

# NON-REVERSIBLE LIFTS OF REVERSIBLE DIFFUSIONS AND THEIR CONVERGENCE TO EQUILIBRIUM

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

**Francis Cameron Lörler**

aus

Erlangen

Bonn, 2026



Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät  
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter und Betreuer: Prof. Dr. Andreas Eberle  
Gutachter: Prof. Dr. Gabriel Stoltz

Tag der Promotion: 15. Juni 2026  
Erscheinungsjahr: 2026



# Summary

Long-time convergence of Markov processes towards their stationary equilibrium distribution is of fundamental importance in a variety of applications. For one, efficient sampling of intractable probability distributions arising, for instance, in Bayesian statistics or computational physics, typically relies on Markov chain Monte Carlo methods that use samples of a Markov process at large time as surrogates for the target measure. Furthermore, many phenomena arising in the modelling of real-world systems in physics or biology as well as in finance or deep learning are described as stochastic processes, and often valuable insight can be gained by understanding their long-time behaviour. Such models frequently lead to dynamics that are not reversible in time and are subject to degenerate noise acting only on a subset of the variables of the model. In Markov chain Monte Carlo (MCMC), one observes a desirable accelerated convergence behaviour due to the interplay of non-reversibility and degenerate noise, which has led to increasing popularity of algorithms based on such processes.

This thesis is concerned with convergence to stationarity of such non-reversible Markov processes with degenerate noise, known as hypocoercivity. From an analytical viewpoint, these features inhibit the use of many standard tools and thus present challenging problems. We address this issue by developing a framework that allows to derive rates of convergence towards the invariant measure in  $L^2$  for a large class of such dynamics, which we term second-order lifts of reversible diffusions. These are motivated by the concept of lifts of Markov chains in discrete time that was introduced to understand acceleration due to non-reversibility in Markov chain Monte Carlo methods. The six works forming the basis for this thesis are included in the appendix as Chapters A–F.

The development of the hypocoercivity framework based on second-order lifts is the main content of Chapters A, B and C. The concept of second-order lifts of reversible diffusions is motivated and introduced in Chapter A, and shown to include several processes commonly used in the MCMC literature. It provides a structural relationship between many non-reversible dynamics with degenerate noise and simple, reversible diffusions. We show how this relationship can be exploited to derive a simple lower bound on the  $L^2$ -relaxation time of arbitrary second-order lifts, and give a first demonstration of how it can be used to obtain quantitative hypocoercivity results for randomised Hamiltonian Monte Carlo, and thereby a matching upper bound on the relaxation time. In Chapter B, we extend this framework to processes on Riemannian manifolds with boundary, and prove the quantitative divergence lemma. The latter is a statement purely on the generator of the underlying reversible diffusion, and is a crucial ingredient in the proof of upper bounds on the relaxation time using the approach of second-order lifts. We conclude in Chapter C by turning these ideas developed for randomised Hamiltonian Monte Carlo and Langevin dynamics into a general framework. It allows to derive quantitative

bounds on rates of convergence to stationarity for second-order lifts by proving a flow Poincaré inequality, a time-averaged Poincaré inequality along trajectories of the associated transition semigroup, under assumptions that are simple to verify in practice. We demonstrate this result by applying it to piecewise-deterministic Markov processes used in sampling applications, as well as run-and-tumble particles with jamming, a model from non-equilibrium statistical mechanics. [Chapter D](#) compares our approach to hypocoercivity with the highly influential one developed by Dolbeault, Mouhot and Schmeiser, uncovering several structural similarities. The approach based on second-order lifts, however, often yields better convergence rates. For Langevin dynamics, for instance, rates of convergence that are sharp in terms of the order of the dependence on model parameters eluded previous approaches, yet can be obtained using the framework of second-order lifts.

[Chapters E](#) and [F](#) focus on two classes of self-interacting processes, namely self-repellent random walks and self-repelling diffusions. These are non-Markovian since they interact with their history through their occupation time measure, yet considering their evolution jointly with their environment or local time profile yields Markov processes. We show that these are second-order lifts of a suitable Ornstein-Uhlenbeck process whose invariant probability measure corresponds to that of the environment process. Self-repellent random walks are closely related to Event Chain Monte Carlo methods—a special class of MCMC methods—applied to a chain of harmonically interacting oscillators, while self-repelling diffusions are related to Brownian polymers, models for polymer growth. For both classes, we show that convergence rates to stationarity of the joint processes can be obtained using the framework of second-order lifts and the results of [Chapter C](#).

We begin with some background on convergence to equilibrium of Markov processes and some examples motivating acceleration through non-reversibility in [Chapter 1](#). The main result of this thesis, the hypocoercivity framework based on second-order lifts and the flow Poincaré inequality, is presented in [Chapter 2](#), focussing on the convergence of Langevin dynamics to highlight the main ideas. Finally, [Chapter 3](#) presents the contributions of the individual projects as well as some open questions for further research.

# Acknowledgements

First and foremost, I would like to thank Andreas Eberle for his support and guidance throughout my PhD and for introducing me to the study of convergence of Markov processes. I benefited immensely from the countless discussions at the blackboard and over lunch and from his expertise and intuition in the field. I am also very thankful to him for introducing me to his academic network and providing me with the opportunity to present our research at conferences and workshops.

Merci à Arnaud Guillin et Manon Michel for hosting me during numerous highly enjoyable visits to Clermont-Ferrand with a very collaborative research environment (and for introducing me to some highlights of French cuisine). In the same spirit, I would like to thank Leo Hahn for our productive and fruitful collaboration, for hosting me in Neuchâtel, and for proofreading parts of this thesis. I would also like to thank my co-authors Giovanni Brigati and Lihan Wang for our enriching collaboration. Thank you to Herbert Koch and Hartmut Monien for agreeing to be part of my doctoral committee, and special thanks to Gabriel Stoltz for being a reviewer of this thesis.

I would like to thank all members with whom I have shared my time in the probability theory group for providing a friendly and engaging environment, with my warmest thanks going to Mei-Ling Wang for her relentless dedication to solving any non-mathematical problem arising in our group. I have had the pleasure of sharing an office with Katharina Schuh, Saeda Mareello and Stefan Oberdörster, all of whom I would like to thank for the very pleasant time spent together. Special thanks go to Katharina Schuh and Stefan Oberdörster for the many fruitful (mathematical and non-mathematical) discussions we have had.

I am very grateful to the Hausdorff Center for Mathematics, the Bonn International Graduate School for Mathematics, and the Collaborative Research Center 1720 for funding my PhD studies at the Institute for Applied Mathematics. They provided me with the opportunity to pursue research here and to travel to numerous conferences and summer schools, of which I have profited greatly.

Finally, I would like to thank my friends and family for supporting me through the ups and downs of the past years. I cannot express my gratitude to Paula and Felix enough.

*Thank you.*



# Contents

|                                                                                                                             |            |
|-----------------------------------------------------------------------------------------------------------------------------|------------|
| <b>1. Introduction</b>                                                                                                      | <b>1</b>   |
| 1.1. Markov chain Monte Carlo . . . . .                                                                                     | 1          |
| 1.2. Convergence to equilibrium of Markov processes . . . . .                                                               | 3          |
| 1.3. Acceleration through non-reversibility . . . . .                                                                       | 8          |
| <b>2. Second-order lifts and hypocoercivity</b>                                                                             | <b>21</b>  |
| 2.1. Second-order lifts of reversible diffusions . . . . .                                                                  | 21         |
| 2.2. Hypocoercivity and flow Poincaré inequality . . . . .                                                                  | 28         |
| <b>3. Conclusion</b>                                                                                                        | <b>37</b>  |
| 3.1. Outline of project A . . . . .                                                                                         | 37         |
| 3.2. Outline of project B . . . . .                                                                                         | 38         |
| 3.3. Outline of project C . . . . .                                                                                         | 40         |
| 3.4. Outline of project D . . . . .                                                                                         | 42         |
| 3.5. Outline of project E . . . . .                                                                                         | 44         |
| 3.6. Outline of project F . . . . .                                                                                         | 46         |
| 3.7. Outlook and open questions . . . . .                                                                                   | 48         |
| <b>A. Non-reversible lifts of reversible diffusion processes and relaxation times</b>                                       | <b>51</b>  |
| A.1. Introduction . . . . .                                                                                                 | 51         |
| A.2. Lifts of reversible diffusions . . . . .                                                                               | 54         |
| A.3. Relaxation times of non-reversible Markov processes . . . . .                                                          | 61         |
| A.4. Lower bounds for relaxation times of lifts . . . . .                                                                   | 65         |
| A.5. Upper bounds for relaxation times of lifts . . . . .                                                                   | 67         |
| <b>B. Space-time divergence lemmas and optimal non-reversible lifts of diffusions on Riemannian manifolds with boundary</b> | <b>79</b>  |
| B.1. Introduction . . . . .                                                                                                 | 79         |
| B.2. Preliminaries and the generalised Reilly formula . . . . .                                                             | 83         |
| B.3. The space-time divergence lemma . . . . .                                                                              | 86         |
| B.4. Optimal lifts on Riemannian manifolds with boundary . . . . .                                                          | 98         |
| B.5. Appendix . . . . .                                                                                                     | 109        |
| <b>C. Convergence of non-reversible Markov processes via lifting and flow Poincaré inequality</b>                           | <b>115</b> |
| C.1. Introduction . . . . .                                                                                                 | 115        |
| C.2. Second-order lifts . . . . .                                                                                           | 117        |
| C.3. Convergence to equilibrium . . . . .                                                                                   | 124        |

|                                                                                        |            |
|----------------------------------------------------------------------------------------|------------|
| C.4. Convergence of the jamming RTP process . . . . .                                  | 138        |
| C.5. PDMPs for sampling . . . . .                                                      | 146        |
| <b>D. Hypocoercivity meets lifts</b>                                                   | <b>155</b> |
| D.1. Introduction . . . . .                                                            | 155        |
| D.2. The general convergence proof . . . . .                                           | 166        |
| D.3. Example: Hypocoercivity of adaptive Langevin dynamics . . . . .                   | 168        |
| D.4. Lower bounds for the generalised Langevin equation . . . . .                      | 173        |
| <b>E. Convergence rates of self-repellent random walks and event chain Monte Carlo</b> | <b>177</b> |
| E.1. Introduction . . . . .                                                            | 177        |
| E.2. Application to Markov Chain Monte Carlo methods . . . . .                         | 186        |
| E.3. Convergence to equilibrium . . . . .                                              | 189        |
| <b>F. Convergence rates of self-repelling diffusions on Riemannian manifolds</b>       | <b>193</b> |
| F.1. Introduction . . . . .                                                            | 193        |
| F.2. Self-interacting diffusions on Riemannian manifolds . . . . .                     | 195        |
| F.3. Relaxation times of self-interacting diffusions . . . . .                         | 198        |
| F.4. Convergence to equilibrium . . . . .                                              | 205        |
| <b>Bibliography</b>                                                                    | <b>209</b> |

# Chapter 1.

## Introduction

Many phenomena arising in nature and applications, ranging from physical or chemical systems or biological dynamics to financial models, are too complex to be modelled exactly and exhibit unavoidable uncertainty or even intrinsic random fluctuations. Therefore, incorporating randomness into their description is often necessary, and such phenomena are conveniently modelled by stochastic processes. A central question is understanding the long-time behaviour of such dynamics, which is often described by an equilibrium distribution that is left invariant by the process. Among other questions such as existence, uniqueness and regularity of such a steady state, the rate at which the law of the process at finite time converges to the limiting distribution is of crucial importance. This question is also of fundamental importance in Markov chain Monte Carlo (MCMC), a cornerstone of modern statistics. In MCMC, one constructs a Markov process whose stationary distribution is a given target and uses samples from this process at sufficiently large time as surrogates for samples from the target measure. In this context, the rate of convergence of the law of the process at finite time towards the target distribution determines the efficiency of the scheme at approximating the target measure.

We begin in [Section 1.1](#) by briefly introducing Markov chain Monte Carlo, which serves as one of the main motivations for the phenomena and processes we consider. We continue in [Section 1.2](#) by introducing the main distances on the set of probability measures that we will use to quantify convergence to equilibrium, focusing on the  $\chi^2$ -divergence and the role of the Poincaré inequality in the case of time-reversible Markov processes. In the subsequent [Section 1.3](#), we illustrate the phenomenon of accelerated convergence due to non-reversibility and degenerate noise that has motivated a variety of popular non-reversible MCMC methods and has attracted substantial attention from the points of view of analysis, probability and statistical physics. The systematic study and exploration of this phenomenon is the goal of this thesis, with the main results presented in [Chapters 2](#) and [3](#).

### 1.1. Markov chain Monte Carlo

In many applications one is interested in computing integrals of the form

$$\mu(f) = \int_S f \, d\mu,$$

where  $\mu \in \mathcal{P}(S)$  is a probability measure on a state space  $S$  and  $f: S \rightarrow \mathbb{R}$  is a function that is integrable with respect to  $\mu$ . For instance,  $\mu$  may be a representation of a large dataset in Bayesian statistics or machine learning, or describe a distribution on a large or high-dimensional configuration space in computational physics or chemistry. Applications are also commonplace throughout other sciences such as mathematical biology, medicine, or economics.

A closely related problem is that of creating samples from the distribution  $\mu$ , since, by the law of large numbers, the expectation  $\mu(f)$  can be approximated by the sample means of a sequence of i.i.d. random variables with distribution  $\mu$ . However, typically, the probability distribution  $\mu$  is complicated due to multimodality or anisotropy, the state space  $S$  is high-dimensional, or the density of  $\mu$  is only available up to a normalising constant, so that neither direct computation of the integrals  $\mu(f)$  nor direct sampling of  $\mu$  is feasible. For this reason, one resorts to Markov chain Monte Carlo methods that construct a time-homogeneous Markov chain  $((X_n)_{n \in \mathbb{N}}, (\mathbb{P}_x)_{x \in S})$  with state space  $S$  that has the target measure  $\mu$  as invariant measure. Assuming ergodicity, the law of the Markov chain for large  $n$  approximates  $\mu$ , and ergodic averages can be used to approximate  $\mu(f)$  by the ergodic theorem, see e.g. [195, 235]. This idea originates in the 1953 work of Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [194] on interacting molecules and hard sphere systems. It was later popularised by the work of Hastings [130], see also [57, 221] for a more complete history and modern introduction to the field.

While practical implementations of such schemes require discrete-time formulations, many Markov chains used in applications arise as discretisations of continuous-time dynamics through discretisations of stochastic differential equations (SDEs) or piecewise-deterministic Markov processes (PDMPs). Such continuous-time Markov processes are often more convenient to work with, as they are more amenable to analytic methods and avoid discrete-time issues such as periodicity. For this reason, we will focus on continuous-time Markov processes  $((X_t)_{t \geq 0}, (\mathbb{P}_x)_{x \in S})$  and their convergence behaviour, and only mention discrete-time counterparts obtained through discretisation. This is further justified by the fact that, at leading order in the step size  $h$ , the variance of time-discretisations typically agrees with that of the continuous-time process [178]. Again assuming ergodicity, ergodic averages

$$A_{b,t}(f) = \frac{1}{t} \int_b^{b+t} f(X_s) \, ds$$

can then be used to approximate  $\mu(f)$ , where  $b \in (0, \infty)$  is a burn-in time and  $t \in (0, \infty)$ . Indeed, by the ergodic theorem,

$$A_{b,t}(f) \xrightarrow{t \rightarrow \infty} \mu(f) \quad \mathbb{P}_x\text{-almost surely}$$

for  $\mu$ -almost every  $x \in S$  [235]. The speed of convergence of  $(X_t)_{t \geq 0}$  to stationarity is of crucial importance, since it, together with the computational complexity of simulation, determines the efficiency of the MCMC scheme.

## 1.2. Convergence to equilibrium of Markov processes

We will study time-homogeneous right-continuous Markov processes  $((X_t)_{t \geq 0}, (\mathbb{P}_x)_{x \in S})$  with a Polish state space  $S$  endowed with its Borel  $\sigma$ -algebra  $\mathcal{B}(S)$ , defined on a measurable space  $(\Omega, \mathcal{A})$ . We will usually assume the process to have an invariant probability measure  $\mu \in \mathcal{P}(S)$ , that is, the transition function  $(p_t)_{t \geq 0}$  describing the time evolution of the law of  $X_t$  satisfies

$$(\mu p_t)(A) = \int_S p_t(x, A) \mu(dx) = \mu(A) \quad \text{for all } t \geq 0 \text{ and } A \in \mathcal{B}(S).$$

The dual action of  $p_t$  on observables in  $L^2(\mu)$  is the transition semigroup  $(P_t)_{t \geq 0}$  on  $L^2(\mu)$  given by

$$P_t f(x) = \mathbb{E}_x[f(X_t)] = \int_S f(y) p_t(x, dy) \quad \text{for all } f \in L^2(\mu) \text{ and } t \geq 0.$$

As motivated above, a question central to this thesis is that of the speed of convergence of the law of the process at finite time to its invariant measure.

### 1.2.1. $f$ -divergences

In order to quantify the convergence of the law  $\nu p_t$  of the process at time  $t$  with initial distribution  $\nu$  to equilibrium  $\mu$ , it is necessary to equip the space  $\mathcal{P}(S)$  of probability measures on  $(S, \mathcal{B}(S))$  with a concept of distance. A popular class of such functions is given by  $f$ -divergences for convex functions  $f: [0, \infty) \rightarrow \mathbb{R} \cup \{\infty\}$  with  $f(1) = 0$ . For two probability measures  $\nu, \mu \in \mathcal{P}(S)$ , the  $f$ -divergence of  $\nu$  with respect to  $\mu$  is

$$D_f(\nu | \mu) = \int_S f(\rho) d\mu + \nu(\{\rho = \infty\}) \lim_{t \rightarrow \infty} \frac{f(t)}{t}, \quad (1.1)$$

where  $\rho = \frac{d\nu}{d\mu}$  is the Lebesgue density and the limit exists due to convexity. Even though  $f$ -divergences are not symmetric in general, they are always non-negative and satisfy  $D_f(\nu | \mu) = 0$  if and only if  $\nu = \mu$  as soon as  $f$  is strictly convex at 1, so that they are suitable distance functions. This class of distance functions was introduced by Rényi [219] as a generalisation of the Kullback-Leibler divergence [159], and later developed further by Csiszár [73].

Several classical distance functions on  $\mathcal{P}(S)$  can be recovered by choosing  $f$  in specific ways. Notably,  $f(t) = |t - 1|/2$  yields the total variation distance,  $f(t) = t \log t$  the relative entropy or Kullback-Leibler divergence, and  $f(t) = (\sqrt{t} - 1)^2/2$  the Hellinger distance. Most important for us is the case  $f(t) = t^2 - 1$ , which yields the  $\chi^2$ -divergence

$$\chi^2(\nu | \mu) = \begin{cases} \int_S \rho^2 d\mu - 1 & \text{if } \nu \ll \mu, \\ +\infty & \text{otherwise.} \end{cases} \quad (1.2)$$

As a consequence of Jensen's inequality, the  $f$ -divergence of the law of  $X_t$  with respect to its invariant measure is a non-increasing function of time, i.e.  $t \mapsto D_f(\nu p_t | \mu)$  is

non-increasing for all  $\nu \in \mathcal{P}(S)$ , see [73]. In the special case of relative entropy, this is known as the H-theorem in statistical physics and was first described by Boltzmann in 1872 [42].

### 1.2.2. Convergence in $L^2$ and relaxation times

The  $\chi^2$ -divergence (1.2) admits the variational characterisation

$$\chi^2(\nu \mid \mu)^{1/2} = \sup \left\{ \int_S f \, d\nu : f \in L_0^2(\mu), \|f\|_{L^2(\mu)} \leq 1 \right\}. \quad (1.3)$$

Here and in the following,  $L_0^2(\mu) = \{f \in L^2(\mu) : \int f \, d\mu = 0\}$  denotes the subspace of  $L^2(\mu)$  consisting of mean-zero functions. This representation yields

$$\int_S f \, d(\nu p_t) = \int_S P_t f \, d\nu \leq \|P_t f\|_{L^2(\mu)} \chi^2(\nu \mid \mu)^{1/2} \quad \text{for all } f \in L_0^2(\mu),$$

so that taking the supremum over  $f \in L_0^2(\mu)$  with  $\|f\|_{L^2(\mu)} \leq 1$  shows that

$$\chi^2(\nu p_t \mid \mu) \leq \|P_t\|_{B(L_0^2(\mu))}^2 \chi^2(\nu \mid \mu), \quad (1.4)$$

where  $\|\cdot\|_{B(L_0^2(\mu))}$  denotes the operator norm of a bounded linear operator on  $L_0^2(\mu)$ .

$\|P_t\|_{B(L_0^2(\mu))}$  denotes the operator norm of  $P_t : L_0^2(\mu) \rightarrow L_0^2(\mu)$ . This motivates the definition

$$t_{\text{rel}}(P, \varepsilon) = \inf \{t \geq 0 : \|P_t f\|_{L^2(\mu)} \leq \varepsilon \|f\|_{L^2(\mu)} \text{ for all } f \in L_0^2(\mu)\}$$

of the  $L^2$   $\varepsilon$ -relaxation time of  $(P_t)_{t \geq 0}$ : After the time  $t_{\text{rel}}(P, \varepsilon)$ , the square-root of the  $\chi^2$ -divergence with respect to the invariant probability measure has decayed by a factor  $\varepsilon$  for any initial distribution with square-integrable density with respect to  $\mu$ . We set  $t_{\text{rel}}(P) = t_{\text{rel}}(P, e^{-1})$  and refer to this quantity simply as the  $L^2$ -relaxation time.

### The Poincaré inequality

A key advantage of  $f$ -divergences is their amenability to analytic tools such as functional inequalities. In particular, for the  $\chi^2$ -divergence, the Poincaré inequality plays a central role in obtaining upper bounds on the  $L^2$ -relaxation time. For a transition semigroup  $(P_t)_{t \geq 0}$  on  $L^2(\mu)$  with generator  $(L, \text{Dom}(L))$ , we say that  $L$  satisfies a Poincaré inequality with constant  $C > 0$  if

$$\int_S f^2 \, d\mu \leq -C \int_S f L f \, d\mu \quad \text{for all } f \in L_0^2(\mu) \cap \text{Dom}(L), \quad (1.5)$$

where the quantity  $\mathcal{E}(f) = -\int_S f L f \, d\mu$  appearing on the right-hand side is the Dirichlet form associated to  $L$ . We denote by  $C_P$  the optimal constant  $C$  in (1.5).

The guiding principle is that bounds for the relaxation time can be obtained by differentiating  $\|P_t f\|_{L^2(\mu)}$  in time, upper bounding the resulting rate of energy dissipation

## 1.2. Convergence to equilibrium of Markov processes

using a functional inequality, and concluding by Grönwall's lemma. More precisely, the Poincaré inequality (1.5) yields

$$\frac{d}{dt} \|P_t f\|_{L^2(\mu)}^2 = -2\mathcal{E}(P_t f) \leq -\frac{2}{C_P} \|P_t f\|_{L^2(\mu)}^2 \quad (1.6)$$

for all  $f \in L_0^2(\mu)$ , so that

$$\|P_t f\|_{L^2(\mu)} \leq e^{-t/C_P} \|f\|_{L^2(\mu)} \quad \text{for all } f \in L_0^2(\mu) \text{ and } t \geq 0.$$

This in particular implies the relaxation time bound

$$t_{\text{rel}}(P, \varepsilon) \leq C_P \log(\varepsilon^{-1}) \quad \text{and} \quad t_{\text{rel}}(P) \leq C_P. \quad (1.7)$$

The Poincaré inequality also has important consequences beyond exponential decay of the associated transition semigroup. For instance, in the special case of a gradient Dirichlet form associated to the generator of an overdamped Langevin diffusion, see [Section 1.3.1](#) below, the Poincaré inequality implies exponential concentration of measure, see e.g. [20, 163].

### The reversible case

In order to better understand (1.7), it is instructive to at first consider the case where the generator  $(L, \text{Dom}(L))$  of the transition semigroup is self-adjoint in  $L^2(\mu)$ . In this case, the corresponding Markov process is reversible with respect to  $\mu$ , i.e.

$$(X_t)_{t \in [0, T]} \sim (X_{T-t})_{t \in [0, T]} \quad \text{under } \mathbb{P}_\mu \text{ for any } T > 0,$$

so that the forward and backward evolutions in time have the same distribution in stationarity. By the spectral theorem, the semigroup can then be written as the exponential  $P_t = \exp(tL)$  defined via spectral theory, see e.g. [104, 217]. In particular, the operator norm of the semigroup can be expressed explicitly as

$$\|P_t\|_{B(L_0^2(\mu))} = \exp(-\text{gap}(L)t),$$

where

$$\text{gap}(L) = \inf \{ \text{Re}(\lambda) : \lambda \in \text{spec}(-L|_{L_0^2(\mu)}) \} \quad (1.8)$$

denotes the spectral gap of the generator  $(L, \text{Dom}(L))$  in  $L^2(\mu)$ . Here  $\text{spec}(-L|_{L_0^2(\mu)})$  is the set of  $\lambda \in \mathbb{C}$  such that the operator  $\lambda I + L : \text{Dom}(L) \cap L_0^2(\mu) \rightarrow L_0^2(\mu)$  does not have a bounded inverse. It describes the gap between the eigenvalue 0 corresponding to the constant functions and the spectrum of  $L$  restricted to their orthogonal complement. By the variational characterisation of the eigenvalues of self-adjoint operators,

$$\text{gap}(L) = \inf \left\{ \frac{\langle f, -Lf \rangle_{L^2(\mu)}}{\|f\|_{L^2(\mu)}^2} : f \in L_0^2(\mu) \cap \text{Dom}(L) \setminus \{0\} \right\}, \quad (1.9)$$

so that, in the reversible case, the inverse spectral gap and the Poincaré constant coincide. We therefore have

$$\frac{1}{\text{gap}(L)} = t_{\text{rel}}(P) = C_P. \quad (1.10)$$

### The non-reversible case

The non-reversible case, where the generator  $(L, \text{Dom}(L))$  of  $(P_t)_{t \geq 0}$  is no longer self-adjoint in the Hilbert space  $L^2(\mu)$ , is of particular interest in applications, see the discussion in [Section 1.3](#) below. Here, the operator norm  $\|P_t\|_{B(L_0^2(\mu))}$  is no longer a pure exponential, and one can only expect bounds of the form

$$\|P_t\|_{B(L_0^2(\mu))} \leq C \exp(-\lambda t) \quad (1.11)$$

with constants  $C \in [1, \infty)$  and  $\lambda \leq \text{gap}(L)$ , see [\[104, Proposition IV.2.2\]](#), due to the lack of a spectral theorem. In contrast to the reversible case, one therefore only has the relation

$$\frac{1}{\text{gap}(L)} \leq t_{\text{rel}}(P) \leq C_P \quad (1.12)$$

in general. Note that the Dirichlet form  $\mathcal{E}(f)$  only depends on the symmetric part  $L_s = \frac{1}{2}(L + L^*)$  of the generator, and cannot measure acceleration of convergence due to the antisymmetric part that we will observe in [Section 1.3](#). In particular, the Poincaré constant may be infinite, while the process still converges exponentially fast, as is for instance the case for Langevin dynamics, see [Section 1.3.3](#) below.

A useful lower bound on the relaxation time in the non-reversible case can be given in terms of the *singular value gap*

$$s(L) = \inf \left\{ \frac{\|Lf\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}} : f \in L_0^2(\mu) \cap \text{Dom}(L) \setminus \{0\} \right\}. \quad (1.13)$$

In the reversible case,  $s(L) = \text{gap}(L)$ , so that the singular value gap and the usual spectral gap coincide. In the non-reversible case, one only has  $|\lambda| \geq s(L)$  for any  $\lambda \in \text{spec}(-L|_{L_0^2(\mu)})$ , which, however, does not yield a direct relation to the spectral gap which is defined in terms of the real part of the spectrum.

**Lemma 1.1.** *For any  $\varepsilon \in (0, 1)$ , we have*

$$t_{\text{rel}}(P, \varepsilon) \geq \frac{1 - \varepsilon}{s(L)}.$$

*Proof.* We have  $\|P_t f - f\|_{L^2(\mu)} \leq \int_0^t \|P_s Lf\|_{L^2(\mu)} \, ds \leq t \|Lf\|_{L^2(\mu)}$  for any  $f \in L_0^2(\mu) \cap \text{Dom}(L)$  and  $t \geq 0$ , so that

$$\|P_t f\|_{L^2(\mu)} \geq \|f\|_{L^2(\mu)} - \|P_t f - f\|_{L^2(\mu)} \geq \|f\|_{L^2(\mu)} - t \|Lf\|_{L^2(\mu)}.$$

For  $\sigma > s(L)$ , we can choose  $f \in L_0^2(\mu) \cap \text{Dom}(L)$  with  $\|Lf\|_{L^2(\mu)} < \sigma \|f\|_{L^2(\mu)}$ , so that

$$\|P_t f\|_{L^2(\mu)} > (1 - \sigma t) \|f\|_{L^2(\mu)},$$

and in particular  $t_{\text{rel}}(P, \varepsilon) \geq (1 - \varepsilon)/\sigma$ . Taking  $\sigma \downarrow s(L)$  finishes the proof.  $\square$

Bounds such as (1.11) correspond hypocoercivity of the generator [248] and can immediately be used to obtain the upper bound

$$t_{\text{rel}}(P, \varepsilon) \leq \frac{\log(C) + \log(\varepsilon^{-1})}{\lambda}$$

on the relaxation time. However, proving (1.11) with quantitative estimates on the constants  $C$  and  $\lambda$  is a challenging task in many applications, since standard tools such as the Poincaré inequality do not yield sharp results or fail completely. Hypocoercivity will be discussed in more detail in Section 2.2, and we will show how a time-integrated Poincaré inequality, the flow Poincaré inequality, can be leveraged to obtain such hypocoercive bounds.

### 1.2.3. Other convergence metrics

Before continuing the study of convergence in  $\chi^2$ -divergence, let us mention some other convergence metrics, as well as their relations and applications. Hence let  $\mu, \nu \in \mathcal{P}(S)$  be probability measures.

Another analytically important  $f$ -divergence is the relative entropy  $H(\nu \mid \mu)$  of  $\nu$  with respect to  $\mu$ . It arises by choosing  $f(t) = t \log t$  in (1.1), i.e.

$$H(\nu \mid \mu) = \int_S \rho \log \rho \, d\mu = \int_S \log \rho \, d\nu$$

if  $\nu \ll \mu$  with relative density  $\rho$ , and  $H(\nu \mid \mu) = \infty$  otherwise. The role of the Poincaré inequality (1.5) in governing decay in  $\chi^2$ -divergence is taken on by the logarithmic Sobolev inequality for decay in relative entropy, see e.g. [225]. The relative entropy can be bounded by the  $\chi^2$ -divergence as

$$H(\nu \mid \mu) \leq \log(1 + \chi^2(\nu \mid \mu)) \leq \chi^2(\nu \mid \mu).$$

Furthermore, the  $\chi^2$ -divergence arises as an approximation of the relative entropy up to second order, since a Taylor expansion of  $x \log(x)$  around 1 yields  $x \log(x) = (x - 1) + \frac{1}{2}(x - 1)^2 + O(|x - 1|^3)$ . Logarithmic Sobolev inequalities and convergence in relative entropy are of particular importance in high dimension, since the relative entropy often scales better with the dimension than the  $\chi^2$ -divergence. For instance, for product measures, the  $\chi^2$ -divergence grows exponentially in the dimension, while the relative entropy only grows linearly.

A further class of distance functions on  $\mathcal{P}(S)$  is given by Wasserstein distances. Given a lower semi-continuous metric  $d: S \times S \rightarrow [0, \infty)$  on  $S$ , the  $L^p$ -Wasserstein distance between probability measures  $\mu, \nu \in \mathcal{P}(S)$  is

$$\mathcal{W}^p(\nu, \mu) = \inf_{\gamma \in \Pi(\nu, \mu)} \left( \int d(x, y)^p \gamma(dx \, dy) \right)^{1/p}, \quad (1.14)$$

where  $\Pi(\nu, \mu)$  denotes the set of all couplings of  $\nu$  and  $\mu$ , i.e. probability measures  $\gamma$  on  $S \times S$  with  $\gamma(A \times S) = \nu(A)$  and  $\gamma(S \times B) = \mu(B)$  for all  $A, B \in \mathcal{B}(S)$ . The

minimisation in (1.14) is a particular case of an optimal transportation problem for the cost function  $c(x, y) = d(x, y)^p$  in the Kantorovich formulation, and a minimising coupling exists, see [247]. The  $L^p$ -Wasserstein distance forms a metric on the space  $\mathcal{P}^p(S)$  of probability measures on  $S$  with finite  $p$ -th moment, which, for  $p = 1$ , is also known as the Kantorovich-Rubinstein distance. A special case arises by choosing  $d(x, y) = 1_{\{x \neq y\}}$  and  $p = 1$ , where the Wasserstein distance coincides with the total variation norm

$$\|\nu - \mu\|_{\text{TV}} = \sup_{A \in \mathcal{B}(S)} |\nu(A) - \mu(A)|, \quad (1.15)$$

see e.g. [247] or [179]. Total variation distances can be bounded via Pinsker's inequality

$$\|\nu - \mu\|_{\text{TV}} \leq \left( \frac{1}{2} H(\nu \mid \mu) \right)^{1/2},$$

see [160].

Since the Wasserstein distances (1.14) are defined as an infimum, upper bounds for  $\mathcal{W}^p(\nu, \mu)$  can be obtained by considering any coupling of  $\nu$  and  $\mu$ . To study convergence of a Markov process to stationarity, one hence constructs couplings of two copies  $(X_t)_{t \geq 0}$  and  $(Y_t)_{t \geq 0}$  of the process with different initial laws  $\nu$  and  $\tilde{\nu}$ . If one can find such a coupling such that there exists a constant  $C > 0$  with

$$\mathbb{E}[d(X_t, Y_t)^p] \leq \exp(-Ct) \mathbb{E}[d(X_0, Y_0)^p] \quad \text{for all } t \geq 0 \text{ and } \nu, \tilde{\nu} \in \mathcal{P}^p(S),$$

then the transition function  $(p_t)_{t \geq 0}$  is a contraction with respect to the  $L^p$ -Wasserstein distance, and the existence of an invariant probability measure  $\mu$  with exponential convergence  $\mathcal{W}^p(\nu p_t, \mu) \leq e^{-Ct} \mathcal{W}^p(\nu, \mu)$  for all  $\nu \in \mathcal{P}^p(S)$  and  $t \geq 0$  follows, see e.g. [95, 96, 183].

### 1.3. Acceleration through non-reversibility

There are many ways to construct a Markov process with given invariant probability measure. In the origins of Markov chain Monte Carlo, this was achieved by constructing a Markov chain satisfying the detailed balance condition, a local symmetry condition on the transitions of the Markov chain implying invariance of the desired distribution. It is, however, not necessary, and recently, there has been growing interest in accelerated convergence to stationarity in absence of the detailed balance condition, leading to non-reversible Markov processes, see for instance [48, 79, 143, 158, 175, 203].

We begin by introducing the detailed balance condition and the paradigmatic example of a continuous-time Markov process satisfying it, the overdamped Langevin diffusion, in Section 1.3.1, before exploring how it can be modified to accelerate convergence to stationarity in Section 1.3.2. The main insight is that acceleration can be achieved through non-reversibility and degenerate noise, and we present two classes of processes based on this observation in Sections 1.3.3 and 1.3.4.

### 1.3.1. Reversible methods and detailed balance

A simple and sufficient, yet not necessary, requirement for a Markov chain with transition kernel  $p$  to have  $\mu$  as invariant measure is the detailed balance condition

$$\mu(dx)p(x, dy) = \mu(dy)p(y, dx) \quad (1.16)$$

that has been used in the context of Markov chain Monte Carlo since the original work of Metropolis and co-authors [194]. To this day, it forms the basis of many MCMC methods, such as Gibbs sampling [114], Glauber dynamics [117], or Hit-and-Run [233, 244], or enters through the inclusion of a Metropolis-Hastings acceptance-rejection step in the spirit of [130] in a vast variety of algorithms, such as the Metropolis-adjusted Langevin algorithm [222] or the No-U-turn sampler [136], to name but a few.

For continuous-time Markov processes with generator  $(L, \text{Dom}(L))$ , the analogous statement is that

$$\int_S f Lg \, d\mu = \int_S g Lf \, d\mu \quad \text{for all } f, g \in \text{Dom}(L), \quad (1.17)$$

which amounts to self-adjointness of the generator in  $L^2(\mu)$ , so that the corresponding Markov process is reversible with respect to  $\mu$ .

In case  $S = \mathbb{R}^d$ , a prototypical reversible diffusion with given invariant probability measure is the overdamped Langevin diffusion. If  $\mu \in \mathcal{P}(\mathbb{R}^d)$  is a probability measure with

$$\mu(dx) \propto \exp(-U(x)) \, dx, \quad (1.18)$$

where  $U \in C^1(\mathbb{R}^d)$  is a potential such that  $U(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ , then the overdamped Langevin diffusion is the process  $(Z_t)_{t \geq 0}$  with state space  $\mathbb{R}^d$  that solves the stochastic differential equation

$$dZ_t = -\frac{1}{2} \nabla U(Z_t) \, dt + dB_t, \quad (1.19)$$

where  $(B_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion. Under mild drift assumptions, such as

$$\lim_{|x| \rightarrow \infty} \frac{x \cdot \nabla U(x)}{|x|^2} > -\infty,$$

a unique non-explosive strong solution exists for any initial distribution, see e.g. [211], and the measure  $\mu$  is the unique invariant probability measure. The generator of the overdamped Langevin diffusion in  $L^2(\mu)$  is given by

$$Lf = -\frac{1}{2} \nabla U \cdot \nabla f + \frac{1}{2} \Delta f \quad \text{for all } f \in C_b^\infty(\mathbb{R}^d), \quad (1.20)$$

and the space  $C_b^\infty(\mathbb{R}^d)$  of smooth functions with bounded derivatives is a core for  $L$  [251]. The generator can also be written as  $L = -\frac{1}{2} \nabla^* \nabla$ , where the adjoint is in  $L^2(\mu)$ . In particular,  $(L, \text{Dom}(L))$  satisfies (1.17) and is self-adjoint. The associated Dirichlet form is the gradient Dirichlet form

$$\mathcal{E}(f, g) = \frac{1}{2} \int_{\mathbb{R}^d} \nabla f \cdot \nabla g \, d\mu.$$

By reversibility, if the measure  $\mu$  satisfies the Poincaré inequality<sup>1</sup>

$$\int_{\mathbb{R}^d} f^2 d\mu \leq \frac{1}{m} \int_{\mathbb{R}^d} |\nabla f|^2 d\mu \quad \text{for all } f \in C_c^\infty(\mathbb{R}^d) \quad (1.21)$$

with some  $m \in (0, \infty)$ , the relaxation time of the associated transition semigroup  $(P_t)_{t \geq 0}$  satisfies  $t_{\text{rel}}(P) \leq \frac{2}{m}$ , with equality if  $m$  is chosen maximally such that (1.21) holds. The inequality (1.21) is for instance satisfied if  $U \in C^2(\mathbb{R}^d)$  is  $m$ -strongly convex, i.e.  $\nabla^2 U(x) \geq mI_d$  for all  $x \in \mathbb{R}^d$  in the sense of quadratic forms by the Bakry-Émery theorem [19], and more generally if  $U \in C^2(\mathbb{R}^d)$  is a bounded perturbation of a strongly convex function by the theorem of Holley-Strook [138].<sup>2</sup> In applications, due to issues such as high-dimensionality, multimodality or non-convexity, the regime of large Poincaré constant, and hence small  $m \in (0, 1)$ , is of particular interest. In these situations, the order of the dependence of the relaxation time on  $m$  is crucial.

**Example 1.2** (Gaussian case). Consider the probability measure  $\mu = \mathcal{N}(0, C^{-1})$  on  $\mathbb{R}^d$  with inverse covariance matrix  $C \in \mathbb{R}^{d \times d}$ , i.e.  $U(x) = \frac{1}{2}x^\top Cx$ . Then the overdamped Langevin diffusion (1.19) becomes a generalised Ornstein-Uhlenbeck process

$$dZ_t = -\frac{1}{2}CZ_t dt + dB_t.$$

In this case, the spectrum of the associated generator is given by

$$\text{spec}(-L) = \left\{ \frac{1}{2} \sum_{i=1}^n a_i \lambda_i : a_i \in \mathbb{N}_0, \lambda_i \in \text{spec}(C) \right\}, \quad (1.22)$$

see [193, Theorem 3.1]. In particular, the spectral gap of  $L$  is  $\text{gap}(L) = \frac{1}{2}m$ , where  $m = \min \text{spec}(C)$  is the minimal eigenvalue of  $C$ , and the relaxation time is  $t_{\text{rel}}(P) = \frac{2}{m}$ .

While the sampling of Gaussian measures can typically be performed directly, not necessitating the use of MCMC methods, they still provide a useful base case and toy example. Furthermore, they arise as natural reference measures in high-dimensional settings as finite-dimensional projections of infinite-dimensional Gaussian measures such as the Wiener measure, and understanding the effect of high-dimensionality and anisotropy on convergence rates and relaxation times is important.

MCMC algorithms based on the overdamped Langevin diffusion include the unadjusted Langevin algorithm (ULA) [213], which is an Euler discretisation of (1.19), and its unbiased counterpart, the Metropolis-adjusted Langevin algorithm (MALA), obtained through inclusion of a Metropolis-Hastings accept-reject step [222]. While these have been widely and effectively used in applications, they exhibit diffusive behaviour due to

<sup>1</sup>Note that we say that a probability measure satisfies a Poincaré inequality with constant  $\frac{1}{m}$  if (1.21) holds, i.e. if the generator  $L$  of an overdamped Langevin diffusion with invariant probability measure  $\mu$  satisfies the Poincaré inequality (1.5) with constant  $C = \frac{2}{m}$ .

<sup>2</sup>Both these results in fact show that a logarithmic Sobolev inequality holds, which implies a Poincaré inequality by a linearisation argument [224].

the detailed balance condition (1.16). Intuitively, the detailed balance condition imposes a local symmetry condition in the sense that, in stationarity, a step from  $x$  to  $y$  is followed by an equally likely step from  $y$  to  $x$ . This inhibits efficient exploration of the state space and thereby good convergence behaviour. We will explore in the following sections how breaking detailed balance and considering non-reversible Markov processes can improve convergence to equilibrium.

### 1.3.2. Measure-preserving diffusions and antisymmetric perturbations

While the overdamped Langevin diffusion (1.19) is arguably the simplest diffusion with  $\mu$  as invariant probability measure, it is by no means unique with this property. More generally, we consider drift and diffusion coefficients  $b \in C^1(\mathbb{R}^d, \mathbb{R}^d)$  and  $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$  and let  $a(x) = \sigma(x)\sigma(x)^\top$ . Then the Itô diffusion

$$\begin{aligned} dZ_t &= \frac{1}{2}(\exp(U) \operatorname{div}(\exp(-U)a))(Z_t) dt + b(Z_t) dt + \sigma(Z_t) dB_t \\ &= -\left(\frac{1}{2}a(Z_t)\nabla U(Z_t) - \operatorname{div}(a)(Z_t)\right) dt + b(Z_t) dt + \sigma(Z_t) dB_t \end{aligned} \quad (1.23)$$

with a  $d$ -dimensional Brownian motion  $(B_t)_{t \geq 0}$  has the generator

$$\begin{aligned} Lf &= \frac{1}{2}(\exp(U) \operatorname{div}(\exp(-U)a \nabla f) + b \cdot \nabla f) \\ &= -\frac{1}{2}\nabla^*(a \nabla f) + b \cdot \nabla f, \end{aligned} \quad (1.24)$$

where the adjoint is in  $L^2(\mu)$ . The first summand in (1.24) is symmetric in  $L^2(\mu)$ , so that the measure  $\mu$  is infinitesimally invariant if and only if the drift is divergence-free, i.e.

$$\int_{\mathbb{R}^d} Lf d\mu = 0 \quad \text{for all } f \in C_c^\infty(\mathbb{R}^d) \iff \operatorname{div}(\exp(-U)b) = 0. \quad (1.25)$$

In this case, the operator  $b \cdot \nabla$  is antisymmetric in  $L^2(\mu)$ . Note that choosing  $\sigma(x) = I_d$  and  $b(x) = 0$  recovers the overdamped Langevin diffusion (1.19). Invariance of the measure  $\mu$  then immediately follows from infinitesimal invariance if  $C_c^\infty(\mathbb{R}^d)$  is a core for the generator. Such divergence-free vector fields can easily be constructed by choosing any antisymmetric matrix  $J \in \mathbb{R}^{d \times d}$  and setting  $b(x) = J\nabla U(x)$ .

In case  $\sigma(x) = I_d$  and  $b$  is divergence-free, the SDE (1.23) becomes

$$dZ_t = -\frac{1}{2}\nabla U(Z_t) dt + b(Z_t) dt + dB_t, \quad (1.26)$$

an antisymmetric perturbation of the classical overdamped Langevin diffusion with invariant measure  $\mu$ . It is shown in [143] that adding such a divergence-free drift to the dynamics cannot decrease both the spectral gap of the generator and the asymptotic rate of convergence in total variation distance. This shows that, at least at the level of the continuous-time overdamped Langevin diffusion, convergence behaviour is improved by including non-reversibility.

**Example 1.3.** Returning to the Gaussian case  $\mu = \mathcal{N}(0, C^{-1})$  introduced in [Example 1.2](#), considering linear drifts and constant diffusion coefficients in [\(1.23\)](#) leads to generalised Ornstein-Uhlenbeck processes of the form

$$dZ_t = -\frac{1}{2}(D + J)CZ_t dt + D^{1/2}dB_t \quad (1.27)$$

with  $D \in \mathcal{S}_+^d(\mathbb{R})$  and  $J \in \mathcal{A}^d(\mathbb{R})$ , where  $\mathcal{S}_+^d(\mathbb{R})$  denotes the cone of symmetric, positive semidefinite real  $d \times d$ -matrices and  $\mathcal{A}^d(\mathbb{R})$  the set of antisymmetric real  $d \times d$ -matrices. Denoting the corresponding generators in  $L^2(\mu)$  by  $L_{D,J}$ , [\[193, Theorem 3.1\]](#) again shows that

$$\text{spec}(-L_{D,J}) = \left\{ \frac{1}{2} \sum_{i=1}^n a_i \lambda_i : a_i \in \mathbb{N}_0, \lambda_i \in \text{spec}((D + J)C) \right\},$$

and in particular  $\text{gap}(L_{D,J}) = \frac{1}{2} \min\{\text{Re}(\lambda) : \lambda \in \text{spec}((D + J)C)\}$ . Recall that for overdamped Langevin dynamics,  $D = I_d$  and  $J = 0$ , and we have  $\text{gap}(L_{I_d,0}) = \frac{1}{2} \min \text{spec}(C)$ .

At first considering only antisymmetric perturbations of the drift yields

$$\max\{\text{gap}(L_{I_d,J}) : J \in \mathcal{A}^d(\mathbb{R})\} = \frac{\text{tr}(C)}{2d}, \quad (1.28)$$

as shown in [\[142\]](#), with constructive procedures to choose such a  $J$  given in [\[175, 253\]](#). The spectral gap can hence be improved from half the minimal eigenvalue to half the mean of the eigenvalues of  $C$ . If, on the other hand, one imposes  $J = 0$  and allows to vary  $D$ , one obtains

$$\max\{\text{gap}(L_{D,0}) : D \in \mathcal{S}_+^d(\mathbb{R}), \text{tr}(D) \leq d\} = \frac{d}{2 \text{tr}(C^{-1})}, \quad (1.29)$$

half the harmonic mean of the eigenvalues of  $C$ , as shown in [\[122\]](#), where the maximum is attained for  $D = \frac{d}{\text{tr}(C^{-1})}C^{-1}$ . A constraint on  $D$  is necessary since otherwise the spectral gap (and rate of convergence) could be made arbitrarily large by multiplying  $D$  by a constant, corresponding to a time acceleration. Finally, combining antisymmetric drift perturbations and non-trivial diffusion matrices,

$$\max\{\text{gap}(L_{D,J}) : D \in \mathcal{S}_+^d(\mathbb{R}), J \in \mathcal{A}^d(\mathbb{R}), \text{tr}(D) \leq d\} = \frac{1}{2} \max \text{spec}(C),$$

where the maximum is attained for  $D = d v v^\top$  with a normalised eigenvector  $v$  to the largest eigenvalue of  $C$ . Hence the spectral gap  $\text{gap}(L_{D,J})$  can be maximised by choosing a highly degenerate diffusion [\(1.27\)](#): The diffusion matrix  $D = d v v^\top$  is only of rank 1. In general, one has

$$\min \text{spec}(C) \leq \frac{d}{\text{tr}(C^{-1})} \leq \frac{\text{tr}(C)}{d} \leq \max \text{spec}(C),$$

with equality holding if and only if all eigenvalues of  $C$  coincide. Typically, probability distributions of interest have many different scales, so that, in this toy example, the

eigenvalues of  $C$  lie over different orders of magnitude. Therefore, adding a suitable antisymmetric drift can already improve the spectral gap by passing from the minimal eigenvalue to the arithmetic mean, while further considering degenerate diffusions with  $\dim \ker(D) > 0$  has the potential to vastly improve it from the minimal to the maximal eigenvalue.

Beyond the Gaussian case, it is unfortunately unclear how to choose an optimal divergence-free vector field  $b$  in practice, and in what way this is to be combined with degenerate noise. However, the principle that non-reversible dynamics with degenerate noise can be used to accelerate convergence remains applicable, and has led to several systematic approaches to the construction of non-reversible Markov processes and associated MCMC algorithms. We detail some of these in the following sections.

### 1.3.3. Processes based on Hamiltonian dynamics

One way of introducing non-reversibility is by considering kinetic, second-order processes. These can be constructed as stochastic perturbations of the Hamiltonian flow

$$\dot{X}_t = V_t, \quad \dot{V}_t = -\nabla U(X_t) \quad (1.30)$$

associated to the Hamiltonian

$$H(x, v) = U(x) + \frac{1}{2}|v|^2, \quad (x, v) \in \mathbb{R}^d \times \mathbb{R}^d \quad (1.31)$$

with unit mass. Since the Hamiltonian flow is symplectic and preserves  $H$  [18], it leaves the Boltzmann-Gibbs distribution

$$\hat{\mu} = \mu \otimes \kappa \quad \text{with } \kappa = \mathcal{N}(0, I_d) \quad (1.32)$$

with density proportional to  $\exp(-H)$  on  $\mathbb{R}^d \times \mathbb{R}^d$  invariant. Here, as before,  $\mu(dx) \propto \exp(-U(x))dx$  with  $U \in C^1(\mathbb{R}^d)$ , and  $X$  and  $V$  are typically interpreted as a position and velocity or momentum component, respectively. Note that the first  $d$ -dimensional marginal of  $\hat{\mu}$  is  $\mu$ , so that samples from  $\hat{\mu}$  yield samples from  $\mu$  by projection. The generator of the deterministic dynamics (1.30) in  $L^2(\hat{\mu})$  is given by

$$\hat{L}_{\text{ham}} f(x, v) = v \cdot \nabla_x f(x, v) - \nabla U(x) \cdot \nabla_v f(x, v),$$

where  $\{f \in H_{\text{loc}}^{1,2}(\mathbb{R}^d \times \mathbb{R}^d) \cap L^2(\hat{\mu}) : \hat{L}_{\text{ham}} f \in L^2(\hat{\mu})\} \subseteq \text{Dom}(\hat{L}_{\text{ham}})$ . Since the Hamiltonian dynamics (1.30) preserves the total energy  $H$ , it is not ergodic. To mitigate this issue, one considers stochastic perturbations of the Hamiltonian flow, where the noise is typically added in the velocity component only. This approach in particular leads to randomised Hamiltonian Monte Carlo and Langevin dynamics.

### Randomised Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) methods are a class of sampling methods that are obtained by interspersing the Hamiltonian dynamics (1.30), or a suitable discretisation,

with resamplings of the velocity component at deterministic or random times. They were first introduced as Hybrid Monte Carlo [89, 141] in the statistical physics community in the context of numerical simulation of lattice field theories, and later popularised in the statistics community by Neal [202, 204]. Today, sampling methods based on HMC such as the No-U-turn Sampler (NUTS) [136] are implemented as the default algorithm in modern statistical programming languages such as Stan [62] and PyMC [2] that are widely used in statistics to perform efficient Bayesian inference.

Here we consider randomised Hamiltonian Monte Carlo (RHMC) [47], where the Hamiltonian dynamics (1.30) is integrated exactly and the velocity is fully resampled from  $\kappa$  after successive integration times with exponential distribution  $\text{Exp}(\gamma)$  with parameter  $\gamma > 0$ . In applications, discretisation schemes such as the leapfrog integrator and the potentially introduced bias play an important role, see [204], while we focus on the idealised dynamics and do not consider these here. The result is a time-homogeneous piecewise-deterministic Markov process  $(X_t, V_t)_{t \geq 0}$  with state space  $\mathbb{R}^d \times \mathbb{R}^d$ , invariant probability measure  $\hat{\mu}$  given by (1.32) and generator

$$\hat{L}^{(\gamma)} f = \hat{L}_{\text{ham}} f + \gamma(\Pi - I)f, \quad (1.33)$$

where the velocity average

$$\Pi f(x, v) = \int_{\mathbb{R}^d} f(x, w) \kappa(dw) \quad (1.34)$$

corresponds to a full resampling of the velocity, and again  $\{f \in H_{\text{loc}}^{1,2}(\mathbb{R}^d \times \mathbb{R}^d) \cap L^2(\hat{\mu}) : \hat{L}_{\text{ham}} f \in L^2(\hat{\mu})\} \subseteq \text{Dom}(\hat{L}^{(\gamma)})$ . We refer to Section 1.3.4 below and [75] for background on piecewise-deterministic Markov processes (PDMPs), and to [47] for details on and a construction of randomised Hamiltonian Monte Carlo. The associated equation  $\partial_t f = (\hat{L}^{(\gamma)})^* f$ , where the adjoint is with respect to  $\hat{\mu}$ , models the motion of a particle under the influence of an external potential  $U$  and random collisions with other particles with Gaussian velocities [198] and is known as the linear Boltzmann equation in physics, see e.g. [64, 128].

### Langevin dynamics

Instead of fully resampling the velocity according to the Gaussian distribution  $\kappa$  with rate  $\gamma > 0$ , the deterministic Hamiltonian dynamics (1.30) can be stochastically perturbed by an Ornstein-Uhlenbeck process acting only on the velocity component leaving  $\kappa$  invariant. This yields the Langevin dynamics  $(X_t, V_t)_{t \geq 0}$  solving the stochastic differential equation

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= -\nabla U(X_t) dt - \gamma V_t dt + \sqrt{2\gamma} dB_t, \end{aligned} \quad (1.35)$$

where  $\gamma \in (0, \infty)$  is interpreted as a friction parameter and  $(B_t)_{t \geq 0}$  is a standard Brownian motion. Such a process describing the movement of a particle subjected to a potential  $U$ , damping with rate  $\gamma$ , and random collisions was first studied by Langevin in 1908

[162], following works of Einstein [102]. If  $U \in C^1(\mathbb{R}^d)$  such that  $U(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$  and  $|\nabla U(x)|$  grows at most polynomially at infinity, a unique global strong solution exists, see e.g. [220, Chapter IX.2] and [148], and  $\hat{\mu}$  is the unique invariant probability measure [215, Proposition 6.1]. The action of the generator of the associated transition semigroup on  $L^2(\hat{\mu})$  is given by

$$\hat{L}^{(\gamma)} f = \hat{L}_{\text{ham}} f + \gamma(-v \cdot \nabla_v + \Delta_v) f \quad \text{for } f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d), \quad (1.36)$$

and the domain of  $\hat{L}^{(\gamma)}$  is given by  $\text{Dom}(\hat{L}^{(\gamma)}) = \{f \in L^2(\hat{\mu}) : \hat{L}^{(\gamma)} f \in L^2(\hat{\mu})\}$  [131]. Langevin dynamics is a paradigmatic example of a non-reversible diffusion process with degenerate noise, and will serve as a useful example in demonstrating how to obtain upper bounds on the relaxation time by proving that it is hypocoercive in Section 2.2. In analogy, the Kolmogorov forward equation associated to Langevin dynamics is known as the kinetic Fokker-Planck equation and is an example of a hypoelliptic equation [140]. It was first studied by Kramers [156] in the context of chemical reactions, following works of Kolmogorov [152]. In practice, various numerically implementable discretisation and splitting schemes for Langevin dynamics can be considered, see e.g. [74, 90, 171, 199, 231], which we will not further elaborate on here.

**Remark 1.4** (Overdamped limit). The reason for the terminology “overdamped” Langevin diffusion for the process given by (1.19) is that it arises from (1.35) in the limit  $\gamma \rightarrow \infty$  of infinite damping coefficient, also known as the Smoluchowski-Kramers limit. More precisely, the position component  $(X_{\gamma t})_{t \geq 0}$  of Langevin dynamics with friction  $\gamma$  sped up by a factor  $\gamma$  converges to a solution  $(Z_t)_{t \geq 0}$  of (1.19) in probability uniformly on compact time intervals, see [215, Section 6.5.1] for an intuitive derivation and [109, 205] for precise statements. To avoid confusion, the Langevin diffusion (1.35) is sometimes referred to as the underdamped or kinetic Langevin diffusion. We prefer to avoid the terminology “underdamped”, as the Langevin dynamics (1.35) is defined for any value of the damping parameter  $\gamma \in (0, \infty)$ .

**Example 1.5.** In the one-dimensional Gaussian case  $\mu = \mathcal{N}(0, m^{-1})$  on  $\mathbb{R}$  with  $m \in (0, \infty)$ , i.e.  $U(x) = \frac{m}{2}x^2$ , the overdamped Langevin diffusion has relaxation time  $t_{\text{rel}}(P) = \frac{2}{m}$ . The Langevin dynamics with friction  $\gamma \in (0, \infty)$  can be written as

$$d \begin{pmatrix} X_t \\ V_t \end{pmatrix} = A \begin{pmatrix} X_t \\ V_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sqrt{2\gamma} dB_t \end{pmatrix} \quad \text{with} \quad A = \begin{pmatrix} 0 & 1 \\ -m & -\gamma \end{pmatrix}.$$

By [17], the corresponding transition semigroup satisfies

$$\|\hat{P}_t\|_{B(L_0^2(\hat{\mu}))} = \|\exp(t\tilde{A})\| \quad \text{with} \quad \tilde{A} = \begin{pmatrix} 0 & \sqrt{m} \\ -\sqrt{m} & -\gamma \end{pmatrix}.$$

The asymptotic decay rate is hence given by

$$-\lim_{t \rightarrow \infty} \frac{1}{t} \log \|\exp(tA)\| = \inf\{-\text{Re}(\lambda) : \lambda \in \text{spec}(A)\} = \begin{cases} \frac{\gamma}{2} & \text{for } \gamma \leq 2\sqrt{m} \\ \frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} - m} & \text{for } \gamma > 2\sqrt{m} \end{cases},$$

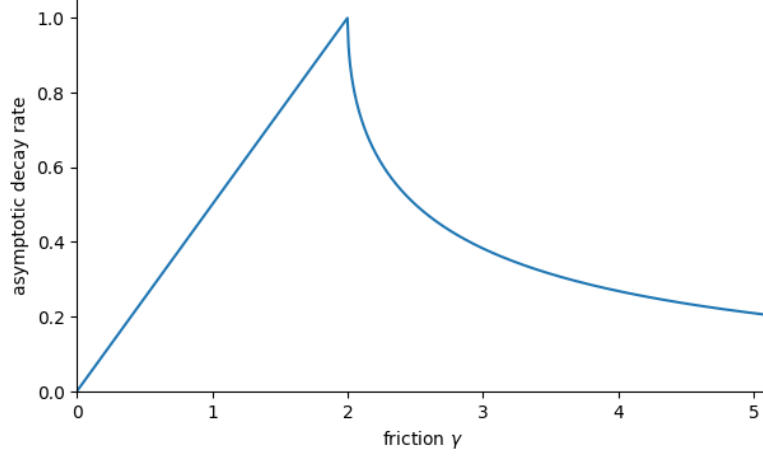


Figure 1.1.: Asymptotic decay rate of Langevin dynamics for  $\mu = \mathcal{N}(0, 1)$ .

which is maximised for the critical choice  $\gamma = 2\sqrt{m}$ , see Figure 1.1. To obtain non-asymptotic bounds and thereby a bound on the relaxation time, an explicit calculation shows that

$$\exp(tA) = e^{-\frac{\gamma}{2}t} \left( c(\omega t)I + \frac{1}{\omega} s(\omega t)B \right) \quad \text{with} \quad B = \begin{pmatrix} \frac{\gamma}{2} & \sqrt{m} \\ -\sqrt{m} & -\frac{\gamma}{2} \end{pmatrix},$$

where

$$\omega = \sqrt{\left| \frac{\gamma^2}{4} - m \right|}, \quad c(x) = \begin{cases} \cos(x) & \text{for } \gamma \leq 2\sqrt{m} \\ \cosh(x) & \text{for } \gamma > 2\sqrt{m} \end{cases}, \quad s(x) = \begin{cases} \sin(x) & \text{for } \gamma \leq 2\sqrt{m} \\ \sinh(x) & \text{for } \gamma > 2\sqrt{m} \end{cases}.$$

Computing the operator norm of  $\exp(tA)$  therefore yields

$$\|\hat{P}_t\|_{B(L_0^2(\hat{\mu}))} = e^{-\frac{\gamma}{2}t} \left( \frac{\gamma}{2\omega} |s(\omega t)| + \sqrt{1 + \frac{\gamma^2}{4\omega^2} s(\omega t)^2} \right),$$

with the understanding that  $s(\omega t)/\omega = t$  in the critically damped case  $\gamma = 2\sqrt{m}$ , where  $\omega = 0$ . In this case, we obtain

$$\|\hat{P}_t\|_{B(L_0^2(\hat{\mu}))} = e^{-\sqrt{m}t} (\sqrt{m}t + \sqrt{1 + mt^2}) \quad \text{for } \gamma = 2\sqrt{m},$$

so that a numerical calculation yields  $t_{\text{rel}}(\hat{P}) \approx 2.73m^{-1/2}$ , i.e. the relaxation time of critically damped Langevin dynamics is of the order  $m^{-1/2}$ .

A factorisation argument, see Example A.14, shows that for arbitrary Gaussian measures on  $\mathbb{R}^d$ , critically damped Langevin dynamics with  $\gamma = 2\sqrt{m}$ , where  $m$  is the smallest eigenvalue of the inverse covariance matrix, also satisfies  $t_{\text{rel}}(\hat{P}) \approx 2.73m^{-1/2}$ .

In applications, due to issues such as high-dimensionality, multimodality or non-convexity, the regime of large Poincaré constant, and hence small  $m \in (0, 1)$ , is of particular interest. At least in the Gaussian case, [Example 1.5](#) shows that convergence to equilibrium as measured by the relaxation time can be accelerated from order  $m^{-1}$  to  $m^{-1/2}$  by considering the non-reversible, degenerate Langevin dynamics with critical friction instead of the reversible overdamped Langevin diffusion. We will see that this significant improvement can be observed much more generally beyond the Gaussian case: A lower bound of order  $m^{-1/2}$  for the relaxation time holds for a large class of processes, namely second-order lifts, and Langevin dynamics and randomised HMC can achieve this lower bound on a class of mildly non-convex potentials by choosing  $\gamma$  appropriately. The proof of these lower and upper bounds is the content of [Section 2.1.3](#) and [Section 2.2.3](#), respectively.

#### 1.3.4. Piecewise-deterministic Markov processes

Piecewise-deterministic Markov processes (PDMPs), introduced by Davis [\[75, 76\]](#), are a class of continuous-time Markov processes that evolve according to a deterministic dynamics between random times at which they perform random jumps. They can easily be designed to leave a given probability measure invariant, are typically non-reversible, and can often be simulated in a straightforward way, all of which has led to their popularity as a basis for MCMC algorithms [\[108, 245\]](#).

We restrict ourselves to PDMPs on a state space  $S$  whose deterministic dynamics are given by the flow of a continuously differentiable vector field  $b \in C^1(S, S)$ . A PDMP is then defined by the flow  $\xi: \mathbb{R} \times S \rightarrow S$  satisfying the ordinary differential equation

$$\frac{d}{dt}\xi_t(x) = b(\xi_t(x)), \quad \xi_0(x) = x,$$

a continuous jump rate  $\lambda \in C(S, [0, \infty))$ , and a probability kernel  $Q$  on  $(S, \mathcal{B}(S))$ . The resulting process  $(X_t)_{t \geq 0}$  evolves according to the flow  $\xi$  between jumps with transition kernel  $Q$  occurring with rate  $\lambda$ . A construction of these processes is given in [\[92\]](#). This defines a time-homogeneous and right-continuous Markov process that solves the martingale problem for  $(L, C_c^1(S))$ , where

$$Lf = b \cdot \nabla f + \lambda(Q - I)f \quad \text{for } f \in C_c^1(S).$$

The domains of the extended and Feller generator can in fact be fully characterised, see [\[75, Theorem 5.5\]](#) and [\[96, Theorem 5.6\]](#).

For a probability measure  $\mu$  on  $S$ , we say that an operator  $(L, \text{Dom}(L))$  on  $L^2(S, \mu)$  satisfies the generalised detailed balance condition with respect to an involution  $\Theta: S \rightarrow S$  if

$$\int_S f Lg \, d\mu = \int_S L(f \circ \Theta) g \circ \Theta \, d\mu \quad \text{for all } f, g \in \text{Dom}(L). \quad (1.37)$$

Typically,  $S = \mathbb{R}^d \times \mathbb{R}^d$  and  $\Theta(x, v) = (x, -v)$ . Note that the usual detailed balance condition [\(1.17\)](#) is obtained by choosing  $\Theta$  to be the identity. Using the generalised

detailed balance condition, one can show that, in case  $S = \mathbb{R}^d \times \mathbb{R}^d$ , a probability measure  $\hat{\mu}(\mathrm{d}x \mathrm{d}v) \propto \exp(-H(x, v)) \mathrm{d}x \mathrm{d}v$  with  $H: S \rightarrow \mathbb{R}$  is infinitesimally invariant for the PDMP with characteristics  $b$ ,  $\lambda$  and  $Q$  as soon as the flow satisfies  $\xi_{-t} \circ \Theta = \Theta \circ \xi_t$ , the jump intensity satisfies

$$\lambda \circ \Theta - \lambda = \exp(H) \operatorname{div}(\exp(-H)b), \quad (1.38)$$

and the operator  $(\lambda Q, C_c^1(S))$  satisfies the generalised detailed balance condition (1.37), see [45, Lemma 7.6]. Comparing with (1.25), the deterministic flow associated to the vector field  $b$  alone need not preserve  $\hat{\mu}$ , since  $\operatorname{div}(\exp(-H)b)$  may not vanish; however (1.38) states that this divergence must exactly be compensated by the jumps. Typically,  $H$  is the Hamiltonian given by (1.31), so that  $\hat{\mu}$  is the Boltzmann-Gibbs measure (1.32). Invariance of  $\hat{\mu}$  can then be established by checking (1.38) and showing that  $C_c^1(S)$  forms a core for the generator, and sufficient conditions for this to hold true are given in [92, 137].

**Example 1.6.** In the following, we present some PDMPs that have been introduced in the sampling literature and are relevant to this thesis.

- (i) Randomised Hamiltonian Monte Carlo (RHMC) with refresh rate  $\gamma \in (0, \infty)$ , see [47], introduced above as a stochastic perturbation of the Hamiltonian flow associated to  $H(x, v) = U(x) + \frac{1}{2}|v|^2$ , is also a PDMP with state space  $\mathbb{R}^d \times \mathbb{R}^d$  corresponding to the choice

$$b(x, v) = (v, -\nabla U(x))^\top, \quad \lambda(x, v) = \gamma, \quad Q = \Pi$$

for all  $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$ , where  $\Pi$  is given by (1.34). One easily verifies that in this case,  $\operatorname{div}(\exp(-H)b) = \lambda \circ \Theta - \lambda = 0$ , and RHMC indeed leaves  $\hat{\mu}$  invariant.

- (ii) The bouncy particle sampler (BPS), introduced by [48] and based on [216], runs in straight lines until it “bounces”, when the velocity is reflected at a level set of the potential. To ensure ergodicity, a velocity refreshment with constant rate  $\gamma \in (0, \infty)$  from  $\kappa = \mathcal{N}(0, I_d)$  is added [77, 91]. The resulting PDMP corresponds to the choice  $b(x, v) = (v, 0)^\top$ , the jump rate  $\lambda(x, v) = (v \cdot \nabla U(x))_+ + \gamma$ , and the jump kernel

$$Qf(x, v) = \frac{(v \cdot \nabla U(x))_+}{\lambda(x, v)} Bf(x, v) + \frac{\gamma}{\lambda(x, v)} \Pi f(x, v), \quad f \in C_c^1(S).$$

Here  $Bf(x, v) = f(x, v - 2n(x)n(x)^\top v)$ , where  $n(x) = \frac{\nabla U(x)}{|\nabla U(x)|}$  if  $\nabla U(x) \neq 0$  and  $n(x) = 0$  otherwise, denotes the specular reflection at a level set of  $\nabla U$  and  $\Pi$  is again given by (1.34).

Variants of the BPS are obtained by considering randomised or partial reflections, or restricting to normalised velocities [252]. For example, the forward event chain process introduced in [196] is obtained by replacing the reflection operator  $B$  by the operator

$$B_- f(x, v) = \frac{1}{Z_x} \int_{\mathbb{R}^d} (w \cdot \nabla U(x))_- f(x, w) \kappa(\mathrm{d}w), \quad Z_x = \frac{1}{\sqrt{2\pi}} |\nabla U(x)|.$$

This corresponds to fully resampling of the velocity component in direction  $\nabla U(x)$  instead of simply performing reflections.

- (iii) The zig-zag process [34, 35] is closely related to the bouncy particle sampler, and is obtained by replacing the continuous velocities by velocities drawn from the uniform distribution  $\kappa$  on the discrete hypercube  $\{-1, +1\}^d$  and splitting the gradient of the potential into its components. For functions  $f \in C_c^1(\mathbb{R}^d \times \{-1, +1\}^d)$ , its generator takes the form

$$Lf(x, v) = v \cdot \nabla_x f(x, v) + \sum_{i=1}^d (v_i \partial_i U(x))_+ (f(x, v - 2e_i v_i) - f(x, v)) \\ + \gamma \left( \frac{1}{2^d} \sum_{w \in \{\pm 1\}^d} f(x, w) - f(x, v) \right).$$

We refer to [Section C.5.2](#) for details. It is particularly suited to high-dimensional models with an underlying product structure that, for instance, arise in big data or Bayesian models with priors in product form [34].

- (iv) Event Chain Monte Carlo (ECMC) algorithms are based on piecewise-deterministic Markov processes that only move one particle or coordinate at a time on a straight line between “events” at which a new particle becomes active. They were originally introduced in the context of systems of hard spheres [30] and have since been generalised to many other models [129, 169, 196, 197]. For more details, we refer to [Section E.2](#).

Rates of convergence to stationarity for RHMC are studied in [Chapter A](#), and for RHMC and Langevin dynamics on Riemannian manifolds with specular reflection at the boundary in [Chapter B](#). The forward event chain process as well as variants of the zig-zag process are studied in [Chapter C](#) together with a PDMP with non-trivial boundary behaviour describing a pair of run-and-tumble particles with jamming. Finally, an ECMC algorithm for a chain of harmonic oscillators is studied in [Chapter E](#).



## Chapter 2.

# Second-order lifts and hypocoercivity

In this chapter, we present the approach to hypocoercivity—the study of convergence to equilibrium of stochastic processes with degenerate noise—based on the concept of second-order lifts and a flow Poincaré inequality. The development of this theory in full generality is the main content of Chapters [A](#), [B](#) and [C](#). We present the core ideas of this approach, using the overdamped Langevin diffusion and Langevin dynamics as the main example, which should allow the reader to easily consult the relevant chapters for further details and more general statements where desired. We further point out connections to Chapters [D](#), [E](#) and [F](#) where they arise.

### 2.1. Second-order lifts of reversible diffusions

Motivated by accelerated convergence observed in non-reversible sampling methods, such as Hamiltonian Monte Carlo [\[141\]](#) or a guided walk Metropolis algorithm [\[124\]](#), Diaconis, Holmes and Neal [\[79\]](#) introduced and analysed a model non-reversible Markov chain based on a one-dimensional random walk. They show theoretically and numerically that the mixing and relaxation time is improved by non-reversibility. This led to the development of the concept of lifts of Markov chains, introduced by Chen, Lovász and Pak [\[66\]](#), that provides a systematic approach to non-reversible Markov chains on an enlarged state space that is obtained by splitting each state into several copies. We summarise this concept of lifts and the example of Diaconis, Holmes and Neal in [Section 2.1.1](#). The definition given by Chen, Lovász and Pak relies on a first-order identity between the transition kernels or generators of the processes. In continuous time and space, it turns out that higher-order terms must be considered, leading to the definition of second-order lifts of reversible diffusions in [Section 2.1.2](#). These provide a continuous-time and -space counterpart to lifts of Markov chains that is conceptually linked but analytically very different. A lower bound on the relaxation time achievable using second-order lifts is presented in [Section 2.1.3](#), generalising a related lower bound in [\[66\]](#). This section mainly presents the content of [Chapter A](#) in a summarised way with some further context.

#### 2.1.1. First-order lifts

Let  $S$  and  $\hat{S}$  be Polish state spaces and let  $\pi: \hat{S} \rightarrow S$  be a measurable surjection. Let  $I = \mathbb{N}$  or  $I = [0, \infty)$ . Consider a time-homogeneous Markov process  $(X_t)_{t \in I}$  with invariant probability measure  $\mu$  on  $S$ , and a time-homogeneous Markov process  $(\hat{Z}_t)_{t \in I}$

with invariant probability measure  $\hat{\mu}$  satisfying  $\hat{\mu} \circ \pi^{-1} = \mu$ . The associated transition semigroups and generators on  $L^2(\mu)$  and  $L^2(\hat{\mu})$ , respectively, are denoted by  $(P_t)_{t \geq 0}$  and  $(L, \text{Dom}(L))$ , as well as  $(\hat{P}_t)_{t \geq 0}$  and  $(\hat{L}, \text{Dom}(\hat{L}))$ , respectively. Typically,  $\hat{S} = S \times V$  is a product space,  $\pi(x, v) = x$ , and  $(\hat{Z}_t)_{t \in I} = (\hat{X}_t, \hat{V}_t)_{t \in I}$ . Recall that

$$\mathcal{E}(f, g) = - \int_S (Lf)g \, d\mu$$

denotes the Dirichlet form associated to  $L$ .

**Definition 2.1** (First-order lifts). The Markov process  $(\hat{Z}_t)_{t \in I}$  is a first-order lift of  $(X_t)_{t \in I}$  if

$$\text{Dom}(L) \subseteq \{f \in L^2(\mu) : f \circ \pi \in \text{Dom}(\hat{L})\} \quad (2.1)$$

and for all  $f, g \in \text{Dom}(L)$  we have

$$- \int_{\hat{S}} \hat{L}(f \circ \pi)(g \circ \pi) \, d\hat{\mu} = \mathcal{E}(f, g). \quad (2.2)$$

Conversely, we refer to the underlying process  $(X_t)_{t \geq 0}$  as the first-order collapse of  $(\hat{Z}_t)_{t \geq 0}$ .

It is in fact sufficient to verify (2.1) and (2.2) for a core  $C$  of  $L$  in place of  $\text{Dom}(L)$ . Definition 2.1 of first-order lifts was first given by [66] in the case of discrete-time Markov chains on finite state spaces, where they are simply referred to as lifts. We employ the term ‘first-order’ to distinguish them from the second-order lifts introduced below.

**Remark 2.2.** In the case  $I = \mathbb{N}$  of discrete-time Markov chains, the generators are  $L = p - I$  and  $\hat{L} = \hat{p} - I$  with transition kernels  $p$  and  $\hat{p}$ , respectively. If  $\hat{S} = S \times V$  is a product space,  $\pi(x, v) = x$ , and  $\hat{\mu}(dx \, dv) = \mu(dx) \kappa(x, dv)$  for some probability kernel  $\kappa$  from  $S$  to  $V$ , the domain condition (2.1) is automatically satisfied, and (2.2) simplifies to

$$\int_V \hat{p}((x, v), A \times V) \kappa(x, dv) = p(x, A) \quad (2.3)$$

for all  $x \in S$  and  $A \in \mathcal{B}(S)$ , which is the form in which the definition of first-order lifts was first given in [66]. It is important to note that, even in this case, corresponding identities for the  $n$ -step transition kernels  $\hat{p}^n$  and  $p^n$  do not hold, so that Definition 2.1 is indeed an infinitesimal property.

The following example is a slight variation of the construction by Diaconis, Holmes and Neal [79] that originally motivated the definition of first-order lifts, obtained by replacing reflecting with periodic boundaries.

**Example 2.3.** Consider a symmetric simple random walk  $(X_n)_{n \in \mathbb{N}}$  on the discrete circle  $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$  with  $N \in \mathbb{N}$  odd. As a lift, consider a Markov chain  $(\hat{X}_n, \hat{V}_n)_{n \in \mathbb{N}}$  on the product space  $\mathbb{Z}_N \times \{-1, +1\}$  that moves deterministically in the  $x$ -coordinate with velocity  $v \in \{-1, +1\}$ . After each step, it flips its velocity with probability  $\gamma \in [0, 1]$ . The

## 2.1. Second-order lifts of reversible diffusions

corresponding transition kernels of  $(X_n)_{n \in \mathbb{N}}$  and  $(\hat{X}_n, \hat{V}_n)_{n \in \mathbb{N}}$  are  $p(x, \cdot) = \frac{1}{2}\delta_{x-1} + \frac{1}{2}\delta_{x+1}$  and

$$\begin{aligned}\hat{p}((x, v), (x + v, v)) &= \gamma, \\ \hat{p}((x, v), (x + v, -v)) &= 1 - \gamma,\end{aligned}$$

respectively. A graphical representation of the lifted random walk  $(\hat{X}_n, \hat{V}_n)_{n \in \mathbb{N}}$  is given in Figure 2.1.

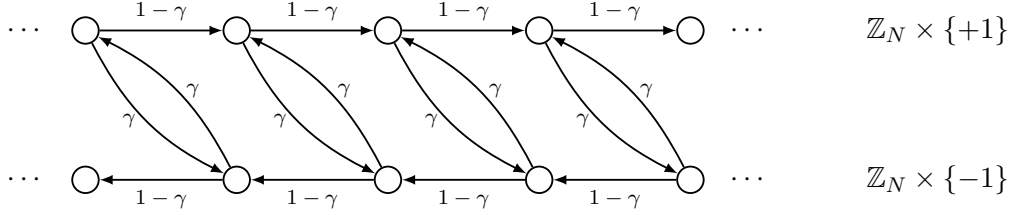


Figure 2.1.: The lifted random walk on  $\mathbb{Z}_N \times \{-1, +1\}$ .

The invariant probability measure of  $(X_n)_{n \in \mathbb{N}}$  is the uniform distribution  $\mu$  on  $\mathbb{Z}_N$ , and the product  $\hat{\mu} = \mu \otimes \text{Unif}(\{-1, +1\})$  is invariant for  $(\hat{X}_n, \hat{V}_n)_{n \in \mathbb{N}}$ . One immediately verifies that for any  $A \subseteq \mathbb{Z}_N$  and  $x \in \mathbb{Z}_N$ ,

$$\frac{1}{2} \sum_{v \in \{\pm 1\}} \hat{p}((x, v), A \times \{\pm 1\}) = \frac{1}{2} \sum_{v \in \{\pm 1\}} \delta_{x+v}(A) = p(x, A),$$

so that  $(\hat{X}_n, \hat{V}_n)_{n \in \mathbb{N}}$  is indeed a first-order lift of the symmetric simple random walk. It is shown in [79] that the mixing and relaxation time of order  $\Theta(N^2)$  for the random walk on  $\mathbb{Z}_N$  can be reduced to  $\Theta(N)$  if the probability of flipping the velocity is chosen as  $\gamma = \frac{1}{N}$ .

**Example 2.4.** Consider an overdamped Langevin diffusion on  $\mathbb{R}^d$  with invariant probability measure  $\mu$  with potential  $U \in C^1(\mathbb{R}^d)$ , perturbed by a divergence-free vector field  $b \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ , as in (1.26), and flip the sign of the vector field with some rate  $\gamma \in [0, \infty)$ . That is, consider

$$dX_t = -\frac{1}{2}\nabla U(X_t)dt + \xi_t b(X_t)dt + dB_t, \quad (2.4)$$

where  $(\xi_t)_{t \geq 0}$  is a continuous-time Markov chain on  $\{-1, +1\}$  that flips its state with rate  $\gamma$ , independent of  $(B_t)_{t \geq 0}$ . The process  $(X_t, \xi_t)_{t \geq 0}$  then admits  $\hat{\mu} = \mu \otimes \kappa$  with  $\kappa = \text{Unif}(\{-1, +1\})$  as invariant probability measure, and the corresponding generator on functions  $f \in C_c^\infty(\mathbb{R}^d \times \{-1, +1\})$  is given by

$$\hat{L}f(x, \xi) = -\frac{1}{2}\nabla_x^* \nabla_x f(x, \xi) + \xi b(x) \cdot \nabla_x f(x, \xi) + \gamma(\Pi f - f)(x, \xi),$$

where  $\Pi f(x, \xi) = \frac{1}{2}(f(x, -1) + f(x, 1))$ . Letting  $\pi(x, \xi) = x$ , functions  $f \circ \pi$  with  $f \in C_c^\infty(\mathbb{R}^d)$  are contained in the domain of  $\hat{L}$ , and

$$\begin{aligned} - \int_{\mathbb{R}^d \times \{\pm 1\}} \hat{L}(f \circ \pi)(g \circ \pi) d\hat{\mu} &= \frac{1}{4} \sum_{\xi \in \{\pm 1\}} \int_{\mathbb{R}^d} (\nabla^* \nabla f + \xi b \cdot \nabla f) g d\mu \\ &= - \int_{\mathbb{R}^d} (Lf) g d\mu, \end{aligned}$$

with  $L = -\frac{1}{2}\nabla^* \nabla$ . Hence  $(X_t, \xi_t)_{t \geq 0}$  is a first-order lift of the overdamped Langevin diffusion (1.19).

### 2.1.2. Second-order lifts

Natural counterparts to Example 2.3 in continuous time and space are a Brownian motion on the circle  $\mathbb{T}_a = \mathbb{R}/a\mathbb{Z}$  of circumference  $a$ , replacing the symmetric simple random walk, and deterministic motion on  $\mathbb{T}_a$  with velocity  $v \in V = \{-1, +1\}$  between velocity flips occurring at rate  $\gamma \in [0, \infty)$ , replacing the lifted random walk. In this case, the corresponding generators of the processes on  $\mathbb{T}_a$  and  $\mathbb{T}_a \times \{-1, +1\}$  are

$$Lf(x) = \frac{1}{2}f''(x) \quad \text{and} \quad \hat{L}f(x, v) = v\partial_x f(x, v) + \gamma(\Pi f - f)(x, v),$$

where  $\Pi f(x, v) = \frac{1}{2}(f(x, -1) + f(x, 1))$ . In analogy to the discrete case, the probability measures  $\mu = \text{Unif}(\mathbb{T}_a)$  and  $\hat{\mu} = \mu \otimes \kappa$  with  $\kappa = \text{Unif}\{-1, +1\}$  are invariant for these processes. One then immediately realises that with  $\pi(x, v) = x$ ,

$$\int_V \hat{L}(f \circ \pi)(x, v) \kappa(dv) = \frac{1}{2} \sum_{v \in \{\pm 1\}} v f'(x) = 0 \quad \text{for all } f \in C^1(\mathbb{T}_a) \text{ and } x \in \mathbb{T}_a,$$

so that a first-order condition analogous to (2.2) cannot be satisfied. The corresponding transition semigroups  $(P_t)_{t \geq 0}$  and  $(\hat{P}_t)_{t \geq 0}$  however satisfy the relation

$$\int_V \hat{P}_t(f \circ \pi)(x, v) \kappa(x, dv) = (P_{t^2} f)(x) + o(t^2) \quad \text{as } t \downarrow 0 \quad (2.5)$$

for all  $f \in \text{Dom}(L)$ . Indeed, this is a direct computation in case  $\gamma = 0$ , and can be verified by conditioning on the jump times if  $\gamma > 0$ , or by checking that the first two time derivatives of the left- and right-hand sides coincide. An analogous relation holds for the transition semigroups  $(P_t)_{t \geq 0}$  of an overdamped Langevin diffusion (1.19) and  $(\hat{P}_t)_{t \geq 0}$  of Langevin dynamics (1.35) or randomised Hamiltonian Monte Carlo (1.33). The relation (2.5) is analogous to (2.3) in the sense that it is an infinitesimal property, yet differs in the linear and quadratic time scaling on the left- and right-hand side. When trying to phrase (2.5) in terms of the infinitesimal generators  $\hat{L}$  and  $L$ , this necessitates considering an expansion of the left-hand side up to second-order terms, see Section A.1, which motivates the following definition of second-order lifts.

Consider the same setup as in [Section 2.1.1](#), where we now restrict ourselves to the case  $I = [0, \infty)$  and assume that  $(X_t)_{t \geq 0}$  is a reversible diffusion. Again, typically  $\hat{S} = S \times V$  is a product space but could also be the tangent bundle of a Riemannian manifold  $S$ , and the components of  $(\hat{Z}_t)_{t \geq 0}$  can be interpreted as position and velocity or momentum.

**Definition 2.5** (Second-order lifts). The process  $(\hat{Z}_t)_{t \geq 0}$  is a second-order lift of  $(X_t)_{t \geq 0}$ , if

$$\text{Dom}(L) \subseteq \{f \in L^2(\mu) : f \circ \pi \in \text{Dom}(\hat{L})\} \quad (2.6)$$

and for all  $f, g \in \text{Dom}(L)$  we have

$$\int_{\hat{S}} \hat{L}(f \circ \pi)(g \circ \pi) d\hat{\mu} = 0 \quad (2.7)$$

and

$$\frac{1}{2} \int_{\hat{S}} \hat{L}(f \circ \pi) \hat{L}(g \circ \pi) d\hat{\mu} = \mathcal{E}(f, g). \quad (2.8)$$

Conversely, we refer to the underlying reversible diffusion  $(X_t)_{t \geq 0}$  as the second-order collapse of  $(\hat{Z}_t)_{t \geq 0}$ .

**Example 2.6.** (i) Many of the processes introduced in [Section 1.3.3](#) and [Section 1.3.4](#) are lifts of an overdamped Langevin diffusion, see [Example A.3](#). In particular, this includes the deterministic Hamiltonian dynamics [\(1.30\)](#), randomised Hamiltonian Monte Carlo [\(1.33\)](#) and Langevin dynamics [\(1.35\)](#) for any choice of  $\gamma \in (0, \infty)$ , as well as the bouncy particle sampler, the forward event chain and the zig-zag process. These all have in common that  $\hat{L}(f \circ \pi)(x, v) = v \cdot \nabla f(x)$  for all functions  $f$  in the domain of the generator of the overdamped Langevin diffusion, which immediately leads to [\(2.7\)](#) and [\(2.8\)](#).

- (ii) Hamiltonian dynamics naturally extend to Riemannian manifolds, and the analogues to RPMC and Langevin dynamics on Riemannian manifolds are second-order lifts of a Riemannian overdamped Langevin diffusion. In presence of a boundary, equipping them with specular reflections at the boundary, they are second-order lifts of a Riemannian overdamped Langevin diffusion with instantaneous reflection at the boundary. These processes are treated in [Chapter B](#).
- (iii) Run-and-tumble particles with jamming are a model from non-equilibrium statistical mechanics that can also be used to model bacterial movement [\[230\]](#). They are a second-order lift of a sticky Brownian motion on an interval, as shown in [Chapter C](#). Their non-trivial boundary behaviour highlights the importance of the domain condition [\(2.6\)](#).
- (iv) Self-repelling random walks and self-repelling diffusions are treated in [Chapter E](#) and [Chapter F](#). These are classes of processes that interact with their history through their local time profiles, encountering a drift that repels them from previously explored regions, so that they are no longer Markovian. However, jointly

considering the local time profile, termed the environment, and the position process yields a Markov process, which is a second-order lift of an Ornstein-Uhlenbeck process with invariant measure corresponding to that of the environment process.

- (v) More examples, such as adaptive or generalised Langevin dynamics, or Langevin dynamics with velocities constrained to the unit sphere, are presented in [Chapter D](#).

### 2.1.3. Lower bound for second-order lifts

An important consequence of the second-order lift property is that it allows to relate the spectrum of the generator of a reversible diffusion to the singular values of an arbitrary second-order lift. In the notation of [Section 2.1.2](#), let  $(P_t)_{t \geq 0}$  and  $(\hat{P}_t)_{t \geq 0}$  be the transition semigroups on  $L^2(\mu)$  and  $L^2(\hat{\mu})$  of a reversible diffusion process and an arbitrary second-order lift, respectively. The second-order lift property (2.8) yields a bound for the singular value gap (1.13) of  $\hat{L}$ , namely

$$\begin{aligned} s(\hat{L}) &= \inf \left\{ \frac{\|\hat{L}f\|_{L^2(\hat{\mu})}}{\|f\|_{L^2(\hat{\mu})}} : f \in L_0^2(\hat{\mu}) \cap \text{Dom}(\hat{L}) \setminus \{0\} \right\} \\ &\leq \inf \left\{ \frac{\|\hat{L}(g \circ \pi)\|_{L^2(\hat{\mu})}}{\|g\|_{L^2(\mu)}} : g \in L_0^2(\mu) \cap \text{Dom}(L) \setminus \{0\} \right\} \\ &= \inf \left\{ \frac{\sqrt{2\mathcal{E}(g)}}{\|g\|_{L^2(\mu)}} : g \in L_0^2(\mu) \cap \text{Dom}(L) \setminus \{0\} \right\} = \sqrt{2 \text{gap}(L)}. \end{aligned} \quad (2.9)$$

Therefore, [Lemma 1.1](#) immediately yields the following lower bound on the relaxation time  $t_{\text{rel}}(\hat{P}, \varepsilon)$ . A similar result with a slightly worse constant is proved in [Theorem A.11](#).

**Theorem 2.7.** *For any  $\varepsilon \in (0, 1)$ , we have*

$$t_{\text{rel}}(\hat{P}, \varepsilon) \geq \frac{1 - \varepsilon}{\sqrt{2 \text{gap}(L)}}.$$

In particular,  $t_{\text{rel}}(\hat{P}) \geq \frac{1-e^{-1}}{\sqrt{2}} \sqrt{t_{\text{rel}}(P)}$ .

Recall that, for a probability measure  $\mu$  on  $\mathbb{R}^d$  satisfying a Poincaré inequality (1.21) with constant  $\frac{1}{m}$ , the reversible overdamped Langevin diffusion (1.19) has relaxation time  $t_{\text{rel}}(P) = \frac{2}{m}$ . [Theorem 2.7](#) hence shows that in this case, any second-order lift of an overdamped Langevin diffusion satisfies

$$t_{\text{rel}}(\hat{P}) \geq (1 - e^{-1})m^{-1/2}. \quad (2.10)$$

In particular, this yields a lower bound on the relaxation times of Langevin dynamics, randomised Hamiltonian Monte Carlo, as well as the piecewise-deterministic Markov processes mentioned in [Example 2.6\(i\)](#).

**Example 2.8.** Recall that for arbitrary Gaussian potentials in  $\mathbb{R}^d$ , critically damped Langevin dynamics with  $\gamma = 2\sqrt{m}$ , where  $m$  is the smallest eigenvalue of the inverse covariance matrix, satisfies  $t_{\text{rel}}(\hat{P}) \leq 2.73m^{-1/2}$ , see [Example 1.5](#). It hence achieves the lower bound of [Theorem 2.7](#) up to a factor of approximately 4.32. We will see in [Section 2.2.3](#) below that, more generally, for probability measures satisfying a Poincaré inequality with constant  $\frac{1}{m}$  with potentials satisfying  $\nabla^2 U \geq -cm$ , where  $c \in [0, \infty)$  is a constant, both Langevin dynamics and RHMC with the critical choice of  $\gamma$  still achieve this lower bound up to a constant factor. In this sense, they are optimal lifts of an overdamped Langevin diffusion.

**Remark 2.9.** A related lower bound for first-order lifts of Markov chains is proved in [[66](#), Theorem 3.1] in terms of the more obscure set-time, which is the worst-case expected time of reaching a set started from stationarity weighted by its stationary measure. [Remark 2.11](#) below shows that the lower bound [Theorem 2.7](#) also holds for first-order lifts of Markov chains, providing a more easily interpretable one than the one in [[66](#)].

**Remark 2.10** (Asymptotic variance of ergodic averages). When using ergodic averages  $A_t(f) = \frac{1}{t} \int_0^t f(X_s) ds$  of a Markov process  $(X_t)_{t \geq 0}$  with invariant probability measure  $\mu$  to approximate integrals  $\mu(f)$  of square-integrable functions, it is desirable to be able to quantify the fluctuations of  $A_t(f)$  around  $\mu(f)$ . In stationarity,  $\mathbb{E}_\mu[A_t(f)] = \mu(f)$ , and the variances of ergodic averages can be bounded using the potential operator  $G$ . Assuming that the transition semigroup satisfies  $\int_0^\infty \|P_s f\|_{L^2(\mu)} ds < \infty$  for all  $f \in L_0^2(\mu)$ , it is the bounded linear inverse of  $-L: L_0^2(\mu) \cap \text{Dom}(L) \rightarrow L_0^2(\mu)$  given by

$$Gf = \int_0^\infty P_s f ds \quad \text{for } f \in L_0^2(\mu).$$

For  $f \in L^2(\mu)$ , the asymptotic variance is then

$$\sigma_f^2 = \lim_{t \rightarrow \infty} t \text{Var}_{\mathbb{P}_\mu}[A_t f] = 2 \langle f - \mu(f), G(f - \mu(f)) \rangle_{L^2(\mu)},$$

and one even has the central limit theorem

$$\frac{1}{\sqrt{t}} \int_0^t (f(X_s) - \mu(f)) ds \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_f^2) \quad \text{under } \mathbb{P}_\mu,$$

see e.g. [[32](#), [153](#)]. Since  $G = (-L|_{L_0^2(\mu)})^{-1}$ , we have  $\|G\|_{B(L_0^2(\mu))} = \frac{1}{s(L)}$ , so that

$$\sup \{ \sigma_f^2 : f \in L^2(\mu), \text{Var}_\mu[f] = 1 \} = \frac{2}{s(L)}$$

and the singular value gap controls the worst-case asymptotic variance in stationarity. Since for a second-order lift,  $s(\hat{L})^{-1} \geq (2 \text{gap}(L))^{-1/2} = (2s(L))^{-1/2}$  by [\(2.9\)](#), the worst-case asymptotic variance can at most be decreased by a square root through lifting, i.e.

$$\sup \{ \hat{\sigma}_f^2 : f \in L^2(\hat{\mu}), \text{Var}_{\hat{\mu}}[f] = 1 \} \geq (\sup \{ \sigma_f^2 : f \in L^2(\mu), \text{Var}_\mu[f] = 1 \})^{1/2}, \quad (2.11)$$

where  $\hat{\sigma}_f^2$  denotes the asymptotic variance of ergodic averages of a second-order lift.

**Remark 2.11** (Lower bounds for first-order lifts). If  $\hat{p}$  is the transition kernel of a Markov chain in discrete or continuous time that is a first-order lift of a Markov chain with transition kernel  $p$ , then the associated generators  $\hat{L} = \hat{p} - I$  and  $L = p - I$  satisfy

$$\begin{aligned} \|\hat{L}(f \circ \pi)\|_{L^2(\hat{\mu})}^2 &= \|f \circ \pi\|_{L^2(\hat{\mu})}^2 - 2\langle f \circ \pi, \hat{p}(f \circ \pi) \rangle_{L^2(\hat{\mu})} + \|\hat{p}(f \circ \pi)\|_{L^2(\hat{\mu})}^2 \\ &\leq 2\|f\|_{L^2(\hat{\mu})}^2 - 2\langle f, pf \rangle_{L^2(\mu)} = 2\langle f, -Lf \rangle_{L^2(\mu)}, \end{aligned}$$

since  $\hat{p}$  is a contraction in  $L^2(\hat{\mu})$ . Therefore, one can conclude that  $s(\hat{L}) \leq \sqrt{2 \operatorname{gap}(L)}$  as above, and obtain bounds analogous to [Theorem 2.7](#) and [Remark 2.10](#).

## 2.2. Hypocoercivity and flow Poincaré inequality

Convergence to equilibrium of degenerate stochastic dynamics, where the noise only acts on a subset of variables, cannot be obtained using a direct argument based on a differential inequality such as (1.6), since the associated generators do not satisfy a Poincaré inequality (1.5), i.e. they are not coercive in the  $L^2$  space with respect to the invariant measure. For example, letting  $(\hat{P}_t)_{t \geq 0}$  denote the transition semigroup on  $L^2(\hat{\mu})$  of Langevin dynamics (1.35) with invariant measure  $\hat{\mu} = \mu \otimes \kappa$ , we have

$$\frac{d}{dt} \|\hat{P}_t f\|_{L^2(\hat{\mu})}^2 = 2\langle \hat{P}_t f, \hat{L} \hat{P}_t f \rangle_{L^2(\hat{\mu})} = -2\gamma \|\nabla_v \hat{P}_t f\|_{L^2(\hat{\mu})}^2 \quad (2.12)$$

for all  $f \in L_0^2(\hat{\mu})$ , which cannot be upper bounded by  $-C \|\hat{P}_t f\|_{L^2(\hat{\mu})}^2$  for some constant  $C > 0$  since the velocity gradient may vanish even when  $\hat{P}_t f$  does not. Nonetheless, exponential convergence to equilibrium still holds, as already shown in [Example 1.5](#) in the Gaussian case.

