

## Non-reversible processes

Duong, Manh Hong; Ottobre, Michela

DOI:

[10.1088/1361-6544/acb47b](https://doi.org/10.1088/1361-6544/acb47b)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Duong, MH & Ottobre, M 2023, 'Non-reversible processes: GENERIC, hypocoercivity and fluctuations', *Nonlinearity*, vol. 36, no. 3, 1617. <https://doi.org/10.1088/1361-6544/acb47b>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



PAPER • OPEN ACCESS

## Non-reversible processes: GENERIC, hypocoercivity and fluctuations

To cite this article: M H Duong and M Ottobre 2023 *Nonlinearity* **36** 1617

View the [article online](#) for updates and enhancements.

### You may also like

- [Absorbing state phase transitions with a non-accessible vacuum](#)  
Omar Al Hammal, Juan A Bonachela and Miguel A Muñoz
- [Existence and stability of 3-site breathers in a triangular lattice](#)  
Vassilis Koukouloyannis and Robert S MacKay
- [Effective dynamics for non-reversible stochastic differential equations: a quantitative study](#)  
Frédéric Legoll, Tony Lelièvre and Upanshu Sharma

# Non-reversible processes: GENERIC, hypocoercivity and fluctuations

M H Duong<sup>1</sup> and M Ottobre<sup>2,\*</sup> 

<sup>1</sup> School of Mathematics, University of Birmingham, Birmingham B15 2TT, United Kingdom

<sup>2</sup> Maxwell Institute for Mathematical Sciences, Department of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

E-mail: [m.ottobre@hw.ac.uk](mailto:m.ottobre@hw.ac.uk)

Received 30 October 2021; revised 11 October 2022

Accepted for publication 18 January 2023

Published 3 February 2023

Recommended by Dr Sandra Cerrai



CrossMark

## Abstract

We consider two approaches to study non-reversible Markov processes, namely the hypocoercivity theory and general equations for non-equilibrium reversible–irreversible coupling; the basic idea behind both of them is to split the process into a reversible component and a non-reversible one, and then quantify the way in which they interact. We compare such theories and provide explicit formulas to pass from one formulation to the other; as a bi-product we give a simple proof of the link between reversibility of the dynamics and gradient flow structure of the associated Fokker–Planck equation. We do this both for linear Markov processes and for a class of nonlinear Markov process as well. We then characterise the structure of the large deviation functional of generalised-reversible processes; this is a class of non-reversible processes of large relevance in applications. Finally, we show how our results apply to two classes of Markov processes, namely non-reversible diffusion processes and a class of piecewise deterministic Markov processes (PDMPs), which have recently attracted the attention of the statistical sampling community. In particular, for the PDMPs we consider we prove entropy decay.

\* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: GENERIC, hypocoercivity, large deviation principles, non-reversible processes, gradient flows, diffusion processes, piecewise deterministic Markov processes

Mathematics Subject Classification numbers: 35Q82, 35Q84, 60J25, 60F10, 60H30, 82B35, 82C31

## 1. Introduction

The study of non-reversible Markov processes has attracted the attention of the mathematics and physics communities for several decades; while significant progress and more understanding has been brought by a large body of research, it is fair to say that several important questions regarding non-reversibility still remain unanswered. Non-reversible Markov processes are fundamental models in non-equilibrium statistical mechanics, for example in the study of so-called *open systems* [48], see e.g. the vast literature on heat baths [21, 30], and in collisional kinetic theory, prominently in connection with the study of the Boltzmann equation, in its many forms and simplifications [49, 52, 53]. More recently, the observation that non-reversible processes might enjoy favourable properties in terms of speed of convergence to equilibrium, has attracted also the statistical sampling community to further investigations in this field, so that the applications-driven need to produce increasingly high performing algorithms has played an important role in pushing forward the theory of non-reversible processes [2, 4, 17, 41], with particular reference to the class of piecewise deterministic Markov process (PDMPs) [11, 23–26, 51].

In this paper we use the word *reversible* to refer to stochastic processes which are *time-reversible*, i.e. for any  $T > 0$  the process  $\{X_t\}_{t \in [0, T]}$  and its time-reversed  $\{X_{T-t}\}_{t \in [0, T]}$  have the same distributions (on the space of continuous paths); because of this time-symmetry, reversible processes give rise to partial differential equations (PDEs) associated with symmetric (i.e. formally self-adjoint) operators [45]. However we point out that the term ‘reversible’ is used with very different meanings throughout the literature, and certainly across the strands of research that we use in this paper; so, to avoid confusion, we clarify matters in section 2.4.

The theory of reversible processes is by far more settled than its non-reversible counterpart, and this is true both of probabilistic and of analytic/functional analytic approaches. Nonetheless conceptual frameworks for the analysis and modelling of non-reversible phenomena do exist; in this paper we consider two of them, the hypocoercivity theory (HT), initiated by Herau [31] and then made systematic by Villani [52], and general equations for non-equilibrium reversible–irreversible coupling (GENERIC), whose first complete exposition can be found in [40]. The premise and purpose of these two theories is quite different: the HT is from its inception a functional analytic theory aimed at studying exponentially fast convergence to equilibrium for non-reversible processes; GENERIC was born in the physics and engineering community, as a framework to help applied scientists model reversible (dissipative) and non-reversible (conservative) contributions to a given dynamics, and only later evolved into a more mathematised theory, by hands of [33, 37, 38], and references therein, which also significantly pushed it forward. One of the main purposes of this paper is to observe that the formulations of such theories, which can look quite different from the outset, are indeed complementary and substantially equivalent, in the sense that we can provide formulas to pass from one formulation to the other, and we do so in sections 3 and 4; we also give examples to show how this fact can be exploited in practice.

To give more background on these two theories let us start with a toy example. Consider the second order Langevin equation, that is, the evolution

$$dx_t = v_t dt, \quad dv_t = -\partial_x V(x_t) dt - v_t dt + \sqrt{2} dW_t, \tag{1}$$

where  $(x_t, v_t) \in \mathbb{R}^2$  (for simplicity),  $W_t$  is one dimensional standard Brownian motion, the potential  $V: \mathbb{R} \rightarrow \mathbb{R}$  is smooth, confining (i.e.  $V(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ ) and grows at least quadratically at infinity. Under these conditions the process (1) admits a unique invariant measure, namely the measure with density  $\mu: \mathbb{R}^2 \rightarrow \mathbb{R}$  given by

$$\mu(x, v) = \frac{1}{Z} e^{-V(x)} e^{-\frac{v^2}{2}}, \tag{2}$$

where  $Z$  is a normalisation constant. To the SDE (1) one can associate two PDEs, namely the Kolmogorov equation (KE), describing the evolution of so-called observables, i.e. of quantities of the form  $\mathbb{E}(f(x_t, v_t) | (x_0, v_0) = (x, v))$ , for any given  $f \in C_b(\mathbb{R}^2)$ <sup>3</sup>, and the Fokker–Planck (FP) equation, describing the evolution of the law of the process; in the case of the diffusion (1) such equations take, respectively, the form

$$\partial_t u_t(x, v) = \mathcal{L}u_t := v\partial_x u_t - \partial_x V(x)\partial_v u_t - v\partial_v u_t + \partial_v^2 u_t, \quad u_0(x, v) = f(x, v), \tag{3}$$

and

$$\partial_t \rho_t(x, v) = \mathcal{L}'\rho_t := -v\partial_x \rho_t + \partial_x V(x)\partial_v \rho_t + \partial_v(v\rho_t) + \partial_v^2 \rho_t, \quad \rho_0 = \text{Law}(x_0, v_0), \tag{4}$$

respectively, where  $\mathcal{L}$  is the operator defined on smooth functions  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  as

$$\mathcal{L}g := v\partial_x g - \partial_x V(x)\partial_v g - v\partial_v g + \partial_v^2 g,$$

usually referred to as the *generator* of the system, and  $'$  denotes (formal) adjoint in  $L^2(\mathbb{R}^2) := \{f: \mathbb{R}^2 \rightarrow \mathbb{R} : \int_{\mathbb{R}^2} |f(x, v)|^2 d\mu < \infty\}$ ; the operator  $\mathcal{L}'$  is also called the *FP operator* (we will give more context on these equations in section 2).

The evolution (1) can be split into a Hamiltonian component, namely

$$dx_t = v_t dt, \quad dv_t = -\partial_x V(x_t) dt,$$

plus an Ornstein–Uhlenbeck (OU) process

$$dv_t = -v_t dt + \sqrt{2} dW_t.$$

The generator of the Hamiltonian dynamics is the Liouville operator  $B := v\partial_x - \partial_x V(x)\partial_v$ ; this operator is antisymmetric in  $L^2_\mu(\mathbb{R}^2) := \{f: \mathbb{R}^2 \rightarrow \mathbb{R} : \int_{\mathbb{R}^2} |f(x, v)|^2 \mu(x, v) dx dv < \infty\}$ . Because of this antisymmetry, along the flow generated by  $B$ , the  $L^2_\mu$ -norm is conserved (see section 2.2). The generator of the OU process is instead the operator  $\mathcal{L}_{OU} = -v\partial_v + \partial_v^2$ ; by setting  $A = \partial_v$ , we have  $\mathcal{L}_{OU} = -A^*A$ , where  $*$  denotes adjoint in  $L^2_\mu$ , so that  $A^* = -\partial_v + v$ . Therefore, overall, equation (3) can be written in the by now classic (linear) hypocoercive form

$$\partial_t u_t = Bu_t - A^*Au_t. \tag{5}$$

While  $B$  is antisymmetric in  $L^2_\mu$ , trivially, the operator  $A^*A$  is symmetric in  $L^2_\mu$ ; moreover, along the flow generated by  $-A^*A$ , the  $L^2_\mu$ -norm is dissipated, see again section 2.2. Writing the dynamics in the above form, is the starting point of the HT and it serves the purpose of emphasising the splitting of the dynamics into its symmetric/dissipative and antisymmetric/-conservative parts. Once the dynamics has been cast in the form (5), dissipation of the  $L^2_\mu$ -norm

<sup>3</sup> Here and throughout  $C_b(\mathbb{R}^d)$  denotes the set of functions  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  which are continuous and bounded.

along the flow generated by  $\mathcal{L}$  becomes trivial to show, at least if one is not after a rate. The aim of the HT is to determine sufficient conditions on  $A$  and  $B$  such that dissipation towards the steady state  $\mu$  is exponentially fast.

The setting of GENERIC is, in principle, analogous, however the theory is more adapted to working with the FP formulation of the dynamics; informally (precise definition in section 2.3), we say that the evolution (4) is in GENERIC form<sup>4</sup> if it can be written in the form

$$\partial_t \rho_t = W\rho_t - M_{\rho_t} \left( \frac{1}{2} \mathbf{dS}(\rho_t) \right), \tag{6}$$

where  $W$  is some operator,  $\mathcal{S}$  is a real-valued functional,  $\mathbf{dS}$  is the functional derivative of  $\mathcal{S}$  (i.e. an appropriate derivative of  $\mathcal{S}$  with respect to its argument  $\rho$ , see section 2) and, for every fixed  $\rho$ ,  $M_\rho$  is a symmetric positive semidefinite operator (in  $L^2$ ); we furthermore require that the following orthogonality condition should hold:

$$(W(\rho), \mathbf{dS}(\rho)) = 0, \tag{7}$$

where  $(\cdot, \cdot)$  denotes scalar product in  $L^2$ . The FP equation (4) can be cast in GENERIC form, upon choosing  $W = -B$ , i.e. taking  $W$  to be the opposite of the Liouville operator, and furthermore choosing  $\mathcal{S}$  to be the *relative entropy* of the system,

$$\begin{aligned} \mathcal{S}_\mu(\rho) &= \int \rho(x, v) \log \left( \frac{\rho(x, v)}{\mu(x, v)} \right) dx dv \\ &\stackrel{(2)}{=} \int \left( \rho(x, v) \log \rho(x, v) + V(x)\rho(x, v) + \frac{v^2}{2} \rho(x, v) \right) dx dv, \end{aligned} \tag{8}$$

so that  $\mathbf{dS}_\mu(\rho)$  denotes the variational derivative (Frechet derivative) of  $\mathcal{S}$  with respect to  $\rho$  (and then calculated in  $\rho$ ), namely

$$\mathbf{dS}_\mu(\rho) = \log \rho + 1 + V(x) + \frac{v^2}{2};$$

and, finally, by taking for each  $\rho$ , the operator  $M_\rho$  (acting on a function  $f$ ) to be defined as

$$M_\rho(f) := -2\partial_v [\rho \partial_v f].$$

It is easy to show that, for each  $\rho$  fixed,  $M_\rho$  is a symmetric operator in  $L^2$ , see section 2.3. By observing that the Liouville operator  $B$  is antisymmetric in  $L^2$  as well, one can see that the decomposition (6) is in spirit analogous to the decomposition (5). The orthogonality condition ensures that the relative entropy  $\mathcal{S}_\mu$  is dissipated along the flow.

In view of the discussion that will follow it is important to point out that the second addend of equation (6), i.e. the part of the equation that can be cast in the form  $M_\rho(\mathbf{dS}(\rho))$ , is the so-called *gradient-flow* part of the equation, see [1, 32, 33, 38], more details on this in section 2. Hence the decomposition (6) can also be viewed as a splitting into a symmetric/dissipative/gradient-flow part and a antisymmetric/conservative/non-gradient flow component. As a bi-product of this discussion, we have the following: the generator  $\mathcal{L}_{OU}$  of the OU process is symmetric in  $L^2_\mu$  and associated to the dissipative/reversible part of the dynamics; its dual,  $\mathcal{L}'_{OU}$ , which, using (4), coincides with  $-M_\rho(\mathbf{dS}(\rho))$ , is symmetric in  $L^2$  and is associated to the gradient flow part of the dynamics.

This is no coincidence and it is indeed a specific instance of a well-known ‘meta-theorem’, substantiated by the work of Onsager and Machlup [39] as well as by e.g. [6, 34, 35, 47],

<sup>4</sup> To be precise, (6) is the pre-GENERIC form of the equation; pre-GENERIC is a more general formulation than GENERIC, we will recall the difference between such formulations in section 2.

according to which reversibility and gradient flow structure are related. One way of expressing and understanding this meta-theorem is as follows: consider a stochastic particle system made (for simplicity) of i.i.d. particles, suppose the process describing each particle is reversible and that the particle system converges to a deterministic limit, in the sense that the empirical measure associated with the particle system converges to a (deterministic) measure which satisfies a deterministic evolution equation. Then such a deterministic equation is in gradient flow form. This meta-theorem, expressed in this form, was settled into an actual theorem in [38], for a large class of reversible processes. This was done by going through a large deviation principle (LDP) for the particle system. The reason why LDPs enter the picture is quite profound: it turns out that the form of the large deviation functional (LDF) appearing in the LDP is directly related to the gradient flow structure and indeed that it determines it. More specifically, if we know the form of the LDF then we can find  $M$  (and  $\mathcal{S}$ ), see section 5 and [38] for precise statements. Note that typically every gradient flow evolution can be written in gradient flow form in more than one way, i.e. one can find several pairs  $M$  and  $\mathcal{S}$  to write the same equation in gradient flow form; so if the objective is to link reversibility and gradient flow structure, the LDP helps ‘select’ one such gradient flow structure. This fact, i.e. the fact that the LDP ‘selects’ one gradient flow form, is not in contradiction with the fact that the deterministic equation can be written in gradient flow form in more than one way: given a deterministic evolution, this evolution can be realised as limit of various particle systems (e.g. the particle system need not be i.i.d.). Two different particle systems that converge to the same deterministic limit will, in general, undergo different fluctuations and hence, assuming they both satisfy a LDP, such LDPs will correspond to different LDFs, which are then associated with different gradient flow structures [47]. This connection between evolutions with gradient flow structure and reversibility was then extended to a connection between evolutions in GENERIC form and ‘non-reversibility’ in [38], again making use of LDPs. In note 4 we elaborate on pros and cons of this approach.

In view of our toy example it should be possible to state and prove the relation between reversibility and gradient flow structure without making use of LDPs. In this paper we state the mentioned ‘meta-theorem’ as follows: assuming a given Markov process has a unique invariant measure  $\mu$ , if the generator of the process is symmetric in  $L^2_\mu$  (which is equivalent to the process being reversible, see section 2.4) then the associated FP operator is in gradient-flow form. This allows us to prove this statement without making explicit use of LDPs.

### 1.1. Main results

We are now in a position to start explaining the main contributions of this paper, section by section. After introducing setting and notation in section 2, in section 3 we compare HT and GENERIC—more precisely, we compare linear HT and so called Wasserstein pre-GENERIC, see sections 2 and 3 for precise statements and definitions. We show that if the KE associated with a given Markov process can be cast in the hypocoercive form (5) then the corresponding FP equation can be written in the form (6), and the orthogonality condition (7) holds as well. We emphasise that the orthogonality condition for the FP evolution is implied by the hypocoercive structure of the KE, without extra assumptions, see note 3.2 on this point. In particular, we provide explicit formulas to pass from one formulation to the other (i.e. given  $A$  and  $B$ , we give formulas to obtain  $W, M$  and  $\mathcal{S}$ , and viceversa). On a practical level, having such ‘conversion formulas’ at hand allows one to leverage results on the KE and potentially use them to produce results (almost for free) on the FP evolution, and viceversa. Since equations in GENERIC form dissipate relative entropy, a simple example of how one can produce results by exploiting such connection is the following: if the KE is in hypocoercive

form (5) then relative entropy is dissipated along the corresponding FP equation; see again note 3.2.

More importantly from a conceptual standpoint, the work in section 3 provides a simple connection between ‘non-reversibility’ and GENERIC structure; hence, by comparing the symmetric part of both formulations, it provides a simple proof of the fact that reversibility gives rise to gradient flow structure, i.e. of the meta-theorem we mentioned above, see note 3.2.

The results of section 3 only refer to Markov processes generated by *linear*<sup>5</sup> Markov operators, which are typically simple to cast in the form (5). As we have already pointed out, the form (5) is simply a symmetric–antisymmetric splitting (in  $L^2_\mu$ ), where the symmetric part is required to have a certain form, i.e. the form  $-A^*A$ . In section 4, we relax this assumption on the generator of the dynamics and produce analogous results to those in section 3, this time simply splitting the generator into its symmetric and antisymmetric part (again, symmetric and antisymmetric in  $L^2_\mu$ ) and assume that the symmetric part is linear, while the antisymmetric part is allowed to be non-linear. While the main result of section 4 is inspired by techniques used in large deviation theory, and in particular the approach of the proof is quite different from the one used in section 3, again no explicit use of LDP is made. On this matter of linear vs non-linear generators, we point out that GENERIC, in the way we present it in section 2 (which is a slight modification of the presentation in [33]), is perfectly well adapted to include both linear and nonlinear dynamics. The hypocoercive form (5) is instead typically only used for linear evolutions. Extensions of the HT to non-linear generators certainly exist, see e.g. [52, part III]. For purposes different from those of this paper it would be important to compare such non-linear formulations of HT with GENERIC; this will be subject of future work. For the time being we point out that, remarkably, the same orthogonality condition (7) which plays an important role in GENERIC, is also explicitly used in Villani’s memoir (see [52, remark 38]) to simplify verification of some of the (many) assumptions needed there.

Coming back to LDPs, in view of the fact that in the reversible case the gradient flow structure prescribes the form of the LDF (and viceversa), it is natural to ask whether a similar connection can be established in the non-reversible case. For this reason, in section 5, we consider generalised-reversible processes, which are a class of non-reversible processes, and we characterise the form of the LDF of generalised reversible processes. The only other result in this spirit which we know of is contained in [33], see also [8, 55, 56] and a recent preprint [62] for similar works in this direction for some other non-reversible processes (but these papers employ a different decomposition, namely decomposing fluxes/forces instead of the generator, and do not make connections to the HT).

Finally, in section 6 we show how the results of this paper can be applied to diffusion processes, section 6.1, and to a class of PDMPs, section 6.2, the so-called Hamiltonian Piecewise Deterministic Markov Chain Monte Carlo processes (Ham-PD-MCMC). We consider diffusion processes for purely expository purposes, to show in a simple setting how to use the results of this paper, but all the results of section 6.1 are known already, though perhaps not in the perspective in which we present them here. The PDMPs considered in section 6.2 are considerably less standard; while the HT for Ham-PD-MCMC has already been developed e.g. in [3], to the best of our knowledge GENERIC has not been applied to such processes yet. So, we first cast Ham-PD-MCMC in pre-GENERIC form and then we show that such dynamics constitute an example for which the orthogonality condition (7) does not hold, and yet one is able to show entropy decay.

<sup>5</sup> Here by *linear* we mean linear in the sense of McKean; i.e. the coefficients of the operator  $\mathcal{L}$  can be nonlinear in the state-space variable, as long as the action of the operator on functions is linear.



To summarise, the paper is organised as follows: in section 2 we first clarify our setting, notation and standing assumptions (section 2.1); we then give a concise exposition of the HT, gradient flows and GENERIC, in sections 2.2 and 2.3. We also point out that, while GENERIC is usually formulated on manifolds, we rephrase it here in a function space setting, mostly for ease of comparison with HT. Besides this small modification, the content of section 2 is well-known. In section 3 we investigate the relation between hypocoercive and GENERIC formulation of the dynamics in the case in which the generator  $\mathcal{L}$  is linear. The content of this section and of all subsequent sections (with the exception of section 6.1) is new, to the best of our knowledge. In section 4 we relax the class of generators for which it is possible to establish this connection and consider a class of non-linear operators. In section 5 we characterise the structure of LDFs of generalised-reversible Markov processes. Finally, in section 6 we apply the results of this paper to two classes of Markov processes, namely diffusion processes and PDMPs. Overall, because this paper makes use of the work of two very different communities, we have tried to make it as self-contained as possible.

## 2. Background: hypocoercivity and GENERIC

In this section we recall the main facts about the HT and GENERIC.

