MATHEMATICS OF COMPUTATION
Volume 81, Number 280, October 2012, Pages 2071–2125
S 0025-5718(2012)02594-4
Article electronically published on February 22, 2012

LANGEVIN DYNAMICS WITH CONSTRAINTS AND COMPUTATION OF FREE ENERGY DIFFERENCES

TONY LELIÈVRE, MATHIAS ROUSSET, AND GABRIEL STOLTZ

Abstract. In this paper, we consider Langevin processes with mechanical constraints. The latter are a fundamental tool in molecular dynamics simulation for sampling purposes and for the computation of free energy differences. The results of this paper can be divided into three parts. (i) We propose a simple discretization of the constrained Langevin process based on a splitting strategy. We show how to correct the scheme so that it samples *exactly* the canonical measure restricted on a submanifold, using a Metropolis-Hastings correction in the spirit of the Generalized Hybrid Monte Carlo (GHMC) algorithm. Moreover, we obtain, in some limiting regime, a consistent discretization of the overdamped Langevin (Brownian) dynamics on a submanifold, also sampling exactly the correct canonical measure with constraints. (ii) For free energy computation using thermodynamic integration, we rigorously prove that the longtime average of the Lagrange multipliers of the constrained Langevin dynamics yields the gradient of a rigid version of the free energy associated with the constraints. A second order time discretization using the Lagrange multipliers is proposed. (iii) The Jarzynski-Crooks fluctuation relation is proved for Langevin processes with mechanical constraints evolving in time. An original numerical discretization without time discretization error is proposed, and its overdamped limit is studied. Numerical illustrations are provided for (ii) and (iii).

1. Introduction and main results

Free energy is a central concept in thermodynamics and in modern works on biochemical and physical systems. Typical examples studied by computer simulations include the solvation free energies (which is the free energy difference between a molecule *in vacuo* and its counterpart surrounded by solvent molecules) and the binding free energy of two molecules (which determines whether a new drug can have an efficient action on a given protein). In many applications, it is actually the free energy difference profile between the initial and the final state which is a quantity of paramount importance. It is observed by practitioners that free energy barriers are a very important element to describe transition kinetics from one state to the other. For instance, the chemical kinetics of reactions happening in solvents (such as in the cells of our bodies) are limited by free energy barriers, and can take place only when the free energy difference between the initial and the final

Received by the editor June 23, 2010 and, in revised form, April 18, 2011 and June 24, 2011.
2010 *Mathematics Subject Classification.* Primary 82B80, 65C30; Secondary 82B35.
Key words and phrases. Constrained stochastic differential equations, free energy computations, nonequilibrium dynamics.
We would like to thank the anonymous referee for a careful reading of the manuscript and useful suggestions. This work was supported by the Agence Nationale de la Recherche, under the grant ANR-09-BLAN-0216-01 (MEGAS).

state is negative, or at least less than the typical thermal energy. It is therefore very important to accurately compute free energy differences in order to assess the likelihood of a certain physical event to happen.

Beside these physical motivations to compute free energy differences, a more abstract motivation is to overcome sampling barriers encountered when computing canonical averages (see the discussion in [35, Section 1.3.3]). Indeed, it is often the case in practice that the trajectories generated by the numerical methods at hand remain stuck for a long time in some region of the phase space, and hop only occasionally to another region, where they also remain stuck—a behavior known as metastability. Chemical and physical intuitions may guide the practitioners of the field towards the choice of some slowly evolving degree of freedom, called *reaction coordinate* in the following, responsible for the metastable behavior of the system. In this case, free energy techniques can be used to accelerate the sampling. This viewpoint allows us to consider applications which are not motivated by physical or biological problems, such as curing sampling issues in Bayesian statistics [7].

In this introductory section, we present the main results and give the outline of the paper, highlighting the three main contributions of this work. We only briefly define the concepts we need in this general introduction, and refer the reader to the following sections (in particular Section 2) for further precisions on the mathematical objects at hand.

1.1. **General setting for molecular dynamics with constraints.** We consider mechanical systems with constraints. The configuration of a classical N-body system is denoted by $(q,p) \in \mathbb{R}^{6N}$. The results of the paper can be generalized *mutatis mutandis* to periodic boundary conditions ($q \in \mathbb{T}^{3N}$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the one-dimensional torus), or to systems with positions confined in a domain $q \in \mathcal{D} \subset \mathbb{R}^{3N}$. The mass matrix of the system is assumed to be a constant strictly positive symmetric matrix M. One could typically think of a diagonal matrix $M = \mathrm{Diag}(m_1 \mathrm{Id}_3, \ldots, m_N \mathrm{Id}_3) \in \mathbb{R}^{3N \times 3N}$. The interaction potential is a smooth function $V : \mathbb{R}^{3N} \to \mathbb{R}$. The Hamiltonian of the system is assumed to be separable:

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

In the present paper, the focus is on the canonical ensemble, which is the equilibrium probability distribution of microscopic states of a system at fixed temperature (fixed average energy). For systems without constraints, this ensemble is characterized by the probability distribution

$$\mu(dq\,dp) = Z^{-1}\,\mathrm{e}^{-\beta H(q,p)}\,dq\,dp, \qquad Z = \int_{\mathbb{R}^{6N}} \mathrm{e}^{-\beta H}, \tag{1.1}$$

where Z is the normalizing constant[1] ensuring that μ is indeed a probability distribution, and $\beta = (k_{\mathrm{B}}T)^{-1}$ is proportional to the inverse temperature. One dynamics which admits the canonical measure (1.1) as an invariant measure is the Langevin dynamics (see for instance [35, Section 2.2.3] and references therein):

$$\begin{cases} dq_t = M^{-1} p_t\,dt, \\ dp_t = -\nabla V(q_t)\,dt - \gamma(q_t) M^{-1} p_t\,dt + \sigma(q_t)\,dW_t, \end{cases} \tag{1.2}$$

[1]The potential V is assumed to be such that $Z < \infty$.

where W_t is a standard $3N$-dimensional Brownian motion, and $\gamma(q), \sigma(q)$ are $3N \times 3N$ position dependent real matrices which are assumed to satisfy the fluctuation-dissipation identity

$$\sigma(q)\,\sigma^T(q) = \frac{2}{\beta}\gamma(q). \tag{1.3}$$

The Langevin dynamics can be seen as some modification of the Hamiltonian dynamics with two added components: a damping term $-\gamma(q_t)M^{-1}p_t\,dt$ and a random forcing term $\sigma(q_t)\,dW_t$. The energy dissipation due to the damping is compensated by the random forcing in such a way that the temperature of the system is $T = (k_B\beta)^{-1}$ (with $k_{\rm B}$ Boltzmann's constant).

We will consider positions subject to an m-dimensional mechanical constraint denoted by

$$\xi(q) = (\xi_1(q), \ldots, \xi_m(q))^T = z \in \mathbb{R}^m.$$

As will become clear below, constrained systems appear in computational statistical physics in two kinds of contexts (see *e.g.* [43, Chapter 10], and [12, 35] for applications to the computation of free energy differences, and [3, 32] for mathematical textbooks dealing with constrained Hamiltonian dynamics):

(i) for free energy computations, where ξ is a given reaction coordinate parameterizing a transition between "states" of interest;
(ii) when the system is subject to molecular constraints such as rigid covalent bonds, or rigid bond angles in molecular systems.

In the sequel, ξ may be thought of at first reading as a reaction coordinate (case (i)). Section 4.1 explains how to handle additional molecular constraints (case (ii)) within the same formalism. In any case, the position of the system is constrained onto the submanifold of codimension m:

$$\Sigma(z) = \Big\{ q \in \mathbb{R}^{3N} \;\Big|\; \xi(q) = z \Big\}, \tag{1.4}$$

and the associated phase space is the so-called cotangent bundle denoted by

$$T^*\Sigma(z) = \Big\{ (q,p) \in \mathbb{R}^{6N} \;\Big|\; q \in \Sigma(z),\ \nabla\xi(q)^T M^{-1} p = 0 \Big\}. \tag{1.5}$$

For a given $q \in \Sigma(z)$, the set of cotangent momenta is denoted by

$$T^*_q\Sigma(z) = \Big\{ p \in \mathbb{R}^{3N} \;\Big|\; \nabla\xi(q)^T\, M^{-1} p = 0 \Big\}. \tag{1.6}$$

The orthogonal projection on $T^*_q\Sigma(z)$ with respect to the scalar product induced by M^{-1} is denoted

$$P_M(q) = \mathrm{Id} - \nabla\xi(q)\, G_M^{-1}(q) \nabla\xi(q)^T M^{-1}, \tag{1.7}$$

where $G_M(q)$ is the so-called Gram matrix associated with the constraints

$$G_M(q) = \nabla\xi(q)^T M^{-1}\, \nabla\xi(q). \tag{1.8}$$

Throughout the paper, we assume that G_M is invertible everywhere on $\Sigma(z)$ (for all z). It is easily checked that P_M satisfies the projector property $P_M(q)^2 = P_M(q)$, and the orthogonality property

$$M^{-1} P_M(q) = P_M(q)^T M^{-1}.$$

1.2. **The constrained Langevin dynamics.** For constrained systems, the associated canonical distribution is defined by

$$\mu_{T^*\Sigma(z)}(dq\,dp) = Z_{z,0}^{-1}\,\mathrm{e}^{-\beta H(q,p)}\,\sigma_{T^*\Sigma(z)}(dq\,dp), \tag{1.9}$$

where $\sigma_{T^*\Sigma(z)}(dq\,dp)$ is the phase space Liouville measure of $T^*\Sigma(z)$, and $Z_{z,0}$ the normalizing constant (z refers to the position constraint, and 0 to the velocity or momentum constraint, see (2.15) below). See Section 2.3 for precise definitions.

A dynamics admitting the constrained canonical measure (1.9) as an invariant equilibrium measure is the following Langevin process ("CL" stands for "constrained Langevin"): For a given initial condition $(q_0, p_0) \in T^*\Sigma(z)$,

$$\text{(CL)} \qquad \begin{cases} dq_t = M^{-1}p_t\,dt, \\ dp_t = -\nabla V(q_t)\,dt - \gamma(q_t)M^{-1}p_t\,dt + \sigma(q_t)\,dW_t + \nabla\xi(q_t)\,d\lambda_t, \\ \xi(q_t) = z, \qquad (C_q) \end{cases}$$

where the $\mathbb{R}^m$-valued adapted[2] process $t \mapsto \lambda_t$ is the Lagrange multiplier associated with the (vectorial) constraint (C_q), and $\gamma(q), \sigma(q)$ are again assumed to satisfy (1.3). Note that $(q_t, p_t) \in T^*\Sigma(z)$ for all $t \geq 0$. Then, averages of an observable $A : \mathbb{R}^{6N} \to \mathbb{R}$ with respect to the distribution (1.9) can be obtained as longtime averages along any trajectory of the dynamics (CL) (when $P_M\gamma P_M^T$ is symmetric positive on $\Sigma(z)$):

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T A(q_t, p_t)\,dt = \int_{T^*\Sigma(z)} A\,d\mu_{T^*\Sigma(z)} \qquad \text{a.s.} \tag{1.10}$$

This is made precise in Section 3. Several recent studies (*e.g.* [24, 25, 47, 9]) have analyzed dynamics similar to (CL) and some appropriate discretization of the process in order to approximate the left-hand side of (1.10).

The first contribution of our work is to propose a simple discretization of the dynamics (CL) and to highlight its remarkable properties. The numerical scheme is based on a splitting strategy between the Hamiltonian and the thermostat part; see equations (3.16)-(3.17)-(3.18) below (in the spirit of the scheme proposed in [4] in the unconstrained case). The Hamiltonian part is discretized using a Verlet scheme with position and momentum constraints (the so-called RATTLE scheme, see [33]). We show that this discretization enjoys the following properties: (i) for some choice of the parameters, an Euler discretization of the *overdamped* Langevin dynamics (also called Brownian dynamics) with a projection step associated with the constraints is obtained (see equation (3.23) and Proposition 3.6); (ii) it can be completed by a Metropolis-Hastings correction to obtain a Generalized Hybrid Monte Carlo (GHMC) method sampling *exactly* (*i.e.* without any bias due to time-discretization) the constrained canonical distribution (1.9) (see Algorithm 3.5 below). The so-obtained numerical scheme is close to the ones proposed in [25, 24, 23]. See also [16, 37] for historic references on Hybrid Monte Carlo methods, and [28] for GHMC. One output of this part is thus a new Metropolization procedure for overdamped Langevin dynamics to sample, without bias, measures with support a submanifold.

[2] *I.e.* a random variable depending only on the past values of the Brownian motion.

1.3. **Free energy computations.** The free energy $F : \mathbb{R}^m \to \mathbb{R}$ associated with the reaction coordinates $\xi : \mathbb{R}^{3N} \to \mathbb{R}^m$ is defined as $-\beta^{-1}$ times the log-density of the marginal probability distribution of the reaction coordinates ξ under the canonical distribution (1.1). Explicitly, it is defined through the following relation: for any test function $\phi : \mathbb{R}^m \to \mathbb{R}$,

$$\int_{\mathbb{R}^m} \phi(z)\, \mathrm{e}^{-\beta F(z)}\, dz = \int_{\mathbb{R}^{6N}} \phi(\xi(q))\, \mu(dq\, dp). \tag{1.11}$$

In other words, $\mathrm{e}^{-\beta F(z)}\, dz$ is the image of the measure μ by ξ, and F can be seen as an "effective potential energy" associated to ξ.

Computing the free energy profile $z \mapsto F(z)$ (up to an additive constant independent of z), or free energy differences between two states $F(z_2) - F(z_1)$ is a way to compare the relative probabilities of different "states" parameterized by ξ. This is a very important calculation for practical applications; see [6, 35]. A state should be understood here as the collection of all possible microscopic configurations (q, p), distributed according to the canonical measure (1.1), and satisfying the macroscopic constraint $\xi(q) = z$. Since we only focus on computing free energy differences, F is defined up to an additive constant (independent of z, denoted by C below, and whose value may vary from line to line) and can be rewritten as:

$$\begin{aligned} F(z) &= -\frac{1}{\beta} \ln \int_{\Sigma(z)\times\mathbb{R}^{3N}} \mathrm{e}^{-\beta H(q,p)}\, \delta_{\xi(q)-z}(dq)\, dp \\ &= -\frac{1}{\beta} \ln \int_{\Sigma(z)} \mathrm{e}^{-\beta V(q)}\, \delta_{\xi(q)-z}(dq) + \mathrm{C}, \end{aligned} \tag{1.12}$$

where $\delta_{\xi(q)-z}$ denotes the conditional measure on $\Sigma(z)$ satisfying the following identity of measures in $\mathbb{R}^{3N}$: $dq = \delta_{\xi(q)-z}(dq)\, dz$ (see Section 2.3 for more precisions on this relation).

However, when using constrained simulations in phase space, the momentum variable of the dynamical system is also constrained, and a modified free energy (called "rigid free energy" in the sequel, see Remark 3.3 below for a justification of the term "rigid") is more naturally computed; see Section 3. The latter is defined as

$$F^M_{\mathrm{rgd}}(z) = -\frac{1}{\beta} \ln \int_{T^*\Sigma(z)} \mathrm{e}^{-\beta H(q,p)} \sigma_{T^*\Sigma(z)}(dq\, dp). \tag{1.13}$$

The superscript M indicates that this free energy depends on the considered mass matrix, even though this is not clear at this stage (see (4.3) below). The above two definitions of free energy are related through the identity:

$$F(z) - F^M_{\mathrm{rgd}}(z) = -\frac{1}{\beta} \ln \int_{T^*\Sigma(z)} (\det G_M)^{-1/2} d\mu_{T^*\Sigma(z)} + \mathrm{C}, \tag{1.14}$$

where $\mu_{T^*\Sigma(z)}$ is the equilibrium distribution with constraints (1.9). The relation (1.14), already proposed in [12] (see also [14, 18, 45, 26] for related formulas), is proved at the beginning of Section 4. For any value of the reaction coordinate, the difference $F(z) - F^M_{\mathrm{rgd}}(z)$ can then be easily computed with any method sampling the probability distribution $\mu_{T^*\Sigma(z)}$, such as (CL).

Several methods have been suggested in the literature to compute either F or F^M_{rgd} from the Lagrange multipliers of a constrained process similar to (CL). We refer, for instance, to [12] (and references therein) for the Hamiltonian case, and

to [9] (and references therein) for the overdamped case. The second contribution of this paper is twofold: (i) we rigorously prove that the longtime average of the Lagrange multipliers in (CL) converges to the gradient of the rigid free energy (1.13) (the so-called *mean force*); (ii) we then show that the latter mean-force can be computed with second order accuracy (*i.e.* up to $\mathrm{O}(\Delta t^2)$ error terms, where Δt is the time-step) using the Lagrange multipliers involved in the Hamiltonian part of the splitting scheme.

More precisely, the first point (i) amounts to showing that

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T d\lambda_t = \nabla_z F^M_{\mathrm{rgd}}(z) \qquad \text{a.s.} \tag{1.15}$$

As compared to [12], where a formal proof for the Hamiltonian case is proposed, we use an explicit calculation that does not require the use of the Lagrangian structure of the problem, or a change of coordinates. Once $\nabla_z F^M_{\mathrm{rgd}}(z)$ is obtained, $F^M_{\mathrm{rgd}}(z)$ can be computed (up to an additive constant) by integration. This procedure is known as *thermodynamic integration*. Note that using (1.15) and thermodynamic integration, together with (1.14), allows us to obtain $F(z)$ without computing second order derivatives of ξ. This is a desirable property since computing such high derivatives may be cumbersome for some reaction coordinates used in practice. Straightforward computations of the mean force using analytical expressions (see for instance (4.11)-(4.12)) usually involve such high order derivatives.

The second point (ii) is then based on a discretization of a variant of (1.15), obtained by subtracting the martingale part of the Lagrange multipliers. This amounts to averaging the two Lagrange multipliers involved in the RATTLE part of the scheme; see (4.17)).

We also discuss how these techniques can be generalized to compute the free energy for systems with molecular constraints; see Section 4.1.

1.4. **Jarzynski-Crooks relations and nonequilibrium computations of the free energy.** The last part of this article is devoted to nonequilibrium methods for free energy computations, based on a Hamiltonian or Langevin dynamics with constraints subject to a predetermined time evolution. Such methods rely on a nonequilibrium fluctuation equality, the so-called Jarzynski-Crooks relation. See [29] for a pioneering work, as well as [10, 11] for an extension. They are termed "nonequilibrium" since the transition from one value of the reaction coordinate ξ to another one is imposed *a priori*, in a finite time T, and with a given smooth deterministic schedule $t \in [0,T] \mapsto z(t) \in \mathbb{R}^m$. In particular, it may be arbitrarily fast. Therefore, even if the system starts at equilibrium, it does not remain at equilibrium. The out-of-equilibrium Langevin process we consider to this end is given by the following equations of motion ("SCL" stands for "switched constrained Langevin"):

$$\text{(SCL)} \qquad \begin{cases} dq_t = M^{-1}p_t\,dt, \\ dp_t = -\nabla V(q_t)\,dt - \gamma_P(q_t)M^{-1}p_t\,dt + \sigma_P(q_t)\,dW_t + \nabla\xi(q_t)\,d\lambda_t, \\ \xi(q_t) = z(t), \qquad (C_q(t)) \end{cases}$$

where $t \mapsto \lambda_t \in \mathbb{R}^m$ is an adapted process enforcing the constraints $(C_q(t))$ (the Lagrange multipliers). Initial conditions are sampled from the phase-space canonical distribution defined by the constraints $\xi(q) = z(0)$ and $v_\xi(q,p) = \nabla\xi(q)^T M^{-1}p =$

$\dot{z}(0)$ (see (2.15)). We restrict ourselves to projected fluctuation-dissipation matrices of the specific form

$$(\sigma_P, \gamma_P) := (P_M\,\sigma, P_M\,\gamma\,P_M^T), \tag{1.16}$$

where $\gamma(q), \sigma(q) \in \mathbb{R}^{3N\times 3N}$ satisfy the fluctuation-dissipation identity (1.3). Note that γ_P, σ_P also satisfy (1.3). Our analysis also applies to deterministic Hamiltonian dynamics upon choosing $\gamma = 0$. The dynamics (SCL) is a natural extension of the constrained Langevin dynamics (CL). It is different from the dynamics proposed in [31], which is a Langevin dynamics associated with a modified Hamiltonian with projected momenta, driven by a forcing term along $\nabla\xi$ which acts directly on the position variable. As explained below (see (5.3) and the discussion following this equation), the specific choice (1.16) (rather than considering unprojected matrices $\gamma(q), \sigma(q) \in \mathbb{R}^{3N\times 3N}$) leads to a simpler analysis and more natural numerical schemes, based again on a splitting procedure.

As explained in Section 5.3, it is possible to define the work associated with the constraints exerted on the system between time 0 and T as the displacement multiplied by the constraining force:

$$\mathcal{W}_{0,t}\left(\{q_s, p_s\}_{0\le s\le t}\right) := \int_0^t \dot{z}^T(s)\, d\lambda_s. \tag{1.17}$$

The third contribution of the present paper is twofold: (i) We derive a new general Crooks-Jarzynski relation (see Theorem 5.3 below) based on the nonequilibrium constrained dynamics (SCL) and the associated work defined in (1.17); and (ii) An original numerical scheme is proposed, which allows us to compute free energy differences *without time discretization error* (see Theorem 5.5 below). More precisely, concerning the first point, the main corollary is given by the following result. Consider the corrector

$$C(t,q) = \frac{1}{2\beta}\ln\Big(\det G_M(q)\Big) - \frac{1}{2}\dot{z}(t)^T G_M^{-1}(q)\dot{z}(t), \tag{1.18}$$

where $\frac{1}{2\beta}\ln\det G_M(q)$ is the so-called Fixman term due to the geometry of the position constraints (see (1.14) and Remark 3.3), and $\frac{1}{2}\dot{z}(t)^T G_M^{-1}(q)\dot{z}(t)$ is the kinetic energy term due to the velocity of the switching. Then, the free energy profile can be computed through the following fluctuation identity (see (5.22)):

$$F(z(T)) - F(z(0)) = -\frac{1}{\beta}\ln\left(\frac{\mathbb{E}\left(\mathrm{e}^{-\beta\left[\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{0\le t\le T}\right)+C(T,q_T)\right]}\right)}{\mathbb{E}\left(\mathrm{e}^{-\beta C(0,q_0)}\right)}\right), \tag{1.19}$$

where the expectation is with respect to canonical (equilibrium) initial conditions and for all realizations of the dynamics (SCL).

The numerical scheme mentioned in the second point (ii) above is based on a modification of the splitting scheme used to discretize the constrained Langevin dynamics (CL). This modification allows us to take into account the evolving constraints. Using the symplecticity of the modified RATTLE scheme, we are able to prove a discrete-in-time version of the Crooks relation, and of the associated Jarzynski free energy estimator (1.19). Moreover, for some choice of the parameters, the latter scheme yields a Jarzynski-Crooks relation for an Euler discretization of the overdamped Langevin (Brownian) dynamics with a projection step associated with the evolving constraints, without time discretization error. This can be seen

as an extension of the scheme formerly proposed in [34] (see equation (5.52) and Proposition 5.6). We also check the consistency of the various free energy estimators we introduce.

1.5. **Organization of the paper.** We start with an introduction to the mathematical concepts required for mechanically constrained systems in Section 2. Section 3 is devoted to the properties and the discretization of mechanically constrained Langevin processes defined by (CL), and the problem of sampling the canonical distribution (1.9). Thermodynamic integration with constrained Langevin processes is presented in Section 4. Section 5 discusses nonequilibrium constrained Langevin processes (SCL) and the associated Jarzynski-Crooks fluctuation identity (1.19).

2. Preliminaries

After making precise our notation for matrices and matrix valued functions in Section 2.1, we introduce some additional concepts required to describe constrained systems in Section 2.2, and define the phase space measures with constraints in Section 2.3.

2.1. **Notation.** Throughout the paper, the following notation is used:

- Vectors and vector fields are by convention of column type. When vectors are written as a line, they should be understood as the corresponding column version. For instance, $(q,p) \in \mathbb{R}^{6N}$ should be understood as $(q^T,p^T)^T$, where $q,p \in \mathbb{R}^{3N}$ are both column vectors.
- Gradients in $\mathbb{R}^{3N}$ (or $\mathbb{R}^{6N}$) of m-dimensional vector fields are by convention $3N \times m$-matrices, for instance,

$$\nabla \xi(q) = \Big(\nabla \xi_1(q), \ldots, \nabla \xi_m(q) \Big) \in \mathbb{R}^{3N \times m},$$

  where $\nabla \xi_i(q) \in \mathbb{R}^{3N}$ is a column vector for any $i = 1, \ldots, m$. Gradients in the space of constraints parameters $z \in \mathbb{R}^m$ or $\zeta \in \mathbb{R}^{2m}$ are denoted with the associated subscripts, namely ∇_z and ∇_ζ.
- Second order derivatives in $\mathbb{R}^{3N}$ of m-dimensional vector fields are characterized through the Hessian bilinear form

$$\mathrm{Hess}_q(\xi)\big(v_1,v_2\big) = \begin{pmatrix} v_1^T \nabla^2 \xi_1(q) v_2 \\ \vdots \\ v_1^T \nabla^2 \xi_m(q) v_2 \end{pmatrix} \in \mathbb{R}^m, \tag{2.1}$$

  where $v_1, v_2 \in \mathbb{R}^{3N}$ are test vectors.
- The canonical symplectic matrix is denoted by

$$J := \begin{pmatrix} 0 & \mathrm{Id}_{3N} \\ -\mathrm{Id}_{3N} & 0 \end{pmatrix} \in \mathbb{R}^{6N \times 6N}. \tag{2.2}$$

  For any smooth test functions $\varphi_1 : \mathbb{R}^{6N} \to \mathbb{R}^{n_1}$ and $\varphi_2 : \mathbb{R}^{6N} \to \mathbb{R}^{n_2}$, the Poisson bracket is the $n_1 \times n_2$ matrix

$$\{\varphi_1, \varphi_2\} = \big(\nabla \varphi_1\big)^T J \, \nabla \varphi_2 \in \mathbb{R}^{n_1 \times n_2}. \tag{2.3}$$

- For two matrices $A, B \in \mathbb{R}^{n \times n}$, $A : B = \mathrm{Tr}(A^T B)$.

2.2. **Constraints.** Contrarily to what is often done in the literature, we avoid global changes of variables and the use of generalized coordinates. We observe that global changes of variables are not required for the proofs of the theoretical results we present, and they are definitely to be avoided in practical numerical computations whenever possible.