This behaviour is observed much more generally, beyond the Gaussian case and for various non-reversible, degenerate dynamics, and is facilitated by the interplay between the dissipation provided by the degenerate noise and an antisymmetric transport term. In analogy to the theory of hypoellipticity, where smoothness of solutions to degenerate partial differential equations in the absence of ellipticity can be shown using Hörmander's bracket condition [140], the study of exponential convergence to equilibrium of degenerate stochastic dynamics in the absence of coercivity was termed hypocoercivity by Villani and Gallay [248].

The idea underlying most works in the area of hypocoercivity is to define a new norm or metric equivalent to the one of interest that is well-behaved along the dynamics, so that an exponential decay can be established. Building on work of Hérau and Nier [134] and inspired by [123, 238], Villani achieved this by considering a carefully chosen  $H^1$ -type Sobolev norm involving terms with mixed derivatives [248]. Under commutator conditions resembling those of Hörmander's hypoellipticity theorem, this allowed to unify previous works and to show exponential convergence to equilibrium for many hypocoercive processes. Based on ideas of Hérau [132], Dolbeault, Mouhot and Schmeiser [86, 87] developed an alternative approach based on an  $L^2$ -norm that is again modified by a cross-term. This approach is simpler and widely applicable, and has even been generalised to

various MCMC methods [12]. Nonetheless, while it yields more quantitative bounds, it still fails to capture the optimal dependence on parameters of the invariant measure such as its Poincaré inequality in the case of Langevin dynamics. While probabilistic approaches using coupling techniques do not rely on reversibility, they are challenging in the case of degenerate noise. In some cases, they have successfully been developed using model-specific couplings and carefully designed distance functions, see e.g. [21, 40, 46, 74, 98]. For example, for Langevin dynamics, Eberle, Guillin and Zimmer prove Wasserstein contractivity in a twisted metric using a combination of reflection and synchronous coupling [98]. However, such constructions are model-dependent and typically require strong assumptions on the underlying potential or parameter regimes.

The approaches outlined above have in common that the change in norm or metric relies on a trick, such as twisting or perturbing the norm by some quantity, that is not necessarily intrinsic or natural to the dynamics at hand. Intuitively, the process needs some time for the dissipation present in a subset of variables such as the velocity to pass to the other variables, such as the position. This observation leads to considering time-averages of the  $L^2$ -norm and the Dirichlet form along trajectories of the semigroup. Exponential decay of such time-averages can then be established by a Poincaré-type inequality, termed a flow Poincaré inequality. This approach was initiated by the work of Albritton, Armstrong, Mourrat and Novack [6] for the kinetic Fokker-Planck equation, the Kolmogorov forward equation associated to Langevin dynamics, and later refined more quantitatively and applied to Langevin dynamics by Cao, Lu and Wang in [61].

We present this functional inequality in the following Section 2.2.1. For second-order lifts, such a flow Poincaré inequality can be established using the so-called divergence lemma, a space-time replacement for the Poisson equation that only involves the generator of the underlying reversible diffusion. We present this divergence lemma for overdamped Langevin diffusions on  $\mathbb{R}^d$  in Section 2.2.2, while a more general statement on Riemannian manifolds with boundary is the subject of Chapter B. In Section 2.2.3, we demonstrate how to deduce a flow Poincaré inequality from this divergence lemma for Langevin dynamics, serving as a useful base case, while the development of these ideas into a general framework is the content of Chapter C.

### 2.2.1. Flow Poincaré inequality

Consider a time-homogeneous, right-continuous Markov process with invariant probability measure  $\mu$ , associated transition semigroup  $(P_t)_{t \geq 0}$  on  $L^2(\mu)$  with generator  $(L, \text{Dom}(L))$ , and associated Dirichlet form  $\mathcal{E}$ .

**Definition 2.12** (Flow Poincaré inequality). The semigroup  $(P_t)_{t \geq 0}$  satisfies a flow Poincaré inequality with constant  $\frac{1}{\nu} \in (0, \infty)$  and duration  $T \in (0, \infty)$  if

$$\int_0^T \|P_t f\|_{L^2(\mu)}^2 dt \leq \frac{1}{\nu} \int_0^T \mathcal{E}(P_t f) dt \quad (2.13)$$

for all  $f \in L_0^2(\mu) \cap \text{Dom}(L)$ .

Note that if  $L$  satisfies a Poincaré inequality (1.5) with constant  $\frac{1}{\nu}$ , then (2.13) is satisfied for any  $T \in (0, \infty)$ , so that the flow Poincaré inequality generalises the usual Poincaré inequality. The flow Poincaré inequality can also be written as

$$\langle f, f \rangle_T \leq \frac{1}{\nu} \langle f, -Lf \rangle_T \quad \text{for all } f \in L_0^2(\mu) \cap \text{Dom}(L),$$

where  $\langle f, g \rangle_T = \int_0^T \langle P_t f, P_t g \rangle_{L^2(\mu)} dt$  for  $f, g \in L^2(\mu)$  defines an inner product on  $L^2(\mu)$ . It is thus equivalent to coercivity of the generator with respect to this new inner product, and therefore equivalent to the exponential decay

$$\int_t^{t+T} \|P_s f\|_{L^2(\mu)}^2 ds \leq e^{-2\nu t} \int_0^T \|P_s f\|_{L^2(\mu)}^2 ds \quad \text{for all } t \geq 0 \text{ and } f \in L_0^2(\mu)$$

of time-averaged  $L^2$ -norms, see Theorem C.9. This in turn implies

$$\|P_t f\|_{L^2(\mu)} \leq e^{-\nu(t-T)} \|f\|_{L^2(\mu)} \quad \text{for all } t \geq 0 \text{ and } f \in L_0^2(\mu),$$

and thereby the relaxation time bound

$$t_{\text{rel}}(P, \varepsilon) \leq \frac{\log(\varepsilon^{-1})}{\nu} + T. \quad (2.14)$$

**Remark 2.13.** A related functional inequality, termed a space-time Poincaré inequality, was first proved by Albritton, Armstrong, Mourrat and Novack [6] for the kinetic Fokker-Planck equation, and generalised successfully by Cao, Lu and Wang to Langevin dynamics and some piecewise deterministic Markov processes [61, 186], see Remark 2.19 below. It involves a natural norm in the kinetic solution space, and reduces to the flow Poincaré inequality when applied to solutions of the kinetic Fokker-Planck equation. The formulation (2.13) using time-averages is weaker in this case, yet generalises well to other dynamics.

### 2.2.2. The divergence lemma

The key tool in proving a flow Poincaré inequality for second-order lifts is a space-time divergence lemma. It is a statement solely on the generator of the underlying reversible diffusion. A general statement and proof on Riemannian manifolds with boundary is presented in Chapter B, while, in the following, we focus on the Euclidean case without boundary for the sake of exposition, pointing out which modifications are to be made in the general case.

Let  $(L, \text{Dom}(L))$  denote the generator in  $L^2(\mu)$  of the overdamped Langevin diffusion (1.19) with invariant probability measure  $\mu \in \mathcal{P}(\mathbb{R}^d)$  with density proportional to  $\exp(-U(x))$ .

#### Assumption 2.1.

- (i) **Lower curvature bound.** The potential  $U$  satisfies  $U \in C^2(\mathbb{R}^d)$  and there is a constant  $\rho \in [0, \infty)$  such that

$$v^\top \nabla^2 U(x) v \geq -\rho |v|^2 \quad \text{for all } x, v \in \mathbb{R}^d.$$

## 2.2. Hypocoercivity and flow Poincaré inequality

(ii) **Poincaré inequality.** The probability measure  $\mu$  satisfies a Poincaré inequality with constant  $m^{-1}$ , i.e.

$$\int_{\mathbb{R}^d} f^2 d\mu \leq \frac{1}{m} \int_{\mathbb{R}^d} |\nabla f|^2 d\mu \quad \text{for all } f \in H^{1,2}(\mu) \text{ with } \int_{\mathbb{R}^d} f d\mu = 0.$$

(iii) **Discrete spectrum.** The spectrum of the operator  $(L, \text{Dom}(L))$  on  $L^2(\mu)$  is discrete.

Let  $T \in (0, \infty)$  be fixed,  $\lambda$  denote the Lebesgue measure on  $[0, T]$ , and  $\bar{\nabla} = (\partial_t, \nabla)$  denote the space-time gradient in  $[0, T] \times \mathbb{R}^d$ . For  $m \in \{1, 2\}$ , the Sobolev space with Dirichlet boundary conditions in time  $H_D^{m,2}(\lambda \otimes \mu)$  is the closure of

$$\{u \in C^\infty([0, T] \times \mathbb{R}^d) \cap L^2(\lambda \otimes \mu) : u(0, \cdot) = u(T, \cdot) = 0\}$$

in  $L^2(\lambda \otimes \mu)$  with respect to the norm given by

$$\|u\|_{H_D^{m,2}(\lambda \otimes \mu)}^2 = \sum_{k=0}^m \|\bar{\nabla}^k u\|_{L^2(\lambda \otimes \mu)}^2.$$

**Theorem 2.14** (Quantitative divergence lemma). *Let [Assumption 2.1](#) be satisfied. Then for any  $T \in (0, \infty)$ , there exist constants  $c_0(T), c_1(T) \in (0, \infty)$  such that for any  $f \in L_0^2(\lambda \otimes \mu)$  there are functions*

$$h \in H_D^{1,2}(\lambda \otimes \mu) \quad \text{and} \quad g \in H_D^{1,2}(\lambda \otimes \mu) \quad (2.15)$$

satisfying

$$g(t, \cdot) \in \text{Dom}(L) \quad \text{for } \lambda\text{-almost all } t \in [0, T]$$

such that

$$f = \partial_t h - 2Lg, \quad (2.16)$$

and the functions  $h$  and  $g$  satisfy

$$\begin{aligned} \|h\|_{L^2(\lambda \otimes \mu)}^2 + \|\nabla g\|_{L^2(\lambda \otimes \mu)}^2 &\leq c_0(T) \|f\|_{L^2(\lambda \otimes \mu)}^2, \\ \|\bar{\nabla} h\|_{L^2(\lambda \otimes \mu)}^2 + \|\bar{\nabla} \nabla g\|_{L^2(\lambda \otimes \mu)}^2 &\leq c_1(T) \|f\|_{L^2(\lambda \otimes \mu)}^2, \end{aligned} \quad (2.17)$$

where

$$\begin{aligned} c_0(T) &= 2T^2 + 43 \frac{1}{m} \quad \text{and} \\ c_1(T) &= 290 + \frac{991}{mT^2} + 43 \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \rho. \end{aligned} \quad (2.18)$$

**Remark 2.15.** Let us comment on [Assumption 2.1](#) and [Theorem 2.14](#).

- (i) An analogous divergence lemma holds for the generator of an overdamped Langevin diffusion on Riemannian manifolds, equipped with Neumann boundary conditions in case of a boundary, see [Theorem B.5](#). In this context, the lower bound on the Hessian of the potential in [Assumption 2.1\(i\)](#) is to be replaced by a lower bound on the generalised curvature tensor, namely  $\text{Ric} + \nabla^2 U \geq -\rho$ . This condition is also known as the curvature-dimension condition  $\text{CD}(-\rho, \infty)$ , see [\[20, Definition 1.16.1\]](#). In case of a boundary, an additional assumption on the second fundamental form ensuring local convexity must be made, see [Assumption B.1\(i\)](#).
- (ii) [Assumption 2.1\(ii\)](#) of a Poincaré inequality for the measure  $\mu$  yields a spectral gap and thus exponential decay for the associated overdamped Langevin diffusion, and allows us to conveniently compare the convergence rate to that obtained for Langevin dynamics in the following section. In the absence of a spectral gap, this assumption can however be relaxed to a weak or weighted Poincaré inequality to obtain algebraic convergence to equilibrium, see [\[56, 223\]](#).
- (iii) [Assumption 2.1\(iii\)](#) of a purely discrete spectrum yields a spectral decomposition that allows for simple calculations in the proof of [Theorem 2.14](#). While this assumption was also considered in previous works [\[61, 99\]](#), it is of technical nature and can be relaxed, see [\[55\]](#).
- (iv) The terminology “divergence lemma” stems from the fact that [\(2.16\)](#) can be written in divergence form as

$$f = \bar{\nabla}^* X \quad \text{with} \quad X = \begin{pmatrix} -h \\ \nabla g \end{pmatrix},$$

where the adjoint is in  $L^2(\lambda \otimes \mu)$ . In the classical Lions’ divergence lemma [\[10, 94\]](#) that is of fundamental importance in fluid mechanics [\[184\]](#), the spatial domain is typically bounded and Dirichlet boundary conditions for the vector field are assumed both in time and space. Here, however, in case of bounded domains, the vector field is only assumed to vanish in the direction normal to the boundary, see [Theorem B.5](#), but an additional gradient structure is obtained.

- (v) A key tool in the proof of [Theorem 2.14](#) is the generalised Bochner-Lichnerowicz-Weitzenböck formula

$$L|\nabla u|^2 = |\nabla^2 u|^2 + 2\langle \nabla u, \nabla Lu \rangle + (\text{Ric} + \nabla^2 U)(\nabla u, \nabla u) \quad (2.19)$$

for all  $u \in C^\infty(M)$  that holds for the weighted Laplace-Beltrami operator  $Lu = \frac{1}{2}\Delta u - \frac{1}{2}\langle \nabla U, \nabla u \rangle$  on Riemannian manifolds  $M$  with Ricci curvature tensor  $\text{Ric}$ , see [\[19, Proposition 3\]](#) and [Section B.2](#). In case  $M = \mathbb{R}^d$ , the operator  $L$  is the generator of an overdamped Langevin diffusion with invariant measure  $\mu$  with potential  $U$ , and  $(\text{Ric} + \nabla^2 U)(\nabla u, \nabla u) = \nabla u^\top \nabla^2 U \nabla u$ . Integrating [\(2.19\)](#) by parts with respect to  $\mu$  yields the identity

$$\int_{\mathbb{R}^d} |\nabla^2 u|^2 d\mu = \int_{\mathbb{R}^d} (2Lu)^2 d\mu - \int_{\mathbb{R}^d} \nabla u^\top \nabla^2 U \nabla u d\mu, \quad (2.20)$$

which, together with the lower bound [Assumption 2.1\(i\)](#) on the curvature of  $U$ , can be used to upper bound the  $L^2(\mu)$ -norm of the Hessian of functions by the norm of the generator applied to them, leading to the term involving  $\rho$  in the constant  $c_1(T)$ . On Riemannian manifolds with boundary, an analogous integration-by-parts and careful treatment of the boundary terms leads to the generalised Reilly formula, see [\[187\]](#) and [Lemma B.1](#).

- (vi) For right-hand sides  $f \in L_0^2(\lambda \otimes \mu)$  that are orthogonal to the space-time harmonic functions for  $\partial_t^2 + 2L$ , a solution to [\(2.16\)](#) can be obtained by solving  $(\partial_t^2 + 2L)u = f$  and setting  $h = \partial_t u$  and  $g = -u$ . The bounds [\(2.17\)](#) can then be established in a straightforward way using a Poincaré inequality for the measure  $\lambda \otimes \mu$ . However, for space-time harmonic right-hand sides, these solutions will not satisfy the Dirichlet boundary conditions in time required in [\(2.15\)](#). Therefore, the harmonic functions are considered separately and expanded in the orthogonal spectral basis provided by [Assumption 2.1\(iii\)](#). A careful splitting into high and low modes as well as into symmetric and antisymmetric functions then allows for explicit constructions of solutions on these subspaces. For more intuition on the proof, see the discussion before the proof of [Theorem B.5](#).

### 2.2.3. Flow Poincaré inequality for second-order lifts

The divergence lemma allows to prove a flow Poincaré inequality for second-order lifts. In this section, we demonstrate the argument for Langevin dynamics [\(1.35\)](#). Let  $(\hat{P}_t)_{t \geq 0}$  denote the transition semigroup on  $L^2(\hat{\mu})$  of Langevin dynamics with friction  $\gamma \in (0, \infty)$ , where  $\hat{\mu} = \mu \otimes \kappa$ . Recall that its generator is given by the sum

$$\hat{L}f = \hat{L}_{\text{ham}}f + \gamma(-v \cdot \nabla_v + \Delta_v)f$$

of the antisymmetric generator of the Hamiltonian flow and the symmetric generator of an Ornstein-Uhlenbeck process in the velocity, see [\(1.36\)](#). The associated Dirichlet form is

$$\hat{\mathcal{E}}(f) = \gamma \|\nabla_v f\|_{L^2(\hat{\mu})}^2.$$

**Remark 2.16** (Poisson equation). For the overdamped Langevin diffusion with generator  $L = -\frac{1}{2}\nabla^* \nabla$ , a Poincaré inequality can be established by solving the Poisson equation

$$-2Lg = f$$

for all  $f \in L_0^2(\mu)$  with a quantitative bound  $\|\nabla g\|_{L^2(\mu)}^2 \leq C\|f\|_{L^2(\mu)}^2$ . Then an integration by parts and the Cauchy-Schwarz inequality yield

$$\begin{aligned} \|f\|_{L^2(\mu)}^2 &= \langle f, -2Lg \rangle_{L^2(\mu)} = \langle \nabla f, \nabla g \rangle_{L^2(\mu)} \leq \|\nabla f\|_{L^2(\mu)} \|\nabla g\|_{L^2(\mu)} \\ &\leq C \|\nabla f\|_{L^2(\mu)} \|f\|_{L^2(\mu)} \end{aligned} \tag{2.21}$$

for all  $f \in L_0^2(\mu) \cap \text{Dom}(L)$ , and the Poincaré inequality

$$\int_{\mathbb{R}^d} f^2 d\mu \leq C^2 \int_{\mathbb{R}^d} |\nabla f|^2 d\mu = -2C^2 \int_{\mathbb{R}^d} f Lf d\mu \quad \text{for all } f \in L_0^2(\mu) \cap \text{Dom}(L)$$

follows. The divergence lemma should be seen as a replacement for the Poisson equation within the framework of second-order lifts, since it allows to prove a flow Poincaré inequality for second-order lifts in a manner mirroring (2.21). Indeed, in the proof of Theorem 2.17, after projecting out the velocity variable, the divergence lemma takes on the role of the Poisson equation, and the integration-by-parts is replaced by the second-order lift property. The resulting terms are then bounded using moments of the standard normal distribution and the estimates (2.17) on the functions provided by the divergence lemma.

**Theorem 2.17** (Flow Poincaré inequality for Langevin dynamics). *For any  $T \in (0, \infty)$ , the transition semigroup  $(\hat{P}_t)_{t \geq 0}$  of Langevin dynamics with friction  $\gamma \in (0, \infty)$  satisfies the flow Poincaré inequality*

$$\int_0^T \|\hat{P}_t f_0\|_{L^2(\hat{\mu})}^2 dt \leq \frac{1}{\nu} \int_0^T \hat{\mathcal{E}}(\hat{P}_t f_0) dt \quad \text{for all } f_0 \in L_0^2(\hat{\mu}) \cap \text{Dom}(\hat{L}) \quad (2.22)$$

with

$$\nu = \frac{\gamma}{2 + 4c_1(T) + 2\gamma^2 c_0(T)}, \quad (2.23)$$

where  $c_0(T)$  and  $c_1(T)$  are the constants from Theorem 2.14.

*Proof.* For  $f_0 \in L_0^2(\hat{\mu}) \cap \text{Dom}(\hat{L})$ , letting  $f(t, x, v) = (\hat{P}_t f_0)(x, v)$  yields  $f \in L_0^2(\lambda \otimes \hat{\mu})$ , and the flow Poincaré inequality (2.22) can be written as

$$\|f\|_{L^2(\lambda \otimes \hat{\mu})}^2 \leq \frac{\gamma}{\nu} \|\nabla_v f\|_{L^2(\lambda \otimes \hat{\mu})}^2.$$

Recall the velocity average operator  $\Pi$  given in (1.34), which we extend naturally to  $L^2(\lambda \otimes \hat{\mu})$ . Let  $\tilde{f} \in L_0^2(\lambda \otimes \mu)$  such that  $\tilde{f} \circ \pi = \Pi f$ . An orthogonal decomposition and the Gaussian Poincaré inequality then yield

$$\begin{aligned} \|f\|_{L^2(\lambda \otimes \hat{\mu})}^2 &= \|\Pi f\|_{L^2(\lambda \otimes \hat{\mu})}^2 + \|f - \Pi f\|_{L^2(\lambda \otimes \hat{\mu})}^2 \\ &\leq \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2 + \|\nabla_v f\|_{L^2(\lambda \otimes \hat{\mu})}^2. \end{aligned} \quad (2.24)$$

It thus remains to bound the first term. Let  $h$  and  $g$  be as in Theorem 2.14 such that  $\tilde{f} = \partial_t h - 2Lg$ . By the second-order lift properties (2.7) and (2.8) and an integration by parts in the time variable, we see that

$$\begin{aligned} \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2 &= \langle \tilde{f}, \partial_t h - 2Lg \rangle_{L^2(\lambda \otimes \mu)} \\ &= \langle (-\partial_t + \hat{L})(\tilde{f} \circ \pi), h \circ \pi + \hat{L}(g \circ \pi) \rangle_{L^2(\lambda \otimes \hat{\mu})} \\ &= \langle (-\partial_t + \hat{L})(\tilde{f} \circ \pi - f), h \circ \pi + \hat{L}(g \circ \pi) \rangle_{L^2(\lambda \otimes \hat{\mu})}, \end{aligned}$$

where we used that  $\partial_t f = \hat{L}f$  in the last step. The boundary conditions imposed on  $h$  and  $g$  in Theorem 2.14 ensure that  $h \circ \pi + \hat{L}(g \circ \pi) \in \text{Dom}((-\partial_t + \hat{L})^*)$ , and the

Cauchy-Schwarz inequality yields

$$\begin{aligned}
 \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2 &= \langle (-\partial_t + \hat{L})(\tilde{f} \circ \pi - f), h \circ \pi + \hat{L}(g \circ \pi) \rangle_{L^2(\lambda \otimes \hat{\mu})} \\
 &= \langle \tilde{f} \circ \pi - f, (-\partial_t + \hat{L})^*(h \circ \pi + \hat{L}(g \circ \pi)) \rangle_{L^2(\lambda \otimes \hat{\mu})} \\
 &\leq \|f - \Pi_v f\|_{L^2(\lambda \otimes \hat{\mu})} \|(-\partial_t + \hat{L})^*(h \circ \pi + \hat{L}(g \circ \pi))\|_{L^2(\lambda \otimes \hat{\mu})}.
 \end{aligned} \tag{2.25}$$

The first factor can again be bounded by the Gaussian Poincaré inequality, while the second can be bounded using the estimates (2.17), since only first-order derivatives of  $h$  as well as first- and second-order derivatives of  $g$  appear. Indeed,

$$(-\partial_t + \hat{L})^*(h \circ \pi + \hat{L}(g \circ \pi)) = \partial_t h - v \cdot \nabla h + v \cdot \partial_t \nabla g - v \cdot \nabla^2 g v + \nabla U \cdot \nabla g - \gamma v \cdot \nabla g,$$

and therefore, by orthogonality and since  $\partial_t h - 2Lg = \tilde{f}$ ,

$$\begin{aligned}
 \|(-\partial_t + \hat{L})^*(h \circ \pi + \hat{L}(g \circ \pi))\|_{L^2(\lambda \otimes \hat{\mu})}^2 &= \|\partial_t h - v \cdot \nabla^2 g v + \nabla U \cdot \nabla g\|_{L^2(\lambda \otimes \hat{\mu})}^2 + \|v \cdot (-\nabla h + \partial_t \nabla g - \gamma \nabla g)\|_{L^2(\lambda \otimes \hat{\mu})}^2 \\
 &= \|\tilde{f} + \Delta g - v \cdot \nabla^2 g v\|_{L^2(\lambda \otimes \hat{\mu})}^2 + \|-\nabla h + \partial_t \nabla g - \gamma \nabla g\|_{L^2(\lambda \otimes \mu)}^2.
 \end{aligned}$$

The first summand satisfies

$$\|\tilde{f} + \Delta g - v \cdot \nabla^2 g v\|_{L^2(\lambda \otimes \hat{\mu})}^2 = \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2 + 2\|\nabla^2 g\|_{L^2(\lambda \otimes \mu)}^2$$

while the second can be bounded as

$$\|-\nabla h + \partial_t \nabla g - \gamma \nabla g\|_{L^2(\lambda \otimes \mu)}^2 \leq 4\|\nabla h\|_{L^2(\lambda \otimes \mu)}^2 + 4\|\partial_t \nabla g\|_{L^2(\lambda \otimes \mu)}^2 + 2\gamma^2 \|\nabla g\|_{L^2(\lambda \otimes \mu)}^2,$$

so that

$$\|(-\partial_t + \hat{L})^*(h \circ \pi + \hat{L}(g \circ \pi))\|_{L^2(\lambda \otimes \hat{\mu})}^2 \leq (1 + 4c_1(T) + 2\gamma^2 c_0(T)) \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2.$$

Inserting this into (2.25) and ultimately (2.24) yields

$$\|f\|_{L^2(\lambda \otimes \hat{\mu})}^2 \leq (2 + 4c_1(T) + 2\gamma^2 c_0(T)) \|\nabla_v f\|_{L^2(\lambda \otimes \hat{\mu})}^2,$$

and thus the claim.  $\square$

The proof of [Theorem 2.17](#) works almost word-for-word for randomised HMC. It further translates well to arbitrary second-order lifts, and a general framework along these lines is developed in [Chapter C](#), see also the corresponding summary in [Section 3.3](#) and the main result [Theorem 3.1](#). The key quantity to bound for arbitrary second-order lifts with generator  $\hat{L}$  is the term  $\|\hat{L}^* \hat{L}(g \circ \pi)\|_{L^2(\lambda \otimes \hat{\mu})}$  that appears in (2.25), see Assumption (C1).

**Example 2.18.** To obtain quantitative upper bounds on the relaxation time for Langevin dynamics, a value for  $T$  must be chosen in (2.23). With  $T = 3/\sqrt{m}$ , we have

$$c_0(T) = \frac{51}{m} \quad \text{and} \quad c_1(T) = \frac{3601}{9} + 43 \frac{\rho}{m},$$

and the bound (2.14) yields

$$t_{\text{rel}}(\hat{P}) \leq \frac{1}{\nu} + T = \frac{102}{m} \gamma + \left( \frac{14422}{9} + 172 \frac{\rho}{m} \right) \gamma^{-1} + \frac{3}{\sqrt{m}}.$$

- (i) For convex potentials, one has  $\rho = 0$ , and the upper bound on the relaxation time is minimised by taking  $\gamma_* = \sqrt{14422m/918} \approx 3.96\sqrt{m}$ , yielding

$$t_{\text{rel}}(\hat{P}) \leq \frac{811.58}{\sqrt{m}} \quad \text{for Langevin dynamics with friction } \gamma_*.$$

This is larger than the lower bound  $t_{\text{rel}}(\hat{P}) \geq \frac{1-e^{-1}}{\sqrt{m}}$  obtained from [Theorem 2.7](#) by a factor of approximately 1283.9, so that, on the class of convex potentials, critically damped Langevin dynamics achieves the optimal acceleration of any second-order lift, up to this constant factor. In particular, the relaxation time is of the order  $m^{-1/2}$ , as opposed to the order  $m^{-1}$  for the corresponding overdamped Langevin diffusion.

- (ii) More generally, for the class of potentials satisfying [Assumption 2.1](#) with  $\rho \leq cm$  for some constant  $c > 0$ , choosing the friction  $\gamma$  of the order of the optimal value  $\gamma_*$  from the convex case yields the bound

$$t_{\text{rel}}(\hat{P}) \leq \frac{(404.29 + 43.4c)A + 407.29}{\sqrt{m}} \quad \text{for } \frac{1}{A} \leq \frac{\gamma}{\gamma_*} \leq A, \quad A \in [1, \infty).$$

Thus an optimal acceleration of critically damped Langevin dynamics compared to the overdamped Langevin diffusion can be observed beyond the convex case, as long as the non-convexity measured by the lower bound  $-\rho$  on the Hessian of the potential is of the same order as the inverse Poincaré constant.

**Remark 2.19.** A Poincaré-type inequality in a suitable extended space-time sense was first introduced in [\[6\]](#) and [\[61\]](#) to obtain exponential decay for the kinetic Fokker-Planck equation and Langevin dynamics. They prove that there exists a constant  $C > 0$  such that

$$\|f\|_{L^2(\lambda \otimes \hat{\mu})} \leq C(\|\nabla_v f\|_{L^2(\lambda \otimes \hat{\mu})} + \|(\partial_t - \hat{L}_{\text{ham}})f\|_{L^2(\lambda \otimes \mu)H^{-1}(\kappa)}) \quad (2.26)$$

for a class of functions  $f \in L_0^2(\lambda \otimes \hat{\mu})$  in a suitable kinetic solution space. This functional inequality is termed a space-time Poincaré inequality. Applying it to  $f(t, \cdot) = \hat{P}_t f_0$  for some  $f_0 \in L_0^2(\hat{\mu})$  as above and using that  $(\partial_t - \hat{L}_{\text{ham}})f = -\gamma \nabla_v^* \nabla_v f$  then yields the flow Poincaré inequality as stated here. While the flow Poincaré inequality therefore differs from the space-time Poincaré inequality, as it is only valid along trajectories of the semigroup, this makes it simpler to prove and much more general. Indeed, while the space-time Poincaré inequality is applicable to several kinetic dynamics, as demonstrated in the study of convergence rate of some piecewise deterministic Markov processes [\[186\]](#), the flow Poincaré inequality can be proved in a unified way for second-order lifts, as shown in [Chapter C](#). We demonstrate the versatility of this approach by applying it to a variety of processes in the projects constituting this thesis: After beginning with Langevin dynamics and RHMC, we study piecewise-deterministic Markov processes used in sampling applications, run-and-tumble particles with jamming, self-repellent random walks on the discrete torus, and self-repelling diffusions on Riemannian manifolds. Quantitative bounds on the relaxation time can be derived in all of these cases by applying [Theorem C.12](#).

## Chapter 3.

## Conclusion

Building on the context provided by [Chapters 1 and 2](#), we begin this chapter by including a summary of the contributions of the individual projects constituting this thesis. We remain brief where most of the material has already been covered in [Chapter 2](#) and provide more details where necessary. Finally, we conclude by discussing some open questions and directions for further research suggested by this thesis.

### 3.1. Outline of project A

This project is joint work with Andreas Eberle and included as [Chapter A](#). It has been published as

**A. Eberle and F. Lörler.** ‘Non-reversible lifts of reversible diffusion processes and relaxation times’. In: *Probab. Theory Relat. Fields* 194 (Aug. 2024). DOI: [10.1007/s00440-024-01308-x](https://doi.org/10.1007/s00440-024-01308-x).

The concept of lifts of Markov chains in discrete time introduced in [\[66, 79\]](#), see [Section 2.1.1](#), is crucial to the understanding of accelerated convergence to stationarity of Markov chains due to non-reversibility. A systematic understanding of this phenomenon in continuous time and space was, however, still missing. This project addresses this issue by introducing the concept of second-order lifts of reversible diffusions. The key observation is that, in continuous time and in the notation of [Section 2.1](#), the transition semigroups  $(P_t)_{t \geq 0}$  and  $(\hat{P}_t)_{t \geq 0}$  of a reversible diffusion and a second-order lift should satisfy

$$\int_{\mathcal{V}} \hat{P}_t(f \circ \pi)(x, v) \kappa(x, dv) = (P_{t^2} f)(x) + o(t^2) \quad \text{as } t \downarrow 0$$

for all  $f \in \text{Dom}(L)$ . In contrast, in discrete time, the transition kernels satisfy a similar identity, yet the timescales  $t$  and  $t^2$  are not present. A formulation using the associated generators leads to the formal definition of second-order lifts in [Definition A.1](#). Several non-reversible dynamics used in sampling applications are shown to be second-order lifts of an overdamped Langevin diffusion in [Example A.3](#)

A crucial consequence is that the non-asymptotic relaxation time can at most be reduced by a square root through lifting, i.e.

$$t_{\text{rel}}(\hat{P}) \geq \frac{1 - e^{-1}}{\sqrt{2}} \sqrt{t_{\text{rel}}(P)}$$

for any second-order lift  $(\hat{P}_t)_{t \geq 0}$  of  $(P_t)_{t \geq 0}$ . This result relies on the singular value gap, and has been shown in [Theorem 2.7](#) above, albeit with a slightly better constant than the one shown in [Theorem A.11](#).

Finally, with this lower bound on the relaxation time at hand, we demonstrate in [Section A.5](#) how the variational approach to hypocoercivity introduced by Albritton, Armstrong, Mourrat and Novack [\[6\]](#) can be reformulated and simplified in the language of second-order lifts, leading to quantitative upper bounds on the relaxation time. In particular, we show that, for target measures satisfying a Poincaré inequality with mildly non-convex potential, randomised Hamiltonian Monte Carlo with critically chosen refresh rate achieves the lower bound up to a universal factor, see [Corollary A.27](#). The proof is based on the divergence lemma and the space-time Poincaré inequality and essentially follows [\[61\]](#). It is, however, significantly simplified by the language of second-order lifts, and opens the door to the various applications and generalisations studied in the subsequent projects.

**Contribution of the author.** The idea of trying to systematically understand and generalise the space-time approach introduced by [\[6\]](#) was given to me by my advisor Andreas Eberle. Independently, I realised that for defining lifts of diffusions in continuous time, a second-order expansion must be considered, and that terms occurring in this expansion coincide with those arising in [\[6\]](#). This led to the collaborative development of the concept of second-order lifts and the phrasing of the proof of upper bounds in this language. Technical details and writing were mostly carried out by me with advice and assistance from Andreas Eberle.

## 3.2. Outline of project B

This project is joint work with Andreas Eberle and included as [Chapter B](#). It has been published as

**A. Eberle and F. Lörler.** ‘Space-time divergence lemmas and optimal non-reversible lifts of diffusions on Riemannian manifolds with boundary’. In: *J. Funct. Anal.* **291.9** (2026). DOI: [10.1016/j.jfa.2026.111605](https://doi.org/10.1016/j.jfa.2026.111605).

This project extends the study of second-order lifts to the setting of Riemannian manifolds with boundary and provides a proof of the divergence lemma introduced in [Section 2.2.2](#) in this geometric context. On a complete connected and oriented Riemannian manifold  $(M, g)$  with smooth, potentially empty, boundary  $\partial M$ , we consider an over-damped Langevin diffusion with generator

$$Lf = -\frac{1}{2}\langle \nabla U, \nabla f \rangle + \frac{1}{2}\Delta f$$

and invariant probability measure

$$\mu(dx) = \exp(-U(x))\nu_M(dx),$$

where  $U \in C^2(M)$ ,  $\nu_M$  denotes the Riemannian volume measure on  $M$ , and  $\nabla$  and  $\Delta$  are the Riemannian gradient and Laplace-Beltrami operator on  $M$ . In case  $\partial M \neq \emptyset$ , the generator  $L$  is equipped with Neumann boundary conditions, corresponding to an instantaneous reflection at the boundary.

Arising from the generalised Bochner-Lichnerowicz-Weitzenböck formula

$$L|\nabla u|^2 = |\nabla^2 u|^2 + 2\langle \nabla u, \nabla Lu \rangle + (\text{Ric} + \nabla^2 U)(\nabla u, \nabla u)$$

by an integration-by-parts, we prove the generalised Reilly formula

$$\int_M (|\nabla^2 u|^2 - (2Lu)^2 + (\text{Ric} + \nabla^2 U)(\nabla u, \nabla u)) \, d\mu = \int_{\partial M} B(u) \, d\mu_{\partial M}. \quad (3.1)$$

Here  $\mu_{\partial M}$  is the measure with density  $\exp(-U)$  with respect to the induced boundary measure, and the boundary term  $B(u)$  is non-positive for functions  $u$  satisfying Neumann boundary conditions on manifolds that are locally convex in the sense that the second fundamental form on the boundary is negative semi-definite. In case  $M \subseteq \mathbb{R}^d$  with the Euclidean metric, this condition simply means that  $M$  is convex. Together with [Assumption 2.1](#), where condition (i) is replaced with the lower curvature bound  $\text{Ric} + \nabla^2 U \geq -\rho g$  on the Bakry-Émery tensor, (3.1) shows that

$$\|\nabla^2 u\|_{L^2(\mu)}^2 \leq \|2Lu\|_{L^2(\mu)}^2 + \rho \|\nabla u\|_{L^2(\mu)}^2 \quad \text{for all } u \in \text{Dom}(L).$$

This regularity estimate allows us to prove the space-time divergence lemma on Riemannian manifolds with boundary in [Theorem B.5](#) under the assumption that  $(L, \text{Dom}(L))$  has discrete spectrum on  $L^2(\mu)$ , in analogy to [Theorem 2.14](#) above.

Second-order lifts are then given by randomised Hamiltonian Monte Carlo and Langevin dynamics based on the Hamiltonian dynamics with Hamiltonian  $H(x, v) = U(x) + \frac{1}{2}|v|_x^2$  on the tangent bundle  $\text{TM}$ , undergoing specular reflection at the boundary  $\partial M$  corresponding to the involution

$$(x, v) \mapsto (x, v - 2N(x)\langle N(x), v \rangle), \quad x \in \partial M, v \in \text{T}_x M,$$

where  $N(x)$  denotes the outward-facing normal, see [Section B.4.1](#). We finally show that, as in the unconstrained Euclidean case, upper bounds on the relaxation time can be obtained for these processes using a space-time Poincaré inequality and the framework of second-order lifts. In particular, under [Assumptions B.1](#) and [B.2](#), and assuming that the lower curvature bound  $\rho$  is of the same order as the inverse Poincaré constant of  $\mu$ , both Riemannian Langevin dynamics and Riemannian randomised Hamiltonian Monte Carlo with reflection at the boundary and critical choice of  $\gamma$  achieve the lower bound up to a universal factor, see [Corollary B.17](#).

**Contribution of the author.** Closely examining the proof of the divergence lemma in [\[61\]](#), I realised that there were some gaps and that the order of dependence of the constants on the parameters was not optimal. I had the idea of splitting right-hand sides into symmetric and antisymmetric functions, while Andreas Eberle had the idea

of additionally splitting the right-hand side into high and low modes. The role of the Reilly formula in possibly extending to Riemannian manifolds was pointed out to me by Andreas Eberle. The technical details, writing and generalisation to the case with specular reflection at the boundary using the generalised Reilly formula were then carried out by me with advice from my supervisor Andreas Eberle.

### 3.3. Outline of project C

This project is joint work with Arnaud Guillin, Leo Hahn, Andreas Eberle and Manon Michel and included as [Chapter C](#). It is available as the preprint

**A. Eberle, A. Guillin, L. Hahn, F. Lörler and M. Michel.** *Convergence of non-reversible Markov processes via lifting and flow Poincaré inequality.* Mar. 2025. arXiv: [2503.04238 \[math.AP\]](#).

This project develops the ideas presented in [Chapters A](#) and [B](#) into a general framework for quantitative convergence analysis of second-order lifts of reversible diffusions by providing conditions on the generator of a second-order lift that allow to derive a flow Poincaré inequality. We present the main result [Theorem C.12](#) in the following.

Assume that for any  $\gamma \in [0, \infty)$ , the transition semigroup  $(\hat{P}_t)_{t \geq 0}$  on  $L^2(\hat{\mu})$  with  $\hat{\mu} = \mu \otimes \kappa$  is a second-order lift of  $(P_t)_{t \geq 0}$  with generator

$$\hat{L}^{(\gamma)} f = \hat{L} f + \gamma R f \quad \text{for all } f \in \text{Dom}(\hat{L}^{(\gamma)}), \quad (3.2)$$

where  $\hat{L}$  and  $R$  are typically interpreted as a transport and refreshment term, respectively. Let  $(L, \text{Dom}(L))$  denote the generator of  $(P_t)_{t \geq 0}$  in  $L^2(\mu)$ , and let  $\Pi$  be given by [\(1.34\)](#). Such a splitting is natural in many cases, including the second-order lifts of overdamped Langevin given in [Example 2.6](#). For instance, in the case of Langevin dynamics or randomised Hamiltonian Monte Carlo,  $\hat{L} = \hat{L}_{\text{ham}}$  is the generator of the Hamiltonian flow and  $R = -\nabla_v^* \nabla_v$  is the generator of an Ornstein-Uhlenbeck process in the velocity or  $R = \Pi - I$  corresponds to a full velocity refreshment, respectively.

**Assumption (A).** The operator  $(\hat{L}, \text{Dom}(\hat{L}))$  is weakly antisymmetric, i.e.

$$f \circ \pi \in \text{Dom}(\hat{L}^*) \quad \text{and} \quad \hat{L}^*(f \circ \pi) = -\hat{L}(f \circ \pi)$$

for all  $f \in \text{Dom}(\hat{L})$ .

**Assumption (B).** The operator  $(L, \text{Dom}(L))$  has purely discrete spectrum on  $L^2(\mu)$  and satisfies a Poincaré inequality [\(1.5\)](#) with constant  $\frac{1}{m}$ .

**Assumption (C).** (i) There exist constants  $C_1, C_2 > 0$  such that

$$|\langle \hat{L}(g \circ \pi), \hat{L}(f - \Pi f) \rangle_{L^2(\hat{\mu})}| \leq C_1 \|Lg\|_{L^2(\mu)} \langle f, -Rf \rangle_{L^2(\hat{\mu})}^{1/2}, \quad (3.3)$$

and

$$|\langle \hat{L}(g \circ \pi), Rf \rangle_{L^2(\hat{\mu})}| \leq C_2 \|Lg\|_{L^2(\mu)} \langle f, -Rf \rangle_{L^2(\hat{\mu})}^{1/2}, \quad (3.4)$$

for all  $f \in \text{Dom}(\hat{L}^{(\gamma)})$  and  $g \in \text{Dom}(L)$ .

(ii) The operator  $R$  leaves  $\kappa$  invariant and there exists a constant  $m_v$  such that

$$\int_0^T \|\hat{P}_t f - \Pi \hat{P}_t f\|_{L^2(\hat{\mu})}^2 dt \leq \frac{1}{m_v} \int_0^t \langle \hat{P}_t f, -R \hat{P}_t f \rangle_{L^2(\hat{\mu})} dt \quad (3.5)$$

for all  $f \in \text{Dom}(\hat{L})$ .

The main result of this chapter is [Theorem C.12](#) which can be formulated as the following.

**Theorem 3.1.** *Let Assumptions (A)–(C) be satisfied. Then there exists a universal constant  $C > 0$  such that, for any  $T \in (0, \infty)$  and  $\gamma \in (0, \infty)$ , the transition semigroup  $(\hat{P}_t)_{t \geq 0}$  satisfies the flow Poincaré inequality*

$$\int_0^T \|\hat{P}_t f\|_{L^2(\hat{\mu})}^2 dt \leq \frac{1}{\nu} \int_0^T \hat{\mathcal{E}}(\hat{P}_t f) dt \quad \text{for all } f \in \text{Dom}(\hat{L}^{(\gamma)}) \cap L_0^2(\hat{\mu}),$$

where

$$\nu = C \frac{\gamma}{\gamma^2 C_2^2 + C_1^2 + \frac{1}{m_v} \left(1 + \frac{1}{mT^2}\right)}.$$

The exponential decay

$$\|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq e^{-\nu(t-T)} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } f \in L_0^2(\hat{\mu})$$

then immediately follows by [Theorem C.9](#).

In applications, (3.4) and (3.5) are often directly satisfied if  $R$  is the generator of a Markov chain and thus bounded. For instance, if  $R = \Pi - I$  corresponding to a full velocity refreshment, one can choose  $C_2 = \frac{1}{\sqrt{2m}}$  and  $m_v = 1$ . Furthermore, if  $\hat{L}(g \circ \pi) \in \text{Dom}(\hat{L}^*)$  for all  $g \in \text{Dom}(L)$ , then condition (3.3) can be verified by checking that

$$\|\hat{L}^* \hat{L}(g \circ \pi)\|_{L^2(\hat{\mu})} \leq C_1 m_v^{1/2} \|Lg\|_{L^2(\mu)},$$

as already noted in [Section 2.2.3](#). Note also that (3.5) resembles a flow Poincaré inequality, where the left-hand side is only centred in the velocity, and the right-hand side only uses the Dirichlet form associated to  $R$ . Therefore, (3.5) is of course satisfied if  $R$  satisfies a Poincaré inequality, yet the weaker time-integrated version may be of use in applications where the refreshment term is again degenerate.

The proof of this result is based on the divergence lemma, as for Langevin dynamics and randomised Hamiltonian Monte Carlo in [Chapters A](#) and [B](#). In [Lemma C.16](#), it is formulated slightly differently than in [Theorem B.5](#), enabling the treatment of processes with more general boundary behaviour by bounding the Dirichlet form  $\mathcal{E}$  instead of the gradient.

Finally, we apply this result to some piecewise-deterministic Markov processes: the forward event chain process and the zig-zag process, see [Example 1.6](#), that are second-order lifts of an overdamped Langevin diffusion, and a pair of run-and-tumble particles

(RTPs) with jamming. The latter is a model for the movement of bacteria such as *E. coli* [28, 230] that alternate between straight runs and reorientations (‘tumbles’). Here, in the two-particle case, it consists of two point particles on the one-dimensional circle of circumference  $L$  with velocities in  $\{-1, +1\}$ . The velocities are independent jump processes that switch between the states  $+1$  and  $-1$  with rate  $\omega \in (0, \infty)$ . Between velocity switches, the particles move with their respective velocity, until they collide, causing them to jam and stop moving until a velocity flip allows them to separate. We show that the piecewise-deterministic Markov process corresponding to the distance between the two particles is a second-order lift of a sticky Brownian motion on  $[0, L]$  with stickiness  $1/\omega$ . This observation allows us to deduce exponential convergence to equilibrium of this process with rate of the order  $\frac{\omega}{1+(\omega L)^2}$  via a flow Poincaré inequality, see Corollary C.20.

**Contribution of the author.** We developed the main ideas of the general framework collaboratively with my supervisor Andreas Eberle and co-authors Arnaud Guillin, Manon Michel and Leo Hahn during several research visits in Clermont-Ferrand and Bonn. The core mathematics and presentation of the results as well as the examples were then written by Leo Hahn and myself under the supervision of our advisors Andreas Eberle (FL) and Arnaud Guillin and Manon Michel (LH).

### 3.4. Outline of project D

This project is joint work with Giovanni Brigati and Lihan Wang and included as Chapter D. It has been published as

G. Brigati, F. Lörler and L. Wang. ‘Hypocoercivity meets lifts’. In: *Kinet. Relat. Models* 20.0 (2026), pp. 34–55. DOI: [10.3934/krm.2025020](https://doi.org/10.3934/krm.2025020).

The main goal of this project is to present the approach to hypocoercivity based on second-order lifts developed in Chapters A to C in the context of kinetic equations and in a language more akin to that of the approach by Dolbeault, Mouhot and Schmeiser [86, 87] by focusing on the formulation at the level of the associated Kolmogorov forward equations. Recall that the approach of [87] is based on an  $L^2$ -norm that is perturbed by a cross-term, while the approach using a flow Poincaré inequality and the divergence lemma is based on time-averages.

A key observation is that the assumptions (H2) and (H3) in [87] are equivalent to the second-order lift condition and the existence of a spectral gap of the collapse. While this shows that the two approaches apply to structurally similar models, the new approach yields sharper convergence rates in many examples. Furthermore, the assumption (H1) in [87] corresponds to the assumption of a spectral gap for the refreshment term, as also present in Theorem C.12. Finally, assumption (H4) in [87] concerns the relative boundedness of some auxiliary operators. It is used to bound error terms that occur when differentiating the modified  $L^2$ -norm along solutions. In a similar fashion, bounds

on the solution to the divergence equation and estimates such as those in [Assumption \(C\)](#) are used to bound the terms directly arising from the second-order lift structure.

We present a further application of this framework, namely adaptive Langevin dynamics, introduced by [\[145\]](#). Inspired by the Nosé–Hoover method [\[210\]](#), it views the friction in classical Langevin dynamics as an additional dynamic variable by incorporating a negative feedback loop that encourages the norm of the velocity to be close to  $\sqrt{d}$ , as is the case in stationarity. Given a potential  $U \in C^1(\mathbb{R}^d)$ , it is the solution  $(X_t, V_t, Z_t)_{t \geq 0}$  on  $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$  to the SDE

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= \left( \nabla U(X_t) - \left( \gamma + \frac{Z_t}{\varepsilon} \right) V_t \right) dt + \sqrt{2\gamma} dB_t, \\ dZ_t &= \frac{1}{\varepsilon} (|V_t|^2 - d) dt, \end{aligned}$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion,  $\gamma \in (0, \infty)$  is the friction and  $\varepsilon \in (0, \infty)$  determines the strength of the feedback loop. The use of adaptive Langevin dynamics is motivated by situations in which the computation of the gradient  $\nabla U$  may be corrupted due to noise or approximation errors, see [\[172\]](#) for more details. The invariant probability measure of adaptive Langevin dynamics is

$$\hat{\mu}(dx dv dz) = \mu(dx dz) \kappa(dv),$$

where  $\mu(dx dz) \propto \exp(-U(x) - \frac{1}{2}|z|^2)$  and  $\kappa = \mathcal{N}(0, I_d)$ . By observing that adaptive Langevin dynamics is a second-order lift of an overdamped Langevin diffusion on  $\mathbb{R}^d \times \mathbb{R}$  with invariant probability measure  $\mu$ , we obtain quantitative rates of convergence to stationarity using our framework, see [Section D.3](#). These improve upon those of [\[172\]](#) obtained using the framework of Dolbeault, Mouhot and Schmeiser by weakening the assumptions on  $U$  and making the dependence on the parameters explicit.

Finally, we consider generalised Langevin dynamics, which, instead of adding noise in the velocity component as in Langevin dynamics, considers a position, velocity, and acceleration, and only adds noise in the acceleration. Again given a potential  $U \in C^1(\mathbb{R}^d)$ , it is the solution  $(X_t, V_t, Z_t)_{t \geq 0}$  on  $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$  to the SDE

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= -\nabla U(X_t) dt + \lambda Z_t dt, \\ dZ_t &= -\lambda V_t dt - \gamma Z_t dt + \sqrt{2\gamma} dB_t, \end{aligned}$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion and  $\lambda, \gamma \in (0, \infty)$ . It arises as a Markovian approximation of a more general model for the motion of a particle interacting with a heat bath at equilibrium originating in molecular dynamics, see [\[214\]](#), and was also considered independently as a third-order Langevin dynamics in the sampling community [\[200\]](#), with the hope that the increased smoothness of trajectories yields better control of discretisation errors. Its invariant probability measure is  $\hat{\mu} = \mu \otimes \kappa$  with  $\mu(dx) \propto \exp(-U(x)) dx$  and  $\kappa = \mathcal{N}(0, I_{2d})$ .

A key observation is that the generalised Langevin dynamics is still a second-order lift of an overdamped Langevin diffusion on  $\mathbb{R}^d$  with invariant probability measure  $\mu$ , just as the usual Langevin dynamics is. In particular, if  $\mu$  satisfies a Poincaré inequality with constant  $\frac{1}{m}$ , the lower bound from [Theorem 2.7](#) applies, and the relaxation time cannot improve beyond the order  $m^{-1/2}$ . Since this order is already obtained by Langevin dynamics for a large class of potentials, as discussed in [Example 2.18](#), no substantial further acceleration is possible by passing to the generalised counterpart. Of course, from a numerical viewpoint, the increased smoothness of trajectories is desirable and may still lead to discretisation benefits, see [\[200\]](#). We demonstrate the observation of the lower bound through an explicit computation in the Gaussian case by showing that generalised Langevin dynamics with critical choices of the parameters  $\lambda$  and  $\gamma$  achieves the relaxation time  $t_{\text{rel}} \approx 0.964m^{-1/2}$ , where  $m$  is the smallest eigenvalue of the inverse covariance matrix. In comparison, the critically damped Langevin dynamics achieves the relaxation time  $2.73m^{-1/2}$ , see [Example 1.5](#), reflecting an improvement in the polynomial prefactor in front of the exponential term in the operator norms of the respective transition semigroups. Let us finally note that general upper bounds on the relaxation time of generalised Langevin dynamics can currently not be obtained using this framework, due to the additional level of degeneracy in the noise compared to the usual Langevin dynamics.

**Contribution of the author.** The similarities between the definition of second-order lifts and the assumptions in the approach of Dolbeault, Mouhot and Schmeiser were originally pointed out by Lihan Wang. A precise comparison was then carried out collaboratively during a research visit of Giovanni Brigati and Lihan Wang to Bonn. The proof of hypocoercivity for adaptive Langevin dynamics was mainly done and written down by Giovanni Brigati and Lihan Wang, with assistance from myself. We observed together that generalised Langevin dynamics is a second-order lift of overdamped Langevin, while the explicit computations of the relaxation time in a Gaussian potential were done by me.

### 3.5. Outline of project E

This project is joint work with Andreas Eberle and included as [Chapter E](#). It is available as the preprint

**A. Eberle and F. Lörler.** *Convergence rates of self-repellent random walks, their local time and Event Chain Monte Carlo*. Nov. 2025. [arXiv: 2511.23453 \[math.PR\]](#).

The continuous-time self-repellent random walk is a continuous-time stochastic process  $(X(t))_{t \geq 0}$  on the discrete circle  $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ . Its jump rates are determined by its

projected local time process  $\mathcal{L}(t) = (\mathcal{L}_i(t))_{i \in \mathbb{Z}_n}$ , where

$$\mathcal{L}_i(t) = \int_0^t 1_{\{i\}}(X(s)) ds - \frac{t}{n}, \quad t \in [0, \infty),$$

and are given as the nearest-neighbour jumps

$$\begin{aligned} i \rightarrow i+1 & \quad \text{with rate} \quad q^+(\ell, i) = (\ell_{i+1} - \ell_i)_-, \quad \text{and} \\ i \rightarrow i-1 & \quad \text{with rate} \quad q^-(\ell, i) = (\ell_{i-1} - \ell_i)_-, \end{aligned}$$

where  $\ell$  is the current value of  $\mathcal{L}(t)$ . The jump rates are chosen such that the process prefers to jump towards neighbouring states at which it has spent less time. The subtraction of  $\frac{t}{n}$  from the local time process corresponds to a projection onto the subspace  $\mathcal{S} = \{\ell \in \mathbb{R}^n : \sum_{i=1}^n \ell_i = 0\}$  of mean-zero vectors and allows to identify a process that converges to stationarity. Indeed, while  $(X(t))_{t \geq 0}$  is clearly non-Markovian, the joint process  $(\mathcal{L}(t), X(t))_{t \geq 0}$  is a piecewise-deterministic Markov process on  $\mathcal{S} \times \mathbb{Z}_n$  that admits

$$\hat{\mu} = \mu \otimes \kappa \quad \text{with} \quad \mu = \mathcal{N}(0, -\Delta_n^{-1}) \quad \text{and} \quad \kappa = \text{Unif}(\mathbb{Z}_n)$$

as an invariant probability measure, where  $\Delta_n$  is the periodic discrete Laplacian on  $\mathbb{Z}_n$ , and the inverse is well-defined on the subspace of mean-zero vectors  $\mathcal{S}$  on which  $\mu$  is supported. In particular,  $\mu$  has a density proportional to  $\exp(-U)$  with  $U(\ell) = \frac{1}{2} \sum_{i \in \mathbb{Z}_n} (\ell_{i+1} - \ell_i)^2$  with respect to the  $(n-1)$ -dimensional Haar measure on  $\mathcal{S}$ . Such self-repellent random walks were first introduced as a model for polymer growth by Amit, Parisi and Peliti [9] in discrete time. The continuous-time dynamics was introduced by Tóth and Vető in [242, Section 1.2] on  $\mathbb{Z}$  and termed the continuous-time ‘true’ self-avoiding random walk. In continuous space and time, the true self-repellent motion was constructed by Tóth and Werner [243] as a continuous space counterpart arising as a scaling limit of such self-repellent random walks.

In this project, we study quantitative convergence to equilibrium of  $(\mathcal{L}(t), X(t))_{t \geq 0}$ , i.e. the self-repellent random walk jointly with its local time process. The key insight is the identification of this joint process as a second-order lift of a discrete stochastic heat equation

$$dZ(t) = \frac{1}{2n} \Delta_n Z(t) dt + \frac{1}{\sqrt{n}} dW(t),$$

where  $W$  is a Brownian motion on  $\mathcal{S}$ . The reversible diffusion  $(Z(t))_{t \geq 0}$  has  $\mu$  as invariant probability measure, and the spectral gap of the associated generator in  $L^2(\mu)$  is of the order  $\Theta(n^{-3})$ . As a direct consequence of the lower bound on the  $L^2$ -relaxation time of second-order lifts, see Theorem 2.7, we obtain a lower bound of order  $\Omega(n^{-3/2})$  for the relaxation time of  $(\mathcal{L}(t), X(t))_{t \geq 0}$ , see Theorem E.4.

In order to obtain upper bounds on the relaxation time using the framework of second-order lifts, we consider a modified dynamics that includes additional particle moves occurring with exponential rate  $\gamma \in (0, \infty)$ . If these additional moves correspond to a full resampling of the position of  $X(t)$  from the uniform distribution on  $\mathbb{Z}_n$ , we obtain an upper bound of order  $O(\gamma n^3 + \gamma^{-1} n)$  for the modified dynamics. Choosing  $\gamma \propto n^{-1}$

then yields a relaxation time of the order  $O(n^2)$ . Based on numerical and heuristic evidence [106, 158, 192] and a Bethe ansatz calculation [105] for related models, it has been conjectured in the physics literature that the relaxation time is of order  $O(n^2)$  even for  $\gamma = 0$ . Our bound unfortunately degenerates in the limit  $\gamma \rightarrow 0$ , yet provides one of the first mathematically rigorous results on convergence rates of self-repellent random walks.

Beyond self-repellent random walks, our analysis has implications for non-reversible Markov chain Monte Carlo methods, which have become popular due to their accelerated convergence behaviour. As has recently been observed, the self-repellent random walk and its local time process coincide with a specific non-reversible MCMC method, an Event chain Monte Carlo (ECMC) algorithm applied to a harmonic chain, where  $n$  particles on a ring interact harmonically with their two neighbours. In this context, the local time process corresponds to the particle configuration, and the position of the self-repellent random walk corresponds to the index of an ‘active’ particle. The characteristic property of ECMC methods, namely that only one coordinate or particle changes its position at a time, is reflected by the configuration or local time only increasing in the coordinate corresponding to the active particle.

The derived relaxation time bounds allow to compare the computational complexity of ECMC and more standard MCMC methods such as randomised Hamiltonian Monte Carlo on this toy model. While randomised Hamiltonian Monte Carlo is an optimal lift in the sense that it achieves the lower bound on the relaxation time for second-order lifts, see Section 2.2.3, practical implementations require discretisations of the Hamiltonian flow, leading to a computational complexity of the order  $\Omega(n^{9/4})$  for this model. The ECMC algorithm with refresh rate  $\gamma \propto n^{-1}$ , on the contrary, has computational complexity of order  $\Theta(1)$  per time unit, and thus yields an overall computational complexity of order  $O(n^2)$ . Our results thus rigorously demonstrate a computational advantage of an Event chain Monte Carlo algorithm over randomised Hamiltonian Monte Carlo in this setting. Similar gains have been observed empirically for chains of oscillators with Lennard-Jones interactions [170] and hard sphere models [169], so that this behaviour is conjectured to hold more generally [158].

**Contribution of the author.** The idea of studying rates of convergence for an Event chain Monte Carlo algorithm for the harmonic chain was conceived by Andreas Eberle and myself after attending a talk of Werner Krauth as the Isaac Newton Institute in Cambridge. The precise connection to self-repelling random walks, the details of the proofs and the writing were then carried out collaboratively by Andreas Eberle and myself.

### 3.6. Outline of project F

This project is included as Chapter F and is available as the preprint

**F. Lörler.** *Convergence rates of self-repelling diffusions on Riemannian*

*manifolds*. Nov. 2025. arXiv: 2511.23333 [math.PR].

We study a class of self-interacting diffusions  $(X_t)_{t \geq 0}$  on compact, connected and oriented smooth Riemannian manifolds  $M$ . They solve the stochastic differential equation

$$dX_t = -\nabla_M V_t(X_t) dt + \sigma \circ dB_t, \quad (3.6)$$

where  $\sigma \geq 0$  is the diffusivity,  $V: M \times M \rightarrow \mathbb{R}$  is a smooth and symmetric interaction potential,  $(B_t)_{t \geq 0}$  is a Brownian motion on  $M$  and  $\nabla_M$  denotes the Riemannian gradient on  $M$ . The drift is the negative gradient of the interaction potential accumulated along the trajectory up to time  $t$ , i.e.

$$V_t(x) = \int_0^t V(X_s, x) ds = \int_M V(y, x) \mu_t(dy), \quad \text{where } \mu_t = \int_0^t \delta_{X_s} ds.$$

We assume that  $V$  takes the form

$$V(x, y) = \sum_{k=1}^n a_k e_k(x) e_k(y), \quad (3.7)$$

where  $e_1, \dots, e_n$  is a system of eigenfunctions of the Laplace-Beltrami operator  $\Delta_M$  that is orthonormal with respect to the Riemannian volume measure  $\nu$ , and  $a_1, \dots, a_n \in [0, \infty)$ . This assumption ensures that  $V$  is a positive semi-definite kernel, so that  $(X_t)_{t \geq 0}$  is self-repelling.

Such self-repelling processes were originally introduced as models for polymer growth, notably by Norris, Rogers and Williams [208, 209], where the asymptotic growth of the polymer is of particular interest. In case  $M = \mathbb{R}^d$ , the asymptotic behaviour in the form of strong laws of large numbers has been studied in the works [72, 93, 201]. In case that  $M$  is compact, as considered here, one expects the law of  $X_t$  to converge to an invariant probability measure. Indeed, under a variety of different assumptions, such a convergence has been shown using the normalised occupation time measure  $\frac{1}{t} \mu_t$  in place of  $\mu_t$  in the series of works [25, 26, 27], and with the non-normalised occupation time measure in [23, 24].

In the special case that  $M = \mathbb{T}^d$  is the  $d$ -dimensional torus, if  $\mu_t$  has a density  $L_t$  with respect to the Lebesgue measure and  $V(x, y) = W(x - y)$  for some  $W: \mathbb{T}^d \rightarrow \mathbb{R}$ , the SDE (3.6) simplifies to

$$dX_t = -\nabla(W * L_t)(X_t) + \sigma dB_t,$$

a diffusion with drift given by the mollified negative gradient of its own local time density. In this case, it resembles a continuous counterpart to the self-repellent random walk introduced in Chapter E, suggesting a similar analysis, which motivated the approach taken in this project.

The decomposition (3.7) of  $V$  into finitely many modes allows to reduce the self-repelling diffusion  $(X_t)_{t \geq 0}$  and its so-called environment process  $(\nabla_M V_t)_{t \geq 0}$  to a finite-dimensional process  $(U_t, X_t)_{t \geq 0}$  on  $\mathbb{R}^d \times M$  satisfying a stochastic differential equation

in  $\mathbb{R}^d \times M$ , as originally observed by Benaïm and Gauthier [24]. It admits an invariant probability measure  $\hat{\mu} = \mu \otimes \kappa$  with a Gaussian measure  $\mu$  on  $\mathbb{R}^n$  whose covariance is determined by  $V$  and the uniform distribution  $\kappa$  on  $M$ . Even though the invariant probability measure admits such a simple and explicit expression, obtaining rates of convergence of the process to stationarity is challenging, due to the high degeneracy of  $(U_t, X_t)_{t \geq 0}$ , where the noise only acts directly on the  $M$ -valued component  $X_t$ .

We show that  $(U_t, X_t)_{t \geq 0}$  is a second-order lift of an Ornstein-Uhlenbeck process with invariant measure  $\mu$ , corresponding to that of the environment process, which is to be seen as the analogue to the discrete stochastic heat equation considered in Chapter E. Using the framework of second-order lifts, we obtain an immediate lower bound on the  $L^2$ -relaxation time by Theorem 2.7, see Theorem F.6. Similarly, upper bounds on the relaxation time are then developed using the framework of Chapter C, see Theorem F.7. While hypocoercivity of such self-repelling diffusions was already proved in [24] using the approach [87] of Dolbeault, Mouhot and Schmeiser, our approach is simpler and provides sharper convergence rates. Indeed, we demonstrate in the periodic case  $M = \mathbb{T}_L = \mathbb{R}/(2\pi L\mathbb{Z})$ , that our upper bounds improve on those of [24], and even match the order of the lower bound in some cases.

### 3.7. Outlook and open questions

Let us conclude by mentioning some natural questions motivated by the work presented in this thesis that may lead to interesting directions for further research.

**First-order lifts.** While the theory of first-order lifts significantly predates that of second-order lifts, there is currently no general framework comparable to that presented here that allows to prove quantitative rates of convergence for first-order lifts. Existing results on such convergence rates typically rely on explicit spectral calculations that are only possible in simple models. A development of such a framework would be significant, since discretisations of second-order lifts employed in algorithms typically lead to Markov chains that can be seen as either first-order lifts or mixed first- and second-order lifts.

**Acceleration of discretised schemes.** We have shown in Section 2.2.3 that for strongly convex potentials  $U$  with  $\nabla^2 U(x) \geq m > 0$  for all  $x \in \mathbb{R}^d$ , Langevin dynamics and randomised HMC are optimal lifts in the sense that they achieve a relaxation time of the order  $m^{-1/2}$ . Discretisations typically necessitate a step size to be chosen depending on an upper bound  $L$  on the Hessian of the potential, i.e.  $\nabla^2 U(x) \leq L$  for all  $x \in \mathbb{R}^d$ . For instance, an explicit Euler discretisation of the overdamped Langevin dynamics, leading to the unadjusted Langevin algorithm, requires the step size to be chosen of order  $1/L$ , leading to an overall  $L^2$ -relaxation time of order  $L/m$ . This ratio is known as the generalised condition number and is typically very large in applications, as the eigenvalues of  $\nabla^2 U$  can lie over different orders of magnitude. A natural question is hence whether discretisations of second-order lifts such as randomised HMC or Langevin dynamics can achieve a reduction of the relaxation time to order  $(L/m)^{1/2}$ , which would

present a significant speed-up compared to the reversible case. This question has not yet been answered, with the first sublinear dependence of order  $(L/m)^{5/6}$  on the generalised condition number for a discretisation of Langevin dynamics proved only recently [8].

**Weak confinement.** The work [56] shows how to prove convergence to stationarity for Langevin dynamics in the case of so-called weak confinement, where the invariant probability measure does not satisfy a Poincaré inequality. The approach is similar to the one presented here and applies when the position component of the invariant measure satisfies a weak Poincaré inequality, in which case algebraic convergence to equilibrium for bounded observables can be established. Under the same condition, the overdamped Langevin diffusion relaxes bounded observables to equilibrium algebraically. A natural question is hence if there is a bound on the optimal algebraic convergence achievable by any second-order lift, in analogy to Theorem 2.7.

**Infinite-dimensional settings.** The framework developed in this thesis has currently been applied to finite-dimensional Markov processes. An important future development will be the extension to infinite-dimensional models, for instance governed by stochastic partial differential equations, where aspects such as singularity and renormalisation may play an important role. An alternative direct approach based on works such as [115] to Brownian polymers akin to Chapter F without assuming a finite spectral decomposition of the interaction kernel could provide a first step in this direction.

**Upper bounds for generalised Langevin dynamics.** As noted in Chapter D, the framework developed in this thesis cannot currently be applied to obtain upper bounds on the relaxation time of generalised Langevin dynamics, due to the additional level of degeneracy in the noise. However, in Assumption (C) (ii), the Poincaré inequality in the velocity variables is not assumed to hold pointwise in time but only in average over a time interval. This resembles a flow Poincaré inequality in the velocity variable, where the Dirichlet form appearing on the right-hand side however only corresponds to the generator of the refreshment term and not the full semigroup, and the left-hand side is only centred in the velocity variable. Since the noise in the velocity in the generalised Langevin dynamics is itself an integrated Ornstein-Uhlenbeck process, it is plausible that an inequality such as (3.5) should hold, yet it is currently unclear how to prove it and if such an argument can potentially be iterated.

**Hypocoercivity in relative entropy.** The results of this thesis concern the  $L^2$ -relaxation time. An important but challenging step would be the development of analogous bounds for the relative entropy relaxation time, in analogy to moving from the Poincaré inequality to the logarithmic Sobolev inequality for reversible diffusions. While hypocoercivity in relative entropy has been proved by Villani [248], the developed bounds are not sharp. A natural question is thus whether the idea of time-averaging employed in the  $L^2$ -case can be of use for the relative entropy, and whether a space-time or flow logarithmic Sobolev inequality can be obtained.



# Appendix A.

## Non-reversible lifts of reversible diffusion processes and relaxation times

A. Eberle and F. Lörler. ‘Non-reversible lifts of reversible diffusion processes and relaxation times’. In: *Probab. Theory Relat. Fields* 194 (Aug. 2024). DOI: [10.1007/s00440-024-01308-x](https://doi.org/10.1007/s00440-024-01308-x).

**Abstract:** We propose a new concept of lifts of reversible diffusion processes and show that various well-known non-reversible Markov processes arising in applications are lifts in this sense of simple reversible diffusions. Furthermore, we introduce a concept of non-asymptotic relaxation times and show that these can at most be reduced by a square root through lifting, generalising a related result in discrete time. Finally, we demonstrate how the recently developed approach to quantitative hypocoercivity based on space-time Poincaré inequalities can be rephrased and simplified in the language of lifts and how it can be applied to find optimal lifts.

**Keywords:** Lift; non-reversible Markov process; diffusion process; relaxation time; hypocoercivity; Langevin dynamics; Hamiltonian Monte Carlo.

**MSC Subject Classification:** 60J25, 60J60, 60J35, 60H10, 65C40.

### A.1. Introduction

Recently, there has been growing interest in accelerated convergence to stationarity of Markov processes due to non-reversibility [66, 67, 79, 98, 122, 142, 143, 175, 203]. In parallel, Markov chain Monte Carlo methods based on non-reversible Markov processes have become popular in applications [34, 35, 48, 74, 157, 169, 199, 204]. Despite various important results on convergence of non-reversible Markov processes and diffusive-to-ballistic speed-ups in specific models [46, 61, 98, 122, 175, 186], a systematic understanding of these phenomena is still missing. Indeed, it is not even clear how crucial concepts such as lifts and relaxation times should be adequately defined for general Markov processes in continuous time. The goal of this paper is to propose a concept of non-reversible lifts for reversible diffusion processes and to show that many non-reversible processes considered recently in applications are actually lifts in this sense of very simple reversible diffusions. Moreover, we will introduce a concept of non-asymptotic relaxation times and show how lower and upper bounds on relaxation times of lifts can be derived.

## Appendix A. Non-reversible lifts of reversible diffusion processes and relaxation times

Let us first recall facts from the discrete time case. Here a concept of lifts of Markov chains has been developed by [79] and [66], and a definition of asymptotic relaxation times of non-reversible Markov chains has very recently been given by [65]. Suppose that  $\mathcal{S}$  and  $\mathcal{V}$  are measurable state spaces (usually interpreted as position and velocity),  $\mu$  is a probability measure on  $\mathcal{S}$  and  $\hat{\mu}(\mathrm{d}x \mathrm{d}v) = \mu(\mathrm{d}x)\kappa(x, \mathrm{d}v)$  for some Markov kernel  $\kappa$  from  $\mathcal{S}$  to  $\mathcal{V}$ . In [66], a Markov chain on  $\mathcal{S} \times \mathcal{V}$  with transition kernel  $\hat{p}$  is called a lift of a Markov chain on  $\mathcal{S}$  with transition kernel  $p$  and invariant measure  $\mu$  if for any  $x \in \mathcal{S}$  and any measurable subset  $A \subseteq \mathcal{S}$

$$\int_{\mathcal{V}} \hat{p}((x, v), A \times \mathcal{V}) \kappa(x, \mathrm{d}v) = p(x, A).$$

Equivalently, if  $\pi: \mathcal{S} \times \mathcal{V} \rightarrow \mathcal{S}$ ,  $(x, v) \mapsto x$  denotes the projection on the first coordinate, then for any bounded measurable function  $f: \mathcal{S} \rightarrow \mathbb{R}$  and  $x \in \mathcal{S}$

$$\int_{\mathcal{V}} \hat{p}(f \circ \pi)(x, v) \kappa(x, \mathrm{d}v) = (pf)(x).$$

Conversely, associating a reduced Markov kernel  $p$  to a Markov kernel  $\hat{p}$  in this way is a special case of a procedure known as lumping [58, 147]. Note that the definition of a lift depends only on the one-step transition kernel, i.e. on the instantaneous change, and that corresponding identities do not hold for the  $n$ -step transition kernels. Similarly, in the continuous time case introduced below, corresponding identities will only hold in the infinitesimal limit.

For example, a lift of the symmetric random walk on the discrete circle  $\mathcal{S} = \mathbb{Z}/(n\mathbb{Z})$  is given by a Markov chain on the product space  $\mathcal{S} \times \{+1, -1\}$  that moves deterministically in the  $x$ -coordinate with velocity  $v = +1$  or  $v = -1$ , respectively, and at each step flips its velocity with a given probability  $\varepsilon \in [0, 1]$ . For  $\varepsilon$  of order  $\Theta(\frac{1}{n})$ , the mixing time of the lifted process is of order  $\Theta(n)$  whereas the mixing time of the random walk is of order  $\Theta(n^2)$  if  $n$  is odd and infinite if  $n$  is even [79]. This corresponds to a diffusive to ballistic speed-up of convergence to equilibrium. In general, based on a conductance argument, [66] have shown a lower bound on mixing times of lifted Markov chains, and for reversible Markov chains on finite graphs, they have shown the existence of “optimal” lifts with a mixing time that is at most a constant factor away from the lower bound. However, this construction relies on adding many edges corresponding to non-local interactions to the underlying graph. Such a construction is not very natural and it is not clear how this approach can be extended to the continuous time setting.

As a motivating example for the continuous time case, let us consider Brownian motion on a circle  $\mathbb{T}_a = \mathbb{R}/(a\mathbb{Z})$  with circumference  $a > 0$ . This is a reversible diffusion process, and natural candidates for lifts are piecewise deterministic Markov processes  $(X_t, V_t)$  with state space  $\mathbb{T}_a \times \mathbb{R} = \{(x, v): x \in \mathbb{T}_a, v \in \mathbb{R}\}$  that move deterministically with velocity  $v$  in the first coordinate and possibly update the velocity in the second coordinate after independent exponentially distributed waiting times to new values drawn from  $\kappa = \mathcal{N}(0, 1)$ . These processes have invariant measure  $\hat{\mu} = \mu \otimes \kappa$  where  $\mu = \text{Unif}(\mathbb{T}_a)$ .

Let  $(P_t)_{t \geq 0}$  and  $(\hat{P}_t)_{t \geq 0}$  denote the transition semigroups of Brownian motion and the lifted Markov process on  $L^2(\mathbb{T}_a, \mu)$  and  $L^2(\mathbb{T}_a \times \mathbb{R}, \mu \otimes \kappa)$ , respectively, and suppose that

$(L, \text{Dom}(L))$  and  $(\hat{L}, \text{Dom}(\hat{L}))$  are the corresponding generators. Explicitly,  $Lf(x) = \frac{1}{2}f''(x)$  and  $\hat{L}f(x, v) = v\partial_x f(x, v) + \lambda(\int_{\mathbb{R}} f(x, w)\kappa(dw) - f(x, v))$ . It can be shown in various ways that for any function  $f \in \text{Dom}(L)$  and  $x \in \mathbb{T}_a$ , in the limit  $t \rightarrow 0$ ,

$$\int_{\mathbb{R}} \hat{P}_t(f \circ \pi)(x, v)\kappa(dv) = (P_{t^2}f)(x) + o(t^2), \quad (\text{A.1})$$

or equivalently

$$\lim_{t \rightarrow 0} \frac{1}{t^2} \int_{\mathbb{R}} (\hat{P}_t(f \circ \pi) - f \circ \pi)(\cdot, v)\kappa(dv) = \lim_{t \rightarrow 0} \frac{1}{t^2} (P_{t^2}f - f) = Lf. \quad (\text{A.2})$$

The quadratic and linear scalings in (A.1) are characteristic of diffusive and ballistic motion and play a central role for the second order lifts we consider. To verify (A.1) and (A.2), note that if the jump intensity  $\lambda$  is zero, then  $X_t = x + vt \pmod{a}$ , where  $x$  and  $v$  are the initial position and velocity. If  $v$  is drawn from  $\kappa$ , then  $X_t$  has law  $\mathcal{N}(x, t^2)$  and (A.1) holds without the  $o(t^2)$  error term. For  $\lambda > 0$ , (A.1) can also be verified by an explicit computation of  $\hat{P}_t$  by conditioning on the jump times. Alternatively, (A.1) follows by verifying that the time derivatives up to order two at  $t = 0$  agree on both sides. Indeed,  $f \in \text{Dom}(L)$  implies  $f \circ \pi \in \text{Dom}(\hat{L})$ ,

$$\begin{aligned} \int_{\mathbb{R}} \hat{L}(f \circ \pi)(x, v)\kappa(dv) &= \int_{\mathbb{R}} v f'(x)\kappa(dv) = 0, \quad \text{and} \\ \int_{\mathbb{R}} \hat{L}^2(f \circ \pi)(x, v)\kappa(dv) &= \int_{\mathbb{R}} (v^2 f''(x) - \lambda v f'(x))\kappa(dv) = f''(x) = 2Lf(x). \end{aligned}$$

For the applications we are interested in, it will be useful to rephrase (A.1) and (A.2) in terms of the infinitesimal generators. To this end, we observe that for  $f \in L^2(\mathcal{S}, \mu)$ ,

$$\int_{\mathbb{T}_a \times \mathbb{R}} (\hat{P}_t(f \circ \pi))^2 d\hat{\mu} = \int_{\mathbb{T}_a \times \mathbb{R}} (f \circ \pi)^2 d\hat{\mu} + o(t^2). \quad (\text{A.3})$$

Indeed, this follows from a computation of the first two derivatives, noting that the noise acts only on the velocity and the generator of the deterministic part of the motion is antisymmetric. Therefore, (A.2) implies that for any  $f \in \text{Dom}(L)$  and  $g \in L^2(\mathcal{S}, \mu)$ ,

$$\int_{\mathbb{T}_a \times \mathbb{R}} \hat{L}(f \circ \pi)(g \circ \pi) d\hat{\mu} = \lim_{t \rightarrow 0} \frac{1}{t} \int_{\mathbb{T}_a \times \mathbb{R}} (\hat{P}_t(f \circ \pi) - f \circ \pi)(g \circ \pi) d\hat{\mu} = 0, \quad (\text{A.4})$$

and

$$\begin{aligned} \int_{\mathbb{T}_a \times \mathbb{R}} \hat{L}(f \circ \pi)^2 d\hat{\mu} &= \lim_{t \rightarrow 0} \frac{1}{t^2} \int_{\mathbb{T}_a \times \mathbb{R}} (\hat{P}_t(f \circ \pi) - f \circ \pi)^2 d\hat{\mu} \\ &= \lim_{t \rightarrow 0} \frac{2}{t^2} \int_{\mathbb{T}_a \times \mathbb{R}} (f \circ \pi)(f \circ \pi - \hat{P}_t(f \circ \pi)) d\hat{\mu} = -2 \int_{\mathbb{T}_a} f Lf d\mu. \end{aligned} \quad (\text{A.5})$$

The identities (A.4) and (A.5) motivate the general definition of second-order lifts in Definition A.1 below.

In the following section we will give a formal definition of a second-order lift of a reversible diffusion process. We will then consider examples and show in particular that various well-known non-reversible Markov processes are second-order lifts of overdamped Langevin dynamics. In Section A.3 we introduce a concept of non-asymptotic relaxation times for non-reversible Markov processes in continuous time, and we show that relaxation times in this sense are lower bounded by the inverse of the singular value gap of the generator. Our concept turns out to be related but not identical to the definition of relaxation times for ergodic averages of discrete Markov chains given by Chatterjee in a preprint that appeared during the preparation of this paper [65]. In Section A.4 we show that relaxation times of second-order lifts in our sense are always lower bounded by the square root of the relaxation times of the underlying reversible diffusion process. This generalises a related result in [66] to the continuous time case. Similarly to the discrete time case, we introduce a concept of optimal lifts which achieve the lower bound on the relaxation time up to a constant factor, and we show that in the Gaussian case, both critical Langevin dynamics and critical randomised Hamiltonian Monte Carlo are optimal lifts in this sense. Finally, we show in Section A.5 that the recently developed approach to quantitative hypocoercivity based on space-time Poincaré inequalities [6, 61, 186] can be rephrased and simplified in the language of lifts and be applied to find optimal lifts in more general scenarios.

## A.2. Lifts of reversible diffusions

In the following, we consider a reversible diffusion process with state space  $\mathcal{S}$  and a time-homogenous Markov process with state space  $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{V}$  and invariant probability measures  $\mu(dx)$  and  $\hat{\mu}(dx dv) = \mu(dx)\kappa(x, dv)$ , respectively. We frequently think of the components of the latter as a position and a velocity or momentum variable. The associated Markov semigroups acting on  $L^2(\mu)$  and  $L^2(\hat{\mu})$  are  $(P_t)_{t \geq 0}$  and  $(\hat{P}_t)_{t \geq 0}$ , and their infinitesimal generators are denoted by  $(L, \text{Dom}(L))$  and  $(\hat{L}, \text{Dom}(\hat{L}))$ , respectively. While the underlying diffusion is assumed to be reversible, the second process will not be reversible in general, as an improvement in convergence rate is usually achieved through non-reversibility. The Dirichlet form associated to  $(L, \text{Dom}(L))$  is the extension of

$$\mathcal{E}(f, g) = - \int_{\mathcal{S}} f Lg \, d\mu.$$

to a closed symmetric bilinear form with domain  $\text{Dom}(\mathcal{E})$  given by the closure of  $\text{Dom}(L)$  with respect to the norm  $\|f\|_{L^2(\mu)} + \mathcal{E}(f, f)^{1/2}$ .

Motivated by the introductory example, we define second-order lifts of reversible diffusions in the following way.

**Definition A.1** (Second-order lift).

- (i) The semigroup  $(\hat{P}_t)_{t \geq 0}$  is a *second-order lift* of  $(P_t)_{t \geq 0}$  if

$$\text{Dom}(L) \subseteq \{f \in L^2(\mu) : f \circ \pi \in \text{Dom}(\hat{L})\} \quad (\text{A.6})$$

and for all  $f, g \in \text{Dom}(L)$  we have

$$\int_{\hat{\mathcal{S}}} \hat{L}(f \circ \pi)(g \circ \pi) d\hat{\mu} = 0 \quad (\text{A.7})$$

and

$$\frac{1}{2} \int_{\hat{\mathcal{S}}} \hat{L}(f \circ \pi) \hat{L}(g \circ \pi) d\hat{\mu} = \mathcal{E}(f, g). \quad (\text{A.8})$$

- (ii) The semigroup  $(\hat{P}_t)_{t \geq 0}$  is a *strong second-order lift* of  $(P_t)_{t \geq 0}$  if

$$\text{Dom}(\mathcal{E}) = \{f \in L^2(\mu) : f \circ \pi \in \text{Dom}(\hat{L})\} \quad (\text{A.9})$$

and (A.7) and (A.8) are satisfied for all  $f, g \in \text{Dom}(\mathcal{E})$ .

We also say that  $(\hat{L}, \text{Dom}(\hat{L}))$  is a (strong) second-order lift of  $(L, \text{Dom}(L))$ . Since we only consider second-order lifts, we may also simply refer to second-order lifts as lifts. In [Theorem A.11](#) below, we will give a lower bound on the relaxation time that holds for arbitrary second-order lifts.

**Remark A.2.** (i) In general, a non-reversible Markov process can be a lift in the sense of [Definition A.1](#) (i) of several reversible diffusion processes (for example with different boundary behaviours). On the other hand, if  $(\hat{P}_t)_{t \geq 0}$  is a strong lift of  $(P_t)_{t \geq 0}$ , then the Dirichlet form  $(\mathcal{E}, \text{Dom}(\mathcal{E}))$  and, correspondingly, the semigroup  $(P_t)_{t \geq 0}$  are uniquely determined by  $(\hat{P}_t)_{t \geq 0}$  through the conditions (A.8) and (A.9). In this case, the semigroup  $(P_t)_{t \geq 0}$  is called the *collapse* of  $(\hat{P}_t)_{t \geq 0}$ .

- (ii) In [Definition A.1](#), it is sufficient to verify conditions (A.7) and (A.8) for functions  $f$  and  $g$  from a core of the generator  $(L, \text{Dom}(L))$  or from a core of the associated Dirichlet form. Similarly, it is sufficient to check that

$$\int_{\mathcal{V}} \hat{L}(f \circ \pi)(x, v) \kappa(x, dv) = 0$$

and

$$\frac{1}{2} \int_{\hat{\mathcal{S}}} \left( \hat{L}(f \circ \pi) \right)^2 d\hat{\mu} = \mathcal{E}(f, f)$$

by restricting to the diagonal.

- (iii) In direct generalisation of the notion of lifts in discrete time, the semigroup  $(\hat{P}_t)_{t \geq 0}$  may be called a *first-order lift* of  $(P_t)_{t \geq 0}$  if the associated generators satisfy the conditions  $\text{Dom}(L) \circ \pi \subseteq \text{Dom}(\hat{L})$  and

$$\int_{\hat{\mathcal{S}}} (f \circ \pi) \hat{L}(g \circ \pi) d\hat{\mu} = \int_{\mathcal{S}} f Lg d\mu \quad (\text{A.10})$$

for all  $f, g \in \text{Dom}(L)$ . This is for instance the case if the underlying discrete-time Markov chain corresponding to a pure Markov jump process is a lift in the sense of [66] of another.

- (iv) In practice, it is often convenient to split the generator  $\hat{L}$  into a part that acts only on the velocity variable and a part that acts on both variables, usually the deterministic term in the generator. If the part acting only on the velocity leaves  $\hat{\mu}$  invariant, it is sufficient to check the conditions of Definition A.1 for the latter part of the generator, since  $\hat{L}$  is only applied to functions of the position variable in Definition A.1.

- (v) Assume that the lift  $(\hat{P}_t)_{t \geq 0}$  of  $(P_t)_{t \geq 0}$  satisfies

$$\|\hat{P}_t(f \circ \pi)\|_{L^2(\hat{\mu})}^2 = \|f \circ \pi\|_{L^2(\hat{\mu})}^2 + o(t^2) \quad (\text{A.11})$$

as  $t \rightarrow 0$  and

$$(\hat{P}_t(f \circ \pi), g \circ \pi)_{L^2(\hat{\mu})} = (f \circ \pi, \hat{P}_t(g \circ \pi))_{L^2(\hat{\mu})} \quad (\text{A.12})$$

for all  $f, g \in \text{Dom}(L)$ . The first condition (A.11) is true if  $\hat{L}$  is antisymmetric and, more generally, if the symmetric part acts only on the velocity, as can be verified by computing the first two time derivatives. The second condition (A.12) is true under a generalised detailed balance condition, i.e. that there is a measurable involution  $\Theta: \hat{\mathcal{S}} \rightarrow \hat{\mathcal{S}}$  leaving  $\hat{\mu}$  invariant such that  $\pi \circ \Theta = \pi$  and

$$\int_{\hat{\mathcal{S}}} (f \circ \Theta) \hat{L}(g \circ \Theta) d\hat{\mu} = \int_{\hat{\mathcal{S}}} \hat{L}f g d\hat{\mu}$$

for all  $f, g \in \text{Dom}(\hat{L})$ . Typically  $\Theta(x, v) = (x, -v)$  is the velocity flip, in which case the associated process with initial distribution  $\hat{\mu}$  is invariant under time reversal up to a flip of the velocity [96]. Then

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{1}{t^2} (f, g - P_{t^2}g)_{L^2(\mu)} &= \mathcal{E}(f, g) = \frac{1}{2} (\hat{L}(f \circ \pi), \hat{L}(g \circ \pi))_{L^2(\hat{\mu})} \\ &= \lim_{t \rightarrow 0} \frac{1}{2t^2} (\hat{P}_t(f \circ \pi) - f \circ \pi, \hat{P}_t(g \circ \pi) - g \circ \pi)_{L^2(\hat{\mu})} \\ &= \lim_{t \rightarrow 0} \frac{1}{t^2} (f \circ \pi, g \circ \pi - \hat{P}_t(g \circ \pi))_{L^2(\hat{\mu})}. \end{aligned}$$

In particular,

$$\lim_{t \rightarrow 0} \frac{1}{t^2} (f, P_{t^2}g)_{L^2(\mu)} = \lim_{t \rightarrow 0} \frac{1}{t^2} (f \circ \pi, \hat{P}_t(g \circ \pi))_{L^2(\hat{\mu})} \quad (\text{A.13})$$

for all  $f, g \in \text{Dom}(L)$ . Hence under the above assumptions (A.11) and (A.12),

$$\int_{\mathcal{V}} \hat{P}_t(g \circ \pi)(x, v) \kappa(x, dv) = (P_{t^2}g)(x) + o(t^2) \quad \text{in } L^2(\mu) \text{ as } t \rightarrow 0,$$

is equivalent to the lift conditions (A.7) and (A.8). If  $(\hat{P}_t)_{t \geq 0}$  and  $(P_t)_{t \geq 0}$  are the transition semigroups of stationary Markov processes  $(\hat{X}_t, \hat{V}_t)_{t \geq 0}$  and  $(X_t)_{t \geq 0}$ , respectively, then, in view of (A.13), this is also equivalent to

$$\mathbb{E}[f(\hat{X}_0)g(\hat{X}_t)] = \mathbb{E}[f(X_0)g(X_{t^2})] + o(t^2)$$

for all  $f, g \in \text{Dom}(L)$  as  $t \rightarrow 0$ .

**Example A.3** (Lifts of overdamped Langevin diffusions). Our main running example is provided by lifts of overdamped Langevin diffusions. Given a potential  $U \in C^1(\mathbb{R}^d)$  such that  $U(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$  and  $\int \exp(-U(x)) \, dx < \infty$ , the overdamped Langevin diffusion is the process with state space  $\mathcal{S} = \mathbb{R}^d$  given by the SDE

$$dZ_t = -\frac{1}{2} \nabla U(Z_t) dt + dB_t,$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional standard Brownian motion. One can show that a unique non-explosive strong solution exists for any initial distribution, its invariant measure is the Boltzmann-Gibbs measure

$$\mu(dx) \propto \exp(-U(x)) \, dx,$$

and the generator on  $L^2(\mu)$  is given by

$$Lg = -\frac{1}{2} \nabla U \cdot \nabla g + \frac{1}{2} \Delta g = -\frac{1}{2} \nabla^* \nabla g$$

for all  $g \in C_0^\infty(\mathbb{R}^d)$ , which is a core of the generator [234, 236, 251]. Here  $\nabla^*$  denotes the adjoint of  $\nabla$  in  $L^2(\mu)$ . The following Markov processes are all lifts of overdamped Langevin diffusions:

- (i) *Deterministic Hamiltonian dynamics* associated to the Hamiltonian  $H(x, v) = U(x) + \frac{1}{2}|v|^2$  is the Markov process  $(X, V)$  with state space  $\hat{\mathcal{S}} = \mathbb{R}^d \times \mathbb{R}^d$  given by the ordinary differential equation

$$\begin{aligned} dX_t &= V_t \, dt, \\ dV_t &= -\nabla U(X_t) \, dt, \end{aligned}$$

together with an initial distribution for  $(X_0, V_0)$ . As Hamiltonian dynamics preserves both the volume on  $\hat{\mathcal{S}}$  and the total energy  $H(x, v)$ , the process is non-explosive and the probability measure

$$\hat{\mu}(dx \, dv) = \mu(dx) \kappa(x, dv) \propto \exp(-H(x, v)) \, dx \, dv \quad \text{with } \kappa = \mathcal{N}(0, I_d)$$

is invariant. The generator on  $L^2(\hat{\mu})$  is

$$\hat{L}f(x, v) = v \cdot \nabla_x f(x, v) - \nabla U(x) \cdot \nabla_v f(x, v)$$

with domain  $\text{Dom}(\hat{L}) \supseteq \{f \in H_{\text{loc}}^{1,2}(\mathbb{R}^d \times \mathbb{R}^d) \cap L^2(\hat{\mu}) : \hat{L}f \in L^2(\hat{\mu})\}$ . Clearly we have  $C_0^\infty(\mathbb{R}^d) \circ \pi \subseteq \text{Dom}(\hat{L})$  and for  $f \in C_0^\infty(\mathbb{R}^d)$ ,

$$\int_{\mathbb{R}^d} \hat{L}(f \circ \pi) d\kappa = \int_{\mathbb{R}^d} v \cdot \nabla f \kappa(dv) = 0$$

as well as

$$\begin{aligned} \frac{1}{2} \int_{\mathbb{R}^{2d}} (\hat{L}(f \circ \pi))^2 d\hat{\mu} &= \frac{1}{2} \int_{\mathbb{R}^{2d}} (v \cdot \nabla f)^2 d\hat{\mu} = \frac{1}{2} \int_{\mathbb{R}^d} \nabla f^\top \int_{\mathbb{R}^d} vv^\top \kappa(dv) \nabla f d\mu \\ &= \frac{1}{2} \int_{\mathbb{R}^d} \nabla f \cdot \nabla f d\mu = \frac{1}{2} \int_{\mathbb{R}^d} f \nabla^* \nabla f d\mu = - \int_{\mathbb{R}^d} f Lf d\mu. \end{aligned}$$

Hence Hamiltonian dynamics is a second-order lift of the overdamped Langevin diffusion.

- (ii) *Langevin dynamics* is the diffusion process  $(X, V)$  solving the SDE

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= -\nabla U(X_t) dt - \gamma V_t dt + \sqrt{2\gamma} dB_t, \end{aligned}$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion and the friction parameter  $\gamma$  is a non-negative constant, see e.g. [215]. Its invariant measure is  $\hat{\mu} = \mu \otimes \kappa$ . Its generator is given by  $\hat{L}^{(\gamma)} = \hat{L} + \gamma L_{\text{OU}}$  where  $L_{\text{OU}} = -v \cdot \nabla_v + \Delta_v$  is the generator of an Ornstein-Uhlenbeck process in the velocity variable which again leaves  $\hat{\mu}$  invariant. Since  $L_{\text{OU}}(f \circ \pi) = 0$  for all  $f \in \text{Dom}(L)$ , Langevin dynamics is a lift of the overdamped Langevin diffusion for any choice of the friction parameter  $\gamma \geq 0$ .

- (iii) *Randomised Hamiltonian Monte Carlo (RHMC)* runs the Hamiltonian dynamics for an exponentially distributed time  $T \sim \text{Exp}(\gamma)$  and then performs a complete velocity refreshment according to  $\kappa = \mathcal{N}(0, I_d)$ , see [47]. It leaves invariant the probability measure  $\hat{\mu} = \mu \otimes \kappa$ , and its generator on  $L^2(\hat{\mu})$  is given by  $\hat{L}^{(\gamma)} = \hat{L} + R$  where  $R = \gamma(\Pi_v - I)$  with refresh rate  $\gamma \geq 0$  and

$$(\Pi_v f)(x, v) = \int_{\mathbb{R}^d} f(x, w) \kappa(dw).$$

The associated equation  $\partial_t f = (\hat{L}^{(\gamma)})^* f$ , where the adjoint is with respect to  $\hat{\mu}$ , modelling the motion of a particle influenced by an external potential  $U$  and by random collisions with other particles with Gaussian velocities [198], is also known as the linear Boltzmann equation in statistical physics, see e.g. [128, 132]. Clearly  $R$  leaves  $\hat{\mu}$  invariant. Again we see that  $R(f \circ \pi) = 0$  for all  $f \in \text{Dom}(L)$ , so that RHMC is a lift of the overdamped Langevin diffusion for any refresh rate  $\gamma \geq 0$ .

- (iv) The *Bouncy Particle Sampler (BPS)*, see e.g. [48], runs in a straight line until its velocity is reflected at a level set of the potential  $U$ . To ensure ergodicity, one adds

velocity refreshments according to  $\kappa = \mathcal{N}(0, I_d)$  with rate  $\gamma$  as for RHMC. Under additional assumptions, see [92],  $C_0^\infty(\mathbb{R}^d \times \mathbb{R}^d)$  is a core for the resulting generator

$$\hat{L}^{(\gamma)} f = v \cdot \nabla_x f + (v \cdot \nabla U(x))_+ (B - I) f + \gamma(\Pi_v - I) f.$$

Here  $Bf(x, v) = f(x, v - 2n(x)n(x)^\top v)$  where  $n(x) = \frac{\nabla U(x)}{|\nabla U(x)|}$  if  $\nabla U(x) \neq 0$  and  $n(x) = 0$  otherwise. The probability measure  $\hat{\mu} = \mu \otimes \kappa$  is invariant for this process. Since  $\hat{L}^{(\gamma)}(f \circ \pi) = v \cdot (\nabla f) \circ \pi$ , one again immediately sees that the BPS is a lift of the overdamped Langevin diffusion.

Similarly, other piecewise deterministic Markov processes introduced recently in the MCMC literature are also lifts of an overdamped Langevin diffusion, for example Event Chain Monte Carlo [30, 157, 197] and the Zig-zag process [34, 35, 186].