### 2.1. Setup, notation and preliminaries

Both the HT and GENERIC address the study of (stochastic) dynamics, more precisely of their associated Kolmogorov and/or FP PDEs. Unless otherwise stated, all the stochastic processes  $\{X_t\}_{t \geq 0}$  considered in this paper will be time-homogeneous Markov evolutions with finite dimensional state space; to fix ideas we will assume that the state space is  $\mathbb{R}^d$ . We recall that the semigroup  $P_t$  associated with the Markov process  $\{X_t\}_{t \geq 0}$  is defined on the set of functions  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  which are continuous and bounded by  $(P_t f)(x) := \mathbb{E}[f(X_t) | X_0 = x]$ . Given  $f$ ,  $(P_t f)(x)$  is a real-valued function of  $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ ; by Ito formula, such a function solves a differential equation of the form

$$\partial_t u(t, x) = \mathcal{L}u(t, x), \quad u(0, x) = f(x), \tag{9}$$

where  $\mathcal{L}$  is an appropriate operator (e.g. a second order differential operator in the case in which  $X_t$  is a diffusion process), called the Kolmogorov operator of the process  $X_t$ ; correspondingly, the differential equation (9) is called the KE associated with  $X_t$ . The dual equation, i.e. the equation

$$\partial_t \rho(t, x) = \mathcal{L}'\rho(t, x), \quad \rho(0, x) = \rho_0(x), \tag{10}$$

where  $\mathcal{L}'$  is the (formal)  $L^2$ -adjoint of  $\mathcal{L}$ , is referred to as the FP equation associated with  $X_t$  and it describes the evolution of the law of  $X_t$ . A measure  $\mu$  on  $\mathbb{R}^d$  is invariant for  $P_t$  (or, equivalently, invariant for  $X_t$ ) iff

$$\int_{\mathbb{R}^d} (P_t f)(x) \mu(x) dx = \int_{\mathbb{R}^d} f(x) \mu(x) dx \quad \text{for every } f \in C_b(\mathbb{R}^d). \tag{11}$$

There are various ways of specifying the domain of the (typically unbounded) operator  $\mathcal{L}$ . If there exists an invariant measure then the semigroup extends to a strongly continuous semigroup on  $L^2_\mu$ ; in this case  $\mathcal{L}$  is the generator of such a semigroup and we take as a domain

for  $\mathcal{L}$  the domain  $\mathcal{D}_2(\mathcal{L})$  of  $\mathcal{L}$  viewed as generator of  $\mathcal{P}_t$  in  $L^2_\mu$ , see [29, definition 1.5]<sup>6</sup>. One can show that (11) is equivalent to (see [29, section 1.2])  $\int (\mathcal{L}f)(x)\mu(x)dx = 0$  for every  $f \in \mathcal{D}_2(\mathcal{L})$ . In short, this can be written as

$$\mathcal{L}'\mu = 0. \tag{12}$$

For simplicity, we make the following **standing assumptions**, which will hold throughout the paper, unless otherwise specified:

**Hypothesis 2.1.** Standing assumptions:

- [SA.1] All the probability measures which we consider have a density with respect to the Lebesgue measure and, with abuse of notation, we denote the measure and its density with the same letter, i.e.  $\mu(dx) = \mu(x)dx$ .
- [SA.2] We only consider *ergodic* Markov processes, i.e. the underlying process  $X_t$  admits a unique invariant measure,  $\mu$ , which has a (smooth enough) density with respect to the Lebesgue measure.
- [SA.3] We assume that  $\mathcal{L}'$  and the initial datum  $\rho_0$  are such that the FP equation (10) admits strictly positive classical solutions, which additionally belong to  $L^2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ <sup>7</sup>.
- [SA.4] The operator  $\mathcal{L}$  is the generator of a strongly continuous contraction semigroup on a Hilbert space  $\mathcal{H}$  (which will typically be  $L^2_\mu$ ).

Let us motivate the above assumptions and point out explicitly what they imply.

- Under assumptions [SA.1] and [SA.2], equation (12) has a unique classical solution; without loss of generality, we can then assume  $\mu(x) > 0$  for every  $x \in \mathbb{R}^d$  and we shall do so throughout. This will allow us to consider the equation solved by the function  $h(x, t) = \rho(x, t)\mu(x)^{-1}$ , which we call the *modified Kolmogorov equation* (m-KE)—we will be more precise on the relation between KE, FP equation and m-KE in note 2.1 below. Let us also recall that if  $\mu$  is the unique invariant measure for the semigroup then it is an ergodic measure. If  $\mathcal{L}$  is linear and the semigroup is stochastically continuous [10, section 2.1] this implies that the kernel of the operator  $\mathcal{L}$  in  $L^2_\mu$  is made only of functions which are  $\mu$ -a.s. constant<sup>8</sup>.
- In general, the solution of the m-KE and of the FP equation do not lie in the same space; it is usually more natural to study the KE or the m-KE in the weighted space  $L^2_\mu := L^2_\mu(\mathbb{R}^d; \mathbb{R}) := \{f: \mathbb{R}^d \rightarrow \mathbb{R} \text{ s.t. } \int f^2(x)\mu(dx) < \infty\}$ , and the FP equation in  $L^2 := L^2(\mathbb{R}^d; \mathbb{R})$  (or in  $L^2$  weighted with an appropriate polynomial or, better yet, simply in measure-space, see for example [52, section 2.4]). For this reason we introduce two Hilbert spaces of

<sup>6</sup> We clarify that we are abusing notation by denoting by  $\mathcal{L}$  both the Kolmogorov operator and the generator of the semigroup, as these two objects only formally coincide; indeed the former is just a formal expression given by the Itô formula and most complications in this context arise precisely from trying to reconcile the two objects, i.e. from finding appropriate domains of definition see [5, chapters 1 and 3] for details.

<sup>7</sup> Note that typically FP type equations preserve positivity, see [7].

<sup>8</sup> The fact that  $\mathcal{L}\mathbf{1} = 0$  follows simply by the Hille-Yoside theorem. Viceversa, the ergodicity of the measure is equivalent (see [10, theorem 3.2.4]) to either one of the following statements: (a) if  $f \in L^2_\mu$  then  $P_t f = f \Rightarrow f = \text{constant}$ ,  $\mu$ -a.s.; (b) if  $f \in \mathcal{D}_2(\mathcal{L})$ , then  $\mathcal{L}f = 0 \Rightarrow f = \text{constant}$ ,  $\mu$ -a.s.. To see that statement (a) and (b) are equivalent use [29, equation (1.1.6)] and recall that the domain of the generator is dense in  $L^2_\mu$ .

real-valued functions,  $(\mathcal{H}, \|\cdot\|, \langle \cdot, \cdot \rangle)$  and  $(\mathcal{Z}, ((\cdot)), (\cdot, \cdot))$ , to be thought of as  $L^2_\mu$  and  $L^2$ , respectively. In particular, unless otherwise specified,  $(\cdot, \cdot)$  will denote  $L^2$  scalar product (or duality pairing, we will specify which in context, when needed). For the KE and the m-KE we do not necessarily restrict to considering classical solutions and, since we assume that  $\mathcal{L}$  generates a strongly continuous contraction semigroup in  $L^2_\mu$ , the semigroup itself is the solution in  $L^2_\mu$  of the KE.

- Technical matters about Kolmogorov and FP equations are often quite involved (see e.g. [5, chapter 3]) and are better treated on a case by case basis; GENERIC is unsurprisingly plagued by the same problems (see e.g. [18, remark 1.1] and references therein). In this paper we do not restrict to a particular class of Markov processes (e.g. we do not restrict to diffusion processes); on the contrary, we would like to include a broad range of Markov dynamics (see examples in section 6) and at the same time we want to avoid encumbering the exposition with excessive technicalities. For this reason, we assume the existence of an appropriate set  $\mathcal{V}$  which, unless otherwise stated, is dense both in  $\mathcal{D}_2(\mathcal{L}) \cap \mathcal{D}_2(\mathcal{L}^*)$  and in the space  $\mathcal{Cl} := \{ \text{positive classical solutions of the FP equation which live in } L^2 \cap L^1 \}$ . In practical examples  $\mathcal{V}$  can often be taken to be the set of (positive) Schwartz functions; this way we can perform all our calculations on  $\mathcal{V}$ . In this sense—i.e. in the sense that we only work on  $\mathcal{V}$ —some calculations in this section and the next will be somehow formal. We flag up now and we will come back to this (see note 2.3.1) that even doing this does not completely solve the problem for GENERIC and some remarks will be necessary. In the examples of section 6 we will give indications on the correct functional framework.

For any operator or functional, say  $\mathcal{T}$ ,  $\mathcal{D}(\mathcal{T})$  denotes a domain of definition of  $\mathcal{T}$ <sup>9</sup> and we will use interchangeably  $\mathcal{T}h$  and  $\mathcal{T}(h)$  to denote the action of  $\mathcal{T}$  on an element  $h \in \mathcal{D}(\mathcal{T})$ . Moreover, for any function  $f(t, x)$  depending on both time and space, the notations  $f_t(x)$  and  $f(t, x)$  will be used interchangeably. With this in mind we clarify the relation between the FP equation, the KE and the m-KE.

**Note 2.2 (relation between KE, m-KE and FP equation).** Suppose that the Kolmogorov operator  $\mathcal{L}$  in (9) is of the form

$$\mathcal{L} = B + \tilde{A},$$

where  $B$  and  $\tilde{A}$  are linear operators with  $B$  antisymmetric in  $L^2_\mu(\mathbb{R}^d)$  and  $\tilde{A}$  symmetric in  $L^2_\mu(\mathbb{R}^d)$ . We also assume that  $B$  enjoys the chain and product rule. Let  $\rho_t$  be the solution of the FP equation associated with the process,  $\partial_t \rho_t = \mathcal{L}' \rho_t$ , and set  $\rho_t = h_t \mu$ . Then the function  $h_t$  solves the m-KE, namely

$$\partial_t h_t = \mathfrak{L} h_t, \quad \text{where } \mathfrak{L} = (-B + \tilde{A}). \tag{13}$$

To see the above it suffices to show what follows:

$$\mathcal{L}'(h_t \mu) = \mu(\mathfrak{L} h_t). \tag{14}$$

Indeed if (14) holds, we then have

$$\partial_t \rho_t = \mu \partial_t h_t = \mathcal{L}' \rho_t = \mathcal{L}'(h_t \mu) = \mu \mathfrak{L} h_t,$$

<sup>9</sup> When we want to refer to a specific domain we will do so, see e.g. the difference between  $\mathcal{D}_2$  introduced earlier in this section and  $\mathcal{D}_b$ , used in section 5. Otherwise  $\mathcal{D}$  generically denotes a set on which the operator is well defined.

which directly implies (13). To show (14), notice that for every  $f, h \in \mathcal{V}$ , we have

$$\int B'(h\mu)(x)f(x) \, dx = \int h(x)\mu(x)(Bf)(x) \, dx = - \int f(x)(Bh)(x)\mu(x) \, dx,$$

and, similarly,

$$\int \tilde{A}'(h\mu)(x)f(x) \, dx = \int h(x)\mu(x)(\tilde{A}f)(x) \, dx = \int f(x)(\tilde{A}h)(x)\mu(x) \, dx,$$

which imply

$$B'(h\mu) = -\mu Bh \quad \text{and} \quad \tilde{A}'(h\mu) = \mu \tilde{A}h, \tag{15}$$

from which (14) follows, as  $\mathcal{L}' = B' + \tilde{A}'$ . Note that, from (14) one can formally see that in this setting  $\mu$  is a solution of  $\mathcal{L}'\mu = 0$  if and only if the kernel of  $\mathfrak{L}$  contains constants. In short,

$$\mathcal{L}'\mu = 0 \Leftrightarrow \mathfrak{L}\mathbf{1} = 0, \tag{16}$$

where  $\mathbf{1}$  denotes the function identically equal to one.

We recall that if  $\mathcal{S}$  is a functional on the Hilbert space  $(\mathcal{Z}(\cdot, \cdot), (\cdot, \cdot))$ ,  $\mathcal{S} : \mathcal{D}(\mathcal{S}) \subseteq \mathcal{Z} \rightarrow \mathbb{R}$ , the (directional) Gateaux derivative of  $\mathcal{S}$  at  $h$  in the direction  $g \in \mathcal{Z}$  is given by

$$(\mathbf{d}\mathcal{S}(h), g) = \left. \frac{d}{d\epsilon} \mathcal{S}(h + \epsilon g) \right|_{\epsilon=0}. \tag{17}$$

The Frechet derivative of  $\mathcal{S}$  at  $h$  is the linear functional  $\mathbf{d}\mathcal{S}(h)$  mapping  $g$  into the directional derivative of  $\mathcal{S}$  at  $h$  in the direction  $g$ . When  $\mathcal{S}$  is defined on the whole Hilbert space  $\mathcal{Z}$  then, for each fixed  $h$ , one can view  $\mathbf{d}\mathcal{S}(h)$  as an element of  $\mathcal{Z}$  and on the LHS of (17) the notation  $(\cdot, \cdot)$  is an actual scalar product. This is easier to see in finite dimensions. If  $\mathcal{Z}$  is finite dimensional, i.e. if  $\mathcal{Z} = \mathbb{R}^d$  and  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , then the Frechet derivative of  $f$  coincides with the differential and it will be denoted, as customary, by  $\nabla f$ . In this case, for each  $x \in \mathbb{R}^d$ ,  $\nabla f(x) \in \mathbb{R}^d$  while the map  $\nabla f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  may be viewed as an element of  $\mathcal{Z}^d := \{u = (u^1, \dots, u^d) : u^j \in \mathcal{Z} \text{ for all } j\}$ . However in practical examples, for our choice of the infinite dimensional space  $\mathcal{Z}$  (typically  $\mathcal{Z} = L^2$ ) it is almost never the case that  $\mathcal{S}$  is defined on the whole of  $\mathcal{Z}$  and, for fixed  $h \in \mathcal{Z}$ ,  $\mathbf{d}\mathcal{S}$  is not in  $\mathcal{Z}$  but in a bigger space. In this case the scalar product of  $L^2(\mathcal{Z})$  will only act as a formal duality pairing, see section 2.3.

We will use interchangeably the notation  $\nabla \cdot$  and  $\text{div}$  for the divergence of a  $\mathbb{R}^d$ -valued function. The Euclidean norm of  $\mathbb{R}^d$  will be denoted by  $|\cdot|$ .

### 2.2. Hypocoercivity

In this section we fix  $\mathcal{H} := L^2_\mu(\mathbb{R}^d)$ , where  $\mu$  is assumed to be the only invariant measure of the process with generator  $\mathcal{L}$ , FP operator  $\mathcal{L}'$  and modified Kolmogorow operator  $\mathfrak{L}$ . We say that a Markov evolution is in *linear hypocoercive form* if its generator  $\mathcal{L}$  is linear and can be written in the form

$$\mathcal{L} = B - A^*A, \tag{18}$$

where  $A : \mathcal{V} \rightarrow (L^2_\mu)^d$  and  $B : \mathcal{V} \rightarrow L^2_\mu$  are linear operators,  $A$  obeys the chain rule,  $B$  is anti-symmetric in  $L^2_\mu$ , i.e.  $B^* = -B$ , and obeys both chain and product rule. We recall that under our

standing assumptions  $\ker \mathcal{L} = \{\text{constants}\}$ . Because  $A^*$  denotes adjoint in  $L^2_\mu$ ,  $A^*A$  is necessarily symmetric in  $L^2_\mu$ .<sup>10</sup>

If  $\mathcal{L}$  is of the form (18) the m-KE (13) takes the form

$$\partial_t h_t = -(B + A^*A)h_t. \tag{19}$$

Let us look at the two parts of the dynamics: along the flow generated by  $-B$ , i.e. along the solution of the equation  $\partial_t f_t = -Bf_t$ , the norm is conserved; indeed, since  $B$  is antisymmetric in  $L^2_\mu$ ,  $\langle Bf, f \rangle = 0$  for every  $f \in \mathcal{V}$ , hence

$$\partial_t \|f_t\|^2 = -2\langle Bf_t, f_t \rangle = 0.$$

On the other hand, along the flow generated by  $-A^*A$ , the norm is dissipated:

$$\partial_t \|f_t\|^2 = -2\langle f_t, A^*A f_t \rangle = -2\|A f_t\|^2 \leq 0. \tag{20}$$

Overall, with similar calculations, the  $L^2_\mu$ - norm of  $h_t$  is dissipated:

$$\partial_t \|h_t\|^2 = -2\|A h_t\|^2 \leq 0. \tag{21}$$

For later comparisons with GENERIC we observe that this formalism does not imply the existence of any conserved quantity. Note that if the generator is in hypocoercive form then

$$\ker \mathcal{L} = \ker \mathfrak{L} = \ker A \cap \ker B, \tag{22}$$

see [52, proposition 2].

We gather in the next lemma some straightforward facts which will be useful later on.

**Lemma 2.3.** *With the notation introduced so far, let  $W : \mathcal{V} \rightarrow L^2_\mu$  be a linear operator such that  $W$  and  $W'$  satisfy the chain and product rule (e.g. a first order differential operator) and both  $W'$  and  $W^*$  are well-defined on  $\mathcal{V}$ .*

(a) *If  $W'\mu = 0$ ,<sup>11</sup> then*

$$W'h = W^*h = -Wh, \quad \text{for every } h \in \mathcal{V},$$

*i.e.  $W$  is antisymmetric both in  $L^2$  and in  $L^2_\mu$ ;*

(b) *If the generator is in hypocoercive form (18) and  $\ker \mathcal{L} = \{\text{constants}\}$  then  $B'\mu = 0$ . By using point i), if both  $B$  and  $B'$  satisfy chain and product rule, this implies  $B' = B^* = -B$  and hence also  $B\mu = 0$ .*

**Proof of lemma 2.3.** To prove (a), we first prove that  $W'h = W^*h$  for every  $h \in \mathcal{V}$ ; trivially, if  $W'\mu = 0$  then for every  $f, g \in \mathcal{V}$

$$\begin{aligned} \int_{\mathbb{R}^d} (Wf)(x)g(x)\mu(x)dx &= \int_{\mathbb{R}^d} f(x)(W^*g)(x)\mu(x)dx \\ &= \int_{\mathbb{R}^d} f(x)W'(g\mu)(x)dx = \int_{\mathbb{R}^d} f(x)(W'g)(x)\mu(x)dx. \end{aligned}$$

<sup>10</sup> We clarify that  $A$  should be thought of as a  $d$ -dimensional vector of operators  $A = (A_1, \dots, A_d)$ , e.g.  $A = \nabla = (\partial_{x_1}, \dots, \partial_{x_d})$ . We also adopt the same understanding as in [52], namely  $A^*A$  is a short notation for  $A^*A = \sum_1^d A_i^*A_i$ .

<sup>11</sup> Here we are implicitly assuming that  $\mu$  is smooth enough that the operator  $W'$  can be applied to  $\mu$ .

We now show  $W^* = -W$  (on  $\mathcal{V}$ ) by proving that  $\langle Wg, g \rangle = \int (Wg)(x)g(x)\mu(x)dx = 0$  for every  $g \in \mathcal{V}$ . Indeed:

$$\langle Wg, g \rangle = \frac{1}{2} \int_{\mathbb{R}^d} [W(g^2)](x)\mu(x)dx = \frac{1}{2} \int_{\mathbb{R}^d} g^2(x)(W'\mu)(x)dx = 0,$$

which concludes the proof of (a). As for (b), if  $\mathcal{L}\mathbf{1} = 0$ , by (22),  $B\mathbf{1} = 0$  as well, hence

$$0 = \int_{\mathbb{R}^d} (B\mathbf{1})\mu(x)g(x)dx = - \int_{\mathbb{R}^d} \mu(x)(Bg)(x)dx = - \int_{\mathbb{R}^d} (B'\mu)(x)g(x)dx, \quad \text{for every } g \in \mathcal{V},$$

where the first equality comes from using  $B^* = -B$ . □

An example which notoriously satisfies the assumptions of the above theorem is the Liouville operator  $B$  discussed in the introduction, which is indeed antisymmetric both in  $L^2$  and in  $L^2_\mu$ , with  $\mu$  the Boltzmann distribution.

**Note 2.4.** The aim of the HT [31, 52] is to study exponentially fast decay to equilibrium when the operator  $\mathcal{L}$  is not coercive in the  $L^2_\mu$  norm, but it is instead coercive in a modified norm. In the HT writing the splitting (18) is only a starting point, and for this reason we refer to equations driven by operators in the form (18) as being in linear hypocoercive form; however for the sake of clarity we emphasise that the fact that the operator  $\mathcal{L}$  is in the form (18) does not mean per se that it is hypocoercive (see [52, section 3] for a definition of hypocoercive operator) and hence that exponentially fast decay holds, as further quantitative assumptions on  $A$  and  $B$  are needed to that end. In this paper we compare primarily the structure of the involved equations, so we do not make any such quantitative assumptions. Understanding the meaning of such assumptions in the context of GENERIC will be the object of future work.

### 2.3. Gradient flows, GENERIC and pre-GENERIC

The classical formulation of GENERIC is on manifolds [33] (and this is perhaps a more insightful way of understanding the theory). Here, to compare more easily with the HT, we start by reformulating GENERIC and pre-GENERIC in a function-space setting and we work on  $\mathcal{V} \subseteq \mathcal{Z}$ , where again  $\mathcal{Z}$  is to be thought of  $L^2$ , although the scalar product of  $L^2$  will often only act as a formal duality pairing. We make remarks on why this is the case and on how this formulation compares with the classical one on manifolds throughout the paper, see in particular note 2.3.1 and note 5.2.

**2.3.1. Quadratic gradient flows.** Let  $\{\rho_t\}_{t \geq 0} \subset \mathcal{V}$  and  $\mathcal{S} : \mathcal{V} \rightarrow \mathbb{R}$ . An evolution equation of the form

$$\partial_t \rho_t = -M_{\rho_t} \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) \tag{23}$$

is a (quadratic) gradient flow if for each  $\rho \in \mathcal{V}$  s.t.  $\rho > 0$ ,  $M_\rho : \mathcal{D}(M_\rho) \supseteq \mathcal{V} \rightarrow \mathcal{V}$  is a symmetric and positive semidefinite operator, i.e.

$$(M_\rho(h), g) = (h, M_\rho(g)), \quad \text{and} \quad (M_\rho(g), g) \geq 0 \quad \forall h, g \in \mathcal{V}, \tag{24}$$

$\mathcal{S}$  is Frechet differentiable and  $\mathbf{d}\mathcal{S}(\rho) \in \mathcal{D}(M_\rho)$  for every  $\rho \in \mathcal{V}$ . If  $M_\rho(\cdot) = -2 \operatorname{div}(\rho \nabla \cdot)$ , then (23) is a Wasserstein gradient flow, see [1]. If we can write  $M_\rho$  as

$$M_\rho(\cdot) = 2A'(\rho A(\cdot)) \tag{25}$$

for some operator  $A$  and  $L^2$ -dual  $A'$ , then we say that (23) is a *generalised Wasserstein gradient flow*. Note that the operator  $M_\rho$  in (25) is symmetric and positive definite (we show this explicitly after (47)); moreover, when  $A = \nabla$  we recover  $M_\rho(\cdot) = -2 \operatorname{div}(\rho \nabla \cdot)$ .

As a consequence of the gradient-flow structure, the *entropy*  $\mathcal{S}$  is dissipated along the solution of (23):

$$\frac{d\mathcal{S}(\rho_t)}{dt} = (\mathbf{d}\mathcal{S}(\rho_t), \partial_t \rho_t) = - \left( \mathbf{d}\mathcal{S}(\rho_t), M_{\rho_t} \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) \right) \leq 0, \tag{26}$$

having used the positivity of  $M_\rho$ . We emphasise the analogy between the above calculation and the one in (20). Finally, there is no natural conserved quantity associated with the flow (23).

**Note 2.5.** If  $\mathcal{Z}$  was a general manifold (as in [33, 38]) instead of a Hilbert space and  $\mathcal{S} : \mathcal{Z} \rightarrow \mathbb{R}$  then, by definition (17) of variational derivative,  $\mathbf{d}\mathcal{S}(\rho) \in T_\rho^* \mathcal{Z}$ , where  $T_\rho^* \mathcal{Z}$  is the cotangent space of  $\mathcal{Z}$  at  $\rho$ , and  $M_\rho : T_\rho^* \mathcal{Z} \rightarrow T_\rho \mathcal{Z}$ , where  $T_\rho \mathcal{Z}$  is the tangent space of  $\mathcal{Z}$  at  $\rho$ , for every  $\rho \in \mathcal{Z}$ . (Note that any Hilbert space  $\mathcal{Z}$  is a Hilbert manifold and both  $T_\rho^* \mathcal{Z}$  and  $T_\rho \mathcal{Z}$  are isomorphic to  $\mathcal{Z}$ .) This approach is somehow neater and, as we said, possibly more insightful. The difficulty here becomes to first identify the correct manifold  $\mathcal{Z}$  and then describe explicitly  $T\mathcal{Z}$  and  $T^* \mathcal{Z}$ . This is not straightforward when e.g.  $\mathcal{Z} = \{ \text{finite positive Borel probability measures} \}$ , see [38, section 3]. Further issues then arise when introducing dissipation potentials, as we will do in the next subsection on gradient flows, as such functionals are then typically not defined on the whole tangent and cotangent space. To avoid these complications we simply work on a very good set, where everything is well posed. However also this approach needs some extra comments, as it is clear from the example we gave in the Introduction, where  $(\mathbf{d}\mathcal{S}(\rho))(x, v) = \log \rho + V(x) + 1 + v^2/2$ , that the function  $\mathbf{d}\mathcal{S}(\rho)$  does not belong to  $L^2$  but for example it belongs to  $L^2_\mu$  instead. In that example all would work by choosing  $\mathcal{D}(M_\rho)$  to be the set of smooth functions in  $L^2_\mu$ . Hence in (24) or in (26) the scalar product of  $L^2$  acts only as a formal pairing, in the sense that, upon making a good choice of  $\mathcal{D}(M_\rho)$ , all the integrations implied by the formal  $L^2$  scalar product make sense and this is how we will understand them throughout (including in the expressions (27) and (28) below).