Two useful concepts to study constrained Hamiltonian systems (in particular, to use the co-area formula in phase space, as well as the Poisson bracket formulation of the Liouville equation) are the effective velocity v_ξ and the effective momentum p_ξ associated with the constrained degrees of freedom ξ:

$$v_\xi(q,p) = \nabla\xi(q)^T M^{-1} p \in \mathbb{R}^m \tag{2.4}$$

and

$$p_\xi(q,p) = G_M^{-1}(q)\, v_\xi(q,p) = G_M^{-1}(q)\nabla\xi(q)^T M^{-1} p \in \mathbb{R}^m. \tag{2.5}$$

Let us emphasize that what we call here the "effective momentum" is in general different from the "conjugate momentum" introduced in Hamiltonian mechanics.

The expression of the effective velocity is obtained by deriving the constraint ξ along an unconstrained trajectory of the Hamiltonian dynamics

$$\begin{cases} \dfrac{d\tilde q_t}{dt} = M^{-1}\tilde p_t, \\ \dfrac{d\tilde p_t}{dt} = -\nabla V(\tilde q_t), \end{cases}$$

since $\frac{d\xi(\tilde q_t)}{dt} = v_\xi(\tilde q_t, \tilde p_t)$. The term $G_M^{-1}(q)$ in the expression (2.5) of the effective momentum may be interpreted as the effective mass of ξ. This can be motivated by a decomposition of the kinetic energy of the system into tangential and orthogonal parts, using the projector (1.7) for a given position $q \in \mathbb{R}^{3N}$:

$$\begin{aligned} E_{\rm kin}(p) &= \frac12\, p^T M^{-1} p \\ &= \frac12\, p^T P_M(q)^T M^{-1} P_M(q) p + \frac12\, p^T \big(\mathrm{Id} - P_M(q)\big)^T M^{-1} \big(\mathrm{Id} - P_M(q)\big) p. \end{aligned}$$

The orthogonal part can be rewritten, for any $(q,p) \in \mathbb{R}^{6N}$, as:

$$\begin{aligned} E_{\rm kin}^\perp(q,p) &:= \frac12\, p^T \Big(\mathrm{Id} - P_M(q)\Big)^T M^{-1} \Big(\mathrm{Id} - P_M(q)\Big) p \\ &= \frac12\, v_\xi(q,p)^T G_M^{-1}(q)\, v_\xi(q,p) = \frac12\, p_\xi(q,p)^T G_M(q)\, p_\xi(q,p). \end{aligned}$$

The last equations allow us to consider G_M^{-1} as some effective mass.

The constraints on a mechanical system can also be reformulated in the more general form

$$\Xi(q,p) = \zeta \in \mathbb{R}^{2m}, \tag{2.6}$$

where either (i) the effective momentum is constrained, in which case $\Xi = (\xi, p_\xi)$ and $\zeta = (z, p_z)$; or (ii) the effective velocity is constrained, in which case $\Xi = (\xi, v_\xi)$ and $\zeta = (z, v_z)$. The phase space associated with such constraints is denoted by

$$\Sigma_\Xi(\zeta) = \Big\{ (q,p) \in \mathbb{R}^{6N} \;\Big|\; \Xi(q,p) = \zeta \Big\}. \tag{2.7}$$

A position $q \in \Sigma(z)$ being given, the affine space of constrained momenta satisfying (2.6) is then denoted by

$$\Sigma_{v_\xi(q,\cdot)}(v_z) = \left\{ p \in \mathbb{R}^{3N} \;\middle|\; v_\xi(q,p) = v_z \right\} \tag{2.8}$$

in the effective velocity case, and by $\Sigma_{p_\xi(q,\cdot)}(p_z)$ in the effective momentum case. This notation is very important for nonequilibrium methods where the constraints evolve in time according to a predefined schedule; see Section 5. Note that the phase space of mechanical constraints, defined by (1.5), is simply $T^*\Sigma(z) = \Sigma_{\xi,v_\xi}(z,0) = \Sigma_{\xi,p_\xi}(z,0)$.

We can now define the so-called Gram tensor associated with the constraints, which is a skew-symmetric matrix of dimension $2m \times 2m$:

$$\Gamma(q,p) = \{\Xi,\Xi\}(q,p) = \nabla\Xi^T(q,p)\, J\, \nabla\Xi(q,p) \in \mathbb{R}^{2m\times 2m}. \tag{2.9}$$

The Gram matrix Γ associated with the generalized constraints (2.6) can be explicitly computed by block. Indeed, for $\Xi = (\xi, p_\xi)^T$,

$$\Gamma = \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & \nabla p_\xi^T J\, \nabla p_\xi \end{pmatrix}. \tag{2.10}$$

Therefore, $\det(\Gamma) = 1$ in this case. In the case $\Xi = (\xi, v_\xi)^T$, the Gram matrix reads

$$\Gamma = \begin{pmatrix} 0 & G_M \\ -G_M & \nabla v_\xi^T\, J\, \nabla v_\xi \end{pmatrix}, \tag{2.11}$$

and $\det(\Gamma) = \det(G_M)^2$. Note that in both cases $\det(\Gamma) > 0$. We then define the skew-symmetric matrix

$$J_\Xi(q,p) = J - J\,\nabla\Xi(q,p)\,\Gamma^{-1}(q,p)\,\nabla\Xi^T(q,p)\,J, \tag{2.12}$$

and the Poisson bracket associated with generalized constraints (2.6) by

$$\{\varphi_1,\varphi_2\}_\Xi = \nabla\varphi_1^T J_\Xi \nabla\varphi_2. \tag{2.13}$$

This Poisson bracket is often called the Dirac bracket in the literature (in reference to the seminal work of Dirac [15, 38]).

It is easily checked that $\{\cdot,\cdot\}_\Xi$ satisfies the characteristic properties of Poisson brackets, namely the skew-symmetry, Jacobi's identity, and Leibniz' rule. Therefore, the flow associated with the evolution equation

$$\frac{d}{dt}\begin{pmatrix} q_t \\ p_t \end{pmatrix} = J_\Xi \nabla H(q_t,p_t), \tag{2.14}$$

defines a symplectic map. Recall that (see [22, Section VII.1.2]) a map $\phi : \Sigma_\Xi(\zeta) \to \Sigma_\Xi(\zeta)$ is symplectic if for any $(q,p) \in \Sigma_\Xi(\zeta)$ and $u, v \in T_{(q,p)}\Sigma_\Xi(\zeta)$,

$$u^T \nabla\phi(q,p)^T J \nabla\phi(q,p) v = u^T J v.$$

A consequence of the symplectic structure is the divergence formula (2.25) relating the phase space measure $\sigma_{\Sigma_\Xi(\zeta)}(dq\,dp)$ on $\Sigma_\Xi(\zeta)$ (defined below), and the Poisson bracket (2.13). The reader is referred to Chapter 8 in [3], and Section VII.1 in [22] for more material on constrained systems.

It will be shown in Proposition 3.1 below that the Poisson system (2.14) is equivalent to (CL) when $(\gamma,\sigma) = (0,0)$.

2.3. **Phase space measures.**

2.3.1. *Definitions.* The phase space measure (also termed Liouville measure) on the phase space $T^*\Sigma(z)$ (or more generally on $\Sigma_\Xi(\zeta)$) of constrained mechanical systems is denoted by $\sigma_{T^*\Sigma(z)}$ (or more generally $\sigma_{\Sigma_\Xi(\zeta)}$). The latter is induced by the symplectic, or skew-symmetric 2-form on $\mathbb{R}^{6N}$ defined by the canonical skew-symmetric matrix J in $\mathbb{R}^{6N}$. More precisely, it can be defined through the volume form $|\det \mathcal{G}(u(q,p))|^{1/2}$, where

$$\mathcal{G}_{a,b}(u) = (u_a)^T J u_b, \qquad a,b = 1,\dots,6N-2m,$$

and $(u_1(q,p),\dots,u_{6N-2m}(q,p))$ is a basis of tangential vectors of the submanifold $T^*\Sigma(z)$ (or $\Sigma_\Xi(\zeta)$) at a given point (q,p).

Surface measures induced by scalar products associated with general symmetric definite positive matrices will also be of interest. We denote by $\sigma^M_{\Sigma(z)}(dq)$ the surface measure on $\Sigma(z)$ induced by the scalar product $\langle q,\tilde q\rangle_M = q^T M\tilde q$ on $\mathbb{R}^{3N}$, and, for a given $q\in\Sigma(z)$, by $\sigma^{M^{-1}}_{\Sigma_{p_\xi(q,\cdot)}(p_z)}(dp)$ and $\sigma^{M^{-1}}_{\Sigma_{v_\xi(q,\cdot)}(v_z)}(dp)$ the surface measures on the affine spaces $\Sigma_{p_\xi(q,\cdot)}(p_z)$ and $\Sigma_{v_\xi(q,\cdot)}(v_z)$, respectively, induced by the scalar product $\langle p,\tilde p\rangle_{M^{-1}} = p^T M^{-1}\tilde p$ on $\mathbb{R}^{3N}$. For more precise definitions of these measures, we refer to [35, Sections 3.2.1 and 3.3.2] and the references therein.

It is now possible to define a generalization of the canonical distribution (1.9) as follows:

$$\begin{cases} \mu_{\Sigma_{\xi,v_\xi}(z,v_z)}(dq\,dp) := \dfrac{\mathrm{e}^{-\beta H(q,p)}}{Z_{z,v_z}}\sigma_{\Sigma_{\xi,v_\xi}(z,v_z)}(dq\,dp), \\ Z_{z,v_z} := \displaystyle\int_{\Sigma_{\xi,v_\xi}(z,v_z)} \mathrm{e}^{-\beta H}\, d\sigma_{\Sigma_{\xi,v_\xi}(z,v_z)}. \end{cases} \tag{2.15}$$

The distribution (2.15) is associated with the generalized constraints $\Xi = (\xi, v_\xi)^T$, and is used in Section 5 for nonequilibrium methods. Note that $\mu_{\Sigma_{\xi,v_\xi}(z,0)} = \mu_{T^*\Sigma(z)}$ defined in (1.9).

2.3.2. *Co-area decompositions.* The co-area formula (see [2, 20]) relates the phase space or surface measures, and the conditional measures. Conditional measures are defined in $\mathbb{R}^{6N}$ by the following conditioning formula: for any test function $\phi : \mathbb{R}^{6N}\to\mathbb{R}$,

$$\int_{\mathbb{R}^{6N}} \phi(q,p)\,dq\,dp = \int_{\mathbb{R}^{2m}}\int_{\Sigma_\Xi(\zeta)} \phi(q,p)\,\delta_{\Xi(q,p)-\zeta}(dq\,dp)\,d\zeta. \tag{2.16}$$

In the same way, conditional measures in $\mathbb{R}^{3N}$ are defined, for any test function $\phi : \mathbb{R}^{3N}\to\mathbb{R}$, by

$$\int_{\mathbb{R}^{3N}} \phi(q)\,dq = \int_{\mathbb{R}^m}\int_{\Sigma(z)} \phi(q)\,\delta_{\xi(q)-z}(dq)\,dz. \tag{2.17}$$

A more concise notation for the above equalities is $dq\,dp = \delta_{\Xi(q,p)-\zeta}(dq\,dp)\,d\zeta$ and $dq = \delta_{\xi(q)-z}(dq)\,dz$.

Proposition 2.1 (Co-area)**.** *Let $\Sigma(z)$ be the submanifold* (1.4) *defined by the constraints $\xi(q) = z$, and assume that G_M defined in* (1.8) *is nondegenerate in a neighborhood of $\Sigma(z)$. Then, in the sense of measures on $\mathbb{R}^{3N}$:*

$$\delta_{\xi(q)-z}(dq) = (\det M)^{-1/2}\left|\det G_M(q)\right|^{-1/2}\sigma^M_{\Sigma(z)}(dq). \tag{2.18}$$

Let $\Sigma_\Xi(\zeta)$ be the phase space defined by generalized constraints (2.6)*. Assume that Γ defined in* (2.9) *is nondegenerate in a neighborhood of $\Sigma_\Xi(\zeta)$. Then, in the sense of measures on $\mathbb{R}^{6N}$:*

$$\delta_{\Xi(q,p)-\zeta}(dq\,dp) = \left|\det \Gamma(q,p)\right|^{-1/2} \sigma_{\Sigma_\Xi(\zeta)}(dq\,dp). \tag{2.19}$$

We refer for example to Chapter 3 in [35] for an elementary proof. An equivalent of (2.17)-(2.18) for momenta reads, for constrained effective momenta:

$$dp = \delta_{p_\xi(q,p)-p_z}(dp)\,dp_z = \det(M)^{1/2}\left|\det G_M(q)\right|^{1/2} \sigma^{M^{-1}}_{\Sigma_{p_\xi(q,\cdot)}(p_z)}(dp)\,dp_z, \tag{2.20}$$

and for constrained effective velocities:

$$dp = \delta_{v_\xi(q,p)-v_z}(dp)\,dv_z = \det(M)^{1/2}\left|\det G_M(q)\right|^{-1/2} \sigma^{M^{-1}}_{\Sigma_{v_\xi(q,\cdot)}(v_z)}(dp)\,dv_z.$$

Using the co-area formulas (2.18)-(2.19), and the expressions of the Gram matrices (2.10)-(2.11), we obtain the following expressions of the phase space measures:

(i) The phase space measure on $\Sigma_{\xi,p_\xi}(z,p_z)$ can be identified with the conditional measure defined in (2.16):

$$\sigma_{\Sigma_{\xi,p_\xi}(z,p_z)}(dq\,dp) = \delta_{(\xi(q)-z,p_\xi(q,p)-p_z)}(dq\,dp), \tag{2.21}$$

while the phase space measure on $\Sigma_{\xi,v_\xi}(z,v_z)$ is related to the corresponding conditional measure as

$$\sigma_{\Sigma_{\xi,v_\xi}(z,v_z)}(dq\,dp) = \det(G_M)\,\delta_{(\xi(q)-z,v_\xi(q,p)-v_z)}(dq\,dp). \tag{2.22}$$

(ii) The phase space measures are given by the product of surface measures:

$$\sigma_{\Sigma_{\xi,p_\xi}(z,p_z)}(dq\,dp) = \sigma^{M^{-1}}_{\Sigma_{p_\xi(q,\cdot)}(p_z)}(dp)\,\sigma^{M}_{\Sigma(z)}(dq) \tag{2.23}$$

and

$$\sigma_{\Sigma_{\xi,v_\xi}(z,v_z)}(dq\,dp) = \sigma^{M^{-1}}_{\Sigma_{v_\xi(q,\cdot)}(v_z)}(dp)\,\sigma^{M}_{\Sigma(z)}(dq). \tag{2.24}$$

Equations (2.23)-(2.24) are a consequence of the fact that $\delta_{(\xi(q)-z,p_\xi(q,p)-p_z)}(dq\,dp) = \delta_{p_\xi(q,p)-p_z}(dp)\delta_{\xi(q)-z}(dq)$ (and a similar relation for v_ξ).

2.3.3. *Divergence formulas.* We end this section with an important formula, which is used to show the invariance of the canonical measure in the proof of Proposition 3.2.

Proposition 2.2 (Divergence theorem in phase space)**.** *Consider the Poisson bracket $\{\cdot,\cdot\}_\Xi$ defined by* (2.13)*, and an open neighborhood $\mathcal{O}$ of $\Sigma_\Xi(\zeta) \subset \mathbb{R}^{6N}$ where Γ is invertible. Then for any smooth test functions $\varphi_1, \varphi_2 : \mathbb{R}^{6N} \to \mathbb{R}$ with compact support in $\mathcal{O}$,*

$$\int_{\Sigma_\Xi(\zeta)} \{\varphi_1,\varphi_2\}_\Xi \, d\sigma_{\Sigma_\Xi(\zeta)} = 0. \tag{2.25}$$

The divergence formula (2.25) can be proved using Darboux's theorem and internal coordinates, or directly using the co-area formula (see Section 3.3 in [35]).

We will also need the classical divergence formula on affine spaces (see for instance Section 3.3 in [35]): for a fixed $q \in \mathbb{R}^{3N}$, for any compactly supported smooth vector field $\phi(q,p) \in \mathbb{R}^{3N}$,

$$\int_{\Sigma_{v_\xi(q,\cdot)}(v_z)} \mathrm{div}_p\Big(P_M(q)\phi(q,p)\Big)\sigma^{M^{-1}}_{\Sigma_{v_\xi(q,\cdot)}(v_z)}(dp) = 0. \tag{2.26}$$

2.3.4. *A useful technical result.* The following lemma, whose proof can be read in [36, Section 6] or [35, Section 3.3.6.3], is useful in the proofs of Proposition 4.1 and Theorem 5.3:

Lemma 2.3. *For any compactly supported smooth test function φ on $\mathbb{R}^{6N}$:*

$$\nabla_\zeta \left(\int_{\Sigma_\Xi(\zeta)} \varphi \, d\sigma_{\Sigma_\Xi(\zeta)} \right) = \int_{\Sigma_\Xi(\zeta)} \Gamma^{-1} \left\{ \Xi, \varphi \right\} \, d\sigma_{\Sigma_\Xi(\zeta)},$$

where the phase space $\Sigma_\Xi(\zeta)$ is defined in (2.7)*, and the Gram matrix Γ in* (2.9)*.*

3. Constrained Langevin processes and sampling

We first give some properties of the constrained Langevin equation (CL) in Section 3.1, then propose some numerical schemes to discretize it in Section 3.2, and finally consider the overdamped limit in Section 3.3.

3.1. **Properties of the dynamics.** We consider the dynamics (CL):

$$\begin{cases} dq_t = M^{-1} p_t \, dt, \\ dp_t = -\nabla V(q_t) \, dt - \gamma(q_t) M^{-1} p_t \, dt + \sigma(q_t) \, dW_t + \nabla \xi(q_t) \, d\lambda_t, \\ \xi(q_t) = z, \qquad (C_q). \end{cases}$$

By differentiating with respect to time the constraint $\xi(q_t) = z$, the Lagrange multipliers can be computed explicitly (see for instance Section 3.3 in [35]):

$$\begin{aligned} d\lambda_t &= -G_M^{-1}(q_t) \Big[\mathrm{Hess}_{q_t}(\xi) \big(M^{-1} p_t, M^{-1} p_t \big) \, dt \\ &\quad + \nabla \xi(q_t)^T M^{-1} \Big(- \nabla V(q_t) \, dt - \gamma(q_t) M^{-1} p_t \, dt + \sigma(q_t) \, dW_t \Big) \Big] \\ &= f^M_{\mathrm{rgd}}(q_t, p_t) \, dt + G_M^{-1}(q_t) \nabla \xi(q_t)^T M^{-1} \left(\gamma(q_t) M^{-1} p_t \, dt - \sigma(q_t) \, dW_t \right), \end{aligned} \tag{3.1}$$

where the constraining force $f^M_{\mathrm{rgd}} \in \mathbb{R}^m$ is defined as

$$f^M_{\mathrm{rgd}}(q,p) = G_M^{-1}(q) \nabla \xi(q)^T M^{-1} \nabla V(q) - G_M^{-1}(q) \mathrm{Hess}_q(\xi)(M^{-1}p, M^{-1}p). \tag{3.2}$$

Thus, using the fact that $P_M(q)^T M^{-1} p = M^{-1} p$ when $p \in T^*_q \Sigma(z)$, the dynamics (CL) can be recast in a more explicit form as

$$\begin{cases} dq_t = M^{-1} p_t \, dt, \\ dp_t = -\nabla V(q_t) \, dt + \nabla \xi(q_t) f^M_{\mathrm{rgd}}(q_t, p_t) \, dt - \gamma_P(q_t) M^{-1} p_t \, dt \\ \qquad\quad + \sigma_P(q_t) \, dW_t, \end{cases} \tag{3.3}$$

where we introduced the notation $(\sigma_P, \gamma_P) := (P_M \, \sigma, P_M \, \gamma \, P_M^T)$. The constraint therefore has two effects: (i) the matrices γ, σ in the dissipation and fluctuation terms are replaced by their projected counterparts γ_P, σ_P, and (ii) an orthogonal constraining force $\nabla \xi f^M_{\mathrm{rgd}}$ is introduced.

The generator of this stochastic Langevin dynamics is the operator $\mathcal{L}_\Xi$ which appears in the Kolmogorov evolution equation: for (q_t, p_t) satisfying (CL) or (3.3), and for any smooth test function φ,

$$\frac{d}{dt} \mathbb{E} \left(\varphi(q_t, p_t) \right) = \mathbb{E} \big(\mathcal{L}_\Xi(\varphi)(q_t, p_t) \big).$$

The expression of $\mathcal{L}_\Xi$ can be obtained using Itô calculus, as made precise in the following proposition.

Proposition 3.1. *Consider either the effective momentum* (2.5) *or the effective velocity* (2.4)*, denoted with the general constraints* $\Xi = (\xi, v_\xi)$ *or* $\Xi = (\xi, p_\xi)$ (*see* (2.6)). *The solution of the constrained dynamics* (CL) (*or equivalently* (3.3) *with an initial condition* $(q_0, p_0) \in \Sigma_\Xi(z, 0)$) *belongs to* $\Sigma_\Xi(z, 0) = T^*\Sigma(z)$*, and the generator of this Markov process reads (whatever the value of* z*)*

$$\mathcal{L}_\Xi = \{\cdot, H\}_\Xi + \mathcal{L}_\Xi^{\rm thm}, \tag{3.4}$$

where the fluctuation-dissipation part is

$$\mathcal{L}_\Xi^{\rm thm} = \frac{1}{2}{\rm div}_p\Big(\sigma_P\,\sigma_P^T\nabla_p\,\cdot\,\Big) - p^T M^{-1}\gamma_P\nabla_p,$$

with (σ_P, γ_P) *defined in* (1.16)*. Using the fluctuation-dissipation relation* (1.3)*, the generator* $\mathcal{L}_\Xi^{\rm thm}$ *can be rewritten more compactly as*

$$\mathcal{L}_\Xi^{\rm thm} = \frac{1}{\beta}\,{\rm e}^{\beta H}{\rm div}_p\Big(e^{-\beta H}\,\gamma_P\,\nabla_p\,\cdot\,\Big). \tag{3.5}$$

Proof. We perform the computation in two steps: (i) We compute the generator of the Hamiltonian part of the constrained Langevin dynamics, which is (CL) in the case $(\sigma, \gamma) = (0, 0)$; (ii) we compute the generator of the "thermostat" part of (CL), which is an Ornstein-Uhlenbeck process on momentum variable (corresponding to the second equation in (CL) with $V = 0$).

Let us first consider (i), with $\Xi = (\xi, v_\xi)$ (the case $\Xi = (\xi, p_\xi)$ being similar). Note that

$$\{\Xi, H\}\,(q,p) = \begin{pmatrix} v_\xi(q,p) \\ {\rm Hess}_q(\xi)(M^{-1}p, M^{-1}p) - \nabla\xi(q)^T M^{-1}\nabla V(q) \end{pmatrix}, \tag{3.6}$$

where the Hessian operator Hess is defined in (2.1). Now, (2.11) implies that

$$\Gamma^{-1} = \begin{pmatrix} G_M^{-1}\,\nabla v_\xi^T\,J\,\nabla v_\xi\,G_M^{-1} & -G_M^{-1} \\ G_M^{-1} & 0 \end{pmatrix}. \tag{3.7}$$

Besides, $v_\xi(q_t, p_t) = 0$ along a trajectory, since $(q_t, p_t) \in T^*\Sigma(z)$. Therefore,

$$\forall (q,p) \in T^*\Sigma(z), \qquad \Gamma^{-1}\,\{\Xi, H\}\,(q,p) = \begin{pmatrix} f_{\rm rgd}^M(q,p) \\ 0 \end{pmatrix}, \tag{3.8}$$

where the notation $f_{\rm rgd}^M$ is introduced in (3.2). Consider a test function $\varphi : \mathbb{R}^{6N} \to \mathbb{R}$, and remark that

$$\{\varphi, \Xi\}\begin{pmatrix} a \\ 0 \end{pmatrix} = -a^T\nabla\xi^T\nabla_p\varphi, \tag{3.9}$$

so that, for any $a \in \mathbb{R}^m$,

$$\{\varphi, \Xi\}\,\Gamma^{-1}\,\{\Xi, H\} = -(f_{\rm rgd}^M)^T\nabla\xi^T\nabla_p\varphi.$$

Finally, for all $(q,p) \in T^*\Sigma(z)$,

$$\begin{aligned}\{\varphi, H\}_\Xi\,(q,p) = &-\nabla V(q)^T\nabla_p\varphi(q,p) + f_{\rm rgd}^M(q,p)^T\nabla\xi(q)^T\nabla_p\varphi(q,p) \\ &+ p^T M^{-1}\nabla_q\varphi(q,p).\end{aligned} \tag{3.10}$$

The operator (3.10) is the generator of the Hamiltonian part in (3.3).

We turn to (ii). The diffusive part arises from the fluctuation term $\sigma_P(q_t)\,dW_t$ in (3.3), and its expression

$$\frac{1}{2}\mathrm{div}_p\Big(P_M\,\sigma\sigma^T P_M^T\nabla_p\;\cdot\;\Big)$$

is obtained directly from the standard Itô calculus. Similarly, the dissipation operator is

$$-\Big(\gamma_P M^{-1}p\Big)^T\nabla_p = -p^T M^{-1}P_M\gamma P_M^T\nabla_p.$$

The addition of these two contributions gives the expression of $\mathcal{L}_\Xi^{\mathrm{thm}}$. □

With the expression (3.4) of the generator at hand, it is easily checked that the process (CL) satisfies the following equilibrium properties:

Proposition 3.2. *When the fluctuation-dissipation relation* (1.3) *holds, the constrained Langevin dynamics* (CL) *on $T^*\Sigma(z)$ admits the Boltzmann-Gibbs distribution* (1.9) *as a stationary measure, and is reversible up to momentum reversal with respect to* (1.9)*: If $\mathrm{Law}(q_0,p_0)=\mu_{T^*\Sigma(z)}$, then, for any $T>0$,*

$$\mathrm{Law}(q_t,p_t;0\le t\le T)=\mathrm{Law}(q_{T-t},-p_{T-t};0\le t\le T).$$

Moreover, if $P_M(q)\gamma P_M(q)^T$ is everywhere strictly positive in the sense of symmetric matrices on $T_q^*\Sigma(z)$, then the process (CL) *is ergodic: for any smooth test function φ,*

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T\varphi(q_t,p_t)\,dt=\int_{T^*\Sigma(z)}\varphi\,d\mu_{T^*\Sigma(z)}\qquad\text{a.s.}$$

Proof. The stationarity and reversibility properties follow from the following detailed balance condition up to momentum reversal (see for instance Section 2.2 in [35]): for any test functions φ_1, φ_2,

$$\int_{T^*\Sigma(z)}\varphi_1\,\mathcal{L}_\Xi(\varphi_2)\,d\mu_{T^*\Sigma(z)}=\int_{T^*\Sigma(z)}(\varphi_2\circ S)\,\mathcal{L}_\Xi(\varphi_1\circ S)\,d\mu_{T^*\Sigma(z)},\tag{3.11}$$

where $S:(q,p)\mapsto(q,-p)$ is the momentum flip. In view of the expression (3.4) of the generator, proving (3.11) amounts to proving this property for the operators $\{.,H\}_\Xi$ and $\mathcal{L}_\Xi^{\mathrm{thm}}$.