**Example A.4** (Lifts of overdamped Riemannian Langevin diffusions). We equip  $\mathbb{R}^d$  with a Riemannian metric  $g: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  which for any  $x \in \mathbb{R}^d$  defines the inner product

$$\langle v, w \rangle_x = v \cdot g(x)w$$

for all  $v, w \in \mathbb{R}^d$ . We write

$$\begin{aligned} \text{grad } f &= g^{-1} \nabla f, \\ \text{div } F &= \nabla \cdot F + \frac{1}{2} \sum_{j=1}^d \text{tr} (g^{-1} \partial_j g) F_j, \end{aligned}$$

for the gradient of a smooth function  $f$  and the divergence of a smooth vector field  $F$  on the Riemannian manifold  $(\mathbb{R}^d, g)$ . By Jacobi's formula, these operators satisfy the integration by parts identity

$$\int_{\mathbb{R}^d} \langle \text{grad } h, F \rangle d\nu_g = \int_{\mathbb{R}^d} \nabla h \cdot F d\nu_g = - \int_{\mathbb{R}^d} h \text{div } F d\nu_g$$

for all  $h \in C_0^\infty(\mathbb{R}^d)$  and  $F \in C_0^\infty(\mathbb{R}^d, \mathbb{R}^d)$ , where  $\nu_g(dx) = \sqrt{\det g(x)} dx$  is the associated Riemannian volume measure.

Considering probability measures  $\mu(dx) \propto \exp(-U(x))\nu_g(dx)$  for some continuously differentiable potential  $U: \mathbb{R}^d \rightarrow \mathbb{R}$  leads to the adjoint

$$\text{grad}^*(F) = -\exp(U) \text{div}(F \exp(-U)) = -\text{div } F + \langle \text{grad } U, F \rangle$$

of grad with respect to  $\mu$ . The overdamped Riemannian Langevin diffusion is the solution to the SDE

$$dX_t = \frac{1}{2} \left( -(\text{grad } U)(X_t) + (\text{div } g^{-1})(X_t) \right) dt + \sqrt{g^{-1}(X_t)} dB_t, \quad (\text{A.14})$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional standard Brownian motion and the divergence is applied row-wise to  $g^{-1}$ . Under appropriate assumptions on  $U$  and  $g$ , the process is non-explosive

## Appendix A. Non-reversible lifts of reversible diffusion processes and relaxation times

and reversible with respect to its invariant measure  $\mu$ , see [20, 113]. On smooth compactly supported functions, the generator of the corresponding transition semigroup on  $L^2(\mu)$  is given by

$$\begin{aligned} Lf &= \frac{1}{2} \langle -\operatorname{grad} U + \operatorname{div} g^{-1}, \operatorname{grad} f \rangle + \frac{1}{2} g^{-1} : \nabla^2 f \\ &= \frac{1}{2} \operatorname{div} \operatorname{grad} f - \frac{1}{2} \langle \operatorname{grad} U, \operatorname{grad} f \rangle = -\frac{1}{2} \operatorname{grad}^*(\operatorname{grad} f). \end{aligned}$$

Interpreting  $g$  as a position-dependent mass matrix leads to a Hamiltonian dynamics associated to the Hamiltonian

$$H(x, v) = U(x) + \frac{1}{2} v^\top g^{-1}(x) v$$

on  $\mathbb{R}^d \times \mathbb{R}^d$ , where  $(x, v)$  are position and momentum. It reads

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix} \nabla H(x, v) = \begin{pmatrix} g^{-1}(x) v \\ -\nabla U(x) - \frac{1}{2} \nabla_x (v^\top g^{-1}(x) v) \end{pmatrix}. \quad (\text{A.15})$$

By construction it leaves the measure  $\hat{\mu}(\mathrm{d}x \mathrm{d}v) = \mu(\mathrm{d}x) \mathcal{N}(0, g(x))(\mathrm{d}v)$  invariant. For  $f \in C_0^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ , its generator on  $L^2(\hat{\mu})$  is given by

$$\hat{L}f(x, v) = v^\top g^{-1}(x) \nabla_x f(x, v) - \nabla U(x)^\top \nabla_v f(x, v) - \frac{1}{2} \left( \nabla_x (v^\top g^{-1}(x) v) \right)^\top \nabla_v f(x, v).$$

Denoting  $\kappa(x, \cdot) = \mathcal{N}(0, g(x))$  we have  $\hat{L}(f \circ \pi)(x, v) = v^\top \operatorname{grad} f(x)$  and thus

$$\int_{\mathbb{R}^d} \hat{L}(f \circ \pi)(x, v) \kappa(x, \mathrm{d}v) = 0,$$

and

$$\begin{aligned} \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} (\hat{L}(f \circ \pi))^2 \mathrm{d}\hat{\mu} &= \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \operatorname{grad} f(x)^\top v v^\top \operatorname{grad} f(x) \kappa(x, \mathrm{d}p) \mu(\mathrm{d}x) \\ &= \frac{1}{2} \int_{\mathbb{R}^d} \langle \operatorname{grad} f, \operatorname{grad} g \rangle \mathrm{d}\mu = - \int_{\mathbb{R}^d} f Lf \mathrm{d}\mu \end{aligned}$$

for all  $f \in C_0^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ . Hence the Hamiltonian dynamics (A.15) is a lift of the overdamped Riemannian Langevin diffusion (A.14).

As in Example A.3, the Hamiltonian dynamics can be modified by passing to the generator  $\hat{L} + \gamma R$ , where  $\gamma \geq 0$  and  $R$  acts only the momentum and leaves  $\kappa(x, \mathrm{d}v)$  invariant. The choice  $R = -\nabla_v^* \nabla_v = \Delta_v - g^{-1} \nabla_v$ , where the adjoint is with respect to  $\hat{\mu}$ , leads to Riemannian Langevin dynamics given by the solution  $(X_t, V_t)$  to the SDE

$$\begin{aligned} \mathrm{d}X_t &= g^{-1}(X_t) V_t \mathrm{d}t, \\ \mathrm{d}V_t &= \left( -\nabla U(X_t) - \frac{1}{2} \nabla_x (V_t^\top g^{-1}(X_t) V_t) \right) \mathrm{d}t - \gamma g^{-1}(X_t) V_t \mathrm{d}t + \sqrt{2\gamma} \mathrm{d}B_t. \end{aligned}$$

By the same argument as above, this is again a lift of the underlying overdamped Riemannian Langevin diffusion.

### A.3. Relaxation times of non-reversible Markov processes

Let  $L_0^2(\mu) = \{f \in L^2(\mu) : \int f d\mu = 0\}$  be the orthogonal complement of the constant functions. The transition semigroup  $(P_t f)(x) = \int f(y) p_t(x, dy)$  of a *reversible* Markov process with invariant measure  $\mu$  satisfies

$$\|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)} = \sup_{f \in L_0^2(\mu) \setminus \{0\}} \frac{\|P_t f\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}} = e^{-\lambda t}, \quad (\text{A.16})$$

with the decay rate  $\lambda$  given by the spectral gap

$$\text{gap}(L) = \inf\{\text{Re}(\alpha) : \alpha \in \text{spec}(-L|_{L_0^2(\mu)})\} \quad (\text{A.17})$$

of the generator  $(L, \text{Dom}(L))$  of  $(P_t)_{t \geq 0}$  in the Hilbert space  $L^2(\mu)$ . Equivalently, the  $\chi^2$ -divergence  $\chi^2(\eta p_t | \mu)$  between the law  $\eta p_t$  of the process at time  $t$  with initial law  $\eta$  and the invariant measure  $\mu$  decays exponentially with rate  $\lambda$  if the initial  $\chi^2$ -divergence  $\chi^2(\eta | \mu)$  is finite. Therefore, in this reversible case, decay by a factor of  $\varepsilon > 0$  has taken place after time  $\frac{1}{\lambda} \log(\varepsilon^{-1})$ , so that it is natural to define the  $L^2$ -relaxation time  $t_{\text{rel}}$  as the inverse of the spectral gap [179, 227].

In contrast, for *non-reversible* Markov processes, the operator norm  $\|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)}$  is no longer a pure exponential, and one can only expect bounds of the form

$$\|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)} \leq C e^{-\nu t} \quad (\text{A.18})$$

with constants  $C \in [1, \infty)$  and  $\nu \leq \text{gap}(L)$ , see e.g. [104]. This motivates the following definition of an  $\varepsilon$ -relaxation time.

**Definition A.5** (Non-asymptotic relaxation time). Let  $\varepsilon \in (0, 1)$ . The  $\varepsilon$ -relaxation time  $t_{\text{rel}}(\varepsilon)$  of  $(P_t)_{t \geq 0}$  is

$$t_{\text{rel}}(\varepsilon) = \inf\{t \geq 0 : \|P_t f\|_{L^2(\mu)} \leq \varepsilon \|f\|_{L^2(\mu)} \text{ for all } f \in L_0^2(\mu)\}$$

and the *relaxation time* is  $t_{\text{rel}} = t_{\text{rel}}(e^{-1})$ .

**Remark A.6** (Asymptotic decay rate and relaxation times).

- (i) If  $\|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)} \in (0, 1)$  for all  $t > 0$ , then the  $\varepsilon$ -relaxation time as a function  $\varepsilon \mapsto t_{\text{rel}}(\varepsilon)$  is the inverse function of  $t \mapsto \|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)}$ . Therefore, the asymptotic decay rate

$$\omega = -\lim_{t \rightarrow \infty} \frac{1}{t} \log \|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)}$$

of the semigroup restricted to  $L_0^2(\mu)$  can be expressed as

$$\frac{1}{\omega} = \lim_{\varepsilon \rightarrow 0} \frac{t_{\text{rel}}(\varepsilon)}{\log(\varepsilon^{-1})}, \quad (\text{A.19})$$

where the limit exists since  $t \mapsto \log \|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)}$  is non-increasing and sub-additive, see [104, Proposition IV.2.2]. Hence  $\omega^{-1}$  can also be interpreted as an

*asymptotic relaxation time.* The relation  $\text{gap}(L) = \omega$  only holds true under additional assumptions ensuring a spectral mapping theorem relating the spectrum of  $P_t$  to that of  $L$ , see [104]. In general, only  $\text{gap}(L) \geq \omega$  is true. More importantly, the spectral gap cannot be used to bound the constant  $C$  in (A.18) and can hence only provide asymptotic estimates, whereas  $t_{\text{rel}}(\varepsilon)$  is a non-asymptotic quantity.

- (ii) For reversible processes, (A.16) shows  $t_{\text{rel}}(\varepsilon) = \log(\varepsilon^{-1})/\text{gap}(L)$  and  $t_{\text{rel}}$  hence coincides with the usual definition of the relaxation time as the inverse of the spectral gap. For non-reversible processes, the semigroup property only yields the upper bounds

$$t_{\text{rel}}(\delta) \leq t_{\text{rel}}(\varepsilon) \cdot \frac{\log(\delta^{-1})}{\log(\varepsilon^{-1})} \quad \text{for } \delta \leq \varepsilon.$$

In particular, the limit in (A.19) is monotone, and thus

$$\frac{1}{\omega} \leq \frac{t_{\text{rel}}(\varepsilon)}{\log(\varepsilon^{-1})} \quad \text{for any } \varepsilon \in (0, 1).$$

To obtain quantitative upper bounds on the relaxation time, it will be convenient to reformulate (A.18) in the following way.

**Definition A.7** (Delayed exponential contractivity). Let  $T, \nu \in (0, \infty)$  and let  $(P_t)_{t \geq 0}$  be a Markov semigroup with invariant probability measure  $\mu$ .

- (i) The semigroup  $(P_t)_{t \geq 0}$  is called  *$T$ -delayed exponentially contractive with rate  $\nu$*  if

$$\|P_t f\|_{L^2(\mu)} \leq e^{-\nu(t-T)} \|f\|_{L^2(\mu)} \quad (\text{A.20})$$

for all  $f \in L_0^2(\mu)$  and  $t \geq 0$ .

- (ii) The semigroup  $(P_t)_{t \geq 0}$  is called *exponentially contractive in  $T$ -average with rate  $\nu$*  if

$$\frac{1}{T} \int_t^{t+T} \|P_s f\|_{L^2(\mu)} \, ds \leq e^{-\nu t} \|f\|_{L^2(\mu)} \quad (\text{A.21})$$

for all  $f \in L_0^2(\mu)$  and  $t \geq 0$ .

If  $(P_t)_{t \geq 0}$  is exponentially contractive in  $T$ -average with rate  $\nu$ , then for  $t \geq T$ ,

$$\|P_t f\|_{L^2(\mu)} \leq \frac{1}{T} \int_{t-T}^t \|P_s f\|_{L^2(\mu)} \, ds \leq e^{-\nu(t-T)} \|f\|_{L^2(\mu)}.$$

For  $t < T$  one has  $e^{-\nu(t-T)} > 1$  so that exponential contractivity in  $T$ -average implies  $T$ -delayed exponential contractivity with the same rate. For  $\varepsilon \in (0, 1)$  let

$$t_0(\varepsilon) := \inf \left\{ \frac{1}{\nu} \log(\varepsilon^{-1}) + T : \nu, T \in [0, \infty) \text{ such that (A.20) holds} \right\}.$$

We can compare the  $\varepsilon$ -relaxation time to the time  $t_0(\varepsilon)$ .

**Lemma A.8.** *For any  $\varepsilon \in (0, 1)$ ,*

$$t_{\text{rel}}(\varepsilon) \leq t_0(\varepsilon) \leq 2t_{\text{rel}}(\varepsilon).$$

*Proof.* On the one hand, if (A.20) holds, then the  $L^2$ -norm has decayed by a factor  $\varepsilon$  after time  $T + \frac{1}{\nu} \log(\varepsilon^{-1})$ . This yields the bound  $t_{\text{rel}}(\varepsilon) \leq t_0(\varepsilon)$ . On the other hand, for any  $t \geq 0$  and  $f \in L_0^2(\mu)$ ,

$$\begin{aligned} \|P_t f\|_{L^2(\mu)} &\leq \|P_{\frac{t}{t_{\text{rel}}(\varepsilon)}}^{[t/t_{\text{rel}}(\varepsilon)]} f\|_{L^2(\mu)} \leq \varepsilon^{[t/t_{\text{rel}}(\varepsilon)]} \|f\|_{L^2(\mu)} \\ &\leq e^{-\log(\varepsilon^{-1})t_{\text{rel}}(\varepsilon)^{-1}(t-t_{\text{rel}}(\varepsilon))} \|f\|_{L^2(\mu)}, \end{aligned}$$

so that (A.20) holds with  $\nu = \log(\varepsilon^{-1})t_{\text{rel}}(\varepsilon)$  and  $T = t_{\text{rel}}(\varepsilon)$ . This shows that  $t_0(\varepsilon) \leq 2t_{\text{rel}}(\varepsilon)$ .  $\square$

In the non-reversible case, the spectral gap does not provide a lower bound on the non-asymptotic relaxation times introduced above. However, we will show that lower bounds on  $t_{\text{rel}}$  are given by the inverse of the *singular value gap*

$$s(L) = \inf \left\{ \frac{\|Lf\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}} : f \in L_0^2(\mu) \cap \text{Dom}(L), f \neq 0 \right\}. \quad (\text{A.22})$$

The singular value gap is the infimum of the singular spectrum of the generator  $L$  restricted to the complement of the constant functions, i.e. it is the square root of the spectral gap of  $L^*L$ . In the reversible case,  $s(L) = \text{gap}(L)$ . In the non-reversible case, one only knows that  $|\lambda| \geq s(L)$  for any  $\lambda \in \text{spec}(L) \setminus \{0\}$ , but this does not imply a relation between  $s(L)$  and  $\text{gap}(L)$ .

If the operator  $-L: L_0^2(\mu) \cap \text{Dom}(L) \rightarrow L_0^2(\mu)$  is invertible, then its inverse  $G$  is the potential operator and  $s(L)$  is the inverse of the operator norm of  $G$ , i.e.

$$\frac{1}{s(L)} = \sup \left\{ \frac{\|Gf\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}} : f \in L_0^2(\mu), f \neq 0 \right\}. \quad (\text{A.23})$$

In this case,  $Gf = \int_0^\infty P_t f dt$  for any  $f \in L_0^2(\mu)$ , and in stationarity, the variance of ergodic averages  $\frac{1}{t} \int_0^t f(X_s) ds$  is bounded from above by  $\frac{2}{ts(L)} \|f\|_{L^2(\mu)}^2$ , see e.g. [153] or [96].

**Remark A.9** (Chatterjee's concept of relaxation times). In discrete time and on a finite state space, Chatterjee [65] recently defined the relaxation time for the ergodic averages of a non-reversible Markov chain as the inverse of the singular value gap. Our definition of the relaxation time  $t_{\text{rel}}$  is different and more closely related to mixing times. The inverse of the singular value gap only provides a lower bound for  $t_{\text{rel}}$  in the non-reversible case.

**Lemma A.10.** *Assume  $(P_t)_{t \geq 0}$  is  $T$ -delayed exponentially contractive with rate  $\nu$  or exponentially contractive in  $2T$ -average with rate  $\nu$ . Then*

$$\frac{1}{\nu} \geq \text{gap}(L)^{-1} \quad \text{and} \quad \frac{1}{\nu} + T \geq s(L)^{-1}.$$

*In particular,  $t_{\text{rel}} \geq \frac{1}{2} s(L)^{-1}$ .*

## Appendix A. Non-reversible lifts of reversible diffusion processes and relaxation times

*Proof.* Let  $\alpha \in \mathbb{C}$  with  $\operatorname{Re}(\alpha) < \nu$ . For  $f \in L_0^2(\mu)$ , both the  $T$ -delayed exponential contractivity and exponential contractivity in  $2T$ -average with rate  $\nu$  imply

$$\int_0^\infty |e^{\alpha t}| \|P_t f\|_{L^2(\mu)} dt \leq C(\alpha) \|f\|_{L^2(\mu)},$$

where  $C(\alpha)$  is a finite constant depending only on  $\alpha$ . Therefore, the resolvent  $(-\alpha I - L)^{-1} f = \int_0^\infty e^{\alpha t} P_t f dt$  exists in  $L^2(\mu)$  and defines a bounded linear operator on  $L_0^2(\mu)$ , i.e.  $\alpha \notin \operatorname{spec}(-L|_{L_0^2(\mu)})$  and hence  $\operatorname{gap}(L) \geq \nu$ , showing the first assertion.

For  $f \in L_0^2(\mu)$  the exponential contractivity in  $2T$ -average with rate  $\nu$  yields

$$\begin{aligned} \int_0^\infty \|P_t f\|_{L^2(\mu)} dt &= \int_0^\infty \frac{1}{2T} \int_t^{t+2T} \|P_s f\|_{L^2(\mu)} ds dt + \frac{1}{2T} \int_0^{2T} (2T - t) \|P_t f\|_{L^2(\mu)} dt \\ &\leq \left( \int_0^\infty e^{-\nu t} dt + \frac{1}{2T} \int_0^{2T} (2T - t) dt \right) \|f\|_{L^2(\mu)} \\ &= \left( \frac{1}{\nu} + T \right) \|f\|_{L^2(\mu)}. \end{aligned}$$

This yields convergence of the potential operator  $Gf = \int_0^\infty P_t f dt$  in  $L^2(\mu)$  for all  $f \in L_0^2(\mu)$ , which is the inverse of the operator  $-L: L_0^2(\mu) \cap \operatorname{Dom}(L) \rightarrow L_0^2(\mu)$ , see [229], together with the estimate

$$\|Gf\|_{L^2(\mu)} \leq \left( \frac{1}{\nu} + T \right) \|f\|_{L^2(\mu)}.$$

For  $g \in \operatorname{Dom}(L) \cap L_0^2(\mu)$ , we may set  $f = Lg$  and obtain the bound

$$\|g\|_{L^2(\mu)} \leq \left( \frac{1}{\nu} + T \right) \|Lg\|_{L^2(\mu)}$$

for all  $g \in \operatorname{Dom}(L) \cap L_0^2(\mu)$ , which yields the claim.

In the case of  $T$ -delayed exponential contractivity with rate  $\nu$ , we have that

$$\begin{aligned} \int_0^\infty \|P_t f\|_{L^2(\mu)} dt &\leq \int_T^\infty e^{-\nu(t-T)} \|f\|_{L^2(\mu)} dt + \int_0^T \|f\|_{L^2(\mu)} dt \\ &= \left( \frac{1}{\nu} + T \right) \|f\|_{L^2(\mu)} \end{aligned}$$

for all  $f \in L_0^2(\mu)$ , and we conclude analogously.  $\square$

In contrast to lower bounds, upper bounds on  $t_{\text{rel}}$  of the correct order are much harder to obtain in the degenerate non-reversible case and require quantitative hypocoercivity results [6, 53, 55, 61, 186], see Section A.5 below.

## A.4. Lower bounds for relaxation times of lifts

Suppose that  $(P_t)_{t \geq 0}$  with generator  $(L, \text{Dom}(L))$  is the transition semigroup on  $L^2(\mathcal{S}, \mu)$  of a reversible Markov process and that  $(\hat{P}_t)_{t \geq 0}$  with generator  $(\hat{L}, \text{Dom}(\hat{L}))$  is the transition semigroup on  $L^2(\mathcal{S} \times \mathcal{V}, \hat{\mu})$  of an arbitrary second-order lift. The next theorem shows that relaxation to equilibrium, measured by the relaxation times  $t_{\text{rel}}(P)$  and  $t_{\text{rel}}(\hat{P})$  of the reversible Markov process and its lift, can at most be accelerated by a square root through lifting.

**Theorem A.11.** *If  $(\hat{P}_t)_{t \geq 0}$  is  $T$ -delayed exponentially contractive with rate  $\nu$  or exponentially contractive in  $2T$ -average with rate  $\nu$ , then*

$$\frac{1}{\nu} + T \geq \frac{1}{\sqrt{2 \text{gap}(L)}}.$$

In particular,

$$t_{\text{rel}}(\hat{P}) \geq \frac{1}{2\sqrt{2}} \sqrt{t_{\text{rel}}(P)}.$$

*Proof.* Let  $\lambda > \text{gap}(L)$ . By the variational characterisation of the spectral gap, there exists  $f \in \text{Dom}(L) \cap L_0^2(\mu)$  such that  $(f, -Lf)_{L^2(\mu)} \leq \lambda \|f\|_{L^2(\mu)}^2$ . By [Definition A.1](#),  $f \circ \pi \in \text{Dom}(\hat{L}) \cap L_0^2(\hat{\mu})$  and

$$\|\hat{L}(f \circ \pi)\|_{L^2(\hat{\mu})}^2 = 2(f, -Lf)_{L^2(\mu)} \leq 2\lambda \|f\|_{L^2(\mu)}^2 = 2\lambda \|f \circ \pi\|_{L^2(\hat{\mu})}^2. \quad (\text{A.24})$$

By considering the limit  $\lambda \rightarrow \text{gap}(L)$ , we see that  $s(\hat{L}) \leq \sqrt{2 \text{gap}(L)}$ , which implies the assertion by [Lemma A.10](#).  $\square$

**Remark A.12.** (i) Consider a Markov chain in discrete or continuous time with state space  $\hat{\mathcal{S}}$ , transition kernel  $\hat{p}$  and invariant measure  $\hat{\mu}$  that is a first-order lift (in the sense of [\[66\]](#), or equivalently, [\(A.10\)](#)) of a reversible Markov chain with state space  $\mathcal{S}$ , transition kernel  $p$  and invariant measure  $\mu$ . That is,

$$(f \circ \pi, \hat{p}(g \circ \pi))_{L^2(\hat{\mu})} = (f, pg)_{L^2(\mu)}$$

for all bounded measurable functions  $f, g: \mathcal{S} \rightarrow \mathbb{R}$ . Then the associated generators  $\hat{L} = \hat{p} - I$  and  $L = p - I$  satisfy

$$\begin{aligned} \|\hat{L}(f \circ \pi)\|_{L^2(\hat{\mu})}^2 &= \|(\hat{p} - I)(f \circ \pi)\|_{L^2(\hat{\mu})}^2 \\ &= \|f \circ \pi\|_{L^2(\hat{\mu})}^2 - 2(f \circ \pi, \hat{p}(f \circ \pi))_{L^2(\hat{\mu})} + \|\hat{p}(f \circ \pi)\|_{L^2(\hat{\mu})}^2 \\ &\leq 2\|f\|_{L^2(\mu)}^2 - 2(f, pf)_{L^2(\mu)} = 2(f, -Lf)_{L^2(\mu)}, \end{aligned}$$

since  $\hat{p}$  is a contraction. Hence the second-order lift condition [\(A.8\)](#) holds with an inequality, so that we can conclude  $s(\hat{L}) \leq \sqrt{2 \text{gap}(L)}$  similarly as in [\(A.24\)](#). [Lemma A.10](#) can then be used to obtain a lower bound on the relaxation time of the lifted Markov chain in terms of the spectral gap of  $L$  as in [Theorem A.11](#).

- (ii) It is an open problem whether also  $\frac{1}{\nu}$  can be lower bounded in terms of  $1/\sqrt{\text{gap}(L)}$ . This would follow from an upper bound on  $\text{gap}(\hat{L})$  by a multiple of  $\sqrt{\text{gap}(L)}$ . We have not managed to prove such a bound. This is similar to the discrete time case, where a lower bound for the mixing time is available but an upper bound for the spectral gap of lifts is not known [66].
- (iii) In discrete time, Diaconis and Miclo [81] consider second-order Markov chains on finite state spaces whose transition probabilities depend not only on the current state but also the previous one, thereby adding a notion of discrete-time velocity. The authors conjecture that by adding such a velocity to a random walk on a finite Abelian group, the spectral gap cannot be increased by more than a square root, and that this improvement can always be achieved by considering a particular choice of second order Markov chain. This corresponds to the square-root lower bound Theorem A.11 on the relaxation time of a lift and the notion of  $C$ -optimality introduced in Definition A.13 below.

An acceleration of the relaxation time from  $1/\text{gap}(L)$  to a multiple of  $1/\sqrt{\text{gap}(L)}$  corresponds to a diffusive to ballistic speed-up. Theorem A.11 motivates the following definition.

**Definition A.13** ( $C$ -optimal lift). For a fixed constant  $C \in [1, \infty)$ , a lift as above is called  $C$ -optimal, if

$$t_{\text{rel}}(\hat{P}) \leq C \frac{1}{2\sqrt{2}} \sqrt{t_{\text{rel}}(P)}.$$

**Example A.14** (Gaussian case). Suppose that  $\mu = \mathcal{N}(0, 1)$  is the standard normal distribution on  $\mathbb{R}$ . The overdamped Langevin process solving the SDE

$$dZ_t = -\frac{1}{2}Z_t dt + dB_t$$

driven by a one-dimensional Brownian motion  $(B_t)_{t \geq 0}$  is reversible, has invariant measure  $\mu$  and spectral gap  $1/2$ . The critically damped Langevin dynamics solving

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= -X_t dt - \gamma V_t dt + \sqrt{2\gamma} dB_t \quad \text{with } \gamma = 2 \end{aligned}$$

is a second-order lift of this process with invariant measure  $\hat{\mu} = \mu \otimes \mathcal{N}(0, 1)$ . The choice  $\gamma = 2$  maximises the spectral gap of the generator, see Figure A.1.

Denoting  $A = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}$ , an explicit computation [17] gives

$$\|\hat{P}_t\|_{L_0^2(\hat{\mu}) \rightarrow L_0^2(\hat{\mu})} = \|\exp(-tA)\| = \sqrt{1 + 2t^2 + 2t(1 + t^2)^{1/2}} \cdot e^{-t}$$

for any  $t \geq 0$ , which yields  $t_{\text{rel}}(\hat{P}) \leq 2.73$ . Thus in this case, critical Langevin dynamics is a 5.46-optimal lift of the overdamped Langevin diffusion. Note that the asymptotic

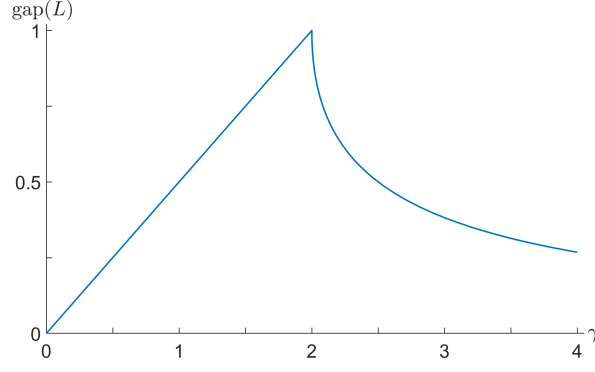


Figure A.1.: Spectral gap for Langevin dynamics.

decay rate of the critically damped Langevin dynamics is exactly 1, the square root of twice the spectral gap of the corresponding reversible diffusion.

More generally, for any  $m > 0$ , the critically damped Langevin dynamics  $(X^{(m)}, V^{(m)})$  for the potential  $U_m(x) = mx^2/2$  with invariant measure  $\hat{\mu}_m = \mathcal{N}(0, m^{-1}) \otimes \mathcal{N}(0, 1)$  and friction  $\gamma = 2\sqrt{m}$  is a second-order lift of the corresponding overdamped Langevin process with spectral gap  $m/2$ . A scaling argument shows that the distributions of  $(X_t^{(m)}, V_t^{(m)})$  and  $(m^{-1/2}X_{\sqrt{m}t}, V_{\sqrt{m}t})$  coincide for all  $t \geq 0$ , so that the transition semigroup  $(\hat{P}_t^{(m)})_{t \geq 0}$  satisfies

$$\|\hat{P}_t^{(m)}\|_{L_0^2(\hat{\mu}_m) \rightarrow L_0^2(\hat{\mu}_m)} = \|\hat{P}_{\sqrt{m}t}\|_{L_0^2(\hat{\mu}) \rightarrow L_0^2(\hat{\mu})}$$

and hence  $t_{\text{rel}}(\hat{P}^{(m)}) = m^{-1/2}t_{\text{rel}}(\hat{P})$ . This shows that also in this case, critical Langevin dynamics is a 5.46-optimal lift of the corresponding overdamped Langevin process.

Finally, a factorisation argument shows that the same results hold true for arbitrary quadratic potentials on  $\mathbb{R}^d$ , i.e., in the Gaussian case, critical Langevin dynamics is *always* a 5.46-optimal lift of the corresponding overdamped Langevin diffusion.

## A.5. Upper bounds for relaxation times of lifts

Based on the approach of Albritton, Armstrong, Mourrat and Novack [6] relying on space-time Poincaré inequalities, Cao, Lu and Wang [61, 186] have derived quantitative upper bounds for relaxation times of various non-reversible Markov processes. The framework of second-order lifts developed above allows us to rephrase and simplify their results.

For any  $\gamma \geq 0$ , we consider a Markov process with state space  $\mathbb{R}^d \times \mathbb{R}^d$  and invariant probability measure  $\hat{\mu} = \mu \otimes \kappa$ , where  $\mu$  is a probability measure on  $\mathbb{R}^d$  and  $\kappa = \mathcal{N}(0, I_d)$ . The associated transition semigroups acting on  $L^2(\hat{\mu})$  are denoted by  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  and their generators by  $(\hat{L}^{(\gamma)}, \text{Dom}(\hat{L}^{(\gamma)}))$ . Recall that  $(\Pi_v f)(x, v) = \int_{\mathbb{R}^d} f(x, w) \kappa(dw)$ . We assume the following decomposition of the generators.

**Assumption A.1.** There is an operator  $(\hat{L}, \text{Dom}(\hat{L}))$  on  $L^2(\hat{\mu})$  such that the following holds:

(i) For all  $\gamma \geq 0$  we have  $\text{Dom}(\hat{L}^{(\gamma)}) \subseteq \text{Dom}(\hat{L})$  and

$$\hat{L}^{(\gamma)} f = \hat{L} f + \gamma(\Pi_v - I)f \quad \text{for all } f \in \text{Dom}(\hat{L}^{(\gamma)}).$$

(ii) The operator  $(\hat{L}, \text{Dom}(\hat{L}))$  is negative semi-definite, i.e.  $(f, \hat{L}f)_{L^2(\hat{\mu})} \leq 0$  for all  $f \in \text{Dom}(\hat{L})$ .

Assumption A.1 is for instance satisfied by the generators of Randomised Hamiltonian Monte Carlo or the Bouncy Particle Sampler with refresh rate  $\gamma$  by choosing

$$\hat{L}f = v \cdot \nabla_x f - \nabla U(x) \cdot \nabla_v f \tag{A.25}$$

to be the generator of the deterministic Hamiltonian flow, or

$$\hat{L}f = v \cdot \nabla_x f + (v \cdot \nabla U(x))_+(R - I)f \tag{A.26}$$

to be the generator of the Bouncy Particle Sampler without refreshment, respectively, see Example A.3. Indeed, in the former case,  $\hat{L}$  is antisymmetric, and in the latter case, the adjoint of  $\hat{L}$  in  $L^2(\hat{\mu})$  is

$$\hat{L}^* f = -v \cdot \nabla_x f + (v \cdot \nabla U(x))_-(R - I)f$$

so that a direct calculation shows

$$(f, \hat{L}f)_{L^2(\hat{\mu})} = \frac{1}{2}(f, (\hat{L} + \hat{L}^*)f)_{L^2(\hat{\mu})} = -\frac{1}{4} \int_{\mathbb{R}^d} |v \cdot \nabla U(x)| (Rf - f)^2 d\hat{\mu}.$$

Hence  $\hat{L}$  is negative definite in both cases.

### A.5.1. Space-time Poincaré inequalities and contractivity

Exponential contractivity in  $T$ -average, and thereby upper bounds on the relaxation times of the transition semigroups  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$ , can be established using a Poincaré-type inequality in space and time, as introduced by [6]. To this end, we fix a finite time  $T > 0$ . It will then be convenient to consider space and time together, and we hence set

$$\bar{\mu} = \lambda|_{[0, T]} \otimes \mu$$

where  $\lambda|_{[0, T]}$  is the Lebesgue measure on  $[0, T]$ . Furthermore, we consider the operator  $(A, \text{Dom}(A))$  on  $L^2([0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \bar{\mu} \otimes \kappa)$  defined by

$$Af = -\partial_t f + \hat{L}f \tag{A.27}$$

with domain consisting of all functions  $f \in L^2(\bar{\mu} \otimes \kappa)$  such that  $f(\cdot, x, v)$  is absolutely continuous for  $\mu \otimes \kappa$ -a.e.  $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$  with  $\partial_t f \in L^2(\bar{\mu} \otimes \kappa)$ , and  $f(t, \cdot) \in \text{Dom}(\hat{L})$  for  $\lambda$ -a.e.  $t \in [0, T]$  with  $\hat{L}f \in L^2(\bar{\mu} \otimes \kappa)$ .

**Definition A.15** (Space-time Poincaré inequality). A *space-time Poincaré inequality* with constants  $C_0(T)$  and  $C_1(T)$  holds for  $\hat{L}$ , if for all functions  $f \in \text{Dom}(A)$

$$\|f - \int f d(\bar{\mu} \otimes \kappa)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq C_0(T) \|Af\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + C_1(T) \|f - \Pi_v f\|_{L^2(\bar{\mu} \otimes \kappa)}^2.$$

We also say that  $\hat{L}$  satisfies  $\text{STPI}(T, C_0(T), C_1(T))$ .

**Remark A.16.** By continuity, it suffices to verify the inequality in [Definition A.15](#) for functions  $f$  from a *core* of  $A$ , i.e. a subset that is dense in the domain of  $A$  with respect to the graph norm  $\|f\|_{L^2(\bar{\mu} \otimes \kappa)} + \|Af\|_{L^2(\bar{\mu} \otimes \kappa)}$ . For example, if  $\mu(dx) \propto \exp(-U(x)) dx$  for some function  $U \in C^1(\mathbb{R}^d)$  such that  $U(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ , and  $\hat{L}$  is the generator of the Hamiltonian flow given by [\(A.25\)](#) associated to the Hamiltonian  $H(x, v) = U(x) + \frac{1}{2}|v|^2$ , then the space  $C_0^\infty([0, T] \times \mathbb{R}^d \times \mathbb{R}^d)$  of smooth compactly supported functions on  $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d$  is a core. Indeed, a function  $f \in \text{Dom}(A)$  can be approximated in the graph norm by the compactly supported functions  $f_n(t, x, v) = g_n(H(x, v))f(t, x, v)$  where  $g_n : [0, \infty) \rightarrow [0, 1]$  is an increasing sequence of smooth compactly supported functions such that  $g_n(s) \rightarrow 1$  for all  $s \in [0, \infty)$ , because  $\hat{L}H = 0$  implies  $(Af_n)(t, x, v) = g_n(H(x, v))(Af)(t, x, v)$ . It thus follows by a standard mollification that  $C_0^\infty([0, T] \times \mathbb{R}^d \times \mathbb{R}^d)$  is a core for  $A$ .

As observed in [\[6, 61, 186\]](#), a space-time Poincaré inequality for  $\hat{L}$  yields the following quantitative hypocoercivity estimate for the semigroups  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$ .

**Theorem A.17** (STPI implies exponential contractivity in  $T$ -average). *Assume that  $\hat{L}$  satisfies  $\text{STPI}(T, C_0(T), C_1(T))$ . Then for any  $\gamma > 0$ , the semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  is exponentially contractive in  $T$ -average with rate*

$$\nu = \frac{\gamma}{\gamma^2 C_0(T) + C_1(T)}.$$

The choice  $\gamma = \sqrt{\frac{C_1(T)}{C_0(T)}}$  maximises the rate and gives  $\nu = \frac{1}{2\sqrt{C_0(T)C_1(T)}}$ . For this choice,

$$t_{\text{rel}}(\varepsilon) \leq T + 2\sqrt{C_0(T)C_1(T)} \log(\varepsilon^{-1}).$$

For the reader's convenience we recall the short proof of the theorem.

*Proof.* Let  $f \in L_0^2(\hat{\mu})$ . Then by [Assumption A.1](#),

$$\begin{aligned} \frac{d}{dt} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds &= \|\hat{P}_{t+T}^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 - \|\hat{P}_t^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 \\ &= \int_t^{t+T} \frac{d}{ds} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \\ &= 2 \int_t^{t+T} \left( \hat{P}_s^{(\gamma)} f, (\hat{L} + \gamma(\Pi_v - I)) \hat{P}_s^{(\gamma)} f \right)_{L^2(\hat{\mu})} ds \end{aligned}$$

$$\begin{aligned}
 &\leq -2\gamma \int_t^{t+T} \left( \hat{P}_s^{(\gamma)} f, (I - \Pi_v) \hat{P}_s^{(\gamma)} f \right)_{L^2(\hat{\mu})} ds \\
 &= -2\gamma \int_t^{t+T} \|(\Pi_v - I) \hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \\
 &= -2\gamma \|(\Pi_v - I) \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2.
 \end{aligned}$$

Since  $\hat{P}_{t+}^{(\gamma)} f \in \text{Dom}(A)$  and  $A\hat{P}_{t+}^{(\gamma)} f = -\gamma(\Pi_v - I)\hat{P}_{t+}^{(\gamma)} f$ , the space-time Poincaré inequality yields

$$\|\hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq (\gamma^2 C_0(T) + C_1(T)) \|(\Pi_v - I) \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2,$$

so that

$$\frac{d}{dt} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \leq \frac{-2\gamma}{\gamma^2 C_0(T) + C_1(T)} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds.$$

Grönwall's inequality yields

$$\begin{aligned}
 \frac{1}{T} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds &\leq \exp\left(\frac{-2\gamma}{\gamma^2 C_0(T) + C_1(T)} t\right) \frac{1}{T} \int_0^T \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \\
 &\leq e^{-2\gamma t} \|f\|_{L^2(\hat{\mu})}^2.
 \end{aligned}$$

Finally, the Cauchy-Schwarz inequality yields

$$\frac{1}{T} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})} ds \leq \left( \frac{1}{T} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \right)^{\frac{1}{2}} \leq e^{-\gamma t} \|f\|_{L^2(\hat{\mu})}$$

and thus the  $\nu$ -exponential contractivity in  $T$ -average. The upper bound on  $t_{\text{rel}}(\varepsilon)$  follows by [Lemma A.8](#).  $\square$

In the following we demonstrate how the framework of second-order lifts can be used to simplify the approach in [\[61, 186\]](#) to proving space-time Poincaré inequalities leading to quantitative hypocoercivity bounds. The main tools are a space-time property of lifts and a quantitative version of the divergence lemma.

### A.5.2. Space-time property of lifts

For  $(t, x, v) \in [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d$  we set  $\pi(t, x, v) = (t, x)$ , whereas  $\pi(x, v) = x$ .

**Lemma A.18** (Space-time property of lifts). *Let  $(\hat{L}, \text{Dom}(\hat{L}))$  be a lift of  $(L, \text{Dom}(L))$ . Then for any  $f, g, h \in L^2(\bar{\mu})$  such that  $f \circ \pi \in \text{Dom}(A)$  and  $g(t, \cdot) \in \text{Dom}(L)$  for a.e.  $t \in [0, T]$  with  $Lg \in L^2(\bar{\mu})$ , we have*

$$\frac{1}{2} \left( A(f \circ \pi), h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} = -\frac{1}{2} (\partial_t f, h)_{L^2(\bar{\mu})} + \mathcal{E}_T(f, g),$$

where  $\mathcal{E}_T(f, g) = \int_0^T \mathcal{E}(f(t, \cdot), g(t, \cdot)) dt$  is the time-integrated Dirichlet form.

*Proof.* By Properties (A.6), (A.7) and (A.8) in Definition A.1,

$$\begin{aligned} \left( \hat{L}(f \circ \pi), h \circ \pi \right)_{L^2(\bar{\mu} \otimes \kappa)} &= 0, \quad \left( -\partial_t(f \circ \pi), \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} = 0, \quad \text{and} \\ \left( \hat{L}(f \circ \pi), \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} &= \int_0^T \left( \hat{L}(f(t, \cdot) \circ \pi), \hat{L}(g(t, \cdot) \circ \pi) \right)_{L^2(\mu \otimes \kappa)} dt \\ &= 2\mathcal{E}_T(f, g). \end{aligned}$$

Moreover,  $(-\partial_t(f \circ \pi), h \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} = (-\partial_t f, h)_{L^2(\bar{\mu})}$ .  $\square$

**Remark A.19.** If we set  $h = -\partial_t g$ , the space-time property reads

$$\frac{1}{2} (A(f \circ \pi), A(g \circ \pi))_{L^2(\bar{\mu} \otimes \kappa)} = \frac{1}{2} (\partial_t f, \partial_t g)_{L^2(\bar{\mu})} + \mathcal{E}_T(f, g).$$

For functions satisfying either periodic or Neumann boundary conditions on  $[0, T]$ , the right hand side is the Dirichlet form of the reversible Markov process with state space  $[0, T] \times \mathbb{R}^d$  given by Brownian motion in the first component and an independent process with generator  $L$  in the other components. The space-time property generalises the observation that the process with generator  $A$  describing deterministic motion with speed  $-1$  in the first (time) component and motion described by  $\hat{L}$  in the other (space and velocity) components is a lift of this reversible process.

### A.5.3. Divergence lemma

We now assume that for any  $\gamma \geq 0$ , the semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  is a second-order lift of the transition semigroup  $(P_t)_{t \geq 0}$  acting on  $L^2(\mu)$  of an overdamped Langevin diffusion with invariant probability measure  $\mu(dx) \propto \exp(-U(x)) dx$ , see Example A.3. Recall that the generator of  $(P_t)_{t \geq 0}$  is given by

$$Lf = -\frac{1}{2} \nabla^* \nabla f = -\frac{1}{2} \nabla U \cdot \nabla f + \frac{1}{2} \Delta f$$

for  $f \in \text{Dom}(L)$ , where the adjoint is in  $L^2(\mu)$ . In particular,

$$\mathcal{E}(f, g) = \frac{1}{2} (\nabla f, \nabla g)_{L^2(\mu)}.$$

#### Assumption A.2.

- (i) The potential satisfies  $U \in C^2(\mathbb{R}^d)$  and there exists a constant  $\kappa_- \in [0, \infty)$  such that  $\nabla^2 U(x) \succeq -\kappa_- I_d$  for all  $x \in \mathbb{R}^d$ .
- (ii) The probability measure  $\mu$  satisfies a Poincaré inequality with constant  $m$ , i.e.

$$\int_{\mathbb{R}^d} f^2 d\mu \leq \frac{1}{m} \int_{\mathbb{R}^d} |\nabla f|^2 d\mu \quad \text{for all } f \in H^1(\mu) \text{ with } \int_{\mathbb{R}^d} f d\mu = 0.$$

- (iii) The spectrum of  $(L, \text{Dom}(L))$  on  $L^2(\mu)$  is discrete.

Assumption A.2(iii) is for instance satisfied if  $\lim_{|x| \rightarrow \infty} \frac{U(x)}{|x|^\alpha} = \infty$  for some  $\alpha > 1$ , cf. [139].

The next statement is a quantitative version of the divergence lemma for the domain  $(0, T) \times \mathbb{R}^d$ . The divergence lemma is equivalent to Lions' lemma, estimating the  $L^2$ -norm by the  $H^{-1}$ -norm of the gradient, see [10]. Let  $\bar{\nabla} f$  denote the gradient in both time and space variables of a function  $f$  defined on  $(0, T) \times \mathbb{R}^d$ , and let  $\bar{\nabla}^* X$  denote the adjoint divergence operator with respect to the measure  $\bar{\mu}$  applied to a vector field  $X: (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^{d+1}$ .

**Lemma A.20** (Quantitative divergence lemma). *Suppose that Assumption A.2 is satisfied. Then for any  $T \in (0, \infty)$ , there exist constants  $c_0(T), c_1(T) \in (0, \infty)$  such that for any  $f \in L_0^2([0, T] \times \mathbb{R}^d, \bar{\mu})$  there is a vector field  $X \in H_0^{1,2}((0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^{d+1}, \bar{\mu})$  with*

$$\|X\|_{L^2(\bar{\mu})}^2 = \sum_{i=0}^d \|X_i\|_{L^2(\bar{\mu})}^2 \leq c_0(T) \|f\|_{L^2(\bar{\mu})}^2, \quad (\text{A.28})$$

$$\|\bar{\nabla} X\|_{L^2(\bar{\mu})}^2 = \sum_{i,j=0}^d \|\partial_i X_j\|_{L^2(\bar{\mu})}^2 \leq c_1(T) \|f\|_{L^2(\bar{\mu})}^2, \quad (\text{A.29})$$

such that

$$f = \bar{\nabla}^* X. \quad (\text{A.30})$$

The constants can be chosen as

$$c_0(T) = 19T^2 + 70\frac{1}{m}, \quad (\text{A.31})$$

$$c_1(T) = 328 + 75\kappa_- \max\left(\frac{1}{\sqrt{m}}, \frac{T}{\pi}\right)^2 + \frac{1821}{mT^2}, \quad (\text{A.32})$$

and the vector field can be represented as

$$X = \begin{pmatrix} -h \\ \nabla_x g \end{pmatrix} \text{ with functions } h \in H_0^{1,2}((0, T) \times \mathbb{R}^d, \bar{\mu}) \text{ and } g \in H_0^{2,2}((0, T) \times \mathbb{R}^d, \bar{\mu}).$$

A corresponding result has first been proven in [61] with constants that are not of the optimal order, see also [55]. An improved bound with constants of the order stated above has been conjectured in [61]. Based on a careful extension of the arguments in this paper, the precise result above is proven in a work in preparation [101].

**Remark A.21.** The bounds resulting from the divergence lemma considered below are optimised when choosing  $T$  proportional to  $1/\sqrt{m}$ . For  $T = 3/\sqrt{m}$ , (A.31) and (A.32) yield

$$c_0(T) = \frac{241}{m} \quad \text{and} \quad c_1(T) = 530 + \frac{1}{3} + 75\frac{\kappa_-}{m}. \quad (\text{A.33})$$

#### A.5.4. Space-time Poincaré inequalities via lifts

For simplicity and clarity of the exposition we now assume as in [Example A.3](#) that  $\hat{L}$  is the generator of the deterministic Hamiltonian flow. We stress, however, that with appropriate modifications, the approach is much more generally applicable, see, for example, [\[101\]](#) for applications to processes with reflection at the boundary of a convex domain and [\[97\]](#) for further applications to sampling algorithms and processes with different boundary behaviours. Our goal is to show how the space-time property of lifts combined with the divergence lemma can be used to prove the space-time Poincaré inequality stated in [Theorem A.23](#) below.

We consider the adjoint  $(A^*, \text{Dom}(A^*))$  in  $L^2([0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \bar{\mu} \otimes \kappa)$  of the operator  $(A, \text{Dom}(A))$  defined in [\(A.27\)](#). By integration by parts and the antisymmetry of  $\hat{L}$ ,

$$(Af, g)_{L^2(\bar{\mu} \otimes \kappa)} = (f, -Ag)_{L^2(\bar{\mu} \otimes \kappa)} + \int_{\mathbb{R}^{2d}} (f(T, \cdot)g(T, \cdot) - f(0, \cdot)g(0, \cdot)) d\hat{\mu}$$

for all  $f, g \in \text{Dom}(A)$ . Thus functions  $g \in \text{Dom}(A)$  satisfying  $g(0, \cdot) = g(T, \cdot) = 0$  are contained in the domain of  $A^*$ , and

$$A^*g = -Ag = \partial_t g - \hat{L}g. \quad (\text{A.34})$$

For  $v \in \mathbb{R}^d$  we set  $\bar{v} = (-1, v) \in \mathbb{R}^{d+1}$ , so that  $A = \bar{v} \cdot \bar{\nabla} - \nabla U(x) \cdot \nabla_v$ .

The space-time property of lifts yields the following.

**Lemma A.22.** *Let  $X = \begin{pmatrix} -h \\ \nabla g \end{pmatrix}$  with  $h \in H_0^{1,2}(\bar{\mu})$  and  $g \in H_0^{2,2}(\bar{\mu})$ . Then*

$$\int_{\mathbb{R}^d} A^*(\bar{v} \cdot X \circ \pi) \kappa(dv) = \bar{\nabla}^* X \circ \pi, \quad \text{and} \quad (\text{A.35})$$

$$\|A^*(\bar{v} \cdot X \circ \pi) - \bar{\nabla}^* X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)} \leq \sqrt{2} \|\bar{\nabla} X\|_{L^2(\bar{\mu})}. \quad (\text{A.36})$$

*Proof.* The Dirichlet boundary values in time of  $X$  give  $\bar{v} \cdot X \circ \pi \in \text{Dom}(A^*)$ . Therefore, by [Lemma A.18](#), for any function  $\varphi \in C_0^\infty((0, T) \times \mathbb{R}^d)$ ,

$$\begin{aligned} (\varphi \circ \pi, \bar{\nabla}^* X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} &= (\varphi, \bar{\nabla}^* X)_{L^2(\bar{\mu})} = (\bar{\nabla} \varphi, X)_{L^2(\bar{\mu})} \\ &= -(\partial_t \varphi, h)_{L^2(\bar{\mu})} + 2\mathcal{E}_T(\varphi, g) = (A(\varphi \circ \pi), h \circ \pi + \hat{L}(g \circ \pi))_{L^2(\bar{\mu} \otimes \kappa)} \\ &= (A(\varphi \circ \pi), \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} = (\varphi \circ \pi, A^*(\bar{v} \cdot X \circ \pi))_{L^2(\bar{\mu} \otimes \kappa)}. \end{aligned}$$

This proves [\(A.35\)](#). For the proof of [\(A.36\)](#) note that since  $A^* = -\bar{v} \cdot \bar{\nabla} + \nabla U(x) \cdot \nabla_v$ ,

$$\|A^*(\bar{v} \cdot X \circ \pi) - \bar{\nabla}^* X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)} = \|\bar{v} \cdot \bar{\nabla}(\bar{v} \cdot X \circ \pi) - \bar{\nabla} \cdot X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)}. \quad (\text{A.37})$$

Moreover, noting that  $\int_{\mathbb{R}^d} v_i^2 v_k^2 d\kappa = \begin{cases} 3 & \text{if } i = k > 0, \\ 1 & \text{otherwise,} \end{cases}$  we obtain

$$\|\bar{v} \cdot \bar{\nabla}(\bar{v} \cdot X \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 = \sum_{i,j,k,l=0}^d (\partial_i X_j, \partial_k X_l)_{L^2(\bar{\mu})} \int_{\mathbb{R}^d} v_i v_j v_k v_l d\kappa$$

$$\begin{aligned}
 &= \sum_{i,k=0}^d (\partial_i X_i, \partial_k X_k)_{L^2(\bar{\mu})} \int v_i^2 v_k^2 d\kappa + \sum_{\substack{i,j=0 \\ i \neq j}}^d ((\partial_i X_j, \partial_i X_j)_{L^2(\bar{\mu})} + (\partial_i X_j, \partial_j X_i)_{L^2(\bar{\mu})}) \\
 &= \sum_{i,j=0}^d ((\partial_i X_i, \partial_j X_j)_{L^2(\bar{\mu})} + (\partial_i X_j, \partial_i X_j)_{L^2(\bar{\mu})} + (\partial_i X_j, \partial_j X_i)_{L^2(\bar{\mu})}) - 2\|\partial_0 X_0\|_{L^2(\bar{\mu})}^2 \\
 &\leq \|\bar{\nabla} \cdot X\|_{L^2(\bar{\mu})}^2 + 2\|\bar{\nabla} X\|_{L^2(\bar{\mu})}^2.
 \end{aligned}$$

Since  $\int_{\mathbb{R}^d} \bar{v} \cdot \bar{\nabla}(\bar{v} \cdot X \circ \pi) d\kappa = \bar{\nabla} \cdot X \circ \pi$  we conclude that

$$\|\bar{v} \cdot \bar{\nabla}(\bar{v} \cdot X \circ \pi) - \bar{\nabla} \cdot X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)}^2 = \|\bar{v} \cdot \bar{\nabla}(\bar{v} \cdot X \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 - \|\bar{\nabla} \cdot X\|_{L^2(\bar{\mu})}^2 \leq 2\|\bar{\nabla} X\|_{L^2(\bar{\mu})}^2.$$

By (A.37), this implies (A.36).  $\square$

**Theorem A.23** (Space-time Poincaré inequality). *Suppose that  $\mu$  satisfies Assumption A.2. Then for any  $T \in (0, \infty)$ , the generator  $\hat{L}$  of the Hamiltonian flow satisfies the inequality  $\text{STPI}(T, 2c_0(T), 3 + 4c_1(T))$  with the constants  $c_0(T)$  and  $c_1(T)$  from Lemma A.20.*

The following proof is an adapted, optimised, and partially simplified version of the proof of [186, Theorem 2.1].

*Proof.* By Remark A.16, it is sufficient to verify the space-time Poincaré inequality for functions  $f_0 \in C_0^\infty([0, T] \times \mathbb{R}^d \times \mathbb{R}^d)$ . Hence fix such a function, let  $f = f_0 - \int f_0 d(\bar{\mu} \otimes \kappa)$ , and let  $\tilde{f}: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  be such that  $\Pi_v f = \tilde{f} \circ \pi$ . First note that

$$\|f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 = \|f - \Pi_v f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + \|\tilde{f}\|_{L^2(\bar{\mu})}^2, \quad (\text{A.38})$$

so that it suffices to bound  $\|\tilde{f}\|_{L^2(\bar{\mu})}^2$  by the quantities  $\|f - \Pi_v f\|_{L^2(\bar{\mu} \otimes \kappa)}^2$  and  $\|Af\|_{L^2(\bar{\mu} \otimes \kappa)}^2$ . By Lemma A.20 applied to  $\tilde{f}$ , there exist functions  $h \in H_0^{1,2}((0, T) \times \mathbb{R}^d, \bar{\mu})$  and  $g \in H_0^{2,2}((0, T) \times \mathbb{R}^d, \bar{\mu})$  such that the vector field  $X = \begin{pmatrix} -h \\ \nabla g \end{pmatrix}$  satisfies (A.28), (A.29) and (A.30) with respect to  $\tilde{f}$ . In particular,

$$\|\tilde{f}\|_{L^2(\bar{\mu})}^2 = (\tilde{f}, \tilde{f})_{L^2(\bar{\mu})} = (\tilde{f}, \bar{\nabla}^* X)_{L^2(\bar{\mu})} = (\bar{\nabla} \tilde{f}, X)_{L^2(\bar{\mu})}. \quad (\text{A.39})$$

Using the space-time property of lifts (Lemma A.18), we see that

$$\begin{aligned}
 (\bar{\nabla} \tilde{f}, X)_{L^2(\bar{\mu})} &= (-\partial_t \tilde{f}, h)_{L^2(\bar{\mu})} + 2\mathcal{E}_T(\tilde{f}, g) \\
 &= (A(\tilde{f} \circ \pi), h \circ \pi + \hat{L}(g \circ \pi))_{L^2(\bar{\mu} \otimes \kappa)} \\
 &= (A(\tilde{f} \circ \pi), \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} \\
 &= (Af, \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} + (A(\Pi_v f - f), \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)}.
 \end{aligned}$$

Let us consider the two summands separately.

The first summand can be estimated as

$$\begin{aligned} (Af, \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} &\leq \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} \|\bar{v} \cdot X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)} = \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} \|X\|_{L^2(\bar{\mu})} \\ &\leq c_0(T)^{\frac{1}{2}} \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} \|\tilde{f}\|_{L^2(\bar{\mu})}, \end{aligned}$$

where we used  $\int_{\mathbb{R}^d} v_i v_j d\kappa = \delta_{ij}$  and  $c_0(T)$  is the constant from [Lemma A.20](#).

The second summand satisfies

$$(A(\Pi_v f - f), \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} = (\Pi_v f - f, A^*(\bar{v} \cdot X \circ \pi))_{L^2(\bar{\mu} \otimes \kappa)},$$

where  $A^*$  is the adjoint of  $A$  in  $L^2(\bar{\mu} \otimes \kappa)$ . Here we have used that by the conditions on  $h$  and  $g$ ,  $\bar{v} \cdot X \circ \pi = h \circ \pi + \hat{L}(g \circ \pi)$  is in the domain of  $A$ , and hence by [\(A.34\)](#) in the domain of  $A^*$ . Since  $\tilde{f} = \bar{\nabla}^* X$ , [Lemma A.22](#) and [Lemma A.20](#) yield

$$\begin{aligned} \|A^*(\bar{v} \cdot X \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 &= \|\tilde{f}\|_{L^2(\bar{\mu})}^2 + \|A^*(\bar{v} \cdot X \circ \pi) - \bar{\nabla}^* X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \\ &\leq \|\tilde{f}\|_{L^2(\bar{\mu})}^2 + 2\|\bar{\nabla} X\|_{L^2(\bar{\mu})}^2 \leq (1 + 2c_1(T)) \|\tilde{f}\|_{L^2(\bar{\mu})}^2, \end{aligned}$$

so that the second summand can be estimated as

$$(A(\Pi_v f - f), \bar{v} \cdot X \circ \pi)_{L^2(\bar{\mu} \otimes \kappa)} \leq (1 + 2c_1(T))^{\frac{1}{2}} \|f - \Pi_v f\|_{L^2(\bar{\mu} \otimes \kappa)} \|\tilde{f}\|_{L^2(\bar{\mu} \otimes \kappa)}.$$

Putting things together, we hence obtain by [\(A.39\)](#),

$$\|\tilde{f}\|_{L^2(\bar{\mu})} \leq c_0(T)^{\frac{1}{2}} \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} + (1 + 2c_1(T))^{\frac{1}{2}} \|f - \Pi_v f\|_{L^2(\bar{\mu} \otimes \kappa)},$$

so that by [\(A.38\)](#),

$$\|f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq 2c_0(T) \|Af\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + (3 + 4c_1(T)) \|f - \Pi_v f\|_{L^2(\bar{\mu} \otimes \kappa)}^2. \quad \square$$

**Remark A.24.** In order to generalise the above proof technique to other second-order lifts of an overdamped Langevin diffusion, the key step is bounding  $\|A^*(\bar{v} \cdot X \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)}$  by  $\|\tilde{f}\|_{L^2(\bar{\mu})}$ . Further care regarding the domain of the operator  $A^*$  must also be taken in more general settings.

### A.5.5. Optimality of randomised Hamiltonian Monte Carlo

By [Theorem A.17](#), the space-time Poincaré inequality for the generator  $\hat{L}$  of the Hamiltonian flow immediately gives the following quantitative hypocoercivity result for the transition semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  of randomised Hamiltonian Monte Carlo with generator

$$\hat{L}^{(\gamma)} = \hat{L} + \gamma(\Pi_v - I).$$

**Corollary A.25** (Exponential contractivity of randomised HMC). *For any  $\gamma, T \in (0, \infty)$ , the semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  is exponentially contractive in  $T$ -average with rate*

$$\nu = \frac{\gamma}{2c_0(T)\gamma^2 + 3 + 4c_1(T)}, \quad (\text{A.40})$$

where  $c_0(T)$  and  $c_1(T)$  are the constants from [Lemma A.20](#).

*Proof.* The expression for  $\nu$  follows immediately from [Theorems A.17](#) and [A.23](#).  $\square$

**Remark A.26** (Optimal refresh rate). For fixed  $T \in (0, \infty)$ , the rate  $\nu$  in [\(A.40\)](#) is maximal when  $\gamma$  is chosen such that the two summands in the denominator coincide, i.e. for  $\gamma = \sqrt{(3 + 4c_1(T))/(2c_0(T))}$ . In this case,

$$\frac{1}{\nu} = \sqrt{8c_0(T)(3 + 4c_1(T))}. \quad (\text{A.41})$$

To obtain explicit bounds, we set  $T = 3/\sqrt{m}$ . By [\(A.40\)](#) and [Remark A.21](#), this yields

$$\frac{1}{\nu} = \frac{482}{\sqrt{m}} \left( \frac{\gamma}{\sqrt{m}} + \left( 2124 + \frac{1}{3} + 300 \frac{\kappa_-}{m} \right) \frac{\sqrt{m}}{482\gamma} \right) < \frac{482}{\sqrt{m}} \left( \frac{\gamma}{\sqrt{m}} + 5 \left( 1 + \frac{\kappa_-}{7m} \right) \frac{\sqrt{m}}{\gamma} \right). \quad (\text{A.42})$$

The decay rate  $\nu$  is maximal for

$$\gamma = \sqrt{((2124 + \frac{1}{3})m + 300\kappa_-)/482}, \quad (\text{A.43})$$

and in this case,

$$\frac{1}{\nu} = 2 \cdot \frac{482\gamma}{m} < \frac{2024}{\sqrt{m}} \cdot \sqrt{1 + \frac{\kappa_-}{7m}}. \quad (\text{A.44})$$

**Corollary A.27** (*C*-optimality of randomised HMC). *Let  $c \in [0, \infty)$ . On the class of all potentials satisfying [Assumption A.2](#) with  $\kappa_- \leq cm$ , randomised Hamiltonian Monte Carlo with  $\gamma$  given by [\(A.43\)](#) is a *C*-optimal lift of the corresponding overdamped Langevin diffusion with*

$$C = 2\sqrt{2}(2024\sqrt{1 + c/7} + 3).$$

*More generally, for  $A \in [1, \infty)$ , randomised HMC is a  $C(A)$ -optimal lift with*

$$C(A) = 2\sqrt{2}(482 \cdot (6 + 5c/7)A + 3)$$

*for any  $\gamma$  satisfying  $1/A \leq \frac{\gamma}{\sqrt{m}} \leq A$ .*

*Proof.* For  $\gamma$  given by [\(A.43\)](#), the bound in [\(A.44\)](#) yields

$$t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq T + \nu^{-1} \leq \frac{1}{\sqrt{m}}(2024\sqrt{1 + c/7} + 3),$$

whereas the relaxation time of the underlying overdamped Langevin diffusion is  $t_{\text{rel}}(P) = \frac{1}{m}$ . This proves the first claim, and the second claim follows similarly from [\(A.42\)](#).  $\square$

**Remark A.28** (Optimal lifts with and without convexity). The explicit values of the constants in [Corollary A.27](#) are certainly not optimal. In particular, we expect that the constants [\(A.33\)](#) in the quantitative divergence lemma can still be improved substantially. Therefore, [Corollary A.27](#) seems to indicate that in the convex or mildly non-convex case, randomised HMC with optimised choice of  $\gamma$  is not too far from being an optimal lift of overdamped Langevin. On the other hand, a corresponding statement cannot be true in

very non-convex cases. For example, for a double well-potential  $U(x) = \beta(x^2 - 1)^2$ , by the Arrhenius law, the relaxation times of overdamped Langevin and randomised HMC are both expected to be asymptotically of exponential order  $\exp(\beta(U(0) - U(1)))$  as  $\beta \rightarrow \infty$ . Therefore, a square root acceleration cannot take place for large  $\beta$ , and the dependence of the bounds on the non-convexity parameter  $c$  is natural.

**Remark A.29** (Hypocoercivity approaches). Based on the work of Hérau [132], Dolbeault, Mouhot and Schmeiser [86, 87] developed a method for proving hypocoercivity for a general class of linear kinetic equations, covering the kinetic Fokker-Planck and the linear Boltzmann equations and thereby Langevin dynamics and randomised HMC. Their method is based on a modification of the  $L^2(\mu \otimes \kappa)$ -norm by a suitable perturbation which allows to recover dissipation in all variables. As pointed out to us by Lihan Wang, if  $\hat{L}$  is the generator of the Hamiltonian flow, then Assumption H3 in [87] resembles our first-order condition (A.7) for a second-order lift, while Assumption H2 in [87] is implied by the second-order condition (A.8) together with Assumption A.2(ii) on the spectral gap of the underlying diffusion. Since the entropy functional in [87] does not directly capture the dynamics of the process, one cannot expect optimal bounds on the convergence rate. Indeed, as calculated in [61, Appendix B], assuming a lower bound on the Hessian of the potential and in the case that  $\gamma$  is proportional to  $\sqrt{m}$ , the approach in [87] seems to yield a rate of order  $m^{5/2}$  in the limit as  $m \rightarrow 0$  instead of the optimal order  $m^{1/2}$ .

## Statements and Declarations

**Funding.** Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – GZ 2047/1, Projekt-ID 390685813. The authors were funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany’s Excellence Strategy – GZ 2047/1, Project-ID 390685813.

**Competing Interests.** The authors declare that they have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

**Acknowledgements.** The authors would like to thank Lihan Wang, Arnaud Guillin and Léo Hahn for many helpful discussions.



## Appendix B.

# Space-time divergence lemmas and optimal non-reversible lifts of diffusions on Riemannian manifolds with boundary

A. Eberle and F. Lörler. ‘Space-time divergence lemmas and optimal non-reversible lifts of diffusions on Riemannian manifolds with boundary’. In: *J. Funct. Anal.* 291.9 (2026). DOI: [10.1016/j.jfa.2026.111605](https://doi.org/10.1016/j.jfa.2026.111605)

**Abstract:** Non-reversible lifts reduce the relaxation time of reversible diffusions at most by a square root. For reversible diffusions on domains in Euclidean space, or, more generally, on a Riemannian manifold with boundary, non-reversible lifts are in particular given by the Hamiltonian flow on the tangent bundle, interspersed with random velocity refreshments, or perturbed by Ornstein-Uhlenbeck noise, and reflected at the boundary. In order to prove that for certain choices of parameters, these lifts achieve the optimal square-root reduction up to a constant factor, precise upper bounds on relaxation times are required. A key tool for deriving such bounds by space-time Poincaré inequalities is a quantitative space-time divergence lemma. Extending previous work of Cao, Lu and Wang, we establish such a divergence lemma with explicit constants for general locally convex domains with smooth boundary in Riemannian manifolds satisfying a lower, not necessarily positive, curvature bound. As a consequence, we prove optimality of the lifts described above up to a constant factor, provided the deterministic transport part of the dynamics and the noise are adequately balanced. Our results show for example that an integrated Ornstein-Uhlenbeck process on a locally convex domain with diameter  $d$  achieves a relaxation time of the order  $d$ , whereas, in general, the Poincaré constant of the domain is of the order  $d^2$ .

**Keywords:** Lift; divergence lemma; Riemannian manifold; hypocoercivity; Langevin dynamics; Hamiltonian Monte Carlo; convex domain; reflection.

**MSC Subject Classification:** 60J25, 60J35, 58J65, 58J05, 35H10.

### B.1. Introduction

Long-time convergence to equilibrium of strongly continuous contraction semigroups plays an important role in many areas of probability and analysis, ranging from con-

vergence of Markov processes with applications in sampling to kinetic equations, ergodic theory, and non-equilibrium statistical physics. The degenerate case in which the generator of the semigroup is not an elliptic but only hypoelliptic operator has attracted much attention in the past years, and the study of convergence of such dynamics is known as hypocoercivity [248]. Recently, a variational approach based on space-time Poincaré inequalities pioneered by Albritton, Armstrong, Mourrat and Novack [6] has proved successful in obtaining sharp and quantitative bounds on the rates of convergence. This approach has further been developed and applied in various scenarios by Cao, Lu and Wang [61, 186]. Furthermore, by casting it in a framework of second-order lifts [99], the authors have simplified the approach, and an associated lower bound on the relaxation time of lifts has raised the question of existence of optimal lifts achieving this lower bound.

The goal of this paper is to prove upper bounds of the optimal order on the relaxation time of Langevin dynamics and randomised Hamiltonian Monte Carlo with appropriately chosen parameters on locally convex Riemannian manifolds with boundary that satisfy a lower curvature bound. To this end, we provide a quantitative space-time divergence lemma, which is the key ingredient in the proof of space-time Poincaré inequalities. By considering Riemannian manifolds with boundary as spatial domains, our result allows for the treatment of non-reversible, degenerate dynamics on Riemannian manifolds, as well as dynamics with reflective boundary behaviour. In particular, it shows that, up to a constant factor, the processes described above are optimal lifts of an overdamped Langevin diffusion with reflection at the boundary, provided the deterministic transport part of the dynamics and the noise are adequately balanced.

Probability distributions supported on Riemannian manifolds arise in many applications; through constrained dynamics in physics [174, 176], as natural parameter spaces in statistics [70], in Bayesian inference [80, 181], through introduction of an artificial Riemannian geometry on  $\mathbb{R}^n$  [167], or in volume computation of convex bodies [112, 166], to name a few. Sampling from such distributions is a challenging task, and theoretical convergence bounds for many proposed algorithms [4, 59, 69, 116, 154] are still lacking. Our quantitative convergence bounds provide a first step towards the theoretical understanding of sampling algorithms based on discretisations of continuous-time processes on Riemannian manifolds. Similarly, dynamics constrained to a convex domain arise in sampling problems with restrictions [5, 161]. The relaxation time of reflected Brownian motion on a convex domain with diameter  $d$  is the Poincaré constant of the uniform distribution on this domain, which is of the order  $d^2$  in general [180]. Our result shows that critically damped Langevin dynamics, which reduces to an integrated Ornstein-Uhlenbeck process reflected at the boundary, and randomised Hamiltonian Monte Carlo with appropriately chosen refresh rate, which reduces to a billiards process with velocity refreshment, achieve a relaxation time of order  $d$ .

Let us first give a brief review of the concept of second-order lifts of reversible diffusions introduced in [99]. Consider a time-homogeneous Markov process  $(X_t, V_t)_{t \geq 0}$  with invariant probability measure  $\hat{\mu}$  on the tangent bundle  $TM$  of a Riemannian manifold  $(M, g)$  with boundary and let  $\pi: TM \rightarrow M$  be the natural projection. Furthermore,

let  $(Z_t)_{t \geq 0}$  be a reversible diffusion process on  $M$  with invariant probability measure  $\mu = \hat{\mu} \circ \pi^{-1}$ . Their associated transition semigroups acting on  $L^2(\hat{\mu})$  and  $L^2(\mu)$  are  $(\hat{P}_t)_{t \geq 0}$  and  $(P_t)_{t \geq 0}$ , and their infinitesimal generators are denoted by  $(\hat{L}, \text{Dom}(\hat{L}))$  and  $(L, \text{Dom}(L))$ , respectively. Then  $(\hat{P}_t)_{t \geq 0}$  is a *second-order lift* of  $(P_t)_{t \geq 0}$  if

$$\text{Dom}(L) \subseteq \{f \in L^2(\mu) : f \circ \pi \in \text{Dom}(\hat{L})\}, \quad (\text{B.1})$$

and for all  $f, g \in \text{Dom}(L)$  we have

$$\int_{\text{TM}} \hat{L}(f \circ \pi)(g \circ \pi) d\hat{\mu} = 0 \quad (\text{B.2})$$

and

$$\frac{1}{2} \int_{\text{TM}} \hat{L}(f \circ \pi) \hat{L}(g \circ \pi) d\hat{\mu} = \mathcal{E}(f, g), \quad (\text{B.3})$$

where  $\mathcal{E}(f, g) = -\int_M f L g d\mu$  is the Dirichlet form associated to  $(L, \text{Dom}(L))$ .

Since the process  $(Z_t)_{t \geq 0}$  is assumed to be reversible, its generator is self-adjoint, and the associated semigroup satisfies

$$\|P_t\|_{L_0^2(\mu) \rightarrow L_0^2(\mu)} = \exp(-\lambda t) \quad (\text{B.4})$$

with the decay rate  $\lambda$  given by the spectral gap

$$\text{gap}(L) = \inf\{\text{Re}(\alpha) : \alpha \in \text{spec}(-L|_{L_0^2(\mu)})\} \quad (\text{B.5})$$

of  $(L, \text{Dom}(L))$  in the Hilbert space  $L^2(\mu)$ . Here  $L_0^2(\mu) = \{f \in L^2(\mu) : \int_M f d\mu = 0\}$  denotes the subspace of  $L^2(\mu)$  of mean-zero functions. In contrast, the strong generator of a *non-reversible* Markov process is no longer self-adjoint and the operator norm of the associated semigroup is no longer a pure exponential. In particular, the spectral gap only coincides with the asymptotic decay rate of the semigroup under additional assumptions ensuring a spectral mapping theorem relating the spectrum of  $P_t$  to that of  $L$ , see [104]. This motivates the introduction of the non-asymptotic relaxation time

$$t_{\text{rel}}(\hat{P}) = \inf\{t \geq 0 : \|\hat{P}_t f\|_{L^2(\mu)} \leq e^{-1} \|f\|_{L^2(\mu)} \text{ for all } f \in L_0^2(\mu)\}$$

of  $(\hat{P}_t)_{t \geq 0}$ . For reversible processes, (B.4) shows that the relaxation time coincides with the usual definition as the inverse of the spectral gap. A crucial consequence of the lift property is the lower bound

$$t_{\text{rel}}(\hat{P}) \geq \frac{1}{2\sqrt{2}} \sqrt{t_{\text{rel}}(P)}$$

on the relaxation time of an arbitrary lift, see [99, Theorem 11], which shows that convergence measured by the relaxation time can at most be accelerated by a square root through lifting. A lift  $(\hat{P}_t)_{t \geq 0}$  is hence called *C-optimal* if

$$t_{\text{rel}}(\hat{P}) \leq \frac{C}{2\sqrt{2}} \sqrt{t_{\text{rel}}(P)}, \quad (\text{B.6})$$

## Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

i.e. if it achieves this lower bound up to a constant factor  $C > 0$ .

Proving  $C$ -optimality of a lift requires precise upper bounds on the relaxation time. The approach recently introduced by Albritton, Armstrong, Mourrat and Novack [6] relies on space-time Poincaré inequalities in order to obtain exponential decay

$$\frac{1}{T} \int_t^{t+T} \|\hat{P}_s f\|_{L^2(\hat{\mu})}^2 ds \leq e^{-\nu t} \frac{1}{T} \int_0^T \|\hat{P}_s f\|_{L^2(\hat{\mu})}^2 ds \quad \text{for all } f \in L_0^2(\hat{\mu}) \text{ and } t \geq 0$$

with rate  $\nu$  of time-averages over some fixed length  $T > 0$ , which yields  $t_{\text{rel}}(\hat{P}) \leq \frac{1}{\nu} + T$ . The key tool in proving a space-time Poincaré inequality is a space-time divergence lemma which can be formulated in terms of the generator  $(L, \text{Dom}(L))$  of the reversible process alone. Namely, letting  $\lambda$  denote the Lebesgue measure on  $[0, T]$ , any  $f \in L_0^2(\lambda \otimes \mu)$  can be written as

$$f = \partial_t h - Lg$$

with functions  $h \in H^{1,2}(\lambda \otimes \mu)$  and  $g \in H^{2,2}(\lambda \otimes \mu)$  satisfying Dirichlet boundary conditions in time, as well as  $g(t, \cdot) \in \text{Dom}(L)$  for  $\lambda$ -almost every  $t \in [0, T]$ , together with the regularity estimates

$$\begin{aligned} \|h\|_{L^2(\bar{\mu})}^2 + \|\nabla g\|_{L^2(\bar{\mu})}^2 &\leq c_0(T) \|f\|_{L^2(\bar{\mu})}^2, \\ \|\bar{\nabla} h\|_{L^2(\bar{\mu})}^2 + \|\bar{\nabla} \nabla g\|_{L^2(\bar{\mu})}^2 &\leq c_1(T) \|f\|_{L^2(\bar{\mu})}^2 \end{aligned}$$

on  $h$  and  $g$ , where  $\bar{\nabla} = (\partial_t, \nabla)$  denotes the space-time gradient. One goal of this work is to provide solutions to these space-time divergence equations with quantitative expressions for  $c_0(T)$  and  $c_1(T)$  in the case where  $\mu$  has a density  $\exp(-U)$  with respect to the Riemannian volume measure and  $(Z_t)_{t \geq 0}$  is a Riemannian overdamped Langevin diffusion with generator

$$L = \Delta - \langle \nabla U, \nabla \rangle$$

with Neumann boundary conditions. In this case,  $f$  can be written as the space-time divergence of the vector field with components  $-h$  and  $\nabla g$ , which motivated the term ‘divergence lemma’ in the literature. This allows for the treatment of second-order lifts of Riemannian overdamped Langevin dynamics, which include Riemannian randomised Hamiltonian Monte Carlo and Langevin dynamics with reflective boundary behaviour.

We begin in the following section by formally introducing the setting and summarising the Riemannian geometry involved. In particular, we provide an extension of the Reilly formula which arises through an integration-by-parts of the Bochner formula for  $L$  and a careful treatment of the boundary terms. Together with a lower curvature bound, it allows to upper bound the Hessian of smooth functions on  $M$ . In [Section B.3](#), we prove the divergence lemma with explicit quantitative bounds on the constructed functions. By considering a splitting of  $L^2(\mu)$  into high and low modes associated to a spectral decomposition of  $L$ , as well as into symmetric and antisymmetric functions, we obtain constants of the optimal order conjectured in [61]. We demonstrate an application of the divergence lemma in [Section B.4](#) by showing that, with critical choice of parameters,

Langevin dynamics and randomised Hamiltonian Monte Carlo on a Riemannian manifold with boundary satisfying a lower curvature bound are optimal lifts of the corresponding overdamped Langevin diffusion. A brief overview of Riemannian geometry on manifolds with boundary is given in [Section B.5](#).

## B.2. Preliminaries and the generalised Reilly formula

Consider a complete connected and oriented Riemannian manifold  $(M, g)$  with (possibly empty) boundary. In particular, the boundary  $\partial M$  of  $M$  is a Riemannian submanifold without boundary, and for simplicity we again denote the induced metric on  $\partial M$  by  $g$ . We equip the boundary with the outward unit-normal vector field  $N$ , and for  $x \in M$  and  $v, w \in T_x M$  we write  $\langle v, w \rangle_x$  for  $g_x(v, w)$ . For a short overview over manifolds with boundary, see [Section B.5.1](#). We equip  $M$  with a probability measure  $\mu$  absolutely continuous with respect to the Riemannian volume measure  $\nu_M$ , i.e.

$$\mu(dx) = \exp(-U(x)) \nu_M(dx)$$

for some function  $U \in C^2(M)$  with  $\int_M \exp(-U) d\nu_M = 1$ .

Let  $\Delta: C^\infty(M) \rightarrow C^\infty(M)$ ,  $u \mapsto \operatorname{div}(\nabla u)$ , where  $\nabla$  is the Riemannian gradient and  $\operatorname{div}$  the Riemannian divergence, be the Laplace-Beltrami operator. Then the weighted Laplace-Beltrami operator  $L$  is defined as

$$Lu = \Delta u - \langle \nabla U, \nabla u \rangle \quad \text{for all } u \in C^\infty(M).$$

The operator  $L$  satisfies the generalised Bochner-Lichnerowicz-Weitzenböck formula

$$\frac{1}{2}L|\nabla u|^2 = |\nabla^2 u|^2 + \langle \nabla u, \nabla Lu \rangle + (\operatorname{Ric} + \nabla^2 U)(\nabla u, \nabla u) \quad (\text{B.7})$$

for all  $u \in C^\infty(M)$ , see [\[19, Proposition 3\]](#), where  $\nabla^2 u$  is the Hessian, or second covariant derivative, of  $u$ , and  $\operatorname{Ric}$  is the Ricci curvature tensor of  $M$ . Letting  $\partial_i$  denote the  $i$ -th coordinate vector field,  $|\nabla^2 u|^2 = \sum_{i,j=1}^d (\nabla_{\partial_i, \partial_j}^2 u)^2$  is the squared Frobenius norm of the tensor  $\nabla^2 u$ . Note that (B.7) reduces to the usual Bochner identity in case  $U$  is constant and thus  $L = \Delta$ , see e.g. [\[146\]](#). Let

$$C_c^\infty(M) = \{u \in C^\infty(M) : \operatorname{supp}(u) \text{ is compact in } M\}$$

denote the set of smooth, compactly supported functions on  $M$ , where elements of  $C_c^\infty(M)$  are *not* required to vanish at the boundary  $\partial M$ . Integrating the Bochner identity by parts with respect to the Riemannian volume  $\nu_M$  yields the Reilly formula [\[218\]](#)

$$\begin{aligned} & \int_M (|\nabla^2 u|^2 - (\Delta u)^2 + \operatorname{Ric}(\nabla u, \nabla u)) d\nu_M \\ &= \int_{\partial M} \left( H(\partial_N u)^2 - 2(\partial_N u)\Delta^{\partial M} u + h(\nabla^{\partial M} u, \nabla^{\partial M} u) \right) d\nu_{\partial M}, \end{aligned}$$

for all  $u \in C_c^\infty(M)$ , see also [71, Lemma 7.10] for a proof. Here  $h$  is the scalar second fundamental form,  $H$  is the mean curvature of the boundary, see Section B.5.2, and  $\partial_N u = \langle \nabla u, N \rangle$ . We use the notation  $\nabla^{\partial M}$  and  $\Delta^{\partial M}$  for the Riemannian gradient and the Laplace-Beltrami operator on the boundary manifold  $(\partial M, g)$ , respectively, and the induced Riemannian volume measure on  $\partial M$  is denoted by  $\nu_{\partial M}$ . Denoting the induced measure on the boundary by

$$\mu_{\partial M}(dx) = \exp(-U(x)) \nu_{\partial M}(dx),$$

the operator  $L$  satisfies the integration-by-parts identity

$$\int_M \varphi L\psi \, d\mu = \int_{\partial M} \varphi \langle \nabla \psi, N \rangle \, d\mu_{\partial M} - \int_M \langle \nabla \varphi, \nabla \psi \rangle \, d\mu \quad (\text{B.8})$$

for all smooth, compactly supported functions  $\varphi, \psi \in C_c^\infty(M)$ . Note that  $\mu_{\partial M}$  is not necessarily a probability measure. Now integrating (B.7) by parts with respect to  $\mu$  yields the following generalisation of the Reilly formula due to Ma and Du [187], which reduces to the usual Reilly formula in case  $U$  is constant. For completeness and the convenience of the reader, we include a proof in Section B.5.4.

**Lemma B.1** (Generalised Reilly formula). *For any  $u \in C_c^\infty(M)$ ,*

$$\begin{aligned} & \int_M (|\nabla^2 u|^2 - (Lu)^2 + (\text{Ric} + \nabla^2 U)(\nabla u, \nabla u)) \, d\mu \\ &= \int_{\partial M} \left( (H + \partial_N U)(\partial_N u)^2 - 2(\partial_N u)L^{\partial M} u + h(\nabla^{\partial M} u, \nabla^{\partial M} u) \right) \, d\mu_{\partial M}, \end{aligned}$$

where  $L^{\partial M} = \Delta^{\partial M} - \langle \nabla^{\partial M} U, \nabla^{\partial M} \rangle$ .

Lemma B.1 has been applied by Kolesnikov and Milman [149, 150, 151] to obtain Brascamp-Lieb-type and log-Sobolev functional inequalities on weighted Riemannian manifolds with boundary.

**Remark B.2.** The set of corner points  $C$  of a Riemannian manifold  $M$  with corners, see Section B.5.1, forms a set of measure 0 with respect to the Riemannian volume measure  $\nu_M$ , and  $M \setminus C$  is a Riemannian manifold with boundary. Hence Lemma B.1 holds analogously for Riemannian manifolds with corners when replacing the domain of integration on the right-hand side with the boundary  $\partial(M \setminus C)$  of  $M \setminus C$ .

The Sobolev space  $H^{1,2}(M, \mu)$  on  $M$  is defined as the closure of  $C^\infty(M) \cap L^2(\mu)$  in  $L^2(\mu)$  with respect to the norm

$$\|u\|_{H^{1,2}(\mu)}^2 = \|u\|_{L^2(\mu)}^2 + \|\nabla u\|_{L^2(\mu)}^2. \quad (\text{B.9})$$

Similarly, one defines the second-order Sobolev space  $H^{2,2}(M, \mu)$  on  $M$  as the closure of  $C^\infty(M) \cap L^2(\mu)$  with respect to the norm

$$\|u\|_{H^{2,2}(\mu)}^2 = \|u\|_{L^2(\mu)}^2 + \|\nabla u\|_{L^2(\mu)}^2 + \|\nabla^2 u\|_{L^2(\mu)}^2, \quad (\text{B.10})$$

where  $\|\nabla^2 u\|_{L^2(\mu)}$  is the  $L^2(\mu)$ -norm of  $|\nabla^2 u|$ .

We make the following assumptions on  $(M, g)$  and the probability measure  $\mu$ .

**Assumption B.1.**

- (i) **Local convexity.** The scalar second fundamental form  $h$  on  $\partial M$  is negative semi-definite, i.e.

$$h(v, v) \leq 0 \quad \text{for all } x \in \partial M \text{ and } v \in T_x \partial M.$$

- (ii) **Lower curvature bound.** There is a constant  $\rho \in [0, \infty)$  such that  $\text{Ric} + \nabla^2 U \geq -\rho$  on  $M$ , i.e.

$$\text{Ric}(v, v) + \nabla^2 U(v, v) \geq -\rho|v|^2 \quad \text{for all } x \in M \text{ and } v \in T_x M.$$

- (iii) **Poincaré inequality.** The probability measure  $\mu$  satisfies a Poincaré inequality with constant  $m^{-1}$ , i.e.

$$\int_M f^2 d\mu \leq \frac{1}{m} \int_M |\nabla f|^2 d\mu \quad \text{for all } f \in H^{1,2}(M, \mu) \text{ with } \int_M f d\mu = 0.$$

The local convexity assumption (i) is implied by geodesic convexity of the manifold, i.e. the existence of a minimising geodesic between any two interior points that lies in the interior, while the converse is not true in general [7, 38, 39]. Note that the lower bound (ii) on the generalised curvature tensor  $\text{Ric} + \nabla^2 U$  is also known as the Bakry-Émery curvature-dimension condition  $\text{CD}(-\rho, \infty)$  [20].

**Remark B.3.** Let  $M$  be a closed, connected subset of  $\mathbb{R}^d$  with smooth boundary. Then  $M$  is a Riemannian manifold with boundary when equipped with the Euclidean metric and Assumption B.1(i) is equivalent to convexity of  $M$ . Since the Euclidean metric is flat, Assumption B.1(ii) simplifies to  $\nabla^2 U \geq -\rho$  on  $M$ . Hence Assumption B.1 is satisfied by probability measures on convex subsets of Euclidean space satisfying a Poincaré inequality and a lower curvature bound on their potential.

By the integration-by-parts identity (B.8), when equipped with the Neumann boundary conditions

$$\mathcal{D} = \{u \in C_c^\infty(M) : \partial_N u = 0 \text{ on } \partial M\}, \quad (\text{B.11})$$

we have  $Lu = -\nabla^* \nabla u$  for all  $u \in \mathcal{D}$ , where the adjoint is in  $L^2(\mu)$ . Therefore, the densely defined linear operator  $(L, \mathcal{D})$  on  $L^2(\mu)$  is symmetric and negative semi-definite.

**Corollary B.4.** Let Assumption B.1 hold.

- (i) For all  $u \in \mathcal{D}$ ,

$$\|\nabla^2 u\|_{L^2(\mu)}^2 \leq \|Lu\|_{L^2(\mu)}^2 + \rho \|\nabla u\|_{L^2(\mu)}^2. \quad (\text{B.12})$$

- (ii) The operator  $(L, \mathcal{D})$  is essentially self-adjoint. Its unique self-adjoint extension  $(L, \text{Dom}(L))$  satisfies  $\text{Dom}(L) \subseteq H^{2,2}(M, \mu)$  and (B.12) holds for all  $u \in \text{Dom}(L)$ .

*Proof.* (i) The estimate (B.12) follows directly from the generalised Reilly formula stated in Lemma B.1, using Assumption B.1(i) and (ii) and the vanishing normal derivative at the boundary.

(ii) In the case of constant potential, the essential self-adjointness of the Laplace-Beltrami operator  $(\Delta, \mathcal{D})$  with Neumann boundary conditions is classical and e.g. shown in [240, Prop. 8.2.5]. Another proof following an approach by Chernoff [68] is given in [33]. Due to the integration-by-parts identity (B.8) and the regularity estimate (B.12) guaranteed by the lower bound on the curvature of  $U$ , the general case  $L = \Delta - \langle \nabla U, \nabla \rangle$  can be treated analogously.

Finally, since  $\|\nabla u\|_{L^2(\mu)}^2 \leq \|u\|_{L^2(\mu)} \|Lu\|_{L^2(\mu)}$  for all  $u \in \mathcal{D}$ , (B.12) yields

$$\|\nabla^2 u\|_{L^2(\mu)}^2 \leq (1 + \rho)(\|u\|_{L^2(\mu)} + \|Lu\|_{L^2(\mu)})^2$$

for all  $u \in \mathcal{D}$ . In particular, this implies  $\text{Dom}(L) \subseteq H^{2,2}(M, \mu)$  and (B.12) holds for all  $u \in \text{Dom}(L)$  by an approximation argument.  $\square$

We make the following technical assumption which allows us to work with a spectral decomposition of  $L$ . It can surely be relaxed, see [55], yet is satisfied in many cases, for instance if  $M$  is compact.

**Assumption B.2.** The spectrum of  $(L, \text{Dom}(L))$  on  $L^2(\mu)$  is discrete.

Let  $L_0^2(\mu) = \{f \in L^2(\mu) : \int_M f d\mu = 0\}$  denote the subspace of  $L^2(\mu)$  of mean-zero functions. By the Poincaré inequality Assumption B.1(iii),

$$-L|_{L_0^2(\mu)} : \text{Dom}(L) \cap L_0^2(\mu) \rightarrow L_0^2(\mu)$$

has a spectral gap and hence admits a bounded linear inverse

$$G : L_0^2(\mu) \rightarrow \text{Dom}(L) \cap L_0^2(\mu) \tag{B.13}$$

with operator norm bounded by  $\frac{1}{m}$ . Assumption B.2 allows us to consider an orthonormal basis  $\{e_k : k \in \mathbb{N}_0\}$  of  $L^2(\mu)$  consisting of eigenfunctions of  $L$  with eigenvalues  $-\alpha_k^2$ , where  $e_0 = 1$  and  $\alpha_0 = 0$ . In particular,  $\inf\{\alpha_k^2 : k \in \mathbb{N}\} \geq m$  and  $e_k \in \text{Dom}(L)$  implies  $\partial_N e_k = 0$  on  $\partial M$ . The operator  $G = (-L|_{L_0^2(\mu)})^{-1}$  satisfies

$$Ge_k = \frac{1}{\alpha_k^2} e_k \quad \text{for all } k \in \mathbb{N}.$$

### B.3. The space-time divergence lemma

We adjoin a time component to the spatial component and hence consider time-space domains  $\bar{M} = [0, T] \times M$ , where  $T \in (0, \infty)$  is fixed and, as before,  $(M, g)$  is a Riemannian manifold with boundary such that Assumption B.1 is satisfied. Note that  $\bar{M}$  is a

### B.3. The space-time divergence lemma

Riemannian manifold with corners with metric  $\bar{g}$  given by the product of the Euclidean metric and  $g$ . Its corner points are  $\{0, T\} \times \partial M$ . The tangent bundle  $T\bar{M}$  is canonically isomorphic to  $([0, T] \times \mathbb{R}) \oplus TM$ , and we identify elements  $V \in T\bar{M}$  of the former with tuples  $(V_0, V_1)$  with  $V_0 \in [0, T] \times \mathbb{R}$  and  $V_1 \in TM$ .

On  $\bar{M}$  we consider the measure

$$\bar{\mu} = \lambda \otimes \mu,$$

where  $\lambda$  is the Lebesgue measure restricted to  $[0, T]$ . We use the notation  $L^2(\bar{\mu}) = L^2(\bar{M}, \bar{\mu})$  and  $L_0^2(\bar{\mu}) = \{f \in L^2(\bar{\mu}) : \int_{\bar{M}} f d\bar{\mu} = 0\}$  for the subspace of  $L^2(\bar{\mu})$  of mean-zero functions. The Riemannian gradient on  $\bar{M}$  is denoted by  $\bar{\nabla}$  and its formal  $L^2(\bar{\mu})$ -adjoint by  $\bar{\nabla}^*$ . The latter acts on a vector field  $X \in \Gamma_{C^1}(T\bar{M})$  by

$$\begin{aligned} \bar{\nabla}^* X &= -\partial_t X_0 + \nabla^* X_1 \\ &= -\partial_t X_0 - \operatorname{div}(X_1) + \langle \nabla U, X_1 \rangle. \end{aligned}$$

For  $m \in \{1, 2\}$ , the Sobolev spaces  $H^{m,2}(\bar{\mu})$  are defined in analogy to (B.9) and (B.10). Similarly, the Sobolev space  $H_D^{m,2}(\bar{\mu})$  with Dirichlet boundary conditions in time is the closure of

$$\{u \in C^\infty(\bar{M}) \cap L^2(\bar{\mu}) : u(0, \cdot) = u(T, \cdot) = 0\}$$

in  $L^2(\bar{\mu})$  with respect to  $\|\cdot\|_{H^{m,2}(\bar{\mu})}$ .

**Theorem B.5** (Quantitative divergence lemma). *Suppose that Assumptions B.1 and B.2 are satisfied. Then for any  $T \in (0, \infty)$ , there exist constants  $c_0(T), c_1(T) \in (0, \infty)$  such that for any  $f \in L_0^2(\bar{\mu})$  there are functions*

$$h \in H_D^{1,2}(\bar{\mu}) \quad \text{and} \quad g \in H_D^{2,2}(\bar{\mu}) \tag{B.14}$$

satisfying

$$g(t, \cdot) \in \operatorname{Dom}(L) \quad \text{for } \lambda\text{-almost all } t \in [0, T] \tag{B.15}$$

such that

$$f = \partial_t h - Lg \tag{B.16}$$

and the functions  $h$  and  $g$  satisfy

$$\|h\|_{L^2(\bar{\mu})}^2 + \|\nabla g\|_{L^2(\bar{\mu})}^2 \leq c_0(T) \|f\|_{L^2(\bar{\mu})}^2, \tag{B.17}$$

$$\|\bar{\nabla} h\|_{L^2(\bar{\mu})}^2 + \|\bar{\nabla} \nabla g\|_{L^2(\bar{\mu})}^2 \leq c_1(T) \|f\|_{L^2(\bar{\mu})}^2, \tag{B.18}$$

where

$$\begin{aligned} c_0(T) &= 2T^2 + 43 \frac{1}{m} \quad \text{and} \\ c_1(T) &= 290 + \frac{991}{mT^2} + 43 \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \rho. \end{aligned} \tag{B.19}$$

**Remark B.6** (Relation to the classical divergence lemma). In Theorem B.5, writing

$$X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix} \quad \text{with } X_0 = -h \text{ and } X_1 = \nabla g$$

yields the divergence equation

$$f = \bar{\nabla}^* X \quad (\text{B.20})$$

together with the bounds

$$\|X\|_{L^2(\bar{\mu})}^2 \leq c_0(T) \|f\|_{L^2(\bar{\mu})}^2, \quad (\text{B.21})$$

$$\|\bar{\nabla} X\|_{L^2(\bar{\mu})}^2 \leq c_1(T) \|f\|_{L^2(\bar{\mu})}^2. \quad (\text{B.22})$$

The vector field  $X$  satisfies Dirichlet boundary conditions on the time boundary, i.e. it vanishes on  $\{0, T\} \times M$ , and it vanishes in the direction normal to the spatial boundary, i.e.  $\langle X_1, N \rangle = 0$  on  $[0, T] \times \partial M$ . In comparison, in the classical Lions' divergence lemma as in [10], one imposes Dirichlet boundary conditions on the solution to (B.20) on the whole boundary. While Theorem B.5 only requires the vector field to vanish in the direction normal to the spatial boundary, we obtain the additional gradient structure of  $X_1$ .

**Remark B.7** (Related works). In the unbounded Euclidean case  $M = \mathbb{R}^d$ , a quantitative divergence lemma has been proved by Cao, Lu and Wang [61, Lemma 2.6]. We obtain the order of the constant  $c_1(T)$  as conjectured in [61, Remark 2.7] by more closely considering the case of space-time harmonic right-hand sides and splitting these into high and low modes, as well as into symmetric and antisymmetric functions. Again on  $\mathbb{R}^d$ , the recent preprint [168] takes a variational approach to the divergence lemma, yet obtains an additional term of the order  $mT^2$  in the constant  $c_1(T)$  in (B.19) due to the estimate of  $\|\nabla h\|_{L^2(\bar{\mu})}^2$ . We avoid this, at the cost of the estimate of  $\|\partial_t \nabla g\|_{L^2(\bar{\mu})}^2$  being of the order  $1 + \frac{1}{mT^2}$  instead of  $\frac{1}{mT^2}$  as in [168]. Brigati and Stoltz [53, 55] prove a related averaging lemma, eliminating the need of discrete spectrum of  $L$ , at the cost of worse scaling in  $T$ . Finally, [97] uses a variation of our proof by formulating the divergence lemma in terms of the associated Dirichlet form, allowing for the treatment of lifts of diffusions with different boundary behaviour.

