ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ - ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
Archimedes Center for Modeling, Analysis & Computation
UNIVERSITY OF CRETE - DEPARTMENT OF APPLIED MATHEMATICS
Archimedes Center for Modeling, Analysis & Computation

ACMAC's PrePrint Repository

Measuring the irreversibility of numerical schemes for reversible stochastic differential equations

Markos A. Katsoulakis and Yannis Pantazis and Luc Rey-Bellet

Original Citation:

Katsoulakis, Markos A. and Pantazis, Yannis and Rey-Bellet, Luc

(2012)

Measuring the irreversibility of numerical schemes for reversible stochastic differential equations.

(Submitted)

This version is available at: http://preprints.acmac.uoc.gr/209/
Available in ACMAC's PrePrint Repository: March 2013

ACMAC's PrePrint Repository aim is to enable open access to the scholarly output of ACMAC.

http://preprints.acmac.uoc.gr/

Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

MEASURING THE IRREVERSIBILITY OF NUMERICAL SCHEMES FOR REVERSIBLE STOCHASTIC DIFFERENTIAL EQUATIONS *

Markos Katsoulakis[1], Yannis Pantazis[1] and Luc Rey-Bellet[1]

Abstract. For a Markov process the detailed balance condition is equivalent to the time-reversibility of the process. For stochastic differential equations (SDE's) time discretization numerical schemes usually destroy the property of time-reversibility. Despite an extensive literature on the numerical analysis for SDE's, their stability properties, strong and/or weak error estimates, large deviations and infinite-time estimates, no quantitative results are known on the lack of reversibility of the discrete-time approximation process. In this paper we provide such quantitative estimates by using the concept of entropy production rate, inspired by ideas from non-equilibrium statistical mechanics. The entropy production rate for a stochastic process is defined as the relative entropy (per unit time) of the path measure of the process with respect to the path measure of the time-reversed process. By construction the entropy production rate is nonnegative and it vanishes if and only if the process is reversible. Crucially, from a numerical point of view, the entropy production rate is an *a posteriori* quantity, hence it can be computed in the course of a simulation as the ergodic average of a certain functional of the process (the so-called Gallavotti-Cohen (GC) action functional). We compute the entropy production for various numerical schemes such as explicit Euler-Maruyama and explicit Milstein's for reversible SDEs with additive or multiplicative noise. Additionally, we analyze the entropy production for the BBK integrator of the Langevin processes. We show that entropy production is an observable that distinguishes between different numerical schemes in terms of their discretization-induced irreversibility. Furthermore, our results show that the type of the noise critically affects the behavior of the entropy production rate.

Keywords and phrases: Stochastic differential equations, Detailed Balance, Reversibility, Relative Entropy, Entropy production, Numerical integration, (overdamped) Langevin processes.

* *We thanks Natesh Pillai for useful comments and suggestions. M. A. K. and Y.P. are partially supported by NSF-CMMI 0835673 and L. R.-B. is partially supported by NSF (DMS-1109316)*

[1] Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA, USA.

© EDP Sciences, SMAI 1999

Résumé. Pour un processus de Markov la condition de balance détaillée est équivalente à la reversibilité du processus par rapport au renversement du temps. Pour les équations différentielles stochastiques, les schémas de discrétisation détruisent en général cette proprieté de reversibilité. En dépit d'une vaste littérature sur l'analyse numérique des équations differentielles stochastiques, leur proprieté de stabilité, les erreurs fortes et/ou faibles, les proprietés de grandes déviations et à long temps, il n'y a pas eu jusqu'à maintenant de résultats quantitatifs sur l'irréversibilité introduite par l'approximation numérique. Dans cet article nous fournissons de telles estimations, en nous basant sur le taux de production d'entropie, inspirés par des idées de mécanique statistique hors-équilibre. Le taux de production d'entropie est, par définition, l'entropie relative (par unité de temps) du processus par rapport au processus renversé en temps. Par construction, le taux de production d'entropie est non-négatif et il est zéro si et seulement si le procesus est réversible. Crucialement, d'un point de vue numérique, le taux de production d'entropie peut être evalué directement comme la moyenne ergodique d'une certaine fonctionnelle du processus (la fonctionelle de Gallavotti-Cohen), sous des conditions d'ergodicité adéquates. Nous calculons la production d'entropie pour le schéma explicite d'Euler-Maruyama et le schéma explicite de Milstein pour des equations différentielles stochastiques reversibles avec des bruit additifs ou multiplicatifs. Nos résultats démontrent que le type de bruit change le comportement la production d'entropie de manière critique. Finalement nous analysons la production d'entropie pour le schéma BBK pour l'equation de Langevin.

1991 Mathematics Subject Classification. ???, ???

The dates will be set by the publisher.

Introduction

In molecular simulations arising in the simulation of systems in materials science, chemical engineering, evolutionary games, computational statistical mechanics, etc. the equilibrium statistics obtained from numerical simulations are of great importance [6,22,28]. For instance, the free energy of the system or free energy differences as well dynamic transitions between metastable states are quantities which are sampled at the stationary regime. In addition, physical processes are often modeled at a microscopic level as interactions between particles which obey a system of stochastic differential equations (SDE's) [6, 12]. To perform equilibrium simulations for the sampling of desirable observables, the solution of the system of SDE's must possess a (unique) ergodic invariant measure. The uniqueness of the invariant measure follows from the ellipticity or hypoellipticity of the generator of the process together with irreducibility, which means that the process can reach at some positive time any open subset of the state space with positive probability [16, 20]. Under such conditions the distribution process converges to the invariant measure (ergodicity) which has a smooth density and the process started in the invariant measure is stationary, i.e. the distribution of the paths of the processes, is invariant under time-shift. Many processes of physical origin, such as diffusion and adsoprtion/desoprtion of interacting particles, satisfy the condition of detailed balance (DB), or equivalently, reversibility, i.e., the distribution of the path of the processes are invariant under time-reversal. It is easy to see that reversibility implies stationarity but is a strictly stronger condition in general. The condition of detailed balance often arises from a gradient-like behavior of the dynamics or from Hamiltonian dynamics if the time-reversal include reversal of the velocities.

However, the numerical simulation of SDE's necessitates the use of numerical discretization schemes. Discretization procedures, except in very special cases, results in the destruction of the DB condition. This affects the approximation process in at least two ways. First, the invariant measure of the approximation process, if it exists at all, is not known explicitly and, second, the time reversibility of the process is lost. Several recent results concerns the existence of the invariant measure for the discrete-time approximation and associated error estimates [2, 3, 14, 15] but, to the best of our knowledge, there is no quantitative assessment of the irreversibility of the approximation process. Of course there exist Metropolized numerical schemes such as MALA [21] and variations thereof which do satisfy the DB condition but they are numerically more expensive, especially in high-dimensional systems, as they require an accept/reject step. Thus, a quantitative understanding of the lack of reversibility for simpler discretization schemes can provide new insights for selecting which schemes are closer to satisfying the DB condition.

The implications of irreversibility are only partially understood, both from the physical and mathematical point of view. These issues have emerged as a main theme in non-equilibrium statistical mechanics and it is well-known that irreversibility introduces a stationary current (net flow) to the system [8, 18, 23] but it is unclear how this current affects the long-time properties (i.e., the dynamics and large deviations) of the process such as exit times, correlation times and phase transitions of metastable states. Reversibility is a natural and fundamental property of physical systems and thus, if numerical simulation results in the destruction of reversibility, one should carefully *quantify the irreversibility of the approximation process* and we do in this paper *using the entropy production rate*. The entropy production rate which is defined as the relative entropy (per unit time) between the path measure of the process and the path measure of the reversed process is widely used in statistical mechanics for the study of non-equilibrium steady states of irreversible systems [5,8,11,13]. A fundamental result on the structure of non-equilibrium steady states is the Gallavotti-Cohen fluctuation theorem that describes the fluctuations (of large deviations type) of the entropy production [5,8,11,13] and this result can be viewed as a generalization of the Kubo-formula and Onsager relations far from equilibrium. For our purpose, it is important to note that the entropy production rate is zero when the process is reversible and positive otherwise making *entropy production rate a sensible quantitative measure of irreversibility.* Furthermore, if we assume ergodicity of the approximation process, the entropy production rate equals the time-average of the Gallavotti-Cohen (GC) action functional which is defined as the logarithm of the Radon-Nikodym derivative between the path measure of the process and the path measure of the reversed process. A key observation of this paper is that an important feature of GC action functional is that it is an *a posteriori* quantity, hence, it is easily computable during the simulation making *the numerical computation of entropy production rate tractable.* We show that entropy production is a computable observable that distinguishes between different numerical schemes in terms of their discretization-induced irreversibility and as such allows us to adjust the discretization in the course of the simulation.

We use entropy production to assess the irreversibility of various numerical schemes for reversible continuous-time processes. A simple class of reversible processes, yet of great interest, is the overdamped Langevin process with gradient-type drift [6, 7, 12]. The discretization of the process is performed using the explicit Euler-Maruyama (EM) scheme and we distinguish between two different cases depending on the kind of the noise. In the case of additive noise, under the assumption of ergodicity of the approximation process [2, 3, 14, 15] we prove that the entropy production rate is of order $O(\Delta t^2)$ where Δt is the time step of the numerical scheme. In the case of multiplicative noise, the results are remarkably different. Indeed, under ergodicity assumption, the entropy production rate for the explicit EM scheme is proved to have a lower positive bound which is independent of Δt. Thus irreversibility is *not* reduced by adjusting Δt, as the approximation process converges to the continuous-time process. The different behavior of entropy production depending on the kind of noise is one of the prominent findings of this paper. As a further step in our study, we formulate and test numerically the explicit Milstein's scheme with multiplicative noise (it is the next higher-order numerical scheme). Simulation results on a wide range of different multiplicative noises show that the entropy production rate of Milstein's scheme decreases as time step decreases with order $O(\Delta t)$.

Finally, we compute both analytically and numerically the entropy production rate for a discretization scheme for Langevin systems which is another important and widely-used class of reversible models [6,12]. The Langevin equation is time-reversible if addition to reversing time, one reverses the sign of the velocity of all particles. The noise is degenerate but the process is hypo-elliptic and under mild conditions the Langevin equation is ergodic [15, 19, 26]. Our discretization scheme is an explicit EM–Verlet (symplectic)–implicit EM scheme also known as BBK integrator [4, 12]. We rigorously prove, under ergodicity assumption of the approximation process, that the entropy rate produced by the numerical scheme for the Langevin process with additive noise is of order $O(\Delta t)$, hence, in terms of irreversibility it can be an acceptable integration scheme.

The paper is organized in four sections. In Section 1 we recall some basic facts about reversible processes and define rigorously the entropy production. Moreover we give the basic assumption necessary for our proofs, namely, the ergodicity of both continuous-time and discrete-time approximation process. In Section 2 we compute the entropy production rate for reversible overdamped Langevin processes. The section is split into two subsections for the additive and multiplicative noise. In Section 3 we compute the entropy production rate for the reversible (up to momenta flip) Langevin process using the BBK integrator. Conclusions and future extensions of the current work are summarized in the fourth and final Section.

1. Reversibility, Gallavotti-Cohen Action Functional, and Entropy Production

Let us consider a d-dimensional system of SDE's written as

$$dX_t = a(X_t)dt + b(X_t)dB_t \tag{1}$$

where $X_t \in \mathbb{R}^d$ is a diffusion Markov process, $a : \mathbb{R}^d \to \mathbb{R}^d$ is the drift vector, $b : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ is the diffusion matrix, and $B_t \in \mathbb{R}^m$ is a standard m-dimensional Brownian motion. We will always assume that a and b are sufficiently smooth and satisfy suitable growth conditions and/or dissipativity conditions at infinity to ensure the existence of global solutions. The generator of the diffusion process is defined by

$$\mathcal{L}f = \sum_{i=1}^{d} a_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} (bb^T)_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{2}$$

for test functions f which are twice continuously differentiable and with bounded derivatives up to second order. We assume that the process X_t has a (unique) invariant measure $\mu(dx)$, and that it satisfies the Detailed Balance (DB) condition, i.e., its generator is symmetric in the Hilbert space $L^2(\mu)$, i.e.

$$< \mathcal{L}f, g >_{L^2(\mu)} = < f, \mathcal{L}g >_{L^2(\mu)} \tag{3}$$

for suitable test functions f, g as above.

A Markov process X_t is said to be reversible if for any n and sequence of times $t_1 < \cdots < t_n$ the finite dimensional distributions of $(X_{t_1}, ..., X_{t_n})$ and of $(X_{t_n}, ..., X_{t_1})$ are identical. More formally, let $\mathbf{P}^\rho_{[0,t]}$ denote the path measure of the process X_t on the time-interval $[0, t]$ with $X_0 \sim \rho$. Let Θ denote the time reversal, i.e. Θ acts on a path $\{X_s\}_{0\le s\le t}$ has

$$(\Theta X)_s = X_{t-s} \tag{4}$$

Then reversibility is equivalent to $\mathbf{P}^\mu_{[0,t]} = \mathbf{P}^\mu_{[0,t]} \circ \Theta$. Additionally, it is well-known that a stationary[1] process which satisfies the DB condition is reversible.