**2.3.2. Gradient flows.** To introduce a more general (non-quadratic) notion of gradient flow, let us first give the definition of *dissipation potential*. For every  $\rho \in \mathcal{V}$  let  $\psi(\rho; \cdot) : \mathcal{D}(\psi_\rho) \subseteq \mathcal{V} \rightarrow \mathbb{R}$  and let  $\psi^*(\rho; \cdot) : \mathcal{D}(\psi^*_\rho) \supseteq \mathcal{V} \rightarrow \mathbb{R}$  denote its Legendre transform, namely

$$\psi^*(\rho; \xi) = \sup_v \{ (v, \xi) - \psi(\rho; v) \}. \tag{27}$$

We recall that, given a convex<sup>12</sup>, lower semicontinuous function  $F$ , not taking the value  $-\infty$ , we have  $F^{**} = F$ , see [22, propositions 3.1 and 4.1] for details. Under such assumptions on  $\psi$ , we have

$$\psi(\rho; v) = \sup_\xi \{ (v, \xi) - \psi^*(\rho; \xi) \}, \tag{28}$$

i.e.  $\psi^{**} = \psi$  (at least on a good enough set of functions). We will use interchangeably the notation  $\psi(\rho; v)$  and  $\psi_\rho(v)$  (same for  $\psi^*$ ). We say that the pair  $\psi, \psi^*$  is a pair of *dissipation potentials* if both  $\psi$  and  $\psi^*$  are strictly convex, continuously differentiable<sup>13</sup> and

$$\psi(\rho; 0) = \psi^*(\rho; 0) = 0, \quad \text{for all } \rho \in \mathcal{V}. \tag{29}$$

<sup>12</sup> We say that a function  $F$  is convex if  $F(\lambda \xi_1 + (1 - \lambda) \xi_2) \leq \lambda F(\xi_1) + (1 - \lambda) F(\xi_2)$  for all  $\lambda \in [0, 1]$ .

<sup>13</sup> This assumptions can be relaxed. If  $\psi^*$  is not differentiable, we can substitute its derivative by its convex subdifferential, see [37, section 2.5].

The functions  $\psi$  and  $\psi^*$  are symmetric if  $\psi(\rho; v) = \psi(\rho; -v)$  and  $\psi^*(\rho; \xi) = \psi^*(\rho; -\xi)$ , for every  $\rho, v, \xi$ . We recall that, by the definition of Legendre transform, i.e. by using (27) and (28),

$$\psi(\rho; 0) = 0 \Leftrightarrow \inf \psi^*(\rho; \cdot) = 0 \Rightarrow \psi^*(\rho; \cdot) \geq 0, \tag{30}$$

$$\psi^*(\rho; 0) = 0 \Leftrightarrow \inf \psi(\rho; \cdot) = 0 \Rightarrow \psi(\rho; \cdot) \geq 0. \tag{31}$$

A gradient flow with respect to a functional  $\mathcal{S} : \mathcal{V} \rightarrow \mathbb{R}$  and dissipation potentials  $\psi, \psi^*$  is an evolution equation of the form

$$\partial_t \rho_t = \mathbf{d}_\xi \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right). \tag{32}$$

Equivalently, one can show (see [33, section 2.1]) that an evolution equation  $\partial_t \rho_t = \mathcal{G}(\rho_t)$  (where  $\mathcal{G} : \mathcal{D}(\mathcal{G}) \supseteq \mathcal{V} \rightarrow \mathcal{V}$  is an operator) is a gradient flow with respect to  $\mathcal{S}, \psi, \psi^*$  iff

$$\psi(\rho_t; \mathcal{G}(\rho_t)) + \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) + \frac{1}{2} (\mathcal{G}(\rho_t), \mathbf{d}\mathcal{S}(\rho_t)) = 0. \tag{33}$$

To clarify the notation in the above, on the RHS of (32)  $\mathbf{d}_\xi$  is the Frechet derivative of  $\psi^*(\rho; \xi)$  with respect to its second argument (then calculated at  $-\mathbf{d}\mathcal{S}(\rho)/2$ ). As we will always regard  $\psi^*(\rho; \xi)$  as a function of  $\xi$  for every  $\rho$  fixed, we could have dropped the subscript  $\xi$ , which is there both for clarity and to keep closer to tradition.

If  $\mathcal{Z} = \mathbb{R}^d$  (which, strictly speaking, is not allowed by our setup, as we assumed that  $\mathcal{Z}$  is a space of real-valued functions) and  $\psi^*_\rho(\xi) = M\xi \cdot \xi$ , where  $M$  is a symmetric positive definite matrix, then (32) boils down to (23), hence the name *quadratic* gradient flow for (23).

A calculation analogous to (26) – this time exploiting the convexity of  $\psi^*$ <sup>14</sup> rather than the positivity of  $M$ —allows again to show that the entropy  $\mathcal{S}$  is dissipated along the flow.

**2.3.3. Pre-GENERIC.** Let  $W : \mathcal{D}(W) \supseteq \mathcal{V} \rightarrow \mathcal{V}$  be an operator and  $\mathcal{S}$  and  $\psi^*$  as above. Then an evolution equation of the form

$$\partial_t \rho_t = W(\rho_t) + \mathbf{d}_\xi \psi^*_{\rho_t} \left( -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) \tag{34}$$

is said to be a *pre-GENERIC* flow (with respect to  $W, \mathcal{S}, \psi, \psi^*$ ) if the following degeneracy condition is satisfied:

$$(W(\rho), \mathbf{d}\mathcal{S}(\rho)) = 0 \quad \text{for all } \rho \in \mathcal{V}. \tag{35}$$

Equivalently (see appendix B) the evolution equation  $\partial_t \rho_t = \mathcal{G}(\rho_t)$  is a *pre-GENERIC* flow (with respect to  $W, \mathcal{S}, \psi, \psi^*$ ) iff (35) holds and

$$\psi(\rho_t; \mathcal{G}(\rho_t) - W(\rho_t)) + \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) + \frac{1}{2} (\mathcal{G}(\rho_t), \mathbf{d}\mathcal{S}(\rho_t)) = 0. \tag{36}$$

**Note 2.6.** Some comments on the above definition

(a) If the gradient flow part of (34) is quadratic, i.e. if the evolution is given by

$$\partial_t \rho_t = W(\rho_t) - M_{\rho_t} \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right), \tag{37}$$

<sup>14</sup> We recall that if  $F$  is convex and continuously differentiable then  $(x - y)(\mathbf{d}F(x) - \mathbf{d}F(y)) \geq 0$  for every  $x, y$ .



still appended with the condition (35), then we talk about *quadratic pre-GENERIC*; if the flow is quadratic pre-generic with  $M_\rho$  given by (25), then we talk about *generalised Wasserstein pre-GENERIC*.

- (b) If (34) and (35) hold, then  $\mathcal{S}$  decays along the flow. However, if the purpose is to have entropy dissipation, then (35) can be replaced with the following dissipativity condition

$$(W(\rho), \mathbf{d}\mathcal{S}(\rho)) \leq 0 \quad \text{for all } \rho \in \mathcal{V}. \tag{38}$$

Also from the point of view of the equation structure, one can dispense with the orthogonality condition (35) at the price of modifying (36). Indeed one can see (see appendix B) that the flow  $\partial_t \rho_t = \mathcal{G}(\rho_t)$  is of the form (34) if and only if

$$\psi(\rho_t; \mathcal{G}(\rho_t) - W(\rho_t)) + \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) + \frac{1}{2} (\mathcal{G}(\rho_t) - W(\rho_t), \mathbf{d}\mathcal{S}(\rho_t)) = 0. \tag{39}$$

However entropy decay does not necessarily hold if (39) holds in place of (36)–(35). If (39) is appended with the dissipativity condition (38) then the entropy functional does decrease along the flow. Indeed, using the non-negativity of  $\psi$  and  $\psi^*$  (see (30) and (31)), (39) and (38), one has

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \mathcal{S}(\rho_t) &= \frac{1}{2} (\dot{\rho}_t, \mathbf{d}\mathcal{S}(\rho_t)) = \frac{1}{2} (\mathcal{G}(\rho_t), \mathbf{d}\mathcal{S}(\rho_t)) \\ &= -\psi(\rho_t, \mathcal{G}(\rho_t) - W(\rho_t)) - \psi^* \left( \rho_t, -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) + \left( W(\rho_t), \frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) \leq 0. \end{aligned}$$

Hence  $\mathcal{S}$  is a Lyapunov functional of the dynamics.

- (c) Since  $\psi^*$  is assumed to be symmetric and  $\psi$  is its Legendre dual, then also  $\psi$  is symmetric.

#### 2.4. The many meanings of the word ‘reversibility’

It is rather unfortunate that the word ‘reversible’ is used by different communities with substantially opposite meanings. This misunderstanding becomes particularly confusing in the context of this paper, so we make some clarifications but refer the reader to [36, 45] for complete statements and proofs. A continuous-time stochastic process  $\{X_t\}_{t \geq 0}$  is called *time-reversible* or simply (*microscopically*) *reversible* if its law is invariant under time-reversal, i.e. if for every fixed  $T > 0$  the process  $\{X_t\}$  and the process  $\{X_{T-t}\}$  have the same distributions on the space of continuous paths. Intuitively, this can be interpreted as follows: the process is time-reversible if by watching a movie of the process run forwards and then backwards, we would not be able to distinguish the two. The process is stationary if for every  $\tau \in \mathbb{R}$  the process  $\{X_t\}$  and the process  $\{X_{t+\tau}\}$  have the same finite dimensional distributions. A time-reversible process is stationary but the converse is not true in general. Importantly to our purposes, a stationary diffusion process with invariant measure  $\mu$  is reversible if and only if its generator is self-adjoint in  $L^2_\mu$  (see [45, theorem 4.5]) i.e. iff

$$\int_{\mathbb{R}^d} f(x) (\mathcal{L}g)(x) d\mu(x) = \int_{\mathbb{R}^d} g(x) (\mathcal{L}f)(x) d\mu(x), \quad \text{for every } f, g, \in \mathcal{D}_2(\mathcal{L}). \tag{40}$$

When the above holds we also say that  $\mu$  satisfies *detailed balance* with respect to  $\mathcal{L}$ . If  $\mathcal{L}$  generates a strongly continuous contraction semigroup, by the Lumer–Phillips theorem [46, section 1.4] it is negative as well hence, with a calculation completely analogous to the one

leading to (21), one can see that reversible dynamics are dissipative. However in the acronym GENERIC-general equations for non-equilibrium reversible–irreversible coupling—the term ‘reversible’ refers to a different property, enjoyed in particular by Hamiltonian dynamics, which are certainly non-reversible according to the definition we have given above. Indeed, Hamiltonian dynamics are associated with the antisymmetric part of the evolution and hence they are conservative. A simple example can be given by considering planar rotations (which are Hamiltonian dynamics), for which one can distinguish whether the process is running forwards or backward in time by looking at the verse of rotation. In the acronym GENERIC the term ‘reversible’ refers to flows  $\phi_t$  on  $\mathbb{R}^{2d}$ ,  $\phi_t : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d}$ , which enjoy the following property:

$$(\phi_t)^{-1}(x, v) = \phi_{-t}(x, v) = (F \circ \phi_t \circ F)(x, v), \quad \text{where } F(x, v) = (x, -v),$$

while the word ‘irreversible’ alludes to the *macroscopic* irreversibility of the dynamics, i.e. to dissipativity. We will avoid using the word ‘reversible’ at all, but when we do we will mean it in the sense of microscopic time-reversibility.

### 3. Relating linear hypocoercivity and Wasserstein pre-GENERIC

#### 3.1. From linear hypocoercivity to Wasserstein pre-GENERIC

Suppose the m-KE is in linear hypocoercive form (19), with both  $B$  and  $B'$  obeying chain and product rule. In this section we show that, if this is the case, then the FP equation for  $\rho_t = h_t \mu$  is in Wasserstein pre-GENERIC form (37) with entropy functional  $\mathcal{S}(\rho)$  given by the *relative entropy*  $\mathcal{S}_\mu(\rho)$  of  $\rho$  with respect to  $\mu$ , i.e.

$$\mathcal{S}(\rho) = \mathcal{S}_\mu(\rho) := \int_{\mathbb{R}^d} \log \left( \frac{\rho(x)}{\mu(x)} \right) \rho(x) dx, \tag{41}$$

and

$$M_\rho(\xi) = 2A'(\rho A \xi), \quad W(\rho) = -B\rho, \tag{42}$$

where the operators  $A$  and  $B$  in the above are as prescribed by the hypocoercive structure (19). Indeed, if the m-KE is in the form (19) then

$$\partial_t \rho_t = \mu \partial_t h_t = -\mu(B + A^* A)h_t = -\mu B(\rho_t/\mu) - \mu A^* A(\rho_t/\mu). \tag{43}$$

For the first term in (43), using the fact that the kernel of  $\mathcal{L}$  is made of constants and the first equality in (15), we get

$$-\mu B(\rho_t/\mu) = -\mu B h_t = B'(h_t \mu) = -B\rho_t,$$

hence

$$\mu B(\rho_t/\mu) = B\rho_t. \tag{44}$$

For the second term in (43), we can see that for every  $f \in \mathcal{V}$  one has

$$\begin{aligned} \int_{\mathbb{R}^d} \mu(x) [A^* A(\rho/\mu)](x) f(x) dx &= \int_{\mathbb{R}^d} [A(\rho/\mu)](x) (Af)(x) \mu(x) dx \\ &= \int_{\mathbb{R}^d} [A'(A(\rho/\mu)\mu)](x) f(x) dx; \end{aligned} \tag{45}$$

using the fact that  $A$  satisfies the chain rule, we can further express the inner integral in (45) as follows

$$A(\rho/\mu) = (\rho/\mu)A(\log(\rho/\mu)).$$

Substituting the above back into (45) we then get

$$\int_{\mathbb{R}^d} \mu(x)[A^*A(\rho/\mu)](x)f(x)dx = \int_{\mathbb{R}^d} A'[\rho A(\log(\rho/\mu))](x)f(x)dx, \quad \text{for every } f \in \mathcal{V}.$$

Now notice that if  $\mathcal{S}$  is as in (41) then  $\mathbf{d}\mathcal{S}(\rho) = \log(\rho/\mu) + 1$ . Again by using the fact that  $\ker \mathcal{L} = \{\text{constants}\}$  and (22), we have  $A\mathbf{1} = 0$ , hence  $A(\mathbf{d}\mathcal{S}(\rho)) = A(\log(\rho/\mu))$ , from which we deduce

$$\mu A^*A(\rho/\mu) = A'(\rho A \log(\rho/\mu)) = A'[\rho A(\mathbf{d}\mathcal{S}(\rho))] = M_\rho \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho) \right). \quad (46)$$

Putting together (43), (44) and (46), we have therefore obtained

$$\begin{aligned} \partial_t \rho_t &= \mathcal{L}' \rho_t = -B\rho_t - A'(\rho A(\mathbf{d}\mathcal{S}(\rho))) \\ &= W\rho_t - M_{\rho_t} \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right). \end{aligned}$$

To show that the above is a pre-GENERIC system we still need to verify that the degeneracy condition (35) holds; to this end, using the chain rule for  $B$ ,  $B\mathbf{1} = 0$  and  $B^* = -B = B'$  (which follows from lemma 2.3), we have

$$\begin{aligned} \int_{\mathbb{R}^d} (B\rho)(x)[\mathbf{d}\mathcal{S}(\rho)](x)dx &= \int_{\mathbb{R}^d} (B\rho)(x)[\log(\rho(x)/\mu(x)) + 1]dx \\ &= - \int_{\mathbb{R}^d} \mu(x)[B(\rho/\mu)](x)dx - \int_{\mathbb{R}^d} \rho(x)B\mathbf{1}dx \\ &= - \int_{\mathbb{R}^d} (B^*\mu)(x)\rho(x)/\mu(x)dx = \int_{\mathbb{R}^d} (B\mu)(x)\rho(x)/\mu(x)dx = 0. \end{aligned} \quad (47)$$

It remains to check that  $M$  is symmetric and positive semidefinite. In fact,

$$(M_\rho \xi, \xi) = 2(A'(\rho A \xi), \xi) = 2 \int_{\mathbb{R}^d} \rho(x)[(A\xi)^2(x)]dx \geq 0, \quad \text{for every } \xi \in \mathcal{V},$$

as  $\rho > 0$  (standing assumption [SA.3]). Similarly, for every  $\xi_1, \xi_2 \in \mathcal{V}$ , we have

$$(M_\rho \xi_1, \xi_2) = 2 \int_{\mathbb{R}^d} [A'(\rho A \xi_1)](x)\xi_2(x)dx = 2 \int_{\mathbb{R}^d} \rho(x)(A\xi_1)(x)(A\xi_2)(x)dx = (M_\rho \xi_2, \xi_1).$$

### 3.2. From Wasserstein pre-GENERIC to linear hypocoercivity

Recalling that  $\mu$  is the unique invariant measure of the underlying process  $X_t$ , we now assume that the FP equation is in pre-GENERIC form (35)–(37), with  $\mathcal{S}(\rho) = \mathcal{S}_\mu(\rho)$  as in (41) and  $W$  is an operator on  $\mathcal{V}$  such that both  $W$  and  $W'$  are defined on  $\mathcal{V}$  and obey the chain and product rule; because  $W'$  obeys the product rule, the kernel of  $W'$  contains constants, i.e.  $W'\mathbf{1} = 0$ . Assuming that that FP equation is of the form (37) means that the FP operator  $\mathcal{L}'$  is precisely given by

$$\mathcal{L}' \rho = W\rho - M_\rho \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho) \right).$$

With this premise, we will show that the dual of  $\mathcal{L}'$ , i.e. the Kolmogorov operator  $\mathcal{L}$ ,<sup>15</sup> can be written in linear hypocoercive form  $\mathcal{L} = B - A^*A$ . In particular we will show that  $W'$  is antisymmetric in  $L^2_\mu$  and it is related to the operator  $B$  appearing in the hypocoercive form by  $W' = B$  while  $M_\rho$  is related to  $A$  through the relations (49) and (50) below.

We start by showing that  $W'$  is antisymmetric. To this end, from the degeneracy condition (35) we have

$$0 = (W(\rho), \mathbf{dS}(\rho)) = \int_{\mathbb{R}^d} \rho(x) [W'(\log(\rho/\mu))](x) dx + \int_{\mathbb{R}^d} \rho(x) (W'\mathbf{1}) dx = \int_{\mathbb{R}^d} \frac{\rho(x)}{\mu(x)} (W\mu)(x) dx; \tag{48}$$

since the degeneracy condition (35) holds for every  $\rho \in \mathcal{V}$ , this implies  $W\mu = 0$ . Using lemma 2.3 we conclude that  $B = W'$  is antisymmetric in  $L^2_\mu$ . Let us notice that by setting  $B = W'$  we also have  $B\mathbf{1} = 0$ .

Regarding the gradient-flow part, because  $M_\rho$  is symmetric and positive definite for every  $\rho > 0$  fixed, one can find a square root of  $M_\rho$ , i.e. an operator  $M_\rho^{1/2}$  such that  $(M_\rho^{1/2})' M_\rho^{1/2} = M_\rho$ ; guided by the calculation in the previous subsection, we assume that one such square root<sup>16</sup> can be written in the form

$$M_\rho^{1/2}(\xi) := \sqrt{2\rho}A(\xi), \tag{49}$$

for some operator  $A$  which does not depend on  $\rho$  (let us point out that any Itô diffusion belongs to this category, see section 6.1, and that expressions analogous to the above also appear in [50]). With this choice, we have  $(M_\rho^{1/2})'(\xi) = \sqrt{2}A'(\sqrt{\rho}\xi)$ , so that

$$M_\rho(\xi) = (M_\rho^{1/2})'(M_\rho^{1/2}(\xi)) = 2A'(\rho A(\xi)). \tag{50}$$

If  $A$  obeys the chain rule then, with calculations completely analogous to those in the previous subsection, one finds that if  $\rho_t$  solves (37) and all of the above assumptions are satisfied, then the m-KE for  $h_t = \rho_t/\mu$  can be written in the form (19).

To conclude, recall that the fact that the kernel of  $\mathcal{L}$  is only made of constants is a consequence of ergodicity, see section 2.1. However, in this setup this can anyway be recovered from the structure of the equations. Indeed, because  $W\mu = 0$  and  $\mu$  is an invariant measure, i.e. a stationary solution of (37), it follows that  $M_\mu(\mathbf{dS}(\mu)) = 0$ ; if this is the case then, since  $\mathbf{dS}(\mu) = 1$ , for every  $g \in \mathcal{V}$  one has

$$0 = (M_\mu(\mathbf{dS}(\mu)), g) = (M_\mu^{1/2}(\mathbf{dS}(\mu)), M_\mu^{1/2}(g)) = \int (A\mathbf{1})(Ag)\mu = \int (A^*A\mathbf{1})g\mu,$$

which implies  $\mathbf{1} \in \ker(A^*A)$ . Because  $\ker(A^*A) = \ker A$ , one has  $A\mathbf{1} = 0$ . This, together with the fact that  $B\mathbf{1} = W'\mathbf{1} = 0$  implies  $\mathcal{L}\mathbf{1} = \mathfrak{L}\mathbf{1} = 0$ .

**Note 3.1.** A couple of observations on sections 3.1 and 3.2.

- The calculation in (47) shows that, in the framework of section 3.1, i.e. if the generator  $\mathcal{L}$  is in linear hypocoercive form, the degeneracy condition (35) is ultimately a consequence of the antisymmetry in  $L^2_\mu$  of  $B$  and of  $B\mathbf{1} = 0$ , which implies  $B\mu = 0$  (by lemma 2.3). Viceversa, in section 3.2 the degeneracy condition implies  $W\mu = 0$  (hence  $B\mu = 0$  as well), but it does not seem to imply  $W'\mathbf{1} = B\mathbf{1} = 0$  directly.

<sup>15</sup> We are implicitly assuming that  $(\mathcal{L}')' = \mathcal{L}$ , at least on  $\mathcal{V}$ , similarly for  $B$  and  $W$ .

<sup>16</sup> As is well known, the square root operator is not unique.

- On a practical level one can see the content of these subsections as giving explicit formulas to go from linear hypocoercivity to pre-GENERIC formulation and viceversa. The potential advantage is that one could obtain some results almost for free; for example, suppose a particular equation is known to have hypocoercive form. Using section 3.1 one can then write it in pre-GENERIC form and hence obtain relative entropy dissipation. An indeed a **Corollary** of the content of section 3.1 is the following: if the generator  $\mathcal{L}$  is in linear hypocoercive form, then the relative entropy is dissipated along the flow of the associated FP equation.
- When  $B = 0$ , a straightforward consequence of section 3.1 is the following: if the generator of the process is self-adjoint in  $L^2_\mu$ , which is equivalent to the underlying process being reversible, the FP operator is in gradient flow form. In this sense the content of section 3.1 can be seen as providing a simple proof of the link between reversibility and gradient flow structure.