Let us first give a sketch of the proof of Theorem B.5. A first idea would be to solve the equation

$$(\partial_t^2 + L)u = f \quad (\text{B.23})$$

with Neumann boundary conditions and to set  $X = -\bar{\nabla} u$ , or, equivalently,  $h = \partial_t u$  and  $g = -u$ . However, by definition, Neumann boundary conditions only lead to  $h$  and  $\partial_t g$  vanishing on the time boundary and not  $g$  as required by (B.14). More precisely, let

$$\begin{aligned} \bar{\mathcal{D}} &= \{u \in C^\infty(\bar{M}) \cap L^2(\bar{\mu}) : \partial_N u = 0 \text{ on } [0, T] \times \partial M, \partial_t u = 0 \text{ on } \{0, T\} \times M\}, \\ \bar{L}u &= (\partial_t^2 + L)u = -\bar{\nabla}^* \bar{\nabla} u \quad \text{for } u \in \bar{\mathcal{D}}. \end{aligned}$$

Then, as in Corollary B.4, the operator  $(\bar{L}, \bar{\mathcal{D}})$  on  $L^2(\bar{\mu})$  is essentially self-adjoint and its unique self-adjoint extension  $(\bar{L}, \text{Dom}(\bar{L}))$  satisfies  $\text{Dom}(\bar{L}) \subseteq H^{2,2}(\bar{\mu})$ . In particular,  $\text{Dom}(\bar{L})$  contains functions  $u \in L^2(\bar{\mu})$  such that  $u(t, \cdot) \in \text{Dom}(L)$  for  $\lambda$ -a.e.  $t \in [0, T]$  and  $u(\cdot, x) \in H^{2,2}([0, T])$  for  $\bar{\mu}$ -a.e.  $x \in M$  satisfying the Neumann boundary conditions  $\partial_t u(0, \cdot) = \partial_t u(T, \cdot) = 0$  and  $(\partial_t^2 + L)u \in L^2(\bar{\mu})$ .

The operator  $(\bar{L}, \text{Dom}(\bar{L}))$  has discrete spectrum on  $L^2(\bar{\mu})$  and satisfies a Poincaré inequality with constant  $\bar{m} = \min(m, \frac{\pi^2}{T^2})$ . Therefore, the restriction

$$-\bar{L}|_{L_0^2(\bar{\mu})}: \text{Dom}(\bar{L}) \cap L_0^2(\bar{\mu}) \rightarrow L_0^2(\bar{\mu})$$

admits a bounded linear inverse

$$\bar{G}: L_0^2(\bar{\mu}) \rightarrow \text{Dom}(\bar{L}) \cap L_0^2(\bar{\mu})$$

whose operator norm is bounded from above by  $\bar{m}^{-1} = \max(\frac{1}{m}, \frac{T^2}{\pi^2})$ .

Hence  $u = \bar{G}f$  is a solution of (B.23) and bounds as in (B.17) and (B.18) can be derived for  $h = -\partial_t u$  and  $g = u$  in a straightforward way. However, in general  $u$  does not satisfy Dirichlet boundary conditions in time as required in (B.14). Indeed, an integration by parts shows that Dirichlet boundary conditions in time hold for  $u$  if and only if  $f$  is orthogonal to the space  $\mathbb{H}$  of space-time harmonic functions, that is  $\mathbb{H}$  consists of all functions  $u \in L^2(\bar{\mu})$  such that  $u(t, \cdot) \in \text{Dom}(L)$  for  $\lambda$ -a.e.  $t \in [0, T]$  and  $u(\cdot, x) \in H^{2,2}([0, T])$  for  $\bar{\mu}$ -a.e.  $x \in M$  and  $(\partial_t^2 + L)u = 0$ .

Therefore, in the proof of Theorem B.5, we treat right-hand sides in  $\mathbb{H}_0 = \mathbb{H} \cap L_0^2(\bar{\mu})$  and its orthogonal complement  $\mathbb{H}_0^\perp$  in  $L_0^2(\bar{\mu})$  separately. For functions  $f \in \mathbb{H}_0^\perp$ , we can set  $u = \bar{G}f \in H^{2,2}(\bar{\mu})$  and  $X = \bar{\nabla}u$ , or equivalently  $h = -\partial_t u$  and  $g = u$ , to obtain  $\bar{\nabla}^* X = \partial_t h - Lg = f$ . The orthogonality of  $f$  to the space of harmonic functions  $\mathbb{H}_0$  then yields the boundary conditions (B.14), and the bounds (B.17) and (B.18) follow by the Poincaré inequality and the Reilly formula. Indeed, by Remark B.2, we can apply Lemma B.1 on the Riemannian manifold  $\bar{M}$  with corners, so that we obtain the estimate

$$\|\bar{\nabla}^2 u\|_{L^2(\bar{\mu})}^2 \leq \|\bar{L}u\|_{L^2(\bar{\mu})}^2 + \rho \|\bar{\nabla}u\|_{L^2(\bar{\mu})}^2 \quad (\text{B.24})$$

for any function  $u \in \text{Dom}(\bar{L})$  in analogy to Corollary B.4.

In order to treat functions  $f \in \mathbb{H}_0$ , we decompose  $\mathbb{H}_0$  using an orthogonal basis of  $L^2(\mu)$  consisting of eigenfunctions of  $L$ . We then consider the antisymmetric and symmetric parts as well as the high and low modes separately. In the case of antisymmetric functions in the low modes,  $h$  and  $g$  can simply be chosen as the time-integral of  $f$  and zero. The tricky case is that of symmetric functions in the low modes: this is where the estimates degenerate for  $T \rightarrow 0$ . Finally, the high modes are not expected to effect the results. This can indeed be verified by a technical computation.

*Proof of Theorem B.5.* Let  $\mathbb{H}_0 = \mathbb{H} \cap L_0^2(\mu)$  as defined above and decompose

$$L_0^2(\bar{\mu}) = \mathbb{H}_0 \oplus \mathbb{H}_0^\perp.$$

At first consider  $f \in \mathbb{H}_0^\perp$ . Then setting  $h = -\partial_t u$  and  $g = u$  with  $u = \bar{G}f$  yields  $u \in \text{Dom}(\bar{L})$  and

$$f = \partial_t h - Lg.$$

Furthermore, for any  $v \in \mathbb{H}_0$ , an integration by parts yields

$$0 = (f, v)_{L^2(\bar{\mu})} = (\bar{L}u, v)_{L^2(\bar{\mu})}$$

$$\begin{aligned}
&= (u, (\partial_t^2 + L)v)_{L^2(\bar{\mu})} + \int_{(0,T)} \int_{\partial M} v \partial_N u - u \partial_N v \, d\mu_{\partial M} \, d\lambda \\
&\quad + \int_M (u(T, \cdot) \partial_t v(T, \cdot) - u(0, \cdot) \partial_t v(0, \cdot)) \, d\mu \\
&= \int_M (u(T, \cdot) \partial_t v(T, \cdot) - u(0, \cdot) \partial_t v(0, \cdot)) \, d\mu.
\end{aligned}$$

Hence  $u = 0$  on  $\{0, T\} \times M$  since  $v \in \mathbb{H}_0$  is arbitrary. Since by definition  $u$  also satisfies Neumann boundary conditions,  $h$  and  $g$  and the boundary conditions (B.14) and (B.15). We obtain the estimates

$$\|h\|_{L^2(\bar{\mu})}^2 + \|\nabla g\|_{L^2(\bar{\mu})}^2 = \|\bar{\nabla} u\|_{L^2(\bar{\mu})}^2 = (u, f)_{L^2(\bar{\mu})} \leq \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \|f\|_{L^2(\bar{\mu})}^2$$

and

$$\begin{aligned}
&\|\partial_t h\|_{L^2(\bar{\mu})}^2 + \|\partial_t \nabla g\|_{L^2(\bar{\mu})}^2 + \|\nabla h\|_{L^2(\bar{\mu})}^2 + \|\nabla^2 g\|_{L^2(\bar{\mu})}^2 \\
&= \|\bar{\nabla}^2 u\|_{L^2(\bar{\mu})}^2 \leq \|\bar{L}u\|_{L^2(\bar{\mu})}^2 + \rho \|\bar{\nabla} u\|_{L^2(\bar{\mu})}^2 \\
&\leq \left(1 + \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \rho\right) \|f\|_{L^2(\bar{\mu})}^2
\end{aligned}$$

by (B.24) and  $\|u\|_{L^2(\bar{\mu})} \leq \max(\frac{1}{m}, \frac{T^2}{\pi^2}) \|f\|_{L^2(\bar{\mu})}$ . Therefore, for functions  $f \in \mathbb{H}_0^\perp$ , the bounds (B.17) and (B.18) are satisfied with the constants

$$c_0^\perp = \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \quad \text{and} \quad c_1^\perp = 1 + \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \rho.$$

Now consider  $f \in \mathbb{H}_0$ . In terms of the orthonormal basis  $(e_k)_{k \in \mathbb{N}_0}$  of eigenfunctions of  $L$ , any function  $f \in \mathbb{H}_0$  can be represented as

$$f(t, x) = \sum_{k \in \mathbb{N}_0} f_k(t) e_k(x),$$

where  $f_k = (f, e_k)_{L^2(\mu)}$ . This yields

$$0 = (\partial_t^2 f + Lf)(t, x) = f_0''(t) + \sum_{k \in \mathbb{N}} (f_k''(t) - \alpha_k^2 f_k(t)) e_k(x),$$

so that the coefficients  $f_k$  satisfy the ordinary differential equations  $f_k'' = \alpha_k^2 f_k$ . Hence for any  $k \in \mathbb{N}$ , the function  $f_k$  can be expressed as a linear combination of  $e^{-\alpha_k t}$  and  $e^{-\alpha_k(T-t)}$ , and the functions

$$\begin{aligned}
H_0^a(t, x) &= (t - (T - t)) e_0(x), \\
H_k^a(t, x) &= (e^{-\alpha_k t} - e^{-\alpha_k(T-t)}) e_k(x) \quad \text{for } k \in \mathbb{N}, \\
H_k^s(t, x) &= (e^{-\alpha_k t} + e^{-\alpha_k(T-t)}) e_k(x) \quad \text{for } k \in \mathbb{N}
\end{aligned}$$

### B.3. The space-time divergence lemma

define an orthogonal basis  $\{H_k^a: k \in \mathbb{N}_0\} \cup \{H_k^s: k \in \mathbb{N}\}$  of  $\mathbb{H}_0$ . The functions  $H_k^a$  and  $H_k^s$  are in  $H^{2,2}(\bar{\mu})$  since  $e_k \in \text{Dom}(L) \subseteq H^{2,2}(\mu)$  for all  $k \in \mathbb{N}_0$ . We decompose  $\mathbb{H}_0$  into the symmetric and antisymmetric functions as well as into the high and low modes, i.e.

$$\mathbb{H}_0 = \mathbb{H}_{l,a} \oplus \mathbb{H}_{l,s} \oplus \mathbb{H}_{h,a} \oplus \mathbb{H}_{h,s},$$

where

$$\begin{aligned}\mathbb{H}_{l,a} &= \text{span}\{H_k^a: k \in \mathbb{N}_0 \text{ with } \alpha_k \leq \frac{\beta}{T}\}, \\ \mathbb{H}_{l,s} &= \text{span}\{H_k^s: k \in \mathbb{N} \text{ with } \alpha_k \leq \frac{\beta}{T}\}, \\ \mathbb{H}_{h,a} &= \overline{\text{span}}\{H_k^a: k \in \mathbb{N} \text{ with } \alpha_k > \frac{\beta}{T}\}, \\ \mathbb{H}_{h,s} &= \overline{\text{span}}\{H_k^s: k \in \mathbb{N} \text{ with } \alpha_k > \frac{\beta}{T}\},\end{aligned}$$

and consider these four subspaces separately. Here  $\beta \geq 2$  is the cutoff between high and low modes. For  $f \in \mathbb{H}_{l,a} \cup \mathbb{H}_{l,s}$  we have  $(f, -Lf)_{L^2(\bar{\mu})} \leq \frac{\beta^2}{T^2} \|f\|_{L^2(\bar{\mu})}^2$ . We will later choose  $\beta = 2$ .

*Case 1:* Let  $f \in \mathbb{H}_{l,a}$ . Set  $h(t, x) = \int_0^t f(s, x) ds$  and  $g(t, x) = 0$ . Then  $h \in H_D^{1,2}(\bar{\mu})$  due to the antisymmetry of  $f$ , so that (B.14) and (B.15) are satisfied. Since  $\partial_t h = f$ , the Poincaré inequality on  $[0, T]$  yields

$$\begin{aligned}\|h\|_{L^2(\bar{\mu})}^2 &\leq \frac{T^2}{\pi^2} \|\partial_t h\|_{L^2(\bar{\mu})}^2 = \frac{T^2}{\pi^2} \|f\|_{L^2(\mu)}^2, \\ \|\nabla h\|_{L^2(\bar{\mu})}^2 &\leq \frac{T^2}{\pi^2} \|\nabla f\|_{L^2(\bar{\mu})}^2 = \frac{T^2}{\pi^2} \int_0^T (f(t, \cdot), -Lf(t, \cdot))_{L^2(\mu)} dt \\ &\leq \frac{T^2}{\pi^2} \int_0^T \frac{\beta^2}{T^2} \|f(t, \cdot)\|_{L^2(\mu)}^2 dt = \frac{\beta^2}{\pi^2} \|f\|_{L^2(\bar{\mu})}^2.\end{aligned}$$

Hence for  $f \in \mathbb{H}_{l,a}$  we obtain (B.17) and (B.18) with the constants

$$c_0^{l,a} = \frac{T^2}{\pi^2} \quad \text{and} \quad c_1^{l,a} = 1 + \frac{\beta^2}{\pi^2}.$$

*Case 2:* Let  $f \in \mathbb{H}_{l,s}$ . Since  $f$  is symmetric, it does not necessarily integrate to 0 in time and we cannot argue as in the previous case. Instead, we consider the decomposition  $f = f_0 + f_1$ ,

$$\begin{aligned}f_0(t, x) &= f(0, x) \cdot \cos\left(\frac{2\pi t}{T}\right), \\ f_1(t, x) &= f(t, x) - f_0(t, x)\end{aligned}$$

Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

of  $f$  into a part  $f_0$  that integrates to 0 in time and a part  $f_1$  with Dirichlet boundary values in time. Note that  $f(t, \cdot), f_0(t, \cdot), f_1(t, \cdot)$  are contained in  $\text{span}\{e_k: k \in \mathbb{N} \text{ with } \alpha_k \leq \frac{\beta}{T}\}$  for all  $t \in [0, T]$ . We now set

$$h(t, x) = \int_0^t f_0(s, x) \, ds \quad \text{and} \quad g(t, x) = Gf_1(t, \cdot)(x),$$

so that  $f_0 = \partial_t h$  and  $f_1 = -Lg$ . The boundary conditions (B.14) are satisfied due to the boundary conditions of  $f_0$  and  $f_1$ , and (B.15) is satisfied by the definition (B.13) of  $G$ . In order to obtain the required estimates, we bound  $\|f_0\|_{L^2(\bar{\mu})}$  by  $\|f\|_{L^2(\bar{\mu})}$ . First note that

$$\|f_0\|_{L^2(\bar{\mu})}^2 = \|f(0, \cdot)\|_{L^2(\mu)}^2 \int_0^T \cos\left(\frac{2\pi t}{T}\right)^2 \, dt = \frac{T}{2} \|f(0, \cdot)\|_{L^2(\mu)}^2.$$

Expanding  $f(t, x) = \sum_{\alpha_k \leq \frac{\beta}{T}} b_k H_k^s(t, x)$  yields

$$\|f(0, \cdot)\|_{L^2(\mu)}^2 = \sum_{\alpha_k \leq \frac{\beta}{T}} b_k^2 (1 + e^{-\alpha_k T})^2 \leq 4 \sum_{\alpha_k \leq \frac{\beta}{T}} b_k^2$$

and

$$\begin{aligned} \|f\|_{L^2(\bar{\mu})}^2 &= \sum_{\alpha_k \leq \frac{\beta}{T}} b_k^2 \int_0^T \left( e^{-\alpha_k t} + e^{-\alpha_k(T-t)} \right)^2 \, dt \\ &\geq \sum_{\alpha_k \leq \frac{\beta}{T}} T \min_{t \in [0, T]} \left( e^{-\alpha_k t} + e^{-\alpha_k(T-t)} \right)^2 b_k^2 \\ &= \sum_{\alpha_k \leq \frac{\beta}{T}} 4T e^{-\alpha_k T} b_k^2 \geq 4T e^{-\beta} \sum_{\alpha_k \leq \frac{\beta}{T}} b_k^2, \end{aligned}$$

so that

$$\|f_0\|_{L^2(\bar{\mu})}^2 \leq \frac{T}{2} \|f(0, \cdot)\|_{L^2(\mu)}^2 \leq \frac{1}{2} e^\beta \|f\|_{L^2(\bar{\mu})}^2 \quad (\text{B.25})$$

and

$$\|f_1\|_{L^2(\bar{\mu})}^2 \leq (\|f\|_{L^2(\bar{\mu})} + \|f_0\|_{L^2(\bar{\mu})})^2 \leq \left(1 + \frac{1}{\sqrt{2}} e^{\beta/2}\right)^2 \|f\|_{L^2(\bar{\mu})}^2 \leq (2 + e^\beta) \|f\|_{L^2(\bar{\mu})}^2.$$

Similarly to the previous case, this gives the estimates

$$\begin{aligned} \|h\|_{L^2(\bar{\mu})}^2 &\leq \frac{T^2}{\pi^2} \|f_0\|_{L^2(\bar{\mu})}^2 \leq \frac{T^2}{2\pi^2} e^\beta \|f\|_{L^2(\bar{\mu})}^2, \\ \|\partial_t h\|_{L^2(\bar{\mu})}^2 &= \|f_0\|_{L^2(\bar{\mu})}^2 \leq \frac{1}{2} e^\beta \|f\|_{L^2(\bar{\mu})}^2, \\ \|\nabla h\|_{L^2(\bar{\mu})}^2 &\leq \frac{\beta^2}{\pi^2} \|f_0\|_{L^2(\bar{\mu})}^2 \leq \frac{\beta^2}{2\pi^2} e^\beta \|f\|_{L^2(\bar{\mu})}^2. \end{aligned}$$

For the bounds on  $g$  we use the fact that

$$\|\nabla Gu\|_{L^2(\mu)}^2 = (u, Gu)_{L^2(\mu)} \leq \frac{1}{m} \|u\|_{L^2(\mu)}^2$$

for any  $u \in L_0^2(\mu)$ . Then

$$\|\nabla g\|_{L^2(\bar{\mu})}^2 = \int_0^T \|\nabla G f_1(t, \cdot)\|_{L^2(\mu)}^2 dt \leq \frac{1}{m} \|f_1\|_{L^2(\bar{\mu})}^2 \leq \frac{1}{m} \left(1 + \frac{1}{\sqrt{2}} e^{\beta/2}\right)^2 \|f\|_{L^2(\bar{\mu})}^2.$$

Similarly, commuting the derivatives,

$$\|\partial_t \nabla g\|_{L^2(\bar{\mu})}^2 = \|\nabla \partial_t g\|_{L^2(\bar{\mu})}^2 \leq \frac{1}{m} \|\partial_t f_1\|_{L^2(\bar{\mu})}^2 \leq \frac{1}{m} (\|\partial_t f\|_{L^2(\bar{\mu})} + \|\partial_t f_0\|_{L^2(\bar{\mu})})^2. \quad (\text{B.26})$$

Again using the expansion  $f(t, x) = \sum_{\alpha_k \leq \frac{\beta}{T}} b_k h_k^s(t, x)$  we obtain

$$\begin{aligned} \|\partial_t f\|_{L^2(\bar{\mu})}^2 &= \sum_{\alpha_k \leq \frac{\beta}{T}} \alpha_k^2 b_k^2 \int_0^T \left(-e^{-\alpha_k t} + e^{-\alpha_k(T-t)}\right)^2 dt \\ &\leq \frac{\beta^2}{T^2} \sum_{\alpha_k \leq \frac{\beta}{T}} b_k^2 \int_0^T \left(e^{-\alpha_k t} + e^{-\alpha_k(T-t)}\right)^2 dt = \frac{\beta^2}{T^2} \|f\|_{L^2(\bar{\mu})}^2 \end{aligned}$$

as well as

$$\|\partial_t f_0\|_{L^2(\bar{\mu})}^2 = \frac{2\pi^2}{T} \|f(0, \cdot)\|_{L^2(\mu)}^2 \leq \frac{2\pi^2}{T^2} e^\beta \|f\|_{L^2(\bar{\mu})}^2$$

by (B.25). Hence (B.26) yields

$$\|\partial_t \nabla g\|_{L^2(\bar{\mu})}^2 \leq \frac{(\beta + \sqrt{2}\pi e^{\beta/2})^2}{mT^2} \|f\|_{L^2(\bar{\mu})}^2.$$

Finally, Corollary B.4 yields

$$\begin{aligned} \|\nabla^2 g\|_{L^2(\bar{\mu})}^2 &\leq \|Lg\|_{L^2(\bar{\mu})}^2 + \rho \|\nabla g\|_{L^2(\bar{\mu})}^2 \leq \left(1 + \frac{\rho}{m}\right) \|f_1\|_{L^2(\bar{\mu})}^2 \\ &\leq \left(1 + \frac{\rho}{m}\right) \left(1 + \frac{1}{\sqrt{2}} e^{\beta/2}\right)^2 \|f\|_{L^2(\bar{\mu})}^2. \end{aligned}$$

For  $f \in \mathbb{H}_{l,s}$  we therefore obtain (B.17) and (B.18) with the constants

$$\begin{aligned} c_0^{l,s} &= \frac{T^2}{2\pi^2} e^\beta + \frac{1}{m} \left(1 + \frac{1}{\sqrt{2}} e^{\beta/2}\right)^2 \quad \text{and} \\ c_1^{l,s} &= \left(\frac{1}{2} + \frac{\beta^2}{2\pi^2}\right) e^\beta + \frac{(\beta + \sqrt{2}\pi e^{\beta/2})^2}{mT^2} + \left(1 + \frac{\rho}{m}\right) \left(1 + \frac{1}{\sqrt{2}} e^{\beta/2}\right)^2. \end{aligned}$$

## Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

*Case 3:* Let  $f \in \mathbb{H}_{h,a}$ . We will use the expansion  $f(t, x) = \sum_{\alpha_k > \frac{\beta}{T}} b_k H_k^a(t, x)$ . As we will show below, it is sufficient to derive a representation

$$H_k^a = \partial_t h_k - L g_k$$

for each of the basis functions  $H_k^a$  with  $k \in \mathbb{N}$  fixed. To this end, we write  $u_k(t) = e^{-\alpha_k t} - e^{-\alpha_k(T-t)}$ , so that  $H_k^a(t, x) = u_k(t)e_k(x)$ . We employ the ansatz

$$u_k = \dot{v}_k + w_k \quad \text{with } v_k(0) = v_k(T) = w_k(0) = w_k(T) = 0. \quad (\text{B.27})$$

Then setting

$$h_k(t, x) = v_k(t)e_k(x) \quad \text{and} \quad g_k(t, x) = \frac{1}{\alpha_k^2} w_k(t)e_k(x)$$

yields  $H_k^a = \partial_t h_k - L g_k$ .

We now construct such functions  $v_k$  and  $w_k$  for a fixed  $k \in \mathbb{N}$ . Let

$$\varphi_k(t) = (\alpha_k t - 1) 1_{[0, \alpha_k^{-1}]}(t).$$

This function is in  $C^1([0, T])$  and piecewise  $C^2([0, T])$ . Moreover, since  $\alpha_k > \frac{2}{T}$ , it satisfies  $\varphi_k(\frac{T}{2}) = 0$ . For  $t \in [0, \frac{T}{2}]$  we set

$$\begin{aligned} v_k(t) &= \varphi_k(t) \int_0^t u_k(s) \, ds \quad \text{and} \\ w_k(t) &= u_k(t) - \dot{v}_k(t) = (1 - \varphi_k(t))u_k(t) - \dot{\varphi}_k(t) \int_0^t u_k(s) \, ds. \end{aligned}$$

Analogously, for  $t \in [\frac{T}{2}, T]$  we set

$$\begin{aligned} v_k(t) &= -\varphi_k(T-t) \int_t^T u_k(s) \, ds \quad \text{and} \\ w_k(t) &= u_k(t) - \dot{v}_k(t) = (1 - \varphi_k(T-t))u_k(t) - \dot{\varphi}_k(T-t) \int_t^T u_k(s) \, ds. \end{aligned}$$

Then  $v_k, w_k$  are in  $C^0([0, T])$  and piecewise  $C^1([0, T])$  with  $v_k(0) = v_k(T) = 0$ ,  $v_k(\frac{T}{2}) = \dot{v}_k(\frac{T}{2}) = 0$ , and

$$\begin{aligned} \dot{v}_k(0) &= \varphi_k(0)u_k(0) = u_k(0), & \dot{v}_k(T) &= \varphi_k(0)u_k(T) = u_k(T), \\ w_k(0) &= u_k(0) - \dot{v}_k(0) = 0, & w_k(T) &= u_k(T) - \dot{v}_k(T) = 0, \end{aligned}$$

so that (B.27) is satisfied. To obtain the required bounds, notice that  $\|H_k^a\|_{L^2(\bar{\mu})}^2 = \int_0^T u_k(t)^2 \, dt$  and

$$\begin{aligned} \|h_k\|_{L^2(\bar{\mu})}^2 &= \int_0^T v_k(t)^2 \, dt, & \|\nabla g_k\|_{L^2(\bar{\mu})}^2 &= \frac{1}{\alpha_k^2} \int_0^T w_k(t)^2 \, dt, \\ \|\partial_t h_k\|_{L^2(\bar{\mu})}^2 &= \int_0^T \dot{v}_k(t)^2 \, dt, & \|\partial_t \nabla g_k\|_{L^2(\bar{\mu})}^2 &= \frac{1}{\alpha_k^2} \int_0^T \dot{w}_k(t)^2 \, dt, \\ \|\nabla h_k\|_{L^2(\bar{\mu})}^2 &= \alpha_k^2 \int_0^T v_k(t)^2 \, dt, & \|L g_k\|_{L^2(\bar{\mu})}^2 &= \int_0^T w_k(t)^2 \, dt, \end{aligned}$$

where we used that  $\|\nabla g_k\|_{L^2(\bar{\mu})}^2 = (g_k, -Lg_k)_{L^2(\bar{\mu})}$  and  $\|\partial_t \nabla g_k\|_{L^2(\bar{\mu})}^2 = (\partial_t g_k, -L\partial_t g_k)_{L^2(\bar{\mu})}$ . One estimates

$$\begin{aligned} \int_0^{\frac{T}{2}} v_k(t)^2 dt &= \int_0^{\frac{1}{\alpha_k}} \varphi_k(t)^2 \left( \int_0^t u_k(s) ds \right)^2 dt \leq \int_0^{\frac{1}{\alpha_k}} (\alpha_k t - 1)^4 t dt \int_0^{\frac{T}{2}} u_k(t)^2 dt \\ &= \frac{1}{30\alpha_k^2} \int_0^{\frac{T}{2}} u_k(t)^2 dt, \end{aligned}$$

and analogously on  $[\frac{T}{2}, T]$ , so that

$$\int_0^T v_k(t)^2 dt \leq \frac{1}{30\alpha_k^2} \int_0^T u_k(t)^2 dt.$$

For  $\dot{v}_k = \dot{\varphi}_k \int_0^\cdot u_k(s) ds + \varphi_k u_k$  we similarly estimate

$$\begin{aligned} \int_0^{\frac{T}{2}} \dot{v}_k(t)^2 dt &\leq \left( \left( \int_0^{\frac{1}{\alpha_k}} \dot{\varphi}_k(t)^2 \left( \int_0^t u_k(s) ds \right)^2 dt \right)^{\frac{1}{2}} + \left( \int_0^{\frac{T}{2}} u_k(t)^2 dt \right)^{\frac{1}{2}} \right)^2 \\ &\leq \left( 1 + \left( \int_0^{\frac{1}{\alpha_k}} (2\alpha_k(\alpha_k t - 1))^2 t dt \right)^{\frac{1}{2}} \right)^2 \int_0^{\frac{T}{2}} u_k(t)^2 dt \\ &= \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \int_0^{\frac{T}{2}} u_k(t)^2 dt < \frac{5}{2} \int_0^{\frac{T}{2}} u_k(t)^2 dt, \end{aligned}$$

so that

$$\int_0^T \dot{v}_k(t)^2 dt \leq \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \int_0^T u_k(t)^2 dt.$$

Similarly, since  $w_k = -\dot{\varphi}_k \int_0^\cdot u_k(s) ds + (1 - \varphi_k)u_k$  on  $[0, \frac{T}{2}]$  and correspondingly on  $[\frac{T}{2}, T]$ , one obtains

$$\int_0^T w_k(t)^2 dt \leq \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \int_0^T u_k(t)^2 dt.$$

For  $\dot{w}_k = (1 - \varphi_k)\dot{u}_k - 2\dot{\varphi}_k u_k - \ddot{\varphi}_k \int_0^\cdot u_k(s) ds$  we first estimate  $\int_0^T \dot{u}_k(t)^2 dt$  by  $\int_0^T u_k(t)^2 dt$ . A direct computation yields

$$\begin{aligned} \int_0^{\frac{T}{2}} \dot{u}_k(t)^2 dt &= \alpha_k^2 \int_0^{\frac{T}{2}} \left( e^{-\alpha_k t} + e^{-\alpha_k(T-t)} \right)^2 dt \\ &= \alpha_k e^{-\alpha_k T} (\sinh(\alpha_k T) + \alpha_k T) \end{aligned}$$

and similarly

$$\int_0^{\frac{T}{2}} u_k(t)^2 dt = \alpha_k^{-1} e^{-\alpha_k T} (\sinh(\alpha_k T) - \alpha_k T),$$

so that

$$\int_0^{\frac{T}{2}} \dot{u}_k(t)^2 dt \leq \alpha_k^2 \frac{\sinh(\beta) + \beta}{\sinh(\beta) - \beta} \int_0^{\frac{T}{2}} u_k(t)^2 dt$$

Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

by monotonicity of  $x \mapsto \frac{\sinh(x)+x}{\sinh(x)-x}$  on  $[0, \infty)$ . Therefore, by similar arguments as before,

$$\begin{aligned} \int_0^{\frac{T}{2}} \dot{w}_k(t)^2 dt &\leq \left( 4\alpha_k + \alpha_k \left( \frac{\sinh(\beta) + \beta}{\sinh(\beta) - \beta} \right)^{\frac{1}{2}} + \left( \int_0^{\frac{1}{\alpha_k}} 4\alpha_k^4 t dt \right)^{\frac{1}{2}} \right)^2 \int_0^{\frac{T}{2}} u_k(t)^2 dt \\ &= \alpha_k^2 \left( 4 + \sqrt{2} + \left( \frac{\sinh(\beta) + \beta}{\sinh(\beta) - \beta} \right)^{\frac{1}{2}} \right)^2 \int_0^{\frac{T}{2}} u_k(t)^2 dt, \end{aligned}$$

and analogously on  $[\frac{T}{2}, T]$ , yielding

$$\int_0^T \dot{w}_k(t)^2 dt \leq \alpha_k^2 \left( 4 + \sqrt{2} + \left( \frac{\sinh(\beta) + \beta}{\sinh(\beta) - \beta} \right)^{\frac{1}{2}} \right)^2 \int_0^T u_k(t)^2 dt.$$

Noting that  $\alpha_k^2 \geq \max(m, \frac{\beta^2}{T^2})$ , we hence obtain the bounds

$$\begin{aligned} \|h_k\|_{L^2(\bar{\mu})}^2 &\leq \frac{1}{30} \min\left(\frac{1}{m}, \frac{T^2}{\beta^2}\right) \|H_k^a\|_{L^2(\bar{\mu})}^2, \\ \|\partial_t h_k\|_{L^2(\bar{\mu})}^2 &\leq \left(1 + \frac{1}{\sqrt{3}}\right)^2 \|H_k^a\|_{L^2(\bar{\mu})}^2, \\ \|\nabla h_k\|_{L^2(\bar{\mu})}^2 &\leq \frac{1}{30} \|H_k^a\|_{L^2(\bar{\mu})}^2, \\ \|\nabla g_k\|_{L^2(\bar{\mu})}^2 &\leq \min\left(\frac{1}{m}, \frac{T^2}{\beta^2}\right) \left(1 + \frac{1}{\sqrt{3}}\right)^2 \|H_k^a\|_{L^2(\bar{\mu})}^2, \\ \|\partial_t \nabla g_k\|_{L^2(\bar{\mu})}^2 &\leq \left(4 + \sqrt{2} + \left( \frac{\sinh(\beta) + \beta}{\sinh(\beta) - \beta} \right)^{\frac{1}{2}}\right)^2 \|H_k^a\|_{L^2(\bar{\mu})}^2, \\ \|Lg_k\|_{L^2(\bar{\mu})}^2 &\leq \left(1 + \frac{1}{\sqrt{3}}\right)^2 \|H_k^a\|_{L^2(\bar{\mu})}^2. \end{aligned}$$

and thus also

$$\|\nabla^2 g_k\|_{L^2(\bar{\mu})}^2 \leq \left(1 + \frac{1}{\sqrt{3}}\right)^2 \left(1 + \min\left(\frac{1}{m}, \frac{T^2}{\beta^2}\right) \rho\right) \|H_k^a\|_{L^2(\bar{\mu})}^2.$$

For a general function  $f = \sum_{\alpha_k > \frac{\beta}{T}} b_k H_k^a$  in  $\mathbb{H}_{h,a}$ , we set

$$h = \sum_{\alpha_k > \beta/T} b_k h_k \quad \text{and} \quad g = \sum_{\alpha_k > \beta/T} b_k g_k$$

to obtain  $f = \partial_t h - Lg$ . By orthogonality of the functions  $e_k$ , the functions  $h_k$  and  $\partial_t h_k$  are also orthogonal systems. Hence

$$\|h\|_{L^2(\bar{\mu})}^2 = \sum_{\alpha_k > \beta/T} \|h_k\|^2, \quad \|\partial_t h\|_{L^2(\bar{\mu})}^2 = \sum_{\alpha_k > \beta/T} \|\partial_t h_k\|^2$$

and

$$\begin{aligned}\|\nabla h\|_{L^2(\bar{\mu})}^2 &= -(h, Lh)_{L^2(\bar{\mu})} = \sum_{\alpha_k > \beta/T} \alpha_k^2 \|h_k\|^2 \\ &= \sum_{\alpha_k > \beta/T} (h_k, Lh_k)_{L^2(\bar{\mu})} = \sum_{\alpha_k > \beta/T} \|\nabla h_k\|_{L^2(\bar{\mu})}^2.\end{aligned}$$

The same is true for the functions  $\nabla g_k$ ,  $\partial_t \nabla g_k$  and  $Lg_k$ , so that analogous identities hold. Since  $\|f\|_{L^2(\bar{\mu})}^2 = \sum_{\alpha_k \geq \frac{\beta}{T}} b_k^2 \|H_k^a\|_{L^2(\bar{\mu})}^2$ , this yields (B.17) and (B.18) with constants

$$\begin{aligned}c_0^{h,a} &= \left( \frac{1}{30} + \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \right) \min \left( \frac{1}{m}, \frac{T^2}{\beta^2} \right) \quad \text{and} \\ c_1^{h,a} &= \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \left( 2 + \min \left( \frac{1}{m}, \frac{T^2}{\beta^2} \right) \rho \right) + \left( 4 + \sqrt{2} + \left( \frac{\sinh(\beta) + \beta}{\sinh(\beta) - \beta} \right)^{\frac{1}{2}} \right)^2 + \frac{1}{30}.\end{aligned}$$

*Case 4:* Let  $f \in \mathbb{H}_{l,s}$ . Similarly to the previous case, we now set  $u_k(t) = e^{-\alpha_k t} + e^{-\alpha_k(T-t)}$ , so that  $H_k^s(t, x) = u_k(t)e_k(x)$ . We again derive a representation

$$H_k^s = \partial_t h_k - Lg_k$$

for each of the basis functions  $H_k^s$  with  $k \in \mathbb{N}$  fixed by employing the ansatz

$$u_k = \dot{v}_k + w_k \quad \text{with } v_k(0) = v_k(T) = w_k(0) = w_k(T) = 0. \quad (\text{B.28})$$

Then setting

$$h_k(t, x) = v_k(t)e_k(x) \quad \text{and} \quad g_k(t, x) = \frac{1}{\alpha_k^2} w_k(t)e_k(x)$$

yields  $H_k^s = \partial_t h_k - Lg_k$ . The only thing that changes compared to the previous case is the bound on  $\int_0^T \dot{u}_k(t)^2 dt$ . In this case, one has

$$\begin{aligned}\int_0^{\frac{T}{2}} \dot{u}_k(t)^2 dt &= \alpha_k^2 \int_0^T \left( -e^{-\alpha_k t} + e^{-\alpha_k(T-t)} \right)^2 dt \\ &= \alpha_k e^{-\alpha_k T} (\sinh(\alpha_k T) - \alpha_k T),\end{aligned}$$

whereas

$$\int_0^{\frac{T}{2}} u_k(t)^2 dt = \alpha_k^{-1} e^{-\alpha_k T} (\sinh(\alpha_k T) + \alpha_k T).$$

Hence

$$\int_0^T \dot{u}_k(t)^2 dt = \alpha_k^2 \frac{\sinh(\alpha_k T) - \alpha_k T}{\sinh(\alpha_k T) + \alpha_k T} \int_0^T u_k(t)^2 dt \leq \alpha_k^2 \int_0^T u_k(t)^2 dt.$$

## Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

Again setting

$$h = \sum_{\alpha_k > \beta/T} b_k h_k \quad \text{and} \quad g = \sum_{\alpha_k > \beta/T} b_k g_k$$

yields  $f = \partial_t h - Lg$ . As in the previous case, the bounds (B.17) and (B.18) follow with the constants

$$\begin{aligned} c_0^{h,s} &= \left( \frac{1}{30} + \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \right) \min \left( \frac{1}{m}, \frac{T^2}{\beta^2} \right) \quad \text{and} \\ c_1^{h,s} &= \left( 1 + \frac{1}{\sqrt{3}} \right)^2 \left( 2 + \min \left( \frac{1}{m}, \frac{T^2}{\beta^2} \right) \rho \right) + \left( 5 + \sqrt{2} \right)^2 + \frac{1}{30}. \end{aligned}$$

For any  $f \in L_0^2(\bar{\mu})$ , combining the results on each of the subspaces using linearity, we can hence write  $f = \partial_t h - Lg$  with  $h$  and  $g$  satisfying (B.14) and (B.15). Choosing  $\beta = 2$ , the bounds (B.17) and (B.18) then follow with

$$\begin{aligned} c_0(T) &= 5 \max \left( c_0^{l,a}, c_0^{l,s}, c_0^{h,a}, c_0^{h,s}, c_0^\perp \right) = 5c_0^{l,s} \\ &\leq 2T^2 + 43 \frac{1}{m} \end{aligned}$$

and

$$\begin{aligned} c_1(T) &= 5 \max \left( c_1^{l,a}, c_1^{l,s}, c_1^{h,a}, c_1^{h,s}, c_1^\perp \right) = 5 \max \left( c_1^{l,s}, c_1^{h,a}, c_1^\perp \right) \\ &\leq 290 + \frac{991}{mT^2} + 43 \max \left( \frac{1}{m}, \frac{T^2}{\pi^2} \right) \rho. \end{aligned}$$

□

### B.4. Optimal lifts on Riemannian manifolds with boundary

In the following, let  $(M, g)$  be a complete connected and oriented Riemannian manifold with boundary and let

$$\mu(dx) = \exp(-U(x)) \nu_M(dx)$$

with  $U \in C^2(M)$  be a probability measure on  $M$  such that Assumptions B.1 and B.2 are satisfied.

The operator  $(\frac{1}{2}L, \text{Dom}(L))$  defined by Corollary B.4 is the generator in  $L^2(\mu)$  of an overdamped Langevin diffusion on  $M$  with reflection at the boundary  $\partial M$  and invariant probability measure  $\mu$ . Explicitly, the set of compactly supported smooth functions

$$\mathcal{D} = \{u \in C_c^\infty(M) : \partial_N u = 0 \text{ on } \partial M\}$$

with Neumann boundary conditions is a core for the self-adjoint operator  $(\frac{1}{2}L, \text{Dom}(L))$  and

$$Lf = -\langle \nabla U, \nabla f \rangle + \Delta f \quad \text{for } f \in \mathcal{D}.$$

#### B.4. Optimal lifts on Riemannian manifolds with boundary

The corresponding diffusion process solves the Stratonovich stochastic differential equation

$$dZ_t = -\frac{1}{2}\nabla U(Z_t) dt + \circ dB_t - N(Z_t) dL_t,$$

where  $(B_t)_{t \geq 0}$  is a Brownian motion on  $M$  and  $(L_t)_{t \geq 0}$  is the local time of  $(Z_t)_{t \geq 0}$  at the boundary  $\partial M$ , see [144, 250]. By Corollary B.4, the associated Dirichlet form is

$$\mathcal{E}(f, g) = \frac{1}{2} \int_M \langle \nabla f, \nabla g \rangle d\mu$$

with domain  $\text{Dom}(\mathcal{E}) = H^{1,2}(M, \mu)$ .

##### B.4.1. Lifts based on Hamiltonian dynamics on Riemannian manifolds

At first assume that  $\partial M = \emptyset$ . We denote elements of the tangent bundle  $TM$  by  $(x, v)$  with  $x \in M$  and  $v \in T_x M$ . Defining the Hamiltonian

$$H(x, v) = U(x) + \frac{1}{2}\langle v, v \rangle_x$$

on  $TM$  yields the associated Hamiltonian flow  $(X_t, V_t)_{t \geq 0}$  on  $TM$ , see [1, Section 3.7]. It preserves the Hamiltonian  $H$  and satisfies the equations

$$\frac{d}{dt} X_t = V_t, \quad \frac{\nabla}{dt} V_t = -\nabla U(X_t).$$

Using horizontal and vertical lifts of vector fields, the tangent space  $T_{(x,v)} TM$  of  $TM$  in  $(x, v) \in TM$  can be canonically identified with the direct sum

$$T_{(x,v)} TM = T_x M \oplus T_x M, \tag{B.29}$$

see [226, Section II.4]. Using this identification, the vector field generating the Hamiltonian flow is  $(x, v) \mapsto (v, -\nabla U(x))$ . Furthermore, the Sasaki metric

$$\tilde{g}_{(x,v)}((\zeta, \xi), (\zeta', \xi')) = g_x(\zeta, \zeta') + g_x(\xi, \xi') \quad \text{for } (x, v) \in TM, (\zeta, \xi), (\zeta', \xi') \in T_{(x,v)} TM$$

on  $TM$  turns the tangent bundle  $(TM, \tilde{g})$  into a Riemannian manifold [226, 228]. By Liouville's theorem, the Hamiltonian flow leaves the associated Riemannian volume form  $d\nu_{\tilde{g}}$  on  $TM$  invariant. Therefore, the Hamiltonian flow preserves the probability measure  $\hat{\mu}$  on  $TM$  with density proportional to  $\exp(-H)$  with respect to the volume  $d\nu_{\tilde{g}}$ , which can be disintegrated as

$$\hat{\mu}(dx dv) = \mu(dx) \kappa(x, dv)$$

with  $\kappa(x, \cdot) = \mathcal{N}(0, g_x^{-1})$ . The generator of the associated transition semigroup on  $L^2(\hat{\mu})$  is then given by

$$\hat{L}f(x, v) = \langle v, \nabla_x f(x, v) \rangle - \langle \nabla U(x), \nabla_v f(x, v) \rangle, \tag{B.30}$$

and its domain satisfies

$$\text{Dom}(\hat{L}) \supseteq \{f \in C^\infty(TM) \cap L^2(\hat{\mu}) : \langle v, \nabla_x f \rangle - \langle \nabla U, \nabla_v f \rangle \in L^2(\hat{\mu})\}.$$

**Lemma B.8.** *The Hamiltonian dynamics  $(X_t, V_t)_{t \geq 0}$  is a second-order lift of the overdamped Langevin diffusion  $(Z_t)_{t \geq 0}$ .*

*Proof.* Note that it suffices to verify (B.1)–(B.3) on the core  $\mathcal{D}$  for  $(\frac{1}{2}L, \text{Dom}(L))$  defined in (B.11). Hence let  $f, g \in \mathcal{D}$ . Then  $\langle v, \nabla_x(f \circ \pi) \rangle = \langle v, \nabla f \circ \pi \rangle \in L^2(\hat{\mu})$ , so that  $f \circ \pi \in \text{Dom}(\hat{L})$  and (B.1) is satisfied. Furthermore,

$$\int_{\text{TM}} \hat{L}(f \circ \pi)(g \circ \pi) d\hat{\mu} = \int_M \int_{\text{T}_x M} \langle v, \nabla f(x) \rangle \kappa(x, dv) g(x) \mu(dx) = 0,$$

so that (B.2) is satisfied. Finally, (B.3) is satisfied by

$$\begin{aligned} \frac{1}{2} \int_{\text{TM}} (\hat{L}(f \circ \pi))^2 d\hat{\mu} &= \frac{1}{2} \int_M \int_{\text{T}_x M} \langle v, \nabla f(x) \rangle^2 \kappa(x, dv) \mu(dx) \\ &= \frac{1}{2} \int_M \langle \nabla f, \nabla f \rangle d\mu = \mathcal{E}(f, f). \end{aligned} \quad \square$$

Now let us consider a complete Riemannian manifold  $(M, g)$  with boundary  $\partial M \neq \emptyset$ . For  $x \in \partial M$  and  $v \in \text{T}_x M$  let

$$R_x v = v - 2N(x) \langle N(x), v \rangle_x$$

be the specular reflection at the boundary. Note that, if  $v$  is outward-pointing, then  $R_x v$  is inward-pointing, and  $R_x v$  preserves  $\kappa(x, dv)$ . The generator of Hamiltonian dynamics with reflection at the boundary is then given by the closure  $(\hat{L}, \text{Dom}(\hat{L}))$  of  $(\hat{L}, \mathcal{R})$  in  $L^2(\hat{\mu})$  with

$$\begin{aligned} \mathcal{R} &= \{f \in C^\infty(\text{TM}) \cap L^2(\hat{\mu}) : \langle v, \nabla_x f \rangle - \langle \nabla U, \nabla_v f \rangle \in L^2(\hat{\mu}) \text{ and} \\ &\quad f(x, R_x v) = f(x, v) \text{ for all } x \in \partial M \text{ and } v \in \text{T}_x M\}, \end{aligned}$$

and  $\hat{L}$  given by (B.30). This definition clearly coincides with the definition above in case  $\partial M = \emptyset$ . The associated process  $(X_t, V_t)_{t \geq 0}$  follows the Hamiltonian flow on the interior of  $M$ , and the normal component of the velocity is reflected when hitting the boundary.

**Lemma B.9.** *The Hamiltonian dynamics  $(X_t, V_t)_{t \geq 0}$  with reflection at the boundary is a second-order lift of the overdamped Langevin diffusion  $(Z_t)_{t \geq 0}$  with reflection at the boundary.*

*Proof.* Since  $f \in \mathcal{D}$  implies  $f \circ \pi \in \mathcal{R}$ , the domain condition (B.1) is satisfied. The remaining conditions (B.2) and (B.3) follow as in Lemma B.8.  $\square$

**Lemma B.10** (Adjoint of  $\hat{L}$ ). *The adjoint of  $(\hat{L}, \text{Dom}(\hat{L}))$  in  $L^2(\hat{\mu})$  satisfies  $\text{Dom}(\hat{L}) \subseteq \text{Dom}(\hat{L}^*)$  and  $\hat{L}^* g = -\hat{L} g$  for all  $g \in \text{Dom}(\hat{L})$ .*

*Proof.* Let  $f, g \in \mathcal{R}$ . Integration by parts gives

$$\int_{\text{TM}} \hat{L} f g d\hat{\mu} = - \int_{\text{TM}} f \hat{L} g d\hat{\mu} + \int_{\partial M} \int_{\text{T}_x M} f g \langle v, N(x) \rangle \kappa(x, dv) \mu(dx).$$

#### B.4. Optimal lifts on Riemannian manifolds with boundary

Denoting  $h(x, v) = f(x, v)\langle v, N(x) \rangle$  gives  $h(x, R_x v) = -h(x, v)$ , so that

$$\begin{aligned} \int_{T_x M} h(x, v) g(x, v) \kappa(x, dv) &= \int_{T_x M} h(x, R_x v) g(x, R_x v) \kappa(x, dv) \\ &= - \int_{T_x M} h(x, v) g(x, v) \kappa(x, dv) \end{aligned}$$

since  $R_x$  leaves  $\kappa(x, \cdot)$  invariant, so that the boundary terms vanish. Hence the operator  $(-\hat{L}^*, \text{Dom}(\hat{L}^*))$  extends  $(\hat{L}, \mathcal{R})$  and thus also  $(\hat{L}, \text{Dom}(\hat{L}))$ .  $\square$

*Randomised Riemannian Hamiltonian Monte Carlo* with refresh rate  $\gamma > 0$  is obtained by interspersing Hamiltonian dynamics with complete velocity refreshments according to  $\kappa(x, dv)$  after random times with exponential distribution  $\text{Exp}(\gamma)$ . By successively conditioning on the refreshment times, one can show that randomised Riemannian Hamiltonian Monte Carlo again leaves  $\hat{\mu}$  invariant [47], and its generator in  $L^2(\hat{\mu})$  is given by

$$\hat{L}_{\text{RPMC}}^{(\gamma)} = \hat{L} + \gamma(\Pi - I) \quad \text{and} \quad \text{Dom}(\hat{L}_{\text{RPMC}}^{(\gamma)}) = \text{Dom}(\hat{L}),$$

where

$$(\Pi f)(x, v) = \int_{\mathbb{R}^d} f(x, w) \kappa(x, dw).$$

*Riemannian Langevin dynamics* is obtained by adding Ornstein-Uhlenbeck noise instead of discrete jumps in the velocity, i.e. it is the solution  $(X_t, V_t)_{t \geq 0}$  to the stochastic differential equation

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= -\nabla U(X_t) dt - \gamma V_t dt + \sqrt{2\gamma} \circ dB_t + dK_t, \end{aligned}$$

where

$$K_t = -2 \sum_{0 < s \leq t} \langle N(X_s), V_{s-} \rangle N(X_s) \cdot 1_{\partial M}(X_s)$$

is a confinement process modelling specular reflections the boundary. Well-posedness of this equation and existence of solutions, in particular non-accumulation of boundary hits, is considered in [43, 44] in the Euclidean case, and the associated kinetic Fokker-Planck equation with specular boundary conditions is considered in [88]. The invariant probability measure of  $(X_t, V_t)_{t \geq 0}$  is  $\hat{\mu}$ , see [119]. We denote the associated generator in  $L^2(\hat{\mu})$  by  $(\hat{L}_{\text{LD}}^{(\gamma)}, \text{Dom}(\hat{L}_{\text{LD}}^{(\gamma)}))$ . Let

$$L_v f = -\langle v, \nabla_v f \rangle + \Delta_v f, \tag{B.31}$$

where  $\Delta = \Delta_x + \Delta_v$  is the splitting of the Laplace-Beltrami operator  $\Delta$  on  $TM$  associated to the decomposition (B.29), i.e.  $\Delta_v = \text{div}_v \nabla_v$  is the Laplace-Beltrami operator in  $v$ . Then, for functions  $f \in \mathcal{R}$  with  $L_v f \in L^2(\hat{\mu})$ , the generator is given by

$$\hat{L}_{\text{LD}}^{(\gamma)} f = \hat{L} f + \gamma L_v f. \tag{B.32}$$

**Lemma B.11.** *Both  $(\hat{L}_{\text{RHMC}}^{(\gamma)}, \text{Dom}(\hat{L}_{\text{RHMC}}^{(\gamma)}))$  and  $(\hat{L}_{\text{LD}}^{(\gamma)}, \text{Dom}(\hat{L}_{\text{LD}}^{(\gamma)}))$  are second-order lifts of  $(\frac{1}{2}L, \text{Dom}(L))$  for any choice of  $\gamma > 0$ .*

*Proof.* Firstly, since  $\mathcal{D}$  forms a core for  $(L, \text{Dom}(L))$ , the condition (B.1) is satisfied for both  $\hat{L}_{\text{RHMC}}^{(\gamma)}$  and  $\hat{L}^{(\gamma)}$ . Secondly, both generators are obtained from  $\hat{L}$  by adding a term that only acts on the velocity and leaves  $\kappa(x, \cdot)$  invariant. This means that

$$\hat{L}_{\text{RHMC}}^{(\gamma)}(f \circ \pi) = \hat{L}_{\text{LD}}^{(\gamma)}(f \circ \pi) = \hat{L}(f \circ \pi)$$

for any  $f \in \mathcal{D}$ , so that (B.2) and (B.3) are immediately satisfied.  $\square$

**Remark B.12.** Suppose that  $M \subseteq \mathbb{R}^d$  is closed and convex with smooth boundary, so that  $M$  is a Riemannian manifold with boundary when equipped with the flat Euclidean metric, as in Remark B.3. The construction of RHMC and Langevin dynamics on  $M$  then reduces to the usual processes on  $\mathbb{R}^d$  constrained to  $M$  via a specular reflection at the boundary. In case  $U$  is constant, the invariant probability measure  $\hat{\mu}$  is the product of the uniform distribution on  $M$  and a standard normal distribution in the velocity. In this case, the Hamiltonian dynamics on  $M$  reduces to billiards in a convex set, see [237].

#### B.4.2. Space-time Poincaré inequality on Riemannian manifolds

Albritton, Armstrong, Mourrat and Novack [6] developed an approach to proving exponential decay for the kinetic Fokker-Planck equation using space-time Poincaré inequalities. Quantitative upper bounds for various non-reversible Markov processes were then obtained by Cao, Lu and Wang [61, 186] by proving such space-time Poincaré inequalities for the Kolmogorov backward equations associated to these processes. In [99], the authors show how the framework of second-order lifts together with a divergence lemma as proved in Section B.3 can be used to prove such space-time Poincaré inequalities and thereby obtain quantitative bounds on the relaxation time of non-reversible Markov processes. In this section, we prove a space-time Poincaré inequality for RHMC and Langevin dynamics on a Riemannian manifold  $M$  with reflection at the boundary.

In the following, we fix  $T \in (0, \infty)$  and write  $L^2(\bar{\mu} \otimes \kappa) = L^2([0, T] \times TM, \bar{\mu} \otimes \kappa)$ . We consider the operator  $(A, \text{Dom}(A))$  on  $L^2(\bar{\mu} \otimes \kappa)$  defined by

$$Af = -\partial_t f + \hat{L}f \tag{B.33}$$

with domain consisting of all functions  $f \in L^2(\bar{\mu} \otimes \kappa)$  such that  $f(\cdot, x, v)$  is absolutely continuous for  $\hat{\mu}$ -a.e.  $(x, v) \in TM$  with  $\partial_t f \in L^2(\bar{\mu} \otimes \kappa)$ , and  $f(t, \cdot) \in \text{Dom}(\hat{L})$  for  $\lambda$ -a.e.  $t \in [0, T]$  with  $\hat{L}f \in L^2(\bar{\mu} \otimes \kappa)$ . Here  $\hat{L}$  is the generator (B.30) of the Hamiltonian flow.

We consider the adjoint  $(A^*, \text{Dom}(A^*))$  in  $L^2(\bar{\mu} \otimes \kappa)$  of the operator  $(A, \text{Dom}(A))$ . An integration by parts and Lemma B.10 yield

$$(Af, g)_{L^2(\bar{\mu} \otimes \kappa)} = (f, -Ag)_{L^2(\bar{\mu} \otimes \kappa)} + \int_{TM} (f(T, \cdot)g(T, \cdot) - f(0, \cdot)g(0, \cdot)) d\hat{\mu}$$

#### B.4. Optimal lifts on Riemannian manifolds with boundary

for all  $f, g \in \text{Dom}(A)$ . Thus functions  $g \in \text{Dom}(A)$  satisfying  $g(0, \cdot) = g(T, \cdot) = 0$  are contained in the domain of  $A^*$ , and

$$A^*g = -Ag = \partial_t g - \hat{L}g. \quad (\text{B.34})$$

In the following, for  $(t, x, v) \in [0, \infty) \times \text{TM}$  we set  $\pi(t, x, v) = (t, x)$ , whereas  $\pi(x, v) = x$ .

**Remark B.13.** For  $h$  and  $g$  as in [Theorem B.5](#) and  $(x, v) \in \text{TM}$  with  $x \in \partial M$ , we have

$$\begin{aligned} (h \circ \pi + \hat{L}(g \circ \pi))(\cdot, x, R_x v) &= h(\cdot, x) + \langle v - 2N(x)\langle N(x), v \rangle, \nabla_x g(\cdot, x) \rangle \\ &= (h \circ \pi + \hat{L}(g \circ \pi))(\cdot, x, v), \end{aligned}$$

since  $\langle N, \nabla_x g \rangle = \partial_N g = 0$  on  $\partial M$ , and

$$(h \circ \pi + \hat{L}(g \circ \pi))(0, \cdot) = (h \circ \pi + \hat{L}(g \circ \pi))(T, \cdot) = 0.$$

so that  $h \circ \pi + \hat{L}(g \circ \pi) \in \text{Dom}(A^*)$ .

**Lemma B.14.** Let  $X = \begin{pmatrix} -h \\ \nabla_x g \end{pmatrix}$  as in [Theorem B.5](#) and  $x \in M$ . Then

$$\int_{\text{T}_x M} A^*(h \circ \pi + \hat{L}(g \circ \pi)) \kappa(x, dv) = \bar{\nabla}^* X \circ \pi, \quad \text{and} \quad (\text{B.35})$$

$$\|A^*(h \circ \pi + \hat{L}(g \circ \pi)) - \bar{\nabla}^* X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)} \leq \sqrt{2} \|\bar{\nabla} X\|_{L^2(\bar{\mu})}. \quad (\text{B.36})$$

*Proof.* The proof in the Riemannian setting works as in [\[99, Lemma 22\]](#).  $\square$

We are now ready to state and prove the space-time Poincaré inequality that is the key ingredient for deriving the relaxation time of lifts.

For any  $x \in M$  we let  $H^{-1}(\kappa_x)$  denote the dual of the Gaussian Sobolev space  $H^{1,2}(\kappa_x) = H^{1,2}(\text{T}_x M, \kappa(x, \cdot))$ . Note that

$$\|f\|_{H^{1,2}(\kappa_x)}^2 = \|f\|_{L^2(\kappa_x)}^2 + \|\nabla f\|_{L^2(\kappa_x)}^2 = (f, (I - L_v)f)_{L^2(\kappa_x)},$$

where  $L_v$  is given by [\(B.31\)](#). For  $f \in L^2(\kappa_x)$ , the  $H^{-1}(\kappa_x)$ -norm is then

$$\|f\|_{H^{-1}(\kappa_x)}^2 = (f, (I - L_v)^{-1}f)_{L^2(\kappa_x)},$$

where the inverse exists by the spectral decomposition. In particular, a spectral decomposition shows that for  $f \in H^{2,2}(\kappa_x)$ ,

$$\|L_v f\|_{H^{-1}(\kappa_x)}^2 = (L_v f, (I - L_v)^{-1} L_v f)_{L^2(\kappa_x)} \leq (f, -L_v f)_{L^2(\kappa_x)} \quad (\text{B.37})$$

and

$$\|L_v f\|_{H^{-1}(\kappa_x)}^2 \geq \frac{1}{2} (f, -L_v f)_{L^2(\kappa_x)}. \quad (\text{B.38})$$

## Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

Let  $L^2(\lambda \otimes \mu; H^{1,2}(\kappa))$  denote the subspace of functions  $f \in L^2(\bar{\mu} \otimes \kappa)$  such that  $f(t, x, \cdot) \in H^{1,2}(\kappa_x)$  for  $\lambda \otimes \mu$ -a.e.  $(t, x) \in [0, T] \times M$  and

$$\|f\|_{L^2(\lambda \otimes \mu; H^{1,2}(\kappa))}^2 = \int_{[0, T] \times M} \|f(t, x, \cdot)\|_{H^{1,2}(\kappa_x)}^2 \mu(dx) \lambda(dt) < \infty.$$

Its dual is the similarly defined  $L^2(\lambda \otimes \mu; H^{-1}(\kappa))$ . In particular, for  $f \in L^2(\lambda \otimes \mu; H^{-1}(\kappa))$  and  $g \in L^2(\lambda \otimes \mu; H^{1,2}(\kappa))$ ,

$$(f, g)_{L^2(\bar{\mu} \otimes \kappa)} \leq \|f\|_{L^2(\lambda \otimes \mu; H^{-1}(\kappa))} \|g\|_{L^2(\lambda \otimes \mu; H^{1,2}(\kappa))}.$$

**Theorem B.15** (Space-time Poincaré inequality). *For any  $T > 0$ , there exist constants  $C_0(T), C_1(T) > 0$  such that for all functions  $f \in \text{Dom}(A)$ ,*

$$\|f - \int f d(\bar{\mu} \otimes \kappa)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq C_0(T) \|Af\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + C_1(T) \|(\Pi - I)f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \quad (\text{B.39})$$

and

$$\|f - \int f d(\bar{\mu} \otimes \kappa)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq 2C_0(T) \|Af\|_{L^2(\lambda \otimes \mu; H^{-1}(\kappa))}^2 + C_1(T) \|(\Pi - I)f\|_{L^2(\bar{\mu} \otimes \kappa)}^2, \quad (\text{B.40})$$

where

$$C_0(T) = 2c_0(T) = 4T^2 + 86 \frac{1}{m},$$

$$C_1(T) = 3 + 4c_1(T) = 1163 + \frac{3964}{mT^2} + 172 \max\left(\frac{1}{m}, \frac{T^2}{\pi^2}\right) \rho.$$

The proof of (B.39) is an adaptation of [99, Theorem 23] to the Riemannian setting with boundary. An expression similar to (B.40) has first been shown in the Euclidean setting in [61] but with non-optimal constants. The treatment of Langevin dynamics requires the  $L^2(\lambda \otimes \mu; H^{-1}(\kappa))$ -norm of  $Af$  in (B.40).

In the proof we apply the following *space-time property of lifts*, see [99, Lemma 18]. It states that for any  $f, g, h \in L^2(\lambda \otimes \mu)$  such that  $f \circ \pi \in \text{Dom}(A)$  and  $g(t, \cdot) \in \text{Dom}(L)$  for a.e.  $t \in [0, T]$  with  $Lg \in L^2(\lambda \otimes \mu)$ , we have

$$\frac{1}{2} \left( A(f \circ \pi), h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} = -\frac{1}{2} (\partial_t f, h)_{L^2(\lambda \otimes \mu)} + \mathcal{E}_T(f, g), \quad (\text{B.41})$$

where  $\mathcal{E}_T(f, g) = \int_0^T \mathcal{E}(f(t, \cdot), g(t, \cdot)) dt$  is the time-integrated Dirichlet form associated to  $(\frac{1}{2}L, \text{Dom}(L))$ .

*Proof of Theorem B.15.* Fix a function  $f_0 \in \text{Dom}(A)$ , let  $f = f_0 - \int f_0 d(\bar{\mu} \otimes \kappa)$ , and let  $\tilde{f}: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  be such that  $\Pi f = \tilde{f} \circ \pi$ . First note that

$$\|f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 = \|f - \Pi f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + \|\tilde{f}\|_{L^2(\bar{\mu})}^2, \quad (\text{B.42})$$

#### B.4. Optimal lifts on Riemannian manifolds with boundary

so that it suffices to bound  $\|\tilde{f}\|_{L^2(\bar{\mu})}^2$  by  $\|f - \Pi f\|_{L^2(\bar{\mu} \otimes \kappa)}^2$  and  $\|Af\|_{L^2(\bar{\mu} \otimes \kappa)}^2$  to obtain (B.39), and by  $\|f - \Pi f\|_{L^2(\bar{\mu} \otimes \kappa)}^2$  and  $\|Af\|_{L^2(\bar{\mu}, H^{-1}(\kappa))}^2$  to obtain (B.40).

By Theorem B.5 applied to  $\tilde{f}$ , there exist functions  $h \in H_D^{1,2}(\bar{\mu})$  and  $g \in H_D^{2,2}(\bar{\mu})$  satisfying (B.15) such that

$$\tilde{f} = \partial_t h - Lg.$$

In particular,

$$\|\tilde{f}\|_{L^2(\bar{\mu})}^2 = (\tilde{f}, \tilde{f})_{L^2(\bar{\mu})} = (\tilde{f}, \partial_t h - Lg)_{L^2(\bar{\mu})}. \quad (\text{B.43})$$

Using the space-time property of lifts (B.41), we see that

$$\begin{aligned} (\tilde{f}, \partial_t h - Lg)_{L^2(\bar{\mu})} &= (-\partial_t \tilde{f}, h)_{L^2(\bar{\mu})} + 2\mathcal{E}_T(\tilde{f}, g) \\ &= \left( A(\tilde{f} \circ \pi), h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} \\ &= \left( Af, h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} \\ &\quad + \left( A(\Pi f - f), h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)}. \end{aligned}$$

Let us consider the two summands separately.

Since  $h \circ \pi + \hat{L}(g \circ \pi) \in L^2(\lambda \otimes \mu; H^{1,2}(\kappa))$ , the first summand can be estimated as

$$\begin{aligned} \left( Af, h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} &\leq \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} \|h \circ \pi + \hat{L}(g \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)} \\ &\leq c_0(T)^{\frac{1}{2}} \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} \|\tilde{f}\|_{L^2(\bar{\mu})}, \end{aligned}$$

where we used that

$$\|h \circ \pi + \hat{L}(g \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 = \|h\|_{L^2(\bar{\mu})}^2 + \|\nabla g\|_{L^2(\bar{\mu})}^2 \leq c_0(T) \|\tilde{f}\|_{L^2(\bar{\mu})}^2$$

by (B.17).

By Remark B.13,  $h \circ \pi + \hat{L}(g \circ \pi) \in \text{Dom}(A^*)$ , so that the second summand satisfies

$$\left( A(\Pi f - f), h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} = \left( \Pi f - f, A^*(h \circ \pi + \hat{L}(g \circ \pi)) \right)_{L^2(\bar{\mu} \otimes \kappa)},$$

Since  $\tilde{f} = \bar{\nabla}^* X$ , Lemma B.14 and Theorem B.5 yield

$$\begin{aligned} \|A^*(h \circ \pi + \hat{L}(g \circ \pi))\|_{L^2(\bar{\mu} \otimes \kappa)}^2 &= \|\tilde{f}\|_{L^2(\bar{\mu})}^2 + \|A^*(h \circ \pi + \hat{L}(g \circ \pi)) - \bar{\nabla}^* X \circ \pi\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \\ &\leq \|\tilde{f}\|_{L^2(\bar{\mu})}^2 + 2\|\bar{\nabla} X\|_{L^2(\bar{\mu})}^2 \leq (1 + 2c_1(T)) \|\tilde{f}\|_{L^2(\bar{\mu})}^2, \end{aligned}$$

so that the second summand can be estimated as

$$\left( A(\Pi f - f), h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} \leq (1 + 2c_1(T))^{\frac{1}{2}} \|f - \Pi f\|_{L^2(\bar{\mu} \otimes \kappa)} \|\tilde{f}\|_{L^2(\bar{\mu} \otimes \kappa)}.$$

Putting things together, we hence obtain by (B.43),

$$\|\tilde{f}\|_{L^2(\bar{\mu})} \leq c_0(T)^{\frac{1}{2}} \|Af\|_{L^2(\bar{\mu} \otimes \kappa)} + (1 + 2c_1(T))^{\frac{1}{2}} \|f - \Pi f\|_{L^2(\bar{\mu} \otimes \kappa)}$$

so that by (B.42),

$$\|f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq 2c_0(T) \|Af\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + (3 + 4c_1(T)) \|f - \Pi f\|_{L^2(\bar{\mu} \otimes \kappa)}^2$$

which proves (B.39).

Similarly, to obtain (B.40), the first summand can be estimated as

$$\begin{aligned} \left( Af, h \circ \pi + \hat{L}(g \circ \pi) \right)_{L^2(\bar{\mu} \otimes \kappa)} &\leq \|Af\|_{L^2(\bar{\mu}, H^{-1}(\kappa))} \|h \circ \pi + \hat{L}(g \circ \pi)\|_{L^2(\bar{\mu}, H^{1,2}(\kappa))} \\ &\leq (2c_0(T))^{\frac{1}{2}} \|Af\|_{L^2(\bar{\mu}, H^{-1}(\kappa))} \|\tilde{f}\|_{L^2(\bar{\mu})}, \end{aligned}$$

where we used that

$$\begin{aligned} &\|h \circ \pi + \hat{L}(g \circ \pi)\|_{L^2(\bar{\mu}, H^{1,2}(\kappa))}^2 \\ &= \|h \circ \pi + \hat{L}(g \circ \pi)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + \|\nabla_v(h \circ \pi + \hat{L}(g \circ \pi))\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \\ &= \|h\|_{L^2(\bar{\mu})}^2 + \|\nabla g\|_{L^2(\bar{\mu})}^2 + \|v \cdot \nabla g\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \\ &\leq 2c_0(T) \|\tilde{f}\|_{L^2(\bar{\mu})}^2 \end{aligned}$$

and  $c_0(T)$  is the constant from (B.17). One then concludes analogously.  $\square$

As demonstrated in [99], such a space-time Poincaré inequality yields the following exponential decay in  $T$ -average of the associated semigroup.

**Theorem B.16** (Exponential Decay). *Let  $\gamma > 0$  and  $T > 0$  and set*

$$\nu = \frac{\gamma}{\gamma^2 C_0(T) + C_1(T)}.$$

- (i) *The transition semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  of Riemannian RHMC with reflection at the boundary and refresh rate  $\gamma$  satisfies*

$$\frac{1}{T} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})} ds \leq e^{-\nu t} \|f\|_{L^2(\hat{\mu})}$$

*for all  $f \in L_0^2(\hat{\mu})$ .*

- (ii) *The transition semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  of Riemannian Langevin dynamics with reflection at the boundary and friction  $\gamma$  satisfies*

$$\frac{1}{T} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})} ds \leq e^{-\frac{1}{2}\nu t} \|f\|_{L^2(\hat{\mu})}$$

*for all  $f \in L_0^2(\hat{\mu})$ .*

*Proof.* (i) The proof in this case uses (B.39) and is identical to that of [99, Theorem 17].

(ii) Let  $f \in L_0^2(\hat{\mu})$ . Then by (B.32) and the antisymmetry of  $\hat{L}$ ,

$$\begin{aligned} \frac{d}{dt} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds &= \|\hat{P}_{t+T}^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 - \|\hat{P}_t^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 \\ &= \int_t^{t+T} \frac{d}{ds} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds = 2 \int_t^{t+T} \left( \hat{P}_s^{(\gamma)} f, (\hat{L} + \gamma L_v) \hat{P}_s^{(\gamma)} f \right)_{L^2(\hat{\mu})} ds \\ &= -2\gamma \int_t^{t+T} \left( \hat{P}_s^{(\gamma)} f, -L_v \hat{P}_s^{(\gamma)} f \right)_{L^2(\hat{\mu})} ds \leq -2\gamma \|L_v \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\lambda \otimes \mu, H^{-1}(\kappa))}^2, \end{aligned}$$

where we used (B.37) in the last step. Since  $\hat{P}_{t+}^{(\gamma)} f \in \text{Dom}(A)$  and  $A\hat{P}_{t+}^{(\gamma)} f = \gamma L_v \hat{P}_{t+}^{(\gamma)} f$ , the space-time Poincaré inequality (B.40) yields

$$\|\hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq 2\gamma^2 C_0(T) \|L_v \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\lambda \otimes \mu, H^{-1}(\kappa))}^2 + C_1(T) \|(\Pi - I) \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2.$$

The Gaussian Poincaré inequality and (B.38) yield

$$\|(\Pi - I) \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq 2 \|L_v \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu}, H^{-1}(\kappa))}^2,$$

so that

$$\|\hat{P}_{t+}^{(\gamma)} f\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \leq (2\gamma^2 C_0(T) + 2C_1(T)) \|L_v \hat{P}_{t+}^{(\gamma)} f\|_{L^2(\lambda \otimes \mu, H^{-1}(\kappa))}^2.$$

Therefore, Grönwall's inequality yields

$$\frac{1}{T} \int_t^{t+T} \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \leq e^{-\nu t} \frac{1}{T} \int_0^T \|\hat{P}_s^{(\gamma)} f\|_{L^2(\hat{\mu})}^2 ds \leq e^{-\nu t} \|f\|_{L^2(\hat{\mu})}^2,$$

and we obtain the exponential contractivity in  $T$ -average with rate  $\frac{1}{2}\nu$  by the Cauchy-Schwarz inequality.  $\square$

### B.4.3. Optimality of randomised Hamiltonian Monte Carlo and Langevin dynamics on Riemannian manifolds with boundary

The exponential decay of  $T$ -averages in Theorem B.16 immediately yields the bounds

$$t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq \frac{1}{\nu} + T \quad \text{or} \quad t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq \frac{2}{\nu} + T$$

for the relaxation times of Riemannian RHMC with reflection at the boundary and refresh rate  $\gamma$ , as well as Riemannian Langevin dynamics with reflection at the boundary and friction  $\gamma$ , respectively, see [99, Lemma 8]. For a fixed  $T > 0$ , the rate  $\nu$  is maximised by choosing  $\gamma = \sqrt{C_1(T)/C_0(T)}$ , and this choice yields the upper bounds

$$t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq 2\sqrt{C_0(T)C_1(T)} + T \quad \text{and} \quad t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq 4\sqrt{C_0(T)C_1(T)} + T,$$

respectively. To obtain explicit bounds, we set  $T = \frac{\pi}{\sqrt{m}}$ , which yields

$$C_0(T) = (4\pi^2 + 86) \frac{1}{m} \quad \text{and} \quad C_1(T) = 1163 + \frac{3964}{\pi^2} + 172 \frac{\rho}{m}. \quad (\text{B.44})$$

**Corollary B.17** (*C*-optimality of RHMC and Langevin dynamics). *Let  $c \geq 0$ . On the class of potentials satisfying [Assumption B.1](#) with  $\rho \leq cm$  and [Assumption B.2](#), Riemannian randomised Hamiltonian Monte Carlo with reflection at the boundary with refresh rate*

$$\gamma = \sqrt{\frac{(1163 + \frac{3964}{\pi^2})m + 172\rho}{4\pi^2 + 86}} \quad (\text{B.45})$$

*is a  $C$ -optimal lift of the corresponding overdamped Langevin diffusion with reflection at the boundary with*

$$C = 2 \left( 887\sqrt{1 + c/9} + \pi \right).$$

*Similarly, on the same class of potentials, Langevin dynamics with reflection at the boundary with friction  $\gamma$  given by (B.45) is a  $C$ -optimal lift of the same process with*

$$C = 2 \left( 1773\sqrt{1 + c/9} + \pi \right).$$

*More generally, for  $A \in [1, \infty)$ , Riemannian RHMC is a  $C(A)$ -optimal lift with*

$$C(A) = (3381 + 344c)A + 2\pi$$

*for any  $\gamma$  satisfying  $1/A \leq \frac{\gamma}{\sqrt{m}} \leq A$ . Similarly, for the same choices of  $\gamma$ , Riemannian Langevin dynamics is a  $C(A)$ -optimal lift with*

$$C(A) = (6761 + 344c)A + 2\pi.$$

*Proof.* The relaxation time of the corresponding overdamped Langevin diffusion with reflection at the boundary is  $\frac{2}{m}$ . Choosing  $T = \frac{\pi}{\sqrt{m}}$  and plugging (B.44) into the expression for  $\nu$  gives

$$\frac{1}{\nu} = 2\sqrt{(4\pi^2 + 86) \left( 1163 + \frac{3964}{\pi^2} \right)} \frac{1}{\sqrt{m}} \sqrt{1 + \frac{172}{1163 + \frac{3964}{\pi^2}} \frac{\rho}{m}},$$

so that

$$t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq \frac{1}{\sqrt{m}} \left( 887\sqrt{1 + \frac{\rho}{9m}} + \pi \right) \quad \text{and} \quad t_{\text{rel}}(\hat{P}^{(\gamma)}) \leq \frac{1}{\sqrt{m}} \left( 1773\sqrt{1 + \frac{\rho}{9m}} + \pi \right)$$

for RHMC and Langevin dynamics, respectively. The final two statements follow similarly from the expression

$$\frac{1}{\nu} = \gamma C_0(T) + \frac{1}{\gamma} C_1(T) \leq \sqrt{m} A C_0(T) + \frac{A}{\sqrt{m}} C_1(T). \quad \square$$

**Remark B.18.** The explicit values of the constants in the divergence lemma [Theorem B.5](#) are surely not optimal, so that we expect [Corollary B.17](#) to hold with better constants. Nonetheless, in the case of mildly negative Bakry-Émery curvature  $\text{Ric} + \nabla^2 U$ , i.e. if the lower bound  $\rho$  in [Assumption B.1](#) (ii) is of the same order as the inverse Poincaré constant  $m$ , the result shows that Riemannian RHMC and Riemannian Langevin dynamics with reflection at the boundary are close to being optimal lifts.

## B.5. Appendix

In this appendix, for the convenience of the reader, we recall some concepts from Riemannian geometry and fix notation. Particular emphasis is placed on manifolds with boundary and curvature properties of the boundary, measured in terms of the second fundamental form. We finally provide a self-contained proof for the generalisation of the Reilly formula stated in [Lemma B.1](#).

We mainly follow [\[164\]](#) and [\[165\]](#) as well as [\[111\]](#), and point to these sources for a more comprehensive overview of the topic.

### B.5.1. Manifolds with boundary and manifolds with corners

A *d-dimensional topological manifold with boundary* is a second-countable Hausdorff space such that every point has a neighbourhood that is homeomorphic to an open subset of  $\mathbb{R}^n$  or a relatively open subset of  $\mathbb{R}_+^d = \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_d \geq 0\}$ . A chart  $(U, \varphi)$  is an open subset  $U \subset M$  together with a map  $\varphi: U \rightarrow \mathbb{R}^d$  that is a homeomorphism onto an open subset of  $\mathbb{R}^d$  or  $\mathbb{R}_+^d$ . It is called a *boundary chart* if  $\varphi(U)$  is an open subset of  $\mathbb{R}_+^d$  such that  $\varphi(U) \cap \partial\mathbb{R}_+^d \neq \emptyset$  and an *interior chart* otherwise. The *boundary*  $\partial M$  of  $M$  is the set of all  $p \in M$  that are in the domain of some boundary chart  $(U, \varphi)$  with  $\varphi(p) \in \partial\mathbb{R}_+^d$ . In case  $\partial M = \emptyset$ ,  $M$  is just a conventional  $d$ -dimensional topological manifold (without boundary). The interior  $\mathring{M} = M \setminus \partial M$  of  $M$  is an open subset of  $M$  and a  $d$ -dimensional topological manifold without boundary, whereas  $\partial M$  is a closed subset of  $M$  and a  $d - 1$ -dimensional topological manifold without boundary.

A *d-dimensional smooth manifold with boundary* is a  $d$ -dimensional topological manifold with boundary such that any two charts  $(U, \varphi)$  and  $(V, \psi)$  are smoothly compatible, i.e.

$$\varphi \circ \psi^{-1}: \psi(U \cap V) \rightarrow \varphi(U \cap V)$$

is a smooth diffeomorphism. Here smoothness of a function  $f$  defined on a subset  $A$  of  $\mathbb{R}^n$  that may not be open means that it is the restriction to  $A$  of some smooth function  $\tilde{f}$  defined on an open set  $\tilde{A} \subset \mathbb{R}^n$  containing  $A$ . A map  $f$  between two smooth manifolds  $M$  and  $N$  is called smooth if for any charts  $(U, \varphi)$  in  $M$  and  $(V, \psi)$  in  $N$  with  $f(U) \subset V$ , the composition

$$\psi \circ f \circ \varphi^{-1}: \varphi(U) \rightarrow \psi(V)$$

is smooth. The set of smooth functions from  $M$  to  $N$  is denoted  $C^\infty(M, N)$ , and we set  $C^\infty(M) = C^\infty(M, \mathbb{R})$ . The boundary  $\partial M$  together with the smooth structure given by the restrictions of the charts of  $M$  to  $\partial M$  is a  $(d - 1)$ -dimensional smooth manifold (without boundary).

In the following, let  $M$  be a  $d$ -dimensional smooth manifold with boundary. As in the case of smooth manifolds without boundary, for any point  $p \in M$ , the tangent space  $T_p M$  is the vector space of derivations at  $p$ , i.e. linear maps  $v: C^\infty(M) \rightarrow \mathbb{R}$  such that

$$v(fg) = f(p)v(g) + g(p)v(f) \quad \text{for all } f, g \in C^\infty(M).$$

## Appendix B. Space-time divergence lemmas on Riemannian manifolds with boundary

Note that  $T_p M$  is a  $d$ -dimensional vector space, regardless of whether  $p \in \partial M$  or  $p \in \overset{\circ}{M}$ . As usual, the *tangent bundle*  $TM$  is the disjoint union

$$TM = \coprod_{p \in M} T_p M$$

of the tangent spaces and is again a smooth manifold. A *vector field* on  $M$  is a Borel-measurable section of  $TM$ , and the space of vector fields on  $M$  is denoted by  $\Gamma(TM)$ . For  $k \in \mathbb{N}_0 \cup \{\infty\}$ , a  $C^k$ -*vector field* on  $M$  is a  $C^k$ -section of  $TM$ . The space of  $C^k$ -vector fields is denoted by  $\Gamma_{C^k}(TM)$ , and we denote the space of smooth vector fields by  $\mathfrak{X}(M) = \Gamma_{C^\infty}(TM)$ .

As in the case without boundary, a *Riemannian metric*  $g$  on  $M$  is a smooth covariant 2-tensor field on  $M$  that is symmetric and positive definite. This is nothing else than an inner product  $g_p$  on each tangent space  $T_p M$  that varies smoothly in  $p$ . Such a pair  $(M, g)$  is called a *d-dimensional Riemannian manifold with boundary*. The smooth submanifold  $\partial M$  of  $M$  can be endowed with a Riemannian metric induced by  $g$  in the following way. Letting  $\iota: \partial M \rightarrow M$  denote the inclusion map and  $d\iota_p: T_p \partial M \rightarrow T_{\iota(p)} M$  its differential at  $p \in \partial M$ , the smooth submanifold  $\partial M$  of  $M$  can be endowed with the *induced Riemannian metric*

$$\tilde{g}_p(v, w) = g_{\iota(p)}(d\iota_p(v), d\iota_p(w)).$$

Unwinding the definition,  $d\iota_p(v)$  is simply the derivation that acts as  $d\iota_p(v)(f) = v(f|_{\partial M})$  for all  $f \in C^\infty(M)$ . For sake of simplicity, we usually identify  $T_p \partial M$  with its image  $d\iota_p(T_p \partial M)$  in  $T_{\iota(p)} M = T_p M$  and again denote  $\tilde{g}$  by  $g$ . Thus the induced metric is just the restriction of  $g$  to vectors tangent to  $\partial M$ . Once a Riemannian metric  $g$  is fixed, for any  $p \in M$  and  $v, w \in T_p M$  we also write  $\langle v, w \rangle = g(v, w)$  and  $|v| = \langle v, v \rangle^{1/2}$  for the inner product and norm on  $T_p M$  given by  $g$ .

In the following, let  $(M, g)$  be a  $d$ -dimensional Riemannian manifold with boundary. The normal space  $N_p \partial M$  at  $p \in \partial M$  is the orthogonal complement of  $T_p \partial M$  in  $T_p M$  with respect to the inner product  $\langle \cdot, \cdot \rangle$ , and the normal bundle is the disjoint union

$$N\partial M = \coprod_{p \in \partial M} N_p \partial M$$

of the normal spaces. The space of smooth sections of  $N\partial M$  is denoted by  $\mathfrak{N}(\partial M)$ . There exists a unique smooth outward-pointing unit normal vector field  $N \in \mathfrak{N}(\partial M)$  along all of  $\partial M$ , see [165, Prop. 2.17].

Since the product of two smooth manifolds with boundary is not necessarily again a smooth manifold with boundary, it is necessary to introduce *smooth manifolds with corners*, see [164, Chapter 16]. They are defined analogously to smooth manifolds with boundaries, with the difference being that every point has a neighbourhood that is homeomorphic to a subset of  $\mathbb{R}_+^d = \{(x_1, \dots, x_d) \in \mathbb{R}^d: x_1 \geq 0, \dots, x_d \geq 0\}$ . In particular, any smooth manifold with boundary is a smooth manifold with corners. The corner points of  $\mathbb{R}_+^d$  are all the points at which more than one coordinate vanishes, and the corner points

of a smooth manifold with corners are all the points  $p$  that are in the domain of some corner chart mapping  $p$  to a corner point of  $\overline{\mathbb{R}}_+^d$ . The tangent bundle, vector fields and Riemannian metrics are defined analogously to the case with boundary. The product of a finite number of smooth manifolds with corners is again a smooth manifold with corners, and removing the corner points from a smooth manifold with corners yields a smooth manifold with boundary.

### B.5.2. The Hessian and the second fundamental form

Let  $(M, g)$  be a  $d$ -dimensional Riemannian manifold with boundary. The *Riemannian gradient* of a smooth function  $f \in C^\infty(M)$  is the smooth vector field  $\nabla f$  defined by

$$\langle \nabla f, v \rangle = df(v) \quad \text{for all } v \in TM,$$

where  $df$  is the differential of  $f$ .

The Levi-Civita connection

$$\nabla: \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M), \quad (X, Y) \mapsto \nabla_X Y$$

defines the *covariant derivative* of a vector field  $Y$  in direction of the vector field  $X$ . Since  $\nabla_X Y|_p$  only depends on  $X$  through the value of  $X$  at  $p$ , the derivative  $\nabla_v Y$  of  $Y$  at  $p$  in direction  $v$  is defined for any  $v \in T_p M$ . Furthermore,  $\nabla$  uniquely determines a connection in each tensor bundle  $T^{k,l}TM$  that agrees with the usual Levi-Civita connection on  $T^{1,0}TM = TM$  and the Riemannian gradient on  $T^{0,0} = M \times \mathbb{R}$  in the sense that  $\nabla_X f = \langle \nabla f, X \rangle$  for all  $f \in C^\infty(M)$  and  $X \in \mathfrak{X}(M)$ , see [165, Prop. 4.15].

The *Hessian* of a smooth function  $f \in C^\infty(M)$  is the symmetric 2-tensor  $\nabla^2 f$ . Its action on vector fields  $X, Y \in \mathfrak{X}(M)$  is given by

$$\nabla^2 f(X, Y) = \nabla_{X,Y}^2 f = \nabla_X(\nabla_Y f) - \nabla_{\nabla_X Y} f.$$

For any  $p \in \partial M$  and  $Z \in T_p M$  denote by  $Z_p^\top$  and  $Z_p^\perp$  the projections of  $Z_p$  onto  $T_p \partial M$  and  $N_p \partial M$ , respectively. Extending vector fields  $X$  and  $Y$  on  $\partial M$  arbitrarily to vector fields on  $M$ , we can decompose

$$\nabla_X Y = (\nabla_X Y)^\top + (\nabla_X Y)^\perp.$$

The Gauß formula, see [165, Theorem 8.2], identifies  $(\nabla_X Y)^\top$  as the covariant derivative  $\nabla_X^{\partial M} Y$  with respect to the Levi-Civita connection  $\nabla^{\partial M}$  on  $T\partial M$  of the induced metric  $g$  on  $\partial M$ . This leads to the *second fundamental form*  $\mathbb{I}: \mathfrak{X}(\partial M) \times \mathfrak{X}(\partial M) \rightarrow \mathfrak{N}(\partial M)$  defined by

$$\mathbb{I}(X, Y) = (\nabla_X Y)^\perp = \nabla_X Y - \nabla_X^{\partial M} Y. \quad (\text{B.46})$$

The *scalar second fundamental form*  $h: \mathfrak{X}(\partial M) \times \mathfrak{X}(\partial M) \rightarrow C^\infty(\partial M)$  is then defined by

$$h(X, Y) = \langle \mathbb{I}(X, Y), N \rangle = \langle \nabla_X Y, N \rangle.$$

Both  $\mathbb{I}$  and  $h$  are bilinear over  $C^\infty(M)$  and symmetric, and the values of  $h(X, Y)$  and  $\mathbb{I}(X, Y)$  at  $p \in \partial M$  only depend on the values of  $X$  and  $Y$  at  $p$ . The *mean curvature*  $H$  is

$$H = \text{tr}(h),$$

where  $\text{tr}$  is the trace operator.

### B.5.3. The Laplace-Beltrami operator

Let  $(M, g)$  be a  $d$ -dimensional Riemannian manifold with corners. The *Riemannian volume measure*  $\nu_M$  is the Borel measure on  $M$  given by

$$d\nu_M = \sqrt{\det g} |dx^1 \wedge \cdots \wedge dx^d|$$

in any local coordinates  $(x^1, \dots, x^d)$ . The *divergence* of a vector field  $X \in \Gamma_{C^1}(TM)$  is  $\text{div}(X) = \text{tr}(\nabla X)$ . In local coordinates  $X = \sum_{i=1}^d X^i \frac{\partial}{\partial x^i}$  it takes the form

$$\text{div}(X) = \frac{1}{\sqrt{\det g}} \sum_{i=1}^d \frac{\partial}{\partial x^i} \left( X^i \sqrt{\det g} \right).$$

It satisfies the integration-by-parts identity

$$\int_M \langle \nabla f, X \rangle d\nu_M = \int_{\partial M} f \langle X, N \rangle d\nu_{\partial M} - \int_M (f \text{div} X) d\nu_M,$$

where  $\nu_{\partial M}$  is the Riemannian volume measure on  $(\partial M, g)$ . The *Laplace-Beltrami operator* is the linear operator  $\Delta: C^\infty(M) \rightarrow C^\infty(M)$  defined by

$$\Delta u = \text{div}(\nabla u) = \text{tr}(\nabla^2 u).$$

In coordinates it takes the form

$$\Delta u = \frac{1}{\sqrt{\det g}} \sum_{i,j=1}^d \frac{\partial}{\partial x^i} \left( (g^{-1})_{ij} \sqrt{\det g} \frac{\partial u}{\partial x^j} \right).$$

If  $M \subset \mathbb{R}^d$  with the Euclidean metric, these definitions reduce to the usual divergence and Laplacian. The definition of the Laplace-Beltrami operator  $\Delta$  immediately implies the integration-by-parts identity

$$\int_M \varphi \Delta \psi d\nu_M = \int_{\partial M} \varphi \langle \nabla \psi, N \rangle d\nu_{\partial M} - \int_M \langle \nabla \varphi, \nabla \psi \rangle d\nu_M$$

for all smooth functions  $\varphi, \psi \in C^\infty(M)$ .

### B.5.4. A generalisation of the Reilly formula

Let  $(M, g)$  be a  $d$ -dimensional Riemannian manifold with boundary. In this section, we prove [Lemma B.1](#).

**Lemma B.1** (Generalised Reilly formula). *For any  $u \in C_c^\infty(M)$ ,*

$$\begin{aligned} & \int_M (|\nabla^2 u|^2 - (Lu)^2 + (\text{Ric} + \nabla^2 U)(\nabla u, \nabla u)) \, d\mu \\ &= \int_{\partial M} \left( (H + \partial_N U) (\partial_N u)^2 - 2(\partial_N u) L^{\partial M} u + h(\nabla^{\partial M} u, \nabla^{\partial M} u) \right) d\mu_{\partial M}, \end{aligned}$$

where  $L^{\partial M} = \Delta^{\partial M} - \langle \nabla^{\partial M} U, \nabla^{\partial M} \rangle$ .

*Proof.* Let  $u \in C^\infty(M)$ . The integration-by-parts identity [\(B.8\)](#) yields

$$\frac{1}{2} \int_M L |\nabla u|^2 \, d\mu = \frac{1}{2} \int_{\partial M} \langle \nabla |\nabla u|^2, N \rangle \, d\mu_{\partial M}$$

and

$$\int_M \langle \nabla u, \nabla Lu \rangle \, d\mu = - \int_M (Lu)^2 \, d\mu + \int_{\partial M} Lu \langle \nabla u, N \rangle \, d\mu_{\partial M}.$$

Thus integrating the Bochner identity [\(B.7\)](#) shows

$$\int_M (|\nabla^2 u|^2 - (Lu)^2 + (\text{Ric} + \nabla^2 U)(\nabla u, \nabla u)) \, d\mu = \int_{\partial M} \left\langle \frac{1}{2} \nabla |\nabla u|^2 - \Delta u \nabla u, N \right\rangle \, d\mu_{\partial M}.$$

By splitting  $\nabla u = \nabla^{\partial M} u + (\nabla u)^\perp$ , one sees that

$$\begin{aligned} \frac{1}{2} \langle \nabla |\nabla u|^2, N \rangle &= \langle \nabla \nabla u, \nabla u, N \rangle \\ &= \langle \nabla_{\nabla^{\partial M} u} \nabla^{\partial M} u, N \rangle + \langle \nabla_{\nabla^{\partial M} u} (\nabla u)^\perp, N \rangle + \langle \nabla_{(\nabla u)^\perp} \nabla u, N \rangle. \end{aligned}$$

The first summand is just the scalar second fundamental form  $h(\nabla^{\partial M} u, \nabla^{\partial M} u)$ . For the second summand, note that any vector field  $X$  defined on the boundary  $\partial M$  satisfies

$$\nabla_{X^\top} X^\perp = \langle X, N \rangle \nabla_{X^\top} N + \langle X^\top, \nabla \langle X, N \rangle \rangle N,$$

so that

$$\langle \nabla_{X^\top} X^\perp, N \rangle = \langle X^\top, \nabla \langle X, N \rangle \rangle = \langle X^\top, \nabla^{\partial M} \langle X, N \rangle \rangle.$$

This yields

$$\langle \nabla_{\nabla^{\partial M} u} (\nabla u)^\perp, N \rangle = \langle \nabla^{\partial M} u, \nabla^{\partial M} \langle \nabla u, N \rangle \rangle$$

and thus

$$\frac{1}{2} \langle \nabla |\nabla u|^2, N \rangle = h(\nabla^{\partial M} u, \nabla^{\partial M} u) + \langle \nabla^{\partial M} u, \nabla^{\partial M} \langle \nabla u, N \rangle \rangle + \langle \nabla_{(\nabla u)^\perp} \nabla u, N \rangle.$$

Furthermore, in an orthonormal boundary frame  $(E_1, \dots, E_{d-1}, N)$ , one has

$$\begin{aligned}\Delta u &= \text{tr}(\nabla^2 u) = \langle \nabla_N \nabla u, N \rangle + \sum_{i=1}^{d-1} \langle \nabla_{E_i} \nabla u, E_i \rangle \\ &= \langle \nabla_N \nabla u, N \rangle + \sum_{i=1}^{d-1} \left( \langle \nabla_{E_i} \nabla^{\partial M} u, E_i \rangle + \langle \nabla_{E_i} (\nabla u)^\perp, E_i \rangle \right) \\ &= \langle \nabla_N \nabla u, N \rangle + \Delta^{\partial M} u + \sum_{i=1}^{d-1} \langle \nabla_{E_i} (\nabla u)^\perp, E_i \rangle\end{aligned}$$

on  $\partial M$ . For any tangential vector field  $X$  defined on the boundary  $\partial M$ , one has

$$\begin{aligned}0 &= X \langle N, X \rangle = \langle \nabla_X N, X \rangle + \langle N, \nabla_X X \rangle \\ &= \langle \nabla_X N, X \rangle + \langle N, \mathbb{I}(X, X) + \nabla_X^{\partial M} X \rangle \\ &= \langle \nabla_X N, X \rangle + h(X, X).\end{aligned}$$

Hence  $\langle \nabla_{E_i} (\nabla u)^\perp, E_i \rangle = -\langle \nabla u, N \rangle h(E_i, E_i)$  and thus

$$\begin{aligned}Lu &= \Delta u - \langle \nabla U, \nabla u \rangle \\ &= \langle \nabla_N \nabla u, N \rangle + \Delta^{\partial M} u - \langle \nabla u, N \rangle H - \langle \nabla^{\partial M} U, \nabla^{\partial M} u \rangle - \langle (\nabla U)^\perp, (\nabla u)^\perp \rangle \\ &= \langle \nabla_N \nabla u, N \rangle + L^{\partial M} u - (H + \langle \nabla U, N \rangle) \langle \nabla u, N \rangle\end{aligned}$$

on the boundary  $\partial M$ .

Putting things together and integrating by parts on  $\partial M$  finally yields

$$\begin{aligned}&\int_{\partial M} \left\langle \frac{1}{2} \nabla |\nabla u|^2 - Lu \nabla u, N \right\rangle d\mu_{\partial M} \\ &= \int_{\partial M} \left( h(\nabla^{\partial M} u, \nabla^{\partial M} u) + \langle \nabla^{\partial M} u, \nabla^{\partial M} \langle \nabla u, N \rangle \rangle + \langle \nabla_{(\nabla u)^\perp} \nabla u, N \rangle \right) d\mu_{\partial M} \\ &\quad - \int_{\partial M} \left( \langle \nabla_N \nabla u, N \rangle + L^{\partial M} u - (H + \partial_N U) \langle \nabla u, N \rangle \right) \langle \nabla u, N \rangle d\mu_{\partial M} \\ &= \int_{\partial M} \left( h(\nabla^{\partial M} u, \nabla^{\partial M} u) - 2(\partial_N u) L^{\partial M} u + (H + \partial_N U)(\partial_N u)^2 \right) d\mu_{\partial M},\end{aligned}$$

completing the proof.  $\square$

## Statements and Declarations

**Funding.** Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – GZ2047/1, Projekt-ID 390685813. The authors were funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany’s Excellence Strategy – GZ 2047/1, Project-ID 390685813.

**Acknowledgements.** The authors would like to thank Lihan Wang, Arnaud Guillin and Léo Hahn for many helpful discussions.