The condition of reversibility can be also expressed in terms of relative entropy as follows. Recall that for two probability measure π_1, π_2 on some measurable space, the relative entropy of π_1 with respect to π_2 is given by $R(\pi_1|\pi_2) \equiv \int d\pi_1 \log \frac{d\pi_1}{d\pi_2}$ if π_1 is absolutely continuous with respect to π_2 and $+\infty$ otherwise. The relative entropy is nonnegative, $R(\pi_1|\pi_2) \ge 0$ and $R(\pi_1|\pi_2) = 0$ if and only if $\pi_1 = \pi_2$. The entropy production rate of a Markov process X_t is defined by

$$EP_{cont} := \lim_{t\to\infty} \frac{1}{t} R(\mathbf{P}^\rho_{[0,t]}|\mathbf{P}^\rho_{[0,t]} \circ \Theta) = \lim_{t\to\infty} \frac{1}{t} \int d\mathbf{P}^\rho_{[0,t]} \log \frac{d\mathbf{P}^\rho_{[0,t]}}{d\mathbf{P}^\rho_{[0,t]} \circ \Theta} \tag{5}$$

If X_t satisfies DB and $X_0 \sim \mu$ then $R(\mathbf{P}^\mu_{[0,t]}|\mathbf{P}^\mu_{[0,t]} \circ \Theta)$ is identically 0 for all t and the entropy production rate is 0. Note that if $X_0 \sim \rho \ne \mu$ then $R(\mathbf{P}^\rho_{[0,t]}|\mathbf{P}^\rho_{[0,t]} \circ \Theta)$ is a boundary term, in the sense that it is $O(1)$

[1]Stationarity is equivalent to starting the process X_t from its invariant measure, i.e., $X_0 \sim \mu$.

and so the entropy rate vanishes in this case in the large time limit (under suitable ergodicity assumptions). Conversely when $EP_{cont} \neq 0$ the process is truly irreversible. The entropy production rate for Markov processes and stochastic differential equations is discussed in more detail in [11,13].

Let us consider a numerical integration scheme for the SDE (1) which is written in the general form

$$x_{i+1} = F(x_i, \Delta t, \Delta W_i) \quad i = 1, 2, ... \tag{6}$$

where $x_i \in \mathbb{R}^d$ is a discrete-time continuous state-space Markov process, Δt is the time-step and $\Delta W_i \in \mathbb{R}^m$, $i = 1, 2, ...$ are i.i.d. Gaussian random variables with mean 0 and variance $\Delta t I_m$. We assume that the Markov process x_i has transition probabilities which are absolutely continuous with respect to Lebesgue measure with everywhere positive densities $\Pi(x_i, x_{i+1}) := \Pi_{F(x,\Delta t,\Delta W)}(x_{i+1}|x_i)$ and we also assume that x_i has a invariant measure which we denote $\bar{\mu}(dx)$ and which is then unique and has a density with respect to Lebesgue. In general the invariant measure for X_t and x_i differ, $\mu \neq \bar{\mu}$ and x_i does not satisfy a DB condition. Note also that the very existence of $\bar{\mu}$ is not guaranteed in general. Results on the existence of $\bar{\mu}$ do exist however and typically require that the SDE is elliptic or hypoellitptic and that the state space of X_t is compact or that a global Lipschitz condition on the drift holds [2,3,14,15].

Proceeding as in the continuous case we introduce an entropy production rate for the Markov process x_i. Let us assume that the process starts from some distribution $\rho(x)dx$, then the finite dimensional distribution on the time window $[0, t]$ where $t = n\Delta t$ is given by

$$\bar{\mathbf{P}}_{[0,t]}(dx_0, \cdots, dx_n) = \rho(x_0)\Pi(x_0, x_1) \cdots \Pi(x_{n-1}, x_n) dx_0 \cdots dx_n\,. \tag{7}$$

For the time reversed path $\Theta(x_0, \cdots x_n) = (x_n, \cdots, x_0)$ we have then

$$\bar{\mathbf{P}}_{[0,t]} \circ \Theta(dx_0, ..., dx_n) = \rho(x_n)\Pi(x_n, x_{n-1}) \cdots \Pi(x_1, x_0) dx_0 \cdots dx_n \tag{8}$$

and the Radon-Nikodym derivative takes the form

$$\frac{d\bar{\mathbf{P}}_{[0,t]}}{d\bar{\mathbf{P}}_{[0,t]} \circ \Theta} = \exp(W(t)) \frac{\rho(x_0)}{\rho(x_n)} \tag{9}$$

where $W(t)$ is the Gallavotti-Cohen (GC) action functional given by

$$W(t) = W(n; \Delta t) := \sum_{i=0}^{n-1} \log \frac{\Pi(x_i, x_{i+1})}{\Pi(x_{i+1}, x_i)}\,. \tag{10}$$

Note that $W(t)$ is an additive functional of the paths and thus if x_i is ergodic, by the ergodic theorem the following limit exists

$$EP(\Delta t) = \lim_{t\to\infty} \frac{1}{t} W(t) = \lim_{n\to\infty} \frac{1}{n\Delta t} W(n; \Delta t) \quad \bar{P} - a.s.. \tag{11}$$

We call the quantity $EP(\Delta t)$ the entropy production rate associated to the numerical scheme. Note that we have, almost surely,

$$EP(\Delta t) = \frac{1}{\Delta t} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \frac{\Pi(x_i, x_{i+1})}{\Pi(x_{i+1}, x_i)} = \frac{1}{\Delta t} \int \log \frac{\Pi(x, y)}{\Pi(y, x)} \Pi(x, y)\bar{\mu}(x)\, dxdy \tag{12}$$

and for concrete numerical schemes we will compute fairly explicitly the entropy production in the next sections. Since we are interested in the ergodic average we will systematically omit boundary terms which do not contribute to ergodic averages and we will use the notation

$$W_1(t) \doteq W_2(t) \quad \text{if} \quad \lim_{t\to\infty} \frac{1}{t}(W_1(t) - W_2(t)) = 0\,. \tag{13}$$

For example we have

$$W(t) \doteq \log \frac{d\bar{\mathbf{P}}_{[0,t]}}{d\bar{\mathbf{P}}_{[0,t]} \circ \Theta}\,. \tag{14}$$

Note also that using (11) and (10), entropy production rate is tractable numerically and it can be easily calculated "on-the-fly" once the transition probability density function $\Pi(\cdot,\cdot)$ is provided.

In the following sections we investigate the behavior of the entropy production rate for different discretization schemes of various reversible processes in the stationary regime. However, before proceeding with our analysis, let us state formally the basic assumptions necessary for our results to apply.

Assumption 1.1. *We have*

- *The drift a and the diffusion b in (1) as well as the vector F in (6) are C^∞ and all their derivatives have at most polynomial growth at infinity.*
- *The generator $\mathcal{L}$ is elliptic or hypo-elliptic, in particular the transition probabilities and the invariant measure (if it exists) are absolutely continuous with respect to Lebesgue with smooth densities. For the discretized scheme we assume that x_i has smooth transition probabilities.*
- *Both the continuous-time process X_t and discrete-time process x_i are ergodic with unique invariant measures μ and $\bar{\mu}$, respectively. Furthermore for sufficiently small Δt we have*

$$|\mathbb{E}_\mu[f] - \mathbb{E}_{\bar{\mu}}[f]| = O(\Delta t) \tag{15}$$

for functions f which are C^∞ with at most polynomial growth at infinity.

Notice that inequality (15) is an error estimate for the invariant measures of the processes X_t and x_i. The rate of convergence in terms of Δt depends on the particular numerical scheme [14,25]. Ergodicity results for (numerical) SDEs can be found in [2,3,9,14,15,21,25–27]. For instance, if both drift term $a(x)$ and diffusion term $b(x)$ have bounded derivatives of any order, the covariance matrix $(bb^T)(x)$ is elliptic for all $x \in \mathbb{R}^d$ and there is a compact set outside of which holds $x^T a(x) < -C|x|^2$ for all $x \in \mathbb{R}^d$ (Lyapunov exponent) then it was shown in [25] that the continuous-time process as well both Euler and Milstein numerical schemes are ergodic and error estimate (15) holds. Another less restrictive example where ergodicity properties were proved is for SDE systems with degenerate noise and particularly for Langevin processes [15,26]. Again, a Lyapunov functional is the key assumption in order to handle the stochastic process at the infinity. More recently, Mattingly et al. [14] showed ergodicity for SDE-driven processes restricted on a torus as well their discretizations utilizing only the assumptions of ellipticity or hypoellipticity and the assumption of local Lipschitz continuity for both drift and diffusion terms.

2. Entropy Production for the Overdamped Langevin Processes

The overdamped Langevin process, $X_t \in \mathbb{R}^d$, is the solution of the following system of SDE's

$$dX_t = -\frac{1}{2}\Sigma(X_t)\nabla V(X_t)dt + \frac{1}{2}\nabla\Sigma(X_t)dt + \sigma(X_t)dB_t \tag{16}$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a smooth potential function, $\sigma : R^d \to \mathbb{R}^{d\times m}$ is the diffusion matrix, $\Sigma := \sigma\sigma^T : R^d \to \mathbb{R}^{d\times d}$ is the covariance matrix and B_t is a standard m-dimensional Brownian motion. We assume from now on that $\Sigma(x)$ is invertible for any x so that the process is elliptic. It is straightforward to show that the generator of the process X_t satisfies the DB condition (3) with invariant measure

$$\mu(dx) = \frac{1}{Z}\exp(-V(x))dx \tag{17}$$

where $Z = \int_{\mathbb{R}^d} \exp(-V(x))dx$ is the normalization constant and thus if $X_0 \sim \mu$ then the Markov process X_t is reversible.

The explicit Euler-Maruyama (EM) scheme for numerical integration of (16) is given by

$$x_{i+1} = x_i - \frac{1}{2}\Sigma(x_i)\nabla V(x_i)\Delta t + \frac{1}{2}\nabla\Sigma(x_i)\Delta t + \sigma(x_i)\Delta W_i \tag{18}$$

with $\Delta W_i \sim N(0, \Delta t I_m)$, $i = 1, 2, ...$ are m-dimensional iid Gaussian random variables. The process x_i is a discrete-time Markov process with transition probability density given by

$$\begin{aligned}\Pi(x_i, x_{i+1}) = \frac{1}{Z(x_i)} \exp\Bigg(& \frac{1}{2\Delta t}(\Delta x_i + \frac{1}{2}\Sigma(x_i)\nabla V(x_i)\Delta t - \frac{1}{2}\nabla\Sigma(x_i)\Delta t)^T \\ & \Sigma^{-1}(x_i)(\Delta x_i + \frac{1}{2}\Sigma(x_i)\nabla V(x_i)\Delta t - \frac{1}{2}\nabla\Sigma(x_i)\Delta t)\Bigg)\end{aligned} \tag{19}$$

where $\Delta x_i = x_{i+1} - x_i$ and $Z(x_i) = (2\pi)^{m/2}|\det\Sigma(x_i)|^{1/2}$ is the normalization constant for the multidimensional Gaussian distribution. The following lemma provides the GC action functional for the explicit EM time-discretization scheme of the overdamped Langevin process.

Lemma 2.1. *Assume that* $\det\Sigma(x) \neq 0\ \forall x \in \mathbb{R}^d$. *Then the GC action functional of the process* x_i *solving (18) is*

$$\begin{aligned}W(n;\Delta t) \doteq & -\frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\nabla V(x_{i+1}) + \nabla V(x_i)] + \frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\Sigma^{-1}(x_{i+1})\nabla\Sigma(x_{i+1}) + \Sigma^{-1}(x_i)\nabla\Sigma(x_i)] \\ & + \frac{1}{2\Delta t}\sum_{i=0}^{n-1}\Delta x_i^T\left[\Sigma^{-1}(x_{i+1}) - \Sigma^{-1}(x_i)\right]\Delta x_i\end{aligned} \tag{20}$$

where $\doteq$ *means equality up to boundary terms, as defined in* (13).