#### 4. Relation between pre-GENERIC and symmetric–antisymmetric decomposition

In section 3 we worked with linear Markov generators in hypocoercive form (18). In this section we slightly relax the class of Markov generators we consider; we work again under the standing assumptions of hypothesis 2.1 but this time we simply split the generator  $\mathcal{L}$  as follows

$$\mathcal{L} = \mathcal{L}_a + \mathcal{L}_s, \tag{51}$$

where  $\mathcal{L}_s$  and  $\mathcal{L}_a$  are, respectively, the symmetric and anti-symmetric parts in  $L^2_\mu$  of  $\mathcal{L}$ . We will assume that  $\mathcal{L}_s$  is linear, but no such assumption will be needed for  $\mathcal{L}_a$ . Moreover, we do not explicitly assume that  $\mathcal{L}_s$  should be of the form  $-A^*A$ , as in section 3.17. With the splitting (51) in mind, for every  $\rho \in \mathcal{V}$  we consider the following Hamiltonian functional  $\mathcal{H}(\rho; \cdot) : \mathcal{D}(\mathcal{H}_\rho) \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \mathcal{H}(\rho; \xi) &:= \int_{\mathbb{R}^d} e^{-\xi(x)} (\mathcal{L}e^\xi)(x)\rho(x)dx \\ &= \int_{\mathbb{R}^d} e^{-\xi(x)} [(\mathcal{L}_a + \mathcal{L}_s)e^\xi](x)\rho(x)dx =: \mathcal{H}_a(\rho; \xi) + \mathcal{H}_s(\rho; \xi), \end{aligned} \tag{52}$$

where, for every  $\rho \in \mathcal{V}$ ,  $\mathcal{H}_s(\rho; \cdot) : \mathcal{D}(\mathcal{H}_{s,\rho}) \rightarrow \mathbb{R}$ ,  $\mathcal{H}_a(\rho; \cdot) : \mathcal{D}(\mathcal{H}_{a,\rho}) \rightarrow \mathbb{R}$ ,

$$\mathcal{H}_a(\rho; \xi) := \int_{\mathbb{R}^d} e^{-\xi} (\mathcal{L}_a e^\xi) d\rho, \quad \mathcal{H}_s(\rho; \xi) := \int_{\mathbb{R}^d} e^{-\xi} (\mathcal{L}_s e^\xi) d\rho, \tag{53}$$

and  $\mathcal{D}(\mathcal{H}_\rho)$  is the set of functions  $\xi$  such the integrand in (52) makes sense and the integral in (52) is finite (similarly for  $\mathcal{D}(\mathcal{H}_{s,\rho}), \mathcal{D}(\mathcal{H}_{a,\rho})$ ). We will also make the following technical assumption.

**Assumption 4.1** For every  $\rho \in \mathcal{V}$ ,  $\mathcal{D}(\mathcal{H}_{s,\rho})$  is such that for any  $\xi, \bar{\xi}, \tilde{\xi} \in \mathcal{D}(\mathcal{H}_{s,\rho})$ ,  $\xi + \gamma \bar{\xi} \in \mathcal{D}(\mathcal{H}_{s,\rho})$  for any  $\gamma \in (0, 1)$  and the integral

<sup>17</sup> We note that for less than regularity issues related to the domain of the adjoint operator and for the potential practical difficulty of finding  $A$ , any non-negative symmetric operator can be written in the form  $-A^*A$ . But this might not be easy in practice and indeed this is inconvenient for the PDMPs we study in section 6.2.

$$\int_{\mathbb{R}^d} e^{-\xi(x)} \alpha \xi^p(x) \left[ \mathcal{L}_s(\beta \bar{\xi}^p e^{\bar{\xi}}) \right](x) \rho(x) dx$$

is finite, for any  $p \in \{1, 2\}, \alpha, \beta \in \{0, 1\}$ .

The above assumption is easily satisfied in practice, as in examples  $\mathcal{D}(\mathcal{H}_{s,\rho})$  will typically contain functions that grow polynomially while  $\rho$  decays exponentially fast at infinity. Finally, before stating our main theorem, we recall (see (36)) that the FP equation  $\dot{\rho}_t = \mathcal{L}'(\rho_t)$  is in pre-GENERIC form with respect to  $(\mathcal{S}, \psi, \psi^*, W)$  if and only if

$$\psi(\rho_t; \mathcal{L}'(\rho_t) - W(\rho_t)) + \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho_t) \right) + \frac{1}{2} (\mathcal{L}'(\rho_t), \mathbf{d}\mathcal{S}(\rho_t)) = 0 \quad (54)$$

and the orthogonality condition (35) holds.

With this in mind, the purpose of theorem 4.1 below is to show that if the symmetric–antisymmetric decomposition (51) holds for the generator then the FP equation can be written in pre-GENERIC form, and vice versa. The theorem also provides a constructive way to pass from one formulation to the other.

**Theorem 4.1.** *With the notation introduced so far, the following holds:*

- **From symmetric–antisymmetric decomposition of the generator to pre-GENERIC structure of the FP equation.** *Let  $\mathcal{L}$  be the generator of a time-homogeneous Markov process with unique invariant measure  $\mu$  and consider a symmetric–antisymmetric splitting of  $\mathcal{L}$  as in (51) where we assume that  $\mathcal{L}_s$  is a linear operator. Given  $\mathcal{L}$ , we consider  $\mathcal{H}_s$  as in (53) and assume that  $\mathcal{H}_s(\rho; \cdot)$  is convex and once Frechet differentiable in the second argument, for every  $\rho$ . For every  $\rho \in \mathcal{V}$  we then let  $\Psi^*(\rho; \cdot) : \mathcal{D}(\mathcal{H}_{s,\rho}) \rightarrow \mathbb{R}$  be the functional*

$$\Psi^*(\rho; \xi) = \mathcal{H}_s \left( \rho; \frac{1}{2} \mathbf{d}\mathcal{S}_\mu(\rho) + \xi \right) - \mathcal{H}_s \left( \rho; \frac{1}{2} \mathbf{d}\mathcal{S}_\mu(\rho) \right), \quad (55)$$

where  $\mathcal{S}_\mu$  is the relative entropy between  $\rho$  and  $\mu$ . Suppose  $\mathcal{L}_s \mathbf{1} = 0$  and let assumption 4 and the following orthogonality condition hold,

$$(\mathcal{L}'_a(\rho), \mathbf{d}\mathcal{S}_\mu(\rho)) = 0 \quad \text{for every } \rho \in \mathcal{V}. \quad (56)$$

Then the FP equation  $\dot{\rho} = \mathcal{L}'(\rho)$  is a  $(\mathcal{S}_\mu, \Psi, \Psi^*, \mathcal{L}'_a)$  pre-GENERIC flow. That is, the operator  $\mathcal{L}_s$  is associated with the gradient flow part of the dynamics, i.e.

$$\mathcal{L}'_s \rho = \mathbf{d}_\xi \Psi^* \left( \rho; -\frac{1}{2} \mathbf{d}\mathcal{S}_\mu(\rho) \right) \quad (57)$$

and the flow generated by  $\mathcal{L}'$  can be written as

$$\partial_t \rho_t = W(\rho_t) + \mathbf{d}_\xi \Psi^* \left( \rho_t; -\frac{1}{2} \mathbf{d}\mathcal{S}_\mu(\rho_t) \right), \quad \text{with } W\rho = \mathcal{L}'_a \rho. \quad (58)$$

- **From pre-GENERIC form of the FP equation to symmetric–antisymmetric splitting of the generator.** *Vice versa, suppose that there exist  $\psi, \psi^*, W$  and  $\mathcal{S} \in C^1$  such that the FP equation  $\dot{\rho}_t = \mathcal{L}'(\rho_t)$  is a  $(\mathcal{S}, \psi, \psi^*, W)$  pre-GENERIC system; that is, suppose*

$$\partial_t \rho_t = \mathcal{L}' \rho_t = W(\rho_t) + \mathbf{d}_\xi \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{dS}(\rho_t) \right), \tag{59}$$

with  $W$  and  $S$  such that  $(W(\rho), \mathbf{dS}(\rho)) = 0$  for every  $\rho \in \mathcal{V}$ . Let

$$\tilde{\mathcal{H}}(\rho; \xi) := \psi^* \left( \rho; \xi - \frac{1}{2} \mathbf{dS}(\rho) \right) - \psi^* \left( \rho; -\frac{1}{2} \mathbf{dS}(\rho) \right)$$

and assume that  $\tilde{\mathcal{H}}$  is convex and  $\tilde{\mathcal{H}}(\rho; 0) = 0$  for every  $\rho \in \mathcal{V}$ . Assume moreover that  $\tilde{\mathcal{L}}$ , the Legendre transform of  $\tilde{\mathcal{H}}$ , is twice differentiable at the point  $(\mu; 0)$ . Then the operator  $\mathcal{L} - W'$  is symmetric with respect to  $L_\mu^2$  and

$$\mathcal{L} = (\mathcal{L} - W') + W'$$

constitutes a symmetric–antisymmetric decomposition of  $\mathcal{L}$ .

**Proof.** We first show that  $\Psi^*$  defined in (55) is a dissipation potential, i.e. that it is non-negative, symmetric, convex in the second argument and attains its minimum at 0. In fact, the convexity of  $\Psi^*$  follows immediately from the convexity of  $\mathcal{H}_s$ . We will show in note 5.3 (take  $\mathcal{F}$  to be the identity there) that the symmetry in  $L_\mu^2$  of  $\mathcal{L}_s$  implies

$$\mathcal{H}_s \left( \rho; \frac{1}{2} \mathbf{dS}_\mu(\rho) + \xi \right) = \mathcal{H}_s \left( \rho; \frac{1}{2} \mathbf{dS}_\mu(\rho) - \xi \right), \tag{60}$$

which in turn implies the symmetry of  $\Psi^*$ . From the definition of  $\Psi^*$ , equation (55), we have

$$\Psi^*(\rho, 0) = 0 \quad \text{and} \quad \mathbf{d}_\xi \Psi^*(\rho; \xi) = (\mathbf{d}_\xi \mathcal{H}_s) \left( \rho; \frac{1}{2} \mathbf{dS}_\mu(\rho) + \xi \right),$$

thus,

$$\mathbf{d}_\xi \Psi^*(\rho; 0) = (\mathbf{d}_\xi \mathcal{H}_s) \left( \rho; \frac{1}{2} \mathbf{dS}_\mu(\rho) \right) \quad \text{and} \quad \mathbf{d}_\xi \Psi^* \left( \rho; -\frac{1}{2} \mathbf{dS}_\mu(\rho) \right) = (\mathbf{d}_\xi \mathcal{H}_s)(\rho, 0).$$

To clarify notation, the RHS of the above formulas is the Frechet derivative of  $\mathcal{H}_s$  with respect to its second argument, calculated at  $\frac{1}{2} \mathbf{dS}_\mu(\rho)$  and 0 respectively. By taking the derivative with respect to  $\xi$  on both sides of (60) and then letting  $\xi = 0$ , we get

$$\mathbf{d}_\xi \Psi^*(\rho; 0) = (\mathbf{d}_\xi \mathcal{H}_s) \left( \rho; \frac{1}{2} \mathbf{dS}_\mu(\rho) \right) = 0.$$

Using the convexity and differentiability of  $\Psi^*$ , we have

$$\Psi^*(\rho; \xi) \geq \Psi^*(\rho; 0) + (\mathbf{d}_\xi \Psi^*(\rho; 0), \xi) = 0, \text{ for every } \rho, \xi.$$

We have thus proved that  $\Psi^*$  is a dissipation potential. Next we will establish (58). For every  $\bar{\xi} \in \mathcal{D}(\mathcal{H}_s, \rho)$ , we have

$$(\mathbf{d}_\xi \mathcal{H}_s(\rho; \xi), \bar{\xi}) = \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_s(\rho; \xi + \varepsilon \bar{\xi}) - \mathcal{H}_s(\rho; \xi)}{\varepsilon} = \int_{\mathbb{R}^d} \left[ -\bar{\xi} e^{-\xi} \mathcal{L}_s e^\xi + e^{-\xi} \mathcal{L}_s(\bar{\xi} e^\xi) \right] d\rho. \tag{61}$$

We postpone showing how the above expression is obtained and proceed with the main argument. From the above,

$$(\mathbf{d}_\xi \mathcal{H}_s(\rho; 0), \bar{\xi}) = \int_{\mathbb{R}^d} \left[ -\bar{\xi}(\mathcal{L}_s 1) + (\mathcal{L}_s \bar{\xi}) \right] d\rho = \int_{\mathbb{R}^d} (\mathcal{L}_s \bar{\xi}) d\rho = (\mathcal{L}'_s(\rho), \bar{\xi}),$$

where we have used  $\mathcal{L}_s 1 = 0$ . From this, we deduce

$$\mathcal{L}'_s(\rho) = \mathbf{d}_\xi \mathcal{H}_s(\rho; 0) = \mathbf{d}_\xi \Psi^* \left( \rho; -\frac{1}{2} \mathbf{dS}(\rho) \right),$$

which is precisely (57). That is,  $\mathcal{L}'_s$  is in gradient flow form; using (33) and writing  $\mathcal{L}'_s(\rho) = \mathcal{L}'(\rho) - \mathcal{L}'_a(\rho)$ , we then have

$$\Psi(\rho; \mathcal{L}'(\rho) - \mathcal{L}'_a(\rho)) + \Psi^* \left( \rho; -\frac{1}{2} \mathbf{dS}(\rho) \right) = \left( \mathcal{L}'(\rho) - \mathcal{L}'_a(\rho), -\frac{1}{2} \mathbf{dS}(\rho) \right) \stackrel{(56)}{=} \left( \mathcal{L}'(\rho), -\frac{1}{2} \mathbf{dS}(\rho) \right),$$

which, is equivalent to (58) (see (36)). We are now only left with showing the limit (61). To this end, for some  $\theta = \theta(x) \in (0, 1)$  we have:

$$\begin{aligned} \mathcal{H}_s(\rho; \xi + \varepsilon \bar{\xi}) &= \int_{\mathbb{R}^d} e^{-(\xi + \varepsilon \bar{\xi})} [\mathcal{L}_s(e^{\xi + \varepsilon \bar{\xi}})] d\rho \\ &= \int_{\mathbb{R}^d} \left( e^{-\xi} - \varepsilon \bar{\xi} e^{-\xi} + \frac{1}{2} \varepsilon^2 (\bar{\xi}(x))^2 e^{-\xi(x) - \theta \varepsilon \bar{\xi}(x)} \right) \\ &\quad \times \left[ \mathcal{L}_s(e^\xi + \varepsilon \bar{\xi} e^\xi + \frac{1}{2} \varepsilon^2 (\bar{\xi}(x))^2 e^{\xi(x) + \theta \varepsilon \bar{\xi}(x)}) \right] d\rho \\ &= \int_{\mathbb{R}^d} \left( e^{-\xi} \mathcal{L}_s e^\xi + \varepsilon [-\bar{\xi} e^{-\xi} \mathcal{L}_s(e^\xi) + e^{-\xi} \mathcal{L}_s(\bar{\xi} e^\xi)] \right) d\rho + \varepsilon^2 I \\ &= \mathcal{H}_s(\rho, \xi) + \varepsilon \int_{\mathbb{R}^d} [-\bar{\xi} e^{-\xi} \mathcal{L}_s(e^\xi) + e^{-\xi} \mathcal{L}_s(\bar{\xi} e^\xi)] d\rho + \varepsilon^2 I, \end{aligned}$$

where the term  $I$  gathers all the terms which are multiplied by a power of  $\varepsilon$  bigger or equal than two. In view of assumption 4 all the integrals appearing in the above are finite, hence we can take the limit  $\varepsilon \rightarrow 0$  and obtain the result.

Now suppose that (59) holds true or, equivalently,

$$\psi(\rho_t; \mathcal{L}'(\rho_t) - W(\rho_t)) + \psi^* \left( \rho_t; -\frac{1}{2} \mathbf{dS}(\rho_t) \right) + \frac{1}{2} (\mathcal{L}'(\rho_t), \mathbf{dS}(\rho_t)) = 0, \quad (W(\rho), \mathbf{dS}(\rho)) = 0 \quad \forall \rho \in \mathcal{V}. \tag{62}$$

By subtracting  $\frac{1}{2} (W(\rho_t), \mathbf{dS}(\rho_t)) = 0$  from the LHS of (62), we get

$$\psi(\rho; \mathcal{L}'(\rho) - W(\rho)) + \psi^* \left( \rho; -\frac{1}{2} \mathbf{dS}(\rho) \right) + (\mathcal{L}'(\rho) - W(\rho), \frac{1}{2} \mathbf{dS}(\rho)) = 0. \tag{63}$$

The above implies that the operator  $\mathcal{L}' - W$  is a gradient flow. From [33, theorem 3.4] the Hamiltonian  $\mathcal{H}$  is reversible with respect to  $\mathcal{S}$ , i.e.  $\tilde{\mathcal{H}}(\rho; \xi) = \tilde{\mathcal{H}}(\rho; \mathbf{dS}(\rho) - \xi)$  for every  $\xi$ . This is equivalent to  $\tilde{\mathcal{L}}$ , the Legendre transform of  $\mathcal{H}$ , satisfying equality (73) (that is, the relation [38, (2.6)]). This implies, by [38, theorem 3.3], to the fact that the operator  $(\mathcal{L}' - W)'$  satisfies detailed balance with respect to  $\mu$ . This concludes the proof.  $\square$



**Note 4.2.** Some comments on theorem 4.1.

- The first part of the theorem implies that if  $\mathcal{L}_s$  is symmetric in  $L^2_\mu$  then  $\mathcal{L}'_s$  is in gradient flow form. We prove this fact without making use of LDPs. It must be emphasised that we find *one possible* gradient flow structure for  $\mathcal{L}'_s$ , the one that corresponds to using the relative entropy as entropy functional. In other words the first part of the theorem says that we can always look at  $\mathcal{L}'_s$  as a gradient flow for the relative entropy. It is this ‘a-priori choice’ of entropy functional that allows one to bypass the use of LDPs. Avoiding the use of LDPs makes proofs simpler, and this is particularly true in the linear case of section 3, where only straightforward arguments are used. However it also conceals important physical considerations (summarised in the introduction) which shed a more profound light on the microscopic origin of the gradient flow structure.
- Condition (56) can be dropped by using the observation in note 2.3.3, point (b).
- In the proof of the second part of the theorem we used [38, theorem 3.3], which assumes the validity of a LDP. However the proof of the specific implication of that theorem that we use here does not require assuming the validity of a LDP.
- Suppose that  $\mathcal{H}_s(\rho; \cdot)$  is differentiable. Then it is convex in the second argument, for every  $\rho$  fixed, if and only if

$$\mathcal{H}_s(\rho; \xi + \bar{\xi}) \geq \mathcal{H}_s(\rho; \xi) + (\mathbf{d}_\xi \mathcal{H}_s(\rho; \xi), \bar{\xi}),$$

for all functions  $\xi, \bar{\xi}$  in the appropriate domains. Let  $\Gamma_s$  be the carré du champ operator associated to  $\mathcal{L}_s$ , namely

$$\Gamma_s(\xi, \bar{\xi}) := \frac{1}{2} (\mathcal{L}_s(\xi\bar{\xi}) - \xi\mathcal{L}_s\bar{\xi} - \bar{\xi}\mathcal{L}_s\xi).$$

Then

$$\begin{aligned} \mathcal{H}_s(\rho; \xi + \bar{\xi}) &= \int_{\mathbb{R}^d} \left( e^{-\xi}\mathcal{L}_s e^\xi + e^{-\bar{\xi}}\mathcal{L}_s e^{\bar{\xi}} + 2e^{-(\xi+\bar{\xi})}\Gamma_s(e^\xi, e^{\bar{\xi}}) \right) d\rho(x), \\ \mathcal{H}_s(\rho; \xi) + (\mathbf{d}_\xi \mathcal{H}_s(\rho; \xi), \bar{\xi}) &= \int_{\mathbb{R}^d} e^{-\xi}\mathcal{L}_s e^\xi d\rho(x) + \int (\mathcal{L}_s\bar{\xi} + 2e^{-\xi}\Gamma_s(e^\xi, \bar{\xi})) d\rho(x). \end{aligned}$$

Thus  $\mathcal{H}_s$  is convex if and only if

$$\int_{\mathbb{R}^d} e^{-\bar{\xi}} (\mathcal{L}_s e^{\bar{\xi}} + 2e^{-\xi}\Gamma_s(e^\xi, e^{\bar{\xi}})) d\rho(x) \geq \int_{\mathbb{R}^d} (\mathcal{L}_s\bar{\xi} + 2e^{-\xi}\Gamma_s(e^\xi, \bar{\xi})) d\rho(x) \text{ for all } \xi, \bar{\xi}. \tag{64}$$

For diffusion processes,  $\mathcal{L}_s$  and  $\Gamma_s$  satisfy the chain rule

$$\mathcal{L}_s\phi(\xi) = \partial_\xi\phi(\xi)\mathcal{L}_s\xi + \partial_\xi^2\phi(\xi)\Gamma_s(\xi, \xi) \quad \text{and} \quad \Gamma_s(\phi(\xi), \bar{\xi}) = \partial_\xi\phi(\xi)\Gamma_s(\xi, \bar{\xi}),$$

thus

$$\mathcal{L}_s e^{\bar{\xi}} = e^{\bar{\xi}}(\mathcal{L}_s\bar{\xi} + \Gamma_s(\bar{\xi}, \bar{\xi})), \quad e^{-(\xi+\bar{\xi})}\Gamma_s(e^\xi, e^{\bar{\xi}}) = e^{-\xi}\Gamma_s(e^\xi, \bar{\xi}) = \Gamma_s(\xi, \bar{\xi}). \tag{65}$$

In this case the convexity condition (64) reduces to

$$\int_{\mathbb{R}^d} \Gamma_s(\bar{\xi}, \bar{\xi}) \, d\rho(x) \geq 0,$$

which is always true as  $\Gamma_s(\bar{\xi}, \bar{\xi}) \geq 0$  (see for instance [5, section 1.4.2]).

### 5. LDPs and generalised-reversibility

In this section we consider a particular class of non-reversible processes, the class of so-called *generalised-reversible* processes, and we characterise the form of their large deviation rate functional. To this end, in sections 5.1 and 5.2, respectively, we give a summary background on generalised reversibility and LDPs, respectively. In section 5.3 we state and prove the main result of this section, proposition 5.2.

#### 5.1. Generalised reversibility

We start by briefly recalling the definition and main facts about generalised reversibility, more details can be found in [36, section 2.2.1.2], but this definition goes back to at least the work of Yaglom [54] (and for this reason generalised reversible processes are also called Yaglom-reversible at times). Let  $\mu$  be a probability measure on  $\mathbb{R}^d$  and  $\mathcal{F}$  be an involutive transformation (i.e.  $\mathcal{F} = \mathcal{F}^{-1}$ ) on  $\mathbb{R}^d$  leaving  $\mu$  invariant, that is,

$$\mu(\mathcal{F}(dx)) = \mu(\mathcal{F}^{-1}(dx)) = \mu(dx). \tag{66}$$

An  $\mathbb{R}^d$ -valued time-homogeneous Markov process  $(X(t))_{t \geq 0}$  is said to be *generalised reversible with respect to  $\mu$  up to  $\mathcal{F}$*  iff whenever  $X(0)$  is distributed according to  $\mu$  then  $(X(t))_{t \in [0, T]}$  and the time-reversed process  $(\mathcal{F}(X(T-t)))_{t \in [0, T]}$  have the same distribution (on the space of continuous paths), for every fixed  $T > 0$ .