## Appendix C.

# Convergence of non-reversible Markov processes via lifting and flow Poincaré inequality

A. Eberle, A. Guillin, L. Hahn, F. Lörler and M. Michel. *Convergence of non-reversible Markov processes via lifting and flow Poincaré inequality*. Mar. 2025. arXiv: [2503.04238](#) [math.AP]

**Abstract:** We propose a general approach for quantitative convergence analysis of non-reversible Markov processes, based on the concept of second-order lifts and a variational approach to hypocoercivity. To this end, we introduce the flow Poincaré inequality, a space-time Poincaré inequality along trajectories of the semigroup, and a general divergence lemma based only on the Dirichlet form of an underlying reversible diffusion. We demonstrate the versatility of our approach by applying it to a pair of run-and-tumble particles with jamming, a model from non-equilibrium statistical mechanics, and several piecewise deterministic Markov processes used in sampling applications. Our framework in particular includes processes with general stochastic jump kernels.

**Keywords:** Lift; non-reversible Markov process; run-and-tumble particles; piecewise deterministic Markov process.

### C.1. Introduction

For reversible Markov processes, the inverse Poincaré constant or spectral gap is a powerful tool for deriving quantitative convergence bounds, as it coincides with the rate of convergence and the asymptotic decorrelation time. For non-reversible processes, however, the inverse Poincaré constant can be much lower than the convergence rate, and the process may not even satisfy a Poincaré inequality but still converge rapidly. In fact, acceleration of convergence stemming from non-reversibility has been shown in different models [66, 67, 79, 98, 122, 142, 143, 175] and observed in many more [34, 141, 197, 204, 207]. The design of fast non-reversible processes often relies on the lifting procedure, which introduces an enlarged position-velocity state space to incorporate persistence, as introduced in [79] for discrete chains. In the continuous-time and -space setting, this approach has led to the use of piecewise deterministic Markov processes (PDMPs) [75].

These processes combine deterministic motion with discrete velocity jumps at event times, along with velocity refreshments typically occurring after exponentially distributed waiting times [34, 48, 197]. Despite their impressive numerical performance, the analytical understanding of these processes remains limited, posing challenges for rigorous theoretical analysis. Therefore, obtaining quantitative bounds on the rate of convergence of non-reversible processes has attracted a lot of attention, while being much more involved than in the reversible case. For degenerate diffusions, this active area of research has become known as hypocoercivity [87, 133, 248]. Recently, a variational approach to hypocoercivity pioneered by Albritton, Armstrong, Mourrat and Novack [6] was applied by Cao, Lu and Wang [61] for Langevin dynamics, for the first time obtaining rates of convergence of the correct order. This approach has further been applied to discretisations of Langevin dynamics [168] and to show that some PDMPs exhibit sharp dimensional dependence [186].

Inspired by the notion of lifts of discrete Markov chains [66, 79], second-order lifts of reversible diffusions [99] were introduced as their counterpart in continuous time and space. This enables the analysis of the convergence rate of non-reversible Markov processes by yielding an upper bound in terms of the spectral gap of an associated reversible collapsed process, generalising the related result for discrete Markov chains [66] by showing that the relaxation time can at most be improved by a square root through lifting. Furthermore, the variational approach to hypocoercivity [6, 61, 186] can be simplified by phrasing it in the language of second-order lifts, allowing for a systematic derivation of lower bounds on the convergence rates. The second-order lift property enables the exploitation of properties of the simpler collapsed process to obtain bounds on the convergence rate of the lifted process.

The principle of energy dissipation, i.e. differentiating the  $L^2$ -norm along the semigroup and controlling this dissipation by the original  $L^2$ -norm, is often used to get an exponential decay in  $L^2$ . In the elliptic setting, where  $(P_t)_{t \geq 0}$  is the transition semigroup of a reversible Markov process with invariant probability measure  $\mu$  and generator  $\mathcal{L}$ , this can be achieved through the Poincaré inequality

$$\|f\|_{L^2(\mu)}^2 \leq \frac{1}{\nu} \langle f, -\mathcal{L}f \rangle_{L^2(\mu)}$$

for all  $f \in L^2(\mu)$  with mean zero, as it implies that  $\|P_t f\|_{L^2(\mu)} \leq e^{-\nu t} \|f\|_{L^2(\mu)}$  for all  $t \geq 0$ . However, such a Poincaré inequality is no longer valid in the hypoelliptic, non-reversible setting. Instead, to obtain exponential decay of the transition semigroup of a possibly non-reversible Markov process, we introduce the *flow Poincaré inequality*

$$\int_0^T \|P_t f\|_{L^2(\mu)}^2 dt \leq \frac{1}{\nu} \int_0^T \langle P_t f, -\mathcal{L}P_t f \rangle_{L^2(\mu)} dt$$

for all  $f \in L^2(\mu)$  with mean zero, where  $T > 0$  is a fixed time length. Such a flow Poincaré inequality amounts to a time-integrated Poincaré inequality along trajectories of the semigroup, and is equivalent to dissipation of time-averaged  $L^2$ -norms, as considered in [6, 61, 99]. In particular, it implies the existence of a constant  $C > 0$  such that

$$\|P_t f\|_{L^2(\mu)} \leq C e^{-\nu t} \|f\|_{L^2(\mu)} \quad \text{for all } t \geq 0$$

and all mean-zero square-integrable observables  $f$ . In [87, 248], a modified  $L^2$ -norm based on an additional mixing term is introduced to ensure coercivity and hence control of the dissipation. The strength of an approach based on the flow Poincaré inequality lies in the more intrinsic nature of the time-averaged  $L^2$ -norm as opposed to adding an artificial mixing term to the norm, allowing for sharper rates. By exploiting the second-order lift property implicitly considered in [6, 61, 186], we prove such a flow Poincaré inequality in Theorem C.12. The proof is based on a divergence lemma that only relies on the Dirichlet form of the underlying collapsed process, see Lemma C.16. Previously, such divergence lemmas were only considered for the overdamped Langevin diffusion. Our result is more general and enables the treatment of non-reversible processes whose collapse is not an overdamped Langevin process. This is the case when incorporating non-trivial boundary conditions, as is often the case in non-equilibrium statistical mechanics. Finally, our framework allows for a simple treatment of PDMPs with more general refreshment and jump mechanisms.

We demonstrate the versatility of our approach by applying it to two first examples. Firstly, we consider a pair of jamming run-and-tumble particles (RTPs), a model of active particles which maintain persistent movement, lying at the interface of out-of-equilibrium statistical mechanics and biology. Their speed of convergence to equilibrium is physically relevant, as it is linked to the time scale of the onset of out-of-equilibrium phenomena [103, 246], such as motility-induced phase separation [63]. The prevailing method in the statistical physics literature to address this question is the spectral approach [13, 190, 191], leading to convergence speeds which are only valid asymptotically due to non-reversibility. Hence the value of the non-asymptotic total variation mixing time results [121, 125], extended to the  $L^2$  relaxation time framework in the present article. Secondly, we consider piecewise deterministic Markov processes (PDMPs) used in sampling applications [34, 196]. Due to the degeneracy introduced through their definition on an enlarged position-velocity state space, the analysis of their convergence is delicate [36, 37, 91, 186].

The paper is divided into four more sections. In Section C.2, we begin by summarising and illustrating the concept of second-order lifts using overdamped Langevin diffusions as a simple example. We then focus on the more intricate case of a pair of jamming RTPs, demonstrating how this system can be understood as a lift of sticky Brownian motion. Then Section C.3 presents the main results on convergence to equilibrium of lifts via the flow Poincaré inequality. Section C.4 applies our framework to a pair of RTPs with jamming. Finally, Section C.5 presents applications to PDMPs for sampling, namely the Forward process and variants of the Zig-Zag process.

## C.2. Second-order lifts

We begin by summarising the definition of second-order lifts and the related upper bound on the convergence rate and then illustrate the concept through lifts of overdamped Langevin diffusions. These include Langevin dynamics, randomised Hamiltonian Monte Carlo and many PDMPs introduced in sampling applications. We subsequently intro-

duce the system formed by a pair of jamming RTPs on the torus and show how it can be understood as a lift of a sticky Brownian motion. Such a system presents unique challenges within the second-order lift framework. Specifically, the presence of boundaries at which the process spends a non-zero amount of time complicates the identification of the underlying collapsed process. Additionally, the RTP velocity dynamics differ from previously considered mechanisms, requiring distinct proof techniques. The convergence of such systems will be explored in greater detail in Section C.4.

### C.2.1. Definition and examples

Consider a time-homogenous Markov process  $(\hat{Z}_t)$  with invariant probability measure  $\hat{\mu}$  on a Polish space  $\hat{\mathcal{S}}$ . Let  $\pi: \hat{\mathcal{S}} \rightarrow \mathcal{S}$  be a measurable surjection onto another Polish space, and let  $(Z_t)$  be a reversible diffusion process with state space  $\mathcal{S}$  whose invariant probability measure is the pushforward  $\mu = \hat{\mu} \circ \pi^{-1}$  of  $\hat{\mu}$  under  $\pi$ . We denote their associated transition semigroups on  $L^2(\hat{\mu})$  and  $L^2(\mu)$  by  $(\hat{P}_t)$  and  $(P_t)$ , and the associated generators by  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  and  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$ , respectively. The Dirichlet form associated to  $(Z_t)$  is the extension of

$$\mathcal{E}(f, g) = - \int_{\mathcal{S}} f \mathcal{L} g \, d\mu$$

to a closed symmetric bilinear form with domain  $\text{Dom}(\mathcal{E})$  given by the closure of  $\text{Dom}(\mathcal{L})$  with respect to the norm  $\|f\|_{L^2(\mu)} + \mathcal{E}(f, f)^{1/2}$ .

**Definition C.1** (Second-order lifts [99]). The process  $(\hat{Z}_t)$  is a *second-order lift* of  $(Z_t)$  if there exists a core  $C$  of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  such that

$$f \circ \pi \in \text{Dom}(\hat{\mathcal{L}}) \text{ for all } f \in C \quad (\text{C.1})$$

and for all  $f, g \in C$  we have

$$\langle \hat{\mathcal{L}}(f \circ \pi), g \circ \pi \rangle_{L^2(\hat{\mu})} = 0, \quad (\text{C.2})$$

and

$$\frac{1}{2} \langle \hat{\mathcal{L}}(f \circ \pi), \hat{\mathcal{L}}(g \circ \pi) \rangle_{L^2(\hat{\mu})} = - \langle f, \mathcal{L} g \rangle_{L^2(\mu)} = \mathcal{E}(f, g). \quad (\text{C.3})$$

If in addition

$$\{f \in L^2(\mu) : f \circ \pi \in \text{Dom}(\hat{\mathcal{L}})\} = \text{Dom}(\mathcal{E}),$$

then  $(\hat{Z}_t)$  is called a *strong second-order lift* of  $(Z_t)$ . Conversely, we refer to the process  $(Z_t)$  as the *collapse* of  $(\hat{Z}_t)$ .

We also say that  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  or  $(\hat{P}_t)$  is a second-order lift of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  or  $(P_t)$ , respectively. In the rest of this paper, we simply refer to second-order lifts as lifts. Usually,  $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{V}$  will be a product space,  $\pi(x, v) = x$  is the projection on  $\mathcal{S}$ , and we think of  $(x, v) \in \hat{\mathcal{S}}$  as a position and a velocity. More generally,  $\mathcal{S}$  could for example be a Riemannian manifold and  $\hat{\mathcal{S}}$  the tangent bundle, see [101].

To get a better understanding of the definition of lifts, disintegrate the measure  $\hat{\mu}$  as

$$\hat{\mu}(A) = \int_{\mathcal{S}} \kappa_x(A) \mu(dx) \quad \text{for all measurable } A \subseteq \hat{\mathcal{S}},$$

i.e.  $\hat{\mu}(dx dv) = \mu(dx) \kappa_x(dv)$  in case  $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{V}$ . Then, under minor technical assumptions, see [99, Remark 2 (v)], the first- and second-order conditions (C.2) and (C.3) are equivalent to the limit

$$\int_{\mathcal{V}} \hat{P}_t(g \circ \pi)(x, v) \kappa_x(dv) = (P_{t^2}g)(x) + o(t^2)$$

in  $L^2(\mu)$  as  $t$  goes to 0. Hence, for small values of  $t$ , the transition semigroup  $(P_{t^2})$  of the collapse behaves approximately as an averaged version of the transition semigroup  $(\hat{P}_t)$ . Compared to lifts of Markov chains as introduced in [66, 79], the interplay between the quadratic and linear timescales can be directly obtained in the continuous-time case.

By the spectral theorem, the rate of convergence to equilibrium in  $L^2(\mu)$  of the reversible process  $(Z_t)$  is accurately measured by the inverse of the spectral gap

$$\text{Gap}(\mathcal{L}) = \inf \{ \text{Re}(\alpha) : \alpha \in \text{spec}(-\mathcal{L}|_{L_0^2(\mu)}) \}$$

of its generator  $\mathcal{L}$ , where  $L_0^2(\mu) = \{f \in L^2(\mu) : \int f d\mu = 0\}$  denotes the subspace of mean-zero functions. This is no longer true for non-reversible Markov processes since their generators are not self-adjoint. Even in the exponentially ergodic case, the associated transition semigroup only satisfies

$$\|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq C e^{-\nu t} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } f \in L_0^2(\hat{\mu}) \quad (\text{C.4})$$

with  $\nu \in [0, \infty)$  and some constant  $C \in [1, \infty)$ . However, the rate of convergence to equilibrium of a lift can be bounded above by the square root of that of the collapsed process in the following way.

**Theorem C.2** ([99, Theorem 11]). *Suppose that  $(\hat{P}_t)$  is a second-order lift of  $(P_t)$ . Assume that there exist constants  $C \in [1, \infty)$  and  $\nu \in [0, \infty)$  such that (C.4) is satisfied. Then*

$$\nu \leq (1 + \log C) \sqrt{2 \text{Gap}(\mathcal{L})}. \quad (\text{C.5})$$

Theorem C.2 provides a bound on the rate of convergence achievable through lifting in terms of the spectral gap of the collapsed process, generalising a related result [66, Theorem 3.1] in discrete time and space. This motivates the search for optimal lifts, in the sense that they satisfy (C.4) with  $\nu$  of order  $\sqrt{\text{Gap}(\mathcal{L})}$  when  $\text{Gap}(\mathcal{L})$  is small. Proving such a statement requires quantitative lower bounds on the rate of convergence of non-reversible Markov processes. We show in Section C.3 how to obtain such bounds using the framework of lifts under general assumptions on the collapsed process and the structure of the lift. The result is demonstrated in several applications in Sections C.4 and C.5.

**Example C.3** (Lifts of overdamped Langevin diffusions). Prototypical examples of second-order lifts are given by lifts of overdamped Langevin diffusions. Let  $\mathcal{S} = \mathbb{R}^d$  and consider a potential  $U \in C^1(\mathbb{R}^d)$  such that  $U(x) \rightarrow +\infty$  when  $|x| \rightarrow +\infty$  and  $\int e^{-U(x)} dx < +\infty$ .

The *overdamped Langevin diffusion* is the process  $(Z_t)$  with state space  $\mathcal{S}$  given by the stochastic differential equation

$$dZ_t = -\frac{1}{2}\nabla U(Z_t) dt + dB_t,$$

where  $(B_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion. Its unique invariant probability measure is the Boltzmann-Gibbs measure

$$\mu(dx) \propto e^{-U(x)} dx,$$

and the associated generator in  $L^2(\mu)$  is

$$\mathcal{L}f = -\frac{1}{2}\nabla U \cdot \nabla f + \frac{1}{2}\Delta f = -\frac{1}{2}\nabla^* \nabla f$$

for all  $f \in C_c^\infty(\mathbb{R}^d)$ , which is a core of the generator [251, Theorem 3.1]. Here  $\nabla^*$  is the adjoint of  $\nabla$  in  $L^2(\mu)$ .

Lifts of overdamped Langevin diffusions are now given by Markov processes on  $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{V} = \mathbb{R}^d \times \mathbb{R}^d$ , see [99, Example 3]. They include processes based on Hamiltonian dynamics, such as the Langevin diffusion or randomised Hamiltonian Monte Carlo (RHMC), but also other piecewise deterministic Markov processes (PDMPs) used in sampling applications, that we will treat in Section C.5. These processes have in common that their invariant probability measure is a product measure  $\hat{\mu} = \mu \otimes \kappa$ , where  $\kappa$  is for example the  $d$ -dimensional standard normal distribution, the uniform distribution on the hypercube  $\{-1, +1\}^d$ , or the uniform distribution on the coordinate directions  $\{\pm e_i : i = 1, \dots, n\}$ . For functions  $f \in \text{Dom}(\mathcal{L})$ , their generators take the form  $\hat{\mathcal{L}}(f \circ \pi)(x, v) = v \cdot \nabla f(x)$ , so that (C.2) and (C.3) can immediately be verified.

### C.2.2. Jamming run-and-tumble particles as a lift of sticky Brownian motion

Run-and-tumble particles (RTPs), a close approximation for the movement of bacteria such as *E. coli* [28, 230], alternate between long straight runs and rapid reorientations. In addition to their use in biological modelling, they are a topic of investigation in statistical mechanics as they break time reversibility at the particle level and achieve persistent motion. This can drive them out of thermodynamic equilibrium and lead to interesting phenomena not found in passive systems, such as motility-induced phase separation [63] where the particles form clusters. Even a single RTP displays rich long-term behaviour not observed in passive systems, such as accumulation at boundaries [14, 78, 190]. From a mathematical perspective, systems of RTPs can be framed as piecewise deterministic Markov processes [121, 126], for which many fundamental questions, such as the regularity of invariant measures, remain open.

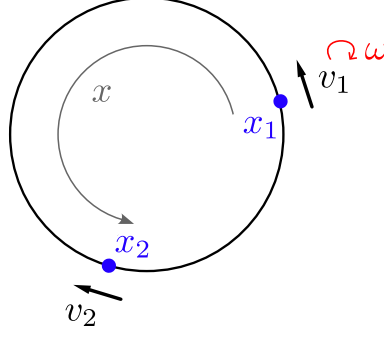
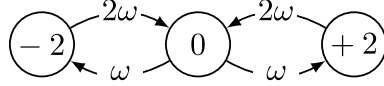


Figure C.1.: Two run-and-tumble particles on the 1D torus

Here, we focus on the two-particle case and start with an informal description of the model. Consider two point particles on the 1D torus of length  $L$  described by their positions  $x_1, x_2 \in \mathbb{R}/L\mathbb{Z}$  and their velocities  $v_1, v_2 \in \{\pm 1\}$  (see Figure C.1). The velocities are independent Markov jump processes switching between  $+1$  and  $-1$  with rate  $\omega > 0$ , corresponding to the *instantaneous* tumble mechanism. Between two velocity switches, each particle  $x_i$  moves around the 1D torus with constant velocity  $v_i$  until the two particles collide. Colliding particles jam and stop moving until a velocity flip allows them to separate.

A discrete version of this process was introduced in [232] where jammed configurations were found to have increased mass under the invariant measure. This is a form of clustering hinting at motility-induced phase separation. A continuous-space version was later considered in [126] for velocities following an arbitrary Markov jump process, revealing two universality classes for the invariant measure.

Figure C.2.: Jump rates of the relative velocity  $v = v_2 - v_1$ 

When modelling the two-particle system as a Markov process, it suffices to keep track of the separation from the first to the second particle  $x \in \mathcal{S} = [0, L]$  (see Figure C.1) and the relative velocity  $v = v_2 - v_1 \in \mathcal{V} = \{-2, 0, 2\}$  (see Figure C.2). Figure C.3 shows a realisation of the process.

**Definition C.4** (Jamming RTP process [121]). Let  $(x_0, v_0) \in [0, L] \times \{-2, 0, +2\}$  and  $\omega \in (0, \infty)$  be given. Let the relative velocity  $v(t)$  be a Markov jump process with initial state  $v_0$  and the jump rates of Figure C.2. Denote  $0 = T_0 < T_1 < \dots$  its jump times, i.e.  $v(t) = v(T_n)$  for  $t \in [T_n, T_{n+1})$ . Recursively define the particle separation  $x(t)$  by  $x(0) = x_0$  and

$$x(t) = \max[0, \min[L, x(T_n) + v(T_n)(t - T_n)]] \quad \text{for } t \in [T_n, T_{n+1}).$$

We call  $X(t) = (x(t), v(t))$  the jamming RTP process on  $[0, L]$  with parameter  $\omega$ .

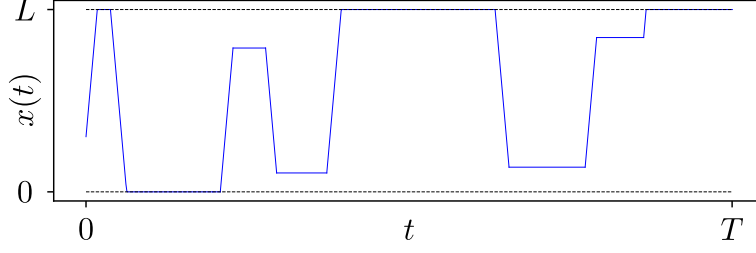


Figure C.3.: Realisation of the jamming RTP process

Henceforth, for conciseness, and when no confusion arises, we will occasionally refer to the jamming RTP process simply as the RTP process. The following theorem gives the invariant measure and the generator of the jamming RTP process. Its proof is relegated to Section C.4.2.

**Theorem C.5** (Invariant measure and generator of the jamming RTP process). *Let  $\omega, L \in (0, \infty)$ .*

- (i) ([126]) *The jamming RTP process on  $[0, L]$  with parameter  $\omega$  takes its values in  $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{V} = [0, L] \times \{-2, 0, +2\}$  and its unique invariant probability is given by*

$$\hat{\mu} = \sum_{v \in \mathcal{V}} (d_v^0 \delta_0 + a_v dx + d_v^L \delta_L) \otimes \delta_v, \quad \text{where}$$

$$\begin{aligned} d_2^0 &= 0, & a_2 &= \frac{\omega/4}{2 + \omega L}, & d_2^L &= \frac{1/2}{2 + \omega L}, \\ d_0^0 &= \frac{1/2}{2 + \omega L}, & a_0 &= \frac{\omega/2}{2 + \omega L}, & d_0^L &= \frac{1/2}{2 + \omega L}, \\ d_{-2}^0 &= \frac{1/2}{2 + \omega L}, & a_{-2} &= \frac{\omega/4}{2 + \omega L}, & d_{-2}^L &= 0. \end{aligned}$$

- (ii) *The domain  $\text{Dom}(\hat{\mathcal{L}})$  of its  $L^2(\hat{\mu})$ -generator  $\hat{\mathcal{L}}$  is the set of functions  $f \in L^2(\hat{\mu})$  s.t.*

$$f(x, \pm 2) \in H^1(0, L), \quad \lim_{x \rightarrow L} f(x, 2) = f(L, 2), \quad \lim_{x \rightarrow 0} f(x, -2) = f(0, -2).$$

For  $f \in \text{Dom}(\hat{\mathcal{L}})$  one has

$$\hat{\mathcal{L}}f(x, v) = v 1_{\{0 < x < L\}} \partial_x f(x, v) + \sum_{\tilde{v} \in \mathcal{V}} \lambda_{v\tilde{v}} f(x, \tilde{v}), \quad (\text{C.6})$$

where

$$\begin{pmatrix} \lambda_{+2,+2} & \lambda_{+2,0} & \lambda_{+2,-2} \\ \lambda_{0,+2} & \lambda_{0,0} & \lambda_{0,-2} \\ \lambda_{-2,+2} & \lambda_{-2,0} & \lambda_{-2,-2} \end{pmatrix} = \begin{pmatrix} -2\omega & 2\omega & 0 \\ \omega & -2\omega & \omega \\ 0 & 2\omega & -2\omega \end{pmatrix}$$

is the discrete generator of the relative velocity (see Figure C.2).

Regarding the second-order lift framework, this process differs from previously studied PDMPs and is interesting for two reasons. Firstly, the state space has boundaries and the process spends a non-zero amount of time on them. This is particularly interesting as the behaviour at boundaries usually manifests itself not in the formula for the generator but rather in its domain. Because second-order lifts are defined in terms of generators, even identifying the associated collapsed process becomes delicate. Secondly, different velocity processes necessitate different proofs, as was previously done when the velocity is an Ornstein–Uhlenbeck process [61, 101] and when the velocity is completely resampled at Poisson times [99, 186]. The jamming RTP process’s velocity, however, is a Markov jump process with the rates of Figure C.2, requiring a new approach.

In [126], this particular jamming RTP process is shown to belong to a class of near-equilibrium systems, whose steady states differ from those of equilibrium diffusive particles only by the presence of positive mass on the boundaries. This suggests that the jamming RTP process can be understood as the lift of a sticky diffusion, whose invariant measure is the image measure

$$\mu = \hat{\mu} \circ \pi^{-1} = \frac{1}{2 + \omega L} \delta_0 + \frac{\omega}{2 + \omega L} dx + \frac{1}{2 + \omega L} \delta_L$$

under the projection  $\pi(x, v) = x$ . An integration by parts argument further supports this interpretation by pointing to a diffusion, as

$$\int \hat{\mathcal{L}}(f \circ \pi) \hat{\mathcal{L}}(g \circ \pi) d\hat{\mu} = -\frac{2\omega}{2 + \omega L} \int_0^L f'' g dx$$

for all  $f, g \in C^\infty([0, L])$  with  $\text{supp}(f), \text{supp}(g) \subset (0, L)$ . These heuristic arguments are made rigorous in Theorem C.7.

**Definition C.6** (Sticky Brownian motion). For  $\omega \in (0, \infty)$ , sticky Brownian motion on  $[0, L]$  with parameter  $\omega$  is the  $[0, L]$ -valued Feller process with generator

$$\mathcal{L}_{C^0} f = f'', \quad \text{Dom}(\mathcal{L}_{C^0}) = \{f \in C^2([0, L]): f''(0) = \omega f'(0) \text{ and } f''(L) = -\omega f'(L)\}.$$

Conventionally, sticky Brownian motion is defined with  $\mathcal{L}_{C^0} f = \frac{1}{2} f''$ , see [182, Section 3.5.2]. The process defined in Definition C.6 would be a sticky Brownian motion accelerated by a factor 2. However, leaving out the factor  $\frac{1}{2}$  is more convenient for our purpose. In this setting, the tumble rate  $\omega$  quantifies the *stickiness* at the boundaries. The limits  $\omega \rightarrow 0$  and  $\omega \rightarrow +\infty$  describe the asymptotic *persistent* and *diffusive* dynamical regimes [121, 126] and correspond here to absorption and reflection at the boundaries, respectively.

**Theorem C.7** (Jamming RTP is a lift of sticky Brownian motion). *Let  $\omega, L \in (0, \infty)$ .*

- (i) *Sticky Brownian motion on  $[0, L]$  with parameter  $\omega$  has a unique invariant probability measure given by*

$$\mu = \frac{1}{2 + \omega L} \delta_0 + \frac{\omega}{2 + \omega L} dx + \frac{1}{2 + \omega L} \delta_L.$$

## Appendix C. Convergence of Markov processes via lifting and flow Poincaré inequality

- (ii) Its  $C^0$ -generator  $(\mathcal{L}_{C^0}, \text{Dom}(\mathcal{L}_{C^0}))$  is a core of its  $L^2(\mu)$ -generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$ .
- (iii) The jamming RTP process on  $[0, L]$  with parameter  $\omega$  is a lift of sticky Brownian motion on  $[0, L]$  with parameter  $\omega$ .

*Proof.* (i) Follows from the generator characterisation of invariance [182, Theorem 3.37].

(ii) The set  $\text{Dom}(\mathcal{L}_{C^0})$  is dense in  $L^2(\mu)$  and invariant by the  $L^2$ -semigroup. Hence the claim follows from [104, Proposition II.1.7].

(iii) For  $f \in \text{Dom}(\mathcal{L}_{C^0})$  one has  $f \circ \pi \in \text{Dom}(\hat{\mathcal{L}})$  and

$$\hat{\mathcal{L}}(f \circ \pi)(x, v) = v 1_{\{0 < x < L\}} f'(x).$$

Hence, for  $f, g \in \text{Dom}(\mathcal{L}_{C^0})$ , integrating first with respect to  $v$  and then with respect to  $x$  yields

$$\langle \hat{\mathcal{L}}(f \circ \pi), g \circ \pi \rangle_{L^2(\hat{\mu})} = \frac{\omega}{2 + \omega L} \int_0^L g(x) \left( \frac{1}{4}(2f'(x)) + \frac{1}{2}(0f'(x)) + \frac{1}{4}(-2f'(x)) \right) dx = 0.$$

Again integrating first with respect to  $v$  and then  $x$  yields

$$\begin{aligned} \frac{1}{2} \langle \hat{\mathcal{L}}(f \circ \pi), \hat{\mathcal{L}}(g \circ \pi) \rangle_{L^2(\hat{\mu})} &= \frac{\omega}{2 + \omega L} \int_0^L f'(x) g'(x) dx \\ &= \frac{\omega}{2 + \omega L} \left( f'(L)g(L) - f'(0)g(0) - \int_0^L f''g dx \right) \\ &= \frac{\omega}{2 + \omega L} \left( -\frac{1}{\omega} f''(L)g(L) - \frac{1}{\omega} f''(0)g(0) - \int_0^L f''g dx \right) \\ &= \int_{[0, L]} (-\mathcal{L}f)g d\mu. \quad \square \end{aligned}$$

### C.3. Convergence to equilibrium

This section presents the variational approach to hypocoercivity using a flow Poincaré inequality and second-order lifts, extending the results in [6, 61, 99, 186] and unifying a broad class of processes, including several examples from the sampling and statistical mechanics literature. The key idea is to characterise the collapsed process solely through its Dirichlet form.

#### C.3.1. Convergence via flow Poincaré inequality

For the transition semigroup  $(P_t)_{t \geq 0}$  of a potentially non-reversible Markov process with invariant probability measure  $\mu$  and generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  in  $L^2(\mu)$ , as explained in Section C.2, the spectral gap or Poincaré inequality does not suffice in general to characterise the rate of convergence to equilibrium. In particular, the process may not satisfy a Poincaré inequality at all yet still converge exponentially fast. However, the following flow Poincaré inequality, amounting to a Poincaré inequality time-integrated over a fixed time length  $T \in (0, \infty)$  along trajectories of the semigroup, is equivalent to exponential decay of the  $L^2(\mu)$ -norm averaged over time  $T$  along trajectories of the semigroup.

**Definition C.8** (Flow Poincaré inequality). The semigroup  $(P_t)_{t \geq 0}$  satisfies a *flow Poincaré inequality* with constant  $\nu^{-1} \in (0, \infty)$  and duration  $T \in (0, \infty)$  if

$$\int_0^T \|P_t f\|_{L^2(\mu)}^2 dt \leq \frac{1}{\nu} \int_0^T \langle P_t f, -\mathcal{L} P_t f \rangle_{L^2(\mu)} dt \quad (\text{C.7})$$

for all  $f \in L_0^2(\mu) \cap \text{Dom}(\mathcal{L})$ .

In fact, defining the inner product  $\langle f, g \rangle_T = \int_0^T \langle P_t f, P_t g \rangle_{L^2(\mu)} dt$  on  $L^2(\mu)$ , the flow Poincaré inequality is equivalent to

$$\langle f, -\mathcal{L} f \rangle_T \geq \nu \langle f, f \rangle_T \quad \text{for all } f \in L_0^2(\mu) \cap \text{Dom}(\mathcal{L}).$$

Similarly to the reversible case, in which the Poincaré inequality is equivalent to the decay of the  $L^2$ -norm, the flow Poincaré inequality is equivalent to decay of time-averaged  $L^2$ -norms.

**Theorem C.9** (Exponential decay). *The following assertions are equivalent:*

- (i) *The semigroup  $(P_t)_{t \geq 0}$  satisfies a flow Poincaré inequality with constant  $\nu^{-1}$  and duration  $T$ .*
- (ii) *For all  $f \in L_0^2(\mu)$  and  $t \geq 0$  one has*

$$\int_t^{t+T} \|P_s f\|_{L^2(\mu)}^2 ds \leq e^{-2\nu t} \int_0^T \|P_s f\|_{L^2(\mu)}^2 ds. \quad (\text{C.8})$$

Furthermore, both (i) and (ii) imply

$$\|P_t f\|_{L^2(\mu)} \leq C e^{-\nu t} \|f\|_{L^2(\mu)} \quad \text{with } C = e^{\nu T}$$

for all  $f \in L_0^2(\mu)$  and  $t \geq 0$ .

The proof of the decay estimate in Theorem C.9 using variants of a Poincaré-type inequality in time and space is well-established, see e.g. [6, 61, 99], whereas the converse relies on our formulation of the flow Poincaré inequality.

*Proof of Theorem C.9.* Assume that a flow Poincaré inequality with constant  $\nu^{-1}$  and duration  $T$  holds and let  $f \in L_0^2(\mu) \cap \text{Dom}(\mathcal{L})$ . Then

$$\begin{aligned} \frac{d}{dt} \int_t^{t+T} \|P_s f\|_{L^2(\mu)}^2 ds &= \|P_{t+T} f\|_{L^2(\mu)}^2 - \|P_t f\|_{L^2(\mu)}^2 \\ &= \int_t^{t+T} \frac{d}{ds} \|P_s f\|_{L^2(\mu)}^2 ds = 2 \int_t^{t+T} \langle P_s f, \mathcal{L} P_s f \rangle_{L^2(\mu)} ds. \end{aligned}$$

By the semigroup property, the flow Poincaré inequality implies

$$\int_0^T \langle P_{t+s} f, \mathcal{L} P_{t+s} f \rangle_{L^2(\mu)} ds \leq -\nu \int_0^T \|P_{t+s} f\|_{L^2(\mu)}^2 ds$$

so that

$$\frac{d}{dt} \int_t^{t+T} \|P_s f\|_{L^2(\mu)}^2 ds \leq -2\nu \int_t^{t+T} \|P_s f\|_{L^2(\mu)}^2 ds.$$

Grönwall's inequality yields

$$\int_t^{t+T} \|P_s f\|_{L^2(\mu)}^2 ds \leq e^{-2\nu t} \int_0^T \|P_s f\|_{L^2(\mu)}^2 ds.$$

The claim then follows since  $\text{Dom}(\mathcal{L})$  is dense in  $L^2(\mu)$ .

Conversely, assume that (C.8) holds and let  $f \in L_0^2(\mu) \cap \text{Dom}(\mathcal{L})$ . Then

$$\begin{aligned} \int_0^T \langle P_t f, -\mathcal{L} P_t f \rangle_{L^2(\mu)} dt &= \lim_{h \rightarrow 0} \frac{1}{h} \int_0^T \langle P_t f, P_t f - P_{t+h} f \rangle_{L^2(\mu)} dt \\ &\geq \lim_{h \rightarrow 0} \frac{1}{h} (1 - e^{-\nu h}) \int_0^T \|P_t f\|_{L^2(\mu)}^2 dt = \nu \int_0^T \|P_t f\|_{L^2(\mu)}^2 dt, \end{aligned}$$

which yields the claim.

Finally, if (C.8) holds, then for  $t \geq T$ ,

$$\|P_t f\|_{L^2(\mu)}^2 \leq \frac{1}{T} \int_{t-T}^t \|P_s f\|_{L^2(\mu)}^2 ds \leq e^{-2\nu(t-T)} \frac{1}{T} \int_0^T \|P_s f\|_{L^2(\mu)}^2 ds \leq e^{-2\nu(t-T)} \|f\|_{L^2(\mu)}^2,$$

which yields the claim. For  $t < T$ , the claim is trivially satisfied since  $e^{-\nu(t-T)} \geq 1$ .  $\square$

**Remark C.10.** A Poincaré inequality in a suitable extended space-time sense was first introduced in [6] to obtain exponential decay for the kinetic Fokker-Planck equation, and later applied to study the convergence rate of Langevin dynamics [61] and some piecewise deterministic Markov processes [186]. The flow Poincaré inequality in Definition C.8 differs from the space-time Poincaré inequality as in [6, 61, 186], as it is only valid along trajectories  $t \mapsto P_t f$  of the semigroup. However, this is not a restriction, since it is specifically intended for application to solutions of the corresponding Kolmogorov backward equation, as in the proof of Theorem C.9. More importantly, this formulation allows for its extension to a broad class of refreshment mechanism, which we will make use of in Section C.4, and avoids the use of the  $H^{-1}$ -norm in the velocity that is present in the formulation of a space-time Poincaré inequality as in [6, 61].

### C.3.2. Flow Poincaré inequality for second-order lifts

We now demonstrate how to prove flow Poincaré inequalities for non-reversible Markov processes using the framework of second-order lifts. As in Section C.2, we consider a time-homogenous Markov process with state space  $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{V}$ , where  $\mathcal{S} \subseteq \mathbb{R}^d$ , and invariant probability measure  $\hat{\mu}$  and a reversible diffusion on  $\mathcal{S}$  with invariant measure  $\mu = \hat{\mu} \circ \pi^{-1}$ , where  $\pi(x, v) = x$  is the projection from  $\hat{\mathcal{S}}$  to  $\mathcal{S}$ . The associated transition semigroups acting on  $L^2(\hat{\mu})$  and  $L^2(\mu)$  and their generators are denoted by  $(\hat{P}_t)$  and  $(P_t)$ , as well

as  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  and  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$ , respectively. Disintegrating  $\hat{\mu}(\mathrm{d}x \mathrm{d}v) = \mu(\mathrm{d}x)\kappa_x(\mathrm{d}v)$  yields the conditional expectation

$$\Pi_v f(x, v) = \int_{\mathcal{V}} f(x, w) \kappa_x(\mathrm{d}w).$$

We split the generator of  $(\hat{P}_t)$  into

$$\hat{\mathcal{L}}f = \hat{\mathcal{L}}_{\text{tr}}f + \gamma \hat{\mathcal{L}}_v f \quad \text{for all } f \in \text{Dom}(\hat{\mathcal{L}}), \quad (\text{C.9})$$

where  $\gamma > 0$ , and  $(\hat{\mathcal{L}}_{\text{tr}}, \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}))$  and  $(\hat{\mathcal{L}}_v, \text{Dom}(\hat{\mathcal{L}}_v))$  are two linear operators on  $L^2(\hat{\mu})$  such that  $\text{Dom}(\hat{\mathcal{L}}) \subseteq \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}) \cap \text{Dom}(\hat{\mathcal{L}}_v)$ . The operators  $\hat{\mathcal{L}}_{\text{tr}}$  and  $\hat{\mathcal{L}}_v$  are usually interpreted as transport and refreshment terms, respectively. We denote by

$$\mathcal{E}_v(f, g) = \langle f, -\hat{\mathcal{L}}_v g \rangle_{L^2(\hat{\mu})}$$

the velocity Dirichlet form and set  $\mathcal{E}_v(f) = \mathcal{E}_v(f, f)$ .

**Assumption (A).** The operator  $(\hat{\mathcal{L}}_{\text{tr}}, \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}))$  is a lift of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  that is weakly antisymmetric, i.e. there exists a core  $C$  of  $\mathcal{L}$  such that

$$f \circ \pi \in \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}^*) \quad \text{and} \quad \hat{\mathcal{L}}_{\text{tr}}^*(f \circ \pi) = -\hat{\mathcal{L}}_{\text{tr}}(f \circ \pi)$$

for all  $f \in C$ .

**Assumption (B).** The operator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  has purely discrete spectrum on  $L^2(\mu)$  and satisfies a Poincaré inequality with constant  $\frac{1}{m}$ , i.e.

$$\|f - \mu(f)\|_{L^2(\mu)}^2 \leq \frac{1}{m} \mathcal{E}(f) \quad (\text{C.10})$$

for all  $f \in \text{Dom}(\mathcal{E})$ .

**Assumption (C).** (i) There exist constants  $C_1, C_2 > 0$  such that

$$|\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\hat{\mu})}| \leq C_1 \|2\mathcal{L}g\|_{L^2(\mu)} (\mathcal{E}_v(f))^{1/2}, \quad (\text{C1})$$

and

$$|\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_v f \rangle_{L^2(\hat{\mu})}| \leq C_2 \|2\mathcal{L}g\|_{L^2(\mu)} (\mathcal{E}_v(f))^{1/2}, \quad (\text{C2})$$

for all functions  $f \in \text{Dom}(\hat{\mathcal{L}})$  with  $\Pi_v f \in \text{Dom}(\hat{\mathcal{L}}_{\text{tr}})$  and  $g$  in a core of  $\mathcal{L}$ .

(ii) The operator  $\hat{\mathcal{L}}_v$  leaves  $\kappa_x$  invariant, i.e.

$$\int_{\mathcal{V}} \hat{\mathcal{L}}_v f(x, v) \mathrm{d}\kappa_x(v) = 0 \quad \text{for all } x \in \mathcal{S} \text{ and } f \in \text{Dom}(\hat{\mathcal{L}}),$$

and there exists a constant  $m_v > 0$  such that

$$\int_0^T \|\hat{P}_t f - \Pi_v \hat{P}_t f\|_{L^2(\hat{\mu})}^2 \mathrm{d}t \leq \frac{1}{m_v} \int_0^T \mathcal{E}_v(\hat{P}_t f) \mathrm{d}t \quad (\text{C.11})$$

for all  $f \in \text{Dom}(\hat{\mathcal{L}})$ .

Let us comment on these assumptions before turning to the main result in Theorem C.12.

**Remark C.11.** (i) A splitting as in (C.9) into two distinct dynamics, interpreted as a transport and velocity refreshment, is natural in many examples. For instance, in the case of Langevin dynamics or randomised Hamiltonian Monte Carlo, one chooses the transport operator

$$\hat{\mathcal{L}}_{\text{tr}} f(x, v) = v \cdot \nabla_x f(x, v) - \nabla U(x) \cdot \nabla_v f(x, v)$$

to be the generator of the Hamiltonian flow associated to the Hamiltonian  $H(x, v) = U(x) + \frac{1}{2}|v|^2$ . The velocity refreshment operator  $\hat{\mathcal{L}}_v$  could be the full refreshment  $\hat{\mathcal{L}}_v = \Pi_v - I$  with respect to the conditional distribution  $\kappa_x$  but much more general mechanisms are possible, as discussed in Section C.4. Regarding the jamming RTP process, the transport and velocity refreshments terms correspond to the run-and-tumble mechanisms, respectively, with the latter differing from a complete refreshment.

- (ii) In Assumption (A), if  $\hat{\mathcal{L}}_v$  only acts on  $v$  in the sense that there exists a core  $C$  of  $\mathcal{L}$  such that  $\hat{\mathcal{L}}_v(f \circ \pi) = 0$  for all  $f \in C$ , then  $\hat{\mathcal{L}}$  is a lift of  $\mathcal{L}$  if and only if  $\hat{\mathcal{L}}_{\text{tr}}$  is a lift of  $\mathcal{L}$ .
- (iii) The Assumption (B) of a purely discrete spectrum, which was also considered in [61, 99], allows for simple calculations using a spectral decomposition that enables us to prove a divergence lemma not limited to the generator of an overdamped Langevin diffusion. This technical assumption can be relaxed, see [55]. The Poincaré inequality (C.10) is equivalent to existence of a spectral gap  $m = \text{Gap}(\mathcal{L})$  and to exponential decay of the semigroup  $(P_t)$  generated by  $\mathcal{L}$  with rate  $m$ . It thus allows us to directly compare the decay of  $(\hat{P}_t)$  to that of  $(P_t)$ . In the absence of a spectral gap, Assumption (A) can however also be relaxed to a weak Poincaré inequality, yielding subexponential convergence to equilibrium, see [56].
- (iv) Assumption (C1) can often be verified in the following way in applications. If  $\hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) \in \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}^*)$  for all  $g$  in a core of  $\mathcal{L}$ , the second order lift condition becomes

$$\begin{aligned} \Pi_v \hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) &= -(2\mathcal{L}g) \circ \pi, \quad \text{and thus} \\ \left| \langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\hat{\mu})} \right| &\leq \|(I - \Pi_v) \hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)\|_{L^2(\hat{\mu})} \|f - \Pi_v f\|_{L^2(\hat{\mu})} \\ &\leq m_v^{-1/2} \|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + (2\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})} (\mathcal{E}_v(f))^{1/2} \end{aligned}$$

Hence (C1) is satisfied if

$$\|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + (2\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})} \leq C_1 m_v^{1/2} \|2\mathcal{L}g\|_{L^2(\mu)},$$

i.e. if the  $L^2(\mu)$ -norm of the standard deviation of  $\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)$  with respect to  $\kappa_x$  is bounded by the  $L^2(\mu)$ -norm of its expectation  $-2\mathcal{L}g$ .

- (v) Assumption (C2) can be verified immediately if  $\hat{\mathcal{L}}_v$  is a bounded operator, since then there exists a constant  $c > 0$  such that  $\|\hat{\mathcal{L}}_v f\|_{L^2(\hat{\mu})} \leq c(\mathcal{E}_v(f))^{1/2}$ . By the Cauchy-Schwarz inequality and the second-order lift condition (C.3),

$$|\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_v f \rangle_{L^2(\hat{\mu})}| \leq c(2\mathcal{E}(g))^{1/2}(\mathcal{E}_v(f))^{1/2} \leq \frac{c}{\sqrt{2m}} \|2\mathcal{L}g\|_{L^2(\mu)} (\mathcal{E}_v(f))^{1/2},$$

so that (C2) holds with  $C_2 = \frac{c}{\sqrt{2m}}$ . In particular, if  $\hat{\mathcal{L}}_v = \Pi_v - I$  is the full refreshment with respect to the conditional distribution  $\kappa_x(dv)$ , Assumptions (C2) and (C.11) hold with  $C_2 = \frac{1}{\sqrt{2m}}$  and  $m_v = 1$ .

- (vi) In Assumption (C), (C.11) resembles a flow Poincaré inequality where the left-hand side is only centred in the velocity variable, and the right-hand side uses the velocity Dirichlet form. The inequality in particular holds if the refreshment term satisfies a Poincaré inequality in  $v$ , i.e. if

$$\|f - \Pi_v f\|_{L^2(\hat{\mu})}^2 \leq \frac{1}{m_v} \mathcal{E}_v(f)$$

for all  $f \in \text{Dom}(\hat{\mathcal{L}})$ . This is usually the case, as, in line with the general principles of hypocoercivity, the dynamics in  $v$  has strong relaxation properties that are then propagated to the variable  $x$  by the rest of the dynamics. Nonetheless, the weaker time-integrated version (C.11) may be of use in applications where the refreshment term is degenerate.

- (vii) Assumptions (A), (B) and (C)(ii) are closely related to Assumptions (H1), (H2), and (H3) in the seminal work [87], while Assumption (C)(i) ensuring good interaction between  $\hat{\mathcal{L}}_{\text{tr}}$  and  $\hat{\mathcal{L}}_v$  is reminiscent of Assumption (H4). The connection between second-order lifts and the approach to hypocoercivity in [87] is detailed in [54].

We can now establish the following flow Poincaré inequality.

**Theorem C.12** (Flow Poincaré inequality for lifts). *Let Assumptions (A), (B) and (C) hold. Then there exists a universal constant  $C > 0$  such that for any  $T \in (0, \infty)$  and  $f \in \text{Dom}(\hat{\mathcal{L}}) \cap L_0^2(\hat{\mu})$ ,*

$$\int_0^T \|\hat{P}_t f\|_{L^2(\hat{\mu})}^2 dt \leq C \left( \gamma^2 C_2^2 + C_1^2 + \frac{1}{m_v} \left( 1 + \frac{1}{mT^2} \right) \right) \int_0^T \mathcal{E}_v(\hat{P}_t f) dt.$$

In particular,  $(\hat{P}_t)_{t \geq 0}$  satisfies a flow Poincaré inequality with duration  $T$  and constant

$$\frac{1}{\nu} = C\gamma^{-1} \left( \gamma^2 C_2^2 + C_1^2 + \frac{1}{m_v} \left( 1 + \frac{1}{mT^2} \right) \right). \quad (\text{C.12})$$

We defer the proof to Section C.3.4. The constant  $C$  in Theorem C.12 does not depend on any parameters and can easily be made explicit. By Theorem C.9, Theorem C.12 immediately yields the following corresponding decay estimate.

**Corollary C.13.** *Let Assumptions (A), (B) and (C) hold. Then the semigroup  $(\hat{P}_t)$  generated by  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  satisfies*

$$\|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq C e^{-\nu t} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } t \geq 0$$

and  $f \in L_0^2(\hat{\mu})$ , where the rate  $\nu$  is given by (C.12) and  $C = e^{\nu T}$ .

One of the main tools in the proof of a flow Poincaré inequality using the framework of second-order lifts is a so-called divergence lemma, see Section C.3.3. There are now several different works [6, 55, 61, 101] which aim at providing such a divergence lemma under general hypotheses. The Dirichlet form approach taken here enables general assumptions on the collapsed process, imposing only the Assumption (B) of a spectral gap and discrete spectrum.

**Remark C.14.** To obtain an explicit rate in Corollary C.13 from the expression (C.12), one can choose  $T = m^{-1/2}$ . If  $m_v^{-1} = O(1)$  and  $C_2 = O(m^{-1/2})$ , as is the case if  $\hat{\mathcal{L}}_v = \Pi_v - I$  is the full refreshment with respect to  $\kappa_x$  by Remark C.11(v), this yields

$$\nu = \Omega\left(\frac{\gamma m}{\gamma^2 + (1 + C_1^2) m}\right).$$

Optimising by choosing  $\gamma = (1 + C_1)\sqrt{m}$  yields  $\nu = \Omega(\sqrt{m}/(1 + C_1))$ . For many examples, including Langevin dynamics, randomised HMC, and the RTP process considered here, one can choose  $C_1$  of order 1, see Section C.4. The expression then further simplifies to

$$\nu = \Omega\left(\frac{\gamma m}{\gamma^2 + m}\right),$$

and if  $\gamma = \Theta(\sqrt{m})$  then  $\nu = \Omega(\sqrt{m})$ , i.e. we obtain a square-root speed-up with the optimal choice of  $\gamma$ .

The usual example of Langevin dynamics (e.g. [61, 99]) serves as a useful base case in showcasing the arguments involved in applying Theorem C.12.

**Example C.15** (Langevin dynamics). Consider a probability measure  $\mu$  on  $\mathbb{R}^d$  with density proportional to  $\exp(-U)$ , where  $U: \mathbb{R}^d \rightarrow \mathbb{R}$  is such that  $\int_{\mathbb{R}^d} e^{-U(x)} dx < \infty$  and  $\lim_{|x| \rightarrow +\infty} U(x) = +\infty$ . Let us assume that  $\mu$  satisfies a Poincaré inequality with constant  $\frac{1}{m}$ , and that the generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  of the associated overdamped Langevin diffusion, see Example C.3, has discrete spectrum. Furthermore, let the potential  $U$  satisfy the lower curvature bound

$$\nabla^2 U(x) \geq -K m \cdot I_d \quad \text{for all } x \in \mathbb{R}^d \tag{C.13}$$

for some  $K \geq 0$ . Langevin dynamics is the  $\mathbb{R}^d \times \mathbb{R}^d$ -valued diffusion process  $(X_t, V_t)$  solving the stochastic differential equation

$$dX_t = V_t dt,$$

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

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion and  $\gamma \geq 0$  is the friction parameter, see e.g. [215]. Its invariant measure is  $\hat{\mu} = \mu \otimes \kappa$  with the standard normal distribution  $\kappa = \mathcal{N}(0, I_d)$ , and its generator in  $L^2(\hat{\mu})$  is given by  $\hat{\mathcal{L}}f = \hat{\mathcal{L}}_{\text{tr}}f + \gamma \hat{\mathcal{L}}_v f$ , where

$$\begin{aligned}\hat{\mathcal{L}}_{\text{tr}}f(x, v) &= v \cdot \nabla_x f(x, v) - \nabla U(x) \cdot \nabla_v f(x, v), \\ \hat{\mathcal{L}}_v f(x, v) &= -v \cdot \nabla_v f(x, v) + \Delta_v f(x, v),\end{aligned}$$

are the generators of the Hamiltonian flow associated to the Hamiltonian  $H(x, v) = U(x) + \frac{1}{2}|v|^2$ , and of an Ornstein-Uhlenbeck process in the velocity, respectively. Let us check Assumptions (A)–(C). The generator  $(\hat{\mathcal{L}}_{\text{tr}}, \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}))$  of the Hamiltonian flow is antisymmetric in  $L^2(\hat{\mu})$  and is a lift of an overdamped Langevin diffusion with generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  by [99, Example 3(i)]. This can be checked immediately since  $\hat{\mathcal{L}}_{\text{tr}}(f \circ \pi)(x, v) = v \cdot \nabla f(x)$  for all  $f \in C_c^\infty(\mathbb{R}^d)$ , so that Assumption (A) is satisfied. Assumption (B) is satisfied by our assumptions on the potential. It remains to check Assumption (C). For (C1), since the Ornstein-Uhlenbeck process with generator  $\hat{\mathcal{L}}_v$  leaves  $\kappa$  invariant and has spectral gap  $m_v = 1$ , see for example [20], by Remark C.11(iv), it is enough to prove

$$\|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + (2\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})} \leq C_1 \|2\mathcal{L}g\|_{L^2(\mu)}.$$

We have

$$\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)(x, v) = -v^\top \nabla^2 g(x) v - \nabla U(x) \cdot \nabla g(x)$$

and  $2\mathcal{L}g = \Delta g + \nabla U \cdot \nabla g$ , so that

$$\begin{aligned}\int_{\mathbb{R}^d} \left( \hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)(\cdot, v) + 2\mathcal{L}g \right)^2 d\kappa(v) &= \int_{\mathbb{R}^d} (v^\top \nabla^2 g v - \Delta g)^2 d\kappa(v), \\ &= \int_{\mathbb{R}^d} (v^\top \nabla^2 g v)^2 d\kappa(v) - (\Delta g)^2,\end{aligned}$$

using the fact that  $\int v^\top \nabla^2 g v d\kappa(v) = \Delta g$ . Furthermore, since  $v_i, v_j, v_k, v_l$  are independent standard Gaussians,

$$\begin{aligned}\int_{\mathbb{R}^d} (v^\top \nabla^2 g v)^2 d\kappa(v) &= \sum_{i,j,k,l=1}^d (\partial_{ij} g)(\partial_{kl} g) \int_{\mathbb{R}^d} v_i v_j v_k v_l d\kappa(v) \\ &= (\Delta g)^2 + 2\|\nabla^2 g\|_F^2,\end{aligned}$$

where  $\|\nabla^2 g\|_F^2 = \sum_{i,j=1}^d (\partial_{ij} g)^2$  is the squared Frobenius norm of  $\nabla^2 g$ . By Bochner's formula, using the lower curvature bound (C.13), we have

$$\|\|\nabla^2 g\|_F\|_{L^2(\mu)}^2 = \|2\mathcal{L}g\|_{L^2(\mu)}^2 - \int_{\mathbb{R}^d} \nabla g^\top \nabla^2 U \nabla g d\mu \leq \|2\mathcal{L}g\|_{L^2(\mu)}^2 + 2Km \mathcal{E}(g),$$

Since  $\mathcal{E}(g) \leq \frac{1}{m} \|\mathcal{L}g\|_{L^2(\mu)}^2$ , overall we obtain

$$\|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + (2\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})}^2 = 2\|\|\nabla^2 g\|_F\|_{L^2(\mu)}^2 \leq (2+K)\|2\mathcal{L}g\|_{L^2(\mu)}^2$$

and (C1) holds with  $C_1 = \sqrt{2+K}$ . Finally, Assumption (C2) is satisfied with  $C_2 = \frac{1}{\sqrt{2m}}$  since

$$\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_v f \rangle_{L^2(\hat{\mu})} = \langle v \cdot \nabla_x g, -\nabla_v^* \nabla_v f \rangle_{L^2(\hat{\mu})} = \langle \nabla_x g, -\nabla_v f \rangle_{L^2(\hat{\mu})}.$$

Hence, by Corollary C.13, when choosing  $T = m^{-1/2}$ , the transition semigroup  $(\hat{P}_t)$  associated to Langevin dynamics is exponentially contractive in  $T$ -average with rate

$$\nu = \Omega\left(\frac{\gamma m}{\gamma^2 + (1+K)m}\right).$$

Choosing  $\gamma = \sqrt{(1+K)m}$  yields the rate  $\nu = \Omega(\sqrt{m}/\sqrt{1+K})$ . We thus recover the result of [61, Theorem 1].

### C.3.3. Divergence lemma and space-time property of lifts

In the classical elliptical setting, a Poincaré inequality can be established by solving the Poisson equation  $-\mathcal{L}g = f$  with explicit bounds on  $g$ . Similarly, in the hypoelliptic setting, a generalised time-dependent Poisson equation may be used to obtain a flow Poincaré inequality. The divergence Lemma C.16 provides solutions to such an equation with explicit bounds on the solution and is at the heart of the proof of the flow Poincaré inequality Theorem C.12. Different versions of this divergence lemma have been studied recently by [55, 61, 101, 168], and a related divergence equation with different boundary conditions is of fundamental importance in fluid mechanics [110]. We take an approach using Dirichlet forms, adapting the proof of [101] and enabling minimal structural assumptions on  $\mathcal{L}$ .

Let us assume that  $\mathcal{S} \subseteq \mathbb{R}^d$  and that  $C^\infty(\mathcal{S}) \cap L^2(\mu) \subseteq \text{Dom}(\mathcal{L})$ . For  $g \in L^2(\lambda \otimes \mu)$  such that  $g(t, \cdot) \in \text{Dom}(\mathcal{E})$  for  $\lambda$ -almost every  $t \in [0, T]$ , denote by

$$\mathcal{E}_T(g) = \int_0^T \mathcal{E}(g(t, \cdot)) dt$$

the time-integrated Dirichlet form. Let the Sobolev space  $H_D^{1,2}(\lambda \otimes \mu)$  with Dirichlet boundary conditions in time be the closure of

$$\{u \in C^\infty([0, T] \times \mathcal{S}) \cap L^2(\lambda \otimes \mu) : u(0, \cdot) = u(T, \cdot) = 0\}$$

with respect to the norm given by

$$\|u\|_{1,2} = \|u\|_{L^2(\lambda \otimes \mu)} + \mathcal{E}_T(u)^{1/2} + \|\partial_t u\|_{L^2(\lambda \otimes \mu)},$$

and similarly let  $H_D^{2,2}(\lambda \otimes \mu)$  be the closure of the same set with respect to

$$\|u\|_{2,2} = \|u\|_{L^2(\lambda \otimes \mu)} + \mathcal{E}_T(u)^{1/2} + \|\partial_t u\|_{L^2(\lambda \otimes \mu)} + \|\mathcal{L}u\|_{L^2(\lambda \otimes \mu)} + \|\partial_{tt} u\|_{L^2(\lambda \otimes \mu)} + \mathcal{E}_T(\partial_t u)^{1/2}.$$

In particular, if  $u \in H_D^{2,2}(\lambda \otimes \mu)$ , then  $u(t, \cdot) \in \text{Dom}(\mathcal{L})$  for  $\lambda$ -a.e.  $t \in [0, T]$  since  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  is closed. The assumption that  $\mathcal{S} \subseteq \mathbb{R}^d$  can be relaxed to  $\mathcal{S}$  being a Riemannian manifold with boundary, see [101].

**Lemma C.16** (Divergence lemma). *Suppose that Assumption (B) is satisfied. Then for any  $T \in (0, \infty)$ , there exists a constant  $C \in (0, \infty)$  such that for any  $f \in L_0^2(\lambda \otimes \mu)$  there are functions*

$$h \in H_D^{1,2}(\lambda \otimes \mu) \quad \text{and} \quad g \in H_D^{2,2}(\lambda \otimes \mu)$$

such that

$$\partial_t h - 2\mathcal{L}g = f,$$

and the functions  $h$  and  $g$  satisfy

$$\begin{aligned} \|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} &\leq C\|f\|_{L^2(\lambda \otimes \mu)}, \\ \sqrt{2\mathcal{E}_T(h)} &\leq C\|f\|_{L^2(\lambda \otimes \mu)}, \\ \sqrt{2\mathcal{E}_T(\partial_t g)} &\leq C(1 + T^{-1}m^{-1/2})\|f\|_{L^2(\lambda \otimes \mu)}. \end{aligned}$$

The proof of Lemma C.16 is delayed until Section C.3.5. The final ingredient for the proof of the flow Poincaré inequality is the following space-time property of lifts. We consider the operator  $(A, \text{Dom}(A))$  on  $L^2([0, T] \times \hat{\mathcal{S}}, \lambda \otimes \hat{\mu})$  defined by

$$Af = -\partial_t f + \hat{\mathcal{L}}_{\text{tr}} f \tag{C.14}$$

with domain consisting of all functions  $f \in L^2(\lambda \otimes \hat{\mu})$  such that  $f(\cdot, x, v)$  is absolutely continuous for  $\hat{\mu}$ -a.e.  $(x, v) \in \hat{\mathcal{S}}$  with  $\partial_t f \in L^2(\lambda \otimes \hat{\mu})$ , and  $f(t, \cdot) \in \text{Dom}(\hat{\mathcal{L}}_{\text{tr}})$  for  $\lambda$ -a.e.  $t \in [0, T]$  with  $\hat{\mathcal{L}}_{\text{tr}} f \in L^2(\lambda \otimes \hat{\mu})$ .

**Lemma C.17** (Space-time property of lifts). *Let  $(\hat{\mathcal{L}}_{\text{tr}}, \text{Dom}(\hat{\mathcal{L}}_{\text{tr}}))$  be a lift of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$ . Then, for any  $f, g, h \in L^2(\lambda \otimes \mu)$  such that  $f \circ \pi \in \text{Dom}(A)$  and  $g(t, \cdot) \in \text{Dom}(\mathcal{L})$  for a.e.  $t \in [0, T]$  with  $\mathcal{L}g \in L^2(\lambda \otimes \mu)$ , we have*

$$\langle A(f \circ \pi), h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) \rangle_{L^2(\lambda \otimes \hat{\mu})} = -\langle h, \partial_t f \rangle_{L^2(\lambda \otimes \mu)} + 2\mathcal{E}_T(f, g).$$

*Proof.* This follows directly from the defining properties (C.2) and (C.3) of second-order lifts, see [99, Lemma 18].  $\square$

### C.3.4. Proof of the flow Poincaré inequality for second-order lifts

We now prove Theorem C.12. The proof is similar to that of [99, Theorem 23]. However, bounding  $\int_0^T \|\hat{P}_t f\|_{L^2(\hat{\mu})}^2 dt$  directly by the time-integrated velocity Dirichlet form  $\int_0^T \mathcal{E}_v(\hat{P}_t f) dt$  in this general setting requires more refined estimates.

*Proof of Theorem C.12.* Let  $f_0 \in L_0^2(\hat{\mu})$  and let  $f \in L^2(\lambda \otimes \hat{\mu})$  be defined by

$$f(t, x, v) = (\hat{P}_t f_0)(x, v).$$

### Appendix C. Convergence of Markov processes via lifting and flow Poincaré inequality

Further let  $\tilde{f} \in L^2(\lambda \otimes \mu)$  be such that  $\Pi_v f = \tilde{f} \circ \pi$ . Let  $\tilde{f} = \partial_t h - 2\mathcal{L}g$  with  $h$  and  $g$  as in Lemma C.16. Because  $\Pi_v$  is an orthogonal projection, we have

$$\|f\|_{L^2(\lambda \otimes \hat{\mu})}^2 = \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2 + \|f - \Pi_v f\|_{L^2(\lambda \otimes \hat{\mu})}^2.$$

Furthermore, by an integration by parts and the space-time property of lifts Lemma C.17,

$$\begin{aligned} \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2 &= \langle \partial_t h - 2\mathcal{L}g, \tilde{f} \rangle_{L^2(\lambda \otimes \mu)} = -\langle h, \partial_t \tilde{f} \rangle_{L^2(\lambda \otimes \mu)} + 2\mathcal{E}_T(\tilde{f}, g) \\ &= \langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), A(\tilde{f} \circ \pi) \rangle_{L^2(\lambda \otimes \hat{\mu})} \\ &= \langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), Af \rangle_{L^2(\lambda \otimes \hat{\mu})} + \langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), A(\Pi_v f - f) \rangle_{L^2(\lambda \otimes \hat{\mu})} \end{aligned} \quad (\text{C.15})$$

The two summands on the right-hand side of (C.15) can be bounded separately.

Firstly, we have  $Af = -\partial_t f + \hat{\mathcal{L}}_{\text{tr}} f = -\gamma \hat{\mathcal{L}}_v f$ , so that  $\langle h \circ \pi, Af \rangle_{L^2(\lambda \otimes \hat{\mu})} = 0$ , and thus

$$\begin{aligned} \langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), Af \rangle_{L^2(\lambda \otimes \hat{\mu})} &= -\gamma \langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_v f \rangle_{L^2(\lambda \otimes \hat{\mu})} \\ &\leq \gamma C_2 \int_0^T \|2\mathcal{L}g\|_{L^2(\mu)} (\mathcal{E}_v(f))^{1/2} dt \leq \gamma C \cdot C_2 \int_0^T \|\tilde{f}\|_{L^2(\mu)} (\mathcal{E}_v(f))^{1/2} dt \\ &\leq \gamma C \cdot C_2 \|\tilde{f}\|_{L^2(\lambda \otimes \mu)} \left( \int_0^T \mathcal{E}_v(f) dt \right)^{1/2}, \end{aligned}$$

using Assumption (C2) for the first inequality and Lemma C.16 for the second.

The second summand in (C.15) can be estimated using (C.14). The Dirichlet boundary conditions in time of  $h$  and  $g$  give

$$\begin{aligned} \langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), -\partial_t(\Pi_v f - f) \rangle_{L^2(\lambda \otimes \hat{\mu})} &= \langle \partial_t h \circ \pi + \partial_t \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \Pi_v f - f \rangle_{L^2(\lambda \otimes \hat{\mu})} \\ &= \langle \partial_t \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \Pi_v f - f \rangle_{L^2(\lambda \otimes \hat{\mu})} \leq \|\partial_t \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)\|_{L^2(\lambda \otimes \mu)} \|f - \Pi_v f\|_{L^2(\lambda \otimes \hat{\mu})}. \end{aligned}$$

Furthermore, using the fact that  $\partial_t$  commutes with  $\hat{\mathcal{L}}_{\text{tr}}(\cdot \circ \pi)$  and the second-order lift property, we get

$$\|\partial_t \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)\|_{L^2(\lambda \otimes \hat{\mu})} = (2\mathcal{E}_T(\partial_t g))^{1/2} \leq C \left( 1 + \frac{1}{\sqrt{mT}} \right) \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}.$$

We deduce that

$$\begin{aligned} &\langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), -\partial_t(\Pi_v f - f) \rangle_{L^2(\lambda \otimes \hat{\mu})} \\ &\leq \frac{C}{\sqrt{m_v}} \left( 1 + \frac{1}{\sqrt{mT}} \right) \left( \int_0^T \mathcal{E}_v(f) dt \right)^{1/2} \|\tilde{f}\|_{L^2(\lambda \otimes \mu)}^2. \end{aligned}$$

Finally,

$$\begin{aligned} &\langle h \circ \pi + \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\lambda \otimes \hat{\mu})} \\ &= -\langle \hat{\mathcal{L}}_{\text{tr}}(h \circ \pi), f - \Pi_v f \rangle_{L^2(\lambda \otimes \hat{\mu})} + \langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\lambda \otimes \hat{\mu})} \end{aligned}$$

$$\begin{aligned}
&\leq \|\hat{\mathcal{L}}_{\text{tr}}(h \circ \pi)\|_{L^2(\lambda \otimes \hat{\mu})} \|f - \Pi_v f\|_{L^2(\lambda \otimes \hat{\mu})} + C_1 \int_0^T \|2\mathcal{L}g\|_{L^2(\mu)} (\mathcal{E}_v(f))^{1/2} dt \\
&\leq \left( \frac{1}{\sqrt{m_v}} (2\mathcal{E}_T(h))^{1/2} + C_1 \|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} \right) \left( \int_0^T \mathcal{E}_v(f) dt \right)^{1/2} \\
&\leq C \left( \frac{1}{\sqrt{m_v}} + C_1 \right) \|\tilde{f}\|_{L^2(\lambda \otimes \mu)} \left( \int_0^T \mathcal{E}_v(f) dt \right)^{1/2}
\end{aligned}$$

using the weak antisymmetry of  $\hat{\mathcal{L}}_{\text{tr}}$ , Assumption (C1) and Lemma C.16.

Combining everything finally leads to

$$\begin{aligned}
\|f\|_{L^2(\lambda \otimes \hat{\mu})}^2 &\leq \|\Pi_v f - f\|_{L^2(\lambda \otimes \hat{\mu})}^2 \\
&\quad + C^2 \left( \gamma C_2 + \frac{1}{\sqrt{m_v}} \left( 1 + \frac{1}{\sqrt{mT}} \right) + \left( \frac{1}{\sqrt{m_v}} + C_1 \right) \right)^2 \int_0^T \mathcal{E}_v(f) dt \\
&\leq \tilde{C} \left( \gamma^2 C_2^2 + C_1^2 + \frac{1}{m_v} \left( 1 + \frac{1}{mT^2} \right) \right) \int_0^T \mathcal{E}_v(f) dt,
\end{aligned}$$

where  $\tilde{C}$  denotes an absolute constant, as claimed. □

### C.3.5. Proof of the divergence lemma

We now prove Lemma C.16. While some differences arise due to the fact that we estimate slightly different quantities, the proof closely follows that of [101, Theorem 5], which is a careful refinement of [61, Lemma 2.6]. For the convenience of the reader, we provide an overview of the proof, omitting the details where they are identical to [101].

*Proof of Lemma C.16.* Consider an orthonormal basis  $\{e_k : k \geq 0\}$  of eigenfunctions of  $\mathcal{L}$  with eigenvalues  $-\alpha_k^2$ , where  $e_0$  is constant and  $\alpha_0 = 0$ . In particular,  $\inf\{\alpha_k^2 : k > 0\} = m$  is the spectral gap of  $\mathcal{L}$ . Then  $\{H_k^a : k \geq 0\} \cup \{H_k^s : k > 0\}$  given by

$$\begin{aligned}
H_0^a(t, x) &= (t - (T - t))e_0(x), \\
H_k^a(t, x) &= (e^{-\alpha_k t} - e^{-\alpha_k(T-t)})e_k(x) \text{ for } k > 0, \\
H_k^s(t, x) &= (e^{-\alpha_k t} + e^{-\alpha_k(T-t)})e_k(x) \text{ for } k > 0,
\end{aligned}$$

defines an orthogonal basis of the space

$$\mathbb{H} = \left\{ u \in L_0^2(\lambda \otimes \mu) : u(\cdot, x) \in H^2((0, T)) \text{ for } \mu\text{-a.e. } x, u(t, \cdot) \in \text{Dom}(\mathcal{L}) \text{ for } \lambda\text{-a.e. } t, \text{ and } \partial_{tt}u + 2\mathcal{L}u = 0 \right\}$$

of mean-zero harmonic functions for  $\partial_{tt} + 2\mathcal{L}$ . We consider a simultaneous decomposition of  $\mathbb{H}$  into high and low modes and symmetric and antisymmetric functions, i.e.

$$L_0^2(\lambda \otimes \mu) = \mathbb{H} \oplus \mathbb{H}^\perp = \mathbb{H}_{l,a} \oplus \mathbb{H}_{l,s} \oplus \mathbb{H}_{h,a} \oplus \mathbb{H}_{h,s} \oplus \mathbb{H}^\perp,$$

where

$$\begin{aligned}\mathbb{H}_{l,a} &= \text{span} \left\{ H_k^a : \alpha_k \leq \frac{2}{T} \right\}, & \mathbb{H}_{l,s} &= \text{span} \left\{ H_k^s : \alpha_k \leq \frac{2}{T} \right\}, \\ \mathbb{H}_{h,a} &= \overline{\text{span}} \left\{ H_k^a : \alpha_k > \frac{2}{T} \right\}, & \mathbb{H}_{h,s} &= \overline{\text{span}} \left\{ H_k^s : \alpha_k > \frac{2}{T} \right\}.\end{aligned}$$

Due to linearity, it suffices to prove the statement separately for right-hand sides  $f$  in these subspaces and in  $\mathbb{H}^\perp$ .

*Case 1:* Let  $f \in \mathbb{H}^\perp$ . Consider the operator  $(\bar{\mathcal{L}}, \text{Dom}(\bar{\mathcal{L}}))$  on  $L^2(\lambda \otimes \mu)$  given by  $\bar{\mathcal{L}}f = \partial_{tt}f + 2\mathcal{L}f$  with Neumann boundary conditions in time, i.e.

$$\begin{aligned}\text{Dom}(\bar{\mathcal{L}}) &= \left\{ f \in L^2(\lambda \otimes \mu) : f(t, \cdot) \in \text{Dom}(\mathcal{L}) \text{ for } \lambda\text{-a.e. } t \in (0, T), \right. \\ &\quad \left. f(\cdot, x) \in H^{2,2}((0, T)) \text{ for } \mu\text{-a.e. } x \in \mathcal{S} \text{ and } \partial_t f = 0 \text{ on } \{0, T\} \times \mathcal{S}, \right. \\ &\quad \left. \partial_{tt}f + \mathcal{L}f \in L^2(\lambda \otimes \mu) \right\}.\end{aligned}$$

By tensorisation,  $\bar{\mathcal{L}}$  satisfies a Poincaré inequality with constant  $\min(2m, \frac{\pi^2}{T})$ . Hence  $\bar{\mathcal{L}}$  has a spectral gap, so that

$$-\bar{\mathcal{L}}|_{L_0^2(\lambda \otimes \mu)} : \text{Dom}(\bar{\mathcal{L}}) \cap L_0^2(\lambda \otimes \mu) \rightarrow L_0^2(\lambda \otimes \mu)$$

admits a bounded linear inverse

$$\bar{G} : L_0^2(\lambda \otimes \mu) \rightarrow L_0^2(\lambda \otimes \mu) \cap \text{Dom}(\bar{\mathcal{L}})$$

whose operator norm is bounded above by  $\max(\frac{1}{2m}, \frac{T^2}{\pi^2})$ . We hence set  $g = \bar{G}f$  and  $h = -\partial_t g$ . The Dirichlet boundary conditions in time for  $h$  are satisfied by construction, and the Dirichlet boundary conditions for  $g$  follow from an integration by parts as in [101]. On  $L_0^2(\lambda \otimes \mu) \cap \text{Dom}(\bar{\mathcal{L}})$ , the operators  $-\partial_{tt}$  and  $-2\mathcal{L}$  commute and have discrete spectrum, so that the operators norms of

$$2\mathcal{L}\bar{G} = 2\mathcal{L}(-\partial_{tt} - 2\mathcal{L})^{-1} \quad \text{and} \quad \partial_{tt}\bar{G} = \partial_{tt}(-\partial_{tt} - 2\mathcal{L})^{-1}$$

are bounded by one. Since

$$2\mathcal{E}_T(h) = 2\mathcal{E}_T(\partial_t g) = \langle -\partial_{tt}\bar{G}f, -2\mathcal{L}\bar{G}f \rangle_{L^2(\lambda \otimes \mu)} \leq \|f\|_{L^2(\lambda \otimes \mu)}^2,$$

for right-hand sides  $f \in \mathbb{H}^\perp$ , we obtain the claim with the estimates

$$\|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} \leq \|f\|_{L^2(\lambda \otimes \mu)} \quad \text{and} \quad \sqrt{2\mathcal{E}_T(h)} = \sqrt{2\mathcal{E}_T(\partial_t g)} \leq \|f\|_{L^2(\lambda \otimes \mu)}.$$

*Case 2:* Let  $f \in \mathbb{H}_{l,a}$ . Here, as in [101], we set  $h(t, x) = \int_0^t f(s, x) ds$  and  $g(t, x) = 0$ . Then the boundary conditions are satisfied due to the antisymmetry of  $f$  and we obtain

$$\begin{aligned}\|h\|_{L^2(\lambda \otimes \mu)}^2 &\leq \frac{T^2}{2} \|f\|_{L^2(\lambda \otimes \mu)}^2, \\ \|\mathcal{L}h\|_{L^2(\lambda \otimes \mu)}^2 &\leq \int_0^T t \int_0^t \frac{16}{T^4} \|f(s, \cdot)\|_{L^2(\lambda \otimes \mu)}^2 ds dt \leq \frac{8}{T^2} \|f\|_{L^2(\lambda \otimes \mu)}^2,\end{aligned}$$

where we used that  $\|\mathcal{L}f(s, \cdot)\|_{L^2(\lambda \otimes \mu)} \leq \frac{4}{T^2} \|f(s, \cdot)\|_{L^2(\lambda \otimes \mu)}$  since  $f \in \mathbb{H}_{l,a}$ . By Cauchy-Schwarz, this yields  $\mathcal{E}_T(h) \leq 2\|f\|_{L^2(\lambda \otimes \mu)}$ , so that, for right-hand sides  $f \in \mathbb{H}_{l,a}$ , we obtain the statement with the estimates

$$\|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} = \sqrt{2\mathcal{E}_T(\partial_t g)} = 0 \quad \text{and} \quad \sqrt{2\mathcal{E}_T(h)} \leq 2\|f\|_{L^2(\lambda \otimes \mu)}.$$

*Case 3:* Let  $f \in \mathbb{H}_{l,s}$ . Since  $f$  is symmetric, it no longer necessarily integrates to 0 in time and we cannot argue as previously. We consider the decomposition  $f = f_0 + f_1$  with

$$f_0(t, x) = f(0, x) \cos\left(\frac{2\pi t}{T}\right) \quad \text{and} \quad f_1(t, x) = f(t, x) - f_0(t, x)$$

into a part  $f_0$  that integrates to 0 in time and a part  $f_1$  with Dirichlet boundary values. We then set  $h(t, x) = \int_0^t f_0(s, x) ds$  and  $g(x) = Gf_1(t, \cdot)(x)$ , where  $G = -(2\mathcal{L})^{-1}$  on  $\text{span}\{e_k : \alpha_k \leq \frac{2}{T}\}$ , i.e.  $Ge_k = \frac{1}{2\alpha_k^2} e_k$ , so that indeed  $f = \partial_t h - 2\mathcal{L}g$ . As in [101], one shows that

$$\|f_0\|_{L^2(\lambda \otimes \mu)}^2 \leq e^2 \|f\|_{L^2(\lambda \otimes \mu)}^2 \quad \text{and} \quad \|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)}^2 = \|f_1\|_{L^2(\lambda \otimes \mu)}^2 \leq (1+e)^2 \|f\|_{L^2(\lambda \otimes \mu)}^2,$$

and the same estimate as previously yields  $2\mathcal{E}_T(h) \leq 2\|f_0\|_{L^2(\lambda \otimes \mu)}^2 \leq 4e^2 \|f\|_{L^2(\lambda \otimes \mu)}^2$ . The term  $2\mathcal{E}_T(\partial_t g)$  can be bounded as in [101], giving

$$2\mathcal{E}_T(\partial_t g) \leq \frac{(2 + \sqrt{2\pi}e)^2}{2mT^2} \|\partial_t f_1\|_{L^2(\lambda \otimes \mu)}^2.$$

Hence, for right-hand sides  $f \in \mathbb{H}_{l,s}$ , the statement is satisfied with the estimates

$$\begin{aligned} \|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} &\leq (1+e)\|f\|_{L^2(\lambda \otimes \mu)}, \quad \sqrt{2\mathcal{E}_T(h)} \leq 2e\|f\|_{L^2(\lambda \otimes \mu)}, \\ \sqrt{2\mathcal{E}_T(\partial_t g)} &\leq \frac{10}{\sqrt{mT}}\|f\|_{L^2(\lambda \otimes \mu)}. \end{aligned}$$

*Case 4:* Let  $f \in \mathbb{H}_{h,a}$ . As in [101], we use the expansion  $f(t, x) = \sum_{\alpha_k > \frac{\beta}{T}} b_k H_k^a(t, x)$ , and it is sufficient to derive a representation

$$H_k^a = \partial_t h_k - 2\mathcal{L}g_k$$

for each of the basis functions  $H_k^a$  with  $k \geq 1$  fixed. We write  $u_k(t) = e^{-\alpha_k t} - e^{-\alpha_k(T-t)}$ , so that  $H_k^a(t, x) = u_k(t)e_k(x)$  and employ the ansatz

$$u_k = \dot{v}_k + w_k \quad \text{with} \quad v_k(0) = v_k(T) = w_k(0) = w_k(T) = 0. \quad (\text{C.16})$$

Then setting  $h_k(t, x) = v_k(t)e_k(x)$  and  $g_k(t, x) = \frac{1}{2\alpha_k^2} w_k(t)e_k(x)$  yields  $H_k^a = \partial_t h_k - 2\mathcal{L}g_k$  as desired. The construction of such  $v_k$  and  $w_k$  and the associated bounds follow as in [101], we finally obtain the claim for right-hand sides  $f \in \mathbb{H}_{h,a}$  with the estimates

$$\|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} \leq \left(1 + \frac{1}{\sqrt{3}}\right)\|f\|_{L^2(\lambda \otimes \mu)}, \quad \sqrt{2\mathcal{E}_T(h)} \leq \frac{1}{\sqrt{15}}\|f\|_{L^2(\lambda \otimes \mu)},$$

$$\sqrt{2\mathcal{E}_T(\partial_t g)} \leq 8\|f\|_{L^2(\lambda \otimes \mu)}.$$

*Case 5:* Let  $f \in \mathbb{H}_{h,s}$ . This can be treated as in the previous case, the bounds even improve slightly, see [101]. We obtain the claim for  $f \in \mathbb{H}_{h,s}$  with the estimates

$$\begin{aligned} \|2\mathcal{L}g\|_{L^2(\lambda \otimes \mu)} &\leq \left(1 + \frac{1}{\sqrt{3}}\right)\|f\|_{L^2(\lambda \otimes \mu)}, \quad \sqrt{2\mathcal{E}_T(h)} \leq \frac{1}{\sqrt{15}}\|f\|_{L^2(\lambda \otimes \mu)}, \\ \sqrt{2\mathcal{E}_T(\partial_t g)} &\leq (5 + \sqrt{2})\|f\|_{L^2(\lambda \otimes \mu)}. \end{aligned}$$

Finally, the statement follows with constants given by 5 times the maximum of the constants on the five subspaces.  $\square$

## C.4. Convergence of the jamming RTP process

The challenges posed by the RTP process are twofold, namely the non-trivial boundary behaviour with positive time spent on the boundary, and the intricate refreshment mechanism described by the velocity jumps. Addressing the first difficulty, we have already seen in Section C.2.2 that the RTP process on  $[0, L]$  with parameter  $\omega$  is a lift of a time-changed sticky Brownian motion on  $[0, L]$  with parameter  $\omega$ . We now begin by proving the main result of convergence to equilibrium in Section C.4.1 and continue in Section C.4.2 by completing the proof of Theorem C.5 which is necessary to characterise the  $L^2$ -generator of the RTP process.

### C.4.1. Convergence to equilibrium

We have to verify Assumptions (A)–(C) for the RTP process. To do so, the first step is splitting the generator

$$\hat{\mathcal{L}}f(x, v) = v1_{\{0 < x < L\}}\partial_x f(x, v) + \sum_{\tilde{v} \in \mathcal{V}} \lambda_{v\tilde{v}}f(x, \tilde{v}),$$

of the RTP process, see (C.6), into  $\hat{\mathcal{L}} = \hat{\mathcal{L}}_{\text{tr}} + \gamma\hat{\mathcal{L}}_v$ . By analogy with Langevin dynamics, it would be tempting to choose

$$\hat{\mathcal{L}}_{\text{tr}}f(x, v) = v1_{\{0 < x < L\}}\partial_x f(x, v) \quad \text{and} \quad \gamma\hat{\mathcal{L}}_vf(x, v) = \sum_{\tilde{v} \in \mathcal{V}} \lambda_{v\tilde{v}}f(x, \tilde{v})$$

to be the deterministic part of the dynamics and the random jumps, respectively. Unfortunately, that splitting does not take into account the interplay with the boundaries properly and  $\hat{\mu}$  is invariant for neither  $\hat{\mathcal{L}}_{\text{tr}}$  nor  $\hat{\mathcal{L}}_v$ . Indeed,  $\gamma\hat{\mathcal{L}}_v$  should leave

$$\kappa_x = \begin{cases} \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{-2} & \text{if } x = 0, \\ \frac{1}{4}\delta_2 + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_{-2} & \text{if } x \in (0, L), \\ \frac{1}{2}\delta_2 + \frac{1}{2}\delta_0 & \text{if } x = L, \end{cases}$$

invariant. Therefore,  $\gamma\hat{\mathcal{L}}_v$  is taken to be jumps in  $v$  with the rates of Figure C.2 when  $x \in (0, L)$ , the rates of Figure C.4a when  $x = 0$  and the rates of Figure C.4b when  $x = L$ .



Figure C.4.: Transition rates of  $v$  under  $\gamma\hat{\mathcal{L}}_v$ .

Choosing  $\gamma = \omega$  leads to

$$\begin{aligned}\hat{\mathcal{L}}_v f(x, +2) &= 1_{\{x < L\}} 2(f(x, 0) - f(x, +2)) + 1_{\{x = L\}} (f(x, 0) - f(x, +2)), \\ \hat{\mathcal{L}}_v f(x, 0) &= 1_{\{x > 0\}} (f(x, +2) - f(x, 0)) + 1_{\{x < L\}} (f(x, -2) - f(x, 0)), \\ \hat{\mathcal{L}}_v f(x, -2) &= 1_{\{x > 0\}} 2(f(x, 0) - f(x, -2)) + 1_{\{x = 0\}} (f(x, 0) - f(x, -2)),\end{aligned}$$

and

$$\begin{aligned}\hat{\mathcal{L}}_{\text{tr}} f(x, +2) &= 1_{\{x < L\}} (+2) \partial_x f(x, +2) + 1_{\{x = L\}} \omega (f(x, 0) - f(x, +2)), \\ \hat{\mathcal{L}}_{\text{tr}} f(x, 0) &= 1_{\{x = 0\}} \omega (f(x, +2) - f(x, 0)) + 1_{\{x = L\}} \omega (f(x, -2) - f(x, 0)), \\ \hat{\mathcal{L}}_{\text{tr}} f(x, -2) &= 1_{\{x > 0\}} (-2) \partial_x f(x, -2) + 1_{\{x = 0\}} \omega (f(x, 0) - f(x, -2)).\end{aligned}$$

The transport term  $\hat{\mathcal{L}}_{\text{tr}}$  hence describes motion with constant velocity  $v$  in the interior  $(0, L)$  combined with velocity jumps with rate  $\omega$  on the boundary.

Since the RTP process on  $[0, L]$  with parameter  $\omega$  is a lift of a sticky Brownian motion with parameter  $\omega$  by Theorem C.7, we begin by showing that the generator of sticky Brownian motion satisfies a Poincaré inequality and has discrete spectrum.

**Lemma C.18** (Poincaré inequality for sticky Brownian motion). *There exists  $C > 0$  such that*

$$\|f - \mu(f)\|_{L^2(\mu)}^2 \leq C \frac{\omega L + (\omega L)^2}{\omega^2} \langle f, -\mathcal{L}f \rangle_{L^2(\mu)}$$

for all  $f \in \text{Dom}(\mathcal{L})$ .

*Proof.* Note that

$$\langle f, (-\mathcal{L}f) \rangle_{L^2(\mu)} = \frac{\omega}{2 + \omega L} \int_0^L (f')^2 dx$$

hence the goal is to establish an inequality of the form

$$\text{Var}_\mu(f) \leq \frac{1}{m} \frac{\omega}{2 + \omega L} \int_0^L (f')^2 dx.$$

Recall that if  $\mu_p = \int \Pi_\theta dp(\theta)$  with probability measures  $\Pi_\theta$  and  $p$  then

$$\text{Var}_{\mu_p}(f) = \int \text{Var}_{\Pi_\theta}(f) p(d\theta) + \text{Var}_p \left( \int f d\Pi_\theta \right).$$

Recall also that

$$\mathrm{Var}_{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_L}(g) = \frac{1}{4}(g(0) - g(L))^2 = \frac{1}{4} \left( \int_0^L g'(x) dx \right)^2.$$

Rather than applying this to the triple mixture we have, we apply it twice on a mixture of two components by rewriting

$$\mu = \frac{\omega L}{2 + \omega L} \mathcal{U}_{[0,L]} + \frac{2}{2 + \omega L} \left( \frac{1}{2} \delta_0 + \frac{1}{2} \delta_L \right),$$

where  $\mathcal{U}_{[0,L]}$  is the uniform distribution on  $[0, L]$ . Then

$$\begin{aligned} \mathrm{Var}_\mu(f) &= \frac{\omega L}{2 + \omega L} \mathrm{Var}_{\mathcal{U}_{[0,L]}}(f) + \frac{2}{2 + \omega L} \mathrm{Var}_{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_L}(f) \\ &\quad + \frac{2\omega L}{(2 + \omega L)^2} \left( \frac{1}{L} \int_0^L f(x) dx - \frac{1}{2}f(0) - \frac{1}{2}f(L) \right)^2 \\ &= \frac{\omega L}{2 + \omega L} \mathrm{Var}_{\mathcal{U}_{[0,L]}}(f) + \frac{1/2}{2 + \omega L} \left( \int_0^L f'(x) dx \right)^2 \\ &\quad + \frac{2\omega L}{(2 + \omega L)^2} \left( \int_0^L f'(x) \frac{x - L/2}{L} dx \right)^2. \end{aligned}$$

We apply the Poincaré inequality for the uniform measure and Cauchy-Schwarz to get

$$\mathrm{Var}_\mu(f) \leq \left( \frac{L}{\pi^2} \frac{\omega L}{2 + \omega L} + \frac{L}{2(2 + \omega L)} + \frac{2\omega L}{(2 + \omega L)^2} \cdot \frac{L}{12} \right) \int_0^L (f'(x))^2 dx$$

using  $\int_0^L \left( \frac{x-L/2}{L} \right)^2 dx = \frac{L}{12}$ . The claim follows since

$$\left( \frac{L}{\pi^2} \frac{\omega L}{2 + \omega L} + \frac{L}{2(2 + \omega L)} + \frac{2\omega L}{(2 + \omega L)^2} \cdot \frac{L}{12} \right) \frac{2 + \omega L}{\omega} = \frac{L^2}{\pi^2} + \frac{L}{2\omega} + \frac{L^2}{6(2 + \omega L)}. \quad \square$$

Let us now prove that the generator of sticky Brownian motion has a discrete spectrum. While this result is surely known, we were unable to find a suitable reference.

**Lemma C.19.** *The generator  $(\mathcal{L}, \mathrm{Dom}(\mathcal{L}))$  of sticky Brownian motion has discrete spectrum.*

*Proof.* We will prove a Nash inequality for the Dirichlet form associated to the sticky Brownian motion and invariant measure  $\mu$ . Starting from [52, Chap. 8 Eq. (42)], a Nash inequality holds for any interval in dimension 1 for the Lebesgue measure, i.e. there exists  $c > 0$  such that

$$\left( \int_0^L f^2 dx \right)^{1/2} \leq c \left( \int_0^L |f| dx \right)^{2/3} \left( \left( \int_0^L f^2 dx \right)^{1/2} + \left( \int_0^L (f')^2 dx \right)^{1/2} \right)^{1/3}.$$

By [249] or [52, Chap. 8 Eq. (43)], this Nash inequality is equivalent to a super Poincaré inequality: For all  $s > 0$

$$\int_0^L f^2 dx \leq s \int_0^L (f')^2 dx + \beta(s) \left( \int_0^L |f| dx \right)^2,$$

with  $\beta(s)$  going to infinity of order  $\sim s^{-2}$  when  $s$  goes to 0, so that we may assume  $\beta(s) > 1$ . Now, let us prove that it implies the same type of super Poincaré inequality for the Dirichlet form associated to sticky Brownian motion. Denoting  $a_0 = \frac{1}{2+\omega L}$  and  $b_0 = \omega a_0$ , for all  $s > 0$  we have

$$\begin{aligned} \int f^2 d\mu &= a_0(f(0)^2 + f(L)^2) + b_0 \int_0^L f^2 dx, \\ &\leq a_0(f(0)^2 + f(L)^2) + b_0 s \int_0^L (f')^2 dx + b_0 \beta(s) \left( \int_0^L |f| dx \right)^2, \\ &\leq b_0 s \int_0^L (f')^2 dx + b_0 \max(b_0^{-2}, a_0^{-1}) \beta(s) \left( \int |f| d\mu \right)^2. \end{aligned}$$

Therefore, the sticky Brownian satisfies a super Poincaré inequality. Then by [249, Th. 5.1], it has an empty essential spectrum. Now, by [20, Th. A.6.4], the resolvent is compact and thus the generator has discrete spectrum.  $\square$

**Corollary C.20.** *There exists a universal constant  $K > 0$  s.t. for all  $\omega, L > 0$  there exists  $C = C(\omega, L) > 0$  such that the transition semigroup  $(\hat{P}_t)$  of the RTP process on  $[0, L]$  with parameter  $\omega$  satisfies*

$$\|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq C e^{-\frac{K\omega}{1+(\omega L)^2} t} \|f\|_{L^2(\hat{\mu})}$$

for all  $t \geq 0$  and  $f \in L_0^2(\hat{\mu})$ .

Note that the relaxation time corresponding to this decay rate is of the same order as the mixing time obtained in [121]. It reveals the existence of two regimes controlled by the parameter  $\omega L$ . In the ballistic regime  $\omega L \ll 1$ , velocity flips are rare, leading to a fast exploration of the position space  $\mathcal{S}$  and a comparatively slow exploration of the velocity space  $\mathcal{V}$ . This results in the scaling  $\nu \propto \omega$ . On the contrary, in the diffusive regime  $\omega L \gg 1$ , the high frequency of velocity flips makes the exploration of  $\mathcal{V}$  faster than the exploration of  $\mathcal{S}$ . This leads to the scaling  $\nu \propto \omega^{-1} L^{-2}$ .

*Proof.* It suffices to check Assumptions (A), (B) and (C) and apply Corollary C.13.

Assumption (A): Recall that  $\text{Dom}(\mathcal{L}_{C^0})$  is a core of  $\mathcal{L}$  by Theorem C.7. For all  $f \in \text{Dom}(\mathcal{L}_{C^0})$  we have  $\hat{\mathcal{L}}_v(f \circ \pi) = 0$  hence  $\hat{\mathcal{L}}_{\text{tr}}$  is a lift of  $\mathcal{L}$  by Remark C.11(ii). Furthermore, for  $f \in \text{Dom}(\mathcal{L}_{C^0})$  one has

$$\hat{\mathcal{L}}_{\text{tr}}^*(f \circ \pi)(x, v) = -v 1_{\{0 < x < L\}} f'(x) = -\hat{\mathcal{L}}_{\text{tr}}(f \circ \pi)(x, v).$$

Assumption (B): Follows from Lemmas C.18 and C.19.

### Appendix C. Convergence of Markov processes via lifting and flow Poincaré inequality

Assumptions (C) (ii): A straightforward computation yields

$$\int_{\mathcal{V}} \hat{\mathcal{L}}_v f(x, v) d\kappa_x(v) = 0 \text{ for all } x \in \mathcal{S} \text{ and } f \in \text{Dom}(\hat{\mathcal{L}}).$$

Let us we prove  $\int_0^T \|P_t f - \Pi_v P_t f\|_{L^2(\hat{\mu})}^2 dt \leq \frac{1}{m_v} \int_0^T \mathcal{E}_v(P_t f) dt$  with  $m_v = 2$ . Define the matrices

$$S = \begin{pmatrix} 1/4 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{pmatrix}, \quad \mathcal{Q} = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -2 & 1 \\ 0 & 2 & -2 \end{pmatrix},$$

as well as the scalar product  $\langle x, y \rangle_S = x^\top S y$  and let  $\Pi$  be the orthogonal projection on the kernel of  $\mathcal{Q}$  with respect to  $\langle \cdot, \cdot \rangle_S$ . The matrix  $\mathcal{Q}$  is symmetric w.r.t. the scalar product  $\langle \cdot, \cdot \rangle_S$  (i.e.  $\mathcal{Q} = S^{-1} \mathcal{Q}^\top S$ ) so it admits an orthonormal basis of eigenvectors. Its eigenvalues are 0, -2, -4 and thus

$$\|\xi - \Pi \xi\|_S^2 \leq \frac{1}{2} \langle \xi, (-\mathcal{Q}) \xi \rangle_S \quad \text{for all } \xi \in \mathbb{R}^3.$$

For  $f \in \text{Dom}(\hat{\mathcal{L}})$ , taking  $\xi = (P_t f(x, 2), P_t f(x, 0), P_t f(x, -2))^\top$  in the inequality above yields

$$\kappa_x(P_t f(x, \cdot)^2) - \kappa_x(P_t f(x, \cdot))^2 \leq \frac{1}{2} \int_{\mathcal{V}} P_t f(x, v) (-\hat{\mathcal{L}}_v P_t f)(x, v) d\kappa_x(v) \quad (\text{C.17})$$

for  $x \in (0, L)$ . A similar argument leads to the same inequality for  $x = 0$  and  $x = L$ . Hence

$$\begin{aligned} \int_0^T \|P_t f - \Pi_v P_t f\|_{L^2(\hat{\mu})}^2 dt &= \int_0^T \int_{[0, L]} [\kappa_x(P_t f(x, \cdot)^2) - \kappa_x(P_t f(x, \cdot))^2] d\mu(x) dt \\ &\leq \frac{1}{2} \int_0^T \int_{[0, L]} \int_{\mathcal{V}} P_t f(x, v) (-\hat{\mathcal{L}}_v P_t f)(x, v) d\kappa_x(v) d\mu(x) dt. \end{aligned}$$

Therefore, (C.11) holds with  $m_v = 2$ .