Proof. The assumption for non-zero determinant is imposed so that the transition probabilities and hence the GC action functional are non-singular. The proof is then a straightforward computation using (19) and (10).

$$
\begin{aligned}
&W(n;\Delta t) := \sum_{i=0}^{n-1} [\log \Pi(x_i, x_{i+1}) - \log \Pi(x_{i+1}, x_i)] = \sum_{i=0}^{n-1} [\log Z(x_{i+1}) - \log Z(x_i)] \\
&- \frac{1}{2\Delta t} \sum_{i=0}^{n-1} \Big[(\Delta x_i + \frac{1}{2}\Sigma(x_i)\nabla V(x_i)\Delta t - \frac{1}{2}\nabla\Sigma(x_i)\Delta t)^T \Sigma^{-1}(x_i)(\Delta x_i + \frac{1}{2}\Sigma(x_i)\nabla V(x_i)\Delta t - \frac{1}{2}\nabla\Sigma(x_i)\Delta t) \\
&-(-\Delta x_i + \frac{1}{2}\Sigma(x_{i+1})\nabla V(x_{i+1})\Delta t - \frac{1}{2}\nabla\Sigma(x_{i+1})\Delta t)^T \Sigma^{-1}(x_{i+1})(-\Delta x_i + \frac{1}{2}\Sigma(x_{i+1})\nabla V(x_{i+1})\Delta t - \frac{1}{2}\nabla\Sigma(x_{i+1})\Delta t) \Big] \\
&\doteq -\frac{1}{2\Delta t} \sum_{i=0}^{n-1} \big[\Delta x_i^T \Sigma^{-1}(x_i)\Delta x_i + \frac{1}{4}\nabla V(x_i)^T \Sigma(x_i)\nabla V(x_i)\Delta t^2 + \frac{1}{4}\nabla\Sigma(x_i)^T\Sigma^{-1}(x_i)\nabla\Sigma(x_i)\Delta t^2 \\
&+ \Delta x_i^T \nabla V(x_i)\Delta t - \Delta x_i^T \Sigma^{-1}(x_i)\nabla\Sigma(x_i)\Delta t - \frac{1}{2}\nabla V(x_i)^T \nabla\Sigma(x_i)\Delta t^2 \\
&- \Delta x_i^T \Sigma^{-1}(x_{i+1})\Delta x_i - \frac{1}{4}\nabla V(x_{i+1})^T \Sigma(x_{i+1})\nabla V(x_{i+1})\Delta t^2 - \frac{1}{4}\nabla\Sigma(x_{i+1})^T \Sigma^{-1}(x_{i+1})\nabla\Sigma(x_{i+1})\Delta t^2 \\
&+\Delta x_i^T \nabla V(x_{i+1})\Delta t - \Delta x_i^T \Sigma^{-1}(x_{i+1})\nabla\Sigma(x_{i+1})\Delta t + \frac{1}{2}\nabla V(x_{i+1})^T\nabla\Sigma(x_{i+1})\Delta t^2 \Big] \\
&\doteq -\frac{1}{2\Delta t}\sum_{i=0}^{n-1} \Delta x_i^T \left[\Sigma^{-1}(x_i) - \Sigma^{-1}(x_{i+1})\right]\Delta x_i - \frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\nabla V(x_{i+1}) + \nabla V(x_i)] \\
&+ \frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T [\Sigma^{-1}(x_{i+1})\nabla\Sigma(x_{i+1}) + \Sigma^{-1}(x_i)\nabla\Sigma(x_i)]
\end{aligned}
$$

where all the terms of the general form $G(x_i) - G(x_{i+1})$ in the sums were cancelled out since they form telescopic sums which become boundary terms. □

Three important remarks can readily be made from the above computation.

Remark 2.2. The numerical computation of entropy production rate as the time-average of the GC action functional on the path space (i.e., based on (9)) at first sight seems computationally intractable due to the large dimension of the path space. However, due to ergodicity, the numerical computation of the entropy production can be performed as a time-average based on (11) and (20) for large n. Additionally, this computation can be done for free and "on-the-fly" since the quantities involved are already computed in the simulation of the process. The numerical entropy production rate shown in the following figures is computed using this approach.

Remark 2.3. It was shown in [13] that the GC action functional of the *continuous-time* process driven by (16) equals the Stratonovich integral

$$W_{cont}(t) = -\int_0^t \nabla V(X_s) \circ dX_s = V(x_0) - V(x_t) \tag{21}$$

which reduces to a boundary term as expected. This functional has the discretization

$$W_{cont}(t) \approx \frac{1}{2}\sum_{i=0}^{n-1} \Delta x_i^T [\nabla V(x_{i+1}) + \nabla V(x_i)] \tag{22}$$

and this is exactly the first term in the GC action functional $W(n;\Delta t)$ for the explicit EM approximation process (see (20)). However, the discretization scheme introduces two additional terms to the GC action functional which may greatly affect the asymptotic behavior of entropy production as Δt goes to zero, as we demonstrate in Section 2.2. Notice that when the noise is additive, i.e., when the diffusion matrix is constant, then these two additional terms vanish and taking the limit $\Delta t \to 0$, the GC action functional $W(n;\Delta t)$, if exists, becomes the Stratonovich integral $W_{cont}(t)$ which is a boundary term.

Remark 2.4. The GC action functional $W(n;\Delta t)$ consists of three terms (see (20)), each of which stems from a particular term in the SDE. Thus, each term in the SDE contributes to the entropy production functional a component which is totally decoupled to the other terms. The reason for this decomposition lies in the particular form of the transition probabilities for the explicit EM scheme which are exponentials with quadratic argument. This feature can be exploited for the study of entropy production of numerical schemes for processes with irreversible dynamics. Indeed, if a non-gradient term of the form $a(X_t)dt$ is added to the drift of (16), the process is irreversible and its GC action functional is not anymore a boundary term and is given by [13]

$$W_{cont}(t) \doteq -\int_0^t \Sigma^{-1}(X_t)a(X_t)\circ dX_t \approx \frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\Sigma^{-1}(x_i)a(x_i)+\Sigma^{-1}(x_{i+1})a(x_{i+1})] \tag{23}$$

On the other hand, due to the separation property of the explicit EM scheme, the GC action functional of the discrete-time approximation process $W(n;\Delta t)$ has the additional term

$$\frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\Sigma^{-1}(x_i)a(x_i)+\Sigma^{-1}(x_{i+1})a(x_{i+1})]. \tag{24}$$

Evidently, the discretization of $W_{cont}(t)$ equals the additional term of the GC functional $W(n;\Delta t)$. Thus, GC action functional $W(n;\Delta t)$ is decomposed into two components, one stemming from the irreversibility of the continuous-time process and another one stemming from the irreversibility of the discretization procedure.

2.1. Entropy Production for the Additive Noise

An important special case of (16) is the case of additive noise, i.e., when the covariance matrix does not depend in the process, $\Sigma(x)\equiv\Sigma$. In this case, the SDE system becomes

$$\begin{aligned} dX_t &= -\frac{1}{2}\Sigma\nabla V(X_t)dt+\sigma dB_t \\ X_0 &\sim \mu \end{aligned} \tag{25}$$

and the GC action functional is simply given by

$$W(n;\Delta t) \doteq -\frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\nabla V(x_{i+1})+\nabla V(x_i)] \tag{26}$$

In this section we prove an upper bound for the entropy production of the explicit EM scheme. The proof uses several lemmas stated and proved in Appendix A.

Theorem 2.5. *Let Assumption 1.1 hold. Assume also that the potential function V has bounded fifth-order derivative and that the covariance matrix Σ is invertible. Then, for sufficiently small Δt, there exists $C = C(V,\Sigma) > 0$ such that*

$$EP(\Delta t) \le C\Delta t^2 \tag{27}$$

Proof. Utilizing the generalized trapezoidal rule (75) for $k=3$, the GC action function is rewritten as

$$\begin{aligned} W(n;\Delta t) \doteq & -\frac{1}{2}\sum_{i=0}^{n-1}\Delta x_i^T[\nabla V(x_{i+1})+\nabla V(x_i)] \\ &= \sum_{i=0}^{n-1}\Bigg\{-(V(x_{i+1})-V(x_i))+\sum_{|\alpha|=3}C_\alpha[D^\alpha V(x_{i+1})+D^\alpha V(x_i)]\Delta x_i^\alpha \\ &+\sum_{|\alpha|=1,3,5}\sum_{|\beta|=5-|\alpha|}B_\beta[R_\alpha^\beta(x_i,x_{i+1})+R_\alpha^\beta(x_{i+1},x_i)]\Delta x_i^{\alpha+\beta}\Bigg\} \\ &\doteq \sum_{i=0}^{n-1}\sum_{|\alpha|=3}C_\alpha[D^\alpha V(x_{i+1})+D^\alpha V(x_i)]\Delta x_i^\alpha \\ &+\sum_{i=0}^{n-1}\sum_{|\alpha|=1,3,5}\sum_{|\beta|=5-|\alpha|}B_\beta[R_\alpha^\beta(x_i,x_{i+1})+R_\alpha^\beta(x_{i+1},x_i)]\Delta x_i^{\alpha+\beta}\,. \end{aligned} \tag{28}$$

Applying, once again, Taylor series expansion to $D^\alpha V(x_{i+1})$, the GC action functional becomes

$$\begin{aligned} W(n;\Delta t) \doteq & \sum_{i=0}^{n-1}\Bigg\{\sum_{|\alpha|=3}2C_\alpha D^\alpha V(x_i)\Delta x_i^\alpha+\sum_{|\alpha|=3}C_\alpha\sum_{|\beta|=1}D^{\alpha+\beta}V(x_i)\Delta x_i^{\alpha+\beta}\Bigg\} \\ &+\sum_{i=0}^{n-1}\sum_{|\alpha|=1,3,5}\sum_{|\beta|=5-|\alpha|}\bar{R}_\alpha^\beta(x_i,x_{i+1})\Delta x_i^{\alpha+\beta} \end{aligned} \tag{29}$$

where $\bar{R}_\alpha^\beta(x_i,x_{i+1}) = B_\beta[R_\alpha^\beta(x_i,x_{i+1})+R_\alpha^\beta(x_{i+1},x_i)]+\mathbb{1}_{|\alpha|=3}R_\beta^\alpha(x_i,x_{i+1})$. Moreover, expanding Δx_i^α using the multi-binomial formula

$$\Delta x_i^\alpha = (-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t+\sigma\Delta W_i)^\alpha = \sum_{\nu\le\alpha}\binom{\alpha}{\nu}(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu(\sigma\Delta W_i)^{\alpha-\nu}\,. \tag{30}$$

Then, the GC action functional becomes

$$\begin{aligned} W(n;\Delta t) \doteq & 2\sum_{i=0}^{n-1}\sum_{|\alpha|=3}\sum_{\nu\le\alpha}C_\alpha\binom{\alpha}{\nu}D^\alpha V(x_i)(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu(\sigma\Delta W_i)^{\alpha-\nu} \\ &+\sum_{i=0}^{n-1}\sum_{|\alpha|=3}\sum_{|\beta|=1}\sum_{\nu\le\alpha+\beta}C_\alpha\binom{\alpha+\beta}{\nu}D^{\alpha+\beta}V(x_i)(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu(\sigma\Delta W_i)^{\alpha+\beta-\nu} \\ &+\sum_{i=0}^{n-1}\sum_{|\alpha|=1,3,5}\sum_{|\beta|=5-|\alpha|}\sum_{\nu\le\alpha+\beta}\binom{\alpha+\beta}{\nu}\bar{R}_\alpha^\beta(x_i,x_{i+1})(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu(\sigma\Delta W_i)^{\alpha+\beta-\nu}\,. \end{aligned} \tag{31}$$

From (11), the entropy production rate is the time-averaged GC action functional as $n \to \infty$. Thus,

$$
\begin{aligned}
EP(\Delta t) &= \lim_{n\to\infty} \frac{W(n;\Delta t)}{n\Delta t} \\
&= \frac{2}{\Delta t}\sum_{|\alpha|=3}\sum_{\nu\le\alpha} C_\alpha \binom{\alpha}{\nu} \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} D^\alpha V(x_i)(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu (\sigma\Delta W_i)^{\alpha-\nu} \\
&+ \frac{1}{\Delta t}\sum_{|\alpha|=3}\sum_{|\beta|=1}\sum_{\nu\le\alpha+\beta} C_\alpha \binom{\alpha+\beta}{\nu} \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} D^{\alpha+\beta} V(x_i)(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu (\sigma\Delta W_i)^{\alpha+\beta-\nu} \\
&+ \frac{1}{\Delta t}\sum_{|\alpha|=1,3,5}\sum_{|\beta|=5-|\alpha|}\sum_{\nu\le\alpha+\beta} \binom{\alpha+\beta}{\nu} \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} \bar{R}_\alpha^\beta(x_i,x_{i+1})(-\frac{1}{2}\Sigma\nabla V(x_i)\Delta t)^\nu (\sigma\Delta W_i)^{\alpha+\beta-\nu}\,.
\end{aligned}
\tag{32}
$$

The ergodicity of x_i as well the Gaussianity of ΔW_i guarantees that the first two limits in the entropy production formula exist. Additionally, the residual terms, $\bar{R}_\alpha^\beta(x_i,x_{i+1})$, are bounded due to the assumption on bounded fifth-order derivative of V, hence, the third limit also exists. Note here that this assumption could be changed by assuming boundedness of a higher order derivative and performing a higher-order Taylor expansion. Appendix A gives rigorous proofs of these ergodicity statements. Hence,

$$
\begin{aligned}
EP(\Delta t) &= \frac{2}{\Delta t}\sum_{|\alpha|=3}\sum_{\nu\le\alpha} C_\alpha \binom{\alpha}{\nu} \mathbb{E}_{\bar\mu}[D^\alpha V(x)(-\frac{1}{2}\Sigma\nabla V(x)\Delta t)^\nu]\mathbb{E}_\rho[(\sigma y)^{\alpha-\nu}] \\
&+ \frac{1}{\Delta t}\sum_{|\alpha|=3}\sum_{|\beta|=1}\sum_{\nu\le\alpha+\beta} C_\alpha \binom{\alpha+\beta}{\nu} \mathbb{E}_{\bar\mu}[D^{\alpha+\beta} V(x)(-\frac{1}{2}\Sigma\nabla V(x)\Delta t)^\nu]\mathbb{E}_\rho[(\sigma y)^{\alpha+\beta-\nu}] \\
&+ \frac{1}{\Delta t}\sum_{|\alpha|=1,3,5}\sum_{|\beta|=5-|\alpha|}\sum_{\nu\le\alpha+\beta} \binom{\alpha+\beta}{\nu} \mathbb{E}_{\bar\mu\times\rho}[\bar{R}_\alpha^\beta(x,y)(-\frac{1}{2}\Sigma\nabla V(x)\Delta t)^\nu]\mathbb{E}_\rho[(\sigma y)^{\alpha+\beta-\nu}]
\end{aligned}
\tag{33}
$$

where $\bar\mu$ is the equilibrium measure for x_i while ρ is the Gaussian measure of ΔW_i. Using the Isserlis-Wick formula we can compute the higher moments of multivariate Gaussian random variable from the second-order moments. Indeed, we have

$$
\mathbb{E}[y^\nu] = \mathbb{E}[y_1^{\nu_1}...y_d^{\nu_d}] = \mathbb{E}[z_1 z_2 ... z_{|\nu|}] = \begin{cases} 0 & \text{if } |\nu| \text{ odd} \\ \sum\prod \mathbb{E}[z_i z_j] & \text{if } |\nu| \text{ even} \end{cases}
\tag{34}
$$

where $\sum\prod$ means summing over all distinct ways of partitioning $z_1,...,z_{|\nu|}$ into pairs. Moreover, $\mathbb{E}[z_i z_j] = \Sigma_{ij}\Delta t$, hence, applying (34) into (33) and changing the multi-index notation to the usual notation, the entropy production rate becomes

$$
\begin{aligned}
EP(\Delta t) &= \frac{2}{\Delta t}\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d} C_{k_1k_2k_3}\Big\{\mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_3}}(-\frac{1}{2}\Sigma\nabla V)_{k_1}]\Sigma_{k_2k_3}\Delta t^2 \\
&+\mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_3}}(-\frac{1}{2}\Sigma\nabla V)_{k_2}]\Sigma_{k_1k_3}\Delta t^2 + \mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_3}}(-\frac{1}{2}\Sigma\nabla V)_{k_3}]\Sigma_{k_1k_2}\Delta t^2 + O(\Delta t^3)\Big\} \\
&+ \frac{1}{\Delta t}\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d} C_{k_1k_2k_3}\Big\{\mathbb{E}_{\bar\mu}[\frac{\partial^4 V}{\partial x_{k_1}...\partial x_{k_4}}][\Sigma_{k_1k_2}\Sigma_{k_3k_4}+\Sigma_{k_1k_3}\Sigma_{k_2k_4}+\Sigma_{k_1k_4}\Sigma_{k_2k_3}]\Delta t^2 + O(\Delta t^3)\Big\} \\
&+ \frac{1}{\Delta t}O(\Delta t^3)\,.
\end{aligned}
\tag{35}
$$