For the Hamiltonian part $\{.,H\}_\Xi$, the expression (3.10) yields

$$\{\varphi\circ S,H\}_\Xi(q,p)=-\{\varphi,H\}_\Xi(q,-p)=-\{\varphi,H\}_\Xi(S(q,p)),$$

which states the time symmetry under momentum reversal of the Hamiltonian part of the equations of motion (CL). On the other hand,

$$\mathrm{e}^{-\beta H}\{\cdot,H\}_\Xi=-\frac{1}{\beta}\{\cdot,\mathrm{e}^{-\beta H}\}_\Xi,$$

so that

$$\begin{aligned}\mathrm{e}^{-\beta H}(\varphi_2\circ S)\{\varphi_1\circ S,H\}_\Xi&=-\big(\mathrm{e}^{-\beta H}\varphi_2\{\varphi_1,H\}_\Xi\big)\circ S\\&=\left(\mathrm{e}^{-\beta H}\varphi_1\{\varphi_2,H\}_\Xi+\left\{\varphi_2\varphi_1,\frac{\mathrm{e}^{-\beta H}}{\beta}\right\}_\Xi\right)\circ S,\end{aligned}$$

and the divergence formula (2.25) yields the balance condition (3.11) for the Hamiltonian part, in view of the invariance of the distribution $\sigma_{T^*\Sigma(z)}$ under the momentum flip S.

For the thermostat part, it is easily checked, that $\mathcal{L}_\Xi^{\rm thm}(\varphi \circ S) = \mathcal{L}_\Xi^{\rm thm}(\varphi) \circ S$ for any smooth test function φ, so that the detailed balance condition up to momentum reversal (3.11) follows from the following more general detailed balance condition, in the case $v_z = 0$ ($\mu_{\Sigma_{\xi,v_\xi}(z,v_z)}$ being defined in (2.15)):

$$\int_{\Sigma_{\xi,v_\xi}(z,v_z)} \varphi_1 \, \mathcal{L}_\Xi^{\rm thm}(\varphi_2) \, d\mu_{\Sigma_{\xi,v_\xi}(z,v_z)} = \int_{\Sigma_{\xi,v_\xi}(z,v_z)} \varphi_2 \, \mathcal{L}_\Xi^{\rm thm}(\varphi_1) \, d\mu_{\Sigma_{\xi,v_\xi}(z,v_z)}. \tag{3.12}$$

It is interesting to prove (3.12) for a general $v_z \in \mathbb{R}$ since it will be used in the proof of Theorem 5.3 below. Consider the divergence formula (2.26) in the affine space for the variable p (the position q being fixed), with $\phi = \gamma P_M^T \nabla_p(\varphi_2) \, \mathrm{e}^{-\beta H} \, \varphi_1$. After integration in q, using the formula (3.5) for $\mathcal{L}_\Xi^{\rm thm}$ and (2.24), an expression symmetric in (φ_1, φ_2) is obtained:

$$\begin{aligned} &\int_{\Sigma_{\xi,v_\xi}(z,v_z)} \varphi_1 \, \mathcal{L}_\Xi^{\rm thm}(\varphi_2) \, d\mu_{\Sigma_{\xi,v_\xi}(z,v_z)} \\ &\qquad = - \int_{\Sigma_{\xi,v_\xi}(z,v_z)} \nabla_p^T \varphi_1 P_M \gamma P_M^T \nabla_p \varphi_2 \, d\mu_{\Sigma_{\xi,v_\xi}(z,v_z)}, \end{aligned}$$

hence the detailed balance condition (3.12).

Ergodicity is a consequence of the hypo-ellipticity of the operator $\mathcal{L}_\Xi$ on $T^*\Sigma(z)$ (Hörmander's criterion is satisfied, see [27]), which is itself a consequence of the fact that $P_M(q)\gamma P_M(q)^T$ is strictly positive on each $T_q^*\Sigma(z)$. The proof can be carried out using local coordinates and the results from [30]. □

Remark 3.3 (Infinite stiffness limit). We discuss here the relation between softly and rigidly imposed constraints on the canonical phase-space measures and on the Langevin dynamics. A similar discussion can be read in [9] for canonical measures on the position variables only and overdamped dynamics.

We have considered in this section the Langevin dynamics (3.3) with constraints rigidly imposed by a projection onto the submanifold $T^*\Sigma(z)$. This dynamics samples the canonical distribution (1.9) $\mu_{T^*\Sigma(z)}(dq\,dp) = Z_{z,0}^{-1} \, \mathrm{e}^{-\beta H(q,p)} \, \sigma_{T^*\Sigma(z)}(dq\,dp)$, with constraints on both positions and momenta. The marginal on positions of this distribution is, in view of (2.23), proportional to $\mathrm{e}^{-\beta V(q)} \sigma^M_{\Sigma(z)}(dq)$. This is what we call in the following rigidly imposed constraints. The canonical distribution (1.9) with rigid constraints is naturally associated to the rigid free energy (1.13) (and this is what justifies the qualification "rigid"), since, we recall,

$$F^M_{\rm rgd}(z) = -\frac{1}{\beta} \ln \int_{T^*\Sigma(z)} \mathrm{e}^{-\beta H(q,p)} \sigma_{T^*\Sigma(z)}(dq\,dp).$$

Another way to impose some constraints on a system is to add a penalization term. In the present context, this could be done by changing the potential energy V to $V_\varepsilon(q) = V(q) + \varepsilon^{-1} \, |\xi(q) - z|^2$. It is easy to check that, in the limit $\varepsilon \to 0$ (infinite stiffness limit), the canonical measure associated to this potential is the canonical distribution (1.1) with positions conditioned by $\xi(q) = z$. This distribution is proportional to $\mathrm{e}^{-\beta H(q,p)} \, \delta_{\xi(q)-z}(dq) \, dp$ and its marginal on positions is proportional to $\mathrm{e}^{-\beta V(q)} \delta_{\xi(q)-z}(dq)$. This is what we call in the following softly imposed constraints. The canonical distribution with soft constraints is naturally

associated with the standard free energy, since, we recall,

$$F(z) = -\frac{1}{\beta} \ln \int_{\Sigma(z)\times\mathbb{R}^{3N}} \mathrm{e}^{-\beta H(q,p)}\, \delta_{\xi(q)-z}(dq)\, dp.$$

Note that, in view of (2.18), the marginal on positions for softly imposed constraints can be written in terms of rigidly imposed constraints through a modification of the potential:

$$\mathrm{e}^{-\beta V(q)} \delta_{\xi(q)-z}(dq) = \mathrm{e}^{-\beta (V+V_{\mathrm{fix}})(q)} \sigma^M_{\Sigma(z)}(dq)$$

where

$$V_{\mathrm{fix}}(q) = \frac{1}{2\beta} \ln \Big(\det G_M(q) \Big), \tag{3.13}$$

is sometimes called the Fixman corrector (see [21]). Thus, if (q_t, p_t) satisfies (3.3) with the modified potential $V + V_{\mathrm{fix}}$, then q_t samples (in the longtime limit) the probability measure proportional to $\mathrm{e}^{-\beta V(q)} \delta_{\xi(q)-z}(dq)$, and we thus refer to this dynamics as the softly constrained Langevin dynamics.

These concepts will be used in Section 4.1 to describe the computation of free energy differences for systems with molecular constraints. Finally, let us mention that the infinite stiffness limit $\varepsilon \to 0$ of the Langevin dynamics (1.2) with the potential V_ε is not (except for very specific forms of constraints) the softly constrained Langevin dynamics, as one would expect; see for example [44]. □

3.2. **Numerical implementation.** We consider in this section a numerical scheme based on a splitting of the Langevin dynamics (CL) into a Hamiltonian part (Section 3.2.1) and a fluctuation-dissipation part acting only on the momentum (Section 3.2.2). Such a splitting is standard for unconstrained systems, but other splitting strategies for the Langevin equation can be considered as well (see [39, 40]).

For simplicity, we restrict ourselves to constant matrices γ and σ. Generalizations to position dependent matrices are straightforward.

The Hamiltonian part of the Langevin dynamics (CL) (namely (CL) with $(\sigma, \gamma) = (0, 0)$) is discretized using a velocity-Verlet scheme with constraints, which yields (3.17) below. The fluctuation-dissipation part on momentum variable in (CL) is the following Ornstein-Uhlenbeck process (for a fixed given $q \in \Sigma(z)$):

$$\begin{cases} dp_t = -\gamma M^{-1} p_t\, dt + \sigma\, dW_t + \nabla\xi(q)\, d\lambda^{\mathrm{OU}}_t, \\ \nabla\xi(q) M^{-1} p_t = 0, \qquad (C_p) \end{cases} \tag{3.14}$$

which can be rewritten as (see (3.3)) $dp_t = -\gamma_P(q) M^{-1} p_t\, dt + \sigma_P(q)\, dW_t$. This equation can be explicitly integrated on $[0, t]$ to obtain

$$p_t = \mathrm{e}^{-t\, \gamma_P(q) M^{-1}} p_0 + \int_0^t \mathrm{e}^{-(t-s)\gamma_P(q) M^{-1}}\, \sigma_P(q)\, dW_s. \tag{3.15}$$

However, the matrix exponential $\mathrm{e}^{-t\, \gamma_P(q) M^{-1}}$ may be difficult to compute in practice (except for certain choices of γ and M; see the discussion at the end of Section 3.2.2). Instead of performing an exact integration, (3.14) can be discretized using a midpoint Euler scheme, which yields (3.16) and (3.18) below.

The numerical scheme we investigate, termed midpoint Euler-Verlet-midpoint Euler splitting, is therefore the following:

$$\begin{cases} p^{n+1/4} = p^n - \dfrac{\Delta t}{4}\gamma\, M^{-1}(p^n + p^{n+1/4}) + \sqrt{\dfrac{\Delta t}{2}}\,\sigma\,\mathcal{G}^n \\ \qquad\qquad + \nabla\xi(q^n)\,\lambda^{n+1/4}, \\ \nabla\xi(q^n)^T M^{-1} p^{n+1/4} = 0, \qquad (C_p), \end{cases} \tag{3.16}$$

$$\begin{cases} p^{n+1/2} = p^{n+1/4} - \dfrac{\Delta t}{2}\nabla V(q^n) + \nabla\xi(q^n)\,\lambda^{n+1/2}, \\ q^{n+1} = q^n + \Delta t\, M^{-1}\, p^{n+1/2}, \\ \xi(q^{n+1}) = z, \qquad (C_q), \\ p^{n+3/4} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\,\lambda^{n+3/4}, \\ \nabla\xi(q^{n+1})^T M^{-1} p^{n+3/4} = 0, \qquad (C_p) \end{cases} \tag{3.17}$$

$$\begin{cases} p^{n+1} = p^{n+3/4} - \dfrac{\Delta t}{4}\gamma\, M^{-1}(p^{n+3/4} + p^{n+1}) + \sqrt{\dfrac{\Delta t}{2}}\,\sigma\,\mathcal{G}^{n+1/2} \\ \qquad\qquad + \nabla\xi(q^{n+1})\,\lambda^{n+1}, \\ \nabla\xi(q^{n+1})^T M^{-1} p^{n+1} = 0, \qquad (C_p) \end{cases} \tag{3.18}$$

where $(\mathcal{G}^n)_{n\geq 0}$ and $(\mathcal{G}^{n+1/2})_{n\geq 0}$ are sequences of independently and identically distributed (i.i.d.) Gaussian random variables of mean 0 and covariance matrix Id_{3N}.

Note that when $\gamma = 0$ and $\sigma = 0$, the scheme (3.16)-(3.17)-(3.18) becomes deterministic, and reduces to (3.17), which is a scheme for the deterministic Hamiltonian equations of motion with position constraints $\xi(q) = z$. The latter scheme is referred to as the "Hamiltonian scheme (3.17)" below.

3.2.1. *Comments on the Hamiltonian scheme* (3.17). The Hamiltonian part (3.17) of the scheme, often called "RATTLE" in the literature, is an explicit integrator, and is a modification of the classical "SHAKE" algorithm (see Chapter VII.1 in [22], or Chapter 7 in [32] for more precisions and historical references). In (3.17), $\lambda^{n+1/2} \in \mathbb{R}^m$ are the Lagrange multipliers associated with the position constraints (C_q), and $\lambda^{n+3/4} \in \mathbb{R}^m$ are the Lagrange multipliers associated with the velocity constraints (C_p). The nonlinear constraints (C_q) are typically enforced using Newton's algorithm. In (3.17), the (linear) momentum projection (C_p) is always well defined since we assumed that the Gram matrix $G_M(q)$ is invertible. On the other hand, the nonlinear projection used to enforce the position constraints $\xi(q^{n+1}) = z$ is in general well defined only on a subset of phase space.

Definition 3.4 (Domain $D_{\Delta t}$)**.** The domain $D_{\Delta t} \subset T^*\Sigma(z)$ is defined as the set of configurations $(q^n, p^{n+1/4}) \in T^*\Sigma(z)$ such that there is a unique solution $(q^{n+1}, p^{n+3/4})$ satisfying (3.17).

Solving the position constraints (C_q) consists in projecting onto $\Sigma(z)$ a point in a Δt-neighborhood of q^n. Thus, by the implicit function theorem, the domain $D_{\Delta t}$ satisfies

$$\lim_{\Delta t \to 0} D_{\Delta t} = T^*\Sigma(z).$$

It may happen that there is no solution if the time-step is too large, and, even for small time-steps, that several projections exist; see, for instance, Example 2 in Chapter 7 of [32]. In practice, $D_{\Delta t}$ can be chosen to be the set of $(q^n, p^{n+1/4})$ such that the Newton algorithm enforcing the constraints (C_q) has converged within a given precision threshold and a limited number of iterations.

As for the Verlet scheme in the unconstrained case, the associated numerical flow shares two important qualitative properties with the exact flow: It is time reversible and symplectic (see [33]). This implies quasi-conservation of energy, in the sense that energy is conserved within a given precision threshold over exponentially long times; see [22, 32].

3.2.2. *Comments on the fluctuation-dissipation part* (3.16) *and* (3.18). The new momentum $p^{n+1/4} \in T^*\Sigma(z)$ in (3.16) (or p^{n+1} in (3.18)) may be obtained by first integrating the unconstrained dynamics with a midpoint scheme, and then computing the Lagrange multiplier $\lambda^{n+1/4}$ (or λ^{n+1}) by solving the following linear system implied by the constraints (C_p):

$$\nabla\xi(q^n)^T M^{-1} \left(\mathrm{Id} + \frac{\Delta t}{4}\gamma M^{-1}\right)^{-1} \left[\left(\mathrm{Id} - \frac{\Delta t}{4}\gamma M^{-1}\right) p^n + \sqrt{\frac{\Delta t}{2}}\,\sigma\,\mathcal{G}^n + \nabla\xi(q^n)\,\lambda^{n+1/4}\right] = 0.$$

A sufficient criteria for stability is

$$\frac{\Delta t}{4}\gamma \leq M.$$

Besides, it can be checked (see Sections 2.3.2 and 3.3.5 in [35]) that the Markov chain induced by the fluctuation-dissipation part of the scheme (3.16) (or (3.18)) satisfies a detailed balance equation (both in the plain sense and up to momentum reversal) with respect to the stationary measure $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$. The latter is defined as the kinetic probability distribution

$$\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp) = \left(\frac{\beta}{2\pi}\right)^{(3N-m)/2} \exp\left(-\beta\frac{p^T M^{-1} p}{2}\right) \sigma^{M^{-1}}_{T^*_q\Sigma(z)}(dp), \tag{3.19}$$

and is the marginal in the p-variable of the canonical distribution $\mu_{T^*\Sigma(z)}(dq\,dp)$ conditioned by a given $q \in \Sigma(z)$. Moreover, if $\gamma_P := P_M\,\gamma\,P_M^T$ is strictly positive in the sense of symmetric linear transformations of $T^*_q\Sigma(z)$, then the Markov chain on momentum variable induced by (3.16) (or (3.18)) alone is ergodic with respect to $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$.

Finally, an important simplification occurs in the integration of (3.14) in the special case when γ and M are equal up to a multiplicative constant (so that γM^{-1} is proportional to identity). Indeed, in this case, the equality $\left(\gamma_P(q)M^{-1}\right)^n = P_M(q)\left(\gamma\,M^{-1}\right)^n$ holds for any $n \geq 0$, and (3.15) simplifies to

$$p_t = P_M(q)\left(\mathrm{e}^{-t\,\gamma M^{-1}} p_0 + \int_0^t \mathrm{e}^{-(t-s)\gamma M^{-1}}\,\sigma\,dW_s\right). \tag{3.20}$$

The numerical integration of (3.14) can thus be carried out in two steps: (i) exactly integrating (3.14) without constraint, and then (ii) projecting the result onto $T^*_q\Sigma(z)$.

3.2.3. *Metropolis-Hastings correction.* Usually, the invariant probability distribution sampled by the solution of a numerical scheme is biased by the time discretization. Relying on (i) the time symmetry (up to momentum reversal) and (ii) the preservation of the phase space measure $\sigma_{T^*\Sigma(z)}(dq\,dp)$ by the solution of the RATTLE scheme (3.17), it is possible to eliminate the time discretization error in the splitting scheme (3.16)-(3.17)-(3.18) by resorting to a Generalized Hybrid Monte Carlo algorithm.

Algorithm 3.5 (GHMC with constraints)**.** *Consider an initial configuration* $(q^0, p^0) \in T^*\Sigma(z)$*, and a sequence* $(\mathcal{G}^n, \mathcal{G}^{n+1/2})_{n\geq 0}$ *of independently and identically distributed standard Gaussian vectors. Iterate on* $n \geq 0$*:*

(1) Evolve the momentum according to the midpoint Euler scheme (3.16)*, and compute the energy* $E^n = H(q^n, p^{n+1/4})$ *of the new configuration.*

(2) Integrate the Hamiltonian part according to the RATTLE scheme (3.17)*, denote* $(\widetilde{q}^{n+1}, \widetilde{p}^{n+3/4})$ *the resulting state, and set* $E^{n+1} = H(\widetilde{q}^{n+1}, \widetilde{p}^{n+3/4})$*.*

(3) Accept the proposal $(q^{n+1}, p^{n+3/4}) := (\widetilde{q}^{n+1}, \widetilde{p}^{n+3/4})$ *with probability*

$$\min\left(e^{-\beta(E^{n+1}-E^n)}, 1\right).$$

Otherwise, reject and flip the momentum: $(q^{n+1}, p^{n+3/4}) = (q^n, -p^{n+1/4})$*.*

(4) Evolve the momentum according to the midpoint Euler scheme (3.18)*.*

By construction, the GHMC algorithm with constraints leaves invariant the equilibrium distribution $\mu_{T^*\Sigma(z)}(dq\,dp)$ (see Section 3.3.5 in [35]).

To understand the momentum reversal required upon rejection, it is useful to write more explicitly the Markov chain as the composition of a Metropolis-Hastings part, where the proposal is obtained by a RATTLE step followed by a momentum reversal (the latter operation is needed to ensure the symmetry of the proposition), which is then accepted or rejected; and another momentum reversal (which leaves invariant the targeted probability measure $\mu_{T^*\Sigma(z)}(dq\,dp)$). When the proposal is accepted, the two momentum reversals cancel out each other. On the other hand, when the proposal is rejected, momenta are actually reversed. See [35, Section 2.1.4] for more background on generalized Metropolis-Hastings algorithms.

In the above, we implicitly assume that the RATTLE scheme (3.17) is everywhere well defined. In practice however, it is necessary to modify Algorithm 3.5 by restricting the sampled configurations to $D_{\Delta t}$. This can be achieved by introducing additional tests in steps (1), (2) and (4), and rejecting the states that have gone outside the set $D_{\Delta t} \subset T^*\Sigma(z)$ where the position constraint (C_q) is well defined. By doing so, the global algorithm has an invariant equilibrium distribution given by $\mu_{T^*\Sigma(z)}(dq\,dp)$ conditioned on the set of states $D_{\Delta t}$. This invariant distribution can be written explicitly as follows:

$$\frac{1}{Z_{z,0,\Delta t}} \mathrm{e}^{-\beta H(q,p)} \mathbf{1}_{(q,p)\in D_{\Delta t}} \sigma_{T^*\Sigma(z)}(dq\,dp). \tag{3.21}$$

Alternatively, the rejection tests in steps (1), (2) and (4) of Algorithm 3.5 can be performed with a cut-off parameter $R_{\Delta t} > 0$ on the momentum variable, chosen so that the position constraint (C_q) in (3.17) is everywhere well defined when $\frac{1}{2}p^T M^{-1} p \leq R_{\Delta t}$. This can be achieved when there exists $R_{\Delta t} > 0$ small enough so that $\Sigma(z) \times \{\frac{1}{2}p^T M^{-1} p \leq R_{\Delta t}\} \subset D_{\Delta t} \subset T^*\Sigma(z)$. Since this is useful for later purposes (see the discussion at the end of Section 4.3), we provide a rough estimate

of $R_{\Delta t}$ in terms of Δt, assuming for simplicity that $\Sigma(z)$ is compact. First, by the implicit function theorem, there exists $\alpha > 0$ such that, for all $q \in \Sigma(z)$ and δq with norm $\|\delta q\| < \alpha$, there is a unique $\lambda \in \mathbb{R}^m$ satisfying

$$\xi\left(q + M^{-1}(\delta q + \nabla\xi(q)\lambda)\right) = z.$$

Therefore, there exists $a > 0$ small enough such that, when $\|p^{n+1/4}\| \leq a/\Delta t$, the RATTLE scheme in (3.17) is well defined, namely there exists a unique q^{n+1} satisfying the constraint (C_q). This shows that

$$R_{\Delta t} \geq A\Delta t^{-2} \tag{3.22}$$

for some $A > 0$.

The invariant probability distribution of the Markov chain generated by GHMC with the additional rejection steps ensuring $\frac{1}{2}p^T M^{-1}p \leq R_{\Delta t}$, is given by (3.21), and actually reads

$$\frac{1}{Z_{z,0,\Delta t}} \mathrm{e}^{-\beta H(q,p)} \mathbf{1}_{\frac{1}{2}p^T M^{-1}p \leq R_{\Delta t}} \sigma_{T^*\Sigma(z)}(dq\, dp).$$

Its marginal distribution in the position variable is then exactly given by

$$\frac{1}{Z_z} \mathrm{e}^{-\beta V(q)} \sigma^M_{\Sigma(z)}(dq).$$

This is also the marginal distribution in the position variable of $\mu_{T^*\Sigma(z)}$. Note, however, that if $R_{\Delta t}$ is too small, only small momenta will be sampled in step (1) of Algorithm 3.5, and the correlation time of the sampling will be large. In practice, the threshold $R_{\Delta t}$ should be tuned in preliminary computations so that: (i) $R_{\Delta t}$ is small enough so that the maximal number $N_{\max}$ of iterations for the Newton algorithm used to enforce (C_q) in (3.17) is never reached; (ii) $R_{\Delta t}$ is large enough so that the correlation time of the sampling is as small as possible.

Let us end this section with a warning: It is now known that the correction of the bias in discretizations of the Langevin dynamics by a Metropolization of the scheme may reduce the efficiency of the sampling; see for instance [1].

3.3. **Exact sampling on a submanifold with overdamped dynamics.** Constrained overdamped Langevin processes (or Brownian dynamics) are solutions of the stochastic differential equation (see also [35, 9])

$$\begin{cases} dq_t = -\nabla V(q_t)\, dt + \sqrt{\dfrac{2}{\beta}}\, dW_t + \nabla\xi(q_t)\, d\lambda_t, \\ \xi(q_t) = z, \end{cases} \tag{3.23}$$

where λ_t is an adapted stochastic process. Equivalently, (3.23) can be rewritten in the Stratonovitch form as

$$dq_t = -P(q_t)\nabla V(q_t)\, dt + \sqrt{\frac{2}{\beta}} P(q_t) \circ dW_t,$$

where $\circ$ denotes the Stratonovitch integration, and P is the projector defined by (1.7) with the choice $M = \mathrm{Id}$:

$$P(q) = \mathrm{Id} - \nabla\xi(q)\, G^{-1}(q)\nabla\xi(q)^T, \qquad G(q) = \nabla\xi(q)^T\nabla\xi(q). \tag{3.24}$$

It can be shown that (3.23) satisfies the detailed balance condition for (and is ergodic with respect to) the invariant distribution

$$Z_z^{-1}\,\mathrm{e}^{-\beta V(q)}\sigma_{\Sigma(z)}^{\mathrm{Id}}(dq), \tag{3.25}$$

which is the marginal in the q-variable of the canonical distribution with constraints (1.9) for the choice $M = \mathrm{Id}$. It is easy to generalize all our results to scalar products associated with a general symmetric definite positive matrix M upon considering (3.32); see Remark 3.7 below.

The constrained overdamped Langevin process (3.23) may be obtained from a scaling limit of the constrained Langevin dynamics (CL) (in the limit when either the mass goes to zero, or the damping γ goes to infinity); see Propositions 2.14 and 2.15 in [35].

Likewise, at the discrete level, an Euler-Maruyama discretization of the overdamped process (3.23) can be obtained as a particular case of the numerical discretization (3.16)-(3.17)-(3.18) for the Langevin equation (CL), yielding a Markov chain $(q^n)_{n\geq 0}$ on positions. The condition that the mass goes to 0 is replaced by the condition that the mass is proportional to the time-step. This is the content of the following proposition.