Assumption (C) (i): Let  $f \in \text{Dom}(\hat{\mathcal{L}}_{\text{tr}})$  and  $g \in \text{Dom}(\mathcal{L}_{C^0})$ . We have that  $u = f - \Pi_v f$  satisfies  $\Pi_v u = 0$ , so, in particular,

$$u(x, 2) + u(x, -2) = -2u(x, 0) \quad \text{for all } x \in (0, L).$$

Therefore

$$\begin{aligned} &\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\hat{\mu})} \\ &= \frac{\omega/4}{2 + \omega L} \int_0^L (+2) \partial_x g(x) (+2) \partial_x u(x, +2) dx \\ &\quad + \frac{\omega/4}{2 + \omega L} \int_0^L (-2) \partial_x g(x) (-2) \partial_x u(x, -2) dx \end{aligned}$$

$$\begin{aligned}
 &= \frac{\omega}{2 + \omega L} \int_0^L \partial_x g(x) (\partial_x u(x, +2) + \partial_x u(x, -2)) \, dx \\
 &= \frac{-2\omega}{2 + \omega L} \int_0^L \partial_x g(x) \partial_x u(x, 0) \, dx, \\
 &|\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\hat{\mu})}| \\
 &= |\langle 2\mathcal{L}g, u(\cdot, 0) \rangle_{L^2(\mu)}| \leq \|2\mathcal{L}g\|_{L^2(\mu)} \|u(\cdot, 0)\|_{L^2(\mu)}.
 \end{aligned}$$

Finally,

$$\begin{aligned}
 \|u(\cdot, 0)\|_{L^2(\mu)}^2 &= \frac{1}{2 + \omega L} u(0, 0)^2 + \frac{\omega}{2 + \omega L} \int_0^L u(x, 0)^2 \, dx + \frac{1}{2 + \omega L} u(L, 0)^2 \\
 &\leq 2 \left[ \frac{1}{2 + \omega L} \left( \frac{1}{2} u(0, -2)^2 + \frac{1}{2} u(0, 0)^2 \right) \right. \\
 &\quad \left. + \frac{\omega}{2 + \omega L} \int_0^L \left( \frac{1}{4} u(x, -2)^2 + \frac{1}{2} u(x, 0)^2 + \frac{1}{4} u(x, +2)^2 \right) \, dx \right. \\
 &\quad \left. + \frac{1}{2 + \omega L} \left( \frac{1}{2} u(L, +2)^2 + \frac{1}{2} u(L, 0)^2 \right) \right] \\
 &\leq 2 \|u\|_{L^2(\hat{\mu})}^2,
 \end{aligned}$$

and hence

$$|\langle \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi), \hat{\mathcal{L}}_{\text{tr}}(f - \Pi_v f) \rangle_{L^2(\hat{\mu})}| \leq \sqrt{2} \|2\mathcal{L}g\|_{L^2(\mu)} \|f - \Pi_v f\|_{L^2(\hat{\mu})} \leq \|2\mathcal{L}g\|_{L^2(\mu)} \sqrt{\mathcal{E}_v(f)}$$

using (C.17) for the last inequality. Therefore, (C1) is satisfied with  $C_1 = 1$ . The arguments used to check (C.11) show that  $\hat{\mathcal{L}}_v$  is a bounded operator and hence (C2) is satisfied with  $C_2 = O(m^{-1/2})$  by Remark C.11(v).

Remark C.14 thus yields the desired decay rate.  $\square$

### C.4.2. Proof of Theorem C.5

The following lemma is surely known, or at least folklore, but we did not find a reference, so the proof is included.

**Lemma C.21.** *Let  $(\mathcal{L}_{\text{ext}}, \text{Dom}(\mathcal{L}_{\text{ext}}))$  (resp.  $(\mathcal{L}_{C^0}, \text{Dom}(\mathcal{L}_{C^0}))$ ) be the extended (resp.  $C^0$ ) generator of a Feller process  $X(t)$  taking its values in the compact state space  $\mathcal{S}$ . Then  $\text{Dom}(\mathcal{L}_{C^0}) = \{f \in C^0(\mathcal{S}) : f \in \text{Dom}(\mathcal{L}_{\text{ext}}) \text{ and } \mathcal{L}_{\text{ext}} f \in C^0(\mathcal{S})\}$  and*

$$\mathcal{L}_{C^0} f = \mathcal{L}_{\text{ext}} f \quad \text{for all } f \in \text{Dom}(\mathcal{L}_{C^0}).$$

*Proof.* It follows from [182, Theorem 3.32] that the extended generator is an extension of the  $C^0$ -generator.

Let  $f \in C^0(\mathcal{S})$  be such that  $f \in \text{Dom}(\mathcal{L}_{\text{ext}})$  and  $\mathcal{L}_{\text{ext}} f \in C^0(\mathcal{S})$ . Then

$$M(t) = f(X(t)) - f(X(0)) - \int_0^t \mathcal{L}_{\text{ext}} f(X(s)) \, ds$$

is a local martingale. Hence

$$\mathbb{E} \left[ \sup_{0 \leq s \leq t} \left| f(X(t)) - f(X(0)) - \int_0^t \mathcal{L}_{\text{ext}} f(X(s)) \, ds \right| \right] \leq 2\|f\|_\infty + t\|\mathcal{L}_{\text{ext}} f\|_\infty < +\infty$$

implies that  $M(t)$  is a martingale and thus

$$P_t f(x) - f(x) = \mathbb{E}_x [f(X(t)) - f(X(0))] = \mathbb{E}_x \left[ \int_0^t \mathcal{L}_{\text{ext}} f(X(s)) \, ds \right] = \int_0^t P_s \mathcal{L}_{\text{ext}} f(x) \, ds.$$

The function  $s \mapsto P_s \mathcal{L}_{\text{ext}} f$  is continuous from  $\mathbb{R}_+$  to  $C^0(\mathcal{S})$  hence the integral  $\int_0^t P_s \mathcal{L}_{\text{ext}} f \, ds$  is well-defined in  $C^0$  and the pointwise equality above implies

$$\frac{P_t f - f}{t} = \frac{1}{t} \int_0^t P_s \mathcal{L}_{\text{ext}} f \, ds \rightarrow \mathcal{L}_{\text{ext}} f \quad \text{as } t \rightarrow 0,$$

using the continuity of  $s \mapsto P_s \mathcal{L}_{\text{ext}} f$  to obtain a  $\|\cdot\|_\infty$ -limit.  $\square$

**Lemma C.22** ( $C^0$ -generator of the RTP process).

(i) *The RTP process is Feller.*

(ii) *Its  $C^0$ -generator is given by*

$$\begin{aligned} \hat{\mathcal{L}}_{C^0} f(x, 2) &= 2\partial_x f(x, 2) + 2\omega (f(x, 0) - f(x, 2)), \\ \hat{\mathcal{L}}_{C^0} f(x, 0) &= \omega f(x, 2) + \omega f(x, -2) - 2\omega f(x, 0), \\ \hat{\mathcal{L}}_{C^0} f(x, -2) &= -2\partial_x f(x, -2) + 2\omega (f(x, 0) - f(x, -2)), \end{aligned}$$

and its domain  $\text{Dom}(\hat{\mathcal{L}}_{C^0})$  consists of all  $f \in C^0(\hat{\mathcal{S}})$  such that

$$f(\cdot, \pm 2) \in C^1([0, L]) \quad \text{and} \quad \partial_x f(L, 2) = \partial_x f(0, -2) = 0.$$

*Proof.* (i) It is enough to show that  $\mathbb{P}_{(\tilde{x}_0, v_0)}$  converges weakly to  $\mathbb{P}_{(x_0, v_0)}$  when  $\tilde{x}_0 \rightarrow x_0$ . Following arguments in [121], we consider the coupling of  $\mathbb{P}_{(\tilde{x}_0, v_0)}$  and  $\mathbb{P}_{(x_0, v_0)}$  obtained by taking the same relative velocity process  $v(t)$ . Since  $y \mapsto \min(L, \max(0, y + v(0)t))$  is 1-Lipschitz, we have

$$|x(t) - \tilde{x}(t)| = |\min(L, \max(0, x_0 + v(0)t)) - \min(L, \max(0, \tilde{x}_0 + v(0)t))| \leq |x_0 - \tilde{x}_0|$$

for  $t \leq T_1$ . In particular,  $|x(T_1) - \tilde{x}(T_1)| \leq |x_0 - \tilde{x}_0|$ . So, by induction,  $|x(t) - \tilde{x}(t)| \leq |x_0 - \tilde{x}_0|$  for all  $t \geq 0$ , and the Feller property follows.

(ii) The claim follows from the derivation of the extended generator in [121] and Lemma C.21.  $\square$

The invariance of  $\hat{\mu}$  (Theorem C.5 (i)) is already proven in [126] and can be directly checked here using the generator [182, Theorem 3.37]. We now provide the proof for the  $L^2(\hat{\mu})$  generator and its domain.

*Proof of Theorem C.5 (ii).* The  $L^2(\hat{\mu})$  generator is the closure of  $(\hat{\mathcal{L}}_{C^0}, \text{Dom}(\hat{\mathcal{L}}_{C^0}))$  in  $L^2(\hat{\mu})$  by [104, Proposition II.1.7]. We concentrate on the case  $v = 2$  as cases  $v = 0$  and  $v = -2$  follow from the same arguments. Denote  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  the operator defined in Theorem C.5 (ii).

We first show  $\overline{(\hat{\mathcal{L}}_{C^0}, \text{Dom}(\hat{\mathcal{L}}_{C^0}))} \subseteq (\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$ : Let  $f_n \in \text{Dom}(\hat{\mathcal{L}}_{C^0})$  be such that

$$f_n \longrightarrow f \quad \text{and} \quad \hat{\mathcal{L}}_{C^0} f_n \longrightarrow g \quad \text{in } L^2(\hat{\mu}).$$

We need to show that  $f \in \text{Dom}(\hat{\mathcal{L}})$  and  $\hat{\mathcal{L}}f = g$ . On  $(0, L)$  one has

$$\partial_x f_n(\cdot, 2) = \frac{1}{2} \left( \hat{\mathcal{L}}_{C^0} f_n(\cdot, 2) - 2\omega [f_n(\cdot, 0) - f_n(\cdot, 2)] \right)$$

hence  $\partial_x f_n(\cdot, 2)$  converges in  $L^2((0, L), dx)$  to  $h = \frac{1}{2} (g(\cdot, 2) - 2\omega (f(\cdot, 0) - f(\cdot, 2)))$ . For  $\varphi \in C_c^\infty((0, L))$  we have

$$-\int_0^L f(\cdot, 2) \varphi' dx = -\lim_{n \rightarrow \infty} \int_0^L f_n(\cdot, 2) \varphi' dx = \lim_{n \rightarrow \infty} \int_0^L \partial_x f_n(\cdot, 2) \varphi dx = \int_0^L h \varphi dx.$$

Hence  $f(\cdot, 2) \in H^1((0, L))$ ,  $f_n(\cdot, 2) \rightarrow f(\cdot, 2)$  in  $H^1((0, L))$ , and  $g(x, 2) = 2\partial_x f(x, 2) + 2\omega (f(x, 0) - f(x, 2))$  for  $x \in (0, L)$ . Since function evaluations at  $(L, 2)$  and  $(L, 0)$  are well-defined and continuous from  $L^2(\hat{\mu})$  to  $\mathbb{R}$ , we obtain  $g(L, 2) = \omega (f(L, 0) - f(L, 2))$ . Furthermore,  $F \mapsto F(L-, 2)$  is the trace operator which is continuous on  $H^1((0, L))$ , so that

$$f_n(L-, 2) = f_n(L, 2) \text{ for all } n \geq 0 \implies f(L-, 2) = f(L, 2),$$

which yields  $f \in \text{Dom}(\hat{\mathcal{L}})$  and  $\hat{\mathcal{L}}f = g$ .

It remains to show  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}})) \subseteq \overline{(\hat{\mathcal{L}}_{C^0}, \text{Dom}(\hat{\mathcal{L}}_{C^0}))}$ : Let  $f \in \text{Dom}(\hat{\mathcal{L}})$  be given. We now construct a sequence  $f_n \in \text{Dom}(\hat{\mathcal{L}}_{C^0})$  such that  $f_n \longrightarrow f$  and  $\hat{\mathcal{L}}_{C^0} f_n \longrightarrow \hat{\mathcal{L}}f$  in  $L^2(\hat{\mu})$ .

Let  $h_n \in C^\infty((0, L))$  be a sequence converging to  $f(\cdot, 2)$  in  $H^1((0, L))$ . Define

$$\chi_n(x) = \begin{cases} 0 & \text{if } x \leq L(1 - 1/n), \\ \frac{x - L(1 - 1/n)}{L(1 - 1/(2n)) - L(1 - 1/n)} & \text{if } x \in (L(1 - 1/n), L(1 - 1/(2n))), \\ 1 & \text{if } x \geq L(1 - 1/(2n)). \end{cases}$$

which is in  $C^0([0, L])$ . Set  $g_n(x) = h_n(x) - \int_0^x h'_n(y) \chi_n(y) dy$ . One has  $g_n \in C^1([0, L])$  as well as

$$g'_n(L) = h'_n(L) - h'_n(L) \chi_n(L) = 0.$$

Furthermore,

$$\left( \int_0^x h'_n \chi_n dy \right)^2 \leq \left( \int_0^L (h'_n)^2 dy \right) \left( \int_0^L \chi_n^2 dy \right) \leq \left( \int_0^L (h'_n)^2 dy \right) \cdot \frac{L}{n}$$

so that  $x \mapsto \int_0^x h'_n(y) \chi_n(y) dy$  converges to 0 uniformly and thus in  $L^2((0, L), dx)$ . For all  $a \in (0, L)$  we have

$$\lim_{n \rightarrow \infty} \int_0^L (h'_n \chi_n)^2 dy \leq \lim_{n \rightarrow \infty} \int_a^L (h'_n)^2 dy = \int_a^L (\partial_x f(\cdot, +2))^2 dy,$$

so taking  $a \rightarrow L$  yields

$$\lim_{n \rightarrow \infty} \int_0^L (h'_n \chi_n)^2 dy = 0.$$

Hence  $x \mapsto \int_0^x h'_n(y) \chi_n(y) dy$  converges to 0 in  $H^1((0, L))$ . We deduce that  $g_n$  converges to  $f(\cdot, 2)$  in  $H^1((0, L))$ . Because the trace  $F \mapsto F(L-, 2)$  is continuous from  $H^1((0, L))$  to  $\mathbb{R}$  and  $f(L-, 2) = f(L, 2)$  we have that  $g_n \rightarrow f(\cdot, 2)$  in  $L^2([0, L], \lambda_{[0, L]} + \delta_L)$  as required. Thus taking  $f_n(x, +2) = g_n(x)$  yields convergence of  $(\hat{\mathcal{L}}_{C^0} f_n)(\cdot, +2)$  to  $\hat{\mathcal{L}} f(\cdot, +2)$  in  $L^2(\lambda_{[0, L]} + \delta_L)$ . The cases  $v = 0$  and  $v = -2$  can be treated similarly to obtain an approximating sequence for  $f \in \text{Dom}(\hat{L})$  that converges in  $L^2(\hat{\mu})$ .  $\square$

## C.5. PDMPs for sampling

In this section, we consider a probability measure  $\mu$  on  $\mathbb{R}^d$  with density proportional to  $\exp(-U)$ , where  $U: \mathbb{R}^d \rightarrow \mathbb{R}$  is such that  $\int_{\mathbb{R}^d} e^{-U(x)} dx < \infty$ . We assume that  $\mu$  satisfies a Poincaré inequality with constant  $\frac{1}{m}$  and that the generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  of the overdamped Langevin diffusion with invariant probability measure  $\mu$ , see Example C.3, has discrete spectrum. The latter is for instance satisfied if there exists  $\beta > 1$  such that

$$\liminf_{|x| \rightarrow \infty} \frac{\nabla U(x) \cdot x}{|x|^\beta} > 0,$$

see [41, Theorem 5]. As in Example C.15, we assume the lower curvature bound

$$\nabla^2 U(x) \geq -Km \cdot I_d \quad \text{for all } x \in \mathbb{R}^d \quad (\text{C.18})$$

for some  $K \geq 0$ .

Piecewise deterministic Markov processes (PDMPs) have recently been studied as highly efficient non-reversible samplers in physics and statistics, where proving exponential ergodicity is crucial. Depending on communities or the specific choice of dynamics, these processes appear under different algorithmic names, including (Forward) Event Chain Monte Carlo [196, 197], the Zig-Zag sampler [36] and the Bouncy Particle sampler [91]. Very recently, using the Dolbeault-Mouhot-Schmeiser approach [87], results with well-controlled dimensional dependence were established in [12]. These results were refined in [186] using the variational approach by Albritton, Armstrong, Mourrat and Novack for the Zig-Zag process and the Bouncy Particle sampler. Here, we extend these results using our method, treating in particular two previously unaddressed cases: the true Zig-Zag process, without the condition of Gaussian velocities, and the Forward generalisation that incorporates stochastic jumps at events.

### C.5.1. Forward process

The forward Event Chain process introduced in [196] is a PDMP  $(X_t, V_t)$  with state space  $\hat{\mathcal{S}} = \mathbb{R}^d \times \mathbb{R}^d$  and invariant probability measure  $\hat{\mu} = \mu \otimes \kappa$ , where  $\kappa$  is the  $d$ -dimensional standard Gaussian. It extends previously developed algorithms, as the Bouncy Particle

Sampler and the Zig-Zag process, as the velocity update at an event is now governed more generally by a Markov kernel. We consider the *full-orthogonal refresh* variant whose generator is given by

$$\hat{\mathcal{L}}f(x, v) = v \cdot \nabla_x f(x, v) + (v \cdot \nabla U(x))_+(B_-f - f)(x, v) + \gamma(\Pi_v f - f)(x, v)$$

for  $f \in C_c^1(\mathbb{R}^{2d})$  and  $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$ , where

$$B_{\pm}f(x, v) = \frac{1}{Z_x} \int_{\mathbb{R}^d} (w \cdot \nabla U(x))_{\pm} f(x, w) \kappa(dw)$$

and  $Z_x = \int_{\mathbb{R}^d} (w \cdot \nabla U(x))_+ \kappa(dw) = \frac{1}{\sqrt{2\pi}} |\nabla U(x)|$ . It moves along straight lines between events occurring with rate  $(v \cdot \nabla U)_+$ . At an event, the velocity  $V_t$  is resampled according to the distribution  $Z_{X_t}^{-1} (v \cdot \nabla U(X_t))_- \kappa(dw)$ , corresponding to the distribution for the reflection of the new direction to trigger an event. This corresponds to a resampling of the whole vector  $V_t$ , including its component along the direction  $\nabla U(X_t)$ , as opposed to a simple reflection or flip as respectively for the Bouncy Particle Sampler or Zig-Zag process. The forward Event Chain process admits a natural splitting of its generator. For  $f \in C_c^1(\mathbb{R}^{2d})$ , we set

$$\begin{aligned} \hat{\mathcal{L}}_{\text{tr}} f(x, v) &= v \cdot \nabla_x f(x, v) + (v \cdot \nabla U(x))_+(B_-f - f)(x, v) \quad \text{and} \\ \hat{\mathcal{L}}_v f(x, v) &= \int_{\mathbb{R}^d} f(x, w) \kappa(dw) - f(x, v) = \Pi_v f(x, v) - f(x, v), \end{aligned}$$

so that  $\hat{\mathcal{L}}f = \hat{\mathcal{L}}_{\text{tr}}f + \gamma\hat{\mathcal{L}}_vf$ .

**Assumption C.1.** There exist  $a \in [0, 1)$  and  $b \geq 0$  such that

$$\Delta U(x) \leq a|\nabla U(x)|^2 + b \quad \text{for all } x \in \mathbb{R}^d. \quad (\text{C.19})$$

For example, Assumption C.1 is satisfied with  $a = 0$  and  $b = dL$  if  $\nabla^2 U(x) \leq L \cdot I_d$  for all  $x \in \mathbb{R}^d$ .

**Corollary C.23.** *Let Assumption C.1 be satisfied. Then, choosing  $T = m^{-1/2}$ , the transition semigroup of the Forward process is exponentially contractive in  $T$ -average with rate*

$$\nu = \Omega \left( \frac{\gamma m}{\gamma^2 + \frac{m(1+K)+b}{(1-a)^2}} \right).$$

*In particular, if  $\gamma = \Theta \left( \frac{\sqrt{m(1+K)+b}}{1-a} \right)$  then  $\nu = \Omega \left( \sqrt{m}(1-a) / \sqrt{1+K+b/m} \right)$ .*

**Remark C.24.** In case  $m \cdot I_d \leq \nabla^2 U \leq L \cdot I_d$ , and with  $\gamma = \Theta(\sqrt{m+dL})$ , we obtain  $\nu = \Omega(\sqrt{m}/\sqrt{dL/m})$ . We thus recover the same dimension dependence as for the Bouncy Particle sampler in [186] but an additional dependence on the condition number.

*Proof of Corollary C.23.* Assumptions (A) and (B) can be verified as for the Langevin dynamics in Example C.15. Indeed, the forward Event Chain process is a lift of an overdamped diffusion with generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$ , and Assumption (C)(ii) is satisfied with  $m_v = 1$ . Assumption (C2) holds with  $C_2 = \frac{1}{\sqrt{2m}}$  by Remark C.11(v), we hence focus on Assumption (C1).

It is enough to estimate  $\|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + (\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})}$  by Remark C.11(iv). Noticing

$$\hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) = v \cdot \nabla_x g \quad \text{and} \quad \hat{\mathcal{L}}_{\text{tr}}^* f = -v \cdot \nabla_x f(x, v) + (v \cdot \nabla U(x))_-(B_+ f - f)(x, v)$$

for  $f \in C_c^1(\mathbb{R}^{2d})$ , one has

$$\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)(x, v) = -v^\top \nabla^2 g(x) v + \frac{1}{Z_x} (v \cdot \nabla U(x))_- \left( \frac{1}{2} \nabla U(x) \cdot \nabla g(x) - Z_x v \cdot \nabla g(x) \right).$$

By Cauchy-Schwarz,

$$\begin{aligned} \int (\hat{\mathcal{L}}_{\text{tr}}^* (v \cdot \nabla_x g))^2 \kappa(dv) &\leq 2 \int (v \cdot \nabla^2 g v)^2 \kappa(dv) \\ &\quad + \frac{2}{Z^2} \int (v \cdot \nabla U)_-^2 \left( \frac{1}{2} \nabla U \cdot \nabla g - Z v \cdot \nabla g \right)^2 \kappa(dv) \\ &= 4 \|\nabla^2 g\|_F^2 + 2(\Delta g)^2 \\ &\quad + 2(3 + \frac{\pi}{4})(\nabla U \cdot \nabla g)^2 + |\nabla U|^2 |\nabla g|^2. \end{aligned}$$

where  $\|\cdot\|_F$  is the Frobenius norm. Using (C.19) we estimate

$$\begin{aligned} \int_{\mathbb{R}^d} |\nabla U|^2 |\nabla g|^2 d\mu &= \int_{\mathbb{R}^d} \Delta U |\nabla g|^2 d\mu + 2 \int_{\mathbb{R}^d} \nabla U^\top \nabla^2 g \nabla g d\mu \\ &\leq a \int_{\mathbb{R}^d} |\nabla U|^2 |\nabla g|^2 d\mu + b \int_{\mathbb{R}^d} |\nabla g|^2 d\mu + 2 \int_{\mathbb{R}^d} \nabla U^\top \nabla^2 g \nabla g d\mu. \end{aligned}$$

Now using  $2\nabla U^\top \nabla^2 g \nabla g \leq 2|\nabla U| |\nabla g| \|\nabla^2 g\|_F \leq p|\nabla U|^2 |\nabla g|^2 + \frac{1}{p} \|\nabla^2 g\|_F^2$  for any  $p > 0$ , choosing  $p = \frac{1-a}{2}$  and rearranging yields

$$\begin{aligned} \int_{\mathbb{R}^d} |\nabla U|^2 |\nabla g|^2 d\mu &\leq \frac{2b}{1-a} \int_{\mathbb{R}^d} |\nabla g|^2 d\mu + \frac{4}{(1-a)^2} \int_{\mathbb{R}^d} \|\nabla^2 g\|_F^2 d\mu \\ &\leq \left( \frac{2b}{1-a} + \frac{4Km}{(1-a)^2} \right) \int_{\mathbb{R}^d} |\nabla g|^2 d\mu + \frac{4}{(1-a)^2} \|2\mathcal{L}g\|_{L^2(\mu)}^2. \end{aligned} \tag{C.20}$$

Finally, the bounds  $(\Delta g)^2 \leq 2(2\mathcal{L}g)^2 + (\nabla U \cdot \nabla g)^2$  and  $(\nabla U \cdot \nabla g)^2 \leq |\nabla U|^2 |\nabla g|^2$  as well as Bochner's identity yield

$$\begin{aligned} \|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)\|_{L^2(\hat{\mu})}^2 &\leq C \left( \|\|\nabla^2 g\|_F\|_{L^2(\mu)}^2 + \|\Delta g\|_{L^2(\mu)}^2 + \|\nabla U \cdot \nabla g\|_{L^2(\mu)}^2 + \|\|\nabla U\| |\nabla g|\|_{L^2(\mu)}^2 \right) \\ &\leq C \left( \|2\mathcal{L}g\|_{L^2(\mu)}^2 + 2Km\mathcal{E}(g) + \|\|\nabla U\| |\nabla g|\|_{L^2(\mu)}^2 \right) \end{aligned}$$

$$\begin{aligned}
&\leq C \left( \frac{1}{(1-a)^2} \|2\mathcal{L}g\|_{L^2(\mu)}^2 + \frac{b(1-a) + Km}{(1-a)^2} (2\mathcal{E}(g)) \right) \\
&\leq C \left( \frac{b(1-a) + (1+K)m}{m(1-a)^2} \right) \|2\mathcal{L}g\|_{L^2(\mu)}^2
\end{aligned}$$

where  $C > 0$  denotes an absolute constant changing from line to line. Therefore, overall, one can choose  $C_1^2 = C \left( \frac{b+(1+K)m}{m(1-a)^2} \right)$  and the claim follows by Remark C.14.  $\square$

### C.5.2. Zig-Zag process

The convergence to equilibrium for the Zig-Zag process with Gaussian velocities is considered in [186]. In practice, one usually considers velocities chosen uniformly among the (scaled) coordinate directions  $\{\pm\sqrt{d}e_i : i = 1, \dots, n\}$  or from the hypercube  $\{-1, +1\}^d$  [36]. We consider any of these dynamics, i.e. let  $\mathcal{V}$  be  $\mathbb{R}^d$ ,  $\{\pm\sqrt{d}e_i : i = 1, \dots, n\}$ , or  $\{-1, +1\}^d$ , and consider  $\kappa = \mathcal{N}(0, I_d)$  or  $\kappa = \text{Unif}(\mathcal{V})$ , correspondingly.

The Zig-Zag process  $(X_t, V_t)$  is a PDMP with state space  $\hat{\mathcal{S}} = \mathbb{R}^d \times \mathcal{V}$  that, between events, moves with constant velocity  $v \in \mathcal{V}$ . The  $k$ -th component  $v_k$  of the velocity is then flipped with rate  $(v_k \partial_k U(x))_+$ . Additionally, with rate  $\gamma > 0$ , a new velocity is chosen from  $\mathcal{V}$  according to the distribution  $\kappa$ . Note that this velocity refreshment mechanism may not be needed for ergodicity in the hypercube case [36], but it is essential in our approach. This process has  $\hat{\mu} = \mu \otimes \kappa$  as its invariant probability measure [36], and the generator of the associated transition semigroup in  $L^2(\hat{\mu})$  is given by

$$\begin{aligned}
\hat{\mathcal{L}}f(x, v) &= v \cdot \nabla_x f(x, v) + \sum_{i=1}^d (v_i \partial_i U(x))_+ (f(x, v - 2e_i v_i) - f(x, v)) \\
&\quad + \gamma \left( \int f(x, w) \kappa(dw) - f(x, v) \right)
\end{aligned}$$

for compactly supported functions  $f \in C_c^1(\mathbb{R}^{2d})$  and  $(x, v) \in \mathbb{R}^d \times \mathcal{V}$ . Under additional assumptions, see [92],  $C_c^1(\mathbb{R}^d)$  forms a core for  $\hat{\mathcal{L}}$ . For  $f \in \text{Dom}(\hat{\mathcal{L}})$ , we set

$$\begin{aligned}
\hat{\mathcal{L}}_{\text{tr}}f(x, v) &= v \cdot \nabla_x f(x, v) + \sum_{i=1}^d (v_i \partial_i U(x))_+ (f(x, v - 2e_i v_i) - f(x, v)) \quad \text{and} \\
\hat{\mathcal{L}}_vf(x, v) &= \int_{\mathcal{V}} f(x, w) \kappa(dw) - f(x, v) = (\Pi_v f)(x, v) - f(x, v),
\end{aligned}$$

so that  $\hat{\mathcal{L}}f = \hat{\mathcal{L}}_{\text{tr}}f + \gamma \hat{\mathcal{L}}_vf$ .

**Assumption C.2.** There exists  $L \geq 0$  such that  $\partial_{kk}U(x) \leq L$  for all  $k \in \{1, \dots, d\}$  and  $x \in \mathbb{R}^d$ .

In particular, Assumption C.2 holds if  $\nabla^2 U(x) \leq L \cdot I_d$  for all  $x \in \mathbb{R}^d$ .

**Corollary C.25.** Let Assumption C.2 be satisfied.

### Appendix C. Convergence of Markov processes via lifting and flow Poincaré inequality

- (i) Let  $\kappa = \mathcal{N}(0, I_d)$  or  $\kappa = \text{Unif}(\{-1, +1\}^d)$ . Then, choosing  $T = m^{-1/2}$ , the transition semigroup  $(\hat{P}_t)$  of the Zig-Zag process is exponentially contractive in  $T$ -average with rate

$$\nu = \Omega\left(\frac{\gamma m}{\gamma^2 + m(1 + K) + L}\right).$$

If  $\gamma = \Theta\left(\sqrt{m(1 + K) + L}\right)$  then  $\nu = \Omega\left(\sqrt{m} / \sqrt{(1 + K) + L/m}\right)$ .

- (ii) Let  $\kappa = \text{Unif}(\{\pm\sqrt{d}e_i : i = 1, \dots, n\})$ . Then, choosing  $T = m^{-1/2}$ , the transition semigroup  $(\hat{P}_t)$  of the Zig-Zag process is exponentially contractive in  $T$ -average with rate

$$\nu = \Omega\left(\frac{\gamma m}{\gamma^2 + md(1 + K) + Ld}\right).$$

If  $\gamma = \Theta\left(\sqrt{(m(1 + K) + L)d}\right)$  then  $\nu = \Omega\left(\sqrt{m} / \sqrt{((1 + K) + L/m)d}\right)$ .

In [186], quantitative convergence rates for the Zig-Zag process with Gaussian velocities are shown. Corollary C.25 recovers the same scaling in  $m$ ,  $d$  and  $L$  under slightly weaker assumptions in the practically relevant case of velocities chosen from the hypercube [34].

*Proof of Corollary C.25.* Assumptions (A) and (B) can be verified as for the Langevin dynamics in Example C.15. Indeed, the Zig-Zag process satisfies  $\hat{\mathcal{L}}(f \circ \pi)(x, v) = v \cdot \nabla f(x)$  for all  $f \in C_c^\infty(\mathbb{R}^d)$ . Since the distributions  $\kappa$  we consider are centred and have identity covariance matrix, the Zig-Zag process is a lift of the overdamped Langevin diffusion with invariant measure  $\mu$  and generator  $\mathcal{L} = -\frac{1}{2}\nabla^* \nabla$  in  $L^2(\mu)$  and core  $C_c^\infty(\mathbb{R}^d)$  for any of these choices of velocity distribution, and Assumption (C)(ii) is satisfied with  $m_v = 1$ . Assumption (C2) holds with  $C_2 = 1$  by Remark C.11(v), we hence focus on Assumption (C1).

We have

$$\hat{\mathcal{L}}_{\text{tr}}^* f(x, v) = -v \cdot \nabla_x f + \sum_{k=1}^d (v_k \partial_k U(x))_- (f(x, v - 2e_k v_k) - f(x, v))$$

for all  $f \in C_c^1(\mathbb{R}^{2d})$  and thus

$$\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) = -v \cdot \nabla^2 g(x) v - 2 \sum_{k=1}^d (v_k \partial_k U(x))_- v_k \partial_k g(x).$$

for all  $g \in C_c^\infty(\mathbb{R}^d)$ . This gives

$$\text{Var}_\kappa(\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi)) \leq 2 \text{Var}_\kappa(v \cdot \nabla^2 g v) + 8 \text{Var}_\kappa\left(\sum_{k=1}^d (v_k \partial_k U)_- v_k \partial_k g\right).$$

This can be bounded using the moments of  $\kappa$  in the following way.

(i) If  $\kappa = \mathcal{N}(0, I_d)$  or  $\kappa = \text{Unif}(\{-1, +1\}^d)$ , then

$$\begin{aligned}
\text{Var}_\kappa(v \cdot \nabla^2 g v) &= \sum_{i=1}^d \mathbb{E}_\kappa(v_i^4)(\partial_{ii}g)^2 + \sum_{i \neq j} \mathbb{E}_\kappa(v_i^2 v_j^2)(2(\partial_{ij}g)^2 + \partial_{ii}g \partial_{jj}g) - (\Delta g)^2 \\
&\leq 2\|\nabla^2 g\|_F^2, \quad \text{and} \\
\text{Var}_\kappa\left(\sum_{k=1}^d (v_k \partial_k U) - v_k \partial_k g\right) &= \sum_{k=1}^d \text{Var}_\kappa((v_k \partial_k U) - v_k \partial_k g) \\
&= \sum_{k=1}^d \mathbb{E}_\kappa((v_k \partial_k U)^2 - (v_k \partial_k g)^2) - (\mathbb{E}_\kappa((v_k \partial_k U) - v_k \partial_k g))^2 \\
&= \sum_{k=1}^d \partial_k U^2 \partial_k g^2 \left( \frac{1}{2} \mathbb{E}_\kappa(v_k^4) - \left( \frac{1}{2} \mathbb{E}_\kappa(v_k^2) \right)^2 \right) \\
&= C_\kappa \sum_{k=1}^d (\partial_k U)^2 (\partial_k g)^2
\end{aligned}$$

with  $C_\kappa = \frac{5}{4}$  in the Gaussian case, and  $C_\kappa = \frac{1}{4}$  in the hypercube case. Hence

$$\|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + 2(\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})}^2 \leq 4\|\nabla^2 g\|_F^2 + 8C_\kappa \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu.$$

(ii) If  $\kappa = \text{Unif}(\{\pm\sqrt{d}e_i : i = 1, \dots, n\})$  is the uniform distribution on the scaled coordinate directions, then

$$\begin{aligned}
\text{Var}_\kappa(v \cdot \nabla^2 g v) &= d \sum_{k=1}^d (\partial_{kk}g)^2 - (\Delta g)^2 \leq d\|\nabla^2 g\|_F^2, \quad \text{and} \\
\text{Var}_\kappa\left(\sum_{k=1}^d (v_k \partial_k U) - v_k \partial_k g\right) &= \mathbb{E}_\kappa\left(\sum_{k=1}^d (v_k \partial_k U)^2 - (v_k \partial_k g)^2\right) - \left(\mathbb{E}_\kappa\left(\sum_{k=1}^d (v_k \partial_k U) - v_k \partial_k g\right)\right)^2 \\
&= \frac{d}{2} \sum_{k=1}^d (\partial_k U)^2 (\partial_k g)^2 - \frac{1}{4}(\nabla U \cdot \nabla g)^2 \leq \frac{d}{2} \sum_{k=1}^d (\partial_k U)^2 (\partial_k g)^2
\end{aligned}$$

This gives

$$\|\hat{\mathcal{L}}_{\text{tr}}^* \hat{\mathcal{L}}_{\text{tr}}(g \circ \pi) + 2(\mathcal{L}g) \circ \pi\|_{L^2(\hat{\mu})}^2 \leq 2d\|\nabla^2 g\|_F^2 + 4d \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu.$$

### Appendix C. Convergence of Markov processes via lifting and flow Poincaré inequality

It thus remains to bound  $\sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu$  in all three cases. An integration by parts and  $2ab \leq \frac{1}{2}a^2 + 2b^2$  gives

$$\begin{aligned} \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu &= \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k g)^2 \partial_{kk} U d\mu + 2 \sum_{k=1}^d \int_{\mathbb{R}^d} \partial_k U \partial_{kk} g \partial_k g d\mu \\ &\leq \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k g)^2 \partial_{kk} U d\mu + \sum_{k=1}^d \int_{\mathbb{R}^d} \left( \frac{1}{2} (\partial_k U)^2 (\partial_k g)^2 + 2 (\partial_{kk} g)^2 \right) d\mu, \end{aligned}$$

so that rearranging yields

$$\sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu \leq 2 \sum_{k=1}^d \int_{\mathbb{R}^d} \partial_{kk} U (\partial_k g)^2 d\mu + 4 \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_{kk} g)^2 d\mu.$$

Therefore, by Assumption C.2,

$$\sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu \leq 2L \|\nabla g\|_{L^2(\mu)}^2 + 4 \|\nabla^2 g\|_F^2_{L^2(\mu)}. \quad (\text{C.21})$$

Hence, using Bochner's identity, (C.18) and  $\|\nabla g\|_{L^2(\mu)}^2 \leq \frac{1}{m} \|2\mathcal{L}g\|_{L^2(\mu)}^2$ , Assumption (C1) is satisfied with

$$C_1^2 = 4 + 32C_\kappa + \frac{16LC_\kappa + (4 + 32C_\kappa)Km}{m} \leq 44(1 + K) + 20\frac{L}{m}$$

in the Gaussian and hypercube case, and with

$$C_1^2 = 18d + \frac{8dL + 18dKm}{m} = 18d(1 + K) + 8d\frac{L}{m}$$

in the case of uniform distribution on the scaled coordinate directions. One then concludes by Remark C.14.  $\square$

If instead of Assumption C.2 we assume the weaker Assumption C.1, we obtain the same rate as for the Forward process in Corollary C.23, at least for Gaussian or uniform on the hypercube velocities.

**Corollary C.26.** *Let Assumption C.1 be satisfied.*

- (i) *Let  $\kappa = \mathcal{N}(0, I_d)$  or  $\kappa = \text{Unif}(\{-1, +1\}^d)$ . Then, choosing  $T = m^{-1/2}$ , the transition semigroup  $(\hat{P}_t)$  of the Zig-Zag process is exponentially contractive in  $T$ -average with rate*

$$\nu = \Omega \left( \frac{\gamma m}{\gamma^2 + \frac{m(1+K)+b}{(1-a)^2}} \right).$$

*If  $\gamma = \Theta \left( \frac{\sqrt{m(1+K)+b}}{1-a} \right)$  then  $\nu = \Omega \left( (1-a)\sqrt{m} / \sqrt{(1+K)+b/m} \right)$ .*

(ii) Let  $\kappa = \text{Unif}(\{\pm\sqrt{d}e_i : i = 1, \dots, n\})$ . Then, choosing  $T = m^{-1/2}$ , the transition semigroup  $(\hat{P}_t)$  of the Zig-Zag process is exponentially contractive in  $T$ -average with rate

$$\nu = \Omega \left( \frac{\gamma m}{\gamma^2 + \frac{md(1+K)+bd}{(1-a)^2}} \right).$$

$$\text{If } \gamma = \Theta \left( \frac{\sqrt{md(1+K)+bd}}{(1-a)^2} \right) \text{ then } \nu = \Omega \left( (1-a)\sqrt{m} / \sqrt{((1+K)+b/m)d} \right).$$

*Proof.* The only difference to the proof of Corollary C.25 is the treatment of the term  $\sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu$  in the verification of Assumption (C1). We can use (C.20) to get

$$\begin{aligned} \sum_{k=1}^d \int_{\mathbb{R}^d} (\partial_k U)^2 (\partial_k g)^2 d\mu &\leq \| |\nabla U| |\nabla g| \|_{L^2(\mu)}^2 \\ &\leq \frac{2b(1-a) + 4Km}{(1-a)^2} 2\mathcal{E}(g) + \frac{4}{(1-a)^2} \|2\mathcal{L}g\|_{L^2(\mu)}^2 \\ &\leq \frac{1}{(1-a)^2} \left( 4(1+K) + \frac{2b(1-a)}{m} \right) \|2\mathcal{L}g\|_{L^2(\mu)}^2. \end{aligned}$$

Assumption (C1) then follows with

$$\begin{aligned} C_1^2 &= 4 + 4K + \frac{8C_\kappa}{(1-a)^2} \left( 4(1+K) + \frac{2b(1-a)}{m} \right) \\ &\leq \frac{44(1+K)}{(1-a)^2} + \frac{20b}{(1-a)m} \leq \frac{44(1+K)m + 20b}{(1-a)^2 m} \end{aligned}$$

in the Gaussian and hypercube case, and with

$$C_1^2 = 2d + 2dK + \frac{4d}{(1-a)^2} \left( 4(1+K) + \frac{2b(1-a)}{m} \right) \leq \frac{18d(1+K)m + 8db}{(1-a)^2 m}$$

in the case of uniform distribution on the scaled coordinate directions. One then concludes by Remark C.14.  $\square$

## Acknowledgements

A. Eberle and F. Lörler were supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy – GZ 2047/1, Project-ID 390685813. A. Eberle and F. Lörler wurden gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – GZ 2047/1, Projekt-ID 390685813. All the authors acknowledge the support of the French Agence nationale de la recherche under the grant ANR-20-CE46-0007 (SuSa project). A. Guillin and M. Michel are supported by the ANR-23-CE-40003, Conviviality. A. Guillin has benefited from a government grant managed by the Agence Nationale de la Recherche under the France 2030 investment plan ANR-23-EXMA-0001.



# Appendix D.

## Hypocoercivity meets lifts

G. Brigati, F. Lörler and L. Wang. ‘Hypocoercivity meets lifts’. In: *Kinet. Relat. Models* 20.0 (2026), pp. 34–55. DOI: [10.3934/krm.2025020](https://doi.org/10.3934/krm.2025020).

**Abstract:** We unify the variational hypocoercivity framework established by D. Albritton, S. Armstrong, J.-C. Mourrat, and M. Novack [6], with the notion of second-order lifts of reversible diffusion processes, recently introduced by A. Eberle and the second author [99]. We give an abstract, yet fully constructive, presentation of the theory, so that it can be applied to a large class of linear kinetic equations. As this hypocoercivity technique does not twist the reference norm, we can recover accurate and sharp convergence rates in various models. Among those, adaptive Langevin dynamics (ALD) is discussed in full detail and we show that for near-quadratic potentials, with suitable choices of parameters, it is a near-optimal second-order lift of the overdamped Langevin dynamics. As a further consequence, we observe that the Generalised Langevin Equation (GLE) is also a second-order lift, as the standard (kinetic) Langevin dynamics are, of the overdamped Langevin dynamics. Then, convergence of (GLE) cannot exceed ballistic speed, i.e. the square root of the rate of the overdamped regime. We illustrate this phenomenon with explicit computations in a benchmark Gaussian case.

### D.1. Introduction

#### D.1.1. Setting and main goal

In this note, we obtain long-time convergence results for a large class of linear kinetic equations that can be written as

$$\partial_t f + \mathcal{T}f = L_v f, \quad f = f(t, x, v), \quad (t, x, v) \in \mathbb{R}_+ \times \mathcal{X} \times \mathcal{V}. \quad (\text{D.1})$$

Here  $\mathcal{T}$  is the transport operator, and the operator  $L_v$  describes the energy dissipation in the  $v$ -variable, with precise assumptions and examples stated below. Usually  $x$  and  $v$  are interpreted as position and velocity variables, respectively, although the situations we consider are general, and  $\mathcal{X}$  and  $\mathcal{V}$  denote general measurable state spaces that may have different dimensions.

We focus in particular on scenarios where (D.1) has a unique normalised steady state of the form

$$\hat{\mu}(\mathrm{d}x \, \mathrm{d}v) = \mu(\mathrm{d}x) \kappa(x, \mathrm{d}v), \quad (\text{D.2})$$

where  $\hat{\mu}$  is a probability distribution,  $\mu$  is the marginal distribution of  $\hat{\mu}$  on  $\mathcal{X}$ , and  $\kappa$  denotes a Markov kernel. In many examples,  $\hat{\mu}$  is a product measure  $\hat{\mu} = \mu \otimes \kappa$  where  $\kappa$  is independent of  $x$ , but this framework does not require  $\hat{\mu}$  to be a product measure; in fact it can also be applied to nonequilibrium dynamics, where only the existence but not the expression of  $\hat{\mu}$  is known [82]. The primary goal of this note is to highlight the link between variational hypocoercivity and non-reversible lifts, in a unified language, which is extensively explained in Sections D.1.2-D.1.3. Then, we state our main result Theorem D.5 in Section D.1.4, which yields a hypocoercive estimate in the form

$$\|f(t, \cdot, \cdot)\|_{L^2(\hat{\mu})} \leq C e^{-\lambda t} \|f(t=0, \cdot, \cdot)\|_{L^2(\hat{\mu})}, \quad (\text{D.3})$$

for all solutions  $f$  to (D.1) with zero average initial data. We construct a fully explicit, accurate decay rate  $\lambda > 0$ , and an explicit constant  $C > 1$ . In Section D.1.5, we list a few important examples which are covered by our analysis.

### D.1.2. From hypocoercivity to non-reversible lifts

In many cases of interest, the operator  $L_v$  is degenerate dissipative, i.e.  $L_v \leq 0$ , but  $\ker(L_v)$  is larger than  $\ker(L_v - \mathcal{T})$ . In other words, equation (D.1) admits local equilibria which are not global ones. Convergence to global equilibrium still holds, as long as the conservative operator  $\mathcal{T}$  is able to force the solution to (D.1) outside  $\text{Ker } L_v$ . This is the heart of hypocoercivity, stemming from [140, 152], and C. Villani's monograph [248]. To counter the degeneracy of  $L_v$ , a modified  $H^1$ -norm  $\tilde{\mathcal{E}}$  is built in [248], such that  $\tilde{\mathcal{E}} \approx \|\cdot\|_{H^1(dx dv)}$ , and  $\tilde{\mathcal{E}}$  acts as a strict entropy functional for (D.1), i.e. for all solutions  $f$ , and a suitable global equilibrium  $f_\infty$ , we have  $\frac{d}{dt} \tilde{\mathcal{E}}(f(t, \cdot, \cdot) - f_\infty) \leq -\lambda \tilde{\mathcal{E}}(f(t, \cdot, \cdot) - f_\infty)$ , for some  $\lambda > 0$ . Thus, for a constant  $C > 1$ ,

$$\|f(t, \cdot, \cdot) - f_\infty\|_{H^1(dx dv)} \leq C e^{-\lambda t} \|f(0, \cdot, \cdot) - f_\infty\|_{H^1(dx dv)}. \quad (\text{D.4})$$

This scheme has been fruitfully exploited in many works, including [22], where a twisted  $\Gamma_2$  calculus, a combination of [19] and [248], is established. We refer the reader to [29] for a review of the literature. Later on, an  $L^2$ -hypocoercivity approach, inspired by diffusion limits, was outlined by Dolbeault, Mouhot, and Schmeiser in [87], with ideas from earlier works [86, 132]. In this influential work, an abstract framework is laid out with several assumptions that can be directly verified for a large class of kinetic equations (even in presence of non-regularising effects, such as scattering), with a wide range of applications. Inequalities in the form (D.4), and extension of those in case of weak confinement [49, 50, 51], are obtained in the  $L^2$ -norm.

As the  $L^2$ -framework of hypocoercivity provides explicit estimates for the constants  $C > 1, \lambda > 0$ , the problem of finding optimal values is posed. This can be done via mode-by-mode Fourier analysis in special situations [16]. In general, providing good estimates for the rate  $\lambda$  is delicate on its own, optimising  $\lambda$  and  $C$  at the same time is problematic, and the two constants can even compete [3].

This motivates a new *variational* approach to  $L^2$ -hypocoercivity, where the  $L^2$ -norm is not twisted, first established by Albritton, Armstrong, Mourrat and Novack in [6] and

later developed in [53, 55, 56, 61, 82, 83, 186]. The method relies on the *time-averaged*  $L^2$ -energy,

$$\mathcal{H}(t) = \int_t^{t+T} \|f(s, \cdot, \cdot)\|_{L^2(\hat{\mu})}^2 ds,$$

which can be related to its dissipation along (D.1)

$$\mathcal{I}(t) := -\frac{d}{dt}\mathcal{H}(t) = -\int_t^{t+T} \frac{d}{ds} \|f(s, \cdot, \cdot)\|_{L^2(\hat{\mu})}^2 ds,$$

via an adapted Poincaré inequality

$$\mathcal{I}(t) \geq \lambda_T \mathcal{H}(t). \quad (\text{D.5})$$

As in the classical parabolic case, exponential decay for  $\mathcal{H}(t)$  is recovered. This directly implies a hypocoercivity estimate (D.4), see Theorem D.5, with the same rate  $\lambda_T$ . The leading idea is that the time average allows the transport operator to carry relaxation from the  $v$  to the  $x$  variables, which is a non-instantaneous effect, as witnessed also in the kinetic De Giorgi theory [11]. The parameter  $T$  can be tuned to optimise the rate  $\lambda_T$ , without enlarging any other constant.

The adapted Poincaré inequality (D.5) is usually proved in two steps, namely a  $(t, x)$ -Poincaré–Lions inequality for the velocity average  $\Pi f$  defined in (D.6), and an averaging lemma (variants of which can be traced back to at least [118]), relating weak norms of  $\Pi f$  to the dissipation term  $\mathcal{I}(t)$ . It was unclear whether this was the only reasonable path leading to (D.5), until the results of [99]. In that contribution, a new structure of *non-reversible lifts of reversible diffusion processes* is formulated, of which (D.1) is a particular case under our assumptions below. The abstract theory of [99], once rewritten for (D.1), is reminiscent of [6]. A novelty is that the intermediate steps for (D.5) (i.e. the Poincaré–Lions inequality and the averaging lemma) are endowed with a clear interpretation and identified by the lift structure, as highlighted by the proof of Theorem D.5. As an additional result, the work [99] allows to derive both upper and lower abstract bounds for the rate  $\lambda_T$ , which match well with the concrete computations of [53, 61].

### D.1.3. Assumptions and discussion

Following the lift structure, we provide a general framework for the variational approach [6], adapting the notation and the style of presentation of the DMS (Dolbeault–Mouhot–Schmeiser) framework [87].

For any function  $f(t, x, v)$ , we define its projection

$$\Pi f(t, x) := \int_{\mathcal{V}} f(t, x, v) \kappa(x, dv), \quad (\text{D.6})$$

which yields a self-adjoint operator on  $L^2(\hat{\mu})$ , where  $\hat{\mu}$  is given by (D.2). We assume that the operator  $\mathcal{L} = L_v - \mathcal{T}$  is the generator of some Markov process with transition semigroup  $(\hat{P}_t)_{t \geq 0}$ , so that the dynamics (D.1) is mass-preserving. Moreover, we impose the following assumptions on  $\mathcal{T}$ .

**Assumption D.1.**  $\Pi\mathcal{T}\Pi = 0$ .

**Assumption D.2.** The operator  $L_x := -(\mathcal{T}\Pi)^*(\mathcal{T}\Pi) = -\Pi\mathcal{T}^*\mathcal{T}\Pi$  is self-adjoint with respect to  $L^2(\hat{\mu})$  and corresponds to the generator of some reversible diffusion process on  $\mathcal{X}$  with transition semigroup  $(P_t)_{t \geq 0}$ . Moreover, we have the Poincaré inequality (macroscopic coercivity)

$$P_x := \inf_{f \in L_0^2(\hat{\mu}) \setminus \{0\}} \frac{\|\mathcal{T}\Pi f\|_{L^2(\hat{\mu})}^2}{\|\Pi f\|_{L^2(\mu)}^2} > 0, \quad (\text{D.7})$$

or, equivalently,  $L_x$  has spectral gap  $P_x$ .

It is straightforward from the expression  $L_x = -(\mathcal{T}\Pi)^*(\mathcal{T}\Pi)$  that  $L_x$  is self-adjoint with respect to  $L^2(\hat{\mu})$ . Note that  $\mu$  is the  $x$ -marginal of  $\hat{\mu}$ , and  $L_x$  does not involve the  $v$ -variable, thus it makes sense to regard  $L_x$  as a self-adjoint operator on  $L^2(\mu)$ .

**Definition D.1** ([99, Definition 1]). The semigroup  $(\hat{P}_t)_{t \geq 0}$  is a second-order lift of  $(P_{t/2})_{t \geq 0}$  if Assumptions D.1–D.2 are satisfied.

Let us comment on this definition. It was observed by [79] that, in discrete time and space, enlarging the state space of a reversible Markov chain by adding a velocity component and moving to a suitable non-reversible kinetic dynamics can accelerate convergence to equilibrium and reduce the relaxation time by a square root. Motivated by this acceleration, a concept of lifts of Markov chains in discrete time was introduced by [79] and later formalised by [66]. Intuitively, a non-reversible Markov chain with transition kernel  $\hat{p}$  on a product space  $\mathcal{X} \times \mathcal{V}$  with invariant probability measure  $\hat{\mu} = \mu \otimes \kappa$  is a lift of a Markov kernel  $p$  on  $\mathcal{X}$  that is reversible with respect to  $\mu$  if

$$\int_{\mathcal{V}} \hat{p}((x, v), A \times \mathcal{V}) \kappa(x, dv) = p(x, A)$$

for all measurable subsets  $A \subseteq \mathcal{X}$ .

A related yet distinct concept of second-order lifts in continuous time and space was introduced in [99], providing a suitable framework for the study of many non-reversible Markov processes by viewing them as second-order lifts of simpler reversible diffusions. Intuitively, the process on  $\mathcal{X} \times \mathcal{V}$  with transition semigroup  $(\hat{P}_t)_{t \geq 0}$  is a second-order lift of the process on  $\mathcal{X}$  with transition semigroup  $(P_{t/2})_{t \geq 0}$  if, as  $t \downarrow 0$ ,

$$\int_{\mathcal{V}} \hat{P}_t(\Pi f)(x, v) \kappa(x, dv) = P_{(t/2)^2} \Pi f(x) + o(t^2) \quad \text{for all } f \in L^2(\hat{\mu}). \quad (\text{D.8})$$

The different linear and quadratic timescales appearing on the left- and right-hand side are due to the kinetic versus diffusive behaviour of  $(\hat{P}_t)_{t \geq 0}$  and  $(P_t)_{t \geq 0}$ , respectively. This difference leads to the property (D.8) carrying second-order information about the infinitesimal behaviour of  $(\hat{P}_t)_{t \geq 0}$ , which is formalised in Assumptions D.1–D.2. Indeed, a short-time Taylor expansion of (D.8) and a comparison of the first- and second-order terms leads to the first order condition  $\Pi\mathcal{T}\Pi = 0$  of Assumption D.1 and the expression

for  $L_x$  given in Assumption D.2. Precise conditions under which this property is equivalent to Definition D.1 are given in [99, Remark 2 (v)]. However, since this property is only infinitesimal, a reversible diffusion and its lift may behave completely differently for long times. Strikingly, the lift property can still be used to relate convergence to equilibrium of the lift  $(\hat{P}_t)_{t \geq 0}$  to spectral properties of the generator  $L_x$  of  $(P_t)_{t \geq 0}$ , as the following Remark shows.

**Remark D.2** (Lower bounds for lifts). Using the framework of lifts, under Assumptions D.1-D.2 only, we can recover a general lower bound for the convergence rate of (D.1). Denoting the relaxation time of  $(\hat{P}_t)_{t \geq 0}$  as

$$t_{\text{rel}}(\hat{P}) = \inf\{t \geq 0: \|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq e^{-1} \|f\|_{L^2(\hat{\mu})} \text{ for all } f \in L_0^2(\hat{\mu})\},$$

and the singular value gap of the generator of  $(\hat{P}_t)_{t \geq 0}$  as

$$s(\mathcal{L}) = \inf_{f \in L_0^2(\hat{\mu}) \setminus \{0\}} \frac{\|\mathcal{L}f\|_{L^2(\hat{\mu})}}{\|f\|_{L^2(\hat{\mu})}},$$

we can make use of the general bound  $t_{\text{rel}}(\hat{P}) \geq \frac{1}{2} s(\mathcal{L})^{-1}$ , see [99, Lemma 10]. Using Assumption D.2, the singular value gap  $s(\mathcal{L})$  can be bounded from above as

$$s(\mathcal{L}) \leq \inf_{f \in L_0^2(\hat{\mu}) \setminus \{0\}} \frac{\|\mathcal{L}\Pi f\|_{L^2(\hat{\mu})}}{\|\Pi f\|_{L^2(\hat{\mu})}} = \sqrt{P_x},$$

which yields the lower bound

$$t_{\text{rel}}(\hat{P}) \geq \frac{1}{2} P_x^{-1/2}$$

for the relaxation time of the lift. Since the inverse spectral gap  $P_x^{-1}$  of  $L_x$  is exactly the relaxation time of the reversible transition semigroup  $(P_t)_{t \geq 0}$ , this shows that convergence to equilibrium as measured by the relaxation time cannot be accelerated by more than a square root through lifting. In particular, if (D.4) is satisfied for some  $C \geq 1$  and  $\lambda > 0$ , then the asymptotic decay rate  $\lambda$  can be bounded as

$$\lambda \leq (1 + \log C) \sqrt{P_x}.$$

Intuitively, one should think of  $-\mathcal{T}$  as being a square root of  $-L_x$  on an extended state space, so that a square-root lower bound on the relaxation time of the semigroup generated by  $\mathcal{L} = L_v - \mathcal{T}$  is immediately plausible.

The following Assumptions D.3-D.5 are not needed for  $(\hat{P}_t)$  to be a second-order lift of  $(P_t)$ , and therefore only affect the upper bound of the convergence analysis.

**Assumption D.3** ([99, Assumption 1 (ii)]). The operator  $\mathcal{T}$  is non-negative in the sense that for any  $f \in \text{Dom}(\mathcal{T})$ ,

$$\iint_{\mathcal{X} \times \mathcal{V}} f(x, v) \mathcal{T} f(x, v) d\hat{\mu}(x, v) \geq 0. \quad (\text{D.9})$$

#### Appendix D. Hypocoercivity meets lifts

Assumption D.3 holds automatically with equality when  $\mathcal{T}$  is antisymmetric, but follows from a direct calculation in some other examples. We define

$$\|f\|_{L^2(\mu; H^1(\kappa))}^2 := \iint_{\mathcal{X} \times \mathcal{V}} (|f|^2 + |\nabla_v f|^2) d\hat{\mu}(x, v),$$

and  $\|f\|_{L^2(\mu; H^{-1}(\kappa))}$  by  $L^2$ -duality

$$\|f\|_{L^2(\mu; H^{-1}(\kappa))} := \sup_{\|g\|_{L^2(\mu; H^1(\kappa))}=1} \iint_{\mathcal{X} \times \mathcal{V}} fg d\hat{\mu}(x, v).$$

**Assumption D.4.** The operator  $L_v$  is self-adjoint in  $L^2(\hat{\mu})$ . In addition, its nullspace  $\ker(L_v)$  is spanned by functions that only depend on the  $x$ -variable, i.e.

$$\ker(L_v) = \text{Im}(\Pi).$$

Moreover, there exists a constant  $P_v > 0$  such that we have microscopic coercivity [87, Assumption (H1)]:

$$-\langle f, L_v f \rangle_{L^2(\hat{\mu})} \geq P_v \|(\text{Id} - \Pi)f\|_{L^2(\hat{\mu})}^2. \quad (\text{D.10})$$

Finally, there exists a constant  $R > 0$  such that

$$\|L_v f\|_{L^2(\mu; H^{-1}(\kappa))}^2 \leq -R \langle f, L_v f \rangle_{L^2(\hat{\mu})}. \quad (\text{D.11})$$

**Remark D.3** (DMS Hypocoercivity and lifts). Assumptions D.1-D.2 correspond to assumptions (H3) and (H2) of the DMS method [87] respectively. At the same time, Assumptions D.1 and D.2 are precisely the two assumptions in [99] which indicate that  $(\hat{P}_t)_{t \geq 0}$  is a second-order lift of  $(P_{t/2})_{t \geq 0}$  as defined in [99, Definition 1], aside from a domain condition. If we disregard (D.11), Assumption D.4 is precisely [87, Assumption (H1)]. The inequality (D.11) holds true with  $R = \gamma$ , under the most common choices  $L_v = \gamma(\Pi - \text{Id})$  or  $L_v = -\gamma \nabla_v^* \nabla_v$ . The coercivity assumptions (D.7) and (D.10) can be weakened into some weak or weighted version of the Poincaré inequality, see [53, 55, 56, 82].

As for our final assumption, readers familiar with the works [53, 61, 99] will recognise that it resembles the solution of the divergence equation, which is in duality with the aforementioned Poincaré–Lions inequality. There are various ways in the literature to construct such solutions with required regularity estimates, see also [97, 101] and [168].

**Assumption D.5.** Let  $T > 0$ ,  $\lambda(dt) := \frac{1}{T} \mathbf{1}_{[0, T]} dt$  denote the normalised Lebesgue measure on  $[0, T]$ , and denote  $\bar{\mu} := \lambda \otimes \mu$ . Then for any  $g(t, x) \in L^2(\bar{\mu})$  with  $\iint g(t, x) d\bar{\mu}(t, x) = 0$ , there exist functions  $\varphi_0, \varphi_1$  with boundary conditions

$$\varphi_0(0, \cdot) = \varphi_0(T, \cdot) = 0, \quad \mathcal{T}\varphi_1(0, \cdot) = \mathcal{T}\varphi_1(T, \cdot) = 0$$

such that

$$-\partial_t \varphi_0 - L_x \varphi_1 = g. \quad (\text{D.12})$$

Moreover, there exist constants  $C_{0,T}, C_{1,T}$  depending on  $T$  such that  $\varphi_0, \varphi_1$  satisfy the regularity estimates

$$\|\varphi_0 + \mathcal{T}\varphi_1\|_{L^2(\bar{\mu}; H^1(\kappa))} \leq C_{0,T} \|g\|_{L^2(\bar{\mu})}, \quad (\text{D.13})$$

and

$$\|(\partial_t - \mathcal{T}^*)(\varphi_0 + \mathcal{T}\varphi_1)\|_{L^2(\lambda \otimes \hat{\mu})} \leq C_{1,T} \|g\|_{L^2(\bar{\mu})}. \quad (\text{D.14})$$

The equation (D.12) is a time-dependent generalisation of the Poisson equation, which in the classical elliptic setting would be a direct consequence of the Poincaré inequality. Here, due to the moving time window, we need to allow functions to depend on time and impose Dirichlet boundary conditions.

In practice, for most of the examples we consider,  $\kappa$  only depends on  $v$  and is standard Gaussian, thus Assumption D.5 can be verified since any finite moments of  $v$  are bounded, and the estimates (D.13), (D.14) boil down to

$$\|\varphi_0\|_{H^1(\bar{\mu})} + \|\partial_t \varphi_1\|_{H^1(\bar{\mu})} + \|\nabla_x \varphi_1\|_{H^1(\bar{\mu})} \lesssim \|g\|_{L^2(\bar{\mu})}. \quad (\text{D.15})$$

In previous works [53, 61, 99, 168], this can be achieved if  $\mu$  satisfies a Poincaré inequality (D.7), and for  $d\mu(x) \propto \exp(-U(x)) dx$ , the Hessian of  $U$  satisfies a (not necessarily positive) lower bound. In fact, since (D.12) only depends on  $L_x$ , any lift of  $L_x$  shares the same estimate (D.15). Then only the terms in (D.13) and (D.14) need to be computed for a specific  $\mathcal{T}$ . Additionally, one may notice from previous works that the verification of Assumption D.5 is the only place where macroscopic coercivity (D.7) is used. We nevertheless write the two assumptions separately as (D.7) is crucial for the definition of lifts and obtaining convergence lower bounds.

**Remark D.4** (Variational hypocoercivity and DMS). In light of Remark D.3, it is natural to compare Assumption D.5 with [87, Assumption (H4)]. In [87, Section 2], the authors show that [87, Assumption (H4)] is related to elliptic regularity theory. On the other side, Assumption D.5 requires regularity for a divergence equation, which can be completed to an elliptic one [53, 61]. The two assumptions differ in their spirit, though, since [87, Assumption (H4)] is used to control error terms arising from a modified  $L^2$ -norm, whereas the bounds (D.13)-(D.14) arise from the lift structure via an integration by parts.

#### D.1.4. Main result

Under the Assumptions D.1–D.5, we can prove exponential convergence of the solutions of (D.1) to equilibrium.

**Theorem D.5.** *Let  $f(t, x, v)$  be the solution of (D.1) with initial condition  $f(t = 0, x, v) = f_0(x, v) \in L^2(\hat{\mu})$ . Then as  $t \rightarrow \infty$ ,  $f_t$  converges exponentially fast to  $(f_0) := \int f_0(x, v) d\hat{\mu}(x, v)$ . More precisely, there exist constants  $C > 1, \lambda > 0$ , depending on constants in Assumptions D.1–D.5, such that*

$$\|f(t, \cdot, \cdot) - (f_0)\|_{L^2(\hat{\mu})} \leq C e^{-\lambda t} \|f_0 - (f_0)\|_{L^2(\hat{\mu})}.$$

## Appendix D. Hypocoercivity meets lifts

In particular, suppose there exists some  $T_* > 0$  such that

$$T_* = \arg \min_{T > 0} \{C_{1,T} + C_{0,T} \sqrt{RP_v}\},$$

then we have  $C = \exp(T_* \lambda_{T_*})$ , with the optimised choice of convergence rate

$$\lambda_{T_*} = \frac{2P_v}{1 + (C_{1,T_*} + C_{0,T_*} \sqrt{RP_v})^2}. \quad (\text{D.16})$$

A major strength of our framework is that it yields sharp convergence rates for a large number of models. In Example (i) below, we discuss why our framework yields sharp convergence rate for Langevin dynamics that is optimal up to a universal constant.

### D.1.5. Examples

Our main running examples of this work are second-order lifts of overdamped Langevin diffusions on  $\mathcal{X} = \mathbb{R}^d$ , given by the solution  $(X_t)_{t \geq 0}$  to the stochastic differential equation

$$dX_t = -\nabla U(X_t) dt + \sqrt{2} dB_t. \quad (\text{D.17})$$

Here  $(B_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion. The invariant probability measure of (D.17) is the Boltzmann-Gibbs measure  $d\mu(x) \propto \exp(-U(x)) dx$ , and the generator of  $(P_t)$  in  $L^2(\mu)$  can be written as  $L_x = \Delta_x - \nabla U \cdot \nabla_x = -\nabla_x^* \nabla_x$ . Therefore, (D.7) essentially amounts to a (super)linear growth of  $U$  at infinity, see [20, Proposition 4.4.2]. Below are some of the dynamics that are covered in this work, which are all second-order lifts of overdamped Langevin dynamics. For examples (i)–(iii) below, we know from [61, 186] that the variational hypocoercivity approach is applicable.

(i) *Langevin dynamics* is the solution  $(X_t, V_t)_{t \geq 0}$  to the stochastic differential equation

$$\begin{cases} dX_t = V_t dt \\ dV_t = -\nabla U(X_t) dt - \gamma V_t dt + \sqrt{2\gamma} dW_t, \end{cases} \quad (\text{D.18})$$

with a standard  $d$ -dimensional Brownian motion  $(W_t)_{t \geq 0}$ . Here  $\gamma > 0$  is the friction parameter, the state spaces are  $\mathcal{X} = \mathcal{V} = \mathbb{R}^d$ , and the corresponding operators write  $\mathcal{T} = -v \cdot \nabla_x + \nabla U \cdot \nabla_v$  and  $L_v = -\gamma \nabla_v^* \nabla_v$ . In this case,  $\kappa = \kappa(dv) = \mathcal{N}(0, \text{Id})$  is independent of  $x$  and  $\hat{\mu} = \mu \otimes \kappa$  is the product measure. Langevin dynamics can also be seen as a perturbation of the deterministic Hamiltonian dynamics  $\frac{dX_t}{dt} = V_t$ ,  $\frac{dV_t}{dt} = -\nabla U(X_t)$  by an Ornstein-Uhlenbeck process with intensity  $\gamma$  in the velocity component.

Langevin dynamics should serve as the primary example for readers to understand the framework of hypocoercivity and lifts, and for readers' convenience, let us use this example to illustrate how the framework could be applied to the Langevin dynamics, which was already achieved in [61]. To verify Assumption D.1, since  $\Pi f$  is independent of  $v$ , we have

$$\mathcal{T} \Pi f = v \cdot \nabla_x \Pi f$$

and the assumption is true since  $\int_{\mathbb{R}^d} v \, d\kappa(v) = 0$ . Additionally, using the fact that  $\int_{\mathbb{R}^d} v_i v_j \, d\kappa(v) = \delta_{ij}$ , one may compute

$$\langle \mathcal{T}\Pi f, \mathcal{T}\Pi g \rangle_{L^2(\hat{\mu})} = \langle v \cdot \nabla_x \Pi f, v \cdot \nabla_x \Pi g \rangle_{L^2(\hat{\mu})} = \langle \nabla_x \Pi f, \nabla_x \Pi g \rangle_{L^2(\mu)}.$$

Since such equality holds for arbitrary  $f, g$ , we obtain  $-L_x = (\mathcal{T}\Pi)^*(\mathcal{T}\Pi) = \nabla_x^* \nabla_x \Pi$ , which means Assumption D.2 is also satisfied as long as  $\mu$  satisfies a Poincaré inequality. Assumption D.3 is satisfied automatically since  $\mathcal{T}$  is antisymmetric. For this example  $L_v = -\gamma \nabla_v^* \nabla_v$ , the condition  $\ker(L_v) = \text{Im}(\Pi)$  holds since both consist precisely of functions that depend only on  $x$ , and Assumption D.4 then holds with  $R = P_v = \gamma$ . Finally, as we discussed after Assumption D.5, Assumption D.5 can be proved, and we refer the readers to [53, 61, 101, 168] for different proofs. Hence, Theorem D.5 as well as the lower bound in Remark D.2 can be directly applied.

To verify the optimality of the convergence rate in Theorem D.5, if  $U$  is convex and has Poincaré constant  $P_x$ , then the proof in [101] gives  $C_{0,T} = C_0(T + \frac{1}{\sqrt{P_x}})$  and  $C_{1,T} = C_1(1 + \frac{1}{T\sqrt{P_x}})$  where  $C_0, C_1$  denote absolute constants which can be made explicit. Substituting these into (D.16), we get

$$\lambda \sim \frac{2\gamma}{\inf_{T>0} \left(1 + \frac{1}{T\sqrt{P_x}} + T\gamma + \frac{\gamma}{\sqrt{P_x}}\right)^2}.$$

Optimising first in  $T_* \sim P_x^{-1/4} \gamma^{-1/2}$  and then in  $\gamma \sim \sqrt{P_x}$ , we have  $\lambda_{T_*} \sim \sqrt{P_x}$  which recovers the result in [61, 99], and thanks to Remark D.2, the convergence rate is optimal up to an absolute constant. The constant  $C = \exp(\lambda_{T_*} T_*)$  is bounded above by an absolute constant.

- (ii) *Randomised Hamiltonian Monte Carlo* (RHMC) [47] is obtained by interspersing the Hamiltonian dynamics  $\frac{dX_t}{dt} = V_t$ ,  $\frac{dV_t}{dt} = -\nabla U(X_t)$  with complete velocity refreshments according to  $\kappa = \mathcal{N}(0, \text{Id})$  after exponentially distributed waiting times. Here the choice of  $\mathcal{X}$ ,  $\mathcal{V}$  and  $\mathcal{T}$  is identical to case (i) but  $L_v = \gamma(\Pi - \text{Id})$ . The corresponding backward Kolmogorov equation (D.1) is also known as the *linear Boltzmann equation*.
- (iii) The *Zig-Zag process* [34] runs in straight lines, interspersed with flips of one of the components of the velocity at random times. Ergodicity of the Zig-Zag process is known without velocity refreshment [36] but refreshment as in the case of RHMC can still be added to obtain quantitative convergence upper bounds. The operators have the form  $\mathcal{T} = -v \cdot \nabla_x - \sum_{k=1}^d (v_k \partial_{x_k} U)_+ (\mathcal{B}_k - \text{Id})$  where the  $\mathcal{B}_k$  are reflection operators  $\mathcal{B}_k f(x, v) = f(x, v - 2v_k e_k)$  with  $e_k$  being the  $k$ -th cartesian coordinate, and  $L_v = \gamma(\Pi - \text{Id})$ . The notable feature in this example is that, unlike the previous two examples,  $\mathcal{T}$  is not antisymmetric in  $L^2(\hat{\mu})$ . Nevertheless, Assumptions D.1–D.4 can all be verified. Various other piecewise deterministic Markov processes introduced in the MCMC literature [48, 197] also fit into the scope of this framework.

#### Appendix D. Hypocoercivity meets lifts

- (iv) One example, which this hypocoercivity framework has not yet been applied to, is the *adaptive Langevin* dynamics [84, 145]

$$\begin{cases} dQ_t = V_t dt, \\ dV_t = -\nabla U(Q_t) dt - \frac{Z_t}{\varepsilon} V_t dt - \gamma V_t dt + \sqrt{2\gamma} dW_t, \\ dZ_t = \frac{1}{\varepsilon} (|V_t|^2 - d) dt, \end{cases} \quad (\text{D.19})$$

where  $(W_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion. The motivation for such dynamics is that in practice, the computation of  $\nabla U$  is challenging and may be effectively corrupted, which may lead to a significant bias in the invariant measure when using the Langevin dynamics (D.18). Hence we define  $Z_t$  to be the friction correction variable which is defined by a negative feedback loop control law (as in the Nosé–Hoover method [210]), so that  $|V_t|^2$  is encouraged to stay around  $d$ , and the true friction  $\frac{Z_t}{\varepsilon} + \gamma$  may adjust with different noise levels of the gradient without knowing the strength of the gradient noise. We refer to [172] for a more comprehensive motivation of adaptive Langevin dynamics.

The correct way to view this example is that it is a second-order lift of the overdamped Langevin dynamics jointly in  $(q, z)$ . In particular, let  $x = (q, z) \in \mathbb{R}^d \times \mathbb{R} = \mathcal{X}$  and  $v \in \mathbb{R}^d = \mathcal{V}$ , the invariant measure of (D.19) writes

$$\begin{aligned} d\mu(q, z) &\propto e^{-U(q) - \frac{|z|^2}{2}} dq dz, \quad d\kappa(v) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{|v|^2}{2}} dv, \\ d\hat{\mu}(q, v, z) &= d\mu(q, z) \otimes d\kappa(v). \end{aligned}$$

The corresponding operators of (D.19) write [172, (2.5)]

$$-\mathcal{T} = \nabla_v^* \nabla_q - \nabla_q^* \nabla_v + \frac{1}{\varepsilon} \mathcal{L}_{\text{NH}}, \quad L_v = -\gamma \nabla_v^* \nabla_v \quad (\text{D.20})$$

with the  $\mathcal{L}_{\text{NH}}$  part resembling the Nosé–Hoover feedback control

$$\mathcal{L}_{\text{NH}} = (|v|^2 - d) \partial_z - zv \cdot \nabla_v = (\partial_z - \partial_z^*) \nabla_v^* \nabla_v + \Delta_v^* \partial_z - \Delta_v \partial_z^* \quad (\text{D.21})$$

with  $\partial_z^* = -\partial_z + z$ , so that  $\mathcal{L}_{\text{NH}}$ , and therefore  $\mathcal{T}$ , is antisymmetric. A straightforward computation yields

$$-L_x = (\mathcal{T}\Pi)^*(\mathcal{T}\Pi) = \nabla_q^* \nabla_q + \frac{2d}{\varepsilon^2} \partial_z^* \partial_z,$$

which, by a rescaling, is the generator of an overdamped Langevin diffusion in  $(q, z)$  variables

$$d \begin{pmatrix} Q_t \\ Z_t \end{pmatrix} = - \begin{pmatrix} \nabla U(Q_t) \\ Z_t \end{pmatrix} dt + \sqrt{2} dB_t,$$

where  $(B_t)_{t \geq 0}$  is a standard  $(d+1)$ -dimensional Brownian motion. In particular, by tensorisation, the Poincaré inequality (D.7) holds true as long as  $\nabla_q^* \nabla_q$  has

a spectral gap. Although exponential convergence has been shown in [135, 172], in Section D.3 we will apply this framework to obtain sharper convergence upper bounds. Furthermore, we will determine the choice of the parameters  $\varepsilon, \gamma$  that optimises the convergence rate.

- (v) The generalised Langevin dynamics is the Markov process  $(X_t, V_t, Z_t)_{t \geq 0}$  on  $\mathcal{X} \times \mathcal{V}$  with  $\mathcal{X} = \mathbb{R}^d$  and  $\mathcal{V} = \mathbb{R}^d \times \mathbb{R}^d$  given by the solution to the stochastic differential equation

$$\begin{cases} dX_t = V_t dt \\ dV_t = -\nabla U(X_t) dt + \lambda Z_t dt, \\ dZ_t = -\lambda V_t dt - \gamma Z_t dt + \sqrt{2\gamma} dW_t, \end{cases} \quad (\text{D.22})$$

where  $(W_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion and  $\lambda, \gamma > 0$ . It arises as a quasi-Markovian approximation of a more general model for the motion of a particle interacting with a heat bath at equilibrium originating in molecular dynamics [214]. It was also considered independently by computer scientists in [200] as a third-order Langevin dynamics. Equation (D.22) is known as the generalised Langevin equation (GLE), and hypocoercivity for this dynamics has been shown in [212, 214]. Its invariant probability measure is

$$\hat{\mu}(dx dv dz) = \mu(dx) \kappa(dv dz) \quad (\text{D.23})$$

with  $d\mu(x) \propto e^{-U(x)} dx$  and  $\kappa = \mathcal{N}(0, I_d) \otimes \mathcal{N}(0, I_d)$ . The generator of the associated transition semigroup on  $L^2(\hat{\mu})$  is  $\mathcal{L} = L_v - \mathcal{T}$  with the transport and dissipation operators

$$-\mathcal{T} = v \cdot \nabla_x - \nabla U \cdot \nabla_v + \lambda(z \cdot \nabla_v - v \cdot \nabla_z)$$

and

$$L_v = \gamma(-z \cdot \nabla_z + \Delta_z) = -\gamma \nabla_z^* \nabla_z,$$

respectively, where the adjoint is with respect to  $\kappa$ . In this example, the role of  $v$  is played by  $(v, z)$ , and  $L_v$  in fact only acts on  $z$ , so that we have an additional level of degeneracy.

The work of [61] finds that, despite the degeneracy in ellipticity, Langevin dynamics (D.18) accelerates the convergence rate by (up to a universal constant) a square root if the potential is convex, which is particularly relevant in case  $P_x$  is very small. This raises the question of whether generalised Langevin dynamics, which adds another level of elliptic degeneracy, can converge with even more acceleration than a square root. In Section D.4, we will give a negative answer to this question by showing the rather surprising observation that GLE is also a second-order lift of an overdamped Langevin diffusion. Meanwhile, due to the extra degeneracy in terms of ellipticity compared to examples (i)–(iv), GLE violates Assumption D.4, since  $L_v$  is only coercive in the  $z$ -variable. Hence, establishing a sharp upper bound for its convergence rate using the framework covered in this work is much more involved. This is an ongoing work of the authors.

- (vi) Boundary conditions can also be incorporated into this framework. For example, an application to run-and-tumble particles confined by hard walls is given in [97]. In this case, the domain of  $\mathcal{T}$  incorporates boundary conditions, and the stochastic process associated to  $L_x$  is a sticky Brownian motion. This necessitates a careful treatment of the associated domains in the divergence equation in Assumption D.5, which we will not cover here.
- (vii) The equilibrium distribution  $\hat{\mu}$  need not be separable, i.e. a product measure. Such situations can for example arise naturally in case  $(\mathcal{X}, g)$  is a Riemannian manifold. Choosing  $\kappa(x, \cdot) = \mathcal{N}(0, g(x))$  allows one to study generalisations of many of the dynamics introduced above to Riemannian manifolds and to interpret them as lifts of an overdamped Riemannian Langevin diffusion, see [99, Example 4] and [101]. Such processes include, for example, Riemannian Hamiltonian Monte Carlo [116] or Riemannian Langevin dynamics [101]. Non-separable Hamiltonians may also appear due to physical or numerical purposes, see for example [127, Section XIII.10], [107] and [177].
- (viii) Finally, let us consider an example, inspired by mathematical physics, where the velocity of  $x$  lies on a Riemannian manifold with non-zero curvature. Here we follow [85] and consider a lay-down process for flexible fibres. Let  $x \in \mathcal{X} = \mathbb{R}^2$ , and denote  $\tau(v) = (\sin v, \cos v)$  for  $v \in \mathcal{V} = \mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$  to be the velocity of  $x$ . Here note that  $v$  is the *angular* variable of the fibre velocity. The random process of creation for a fibre  $t \mapsto x_t$  on a stationary conveyor belt is given by

$$\begin{cases} dx_t = \tau(v_t) dt, \\ dv_t = -\tau(v_t)^\perp \cdot \nabla_x U(x_t) dt + dW_t, \end{cases} \quad (\text{D.24})$$

where the second equation encodes the interplay between a deterministic force, driving the lay-down point to equilibrium, and turbulence, corresponding to the Brownian motion  $(W_t)_t$  on  $\mathbb{T}$ . Despite  $x$  and  $v$  having different dimensions, the analysis is almost identical to Examples (i)–(iii); whenever  $\mu(dx) = e^{-U(x)} dx$  satisfies the Poincaré inequality, the Assumptions D.1–D.5 are verified. Thus, we get exponential convergence to the equilibrium  $\hat{\mu}(dx, dv) = \mu(dx) dv$ , as it was done first—via the technique of [87]—in [85].

We outline how the rest of the paper is organised: In Section D.2 we present the proof of Theorem D.5 under the general framework using Assumptions D.1–D.5. In Section D.3, we provide a new application of the framework to adaptive Langevin dynamics and obtain sharper convergence results compared to [135, 172]. Finally, in Section D.4, we present the observation that GLE is also a second-order lift of (D.17) and illustrate the resulting lower bound in a benchmark Gaussian case.

## D.2. The general convergence proof

The key intermediate step that follows the solution of the divergence equation is the space-time-velocity Poincaré-type inequality, which is also known as the averaging lemma.

Despite  $\hat{\mu}$  potentially being non-separable, it is still convenient to write  $\hat{\mu} = \mu \otimes \kappa$  or  $\lambda \otimes \hat{\mu} = \bar{\mu} \otimes \kappa$ , since we are treating the time and space variables together, and functions may have different regularities in the space-time and velocity variables.

**Lemma D.6.** *Under Assumptions D.1, D.2, D.5, for any  $f(t, x, v) \in L^2(\bar{\mu} \otimes \kappa)$  with  $\iint f d(\bar{\mu} \otimes \kappa) = 0$ , we have the functional inequality*

$$\|\Pi f\|_{L^2(\bar{\mu})} \leq C_{1,T} \|(\text{Id} - \Pi)f\|_{L^2(\bar{\mu} \otimes \kappa)} + C_{0,T} \|(\partial_t + \mathcal{T})f\|_{L^2(\bar{\mu}; H^{-1}(\kappa))}. \quad (\text{D.25})$$

Here  $C_{0,T}, C_{1,T}$  are the same constants that appear in Assumption D.5.

*Proof.* By assumption, we can find  $\varphi_0, \varphi_1$  depending only on  $t, x$  which satisfy the divergence equation (D.12) with  $\Pi f$  playing the role of  $g$ . The key calculation is the following, which, in addition to [61, 99], can also be found in [29, Proposition A.1]:

$$\begin{aligned} \|\Pi f\|_{L^2(\bar{\mu})}^2 &\stackrel{(\text{D.12})}{=} \langle \Pi f, -\partial_t \varphi_0 - \mathcal{L}_x \varphi_1 \rangle_{L^2(\bar{\mu})} \\ &= \langle \Pi f, -\partial_t \varphi_0 + (\mathcal{T}\Pi)^*(\mathcal{T}\Pi)\varphi_1 \rangle_{L^2(\bar{\mu})} \\ &= \langle \Pi f, (-\partial_t + (\mathcal{T}\Pi)^*)(\varphi_0 + \mathcal{T}\varphi_1) \rangle_{L^2(\bar{\mu} \otimes \kappa)} \\ &= \langle (\partial_t + \mathcal{T})\Pi f, \varphi_0 + \mathcal{T}\varphi_1 \rangle_{L^2(\bar{\mu} \otimes \kappa)} \\ &= \langle (\partial_t + \mathcal{T})f, \varphi_0 + \mathcal{T}\varphi_1 \rangle_{L^2(\bar{\mu} \otimes \kappa)} \\ &\quad - \langle (\partial_t + \mathcal{T})(\text{Id} - \Pi)f, \varphi_0 + \mathcal{T}\varphi_1 \rangle_{L^2(\bar{\mu} \otimes \kappa)}, \end{aligned} \quad (\text{D.26})$$

where the second and third equalities are due to Assumption D.2 and D.1 respectively. The first term on the right-hand side can be controlled using the Cauchy-Schwarz inequality

$$\begin{aligned} \langle (\partial_t + \mathcal{T})f, \varphi_0 + \mathcal{T}\varphi_1 \rangle_{L^2(\bar{\mu} \otimes \kappa)} &\leq \|(\partial_t + \mathcal{T})f\|_{L^2(\bar{\mu}; H^{-1}(\kappa))} \|\varphi_0 + \mathcal{T}\varphi_1\|_{L^2(\bar{\mu}; H^1(\kappa))} \\ &\stackrel{(\text{D.13})}{\leq} C_{0,T} \|(\partial_t + \mathcal{T})f\|_{L^2(\bar{\mu}; H^{-1}(\kappa))} \|\Pi f\|_{L^2(\bar{\mu})}. \end{aligned} \quad (\text{D.27})$$

The second term of (D.26) can be controlled by

$$\begin{aligned} & - \langle (\partial_t + \mathcal{T})(\text{Id} - \Pi)f, \varphi_0 + \mathcal{T}\varphi_1 \rangle_{L^2(\bar{\mu} \otimes \kappa)} \\ &= \langle (\text{Id} - \Pi)f, (\partial_t - \mathcal{T}^*)(\varphi_0 + \mathcal{T}\varphi_1) \rangle_{L^2(\bar{\mu} \otimes \kappa)} \\ &\leq \|(\text{Id} - \Pi)f\|_{L^2(\bar{\mu} \otimes \kappa)} \|(\partial_t - \mathcal{T}^*)(\varphi_0 + \mathcal{T}\varphi_1)\|_{L^2(\bar{\mu} \otimes \kappa)} \\ &\stackrel{(\text{D.14})}{\leq} C_{1,T} \|(\text{Id} - \Pi)f\|_{L^2(\bar{\mu} \otimes \kappa)} \|\Pi f\|_{L^2(\bar{\mu})}. \end{aligned} \quad (\text{D.28})$$

We finish the proof by combining (D.26), (D.27) and (D.28).  $\square$

We are now ready to obtain exponential convergence for solutions of (D.1). The proof of Theorem D.5 using Lemma D.6 is well-established by now, see e.g. [6, 53, 61, 99]. We include a proof for completeness and the convenience of the reader.

## Appendix D. Hypocoercivity meets lifts

*Proof of Theorem D.5.* Without loss of generality, we assume  $(f_0) = 0$ , which implies  $(f_t) = 0$  for any  $t > 0$  as the equation (D.1) preserves mass. Define the time-averaged  $L^2$ -energy

$$\mathcal{H}(t) = \int_t^{t+T} \|f(s, \cdot, \cdot)\|_{L^2(\hat{\mu})}^2 ds. \quad (\text{D.29})$$

By a direct energy estimate, using (D.9)

$$\begin{aligned} \frac{d\mathcal{H}(t)}{dt} &= 2 \int_t^{t+T} \iint_{\mathcal{X} \times \mathcal{V}} f(\mathbf{L}_v - \mathcal{T}) f d\hat{\mu}(x, v) ds \\ &\stackrel{(\text{D.9})}{\leq} 2 \int_t^{t+T} \iint_{\mathcal{X} \times \mathcal{V}} f \mathbf{L}_v f d\hat{\mu}(x, v) ds \leq 0. \end{aligned} \quad (\text{D.30})$$

Meanwhile, applying Lemma D.6 on  $f$  between time  $t$  and  $t + T$  yields

$$\begin{aligned} \int_t^{t+T} \|\Pi f\|_{L^2(\hat{\mu})}^2 ds &\stackrel{(\text{D.25})}{\leq} \left( C_{1,T} \|(\text{Id} - \Pi)f\|_{L^2(\bar{\mu} \otimes \kappa)} + C_{0,T} \|(\partial_t + \mathcal{T})f\|_{L^2(\bar{\mu}; H^{-1}(\kappa))} \right)^2 \\ &\stackrel{(\text{D.1}), (\text{D.10}), (\text{D.11})}{\leq} - \left( \frac{C_{1,T}}{\sqrt{P_v}} + C_{0,T} \sqrt{R} \right)^2 \int_t^{t+T} \iint_{\mathcal{X} \times \mathcal{V}} f \mathbf{L}_v f d\hat{\mu}(x, v) ds, \end{aligned}$$

which by an  $L^2$  orthogonal decomposition leads to

$$\begin{aligned} \int_t^{t+T} \|f\|_{L^2(\hat{\mu})}^2 ds &= \int_t^{t+T} \|(\text{Id} - \Pi)f\|_{L^2(\hat{\mu})}^2 ds + \int_t^{t+T} \|\Pi f\|_{L^2(\hat{\mu})}^2 ds \\ &\stackrel{(\text{D.10})}{\leq} - \left( \frac{1}{P_v} + \left( \frac{C_{1,T}}{\sqrt{P_v}} + C_{0,T} \sqrt{R} \right)^2 \right) \int_t^{t+T} \iint_{\mathcal{X} \times \mathcal{V}} f \mathbf{L}_v f d\hat{\mu}(x, v) ds. \end{aligned}$$

Substituting this into (D.30), and we obtain

$$\frac{d\mathcal{H}(t)}{dt} \leq \frac{-2}{\frac{1}{P_v} + \left( \frac{C_{1,T}}{\sqrt{P_v}} + C_{0,T} \sqrt{R} \right)^2} \mathcal{H}(t),$$

which directly gives  $\mathcal{H}(t) \leq e^{-\lambda t} \mathcal{H}(0)$  with  $\lambda$  given in (D.16). We finish the proof as

$$\|f(t)\|_{L^2(\hat{\mu})}^2 \leq \mathcal{H}(t - T) \leq e^{-\lambda(t-T)} \mathcal{H}(0) \leq e^{-\lambda(t-T)} \|f_0\|_{L^2(\hat{\mu})}^2.$$

□

### D.3. Example: Hypocoercivity of adaptive Langevin dynamics

From our discussions in Section D.1, it suffices to verify Assumption D.5, since other assumptions are straightforward to check. Furthermore, from [53, 61], it is already clear that the divergence equation (D.12) has a solution  $(\varphi_0, \varphi_1)$  that satisfies the desired boundary conditions as well as  $\varphi_0, \partial_t \varphi_1, \nabla_x \varphi_1 \in H^1(\bar{\mu})$ . Hence the estimates (D.13) and

(D.14) must hold since their left-hand sides only involve at most first derivatives of  $\varphi_0$ , first and second derivatives of  $\varphi_1$ , and up to eighth moments of  $v$  against the standard Gaussian. Hence the bulk of the section is dedicated to computing the constants  $C_{0,T}$  and  $C_{1,T}$  explicitly.

**Assumption D.6.** Henceforth, the potential  $U(q)$  is assumed to be such that

- (1)  $\nabla^2 U \geq -M \text{Id}$ , for some  $M \geq 0$ ,
- (2)  $\Delta U(q) \leq Ld + a|\nabla U|^2$ , for some  $L > 0$  and  $a \in (0, \frac{1}{2})$ ,
- (3) the spectrum of the operator  $\nabla_q^* \nabla_q$  is discrete, with positive spectral gap  $P_q > 0$ .

Assumption D.6, in particular the assumption of discrete spectrum, can be relaxed, as in [53, 55]. The second part of the assumption can be found in various aforementioned works, for example [61, Assumption 2] and [29, (3.5)], and is needed for the second inequality of (D.34).

### Solving the divergence equation

In view of Assumption D.5, we need to solve

$$-\partial_t \varphi_0 + \left( \nabla_q^* \nabla_q + \frac{2d}{\varepsilon^2} \partial_z^* \partial_z \right) \varphi_1 = g,$$

for  $g \in L^2(\bar{\mu})$  with  $\iint g(t, x) d\bar{\mu}(t, x) = 0$ . We adapt the strategy of [101], after [53, 61]. Let  $\tilde{z} = cz$ , for a constant  $c$  to be adjusted, then the original generator reads

$$\frac{2d}{\varepsilon^2} (-\partial_{zz} + z\partial_z) = -\frac{2dc^2}{\varepsilon^2} \partial_{\tilde{z}\tilde{z}} + \frac{2d}{\varepsilon^2} \tilde{z}\partial_{\tilde{z}}.$$

Since the rescaled equilibrium in  $\tilde{z}$  is  $e^{-\frac{|\tilde{z}|^2}{2c^2}}$ , we know that<sup>1</sup>  $\partial_{\tilde{z}}^\dagger \partial_{\tilde{z}} = -\partial_{\tilde{z}\tilde{z}} + \frac{1}{c^2} \tilde{z}\partial_{\tilde{z}}$ , so choosing  $c = \frac{\varepsilon}{\sqrt{2d}}$  leads to  $\frac{2d}{\varepsilon^2} (-\partial_{zz} + z\partial_z) = \partial_{\tilde{z}}^\dagger \partial_{\tilde{z}}$ . We therefore recover the equations as stated in [101, Theorem 5] and the following estimates by a simple change of variables.

**Proposition D.7.** In the context of (D.19), for all  $g \in L^2(\bar{\mu})$  with  $\iint g(t, x) d\bar{\mu} = 0$ , equation (D.12) admits a solution  $(\varphi_0, \varphi_1)$  with

$$\|\varphi_0\|_{L^2(\bar{\mu})}^2 + \|\partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{2d}{\varepsilon^2} \|\partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 \leq c_0 \|g\|_{L^2(\bar{\mu})}^2 \quad (\text{D.31})$$

$$\begin{aligned} & \|\nabla_{t,q} \varphi_0\|_{L^2(\bar{\mu})}^2 + \|\nabla_{t,q} \nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 \\ & + \frac{2d}{\varepsilon^2} \left( \|\partial_z \varphi_0\|_{L^2(\bar{\mu})}^2 + \|\nabla_{t,q} \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{2d}{\varepsilon^2} \|\partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \right) \leq c_1 \|g\|_{L^2(\bar{\mu})}^2 \end{aligned} \quad (\text{D.32})$$

with

$$\begin{aligned} P_x &= \min \left( P_q, \frac{2d}{\varepsilon^2} \right), \\ c_0 &= (2T^2 + 43P_x^{-1}), \quad c_1 = \left( 290 + \frac{991}{T^2} P_x^{-1} + 43 \max \left( P_x^{-1}, \frac{T^2}{\pi^2} \right) M \right). \end{aligned} \quad (\text{D.33})$$

<sup>1</sup>Here  $\partial_{\tilde{z}}^\dagger$  denotes the  $L^2$ -adjoint of  $\partial_{\tilde{z}}$  with respect to the tilted measure  $e^{-\frac{|\tilde{z}|^2}{2c^2}}$  in the  $\tilde{z}$  variable.