Using that $(-\frac{1}{2}\Sigma\nabla V)_{k_i} = -\frac{1}{2}\sum_{k_4=1}^{d}\Sigma_{k_ik_4}\frac{\partial V}{\partial x_{k_4}}$, entropy production is rewritten as

$$\begin{aligned} EP(\Delta t) = \sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d} C_{k_1k_2k_3}\Big\{\Sigma_{k_1k_2}\Sigma_{k_3k_4}\left(-\mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_3}\partial x_{k_4}}\frac{\partial V}{\partial x_{k_2}}] + \mathbb{E}_{\bar\mu}[\frac{\partial^4 V}{\partial x_{k_1}...\partial x_{k_4}}]\right) \\ + \Sigma_{k_1k_3}\Sigma_{k_2k_4}\left(-\mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_4}}\frac{\partial V}{\partial x_{k_3}}] + \mathbb{E}_{\bar\mu}[\frac{\partial^4 V}{\partial x_{k_1}...\partial x_{k_4}}]\right) \\ +\Sigma_{k_1k_4}\Sigma_{k_2k_3}\left(-\mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_3}}\frac{\partial V}{\partial x_{k_4}}] + \mathbb{E}_{\bar\mu}[\frac{\partial^4 V}{\partial x_{k_1}...\partial x_{k_4}}]\right)\Big\}\Delta t + O(\Delta t^2)\,. \end{aligned} \tag{36}$$

By a simple integration by parts, we observe that for any combination $k_1, ..., k_4 = 1, ..., d$

$$\mathbb{E}_{\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_3}}\frac{\partial V}{\partial x_{k_4}}] = \mathbb{E}_{\mu}[\frac{\partial^4 V}{\partial x_{k_1}...\partial x_{k_4}}] \tag{37}$$

where the expectation is taken with respect of μ which is the invariant measure of the continuous-time process. However, in (36) the expectation is w.r.t. the invariant measure of the discrete-time process (i.e., $\bar\mu$ instead of μ). Nevertheless, Assumption 1.1 guarantees that the alternation of the measure from μ to $\bar\mu$ costs an error of order $O(\Delta t)$. Hence, for any coefficient in (36), we obtain that

$$\left|\mathbb{E}_{\bar\mu}[\frac{\partial^3 V}{\partial x_{k_1}\partial x_{k_2}\partial x_{k_3}}\frac{\partial V}{\partial x_{k_4}}] - \mathbb{E}_{\bar\mu}[\frac{\partial^4 V}{\partial x_{k_1}...\partial x_{k_4}}]\right| \leq 2K\Delta t \tag{38}$$

since the potential V as well its derivatives are sufficiently smooth. Hence, we overall showed that

$$EP(\Delta t) = O(\Delta t^2) \tag{39}$$

which completes the proof. □

Remark 2.6. Depending on the potential function the entropy production could be even smaller. For instance, when the potential V is a quadratic function (i.e. the continuous-time process is an Ornstein-Uhlenbeck process), then, it is easily checked by a trivial calculation of (26) that the GC action function is a boundary term, thus, the entropy production of the explicit EM scheme is zero. However, for a generic potential V we expect that the entropy production rate decays quadratically as a function of Δt but not faster.

2.1.1. *Fourth-order potential on a torus*

Lets now proceed with an important example where the potential is a forth-order polynomial while the process takes values on a torus. Assume $d = 2$ while potential $V = V_\beta$ is given by

$$V_\beta(x) = \beta\left(\frac{|x|^4}{4} - \frac{|x|^2}{2}\right) \tag{40}$$

where β is a positive real number which in statistical mechanics has the meaning of the inverse temperature. The diffusion matrix is set to $\sigma = \sqrt{2\beta^{-1}}I_d$. Based on [15], Assumption 1.1 is satisfied because the domain is restricted to a torus, the potential is locally Lipschitz continuous and the covariance matrix is elliptic. Figure 1 presents both the GC action functional (upper panel) and the entropy production rate (lower panel) as a function of time for fixed $\Delta t = 0.05$. Both quantities are numerically computed while the inverse temperature is set to $\beta = 10$. Even though the variance of the GC action functional is large, entropy production which is the cumulative sum of the GC functional converges due to the law of large numbers to a (positive) value after relatively long time. Additionally, due to the ergodicity assumption, it converges to the correct value. Figure 2 shows the loglog plot of the numerical entropy production rate as a function of Δt for $\beta = 20,\ 40,\ 60$.

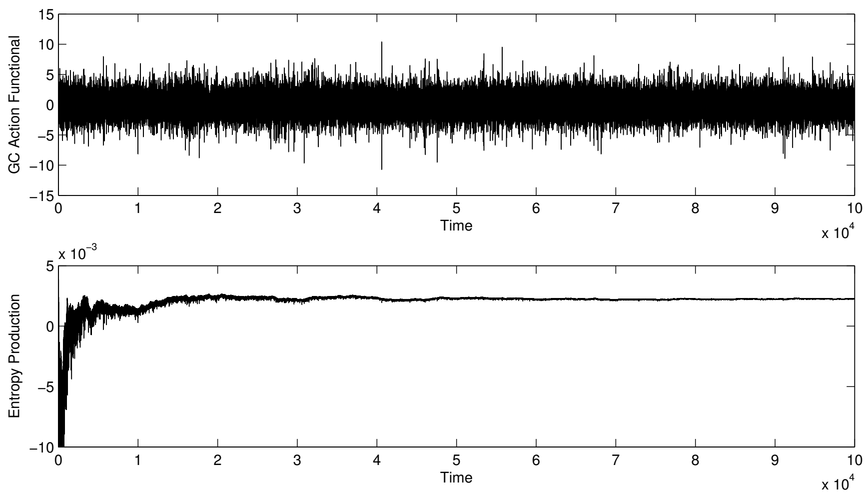

FIGURE 1. Upper panel: The GC action functional as a function of time for fixed $\Delta t = 0.05$. Its variance is large necessitating the use of many samples in order to obtain statistically confident quantities. Lower Panel: The entropy production rate as a function of time for the same Δt. It converges to a positive value as expected.

Final time was set to $t = 2 \cdot 10^6$ while initial point was set to one of the attraction points of the deterministic counterpart. For reader's convenience, the thick black line denotes the $O(\Delta t^2)$ rate of convergence. This plot is in agreement with the theorem's estimate (27) at least for small Δt while for larger time-steps (i.e. $\Delta t > 0.1$) the rate of entropy production is of order $O(\Delta t^3)$. Notice also that, for small Δt, entropy production rate is very close to 0 and even larger final time is needed in order to obtain a statistically confident numerical estimate for the entropy production. Moreover, as it is evident from the figure and the GC action functional in (26), the dependence of the entropy production w.r.t. the inverse temperature is inverse proportional. Thus, from a statistical mechanics point of view, the larger is the temperature the larger –in a linear manner– is the entropy production rate of the numerical scheme.

2.2. **Entropy Production for the Multiplicative Noise in** $1d$

For the multiplicative overdamped Langevin process, we restrict to the 1-dimensional case. The reason for this restriction is that we apply not only the EM scheme but also a higher-order scheme (Milstein's) which becomes complicated for general diffusion matrices in higher dimensions. Nonetheless, the results and conclusions of this subsection for both explicit EM and Milstein's schemes are valid in a more general, multi-dimensional setting where the diffusion matrix $\sigma(x)$ is diagonal.

In order to study the entropy production rate of the explicit EM scheme for the overdamped Langevin process with multiplicative noise, the remainder terms of the GC action functional should be studied. In this direction we can rewrite the GC action function as it is given by the Lemma 2.1 for $1d$

$$\begin{aligned} W(n;\Delta t) \doteq & -\frac{1}{2}\sum_{i=0}^{n-1}[V'(x_{i+1})+V'(x_i)]\Delta x_i + \frac{1}{2}\sum_{i=0}^{n-1}[\Sigma^{-1}(x_{i+1})\Sigma'(x_{i+1})+\Sigma^{-1}(x_i)\Sigma'(x_i)]\Delta x_i \\ & + \frac{1}{2\Delta t}\sum_{i=0}^{n-1}\left[\Sigma^{-1}(x_{i+1})-\Sigma^{-1}(x_i)\right]\Delta x_i^2 =: W_1(n;\Delta t)+W_2(n;\Delta t)+W_3(n;\Delta t)\,. \end{aligned} \tag{41}$$

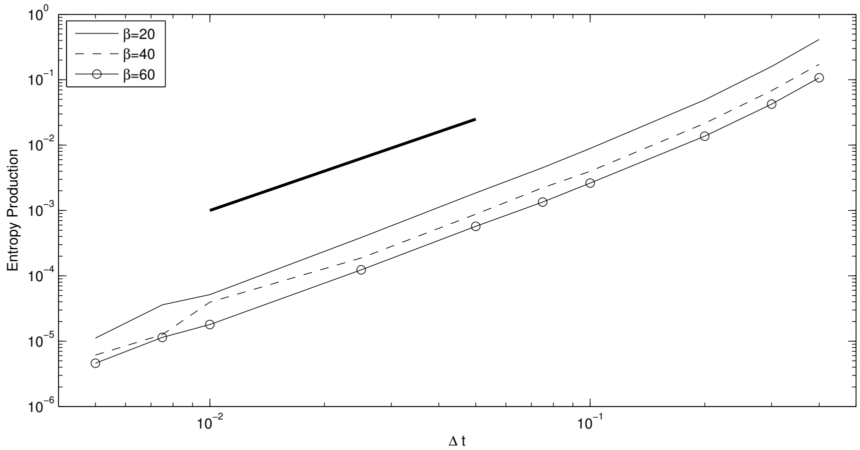

FIGURE 2. Entropy production rate as a function of time step Δt for additive noise. The entropy production rate is of order $O(\Delta t^2)$ for small Δt while it decreases linearly as a function of inverse temperature β.

The entropy produced from $W_1(n;\Delta t)$ was computed in the previous section and after an interesting and rather unexpected cancellation it was proved to be of order $O(\Delta t^2)$. For the multiplicative case, a cancellation also occurs (see (45) and (46) below) but it does not fully eliminate the lower order term. In any case, $W_1(n;\Delta t)$ contributes to the entropy production $O(\Delta t)$. Additionally, $W_2(n;\Delta t)$ is also the sum of a gradient term since variance $\Sigma(x) \in \mathbb{R}$ and holds $\Sigma^{-1}(x)\Sigma'(x) = (\log\Sigma(x))'$. Hence, assuming suitable condition on $\Sigma(x)$, the same computation as for $W_1(n;\Delta t)$ applies and the entropy production rate stemming from $W_2(n;\Delta t)$ is also of order $O(\Delta t)$. However, $W_3(n;\Delta t)$ contributes to the entropy production with a positive term which is of order $O(1)$. The following theorem summarizes the behavior of entropy production rate for the explicit EM scheme for multiplicative noise.

