Springer-Verlag, 1989. MR-0971256
- [20] Michael Reed and Barry Simon. *Methods of mathematical physics I: Functional Analysis*, volume 261. 1956. MR-0493419
- [21] Michael Röckner and Feng-Yu Wang. Weak Poincaré Inequalities and L^2 Convergence Rates of Markov Semigroups. *Journal of Functional Analysis*, 185:564–603, 2001. MR-1856277
- [22] Bálint Tóth. Comment on a theorem of M. Maxwell and M. Woodroffe. *Electronic Communications in Probability*, 18(13):1–4, 2013. MR-3033596

Acknowledgments. We are very grateful to the anonymous reviewer, whose helpful remarks have improved the paper. We would like to thank the Heilbronn Institute for Mathematical Research Research for funding the Hypocoercivity Workshop held at the University of Bristol, March 2020, which initiated this research. We would like to thank Joris Bierkens, Anthony Lee and Sam Power for interesting discussions related to this work.