Proposition 3.6. *Suppose that the following relation is satisfied:*

$$\frac{\Delta t}{4}\gamma = M = \frac{\Delta t}{2}\mathrm{Id}. \tag{3.26}$$

With a slight abuse of notation, the mass matrix and the friction matrix are rewritten as $M\,\mathrm{Id}$ *and* $\gamma\,\mathrm{Id}$ *with* $M, \gamma \in \mathbb{R}$. *Then the splitting scheme* (3.16)-(3.17)-(3.18) *is the following Euler scheme for the overdamped Langevin constrained dynamics* (3.23)*:*

$$\begin{cases} q^{n+1} = q^n - \Delta t \nabla V(q^n) + \sqrt{\dfrac{2\Delta t}{\beta}}\,\mathcal{G}^n + \nabla\xi(q^n)\,\lambda_{\mathrm{od}}^{n+1}, \\ \xi(q^{n+1}) = z, \end{cases} \tag{3.27}$$

where $(\mathcal{G}^n)_{n\geq 0}$ *are independent and identically distributed centered and normalized Gaussian variables, and* $(\lambda_{\mathrm{od}}^n)_{n\geq 1}$ *are the Lagrange multipliers associated with the constraints* $(\xi(q^n) = z)_{n\geq 1}$. *Moreover, the Lagrange multipliers in* (3.17) *satisfy:*

$$2\lambda^{n+1/2} = G^{-1}(q^n)\Big(\nabla\xi(q^n)^T\left(q^{n+1} - q^n\right) + \Delta t \nabla\xi(q^n)^T\nabla V(q^n)\Big) \tag{3.28}$$

$$= \lambda_{\mathrm{od}}^{n+1} + \sqrt{\frac{2\Delta t}{\beta}}G^{-1}(q^n)\nabla\xi(q^n)^T\mathcal{G}^n, \tag{3.29}$$

as well as

$$2\lambda^{n+3/4} = G^{-1}(q^{n+1})\Big(\nabla\xi(q^{n+1})^T\left(q^n - q^{n+1}\right) + \Delta t\nabla\xi(q^{n+1})^T\nabla V(q^{n+1})\Big). \tag{3.30}$$

where G *is defined in* (3.24).

Proof. Irrespective of p^n, the choice (3.26) in the scheme (3.16)-(3.17)-(3.18) leads to

$$p^{n+1/4} = \sqrt{\frac{\Delta t}{8}}\,\sigma\,\mathcal{G}^n + \frac{1}{2}\nabla\xi(q^n)\,\lambda^{n+1/4},$$

where $\lambda^{n+1/4}$ is associated with the constraints $\nabla\xi(q^n)^T p^{n+1/4} = 0$. This gives

$$p^{n+1/2} = -\frac{\Delta t}{2}\nabla V(q^n) + \sqrt{\frac{\Delta t}{8}}\,\sigma\,\mathcal{G}^n + \nabla\xi(q^n)\left(\frac{1}{2}\,\lambda^{n+1/4} + \lambda^{n+1/2}\right),$$

where $\lambda^{n+1/2}$ is such that $\xi(q^{n+1}) = z$. The fluctuation-dissipation relation (1.3) can be reformulated in this context as

$$\sigma\sigma^T = \frac{2}{\beta}\gamma = \frac{4}{\beta}\,\mathrm{Id},$$

and the scheme (3.27) is recovered by taking the associated Lagrange multiplier equal to $\lambda_{\rm od}^{n+1} = \lambda^{n+1/4} + 2\lambda^{n+1/2}$. Finally, remarking that $G_M = \frac{2}{\Delta t}G$ and computing explicitly $\lambda^{n+1/2}$ and $\lambda^{n+3/4}$ in (3.17) yields (3.28)-(3.29)-(3.30). □

This point of view allows us to construct a Metropolis correction to the Euler scheme (3.27), using the Generalized Hybrid Monte Carlo scheme (Algorithm 3.5) with the time-step chosen according to (3.26). In this way, assuming that the position constraint (C_q) in (3.17) is everywhere well defined, we obtain a Markov chain $(q^n)_{n\geq 0}$ discretizing the overdamped dynamics (3.27) which *exactly* samples the invariant distribution (3.25). Deriving such a Metropolis-Hastings correction to the Euler scheme (3.27) without resorting to phase-space dynamics does not seem to be natural.

Remark 3.7 (Discrete overdamped limit). Proposition 3.6 can be seen as a discrete version of the zero-mass limit of the Langevin dynamics. It is also possible to obtain a discrete version of the overdamped limit ($\gamma \to \infty$) of the Langevin dynamics by assuming that the parameters satisfy the relation

$$\frac{\Delta t}{4}\gamma = M \propto \mathrm{Id},$$

which is less restrictive than (3.26). Equation (3.27) is then obtained with Δt replaced by

$$\Delta s = \frac{\Delta t^2}{2M} = \frac{2\Delta t}{\gamma}. \tag{3.31}$$

In this case, the effective discretization time-step Δs for the overdamped Langevin dynamics is thus different from the time-step Δt originally used for the discretization of the Langevin dynamics. This is reminiscent of the fact that the overdamped limit (at the continuous level) of the Langevin dynamics requires a change of timescale to obtain the overdamped Langevin dynamics.

Note also that in the more general case $M = \gamma\Delta t/4$ where M is not supposed to be proportional to the identity, the following numerical scheme is obtained:

$$\begin{cases} q^{n+1} = q^n - \widetilde{\Delta s}\, M^{-1}\nabla V(q^n) + \sqrt{\dfrac{2\widetilde{\Delta s}}{\beta}}\, M^{-1/2}\mathcal{G}^n + M^{-1}\nabla\xi(q^n)\,\lambda_{\rm od}^{n+1}, \\ \xi(q^{n+1}) = z, \end{cases}$$

where $\widetilde{\Delta s} = \Delta t^2/2$. This is a discretization of the overdamped dynamics

$$\begin{cases} dq_s = -M^{-1}\nabla V(q_s)\,ds + \sqrt{\dfrac{2}{\beta}}\, M^{-1/2} dW_s + M^{-1}\nabla\xi(q_s)\,d\lambda_s, \\ \xi(q_s) = z, \end{cases} \tag{3.32}$$

which is a generalization of (3.23) to a scalar product on $\Sigma(z)$ induced by a general positive definite mass matrix M. □

4. Thermodynamic integration with constrained Langevin dynamics

In this section, we focus on the computation of the gradient of the rigid free energy (1.13),

$$F^M_{\mathrm{rgd}}(z) = -\frac{1}{\beta} \ln \int_{T^*\Sigma(z)} \mathrm{e}^{-\beta H(q,p)} \sigma_{T^*\Sigma(z)}(dq\, dp),$$

using a numerical discretization of the constrained Langevin process (CL).

As explained in the introduction, we may indeed concentrate on the computation of the rigid free energy (1.13), since the standard free energy (1.12) can be computed from the latter one using (1.14). The relation (1.14) can be proved with the co-area formula (2.18). Indeed, the free energy defined in (1.12) can be rewritten as (where C denotes a constant which may vary from line to line):

$$\begin{aligned} F(z) &= -\frac{1}{\beta} \ln \int_{\Sigma(z)\times\mathbb{R}^{3N}} \mathrm{e}^{-\beta H(q,p)} \delta_{\xi(q)-z}(dq)\, dp \\ &= -\frac{1}{\beta} \ln \int_{\Sigma(z)} \mathrm{e}^{-\beta V(q)} (\det G_M(q))^{-1/2} \sigma^M_{\Sigma(z)}(dq) + \mathrm{C} \\ &= -\frac{1}{\beta} \ln \int_{T^*\Sigma(z)} \mathrm{e}^{-\beta H(q,p)} (\det G_M(q))^{-1/2} \sigma_{T^*\Sigma(z)}(dq\, dp) + \mathrm{C} \end{aligned} \tag{4.1}$$

hence

$$F(z) = F^M_{\mathrm{rgd}}(z) - \frac{1}{\beta} \ln \int_{T^*\Sigma(z)} (\det G_M)^{-1/2} d\mu_{T^*\Sigma(z)} + \mathrm{C}, \tag{4.2}$$

where surface measures are defined in Section 2.3. Note that the rigid free energy F^M_{rgd} indeed depends explicitly on the mass matrix since

$$F^M_{\mathrm{rgd}}(z) = -\frac{1}{\beta} \ln \int_{\Sigma(z)} \mathrm{e}^{-\beta V(q)} \sigma^M_{\Sigma(z)}(dq) + \mathrm{C}. \tag{4.3}$$

This section is organized as follows. First, we show how systems with molecular constraints and systems with constrained values of the reaction coordinate can be treated in a unified framework (Section 4.1). We then relate the Lagrange multipliers arising in the constrained Langevin dynamics, and the gradient of the rigid free energy (the so-called mean force) in Section 4.2. We consider the numerical computation of the mean force in Section 4.3, where we prove consistency results for the corresponding approximation formulas. Finally, some numerical results on a model system illustrate the approach in Section 4.4.

4.1. Molecular constraints.

We discuss here how to generalize all the computations to systems with molecular constraints, generalizing thereby some results of [8]. This section can be considered as independent of the remainder of the paper and may therefore be omitted in a first reading.

In practice, many systems are subject to molecular constraints, such as fixed lengths for covalent bonds, or fixed angles between covalent bonds. The reader

is referred to [43] for practical aspects related to the simulation of molecular constraints. In the context of free energy computations, two types of constraints are therefore considered: first, the molecular constraints,

$$\xi_{\mathrm{mc}}(q) = (\xi_{\mathrm{mc},1}(q), \dots, \xi_{\mathrm{mc},\overline{m}}(q)) = 0,$$

for $\overline{m} < 3N$, and second, the reaction coordinates denoted in this section by $\xi_{\mathrm{rc}} : \mathbb{R}^{3N} \to \mathbb{R}^m$, with $\overline{m} + m < 3N$. The submanifold of molecular constraints is denoted by

$$\Sigma_{\mathrm{mc}} = \big\{ q \in \mathbb{R}^{3N} \;\big|\; \xi_{\mathrm{mc}}(q) = 0 \big\},$$

and the submanifold associated with the reaction coordinates by

$$\Sigma_{\mathrm{rc}}(z_{\mathrm{rc}}) = \big\{ q \in \mathbb{R}^{3N} \,\big|\, \xi_{\mathrm{rc}}(q) = z_{\mathrm{rc}} \big\}.$$

It is assumed that the full Gram matrix,

$$G_M^{\mathrm{mc,rc}} := \nabla(\xi_{\mathrm{mc}}, \xi_{\mathrm{rc}})^T M^{-1} \nabla(\xi_{\mathrm{mc}}, \xi_{\mathrm{rc}}) \in \mathbb{R}^{(\overline{m}+m)\times(\overline{m}+m)},$$

is everywhere invertible on $\Sigma_{\mathrm{mc}} \cap \Sigma_{\mathrm{rc}}(z_{\mathrm{rc}})$. Likewise, we denote

$$G_M^{\mathrm{rc}} := \nabla \xi_{\mathrm{rc}}^T M^{-1} \nabla \xi_{\mathrm{rc}} \in \mathbb{R}^{m\times m}$$

and

$$G_M^{\mathrm{mc}} := \nabla \xi_{\mathrm{mc}}^T M^{-1} \nabla \xi_{\mathrm{mc}} \in \mathbb{R}^{\overline{m}\times\overline{m}}.$$

Assuming rigid mechanical constraints on the molecular constraints ξ_{mc}, we are led to considering the canonical distribution

$$\begin{aligned} \mu_{T^*\Sigma_{\mathrm{mc}}}(dq\,dp) &= \frac{1}{Z_{\mathrm{mc}}} \mathrm{e}^{-\beta H(q,p)} \delta_{\xi_{\mathrm{mc}}(q), p_{\xi_{\mathrm{mc}}}(p,q)}(dq\,dp) \\ &= \frac{1}{Z_{\mathrm{mc}}} \mathrm{e}^{-\beta H(q,p)} \sigma_{T^*\Sigma_{\mathrm{mc}}}(dq\,dp), \end{aligned} \tag{4.4}$$

to describe systems with molecular constraints at a fixed temperature. The measure $\sigma_{T^*\Sigma_{\mathrm{mc}}}$ denotes the phase space measure on $T^*\Sigma_{\mathrm{mc}}$, equal by (2.21) to the conditional measure $\delta_{\xi_{\mathrm{mc}}(q), p_{\xi_{\mathrm{mc}}}(q,p)}(dq\,dp)$ associated with the constraints $(\xi_{\mathrm{mc}}(q) = 0, p_{\xi_{\mathrm{mc}}}(q,p) = 0)$, where $p_{\xi_{\mathrm{mc}}}$ is the effective momentum (2.5) associated with ξ_{mc}. Note, however, that it is possible to rewrite the remainder of this section by considering softly imposed molecular constraints rather than rigidly imposed molecular constraints (*i.e.* replacing $\delta_{\xi_{\mathrm{mc}}(q), p_{\xi_{\mathrm{mc}}}(p,q)}(dq\,dp)$ by $\delta_{\xi_{\mathrm{mc}}(q)}(dq)\,dp)$, up to an appropriate modification of (4.6) below, with the help of some Fixman corrective potential. Choosing whether molecular constraints should be softly or rigidly imposed is a modeling choice, and there is no clear consensus on this issue in the current literature.

By associativity of the conditioning of measures, the distribution $\mu_{T^*\Sigma_{\mathrm{mc}}}$ conditioned by a value of the reaction coordinates $\xi_{\mathrm{rc}}(q) = z_{\mathrm{rc}}$ is given, up to a normalizing factor, by

$$\mathrm{e}^{-\beta H(q,p)} \delta_{\xi_{\mathrm{mc}}(q), p_{\xi_{\mathrm{mc}}}(q,p), \xi_{\mathrm{rc}}(q) - z_{\mathrm{rc}}}(dq\,dp).$$

Therefore, considering the marginal probability distribution of the reaction coordinates $\xi_{\mathrm{rc}}(q)$ leads to the following definition of the free energy associated with ξ_{rc}:

$$F^{\mathrm{mc}}(z_{\mathrm{rc}}) = -\frac{1}{\beta} \ln \int_{T^*\Sigma_{\mathrm{mc}} \cap (\Sigma_{\mathrm{rc}}(z_{\mathrm{rc}}) \times \mathbb{R}^{3N})} \mathrm{e}^{-\beta H(q,p)} \delta_{\xi_{\mathrm{mc}}(q), p_{\xi_{\mathrm{mc}}}(q,p), \xi_{\mathrm{rc}}(q) - z_{\mathrm{rc}}}(dq\,dp).$$

The conditional distribution can be decomposed as follows, using the co-area formulas (2.18)-(2.20) and the definition of effective momentum (2.5):

$$\begin{aligned}
&\delta_{\xi_{\rm mc}(q),p_{\xi_{\rm mc}}(q,p),\xi_{\rm rc}(q)-z_{\rm rc}}(dq\,dp) \\
&\quad = \delta_{p_{\xi_{\rm mc}}(q,p)}(dp)\delta_{\xi_{\rm mc}(q),\xi_{\rm rc}(q)-z_{\rm rc}}(dq) \\
&\quad = (\det G_M^{\rm mc}(q))^{1/2}\,\sigma^{M^{-1}}_{T_q^*\Sigma_{\rm mc}}(dp)\;(\det G_M^{\rm mc,rc}(q))^{-1/2}\,\sigma^M_{\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc}}(dq).
\end{aligned}$$

Integrating out the momentum in the linear space $T_q^*\Sigma_{\rm mc}$ with scalar product $\langle p_1,p_2\rangle_{M^{-1}} = p_1^T M^{-1} p_2$, the free energy can be rewritten as

$$F^{\rm mc}(z_{\rm rc}) = -\frac{1}{\beta}\ln\int_{\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc}} {\rm e}^{-\beta V(q)}\left(\frac{\det G_M^{\rm mc}(q)}{\det G_M^{\rm mc,rc}(q)}\right)^{1/2}\sigma^M_{\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc}}(dq) + {\rm C}.$$

As a consequence, the free energy $F^{\rm mc}$ can be computed from the generalized rigid free energy:

$$F_{\rm rgd}^{{\rm mc},M}(z_{\rm rc},0) = -\frac{1}{\beta}\ln\int_{T^*(\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc})} {\rm e}^{-\beta H(q,p)}\sigma_{T^*(\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc})}(dp\,dq), \tag{4.5}$$

using the following formula, similar to (1.14):

$$\begin{aligned}
&F^{\rm mc}(z_{\rm rc}) - F_{\rm rgd}^{{\rm mc},M}(z_{\rm rc},0) \\
&\quad = -\frac{1}{\beta}\ln\int_{T^*(\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc})}\frac{(\det G_M^{\rm mc})^{1/2}}{(\det G_M^{\rm mc,rc})^{1/2}}d\mu_{T^*(\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc})} + {\rm C}.
\end{aligned} \tag{4.6}$$

In the above, $\mu_{T^*(\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc})}$ is defined similarly to (1.9). The case of molecular constraints can therefore be treated within the general framework considered in this paper, the sampling of the canonical measure $\mu_{T^*(\Sigma_{\rm rc}(z_{\rm rc})\cap\Sigma_{\rm mc})}$ and the computation of the rigid free energy (4.5) being the problems at hand.

Similar considerations hold for the overdamped Langevin dynamics, upon appropriate modifications (see Remark 4.2 in [36] for further precisions).

4.2. **The mean force and the Lagrange multipliers.** In this section, the average of the constraining force (3.2) is related to the gradient of the rigid free energy (1.13) (or mean force). We also give a similar result for the following generalized rigid free energy:

$$F_{\rm rgd}^{\Xi}(\zeta) = -\frac{1}{\beta}\ln\int_{\Sigma_\Xi(\zeta)} {\rm e}^{-\beta H(q,p)}\sigma_{\Sigma_\Xi(\zeta)}(dq\,dp). \tag{4.7}$$

Proposition 4.1. *The constraining force $f_{\rm rgd}^M : T^*\Sigma(z)\to\mathbb{R}^m$ defined in* (3.2) *as*

$$f_{\rm rgd}^M(q,p) = G_M^{-1}(q)\nabla\xi(q)^T M^{-1}\nabla V(q) - G_M^{-1}(q){\rm Hess}_q(\xi)(M^{-1}p, M^{-1}p),$$

yields on average the rigid free energy derivative:

$$\nabla_z F_{\rm rgd}^M(z) = \int_{T^*\Sigma(z)} f_{\rm rgd}^M(q,p)\,\mu_{T^*\Sigma(z)}(dq\,dp). \tag{4.8}$$

Moreover, for general constraints (2.6) *and the associated generalized free energy* (4.7)*, the formula can be extended as follows: The generalized constraining force is*

$$\begin{pmatrix} f^\Xi \\ g^\Xi \end{pmatrix} := \Gamma^{-1}\left\{\Xi, H\right\}, \tag{4.9}$$

where $\Gamma(q,p) = \{\Xi,\Xi\}(q,p) = \nabla\Xi^T(q,p)\, J\, \nabla\Xi(q,p)$ *is defined in* (2.9)*, and the rigid mean force is*

$$\nabla_\zeta F^\Xi_{\mathrm{rgd}}(\zeta) = \frac{1}{Z_\zeta}\int_{\Sigma_\Xi(\zeta)} \begin{pmatrix} f^\Xi \\ g^\Xi \end{pmatrix} \mathrm{e}^{-\beta H}\, d\sigma_{\Sigma_\Xi(\zeta)}, \tag{4.10}$$

where $Z_\zeta = \displaystyle\int_{\Sigma_\Xi(\zeta)} \mathrm{e}^{-\beta H} d\,\sigma_{\Sigma_\Xi(\zeta)}$*, and* $F^\Xi_{\mathrm{rgd}}(\zeta)$ *is defined in* (4.7)*. When* (q,p) *is such that* $p_\xi(q,p) = v_\xi(q,p) = 0$*, then* $g^\Xi(q,p) = 0$ *and* $f^\Xi(q,p) = f^M_{\mathrm{rgd}}(q,p)$*.*

Proof. Formulas (4.9) and (4.10) are obtained directly by replacing φ by $\mathrm{e}^{-\beta H}$ in Lemma 2.3. The fact that $(f^\Xi(q,p), g^\Xi(q,p)) = (f^M_{\mathrm{rgd}}(q,p), 0)$ in the tangential case (namely when $p_\xi(q,p) = v_\xi(q,p) = 0$) is a consequence of (3.8). □

The following lemma gives a momentum-averaged version of the constraining force (a similar formula exists in the overdamped case; see equations (4.8)-(4.9) in [9], for example).

Lemma 4.2. *The rigid mean force* (4.8) *can be rewritten as:*

$$\nabla_z F_{\mathrm{rgd}}(z) = \int_{T^*\Sigma(z)} \overline{f}^M_{\mathrm{rgd}}(q)\, \mu_{T^*\Sigma(z)}(dq\, dp), \tag{4.11}$$

where

$$\overline{f}^M_{\mathrm{rgd}}(q) = G_M^{-1}(q)\nabla\xi(q)^T M^{-1}\nabla V(q) - \beta^{-1}G_M^{-1}(q)\mathrm{Hess}_q(\xi) : \left(M^{-1}P_M(q)\right). \tag{4.12}$$

Proof. Consider the Gaussian distribution $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$ defined in (3.19):

$$\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp) = \left(\frac{\beta}{2\pi}\right)^{(3N-m)/2} \exp\left(-\beta\frac{p^T M^{-1} p}{2}\right) \sigma^{M^{-1}}_{T^*_q\Sigma(z)}(dp),$$

which is the marginal distribution in the momentum variable of the canonical distribution $\mu_{T^*\Sigma(z)}(dq\, dp)$, conditioned by a given $q \in \Sigma(z)$. Proving Lemma 4.2 amounts to showing that the average of the constraining force f^M_{rgd} with respect to $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$ yields $\overline{f}^M_{\mathrm{rgd}}$:

$$\overline{f}^M_{\mathrm{rgd}}(q) = \int_{T^*_q\Sigma(z)} f^M_{\mathrm{rgd}}(q,p)\, \kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp).$$

First, we compute the covariance matrix $\mathcal{C} := \mathrm{cov}\left(\kappa^{M^{-1}}_{T^*_q\Sigma(z)}\right)$ of the Gaussian distribution $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$. Since $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$ is a centered Gaussian distribution, $\mathcal{C}$ satisfies, for all $p_1, p_2 \in \mathbb{R}^{3N}$,

$$\begin{aligned} p_1^T M^{-1}\mathcal{C}M^{-1}p_2 &:= \int_{T^*_q\Sigma(z)} \left(p^T M^{-1}p_1\right)\left(p^T M^{-1}p_2\right) \kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp) \\ &= \int_{T^*_q\Sigma(z)} \left(p^T M^{-1}P_M(q)p_1\right)\left(p^T M^{-1}P_M(q)p_2\right) \kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp). \end{aligned}$$

Denoting $\langle p_1, p_2 \rangle_{M^{-1}} = p_1^T M^{-1} p_2$, this yields

$$\begin{aligned} & p_1^T M^{-1} \mathcal{C} M^{-1} p_2 \\ & \quad = \int_{T_q^*\Sigma(z)} \langle p, P_M(q) p_1 \rangle_{M^{-1}} \langle p, P_M(q) p_2 \rangle_{M^{-1}} \frac{\mathrm{e}^{-\frac{\beta}{2}\langle p,p\rangle_{M^{-1}}}}{(2\pi/\beta)^{(3N-m)/2}} \, \sigma_{T_q^*\Sigma(z)}^{M^{-1}}(dp) \\ & \quad = \beta^{-1} \langle P_M(q) p_1, P_M(q) p_2 \rangle_{M^{-1}}, \end{aligned}$$

so that $\mathcal{C} = \beta^{-1} P_M(q) M$. This gives

$$\int_{T_q^*\Sigma(z)} \mathrm{Hess}_q(\xi)(M^{-1}p, M^{-1}p) \, \kappa_{T_q^*\Sigma(z)}^{M^{-1}}(dp) = \beta^{-1} \mathrm{Hess}_q(\xi) : \left(M^{-1} P_M(q)\right).$$

Averaging (3.2) over momenta thus leads to the desired result. □

Free energy derivatives can also be obtained from the Lagrange multipliers of the Langevin constrained process (CL). This is very useful in practice since it avoids the computation of second order derivatives of the reaction coordinates which appear in the expressions of f_{rgd}^M and $\overline{f}_{\mathrm{rgd}}^M$ (see the discussion at the beginning of Section 4.3.2):

Theorem 4.3. *Consider the rigidly constrained Langevin process solution of* (CL)*, with associated Lagrange multipliers λ_t. Assume that $\nabla\xi$, G_M^{-1} and σ are bounded functions on $\Sigma(z)$, and γ_P is strictly positive on $T_q^*\Sigma(z)$ (in the sense of symmetric matrices). Then, the almost sure convergence* (1.15) *claimed in the introduction holds:*

$$\lim_{T\to+\infty} \frac{1}{T} \int_0^T d\lambda_t = \nabla_z F_{\mathrm{rgd}}^M(z) \qquad \text{a.s.} \tag{4.13}$$

A similar result holds for the "Hamiltonian part" of the Lagrange multipliers, defined by

$$\begin{aligned} d\lambda_t^{\mathrm{ham}} &= d\lambda_t + G_M^{-1} \nabla\xi(q_t)^T M^{-1} \left(-\gamma(q_t) M^{-1} p_t \, dt + \sigma(q_t) dW_t\right) \\ &= f_{\mathrm{rgd}}^M(q_t, p_t) \, dt. \end{aligned} \tag{4.14}$$

Indeed, the following almost sure convergence holds:

$$\lim_{T\to+\infty} \frac{1}{T} \int_0^T d\lambda_t^{\mathrm{ham}} = \nabla_z F_{\mathrm{rgd}}^M(z) \qquad \text{a.s.} \tag{4.15}$$

The estimator based on (4.15) has a smaller variance than the estimator based on (1.15) (or (4.13) above) since only the bounded variation part is retained, and the martingale part due to the Brownian increments and the dissipation term are subtracted out. Similar results on variance reduction were obtained in the overdamped case in [9].

Proof. Recall the expression (3.1) of the Lagrange multipliers, which can be decomposed as the sum of the constraining force, a dissipation term and a martingale (fluctuation) term:

$$d\lambda_t = f_{\mathrm{rgd}}^M(q_t, p_t) \, dt + G_M^{-1} \nabla\xi(q_t)^T M^{-1} \left(\gamma(q_t) M^{-1} p_t \, dt - \sigma(q_t) dW_t\right). \tag{4.16}$$

The result follows from three facts. First, the process is ergodic with respect to the equilibrium distribution $\mu_{T^*\Sigma(z)}(dq\, dp)$ and averaging f_{rgd}^M yields the rigid free energy derivative in view of Proposition 4.1. This already shows (4.15).