Theorem 2.7. *Let Assumption 1.1 hold. Assume also that the potential function V has bounded fifth-order derivative while there exists $M > 0$ such that $\Sigma(x) > M^{-1}$, $\forall x$.*
(a) Let $c = \frac{3}{4}\mathbb{E}_\mu[(\Sigma^{-1})(x)(\Sigma')^2(x)]$, then, for sufficiently small Δt, there exists $C = C(V,\Sigma) > 0$ independent of Δt such that

$$|EP(\Delta t) - c| \leq C\Delta t \tag{42}$$

(b) Assuming that $\mathbb{E}_\mu[(\Sigma^{-1})(x)(\Sigma')^2(x)] \neq 0$, then, for sufficiently small Δt, there exists a lower bound $c' = c'(V,\Sigma) > 0$ independent of Δt such that

$$c' \leq EP(\Delta t) \tag{43}$$

Proof. Assumption $\Sigma(x) > M^{-1}$ $\forall x$, which is the ellipticity condition applied in $1d$, is necessary because it makes $\Sigma^{-1}(x)$ as well its derivatives bounded around 0. Additionally, both $W_1(n;\Delta t)$ and $W_2(n;\Delta t)$ contribute to the entropy production by a $O(\Delta t)$ amount which does not affect the proof of the theorem hence they are eliminated. Thus, concentrating to $W_3(n;\Delta t)$, after a Taylor series expansion we have

$$W_3(n;\Delta t) = \frac{1}{2\Delta t}\sum_{i=0}^{n-1}\left[(\Sigma^{-1})'(x_i)\Delta x_i^3 + \frac{1}{2}(\Sigma^{-1})''(x_i)\Delta x_i^4 + \frac{1}{2\Delta t}\sum_{i=0}^{n-1}\int_0^1 (1-t)(\Sigma^{-1})'''(tx_{i+1}+(1-t)x_i)dt\Delta x_i^5\right]$$
$$= \frac{1}{2\Delta t}\sum_{i=0}^{n-1}\sum_{k=0}^{3}\binom{3}{k}(\Sigma^{-1})'(x_i)(-\frac{1}{2}\Sigma(x_i)V'(x_i)\Delta t + \frac{1}{2}\Sigma'(x_i)\Delta t)^k(\sigma(x_i)\Delta W_i)^{3-k}$$
$$+ \frac{1}{4\Delta t}\sum_{i=0}^{n-1}\sum_{k=0}^{4}\binom{4}{k}(\Sigma^{-1})''(x_i)(-\frac{1}{2}\Sigma(x_i)V'(x_i)\Delta t + \frac{1}{2}\Sigma'(x_i)\Delta t)^k(\sigma(x_i)\Delta W_i)^{4-k}$$
$$+ \frac{1}{2\Delta t}\sum_{i=0}^{n-1}\sum_{k=0}^{5}\binom{5}{k}\int_0^1 (1-t)(\Sigma^{-1})'''(tx_{i+1}+(1-t)x_i)dt(-\frac{1}{2}\Sigma(x_i)V'(x_i)\Delta t + \frac{1}{2}\Sigma'(x_i)\Delta t)^k(\sigma(x_i)\Delta W_i)^{5-k}.$$

As in Theorem 2.5, applying the ergodic lemmas of the appendix, the entropy production rate stemming from $W_3(n;\Delta t)$ equals to

$$\begin{aligned}
EP_3(\Delta t) &= \lim_{t\to\infty}\frac{W_3(n;\Delta t)}{n\Delta t}\\
&= \frac{1}{2\Delta t^2}\sum_{k=0}^{3}\binom{3}{k}\mathbb{E}_{\bar\mu}[(\Sigma^{-1})'(x)(-\frac{1}{2}\Sigma(x)V'(x)\Delta t + \frac{1}{2}\Sigma'(x)\Delta t)^k\sigma(x)^{3-k}]\mathbb{E}_\rho[\Delta W^{3-k}]\\
&+ \frac{1}{4\Delta t^2}\sum_{k=0}^{4}\binom{4}{k}\mathbb{E}_{\bar\mu}[(\Sigma^{-1})''(x)(-\frac{1}{2}\Sigma(x)V'(x)\Delta t + \frac{1}{2}\Sigma'(x)\Delta t)^k\sigma(x)^{4-k}]\mathbb{E}_\rho[\Delta W^{4-k}]\\
&+ \frac{1}{2\Delta t^2}\sum_{k=0}^{5}\mathbb{E}_{\bar\mu\times\rho}[R(x,y)(-\frac{1}{2}\Sigma(x)V'(x)\Delta t + \frac{1}{2}\Sigma'(x)\Delta t)^k\sigma(x)^{5-k}]\mathbb{E}_\rho[\Delta W^{5-k}]\\
&= \frac{1}{2\Delta t^2}\left[-\frac{3}{2}\mathbb{E}_{\bar\mu}[(\Sigma^{-1})'(x)\Sigma^2(x)V'(x)]\Delta t^2 + \frac{3}{2}\mathbb{E}_{\bar\mu}[(\Sigma^{-1})'(x)\Sigma'(x)\Sigma(x)]\Delta t^2 + O(\Delta t^3)\right]\\
&+ \frac{1}{4\Delta t^2}\left[\mathbb{E}_{\bar\mu}[(\Sigma^{-1})''(x)\Sigma^2(x)]3\Delta t^2 + O(\Delta t^3)\right] + \frac{1}{2\Delta t^2}O(\Delta t^3)\\
&= \frac{3}{4}\left[-\mathbb{E}_{\bar\mu}[(\Sigma^{-1})'(x)\Sigma^2(x)V'(x)] + \frac{1}{2}\mathbb{E}_{\bar\mu}[(\Sigma^{-1})'(x)(\Sigma^2)'(x)] + \mathbb{E}_{\bar\mu}[(\Sigma^{-1})''(x)\Sigma^2(x)]\right] + O(\Delta t)
\end{aligned} \tag{44}$$

On the other hand, it holds for the invariant measure μ that

$$\mathbb{E}_\mu[(\Sigma^{-1})'(x)\Sigma^2(x)V'(x)] = \mathbb{E}_\mu[(\Sigma^{-1})''(x)\Sigma^2(x)] + \mathbb{E}_\mu[(\Sigma^{-1})'(x)(\Sigma^2)'(x)] \tag{45}$$

Thus, using the error estimate (15) of Assumption 1.1 as in the additive case, we obtain that

$$\begin{aligned}
EP_3(\Delta t) &= -\frac{3}{8}\mathbb{E}_{\bar\mu}[(\Sigma^{-1})'(x)(\Sigma^2)'(x)] + O(\Delta t)\\
\Rightarrow EP_3(\Delta t) - \frac{3}{4}\mathbb{E}_{\bar\mu}&[(\Sigma^{-1})(x)(\Sigma')^2(x)] = O(\Delta t)
\end{aligned} \tag{46}$$

which concludes the proof of (a). (b) is a direct consequence of (a). □

2.2.1. *Quadratic potential on* $\mathbb{R}$

Let $V(x) = \frac{x^2}{2}$ be a single-well quadratic potential while the diffusion term is given by

$$\sigma_\epsilon(x) = \sqrt{\frac{1}{1+\epsilon x^2}} \tag{47}$$

The choice of the diffusion term is justified by the fact that we can control its variation in terms of x and sending ϵ to zero, the additive noise case is recovered. The invariant measure of this process is the Gaussian measure with zero mean and variance one. This invariant measure is the simplest measure to be considered. Moreover, all the assumptions of Theorem 2.7 are satisfied thus we expect a $O(1)$ behavior of the entropy production rate at least for small Δt. Indeed, Figure 3 shows the behavior of the numerically-computed entropy production as a function of Δt and it does not decrease to zero as Δt tends to zero. Consequently, explicit EM scheme for multiplicative noise totally destroys the reversibility property of the discrete-time approximation process independently of how small time-step is utilized. Additionally, notice that as ϵ decreases, entropy production decreases, too. This is also expected since $\sigma(x) \to \sigma = \text{const.}$ as $\epsilon \to 0$ and in combination with the quadratic potential V, $EP(\Delta t) \to 0$ as $\epsilon \to 0$ for any Δt sufficiently small.

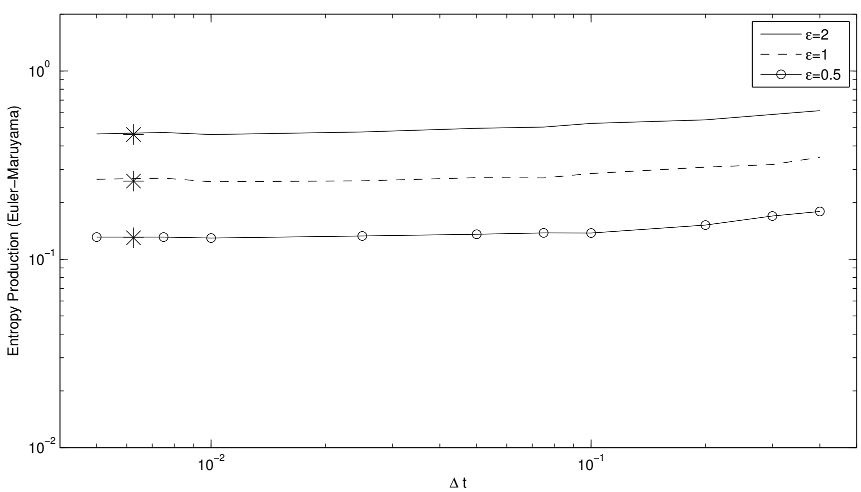

FIGURE 3. Entropy production rate as a function of time step Δt for multiplicative noise and the explicit EM scheme. As Theorem 2.7 asserts, entropy production does not decrease as Δt is decreased. This results in a permanent loss of reversibility which cannot be fixed by reducing the time step. Star symbols denote the theoretical value of the lower bound as it is given by the Theorem (i.e., $c' \approx c = \frac{3}{4}\mathbb{E}_\mu[(\Sigma_\epsilon)^{-1}(x)(\Sigma'_\epsilon)^2(x)]$). The agreement between the theoretical and the numerical values is excellent.

2.2.2. *Milstein's scheme*

Since the EM scheme has entropy production rate which does not decrease as Δt decreases, an immediate question to ask is what happens when a higher-order scheme is applied. Milstein's scheme is the next higher-order scheme [10, 17] and its explicit version is given by

$$x_{i+1} = x_i - \frac{1}{2}\Sigma(x_i)V'(x_i)\Delta t + \frac{1}{2}\Sigma'(x_i)\Delta t + \sigma(x_i)\Delta W_i + \frac{1}{2}\sigma(x_i)\sigma'(x_i)(\Delta W_i^2 - \Delta t) \tag{48}$$

which is rewritten as

$$\Delta x_i = -\frac{1}{2}\Sigma(x_i)V'(x_i)\Delta t + \frac{1}{4}\Sigma'(x_i)\Delta t + \sigma(x_i)\Delta W_i + \frac{1}{4}\Sigma'(x_i)\Delta W_i^2 \,. \tag{49}$$

Since ΔW_i is zero-mean Gaussian random variable with variance Δt, the transition probability for Milstein's scheme is

$$\begin{aligned}\Pi(x_i, x_{i+1}) = \frac{1}{\sqrt{2\pi\Delta t Z(x_i,\Delta x_i)}} &\left[\exp\left(-\frac{1}{2\Delta t}\left|\frac{-\sigma(x_i)+\sqrt{Z(x_i,\Delta x_i)}}{\frac{1}{2}\Sigma'(x_i)}\right|^2\right)\right.\\ &\left.+\exp\left(-\frac{1}{2\Delta t}\left|\frac{\sigma(x_i)+\sqrt{Z(x_i,\Delta x_i)}}{\frac{1}{2}\Sigma'(x_i)}\right|^2\right)\right]\end{aligned} \tag{50}$$

where

$$Z(x_i,\Delta x_i) = \Sigma(x_i) + \Sigma'(x_i)\left(\Delta x_i + \frac{1}{2}\Sigma(x_i)V'(x_i)\Delta t - \frac{1}{4}\Sigma'(x_i)\Delta t\right) \,. \tag{51}$$

Notice also that $Z(x_i,\Delta x_i) = (\sigma(x_i) + \frac{1}{2}\Sigma'(x_i)\Delta W_i)^2 \geq 0$ which is always non-negative while the transition probability density is rewritten as

$$\Pi(x_i, x_{i+1}) = \frac{1}{\sqrt{2\pi\Delta t Z(x_i,\Delta x_i)}} \exp\left(-\frac{2(\Sigma(x_i)+Z(x_i,\Delta x_i))}{\Delta t(\Sigma')^2(x_i)}\right)\cosh\left(\frac{\sqrt{\Sigma(x_i)Z(x_i,\Delta x_i)}}{\Delta t\Sigma'(x_i)}\right) \,. \tag{52}$$

Thus, the GC action functional for Milstein's scheme equals up to boundary terms to

$$\begin{aligned}W(n;\Delta t) \doteq &-\frac{1}{2}\sum_{k=0}^{n-1}[\log\frac{Z(x_i,\Delta x_i)}{Z(x_{i+1},-\Delta x_i)}] - \frac{2}{\Delta t}\sum_{k=0}^{n-1}\left[\frac{Z(x_i,\Delta x_i)}{(\Sigma')^2(x_i)} - \frac{Z(x_{i+1},-\Delta x_i)}{(\Sigma')^2(x_{i+1})}\right]\\ &+\sum_{k=0}^{n-1}\left[\log\cosh\frac{\sqrt{Z(x_i,\Delta x_i)}}{2\Delta t\sigma'(x_i)} - \log\cosh\frac{\sqrt{Z(x_{i+1},-\Delta x_i)}}{2\Delta t\sigma'(x_{i+1})}\right] \,.\end{aligned} \tag{53}$$

We can test the behavior of the entropy production numerically since, as we already stated, averaged GC action functional provides under ergodicity assumption an estimate for the entropy production rate. Figure 4 shows the numerically computed entropy production for the same example shown in Figure 3. Evidently, entropy production rate decreases linearly as time step Δt is decreasing. Additionally, a number of different variance functions which satisfy the condition of Theorem 2.7 were tested and in all cases the decrease of the entropy production for the Milstein's scheme was linear. Thus, we conjecture that entropy production of overdamped Langevin process with multiplicative noise is of order $O(\Delta t)$ for Milstein's scheme.