Let  $\mathcal{L}$  be the generator of  $X_t$ . If  $\mathcal{F}$  is smooth enough—in particular, if it is such that the composition  $f \circ \mathcal{F}$  belongs to the domain of  $\mathcal{L}$  whenever  $f$  does, which we assume from now on—one can see that  $X_t$  is generalised reversible up to  $\mathcal{F}$  with respect to  $\mu$  iff

$$\int f(x)(\mathcal{L}g)(x) \, d\mu(x) = \int (g \circ \mathcal{F})(x)[\mathcal{L}(f \circ \mathcal{F})](x) \, d\mu(x), \quad \forall f, g \in \mathcal{D}_2(\mathcal{L}). \tag{67}$$

From the above it is clear that a generalised-reversible process is non-reversible in the sense that the detailed balance equation (40) does not hold. However when  $\mathcal{F}$  is the identity, the above condition simply reduces to detailed balance. It is easy to see that if  $\mathcal{L}$  is reversible up to  $\mathcal{F}$  with respect to  $\mu$ <sup>18</sup> then  $\mu$  is an invariant measure for  $X_t$  (just take  $f \equiv 1$  in the above and use the fact that constant functions belong to the kernel of  $\mathcal{L}$ ). Note that (67) can be reformulated as

$$(\mathcal{L}^*f)(x) = \mathcal{F}_\#[\mathcal{L}(\mathcal{F}_\#f)](x) = (\mathcal{L}(\mathcal{F}_\#f))(\mathcal{F}(x)), \quad \text{where } \mathcal{F}_\#f := f \circ \mathcal{F}. \tag{68}$$

Indeed, by a change of variable, we have

<sup>18</sup> By this we mean that the process with generator  $\mathcal{L}$  is reversible up to  $\mathcal{F}$  with respect to  $\mu$ .

$$\begin{aligned} \int g(x)\mathcal{F}_\#[\mathcal{L}(\mathcal{F}_\#f)](x) \, d\mu(x) &= \int g(x)[\mathcal{L}(f \circ \mathcal{F})(\mathcal{F}(x))] \, d\mu(x) \\ &\stackrel{(66)}{=} \int g(\mathcal{F}(x))[\mathcal{L}(f \circ \mathcal{F})](x) \, d\mu(x) \\ &\stackrel{(67)}{=} \int f(x)(\mathcal{L}g)(x) \, d\mu(x). \end{aligned}$$

A typical example of generalised reversibility is the Langevin dynamics  $(x_t, v_t)$  of equation (1), which is generalised reversible with respect to the measure  $\mu$  in (2) up to the velocity flip  $\mathcal{F} : (x, v) \mapsto (x, -v)$ , see [36].

5.2. Feng-Kurtz approach to LDP of empirical measures

The functional setting of this subsection and the next is slightly different from the one we used in previous sections; in particular throughout section 5 we drop our standing assumptions hypothesis 2.1.

Let  $t \in [0, T]$  and  $\{X_t^{(n)}\}_{n \in \mathbb{N}}$  be a sequence of time-homogeneous independent Markov processes, each of them with common state space  $\mathbb{R}^d$  and with common generator  $\mathcal{L} : \mathcal{D}_b(\mathcal{L}) \rightarrow C_b(\mathbb{R}^d)$ , where the domain  $\mathcal{D}_b$  of  $\mathcal{L}$  is the set of functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\mathcal{L}f \in C_b(\mathbb{R}^d)$ <sup>19</sup> and  $\mathcal{L}$  is a linear operator. We assume that  $\mathcal{D}_b$  is large enough that if (67) holds for every  $f, g, \in \mathcal{D}_b$  then it also holds for every  $f, g \in \mathcal{D}_2$ .

The empirical process  $\rho^{(n)}$  associated to  $\{X_t^{(n)}\}_n$  is defined by

$$\rho^{(n)} : t \mapsto \frac{1}{n} \sum_{i=1}^n \delta_{X^{(i)}(t)}, \quad t \in [0, T], \tag{69}$$

where  $\delta$  is the dirac delta. Denoting by  $\mathcal{P}(\mathbb{R}^d)$  the space of probability measures on  $\mathbb{R}^d$ , for each  $t \in [0, T]$  and  $n \in \mathbb{N}$  fixed,  $\rho_t^{(n)}$  is a random probability measure on  $\mathbb{R}^d$ . Hence  $\{\rho^{(n)}(\cdot)\}_n$  can be viewed as a sequence of  $D([0, T]; \mathcal{P}(\mathbb{R}^d))$ -valued random variables where  $D([0, T]; \mathcal{P}(\mathbb{R}^d))$  denotes the Skorohod space of paths from  $[0, T]$  to  $\mathbb{R}^d$ . Suppose that  $\rho^{(n)}(0)$ , as a sequence of  $\mathcal{P}(\mathbb{R}^d)$ -valued random variables, satisfies a LDP with a good rate function  $I_0 : \mathcal{P}(\mathbb{R}^d) \rightarrow [0, \infty)$  (see appendix A for basic definitions about LDPs). Then, under rather general conditions (see [28, chapter 13] for non-degenerate diffusion processes, [9, 18] for degenerate diffusion processes, [27, 38] for finite-state continuous time Markov chains or also [12]), the empirical process satisfies a LDP in  $D([0, T]; \mathcal{P}(\mathbb{R}^d))$  of the form <sup>20</sup>

$$\text{Prob}\left(\left(\rho_t^{(n)}\right)_{t=0}^T \approx \{\rho.\}\right) \stackrel{n \rightarrow \infty}{\sim} e^{-n\mathcal{I}_T(\{\rho.\})}, \quad \mathcal{I}_T(\{\rho.\}) := I_0(\rho_0) + \int_0^T \mathcal{L}(\rho_t; \dot{\rho}_t) \, dt, \tag{70}$$

for some function  $\mathcal{L}$  which is commonly referred to as the *Lagrangian*, for reasons that will be clear in few lines. In the above and throughout the notation  $\{\rho.\}$  is to specify that we are referring to the whole path  $\{\rho_t\}_{t \in [0, T]} \subset \mathcal{P}(\mathbb{R}^d)$ . If  $\rho^{(n)}$  satisfies the LDP (70) then the functional  $\mathcal{L}$  is convex in the second variable and

<sup>19</sup> Let us recall that there are in general two approaches to defining the domain of a Markov operator: either one defines it as the largest set where the operator can be seen as the generator of the associated Markov semigroup, and this is substantially what we did in previous sections, or one defines it by fixing the image of  $\mathcal{L}$ , which is what we do here. We use the space of continuous and bounded functions here because it generates the narrow (weak) topology in the space of probability measures, which is a natural topology for large deviation results in this section [12, 28].

<sup>20</sup> The complete explanation of this popular short-hand notation is contained in appendix A.

$$\mathcal{L}(\rho_t; \dot{\rho}_t) = 0 \Leftrightarrow \dot{\rho}_t = \mathcal{L}' \rho_t$$

where the FP-equation on the RHS of the above is to be understood on the space of probability measures [38, proposition 3.1].

Feng and Kurtz [28] provide a formal method to calculate the large-deviation rate functional  $\mathcal{L}$  in a variety of scenarios; in our setting, i.e. in the case in which we consider LDPs for empirical measures of i.i.d processes, such a method proceeds as follows: starting from the generator  $\mathcal{L}$ , we compute the so-called *Hamiltonian*  $\mathcal{H} = \mathcal{H}(\rho; \xi) = \mathcal{H}_\rho(\xi)$ , as follows

$$\mathcal{H}(\rho; \xi) := \int_{\mathbb{R}^d} e^{-\xi(x)} (\mathcal{L}e^\xi)(x) d\rho(x). \tag{71}$$

For every  $\rho \in \mathcal{P}(\mathbb{R}^d)$ , we regard  $\mathcal{H}$  as a function  $\mathcal{H}(\rho; \cdot) : \mathcal{D}_b(\mathcal{H}_\rho) \rightarrow \mathbb{R}$ , where  $\mathcal{D}_b(\mathcal{H}_\rho)$  consists of the set of functions  $\xi : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $e^\xi \in \mathcal{D}_b(\mathcal{L})$  and the right hand side of the above makes sense. Once the Hamiltonian is obtained, if  $\rho^{(n)}$  satisfies a LDP, the Lagrangian  $\mathcal{L}(\rho; \cdot) : \mathcal{D}_b(\mathcal{L}_\rho) \rightarrow \mathbb{R}$  is found as the Legendre transform of  $\mathcal{H}$ , namely

$$\mathcal{L}(\rho; g) = \sup_{\xi \in \mathcal{D}_b(\mathcal{H}_\rho)} \left\{ (\xi, g) - \mathcal{H}(\rho; \xi) \right\}, \tag{72}$$

where in the above  $(\cdot, \cdot)$  is a formal  $L^2$  scalar product.

**Note 5.1.** We emphasise that the approach of Feng and Kurtz gives a formal procedure to find the large deviation rate functional  $\mathcal{L}$ , however to make this procedure rigorous one needs to verify many technical conditions; Feng and Kurtz [28] presents rigorous proofs of LDPs for many Markov processes. Moreover, as one would expect, specifying the appropriate domains of  $\mathcal{H}$  and  $\mathcal{L}$  is better done on a case by case basis, see e.g. [38, section 4.2]. Again to compare with the presentation of this theory on manifolds, if  $\mathcal{Z}$  is a manifold then, for every  $z \in \mathcal{Z}$ ,  $\mathcal{H}(z; \cdot) : T_z^* \mathcal{Z} \rightarrow \mathbb{R}$  and  $\mathcal{L}(z; \cdot) : T_z \mathcal{Z} \rightarrow \mathbb{R}$ .

Assuming that for every  $n \in \mathbb{N}$  the process  $\{X_t^{(n)}\}_{t \geq 0}$  admits a unique invariant measure  $\mu$  (which is the same for every  $n \in \mathbb{N}$ ) and that the associated empirical process  $\rho^{(n)}$  satisfies a LDP of the form (70) with Lagrangian  $\mathcal{L}$ , one of the main results of [38] can be informally stated as follows (see [38, theorem 3.3] for a precise statement): the process  $X_t^{(n)}$  is reversible with respect to  $\mu$  (in the sense that  $\mu$  satisfies detailed balance (40)) if and only if the Lagrangian  $\mathcal{L}$  appearing in the LDP satisfies the following relation

$$\mathcal{L}(\rho; g) - \mathcal{L}(\rho; -g) = (\mathbf{d}\mathcal{S}_\mu(\rho), g), \quad \text{for every } \rho \in \mathcal{P}(\mathbb{R}^d), g \in \mathcal{D}_b(\mathcal{L}_\rho), \tag{73}$$

where  $\mathcal{S}_\mu(\rho)$  has been defined in (41). In the next subsection we show an analogous result for the case in which the invariant measure  $\mu$  is only reversible up to some transformation  $\mathcal{F}$ .

### 5.3. Reversibility, generalised reversibility and LDPs

We keep working in the setting of section 5.2, i.e.  $\{X_t^{(n)}\}_{t \geq 0}$  is a sequence of independent time-homogeneous Markov processes with common generator  $\mathcal{L}$ ; again suppose that for each  $n \in \mathbb{N}$  the process  $X_t^{(n)}$  admits an invariant measure  $\mu$  which is the same for every  $n$ . Consider the map  $\varphi : D([0, T]; \mathcal{P}(\mathbb{R}^d)) \rightarrow D([0, T]; \mathcal{P}(\mathbb{R}^d))$  defined as

$$\varphi(\{\rho_t\}_{t \in [0, T]}) = \{\mathcal{F}_\# \rho_{T-t}\}_{t \in [0, T]}, \tag{74}$$

where we recall  $\mathcal{F}_\# \rho$  denotes the push-forward of the probability measure  $\rho$  under a map  $\mathcal{F} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , i.e.  $\mathcal{F}_\# \rho(B) = \rho(\mathcal{F}^{-1}(B))$  for every Borel set  $B \subset \mathbb{R}^d$ . We make the following assumption.

**Assumption 5.1** Let  $\mathcal{F}$  be an involutive, volume preserving transformation (i.e.  $|J\mathcal{F}| = 1$ )<sup>21</sup> on  $\mathbb{R}^d$  leaving  $\mu$  invariant and satisfying the following regularity assumptions: (a)  $\mathcal{F}$  is smooth enough so that the map  $\varphi$  in (74) is well defined (i.e.  $\{\mathcal{F}_\# \rho_{T-t}\}_{t \in [0, T]}$  belongs to  $D([0, T]; \mathcal{P}(\mathbb{R}^d))$  if the path  $\{\rho_t\}_{t \in [0, T]}$  does) and continuous; (b) if  $g \in \mathcal{D}_b(\mathcal{L})$  ( $\mathcal{D}_b(\mathcal{H})$ , respectively) then  $g \circ \mathcal{F} \in \mathcal{D}_b(\mathcal{L})$  ( $\mathcal{D}_b(\mathcal{H})$ , respectively).

**Proposition 5.2.** *With the setting and notation introduced above, suppose that  $\{X^{(n)}(0)\}_{n \in \mathbb{N}}$  are i.i.d random variables, distributed according to  $\mu$ ; suppose the empirical process  $\rho^{(n)}$  associated with the sequence  $\{X^{(n)}(t)\}_{n \in \mathbb{N}}$  satisfies a LDP on  $D([0, T]; \mathcal{P}(\mathbb{R}^d))$  with a good rate function  $\mathcal{I}_T : D([0, T]; \mathcal{P}(\mathbb{R}^d)) \rightarrow [0, \infty)$  given by*

$$\mathcal{I}_T(\{\rho.\}) = I_0(\rho_0) + \int_0^T \mathcal{L}(\rho_t; \dot{\rho}_t) dt, \tag{75}$$

where  $I_0$  is the large deviation rate functional of  $\{\rho^{(n)}(0)\}_n$ , the empirical process at the initial time. Let  $\mathcal{F} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a transformation on  $\mathbb{R}^d$  satisfying assumption 5.1. If for each  $n \in \mathbb{N}$  the process  $\{X^{(n)}(t)\}_t$  is reversible with respect to  $\mu$  up to  $\mathcal{F}$  then the Lagrangian  $\mathcal{L}$  appearing in the LDP satisfies the following relation:

$$\mathcal{L}(\rho; g) - \mathcal{L}(\mathcal{F}_\# \rho; -g \circ \mathcal{F}) = (\mathbf{d}\mathcal{S}_\mu(\rho), g), \quad \text{for every } \rho \in \mathcal{P}(\mathbb{R}^d), g \in \mathcal{D}_b(\mathcal{L}_\rho), \tag{76}$$

where  $\mathcal{S}_\mu(\rho)$  is as in (41). Moreover, if  $\mathbf{d}\mathcal{S}_\mu(\rho) \in \mathcal{D}_b(\mathcal{H}_\rho)$ , (76) is equivalent to the following relation (expressed in terms of the Hamiltonian)

$$\mathcal{H}(\rho; \mathbf{d}\mathcal{S}_\mu(\rho) + \xi) = \mathcal{H}(\mathcal{F}_\# \rho; -\xi \circ \mathcal{F}). \tag{77}$$

If, additionally, the Lagrangian  $\mathcal{L}$  is twice differentiable near  $(\mu; 0)$ , then the converse statement holds as well, i.e. (76) implies (67).

**Proof.** We build on the proof of [38, theorem 3.3]. Consider the map  $\varphi : D([0, T]; \mathcal{P}(\mathbb{R}^d)) \rightarrow D([0, T]; \mathcal{P}(\mathbb{R}^d))$  defined in (74). By the involutivity of  $\mathcal{F}$ ,  $\varphi$  is invertible and involutive, and indeed the inverse map is given by

$$\varphi^{-1}(\{\rho_t\}_{t \in [0, T]}) = \{\mathcal{F}_\# \rho_{T-t}\}_{t \in [0, T]}, \quad \text{for any } \{\rho_t\}_{t \in [0, T]} \in D([0, T]; \mathcal{P}(\mathbb{R}^d)).$$

Let  $\tilde{\rho}^{(n)}$  be the empirical process associated with the process  $\mathcal{F}(X^{(n)}(T - t))$ , i.e.

$$\tilde{\rho}^{(n)} : t \mapsto \frac{1}{n} \sum_{i=1}^n \delta_{\mathcal{F}(X^{(i)}(T-t))}.$$

Using the identity

$$\int_{\mathbb{R}^d} f(y) (\mathcal{F}_\# \rho)(dy) = \int_{\mathbb{R}^d} f(\mathcal{F}(x)) \rho(dx),$$

which holds for every continuous and bounded  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , one can see that  $\tilde{\rho}^{(n)} = \varphi(\rho^{(n)})$ . Since  $\rho^{(n)}$  satisfies a LDP with the rate functional  $\mathcal{I}_T$ , by the contraction principle [28, lemma 3.11]<sup>22</sup>, also  $\tilde{\rho}^{(n)}$  satisfies a LDP with the rate functional  $\tilde{\mathcal{I}} : D([0, T]; \mathcal{P}(\mathbb{R}^d)) \rightarrow [0, \infty)$  given by<sup>23</sup>

$$\tilde{\mathcal{I}}(\{\rho_t\}_{t \in [0, T]}) = \mathcal{I}(\varphi(\{\rho_t\}_{t \in [0, T]})). \tag{78}$$

<sup>21</sup> Here  $|J\mathcal{F}|$  denotes the determinant of the Jacobian of  $\mathcal{F}$ .

<sup>22</sup> It is in applying the contraction principle that we use the assumption that  $\mathcal{I}$  is a good rate function.

<sup>23</sup> Since  $\varphi$  is invertible, in the application of contraction principle we do not need to take an infimum over the pre-image set

Letting  $\widehat{\{\rho_t\}_t} = \varphi(\{\rho_t\}_t)$ , the above can be rewritten as

$$\tilde{\mathcal{I}}_T(\{\rho_t\}_t) = \mathcal{I}_T(\widehat{\{\rho_t\}_t}) = I_0(\hat{\rho}_0) + \int_0^T \mathcal{L}(\hat{\rho}_t, \dot{\hat{\rho}}_t) dt.$$

We claim that the following identity holds:

$$-\dot{\hat{\rho}}_t(x) = (\hat{\rho}_{T-t} \circ \mathcal{F})(x), \quad \text{for every } x \in \mathbb{R}^d. \tag{79}$$

We will prove this claim below and for the time being we move on with the main argument.

Note now that the generalised reversibility of the underlying particle systems implies that  $\mathcal{I} = \tilde{\mathcal{I}}$ . This is because, by generalised reversibility, for each  $n$ ,  $(X^{(n)}(t))_{t \in [0, T]}$  and the time-reversed process  $(\mathcal{F}(X^{(n)}(T-t)))_{t \in [0, T]}$  have the same distribution, for every fixed  $T > 0$ . Hence,

$$\text{prob} \left( \{\rho_t^{(n)}\}_{t \in [0, T]} \in \Theta \right) = \text{prob} \left( \{\rho_t^{(n)}\}_{t \in [0, T]} \in \varphi(\Theta) \right),$$

for every open set  $\Theta \subseteq D([0, T]; \mathcal{P}(\mathbb{R}^d))$ . From  $\mathcal{I} = \tilde{\mathcal{I}}$  and (78), for any curve  $\{\rho_t\}_t \subset D([0, T]; \mathcal{P}(\mathbb{R}^d))$  we then have

$$\begin{aligned} I_0(\rho_0) + \int_0^T \mathcal{L}(\rho_t; \dot{\rho}_t) dt &= I_0(\hat{\rho}_0) + \int_0^T \mathcal{L}(\hat{\rho}_t; \dot{\hat{\rho}}_t) dt \\ &\stackrel{(79)}{=} I_0(\mathcal{F}_\# \rho_T) + \int_0^T \mathcal{L}(\mathcal{F}_\# \rho_{T-t}; -\dot{\rho}_{T-t} \circ \mathcal{F}) dt \\ &= I_0(\mathcal{F}_\# \rho_T) + \int_0^T \mathcal{L}(\mathcal{F}_\# \rho_t; -\dot{\rho}_t \circ \mathcal{F}) dt. \end{aligned}$$

Since  $\{X^{(n)}(0)\}_n$  are i.i.d. with distribution  $\mu$ , according to Sanov’s theorem [13, theorem 6.2.10] we have  $I_0(\rho_0) = \mathcal{S}_\mu(\rho_0)$ . Thus,

$$\mathcal{S}_\mu(\mathcal{F}_\# \rho_T) - \mathcal{S}_\mu(\rho_0) = \int_0^T [\mathcal{L}(\rho_t; \dot{\rho}_t) - \mathcal{L}(\mathcal{F}_\# \rho_t; -\dot{\rho}_t \circ \mathcal{F})] dt. \tag{80}$$

Differentiating the above with respect to  $T$ , we then get

$$(\mathbf{d}\mathcal{S}_\mu(\mathcal{F}_\# \rho_T), \dot{\rho}_T \circ \mathcal{F}) = \mathcal{L}(\rho_T; \dot{\rho}_T) - \mathcal{L}(\mathcal{F}_\# \rho_T; -\dot{\rho}_T \circ \mathcal{F}).$$

Since  $\{\rho_t\}_t$  is an arbitrary curve in  $D([0, T]; \mathcal{P}(\mathbb{R}^d))$ ,  $\rho_T$  is an arbitrary probability measure, as is arbitrary the element  $\dot{\rho}_T \in \mathcal{D}_b(\mathcal{L}_{\rho_T})$ , hence

$$(\mathbf{d}\mathcal{S}_\mu(\rho), g \circ \mathcal{F}) = \mathcal{L}(\mathcal{F}_\# \rho; g) - \mathcal{L}(\rho; -g \circ \mathcal{F}),$$

for any  $\rho, g$  in the appropriate spaces, which is equivalent to (76). To conclude the proof of (76) we just need to prove the claim (79). This can be seen from the following calculations

$$\begin{aligned} \int \dot{\hat{\rho}}_t(x) f(x) dx &= \frac{d}{dt} \int \hat{\rho}_t(x) f(x) dx \\ &= - \int \dot{\rho}_{T-t}(x) (f \circ \mathcal{F})(x) dx = - \int (\dot{\rho}_{T-t} \circ \mathcal{F})(x) f(x) dx, \end{aligned}$$

where the second equality follows from the definition of  $\hat{\rho}_t$  and having used the fact that  $\mathcal{F}$  is volume preserving. The convexity of  $\mathcal{L}$  (implied by the LDP) now allows us to find the relation (77) as follows:

$$\begin{aligned} \mathcal{H}(\rho; \mathbf{d}\mathcal{S}_\mu(\rho) + \xi) &= \sup_s \{(\mathbf{d}\mathcal{S}_\mu(\rho) + \xi, g) - \mathcal{L}(\rho; g)\} \\ &\stackrel{(76)}{=} \sup_g \{(\xi, g) - \mathcal{L}(\mathcal{F}_\# \rho; -g \circ \mathcal{F})\} \\ &= \sup_g \{(\xi, -g \circ \mathcal{F}) - \mathcal{L}(\mathcal{F}_\# \rho; g)\} \\ &= \sup_g \{(-\xi \circ \mathcal{F}, g) - \mathcal{L}(\mathcal{F}_\# \rho; g)\} \\ &= \mathcal{H}(\mathcal{F}_\# \rho; -\xi \circ \mathcal{F}). \end{aligned}$$