### Estimates on the convergence rates

We start with the moment computations on  $\kappa$  which will be used multiple times in the calculations. Since  $\int |v|^{2k} d\kappa(v) = \prod_{i=0}^{k-1} (d+2i) = d(d+2) \cdots (d+2k-2)$  for any positive integer  $k$ , we have

$$\begin{aligned} \int v_1^2 d\kappa(v) &= 1, & \int v_1^4 d\kappa(v) &= 3, & \int v_1^6 d\kappa(v) &= 15, \\ \int v_1^8 d\kappa(v) &= 105, & \int (|v|^2 - d)^2 d\kappa(v) &= 2d, & \int (|v|^2 - d)^3 d\kappa(v) &= 8d, \\ & & \int (|v|^2 - d)^4 d\kappa(v) &= 12d^2 + 48d. \end{aligned}$$

Let us first estimate  $C_{0,T}$  from the expression in (D.13):

$$\begin{aligned} \|\varphi_0 + \mathcal{T}\varphi_1\|_{L^2(\bar{\mu}, H^1(\kappa))}^2 &= \|\varphi_0 + \mathcal{T}\varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 + \|\nabla_v(\varphi_0 + \mathcal{T}\varphi_1)\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 \\ &= \|\varphi_0 + v \cdot \nabla_q \varphi_1 + \varepsilon^{-1}(|v|^2 - d)\partial_z \varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 + \|\nabla_q \varphi_1 + 2\varepsilon^{-1}v\partial_z \varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 \\ &= \|\varphi_0\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 + \|v \cdot \nabla_q \varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 + \frac{1}{\varepsilon^2} \|(|v|^2 - d)\partial_z \varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 \\ &\quad + \|\nabla_q \varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 + \frac{4}{\varepsilon^2} \|v\partial_z \varphi_1\|_{L^2(\bar{\mu}, L^2(\kappa))}^2 \\ &= \|\varphi_0\|_{L^2(\bar{\mu})}^2 + \|\nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{2d}{\varepsilon^2} \|\partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{4d}{\varepsilon^2} \|\partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\leq 2c_0 \|g\|_{L^2(\bar{\mu})}^2. \end{aligned}$$

Here  $\nabla_v \varphi_0 = 0$  as  $\varphi_0, \varphi_1$  depend only on  $q, z$ , and referring to the explicit expression of the generator (D.20), we have  $\nabla_v(\mathcal{T}\varphi_1) = -(\nabla_q \varphi_1 + 2\varepsilon^{-1}v\partial_z \varphi_1)$ . In the same way, for  $C_{1,T}$  using (D.14), we have to bound  $\|(\partial_t + \mathcal{T})(\varphi_0 + \mathcal{T}\varphi_1)\|_{L^2(\bar{\mu} \otimes \kappa)}^2$  as  $\mathcal{T}$  is antisymmetric. Note that

$$(\partial_t + \mathcal{T})(\varphi_0 + \mathcal{T}\varphi_1) = \text{I} + \text{II} + \text{III} + \text{IV}$$

with

$$\begin{aligned} \text{I} &= \partial_t \varphi_0 + v^\top (\nabla_q^2 \varphi_1) v - \nabla U \cdot \nabla_q \varphi_1 - \frac{2d}{\varepsilon^2} z \partial_z \varphi_1 + \varepsilon^{-2} (|v|^2 - d)^2 \partial_{zz}^2 \varphi_1, \\ \text{II} &= v \cdot \nabla_q (\partial_t \varphi_1 + \varphi_0) - \frac{2}{\varepsilon} \nabla U \cdot v \partial_z \varphi_1 - \frac{1}{\varepsilon} z v \cdot \nabla_q \varphi_1, \\ \text{III} &= \varepsilon^{-1} (|v|^2 - d) (\partial_z \partial_t \varphi_1 + \partial_z \varphi_0) - \frac{2}{\varepsilon^2} (|v|^2 - d) z \partial_z \varphi_1, \\ \text{IV} &= 2\varepsilon^{-1} v (|v|^2 - d) \cdot \nabla_q \partial_z \varphi_1. \end{aligned}$$

Since  $\text{II} + \text{IV}$  is orthogonal to both  $\text{I}$  and  $\text{III}$  in  $L^2(\bar{\mu})$ , we have

$$\begin{aligned} \|(\partial_t + \mathcal{T})(\varphi_0 + \mathcal{T}\varphi_1)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 &\leq 2 \left( \|\text{I}\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + \|\text{III}\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + \|\text{III}\|_{L^2(\bar{\mu} \otimes \kappa)}^2 + \|\text{IV}\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \right). \end{aligned}$$

Using [61, Lemma 2.2], we have that

$$\begin{aligned} \|zf\|_{L^2(\bar{\mu})}^2 &\leq 4\|\partial_z f\|_{L^2(\bar{\mu})}^2 + 2\|f\|_{L^2(\bar{\mu})}^2, \\ \|f\nabla U\|_{L^2(\bar{\mu})}^2 &\leq 16\|\nabla_q f\|_{L^2(\bar{\mu})}^2 + 4Ld\|f\|_{L^2(\bar{\mu})}^2. \end{aligned} \quad (\text{D.34})$$

These allow us to estimate directly

$$\begin{aligned} \|\text{IV}\|_{L^2(\bar{\mu} \otimes \kappa)}^2 &= \frac{4}{\varepsilon^2} \|\nabla_q \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 \int v_1^2 (|v|^2 - d)^2 d\kappa(v) = \frac{8(d+4)}{\varepsilon^2} \|\nabla_q \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2, \\ \|\text{III}\|_{L^2(\bar{\mu} \otimes \kappa)}^2 &= \frac{2d}{\varepsilon^2} \|\partial_z \partial_t \varphi_1 + \partial_z \varphi_0 - \frac{2}{\varepsilon} z \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\leq \frac{6d}{\varepsilon^2} (\|\partial_z \partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\partial_z \varphi_0\|_{L^2(\bar{\mu})}^2 + \frac{4}{\varepsilon^2} \|z \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2) \\ &\stackrel{(\text{D.34})}{\leq} \frac{6d}{\varepsilon^2} \left( \|\partial_z \partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\partial_z \varphi_0\|_{L^2(\bar{\mu})}^2 \right. \\ &\quad \left. + \frac{8}{\varepsilon^2} \|\partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{16}{\varepsilon^2} \|\partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \right), \\ \|\text{I}\|_{L^2(\bar{\mu})}^2 &= \|\partial_t \varphi_0 - \nabla U \cdot \nabla_q \varphi_1 - \frac{2d}{\varepsilon^2} z \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \|v^\top \nabla_q^2 \varphi_1 v\|_{L^2(\bar{\mu})}^2 \\ &\quad + \frac{1}{\varepsilon^4} \|(|v|^2 - d)^2 \partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\quad + 2 \left\langle \partial_t \varphi_0 - \nabla U \cdot \nabla_q \varphi_1 - \frac{2d}{\varepsilon^2} z \partial_z \varphi_1, v^\top \nabla_q^2 \varphi_1 v + \frac{1}{\varepsilon^2} (|v|^2 - d)^2 \partial_{zz}^2 \varphi_1 \right\rangle \\ &\quad + 2 \left\langle v^\top \nabla_q^2 \varphi_1 v, \frac{1}{\varepsilon^2} (|v|^2 - d)^2 \partial_{zz}^2 \varphi_1 \right\rangle \\ &= \|\partial_t \varphi_0 - \nabla U \cdot \nabla_q \varphi_1 - \frac{2d}{\varepsilon^2} z \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + 2\|\nabla_q^2 \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\Delta_q \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\quad + \frac{12d(d+4)}{\varepsilon^4} \|\partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{4(d+4)}{\varepsilon^2} \left\langle \Delta_q \varphi_1, \partial_{zz}^2 \varphi_1 \right\rangle \\ &\quad + 2 \left\langle \partial_t \varphi_0 - \nabla U \cdot \nabla_q \varphi_1 - \frac{2d}{\varepsilon^2} z \partial_z \varphi_1, \Delta_q \varphi_1 + \frac{2d}{\varepsilon^2} \partial_{zz}^2 \varphi_1 \right\rangle \\ &= \|\partial_t \varphi_0 - \nabla U \cdot \nabla_q \varphi_1 - \frac{2d}{\varepsilon^2} z \partial_z \varphi_1 + \Delta_q \varphi_1 + \frac{2d}{\varepsilon^2} \partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\quad + 2\|\nabla_q^2 \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{16}{\varepsilon^2} \left\langle \Delta_q \varphi_1, \partial_{zz}^2 \varphi_1 \right\rangle + \frac{8d(d+6)}{\varepsilon^4} \|\partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\leq \|g\|_{L^2(\bar{\mu})}^2 + 10\|\nabla_q^2 \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{8d(d+7)}{\varepsilon^4} \|\partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2, \end{aligned}$$

as well as

$$\begin{aligned} \|\text{II}\|_{L^2(\bar{\mu})}^2 &= \|\nabla_q (\partial_t \varphi_1 + \varphi_0) - \frac{2}{\varepsilon} \nabla U \partial_z \varphi_1 - \frac{1}{\varepsilon} z \nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 \\ &\leq 4 \left( \|\nabla_q \partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\nabla_q \varphi_0\|_{L^2(\bar{\mu})}^2 \right. \\ &\quad \left. + \frac{4}{\varepsilon^2} \|\nabla U \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{1}{\varepsilon^2} \|z \nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 \right) \end{aligned}$$

## Appendix D. Hypocoercivity meets lifts

$$\begin{aligned}
& \stackrel{(\text{D.34})}{\leq} 4 \left( \|\nabla_q \partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\nabla_q \varphi_0\|_{L^2(\bar{\mu})}^2 + \frac{68}{\varepsilon^2} \|\nabla_q \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 \right. \\
& \quad \left. + \frac{16Ld}{\varepsilon^2} \|\partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{2}{\varepsilon^2} \|\nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 \right).
\end{aligned}$$

To combine everything above, using  $d \geq 1$ , we arrive at

$$\begin{aligned}
& \|(\partial_t + \mathcal{T})(\varphi_0 + \mathcal{T}\varphi_1)\|_{L^2(\bar{\mu} \otimes \kappa)}^2 \\
& \leq \frac{16(d+38)}{\varepsilon^2} \|\nabla_q \partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{12d}{\varepsilon^2} \left( \|\partial_z \partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\partial_z \varphi_0\|_{L^2(\bar{\mu})}^2 \right) \\
& \quad + 2\|g\|_{L^2(\bar{\mu})}^2 + 20\|\nabla_q^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \\
& \quad + \frac{32d}{\varepsilon^4} (3 + 4\varepsilon^2 L) \|\partial_z \varphi_1\|_{L^2(\bar{\mu})}^2 + \frac{16d(d+19)}{\varepsilon^4} \|\partial_{zz}^2 \varphi_1\|_{L^2(\bar{\mu})}^2 \\
& \quad + 8 \left( \|\nabla_q \partial_t \varphi_1\|_{L^2(\bar{\mu})}^2 + \|\nabla_q \varphi_0\|_{L^2(\bar{\mu})}^2 + \frac{2}{\varepsilon^2} \|\nabla_q \varphi_1\|_{L^2(\bar{\mu})}^2 \right) \\
& \stackrel{(\text{D.31}), (\text{D.32})}{\leq} 314 \left( c_1 + \left( \frac{1}{\varepsilon^2} + L \right) c_0 \right) \|g\|_{L^2(\bar{\mu})}^2.
\end{aligned}$$

### Optimisation of the parameters

Now that we have  $C_{0,T}^2 = 2c_0$  and  $C_{1,T}^2 = 314(c_1 + (\frac{1}{\varepsilon^2} + L)c_0)$  with  $c_0, c_1$  given in (D.33), and notice that Assumption D.4 holds with  $R = P_v = \gamma$  in our example, we can directly apply Theorem D.5 and see that the convergence rate

$$\lambda \geq \frac{2\gamma}{1 + (C_{1,T} + C_{0,T}\gamma)^2}$$

holds for any  $T$ . From the expressions in (D.33), it is advantageous to pick  $T^2 \sim P_x^{-1}$ , since the scaling of the rate will worsen in either regime  $T^2 \gg P_x^{-1}$  or  $T^2 \ll P_x^{-1}$ . For simplicity we pick  $T^2 = \pi^2 P_x^{-1}$ , which yields

$$\begin{aligned}
c_0 & \leq 63P_x^{-1}, \quad c_1 \leq 391 + 43P_x^{-1}M, \quad C_{0,T}^2 \leq 126P_x^{-1}, \\
C_{1,T}^2 & \leq 314 \left( 391 + 43P_x^{-1}M + 63\left(\frac{1}{\varepsilon^2} + L\right)P_x^{-1} \right),
\end{aligned}$$

and consequently, writing  $P_x^{-1} = \max\{P_q^{-1}, \frac{\varepsilon^2}{2d}\} \leq P_q^{-1} + \frac{\varepsilon^2}{2d}$ , we obtain

$$\begin{aligned}
\lambda & \geq \frac{2\gamma}{1 + \frac{3}{2}C_{1,T}^2 + 3C_{0,T}^2\gamma^2} \\
& \geq \frac{2\gamma}{61388 + (P_q^{-1} + \frac{\varepsilon^2}{2d}) \left( 378\gamma^2 + 6751M + 9891\left(\frac{1}{\varepsilon^2} + L\right) \right)}.
\end{aligned}$$

From here, we already recover the scaling<sup>2</sup>

$$\lambda \geq \bar{\lambda} \min \left\{ \frac{1}{\gamma}, \frac{1}{\gamma\varepsilon^2}, \gamma\varepsilon^2, \frac{\gamma}{\varepsilon^2} \right\}$$

<sup>2</sup>The scaling in the published version [172, (2.12)] was incorrect. It was corrected later in the third arXiv version of their work [173, (2.12)].

#### D.4. Lower bounds for the generalised Langevin equation

in [173] for fixed  $M$ ,  $d$ ,  $L$  and  $P_q$ . For optimal scaling of  $\lambda$ , we pick  $\varepsilon^2 = \sqrt{\frac{d}{P_q(M+L+\gamma^2)}}$  and then  $\gamma = \sqrt{P_q + M + L}$ , and obtain the rate

$$\lambda \geq \frac{P_q}{66334\sqrt{P_q + M + L}},$$

which is dimension independent. For convex potentials that are not far from quadratic, we have  $M = 0$  and  $L \sim P_q$ , and therefore  $\lambda \sim \sqrt{P_q}$ , which means up to a universal constant, adaptive Langevin dynamics is also an optimal second-order lift of the overdamped Langevin dynamics.

**Remark D.8.** We are unable to completely recover the result in [61] for Langevin dynamics (D.18) in the sense that the rates in [61] do not depend on  $L$  that appears in Assumption D.6 (2). The bottleneck is the estimate of  $\|\Pi\|_{L^2(\bar{\mu})}$ , especially the contribution from  $\nabla U \partial_z \varphi_1$ , where a factor of  $L$  arises after using (D.34), while for the case of (D.18), any term with  $\nabla U$  only appears in a term similar to (I), which can be simplified using the divergence equation (D.12). It is unclear to us whether this is only an artifact of our proof techniques.

#### D.4. Lower bounds for the generalised Langevin equation

Recall the generalised Langevin dynamics  $(X_t, V_t, Z_t)_{t \geq 0}$  on  $\mathcal{X} \times \mathcal{V}$  with  $\mathcal{X} = \mathbb{R}^d$  and  $\mathcal{V} = \mathbb{R}^d \times \mathbb{R}^d$ , which is the solution to the GLE (D.22). Its invariant probability measure is

$$\hat{\mu}(dx \, dv \, dz) = \mu(dx) \kappa(dv \, dz)$$

with  $\mu(dx) \propto e^{-U(x)} dx$  and  $\kappa = \mathcal{N}(0, I_d) \otimes \mathcal{N}(0, I_d)$ . The generator of the associated transition semigroup on  $L^2(\hat{\mu})$  is  $\mathcal{L} = L_v - \mathcal{T}$  with the antisymmetric part

$$-\mathcal{T} = v \cdot \nabla_x - \nabla U \cdot \nabla_v + \lambda(z \cdot \nabla_v - v \cdot \nabla_z)$$

and the symmetric part

$$L_v = \gamma(-z \cdot \nabla_z + \Delta_z) = -\gamma \nabla_z^* \nabla_z,$$

which are the transport and dissipation operators, respectively. We correspondingly set  $\Pi f(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x, v, z) d\kappa(v, z)$ . Note that  $L_v$  acts only on the  $z$ -component and not on  $v$ .

A key observation is that the generalised Langevin dynamics is a second-order lift of an overdamped Langevin diffusion. This may seem counterintuitive at first, since the degeneracy is of higher order in terms of the number of Lie brackets needed to span the whole space. This corresponds to the noise being more degenerate than, e.g., in the case of Langevin dynamics, since it is only present in the second derivative  $Z$  of the spatial component  $X$ . The main difficulty in the treatment of this generalised case is precisely the fact that the ellipticity of  $L_v$  is only present in  $z$ , whereas the lifting variables are  $(v, z)$ . In particular, Assumption D.4 is violated, since  $\ker(L_v) \neq \text{Im}(\Pi)$ .

## Appendix D. Hypocoercivity meets lifts

**Lemma D.9.** *Assume that  $\mu$  satisfies a Poincaré inequality with optimal constant  $P_x^{-1}$ , i.e.*

$$P_x = \inf_{f \in L_0^2(\mu) \setminus \{0\}} \frac{\|\nabla_x f\|_{L^2(\mu)}^2}{\|f\|_{L^2(\mu)}^2}.$$

*Then Assumptions D.1 and D.2 are satisfied with exactly the constant  $P_x$ . In particular, generalised Langevin dynamics is a second-order lift of the overdamped Langevin diffusion with generator  $\frac{1}{2}L_x = -\frac{1}{2}\nabla U \cdot \nabla_x + \frac{1}{2}\Delta_x$ .*

*Proof.* It is immediate that

$$\Pi \mathcal{T} \Pi f = - \int_{\mathbb{R}^d \times \mathbb{R}^d} v \cdot \nabla_x \Pi f(x) d\kappa(v, z) = 0,$$

so that Assumption D.1 is satisfied. Furthermore,

$$-(\mathcal{T} \Pi)^*(\mathcal{T} \Pi) f = \int_{\mathbb{R}^d \times \mathbb{R}^d} v^\top (\nabla_x^2 \Pi f) v - \nabla U \cdot \nabla_x \Pi f d\kappa(v, z) = L_x \Pi f,$$

which, together with the Poincaré inequality for  $\mu$ , yields Assumption D.2.  $\square$

The following lower bound is now an immediate consequence of the general lower bound for second-order lifts in Assumption D.2.

**Corollary D.10** (Lower bound on the relaxation time). *The relaxation time  $t_{\text{rel}}(\hat{P})$  of the generalised Langevin dynamics satisfies*

$$t_{\text{rel}}(\hat{P}) \geq \frac{1}{2} P_x^{-1/2}.$$

In particular, compared to the overdamped Langevin diffusion, the convergence to equilibrium of the generalised Langevin dynamics as measured by the relaxation time is not accelerated by more than a square root. The square-root speed-up to a relaxation time of the order  $P_x^{-1/2}$  compared to the order  $P_x^{-1}$  for the overdamped Langevin diffusion is already achieved by the (second-order) Langevin dynamics [61].

**Example D.11** (Gaussian case). In order to see how sharp the lower bound is, let us consider the simple case where  $\mu = \mathcal{N}(0, m^{-1})$  is a centred normal distribution on  $\mathbb{R}$  with variance  $m^{-1}$ . In this case, Assumption D.2 is satisfied with  $P_x = m$ . The GLE (D.22) then becomes an Ornstein-Uhlenbeck process

$$d \begin{pmatrix} X_t \\ V_t \\ Z_t \end{pmatrix} = A \begin{pmatrix} X_t \\ V_t \\ Z_t \end{pmatrix} dt + \Sigma dW_t$$

with

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -m & 0 & \lambda \\ 0 & -\lambda & -\gamma \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0 \\ 0 \\ \sqrt{2\gamma} \end{pmatrix}.$$

#### D.4. Lower bounds for the generalised Langevin equation

The asymptotic decay rate of the associated transition semigroup  $(\hat{P}_t)_{t \geq 0}$  is given by the spectral gap of its generator in  $L^2(\hat{\mu})$ , which by [193] coincides with the spectral gap

$$\text{gap}(A) = \inf\{-\text{Re}(\lambda) : \lambda \in \text{spec}(A)\}$$

of  $A$ . When fixing the scaling  $\lambda = a\sqrt{m}$  and  $\gamma = b\sqrt{m}$  for some  $a, b > 0$ , an explicit computation of the zeros of the characteristic polynomial shows that the spectrum of  $A$  is given by

$$\begin{aligned}\mu_1 &= \sqrt{m} \left( -\frac{b}{3} - \frac{\beta}{(\sqrt{\alpha^2 + \beta^3} - \alpha)^{1/3}} + (\sqrt{\alpha^2 + \beta^3} - \alpha)^{1/3} \right), \\ \mu_2 &= \sqrt{m} \left( -\frac{b}{3} + \frac{(1 - i\sqrt{3})\beta}{2(\sqrt{\alpha^2 + \beta^3} - \alpha)^{1/3}} - \frac{(1 + i\sqrt{3})(\sqrt{\alpha^2 + \beta^3} - \alpha)^{1/3}}{2} \right), \\ \mu_3 &= \sqrt{m} \left( -\frac{b}{3} + \frac{(1 + i\sqrt{3})\beta}{2(\sqrt{\alpha^2 + \beta^3} - \alpha)^{1/3}} - \frac{(1 - i\sqrt{3})(\sqrt{\alpha^2 + \beta^3} - \alpha)^{1/3}}{2} \right),\end{aligned}$$

where

$$\begin{aligned}\alpha &= \frac{b}{2} + \frac{b^3}{27} - \frac{b(a^2 + 1)}{6}, \\ \beta &= \frac{a^2}{3} - \frac{b^2}{9} + \frac{1}{3}.\end{aligned}$$

The spectral gap is maximised when all three eigenvalues coincide, which is the case for  $\alpha = \beta = 0$ , or equivalently  $a = 2\sqrt{2}$  and  $b = 3\sqrt{3}$ . Correspondingly, the choice  $\lambda = 2\sqrt{2m}$  and  $\gamma = 3\sqrt{3m}$  yields the optimal spectral gap  $\sqrt{3m}$ .

To obtain non-asymptotic bounds, the spectral gap does not suffice, and we consider the relaxation time. By [17], the transition semigroup  $(\hat{P}_t)_{t \geq 0}$  satisfies

$$\|\hat{P}_t\|_{L_0^2(\hat{\mu}) \rightarrow L_0^2(\hat{\mu})} = \|\exp(t\tilde{A})\| \quad \text{for all } t \geq 0$$

with

$$\tilde{A} = \begin{pmatrix} 0 & \sqrt{m} & 0 \\ -\sqrt{m} & 0 & \lambda \\ 0 & -\lambda & -\gamma \end{pmatrix} = \sqrt{m} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 2\sqrt{2} \\ 0 & -2\sqrt{2} & -3\sqrt{3} \end{pmatrix}$$

with the choices of  $\lambda$  and  $\gamma$  given above. Since  $\text{spec}(\tilde{A}) = \{-\sqrt{3m}\}$  and  $\dim \ker(\tilde{A} + \sqrt{3m}I) = 1$ , we obtain

$$\exp(t\tilde{A}) = e^{-\sqrt{3mt}} S^{-1} \begin{pmatrix} 1 & \sqrt{mt} & \frac{1}{2}mt^2 \\ 0 & 1 & \sqrt{mt} \\ 0 & 0 & 1 \end{pmatrix} S.$$

## Appendix D. Hypocoercivity meets lifts

with  $S \in \text{GL}(3, \mathbb{R})$ . An explicit calculation of the singular values then shows that

$$\|\hat{P}_t\|_{L_0^2(\hat{\mu}) \rightarrow L_0^2(\hat{\mu})} = e^{-\sqrt{3mt}} p(\sqrt{mt})^{1/2}$$

$$\text{with } p(s) = 1 + s^2 + \frac{1}{8}s^4 + \left(2s + \frac{1}{2}s^3\right) \sqrt{\frac{s^2}{16} + \frac{1}{2}}.$$

A numerical calculation yields the relaxation time

$$t_{\text{rel}}(\hat{P}) = \inf\{t \geq 0: \|\hat{P}_t\|_{L_0^2(\hat{\mu}) \rightarrow L_0^2(\hat{\mu})} \leq e^{-1}\} \approx 0.964m^{-1/2}$$

of the generalised Langevin dynamics. By a factorisation argument, the same holds true for arbitrary Gaussian probability measures on  $\mathbb{R}^d$ . Comparing this to the lower bound  $0.5m^{-1/2}$  from Theorem D.10, this shows that, in the terminology of [99], on the class of quadratic potentials on  $\mathbb{R}^d$ , the generalised Langevin dynamics with optimal choice of  $\lambda$  and  $\gamma$  is a 1.93-optimal lift of the corresponding overdamped Langevin diffusion. As a comparison, on the same class of potentials, Langevin dynamics (D.18) with optimal choice of friction  $\gamma$  achieves a relaxation time of approximately  $2.73m^{-1/2}$  and is thus only a 5.46-optimal lift, see [99, Example 14].

## Acknowledgments

We would like to thank Andreas Eberle and Gabriel Stoltz for many helpful discussions. GB has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034413. FL wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – GZ2047/1, Projekt-ID 390685813. LW is supported by the National Science Foundation via grant DMS-2407166. He is also indebted to the Mathematical Sciences department at Carnegie Mellon University for partly supporting his visit to Europe in July 2024. Part of this work was completed when GB and LW were visiting the Institute for Applied Mathematics in Bonn. GB and LW would like to thank IAM for their hospitality.

## Appendix E.

# Convergence rates of self-repellent random walks and event chain Monte Carlo

A. Eberle and F. Lörler. *Convergence rates of self-repellent random walks, their local time and Event Chain Monte Carlo*. Nov. 2025. arXiv: [2511.23453 \[math.PR\]](#).

**Abstract:** We study the rate of convergence to equilibrium of the self-repellent random walk and its local time process on the discrete circle  $\mathbb{Z}_n$ . While the self-repellent random walk alone is non-Markovian since the jump rates depend on its history via its local time, jointly considering the evolution of the local time profile and the position yields a piecewise deterministic, non-reversible Markov process. We show that this joint process can be interpreted as a second-order lift of a reversible diffusion process, the discrete stochastic heat equation with Gaussian invariant measure. In particular, we obtain a lower bound on the relaxation time of order  $\Omega(n^{3/2})$ . Using a flow Poincaré inequality, we prove an upper bound for a slightly modified dynamics of order  $O(n^2)$ , matching recent conjectures in the physics literature. Furthermore, since the self-repellent random walk and its local time process coincide with the Event Chain Monte Carlo algorithm for the harmonic chain, a non-reversible MCMC method, we demonstrate that the relaxation time bound confirms the recent empirical observation that Event Chain Monte Carlo algorithms can outperform traditional MCMC methods such as Hamiltonian Monte Carlo.

**Keywords:** Self-repellent random walk; local time; Event Chain Monte Carlo; lift; convergence rate.

**MSC Subject Classification:** 60J25, 60K50, 82C41, 60J22.

## E.1. Introduction

In this paper, we study the rate of convergence to stationarity for the joint process of a self-repellent random walk on a discrete circle and its local time process. This is a stronger statement than convergence to stationarity for the self-repellent random walk on its own. The key new ingredient is the identification of the joint process of local time and the self-repellent random walk as a second order lift of a discrete stochastic heat

equation in the sense of [99]. This will allow us to apply a recently developed approach to quantitative bounds for convergence to equilibrium of non-reversible lifts of Markov processes to self-repellent random walks.

### E.1.1. Self-repellent random walks

We consider the self-repellent random walk as a continuous-time stochastic process  $X(t)$  with state space given by the discrete circle  $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ , and defined on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . For  $t \in [0, \infty)$  let  $L(t) = (L_i(t))_{i \in \mathbb{Z}_n}$ , where

$$L_i(t) = \int_0^t 1_{\{i\}}(X(s)) \, ds$$

is the time the process has spent at position  $i$  up to time  $t$ . The self-repellent random walk performs nearest neighbour jumps

$$\begin{aligned} i \rightarrow i+1 & \quad \text{with rate} \quad q^+(l, i) = (l'_i)_-, \quad \text{and} \\ i \rightarrow i-1 & \quad \text{with rate} \quad q^-(l, i) = (l'_{i-1})_+, \end{aligned}$$

where  $l$  is the current value of the local time process  $L(t)$  and

$$l'_i = l_{i+1} - l_i.$$

Here and in the following, indices are to be understood periodically in  $\mathbb{Z}_n$ . The process  $X(t)$  on its own has memory and is hence not a Markov process, but the process  $(L(t), X(t))_{t \geq 0}$  is a piecewise-deterministic Markov process with state space  $\mathbb{R}^n \times \mathbb{Z}_n$ . The jump rates  $q^+$  and  $q^-$  are chosen such that the walker prefers to jump towards neighbouring states at which the random walk has spent less time. Such self-repellent random walks were first introduced by Amit, Parisi and Peliti [9] in discrete time as a model for polymer growth. The continuous-time dynamics was introduced in [242, Section 1.2] on the infinite grid  $\mathbb{R}^{\mathbb{Z}} \times \mathbb{Z}$  and termed the continuous-time ‘true’ self-avoiding random walk. Note that for any  $t \in [0, \infty)$ ,

$$\sum_{i=1}^n L_i(t) = t. \tag{E.1}$$

In order to obtain a process that converges to stationarity, we project the local time  $l$  onto the subspace

$$\mathcal{S} = \left\{ \ell \in \mathbb{R}^n : \sum_{i=1}^n \ell_i = 0 \right\}$$

of mean-zero vectors by subtracting the average, i.e. we consider the process  $\mathcal{L}(t)$  with state space  $\mathcal{S}$  and components

$$\mathcal{L}_i(t) = L_i(t) - \frac{1}{n} \sum_{j=1}^n L_j(t) = L_i(t) - \frac{t}{n} \quad \text{for } i \in \mathbb{Z}_n. \tag{E.2}$$

Note that functions  $f: \mathcal{S} \rightarrow \mathbb{R}$  can be extended to  $\mathbb{R}^n$  by setting  $\hat{f}(l) = f(l - l \cdot \mathbf{n} \mathbf{n})$ , where  $\mathbf{n} = n^{-1/2} \mathbf{1}$  with  $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^n$  is the unit normal vector to  $\mathcal{S}$ . In this sense, we define

$$\partial_{\ell_i} f(\ell) = \partial_{\ell_i} \hat{f}(\ell) = \partial_{e_i - n^{-1} \mathbf{1}} f(\ell). \quad (\text{E.3})$$

The corresponding vector fields  $\partial_{\ell_i}$  are the projections of the canonical basis vector fields  $e_i \in \mathbb{R}^n$  to  $\mathcal{S}$ . Since the jump rates  $q^+$  and  $q^-$  only depend on the increments of  $l$  which coincide with those of  $\ell$ , the process  $(\mathcal{L}(t), X(t))_{t \geq 0}$  is also a piecewise-deterministic Markov process with state space  $\mathcal{S} \times \mathbb{Z}_n$ . On functions  $f \in C_b^1(\mathcal{S} \times \mathbb{Z}_n)$ , its generator is

$$\begin{aligned} \hat{\mathfrak{L}}f(\ell, i) &= \mathfrak{T}f(\ell, i) + \mathfrak{J}_+ f(\ell, i) + \mathfrak{J}_- f(\ell, i), \quad \text{where} \\ \mathfrak{T}f(\ell, i) &= \partial_{\ell_i} f(\ell, i), \\ \mathfrak{J}_\pm f(\ell, i) &= q^\pm(\ell, i) (f(\ell, i \pm 1) - f(\ell, i)). \end{aligned}$$

Next we identify the stationary distribution. Consider the Laplacian on  $\mathbb{Z}_n$  given by

$$\Delta_n = \begin{pmatrix} -2 & 1 & 0 & \cdots & 1 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -2 \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad (\text{E.4})$$

that is  $(\Delta_n l)_i = l_{i+1} - 2l_i + l_{i-1} = l'_i - l'_{i-1}$ . Since  $\ker \Delta_n = \langle \mathbf{1} \rangle \subset \mathbb{R}^n$ ,  $\Delta_n: \mathcal{S} \rightarrow \mathcal{S}$  is a bijection. Hence the probability measure

$$\mu = \mathcal{N}(0, -\Delta_n^{-1})$$

is well-defined and supported on  $\mathcal{S}$ . Note that  $\mu$  has a density proportional to  $\exp(-U)$  with respect to the  $(n-1)$ -dimensional Haar measure on  $\mathcal{S}$ , where

$$U(\ell) = -\frac{1}{2} \ell^\top \Delta_n \ell = \frac{1}{2} \sum_{i=1}^n (\ell_{i+1} - \ell_i)^2.$$

**Lemma E.1.** (i) The probability measure  $\hat{\mu} = \mu \otimes \text{Unif}(\mathbb{Z}_n)$  is invariant for the Markov process  $(\mathcal{L}(t), X(t))_{t \geq 0}$ .

(ii) Functions  $f \in C_b^1(\mathcal{S} \times \mathbb{Z}_n)$  are contained in the domains of the adjoints of the operators  $\mathfrak{T}$ ,  $\mathfrak{J}_+$ ,  $\mathfrak{J}_-$  and  $\hat{\mathfrak{L}}$  in  $L^2(\hat{\mu})$ , and

$$\begin{aligned} \mathfrak{T}^* f(\ell, i) &= -\mathfrak{T}f(\ell, i) - (\Delta_n \ell)_i f(\ell, i), \\ \mathfrak{J}_\pm^* f(\ell, i) &= q^\pm(\ell, i \mp 1) f(\ell, i \mp 1) - q^\pm(\ell, i) f(\ell, i), \\ \hat{\mathfrak{L}}^* f(\ell, i) &= -\mathfrak{T}f(\ell, i) + (\ell'_i)_+ (f(\ell, i+1) - f(\ell, i)) + (\ell'_{i-1})_- (f(\ell, i-1) - f(\ell, i)). \end{aligned}$$

*Proof.* By an elementary computation based on integration and summation by parts, one verifies that for any  $f, g \in C_b^1(\mathcal{S} \times \mathbb{Z}_n)$  and for  $\mathfrak{A} \in \{\mathfrak{T}, \mathfrak{J}_+, \mathfrak{J}_-\}$ ,

$$\sum_{i=1}^n \int f(\ell, i) (\mathfrak{A}g)(\ell, i) \mu(d\ell) = \sum_{i=1}^n \int (\mathfrak{A}^* f)(\ell, i) g(\ell, i) \mu(d\ell), \quad (\text{E.5})$$

where  $\mathfrak{T}^*$ ,  $\mathfrak{J}_+^*$  and  $\mathfrak{J}_-^*$  are given by the expressions in the statement of the lemma. Consequently, a corresponding identity holds for  $\mathfrak{A} = \hat{\mathfrak{L}}$  with

$$\begin{aligned} \hat{\mathfrak{L}}^* f(\ell, i) &= \mathfrak{T}^* f(\ell, i) + \mathfrak{J}_+^* f(\ell, i) + \mathfrak{J}_-^* f(\ell, i) \\ &= -\mathfrak{T} f(\ell, i) - (\ell'_i - \ell'_{i-1}) f(\ell, i) + (\ell'_i)_+ f(\ell, i+1) - (\ell'_i)_- f(\ell, i) \\ &\quad + (\ell'_{i-1})_- f(\ell, i-1) - (\ell'_{i-1})_+ f(\ell, i) \\ &= -\mathfrak{T} f(\ell, i) + (\ell'_i)_+ (f(\ell, i+1) - f(\ell, i)) + (\ell'_{i-1})_- (f(\ell, i-1) - f(\ell, i)). \end{aligned}$$

In particular, since  $\hat{\mathfrak{L}}^* 1 = 0$ ,

$$\int \hat{\mathfrak{L}} f d\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \int \hat{\mathfrak{L}} f(\ell, i) \mu(d\ell) = 0$$

holds for all  $f \in C_b^1(\mathcal{S} \times \mathbb{Z}_n)$ . This implies invariance of  $\hat{\mu}$  by [92, Theorem 21]. The second part of the claim now follows directly from (E.5).  $\square$

By Lemma E.1, the transition function of the Markov process  $(\mathcal{L}(t), X(t))_{t \geq 0}$  induces a strongly continuous contraction semigroup on  $L^2(\mathcal{S} \times \mathbb{Z}_n, \hat{\mu})$  that we denote by  $(\hat{P}_t)_{t \geq 0}$ . The generator of this semigroup and its domain are denoted by  $(\hat{\mathfrak{L}}, \text{Dom}(\hat{\mathfrak{L}}))$ , it extends the operator  $(\hat{\mathfrak{L}}, C_b^1(\mathcal{S} \times \mathbb{Z}_n))$  introduced above.

**Remark E.2.** Lemma E.1 shows that for  $f \in C_b^1(\mathcal{S} \times \mathbb{Z}_n)$ , the symmetric and antisymmetric part of the generator  $\hat{\mathfrak{L}}$  are given by

$$\begin{aligned} \frac{1}{2}(\hat{\mathfrak{L}} + \hat{\mathfrak{L}}^*) f(\ell, i) &= |\ell'_i| (f(\ell, i+1) - f(\ell, i)) + |\ell'_{i-1}| (f(\ell, i-1) - f(\ell, i)), \\ \frac{1}{2}(\hat{\mathfrak{L}} - \hat{\mathfrak{L}}^*) f(\ell, i) &= \mathfrak{T} f(\ell, i) - \ell'_i (f(\ell, i+1) - f(\ell, i)) + \ell'_{i-1} (f(\ell, i-1) - f(\ell, i)). \end{aligned}$$

Thus the symmetric part is the generator of a jump process on the second component with rates depending on the first component, whereas the antisymmetric part is not the generator of a Markov process.

### E.1.2. The self-repellent random walk as a second-order lift

Inspired by the notion of lifts of Markov chains in discrete time [66, 79], second-order lifts were introduced as a counterpart in continuous time and space in [99]. They provide a suitable framework for the study of convergence to equilibrium of many non-reversible dynamics by viewing them as lifts of reversible diffusions. Intuitively, a Markov process on a product space  $\mathcal{S} \times \mathcal{V}$  with invariant probability measure  $\mu \otimes \kappa$  is a second-order lift

of a Markov process with state space  $\mathcal{S}$  and invariant measure  $\mu$  if in the limit as  $t \downarrow 0$ , the transition semigroups  $(\hat{P}_t)_{t \geq 0}$  and  $(P_t)_{t \geq 0}$  satisfy

$$\int_{\mathcal{V}} \hat{P}_t(f \circ \pi)(\ell, x) \kappa(dx) = (P_t f)(x) + o(t^2) \quad (\text{E.6})$$

for functions  $f$  in the domain of the generator  $\mathfrak{L}$  of  $(P_t)_{t \geq 0}$ , where  $\pi: \mathcal{S} \times \mathcal{V} \rightarrow \mathcal{S}$  is the canonical projection  $\pi(\ell, x) = \ell$ . A precise formal definition is given below.

Although a Markov process and its lift can behave very differently for long times, the lift property (E.6) can be used to study convergence to equilibrium of the lifted process based on spectral properties of  $\mathfrak{L}$  [97, 99, 101].

We will show that the self-repellent random walk considered above is a second-order lift of the reversible diffusion process  $(Z(t))_{t \geq 0}$  with state space  $\mathcal{S}$  that solves the discrete stochastic heat equation

$$dZ(t) = \frac{1}{2n} \Delta_n Z(t) dt + \frac{1}{\sqrt{n}} dW(t). \quad (\text{E.7})$$

Here  $W$  is a Brownian motion on  $\mathcal{S}$ , i.e.  $W(t) = B(t) - n^{-1} B(t) \cdot \mathbf{1} \mathbf{1}$  is the orthogonal projection of a Brownian motion  $B$  in  $\mathbb{R}^n$  to the linear subspace  $\mathcal{S}$ . The process  $(Z(t))_{t \geq 0}$  is a reversible diffusion process with invariant measure  $\mu$ , and its  $L^2(\mu)$  generator  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$  is given by

$$\mathfrak{L}g(\ell) = \frac{1}{2n} \sum_{i=1}^n (\partial_{\ell_i}^2 g(\ell) + (\Delta_n \ell)_i \partial_{\ell_i} g(\ell)) = -\frac{1}{2n} \text{grad}^* \text{grad} g(\ell), \quad (\text{E.8})$$

where  $\text{grad}$  is the gradient on  $\mathcal{S}$ , the adjoint is in  $L^2(\mu)$ , and  $\text{Dom}(\mathfrak{L}) = H^{2,2}(\mathcal{S}, \mu)$ . Note that this is an Ornstein-Uhlenbeck process slowed down by a factor  $1/(2n)$ . In particular, the spectral gap of the generator coincides with that of  $\frac{1}{2n} \Delta_n$ , see [193]. The eigenvectors of  $\Delta_n$  are  $\alpha_k = (\cos(\frac{2\pi k}{n} i))_{i \in \mathbb{Z}_n}$ ,  $k \in \{0, \dots, n-1\}$  with  $\Delta_n \alpha_k = 2(1 - \cos(\frac{2\pi k}{n})) \alpha_k$ . Therefore,

$$\text{gap}(\mathfrak{L}) = \frac{1}{2n} \text{gap}(\Delta_n) = \frac{1}{n} (1 - \cos(2\pi/n)) \in \Theta(n^{-3}). \quad (\text{E.9})$$

We now state the formal definition of a second-order lift that is motivated by a Taylor expansion of (E.6), see [99] for further motivation and examples. The process  $(\mathcal{L}(t), X(t))_{t \geq 0}$  with  $L^2(\hat{\mu})$  generator  $(\hat{\mathfrak{L}}, \text{Dom}(\hat{\mathfrak{L}}))$  is called a *second-order lift* of  $(Z(t))_{t \geq 0}$ , if there exists a core  $C$  of  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$  such that

$$f \circ \pi \in \text{Dom}(\hat{\mathfrak{L}}) \text{ for all } f \in C \quad (\text{E.10})$$

and for all  $f, g \in C$  we have

$$\langle \hat{\mathfrak{L}}(f \circ \pi), g \circ \pi \rangle_{L^2(\hat{\mu})} = 0, \quad (\text{E.11})$$

and

$$\frac{1}{2} \langle \hat{\mathfrak{L}}(f \circ \pi), \hat{\mathfrak{L}}(g \circ \pi) \rangle_{L^2(\hat{\mu})} = -\langle f, \mathfrak{L}g \rangle_{L^2(\mu)}. \quad (\text{E.12})$$

Conversely, we refer to the process  $(Z(t))_{t \geq 0}$  as the *collapse* of  $(\mathcal{L}(t), X(t))_{t \geq 0}$ .

**Theorem E.3.** *The self-repellent random walk  $(\mathcal{L}(t), X(t))$  is a second-order lift of  $(Z(t))$ .*

*Proof.* The smooth compactly supported functions  $C_c^\infty(\mathcal{S})$  form a core for  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$  [251]. For  $f \in C_c^\infty(\mathcal{S})$  we have  $f \circ \pi \in \text{Dom}(\hat{\mathfrak{L}})$  and  $\hat{\mathfrak{L}}(f \circ \pi)(\ell, x) = \partial_{\ell_x} f(\ell)$ , so that

$$\langle \hat{\mathfrak{L}}(f \circ \pi), g \circ \pi \rangle_{L^2(\hat{\mu})} = \frac{1}{n} \sum_{i=1}^n \int \partial_{\ell_i} f(\ell) g(\ell) \mu(d\ell) = 0,$$

which implies (E.11). Here we have used that  $\partial_{\ell_i}$  is the directional derivative in direction  $e_i - n^{-1}\mathbf{1}$ , and these vectors sum to 0. Furthermore,

$$\frac{1}{n} \sum_{i=1}^n (\hat{\mathfrak{L}}(g \circ \pi)(\ell, i))^2 = \frac{1}{n} \sum_{i=1}^n (\partial_{\ell_i} g(\ell))^2 = \frac{1}{n} |\text{grad } g(\ell)|^2,$$

so that  $\frac{1}{2} \int (\hat{\mathfrak{L}}(g \circ \pi))^2 d\hat{\mu} = - \int g \mathfrak{L}g d\mu$ , which implies (E.12) by polarisation.  $\square$

### E.1.3. Bounds for relaxation times

Although the invariant measure is a product of a Gaussian and a uniform distribution, sharp bounds on the exponential rate of convergence to equilibrium are not known for the Markov process  $(\mathcal{L}(t), X(t))$ . Even the order of the dependence of the convergence rate on  $n$  is unknown. The reason is that on the one hand, the process is non-reversible so that standard tools do not apply, and on the other hand, the combined dynamics consisting of jumps in the second component and deterministic moves in the first component does not reflect the special structure of the Gaussian measure in an obvious way.

We first note that the representation of the self-repellent random walk as a lift of the stochastic heat equation enables us to give a lower bound on the  $L^2$ -relaxation time

$$t_{\text{rel}}(\hat{P}) = \inf \left\{ t \geq 0 : \|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq e^{-1} \|f\|_{L^2(\hat{\mu})} \text{ for all } f \in L_0^2(\hat{\mu}) \right\}.$$

Here  $L_0^2(\hat{\mu})$  denotes the orthogonal complements of the constant functions in  $L^2(\hat{\mu})$ . Theorem 11 in [99] provides a lower bound on the relaxation time of arbitrary second-order lifts in terms of the square root of the spectral gap of the collapse. We give a slight improvement in terms of the constant prefactor with a short self-contained proof in Theorem E.4.

**Theorem E.4** (Lower bound on relaxation time). *Let  $(P_t)_{t \geq 0}$  be the transition semigroup of a reversible Markov process and  $(\hat{P}_t)_{t \geq 0}$  an arbitrary second-order lift. Then*

$$t_{\text{rel}}(\hat{P}) \geq \frac{1 - e^{-1}}{\sqrt{2}} \sqrt{t_{\text{rel}}(P)}.$$

*In particular, for the self-repellent random walk,*

$$t_{\text{rel}}(\hat{P}) \geq \frac{1 - e^{-1}}{2\pi} n^{3/2}.$$

*Proof.* For any  $f \in \text{Dom}(\hat{\mathfrak{L}}) \cap L_0^2(\hat{\mu})$  and  $t \geq 0$ , we have

$$\|\hat{P}_t f - f\|_{L^2(\hat{\mu})} \leq \int_0^t \|\hat{P}_s \hat{\mathfrak{L}} f\|_{L^2(\hat{\mu})} ds \leq t \|\hat{\mathfrak{L}} f\|_{L^2(\hat{\mu})},$$

so that

$$\|\hat{P}_t f\|_{L^2(\hat{\mu})} \geq \|f\|_{L^2(\hat{\mu})} - \|\hat{P}_t f - f\|_{L^2(\hat{\mu})} \geq \|f\|_{L^2(\hat{\mu})} - t \|\hat{\mathfrak{L}} f\|_{L^2(\hat{\mu})}.$$

For any  $g \in L_0^2(\mu)$  and  $\lambda > 0$  such that  $\langle g, -\mathfrak{L}g \rangle_{L^2(\mu)} = \lambda \|g\|_{L^2(\mu)}^2$ , we have  $\|\hat{\mathfrak{L}}(g \circ \pi)\| = \sqrt{2\lambda} \|g \circ \pi\|_{L^2(\hat{\mu})}$  by the second-order lift property (E.12), so that

$$\|\hat{P}_t(g \circ \pi)\|_{L^2(\hat{\mu})} \geq (1 - \sqrt{2\lambda}t) \|g \circ \pi\|_{L^2(\hat{\mu})} \quad \text{for all } t \geq 0.$$

In particular, we obtain  $t_{\text{rel}}(\hat{P}) \geq \frac{1-e^{-1}}{\sqrt{2}} \lambda^{-1/2}$ . Considering the limit  $\lambda \rightarrow \text{gap}(\mathfrak{L})$  yields the claim since  $t_{\text{rel}}(P) = \text{gap}(\mathfrak{L})^{-1}$  by reversibility. The lower bound for the self-repellent random walk follows from (E.9) and  $1 - \cos(x) \leq \frac{x^2}{2}$ .  $\square$

**Remark E.5.** Based on numerical and heuristic evidence [158, 192] and a Bethe ansatz calculation [105] for related models, it has been conjectured in the physics literature that the relaxation time is of order  $\Theta(n^2)$ . Mathematically rigorous proofs for quantitative upper and lower bounds on the relaxation time of this order are missing. Physics inspired arguments mostly concern different discrete models such as the lifted totally antisymmetric simple exclusion process [106]. Here the special combinatorial structure may allow for other approaches for bounding the relaxation time. Even if this were possible, the development of a more generally applicable analytic approach is important.

Next, as a first step in the direction of a rigorous proof of upper bounds on relaxation times, we consider a modified dynamics where we include additional particle moves that occur with an exponential rate  $\gamma > 0$ , i.e., we consider the piecewise deterministic Markov process with generator

$$\hat{\mathfrak{L}}^{(\gamma)} f = \hat{\mathfrak{L}} f + \gamma(Qf - f), \quad \text{Dom}(\hat{\mathfrak{L}}^{(\gamma)}) = \text{Dom}(\hat{\mathfrak{L}}), \quad (\text{E.13})$$

where  $Q$  is the transition matrix of a Markov chain on  $\mathbb{Z}_n$  which only acts on the second component of a function  $f: \mathcal{S} \times \mathbb{Z}_n \rightarrow \mathbb{R}$ . We denote the associated transition semi-group acting on  $L^2(\hat{\mu})$  by  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$ . Since  $Q$  only acts on the second component, as in Theorem E.3, this modified dynamics is still a lift of the process  $(Z(t))_{t \geq 0}$  solving the discrete stochastic heat equation (E.7). In particular, the lower bound on the relaxation time in Theorem E.4 still applies to the modified dynamics. If  $Q = \Pi$ , where

$$\Pi f(\ell, x) = \frac{1}{n} \sum_{i=1}^n f(\ell, i),$$

then the second component is resampled from the uniform distribution on  $\mathbb{Z}_n$  after independent  $\text{Exp}(\gamma)$ -distributed waiting times. In this case, we can derive the following bound on convergence to equilibrium.

**Theorem E.6.** *If  $Q = \Pi$ , then the transition semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  satisfies*

$$\|\hat{P}_t^{(\gamma)} f\|_{L^2(\hat{\mu})} \leq e^{-\nu(t-T)} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } f \in L_0^2(\hat{\mu}) \text{ and } t \geq 0 \quad (\text{E.14})$$

with  $T \in O(n)$  and inverse convergence rate

$$\nu^{-1} \in O(\gamma n^3 + \gamma^{-1} n).$$

In particular, for  $\gamma \propto n^{-1}$ , the  $L^2$ -relaxation time is of order  $O(n^2)$ .

We defer the proof of [Theorem E.6](#) to [Section E.3.1](#). The point is that not only the process  $(X^{(\gamma)}(t))$  but the joint process  $(\mathcal{L}^{(\gamma)}(t), X^{(\gamma)}(t))$  converges to equilibrium with rate  $\nu$ .

In the physics literature, it has been conjectured that even for  $\gamma = 0$ , the  $L^2$ -relaxation time of the self-repellent random walk is of order  $O(n^2)$ , see the remark above. Unfortunately, the upper bound in [Theorem E.6](#) degenerates in the limit  $\gamma \rightarrow 0$ .

A dynamics that is somehow closer to the self-repellent random walk is obtained if  $Q$  is the transition matrix of a simple random walk on  $\mathbb{Z}_n$ , i.e.  $Q(i, i \pm 1) = 1/2$  for all  $i \in \mathbb{Z}_n$ .

**Theorem E.7.** *If  $Q$  is the transition matrix of a simple random walk, then the corresponding transition semigroup  $(\hat{P}_t^{(\gamma)})_{t \geq 0}$  satisfies*

$$\|\hat{P}_t^{(\gamma)} f\|_{L^2(\hat{\mu})} \leq e^{-\nu(t-T)} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } f \in L_0^2(\hat{\mu}) \text{ and } t \geq 0$$

with  $T \in O(n^{3/2})$  and inverse convergence rate

$$\nu^{-1} \in O(\gamma n^3 + \gamma^{-1} n^2).$$

In particular, for  $\gamma \propto n^{-1/2}$ , the  $L^2$ -relaxation time is of order  $O(n^{5/2})$ .

[Theorem E.7](#) is proved in a similar way as [Theorem E.6](#), taking into account the spectral gap of  $Q$ , see the proof given in [Section E.3.1](#).

**Remark E.8** (Self-repellent motion). The *true self-repellent motion* was introduced by Tóth and Werner [\[243\]](#) as a locally self-repelling stochastic process in continuous space and time, providing a continuous space counterpart to self-repellent random walks [\[9, 241, 242\]](#). Heuristically, it can be described as a process  $(X(t))_{t \geq 0}$  with drift given by the negative spacial derivative of its own local time density  $L(t)$  at  $X(t)$ , even though the local time profile  $x \mapsto L(t, x)$  is not differentiable. Precise invariance principles, i.e. convergence in distribution of rescaled self-repellent random walks to the true self-repellent motion, have been shown in [\[155, 206\]](#). In the following, we give an informal presentation of how the self-repellent motion arises as a formal scaling limit of the process  $(\mathcal{L}(t), X(t))_{t \geq 0}$  considered above, and show consequences for the relaxation time of the self-repellent motion on the torus.

Consider the above dynamics on a discretisation of the one-dimensional torus  $\mathbb{T}_N = \mathbb{R}/(N\mathbb{Z})$  of length  $N$  with spacing  $h > 0$ , where we assume that  $n = N/h \in \mathbb{N}$ . More precisely, consider the dynamics  $(\mathcal{L}(t), X(t))_{t \geq 0}$  on  $\mathcal{S} \times \mathbb{Z}_n$  with generator

$$\hat{\mathfrak{L}}_h = \frac{1}{h} \mathfrak{T} + \frac{1}{h^2} \mathfrak{J}. \quad (\text{E.15})$$

The invariant probability measure is  $\hat{\mu}_h = \mu_h \otimes \text{Unif}(\mathbb{Z}_n)$  with

$$\mu_h(d\ell) \propto \exp(-U_h(\ell)) d\ell, \quad \text{where} \quad U_h(\ell) = \frac{h}{2} \sum_{i=1}^{N/h} \left( \frac{\ell_{i+1} - \ell_i}{h} \right)^2 = \frac{h}{2} \sum_{i=1}^n (\ell'_i)^2$$

and  $\ell'_i = \frac{1}{h}(\ell_{i+1} - \ell_i)$ . By [Lemma E.1](#), the symmetric and antisymmetric parts of  $\hat{\mathfrak{L}}_h$  are

$$\begin{aligned} \frac{1}{2}(\hat{\mathfrak{L}} + \hat{\mathfrak{L}}^*)f(\ell, i) &= \frac{1}{2h} \left( |\ell'_{i-1}|(f(\ell, i-1) - f(\ell, i)) + |\ell'_i|(f(\ell, i+1) - f(\ell, i)) \right) \quad \text{and} \\ \frac{1}{2}(\hat{\mathfrak{L}} - \hat{\mathfrak{L}}^*)f(\ell, i) &= \frac{1}{2h} \left( \ell'_{i-1}(f(\ell, i-1) - f(\ell, i)) - \ell'_i(f(\ell, i+1) - f(\ell, i)) + \mathfrak{T}f(\ell, i) \right). \end{aligned}$$

Viewing the dynamics  $(\mathcal{L}(t), X(t))_{t \geq 0}$  on  $\mathcal{S} \times \mathbb{Z}_n$  as a discretisation of a process on  $C_0(\mathbb{T}_N) \times \mathbb{T}_N$ , where  $C_0(\mathbb{T}_N)$  denotes the space of continuous functions on  $\mathbb{T}_N$  with mean zero, the above decomposition formally suggests that, in the limit  $h \downarrow 0$ , the process should behave like the solution to the SDE

$$\begin{aligned} d\mathcal{L}_t &= \delta_{X_t} dt \\ dX_t &= -\mathcal{L}'_t(X_t) dt + \sqrt{h\mathcal{L}'_t(X_t)} dB_t, \end{aligned}$$

where  $(B_t)_{t \geq 0}$  is a one-dimensional Brownian motion. In the formal continuum limit  $h \downarrow 0$ , the diffusion coefficient vanishes, and the result is reminiscent of the true-self repellent motion. In this case, since  $U_h(\ell)$  is a discretisation of  $U(\ell) = \int_{\mathbb{T}_N} (\ell'_t)^2 dt$ , the probability measures  $\mu_h$  are expected to converge to the law of a zero-mean Brownian bridge on  $\mathbb{T}_N$ . More rigorously, in the limit  $h \downarrow 0$ , the rescaled process

$$(\tilde{\mathcal{L}}(t), \tilde{X}(t))_{t \geq 0} = (h^{1/2}\mathcal{L}(th^{-3/2}), hX(th^{-3/2}))_{t \geq 0}$$

converges in law to the self-repellent motion on  $\mathbb{T}_N$  and its local time process, see [\[155, 206, 241, 243\]](#). The generator of the rescaled process arises from [\(E.15\)](#) as  $h^{-3/2}(\mathfrak{T} + \mathfrak{J}) = h^{-3/2}\hat{\mathfrak{L}}$  via this change of variables. In particular, suppose that  $(\hat{P}_t^{(h)})$  is the transition semigroup of the self-repellent random walk and its local time on  $\mathcal{S} \times \mathbb{Z}_n$  with  $n = N/h$ . Then one would expect that

$$t_{\text{rel}}^{\text{SRM}} = \lim_{h \rightarrow 0} h^{3/2} t_{\text{rel}}(\hat{P}^{(h)})$$

is the relaxation time of the self-repellent motion on  $\mathbb{T}_N$  and its local time process, provided the limit exists. Therefore, a relaxation time of the order  $t_{\text{rel}}(\hat{P}^{(h)}) \in \Theta(n^{3/2})$  would yield  $t_{\text{rel}}^{\text{SRM}} \in \Theta(N^{3/2})$ . Such a bound on the relaxation time is consistent with

the lower bound in [Theorem E.4](#), yet out of reach of the upper bound of order  $O(n^2)$  in [Theorem E.6](#). While a relaxation time of the order  $N^{3/2}$  for the self-repellent motion on  $\mathbb{T}_N$  is plausible from the superdiffusive scaling  $\mathbb{E}[|X_t|] = t^{2/3}\mathbb{E}[|X_1|]$  of the self-repellent motion on  $\mathbb{R}$ , we stress that a bound on the joint relaxation time of the self-repellent motion and its local time is a stronger statement.

## E.2. Application to Markov Chain Monte Carlo methods

It has recently been noted in the physics literature that the self-repellent random walk can be seen as a special instance of an Event Chain Monte Carlo method [\[188, 189\]](#). In Markov chain Monte Carlo (MCMC) methods, one constructs Markov processes with a given invariant probability measure to generate approximate samples from a target distribution. A standard approach is *Hamiltonian Monte Carlo (HMC)* which combines Hamiltonian dynamics with velocity resamplings [\[47\]](#). MCMC methods based on variants of HMC are widely used across many applications ranging from statistics [\[62\]](#) to molecular dynamics [\[60\]](#). *Event Chain Monte Carlo (ECMC)* is a more recently introduced class of MCMC methods based on non-reversible piecewise deterministic Markov processes that can be simulated in a rejection-free way by a thinning procedure [\[157\]](#). It was originally introduced in the context of hard disks [\[30\]](#) where it enabled the discovery of a previously unknown phase transition. Since then, it has been generalised to many other models [\[129, 169, 196, 197\]](#). The Markov processes underlying ECMC are characterised by moving one particle or coordinate at a time on a straight line, and ensuring invariance of the target probability measure by adding “events” at which the active particle or coordinate changes with some state-dependent rate.

For MCMC methods, the rate of convergence of the law of the process at time  $t$  towards its invariant distribution determines the efficiency of the scheme and is hence of crucial importance. Classical methods are mostly based on reversible Markov processes. Combined with local moves, reversibility limits the process to diffusive behaviour, prohibiting efficient exploration of the state space and hence impeding rapid convergence to stationarity. Non-reversible Markov processes such as ECMC possibly mitigate this issue. They can be constructed by lifting [\[66, 79, 99\]](#) a reversible dynamics through introduction of additional variables and thereby enlarging the state space. Mathematically, the analysis of these processes is challenging due to their non-reversibility and non-gradient nature, and rigorous results remain scarce or out of reach in situations of practical interest. For this reason, one is interested in studying toy models that enable a benchmark comparison of different sampling schemes [\[108, 158, 203\]](#).

### E.2.1. Sampling chains of oscillators

One class of such toy models are chains of oscillators where  $n$  particles are placed on a ring and each particle interacts only with its two neighbours. At inverse temperature

$\beta \in (0, \infty)$ , the corresponding Gibbs measure is

$$\mu_\beta(dy) \propto \exp(-\beta U(y)) dy, \quad \text{where} \quad U(y) = \sum_{i=1}^n W(y_{i+1} - y_i).$$

Here  $W: \mathbb{R} \rightarrow [0, \infty)$  is a continuously differentiable and symmetric interaction potential such that  $\exp(-\beta W)$  is integrable. Without confinement, the Gibbs measure is not normalisable. This can be fixed by pinning the measure, for example by restricting to the subspace  $\mathcal{S} \subset \mathbb{R}^n$  of configurations with mean zero.

We consider a version of the Event Chain Monte Carlo algorithm which, when applied to a Gibbs measure as above, simulates a piecewise deterministic Markov process  $(Y(t), X(t))_{t \geq 0}$  with state space  $\mathbb{R}^n \times \mathbb{Z}_n$ , where the position vector  $Y(t)$  increases in the component corresponding to the active particle  $X(t)$ , i.e.,

$$\frac{d}{dt} Y(t) = e_{X(t)},$$

and  $X(t)$  is resampled from the uniform distribution on  $\mathbb{Z}_n$  with rate  $\gamma \in [0, \infty)$  and performs nearest-neighbour jumps  $i \rightarrow i \pm 1$  with rates  $\beta \cdot (W'(y_{i\pm 1} - y_i))_-$  where  $y$  is the current value of  $Y(t)$ . The process  $\mathcal{Y}(t)$  with invariant distribution given by the Gibbs measure on  $\mathcal{S}$  is again obtained by projection, i.e.,

$$\mathcal{Y}_i(t) = Y_i(t) - n^{-1}t.$$

It can be verified similarly as in the proof of [Lemma E.1](#) that  $(\mathcal{Y}(t), X(t))$  is a piecewise deterministic Markov process with state space  $\mathcal{S} \times \mathbb{Z}_n$  and invariant measure  $\mu_\beta \otimes \text{Unif}(\mathbb{Z}_n)$ . Defining the directional derivatives  $\partial_{y_i}$  as in [\(E.3\)](#), the generator and its adjoint are given on functions  $f \in C_b^1(\mathcal{S} \times \mathbb{Z}_n)$  by

$$\begin{aligned} \hat{\mathcal{L}}f(y, i) &= \partial_{y_i} f(y, i) + \beta (W'(y_{i+1} - y_i))_- (f(y, i+1) - f(y, i)) \\ &\quad + \beta (W'(y_i - y_{i-1}))_+ (f(y, i-1) - f(y, i)) + \frac{\gamma}{n} \sum_{j=1}^n (f(y, j) - f(y, i)), \\ \hat{\mathcal{L}}^* f(y, i) &= -\partial_{y_i} f(y, i) + \beta (W'(y_{i+1} - y_i))_+ (f(y, i+1) - f(y, i)) \\ &\quad + \beta (W'(y_i - y_{i-1}))_- (f(y, i-1) - f(y, i)) + \frac{\gamma}{n} \sum_{j=1}^n (f(y, j) - f(y, i)), \end{aligned}$$

In contrast, randomised Hamiltonian Monte Carlo methods [\[47, 204\]](#) with exponentially distributed integration times are related to piecewise deterministic Markov processes  $(Q(t), V(t))$  where the deterministic motion is given by Hamiltonian dynamics, and the velocity component  $V(t)$  is resampled from a normal distribution after independent waiting times that are exponentially distributed with a fixed parameter  $\gamma \in (0, \infty)$ . For the chain of oscillators, we consider randomised HMC based on the Hamiltonian  $H(q, v) = U(q) + \frac{1}{2}|v|^2$ . The corresponding Hamiltonian dynamics

$$\frac{d}{dt} q_i(t) = v_i(t), \quad \frac{d}{dt} v_i(t) = W'(q_{i+1}(t) - q_i(t)) - W'(q_i(t) - q_{i-1}(t)), \quad i \in \mathbb{Z}_n, \quad (\text{E.16})$$

preserves the Boltzmann-Gibbs distribution

$$\hat{\mu}_\beta(dq dv) = \mu_\beta(dq) \mathcal{N}(0, \beta^{-1} \mathbf{I}_\mathcal{S})(dv) \propto \exp(-\beta H(q, v)) dq dv \quad \text{on } \mathcal{S} \times \mathcal{S}.$$

Therefore, the continuous time Markov process with generator

$$\begin{aligned} \mathfrak{H}f(q, v) &= \sum_{i=1}^n (v_i \partial_{q_i} f(q, v) + (W'(q_{i+1} - q_i) - W'(q_i - q_{i-1})) \partial_{v_i} f(q, v)) \\ &\quad + \gamma \int (f(q, w) - f(q, v)) \mathcal{N}(0, \beta^{-1} \mathbf{I}_\mathcal{S})(dw) \end{aligned}$$

has invariant measure  $\hat{\mu}_\beta$ . Different versions of randomised Hamiltonian Monte Carlo algorithms are obtained by replacing the Hamiltonian dynamics in (E.16) by a time discretisation, usually based on a symplectic integrator, and, possibly, including a Metropolis filter to adjust for the discretisation bias.

Note that since in Event Chain Monte Carlo, only one particle is moving at a time, the complexity for simulating the corresponding Markov process over a time unit can be assumed to be of order  $\Theta(1)$ . In contrast, for Hamiltonian Monte Carlo, the complexity for simulating the process over one time unit is at least of order  $n$ . More precisely, it is known that for the standard discretisation based on the Verlet scheme, the discretisation step size has to be chosen of order  $O(n^{-1/4})$  to adjust for the discretisation bias, resulting in an overall complexity of order  $\Omega(n^{5/4})$  per time unit [15, 31].

### E.2.2. The harmonic chain: Event Chain vs. Hamiltonian Monte Carlo

In order to compare the relaxation times for the continuous time Markov processes corresponding to ECMC and HMC, we consider the case of a harmonic chain of oscillators where  $W(y) = y^2$  and  $\beta = 1$ . In this case, the ECMC algorithm exactly recovers the local time process of the self-repellent random walk considered above. On the other hand, randomised HMC is a second-order lift of the discrete stochastic heat equation

$$dZ(t) = \frac{1}{2} \Delta_n Z(t) dt + dB(t) \tag{E.17}$$

with a Brownian motion  $B$  on  $\mathcal{S}$ , see [99, Example 3]. Recall that the local time process of the self-repellent walk is a second-order lift of (E.7) which describes the same Markov process as (E.17) but slowed down by a factor  $n$ . Due to this factor, the spectral gap  $m$  of the generator of (E.17) is of the order  $\Theta(n^2)$ , see (E.9). Based on the second order lift property, it has been shown in [99, 101, 186] that the  $L^2$  rate of convergence to stationarity of the Markov process with generator  $\mathfrak{H}$  is of the order  $\Theta(\min(\gamma, m\gamma^{-1}))$ . This result holds generally for convex potentials, and in the quadratic case it also follows by an explicit computation. In particular, for any choice of  $\gamma$ , the  $L^2$  relaxation time is at least of order  $\Omega(m^{-1/2}) = \Omega(n)$ , and the optimal order is achieved for  $\gamma \propto 1/n$ .

**Corollary E.9.** *For the harmonic chain of  $n$  particles, the computational complexity of the Event Chain Monte Carlo algorithm with particle refreshment at rate  $\gamma \propto n^{-1}$  is of the order  $O(n^2)$ . On the same model, the computational complexity of randomised HMC is of the order  $\Omega(n^{9/4})$  for any choice of  $\gamma$ .*

*Proof.* The complexity of the ECMC algorithm coincides with its relaxation time since the computational complexity per time unit is of order  $\Theta(1)$ , so that the complexity bound follows from the relaxation time bound in [Theorem E.6](#). Randomised HMC, on the other hand, has an overall complexity of order  $\Omega(n^{5/4})$  per time unit, see the discussion above. The relaxation time of order  $\Omega(n)$  thus yields a complexity of order  $\Omega(n^{9/4})$ .  $\square$

For the harmonic chain, ECMC thus exhibits an efficiency gain compared to more standard MCMC methods such as HMC, and a substantial gain in comparison to traditional MCMC methods based on reversible Markov processes. While similar gains have been observed empirically for chains of oscillators with Lennard-Jones interactions [\[170\]](#) and hard sphere models [\[169\]](#), so that this behaviour is conjectured to hold more generally [\[158\]](#), our result gives the first rigorous proof that ECMC methods can outperform HMC on specific models.

### E.3. Convergence to equilibrium

Convergence to equilibrium of non-reversible Markov processes can be quantified using a *flow Poincaré inequality*, which can be established for second-order lifts using the framework of [\[97, 101\]](#). For the modified dynamics with generator

$$\hat{\mathfrak{L}}^{(\gamma)} f = \hat{\mathfrak{L}} f + \gamma (Qf - f), \quad \text{Dom}(\hat{\mathfrak{L}}^{(\gamma)}) = \text{Dom}(\hat{\mathfrak{L}})$$

introduced in [\(E.13\)](#) above, the general framework simplifies and reduces to [Theorem E.10](#). Denote by  $(\mathcal{E}, \text{Dom}(\mathcal{E}))$  the Dirichlet form associated to  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$ , which is the extension of

$$\mathcal{E}(f, g) = \langle f, -\mathfrak{L}g \rangle_{L^2(\mu)}$$

to a closed symmetric bilinear form with domain  $\text{Dom}(\mathcal{E})$  given by the closure of  $\text{Dom}(\mathfrak{L})$  with respect to the norm  $\|f\|_{L^2(\mu)} + \mathcal{E}(f, f)^{1/2}$ . Recall that the probability measure  $\hat{\mu} = \mu \otimes \text{Unif}(\mathbb{Z}_n)$  is invariant for the Markov process generated by  $\hat{\mathfrak{L}}^{(\gamma)}$ .

**Theorem E.10.** *Assume that*

- (i) *the operator  $(\hat{\mathfrak{L}}, \text{Dom}(\hat{\mathfrak{L}}))$  is a second-order lift of  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$  such that  $g \circ \pi \in \text{Dom}(\hat{\mathfrak{L}}^*)$  and  $\hat{\mathfrak{L}}^*(g \circ \pi) = -\hat{\mathfrak{L}}(g \circ \pi)$  for all  $g \in \text{Dom}(\mathfrak{L})$ ;*
- (ii) *the operator  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$  has purely discrete spectrum on  $L^2(\mu)$  and a spectral gap  $m > 0$ , i.e.*

$$\|g - \mu(g)\|_{L^2(\mu)}^2 \leq \frac{1}{m} \mathcal{E}(g)$$

*for all  $f \in \text{Dom}(\mathcal{E})$ ;*

- (iii) *there exists a constant  $C_1 > 0$  such that*

$$\langle \hat{\mathfrak{L}}(g \circ \pi), \hat{\mathfrak{L}}(f - \Pi f) \rangle_{L^2(\hat{\mu})} \leq C_1 \|\mathfrak{L}g\|_{L^2(\mu)} \|f - Qf\|_{L^2(\hat{\mu})}$$

*for all  $f \in \text{Dom}(\hat{\mathfrak{L}})$  and  $g \in \text{Dom}(\mathfrak{L})$ ;*

(iv) the transition matrix  $Q$  has a spectral gap  $m_Q$ , i.e.

$$x^\top Qx \leq (1 - m_Q)|x|^2$$

for all  $x \in \mathbb{R}^n$  with  $\sum_{i=1}^n x_i = 0$ .

Then there exists a universal constant  $C > 0$  such that, for any  $T > 0$  and  $\gamma > 0$ , the semigroup  $(\hat{P}_t^{(\gamma)})$  generated by  $(\hat{\mathfrak{L}}^{(\gamma)}, \text{Dom}(\hat{\mathfrak{L}}^{(\gamma)}))$  satisfies

$$\|\hat{P}_t^{(\gamma)} f\|_{L^2(\hat{\mu})} \leq e^{-\nu(t-T)} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } t \geq 0 \text{ and } f \in L_0^2(\hat{\mu}),$$

where the rate  $\nu$  is given by

$$\nu = C \frac{\gamma}{\gamma^2/m + C_1^2 + \frac{1}{m_Q}(1 + \frac{1}{mT^2})}. \quad (\text{E.18})$$

*Proof.* The statement is a direct consequence of [97, Theorem 12], simplified by the fact that we consider a refreshment of the active particle determined by the transition matrix  $Q$ .  $\square$

Choosing  $T = m^{-1/2}$  in (E.18) yields the inverse convergence rate

$$\nu^{-1} \in O\left(\gamma m^{-1} + \gamma^{-1}(m_Q^{-1} + C_1^2)\right),$$

and optimising by choosing  $\gamma \propto \sqrt{m(m_Q^{-1} + C_1^2)}$  yields

$$\nu^{-1} \in O\left(m^{-1/2}(m_Q^{-1/2} + C_1)\right).$$

If  $Q = \Pi$ , we can choose  $m_Q = 1$  and obtain  $\nu^{-1} \in O(m^{-1/2}C_1)$  for the choice  $\gamma \propto m^{1/2}C_1$ .

### E.3.1. Upper bounds on relaxation times of self-repellent random walks

We turn to the proof of the upper bounds on the  $L^2$ -relaxation time of the modified self-repellent random walks given in Theorem E.6 and Theorem E.7.

**Lemma E.11.** *We have*

$$\frac{1}{n^2} \int \|\nabla^2 g\|_F^2 d\mu + \frac{1}{n^2} \sum_{i=1}^n \int ((\partial_{\ell_i} - \partial_{\ell_{i-1}})g)^2 d\mu = \int (2\mathfrak{L}g)^2 d\mu$$

where  $\|\nabla^2 g\|_F^2$  denotes the squared Frobenius norm of  $\nabla^2 g$ .

*Proof.* Since  $\mu(d\ell) \propto \exp(\frac{1}{2}\ell^\top \Delta_n \ell) d\ell$ , where  $d\ell$  denotes the Haar measure on  $\mathcal{S}$ , Bochner's identity yields

$$\int (2\mathfrak{L}g)^2 d\mu = \frac{1}{n^2} \int (\nabla^* \nabla g)^2 d\mu = \frac{1}{n^2} \int \|\nabla^2 g\|_F^2 d\mu + \frac{1}{n^2} \int \nabla g^\top \Delta_n \nabla g d\mu.$$

The claim then follows from the relation

$$\sum_{i=1}^n ((\partial_{\ell_i} - \partial_{\ell_{i-1}})g)^2 = \nabla g^\top \Delta_n \nabla g. \quad \square$$

**Lemma E.12.** For any  $v, w \in \mathcal{S}$  and  $g \in C_b^2(\mathcal{S})$ , we have

$$\int (v^\top \ell)^2 (\partial_w g)^2 \mu(d\ell) \leq 2v^\top (-\Delta_n)^{-1} v \int (\partial_w g)^2 d\mu + 4 \int (\partial_{\Delta_n^{-1}v, w}^2 g)^2 d\mu$$

where  $\Delta_n$  is the discrete Laplacian given by (E.4).

*Proof.* An integration by parts with respect to the Gaussian measure  $\mu$  gives

$$-\int v^\top \nabla f d\mu = \int v^\top \Delta_n \ell f(\ell) \mu(d\ell)$$

for all  $f \in C_b^1(\mathcal{S})$ . Applying this with  $f(\ell) = v^\top \Delta_n \ell (\partial_w g)^2$  and the inequality  $2ab \leq \frac{1}{2}a^2 + 2b^2$  yields

$$\begin{aligned} \int (v^\top \Delta_n \ell)^2 (\partial_w g)^2 \mu(d\ell) &= -\int v^\top \nabla (v^\top \Delta_n \ell (\partial_w g)^2) \mu(d\ell) \\ &= v^\top (-\Delta_n) v \int (\partial_w^2 g) d\mu - 2 \int v^\top \Delta_n \ell \partial_{vw}^2 g(\ell) \partial_w g(\ell) \mu(d\ell) \\ &\leq v^\top (-\Delta_n) v \int (\partial_w^2 g) d\mu + \frac{1}{2} \int (v^\top \Delta_n \ell)^2 (\partial_w g)^2 \mu(d\ell) + 2 \int (\partial_{v, w}^2 g)^2 d\mu. \end{aligned}$$

Rearranging and replacing  $v$  by  $\Delta_n^{-1}v$  yields the claim.  $\square$

*Proof of Theorem E.6.* The claim follows from Theorem E.10 if we can show that conditions (i)–(iv) are satisfied with  $m \in \Theta(n^{-3})$ ,  $C_1^2 \in O(n)$ , and  $m_Q = 1$ .

To this end, condition (i) is satisfied by Lemma E.1 and Theorem E.3. The operator  $(\mathfrak{L}, \text{Dom}(\mathfrak{L}))$  has discrete spectrum on  $L^2(\mu)$  and a spectral gap  $m \in \Theta(n^{-3})$  by (E.9), so that condition (ii) is satisfied. Condition (iv) is immediately satisfied with  $m_Q = 1$  in case  $Q = \Pi$ . It remains to verify condition (iii). By the Cauchy-Schwarz inequality, it suffices to bound

$$\|\hat{\mathfrak{L}}^* \hat{\mathfrak{L}}(g \circ \pi)\|_{L^2(\hat{\mu})} \leq C_1 \|\mathfrak{L}g\|_{L^2(\mu)}.$$

By Lemma E.1,

$$\begin{aligned} \hat{\mathfrak{L}}^* \hat{\mathfrak{L}}(g \circ \pi)(\ell, x) &= -\partial_{\ell_x}^2 g(\ell) - (\ell_x - \ell_{x-1})_- (\partial_{\ell_x} g(\ell) - \partial_{\ell_{x-1}} g(\ell)) \\ &\quad + (\ell_{x+1} - \ell_x)_+ (\partial_{\ell_{x+1}} g(\ell) - \partial_{\ell_x} g(\ell)). \end{aligned}$$

Hence

$$\begin{aligned} \|\hat{\mathfrak{L}}^* \hat{\mathfrak{L}}(g \circ \pi)\|_{L^2(\hat{\mu})}^2 &= \frac{1}{n} \sum_{i=1}^n \int ((\hat{\mathfrak{L}}^* \hat{\mathfrak{L}}(g \circ \pi))(\ell, i))^2 \mu(d\ell) \\ &\leq \frac{3}{n} \sum_{i=1}^n \int (\partial_{\ell_i}^2 g(\ell))^2 \mu(d\ell) + \frac{3}{n} \sum_{i=1}^n \int (\ell_i - \ell_{i-1})^2 (\partial_{\ell_i} g(\ell) - \partial_{\ell_{i-1}} g(\ell))^2 \mu(d\ell). \end{aligned}$$

For the second summand, we apply Lemma E.12 with  $v_i = w_i = e_i - e_{i-1}$  to obtain

$$\frac{3}{n} \sum_{i=1}^n \int (\ell_i - \ell_{i-1})^2 (\partial_{\ell_i} g(\ell) - \partial_{\ell_{i-1}} g(\ell))^2 \mu(d\ell)$$

$$\leq \frac{6}{n} \sum_{i=1}^n v_i^\top (-\Delta_n)^{-1} v_i \int ((\partial_{\ell_i} - \partial_{\ell_{i-1}})g)^2 d\mu + \frac{12}{n} \sum_{i=1}^n \int (\partial_{\Delta_n^{-1}v_i, v_i}^2 g)^2 d\mu.$$

Note that  $u_i = \Delta_n^{-1}v_i$  is given by

$$(u_i)_j = \left(1 - \frac{1}{n}\right) \cdot \begin{cases} \frac{1}{2} & \text{if } j = i-1, \\ -\frac{1}{2} & \text{if } j = i, \\ -\frac{1}{2} + \frac{k}{n-1} & \text{if } j = i+k, 1 \leq k \leq n-2. \end{cases}$$

In particular,  $u_{i+1} - u_i = e_i - \frac{1}{n}\mathbf{1} = \tilde{e}_i$  is the projection of the  $i$ -th basis vector in  $\mathbb{R}^n$  onto  $\mathcal{S}$ ,

$$v_i^\top (-\Delta_n)^{-1} v_i = -v_i^\top u_i = (u_i)_i - (u_i)_{i-1} = 1 - \frac{1}{n},$$

and, since  $u_i = \sum_{j=1}^n (u_i)_j e_j = \sum_{j=1}^n (u_i)_j \tilde{e}_j$  and  $v_i = \tilde{e}_i - \tilde{e}_{i-1}$ ,

$$\partial_{u_i, v_i}^2 g = \sum_{j=1}^n (u_i)_j \partial_{\tilde{e}_i, \tilde{e}_j}^2 g - \sum_{j=1}^n (u_i)_j \partial_{\tilde{e}_{i-1}, \tilde{e}_j}^2 g = \sum_{j=1}^n (u_i - u_{i+1})_j \partial_{\tilde{e}_i, \tilde{e}_j}^2 g = -\partial_{\tilde{e}_i, \tilde{e}_i}^2 g.$$

In the final equality we used that  $u_i - u_{i+1} = -\tilde{e}_i$  and  $\sum_{j=1}^n (\tilde{e}_i)_j \tilde{e}_j = \sum_{j=1}^n (\tilde{e}_i)_j e_j = \tilde{e}_i$ . Therefore,

$$\begin{aligned} \|\hat{\mathfrak{L}}^* \hat{\mathfrak{L}}(g \circ \pi)\|_{L^2(\hat{\mu})}^2 &\leq \frac{6}{n} \sum_{i=1}^n \int ((\partial_{\ell_i} - \partial_{\ell_{i-1}})g)^2 d\mu + \frac{15}{n} \sum_{i=1}^n \int (\partial_{\ell_i}^2 g)^2 d\mu \\ &\leq 60n \int (\mathfrak{L}g)^2 d\mu \end{aligned}$$

by [Lemma E.11](#), so that assumption (iii) of [Theorem E.10](#) is satisfied with  $C_1 = \sqrt{60n}$ .  $\square$

*Proof of [Theorem E.7](#).* As in the case of [Theorem E.6](#), the claim follows from [Theorem E.10](#) with  $m_Q \in O(n^{-2})$ .  $\square$

## Acknowledgments

The authors were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2047 – 390685813 as well as under CRC 1720 – 539309657. We would like to thank Werner Krauth, Tony Maggs, Brune Massoulié and Stefan Oberdörster for helpful discussions.

## Appendix F.

# Convergence rates of self-repelling diffusions on Riemannian manifolds

F. Lörler. *Convergence rates of self-repelling diffusions on Riemannian manifolds*. Nov. 2025. arXiv: [2511.23333](#) [math.PR].

**Abstract:** We study a class of self-repelling diffusions on compact Riemannian manifolds whose drift is the gradient of a potential accumulated along their trajectory. When the interaction potential admits a suitable spectral decomposition, the dynamics and its environment are equivalent to a finite-dimensional degenerate diffusion. We show that this diffusion is a second-order lift of an Ornstein-Uhlenbeck process whose invariant law corresponds to the Gaussian invariant measure of the environment, and immediately obtain a general upper bound on the rate of convergence to stationarity using the framework of second-order lifts. Furthermore, using a flow Poincaré inequality, we develop lower bounds on the convergence rate. We show that, in the periodic case, these lower bounds improve upon those of Benaïm and Gauthier [24], and even match the order of the upper bound in some cases.

**Keywords:** Self-repelling diffusions; hypocoercivity; lift; convergence rate.

**MSC Subject Classification:** 58J65, 60J60, 60J55, 60H10, 37A30.

### F.1. Introduction

We consider a class of self-interacting diffusions  $(X_t)_{t \geq 0}$  on compact, connected and oriented smooth Riemannian manifolds  $M$ . Given a smooth and symmetric interaction potential  $V: M \times M \rightarrow \mathbb{R}$ , the process  $(X_t)_{t \geq 0}$  solves the  $M$ -valued stochastic differential equation

$$dX_t = -\nabla_M V_t(X_t) dt + \sigma \circ dB_t, \quad X_0 \sim \mu_0, \quad (\text{F.1})$$

where  $\sigma \geq 0$  is the diffusivity,  $(B_t)_{t \geq 0}$  is a Brownian motion on  $M$ , and  $\nabla_M$  denotes the Riemannian gradient on  $M$ . The drift is the gradient of the accumulated potential along the trajectory up to time  $t$ , i.e.

$$V_t(x) = \int_0^t V(X_s, x) ds = \int_M V(y, x) \mu_t(dy), \quad \text{where } \mu_t = \int_0^t \delta_{X_s} ds, \quad (\text{F.2})$$

and the initial law  $\mu_0$  is a probability measure on  $M$ . The process  $(\nabla_M V_t)_{t \geq 0}$  is also known as the environment process.

Such self-repelling processes were originally motivated by and introduced as models for polymer growth, notably by Norris, Rogers and Williams [208, 209], where the asymptotic growth of the polymer is of particular interest. In case  $M = \mathbb{R}^d$ , the asymptotic behaviour in the form of strong laws of large numbers has been studied in the works [72, 93, 201]. In case that  $M$  is compact, as considered here, one expects the law of  $X_t$  to converge to an invariant probability measure. Indeed, under a variety of different assumptions, such a convergence has been shown using the normalised occupation time measure  $\frac{1}{t}\mu_t$  in place of  $\mu_t$  in the series of works [25, 26, 27], and with the non-normalised occupation time measure in [23, 24].

Self-interacting processes such as the process  $(X_t)_{t \geq 0}$  are not Markovian since they depend on their own history through their occupation measure, rendering their study notoriously difficult. However, the joint process  $(V_t, X_t)_{t \geq 0}$  is a Markov process, and it has been shown by Tarrès, Tóth and Valkó [239] that, under suitable assumptions on the interaction potential  $V$ , the combined process  $(x \mapsto \nabla_M V_t(X_t + x))_{t \geq 0}$ , known as the environment as seen from the particle, admits an ergodic invariant Gaussian measure. Following on this work, Benaïm and Gauthier [24] considered the case that the interaction potential admits a finite diagonal decomposition in terms of eigenfunctions of the Laplace-Beltrami operator  $\Delta_M$ , see [Assumption F.1](#) below. In this case, the process is *self-repelling*, and (F.1) is equivalent to a finite-dimensional system of stochastic differential equations whose solution is a degenerate, non-reversible diffusion process on  $\mathbb{R}^n \times M$ , with invariant probability measure given by the product of a Gaussian and the uniform measure on  $M$ , as shown by [24].

Even though the invariant probability measure admits a simple and explicit description, obtaining rates of convergence of the law of the process to stationarity is challenging due to its non-reversibility and degeneracy. Classical analytical techniques based on Poincaré inequalities and spectral estimates cannot be applied in the non-reversible, degenerate setting, necessitating the development of completely new techniques in this area of research that has become known as hypocoercivity [248]. In the seminal work [87], Dolbeault, Mouhot and Schmeiser introduced a framework for the study of a large class of hypocoercive kinetic equations. This framework was further developed by [120] and then applied to self-repelling diffusions on Riemannian manifolds by Benaïm and Gauthier [24]. The DMS approach of [87], however, fails to provide rates of convergence to stationarity that are sharp in terms of the model parameters in many settings. Indeed, for Langevin dynamics (or the associated kinetic Fokker-Planck equation), a prototypical non-reversible degenerate dynamics, sharp rates of convergence have only recently been established in [61] using a variational approach to hypocoercivity based on space-time Poincaré inequalities, introduced by Albritton, Armstrong, Mourrat and Novack [6]. This approach was further developed and made broadly applicable using the framework of second-order lifts [97, 99] that allows to prove such space-time Poincaré inequalities by relating the non-reversible dynamics to simpler, reversible diffusions.

In this work, we identify the self-repelling diffusion and its environment as a second-order lift of an Ornstein-Uhlenbeck process with invariant law corresponding to the Gaus-

sian measure that is the invariant law of the environment process. This provides a new perspective on these processes and allows us to apply the framework of non-reversible lifts of Markov processes [97, 99] to obtain upper and lower bounds on their rate of convergence to equilibrium. Our upper bounds improve on those of [24] obtained using the DMS approach, as demonstrated by an example on the torus, and are even sharp in some settings.

We begin by introducing self-interacting diffusions and the equivalent finite-dimensional system in Section F.2 and show in Section F.3 that, under Assumption F.1, self-repelling diffusions on Riemannian manifolds are second-order lifts of an Ornstein-Uhlenbeck process on  $\mathbb{R}^n$  with Gaussian invariant probability measure. The resulting upper and lower bounds on the  $L^2$ -relaxation time of the associated transition semigroups are given in Section F.3.3. We compare our upper bounds to those of [24] in the periodic case in Section F.3.4. Finally, we conclude by giving the proof of the upper bounds on the relaxation time in Section F.4.

## F.2. Self-interacting diffusions on Riemannian manifolds

Consider a compact, connected and oriented smooth Riemannian manifold  $(M, g)$  without boundary and a smooth, symmetric function  $V : M \times M \rightarrow \mathbb{R}$ . A *self-interacting diffusion with potential  $V$*  is the  $M$ -valued continuous-time stochastic process  $(X_t)_{t \geq 0}$  on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  that solves the stochastic differential equation

$$dX_t = -\nabla_M V_t(X_t) dt + \sigma \circ dB_t, \quad X_0 \sim \mu_0,$$

Here  $\sigma \geq 0$  is the diffusivity and  $(B_t)_{t \geq 0}$  is a Brownian motion on  $M$ , the drift is the gradient of the accumulated potential as given in (F.2), and the initial law  $\mu_0$  is a probability measure on  $M$ . In the following, we let  $\nu$  denote the Riemannian volume measure on  $M$ . We will work under the following assumption.

**Assumption F.1.**  $V$  takes the form

$$V(x, y) = \sum_{k=1}^n a_k e_k(x) e_k(y), \quad (\text{F.3})$$

where  $(e_1, \dots, e_n)$  is a system of  $L^2(M, \nu)$ -orthonormal eigenfunctions of the Laplace-Beltrami operator  $\Delta_M$  on  $M$ , with associated non-positive eigenvalues  $\lambda_1, \dots, \lambda_n$ , and  $a_1, \dots, a_n \in [0, \infty)$ .

Assumption F.1 ensures that  $V$  is a symmetric and positive semi-definite kernel, i.e.

$$\int_M \int_M V(x, y) f(x) f(y) \nu(dx) \nu(dy) \geq 0$$

for all  $f \in L^2(M, \nu)$ , so that the process  $(X_t)_{t \geq 0}$  is *self-repelling*. Moreover, as shown in [24], the process  $(\nabla_M V_t, X_t)_{t \geq 0}$  admits an explicit invariant probability measure supported on

$$\text{span}\{\nabla_M e_k : k \in \{1, \dots, n\}\} \times M \subseteq \mathfrak{X}(M) \times M,$$

where  $\mathfrak{X}(M)$  denotes the set of all smooth vector fields on  $M$ , see [Section F.2.1](#) below.

**Remark F.1.** If  $M = \mathbb{T}^d$ ,  $\mu_t$  has a density  $L_t$  with respect to the volume measure and  $V(x, y) = W(x - y)$  for some symmetric function  $W: \mathbb{T}^d \rightarrow \mathbb{R}$ , the accumulated potential simplifies to

$$V_t(x) = \int_{\mathbb{T}^d} W(y - x) L_t(y) \nu(dy) = (W * L_t)(x).$$

In this case,  $(X_t)$  is a diffusion driven by the mollified negative gradient of its own local time density  $L_t$ . In the non-compact case  $M = \mathbb{R}^d$ , corresponding processes are known as self-repelling Brownian polymers [\[93\]](#), which have recently been constructed even in the singular case  $W = \delta_0$  [\[115\]](#). They are phenomenologically closely related to the true self-repelling motion introduced by Tóth and Werner [\[243\]](#).

### F.2.1. Reduction to a finite-dimensional system

In this section, we follow the construction in [\[24, Section 2\]](#) that shows how the structure of  $V$  given in [Assumption F.1](#) allows to reduce the self-repelling diffusion and its environment to a system of stochastic differential equations in  $\mathbb{R}^n \times M$ .

[Assumption F.1](#) allows us to decompose

$$\nabla_M V_t(x) = \sum_{k=1}^n a_k \nabla_M e_k(x) U_t^{(k)} \quad \text{with} \quad U_t^{(k)} = \int_0^t e_k(X_s) ds.$$

The self-repelling diffusion [\(F.1\)](#) and its environment are hence equivalent to the system

$$\begin{aligned} dU_t^{(k)} &= e_k(X_t) dt, \quad k = 1, \dots, n \\ dX_t &= - \sum_{k=1}^n a_k \nabla_M e_k(X_t) U_t^{(k)} dt + \sigma \circ dB_t \end{aligned}$$

of stochastic differential equations on  $\mathbb{R}^n \times M$ , where  $\circ dB_t$  denotes Stratonovich integration. Let  $U_t = (U_t^{(k)})_{k=1, \dots, n}$  and  $e(x) = (e_k(x))_{k=1, \dots, n}$ , and introduce the potential

$$\Phi(u) = \frac{1}{2} \sum_{k=1}^n a_k |\lambda_k| u_k^2, \quad u \in \mathbb{R}^n.$$

Then this system can equivalently be written as

$$\begin{aligned} dU_t &= e(X_t) dt \\ dX_t &= -\nabla \Phi(U_t)^\top \Lambda^{-1} \nabla_M e(X_t) dt + \sigma \circ dB_t, \end{aligned} \tag{F.4}$$

where  $\Lambda = \text{diag}(|\lambda_1|, \dots, |\lambda_n|) \in \mathbb{R}^{n \times n}$  and  $\nabla_M e(x) = (\nabla_M e_k(x))_{k=1, \dots, n}$ . The drift is thus to be understood as

$$\nabla \Phi(U_t)^\top \Lambda^{-1} \nabla_M e(X_t) = \sum_{k=1}^n \partial_k \Phi(U_t) |\lambda_k|^{-1} \nabla_M e_k(X_t).$$

Note that, to avoid confusion, we use  $\nabla$  for the Euclidean gradient in  $\mathbb{R}^n$  and  $\nabla_M$  for the gradient on  $M$ .

We summarise some properties of the system (F.4) that are given in [24] in the following lemma.

**Lemma F.2.** (i) *There exists a unique global strong solution  $(U_t, X_t)_{t \geq 0}$  to (F.4) for any initial condition  $(U_0, X_0) = (u, x) \in \mathbb{R}^d \times M$ .*

(ii) *The probability measure*

$$\hat{\mu} = \mu \otimes \kappa$$

with

$$\mu(\mathrm{d}u) = \frac{1}{Z} \exp(-\Phi(u)) \mathrm{d}u \quad \text{and} \quad \kappa(\mathrm{d}x) = \frac{1}{\nu(M)} \nu(\mathrm{d}x),$$

where  $Z$  is a normalising constant, is invariant for this process. If  $\sigma > 0$ , it is the unique invariant probability measure.

*Proof.* This is the content of [24, Proposition 1] and [24, Theorem 5].  $\square$

Note that  $\Phi$  is quadratic, so that

$$\mu = \bigotimes_{k=1}^n \mathcal{N}\left(0, \frac{1}{a_k |\lambda_k|}\right)$$

is a Gaussian distribution. If  $\sigma = 0$  the invariant probability measure is possibly not unique, as there may exist infinitely many ergodic probability measures; see [24, Theorem 3] for an explicit example on the torus.

By Lemma F.2, the process  $(U_t, X_t)$  induces a strongly continuous contraction semi-group  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  on  $L^2(\hat{\mu})$ , and we denote its generator by  $(\hat{\mathcal{L}}^{(\sigma)}, \text{Dom}(\hat{\mathcal{L}}^{(\sigma)}))$ . By [24, Propositions 2 and 3], the set  $C_c^2(\mathbb{R}^n \times M)$  of compactly supported, twice differentiable functions on  $\mathbb{R}^n \times M$  is a core for the generator, and on these functions it acts as

$$\begin{aligned} \hat{\mathcal{L}}^{(\sigma)} f(u, x) &= \sum_{k=1}^n (e_k(x) \partial_{u_k} f(u, x) - a_k u_k \langle \nabla_M e_k(x), \nabla_M f(u, x) \rangle) + \frac{\sigma^2}{2} \Delta_M f(u, x) \\ &= e(x)^\top \nabla f(u, x) - \langle \nabla \Phi(u)^\top \Lambda^{-1} \nabla_M e(x), \nabla_M f(u, x) \rangle + \frac{\sigma^2}{2} \Delta_M f(u, x). \end{aligned} \tag{F.5}$$

**Example F.3.** Consider  $M = \mathbb{T}_L = \mathbb{R}/(2\pi L\mathbb{Z})$  the circle with circumference  $2\pi L$  and

$$W(z) = \frac{1}{\pi L} \sum_{k=1}^n a_k \cos(kz/L), \quad z \in \mathbb{T}_L,$$

where  $a_1, \dots, a_n \in [0, \infty)$ . Then setting  $V(x, y) = W(x - y)$  yields

$$V(x, y) = \frac{1}{\pi L} \sum_{k=1}^n a_k \cos(kx/L) \cos(ky/L) + a_k \sin(kx/L) \sin(ky/L),$$

which is of the form (F.3) with the orthonormal functions  $e_k(x) = (\pi L)^{-1/2} \cos(kx/L)$  and  $f_k(x) = (\pi L)^{-1/2} \sin(kx/L)$ . The associated eigenvalues are  $\lambda_k = -\frac{k^2}{L^2}$ . The potential  $\Phi$  then becomes

$$\Phi(u, v) = \frac{1}{2L^2} \sum_{k=1}^n a_k k^2 (u_k^2 + v_k^2), \quad u, v \in \mathbb{R}^n,$$

where  $u_k$  and  $v_k$  correspond to the eigenfunctions  $e_k$  and  $f_k$ . We will return to this example in Section F.3.4.

### F.3. Relaxation times of self-interacting diffusions

The process  $(U_t, X_t)_{t \geq 0}$  is non-reversible and subject to degenerate noise acting only on  $X_t$ . Therefore, studying the rate of convergence of the law of  $(U_t, V_t)$  to its invariant distribution  $\hat{\mu}$  is a challenging task and falls within the field of hypocoercivity. In this section, we will apply the recently developed framework of second-order lifts and flow Poincaré inequalities to obtain upper and lower bounds on this rate of convergence in the form of bounds on the  $L^2$ -relaxation time.

We begin by recalling the concept of second-order lifts in Section F.3.1, and show how self-repelling diffusions and their environment can be seen as lifts of an Ornstein-Uhlenbeck process with invariant measure  $\mu$  in Section F.3.2. We then apply the framework of [97] to obtain upper and lower bounds on the  $L^2$ -relaxation time of the transition semigroups associated to the self-repelling diffusions in Section F.3.3. These are compared to those obtained by [24] in the periodic case in Section F.3.4.

#### F.3.1. Second-order lifts

Inspired by the notion of lifts of Markov chains in discrete time [66, 79], second-order lifts of reversible diffusions were introduced in [99] as a related counterpart in continuous time and space. They provide a broadly applicable and systematic approach developed in [97, 99, 101] to proving convergence to equilibrium for many non-reversible dynamics by relating them to simpler, reversible diffusions.

Intuitively, a time-homogeneous Markov process  $(\hat{Z}_t)_{t \geq 0}$  on a product space  $\mathcal{S} \times \mathcal{V}$  with invariant probability measure  $\hat{\mu} = \mu \otimes \kappa$  is a second-order lift of another Markov process  $(Z_t)_{t \geq 0}$  with state space  $\mathcal{S}$  and invariant probability measure  $\mu$  if their associated transition semigroups  $(\hat{P}_t)$  and  $(P_t)$  in  $L^2(\hat{\mu})$  and  $L^2(\mu)$ , respectively, satisfy

$$\int_{\mathcal{V}} \hat{P}_t(f \circ \pi)(x, v) \kappa(dv) = (P_t f)(x) + o(t^2) \quad \text{as } t \downarrow 0$$

for functions  $f: \mathcal{S} \rightarrow \mathbb{R}$  in the domain of the generator  $\mathcal{L}$  of  $(P_t)$ . Here  $\pi: \mathcal{S} \times \mathcal{V} \rightarrow \mathcal{S}$  denotes the canonical projection  $\pi(x, v) = x$ . Considering a second-order Taylor expansion around  $t = 0$  leads to the precise definition below, see [99] for more intuition behind this property and some examples. Let  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  and  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  denote the

generators of  $(P_t)$  and  $(\hat{P}_t)$  in  $L^2(\mu)$  and  $L^2(\hat{\mu})$ , respectively. We denote the Dirichlet form associated to  $\mathcal{L}$  by  $\mathcal{E}$ , it is the extension of

$$\mathcal{E}(f, g) = - \int_{\mathcal{S}} f \mathcal{L} g \, d\mu \quad (\text{F.6})$$

to a closed symmetric bilinear form with domain  $\text{Dom}(\mathcal{E})$  given by the closure of  $\text{Dom}(\mathcal{L})$  with respect to the norm  $\|f\|_{L^2(\mu)} + \mathcal{E}(f, f)^{1/2}$ .