3. Entropy Production for Langevin Process

Let us consider another important class of reversible processes, namely the processes driven by the Langevin equation

$$\begin{aligned} dq_t &= M^{-1}p_t dt \\ dp_t &= -\nabla V(q_t)dt - \gamma(q_t)M^{-1}p_t dt + \sigma(q_t)dB_t \end{aligned} \tag{54}$$

where $q_t \in \mathbb{R}^{dN}$ is the position vector of the N particles, $p_t \in \mathbb{R}^{dN}$ is the momentum vector of the particles, M is the mass matrix, V is the potential energy, γ is the friction factor (matrix), σ is the diffusion factor (matrix) and B_t is a dN-dimensional Brownian motion. Even though the Langevin system is degenerate since the noise applies only to the momenta, the process is hypoelliptic and is ergodic under mild conditions on V and σ. The

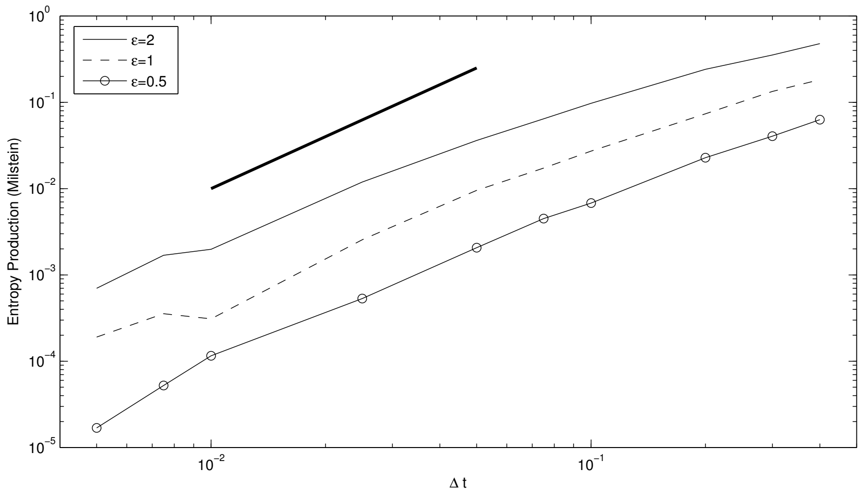

FIGURE 4. Entropy production rate as a function of time step Δt for the explicit Milstein's scheme. The decrease of the entropy production rate for this numerical scheme is linear. Thus, in a loose sense, the reversibility property of the original continuous-time process is restored.

fluctuation-dissipation theorem asserts that friction and diffusion terms are related with the inverse temperature $\beta(\cdot) \in \mathbb{R}$ of the system by

$$(\sigma\sigma^T)(q_t) = 2\beta^{-1}(q_t)\gamma(q_t)\,. \tag{55}$$

If $\beta(q_t) = \beta$ is a constant, the Langevin equation is reversible (modulo momenta flip, see (58)) with invariant measure

$$\mu(dq, dp) = \frac{1}{Z}\exp\left(-\beta H(q,p)\right) dqdp. \tag{56}$$

where $H(q,p)$ is the Hamiltonian of the system given by

$$H(q,p) = V(q) + \frac{1}{2}p^T M^{-1} p\,. \tag{57}$$

Indeed if $\mathcal{L}$ denotes the generator of (54), it is straightforward to verify the following modified DB condition

$$< \mathcal{L}f(q,p), g(q,p) >_{L^2(\mu)} = < f(q,-p), \mathcal{L}g(q,-p) >_{L^2(\mu)} \tag{58}$$

for any test functions f and g which are bounded, twice differentiable with bounded derivatives. This shows that the Langevin process is reversible modulo flipping the momenta of all particles.

An explicit EM–Verlet (symplectic)–implicit EM scheme is applied for the discretization of (54). It is written as

$$\begin{aligned} p_{i+\frac{1}{2}} &= p_i - \nabla V(q_i)\frac{\Delta t}{2} - \gamma(q_i)M^{-1}p_i\frac{\Delta t}{2} + \sigma(q_i)\Delta W_i \\ q_{i+1} &= q_i + M^{-1}p_{i+\frac{1}{2}}\Delta t \\ p_{i+1} &= p_{i+\frac{1}{2}} - \nabla V(q_{i+1})\frac{\Delta t}{2} - \gamma(q_{i+1})M^{-1}p_{i+1}\frac{\Delta t}{2} + \sigma(q_{i+1})\Delta W_{i+\frac{1}{2}} \end{aligned} \tag{59}$$

with $\Delta W_i, \Delta W_{i+\frac{1}{2}} \sim N(0, \frac{\Delta t}{2} I_{dN})$. This numerical scheme also known as BBK integrator [4, 12] utilizes a Strang splitting. Its stability and convergence properties were studied in [4, 12] while its ergodic properties can be found in [14, 15, 26]. An important property of this numerical scheme which simplifies the computation of the transition probabilities is that the transition probabilities are non-degenerate. We rewrite the BBK integrator as

$$q_{i+1} = q_i + M^{-1}[p_i - \nabla V(q_i)\frac{\Delta t}{2} - \gamma(q_i)M^{-1}p_i\frac{\Delta t}{2}]\Delta t + M^{-1}\sigma(q_i)\Delta t\Delta W_i \tag{60a}$$

$$p_{i+1} = (I+\gamma(q_{i+1})M^{-1}\frac{\Delta t}{2})^{-1}[\frac{1}{\Delta t}M(q_{i+1}-q_i) - \nabla V(q_{i+1})\frac{\Delta t}{2}] + (I+\gamma(q_{i+1})M^{-1}\frac{\Delta t}{2})^{-1}\sigma(q_{i+1})\Delta W_{i+\frac{1}{2}} \tag{60b}$$

and thus the transition probabilities of the discrete-time approximation process are given by the product

$$\Pi(q_i, p_i, q_{i+1}, p_{i+1}) = P(q_{i+1}|q_i, p_i)P(p_{i+1}|q_{i+1}, q_i, p_i) \tag{61}$$

where $P(p_{i+1}|q_i, p_i)$ is the propagator of the positions given by

$$\begin{aligned} P(q_{i+1}|q_i, p_i) = \frac{1}{Z_0}\exp\{&\frac{1}{\Delta t^3}(\Delta q_i + M^{-1}(p_i - \nabla V(q_i)\frac{\Delta t}{2} + \gamma(q_i)M^{-1}p_i\frac{\Delta t}{2})\Delta t)^T \\ &(\sigma M^{-T}M^{-1}\sigma^T)^{-1}(q_i)(\Delta q_i + M^{-1}(p_i - \nabla V(q_i)\frac{\Delta t}{2} + \gamma(q_i)M^{-1}p_i\frac{\Delta t}{2})\Delta t)\} \end{aligned} \tag{62}$$

where $\Delta q_i = q_{i+1} - q_i$ while $P(p_{i+1}|q_{i+1}, q_i, p_i)$ is the propagator of the momenta given by

$$\begin{aligned} P(p_{i+1}|q_{i+1}, q_i, p_i) = \frac{1}{Z_1(q_{i+1})}\exp\{&\frac{1}{\Delta t}(p_{i+1} - (I+\gamma(q_{i+1})M^{-1}\frac{\Delta t}{2})^{-1}(\frac{1}{\Delta t}M\Delta q_i - \nabla V(q_{i+1})\frac{\Delta t}{2}))^T \\ (\sigma^T(I+\gamma M)^{-T}(I+\gamma M^{-1})\sigma)^{-1}(q_{i+1})&(p_{i+1} - (I+\gamma(q_{i+1})M^{-1}\frac{\Delta t}{2})^{-1}(\frac{1}{\Delta t}M\Delta q_i - \nabla V(q_{i+1})\frac{\Delta t}{2}))\} \end{aligned} \tag{63}$$

Finally, since the Langevin process is reversible modulo flip of the momenta, the GC action functional takes the form

$$W(n;\Delta t) = \sum_{i=0}^{n-1}\log\frac{\Pi(q_i, p_i, q_{i+1}, p_{i+1})}{\Pi(q_{i+1}, -p_{i+1}, q_i, -p_i)}. \tag{64}$$

3.1. **Langevin Process with Additive Noise**

In the following, even though the general case can be handled, we restrict for clarity to the simpler additive noise case. Thus, we assume that $\sigma(q_i) = \sigma I$, $\gamma(q_i) = \gamma I$ as well that particles have equal masses ($M = mI$). Starting as in the previous section with the GC action functional, the next lemma is stated and proved.

Lemma 3.1. *The GC action functional of the BBK integrator equals to*

$$W(n;\Delta t) \doteq \frac{2\beta}{m\Delta t}\sum_{i=0}^{n-1}\left[\Delta p_i^T\Delta q_i - \nabla V(q_i)^T p_i\frac{\Delta t^2}{2m}\right] \tag{65}$$

Proof. Firstly, (62) and (63) are rewritten as

$$P(q_{i+1}|q_i, p_i) = \frac{1}{Z_0}\exp\left\{\frac{m^2}{\sigma^2\Delta t^3}|\Delta q_i + (p_i - \frac{1}{m}\nabla V(q_i)\frac{\Delta t}{2} + \frac{\gamma}{m}p_i\frac{\Delta t}{2})\Delta t|^2\right\} \tag{66}$$

and

$$P(p_{i+1}|q_{i+1}, q_i, p_i) = \frac{1}{Z_1}\exp\left\{\frac{1}{\sigma^2\Delta t}|(1+\frac{\gamma\Delta t}{2m})p_{i+1} - (\frac{m}{\Delta t}\Delta q_i - \frac{\Delta t}{2}\nabla V(q_{i+1}))|^2\right\} \tag{67}$$

respectively. Then, as in the overdamped Langevin case, the computation of the GC action functional is straightforward,

$$
\begin{aligned}
W(n;\Delta t) &= -\frac{m^2}{\sigma^2\Delta t^3}\sum_{i=0}^{n-1}\left[\left|\Delta q_i + \frac{\Delta t^2}{2m}\nabla V(q_i) - \frac{\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})p_i\right|^2 - \left|-\Delta q_i + \frac{\Delta t^2}{2m}\nabla V(q_{i+1}) + \frac{\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})p_{i+1}\right|^2\right] \\
&\quad - \frac{1}{\sigma^2\Delta t}\sum_{i=0}^{n-1}\left[\left|(1+\frac{\gamma\Delta t}{2m})p_{i+1} - \frac{m}{\Delta t}\Delta q_i + \frac{\Delta t}{2}\nabla V(q_{i+1})\right|^2 - \left|-(1+\frac{\gamma\Delta t}{2m})p_i + \frac{m}{\Delta t}\Delta q_i + \frac{\Delta t}{2}\nabla V(q_i)\right|^2\right] \\
&= -\frac{m^2}{\sigma^2\Delta t^3}\sum_{i=0}^{n-1}\Bigg[|\Delta q_i|^2 + |\frac{\Delta t^2}{2m}\nabla V(q_i)|^2 + |\frac{\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})p_i|^2 + \frac{\Delta t^2}{m}\Delta q_i^T\nabla V(q_i) \\
&\quad - \frac{2\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})\Delta q_i^T p_i - \frac{\Delta t^3}{m^2}(1-\frac{\gamma\Delta t}{2m})\nabla V(q_i)^T p_i \\
&\quad - |\Delta q_i|^2 - |\frac{\Delta t^2}{2m}\nabla V(q_{i+1})|^2 - |\frac{\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})p_{i+1}|^2 + \frac{\Delta t^2}{m}\Delta q_i^T\nabla V(q_{i+1}) \\
&\quad + \frac{2\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})\Delta q_i^T p_{i+1} - \frac{\Delta t^3}{m^2}(1-\frac{\gamma\Delta t}{2m})\nabla V(q_{i+1})^T p_{i+1}\Bigg] \\
&\quad - \frac{1}{\sigma^2\Delta t}\sum_{i=0}^{n-1}\Bigg[|(1+\frac{\gamma\Delta t}{2m})p_{i+1}|^2 + |\frac{m}{\Delta t}\Delta q_i|^2 + |\frac{\Delta t}{2}\nabla V(q_{i+1})|^2 - (1+\frac{\gamma\Delta t}{2m})\frac{2m}{\Delta t}p_{i+1}^T\Delta q_i \\
&\quad + (1+\frac{\gamma\Delta t}{2m})\Delta t p_{i+1}^T\nabla V(q_{i+1}) - m\Delta q_i^T\nabla V(q_{i+1}) \\
&\quad - |(1+\frac{\gamma\Delta t}{2m})p_i|^2 - |\frac{m}{\Delta t}\Delta q_i|^2 - |\frac{\Delta t}{2}\nabla V(q_i)|^2 + (1+\frac{\gamma\Delta t}{2m})\frac{2m}{\Delta t}p_i^T\Delta q_i \\
&\quad + (1+\frac{\gamma\Delta t}{2m})\Delta t p_i^T\nabla V(q_i) - m\Delta q_i^T\nabla V(q_i)\Bigg].
\end{aligned}
$$