Let us now prove the converse implication, i.e. that (76) implies (67). This can be done by a modification of the proof of [38, theorem 3.3], so we do not repeat all the details but we just explain the main steps. Standing the observations of [38, remark 3.4], let us define the function

$$a(\epsilon_1, \epsilon_2, \dots, \epsilon_6) = \mathcal{L}((1 + \epsilon_1 \rho + \epsilon_4 \rho_\#)\mu; \mathcal{L}'((1 + \epsilon_2 \rho + \epsilon_3 g + \epsilon_5 \rho_\# + \epsilon_6 g_\#)\mu)),$$

where  $\epsilon_j \in \mathbb{R}$  for every  $j \in \{1, \dots, 6\}$  and, just for the rest of this proof, we set  $g_\# = \mathcal{F}_\# g$ , similarly for  $\rho_\#$ . Since  $\mathbf{d}^2 \mathcal{S}_\mu(\mu)[g_1, g_2] = \int_{\mathbb{R}^d} \frac{dg_1}{d\mu} \frac{dg_2}{d\mu} d\mu$ , the statement that we want to prove can be rewritten as

$$\mathbf{d}^2 \mathcal{S}_\mu(\mu)[\mu \rho, \mathcal{L}'(\mu g)] = \mathbf{d}^2 \mathcal{S}_\mu(\mu)[\mu g_\#, \mathcal{L}'(\mu \rho_\#)]. \tag{81}$$

To prove the above we make two observations: firstly, by (76) (applied to the functions  $(1 + \epsilon_1 \rho)\mu$  and  $\mathcal{L}'((1 + \epsilon_3 g)\mu)$ ), we have

$$a(\epsilon_1, 0, \epsilon_3, 0, 0, 0) - a(0, 0, 0, \epsilon_1, 0, -\epsilon_3) = (\mathbf{d}\mathcal{S}_\mu((1 + \epsilon_1 \rho)\mu), \mathcal{L}'(\epsilon_3 g \mu)); \tag{82}$$

secondly, by the validity of the LDP, the function  $\mathcal{L}((1 + \epsilon_1 \rho + \epsilon_4 \rho_\#)\mu; \mathcal{L}'[(1 + \epsilon_1 \rho + \epsilon_4 \rho_\#)\mu] + h)$  attains its minimum when  $h = 0$ , hence

$$D_3 a(\epsilon_1, \epsilon_1, 0, \epsilon_4, \epsilon_4, 0) = 0 \text{ and } D_6 a(\epsilon_1, \epsilon_1, 0, \epsilon_4, \epsilon_4, 0) = 0. \tag{83}$$

If we differentiate (82) with respect to  $\epsilon_3$  and then with respect to  $\epsilon_1$  (and then calculate both derivatives in  $\epsilon_1 = \epsilon_3 = 0$ ), we get

$$D_{13} a(0, \dots, 0) + D_{46} a(0, \dots, 0) = \mathbf{d}^2 \mathcal{S}_\mu(\mu)[\rho \mu, \mathcal{L}'(g \mu)].$$

From (83) instead, we have

$$\begin{aligned} D_{13} a(0, \dots, 0) + D_{23} a(0, \dots, 0) &= 0 \\ D_{46} a(0, \dots, 0) + D_{56} a(0, \dots, 0) &= 0. \end{aligned}$$

From the above we deduce

$$-D_{23} a(0, \dots, 0) - D_{56} a(0, \dots, 0) = \mathbf{d}^2 \mathcal{S}_\mu(\mu)[\rho \mu, \mathcal{L}'(g \mu)].$$

If we calculate explicitly the two derivatives on the LHS of the above, we obtain

$$\mathbf{d}_{gg}^2 \mathcal{L}(\mu, 0) ((\mathcal{L}'(g \mu), \mathcal{L}'(\rho \mu)) + (\mathcal{L}'(\mu \rho_\#), \mathcal{L}'(\mu g_\#))) = \mathbf{d}^2 \mathcal{S}_\mu(\mu)[\rho \mu, \mathcal{L}'(g \mu)],$$

where  $\mathbf{d}_{gg}^2 \mathcal{L}$  denotes the second derivative of  $\mathcal{L}$  with respect to its second argument. The LHS of the above expression remains unchanged upon swapping  $g$  with  $\rho_{\#}$  and  $\rho$  with  $g_{\#}$ , hence (81). This concludes the proof.  $\square$

**Note 5.3.** • In proposition 5.2 we started with assuming that the process  $\{X^{(n)}(t)\}_t$  is reversible with respect to  $\mu$  up to  $\mathcal{F}$  and proved that, if this is the case, then the Lagrangian needs to satisfy the relation (75). From the relation for the Lagrangian we then deduced property (77) for the Hamiltonian. An alternative way of proceeding is to derive (77) directly from the generalised reversibility condition (67) expressed in terms of the generator, starting from the explicit formula of  $\mathcal{H}$  in (71):

$$\begin{aligned} \mathcal{H}(\rho, \mathbf{dS}_{\mu}(\rho) + \xi) &= \mathcal{H}\left(\rho, \log \frac{d\rho}{d\mu} + 1 + \xi\right) \\ &= \int_{\mathbb{R}^d} e^{-\log \frac{\rho(x)}{\mu(x)} - 1 - \xi(x)} \mathcal{L}\left(e^{\log \frac{\rho}{\mu} + 1 + \xi}\right)(x) \rho(x) dx \\ &= \int_{\mathbb{R}^d} e^{-\xi(x)} \left(\mathcal{L} e^{\log \frac{\rho}{\mu} + \xi}\right)(x) \mu(x) dx \\ &\stackrel{(66), (67)}{=} \int_{\mathbb{R}^d} \rho(\mathcal{F}(x)) e^{(\xi \circ \mathcal{F})(x)} \left(\mathcal{L} e^{-(\xi \circ \mathcal{F})}\right)(x) dx \\ &= \mathcal{H}(\mathcal{F}_{\#} \rho; -\xi \circ \mathcal{F}). \end{aligned}$$

Note that in the first equality we have used the linearity of  $\mathcal{L}$  to take the constant  $e^{-1}$  out of the argument of  $\mathcal{L}$ . Once (77) is formally obtained, (76) can be derived from (77) as follows (in the second equality below we use the fact that for any function  $f = f(\xi)$  one has  $\sup_{\xi} f(\xi) = \sup_{\xi} f(\xi + a)$ , for any constant  $a$ )

$$\begin{aligned} \mathcal{L}(\rho; g) &= \sup_{\xi} \{(\xi, g) - \mathcal{H}(\rho; \xi)\} \\ &= \sup_{\xi} \{(\xi + \mathbf{dS}_{\mu}(\rho), g) - \mathcal{H}(\rho; \xi + \mathbf{dS}_{\mu}(\rho))\} \\ &= (\mathbf{dS}_{\mu}(\rho), g) + \sup_{\xi} \{(\xi, g) - \mathcal{H}(\rho; \xi + \mathbf{dS}_{\mu}(\rho))\} \\ &\stackrel{(77)}{=} (\mathbf{dS}_{\mu}(\rho), g) + \sup_{\xi} \{(\xi, g) - \mathcal{H}(\rho \circ \mathcal{F}; -\xi \circ \mathcal{F})\} \\ &= (\mathbf{dS}_{\mu}(\rho), g) + \sup_{\xi} \{(\xi \circ \mathcal{F}, g) - \mathcal{H}(\rho \circ \mathcal{F}; -\xi)\} \\ &= (\mathbf{dS}_{\mu}(\rho), g) + \sup_{\xi} \{(\xi, g \circ \mathcal{F}) - \mathcal{H}(\rho \circ \mathcal{F}, -\xi)\} \\ &= (\mathbf{dS}_{\mu}(\rho), g) + \sup_{\xi} \{(-\xi, -g \circ \mathcal{F}) - \mathcal{H}(\rho \circ \mathcal{F}; -\xi)\} \\ &= (\mathbf{dS}_{\mu}(\rho), g) + \sup_{\xi} \{(\xi, -g \circ \mathcal{F}) - \mathcal{H}(\rho \circ \mathcal{F}; \xi)\} \\ &= (\mathbf{dS}_{\mu}(\rho), g) + \mathcal{L}(\rho \circ \mathcal{F}; -g \circ \mathcal{F}). \end{aligned}$$

The observations in [38, remark 3.5] can be seen as a specific instance of the above calculations, in the case in which  $\mathcal{F}$  is the identity.

- By combining theorem 4.1 and proposition 5.2, one can derive a gradient flow structure of the FP equation  $\partial_t \rho = \mathcal{L}' \rho$  from a large deviation rate functional by computing the Hamiltonian  $\mathcal{H}$  from the Lagrangian  $\mathcal{L}$  appearing in the rate functional (75) then the dissipation potential



$\Psi^*$  from  $\mathcal{H}$  using (55) (the entropy functional being the relative entropy). As shown in [38] for a diffusion process with detailed balance, the resulting gradient flow structure is precisely the (quadratic) Wasserstein gradient flow discovered by Jordan–Kinderlehrer–Otto [57, 61] (cf (25)); while for a continuous time Markov chain with detailed balance, it is a non-quadratic gradient flow (cf (32)), corresponding to the dissipation potential given by [38, equation (4.6)]

$$\Psi^*(\rho, \xi) = \frac{1}{2} \sum_{i,j=1}^J \sqrt{\rho_i \rho_j Q_{ij} Q_{ji}} (\cosh(\xi_j - \xi_i) - 1),$$

where  $Q = (Q_{ij}) \in \mathbb{R}^{J \times J}$  denotes the generator of the chain. This is not the gradient flow structure for a (detailed balance) Markov chain found by Maas, Chow *et al* and Mielke [58–60] (although the entropy functional being the same relative entropy, the dissipation potential in these papers is quadratic), see also the Introduction for relevant discussion. In section 6, we derive pre-GENERIC structures for non-reversible processes and PDMP processes using (71) to directly compute the Hamiltonian from the generator.

### 6. Examples: diffusion processes and PDMPs

In this section we consider two classes of processes, diffusion processes and PDMPs, sections 6.1 and 6.2, respectively. We will show how the theory developed in previous sections can be applied to such classes of Markov evolutions and indicate assumptions under which the functional framework of previous sections can be made rigorous.

#### 6.1. Diffusion processes

We demonstrate how the results in sections 3 and 4 can be applied to diffusion processes. Let  $b \in C^1(\mathbb{R}^d; \mathbb{R}^d)$  be globally Lipschitz and  $\sigma \in C^2(\mathbb{R}^d; \mathbb{R}^{d \times m})$  be uniformly elliptic. We consider the following Ito diffusion process

$$dX_t = b(X_t) dt + \sqrt{2} \sigma(X_t) d\beta_t, \quad X_0 = x \in \mathbb{R}^d, \tag{84}$$

where  $(\beta_t)_{t \geq 0}$  is a standard Brownian motion in  $\mathbb{R}^m$ . The generator of this process is given, on suitably smooth functions  $u : \mathbb{R}^d \rightarrow \mathbb{R}$ , by

$$(\mathcal{L}u)(x) = b(x) \cdot \nabla u(x) + D(x) : \nabla^2 u(x), \tag{85}$$

where  $D(x) = \sigma(x)\sigma(x)^\top$  and the notation ‘:’ denotes the Frobenius inner product of two matrices. The FP equation associated to (84) is

$$(\partial_t \rho_t)(x) = (\mathcal{L}' \rho_t)(x) = \operatorname{div}(D(x) \nabla \rho_t(x) - b(x) \rho_t(x)). \tag{86}$$

Assume that the process admits an invariant measure  $\mu(dx) = \mu(x) dx$  (simple sufficient conditions for this can be found e.g. in [43]), which is necessarily unique and smooth in view of our ellipticity assumption. We can also assume that  $\mu$  is positive everywhere. By definition, such an invariant measure satisfies

$$\mathcal{L}' \mu = \operatorname{div}(D \nabla \mu - b \mu) = 0. \tag{87}$$

With this setup we can take  $\mathcal{V}$  to be the set of (positive) Schwartz functions.

6.1.1. *Wasserstein pre-GENERIC and hypocoercive formulations.* Let  $h_t(x) = \rho_t(x)/\mu(x)$ . Then  $h$  satisfies the hypocoercive m-KE (19) where (see [52, proposition 3])<sup>24</sup>

$$B := (b - D\nabla \log \mu) \cdot \nabla, \quad A := \sigma \nabla \quad \text{and} \quad A^*g = -\operatorname{div}(\sigma^\top g) - (\nabla \log \mu, \sigma^\top g).$$

Since  $B$  and  $A$  are derivations on  $\mathbb{R}^d$ , they satisfy both chain and product rules. Hence according to section 3.1, the FP equation (86) can be written in the quadratic pre-GENERIC form (37) with the entropy functional given by the relative entropy of  $\rho$  with respect to  $\mu$  (see (41)) and (see (42))

$$W\rho = B'\rho = \operatorname{div}(\rho D\nabla \log(\mu) - b\rho), \tag{88}$$

$$M_\rho(\xi) = 2A'(\rho A\xi) = -2\operatorname{div}(\rho D\nabla \xi). \tag{89}$$

6.1.2. *Pre-GENERIC formulation of the FPE from the symmetric–antisymmetric decomposition.* In this section, we will cast the FP equation (86) into the pre-GENERIC formulation (59) using the symmetric–antisymmetric decomposition of the generator  $\mathcal{L}$  (see (85)) and show that this is the same as the quadratic pre-GENERIC structure (88) and (89) obtained from the hypocoercive m-KE (19) in the previous section.

First, we find the symmetric and antisymmetric parts in  $L_\mu^2$  of  $\mathcal{L}$ , using the relation

$$\mathcal{L}_s\phi = \frac{\mathcal{L}\phi + \mathcal{L}^*\phi}{2} = \frac{\mathcal{L}\phi + \mu^{-1}\mathcal{L}'(\mu\phi)}{2},$$

where the second equality in the above follows from the following calculation

$$(\varphi, \mathcal{L}^*\phi) = \langle \mu^{-1}\varphi, \mathcal{L}^*\phi \rangle = \langle \phi, \mathcal{L}(\mu^{-1}\varphi) \rangle = (\varphi, \mu^{-1}\mathcal{L}'(\mu\phi)).$$

The symmetric part  $\mathcal{L}_s$  can be found explicitly

$$\begin{aligned} \mathcal{L}_s\phi &= \frac{1}{2}(\mathcal{L}\phi + \mu^{-1}\mathcal{L}'(\mu\phi)) \\ &= \operatorname{div}(D\nabla\phi) + D\nabla\phi \cdot \nabla \log \mu, \end{aligned}$$

we have used standard calculations and the stationary condition (87) of  $\mu$  to obtain the second equality. The anti-symmetric part is then given by

$$\mathcal{L}_a\phi = \mathcal{L}\phi - \mathcal{L}_s\phi = b \cdot \nabla\phi - D\nabla\phi \cdot \nabla \log \mu.$$

The  $L^2$ -dual operator  $\mathcal{L}'_s$  of  $\mathcal{L}_s$  is

$$\mathcal{L}'_s\rho = \operatorname{div}(D\nabla\rho) - \operatorname{div}(\rho D\nabla \log \mu) = \operatorname{div}[\rho D\nabla(\log(\rho/\mu))] = \operatorname{div}[\rho D\nabla(\mathbf{d}\mathcal{S}_\mu(\rho))]. \tag{90}$$

Next, we find the Hamiltonian  $\mathcal{H}$  and its symmetric part  $\mathcal{H}_s$  from  $\mathcal{L}$  and  $\mathcal{L}_s$  (see (52) and (53)). Using (65) we obtain

$$\mathcal{H}(\rho; \xi) = \int e^{-\xi} \mathcal{L}e^\xi \rho = \int e^{-\xi} [e^\xi(\mathcal{L}\xi + \Gamma(\xi, \xi))] \rho = (\xi, \mathcal{L}'\rho) + (D\nabla\xi \cdot \nabla\xi, \rho). \tag{91}$$

Similarly we get

$$\mathcal{H}_s(\rho; \xi) = \int e^{-\xi} \mathcal{L}_s e^\xi \rho = (\xi, \mathcal{L}'_s\rho) + (D\nabla\xi \cdot \nabla\xi, \rho).$$

<sup>24</sup> See the discussion after (18) on the notation  $A^*A$ .

Having obtained  $\mathcal{H}_s$ , we now calculate  $\psi^*$  using (55)

$$\begin{aligned} \psi^*(\rho; \xi) &= \mathcal{H}_s\left(\rho; \xi + \frac{1}{2} \mathbf{dS}_\mu(\rho)\right) - \mathcal{H}_s\left(\rho; \frac{1}{2} \mathbf{dS}_\mu(\rho)\right) \\ &= (\xi, \mathcal{L}'_s \rho) + (D\nabla(\mathbf{dS}_\mu(\rho)) \cdot \nabla \xi, \rho) + (D\nabla \xi \cdot \nabla \xi, \rho) \\ &= (\xi, \mathcal{L}'_s \rho) - \left(\xi, \operatorname{div} [\rho D\nabla(\mathbf{dS}_\mu(\rho))]\right) + (D\nabla \xi \cdot \nabla \xi, \rho) \\ &= (D\nabla \xi \cdot \nabla \xi, \rho), \end{aligned}$$

where we have used (90) to obtain the last equality. The Frechet derivative of  $\psi^*$  with respect to its second argument is given by

$$\mathbf{d}_\xi \psi^*(\rho; \xi) = -2 \operatorname{div}(\rho D\nabla \xi),$$

which is equal to  $M_\rho(\xi)$  in (89). Finally, we have

$$W\rho = \mathcal{L}'_a \rho = \operatorname{div}(\rho D\nabla \log \mu - b\rho),$$

which is the same as (88). Hence, the pre-GENERIC formulation (59) of the FP equation obtained from the symmetric–antisymmetric decomposition of the generator is the same as the quadratic pre-GENERIC structure (88) and (89) obtained from the hypocoercive m-KE (19).

**6.1.3. Kinetic FP equation.** We now consider the kinetic FP equation mentioned in the Introduction demonstrating proposition 5.2. This equation is not elliptic but we can still take  $\mathcal{V}$  to be the set of Schwartz functions, see [52]. We recall that the operator  $\mathcal{L}$  for the kinetic FP equation is given by (here in multidimensional setting)

$$(\mathcal{L}\xi)(x, v) = v \cdot \nabla_x \xi - \nabla V(x) \cdot \nabla_v \xi - v \cdot \nabla_v \xi + \Delta_v \xi = \mathcal{J} \nabla H \cdot \nabla \xi + \operatorname{div}_p \left[ \rho \nabla_p(\mathbf{dS}_\mu(\rho)) \right],$$

where  $\mathcal{S}_\mu$  is the relative entropy (8) and

$$\mathcal{J} := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad H(x, v) := V(x) + \frac{1}{2} |v|^2.$$

The kinetic FP equation is a special case of (84) with

$$b(x, v) = \begin{pmatrix} v \\ -\nabla V(x) - v \end{pmatrix}, \quad \sigma = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

The operator  $\mathcal{L}$  is reversible with respect to  $d\mu(x, v) = Z^{-1} e^{-(V + \frac{1}{2}|v|^2)} dx dv$  up to  $\mathcal{F}(x, v) = (x, -v)$  [36]. According to (91), the Hamiltonian  $\mathcal{H} : \mathcal{D}(\mathcal{H}) \rightarrow \mathbb{R}$  is given by

$$\mathcal{H}(\rho, \xi) = \int (\mathcal{L}\xi + |\nabla_p \xi|^2) \rho = (\mathcal{L}' \rho, \xi) + (|\nabla_p \xi|^2, \rho). \tag{92}$$

We now verify directly the relation (77) using (92). In fact, the LHS of (77) is

$$\begin{aligned} \mathcal{H}(\rho; \mathbf{dS}_\mu(\rho) + \xi) &= (\mathcal{L}' \rho, \mathbf{dS}_\mu(\rho) + \xi) + (|\nabla_p(\mathbf{dS}_\mu(\rho) + \xi)|^2, \rho) \\ &= \left( -\mathcal{J} \nabla H \cdot \nabla \rho + \operatorname{div}_p \left[ \rho \nabla_p(\mathbf{dS}_\mu(\rho)) \right], \mathbf{dS}_\mu(\rho) + \xi \right) \\ &\quad + (|\nabla_p(\mathbf{dS}_\mu(\rho) + \xi)|^2, \rho) \\ &= -(\mathcal{J} \nabla H \cdot \nabla \rho, \xi) + (\nabla_p(\mathbf{dS}_\mu(\rho)) \cdot \nabla_p \xi, \rho) + (|\nabla_p \xi|^2, \rho). \end{aligned}$$

The RHS of (77) is

$$\begin{aligned} \mathcal{H}(\mathcal{F}_{\#}\rho; -\xi \circ \mathcal{F}) &= (\mathcal{L}'(\mathcal{F}_{\#}\rho), -\xi \circ \mathcal{F}) + (|\nabla_p(-\xi \circ \mathcal{F})|^2, \mathcal{F}_{\#}\rho) \\ &= (\mathcal{J}\nabla H \cdot \nabla(\mathcal{F}_{\#}\rho), \xi \circ \mathcal{F}) + \left(\nabla_p(\mathbf{d}\mathcal{S}_{\mu}(\mathcal{F}_{\#}\rho)) \cdot \nabla_p(\xi \circ \mathcal{F}), \mathcal{F}_{\#}\rho\right) \\ &\quad + (|\nabla_p(\xi \circ \mathcal{F})|^2, \mathcal{F}_{\#}\rho) \\ &= -(\mathcal{J}\nabla H \cdot \nabla\rho, \xi) + \left(\nabla_p(\mathbf{d}\mathcal{S}_{\mu}(\rho)) \cdot \nabla_p\xi, \rho\right) + (|\nabla_p\xi|^2, \rho). \end{aligned}$$

Thus  $\mathcal{H}(\rho; \mathbf{d}\mathcal{S}(\rho) + \xi) = \mathcal{H}(\mathcal{F}_{\#}\rho; -\xi \circ \mathcal{F})$  as expected. Finally, we recall that the GENERIC structure of generalised Langevin equations has been studied in [19].

6.2. Hamiltonian-PDMPs

A PDMP  $\{z_t\}_{t \geq 0}$  on some state space say  $Z$  is a continuous-time stochastic process that evolves as follows: between random times (usually called random events), it evolves according to a deterministic dynamics described by an ODE; random events happen at a rate  $\lambda(z)$ ; if before the random event the value of the process was  $z_{t-}$ , when the random event takes place the process jumps to a new position selected according to a Markov kernel  $Q(z_{t-}; \cdot)$  so that  $z_t \sim Q(z_{t-}; \cdot)$ ; the new position  $z_t$  is then used as the new initial condition to start again the ODE evolution. More precisely, a PDMP is described by three ingredients:

- (a) an ODE with drift  $F : Z \rightarrow Z$ , namely

$$\frac{dz_t}{dt} = F(z_t)$$

where  $F$  is a nice enough function to ensure at least that the Cauchy problem for this ODE is well-posed for every initial datum in  $Z$ ;

- (b) a rate function  $\lambda : Z \rightarrow \mathbb{R}_+$ ;<sup>25</sup>
- (c) a family of Markov kernels  $Q$  on  $Z$ ; that is, for every  $z \in Z$ ,  $Q(z; \cdot)$  is a Markov kernel on  $Z$ .

With this notation in place, the generator of the dynamics is given by

$$(\mathcal{L}f)(z) = b(z) \cdot \nabla_z f(z) + \lambda(z) \int_Z [f(z') - f(z)] Q(z, dz') \tag{93}$$

for sufficiently smooth functions  $f : Z \rightarrow \mathbb{R}$ . The associated FP operator  $\mathcal{L}'$  is

$$(\mathcal{L}'f)(z) = -\nabla_z \cdot (b(z)f(z)) - \lambda(z)f(z) + \int_Z \lambda(z')f(z')Q(z', z)dz', \tag{94}$$

having assumed for simplicity that the kernel  $Q(z; dz')$  has a density which, with abuse of notation, we call  $Q(z, z')dz'$ . Clear references on the matter are [26, 51]. In particular in [26] it is shown how some very popular sampling algorithms are of the above form, with  $Z = \{(x, v) \in \mathbb{R}^d \times \{-1, 1\}^d\}$ ,  $Z = \mathbb{R}^d \times \{\text{some discrete set}\}$  or  $Z = \mathbb{R}^d \times \mathbb{R}^d$ . The latter case is the one we will be considering in our example and we will denote  $z = (x, v)$  a point in  $\mathbb{R}^d \times \mathbb{R}^d$ .