Second, the Gaussian distribution of $\mu_{T^*\Sigma(z)}(dq\,dp)$ with respect to momentum variables is centered, which yields:

$$\int_{T^*\Sigma(z)} G_M^{-1}(q)\nabla\xi(q)^T M^{-1}\gamma(q)M^{-1}\,p\,\mu_{T^*\Sigma(z)}(dq\,dp) = 0.$$

Third, the variance of the martingale term can be uniformly bounded as

$$\mathbb{E}\left|\frac{1}{\sqrt{T}}\int_0^T G_M^{-1}(q_t)\nabla\xi^T(q_t)M^{-1}\sigma(q_t)dW_t\right|^2 \leq \left\|\mathrm{Tr}(G_M^{-1}\nabla\xi^T M^{-1}\sigma\sigma^T M^{-1}\nabla\xi G_M^{-1})\right\|_\infty.$$

This implies the almost sure convergence

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T G_M^{-1}(q_t)\nabla\xi(q_t)^T M^{-1}\sigma(q_t)dW_t = 0;$$

see for example Theorem 1.3.15 in [17]. □

The fact that averaging the Lagrange multiplier in (4.13) indeed yields the mean force may not be intuitive. This is actually very much related to the cost interpretation of the Lagrange multipliers in optimization; see [35, Remark 3.29].

4.3. **Numerical discretization of the mean force.** Estimates of the mean force based on either (4.8), (4.11) or (4.15) can be obtained.

4.3.1. *Averaging local rigid mean forces.* Free energy derivatives can be computed by averaging $\overline{f}^M_{\mathrm{rgd}}(q)$ or $f^M_{\mathrm{rgd}}(q,p)$ with respect to the distribution $\mu_{T^*\Sigma(z)}(dq\,dp)$, for instance, using the estimators:

$$\lim_{K\to+\infty}\frac{1}{K}\sum_{k=0}^{K-1}\overline{f}^M_{\mathrm{rgd}}(q^k) \qquad \text{or} \qquad \lim_{K\to+\infty}\frac{1}{K}\sum_{k=0}^{K-1} f^M_{\mathrm{rgd}}(q^k,p^k).$$

The functions $\overline{f}^M_{\mathrm{rgd}}(q)$ and $f^M_{\mathrm{rgd}}(q,p)$ may thus be called "rigid local mean forces". Note that using the momentum-averaged local mean force $\overline{f}^M_{\mathrm{rgd}}$ instead of the original f^M_{rgd} reduces the variance since the fluctuations of the momentum variable have been averaged out analytically. Table 1 below confirms this analysis, although the variance reduction appears to be small in our specific numerical experiment.

Assuming the convergence of the constrained splitting scheme (3.16)-(3.17)-(3.18) in the probability distribution sense[3] to the limiting Langevin process (CL), the convergence of these estimators to $\nabla_z F^M_{\mathrm{rgd}}(z)$ is ensured, when taking first the limit $\Delta t \to 0$ with $K = N_{\Delta t}$ such that $N_{\Delta t}\Delta t \to T$, and then $T \to \infty$.

4.3.2. *Averaging the Lagrange multipliers.* Free energy derivatives can also be computed using the Lagrange multipliers of a Langevin constrained process according to (1.15) or (4.15). This technique avoids the possibly cumbersome computation of second order derivatives $\mathrm{Hess}_q(\xi)$ of the reaction coordinate, which appear in the expressions of f^M_{rgd} or $\overline{f}^M_{\mathrm{rgd}}$. Besides, the Lagrange multipliers are needed anyway for the numerical integration of the dynamics.

[3]This convergence is also called weak convergence in probability theory. The proof of convergence in the present case may be carried out using classical results; see *e.g.* [19].

The computation can be performed as before with a longtime simulation of the splitting scheme (3.16)-(3.17)-(3.18) discretizing the Langevin process with constraints. The following approximation formula can, for instance, be used:

$$\nabla_z F^M_{\rm rgd}(z) \simeq \frac{1}{K\Delta t}\sum_{k=0}^{K-1}(\lambda^{k+1/2}+\lambda^{k+3/4}) \tag{4.17}$$

where $(\lambda^{k+1/2},\lambda^{k+3/4})$ are the Lagrange multipliers in the Hamiltonian part (3.17). The consistency of this estimator is given by the following proposition.

Proposition 4.4 (Consistency)**.** *The approximation formula* (4.17) *is consistent. More precisely, the Lagrange multipliers* $(\lambda^{n+1/2},\lambda^{n+3/4})$ *in* (3.16)-(3.17)-(3.18) *are both equivalent when* $\Delta t\to 0$ *to the constraining force defined in* (3.2)*:*

$$\begin{cases}\lambda^{n+1/2} = f^M_{\rm rgd}(q^n,p^{n+1/2})\dfrac{\Delta t}{2}+{\rm O}(\Delta t^2),\\ \lambda^{n+3/4} = f^M_{\rm rgd}(q^{n+1},p^{n+1/2})\dfrac{\Delta t}{2}+{\rm O}(\Delta t^2).\end{cases}$$

Moreover, the following second order consistency holds for the sum of the Lagrange multipliers:

$$\lambda^{n+1/2}+\lambda^{n+3/4} = \frac{\Delta t}{2}\Big(f^M_{\rm rgd}(q^n,p^{n+1/2})+f^M_{\rm rgd}(q^{n+1},p^{n+1/2})\Big)+{\rm O}(\Delta t^3), \tag{4.18}$$

together with the variant:

$$\lambda^{n+1/2}+\lambda^{n+3/4} = \frac{\Delta t}{2}\Big(f^M_{\rm rgd}(q^n,p^{n+1/4})+f^M_{\rm rgd}(q^{n+1},p^{n+3/4})\Big)+{\rm O}(\Delta t^3). \tag{4.19}$$

The variant (4.19), which involves positions and momenta at the beginning and at the end of the Hamiltonian steps only, is used in (4.21) below to estimate the time discretization error in the thermodynamic integration method based on the estimator (4.15).

Proof. For sufficiently small time-steps Δt, the implicit function theorem ensures that the two projection steps associated with the nonlinear constraints in (3.16)-(3.17)-(3.18) have a unique smooth solution. A Taylor expansion with respect to Δt of the position constraints gives

$$\begin{aligned} z &= \xi(q^{n+1}) = \xi(q^n+\Delta t M^{-1}p^{n+1/2})\\ &= \xi(q^n)+\Delta t\nabla\xi(q^n)^TM^{-1}p^{n+1/2}+\frac{\Delta t^2}{2}{\rm Hess}_{q^n}(\xi)(M^{-1}p^{n+1/2},M^{-1}p^{n+1/2})\\ &\quad+\frac{\Delta t^3}{6}{\rm D}^3_{q^n}(\xi)(M^{-1}p^{n+1/2},M^{-1}p^{n+1/2},M^{-1}p^{n+1/2})+{\rm O}(\Delta t^4),\end{aligned}$$

where $D^3_q(\xi)(x,y,z)\in\mathbb{R}^m$ denotes the order 3 differential of ξ computed at q and evaluated with the vectors $x,y,z\in\mathbb{R}^{3N}$. We denote

$$\alpha^{n+1/2}(q) := G_M^{-1}(q){\rm D}^3_q(\xi)(M^{-1}p^{n+1/2},M^{-1}p^{n+1/2},M^{-1}p^{n+1/2}).$$

Then, the fact that $z=\xi(q^{n+1})=\xi(q^n)$ and the identity

$$\nabla\xi(q^n)^TM^{-1}p^{n+1/2} = -\frac{\Delta t}{2}\nabla\xi(q^n)^TM^{-1}\nabla V(q^n)+G_M(q^n)\lambda^{n+1/2}$$

yield the following expansion of $\lambda^{n+1/2}$ in terms of $(q^n,p^{n+1/2})$:

$$\lambda^{n+1/2} = f^M_{\rm rgd}(q^n,p^{n+1/2})\frac{\Delta t}{2}-\frac{\Delta t^2}{6}\alpha^{n+1/2}(q^n)+{\rm O}(\Delta t^3).$$

By time symmetry, the same computation holds for $\lambda^{n+3/4}$, starting from $(q^{n+1}, p^{n+3/4})$ and by formally replacing Δt by $-\Delta t$. This can be double checked by Taylor expanding with respect to Δt the position constraints, as done above for $\lambda^{n+1/2}$. It thus holds that

$$\lambda^{n+3/4} = f^M_{\rm rgd}(q^{n+1}, p^{n+1/2})\frac{\Delta t}{2} + \frac{\Delta t^2}{6}\alpha^{n+1/2}(q^{n+1}) + \mathrm{O}(\Delta t^3).$$

The sum of the multipliers therefore reads

$$\begin{aligned}\lambda^{n+1/2} + \lambda^{n+3/4} - f^M_{\rm rgd}(q^n, p^{n+1/2})\frac{\Delta t}{2} - f^M_{\rm rgd}(q^{n+1}, p^{n+1/2})\frac{\Delta t}{2}\\ = \frac{\Delta t^2}{6}\left(\alpha^{n+1/2}(q^{n+1}) - \alpha^{n+1/2}(q^n)\right) + \mathrm{O}(\Delta t^3) = \mathrm{O}(\Delta t^3),\end{aligned}$$

which gives (4.18). Now, using the previous calculations, we remark that

$$\begin{cases} p^{n+1/2} = p^{n+1/4} - \dfrac{\Delta t}{2}\nabla V(q^n) + \dfrac{\Delta t}{2}\nabla\xi(q^n) f^M_{\rm rgd}(q^n, p^{n+1/2}) + \mathrm{O}(\Delta t^2),\\ p^{n+1/2} = p^{n+3/4} + \dfrac{\Delta t}{2}\nabla V(q^{n+1}) - \dfrac{\Delta t}{2}\nabla\xi(q^{n+1}) f^M_{\rm rgd}(q^{n+1}, p^{n+1/2}) + \mathrm{O}(\Delta t^2).\end{cases}$$

A simple computation then shows

$$f^M_{\rm rgd}(q^n, p^{n+1/2}) + f^M_{\rm rgd}(q^{n+1}, p^{n+1/2}) = f^M_{\rm rgd}(q^n, p^{n+1/4}) + f^M_{\rm rgd}(q^{n+1}, p^{n+3/4}) + \mathrm{O}(\Delta t^2).$$

This gives the claimed second order consistency of the sum of the Lagrange multipliers (4.19). □

Let us discuss the convergence of the approximation (4.17). Assuming again that the constrained splitting scheme (3.16)-(3.17)-(3.18) converges in the probability distribution sense to the limiting Langevin process (CL), the following convergence in probability distribution occurs when $\Delta t \to 0$ and $N_{\Delta t}\Delta t \to T$:

$$\lim_{\Delta t\to 0} \mathrm{Law}\left(\frac{1}{N_{\Delta t}\Delta t}\sum_{n=0}^{N_{\Delta t}-1}(\lambda^{n+1/2} + \lambda^{n+3/4})\right) = \mathrm{Law}\left(\frac{1}{T}\int_0^T d\lambda_t^{\rm ham}\right). \tag{4.20}$$

This shows the convergence of the estimate (4.17) of the mean force when taking first the limit $\Delta t \to 0$ and then $T \to \infty$.

4.3.3. *Estimates relying on the Metropolized scheme.* When the scheme (3.16)-(3.17)-(3.18) is complemented with a Metropolis step (see Algorithm 3.5), it is possible to prove a result on the longtime limit of trajectorial averages (*i.e.* letting first the number of iterations go to infinity, and then taking the limit $\Delta t \to 0$), upon assuming the irreducibility of the numerical scheme.

Indeed, let us consider the Markov chain (q^k, p^k) generated by the GHMC scheme in Algorithm 3.5, and assume (i) the irreducibility of the Markov chain, and (ii) that appropriate rejections outside the set $\overline{D}_{\Delta t} = \Sigma(z) \times \{\frac{1}{2}p^T M^{-1} p \le R_{\Delta t}\}$ are made in the steps (1), (2), (4) of the algorithm. In particular, the projection steps associated with the nonlinear constraints in Step (2) of Algorithm 3.5 are well defined.

Then, by ergodicity, an average of the analytic expression of the local rigid mean force $\overline{f}^M_{\rm rgd}$ given in (4.12) yields an estimate of the free energy without time

discretization error,

$$\lim_{K\to+\infty} \frac{1}{K} \sum_{k=0}^{K-1} \overline{f}^M_{\rm rgd}(q^k) = \nabla_z F^M_{\rm rgd}(z) \quad \text{a.s.}$$

If $f^M_{\rm rgd}$ is used instead of $\overline{f}^M_{\rm rgd}$, then the mean force is computed with some exponentially small error: almost surely,

$$\begin{aligned}\lim_{K\to+\infty} \frac{1}{K} \sum_{k=0}^{K-1} f^M_{\rm rgd}(q^k,p^k) &= \frac{\displaystyle\int_{\overline{D}_{\Delta t}} f^M_{\rm rgd}(q,p)\mu_{T^*\Sigma(z)}(dq\,dp)}{\displaystyle\int_{\overline{D}_{\Delta t}} \mu_{T^*\Sigma(z)}(dq\,dp)} \\ &= \nabla_z F^M_{\rm rgd}(z) + \mathrm{O}\left(\mathrm{e}^{-\alpha \Delta t^{-2}}\right)\end{aligned}$$

for some $\alpha > 0$. The error arising from replacing the integration over $\overline{D}_{\Delta t}$ by an integration over $T^*\Sigma(z)$ is indeed exponentially small in view of (3.22) (namely $R_{\Delta t} \geq A\,\Delta t^{-2}$) and using the fact that the marginal distribution in the p-variable is Gaussian.

Likewise, for estimates based on Lagrange multipliers, the following longtime averaging holds: almost surely,

$$\begin{aligned}&\lim_{K\to+\infty} \frac{1}{K\Delta t} \sum_{k=0}^{K-1} (\lambda^{k+1/2} + \lambda^{k+3/4}) \\ &\quad = \frac{\displaystyle\int_{\overline{D}_{\Delta t}} f^M_{\rm rgd}(q,p)\mu_{T^*\Sigma(z)}(dq\,dp)}{\displaystyle\int_{\overline{D}_{\Delta t}} \mu_{T^*\Sigma(z)}(dq\,dp)} + \mathrm{O}(\Delta t^2),\end{aligned} \tag{4.21}$$

where we have used the estimate (4.19) on the Lagrange multipliers. The limit $\Delta t \to 0$ is obtained by a dominated convergence argument:

$$\lim_{\Delta t\to 0} \lim_{K\to+\infty} \frac{1}{K\Delta t} \sum_{n=0}^{K-1} (\lambda^{k+1/2} + \lambda^{k+3/4}) = \nabla_z F^M_{\rm rgd}(z) \quad \text{a.s.}$$

Note that, due to the Metropolis correction in Algorithm 3.5, the time discretization error in the sampling of the invariant measure is removed. The only remaining time discretization errors come from (i) the approximation of the local mean force by the Lagrange multipliers (this is a second order error), and (ii) the integration domain being $\overline{D}_{\Delta t}$ instead of $T^*\Sigma(z)$ (as discussed above, this is an exponentially small error in Δt). In conclusion, the left-hand side of (4.21) is an approximation of $\nabla_z F^M_{\rm rgd}(z)$ up to a $\mathrm{O}(\Delta t^2)$ error term.

4.3.4. *Overdamped limit.* Finally, let us emphasize that free energy derivatives can be computed with the estimator (4.17) within the overdamped Langevin framework, using the scheme (3.27) and the expressions (3.28)-(3.29)-(3.30) of Proposition 3.6. Let us recall that the latter are equivalent to the scheme (3.16)-(3.17)-(3.18) with fluctuation-dissipation matrices satisfying $\frac{\Delta t}{4}\gamma = M = \frac{\Delta t}{2}\mathrm{Id}$. This leads to the original free energy estimator (recall that, for the overdamped dynamics, $\mathbb{R}^{3N}$ is

equipped with the scalar product associated with the identity matrix):

$$\nabla_z F^{\mathrm{Id}}_{\mathrm{rgd}}(z) \simeq \frac{1}{K\Delta t}\sum_{k=0}^{K-1}(\lambda^{k+1/2}+\lambda^{k+3/4}), \tag{4.22}$$

which can be seen as a variant of the variance reduced estimator proposed directly for the overdamped scheme (3.27) in [9]:

$$\nabla_z F^{\mathrm{Id}}_{\mathrm{rgd}}(z) \simeq \frac{1}{K\Delta t}\sum_{k=0}^{K-1}\left(\lambda^{k+1}_{\mathrm{od}}+\sqrt{\frac{2\Delta t}{\beta}}G^{-1}(q^k)\nabla\xi(q^k)^T\mathcal{G}^k\right)=\frac{1}{K\Delta t}\sum_{k=0}^{K-1}2\lambda^{k+1/2}.$$

The rigorous justification of the consistency of (4.22) in the limit $\Delta t \to 0$ follows from the results of [9].

4.4. **Numerical illustration.** We consider a system composed of N particles in a 2-dimensional periodic box of side length L, interacting through the purely repulsive WCA pair potential, which is a truncated Lennard-Jones potential:

$$V_{\mathrm{WCA}}(r)=\begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right]+\varepsilon & \text{if } r\le r_0,\\ 0 & \text{if } r>r_0,\end{cases}$$

where r denotes the distance between two particles, ε and σ are two positive parameters and $r_0 = 2^{1/6}\sigma$. Among these particles, two (numbered 1 and 2 in the following) are designated to form a dimer while the others are solvent particles. Instead of the above WCA potential, the interaction potential between the two particles of the dimer is a double-well potential

$$V_{\mathrm{S}}(r)=h\left[1-\frac{(r-r_0-w)^2}{w^2}\right]^2,$$

where h and w are two positive parameters. The total energy of the system is therefore, for $q \in (L\mathbb{T})^{dN}$ with $d = 2$,

$$V(q)=V_{\mathrm{S}}(|q_1-q_2|)+\sum_{3\le i<j\le N}V_{\mathrm{WCA}}(|q_i-q_j|)+\sum_{i=1,2}\sum_{3\le j\le N}V_{\mathrm{WCA}}(|q_i-q_j|).$$

See [13, 46] for instance for other computational studies using this model.

The potential V_{S} exhibits two energy minima, one corresponding to the compact state where the length of the dimer is $r = r_0$, and one corresponding to the stretched state where this length is $r = r_0 + 2w$. The energy barrier separating both states is h. The reaction coordinate used to describe the transition from the compact to the stretched state is the normalized bond length of the dimer molecule:

$$\xi(q)=\frac{|q_1-q_2|-r_0}{2w}, \tag{4.23}$$

where q_1 and q_2 are the positions of the two particles forming the dimer. The compact state (resp. the stretched state) corresponds to the value $z = 0$ (resp. $z = 1$) of the reaction coordinate.

The inverse temperature is set to $\beta = 1$, with $N = 100$ particles ($N - 2$ solvent particles and the dimer) with solvent density $\rho = (1-2/N)a^{-2} = 0.436$, since there are $N - 2$ solvent particles in a square box of side length $L = a\sqrt{N}$ with $a = 1.5$. The parameters describing the WCA interactions are set to $\sigma = 1$ and $\varepsilon = 1$, and the additional parameters for the dimer are $w = 2$ and $h = 2$.

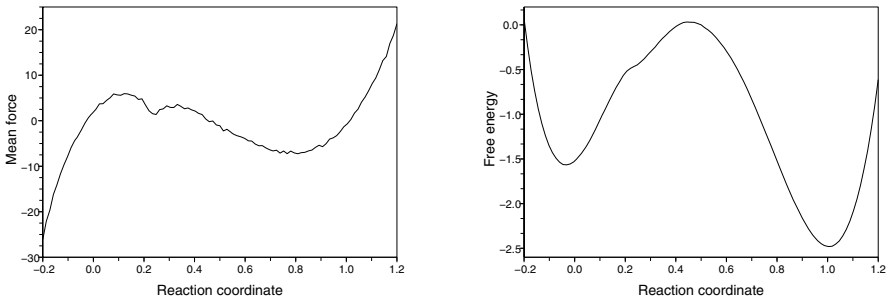

FIGURE 1. Left: Estimated mean force. Right: Corresponding free energy profile.

For this system, $M = \mathrm{Id}$ and $|\nabla\xi|$ is constant, so that the rigid free energy $F^M_{\mathrm{rgd}}(z)$ is equal to the free energy $F(z)$.

The mean force is estimated at the values $z_i = z_{\min} + i\Delta z$, with $z_{\min} = -0.2$, $z_{\max} = 1.2$ and $\Delta z = 0.014$, by ergodic averages obtained with the projected dynamics with Metropolis correction (Algorithm 3.5, where in the simple case considered here, the fluctuation-dissipation part can be integrated exactly). For each value of z, we integrate the dynamics on a time $T = 2 \times 10^4$ with a step size $\Delta t = 0.02$, using a scalar friction coefficient $\gamma = 1$.

The resulting mean force profile is presented in Figure 1, together with the associated free energy profile. Figure 2 compares the analytical constraining force $f^M_{\mathrm{rgd}}(q^n, p^n)$ and the Lagrange multipliers; see Proposition 4.4. In Figure 2, the x-axis represents the blocks of 10^5 simulation steps, concatenated for the 101 different values of z_i. It can be seen that the difference between $f^M_{\mathrm{rgd}}(q^n, p^n)$ and the Lagrange multipliers is small in any cases, though somewhat larger for the lowest values of ξ.

It can be checked numerically that the differences $|\lambda^{n+1/2} - f^M_{\mathrm{rgd}}(q^n, p^{n+1/2})\frac{\Delta t}{2}|$ and $|\lambda^{n+3/4} - f^M_{\mathrm{rgd}}(q^{n+1}, p^{n+1/2})\frac{\Delta t}{2}|$ are indeed of order Δt^2, and that the difference

$$\left|\lambda^{n+1/2} + \lambda^{n+3/4} - f^M_{\mathrm{rgd}}(q^n, p^{n+1/2})\frac{\Delta t}{2} - f^M_{\mathrm{rgd}}(q^{n+1}, p^{n+1/2})\frac{\Delta t}{2}\right|$$

is indeed of order Δt^3 (by computing the average of these elementary differences for various step sizes). The Lagrange multipliers are in any case very good approximations to the constraining force f^M_{rgd}.

Let us finally discuss the efficiency of the different estimators of the mean force, in terms of their variances. They can be written as the empirical average of the following random sequences:

$$\left(f^M_{\mathrm{rgd}}(q^n, p^n), \overline{f}^M_{\mathrm{rgd}}(q^n), \frac{\lambda^{n+1/2} + \lambda^{n+3/4}}{\Delta t}\right),$$

where $q^n, p^n, \lambda^{n+1/2}, \lambda^{n+3/4}$ are given by the numerical scheme (3.16)-(3.17)-(3.18). The correlations in time (between the iterates) are very similar for the three methods, and we therefore simply compute the variance over all the samples. Table 1 compares the so-obtained standard errors over 10^5 time-steps with $\Delta t = 0.02$ (simulation time $T = 2000$ for each value of the reaction coordinate). The results show

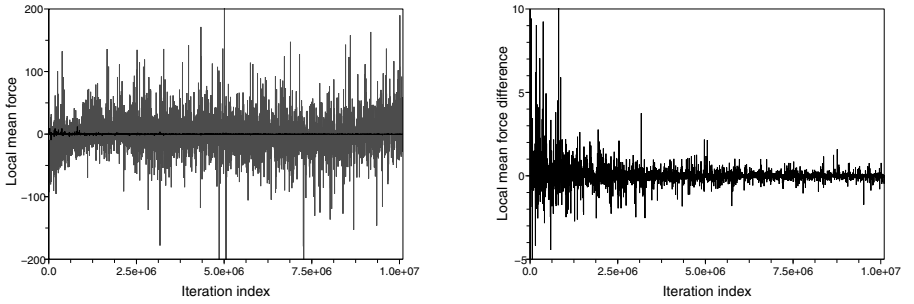

FIGURE 2. Left: The constraining force $f^M_{\mathrm{rgd}}(q^n, p^n)$ (pale line), and the difference between the constraining force and its estimate from the Lagrange multipliers (dark line). Right: Zoom on the difference between the constraining force and the Lagrange multipliers. Note the difference of scales for the y-axis. In all figures, the x-axis represents the blocks of 10^5 simulation steps, concatenated for the 101 different values of z_i.

that the different estimators are more or less equivalent. This is related to the fact that the essential source of variance comes from the sampling of the positions, and not the sampling of the velocities. Note, however, that, for the smallest value of the reaction coordinate, the estimator based on the averaged local mean force $\overline{f}^M_{\mathrm{rgd}}(q^n)$ appears to be better in terms of variance.

TABLE 1. Standard error (square-root of the variance) of three mean force estimators, with correlations in time neglected, for different values z of the reaction coordinate.

z	$f^M_{\mathrm{rgd}}(q^n, p^n)$	$\dfrac{\lambda^{n+1/2} + \lambda^{n+3/4}}{\Delta t}$	$\overline{f}^M_{\mathrm{rgd}}(q^n)$
-0.2	22.1	21.9	14.7
0.0	16.0	15.5	15.4
0.2	23.1	22.5	22.9
0.4	21.1	20.4	21.0
0.6	21.4	20.7	21.3
0.8	21.6	20.9	21.5
1.0	21.4	20.6	21.4
1.2	21.0	20.3	20.9

5. HAMILTONIAN AND LANGEVIN NONEQUILIBRIUM DYNAMICS

This section presents nonequilibrium Hamiltonian and Langevin dynamics with time-evolving constraints. We thus consider (q_t, p_t) solution to the dynamics (SCL),

which we recall for convenience:

$$\text{(SCL)} \qquad \begin{cases} dq_t = M^{-1} p_t \, dt, \\ dp_t = -\nabla V(q_t) \, dt - \gamma_P(q_t) M^{-1} p_t \, dt + \sigma_P(q_t) \, dW_t + \nabla \xi(q_t) \, d\lambda_t, \\ \xi(q_t) = z(t), \qquad (C_q(t)). \end{cases}$$

We prove, in particular, the fluctuation identity (1.19); see (5.22) below. Recall (see (1.16)) that, for simplicity, we assume in this section that the fluctuation-dissipation matrices are assumed to be of the form $(\sigma_P, \gamma_P) = (P_M \, \sigma, P_M \, \gamma \, P_M^T)$ with $\gamma, \sigma \in \mathbb{R}^{3N \times 3N}$. At variance with the previous sections, we do *not* assume that γ_P is strictly positive. Actually, $\gamma_P = 0$ corresponds to an interesting case: the deterministic Hamiltonian dynamics.