Thus we have,

$$
\begin{aligned}
W(n;\Delta t) &\doteq -\frac{m^2}{\sigma^2\Delta t^3}\sum_{i=0}^{n-1}\Bigg[\frac{\Delta t^2}{m}\Delta q_i^T(\nabla V(q_i)+\nabla V(q_{i+1})) + \frac{2\Delta t}{m}(1-\frac{\gamma\Delta t}{2m})\Delta q_i^T\Delta p_i \\
&\quad - \frac{\Delta t^3}{m^2}(1-\frac{\gamma\Delta t}{2m})(\nabla V(q_{i+1})^T p_{i+1} + \nabla V(q_i)^T p_i)\Bigg] \\
&\quad - \frac{1}{\sigma^2\Delta t}\sum_{i=0}^{n-1}\Bigg[-(1+\frac{\gamma\Delta t}{2m})\frac{2m}{\Delta t}\Delta p_i^T\Delta q_i - m\Delta q_i^T(\nabla V(q_i)+\nabla V(q_{i+1})) \\
&\quad + (1+\frac{\gamma\Delta t}{2m})\Delta t(p_i^T\nabla V(q_i) + p_{i+1}^T\nabla V(q_{i+1}))\Bigg] \\
&= -\frac{2m}{\sigma^2\Delta t^2}\sum_{i=0}^{n-1}\Bigg[-(1-\frac{\gamma\Delta t}{2m})\Delta q_i^T\Delta p_i + (1+\frac{\gamma\Delta t}{2m})\Delta q_i^T\Delta p_i \\
&\quad + \frac{\Delta t^2}{2m}(1-\frac{\gamma\Delta t}{2m})(\nabla V(q_{i+1})^T p_{i+1} + \nabla V(q_i)^T p_i) - \frac{\Delta t^2}{2m}(1+\frac{\gamma\Delta t}{2m})(\nabla V(q_{i+1})^T p_{i+1} + \nabla V(q_i)^T p_i)\Bigg] \\
&= -\frac{2\gamma}{m\sigma^2\Delta t}\sum_{i=0}^{n-1}\left[\Delta p_i^T\Delta q_i - \frac{\Delta t^2}{2m}(\nabla V(q_{i+1})^T p_{i+1} + \nabla V(q_i)^T p_i)\right]
\end{aligned}
$$

which is equal, up to boundary terms, with (65). □

Remark 3.2. Proceeding as in Remark 2.3 we can compare the GC action functional of the BBK integrator to the GC functional for the additive Langevin process with constant temperature, which is given, [13], by

$$W_{cont}(t) = \frac{\beta}{m}\int_0^t \nabla V(q_t)p_t dt \approx \frac{\beta \Delta t}{m}\sum_{i=0}^{n-1} \nabla V(q_i)^T p_i \tag{68}$$

and is a boundary term in continuous time. Comparing the GC functionals, it is evident that the discrete version of $W_{cont}(t)$ is contained in the functional $W(n;\Delta t)$ given by (65). This is similar to the overdamped Langevin case when discretized utilizing the explicit EM scheme. In addition the remaining term in the GC action functional $W(n;\Delta t)$ stems from the Strang splitting of the numerical scheme. Moreover, this additional term critically affects the irreversibility of the discrete-time approximation process since it is the leading order term in the entropy production rate, as shown in the following theorem.

Theorem 3.3. *Let Assumption 1.1 hold. Assume also that the potential function V has bounded fifth-order derivative. Then, for sufficiently small Δt, there exists $C = C(N,\gamma,m) > 0$ such that*

$$EP(\Delta t) \le C\Delta t \tag{69}$$

Proof. Solving (60a) for p_i, changing the index from $i+1$ to i in (60b) and adding them, the momenta equal to

$$p_i = \frac{m}{2\Delta t}(\Delta q_i + \Delta q_{i-1}) + \frac{\sigma}{2}(\Delta W_{i-\frac{1}{2}} - \Delta W_i) \tag{70}$$

Then,

$$\Delta p_i = \frac{m}{2\Delta t}(\Delta q_{i+1} - \Delta q_{i-1}) + \frac{\sigma}{2}(-\Delta W_{i+1} - \Delta W_{i+\frac{1}{2}} + \Delta W_i - \Delta W_{i-\frac{1}{2}}) \tag{71}$$

hence the GC action functional becomes

$$\begin{aligned}
W(n;\Delta t) &\doteq \frac{2\beta}{m\Delta t}\sum_{i=0}^{n-1}\left[\Delta p_i^T \Delta q_i - \nabla V(q_i)^T p_i \frac{\Delta t^2}{2m}\right] \\
&= \frac{2\beta}{m\Delta t}\sum_{i=0}^{n-1}\Bigg[\left(\frac{m}{2\Delta t}(\Delta q_{i+1} - \Delta q_{i-1}) + \frac{\sigma}{2}(-\Delta W_{i+1} - \Delta W_{i+\frac{1}{2}} + \Delta W_i - \Delta W_{i-\frac{1}{2}})\right)^T \Delta q_i \\
&\quad -\nabla V(q_i)^T \left(\frac{m}{2\Delta t}(\Delta q_i + \Delta q_{i-1}) + \frac{\sigma}{2}(\Delta W_{i-\frac{1}{2}} - \Delta W_i)\right)\frac{\Delta t^2}{2m}\Bigg] \\
&\doteqdot \frac{2\beta}{m\Delta t}\sum_{i=0}^{n-1}\left[\frac{\sigma}{2}(\Delta W_i - \Delta W_{i-\frac{1}{2}})^T \Delta q_i - \nabla V(q_i)^T(\Delta q_i + \Delta q_{i-1})\Delta t\right] \\
&\doteq \frac{\beta\sigma}{m^2}\sum_{i=0}^{n-1}(\Delta W_i - \Delta W_{i-\frac{1}{2}})^T\left((1-\frac{\gamma\Delta t}{2m})p_i - \nabla V(q_i)\frac{\Delta t}{2} + \sigma\Delta W_i\right) \\
&\quad - \frac{\beta\sigma}{m^2}\sum_{i=0}^{n-1}(\nabla V(q_{i+1})^T p_{i+1} + \nabla V(q_i))^T \Delta q_i
\end{aligned} \tag{72}$$

where $\doteqdot$ means equality not only up to boundary terms but also up to statistical independence which does not affect the value of the entropy production rate, either.

The second sum of GC action functional has exactly the same form as in additive overdamped Langevin equation and adapting the arguments of Theorem 2.5 it can be proved that the entropy production rate for this term is of order $O(\Delta t^2)$. The first term is treated similarly, but since an additional cancellation occurs we provide the details. The first sum in (72) equals to

$$\begin{aligned}
&\frac{\beta\sigma}{m^2}\sum_{i=0}^{n-1}(\Delta W_i - \Delta W_{i-\frac{1}{2}})^T\left((1-\frac{\gamma\Delta t}{2m})p_i - \nabla V(q_i)\frac{\Delta t}{2} + \sigma\Delta W_i\right)\\
&\doteq \frac{\beta\sigma}{m^2}\sum_{i=0}^{n-1}\left[-(1-\frac{\gamma\Delta t}{2m})(1+\frac{\gamma\Delta t}{2m})^{-1}(\frac{m}{\Delta t}\Delta q_{i-1} - \frac{\Delta t}{2}\nabla V(q_i) + \sigma\Delta W_{i-\frac{1}{2}})^T\Delta W_{i-\frac{1}{2}} + \sigma|\Delta W_i|^2\right]\\
&\doteq \frac{\beta\sigma^2}{m^2}\sum_{i=0}^{n-1}\left[-(1-\frac{\gamma\Delta t}{2m})(1+\frac{\gamma\Delta t}{2m})^{-1}|\Delta W_{i-\frac{1}{2}}|^2 + |\Delta W_i|^2\right]
\end{aligned} \tag{73}$$

Hence, the total entropy production rate becomes

$$\begin{aligned}
EP(\Delta t) &= \frac{2\gamma}{m^2\Delta t}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\left[-(1-\frac{\gamma\Delta t}{2m})(1+\frac{\gamma\Delta t}{2m})^{-1}|\Delta W_{i-\frac{1}{2}}|^2 + |\Delta W_i|^2\right] + O(\Delta t^2)\\
&= \frac{2\gamma}{m^2\Delta t}\left[-(1-2\frac{\gamma\Delta t}{2m} + O(\Delta t^2))\frac{N\Delta t}{2} + \frac{N\Delta t}{2}\right] + O(\Delta t^2)\\
&= \frac{N\gamma^2}{m^3}\Delta t + O(\Delta t^2)
\end{aligned} \tag{74}$$

which completes the proof. □

3.1.1. *Quadratic potential on a torus*

The conclusions of the above theorem are validated by a numerical example where the potential function is quadratic, $V(x) = \frac{|x|^2}{2}$. Figure 5 shows the behavior of numerical entropy production rate as a function of Δt computed as the time-average of the GC action functional. Number of particles was set to $N = 5$ while the mass of its particle was set to $m = 1$. The variance of the stochastic term was set $\sigma^2 = 0.01$ while the final time was set to $t = 2 \cdot 10^5$. The initial data was chosen randomly from the zero-mean Gaussian distribution with appropriate variance. Notice also that due to the quadratic potential of this example Gaussian distribution is also the invariant measure of the process. Thus, the simulation is performed at the equilibrium regime. Evidently, the entropy production rate is of order $O(\Delta t)$ as it is expected. Additionally, we plot (stars in the Figure) the leading term of the theoretical value of the entropy production rate as it given by (74). Apparently, the theoretical coefficient, $\frac{N\gamma^2}{m^3}$, is very close to the numerically-computed coefficient. Finally, notice that the entropy production rate is quadratically proportional to the friction factor γ which is in accordance with (74).

4. Summary and Future Work

In this paper, we introduce the entropy production rate as a novel tool to assess quantitatively the (lack of) reversibility of discretization schemes for various reversible SDE's. Reversibility of the discrete-time approximation process is a desirable feature when equilibrium simulations are performed. The entropy production rate which is defined as the time-average of the relative entropy between the path measure of the forward process and the path measure of the time-reversed process is zero when the process is reversible and positive when it is irreversible. Thus, it provides a way to quantify the (ir)reversibility of the approximation process. Moreover, under an ergodicity assumption, entropy production rate can be computed numerically on-the-fly utilizing the GC action functional. This is another attractive feature of the entropy production rate.

We have computed the entropy production rate for overdamped Langevin processes both analytically and numerically when discretized with explicit Euler-Maruyama scheme. One of the main finding in this paper is that depending on the type of the noise –additive vs multiplicative– the entropy production for the explicit EM

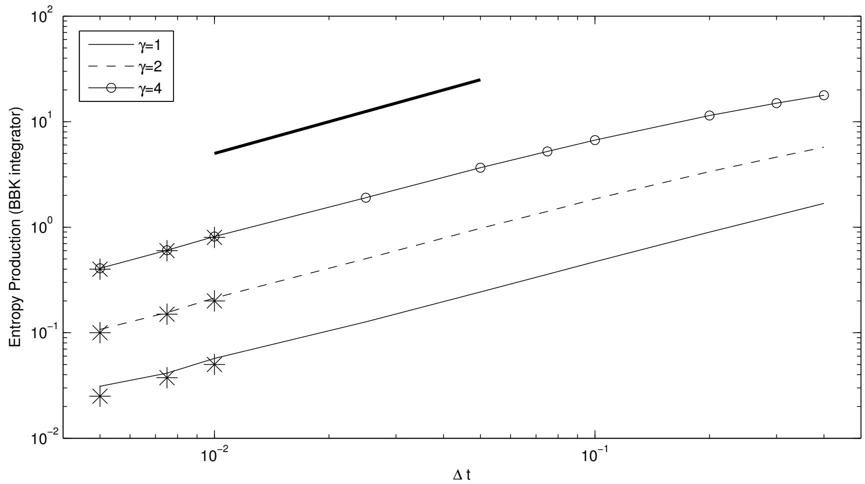

FIGURE 5. Entropy production rate as a function of time step, Δt, for various friction factors γ. The decrease of the entropy production rate is linear as Theorem 3.3 asserts. Additionally, the theoretically-computed entropy production rate (star points) perfectly matches the numerically-computed entropy rate.

scheme had totally different behavior. Indeed, for additive noise entropy production rate is of order $O(\Delta t^2)$ while for multiplicative noise it is of order $O(1)$. Hence, reversibility of the discrete-time approximation process does not depend only on the numerical scheme but also on the intrinsic characteristics of the SDE. The Milstein's scheme improved the convergence rate of the entropy production rate for multiplicative noise as shown in numerical simulations. Furthermore, we have computed the entropy production rate both analytically and numerically for discretization schemes of the Langevin process with additive noise. Specifically, we computed the entropy production rate for the BBK integrator of the Langevin equation which is an explicit EM-symplectic (Verlet)- implicit EM numerical scheme. The rate of entropy production was shown to be of order $O(\Delta t)$.

This paper offers a new conceptual tool for the evaluation of discretization schemes of SDE systems simulated at the equilibrium regime. We consider only the simplest schemes here and we will analyze in future work the behavior of the entropy production for other numerical schemes such as fully implicit EM, drift-implicit EM, higher-order schemes as well as different kind of splitting methods. Moreover, other reversible or even non-reversible processes can be analyzed in the same way, in particular extended, spatially-distributed processes. A particularly interesting example, where the reversibility of the original system is destroyed by numerical schemes in the form of spatio-temporal fractional step approximations of the generator, arises in the (partly asynchronous) parallelization of Kinetic Monte Carlo algorithms [24], [1]. Finally, another possible extension of this work is to develop adaptive schemes based on the *a posteriori* simulation of entropy production rate, which should guarantee the reversibility or the approximate reversibility of the discrete-time approximation process. In this direction, the decomposition of entropy production functional for Metropolis-adjusted Langevin algorithms (MALA) [12, 21] should be further studied and understood.