<sup>25</sup> Because of the way the rate function will be employed, we will have that the probability of an event taking place in the interval  $[t, t + \delta]$  is precisely  $\lambda(z_t)\delta + o(\delta)$ .

A classical example which belongs to this setup is the so-called *Andersen thermostat* (see [32, IIA-Sec2]) which, from a sampling perspective, is the Hybrid Monte Carlo algorithm (see [44] and [14, 42] for related Hamiltonian samplers). The Andersen thermostat can be understood as follows: suppose we want to sample from the measure  $\mu$  defined in (2), or, more precisely, from its multidimensional version,

$$\mu(x, v) = e^{-V(x)} e^{-|v|^2/2}, \tag{95}$$

where the normalisation constant has now been included in the potential  $V: \mathbb{R}^d \rightarrow \mathbb{R}$ . One way of doing so is to employ a deterministic Hamiltonian dynamics, i.e. a dynamics with generator

$$\mathcal{L}_H f := v \cdot \nabla_x f(x, v) - \nabla_x V(x) \cdot \nabla_v f(x, v)$$

which is just the Liouville operator associated with the Hamiltonian  $H: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ <sup>26</sup>  $H(x, v) = V(x) + |v|^2/2$ . This dynamics preserves  $\mu$  but it is clearly not ergodic on  $Z$ ; however it can be modified by resampling the momentum variable according to a standard Gaussian on  $\mathbb{R}^d$ ,  $\pi_0 \sim \mathcal{N}(0, I_d)$ , where  $I_d$  is the  $d$ -dimensional identity matrix. The resulting dynamics is the Andersen thermostat, which evolves according to the following generator:

$$\mathcal{L}_{AT} f := \mathcal{L}_H f + \mathcal{L}_Q f \tag{96}$$

where

$$(\mathcal{L}_Q f)(x, v) := \lambda_r [(\tilde{Q}f)(x, v) - f(x, v)]$$

with the constant  $\lambda_r > 0$  being the rate at which the momentum variable is resampled and

$$(\tilde{Q}f)(x, v) := \int_{\mathbb{R}^d} f(x, \xi) d\pi_0(\xi).$$

The dual  $\mathcal{L}'_Q$  of  $\mathcal{L}_Q$  is often referred to as the BGK collision operator, see [16]; the dual of  $\mathcal{L}_{AT}$ ,  $\mathcal{L}'_{AT}$ , is given by

$$\mathcal{L}'_{AT} = -\mathcal{L}_H + \mathcal{L}'_Q$$

with  $\mathcal{L}'_Q$  as in lemma 6.2 below; such an operator can be put in pre-GENERIC form (34) (see [32, 33]) by setting  $W(f) = -\mathcal{L}_H f$  and using the dissipation potential

$$\psi_f^*(\xi) = \frac{\lambda_r}{(2\pi)^d} \iiint dx dv dv' [\cosh(\xi(x, v') - \xi(x, v)) - 1] e^{-\frac{|v|^2 + |v'|^2}{4}} \sqrt{f(x, v)f(x, v')}$$

and the relative entropy  $\mathcal{S}_\mu(f)$  in (41) as entropy functional; for  $\mu$  as in (95),  $\mathcal{S}_\mu(f)$  becomes

$$\begin{aligned} \mathcal{S}(f) = \mathcal{S}_\mu(f) &= \iint dx dv f(x, v) \log \left( \frac{f(x, v)}{\mu(x, v)} \right) \\ &\stackrel{(41)}{=} \iint f \log f dx dv + \iint f \left( V(x) + \frac{|v|^2}{2} \right) dx dv. \end{aligned} \tag{97}$$

In particular,  $-\mathcal{L}_H$  corresponds to the ‘non gradient-flow’ part of the dynamics, while, with  $\psi^*$  and  $\mathcal{S}_\mu$  as in the above,  $\mathcal{L}'_Q$  is in gradient flow form:

$$\mathcal{L}'_Q f = \mathbf{d}_\xi \psi^* \left( f; -\frac{\mathbf{d}\mathcal{S}_\mu(f)}{2} \right). \tag{98}$$

<sup>26</sup> This is a classical finite-dimensional Hamiltonian function, i.e. here we do not use the word *Hamiltonian* with the same meaning as for the functional  $\mathcal{H}$  introduced in equation (71).

It is straightforward to verify that the degeneracy condition (35) is satisfied, i.e.

$$(W(f), \mathbf{dS}_\mu(f)) = (-\mathcal{L}_H f, \mathbf{dS}_\mu(f)) = 0 \tag{99}$$

so that the entropy (97) is dissipated along the Andersen thermostat flow  $\partial_{t_i} f = \mathcal{L}'_{AT} f$ .

**Note 6.1.** We note now in passing that if in the expression (97) one replaces the potential  $V(x)$  with any other (suitable) function of  $x$  only,  $\tilde{V} = \tilde{V}(x)$ <sup>27</sup> then the representation (98) still holds, and  $\mathcal{L}_{AT}$  can still be cast in pre-GENERIC form (34), but the degeneracy condition (99) holds iff  $\nabla_x \tilde{V}(x) = \nabla_x V(x)$ , coherently with the fact that the dynamics converges to the measure  $\mu$ .

**6.2.1. Casting Hamiltonian PD-MCMC in pre-GENERIC form.** If the Hamiltonian flow is too complicated to simulate numerically, one may wish to use an auxiliary potential  $\tilde{V}(x)$  and the associated Liouville operator

$$(\tilde{\mathcal{L}}_H f)(x, v) := v \cdot \nabla_x f(x, v) - \nabla_x \tilde{V}(x) \cdot \nabla_v f(x, v).$$

However the generator  $\tilde{\mathcal{L}}_H + \mathcal{L}_Q$  does not preserve  $\mu$  (it preserves  $\tilde{\mu} = e^{-\tilde{V}(x)} \times \pi_0(v)$  instead). To correct for this bias one can introduce a further jump process, which restores the invariance of  $\mu$ . The resulting dynamics is a so-called Hamiltonian PD-MCMC, see [51, section 2.4], with generator given by

$$\mathcal{L}f(x, v) = (\tilde{\mathcal{L}}_H f)(x, v) + (\mathcal{L}_Q f)(x, v) + (\mathcal{L}_R f)(x, v) \tag{100}$$

where

$$(\mathcal{L}_R f)(x, v) := \tilde{\lambda}(x, v)[f(x, R_x^{\tilde{U}} v) - f(x, v)],$$

$\tilde{U}(x) = V(x) - \tilde{V}(x)$  and

$$\tilde{\lambda}(x, v) := (v \cdot \nabla_x \tilde{U})_+, \quad R_x^{\tilde{U}} := v - \frac{2(v \cdot \nabla_x \tilde{U})}{\|\nabla_x \tilde{U}\|^2} \nabla_x \tilde{U}. \tag{101}$$

As customary, in the above  $c_+$  denotes the positive part of  $c \in \mathbb{R}$  and  $\cdot$  is just scalar product in  $\mathbb{R}^d$ . For future use we also set

$$a(x, v) = (v \cdot \nabla_x \tilde{U}). \tag{102}$$

The rate of decay to equilibrium in  $\hat{L}_\mu^2 := \{f \in L_\mu^2 : \int f d\mu = 0\}$  for the dynamics generated by (100) can be studied via hypocoercive techniques; this has been done in detail in [3] (see also references therein). Using [3, theorem 1]<sup>28</sup> one can show the following: if the potential  $V(x)$  satisfies assumption **H.1** in [3] and the auxiliary potential  $\tilde{V}$  is such that<sup>29</sup>

$$|\nabla_x \tilde{V}(x)| \leq b(1 + |\nabla_x V(x)|), \quad \text{for some } b > 0,$$

then there exist constants  $C, c > 0$  such that

$$\|e^{t\mathcal{L}} f\|_{L_\mu^2} \leq C e^{-ct} \|f\|_{L_\mu^2}, \quad \text{for every } f \in \hat{L}_\mu^2.$$

<sup>27</sup> Suitable in the sense that the integral in (97) still makes sense.

<sup>28</sup> The generator we are studying is a particular case of the generator  $\mathcal{L}_2$  defined in [3, equation (3)]. To obtain our generator from  $\mathcal{L}_2$  in [3] take  $m_2 = 1, \lambda_{ref}(x) = \lambda_{ref}, F_0(x) = \nabla_x \tilde{V}, F_1(x) = \nabla_x \tilde{U}, K = 1, n_1(x) = \nabla_x \tilde{U} / \|\nabla_x \tilde{U}\|, \lambda_1(x, v) = \tilde{\lambda}(x, v), (\mathcal{B}_1 f)(x, v) = f(x, R_x^{\tilde{U}} v)$  and  $(\Pi_1 f)(x, v) = (\tilde{Q} f)(x, v)$ . In doing so bear in mind that their  $U_0$  is our  $\tilde{V}$  and their  $U$  is our  $V$ .

<sup>29</sup> This is to satisfy [3, assumption **H.2** (c)].

Note that the functional framework for PDMPs is rather involved, and partly still to be developed as generators of PDMPs do not enjoy smoothing properties [3, 20]. However, it has been shown in [3] that, under the above assumptions the operator  $\mathcal{L}$  generates a strongly continuous contraction semigroup in  $L^2_\mu$  and that the set  $C^2_b$  (of bounded functions with bounded derivatives up to order two) is a core for  $\mathcal{D}(\mathcal{L}) \cap \mathcal{D}(\mathcal{L}^*)$ , hence throughout this section we will assume that  $\mathcal{L}$  and  $\mathcal{L}^*$  act on such a set. As domain for  $\mathcal{L}'$  (and all dual operators) we will instead consider the set  $\mathcal{V}$  of positive Schwartz functions.

**Lemma 6.2.** *The  $L^2$  and  $L^2_\mu$  (formal) adjoint of the generator  $\mathcal{L}$  in (100) are given, respectively, by*

$$\begin{aligned} \mathcal{L}' &= \tilde{\mathcal{L}}'_H + \mathcal{L}'_R + \mathcal{L}'_Q \\ \mathcal{L}^* &= \tilde{\mathcal{L}}^*_H + \mathcal{L}^*_R + \mathcal{L}^*_Q, \end{aligned}$$

where

$$\begin{aligned} (\mathcal{L}'_Q f)(x, v) &= \lambda_r \pi_0(v) \left( \int f(x, w) \, dw \right) - \lambda_r f(x, v), \\ (\mathcal{L}'_R f)(x, v) &= \tilde{\lambda}(x, v) [f(x, R_x^{\tilde{U}} v) - f(x, v)] - (v \cdot \nabla_x \tilde{U}) f(x, R_x^{\tilde{U}} v) \\ &= -(v \cdot \nabla_x \tilde{U})_+ f(x, R_x^{\tilde{U}} v) - \tilde{\lambda}(x, v) f(x, v) \end{aligned} \tag{103}$$

and

$$\tilde{\mathcal{L}}'_H = -\tilde{\mathcal{L}}_H, \quad \mathcal{L}^*_R = \mathcal{L}'_R.$$

Moreover,

$$\mathcal{L}^*_Q = \mathcal{L}_Q \quad \text{and} \quad \tilde{\mathcal{L}}^*_H f = -\tilde{\mathcal{L}}_H f + (v \cdot \nabla_x \tilde{U}) f.$$

**Proof of lemma 6.2.** We show how to calculate  $\mathcal{L}'_R$ , the rest is obtained with similar calculations:

$$\int dx dv (\mathcal{L}_R f)(x, v) g(x, v) = \int dx dv \tilde{\lambda}(x, v) [f(x, R_x^{\tilde{U}} v) - f(x, v)] g(x, v).$$

Now make change of variable  $v' = R_x^{\tilde{U}} v$  and then observe that transformation  $R_x^{\tilde{U}} v$  is involutive, so that  $R_x^{\tilde{U}} v' = v$ , and the Jacobian of the transformation is -1. We will also use below the identity

$$a(x, R_x^{\tilde{U}} v) = (R_x^{\tilde{U}} v \cdot \nabla_x \tilde{U}) = -(v \cdot \nabla_x \tilde{U}) = -a(x, v).$$

With this in mind,

$$\begin{aligned} \int dx dv (\mathcal{L}_R f)(x, v) g(x, v) &= \int \tilde{\lambda}(x, v) f(x, R_x^{\tilde{U}} v) g(x, v) \, dx dv - \int \tilde{\lambda}(x, v) f(x, v) g(x, v) \, dx dv \\ &= \int \tilde{\lambda}(x, R_x^{\tilde{U}} v') f(x, v') g(x, R_x^{\tilde{U}} v') \, dx dv' - \int \tilde{\lambda}(x, v) f(x, v) g(x, v) \, dx dv \\ &= \int -(v \cdot \nabla_x \tilde{U})_+ f(x, v) g(x, R_x^{\tilde{U}} v) \, dx dv - \int \tilde{\lambda}(x, v) f(x, v) g(x, v) \, dx dv. \end{aligned}$$

This gives the second expression for  $\mathcal{L}'_R$ . The first is found by observing that  $(-c)_+ = c_- = c_+ - c$ , for every  $c \in \mathbb{R}$ . To find the expression for  $\mathcal{L}^*_R$  one also needs to use

$$|v|^2 = \left| R_x^{\tilde{U}} v \right|^2, \quad \text{for every } v \in \mathbb{R}^d. \tag{104}$$

□

Using the above lemma, the symmetric and antisymmetric parts (in  $L^2_\mu$ ) of the operator (100) are

$$\begin{aligned}
 (\mathcal{L}_S f)(x, v) &= \frac{\mathcal{L} + \mathcal{L}^*}{2} f(x, v) \\
 &= (\mathcal{L}_Q f)(x, v) + \frac{1}{2} ((a_-(x, v)) + a_+(x, v)) \left[ f(x, R_x^{\tilde{U}} v) - f(x, v) \right]
 \end{aligned}
 \tag{105}$$

and

$$(\mathcal{L}_A f)(x, v) = \frac{\mathcal{L} - \mathcal{L}^*}{2} f(x, v) = (\tilde{\mathcal{L}}_H f)(x, v) + \mathcal{L}_{J_A} f,
 \tag{106}$$

where

$$\mathcal{L}_{J_A} f := \frac{1}{2} a(x, v) \left[ f(x, R_x^{\tilde{U}} v) - f(x, v) \right],
 \tag{107}$$

and, for the sake of clarity, we iterate that  $a_+(x, v)$  and  $a_-(x, v)$  are, respectively, the positive and negative parts of the function  $a(x, v)$ . We now come to show that the operator  $\mathcal{L}' = \mathcal{L}'_A + \mathcal{L}'_S$  can be put in pre-GENERIC form (34). To this end, let

$$\mathcal{S}_{\hat{V}}(f) = \iint dx dv f(x, v) \log f(x, v) + \iint dx dv f(x, v) \left( \hat{V}(x) + \frac{|v|^2}{2} \right),
 \tag{108}$$

and note that, when  $\hat{V}(x) = V(x)$  the above entropy functional coincides with  $\mathcal{S}_\mu$ . Then the following holds.

**Proposition 6.3.** *Let  $\mathcal{L}$  be the generator of the Hamiltonian PD-MCMC introduced in (100). Then, for any potential  $\hat{V} = \hat{V}(x)$  such that the second addend of (108) is integrable, the operator  $\mathcal{L}'$  can be written in the form*

$$\mathcal{L}' \rho = \tilde{W}(\rho) + \mathbf{d}_\xi \tilde{\psi}^* \left( \rho; -\frac{\mathbf{d}\mathcal{S}_{\hat{V}}(\rho)}{2} \right)
 \tag{109}$$

where  $\tilde{W}(\rho) = \mathcal{L}'_A \rho$  and, for every  $\rho \in \mathcal{V}$ ,  $\tilde{\psi}^*(\rho; \xi) = \psi^*(\rho; \xi) + \hat{\psi}^*(\rho; \xi)$  with  $\psi^*$  as in (98) and  $\hat{\psi}^*$  given by

$$\hat{\psi}^*(\rho; \xi) := \frac{1}{2} \iint dx dv (a_+ + a_-)(x, v) \sqrt{\rho(x, R_x^{\tilde{U}} v) \rho(x, v)} \left[ \cosh \left( \xi(x, R_x^{\tilde{U}} v) - \xi(x, v) \right) - 1 \right],
 \tag{110}$$

is positive and convex (in the second argument).

Before moving on to the proof, let us point out that the above proposition is not making any statements about the degeneracy condition (35); we will comment on the validity (rather, non validity) of such a condition in note 6.2.3 below.

**Proof of proposition 6.3.** Let us start by decomposing the operator  $\mathcal{L}_S$ :

$$\mathcal{L}_S f = \mathcal{L}_Q f + \mathcal{L}_{J_S} f$$

where

$$(\mathcal{L}_{J_S} \rho)(x, v) = \frac{1}{2} (a(x, v)_+ + a(x, v)_-) \left[ \rho(x, R_x^{\tilde{U}} v) - \rho(x, v) \right].$$



Therefore  $\mathcal{L}'_S = \mathcal{L}'_Q + \mathcal{L}'_{JS}$  as  $\mathcal{L}'_{JS} = \mathcal{L}'_{JS}$ ; then, by (98) and note 6.2, to show (109) one just needs to observe the following

$$(\mathcal{L}_{JS}f)(x, v) = \mathbf{d}_\xi \hat{\psi}^* \left( f; -\frac{\mathbf{d}\mathcal{S}_{\hat{V}}(f)}{2} \right). \tag{111}$$

To see that the above holds, let us first calculate the variational derivative of  $\hat{\psi}^*$ : for every  $\eta \in \mathcal{V}$ ,

$$\begin{aligned} (\mathbf{d}_\xi \hat{\psi}^*(\rho; \xi), \eta) &= \left. \frac{d}{dt} \hat{\psi}^*(\rho; \xi + t\eta) \right|_{t=0} \\ &= \frac{1}{2} \iint (a_+ + a_-)(x, v) \sqrt{\rho(x, R_x^{\tilde{U}}v)\rho(x, v)} \sinh \left( \xi(x, R_x^{\tilde{U}}v) - \xi(x, v) \right) \\ &\quad \times \left( \eta(x, R_x^{\tilde{U}}v) - \eta(x, v) \right) dx dv. \end{aligned}$$

We want to calculate the above expression at  $\xi = -\frac{1}{2}\mathbf{d}\mathcal{S}_{\hat{V}}(\rho)$ . In particular we need to calculate the difference

$$\xi(x, R_x^{\tilde{U}}v) - \xi(x, v) = -\frac{1}{2} \left( \mathbf{d}\mathcal{S}_{\hat{V}}(\rho)(x, v) - \mathbf{d}\mathcal{S}(\rho)(x, v) \right).$$

It is in doing so that one can see that the specific choice of potential  $\hat{V}$  does not matter, as the potential is cancelled in the difference. Indeed, since

$$\mathbf{d}\mathcal{S}_{\hat{V}}(\rho)(x, v) = \log \rho(x, v) + \hat{V}(x) + \frac{|v|^2}{2} + 1,$$

from (104), we have

$$-\frac{1}{2} \left( \mathbf{d}\mathcal{S}_{\hat{V}}(\rho)(x, R_x^{\tilde{U}}v) - \mathbf{d}\mathcal{S}_{\hat{V}}(\rho)(x, v) \right) = \log \left( \sqrt{\frac{\rho(x, v)}{\rho(x, R_x^{\tilde{U}}v)}} \right).$$

Therefore, using the identity  $(a_+ + a_-)(x, v) = (a_+ + a_-)(x, R_x^{\tilde{U}}v)$ , we obtain

$$\begin{aligned} (\mathbf{d}_\xi \hat{\psi}^*(\rho; -\frac{1}{2}\mathbf{d}\mathcal{S}(\rho)), \eta) &= \frac{1}{4} \int (a_+ + a_-)(x, v) \rho(x, v) \left( \eta(x, R_x^{\tilde{U}}v) - \eta(x, v) \right) dx dv \\ &\quad - \frac{1}{4} \int (a_+ + a_-)(x, v) \rho(x, R_x^{\tilde{U}}v) \left( \eta(x, R_x^{\tilde{U}}v) - \eta(x, v) \right) dx dv \\ &= \frac{1}{2} \int (a_+ + a_-)(x, v) \rho(x, v) \left( \eta(x, R_x^{\tilde{U}}v) - \eta(x, v) \right) dx dv \\ &= \frac{1}{2} \int (a_+ + a_-)(x, v) \left( \rho(x, R_x^{\tilde{U}}v) - \rho(x, v) \right) \eta(x, v) dx dv, \end{aligned}$$

which is the desired result. □

**6.2.2. Relation to section 4.** In proposition 6.3 the functional  $\tilde{\psi}^*$  appeared as an ansatz. We show here that  $\tilde{\psi}^*$  does indeed coincide with the dissipation potential that we would obtain using the procedure of section 4, theorem 4.1 in particular. First calculate the Hamiltonian

$$\begin{aligned} \mathfrak{H}_s(\rho; \xi) &:= \iint e^{-\xi(x, v)} (\mathcal{L}_S e^\xi)(x, v) \rho(x, v) dx dv \\ &= \iint e^{-\xi(x, v)} (\mathcal{L}_Q e^\xi)(x, v) \rho(x, v) dx dv + \iint e^{-\xi(x, v)} (\mathcal{L}_{JS} e^\xi)(x, v) \rho(x, v) dx dv. \end{aligned}$$

For general purposes, for every  $\rho \in \mathcal{V}$ , one can take  $\mathcal{D}_\rho(\mathfrak{H}_s)$  to be the set of smooth functions which grow at most quadratically in  $v$  and at most like  $V(x)$  in the space variable  $x$ . We work on the second addend of the above, which will give  $\hat{\psi}^*$ ; applying a similar procedure to the first addend gives  $\psi^*$ .

$$\begin{aligned} \hat{\mathfrak{H}}_s(\rho; \xi) &:= \iint e^{-\xi(x,v)} (\mathcal{L}_{JS} e^\xi)(x, v) \rho(x, v) dx dv \\ &= \iint e^{-\xi(x,v)} \frac{1}{2} [a_+(x, v) + a_-(x, v)] \left[ e^{\xi(x, R_x^U v)} - e^{\xi(x,v)} \right] \rho(x, v) dx dv. \end{aligned}$$

We want to show that (55) holds for this example, i.e. that the following identity holds:

$$\begin{aligned} \hat{\psi}^*(\rho; \xi) &= \hat{\mathfrak{H}}_s \left( \rho; \xi + \frac{1}{2} \mathbf{dS}_V(f) \right) - \hat{\mathfrak{H}}_s \left( \rho; \frac{1}{2} \mathbf{dS}_V(\rho) \right) \\ &= \iint \frac{1}{2} [a_+(x, v) + a_-(x, v)] \left[ e^{\xi(x, R_x^U v) - \xi(x,v)} - 1 \right] \sqrt{\rho(x, v) \rho(x, R_x^U v)} dx dv. \end{aligned}$$

Indeed, by observing that the above integral does not change under the change of variable  $v' = R_x^U v$ , one can easily see that the above does indeed coincide with the expression for  $\hat{\psi}^*$  given in (110).