To our knowledge, the standard work fluctuations derived so far (except for our previous work [34]) apply only to the case of time-dependent Hamiltonians. It is possible to consider transitions in the values of some reaction coordinate in this framework upon resorting to steered molecular dynamics techniques. In this case, a penalty term $\varepsilon^{-1}(\xi(q) - z(t))^2$ (with ε small) is used in the energy of the system to "softly" constrain the system to remain close to the submanifold $\Sigma(z(t)) = \{q \in \mathbb{R}^{3N} \, | \, \xi(q) = z(t)\}$ at time t. However, it is observed in practice that the statistical fluctuations increase with smaller ε (see [42]). We propose instead to replace the stiff constraining potential $\varepsilon^{-1}(\xi(q)-z)^2$ by a projection onto the submanifold $\Sigma(z)$. This is reminiscent of the replacement of stiff constrained Langevin dynamics by rigidly constrained ones; see Remark 3.3.

This section is organized as follows. We first define the generalized free energy which is naturally computed with (SCL), and relate it to the standard free energy (1.12) in Section 5.1. Then, we give some precisions on the nonequilibrium dynamics (SCL) in Section 5.2. Next, we prove an appropriate version of the Jarzynski-Crooks fluctuation equality in Section 5.3. A numerical discretization of the nonequilibrium dynamics is proposed in Section 5.4, together with various approximations of the work. In particular, we propose a numerical strategy to obtain a Jarzynski-Crooks identity without time discretization error (see Section 5.4.5). We then consider the overdamped limit when the mass matrix M goes to 0 (see Section 5.5). Finally, in Section 5.6, we present some numerical results for the model system already considered in Section 4.4.

5.1. **Generalized free energy.** For the (q_t, p_t) solution to the Langevin dynamics (SCL), the reaction coordinate evolution $\xi(q_t) = z(t)$ implies that $v_\xi(q_t, p_t) = \dot{z}(t)$, so that, at each time $t \geq 0$, the system (q_t, p_t) belongs to the state space $\Sigma_{\xi, v_\xi}(z(t), \dot{z}(t))$. As a consequence, the free energy difference computed in this section by the Jarzynski relation without correction (see (5.20) below) is in fact the generalized rigid free energy F^{Ξ}_{rgd} defined in (4.7), in the special case $\Xi = (\xi, v_\xi)^T$:

$$F^{\Xi}_{\mathrm{rgd}}(\zeta) = -\frac{1}{\beta} \ln \int_{\Sigma_\Xi(\zeta)} \mathrm{e}^{-\beta H(q,p)} \sigma_{\Sigma_\Xi(\zeta)}(dq \, dp).$$

The latter free energy is associated to the normalization constant $Z_{z(t), \dot{z}(t)}$ of the distribution $\mu_{\Sigma_{\xi, v_\xi}(z(t), \dot{z}(t))}$ defined by (2.15). The generalized rigid free energy (4.7) can be explicitly related to the usual free energy as follows. First, remark that, for

a fixed q,

$$\begin{aligned}\int_{\Sigma_{v_\xi(q,\cdot)}(v_z)} &\exp\left(-\frac{\beta}{2}p^TM^{-1}p\right)\sigma^{M^{-1}}_{\Sigma_{v_\xi(q,\cdot)}(v_z)}(dp)\\ &= \exp\left(-\frac{\beta}{2}v_z^TG_M^{-1}(q)v_z\right)\int_{T_q^*\Sigma(z)}\exp\left(-\frac{\beta}{2}p^TM^{-1}p\right)\sigma^{M^{-1}}_{T_q^*\Sigma(z)}(dp)\\ &= \left(2\pi\beta^{-1}\right)^{\frac{3N-m}{2}}\exp\left(-\frac{\beta}{2}v_z^TG_M^{-1}(q)v_z\right).\end{aligned}$$

In the above, the change of variable $p \to p - \nabla\xi(q)G_M^{-1}(q)v_z$ has been used, in the space

$$\Sigma_{v_\xi(q,\cdot)}(v_z) = \Big\{p\in\mathbb{R}^{3N}\;\Big|\;\nabla\xi(q)^TM^{-1}\Big(p-\nabla\xi(q)G_M^{-1}(q)v_z\Big)=0\Big\}.$$

Note that $\frac{1}{2}v_z^TG_M^{-1}(q)v_z$ can be interpreted as the kinetic energy of the reaction coordinate ξ. Using the decomposition of measures (2.24) and the above calculations, an alternative expression of the generalized free energy is

$$F_{\mathrm{rgd}}^{\xi,v_\xi}(z,v_z) = -\frac{1}{\beta}\ln\int_{\Sigma(z)}\exp\left(-\beta V(q)-\frac{\beta}{2}v_z^TG_M^{-1}(q)v_z\right)\sigma^M_{\Sigma(z)}(dq)+\mathrm{C}, \tag{5.1}$$

where, as usual, C denotes a generic constant (independent of z) whose value may vary from line to line. As a consequence, the standard free energy (1.12) is easily recovered from the generalized free energy, using relations similar to (1.14). Indeed, using (5.1), and with computations similar to the ones leading to (4.2), the difference of the two free energies writes:

$$\begin{aligned}F(z) &- F_{\mathrm{rgd}}^{\xi,v_\xi}(z,v_z)\\ &= -\frac{1}{\beta}\ln\int_{\Sigma_{\xi,v_\xi}(z,v_z)}(\det G_M(q))^{-1/2}\exp\left(\frac{\beta}{2}v_z^TG_M^{-1}(q)v_z\right)\mu_{\Sigma_{\xi,v_\xi}(z,v_z)}(dq\,dp)+\mathrm{C}.\end{aligned} \tag{5.2}$$

In practical nonequilibrium computations, the profile $t\mapsto F(z(t))$ can then be computed by adding a corrector to the work value in the Jarzynski estimator computing $F_{\mathrm{rgd}}^{\xi,v_\xi}(z(t),\dot z(t))$. This yields the identity (5.22) mentioned in the introduction and proved below (see the discussion after Theorem 5.3).

5.2. **Dynamics and generators.** The explicit expression of the Lagrange multipliers in (SCL) is obtained by a computation similar to (3.1) for the case without switching, by differentiating twice the constraints over time:

$$\frac{d^2}{dt^2}\xi(q_t) = \ddot z(t).$$

In view of the special structure of (σ_P,γ_P), this leads to

$$\begin{aligned}d\lambda_t &= f_{\mathrm{rgd}}^M(q_t,p_t)\,dt + G_M^{-1}(q_t)\ddot z(t)\,dt\\ &\quad + G_M^{-1}\nabla\xi(q_t)^TM^{-1}\left(\gamma_P(q_t)M^{-1}p_t\,dt-\sigma_P(q_t)\,dW_t\right)\\ &= f_{\mathrm{rgd}}^M(q_t,p_t)\,dt + G_M^{-1}(q_t)\ddot z(t)\,dt.\end{aligned} \tag{5.3}$$

The expression (5.3) does not depend on the fluctuation-dissipation tensors (σ_P,γ_P). This leads to simplified computations and motivates the special form of the latter

matrices. The momentum evolution (SCL) thus simplifies as

$$\begin{aligned} dp_t = &-\nabla V(q_t)\,dt + \nabla\xi(q_t) f^M_{\mathrm{rgd}}(q_t,p_t)\,dt + \nabla\xi(q_t) G_M^{-1}(q_t)\ddot{z}(t)\,dt \\ &- \gamma_P(q_t) M^{-1} p_t\,dt + \sigma_P(q_t)\,dW_t. \end{aligned} \tag{5.4}$$

Let us denote by $\mathcal{L}^{\mathrm{f}}_t$ the generator of the forward dynamics $t \mapsto (q_t, p_t)$ defined in (SCL). The latter has a backward switching version, $t' \mapsto (q^{\mathrm{b}}_{t'}, p^{\mathrm{b}}_{t'})$, obtained by using a time reversed switching $t' \mapsto z(T-t')$, and by reversing the momentum first in the initial condition, and then reversing them back after the time evolution (see [5] for more general backward dynamics). More precisely, the backward dynamics can be defined through its generator

$$\mathcal{L}^{\mathrm{b}}_{t'} = \mathcal{R}\,\mathcal{L}^{\mathrm{f}}_{T-t'}\,\mathcal{R}, \tag{5.5}$$

where $\mathcal{L}^{\mathrm{f}}_{T-t'}$ is the generator of the forward process at time $t = T-t'$, and $\mathcal{R} : \phi \mapsto \phi \circ S$ is the momentum flip operator with $S(q,p) = (q,-p)$. Thus $t' \mapsto (q^{\mathrm{b}}_{t'}, -p^{\mathrm{b}}_{t'})$ is solution of the forward evolution equation (SCL) with a switching schedule $t' \mapsto z(T-t')$. Therefore, the time evolution of the backward dynamics is given by

$$\begin{cases} dq^{\mathrm{b}}_{t'} = -M^{-1} p^{\mathrm{b}}_{t'} dt', \\ dp^{\mathrm{b}}_{t'} = \nabla V(q^{\mathrm{b}}_{t'})\,dt' - \gamma_P(q^{\mathrm{b}}_{t'}) M^{-1} p^{\mathrm{b}}_{t'}\,dt' + \sigma_P(q^{\mathrm{b}}_{t'})\,dW^{\mathrm{b}}_{t'} + \nabla\xi(q^{\mathrm{b}}_{t'})\,d\lambda^{\mathrm{b}}_{t'}, \\ \xi(q^{\mathrm{b}}_{t'}) = z(T-t'). \end{cases} \tag{5.6}$$

In the following proposition, the expressions of $\mathcal{L}^{\mathrm{f}}_t$ and $\mathcal{L}^{\mathrm{b}}_{t'}$ are explicitly written.

Proposition 5.1. *Consider $\zeta(t) = (z(t), \dot{z}(t))$. Then, the generator of the forward process* (SCL) *at time $t \in [0,T]$ reads:*

$$\mathcal{L}^{\mathrm{f}}_t = \{\,\cdot\,, H\}_\Xi + \mathcal{L}^{\mathrm{thm}}_\Xi + \{\cdot, \Xi\}\,\Gamma^{-1}\dot{\zeta}(t), \tag{5.7}$$

and the generator of the backward process (5.6) *at time $t' \in [0,T]$ reads:*

$$\mathcal{L}^{\mathrm{b}}_{t'} = -\{\,\cdot\,, H\}_\Xi + \mathcal{L}^{\mathrm{thm}}_\Xi - \{\cdot, \Xi\}\,\Gamma^{-1}\dot{\zeta}(T-t'), \tag{5.8}$$

where

$$\mathcal{L}^{\mathrm{thm}}_\Xi = \frac{1}{\beta}\,\mathrm{e}^{\beta H}\mathrm{div}_p\Big(e^{-\beta H}\,\gamma_P\,\nabla_p\,\cdot\,\Big)$$

is the fluctuation-dissipation operator defined in (3.5).

Proof. First, let us consider the terms in (SCL) arising from the Hamiltonian evolution and from the switching (*i.e.* without fluctuation-dissipation, which amounts to setting $\gamma_P = 0$ and $\sigma_P = 0$ in (SCL)). Since during this dynamics $v_\xi(q_t,p_t) = \dot{z}(t)$, (3.6) yields:

$$\{\Xi, H\}\,(q_t,p_t) = \begin{pmatrix} \dot{z}(t) \\ \mathrm{Hess}_{q_t}(\xi)(M^{-1}p_t, M^{-1}p_t) - \left(\nabla\xi^T M^{-1}\nabla V\right)(q_t) \end{pmatrix},$$

so that, using (3.7),

$$\Gamma^{-1}(q_t,p_t)\left(\{\Xi,H\}\,(q_t,p_t) - \dot{\zeta}(t)\right) = \begin{pmatrix} G_M^{-1}(q_t)\ddot{z}(t) + f^M_{\mathrm{rgd}}(q_t,p_t) \\ 0 \end{pmatrix}. \tag{5.9}$$

With (3.9), we then obtain:

$$\begin{aligned} &\{\varphi,\Xi\}\,\Gamma^{-1}\left(\dot{\zeta}(t) - \{\Xi,H\}\right)(q_t,p_t) \\ &\quad = \left(G_M^{-1}(q_t)\ddot{z}(t) + f^M_{\mathrm{rgd}}(q_t,p_t)\right)^T \nabla\xi(q_t)^T\nabla_p\varphi(q_t,p_t). \end{aligned}$$

Now, the Hamiltonian part of the switched dynamics (SCL) (see also (5.4)) can be recognized in the latter equation, so that the generator $\mathcal{L}_t^{\mathrm{f}}$ when $(\gamma_P, \sigma_P) = (0,0)$ reads: for any smooth test function φ,

$$\begin{aligned}\mathcal{L}_t^{\mathrm{f}}(\varphi) &= \left(\nabla\xi f_{\mathrm{rgd}}^M + \nabla\xi G_M^{-1}\ddot{z}(t)\right)^T \nabla_p\varphi - (\nabla V)^T \nabla_p\varphi + p^T M^{-1}\nabla_q\varphi \\ &= \{\varphi,\Xi\}\,\Gamma^{-1}(\dot{\zeta}(t) - \{\Xi,H\}) + \{\varphi,H\} \\ &= \{\varphi,H\}_\Xi + \{\varphi,\Xi\}\,\Gamma^{-1}\dot{\zeta}(t). \qquad (5.10)\end{aligned}$$

The full expression of the generator $\mathcal{L}_t^{\mathrm{f}}$ is then obtained by adding the terms arising from the fluctuation-dissipation. These terms are directly obtained from the terms involving γ_P and σ_P in (5.4), as in the proof of Proposition 3.1.

The generator of the backward switching process given by (5.6) can be obtained from similar computations. First, the thermostat parts in (5.6) and in (SCL) are the same. Consider now the Hamiltonian part (obtained by taking $(\gamma_P, \sigma_P) = (0,0)$) in the dynamics (5.6). By definition of the backward dynamics, and the expression (5.10) of the forward dynamics, the Hamiltonian part reads

$$\begin{aligned}\mathcal{L}_{t'}^{\mathrm{b}}(\varphi)(q,p) &= \mathcal{R}\mathcal{L}_{T-t'}^{\mathrm{f}}\left(\mathcal{R}(\varphi)\right)(q,p) \\ &= \left(\nabla\xi(q) f_{\mathrm{rgd}}^M(q,p) + \nabla\xi(q)G_M^{-1}(q)\ddot{z}(T-t')\right)^T (-\nabla_p\varphi) \\ &\quad - \nabla V(q)^T(-\nabla_p\varphi) - p^T M^{-1}\nabla_q\varphi,\end{aligned}$$

so that $\mathcal{L}_{t'}^{\mathrm{b}}\varphi = -\mathcal{L}_{T-t'}^{\mathrm{f}}\varphi = -\{\varphi,H\}_\Xi - \{\varphi,\Xi\}\,\Gamma^{-1}\dot{\zeta}(T-t')$. This gives (5.8). □

5.3. **Jarzynski-Crooks identity.** Before stating the main result of this section (Theorem 5.3 below), we need to introduce a notion of work. This quantity is most conveniently defined for deterministic dynamics, but the corresponding definition is also valid for stochastic dynamics.

We define the work $(\mathcal{W}_t)_{t\geq 0}$ associated with the constraining force $\nabla\xi(q_t)\,d\lambda_t$ in (SCL) as the physical displacement multiplied by the force

$$\begin{aligned}d\mathcal{W}_t &:= \left(\frac{dq_t}{dt}\right)^T \circ \left(\nabla\xi(q_t)\,d\lambda_t\right) = \left(\frac{dq_t}{dt}\right)^T \nabla\xi(q_t) \circ d\lambda_t = \dot{z}^T(t)\circ d\lambda_t \\ &= \dot{z}^T(t)\,d\lambda_t. \qquad (5.11)\end{aligned}$$

By convention, $\mathcal{W}_0 = 0$. In the above computations, we used successively the fact that $t \mapsto \xi(q_t)$, and then $t \mapsto z(t)$ are differentiable processes, so that Stratonovitch and Itô integrations are equivalent. Let us introduce the deterministic version of the nonequilibrium process (SCL) (*i.e.* $(\gamma_P, \sigma_P) = (0,0)$):

$$\left\{\begin{array}{l} d\tilde{q}_t = M^{-1}\tilde{p}_t\,dt, \\ \\ d\tilde{p}_t = -\nabla V(\tilde{q}_t)\,dt + \nabla\xi(\tilde{q}_t)\,d\tilde{\lambda}_t, \\ \\ \xi(\tilde{q}_t) = z(t), \qquad (C_q(t)) \end{array}\right. \qquad (5.12)$$

and denote by $\Phi_{t,t+h} : \Sigma_{\xi,v_\xi}(z(t),\dot{z}(t)) \to \Sigma_{\xi,v_\xi}(z(t+h),\dot{z}(t+h))$ the associated flow between time $t \in [0,T]$ and $t+h \in [0,T]$. The work can now be written out more explicitly using the following lemma:

Lemma 5.2. *The infinitesimal variation of the work* (5.11) *reads:*

$$d\mathcal{W}_t = w(t,q_t,p_t)\,dt,$$

where for all $t \in [0,T]$ *and all* $(q,p) \in \Sigma_{\xi,v_\xi}(z(t),\dot z(t))$,

$$w(t,q,p) = \dot\zeta(t)^T \Gamma^{-1} \left\{\Xi, H\right\}(q,p) \tag{5.13}$$

$$= \dot z(t)^T \left(G_M^{-1}(q)\ddot z(t) + f^M_{\rm rgd}(q,p)\right) \tag{5.14}$$

$$= \left(\frac{d}{dh} H \circ \Phi_{t,t+h}\right)\Big|_{h=0} (q,p). \tag{5.15}$$

The total exchanged work is then a time integral associated with the path $t \mapsto (q_t,p_t)$, *and is denoted by* (*recall that* $\mathcal{W}_t$ *is defined in* (5.11)):

$$\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{0\le t\le T}\right) = \mathcal{W}_T = \int_0^T w(t,q_t,p_t)\,dt.$$

Note that the expression (5.15) can be interpreted as the energy variation of the system during the switching when the stochastic thermostat is turned off.

Proof. The expression of the Lagrange multipliers in (5.3) yields (5.14):

$$\dot z(t)^T d\lambda_t = \dot z(t)^T \left(G_M^{-1}(q_t)\ddot z(t) + f^M_{\rm rgd}(q_t,p_t)\right)\,dt.$$

Moreover, (5.9) gives:

$$\begin{aligned}\dot z(t)^T \left(G_M^{-1}(q_t)\ddot z(t) + f^M_{\rm rgd}(q_t,p_t)\right)\,dt &= \dot\zeta(t)^T\Gamma^{-1}(q_t,p_t)\left(\left\{\Xi,H\right\}(q_t,p_t) - \dot\zeta(t)\right)\\ &= \dot\zeta(t)^T\Gamma^{-1}\left\{\Xi,H\right\}(q_t,p_t),\end{aligned}$$

where in the last line we have used $\dot\zeta(t)^T\Gamma^{-1}\dot\zeta(t) = 0$. This gives (5.13). To prove (5.15), we compute the variations of the energy $H(\tilde q_t,\tilde p_t)$ for $(\tilde q_t,\tilde p_t)$ solution of (5.12) with initial condition (q,p):

$$\begin{aligned} dH(\tilde q_t,\tilde p_t) &= \tilde p_t^T M^{-1}\,d\tilde p_t + \tilde p_t^T M^{-1}\nabla V(\tilde q_t)\,dt\\ &= \dot z(t)^T\,d\tilde\lambda_t\\ &= \dot z(t)^T\left(G_M^{-1}(\tilde q_t)\ddot z(t) + f^M_{\rm rgd}(\tilde q_t,\tilde p_t)\right)\,dt.\end{aligned}$$

The last equality is obtained using the computation of the Lagrange multipliers in (5.3). This yields (5.15). □

We are now in position to state the main result of this section.

Theorem 5.3 (Jarzynski-Crooks fluctuation identity)**.** *Consider the normalization* $Z_{z(t),\dot z(t)}$ *for the canonical distribution* $\mu_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}$ *defined in* (2.15). *Denote by* $\{q_t,p_t\}_{0\le t\le T}$ *the solution of the forward Langevin dynamics* (SCL) *with initial conditions distributed according to*

$$(q_0,p_0) \sim \mu_{\Sigma_{\xi,v_\xi}(z(0),\dot z(0))}(dq\,dp), \tag{5.16}$$

and by $\{q^{\rm b}_{t'},p^{\rm b}_{t'}\}_{0\le t'\le T}$ *the solution of the backward Langevin process* (5.6) *with initial conditions distributed according to*

$$(q^{\rm b}_0,p^{\rm b}_0) \sim \mu_{\Sigma_{\xi,v_\xi}(z(T),\dot z(T))}(dq\,dp). \tag{5.17}$$

Then, the following Jarzynski-Crooks identity holds on $[0,T]$*: for any bounded path functional* $\varphi_{[0,T]}$,

$$\frac{Z_{z(T),\dot z(T)}}{Z_{z(0),\dot z(0)}} = \frac{\mathbb{E}\left(\varphi_{[0,T]}\left(\{q_t,p_t\}_{0\le t\le T}\right)\,{\rm e}^{-\beta\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{t\in[0,T]}\right)}\right)}{\mathbb{E}\left(\varphi^{\rm r}_{[0,T]}\left(\{q^{\rm b}_{t'},p^{\rm b}_{t'}\}_{0\le t'\le T}\right)\right)}, \tag{5.18}$$

where $(\,\cdot\,)^{\mathrm{r}}$ *denotes the composition with the operation of time reversal of paths:*

$$\varphi^{\mathrm{r}}_{[0,T]}\Big(\{q^{\mathrm{b}}_{t'},p^{\mathrm{b}}_{t'}\}_{0\le t'\le T}\Big)=\varphi_{[0,T]}\Big(\{q^{\mathrm{b}}_{T-t},p^{\mathrm{b}}_{T-t}\}_{0\le t\le T}\Big). \tag{5.19}$$

Note that the theorem still holds in the Hamiltonian case, *i.e.*, when $(\gamma_P,\sigma_P)=(0,0)$. The choice $\varphi_{[0,T]}=1$ in (5.18) leads to the following work fluctuation identity:

$$F^{\xi,v_\xi}_{\mathrm{rgd}}(z(T),\dot z(T))-F^{\xi,v_\xi}_{\mathrm{rgd}}(z(0),\dot z(0))=-\frac{1}{\beta}\ln\Big[\mathbb{E}\Big(\mathrm{e}^{-\beta\mathcal{W}_{0,T}(\{q_t,p_t\}_{t\in[0,T]})}\Big)\Big]. \tag{5.20}$$

Besides, upon choosing a path functional $\exp(\theta\beta\mathcal{W}_{0,T})$, it is possible to obtain a family of free energy estimators, parameterized by θ and where both forward and backward paths are weighted by the exponential of some work. Moreover, the standard Crooks equality on ratios of probability density functions of work values is also a consequence of (5.18); see Section 4.2.2 in [35].

Note also that the choice $\varphi_{[0,T]}(q,p)=\phi(q_T,p_T)$ leads to the following representation of the canonical distribution $\mu_{\Sigma_{\xi,v_\xi}(z(T),\dot z(T))}$:

$$\begin{aligned}&\frac{\mathbb{E}\Big(\phi(q_T,p_T)\mathrm{e}^{-\beta\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{t\in[0,T]}\right)}\Big)}{\mathbb{E}\Big(\mathrm{e}^{-\beta\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{t\in[0,T]}\right)}\Big)}\\&\qquad=\int_{\Sigma_{\xi,v_\xi}(z(T),\dot z(T))}\phi(q,p)\,\mu_{\Sigma_{\xi,v_\xi}(z(T),\dot z(T))}(dq\,dp).\end{aligned} \tag{5.21}$$

The usual free energy profile $z\mapsto F(z)$ can therefore be computed using the relations (1.18)-(1.19) presented in the introduction. Indeed, equation (1.19) (see (5.22) below) can be proved by combining (5.2) and (5.20)–(5.21) as follows:

$$\begin{aligned}F(z(T))-F(z(0))&=\Big(F(z(T))-F^{\xi,v_\xi}_{\mathrm{rgd}}(z(T),\dot z(T))\Big)\\&\quad-\Big(F(z(0))-F^{\xi,v_\xi}_{\mathrm{rgd}}(z(0),\dot z(0))\Big)\\&\quad-\frac{1}{\beta}\ln\Big[\mathbb{E}\Big(\mathrm{e}^{-\beta\mathcal{W}_{0,T}(\{q_t,p_t\}_{t\in[0,T]})}\Big)\Big]\\&=-\frac{1}{\beta}\ln\mathbb{E}\Big((\det G_M(q_T))^{-1/2}\mathrm{e}^{\frac{\beta}{2}\dot z(T)^TG_M^{-1}(q_T)\dot z(T)}\,\mathrm{e}^{-\beta\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{t\in[0,T]}\right)}\Big)\\&\quad+\frac{1}{\beta}\ln\mathbb{E}\Big((\det G_M(q_0))^{-1/2}\mathrm{e}^{\frac{\beta}{2}\dot z(0)^TG_M^{-1}(q_0)\dot z(0)}\Big),\end{aligned}$$

so that, with the corrector $C(t,q)=\dfrac{1}{2\beta}\ln\Big(\det G_M(q)\Big)-\dfrac{1}{2}\dot z(t)^TG_M^{-1}(q)\dot z(t)$ defined in (1.18),

$$F(z(T))-F(z(0))=-\frac{1}{\beta}\ln\left(\frac{\mathbb{E}\Big(\mathrm{e}^{-\beta\left[\mathcal{W}_{0,T}\left(\{q_t,p_t\}_{0\le t\le T}\right)+C(T,q_T)\right]}\Big)}{\mathbb{E}\Big(\mathrm{e}^{-\beta C(0,q_0)}\Big)}\right). \tag{5.22}$$

Estimators of the free energy based on (5.22) can then be constructed; see Chapter 4 in [35] for a review.

Before turning to the proof of Theorem 5.3, we first give the general lemma which enables to deduce the Jarzynski-Crooks fluctuation identity from a *nonequilibrium detailed balance condition* (similar to the one presented in [5] for switchings arising from a time-dependence in the Hamiltonian).