REFERENCES

[1] G. Arampatzis, M. A. Katsoulakis, P. Plechac, M. Taufer, and L. Xu. Hierarchical fractional-step approximations and parallel kinetic Monte Carlo algorithms. *J. Comp. Phys. (accepted), ArXiv e-prints*, May 2012.

[2] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the density. *Monte Carlo Methods Appl.*, 2:93–128, 1996.

[3] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104:43–60, 1996.
[4] A. Brunger, C. B. Brooks, and M. Karplus. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.*, 105:495–500, 1984.
[5] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, 74:2694–2697, 1995.
[6] C. Gardiner. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer, 1985.
[7] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. NewYork: Academic Press, 1992.
[8] V. Jakšić, C.-A. Pillet, and L. Rey-Bellet. Entropic fluctuations in statistical mechanics: I. classical dynamical systems. *Nonlinearity*, 24(3):699–763, 2011.
[9] R. Khasminskii. *Stochastic Stability of Differential Equations*. Springer, 2nd Edition, 2010.
[10] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 3rd Ed., 1999.
[11] J. L. Lebowitz and H. Spohn. A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.*, 95:333–365, 1999.
[12] T. Lelievre, M. Rousset, and G. Stoltz. *Free energy computations: a mathematical perspective*. Imperial College Press, 2010.
[13] C. Maes, F. Redig, and A. Van Moffaert. On the definition of entropy production, via examples. *J. Math. Phys.*, 41:1528–1553, 2000.
[14] J. C. Mattingly, A.M. Stuart, and M.V. Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, 48:552–577, 2010.
[15] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101:185–232, 2002.
[16] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.
[17] G. Milstein and M. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.
[18] G. Nicolis and I. Prigogine. *Self-Organization in Nonequilibrium Systems*. Wiley, New York, 1977.
[19] L. Rey-Bellet and L.E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.*, 225(2):305–329, 2002.
[20] Luc Rey-Bellet. Ergodic properties of Markov processes. In *Open quantum systems. II*, volume 1881 of *Lecture Notes in Math.*, pages 1–39. Springer, Berlin, 2006.
[21] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110, 1996.
[22] T. Schlick. *Molecular Modeling and Simulation*. Springer, 2002.
[23] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Modern Phys.*, 48(4):571–585, 1976.
[24] Yunsic Shim and Jacques G. Amar. Semirigorous synchronous relaxation algorithm for parallel kinetic Monte Carlo simulations of thin film growth. *Phys. Rev. B*, 71(12):125432, Mar 2005.
[25] D. Talay. Second order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics Stochastics Rep.*, 29:13–36, 1990.
[26] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields*, 8:163–198, 2002.
[27] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl*, 8:483–509, 1990.
[28] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, 2006.

Appendix A. Tools for proving Theorem 3.2

Firstly, a generalization of the trapezoidal rule is stated and proved.

Lemma A.1 (Generalized Trapezoidal Rule)**.** *For k odd,*

$$\begin{aligned} V(x_{i+1}) - V(x_i) = &\sum_{|\alpha|=1,3,\dots}^{k} C_\alpha [D^\alpha V(x_{i+1}) + D^\alpha V(x_i)] \Delta x_i^\alpha \\ &+ \sum_{|\alpha|=1,3,\dots}^{k+2} \sum_{|\beta|=k+2-|\alpha|} B_\beta [R_\alpha^\beta(x_i, x_{i+1}) + R_\alpha^\beta(x_{i+1}, x_i)] \Delta x_i^{\alpha+\beta} \end{aligned} \tag{75}$$

where $\alpha = (\alpha_1, ..., \alpha_d)$ *is a typical* d*-dimensional multi-index vector,* $D^\alpha V(x) = \frac{\partial^{|\alpha|} V}{\partial x_1^{\alpha_1} ... \partial x_d^{\alpha_d}}(x)$ *is the* α*-th partial derivative while* $x^\alpha = x_1^{\alpha_1} ... x_d^{\alpha_d}$*. The coefficients* C_α *are defined recursively by*

$$\begin{aligned} C_\alpha &= \frac{1}{2} \qquad \textit{for} \quad |\alpha| = 1 \\ C_\alpha &= \frac{1}{2}\left(\frac{1}{\alpha!} - \sum_{|\gamma|=1,3,...}^{|\alpha|-2} \frac{1}{(\alpha-\gamma)!} C_\gamma\right) \qquad \textit{for} \quad |\alpha| = 3, 5, ..., k \end{aligned} \tag{76}$$

while the coefficients B_β *are also recursively defined by*

$$\begin{aligned} B_\beta &= \frac{1}{2} \qquad \textit{for} \quad |\beta| = 0 \\ B_\beta &= -\frac{1}{2} \sum_{|\gamma|=2,4,...}^{|\beta|} \frac{1}{\gamma!} B_{\beta-\gamma} \qquad \textit{for} \quad |\beta| = 2, 4, ..., k+1 \end{aligned} \tag{77}$$

Finally, the remainder terms are given by
$R_\alpha^\beta(x_i, x_{i+1}) = \frac{|\alpha|}{\alpha!} \int_0^1 (1-t)^{|\alpha|-1} D^{\alpha+\beta} V((1-t)x_i + t x_{i+1}) dt.$

Proof. The starting point is the usual Taylor series expansion around x_i

$$V(x_{i+1}) - V(x_i) = \sum_{|\alpha|=1}^{k+1} \frac{1}{\alpha!} D^\alpha V(x_i) \Delta x_i^\alpha + \sum_{|\alpha|=k+2} R_\alpha^0(x_i, x_{i+1}) \Delta x_i^\alpha \tag{78}$$

and around x_{i+1}

$$V(x_{i+1}) - V(x_i) = -\sum_{|\alpha|=1}^{k+1} \frac{1}{\alpha!} D^\alpha V(x_{i+1}) (-\Delta x_i)^\alpha - \sum_{|\alpha|=k+2} R_\alpha^0(x_{i+1}, x_i)(-\Delta x_i)^\alpha \tag{79}$$

Adding the two equations we obtain the symmetrized Taylor series expansion for V given by

$$\begin{aligned} V(x_{i+1}) - V(x_i) &= \frac{1}{2} \sum_{|\alpha|=1,3,...}^{k} \frac{1}{\alpha!} [D^\alpha V(x_{i+1}) + D^\alpha V(x_i)] \Delta x_i^\alpha \\ &- \frac{1}{2} \sum_{|\alpha|=2,4,...}^{k+1} \frac{1}{\alpha!} [D^\alpha V(x_{i+1}) - D^\alpha V(x_i)] \Delta x_i^\alpha + \frac{1}{2} \sum_{|\alpha|=k+2} [R_\alpha^0(x_i, x_{i+1}) + R_\alpha^0(x_{i+1}, x_i)] \Delta x_i^\alpha \end{aligned} \tag{80}$$

Moreover, generalized trapezoidal formula (75) for $D^\alpha V$ with $|\alpha|$ even is

$$\begin{aligned} D^\alpha V(x_{i+1}) - D^\alpha V(x_i) &= \sum_{|\gamma|=1,3,...}^{k-|\alpha|} C_\gamma [D^{\alpha+\gamma} V(x_{i+1}) + D^{\alpha+\gamma} V(x_i)] \Delta x_i^\gamma \\ &+ \sum_{|\gamma|=1,3,...}^{k+2-|\alpha|} \sum_{|\beta|=k+2-|\alpha|-|\gamma|} B_\beta [R_\gamma^{\alpha+\beta}(x_i, x_{i+1}) + R_\gamma^{\alpha+\beta}(x_{i+1}, x_i)] \Delta x_i^{\beta+\gamma} \end{aligned} \tag{81}$$

Hence, substituting (81) into (80), a recursive Taylor series expansion

$$
\begin{aligned}
V(x_{i+1}) - V(x_i) &= \frac{1}{2}\sum_{|\alpha|=1,3,\ldots}^{k}\frac{1}{\alpha!}[D^\alpha V(x_{i+1}) + D^\alpha V(x_i)]\Delta x_i^\alpha \\
&- \frac{1}{2}\sum_{|\alpha|=2,4,\ldots}^{k+1}\frac{1}{\alpha!}\sum_{|\gamma|=1,3,\ldots}^{k-|\alpha|} C_\gamma[D^{\alpha+\gamma}V(x_{i+1}) + D^{\alpha+\gamma}V(x_i)]\Delta x_i^{\alpha+\gamma} \\
&- \frac{1}{2}\sum_{|\alpha|=2,4,\ldots}^{k+1}\frac{1}{\alpha!}\sum_{|\gamma|=1,3,\ldots}^{k+2-|\alpha|}\sum_{|\beta|=k+2-|\alpha|-|\gamma|} B_\beta[R_\gamma^{\alpha+\beta}(x_i, x_{i+1}) + R_\gamma^{\alpha+\beta}(x_{i+1}, x_i)]\Delta x_i^{\alpha+\beta+\gamma} \\
&+ \frac{1}{2}\sum_{|\alpha|=k+2}[R_\alpha^0(x_i, x_{i+1}) + R_\alpha^0(x_{i+1}, x_i)]\Delta x_i^\alpha \\
&= \frac{1}{2}\sum_{|\alpha|=1,3,\ldots}^{k}\frac{1}{\alpha!}[D^\alpha V(x_{i+1}) + D^\alpha V(x_i)]\Delta x_i^\alpha \\
&- \frac{1}{2}\sum_{|\alpha|=3,5,\ldots}^{k}\sum_{|\gamma|=1,3,\ldots}^{|\alpha|-2}\frac{1}{(\alpha-\gamma)!}C_\gamma[D^\alpha V(x_{i+1}) + D^\alpha V(x_i)]\Delta x_i^\alpha \\
&+ \frac{1}{2}\sum_{|\alpha|=k+2}\sum_{|\beta|=k+2-|\alpha|}[R_\alpha^\beta(x_i, x_{i+1}) + R_\alpha^\beta(x_{i+1}, x_i)]\Delta x_i^\alpha \\
&- \frac{1}{2}\sum_{|\alpha|=1,3,\ldots}^{k}\sum_{|\beta|=k+2-|\alpha|}\sum_{|\gamma|=2,4,\ldots}^{|\beta|}\frac{1}{\gamma!}B_{\beta-\gamma}[R_\alpha^\beta(x_i, x_{i+1}) + R_\alpha^\beta(x_{i+1}, x_i)]\Delta x_i^{\alpha+\beta}
\end{aligned}
\tag{82}
$$

is obtained after few rearrangements of the sums. Equating the same powers of (82) and (75), the coefficients C_α and B_β are obtained.

Up to now, we present how to compute the coefficients of the generalized trapezoidal formula. A rigorous proof of the lemma is then easily derived by induction on the order, k, of (75) and proceeding on the reverse direction of the above formulas. □

Lemma A.2. *Assume that the discrete-time Markov process x_i driven by*

$$x_{i+1} = F(x_i, \Delta W_i) \tag{83}$$

where ΔW_i are i.i.d. Gaussian random variables is ergodic with invariant measure $\bar{\mu}$. Then,

(i) *For sufficiently smooth function h we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} h(x_i, \Delta W_i) = \mathbb{E}_{\bar{\mu}\times\rho}[h(x, y)] \tag{84}$$

(ii) *For sufficiently smooth functions f and g we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(x_i)g(\Delta W_i) = \mathbb{E}_{\bar{\mu}}[f(x)]\mathbb{E}_\rho[g(y)] \tag{85}$$

(iii) *For sufficiently smooth functions f and g and for bounded f holds that*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(x_i,\Delta W_i)g(\Delta W_i) = \mathbb{E}_{\bar{\mu}\times\rho}[f(x,y)]\mathbb{E}_{\rho}[g(y)] \tag{86}$$

where ρ is always the Gaussian measure.

Proof. Proving (i) is based on showing that the transition density of the joint process $z_i = (x_i, \Delta W_i)$ exists and it is positive. Both are trivial since the transition density is the product of the two densities which are both positive. Thus, irreducibility for the joint process is proved and in combination with stationarity the joint process is ergodic.

(ii) is a direct consequence of (i) for $h(x,y) = f(x)g(y)$.

Denoting $\bar{f} = \mathbb{E}_{\bar{\mu}\times\rho}[f(x,y)]$ and $\bar{g} = \mathbb{E}_{\rho}[g(y)]$, (iii) is proved applying (i) and that

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=0}^{n-1} f(x_i,\Delta W_i)g(\Delta W_i) - \bar{f}\bar{g}\right| \\
&= \left|\frac{1}{n}\sum_{i=0}^{n-1} f(x_i,\Delta W_i)g(\Delta W_i) - \frac{1}{n}\sum_{i=0}^{n-1} f(x_i,\Delta W_i)\bar{g} + \frac{1}{n}\sum_{i=0}^{n-1} f(x_i,\Delta W_i)\bar{g} - \bar{f}\bar{g}\right| \\
&\le M\left|\frac{1}{n}\sum_{i=0}^{n-1} g(\Delta W_i) - \bar{g}\right| + |\bar{g}|\left|\frac{1}{n}\sum_{i=0}^{n-1} f(x_i,\Delta W_i) - \bar{f}\right|
\end{aligned} \tag{87}$$

since f is bounded (i.e., $|f| \le M$). Hence, sending $n\to\infty$, (iii) is proved.

□