**6.2.3. Relative entropy dissipation for Hamiltonian PD-MCMC.** We now want to show that the entropy functional  $\mathcal{S}_\mu$  decays along the flow, so from now on we pick  $\hat{V}(x) = V(x)$  in the representation (109).

**Note 6.4.** Within the pre-GENERIC framework, entropy decay is implied by either the degeneracy condition (35) or the dissipation condition (38). In our case the latter condition would read

$$(\mathcal{L}'_A f, \mathbf{dS}_\mu(f)) \leq 0. \tag{112}$$

Let us calculate explicitly the LHS of the above. The operator  $\mathcal{L}'_A$ , i.e. the  $L^2$  adjoint of the operator  $\mathcal{L}_A$  defined in (106), is

$$(\mathcal{L}'_A f)(x, v) = -\tilde{\mathcal{L}}_H f + \mathcal{L}'_{JA}$$

where

$$\mathcal{L}'_{JA} = -\frac{1}{2} a(x, v) \left[ f(x, R_x^U v) + f(x, v) \right].$$

Because

$$-(\tilde{\mathcal{L}}_H f, \mathbf{dS}_\mu(f)) = \int a(x, v) f(x, v) dx dv,$$

and

$$(\mathcal{L}'_{JA} f, \mathbf{dS}_\mu(f)) = \int (\mathcal{L}'_{JA} f)(x, v) \log f(x, v) dx dv,$$

(the above coming from the fact that  $\mathcal{L}_{JA}(V + \frac{v^2}{2} + 1) = 0$ , where we recall that  $\mathcal{L}_{JA}$  is defined in (107)) one has

$$(\mathbf{dS}_\mu(f), \mathcal{L}'_A f) = \int a(x, v) f(x, v) dx dv + \int (\mathcal{L}'_{JA} f)(x, v) \log f(x, v) dx dv,$$

or, more explicitly,

$$(\mathbf{d}\mathcal{S}_\mu(f), \mathcal{L}'_A f) = \int a(x, v)f(x, v) \, dx \, dv + \frac{1}{2} \int a(x, v)f(x, v) \log \frac{f(x, R_x^{\tilde{U}}v)}{f(x, v)} \, dx \, dv \quad (113)$$

$$= \int a(x, v)f(x, v) \, dx \, dv - \frac{1}{2} \int a(x, v)[f(x, R_x^{\tilde{U}}v) + f(x, v)] \log f(x, v) \, dx \, dv. \quad (114)$$

We can now make some observations on (112)

- Using the expression (113), we can see that the inequality (112) does not hold as a functional inequality, in the sense that it does not hold for every  $f$  (in an appropriate class, to which the counterexample which we are about to exhibit should belong). This is easy to see in one dimension: if  $d = 1$  then  $R_x^{\tilde{U}}v = -v$ ; if we choose  $f$  to be in product form, more specifically,  $f(x, v) = e^{-\tilde{V}(x)}\pi_c(v)$  where  $\pi_c(v)$  is a Gaussian with mean  $c \in \mathbb{R}$  and variance  $\sigma$ ,  $\tilde{V}$  is such that  $e^{-\tilde{V}}$  is normalised in space and  $V, \tilde{V}$  such that  $\nabla_x \tilde{U}(x) = 1$ , then  $a(x, v) = v$  and

$$(\mathbf{d}\mathcal{S}_\mu(f), \mathcal{L}'_A f) = -\frac{c^3}{\sigma},$$

which can have any sign. We note that in order to show entropy decay one needs not prove (112) for every  $f$ , it is sufficient to prove (112) only along the flow, but we have not been able to do so.

- Finally, we observe that (112) holds (with equality) for every function  $f$  which is reflection invariant, i.e. such that  $f(x, R_x^{\tilde{U}}v) = f(x, v)$ . Indeed if this is the case then the second addend in (113) vanishes. For the first, notice that by a change of variables

$$\int a(x, v)f(x, v) \, dx \, dv = - \int a(x, v)f(x, R_x^{\tilde{U}}v) \, dx \, dv,$$

which implies

$$\int a(x, v) \left( f(x, v) + f(x, R_x^{\tilde{U}}v) \right) \, dx \, dv = 0 \quad \text{for every } f;$$

by writing  $f(x, v)$  as  $f(x, v) = (f(x, v) + f(x, R_x^{\tilde{U}}v))/2 + (f(x, v) - f(x, R_x^{\tilde{U}}v))/2$  one then has

$$\int a(x, v)f(x, v) \, dx \, dv = \int a(x, v) \frac{f(x, v) - f(x, R_x^{\tilde{U}}v)}{2} \, dx \, dv \quad \text{for every } f. \quad (115)$$

This suggests that if the flow preserves the invariance of the profile  $f$  under reflections then (112) is true along the flow, assuming  $f_0$  is reflection invariant.

In view of these observations, we prove entropy decay in a different way.

**Proposition 6.5.** *Consider the Hamiltonian PD-MCMC with generator  $\mathcal{L}$  as in (100) and let  $\mu$  be the measure on  $\mathbb{R}^{2d}$  defined in (95). Then the relative entropy  $\mathcal{S}_\mu$  decreases along the solution of the FP equation  $\partial_t f_i(x, v) = (\mathcal{L}' f_i)(x, v)$ .*

**Proof.** In order to prove entropy decay we need to show the following

$$\frac{d\mathcal{S}_\mu(f_t)}{dt} = (\mathbf{d}\mathcal{S}_\mu(f_t), \mathcal{L}'_S f_t) + (\mathbf{d}\mathcal{S}_\mu(f_t), \mathcal{L}'_A f_t) \leq 0.$$

Because  $\mathcal{L}'_S = \mathcal{L}'_Q + \mathcal{L}'_{JS}$  from (98) and (111) we already know

$$(\mathbf{d}\mathcal{S}_\mu, \mathcal{L}'_S f) \leq 0 \quad \text{for all } f;$$

more specifically, thanks to the convexity of the functionals  $\psi^*$  and  $\hat{\psi}^*$ , we know that both

$$(\mathbf{d}\mathcal{S}_\mu(f), \mathcal{L}'_Q f) \leq 0 \tag{116}$$

and

$$(\mathbf{d}\mathcal{S}_\mu(f), \mathcal{L}'_{JS} f) \leq 0. \tag{117}$$

Hence, we know

$$(\mathbf{d}\mathcal{S}_\mu(f), \mathcal{L}'_Q f) = \int \log f(\mathcal{L}'_Q f) dx dv + \lambda_r \int f \left(1 - \frac{v^2}{2}\right) dx dv \leq 0$$

and

$$\begin{aligned} (\mathbf{d}\mathcal{S}_\mu(f), \mathcal{L}'_{JS} f) &= \frac{1}{2} \int |a(x, v)| \log f \left[ f(x, R_x^{\tilde{U}} v) - f(x, v) \right] dx dv \\ &= \frac{1}{2} \int |a(x, v)| \log \left( \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} \right) f(x, v) dx dv \leq 0 \end{aligned}$$

and we want to prove the inequality

$$\begin{aligned} &\int \log f(\mathcal{L}'_Q f) dx dv + \lambda_r \int f \left(1 - \frac{v^2}{2}\right) dx dv + \frac{1}{2} \int |a(x, v)| \log \left( \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} \right) f(x, v) dx dv \\ &+ \int a(x, v) f(x, v) dx dv + \frac{1}{2} \int a(x, v) f(x, v) \log \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} dx dv \leq 0. \end{aligned} \tag{118}$$

In view of (116) and (117), it is sufficient to prove the following inequality, namely

$$\begin{aligned} &\frac{1}{2} \int |a(x, v)| \log \left( \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} \right) f(x, v) dx dv + \int a(x, v) f(x, v) dx dv \\ &+ \frac{1}{2} \int a(x, v) f(x, v) \log \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} dx dv \leq 0. \end{aligned} \tag{119}$$

To this end start by observing that, using (115), the above is equivalent to

$$\begin{aligned} &\frac{1}{2} \int |a(x, v)| \log \left( \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} \right) f(x, v) dx dv + \frac{1}{2} \int a(x, v) \left[ f(x, v) - f(x, R_x^{\tilde{U}} v) \right] dx dv \\ &+ \frac{1}{2} \int a(x, v) f(x, v) \log \frac{f(x, R_x^{\tilde{U}} v)}{f(x, v)} dx dv \leq 0. \end{aligned} \tag{120}$$

In turn, by writing  $a = a_+ - a_-$  and  $|a| = a_+ + a_-$  and observing that for every function  $g$  one has

$$\int a_-(x, v) g(x, v) = \int a_+(x, v) g(x, R_x^{\tilde{U}} v),$$

(120) is equivalent to

$$\int a_+(x, v) [f(x, v) - f(x, R_x^{\tilde{U}})] dx dv + \int a_+(x, v) f(x, v) \log \frac{f(x, R_x^{\tilde{U}})}{f(x, v)} dx dv \leq 0. \tag{121}$$

The above is now easy to prove; indeed, since  $\log u \leq u - 1$ , we have

$$\int a_+(x, v) f(x, v) \log \frac{f(x, R_x^{\tilde{U}})}{f(x, v)} dx dv \leq \int a_+(x, v) [f(x, R_x^{\tilde{U}}) - f(x, v)] dx dv$$

which is precisely (121). □

**Acknowledgment**

The research of M H Duong was supported by the EPSRC Grant EP/V038516/1.

**Appendix A. Basic facts about large deviations**

The notation in this appendix is independent of the notation in the rest of the paper. Let  $(X, d)$  be a metric space,  $\{\chi_n\}_{n \in \mathbb{N}}$  be a sequence of  $X$ -valued random variables and  $\mathcal{I} : X \rightarrow [0, \infty]$  be a lower semicontinuous function. The sequence  $\{\chi_n\}_{n \in \mathbb{N}}$  satisfies a LDP with rate function  $\mathcal{I}$  if the following two conditions are satisfied (see [28, chapter 1]):

- for every open set  $A \subseteq X$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\chi_n \in A) \geq - \inf_{x \in A} \mathcal{I}(x)$$

- for every closed set  $B \subseteq X$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\chi_n \in B) \leq - \inf_{x \in B} \mathcal{I}(x).$$

The rate function  $\mathcal{I}$  is said to be a *good rate function* if for every  $a \in [0, \infty)$  the set  $\{x : \mathcal{I}(x) \leq a\}$  is compact. A short-hand notation to express the above is

$$\text{Prob}(\chi_n \approx \chi) \stackrel{n \rightarrow \infty}{\sim} e^{-n\mathcal{I}(\chi)}.$$

**Appendix B. Some observations about the definition of pre-GENERIC**

For completeness we show here that the definition (34) of pre-GENERIC is equivalent to (36). To this end let us consider the function

$$\varphi(\rho; g) := \varphi_\rho(g) = \psi_\rho(g - W(\rho)) + \psi_\rho^* \left( -\frac{1}{2} \mathbf{d}\mathcal{S}(\rho) \right) + \frac{1}{2} (g, \mathbf{d}\mathcal{S}(\rho)).$$

We look at the above as a function of  $g$ , for every  $\rho$  fixed and we want to prove that  $\varphi_z(g) = 0$  iff  $g = W(z) + (\mathbf{d}_\xi \psi_z^*) \left( -\frac{1}{2} \mathbf{d}\mathcal{S}(z) \right)$ . To this end note that the function  $\varphi$  is the sum of two strictly convex functions plus a linear function and it is therefore strictly convex (in  $g$ ). Moreover, by the Young–Fenchel inequality we have

$$\frac{1}{2} (g, \mathbf{d}\mathcal{S}(\rho)) \leq \psi_\rho(g) + \psi_\rho^* \left( \frac{1}{2} \mathbf{d}\mathcal{S}(\rho) \right),$$

so that, using the symmetry of  $\psi^*$  and the orthogonality condition (35),  $\varphi$  is positive as well and it attains its minimum (zero) iff  $\mathbf{d}_g\varphi_\rho(g) = 0$ . Now,

$$\begin{aligned}\mathbf{d}_v\varphi_\rho(g) &= \mathbf{d}\psi_\rho(g - W(\rho)) + \frac{1}{2}\mathbf{d}\mathcal{S}(\rho) = 0 \\ \Leftrightarrow g &= W(\rho) + (\mathbf{d}_g\psi_\rho)^{-1}\left(-\frac{1}{2}\mathbf{d}\mathcal{S}(\rho)\right) \\ \Leftrightarrow g &= W(\rho) + (\mathbf{d}_\xi\psi_\rho^*)\left(-\frac{1}{2}\mathbf{d}\mathcal{S}(\rho)\right),\end{aligned}$$

having used the fact that  $\psi_\rho = \psi_\rho(g)$  and  $\psi_\rho^* = \psi_\rho^*(\xi)$  are Legendre dual of each other, hence  $(\mathbf{d}_v\psi_\rho)^{-1} = \mathbf{d}_\xi\psi_\rho^*$ .

**Note B.1.** In absence of the orthogonality condition (35) one can observe that the following condition holds

$$\psi_\rho(v - W(\rho)) + \psi_\rho^*\left(-\frac{1}{2}\mathbf{d}\mathcal{S}(\rho)\right) + \frac{1}{2}(v - W(\rho), \mathbf{d}\mathcal{S}(\rho)) = 0$$

if and only if  $v = W(\rho) + (\mathbf{d}_\xi\psi_\rho^*)\left(-\frac{1}{2}\mathbf{d}\mathcal{S}(\rho)\right)$ . To see this, just consider the function

$$\tilde{\varphi}_\rho(g) = \psi_\rho(g - W(\rho)) + \psi_\rho^*\left(-\frac{1}{2}\mathbf{d}\mathcal{S}(\rho)\right) + \frac{1}{2}(g - W(\rho), \mathbf{d}\mathcal{S}(\rho)),$$

which is again convex and positive by Young–Fenchel inequality. The rest of the reasoning is identical to the one done above.

## ORCID iD

M Ottobre  <https://orcid.org/0000-0002-8725-4278>

## References

- [1] Adams S, Dirr N, Peletier M and Zimmer J 2013 Large deviations and gradient flows *Phil. Trans. R. Soc. A* **371** 20120341
- [2] Andrieu C, Dobson P and Wang A Q 2020 Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods (arXiv:2011.09341)
- [3] Andrieu C, Durmus A, Nusken N and Roussel J 2018 Hypocoercivity of piecewise deterministic Markov process-Monte Carlo (arXiv:1808.08592)
- [4] Andrieu C and Livingstone S 2021 Peskun–Tierney ordering for Markovian Monte Carlo: beyond the reversible scenario *Ann. Stat.* **49** 1958–81
- [5] Bakry D, Gentil I and LeDoux M 2013 *Analysis and Geometry of Markov Diffusion Operators* (Berlin: Springer)
- [6] Bertini L, Gabrielli D and Lebowitz J L 2005 Large deviations for a stochastic model of heat flow *J. Stat. Phys.* **121** 843–85
- [7] Bogachev V, Krylov N and Roeckner M 2009 Elliptic and parabolic equations for measures *Russ. Math. Surv.* **64** 973–1078
- [8] Bouchet F 2020 Is the Boltzmann equation reversible? A large deviation perspective on the irreversibility paradox *J. Stat. Phys.* **181** 515–50
- [9] Budhiraja A, Dupuis P and Fischer M 2012 Large deviation properties of weakly interacting processes via weak convergence methods *Ann. Probab.* **40** 74–102
- [10] Da Prato G and Zabczyk J 1996 *Ergodicity for Infinite Dimensional Systems (LMS Lecture Notes Series)* (Cambridge: Cambridge University Press)
- [11] Davis M H A 1984 Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models *J. R. Stat. Soc. B* **46** 353–76

- [12] Dawson D A and Gärtner J 1987 Large deviations from the McKean–Vlasov limit for weakly interacting diffusions *Stochastics* **20** 247–308
- [13] Dembo A and Zeitouni O 2010 *Large Deviations: Techniques and Applications* (Berlin: Springer)
- [14] Dobson P, Fursov I, Lord G and Ottobre M 2020 Reversible and non-reversible Markov chain Monte Carlo algorithms for reservoir simulation problems *Comput. Geosci.* **24** 1301–13
- [15] Dolbeault J, Mouhot C and Schmeiser C 2009 Hypocoercivity for kinetic equations with linear relaxation terms *C. R. Math.* **347** 511–6
- [16] Dolbeault J, Mouhot C and Schmeiser C 2015 Hypocoercivity for linear Kinetic equations conserving mass *Trans. Am. Math. Soc.* **367** 3807–28
- [17] Duncan A, Lelièvre T and Pavliotis G 2016 Variance reduction using nonreversible Langevin samplers *J. Stat. Phys.* **163** 457–91
- [18] Duong M H, Peletier M A and Zimmer J 2013 GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large deviation principle *Nonlinearity* **26** 2951–71
- [19] Duong M H and Pavliotis G 2018 Mean field limits for non-Markovian interacting particles: convergence to equilibrium, GENERIC formalism, asymptotic limits and phase transitions (arXiv:1805.04959)
- [20] Durmus A, Guillin A and Monmarche P 2021 Piecewise deterministic Markov processes and their invariant measure *Ann. Inst. Henri Poincaré B* **57** 1442–75
- [21] Eckmann J P and Hairer M 2000 Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators *Commun. Math. Phys.* **212** 105–64
- [22] Ekeland I and Temam R 1976 *Convex Analysis and Variational Problems* (Amsterdam: North-Holland)
- [23] Faggionato A, Gabrielli D and Ribezzi Crivellari M 2009 Non-equilibrium thermodynamics of piecewise deterministic Markov processes *J. Stat. Phys.* **137** 259
- [24] Faggionato A, Gabrielli D and Ribezzi Crivellari M 2008 Averaging and large deviation principles for fully-coupled piecewise deterministic Markov processes and applications to molecular motors *Markov Process. Relat. Fields* **16** 497–548
- [25] Faggionato A Gabrielli D 2012 A representation formula for large deviations rate functionals of invariant measures on the one dimensional torus *Ann. Inst. Henri Poincaré B* **48** 212–34
- [26] Fearnhead P, Bierkens J, Pollock M and Roberts G O 2018 Piecewise deterministic Markov processes for continuous-time Monte carlo *Stat. Sci.* **33** 386–412
- [27] Feng S 1994 Large deviations for empirical process of mean-field interacting particle system with unbounded jumps *Ann. Probab.* **22** 2122–51
- [28] Feng J and Kurtz T G 2006 *Large Deviations for Stochastic Processes (AMS Mathematical Surveys and Monographs)* (Providence, RI: American Mathematical Society)
- [29] Guionnetand A and Zegarlinski B 2002 Lectures on logarithmic Sobolev inequalities *Seminaire de Probabilités de Strasbourg* (Berlin: Springer)
- [30] Hairer M 2009 How can a heat bath get *Commun. Math. Phys.* **292** 131–77
- [31] Herau F 2007 Short and long time behavior of the Fokker–Planck equation in a confining potential and applications *J. Funct. Anal.* **244** 95–118
- [32] Kraaij R C, Lazarescu A, Maes C and Peletier M 2018 Deriving GENERIC from a generalised fluctuation symmetry *J. Stat. Phys.* **170** 492–508
- [33] Kraaij R C, Lazarescu A, Maes C and Peletier M 2020 Fluctuation symmetry leads to generic equations with non-quadratic dissipation *Stoch. Process. Appl.* **130** 139–70
- [34] Kipnis C, Olla S and Varadhan S R S 1989 Hydrodynamics and large deviations for simple exclusion processes *Commun. Pure. Appl. Math.* **42** 115–37
- [35] Kipnis C and Olla S 1990 Large deviations from the hydrodynamical limit for a system of independent Brownian particles *Stoch. Stoch. Rep.* **33** 17–25
- [36] Lelièvre T, Rousset M and Stoltz G 2010 *Free Energy Computations* (London: Imperial College Press)
- [37] Mielke A 2011 Formulation of thermoelastic dissipative material behavior using GENERIC *Contin. Mech. Thermodyn.* **23** 233–56
- [38] Mielke A, Peletier M A and Renger D R M 2014 Relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion *Potential Anal.* **41** 1293–327
- [39] Onsager L and Machlup S 1953 Fluctuations and irreversible processes *Phys. Rev.* **91** 1505–12
- [40] Ottinger H C 2005 *Beyond Equilibrium Thermodynamics* (New York: Wiley-Interscience)
- [41] Ottobre M 2016 Markov chain Monte Carlo and irreversibility *Rep. Math. Phys.* **77** 267–92

- [42] Ottobre M, Pillai N S, Pinski F J, Stuart A M 2016 A function space HMC algorithm with second order Langevin diffusion limit (Bernoulli) (arXiv:[1308.0543](#))
- [43] Pardoux E and Yu Veretennikov A 2001 On the Poisson equation and diffusion approximation *Ann. Probab.* **29** 1061–85
- [44] Neal R M 2011 MCMC using Hamiltonian dynamics *Handbook of Markov Chain Monte Carlo* (Boca Raton, FL: Chapman & Hall/CRC)
- [45] Pavliotis G A 2014 *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations (Texts in Applied Mathematics)* (New York: Springer)
- [46] Pazy A 1983 *Semigroups of Linear Operators and Applications to Partial Differential Equations* (Berlin: Springer)
- [47] Peletier M A, Redig F and Vafayi K 2014 Large deviations in stochastic heat conduction processes provide a gradient flow structure for heat conduction *J. Math. Phys.* **55** 093301
- [48] Penrose O 1979 Foundations of statistical mechanics *Rep. Prog. Phys.* **42** 1937
- [49] Talay D and Tubaro L (eds) 1995 *Probabilistic Models for Nonlinear Partial Differential Equations* (Berlin: Springer)
- [50] Reina C and Zimmer J 2015 Entropy production and the geometry of dissipative evolution equations *Phys. Rev. E* **92** 052117
- [51] Vanetti P, Bouchard-Cote A, Deligiannidis G and Doucet A Piecewise deterministic Markov chain Monte Carlo (arXiv:[1707.05296](#))
- [52] Villani C 2009 Hypocoercivity *Mem. Am. Math. Soc.* **202** 141
- [53] Villani C 2002 A review of mathematical topics in collisional kinetic theory *Handbook of Mathematical Fluid Dynamics* (Amsterdam: Elsevier Science)
- [54] Yaglom A M 1949 On the statistical reversibility of Brownian motion *Mat. Sb.* **24** 457–92
- [55] Renger D R M and Zimmer J 2021 Orthogonality of fluxes in general nonlinear reaction networks *Discrete Contin. Dyn. Syst. S* **14** 205–17
- [56] Kaiser M, Jack R L and Zimmer J 2018 Canonical structure and orthogonality of forces and currents in irreversible Markov chains *J. Stat. Phys.* **170** 1019–50
- [57] Jordan R, Kinderlehrer D and Otto F 1998 The variational formulation of the Fokker–Planck equation *SIAM J. Math. Anal.* **29** 1–17
- [58] Maas J 2011 Gradient flows of the entropy for finite Markov chains *J. Funct. Anal.* **261** 2250–92
- [59] Chow S-N, Huang W, Li Y and Zhou H 2012 Fokker–Planck equations for a free energy functional or Markov process on a graph *Arch. Ration. Mech. Anal.* **203** 969–1008
- [60] Mielke A 2011 A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems *Nonlinearity* **24** 1329–46
- [61] Ambrosio L, Gigli N and Savaré G 2008 *Gradient Flows: In Metric Spaces and in the Space of Probability Measures (Lectures in Mathematics)* 2nd edn (Basel: ETH Zürich)
- [62] Patterson R I A, Renger D R M and Sharma U 2021 Variational structures beyond gradient flows: a macroscopic fluctuation-theory perspective (arXiv:[2103.14384](#))