Lemma 5.4. *Let $(q_t,p_t)_{0\le t\le T}$ (resp. $(q_t^{\rm b},p_t^{\rm b})_{0\le t\le T}$) be a Markov process with infinitesimal generator $\mathcal{L}_t^{\rm f}$ (resp. $\mathcal{L}_t^{\rm b}$) and initial conditions distributed according to* (5.16) (*resp.* (5.17)). *Let us assume that the following nonequilibrium detailed balance condition is satisfied: for any two smooth test functions φ_1, φ_2,*

$$\begin{aligned}&\int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\left(\varphi_1\mathcal{L}_t^{\rm f}(\varphi_2)-\varphi_2\mathcal{L}_{T-t}^{\rm b}(\varphi_1)\right)\mathrm{e}^{-\beta H}\,d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\\&\quad=\int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\beta w(t,\cdot)\varphi_1\varphi_2\mathrm{e}^{-\beta H}\,d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\\&\qquad+\frac{d}{dt}\left(\int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\varphi_1\varphi_2\mathrm{e}^{-\beta H}\,d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\right).\end{aligned}\tag{5.23}$$

Then the Jarzynski-Crooks fluctuation identity (5.18) *holds.*

Proof. We use in this proof the shorthand notation Z_t, π_t and $\mathcal{S}_t$ for the partition function $Z_{z(t),\dot z(t)}$, the (unnormalized) distribution $Z_{z(t),\dot z(t)}\mu_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}=\mathrm{e}^{-\beta H}\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}$ and the submanifold $\Sigma_{\xi,v_\xi}(z(t),\dot z(t))$, respectively.

Let us introduce the following weighted transition operators: for any bounded test function φ,

$$P_{t,T}^{\rm f}(\varphi)(q,p)=\mathbb{E}\Big(\varphi(q_T,p_T)\,\mathrm{e}^{-\beta\mathcal{W}_{t,T}\left(\{q_s,p_s\}_{s\in[t,T]}\right)}\,\Big|\,(q_t,p_t)=(q,p)\Big),\tag{5.24}$$

$$P_{t',T}^{\rm b}(\varphi)(q,p)=\mathbb{E}\Big(\varphi(q_T^{\rm b},p_T^{\rm b})\,\Big|\,(q_{t'}^{\rm b},p_{t'}^{\rm b})=(q,p)\Big),\tag{5.25}$$

where $(q_t,p_t)_{0\le t\le T}$ (resp. $(q_t^{\rm b},p_t^{\rm b})_{0\le t\le T}$) is a Markov process with infinitesimal generator $\mathcal{L}_t^{\rm f}$ (resp. $\mathcal{L}_t^{\rm b}$), and $\mathcal{W}_{t,T}=\mathcal{W}_{0,T}-\mathcal{W}_{0,t}$.

We assume that these operators are well defined and smooth with respect to time for sufficiently smooth test functions defined in an open neighborhood of $\Sigma_{\xi,v_\xi}(z(t),\dot z(t))$ and $\Sigma_{\xi,v_\xi}(z(t'),\dot z(t'))$, respectively (for any $t,t'\in[0,T]$).

The transition operators satisfy the following backward Kolmogorov evolution equations:

$$\begin{cases}\partial_t P_{t,T}^{\rm f}=-\mathcal{L}_t^{\rm f}P_{t,T}^{\rm f}+\beta\,w(t,\cdot)\,P_{t,T}^{\rm f},\\ P_{T,T}^{\rm f}=\mathrm{Id},\end{cases}\qquad\begin{cases}\partial_{t'}P_{t',T}^{\rm b}=-\mathcal{L}_{t'}^{\rm b}\,P_{t',T}^{\rm b},\\ P_{T,T}^{\rm b}=\mathrm{Id}.\end{cases}$$

Consider now two test functions φ_0 and φ_T. The balance condition (5.23) implies

$$\frac{d}{dt}\left(\int_{\mathcal{S}_t}P_{t,T}^{\rm f}(\varphi_T)\,P_{T-t,T}^{\rm b}(\varphi_0)\,d\pi_t\right)=0.$$

Integrating this equality on $[0,T]$ yields

$$\int_{\mathcal{S}_0}P_{0,T}^{\rm f}(\varphi_T)\,\varphi_0\,d\pi_0=\int_{\mathcal{S}_T}\varphi_T\,P_{0,T}^{\rm b}(\varphi_0)\,d\pi_T,\tag{5.26}$$

which is the Crooks identity (5.18) for path functionals of the form

$$\varphi_{[0,T]}(q,p)=\varphi_0(q_0,p_0)\,\varphi_T(q_T,p_T).$$

Indeed,

$$\int_{\mathcal{S}_0}P_{0,T}^{\rm f}(\varphi_T)\,\varphi_0\,d\pi_0=Z_0\,\mathbb{E}\Big[\varphi_T(q_T,p_T)\,\varphi_0(q_0,p_0)\,\mathrm{e}^{-\beta\mathcal{W}_{0,T}(\{q_s,p_s\}_{s\in[0,T]})}\Big],$$

while

$$\int_{\mathcal{S}_T} \varphi_T \, P^{\rm b}_{0,T}(\varphi_0)\, d\pi_T = Z_T \, \mathbb{E}\Big[\varphi_T(q_0^{\rm b}, p_0^{\rm b})\, \varphi_0(q_T^{\rm b}, p_T^{\rm b})\Big].$$

Then, using the Markov property of the forward and backward processes, Crooks identity (5.18) can be extended to finite-dimensional path functionals of the form

$$\varphi_{[0,T]}(q,p) = \varphi_0(q_0,p_0)\dots\varphi_k(q_{t_k},p_{t_k})\dots\varphi_K(q_T,p_T) \tag{5.27}$$

with $0 = t_0 < t_1 < \cdots < t_K = T$ by repeatedly using a variant of (5.26) on time subintervals $[t_k, t_{k+1}]$ (see the proof of Theorem 4.10 in [35] for further precisions). This allows us to conclude since finite dimensional time marginal laws characterize the distribution on continuous paths; see for instance [19]. □

We are now in position to write:

Proof of Theorem 5.3. By Lemma 5.4, it is sufficient to prove the nonequilibrium detailed balance (5.23) for the Markov processes $(q_t,p_t)_{0\le t\le T}$ and $(q_t^{\rm b},p_t^{\rm b})_{0\le t\le T}$, solutions to (SCL) and (5.6), respectively, with generators $\mathcal{L}_t^{\rm f}$ and $\mathcal{L}_{t'}^{\rm b}$ defined by (5.7) and (5.8).

First, using Lemma 2.3, we compute the variation of the unnormalized canonical equilibrium distribution with constraints with respect to the switching:

$$\begin{aligned}
&\frac{d}{dt}\left(\int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))} \varphi_1\varphi_2\, {\rm e}^{-\beta H}\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\right)\\
&\qquad = \int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))} \dot\zeta(t)^T\Gamma^{-1}\left\{\Xi, \varphi_1\varphi_2\, {\rm e}^{-\beta H}\right\}\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}.
\end{aligned} \tag{5.28}$$

On the other hand, (5.13) and Proposition 5.1 give

$$\begin{aligned}
&\varphi_1\mathcal{L}_t^{\rm f}(\varphi_2) - \varphi_2\mathcal{L}_{T-t}^{\rm b}(\varphi_1) - \beta w(t,\cdot)\varphi_1\varphi_2\\
&\quad = \{\varphi_1\varphi_2, H\}_\Xi + {\rm e}^{\beta H}\left\{\varphi_1\varphi_2 {\rm e}^{-\beta H}, \Xi\right\}\Gamma^{-1}\dot\zeta(t)\\
&\qquad + \varphi_1\frac{1}{\beta}{\rm e}^{\beta H}{\rm div}_p\left({\rm e}^{-\beta H}\gamma_P\nabla_p\varphi_2\right) - \varphi_2\frac{1}{\beta}{\rm e}^{\beta H}{\rm div}_p\left({\rm e}^{-\beta H}\gamma_P\nabla_p\varphi_1\right).
\end{aligned} \tag{5.29}$$

Now, (5.23) can be checked in two steps. First, the last two terms in (5.29) (the "thermostat" terms) cancel out after integration with respect to ${\rm e}^{-\beta H}\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}$ thanks to the detailed balance condition (3.12). Then, an integration of (5.29) with respect to ${\rm e}^{-\beta H}\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}$ gives, in view of (5.28) and (2.25),

$$\begin{aligned}
&\int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))} \Big(\varphi_1\mathcal{L}_t^{\rm f}(\varphi_2) - \varphi_2\mathcal{L}_{T-t}^{\rm b}(\varphi_1) - \beta w(t,\cdot)\varphi_1\varphi_2\Big){\rm e}^{-\beta H}\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\\
&\quad = \int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))} \left\{\varphi_1\varphi_2{\rm e}^{-\beta H}, \Xi\right\}\Gamma^{-1}\dot\zeta(t)\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\\
&\quad = \frac{d}{dt}\left(\int_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))} \varphi_1\varphi_2\, {\rm e}^{-\beta H}\, d\sigma_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}\right),
\end{aligned}$$

which is indeed (5.23). Note that the time-regularity on the evolution semi-groups (5.24)-(5.25) required to make these computations rigorous is proved in the overdamped case in the proof of Theorem A.5 in [34]. A similar proof can be carried out for constrained Langevin equations. □

5.4. **Numerical schemes.** In this section, a numerical scheme for the nonequilibrium dynamics (SCL) and the associated free energy estimator are presented. As for Langevin processes with constraints (Section 3.2), a splitting between the Hamiltonian part and the thermostat part of the dynamics (SCL) leads to a simple and natural scheme (see (5.30)-(5.31)-(5.32) below). Note that a consistent numerical scheme in the case of Hamiltonian dynamics can be obtained by considering only (5.31) (this corresponds to $\gamma = \sigma = 0$). Besides, we propose a discrete Jarzynski-Crooks identity without time discretization error; see Section 5.4.5.

The reaction coordinate path is first discretized as $\{z(0), \dots, z(t_{N_T})\}$ where N_T is the number of time-steps. For simplicity, equal time increments are used, so that $\Delta t = \frac{T}{N_T}$ and $t_n = n\Delta t$. The deterministic Hamiltonian part in the equations of motion (SCL) with switched position constraints $\xi(q) = z(t)$ can be integrated by a velocity-Verlet algorithm with constraints similar to (3.17). The fluctuation-dissipation term in (SCL) can be integrated similarly to the constrained case without switching (3.16)-(3.18), using an Ornstein-Uhlenbeck process on the momentum variable approximated by a midpoint Euler scheme. In conclusion, the splitting scheme for the Langevin dynamics with time-evolving constraints reads as follows: Take initial conditions (q^0, p^0) distributed according to $\mu_{\Sigma_{\xi,v_\xi}\left(z(t_0),\frac{z(t_1)-z(t_0)}{\Delta t}\right)}$ and iterate on $0 \le n \le N_T - 1$:

$$\left\{ p^{n+1/4} = p^n - \frac{\Delta t}{4}\gamma_P(q^n)M^{-1}(p^{n+1/4} + p^n) + \sqrt{\frac{\Delta t}{2}}\sigma_P(q^n)\mathcal{G}^n, \right. \tag{5.30}$$

$$\begin{cases} p^{n+1/2} = p^{n+1/4} - \dfrac{\Delta t}{2}\nabla V(q^n) + \nabla\xi(q^n)\lambda^{n+1/2}, \\ q^{n+1} = q^n + \Delta t\ M^{-1}p^{n+1/2}, \\ \xi(q^{n+1}) = z(t_{n+1}), \qquad (C_q), \\ p^{n+3/4} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\lambda^{n+3/4}, \\ \nabla\xi(q^{n+1})^T M^{-1}p^{n+3/4} = \dfrac{z(t_{n+2}) - z(t_{n+1})}{\Delta t}, \qquad (C_p), \end{cases} \tag{5.31}$$

$$\left\{ \begin{aligned} p^{n+1} = p^{n+3/4} &- \frac{\Delta t}{4}\gamma_P(q^{n+1})M^{-1}(p^{n+3/4} + p^{n+1}) \\ &+ \sqrt{\frac{\Delta t}{2}}\sigma_P(q^{n+1})\mathcal{G}^{n+1/2}, \end{aligned} \right. \tag{5.32}$$

where $(\mathcal{G}^n)$ and $(\mathcal{G}^{n+1/2})$ are sequences of i.i.d. Gaussian random variables of mean 0 and covariance matrix Id_{3N}. Note that the momenta obtained from (5.30)-(5.31)-(5.32) satisfy

$$\begin{aligned} \nabla\xi(q^n)^T M^{-1}p^{n+1/4} = \nabla\xi(q^n)^T M^{-1}p^n = \nabla\xi(q^n)^T M^{-1}p^{n-1/4} \\ = \frac{z(t_{n+1}) - z(t_n)}{\Delta t}, \end{aligned} \tag{5.33}$$

so that constraints on momenta are automatically enforced, and no Lagrange multiplier is needed in (5.30) and (5.32).

We comment in the subsequent sections on the different parts of the scheme.

5.4.1. *Comments on the Hamiltonian scheme* (5.31). The Lagrange multipliers $\lambda^{n+1/2}$ are associated with the position constraints (C_q), and the Lagrange multipliers $\lambda^{n+3/4}$ are associated with the velocity constraints (C_p). In (C_p), the velocity of the switching at time t_{n+1} is discretized as

$$\dot{z}(t_{n+1}) \simeq \frac{z(t_{n+2}) - z(t_{n+1})}{\Delta t}.$$

The latter choice is motivated by the following observation: The position after one step of an unconstrained motion, given by

$$\tilde{q}^{n+1} = q^n + \Delta t\ M^{-1} p^{n+1/4} - \frac{\Delta t^2}{2} M^{-1} \nabla V(q^n),$$

already satisfies (C_q) up to error terms of order two with respect to Δt. Indeed, using (5.33):

$$\xi(\tilde{q}^{n+1}) = \xi(q^n) + \Delta t \nabla \xi(q^n)^T M^{-1} p^{n+1/4} + \mathrm{O}(\Delta t^2) = z(t_{n+1}) + \mathrm{O}(\Delta t^2).$$

This property is useful to ensure a fast convergence of the numerical algorithm solving the nonlinear constraints (C_q).

The numerical flow associated with (5.31) is denoted in the sequel as

(5.34)
$$\Phi^n \ : \ \left\{ \begin{array}{ccc} \Sigma_{\xi,v_\xi}\left(z(t_n), \dfrac{z(t_{n+1}) - z(t_n)}{\Delta t}\right) & \to & \Sigma_{\xi,v_\xi}\left(z(t_{n+1}), \dfrac{z(t_{n+2}) - z(t_{n+1})}{\Delta t}\right), \\ (q^n, p^{n+1/4}) & \mapsto & (q^{n+1}, p^{n+3/4}). \end{array} \right.$$

It can be proven that Φ^n is a symplectic map. The proof is indeed exactly the same as for the symplecticity of the classical RATTLE scheme; see [22, Sections VII.1.3] for an explicit computation for symplectic Euler and [22, Sections VII.1.4] for an extension to RATTLE. As a consequence, Φ^n transports the phase space measure $\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_n), \frac{z(t_{n+1}) - z(t_n)}{\Delta t}\right)}$ to the phase space measure $\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_{n+1}), \frac{z(t_{n+2}) - z(t_{n+1})}{\Delta t}\right)}$.

5.4.2. *Comments on the fluctuation-dissipation part* (5.30)-(5.32). In practice, (5.30) may be rewritten in a form more suited to numerical computations. Of course, similar considerations hold for (5.32). Since $\gamma_P(q) = P_M(q)\gamma P_M(q)^T$ and $\sigma_P(q) = P_M(q)\sigma$, (5.30) is equivalent to:

$$\left\{ \begin{array}{l} p^{n+1/4} = p^n - \dfrac{\Delta t}{4}\gamma M^{-1}\left(p^n + p^{n+1/4} - 2\nabla\xi G_M^{-1}(q^n)\dfrac{z(t_{n+1}) - z(t_n)}{\Delta t}\right) \\ \qquad\qquad + \sqrt{\dfrac{\Delta t}{2}}\,\sigma\,\mathcal{G}^n + \nabla\xi(q^n)^T\,\lambda^{n+1/4}, \\ \nabla\xi(q^n)^T M^{-1} p^{n+1/4} = \dfrac{z(t_{n+1}) - z(t_n)}{\Delta t}, \qquad (C_p) \end{array} \right. \tag{5.35}$$

where the Lagrange multiplier $\lambda^{n+1/4}$ is associated with the constraint (C_p). The equivalence between (5.30) and (5.35) can be checked by multiplying (5.35) by $P_M(q^n)$ and using (5.33).

The Lagrange multiplier $\lambda^{n+1/4}$ in (5.35) is obtained by multiplying the above equation by $\nabla\xi(q^n)^T M^{-1}\left(\mathrm{Id}+\frac{\Delta t}{4}\gamma M^{-1}\right)^{-1}$, and solving the following linear system:

$$\begin{aligned}\frac{z(t_{n+1})-z(t_n)}{\Delta t} &= \nabla\xi(q^n)^T M^{-1}\left(\mathrm{Id}+\frac{\Delta t}{4}\gamma M^{-1}\right)^{-1}\left(\mathrm{Id}-\frac{\Delta t}{4}\gamma M^{-1}\right)p^n\\ &+\nabla\xi(q^n)^T M^{-1}\left(\mathrm{Id}+\frac{\Delta t}{4}\gamma M^{-1}\right)^{-1}\\ &\quad\times\left(\gamma M^{-1}\nabla\xi(q^n)G_M^{-1}(q^n)\frac{z(t_{n+1})-z(t_n)}{2}+\sqrt{\frac{\Delta t}{2}}\,\sigma\,\mathcal{G}^n\right)\\ &+\nabla\xi(q^n)^T M^{-1}\left(\mathrm{Id}+\frac{\Delta t}{4}\gamma M^{-1}\right)^{-1}\nabla\xi(q^n)\,\lambda^{n+1/4}.\end{aligned}$$

This system is well posed. Indeed, the matrix $\nabla\xi(q^n)^T M^{-1}\left(\mathrm{Id}+\frac{\Delta t}{4}\gamma M^{-1}\right)^{-1}\nabla\xi(q^n)$ can be rewritten as $\nabla\xi(q)^T S\nabla\xi(q)$ with $S=M^{-1}\left(\mathrm{Id}+\frac{\Delta t}{4}\gamma M^{-1}\right)^{-1}$. Both M and γ are symmetric and nonnegative, so that S is symmetric, positive and invertible. Finally, the invertibility of $\nabla\xi(q)^T S\nabla\xi(q)$ follows from the invertibility of $G_M(q)$.

In the special case when γ and M are equal up to a multiplicative constant, the numerical integration can be simplified using the explicit formula (3.20) and the method described below (3.20), which still holds for the tangential part of the momentum. See Section 5.6 below for further precisions.

5.4.3. *Discretization of the backward process* (5.6). The splitting scheme for the backward Langevin dynamics with time-evolving constraints (5.6) reads as follows: Denote $n'=N_T-n$, take initial conditions $(q^{\mathrm{b},0},p^{\mathrm{b},0})$ distributed according to $\mu_{\Sigma_\xi,v_\xi}\left(z(t_{N_T}),\frac{z(t_{N_T+1})-z(t_{N_T})}{\Delta t}\right)$ and iterate on $0\le n'\le N_T-1$,

$$\text{(5.36)}\qquad \begin{cases} p^{\mathrm{b},n'+1/4}=p^{\mathrm{b},n'}-\dfrac{\Delta t}{4}\gamma_P(q^{\mathrm{b},n'})M^{-1}(p^{\mathrm{b},n'+1/4}+p^{\mathrm{b},n'})\\ \qquad\qquad +\sqrt{\dfrac{\Delta t}{2}}\sigma_P(q^{\mathrm{b},n'})\mathcal{G}^{\mathrm{b},n'},\end{cases}$$

$$\text{(5.37)}\qquad \begin{cases} p^{\mathrm{b},n'+1/2}=p^{\mathrm{b},n'+1/4}+\dfrac{\Delta t}{2}\nabla V(q^{\mathrm{b},n'})+\nabla\xi(q^{\mathrm{b},n'})\lambda^{\mathrm{b},n'+1/2},\\ q^{\mathrm{b},n'+1}=q^{\mathrm{b},n'}-\Delta t\ M^{-1}p^{\mathrm{b},n'+1/2},\\ \xi(q^{\mathrm{b},n'+1})=z(t_{N_T-n'-1}),\qquad (C_q),\\ p^{\mathrm{b},n'+3/4}=p^{\mathrm{b},n'+1/2}+\dfrac{\Delta t}{2}\nabla V(q^{\mathrm{b},n'+1})+\nabla\xi(q^{\mathrm{b},n'+1})\lambda^{\mathrm{b},n'+3/4},\\ \nabla\xi(q^{\mathrm{b},n'+1})^T M^{-1}p^{\mathrm{b},n'+3/4}=\dfrac{z(t_{N_T-n'})-z(t_{N_T-n'-1})}{\Delta t},\qquad (C_p),\end{cases}$$

$$\text{(5.38)}\qquad \begin{cases} p^{\mathrm{b},n'+1}=p^{\mathrm{b},n'+3/4}-\dfrac{\Delta t}{4}\gamma_P(q^{\mathrm{b},n'+1})M^{-1}(p^{\mathrm{b},n'+3/4}+p^{\mathrm{b},n'+1})\\ \qquad\qquad +\sqrt{\dfrac{\Delta t}{2}}\sigma_P(q^{\mathrm{b},n'+1})\mathcal{G}^{\mathrm{b},n'+1/2},\end{cases}$$

where $(\mathcal{G}^{\mathrm{b},n'})$ and $(\mathcal{G}^{\mathrm{b},n'+1/2})$ are sequences of i.i.d. Gaussian random variables of mean 0 and covariance matrix Id_{3N}. The numerical flow associated with (5.37) is denoted

$$\Phi^{\mathrm{b},n'} : \left\{ \begin{array}{ccc} \Sigma_{\xi,v_\xi}\left(z(t_n), \dfrac{z(t_{n+1})-z(t_n)}{\Delta t}\right) & \to & \Sigma_{\xi,v_\xi}\left(z(t_{n-1}), \dfrac{z(t_n)-z(t_{n-1})}{\Delta t}\right), \\ (q^{\mathrm{b},n'}, p^{\mathrm{b},n'+1/4}) & \mapsto & (q^{\mathrm{b},n'+1}, p^{\mathrm{b},n'+3/4}). \end{array} \right.$$

where we recall $n' = N_T - n$. Assuming that the flow Φ^n given by (5.34) and $\Phi^{\mathrm{b},n'}$ are both well defined, the following reversibility property is easily checked (extending the symmetry property of the standard RATTLE scheme; see for instance [22, Section VII.1.4]):

$$\Phi^{\mathrm{b},N_T-n} \circ \Phi^{n-1} = \mathrm{Id}. \tag{5.39}$$

5.4.4. *Work discretization and free energy computations.* The work (5.11) can be approximated using the Lagrange multipliers in (5.31):

$$\begin{cases} \mathcal{W}^0 = 0, \\ \mathcal{W}^{n+1} = \mathcal{W}^n + \left(\dfrac{z(t_{n+1})-z(t_n)}{\Delta t}\right)^T \left(\lambda^{n+1/2} + \lambda^{n+3/4}\right), \end{cases} \tag{5.40}$$

for $n = 0 \dots N_T - 1$. The (formal) consistency of the work discretization (5.40) in the time continuous limit is a direct consequence of the work expression (5.11).

An estimator of the free energy profile is then obtained by using K independent realizations of the switching process, computing the work $\mathcal{W}^{N_T,k}$ for each realization $k \in \{1, \dots, K\}$ (with the numerical trajectories obtained from the numerical scheme (5.30)-(5.31)-(5.32) and i.i.d. initial conditions sampled according to $\mu_{\Sigma_{\xi,v_\xi}\left(z(t_0), \frac{z(t_1)-z(t_0)}{\Delta t}\right)}$), and approximating (5.22), rewritten up to an unimportant additive constant (independent of T), as

$$F(z(T)) \simeq -\frac{1}{\beta} \ln \mathbb{E}\left(\mathrm{e}^{-\beta\left[\mathcal{W}^{N_T} + C^{N_T}(q^{N_T})\right]}\right),$$

with empirical averages such as

$$-\frac{1}{\beta} \ln\left(\frac{1}{K}\sum_{k=1}^{K} \exp\left[-\beta\left(\mathcal{W}^{N_T,k} + C^{N_T}(q^{N_T,k})\right)\right]\right).$$

In the above, the discretization $C^n(q)$ of the corrector (1.18) is

$$C^n(q) = \frac{1}{2\beta} \ln\left(\det G_M(q)\right) - \frac{1}{2}\left(\frac{z(t_{n+1})-z(t_n)}{\Delta t}\right)^T G_M^{-1}(q)\left(\frac{z(t_{n+1})-z(t_n)}{\Delta t}\right). \tag{5.41}$$

We refer to Chapter 4 in [35] for more background on free energy estimators for nonequilibrium dynamics. In particular, it is possible to compute a work associated with the backward switching from the Lagrange multipliers in (5.6), and to resort to bridge estimators (see Section 4.2.3 in [35]).

However, using approximations such as (5.40) in the Jarzynski-Crooks identity introduces a time discretization error. We show in the next section how to eliminate this error.

5.4.5. *Discrete Jarzynski-Crooks identity.* It turns out that a discrete version of the Jarzynski-Crooks identity (5.18) can be obtained. This enables the estimation of free energy differences using nonequilibrium simulation *without time discretization error*. The discrete equality (5.46) below may be seen as an extension of the corresponding equality obtained for transitions associated with time-dependent Hamiltonians and performed with Metropolis-Hastings dynamics (see [10] and Remark 4.5 in [35]).

For this purpose, we consider a discretization of the work $\mathcal{W}_{0,T}$ using the interpretation (5.15) of the work as the energy variation of the Hamiltonian part of the Langevin dynamics. This leads to the following definition of the work at the discrete level:

$$\begin{cases} \mathcal{W}^0 = 0, \\ \mathcal{W}^{n+1} = \mathcal{W}^n + H(q^{n+1}, p^{n+3/4}) - H(q^n, p^{n+1/4}), \end{cases} \tag{5.42}$$

for $n = 0 \dots N_T - 1$. This work discretization leads to a Jarzynski-Crooks identity without time discretization error.