**Definition F.4** (Second-order lifts [99]). The process  $(\hat{Z}_t)$  is a *second-order lift* of  $(Z_t)$  if there exists a core  $C$  of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  such that

$$f \circ \pi \in \text{Dom}(\hat{\mathcal{L}}) \text{ for all } f \in C \quad (\text{F.7})$$

and for all  $f, g \in C$  we have

$$\langle \hat{\mathcal{L}}(f \circ \pi), g \circ \pi \rangle_{L^2(\hat{\mu})} = 0, \quad (\text{F.8})$$

and

$$\frac{1}{2} \langle \hat{\mathcal{L}}(f \circ \pi), \hat{\mathcal{L}}(g \circ \pi) \rangle_{L^2(\hat{\mu})} = - \langle f, \mathcal{L} g \rangle_{L^2(\mu)} = \mathcal{E}(f, g). \quad (\text{F.9})$$

Conversely, we refer to the process  $(Z_t)$  as the *collapse* of  $(\hat{Z}_t)$ .

We also say that  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  or  $(\hat{P}_t)$  is a second-order lift of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  or  $(P_t)$ , respectively. In the rest of this paper, we simply refer to second-order lifts as lifts.

### F.3.2. Self-interacting diffusions as second-order lifts

In this section, we show that the process  $(U_t, X_t)_{t \geq 0}$  that is equivalent to the self-repelling diffusion and its environment is a second-order lift of an Ornstein-Uhlenbeck process on  $\mathbb{R}^n$  for any choice of diffusivity. To this end, we consider the transition semigroup  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  of  $(U_t, X_t)_{t \geq 0}$  with diffusivity  $\sigma$  on  $L^2(\hat{\mu})$ , and recall that, for compactly supported functions  $f \in C_c^2(\mathbb{R}^n \times M)$ , the associated generator is given by

$$\begin{aligned} \hat{\mathcal{L}}^{(\sigma)} f(u, x) &= \sum_{k=1}^n (e_k(x) \partial_{u_k} f(u, x) - a_k u_k \langle \nabla_M e_k(x), \nabla_M f(u, x) \rangle) + \frac{\sigma^2}{2} \Delta_M f(u, x) \\ &= e(x)^\top \nabla f(u, x) - \langle \nabla \Phi(u)^\top \Lambda^{-1} \langle \nabla_M e(x), \nabla_M f(u, x) \rangle + \frac{\sigma^2}{2} \Delta_M f(u, x), \end{aligned}$$

where  $\Lambda = \text{diag}(|\lambda_1|, \dots, |\lambda_n|)$ . In the following, we let  $\pi(u, x) = u$ ,

$$(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}})) = (\hat{\mathcal{L}}^{(0)}, \text{Dom}(\hat{\mathcal{L}}^{(0)})) \quad \text{and} \quad (\hat{P}_t)_{t \geq 0} = (\hat{P}_t^{(0)})_{t \geq 0}.$$

We will show that the self-repelling diffusion  $(U_t, X_t)_{t \geq 0}$  is a second-order lift of the reversible diffusion process  $(Z_t)_{t \geq 0}$  on  $\mathbb{R}^n$  that solves the stochastic differential equation

$$dZ_t = -\frac{1}{2\nu(M)} \nabla \Phi(Z_t) \, dt + \frac{1}{\sqrt{\nu(M)}} \, dW_t, \quad (\text{F.10})$$

where  $(W_t)_{t \geq 0}$  is an  $n$ -dimensional standard Brownian motion. We denote the transition semigroup of  $(Z_t)_{t \geq 0}$  on  $L^2(\mu)$  by  $(P_t)_{t \geq 0}$ . Its generator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  is given by

$$\mathcal{L}f = -\frac{1}{2\nu(M)} \nabla^* \nabla f = -\frac{1}{2\nu(M)} \nabla \Phi \cdot \nabla f + \frac{1}{2\nu(M)} \Delta f, \quad (\text{F.11})$$

where the adjoint is in  $L^2(\mu)$ , and  $\text{Dom}(\hat{\mathcal{L}}) = H^{2,2}(\mathbb{R}^n, \mu)$ . Note that (F.10) is an Ornstein-Uhlenbeck process on  $\mathbb{R}^n$  with invariant probability measure  $\mu$ , slowed down by a factor  $\frac{1}{2\nu(M)}$ . In particular, the spectral gap of the generator in  $L^2(\mu)$  is

$$\text{gap}(\mathcal{L}) = \frac{1}{2\nu(M)} \min\{a_k |\lambda_k| : k \in \{1, \dots, n\}\}, \quad (\text{F.12})$$

see [193].

**Theorem F.5.** *The semigroup  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  is a second-order lift of  $(P_t)_{t \geq 0}$  for any choice of diffusivity  $\sigma \geq 0$ .*

*Proof.* It suffices to prove the claim for  $\sigma = 0$ . Firstly,  $C_c^\infty(\mathbb{R}^n)$  forms a core for  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  [251]. Hence let  $g \in C_c^\infty(\mathbb{R}^n)$ . This implies  $g \circ \pi \in \text{Dom}(\hat{\mathcal{L}})$  and

$$\int_M \hat{\mathcal{L}}(g \circ \pi)(u, x) \kappa(dx) = \int_M e(x)^\top \nabla f(u) \kappa(dx) = 0$$

since  $M$  has no boundary. Furthermore, since the  $e_k$  are orthonormal with respect to  $\nu$ ,

$$\frac{1}{2} \int_M (\hat{\mathcal{L}}(g \circ \pi)(u, x))^2 \kappa(dx) = \frac{1}{2} \int_M e(x)^\top \nabla g(u) \nabla g(u)^\top e(x) \kappa(dx) = \frac{1}{2\nu(M)} |\nabla g(u)|^2,$$

so that

$$\frac{1}{2} \int_{\mathbb{R}^n \times M} (\hat{\mathcal{L}}(g \circ \pi))^2 d\hat{\mu} = \frac{1}{2\nu(M)} \int_{\mathbb{R}^n} |\nabla g|^2 d\mu = - \int_{\mathbb{R}^n} g \mathcal{L}g d\mu,$$

and (F.9) follows by polarisation.  $\square$

### F.3.3. Bounds for relaxation times

We first note that the representation of the self-repelling diffusions as a second-order lift of the Ornstein-Uhlenbeck process (F.10) directly yields a lower bound on the  $L^2$ -relaxation time

$$t_{\text{rel}}(\hat{P}) = \inf \left\{ t \geq 0 : \|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq e^{-1} \|f\|_{L^2(\hat{\mu})} \text{ for all } f \in L_0^2(\hat{\mu}) \right\}$$

of the semigroup  $(\hat{P}_t)$ . Here  $L_0^2(\hat{\mu})$  denotes the orthogonal complements of the constant functions in  $L^2(\hat{\mu})$ . Let us stress that, since  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  are the transition semigroups of the processes  $(U_t, X_t)_{t \geq 0}$  with diffusivity  $\sigma$ , all relaxation time bounds concern the joint relaxation time of the environment parametrised by  $U_t$  and the self-repelling diffusion  $X_t$  itself.

**Theorem F.6** (Lower bound on relaxation time). *For the transition semigroup  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  associated to the self-repelling diffusion with diffusivity  $\sigma \geq 0$ ,*

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \geq \left( \frac{\nu(M)}{4 \min\{a_k |\lambda_k| : k \in \{1, \dots, n\}\}} \right)^{1/2}.$$

*Proof.* This is a consequence of [Theorem F.5](#) above and [\[99, Theorem 11\]](#) which states that  $t_{\text{rel}}(\hat{P}) \geq \sqrt{t_{\text{rel}}(P)}/(2\sqrt{2})$  holds whenever  $P = (P_t)_{t \geq 0}$  and  $\hat{P} = (\hat{P}_t)_{t \geq 0}$  are the transition semigroups of a Markov process and an arbitrary second order lift. By reversibility, the relaxation time of the Ornstein-Uhlenbeck process [\(F.10\)](#) is the inverse of the spectral gap of its generator  $\mathcal{L}$ . Therefore, [\(F.12\)](#) yields the claim.  $\square$

Obtaining quantitative upper bounds on the relaxation time is a lot more involved. Second-order lifts provide a framework developed in [\[97, 99, 101\]](#) for a systematic approach to this problem, based on a variational approach to hypocoercivity first introduced in [\[6, 61\]](#). Quantitative rates of convergence are established using a flow Poincaré inequality relating the time-averaged  $L^2$ -norm to the time-averaged energy or Dirichlet form. Previous approaches to proving convergence for non-reversible dynamics relied on tricks such as considering Sobolev norms [\[248\]](#), adding a mixing term to create an equivalent norm in which dissipation takes place [\[87\]](#), or specifically tailored couplings and twisted metrics [\[98\]](#). The strength of the approach using second-order lifts lies in the more intrinsic nature of the time-averaged  $L^2$ -norms as opposed to such artificially introduced terms [\[54\]](#). We summarise the resulting bounds for the convergence of self-repelling diffusions in the following, and defer their proofs and an overview of the framework to [Section F.4](#). Since the approach [\[87\]](#) developed by Dolbeault, Mouhot and Schmeiser was already applied to self-repelling diffusions by Benaïm and Gauthier in [\[24\]](#), we compare the result of [Theorem F.7](#) to theirs in [Section F.3.4](#).

We let

$$\chi_{ijkl} = \int_M e_i e_j e_k e_l \, d\nu \quad \text{and} \quad \tilde{\chi}_{ijkl} = \int_M \langle \nabla_M e_i, \nabla_M e_j \rangle \langle \nabla_M e_k, \nabla_M e_l \rangle |\lambda_i|^{-1} |\lambda_k|^{-1} \, d\nu,$$

as well as

$$\chi = \left( \sum_{i,j,k,l=1}^n \chi_{ijkl}^2 \right)^{1/2} \quad \text{and} \quad \tilde{\chi} = \left( \sum_{i,j,k,l=1}^n \tilde{\chi}_{ijkl}^2 \right)^{1/2}.$$

**Theorem F.7** (Upper bound on the relaxation time). *The transition semigroup  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  satisfies*

$$\|\hat{P}_t^{(\sigma)}\|_{L^2(\hat{\mu})} \leq e^{-\nu(t-T)} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } t \geq 0 \text{ and } f \in L_0^2(\hat{\mu}) \quad (\text{F.13})$$

with  $T \in O(m^{-1/2})$  and inverse convergence rate

$$\nu^{-1} \in O\left(\sigma^2 C_2^2 + \sigma^{-2} \left(C_1^2 + \frac{1}{\eta}\right)\right),$$

where

$$\begin{aligned} C_1^2 &= 8\nu(M) \left( \chi + 4\tilde{\chi} + \frac{2\tilde{\chi} \sum_{k=1}^n a_k |\lambda_k|}{\min\{a_k |\lambda_k| : k \in \{1, \dots, n\}\}} \right), \\ C_2^2 &= \frac{4\nu(M)}{\min\{a_k : k \in \{1, \dots, n\}\}}, \\ \eta &= \inf \operatorname{spec}(-\Delta_M) \setminus \{0\}, \\ m &= \frac{\min\{a_k |\lambda_k| : k \in \{1, \dots, n\}\}}{2\nu(M)}. \end{aligned}$$

In particular, there exists a universal constant  $C > 0$  such that

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C \left( \sigma^2 C_2^2 + \sigma^{-2} \left( C_1^2 + \frac{1}{\eta} \right) \right)$$

for any choice of diffusivity  $\sigma > 0$ .

Note that compactness of  $M$  ensures  $\eta > 0$ . Furthermore, if the Ricci curvature of  $M$  is non-negative, the spectral gap satisfies  $\eta \geq \frac{\pi^2}{\operatorname{diam}(M)^2}$ , see [180, Theorem 5.5].

#### F.3.4. The periodic case

As a concrete example, we return to the setting of [Example F.3](#) where  $M = \mathbb{T}_L = \mathbb{R}/(2\pi L\mathbb{Z})$  is the circle with circumference  $2\pi L$  and choose

$$W(z) = \frac{1}{\pi L} \sum_{k=1}^n a_k \cos(kz/L), \quad z \in \mathbb{T}_L. \quad (\text{F.14})$$

Recall that setting  $V(x, y) = W(x - y)$  yields

$$V(x, y) = \frac{1}{\pi L} \sum_{k=1}^n a_k \cos(kx/L) \cos(ky/L) + a_k \sin(kx/L) \sin(ky/L),$$

which is of the form [\(F.3\)](#) with the orthonormal functions  $e_k(x) = (\pi L)^{-1/2} \cos(kx/L)$  and  $f_k(x) = (\pi L)^{-1/2} \sin(kx/L)$ . The associated eigenvalues are  $\lambda_k = -\frac{k^2}{L^2}$  and the potential  $\Phi$  becomes

$$\Phi(u, v) = \frac{1}{2L^2} \sum_{k=1}^n a_k k^2 (u_k^2 + v_k^2), \quad u, v \in \mathbb{R}^n,$$

where  $u_k$  and  $v_k$  correspond to the eigenfunctions  $e_k$  and  $f_k$ . We then have

$$\nu(M) = 2\pi L \quad \text{and} \quad \eta = \frac{1}{L^2}.$$

The lower bound in [Theorem F.6](#) yields

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \geq \left( \frac{\pi L^3}{2 \min\{a_k k^2 : k \in \{1, \dots, n\}\}} \right)^{1/2}. \quad (\text{F.15})$$

The dependence of the relaxation time on the size parameter  $L$  in the limit  $L \rightarrow \infty$  is of particular interest, as it gives information on the rate at which the process explores its ambient space.

**Remark F.8.** As explained in [Remark F.1](#), in this setting  $V_t(x) = (W * L_t)(x)$ , where  $L_t$  is the local time density of  $(X_t)$  at time  $t$ . The finite-dimensional system [\(F.4\)](#) introduced above is hence obtained as the gradient of an approximation of the local time in Fourier domain. Instead considering spatial discretisations of the local time leads to self-repelling random walks on the discrete circle. These can be treated using the framework of second-order lifts in a similar fashion, see [\[100\]](#).

**Example F.9.** As a first simple case, consider

$$V(x, y) = a \cos(k(x - y)/L) = \frac{1}{\pi L} (a \cos(kx/L) \cos(ky/L) + a \sin(kx/L) \sin(ky/L))$$

with  $k \in \mathbb{N}$  and  $a > 0$ . [Theorem F.6](#) then yields the lower bound

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \geq \left( \frac{\pi L^3}{2ak^2} \right)^{1/2}. \quad (\text{F.16})$$

on the relaxation time. Regarding the upper bound on the relaxation time, a direct computation shows that

$$\frac{1}{(\pi L)^2} \int_0^{2\pi L} \cos(kx/L)^j \sin(kx/L)^{4-j} dx = \begin{cases} \frac{1}{4\pi L} & \text{if } j = 2, \\ \frac{3}{4\pi L} & \text{if } j \in \{0, 4\}, \\ 0 & \text{if } j \in \{1, 3\}, \end{cases}$$

so that  $\tilde{\chi} = \chi = \sqrt{2 \cdot (\frac{3}{4\pi L})^2 + 8 \cdot (\frac{1}{4\pi L})^2} = \frac{\sqrt{26}}{4\pi L}$ . This yields

$$C_1^2 = 16\pi L \cdot \frac{7\sqrt{26}}{4\pi L} = 28\sqrt{26} < 143 \quad \text{and} \quad C_2^2 = \frac{8\pi L}{a},$$

so that [Theorem F.7](#) yields

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C \left( \sigma^2 \frac{L}{a} + \sigma^{-2}(1 + L^2) \right). \quad (\text{F.17})$$

Choosing a diffusivity minimising this upper bound yields

$$t_{\text{rel}}(\hat{P}^{(\sigma_*)}) \leq C \sqrt{(L + L^3)/a} \quad \text{with} \quad \sigma_*^2 \propto \sqrt{a/L + aL}.$$

Comparing this to the lower bound [\(F.16\)](#), we see that, for this choice of  $\sigma$ , the upper bound is sharp up to a factor of a constant times  $k$  in case  $L \geq 1$ .

In comparison, [24, Theorem 6] yields the upper bound<sup>1</sup>

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C L^2 \frac{(1+a\lambda)^2}{a^2 \lambda^2} \left( \sigma^2 \lambda^2 + \lambda \left( 1 + \frac{\sqrt{1+a\lambda}}{L} \right) + \sigma^{-2} \left( 1 + \frac{\sqrt{1+a\lambda}}{L} \right)^2 \right),$$

where  $\lambda = \frac{k^2}{L^2}$ . This upper bound is much weaker than the one in (F.17) obtained here, both in the limiting regimes  $L \rightarrow 0$  and  $L \rightarrow \infty$ , as well as when choosing a minimising diffusivity. In particular, it cannot achieve the lower bound (F.16), since when choosing a minimising diffusivity, we have

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C L^2 \frac{(1+a\lambda)^2}{a^2 \lambda^2} \lambda \left( 1 + \frac{\sqrt{1+a\lambda}}{L} \right) \quad \text{with} \quad \sigma_*^2 \propto \lambda^{-1} \left( 1 + \frac{\sqrt{1+a\lambda}}{L} \right).$$

Fixing  $a$  and  $k$ , this yields a relaxation time of the order  $O(L^4)$  as  $L \rightarrow \infty$  and the order  $O(L^{-2})$  as  $L \rightarrow 0$ .

**Example F.10.** Let us further investigate how the bounds on the relaxation time depend on the parameter  $n$  in (F.14). Note that we have  $\chi_{ijkl} \neq 0$  if and only if  $(i, j, k, l)$  admits a partition into two tuples (not necessarily of the same size) of equal sum, and the same holds for  $\tilde{\chi}_{ijkl}$ . Therefore, there exists a universal constant  $C$  such that

$$\chi \leq C n^{3/2} / L \quad \text{and} \quad \tilde{\chi} \leq C n^3 / L.$$

This yields

$$C_1^2 \leq C \frac{n^3 \sum_{k=1}^n a_k k^2}{\min\{a_k k^2 : k \in \{1, \dots, n\}\}} \quad \text{and} \quad C_2^2 \leq C \frac{L}{\min\{a_k : k \in \{1, \dots, n\}\}},$$

and we observe that the dependence on  $L$  does not change compared to the simpler case of Example F.9. In particular, if we choose  $a_k = 1$  for all  $k \in \{1, \dots, n\}$ , we obtain an upper bound

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C (\sigma^2 L + \sigma^{-2} (n^6 + L^2))$$

on the relaxation time. Similarly, with  $a_k = \frac{1}{k^2}$  for all  $k \in \{1, \dots, n\}$ , we have

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C (\sigma^2 L n^2 + \sigma^{-2} (n^4 + L^2)).$$

Again, in comparison [24, Theorem 6] yields the upper bound

$$t_{\text{rel}}(\hat{P}^{(\sigma)}) \leq C L^2 \frac{(1+\Lambda_0)^2}{\Lambda_0^2} \left( \sigma^2 \Lambda_1^2 + \Lambda_1 \left( n + \frac{n^3 \sqrt{1+\Lambda_1}}{L} \right) + \sigma^{-2} \left( n + \frac{n^3 \sqrt{1+\Lambda_1}}{L} \right)^2 \right),$$

where  $\Lambda_0 = \min\{a_k \frac{k^2}{L^2} : k \in \{1, \dots, n\}\}$  and  $\Lambda_1 = \sum_{k=1}^n a_k \frac{k^2}{L^2}$ . In the two cases mentioned above, we have

$$\begin{aligned} \Lambda_0 &= \frac{1}{L^2}, & \Lambda_1 &\leq C \frac{n^3}{L^2} & \text{if } a_k &= 1 \text{ for all } k \in \{1, \dots, n\}, \\ \Lambda_0 &= \frac{1}{L^2}, & \Lambda_1 &= \frac{n}{L^2} & \text{if } a_k &= \frac{1}{k^2} \text{ for all } k \in \{1, \dots, n\}, \end{aligned}$$

yielding worse dependence on  $n$  in both cases.

<sup>1</sup>The statement of [24, Theorem 6] contains an error in the claimed rate  $\lambda$ . A factor of  $\eta$  (the spectral gap, or  $\eta_1$  in the notation of that paper) is missing in  $K_1$ . Note that it is still present in the proof, specifically Proposition 5.

## F.4. Convergence to equilibrium

We consider the family  $(\hat{P}_t^{(\sigma)})_{t \geq 0}$  of semigroups on  $L^2(\hat{\mu})$  associated to the self-repelling diffusion (F.4) with diffusivity  $\sigma \geq 0$ . Recall that their generators admit the decomposition

$$\hat{\mathcal{L}}^{(\sigma)} f = \hat{\mathcal{L}} f + \frac{\sigma^2}{2} \Delta_M f \quad \text{for all } f \in \text{Dom}(\hat{\mathcal{L}}^{(\sigma)})$$

given in (F.5). Convergence to equilibrium of these semigroups can be shown using a flow Poincaré inequality, which can be established for second-order lifts using the framework of [97, 99]. In our setting, the general framework reduces to [Theorem F.11](#). Recall that  $(\mathcal{E}, \text{Dom}(\mathcal{E}))$  denotes the Dirichlet form (F.6) associated to  $\mathcal{L}$ , and let

$$\Pi f(u, x) = \frac{1}{\nu(M)} \int_M f(u, y) \nu(dy) = \int_M f(u, y) \kappa(dy).$$

**Theorem F.11.** *Assume that*

- (i) *the operator  $(\hat{\mathcal{L}}, \text{Dom}(\hat{\mathcal{L}}))$  is a second-order lift of  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  such that  $g \circ \pi \in \text{Dom}(\hat{\mathcal{L}}^*)$  and  $\hat{\mathcal{L}}^*(g \circ \pi) = -\hat{\mathcal{L}}(g \circ \pi)$  for all  $g \in \text{Dom}(\mathcal{L})$ ;*
- (ii) *the operator  $(\mathcal{L}, \text{Dom}(\mathcal{L}))$  has purely discrete spectrum on  $L^2(\mu)$  and a spectral gap  $m > 0$ , i.e.*

$$\|g - \mu(g)\|_{L^2(\mu)}^2 \leq \frac{1}{m} \mathcal{E}(g)$$

*for all  $f \in \text{Dom}(\mathcal{E})$ ;*

- (iii) *there exist constants  $C_1, C_2 > 0$  such that*

$$\langle \hat{\mathcal{L}}(g \circ \pi), \hat{\mathcal{L}}(f - \Pi f) \rangle_{L^2(\hat{\mu})} \leq C_1 \|\mathcal{L}g\|_{L^2(\mu)} \|\nabla_M f\|_{L^2(\hat{\mu})}, \quad (\text{F.18})$$

$$\langle \hat{\mathcal{L}}(g \circ \pi), \Delta_M f \rangle_{L^2(\hat{\mu})} \leq C_2 \|\mathcal{L}g\|_{L^2(\mu)} \|\nabla_M f\|_{L^2(\hat{\mu})} \quad (\text{F.19})$$

*for all  $f \in \text{Dom}(\hat{\mathcal{L}})$  with  $\Pi f \in \text{Dom}(\hat{\mathcal{L}})$  and  $g$  in a core of  $\mathcal{L}$ ;*

- (iv) *the Laplace-Beltrami operator  $\Delta_M$  on  $M$  has a spectral gap*

$$\eta = \inf \text{spec}(-\Delta_M) \setminus \{0\} > 0.$$

*Then there exists a universal constant  $C > 0$  such that, for any  $T > 0$  and  $\sigma > 0$ , the semigroup  $(\hat{P}_t^{(\sigma)})$  generated by  $(\hat{\mathcal{L}}^{(\sigma)}, \text{Dom}(\hat{\mathcal{L}}^{(\sigma)}))$  satisfies*

$$\|\hat{P}_t^{(\sigma)} f\|_{L^2(\hat{\mu})} \leq e^{-\nu(t-T)} \|f\|_{L^2(\hat{\mu})} \quad \text{for all } t \geq 0 \text{ and } f \in L_0^2(\hat{\mu}),$$

*where the rate  $\nu$  is given by*

$$\nu = C \frac{\sigma^2}{\sigma^4 C_2^2 + C_1^2 + \frac{1}{\eta} (1 + \frac{1}{mT^2})}. \quad (\text{F.20})$$

*Proof.* The statement is a direct consequence of [97, Theorem 12].  $\square$

We will now apply this result to prove the upper bound in [Theorem F.7](#) on the relaxation time of self-repelling diffusions.

**Lemma F.12.** *For any  $g \in C_c^\infty(\mathbb{R}^n)$ ,*

$$\int_{\mathbb{R}^n} |\nabla \Phi|^2 |\nabla g|^2 d\mu \leq 2 \left( \sum_{k=1}^n a_k |\lambda_k| \right) \int_{\mathbb{R}^n} |\nabla g|^2 d\mu + 4 \int_{\mathbb{R}^n} \|\nabla^2 g\|_F^2 d\mu,$$

where  $\|\nabla^2 g\|_F$  denotes the Frobenius norm of  $\nabla^2 g$ .

*Proof.* Using integration by parts, we estimate

$$\int_{\mathbb{R}^n} |\nabla \Phi|^2 |\nabla g|^2 d\mu = \int_{\mathbb{R}^n} \Delta \Phi |\nabla g|^2 d\mu + 2 \int_{\mathbb{R}^d} \nabla \Phi^\top \nabla^2 g \nabla g d\mu.$$

Now,

$$2 \nabla \Phi^\top \nabla^2 g \nabla g \leq 2 |\nabla \Phi| |\nabla g| \|\nabla^2 g\|_F \leq p |\nabla \Phi|^2 |\nabla g|^2 + \frac{1}{p} \|\nabla^2 g\|_F^2$$

for any  $p > 0$  by Young's inequality. Choosing  $p = \frac{1}{2}$  and rearranging finishes the proof.  $\square$

The inequality in [Lemma F.12](#) can in fact immediately be sharpened by replacing the Frobenius norm with the spectral norm of  $\nabla^2 g$ . The Frobenius norm is, however, more conveniently bounded by Bochner's identity which we will use below.

**Lemma F.13.** *The operator  $(\hat{\mathcal{L}}, C_c^\infty(\mathbb{R}^n \times M))$  is antisymmetric, i.e.*

$$\int_{\mathbb{R}^n \times M} f \cdot \hat{\mathcal{L}} g d\hat{\mu} = - \int_{\mathbb{R}^n \times M} \hat{\mathcal{L}} f \cdot g d\hat{\mu}$$

for any  $f, g \in C_c^\infty(\mathbb{R}^n \times M)$ .

*Proof.* Note that the adjoint of  $\partial_{u_k}$  with respect to  $\mu$  is given by  $\partial_{u_k}^* f(u) = -\partial_{u_k} f(u) + a_k |\lambda_k| u_k f(u)$ . An integration by parts thus yields

$$\begin{aligned} \int_{\mathbb{R}^n \times M} f \cdot \hat{\mathcal{L}} g d\hat{\mu} &= \int_{\mathbb{R}^n \times M} f(u, x) \sum_{k=1}^n (e_k(x) \partial_{u_k} g(u, x) - a_k u_k \langle \nabla_M e_k(x), \nabla_M g(u, x) \rangle) d\hat{\mu} \\ &= \sum_{k=1}^n \int_{\mathbb{R}^n \times M} (-\partial_{u_k} f(u, k) e_k(x) + a_k |\lambda_k| u_k f(u, x) g(u, x)) d\hat{\mu} \\ &\quad + \sum_{k=1}^n \int_{\mathbb{R}^n \times M} a_k u_k g(u, x) \nabla_M \cdot (\nabla_M e_k(x) f(u, x)) d\hat{\mu} \\ &= \sum_{k=1}^n \int_{\mathbb{R}^n \times M} (-\partial_{u_k} f(u, k) e_k(x) + a_k |\lambda_k| u_k f(u, x) g(u, x)) d\hat{\mu} \end{aligned}$$

$$\begin{aligned}
& + \sum_{k=1}^n \int_{\mathbb{R}^n \times M} a_k u_k g(u, x) (\Delta_M e_k(x) f(u, x) + \langle \nabla_M e_k(x), \nabla_M f(u, x) \rangle) d\hat{\mu} \\
& = - \int_{\mathbb{R}^n \times M} \hat{\mathcal{L}} f \cdot g d\hat{\mu}.
\end{aligned}$$

□

*Proof of Theorem F.7.* The claim follows from Theorem F.11 if we can show that conditions (i)–(iv) are satisfied with  $m$ ,  $C_1$ , and  $C_2$  as claimed. To this end, condition (i) is satisfied by Theorem F.5 and Lemma F.13, and condition (ii) is satisfied with

$$m = \frac{\min\{a_k |\lambda_k| : k \in \{1, \dots, n\}\}}{2\nu(M)}$$

by (F.12). We turn to condition (iii) and the constant  $C_1$ . For  $g \in C_c^\infty(\mathbb{R}^n)$  we have  $\hat{\mathcal{L}}(g \circ \pi) \in \text{Dom}(\hat{\mathcal{L}}^*)$  and

$$\hat{\mathcal{L}}^* \hat{\mathcal{L}}(g \circ \pi)(u, x) = -e(x)^\top \nabla^2 g(u) e(x) + \nabla \Phi(u)^\top \Lambda^{-1} \langle \nabla_M e(x), \nabla_M e(x)^\top \rangle \nabla g(u),$$

where the term  $\langle \nabla_M e(x), \nabla_M e(x)^\top \rangle$  is to be understood as

$$\langle \nabla_M e(x), \nabla_M e(x)^\top \rangle = (\langle \nabla_M e_k(x), \nabla_M e_l(x) \rangle)_{k,l=1}^n \in \mathbb{R}^{n \times n}.$$

Then

$$\begin{aligned}
& \int_M (\hat{\mathcal{L}}^* \hat{\mathcal{L}}(g \circ \pi)(u, x))^2 \kappa(dx) \\
& \leq 2 \int_M \left( e(x)^\top \nabla^2 g(u) e(x) \right)^2 + \left( \nabla \Phi(u)^\top \Lambda^{-1} \langle \nabla_M e(x), \nabla_M e(x)^\top \rangle \nabla g(u) \right)^2 \kappa(dx) \\
& = \frac{2}{\nu(M)} \sum_{i,j,k,l=1}^n \partial_{ij} g(u) \partial_{kl} g(u) \chi_{ijkl} + \frac{2}{\nu(M)} \sum_{i,j,k,l=1}^n \partial_i \Phi(u) \partial_j g(u) \partial_k \Phi(u) \partial_l g(u) \tilde{\chi}_{ijkl} \\
& \leq \frac{2}{\nu(M)} (\|\nabla^2 g(u)\|_F^2 \chi + |\nabla \Phi(u)|^2 |\nabla g(u)|^2 \tilde{\chi})
\end{aligned}$$

By Lemma F.12, we thus obtain

$$\begin{aligned}
& \int_{\mathbb{R}^n \times M} (\hat{\mathcal{L}}^* \hat{\mathcal{L}}(g \circ \pi))^2 d\hat{\mu} \leq \frac{2\chi}{\nu(M)} \int_{\mathbb{R}^n} \|\nabla^2 g\|_F^2 d\mu + \frac{2\tilde{\chi}}{\nu(M)} \int_{\mathbb{R}^n} |\nabla \Phi|^2 |\nabla g|^2 d\mu \\
& \leq \frac{2\chi + 8\tilde{\chi}}{\nu(M)} \int_{\mathbb{R}^n} \|\nabla^2 g\|_F^2 d\mu + \frac{4\tilde{\chi} \sum_{k=1}^n a_k |\lambda_k|}{\nu(M)} \int_{\mathbb{R}^n} |\nabla g|^2 d\mu.
\end{aligned}$$

Bochner's identity yields

$$\int_{\mathbb{R}^n} (\nabla^* \nabla g)^2 d\mu = \int_{\mathbb{R}^n} \|\nabla^2 g\|_F^2 d\mu + \int_{\mathbb{R}^n} \sum_{k=1}^n a_k |\lambda_k| (\partial_{u_k} g)^2 d\mu, \quad (\text{F.21})$$

so that, since  $\mathcal{L} = -\frac{1}{2\nu(M)}\nabla^*\nabla$  by (F.11),

$$\int_{\mathbb{R}^n \times M} (\hat{\mathcal{L}}^* \hat{\mathcal{L}}(g \circ \pi))^2 d\hat{\mu} \leq C_1^2 \int_{\mathbb{R}^n} (\mathcal{L}g)^2 d\mu$$

with

$$C_1^2 = 8\nu(M) \left( \chi + 4\tilde{\chi} + \frac{2\tilde{\chi} \sum_{k=1}^n a_k |\lambda_k|}{\min\{a_k |\lambda_k| : k \in \{1, \dots, n\}\}} \right).$$

Therefore, (F.18) is satisfied with this choice of  $C_1$  by the Cauchy-Schwarz inequality.

Turning to the constant  $C_2$ , integration by parts, the Cauchy-Schwarz inequality and orthonormality of the  $e_k$  yields

$$\begin{aligned} \langle \hat{\mathcal{L}}(g \circ \pi), \Delta_M f \rangle_{L^2(\hat{\mu})} &= \int_{\mathbb{R}^n \times M} \sum_{k=1}^n e_k(x) \partial_{u_k} g(u) \Delta_M f(u, x) \hat{\mu}(du dx) \\ &= \int_{\mathbb{R}^n \times M} \left\langle \sum_{k=1}^n \partial_{u_k} g(u) \nabla_M e_k(x), \nabla_M f(u, x) \right\rangle \hat{\mu}(du dx) \\ &\leq \left( \int_{\mathbb{R}^n \times M} \left| \sum_{k=1}^n \partial_{u_k} g(u) \nabla_M e_k(x) \right|^2 \hat{\mu}(du dx) \right)^{1/2} \left( \int_{\mathbb{R}^n \times M} |\nabla_M f|^2 d\hat{\mu} \right)^{1/2} \\ &= \frac{1}{\sqrt{\nu(M)}} \left( \int_{\mathbb{R}^n} \sum_{k=1}^n |\lambda_k| (\partial_{u_k} g)^2 d\mu \right)^{1/2} \|\nabla_M f\|_{L^2(\hat{\mu})}. \end{aligned}$$

By Bochner's identity (F.21) we thus obtain

$$\langle \hat{\mathcal{L}}(g \circ \pi), \Delta_M f \rangle_{L^2(\hat{\mu})} \leq C_2 \|\mathcal{L}g\|_{L^2(\mu)} \|\nabla_M f\|_{L^2(\hat{\mu})}$$

with

$$C_2^2 = \frac{4\nu(M)}{\min\{a_k : k \in \{1, \dots, n\}\}}.$$

□

## Acknowledgements

The author was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2047 – 390685813 as well as under CRC 1720 – 539309657. I would like to thank Andreas Eberle, Leo Hahn, and Michel Benaïm for helpful discussions and comments, and Michel Benaïm for funding a stay in Neuchâtel in October 2025.

# Bibliography

- [1] R. Abraham and J. E. Marsden. *Foundations of Mechanics*. 2nd ed. Redwood City, CA: Addison-Wesley Publishing Company, Inc., Oct. 1987. ISBN: 0-8053-0102-X.
- [2] O. Abril-Pla et al. ‘PyMC: a modern, and comprehensive probabilistic programming framework in Python’. In: *PeerJ Comput. Sci.* 9.e1516 (2023). DOI: [10 . 7717/peerj-cs.1516](https://doi.org/10.7717/peerj-cs.1516).
- [3] F. Achleitner, A. Arnold and B. Signorello. ‘On optimal decay estimates for ODEs and PDEs with modal decomposition’. In: *International workshop on Stochastic Dynamics out of Equilibrium*. Springer. 2017, pp. 241–264.
- [4] H. M. Afshar and J. Domke. ‘Reflection, Refraction, and Hamiltonian Monte Carlo’. In: *Adv. Neural Inf. Process. Syst.* 28 (2015).
- [5] K. Ahn and S. Chewi. ‘Efficient constrained sampling via the mirror-Langevin algorithm’. In: *Adv. Neural Inf. Process. Syst.* 34 (2021), pp. 28405–28418.
- [6] D. Albritton, S. Armstrong, J.-C. Mourrat and M. Novack. ‘Variational methods for the kinetic Fokker-Planck equation’. In: *Anal. & PDE* (July 2024). DOI: [10 . 2140/apde.2024.17.1953](https://doi.org/10.2140/apde.2024.17.1953).
- [7] S. Alexander. ‘Locally Convex Hypersurfaces of Negatively Curved Spaces’. In: *Proc. Amer. Math. Soc.* 64.2 (1977), pp. 321–325. DOI: [10 . 2307/2041451](https://doi.org/10.2307/2041451).
- [8] J. M. Altschuler, S. Chewi and M. S. Zhang. *Shifted Composition IV: Toward Ballistic Acceleration for Log-Concave Sampling*. 2025. arXiv: [2506 . 23062](https://arxiv.org/abs/2506.23062) [[math.PR](#)].
- [9] D. J. Amit, G. Parisi and L. Peliti. ‘Asymptotic behavior of the “true” self-avoiding walk’. In: *Phys. Rev. B* 27.3 (Feb. 1983), pp. 1635–1645. DOI: [10 . 1103/PhysRevB. 27 . 1635](https://doi.org/10.1103/PhysRevB.27.1635).
- [10] C. Amrouche, P. G. Ciarlet and C. Mardare. ‘On a lemma of Jacques-Louis Lions and its relation to other fundamental results’. In: *J. Math. Pures Appl.* 104.2 (Aug. 2015), pp. 207–226. DOI: [10 . 1016/j.matpur.2014.11.007](https://doi.org/10.1016/j.matpur.2014.11.007).
- [11] F. Anceschi et al. *Poincaré inequality and quantitative De Giorgi method for hypoelliptic operators*. 2024. arXiv: [2401.12194](https://arxiv.org/abs/2401.12194) [[math.AP](#)].
- [12] C. Andrieu, A. Durmus, N. Nüsken and J. Roussel. ‘Hypocoercivity of piecewise deterministic Markov process-Monte Carlo’. In: *Ann. Appl. Probab.* 31.5 (2021), pp. 2478–2517. DOI: [10 . 1214/20-aap1653](https://doi.org/10.1214/20-aap1653).
- [13] L. Angelani and R. Garra. ‘Run-and-tumble motion in one dimension with space-dependent speed’. In: *Phys. Rev. E* 100.5 (2019), p. 052147. DOI: [10 . 1103 / PhysRevE.100.052147](https://doi.org/10.1103/PhysRevE.100.052147).

- [14] L. Angelani. ‘Confined run-and-tumble swimmers in one dimension’. In: *J. Phys. A: Math. Theor.* 50.32 (2017), p. 325601. DOI: [10.1088/1751-8121/aa734c](https://doi.org/10.1088/1751-8121/aa734c).
- [15] S. Apers, S. Gribling and D. Szilágyi. ‘Hamiltonian Monte Carlo for efficient Gaussian sampling: long and random steps’. In: *J. Mach. Learn. Res.* 25.348 (2024), pp. 1–30.
- [16] A. Arnold, J. Dolbeault, C. Schmeiser and T. Wöhrer. ‘Sharpening of decay rates in Fourier based hypocoercivity methods’. In: *Recent Advances in Kinetic Equations and Applications*. Springer, 2021, pp. 1–50. ISBN: 978-3-030-82945-2. DOI: [10.1007/978-3-030-82946-9](https://doi.org/10.1007/978-3-030-82946-9).
- [17] A. Arnold, C. Schmeiser and B. Signorello. ‘Propagator norm and sharp decay estimates for Fokker-Planck equations with linear drift’. In: *Commun. Math. Sci.* 20.4 (2022), pp. 1047–1080. DOI: [10.4310/CMS.2022.v20.n4.a5](https://doi.org/10.4310/CMS.2022.v20.n4.a5).
- [18] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. 2nd ed. Vol. 60. Graduate Texts in Mathematics. Originally published by Nauka, Moscow, 1974. New York, NY: Springer, 1989, p. 520. ISBN: 978-0-387-96890-2. DOI: [10.1007/978-1-4757-2063-1](https://doi.org/10.1007/978-1-4757-2063-1).
- [19] D. Bakry and M. Émery. ‘Diffusions hypercontractives’. In: *Séminaire de Probabilités XIX 1983/84*. Ed. by J. Azéma and M. Yor. Springer, 1985, pp. 177–206. DOI: [10.1007/BFb0075847](https://doi.org/10.1007/BFb0075847).
- [20] D. Bakry, I. Gentil and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Vol. 348. Grundlehren der mathematischen Wissenschaften. Cham: Springer, 2014. ISBN: 978-3-319-00227-9. DOI: [10.1007/978-3-319-00227-9](https://doi.org/10.1007/978-3-319-00227-9).
- [21] S. Banerjee, M. Gordina and P. Mariano. ‘Coupling in the Heisenberg group and its applications to gradient estimates’. In: *Ann. Probab.* 46.6 (2018), pp. 3275–3312. DOI: [10.1214/17-AOP1247](https://doi.org/10.1214/17-AOP1247).
- [22] F. Baudoin. ‘Bakry–Émery meet Villani’. In: *J. Funct. Anal.* 273.7 (2017), pp. 2275–2291.
- [23] M. Benaïm, I. Ciotir and C.-E. Gauthier. ‘Self-repelling diffusions via an infinite dimensional approach’. In: *Stoch. Partial Differ. Equ.: Anal. Comput.* 3 (2015), pp. 506–530. DOI: [10.1007/s40072-015-0059-5](https://doi.org/10.1007/s40072-015-0059-5).
- [24] M. Benaïm and C.-E. Gauthier. ‘Self-repelling diffusions on a Riemannian manifold’. In: *Probab. Theory Relat. Fields* 169 (2016), pp. 63–104. DOI: [10.1007/s00440-016-0717-1](https://doi.org/10.1007/s00440-016-0717-1).
- [25] M. Benaïm, M. Ledoux and O. Raimond. ‘Self-interacting diffusions’. In: *Probab. Theory Relat. Fields* 122.1 (2002), pp. 1–41. DOI: [10.1007/s004400100161](https://doi.org/10.1007/s004400100161).
- [26] M. Benaïm and O. Raimond. ‘Self-interacting diffusions II: convergence in law’. In: *Ann. Inst. H. Poincaré Probab. Stat.* 39.6 (2003), pp. 1043–1055. DOI: [10.1016/S0246-0203\(03\)00028-1/FLA](https://doi.org/10.1016/S0246-0203(03)00028-1/FLA).

- [27] M. Benaïm and O. Raimond. ‘Self-interacting diffusions. III. Symmetric interactions’. In: *Ann. Probab.* 33.5 (2005), pp. 1716–1759. DOI: [10.1214/009117905000000251](#).
- [28] H. C. Berg. *E. Coli in Motion*. Springer, New York, 2004. ISBN: 978-0-387-21638-6. DOI: [10.1007/b97370](#).
- [29] É. Bernard, M. Fathi, A. Levitt and G. Stoltz. ‘Hypocoercivity with Schur complements’. In: *Ann. Henri Lebesgue* 5 (2022), pp. 523–557. DOI: [10.5802/ahl.129](#).
- [30] E. P. Bernard, W. Krauth and D. B. Wilson. ‘Event-chain Monte Carlo algorithms for hard-sphere systems’. In: *Phys. Rev. E* 80.5 (Nov. 2009), p. 056704. DOI: [10.1103/PhysRevE.80.056704](#).
- [31] A. Beskos, N. Pillai, G. Roberts, J.-M. Sanz-Serna and A. Stuart. ‘Optimal tuning of the hybrid Monte Carlo algorithm’. In: *Bernoulli* 19.5A (Nov. 2013), pp. 1501–1534. DOI: [10.3150/12-BEJ414](#).
- [32] R. N. Bhattacharya. ‘On the functional central limit theorem and the law of the iterated logarithm for Markov processes’. In: *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 60.2 (1982), pp. 185–201.
- [33] D. Bianchi, B. Güneysu and A. G. Setti. *Neumann cut-offs and essential self-adjointness on complete Riemannian manifolds with boundary*. 2024. arXiv: [2406.11120 \[math.DG\]](#).
- [34] J. Bierkens, P. Fearnhead and G. Roberts. ‘The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data’. In: *Ann. Stat.* 47.3 (2019), pp. 1288–1320. DOI: [10.1214/18-AOS1715](#).
- [35] J. Bierkens and G. Roberts. ‘A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model’. In: *Ann. Appl. Probab.* 27.2 (2017), pp. 846–882. DOI: [10.1214/16-AAP1217](#).
- [36] J. Bierkens, G. O. Roberts and P.-A. Zitt. ‘Ergodicity of the zigzag process’. In: *Ann. Appl. Probab.* 29.4 (2019), pp. 2266–2301. DOI: [10.1214/18-AAP1453](#).
- [37] J. Bierkens and S. M. Verduyn Lunel. ‘Spectral analysis of the zigzag process’. In: *Ann. Inst. Henri Poincaré (B) Probab. Stat.* 58.2 (2022), pp. 827–860. DOI: [10.1214/21-AIHP1188](#).
- [38] R. L. Bishop. ‘Infinitesimal Convexity Implies Local Convexity’. In: *Indiana Univ. Math. J.* 24.2 (1974), pp. 169–172.
- [39] R. L. Bishop and R. J. Crittenden. *Geometry of Manifolds*. Vol. 344. AMS Chelsea Publishing Series. American Mathematical Society, 1964. ISBN: 978-1-4704-1133-6. DOI: [10.1090/chel/344](#).
- [40] F. Bolley, A. Guillin and F. Malrieu. ‘Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation’. In: *M2AN Math. Model. Numer. Anal.* 44.5 (2010), pp. 867–884. DOI: [10.1051/m2an/2010045](#).

- [41] F. Bolley, A. Guillin and X. Wang. ‘Non ultracontractive heat kernel bounds by Lyapunov conditions’. In: *Discrete Contin. Dyn. Syst.* 35.3 (2015), pp. 857–870. DOI: [10.3934/dcds.2015.35.857](https://doi.org/10.3934/dcds.2015.35.857).
- [42] L. Boltzmann. ‘Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen’. In: *Sitz.-Ber. Akad. Wiss.* 66 (1872), pp. 275–370.
- [43] M. Bossy and J.-F. Jabir. ‘On confined McKean Langevin processes satisfying the mean no-permeability boundary condition’. In: *Stoch. Process. Their Appl.* 121.12 (2011), pp. 2751–2775. DOI: <https://doi.org/10.1016/j.spa.2011.07.006>.
- [44] M. Bossy and J.-F. Jabir. ‘Lagrangian stochastic models with specular boundary condition’. In: *J. Funct. Anal.* 268.6 (2015), pp. 1309–1381. DOI: [10.1016/j.jfa.2014.11.016](https://doi.org/10.1016/j.jfa.2014.11.016).
- [45] N. Bou-Rabee and A. Eberle. *Lecture Notes “Markov Chain Monte Carlo Methods”*. Institute for Applied Mathematics, University of Bonn. <http://wt.iam.uni-bonn.de/eberle/home/>. 2025.
- [46] N. Bou-Rabee, A. Eberle and R. Zimmer. ‘Coupling and convergence for Hamiltonian Monte Carlo’. In: *Ann. Appl. Probab.* 30.3 (June 2020), pp. 1209–1250. DOI: [10.1214/19-AAP1528](https://doi.org/10.1214/19-AAP1528).
- [47] N. Bou-Rabee and J. M. Sanz-Serna. ‘Randomized Hamiltonian Monte Carlo’. In: *Ann. Appl. Probab.* 27.4 (Aug. 2017), pp. 2159–2194. DOI: [10.1214/16-AAP1255](https://doi.org/10.1214/16-AAP1255).
- [48] A. Bouchard-Côté, S. J. Vollmer and A. Doucet. ‘The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method’. In: *J. Am. Stat. Assoc.* 113.522 (2018), pp. 855–867. DOI: [10.1080/01621459.2017.1294075](https://doi.org/10.1080/01621459.2017.1294075).
- [49] E. Bouin, J. Dolbeault and L. Lafleche. ‘Fractional hypocoercivity’. In: *Commun. Math. Phys.* 390.3 (2022), pp. 1369–1411.
- [50] E. Bouin, J. Dolbeault, L. Lafleche and C. Schmeiser. ‘Hypocoercivity and sub-exponential local equilibria’. In: *Monatsh. Math.* 194 (2021), pp. 41–65.
- [51] E. Bouin, J. Dolbeault and L. Ziviani. ‘ $L^2$  Hypocoercivity methods for kinetic Fokker-Planck equations with factorised Gibbs states’. In: *INdAM Meeting: Kolmogorov Operators and their Applications Workshop*. Springer. 2022, pp. 23–56.
- [52] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011. ISBN: 978-0-387-70913-0. DOI: [10.1007/978-0-387-70914-7](https://doi.org/10.1007/978-0-387-70914-7).
- [53] G. Brigati. ‘Time averages for kinetic Fokker-Planck equations’. In: *Kinet. Relat. Models* 16.4 (Aug. 2023), pp. 524–539. DOI: [10.3934/krm.2022037](https://doi.org/10.3934/krm.2022037).
- [54] G. Brigati, F. Lörler and L. Wang. ‘Hypocoercivity meets lifts’. In: *Kinet. Relat. Models* 20.0 (2026), pp. 34–55. DOI: [10.3934/krm.2025020](https://doi.org/10.3934/krm.2025020).
- [55] G. Brigati and G. Stoltz. ‘How to Construct Explicit Decay Rates for Kinetic Fokker-Planck Equations?’ In: *SIAM J. Math. Anal.* 57.4 (July 2025), pp. 3587–3622. DOI: [10.1137/24M1700351](https://doi.org/10.1137/24M1700351).

- [56] G. Brigati, G. Stoltz, A. Q. Wang and L. Wang. *Explicit convergence rates of underdamped Langevin dynamics under weighted and weak Poincaré–Lions inequalities*. July 2024. arXiv: [2407.16033 \[math.PR\]](#).
- [57] S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds. *Handbook of Markov Chain Monte Carlo*. New York: Chapman and Hall/CRC, 2011. ISBN: 978-1-420-07941-8. DOI: [10.1201/b10905](#).
- [58] C. J. Burke and M. Rosenblatt. ‘A Markovian Function of a Markov Chain’. In: *Ann. Math. Stat.* 29.4 (1958), pp. 1112–1122. DOI: [10.1214/aoms/1177706444](#).
- [59] S. Byrne and M. Girolami. ‘Geodesic Monte Carlo on Embedded Manifolds’. In: *Scand. J. Stat.* 40.4 (2013), pp. 825–845. DOI: [10.1111/sjos.12036](#).
- [60] E. Cancès, F. Legoll and G. Stoltz. ‘Theoretical and numerical comparison of some sampling methods for molecular dynamics’. In: *ESAIM: Math. Model. Numer. Anal.* 41.2 (2007), pp. 351–389. DOI: [10.1051/m2an:2007014](#).
- [61] Y. Cao, J. Lu and L. Wang. ‘On explicit  $L^2$ -convergence rate estimate for underdamped Langevin dynamics’. In: *Arch. Ration. Mech. Anal.* 247.90 (Aug. 2023). DOI: [10.1007/s00205-023-01922-4](#).
- [62] B. Carpenter et al. ‘Stan: A Probabilistic Programming Language’. In: *J. Stat. Softw.* 76.1 (2017), pp. 1–32. DOI: [10.18637/jss.v076.i01](#).
- [63] M. E. Cates and J. Tailleur. ‘Motility-induced phase separation’. In: *Annu. Rev. Condens. Matter Phys.* 6.1 (2015), pp. 219–244. DOI: [10.1146/annurev-conmatphys-031214-014710](#).
- [64] C. Cercignani. *The Boltzmann Equation and Its Applications*. 1st ed. Vol. 67. Applied Mathematical Sciences. New York, NY: Springer, 1988, p. 455. ISBN: 978-0-387-96637-3. DOI: [10.1007/978-1-4612-1039-9](#).
- [65] S. Chatterjee. ‘Spectral gap of nonreversible Markov chains’. In: *Ann. Appl. Probab.* 35.4 (Aug. 2025). DOI: [10.1214/25-AAP2183](#).
- [66] F. Chen, L. Lovász and I. Pak. ‘Lifting Markov Chains to Speed up Mixing’. In: *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*. STOC ’99 (May 1999), pp. 275–281. DOI: [10.1145/301250.301315](#).
- [67] T.-L. Chen and C.-R. Hwang. ‘Accelerating reversible Markov chains’. In: *Stat. Probab. Lett.* 83.9 (2013), pp. 1956–1962. DOI: [10.1016/j.spl.2013.05.002](#).
- [68] P. R. Chernoff. ‘Essential self-adjointness of powers of generators of hyperbolic equations’. In: *J. Funct. Anal.* 12.4 (1973), pp. 401–414. DOI: [10.1016/0022-1236\(73\)90003-7](#).
- [69] A. Chevallier, S. Pion and F. Cazals. *Hamiltonian Monte Carlo with boundary reflections, and application to polytope volume calculations*. 2018.
- [70] Y. Chikuse. *Statistics on special manifolds*. Vol. 174. Lecture Notes in Statistics. New York, USA: Springer, 2012. ISBN: 978-0-387-21540-2. DOI: [10.1007/978-0-387-21540-2](#).

- [71] T. Colding and W. Minicozzi. *A course in minimal surfaces*. Vol. 121. Graduate Studies in Mathematics. American Mathematical Society, 2011. ISBN: 978-1-4704-7640-3.
- [72] M. Cranston and T. S. Mountford. ‘The strong law of large numbers for a Brownian polymer’. In: *Ann. Probab.* 24.3 (1996), pp. 1300–1323. DOI: [10.1214/aop/1065725183](#).
- [73] I. Csiszár. ‘Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten’. In: *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* 8.1-2 (1963), pp. 85–108.
- [74] A. S. Dalalyan and L. Riou-Durand. ‘On sampling from a log-concave density using kinetic Langevin diffusions’. In: *Bernoulli* 26.3 (Aug. 2020), pp. 1956–1988. DOI: [10.3150/19-BEJ1178](#).
- [75] M. H. Davis. ‘Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models’. In: *J. R. Stat. Soc.* 46.3 (1984), pp. 353–376. DOI: [10.1111/j.2517-6161.1984.tb01308.x](#).
- [76] M. H. Davis. *Markov models & optimization*. Vol. 49. CRC Press, 1993. DOI: [10.1201/9780203748039](#).
- [77] G. Deligiannidis, A. Bouchard-Côté and A. Doucet. ‘Exponential ergodicity of the bouncy particle sampler’. In: *Ann. Stat.* 47.3 (2019), pp. 1268–1287. DOI: [10.1214/18-AOS1714](#).
- [78] A. Dhar, A. Kundu, S. N. Majumdar, S. Sabhapandit and G. Schehr. ‘Run-and-tumble particle in one-dimensional confining potentials: Steady-state, relaxation, and first-passage properties’. In: *Phys. Rev. E* 99.3 (2019). DOI: [10.1103/PhysRevE.99.032132](#).
- [79] P. Diaconis, S. Holmes and R. M. Neal. ‘Analysis of a nonreversible Markov chain sampler’. In: *Ann. Appl. Probab.* 10.3 (2000), pp. 726–752. DOI: [10.1214/aoap/1019487508](#).
- [80] P. Diaconis, S. Holmes and M. Shahshahani. ‘Sampling from a manifold’. In: *Advances in modern statistical theory and applications: a Festschrift in honor of Morris L. Eaton*. Vol. 10. Institute of Mathematical Statistics Collections. Institute of Mathematical Statistics, 2013, pp. 102–126. ISBN: 978-0-940-60084-3. DOI: [10.1214/12-IMSCOLL1006](#).
- [81] P. Diaconis and L. Miclo. ‘On the spectral analysis of second-order Markov chains’. In: *Annales de la Faculté des sciences de Toulouse : Mathématiques* Ser. 6, 22.3 (2013), pp. 573–621. DOI: [10.5802/afst.1383](#).
- [82] H. Dietert.  *$L^2$ -Hypocoercivity for non-equilibrium kinetic equations*. 2023. arXiv: [2310.13456 \[math.AP\]](#).
- [83] H. Dietert, F. Hérau, H. Hutridurga and C. Mouhot. *Quantitative geometric control in linear kinetic theory*. 2022. arXiv: [2209.09340 \[math.AP\]](#).

- [84] N. Ding et al. ‘Bayesian sampling using stochastic gradient thermostats’. In: *Adv. Neural Inf. Process.* 27 (2014).
- [85] J. Dolbeault, A. Klar, C. Mouhot and C. Schmeiser. ‘Exponential rate of convergence to equilibrium for a model describing fiber lay-down processes’. In: *Appl. Math. Res. eXpress* 2013.2 (2013), pp. 165–175. DOI: [10.1093/amrx/abs015](https://doi.org/10.1093/amrx/abs015).
- [86] J. Dolbeault, C. Mouhot and C. Schmeiser. ‘Hypocoercivity for kinetic equations with linear relaxation terms’. In: *C. R. Acad. Sci. Paris, Ser. I* 347.9 (Apr. 2009), pp. 511–516. DOI: [10.1016/j.crma.2009.02.025](https://doi.org/10.1016/j.crma.2009.02.025).
- [87] J. Dolbeault, C. Mouhot and C. Schmeiser. ‘Hypocoercivity for linear kinetic equations conserving mass’. In: *Trans. Am. Math. Soc.* 367.6 (June 2015), pp. 3807–3828. DOI: [10.1090/S0002-9947-2015-06012-7](https://doi.org/10.1090/S0002-9947-2015-06012-7).
- [88] H. Dong, Y. Yan Guo and T. Yastrzhembskiy. ‘Kinetic Fokker-Planck and Landau equations with specular reflection boundary condition’. In: *Kinet. Relat. Models* 15.3 (June 2022), pp. 467–516. DOI: [10.3934/krm.2022003](https://doi.org/10.3934/krm.2022003).
- [89] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth. ‘Hybrid Monte Carlo’. In: *Phys. Lett. B* 195.2 (1987), pp. 216–222. DOI: [10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- [90] A. Durmus, A. Enfroy, É. Moulines and G. Stoltz. ‘Uniform minorization condition and convergence bounds for discretizations of kinetic Langevin dynamics’. In: *Ann. Inst. H. Poincaré Probab. Stat.* 61.1 (2025), pp. 629–664. DOI: [10.1214/23-AIHP1442](https://doi.org/10.1214/23-AIHP1442).
- [91] A. Durmus, A. Guillin and P. Monmarché. ‘Geometric ergodicity of the Bouncy Particle Sampler’. In: *Ann. Appl. Probab.* 30.5 (Oct. 2020), pp. 2069–2098. DOI: [10.1214/19-AAP1552](https://doi.org/10.1214/19-AAP1552).
- [92] A. Durmus, A. Guillin and P. Monmarché. ‘Piecewise deterministic Markov processes and their invariant measures’. In: *Ann. Inst. Henri Poincaré (B) Probab. Stat.* 57.3 (Aug. 2021), pp. 1442–1475. DOI: [10.1214/20-AIHP1125](https://doi.org/10.1214/20-AIHP1125).
- [93] R. T. Durrett and L. C. G. Rogers. ‘Asymptotic behavior of Brownian polymers’. In: *Probab. Theory Relat. Fields* 92.3 (1992), pp. 337–349. DOI: [/10.1007/BF01300560](https://doi.org/10.1007/BF01300560).
- [94] G. Duvaut and J.-L. Lions. *Inequalities in Mechanics and Physics*. 1st ed. Vol. 219. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 1976. ISBN: 978-3-642-66167-9. DOI: [10.1007/978-3-642-66165-5](https://doi.org/10.1007/978-3-642-66165-5).
- [95] A. Eberle. ‘Reflection couplings and contraction rates for diffusions’. In: *Probab. Theory Relat. Fields* 166.3-4 (2016), pp. 851–886. DOI: [10.1007/s00440-015-0673-1](https://doi.org/10.1007/s00440-015-0673-1).
- [96] A. Eberle. *Lecture Notes “Markov processes”*. Institute for Applied Mathematics, University of Bonn. <http://wt.iam.uni-bonn.de/eberle/home/>. 2025.

- [97] A. Eberle, A. Guillin, L. Hahn, F. Lörler and M. Michel. *Convergence of non-reversible Markov processes via lifting and flow Poincaré inequality*. Mar. 2025. arXiv: [2503.04238 \[math.AP\]](#).
- [98] A. Eberle, A. Guillin and R. Zimmer. ‘Couplings and quantitative contraction rates for Langevin dynamics’. In: *Ann. Probab.* 47.4 (2019), pp. 1982–2010. DOI: [10.1214/18-AOP1299](#).
- [99] A. Eberle and F. Lörler. ‘Non-reversible lifts of reversible diffusion processes and relaxation times’. In: *Probab. Theory Relat. Fields* 194 (Aug. 2024). DOI: [10.1007/s00440-024-01308-x](#).
- [100] A. Eberle and F. Lörler. *Convergence rates of self-repellent random walks, their local time and Event Chain Monte Carlo*. Nov. 2025. arXiv: [2511.23453 \[math.PR\]](#).
- [101] A. Eberle and F. Lörler. ‘Space-time divergence lemmas and optimal non-reversible lifts of diffusions on Riemannian manifolds with boundary’. In: *J. Funct. Anal.* 291.9 (2026). DOI: [10.1016/j.jfa.2026.111605](#).
- [102] A. Einstein. ‘Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen’. In: *Ann. Phys.* 322.8 (1905), pp. 549–560. DOI: [10.1002/andp.19053220806](#).
- [103] J. Elgeti, R. G. Winkler and G. Gompper. ‘Physics of microswimmers—single particle motion and collective behavior: a review’. In: *Rep. Prog. Phys.* 78.5 (2015). DOI: [10.1088/0034-4885/78/5/056601](#).
- [104] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Vol. 194. Graduate Texts in Mathematics. New York: Springer, Oct. 1999. ISBN: 978-0-387-98463-6. DOI: [10.1007/b97696](#).
- [105] F. Essler, J. Gipouloux and W. Krauth. *Lifted TASEP: long-time dynamics, generalizations, and continuum limit*. 2025. arXiv: [2502.16549 \[cond-mat.stat-mech\]](#).
- [106] F. H. L. Essler and W. Krauth. ‘Lifted TASEP: A Solvable Paradigm for Speeding up Many-Particle Markov Chains’. In: *Phys. Rev. X* 14 (Nov. 2024). DOI: [10.1103/PhysRevX.14.041035](#).
- [107] Y. Fang, J.-M. Sanz-Serna and R. D. Skeel. ‘Compressible generalized hybrid Monte Carlo’. In: *J. Chem. Phys.* 140.17 (May 2014). DOI: [10.1063/1.4874000](#).
- [108] P. Fearnhead, J. Bierkens, M. Pollock and G. O. Roberts. ‘Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo’. In: *Stat. Sci.* 33.3 (2018), pp. 386–412. DOI: [10.1214/18-STS648](#).
- [109] M. Freidlin. ‘Some Remarks on the Smoluchowski–Kramers Approximation’. In: *J. Stat. Phys.* 117 (2004), pp. 617–634. DOI: [10.1007/s10955-004-2273-9](#).

- [110] G. P. Galdi. *An introduction to the mathematical theory of the Navier-Stokes equations. Steady-state problems*. 2nd ed. Springer Monographs in Mathematics. New York: Springer, 2011. ISBN: 978-0-387-09619-3. DOI: [10.1007/978-0-387-09620-9](#).
- [111] S. Gallot, D. Hulin and J. Lafontaine. *Riemannian Geometry*. 3rd ed. Universitext. Springer, 2004. ISBN: 978-3-642-18855-8. DOI: [10.1007/978-3-642-18855-8](#).
- [112] K. Gatmiry, J. Kelner and S. S. Vempala. ‘Sampling Polytopes with Riemannian HMC: Faster Mixing via the Lewis Weights Barrier’. In: *Proc. Mach. Learn. Res.* 247 (June 2024), pp. 1796–1881.
- [113] K. Gatmiry and S. S. Vempala. *Convergence of the Riemannian Langevin Algorithm*. Apr. 2022. arXiv: [2204.10818v1 \[cs.LG\]](#).
- [114] S. Geman and D. Geman. ‘Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images’. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 6.6 (1984), pp. 721–741. DOI: [10.1109/TPAMI.1984.4767596](#).
- [115] H. Giles and L. Gräfner. *Construction of the 1d Self-repelling Brownian Polymer*. Sept. 2025. arXiv: [2509.05286 \[math.PR\]](#).
- [116] M. Girolami and B. Calderhead. ‘Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods’. In: *J. R. Stat. Soc.* 73.2 (Mar. 2011), pp. 123–214. DOI: [10.1111/j.1467-9868.2010.00765.x](#).
- [117] R. J. Glauber. ‘Time-Dependent Statistics of the Ising Model’. In: *J. Math. Phys.* 4.2 (Feb. 1963), pp. 294–307. DOI: [10.1063/1.1703954](#).
- [118] F. Golse, P.-L. Lions, B. Perthame and R. Sentis. ‘Regularity of the moments of the solution of a transport equation’. In: *J. Funct. Anal.* 76.1 (Jan. 1988), pp. 110–125. DOI: [10.1016/0022-1236\(88\)90051-1](#).
- [119] M. Grothaus and M. Mertin. ‘Hypocoercivity of Langevin-type dynamics on abstract smooth manifolds’. In: *Stoch. Process. Their Appl.* 146 (Apr. 2022), pp. 22–59. DOI: [10.1016/j.spa.2021.12.007](#).
- [120] M. Grothaus and P. Stilgenbauer. ‘Hypocoercivity for Kolmogorov backward evolution equations and applications’. In: *J. Funct. Anal.* 267.10 (2014), pp. 3515–3556. DOI: [doi.org/10.1016/j.jfa.2014.08.019](#).
- [121] A. Guillin, L. Hahn and M. Michel. *Long-time analysis of a pair of on-lattice and continuous run-and-tumble particles with jamming interactions*. Nov. 2024. arXiv: [2411.13964 \[math.PR\]](#).
- [122] A. Guillin and P. Monmarché. ‘Optimal linear drift for the speed of convergence of an hypoelliptic diffusion’. In: *Electron. Commun. Probab.* 21 (Oct. 2016), pp. 1–14. DOI: [10.1214/16-ECP25](#).
- [123] Y. Guo. ‘The Landau equation in a periodic box’. In: *Comm. Math. Phys.* 231.3 (2002), pp. 391–434.
- [124] P. Gustafson. ‘A guided walk Metropolis algorithm’. In: *Stat. Comput.* 8 (1998), pp. 357–364. DOI: [10.1023/A:1008880707168](#).

- [125] L. Hahn. *Steady state and mixing of two run-and-tumble particles interacting through jamming and attractive forces*. Jan. 2025. arXiv: [2501.11379 \[math.PR\]](#).
- [126] L. Hahn, A. Guillin and M. Michel. ‘Jamming pair of general run-and-tumble particles: exact results, symmetries and steady-state universality classes’. In: *J. Phys. A: Math. Theor.* 58.15 (Apr. 2025). DOI: [10.1088/1751-8121/adc3df](#).
- [127] E. Hairer, C. Lubich and G. Wanner. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. 2nd ed. Vol. 31. Springer Series in Computational Mathematics. Berlin, Heidelberg: Springer, 2006. ISBN: 978-3-540-30663-4. DOI: [10.1007/3-540-30666-8](#).
- [128] D. Han-Kwan and M. Léautaud. ‘Geometric Analysis of the Linear Boltzmann Equation I. Trend to Equilibrium’. In: *Ann. PDE* 1.3 (Dec. 2015). DOI: [10.1007/s40818-015-0003-z](#).
- [129] J. Harland, M. Michel, T. A. Kampmann and J. Kierfeld. ‘Event-chain Monte Carlo algorithms for three- and many-particle interactions’. In: *EPL* 117.3 (Mar. 2017). DOI: [10.1209/0295-5075/117/30001](#).
- [130] W. K. Hastings. ‘Monte Carlo sampling methods using Markov chains and their applications’. In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: [10.1093/biomet/57.1.97](#).
- [131] B. Helffer and F. Nier. *Hypoellipticity and Spectral Theory for Fokker–Planck Operators and Witten Laplacians*. Vol. 1862. Lecture Notes in Mathematics. Berlin: Springer, 2005.
- [132] F. Hérau. ‘Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation’. In: *Asymptot. Anal.* 46.3-4 (2006), pp. 349–359. DOI: [10.3233/ASY-2006-741](#).
- [133] F. Hérau. ‘Short and long time behavior of the Fokker-Planck equation in a confining potential and applications’. In: *J. Funct. Anal.* 244.1 (Mar. 2007), pp. 95–118. DOI: [10.1016/j.jfa.2006.11.013](#).
- [134] F. Hérau and F. Nier. ‘Isotropic hypoellipticity and trend to equilibrium for the Fokker–Planck equation with a high-degree potential’. In: *Arch. Ration. Mech. Anal.* 171.2 (2004), pp. 151–218.
- [135] D. P. Herzog. ‘Exponential relaxation of the Nosé-Hoover equation under Brownian heating’. In: *Commun. Math. Sci.* 16 (2018), pp. 2231–2260. DOI: [10.4310/CMS.2018.v16.n8.a8](#).
- [136] M. D. Hoffman and A. Gelman. ‘The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo’. In: *J. Mach. Learn. Res.* 15 (2014), pp. 1593–1623. DOI: [10.5555/3044805.3045433](#).
- [137] P. Holderrieth. ‘Cores for piecewise-deterministic Markov processes used in Markov chain Monte Carlo’. In: *Electron. Commun. Probab.* 26 (2021), pp. 1–12. DOI: [10.1214/21-ECP430](#).

- [138] R. Holley and D. Stroock. ‘Logarithmic Sobolev inequalities and stochastic Ising models’. In: *J. Stat. Phys.* 46 (Mar. 1987), pp. 1159–1194. DOI: [10.1007/BF01011161](#).
- [139] J. G. Hooton. ‘Compact Sobolev imbeddings on finite measure spaces’. In: *J. Math. Anal. Appl.* 83.2 (Oct. 1981), pp. 570–581. DOI: [10.1016/0022-247X\(81\)90142-6](#).
- [140] L. Hörmander. ‘Hypoelliptic second order differential equations’. In: *Acta Math.* 119 (Dec. 1967), pp. 147–171. DOI: [10.1007/BF02392081](#).
- [141] A. M. Horowitz. ‘A generalized guided Monte Carlo algorithm’. In: *Phys. Lett. B* 268.2 (Oct. 1991), pp. 247–252. DOI: [10.1016/0370-2693\(91\)90812-5](#).
- [142] C.-R. Hwang, S.-Y. Hwang-Ma and S.-J. Sheu. ‘Accelerating Gaussian Diffusions’. In: *Ann. Appl. Prob.* 3.3 (Aug. 1993), pp. 897–913. DOI: [10.1214/aoap/1177005371](#).
- [143] C.-R. Hwang, S.-Y. Hwang-Ma and S.-J. Sheu. ‘Accelerating Diffusions’. In: *Ann. Appl. Prob.* 15.2 (May 2005), pp. 1433–1444. DOI: [10.1214/105051605000000025](#).
- [144] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. 2nd ed. Vol. 24. North Holland Mathematical Library. Amsterdam: North-Holland Publ. Co., 1989. ISBN: 978-0-444-87378-1. DOI: [10.1016/S0924-6509\(08\)70217-4](#).
- [145] A. Jones and B. Leimkuhler. ‘Adaptive stochastic methods for sampling driven molecular systems’. In: *J. Chem. Phys.* 135.8 (2011). DOI: [10.1063/1.3626941](#).
- [146] J. Jost. *Riemannian Geometry and Geometric Analysis*. Universitext. Springer, 2017. ISBN: 978-3-319-61860-9. DOI: [10.1007/978-3-319-61860-9](#).
- [147] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Undergraduate Texts in Mathematics. New York: Springer, 1976. ISBN: 978-0-387-90192-3.
- [148] R. Khasminskii. *Stochastic Stability of Differential Equations*. 2nd ed. Vol. 27. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer, 2012. ISBN: 978-3-642-23279-4. DOI: [10.1007/978-3-642-23280-0](#).
- [149] A. Kolesnikov and E. Milman. ‘Brascamp–Lieb-type inequalities on weighted Riemannian manifolds with boundary’. In: *J. Geom. Anal.* 27 (Sept. 2017), pp. 1680–1702. DOI: [10.1007/s12220-016-9736-5](#).
- [150] A. V. Kolesnikov and E. Milman. ‘Riemannian metrics on convex sets with applications to Poincaré and log-Sobolev inequalities’. In: *Calc. Var.* 55.77 (June 2016). DOI: [10.1007/s00526-016-1018-3](#).
- [151] A. V. Kolesnikov and E. Milman. ‘Poincaré and Brunn-Minkowski inequalities on the boundary of weighted Riemannian manifolds’. In: *Am. J. Math.* 140.5 (Oct. 2018), pp. 1147–1185. DOI: [10.1353/ajm.2018.0027](#).
- [152] A. Kolmogoroff. ‘Zufällige Bewegungen (zur Theorie der Brownschen Bewegung)’. In: *Ann. Math.* 35.1 (Jan. 1934), pp. 116–117. DOI: [10.2307/1968123](#).

- [153] T. Komorowski, C. Landim and S. Olla. *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation*. Vol. 345. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 2012. ISBN: 978-3-642-29880-6. DOI: [10.1007/978-3-642-29880-6](#).
- [154] Y. Kook, Y. T. Lee, R. Shen and S. Vempala. ‘Condition-number-independent Convergence Rate of Riemannian Hamiltonian Monte Carlo with Numerical Integrators’. In: *Proc. Mach. Learn. Res.* 195 (July 2023), pp. 4504–4569.
- [155] E. Kosygina and J. Peterson. *Convergence of rescaled “true” self-avoiding walks to the Tóth-Werner “true” self-repelling motion*. 2025. arXiv: [2502.10960 \[math.PR\]](#).
- [156] H. A. Kramers. ‘Brownian motion in a field of force and the diffusion model of chemical reactions’. In: *Physica* 7.4 (1940), pp. 284–304. DOI: [10.1016/S0031-8914\(40\)90098-2](#).
- [157] W. Krauth. ‘Event-Chain Monte Carlo: Foundations, Applications, and Prospects’. In: *Front. Phys.* 9 (June 2021). DOI: [10.3389/fphy.2021.663457](#).
- [158] W. Krauth. *Hamiltonian Monte Carlo vs. event-chain Monte Carlo: an appraisal of sampling strategies beyond the diffusive regime*. 2024. arXiv: [2411.11690](#).
- [159] S. Kullback and R. A. Leibler. ‘On Information and Sufficiency’. In: *Ann. Math. Stat.* 22.1 (1951), pp. 79–86. DOI: [10.1214/aoms/1177729694](#).
- [160] S. Kullback. ‘A Lower Bound for Discrimination Information in Terms of Variation (Corresp.)’ In: *IEEE Trans. Inf. Theory* 13.1 (1967), pp. 126–127. DOI: [10.1109/TIT.1967.1053968](#).
- [161] S. Lan and L. Kang. ‘Sampling constrained continuous probability distributions: A review’. In: *WIREs Comp. Stat.* 15.6 (2023). DOI: [10.1002/wics.1608](#).
- [162] P. Langevin. ‘Sur la théorie du mouvement brownien’. In: *C. R. Acad. Sci. Paris* 146 (1908), pp. 530–533.
- [163] M. Ledoux. *The Concentration of Measure Phenomenon*. Vol. 89. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2001. ISBN: 978-0-821-83792-4.
- [164] J. Lee. *Introduction to Smooth Manifolds*. Vol. 218. Graduate Texts in Mathematics. New York: Springer, 2013. ISBN: 978-1-4419-9982-5. DOI: [10.1007/978-1-4419-9982-5](#).
- [165] J. Lee. *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer International Publishing, 2018. ISBN: 978-3-319-91755-9. DOI: [10.1007/978-3-319-91755-9](#).
- [166] Y. T. Lee and S. S. Vempala. ‘Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation’. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. New York, USA: Association for Computing Machinery, 2018, pp. 1115–1121. ISBN: 978-1-4503-5559-9. DOI: [10.1145/3188745.3188774](#).

- [167] Y. T. Lee and S. S. Vempala. ‘The Manifold Joys of Sampling’. In: *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Ed. by M. Bojańczyk, E. Merelli and D. P. Woodruff. Vol. 229. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 4:1–4:20. ISBN: 978-3-95977-235-8. DOI: [10.4230/LIPIcs.ICALP.2022.4](https://doi.org/10.4230/LIPIcs.ICALP.2022.4).
- [168] J. Lehec. ‘Convergence in total variation for the kinetic Langevin algorithm’. In: *Math. Stat. Learn.* (2025). DOI: [10.4171/MSL/49](https://doi.org/10.4171/MSL/49).
- [169] Z. Lei and W. Krauth. ‘Mixing and perfect sampling in one-dimensional particle systems’. In: *EPL* 124.2 (Nov. 2018). DOI: [10.1209/0295-5075/124/20003](https://doi.org/10.1209/0295-5075/124/20003).
- [170] Z. Lei, W. Krauth and A. C. Maggs. ‘Event-chain Monte Carlo with factor fields’. In: *Phys. Rev. E* 99 (Apr. 2019). DOI: [10.1103/PhysRevE.99.043301](https://doi.org/10.1103/PhysRevE.99.043301).
- [171] B. Leimkuhler, C. Matthews and G. Stoltz. ‘The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics’. In: *IMA J. Numer. Anal.* 36.1 (Jan. 2015), pp. 13–79. DOI: [10.1093/imanum/dru056](https://doi.org/10.1093/imanum/dru056).
- [172] B. Leimkuhler, M. Sachs and G. Stoltz. ‘Hypocoercivity properties of adaptive Langevin dynamics’. In: *SIAM J. Appl. Math.* 80.3 (2020), pp. 1197–1222.
- [173] B. Leimkuhler, M. Sachs and G. Stoltz. *Hypocoercivity properties of adaptive Langevin dynamics*. 2023. arXiv: [1908.09363v3](https://arxiv.org/abs/1908.09363v3) [[math.PR](https://arxiv.org/archive/math)].
- [174] T. Lelièvre, M. Rousset and G. Stoltz. ‘Langevin dynamics with constraints and computation of free energy differences’. In: *Mathematics of computation* 81.280 (2012), pp. 2071–2125.
- [175] T. Lelièvre, F. Nier and G. Pavliotis. ‘Optimal Non-reversible Linear Drift for the Convergence to Equilibrium of a Diffusion’. In: *J. Stat. Phys.* 152 (June 2013), pp. 237–274. DOI: [10.1007/s10955-013-0769-x](https://doi.org/10.1007/s10955-013-0769-x).
- [176] T. Lelièvre, M. Rousset and G. Stoltz. *Free Energy Computations. A Mathematical Perspective*. London: Imperial College Press, 2010. ISBN: 978-1-908978-75-2. DOI: [10.1142/p579](https://doi.org/10.1142/p579).
- [177] T. Lelièvre, R. Santet and G. Stoltz. ‘Unbiasing Hamiltonian Monte Carlo algorithms for a general Hamiltonian function’. In: *Foundations of Computational Mathematics* (2024), pp. 1–74.
- [178] T. Lelièvre and G. Stoltz. ‘Partial differential equations and stochastic methods in molecular dynamics’. In: *Acta Numer.* 25 (2016), pp. 681–880. DOI: [10.1017/S0962492916000039](https://doi.org/10.1017/S0962492916000039).
- [179] D. A. Levin, Y. Peres and E. L. Wilmer. *Markov Chains and Mixing Times*. 2nd ed. Providence, RI: American Mathematical Society, 2017. ISBN: 978-1-4704-2962-1. DOI: [10.1090/mbk/107](https://doi.org/10.1090/mbk/107).
- [180] P. Li. *Geometric Analysis*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2012. DOI: [10.1017/CB09781139105798](https://doi.org/10.1017/CB09781139105798).

- [181] H. C. Lie, D. Rudolf, B. Sprungk and T. J. Sullivan. ‘Dimension-independent Markov chain Monte Carlo on the sphere’. In: *Scand. J. Stat.* 50.4 (2023), pp. 1818–1858. DOI: [10.1111/sjos.12653](https://doi.org/10.1111/sjos.12653).
- [182] T. M. Liggett. *Continuous Time Markov Processes: An Introduction*. Vol. 113. American Mathematical Soc., 2010. ISBN: 978-1-4704-1175-6. DOI: [10.1090/gsm/113](https://doi.org/10.1090/gsm/113).
- [183] T. Lindvall. *Lectures on the Coupling Method*. Vol. 294. Wiley Series in Probability and Statistics: Applied Probability and Statistics Section. New York: John Wiley & Sons, 1992. ISBN: 978-0-4715-4025-0.
- [184] P.-L. Lions. *Mathematical Topics in Fluid Mechanics. Volume 1: Incompressible Models*. Vol. 3. Oxford Lecture Series in Mathematics and its Applications. Oxford: Clarendon Press, 1996. ISBN: 0-19-851487-5.
- [185] F. Lörler. *Convergence rates of self-repelling diffusions on Riemannian manifolds*. Nov. 2025. arXiv: [2511.23333](https://arxiv.org/abs/2511.23333) [[math.PR](https://arxiv.org/archive/math)].
- [186] J. Lu and L. Wang. ‘On explicit  $L^2$ -convergence rate estimate for piecewise deterministic Markov processes in MCMC algorithms’. In: *Ann. Appl. Probab.* 32.2 (Apr. 2022), pp. 1333–1361. DOI: [10.1214/21-AAP1710](https://doi.org/10.1214/21-AAP1710).
- [187] L. Ma and S.-H. Du. ‘Extension of Reilly formula with applications to eigenvalue estimates for drifting Laplacians’. In: *Compt. Rendus Math.* 348.21-22 (2010), pp. 1203–1206. DOI: [10.1016/j.crma.2010.10.003](https://doi.org/10.1016/j.crma.2010.10.003).
- [188] A. C. Maggs. ‘Non-reversible Monte Carlo: An example of “true” self-repelling motion’. In: *EPL* 147.2 (Aug. 2024). DOI: [10.1209/0295-5075/ad64ff](https://doi.org/10.1209/0295-5075/ad64ff).
- [189] A. C. Maggs. ‘Event-chain Monte Carlo and the true self-avoiding walk’. In: *Phys. Rev. E* 111.5 (May 2025). DOI: [10.1103/PhysRevE.111.054104](https://doi.org/10.1103/PhysRevE.111.054104).
- [190] K. Malakar et al. ‘Steady state, relaxation and first-passage properties of a run-and-tumble particle in one-dimension’. In: *J. Stat. Mech.* 2018.4 (2018). DOI: [10.1088/1742-5468/aab84f](https://doi.org/10.1088/1742-5468/aab84f).
- [191] E. Mallmin, R. A. Blythe and M. R. Evans. ‘Exact spectral solution of two interacting run-and-tumble particles on a ring lattice’. In: *J. Stat. Mech.* 2019.1 (2019). DOI: [10.1088/1742-5468/aaf631](https://doi.org/10.1088/1742-5468/aaf631).
- [192] B. Massoulié, C. Erignoux, C. Toninelli and W. Krauth. ‘Velocity Trapping in the Lifted Totally Asymmetric Simple Exclusion Process and the True Self-Avoiding Random Walk’. In: *Phys. Rev. Lett.* 135 (Sept. 2025). DOI: [10.1103/mqdr-x95j](https://doi.org/10.1103/mqdr-x95j).
- [193] G. Metafune, D. Pallara and E. Priola. ‘Spectrum of Ornstein-Uhlenbeck Operators in  $L^p$  Spaces with Respect to Invariant Measures’. In: *J. Funct. Anal.* 196.1 (2002), pp. 40–60. DOI: [10.1006/jfan.2002.3978](https://doi.org/10.1006/jfan.2002.3978).
- [194] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller. ‘Equation of state calculations by fast computing machines’. In: *J. Chem. Phys.* 21.6 (1953), pp. 1087–1092. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114).

- [195] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. 2nd ed. Cambridge: Cambridge University Press, 2009. ISBN: 978-0-511-62663-0. DOI: [10.1017/CB09780511626630](#).
- [196] M. Michel, A. Durmus and S. S  n  cal. ‘Forward event-chain Monte Carlo: fast sampling by randomness control in irreversible Markov chains’. In: *J. Comput. Graph. Stat.* 29.4 (2020), pp. 689–702. DOI: [10.1080/10618600.2020.1750417](#).
- [197] M. Michel, S. C. Kapfer and W. Krauth. ‘Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps’. In: *J. Chem. Phys.* 140.5 (Feb. 2014), p. 054116. DOI: [10.1063/1.4863991](#).
- [198] P. Monmarch  . *A note on Fisher Information hypocoercive decay for the linear Boltzmann equation*. Oct. 2020. arXiv: [1703.10504 \[math.AP\]](#).
- [199] P. Monmarch  . ‘High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion.’ In: *Electron. J. Stat.* 15.2 (Sept. 2021), pp. 4117–4166. DOI: [10.1214/21-EJS1888](#).
- [200] W. Mou, Y.-A. Ma, M. J. Wainwright, P. L. Bartlett and M. I. Jordan. ‘High-order Langevin diffusion yields an accelerated MCMC algorithm’. In: *J. Mach. Learn. Res.* 22.1 (2021), pp. 1919–1959.
- [201] T. Mountford and P. Tarr  s. ‘An asymptotic result for Brownian polymers’. In: *Ann. Inst. H. Poincar   Probab. Stat.* 44.1 (2008), pp. 29–46. DOI: [10.1214/07-AIHP113](#).
- [202] R. M. Neal. *Bayesian Learning for Neural Networks*. 1st ed. Vol. 118. Lecture Notes in Statistics. New York, NY: Springer, 1996, p. 204. ISBN: 978-0-387-94724-2. DOI: [10.1007/978-1-4612-0745-0](#).
- [203] R. M. Neal. *Improving Asymptotic Variance of MCMC Estimators: Non-reversible Chains are Better*. July 2004. arXiv: [math/0407281 \[math.PR\]](#).
- [204] R. M. Neal. ‘MCMC using Hamiltonian dynamics’. In: *Handbook of Markov Chain Monte Carlo*. Ed. by S. Brooks, A. Gelman, G. Jones and X.-L. Meng. Handbooks of Modern Statistical Methods. New York: CRC Press, 2011, pp. 113–162. ISBN: 978-1-4200-7942-5. DOI: [10.1201/b10905](#).
- [205] E. Nelson. *Dynamical Theories of Brownian Motion*. 1st ed. Princeton, NJ: Princeton University Press, 1967. ISBN: 978-0-691-07950-9. DOI: [10.1515/9780691219615](#).
- [206] C. M. Newman and K. Ravishankar. ‘Convergence of the T  th lattice filling curve to the T  th-Werner plane filling curve’. In: *ALEA, Lat. Am. J. Probab. Math. Stat.* 1 (2006), pp. 333–346.
- [207] Y. Nishikawa, M. Michel, W. Krauth and K. Hukushima. ‘Event-chain algorithm for the Heisenberg model: Evidence for  $z \simeq 1$  dynamic scaling’. In: *Phys. Rev. E* 92 (Dec. 2015). DOI: [10.1103/PhysRevE.92.063306](#).

- [208] J. Norris, L. Rogers and D. Williams. ‘An excluded-volume problem for Brownian motion’. In: *Phys. Lett. A* 112.1 (1985), pp. 16–18. DOI: [10.1016/0375-9601\(85\)90451-7](https://doi.org/10.1016/0375-9601(85)90451-7).
- [209] J. Norris, L. Rogers and D. Williams. ‘Self-avoiding random walk: a Brownian motion model with local time drift’. In: *Probab. Theory Relat. Fields* 74.2 (1987), pp. 271–287. DOI: [10.1007/BF00569993](https://doi.org/10.1007/BF00569993).
- [210] S. Nosé. ‘A unified formulation of the constant temperature molecular dynamics methods’. In: *J. Chem. Phys.* 81.1 (1984), pp. 511–519.
- [211] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. 6th ed. Universitext. Berlin, Heidelberg: Springer, 2003. ISBN: 978-3-540-04758-2. DOI: [10.1007/978-3-642-14394-6](https://doi.org/10.1007/978-3-642-14394-6).
- [212] M. Ottobre and G. A. Pavliotis. ‘Asymptotic analysis for the generalized Langevin equation’. In: *Nonlinearity* 24.5 (2011), p. 1629.
- [213] G. Parisi. ‘Correlation functions and computer simulations’. In: *Nucl. Phys. B* 180.3 (1981), pp. 378–384. DOI: [10.1016/0550-3213\(81\)90056-0](https://doi.org/10.1016/0550-3213(81)90056-0).
- [214] G. A. Pavliotis, G. Stoltz and U. Vaes. ‘Scaling limits for the generalized Langevin equation’. In: *J. Nonlinear Sci.* 31.1 (2021), p. 8.
- [215] G. A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Vol. 60. Texts in Applied Mathematics. New York: Springer, 2014. ISBN: 978-1-4939-1323-7. DOI: [10.1007/978-1-4939-1323-7](https://doi.org/10.1007/978-1-4939-1323-7).
- [216] E. A. J. F. Peters and G. de With. ‘Rejection-free Monte Carlo sampling for general potentials’. In: *Phys. Rev. E* 85 (2 Feb. 2012). DOI: [10.1103/PhysRevE.85.026703](https://doi.org/10.1103/PhysRevE.85.026703).
- [217] M. Reed and B. Simon. *Methods of modern mathematical physics I: Functional analysis*. Academic Press, 1980. ISBN: 978-0-12-585050-6.
- [218] R. Reilly. ‘Applications of the Hessian Operator in a Riemannian Manifold’. In: *Indiana Univ. Math. J.* 26.3 (1977), pp. 459–472. DOI: [10.1512/iumj.1977.26.26036](https://doi.org/10.1512/iumj.1977.26.26036).
- [219] A. Rényi. ‘On measures of entropy and information’. In: *Berkeley Symp. on Math. Stat. and Prob.* 4.1 (1961), pp. 547–561.
- [220] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. 3rd ed. Vol. 293. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 1999, p. 602. ISBN: 978-3-540-64325-8. DOI: [10.1007/978-3-662-06400-9](https://doi.org/10.1007/978-3-662-06400-9).
- [221] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. 2nd ed. Springer Texts in Statistics. New York: Springer, 2004. ISBN: 9780387212395.
- [222] G. O. Roberts and R. L. Tweedie. ‘Exponential Convergence of Langevin Distributions and Their Discrete Approximations’. In: *Bernoulli* 2.4 (1996), pp. 341–363. DOI: [10.2307/3318418](https://doi.org/10.2307/3318418).

- [223] M. Röckner and F.-Y. Wang. ‘Weak Poincaré Inequalities and L2-Convergence Rates of Markov Semigroups’. In: *J. Funct. Anal.* 185.2 (2001), pp. 564–603. DOI: [10.1006/jfan.2001.3776](#).
- [224] O. S. Rothaus. ‘Lower bounds for eigenvalues of regular Sturm-Liouville operators and the logarithmic Sobolev inequality’. In: *Duke Math. J.* 45.2 (1978), pp. 351–362. DOI: [10.1215/S0012-7094-78-04518-0](#).
- [225] G. Royer. *An Initiation to Logarithmic Sobolev Inequalities*. Vol. 14. SMF/AMS Texts and Monographs. Providence, RI: American Mathematical Society and Société Mathématique de France, 2007, p. 119. ISBN: 978-0-8218-4401-4.
- [226] T. Sakai. *Riemannian Geometry*. Vol. 149. Translations of Mathematical Monographs. American Mathematical Society, 1997. ISBN: 0-8218-0284-4.
- [227] L. Saloff-Coste. ‘Lectures on finite Markov chains’. In: *Lectures on Probability Theory and Statistics: École d’Été de Probabilités de Saint-Flour XXVI-1996*. Ed. by P. Bernard. Berlin: Springer, 1997, pp. 301–413. ISBN: 978-3-540-69210-2. DOI: [10.1007/BFb0092621](#).
- [228] S. Sasaki. ‘On the differential geometry of tangent bundles of Riemannian manifolds’. In: *Tohoku Math. J.* 10.3 (1958), pp. 338–354. DOI: [10.2748/tmj/1178244668](#).
- [229] K.-I. Sato. ‘Potential Operators for Markov Processes’. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*. 1972. ISBN: 978-0-520-37591-8. DOI: [10.1525/9780520375918-013](#).
- [230] M. J. Schnitzer. ‘Theory of continuum random walks and application to chemotaxis’. In: *Phys. Rev. E* 48.4 (1993). DOI: [10.1103/PhysRevE.48.2553](#).
- [231] K. Schuh and P. A. Whalley. *Convergence of kinetic Langevin samplers for non-convex potentials*. 2025. arXiv: [2405.09992 \[math.PR\]](#).
- [232] A. Slowman, M. Evans and R. Blythe. ‘Jamming and attraction of interacting run-and-tumble random walkers’. In: *Phys. Rev. Lett.* 116.21 (2016). DOI: [10.1103/PhysRevLett.116.218101](#).
- [233] R. L. Smith. ‘Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions’. In: *Oper. Res.* 32.6 (1984), pp. 1296–1308. DOI: [10.1287/opre.32.6.1296](#).
- [234] W. Stannat. ‘(Nonsymmetric) Dirichlet operators on  $L^1$ : existence, uniqueness and associated Markov processes’. In: *Ann. Sc. Norm. Super. Pisa, Cl. Sci., IV. Ser.* 28.1 (1999), pp. 99–140.
- [235] D. W. Stroock. *Probability Theory: An Analytic View*. 3rd ed. Cambridge: Cambridge University Press, 2024. ISBN: 978-1-009-54903-5. DOI: [10.1017/9781009549035](#).

- [236] K.-T. Sturm. ‘Analysis on local Dirichlet spaces. I. Recurrence, conservativeness and  $L^p$ -Liouville properties’. In: *J. Reine Angew. Math.* 456 (1994), pp. 173–196. DOI: [10.1515/crll.1994.456.173](#).
- [237] S. Tabachnikov. *Geometry and Billiards*. Vol. 30. Student Mathematical Library. American Mathematical Soc., 2005. DOI: [10.1090/stml/030](#).
- [238] D. Talay. ‘Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme’. In: *Markov Process. Relat. Fields* 8.2 (2002), pp. 163–198.
- [239] P. Tarrès, B. Tóth and B. Valkó. ‘Diffusivity bounds for 1D Brownian polymers’. In: *Ann. Probab.* 40.2 (2012), pp. 695–713. DOI: [10.1214/10-AOP630](#).
- [240] M. E. Taylor. *Partial Differential Equations II. Qualitative Studies of Linear Equations*. Applied Mathematical Sciences. Springer Cham, 2023. ISBN: 978-3-031-33700-0. DOI: [10.1007/978-3-031-33700-0](#).
- [241] B. Tóth. ‘The “True” Self-Avoiding Walk with Bond Repulsion on  $\mathbb{Z}$ : Limit Theorems’. In: *Ann. Probab.* 23.4 (Oct. 1995), pp. 1523–1556. DOI: [10.1214/aop/1176987793](#).
- [242] B. Tóth and B. Vető. ‘Continuous time ‘true’ self-avoiding random walk on  $\mathbb{Z}$ ’. In: *ALEA, Lat. Am. J. Probab. Math. Stat.* 8 (2011), pp. 59–75.
- [243] B. Tóth and W. Werner. ‘The true self-repelling motion’. In: *Probab. Theory Relat. Fields* 111 (1998), pp. 375–452. DOI: [10.1007/s004400050172](#).
- [244] V. F. Turchin. ‘On the Computation of Multidimensional Integrals by the Monte-Carlo Method’. In: *Theory Probab. Appl.* 16 (1971), pp. 720–724. DOI: [10.1137/1116083](#).
- [245] P. Vanetti, A. Bouchard-Côté, G. Deligiannidis and A. Doucet. *Piecewise-Deterministic Markov Chain Monte Carlo*. 2017. arXiv: [1707.05296 \[stat.ME\]](#).
- [246] T. Vicsek and A. Zafeiris. ‘Collective motion’. In: *Phys. Rep.* 517.3-4 (2012), pp. 71–140. DOI: [10.1016/j.physrep.2012.03.004](#).
- [247] C. Villani. *Optimal Transport: Old and New*. 1st ed. Vol. 338. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 2008, p. 976. ISBN: 978-3-540-71049-3. DOI: [10.1007/978-3-540-71050-9](#).
- [248] C. Villani. *Hypocoercivity*. Vol. 202. Memoirs of the American Mathematical Society 950. Providence, RI: American Mathematical Society, 2009. ISBN: 978-0-8218-4498-4. DOI: [10.1090/S0065-9266-09-00567-5](#).
- [249] F.-Y. Wang. ‘Functional inequalities for empty essential spectrum’. In: *J. Funct. Anal.* 170.1 (2000), pp. 219–245. DOI: [10.1006/jfan.1999.3516](#).
- [250] F.-Y. Wang. *Analysis for Diffusion Processes on Riemannian Manifolds*. 1st ed. Vol. 18. Advanced Series on Statistical Science and Applied Probability. Singapore: World Scientific, 2014. ISBN: 978-981-4452-64-9.

- [251] N. Wielens. ‘The essential selfadjointness of generalized Schrödinger operators’. In: *J. Funct. Anal.* 61.1 (1985), pp. 98–115. DOI: [10.1016/0022-1236\(85\)90040-0](https://doi.org/10.1016/0022-1236(85)90040-0).
- [252] C. Wu and C. P. Robert. *Generalized Bouncy Particle Sampler*. June 2017. arXiv: [1706.04781](https://arxiv.org/abs/1706.04781) [[stat.CO](#)].
- [253] S.-J. Wu, C.-R. Hwang and M. T. Chu. ‘Attaining the optimal Gaussian diffusion acceleration’. In: *J. Stat. Phys.* 155.3 (2014), pp. 571–590. DOI: [10.1007/s10955-014-0963-5](https://doi.org/10.1007/s10955-014-0963-5).