Theorem 5.5 (Discrete Jarzynski-Crooks fluctuation identity)**.** *Consider the distribution $\mu_{\Sigma_{\xi,v_\xi}(z(t),\dot z(t))}$ and its normalization $Z_{z(t),\dot z(t)}$ defined in* (2.15)*. Denote by $\{q^n, p^n\}_{0\le n\le N_T}$ the solution of the forward discretized Langevin dynamics* (5.30)-(5.31)-(5.32) *with initial conditions distributed according to*

$$(q^0, p^0) \sim \mu_{\Sigma_{\xi,v_\xi}\left(z(t_0), \frac{z(t_1)-z(t_0)}{\Delta t}\right)}(dq\, dp), \tag{5.43}$$

and by $\{q^{\mathrm{b},n'}, p^{\mathrm{b},n'}\}_{0\le n'\le N_T}$ the solution of the discretized backward Langevin dynamics (5.36)-(5.37)-(5.38) *distributed according to*

$$(q^{\mathrm{b},0}, p^{\mathrm{b},0}) \sim \mu_{\Sigma_{\xi,v_\xi}\left(z(t_{N_T}), \frac{z(t_{N_T+1})-z(t_{N_T})}{\Delta t}\right)}(dq\, dp). \tag{5.44}$$

Then, the following Jarzynski-Crooks identity holds on $[0, N_T]$: for any bounded discrete path functional $\varphi_{[0,N_T]}$,

$$\frac{Z_{z(N_T), \frac{z(t_{N_T+1})-z(t_{N_T})}{\Delta t}}}{Z_{z(t_0), \frac{z(t_1)-z(t_0)}{\Delta t}}} = \frac{\mathbb{E}\left(\varphi_{[0,N_T]}\left(\{q^n, p^n\}_{0\le n\le N_T}\right) \mathrm{e}^{-\beta \mathcal{W}^{N_T}}\right)}{\mathbb{E}\left(\varphi^{\mathrm{r}}_{[0,N_T]}\left(\{q^{\mathrm{b},n'}, p^{\mathrm{b},n'}\}_{0\le n'\le N_T}\right)\right)}, \tag{5.45}$$

where $\mathcal{W}^n$ is computed according to (5.42)*, and $(\,\cdot\,)^{\mathrm{r}}$ denotes the composition with the operation of time reversal of paths:*

$$\varphi^{\mathrm{r}}_{[0,N_T]}\left(\{q^{\mathrm{b},n'}, p^{\mathrm{b},n'}\}_{0\le n'\le N_T}\right) = \varphi_{[0,N_T]}\left(\{q^{\mathrm{b},N_T-n}, p^{\mathrm{b},N_T-n}\}_{0\le n\le N_T}\right). \tag{5.46}$$

The (formal) consistency of the work discretization (5.42) in the time continuous limit is a direct consequence of the work expression (5.15). Free energy estimators based on the identity (5.45) are obtained as described in Section 5.4.4. Let us emphasize once again that there is no error related to the finiteness of the time-step Δt in this estimator, and that the only source of approximation is due to the statistical error.

Proof. With a slight abuse of notation, we denote in the same way the random variables (q^n, p^n), $(q^{\mathrm{b},n}, p^{\mathrm{b},n})$, etc. in (5.30)-(5.31)-(5.32) or (5.36)-(5.37)-(5.38), and the integration variables in the definition of probability distributions. We divide the proof into three steps.

Step 1. The phase space conservation of Φ^n and $\Phi^{\mathrm{b},n'}$ and the reversibility property (5.39) imply

$$\begin{aligned}
&\delta_{\Phi^n(q^n,p^{n+1/4})}(dq^{n+1}\,dp^{n+3/4})\,\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_n),\frac{z(t_{n+1})-z(t_n)}{\Delta t}\right)}(dq^n\,dp^{n+1/4})\\
&\quad=\delta_{\Phi^{\mathrm{b},N_T-n-1}(q^{n+1},p^{n+3/4})}(dq^n\,dp^{n+1/4})\\
&\qquad\times\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_{n+1}),\frac{z(t_{n+2})-z(t_{n+1})}{\Delta t}\right)}(dq^{n+1}\,dp^{n+3/4}).
\end{aligned}\tag{5.47}$$

Step 2. The probability distribution of $p^{n+1/4}$ given (q^n,p^n) in the discretization of the fluctuation-dissipation part (5.30) is denoted $K^{\mathrm{OU}}(q^n,p^n,dp^{n+1/4})$. The scheme (5.30) is a midpoint discretization of an Ornstein-Uhlenbeck process, which can be rewritten by decomposing the orthogonal and tangential updates of the momentum:

$$\begin{cases}
p_\parallel^{n+1/4}=p_\parallel^n-\dfrac{\Delta t}{4}\gamma_P(q^n)M^{-1}\left(p_\parallel^{n+1/4}+p_\parallel^n\right)+\sqrt{\dfrac{\Delta t}{2}}\sigma_P(q^n)\mathcal{G}^n,\\
p_\perp^{n+1/4}=p_\perp^n,
\end{cases}\tag{5.48}$$

where $p_\parallel=P_M(q^n)p$, and $p_\perp=(\mathrm{Id}-P_M(q^n))p$. The Markov chain induced by the parallel part of the momentum is the same as the one induced by the scheme (3.16) (or (3.18)) defined in Section 3.2. The latter satisfies a detailed balance equation (both in the plain sense and up to momentum reversal) with respect to the stationary measure $\kappa^{M^{-1}}_{T^*_q\Sigma(z)}(dp)$ defined by (3.19) (see Sections 2.3.2 and 3.3.5 in [35]). We recall that this measure is defined as the kinetic probability distribution in the momentum variable of the canonical distribution $\mu_{T^*\Sigma(z)}(dq\,dp)$ on the tangential space, conditioned by a given $q\in\Sigma(z)$. Adding the (invariant) orthogonal part of the momentum, the following detailed balance condition is satisfied:

$$\begin{aligned}
&\exp\left(-\frac{\beta}{2}(p^n)^TM^{-1}p^n\right)K^{\mathrm{OU}}(q^n,p^n,dp^{n+1/4})\,\sigma^{M^{-1}}_{\Sigma_{v_\xi(q^n,\cdot)}\left(\frac{z(t_{n+1})-z(t_n)}{\Delta t}\right)}(dp^n)\\
&\quad=\exp\left(-\frac{\beta}{2}(p^{n+1/4})^TM^{-1}p^{n+1/4}\right)K^{\mathrm{OU}}(q^n,p^{n+1/4},dp^n)\\
&\qquad\times\sigma^{M^{-1}}_{\Sigma_{v_\xi(q^n,\cdot)}\left(\frac{z(t_{n+1})-z(t_n)}{\Delta t}\right)}(dp^{n+1/4}).
\end{aligned}\tag{5.49}$$

Step 3. Denote by $K^{\mathrm{f}}(q^n,p^n;dq^{n+1},dp^{n+1/4},dp^{n+3/4},dp^{n+1})$ the probability distribution of the variables $(q^{n+1},p^{n+1/4},p^{n+3/4},p^{n+1})$ given the variables (q^n,p^n) in the scheme (5.30)-(5.31)-(5.32); and by $K^{\mathrm{b}}(q^{\mathrm{b},n'},p^{\mathrm{b},n'};dq^{\mathrm{b},n'+1},dp^{\mathrm{b},n'+1/4},dp^{\mathrm{b},n'+3/4},dp^{\mathrm{b},n'+1})$ the probability distribution of the variables $(q^{\mathrm{b},n'+1},p^{\mathrm{b},n'+1/4},p^{\mathrm{b},n'+3/4},p^{\mathrm{b},n'+1})$ given the variables $(q^{\mathrm{b},n'},p^{\mathrm{b},n'})$ in the scheme (5.36)-(5.37)-(5.38). The splitting structure yields:

$$\begin{aligned}
&K^{\mathrm{f},n}(q^n,p^n;dq^{n+1}\,dp^{n+1/4}\,dp^{n+3/4}\,dp^{n+1})\\
&\quad=K^{\mathrm{OU}}(q^{n+1},p^{n+3/4},dp^{n+1})\delta_{\Phi^n(q^n,p^{n+1/4})}(dq^{n+1}\,dp^{n+3/4})K^{\mathrm{OU}}(q^n,p^n,dp^{n+1/4}),
\end{aligned}$$

as well as

$$\begin{aligned}
&K^{\mathrm{b},n'}(q^{\mathrm{b},n'},p^{\mathrm{b},n'};dq^{\mathrm{b},n'+1}\,dp^{\mathrm{b},n'+1/4}\,dp^{\mathrm{b},n'+3/4}\,dp^{\mathrm{b},n'+1})\\
&\quad=K^{\mathrm{OU}}(q^{\mathrm{b},n'+1},p^{\mathrm{b},n'+3/4},dp^{\mathrm{b},n'+1})\delta_{\Phi^{\mathrm{b},n'}(q^{\mathrm{b},n'},p^{\mathrm{b},n'+1/4})}(dq^{\mathrm{b},n'+1}\,dp^{\mathrm{b},n'+3/4})\\
&\qquad\times K^{\mathrm{OU}}(q^{\mathrm{b},n'},p^{\mathrm{b},n'},dp^{\mathrm{b},n'+1/4}).
\end{aligned}$$

Combining the detailed balance conditions (5.47) and (5.49) of Steps 1 and 2, and using the decomposition (2.24) of phase space measures, it follows that

$$\begin{aligned}
&\mathrm{e}^{-\beta\left(H(q^{n+1},p^{n+3/4})-H(q^n,p^{n+1/4})\right)}K^{\mathrm{f},n}(q^n,p^n;dq^{n+1}\,dp^{n+1/4}\,dp^{n+3/4}\,dp^{n+1})\\
&\quad\times\mathrm{e}^{-\beta H(q^n,p^n)}\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_n),\frac{z(t_{n+1})-z(t_n)}{\Delta t}\right)}(dq^n\,dp^n)\\
&=K^{\mathrm{b},N_T-n-1}(q^{n+1},p^{n+1};dq^n\,dp^{n+3/4}\,dp^{n+1/4}\,dp^n)\\
&\quad\times\mathrm{e}^{-\beta H(q^{n+1},p^{n+1})}\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_{n+1}),\frac{z(t_{n+2})-z(t_{n+1})}{\Delta t}\right)}(dq^{n+1}\,dp^{n+1}),
\end{aligned}$$

which can be seen as the Jarzynski-Crooks identity over one time-step. Iterating the argument, it is easy to obtain:

$$\begin{aligned}
&\mathrm{e}^{-\beta\mathcal{W}^{N_T}}K^{\mathrm{f},0}(q^0,p^0;dq^1\,dp^{1/4}\,dp^{3/4}\,dp^1)\ldots\\
&\qquad\qquad K^{\mathrm{f},N_T-1}(q^{N_T-1},p^{N_T-1};dq^{N_T}\,dp^{N_T-3/4}\,dp^{N_T-1/4}\,dp^{N_T})\\
&\quad\times\mathrm{e}^{-\beta H(q^0,p^0)}\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_0),\frac{z(t_1)-z(t_0)}{\Delta t}\right)}(dq^0\,dp^0)\\
&=K^{\mathrm{b},N_T-1}(q^1,p^1;dq^0\,dp^{3/4}\,dp^{1/4}\,dp^0)\ldots\\
&\qquad\qquad K^{\mathrm{b},0}(q^{N_T},p^{N_T};dq^{N_T-1}\,dp^{N_T-1/4}\,dp^{N_T-3/4}\,dp^{N_T-1})\\
&\quad\times\mathrm{e}^{-\beta H(q^{N_T},p^{N_T})}\sigma_{\Sigma_{\xi,v_\xi}\left(z(t_{N_T}),\frac{z(t_{N_T+1})-z(t_{N_T})}{\Delta t}\right)}(dq^{N_T}\,dp^{N_T}),
\end{aligned}$$

which yields (5.45). □

5.5. **The overdamped limit.** The splitting scheme (5.30)-(5.31)-(5.32) can be used in the overdamped regime, using the method of Proposition 3.6, *i.e.*, by choosing

$$\frac{\Delta t}{4}\gamma = M = \frac{\Delta t}{2}\mathrm{Id}, \tag{5.50}$$

which implies $\gamma_P = 2P_M^T P_M$ and $\sigma_P = \frac{2}{\sqrt{\beta}}P_M$. For this choice of parameters, the continuous limit of the numerical scheme is the following variant of the stochastic differential equation (3.23):

$$\begin{cases} dq_t = -\nabla V(q_t)\,dt + \sqrt{\dfrac{2}{\beta}}\,dW_t + \nabla\xi(q_t)\,d\lambda_t^{\mathrm{od}}, \\ \xi(q_t) = z(t), \end{cases} \tag{5.51}$$

where λ_t^{od} is an adapted stochastic process such that $\xi(q_t) = z(t)$. We then obtain the following Jarzynski-Crooks relation for discretized overdamped dynamics, *without time discretization error.*

Proposition 5.6. *Suppose that the relation* (5.50) *is satisfied. With a slight abuse of notation, the mass matrix and the friction matrix are rewritten as* $M\,\mathrm{Id}$ *and* $\gamma\,\mathrm{Id}$ *with* $M,\gamma\in\mathbb{R}$. *Then the splitting scheme* (5.30)-(5.31)-(5.32) *yields the following Euler discretization of the overdamped Langevin constrained dynamics* (5.51)*:*

$$\begin{cases} q^{n+1} = q^n - \Delta t\nabla V(q^n) + \sqrt{\dfrac{2\Delta t}{\beta}}\,\mathcal{G}^n + \nabla\xi(q^n)\,\lambda_{\mathrm{od}}^{n+1}, \\ \xi(q^{n+1}) = z(t_{n+1}), \end{cases} \tag{5.52}$$

where $(\mathcal{G}^n)_{n\geq 0}$ *are independent and identically distributed centered and normalized Gaussian variables, and* $(\lambda_{\rm od}^n)_{n\geq 1}$ *are the Lagrange multipliers associated with the constraints* $(\xi(q^n) = z(t_n))_{0\leq n\leq N_T}$. *In the same way, the backward process* (5.36)-(5.37)-(5.38) *yields the following Euler scheme:*

$$\begin{cases} q^{{\rm b},n'+1} = q^{{\rm b},n'} - \Delta t \nabla V(q^{{\rm b},n'}) + \sqrt{\dfrac{2\Delta t}{\beta}}\, \mathcal{G}^{{\rm b},n'} + \nabla\xi(q^{{\rm b},n'})\, \lambda_{\rm od}^{{\rm b},n'+1}, \\ \xi(q^{{\rm b},n'+1}) = z(t_{N_T-n'-1}). \end{cases} \tag{5.53}$$

Consider the work update

$$\begin{cases} \mathcal{W}^0 = 0, \\ \mathcal{W}^{n+1} = \mathcal{W}^n + V(q^{n+1}) - V(q^n) + \dfrac{1}{\Delta t}\left(\left|p^{n+3/4}\right|^2 - \left|p^{n+1/4}\right|^2 \right), \end{cases} \tag{5.54}$$

for $n = 0 \dots N_T - 1$, *where*

$$\begin{cases} 2p^{n+1/4} = \sqrt{\dfrac{2\Delta t}{\beta}}\, P(q^n)\mathcal{G}^n + \nabla\xi(q^n) G^{-1}(q^n)\left(z(t_{n+1}) - z(t_n)\right), \\ 2\lambda^{n+1/2} = \lambda_{\rm od}^{n+1} - G^{-1}(q^n)\left(z(t_{n+1}) - z(t_n)\right) + \sqrt{\dfrac{2\Delta t}{\beta}} G^{-1}(q^n)\nabla\xi(q^n)^T \mathcal{G}^n, \end{cases}$$

with $G = \nabla\xi^T\nabla\xi$, *and the scheme* (5.52) *is rewritten as*

$$\begin{cases} p^{n+1/2} = p^{n+1/4} - \dfrac{\Delta t}{2}\nabla V(q^n) + \nabla\xi(q^n)\lambda^{n+1/2}, \\ q^{n+1} = q^n + 2p^{n+1/2}, \\ \xi(q^{n+1}) = z(t_{n+1}), \quad (C_q), \\ p^{n+3/4} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\lambda^{n+3/4}, \\ \nabla\xi(q^{n+1})^T p^{n+3/4} = \dfrac{z(t_{n+2}) - z(t_{n+1})}{2}, \quad (C_p). \end{cases} \tag{5.55}$$

Then the Jarzynski-Crooks relation (5.45) *holds under the assumptions* (5.43) *and* (5.44) *on the initial conditions of the schemes* (5.52) *and* (5.53), *respectively.*

The proof is a direct consequence of the reformulation of (5.52) into (5.55), and a direct application of Theorem 5.5 with the parameters (5.50).

Note that the free energy estimator

$$F(z(T)) = -\frac{1}{\beta}\ln \mathbb{E}\left({\rm e}^{-\beta[\mathcal{W}^{N_T} + C^{N_T}(q^{N_T})]} \right), \tag{5.56}$$

based on the work (5.54) and the corrector

$$C^n(q) = \frac{1}{2\beta}\ln\Big(\det G(q)\Big) - \frac{\Delta t}{4}\left(\frac{z(t_{n+1}) - z(t_n)}{\Delta t}\right)^T G^{-1}(q)\left(\frac{z(t_{n+1}) - z(t_n)}{\Delta t}\right) \tag{5.57}$$

is exact (there is no time discretization error). This free energy estimator can be seen as a variant of the estimator proposed in [34], which was derived directly for the scheme (5.52), and reads (up to an unimportant additive constant):

$$F(z(T)) \simeq -\frac{1}{\beta}\ln \mathbb{E}\left({\rm e}^{-\beta[\widetilde{\mathcal{W}}^{N_T} + \widetilde{C}(q^{N_T})]} \right), \tag{5.58}$$

where the work is defined as

$$\begin{cases} \widetilde{\mathcal{W}}^0 = 0, \\ \widetilde{\mathcal{W}}^{n+1} - \widetilde{\mathcal{W}}^n = \left(\dfrac{z(t_{n+1}) - z(t_n)}{\Delta t} \right)^T \tilde{\lambda}_{\rm od}^{n+1}, \end{cases} \tag{5.59}$$

with

$$\tilde{\lambda}_{\rm od}^{n+1} = 2\lambda^{n+1/2} = \lambda_{\rm od}^{n+1} - G^{-1}(q^n)(z(t_{n+1}) - z(t_n)) + \sqrt{\frac{2\Delta t}{\beta}} G^{-1}(q^n) \nabla \xi(q^n)^T \mathcal{G}^n,$$

and the modified corrector is defined without the kinetic energy term:

$$\widetilde{C}(q) = \frac{1}{2\beta} \ln \Big(\det G(q) \Big). \tag{5.60}$$

There is a bias due to the time discretization error in the estimator (5.58), which can be removed upon following the procedure described in Proposition 5.6.

It can be checked that the three free energy estimators introduced above, namely (5.58)-(5.59) (based on the direct discretization of the overdamped dynamics proposed in [34]), (5.56)-(5.40) (which uses the Lagrange multipliers to approximate the work) and (5.56)-(5.54) (based on the discrete Jarzynski equality) are consistent in the limit $\Delta t \to 0$; see [36, Section 5.5.2].

5.6. **Numerical illustration.** We present some free energy profiles obtained with nonequilibrium switching dynamics for the model system and the parameters described in Section 4.4. The switching schedule reads

$$z(t) = z_{\min} + (z_{\max} - z_{\min}) \frac{t}{T}$$

with $z_{\min} = -0.1$ and $z_{\max} = 1.1$. The time-step is $\Delta t = 0.01$. The initial conditions are obtained by first subsampling a constrained dynamics with $\xi(q) = z_{\min}$ and $v_\xi(q,p) = 0$, with a time spacing $T_{\rm sample} = 1$; and then adding the required component $\nabla \xi(q) G_M^{-1} \dot{z}(0)$ to the momentum variable (with $\dot{z}(0) = (z_{\max} - z_{\min})/T$).

In the specific case at hand, the corrector term (1.18) is constant, and free energies differences are equal to differences of rigid free energies. The dynamics used to integrate the nonequilibrium dynamics is based on a splitting strategy, analogous to (5.30)-(5.31)-(5.32), except that the midpoint integration of the Ornstein-Uhlenbeck part is replaced by an exact integration for the unconstrained dynamics, followed by a projection. This can be done here since we choose a friction matrix of the form $\gamma \, \mathrm{Id}$ (recall also that $M = \mathrm{Id}$). More precisely, the corresponding scheme is obtained by replacing (5.30) (and likewise for (5.32)) with

$$\widetilde{p}^{n+1/4} = \alpha p^n + \sqrt{\frac{1-\alpha^2}{\beta}} \, \mathcal{G}^n,$$

where $\alpha = \mathrm{e}^{-\gamma \Delta t}$, and setting $p^{n+1/4} = \widetilde{p}^{n+1/4} + \lambda^{n+1/4} \nabla \xi(q^n)$ with $\lambda^{n+1/4}$ chosen such that

$$\nabla \xi(q^n)^T M^{-1} p^{n+1/4} = \frac{z(t_{n+1}) - z(t_n)}{\Delta t}.$$

Figure 3 presents estimates obtained with M independent realizations of the switching dynamics for different switching times T, using the estimator presented in Section 5.4.4 with the work discretization (5.40). In all cases, the product MT is kept constant. The free energy profile becomes closer to the reference curve as T is

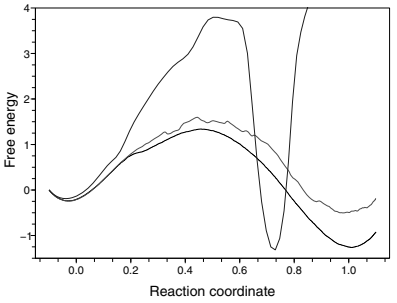

FIGURE 3. Free energy profiles. The top curve corresponds to $T = 1$ with $M = 10^5$, while the two other curves were obtained for $T = 10$ with $M = 10^4$ and $T = 100$ with $M = 10^3$ (smoothest curve).

increased, and the profile obtained for $T = 100$ is in excellent agreement with the result obtained with thermodynamic integration. When the switching time is small, more realizations should be considered to reduce the statistical errors and obtain estimates in better agreement with the reference profile. The fact that the variance is very large when the switching time T is too small is a well-known drawback of estimators based on the Jarzynski-Crooks identity; see the review in Sections 4.1.4 and 4.1.5 in [35]. Roughly speaking, the difficulty is related to the fact that the free energy difference is obtained as an average of $\exp(-\beta\mathcal{W})$, which requires a very good sampling of the small values of the work $\mathcal{W}$. As T decreases, the width of the work distribution increases and the low tail part is more and more difficult to sample. Improved estimates can be obtained with estimators based on combinations of forward and backward trajectories; see for instance [41] and Section 4.2 in [35].

REFERENCES

1. E. Akhmatskaya and S. Reich, *GSHMC: an efficient method for molecular simulation*, J. Comput. Phys. **227** (2008), no. 10, 4934–4954. MR2414842 (2009d:65006)
2. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Science Publications, 2000. MR1857292 (2003a:49002)
3. V. I. Arnol'd, *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics, vol. 60, Springer-Verlag, 1989. MR997295 (90c:58046)
4. N. Bou-Rabee and H. Owhadi, *Long-run behavior of variational integrators in the stochastic context*, SIAM J. Numer. Anal. **48** (2010), 278–297. MR2608370 (2011c:65014)
5. R. Chetrite and K. Gaweçdzki, *Fluctuation relations for diffusion processes*, Commun. Math. Phys. (2008), no. 282, 469–518. MR2421485 (2009f:82026)
6. C. Chipot and A. Pohorille (eds.), *Free energy calculations*, Springer Series in Chemical Physics, vol. 86, Springer, 2007.
7. N. Chopin, T. Lelièvre, and G. Stoltz, *Free energy methods for bayesian inference: Efficient exploration of univariate gaussian mixture posteriors*, Stat. Comput., arXiv:1003.0428v4, (2011).
8. G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, *Blue Moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics*, Chem. Phys. Chem **6** (2005), no. 9, 1809–1814.
9. G. Ciccotti, T. Lelièvre, and E. Vanden-Eijnden, *Projection of diffusions on submanifolds: Application to mean force computation*, Commun. Pure Appl. Math. **61** (2008), no. 3, 371–408. MR2376846 (2008k:82116)

10. G. E. Crooks, *Nonequilibrium measurements of free energy-differences for microscopically reversible markovian systems*, J. Stat. Phys. **90** (1998), no. 5, 1481–1487. MR1628273 (99e:82056)
11. ———, *Entropy production fluctuation theorem and the nonequilibrium work relation for free-energy differences*, Phys. Rev. E **60** (1999), no. 3, 2721–2726.
12. E. Darve, *Thermodynamic integration using constrained and unconstrained dynamics*, Free Energy Calculations (C. Chipot and A. Pohorille, eds.), Springer, 2007, pp. 119–170.
13. C. Dellago, P. G. Bolhuis, and D. Chandler, *On the calculation of reaction rate constants in the transition path ensemble*, J. Chem. Phys. **110** (1999), no. 14, 6617–6625.
14. W. K. den Otter, *Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates*, J. Chem. Phys. **112** (2000), no. 17, 7283–7292.
15. P. A. M. Dirac, *Generalized Hamiltonian dynamics*, Canadian J. Math. **2** (1950), 129–148. MR0043724 (13:306b)
16. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte-Carlo*, Phys. Lett. B **195** (1987), no. 2, 216–222.
17. M. Duflo, *Random iterative models*, Springer, 1997. MR1485774 (98m:62239)
18. W. E and E. Vanden-Eijnden, *Metastability, conformation dynamics, and transition pathways in complex systems*, Multiscale Modelling and Simulation (S. Attinger and P. Koumoutsakos, eds.), Lect. Notes Comput. Sci. Eng., vol. 39, Springer, Berlin, 2004, pp. 35–68. MR2089952 (2005e:82109)
19. S. N. Ethier and T. G. Kurtz, *Markov processes: Characterization and convergence*, Wiley Series in Probability and Statistics, John Wiley & Sons, 1986. MR838085 (88a:60130)
20. L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, 1992. MR1158660 (93f:28001)
21. M. Fixman, *Simulation of polymer dynamics. I. General theory*, J. Chem. Phys. **69** (1978), 1527–1537.
22. E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration: Structure-preserving algorithms for ordinary differential equations*, Springer Series in Computational Mathematics, vol. 31, Springer-Verlag, 2006. MR2221614 (2006m:65006)
23. C. Hartmann, *An ergodic sampling scheme for constrained Hamiltonian systems with applications to molecular dynamics*, J. Stat. Phys. **130** (2008), no. 4, 687–711. MR2387561 (2008m:37140)
24. C. Hartmann and C. Schütte, *A constrained Hybrid Monte Carlo algorithm and the problem of calculating the free energy in several variables*, Z. Angew. Math. Mech. **85** (2005), no. 10, 700–710. MR2172065 (2006e:82095)
25. ———, *A geometric approach to constrained molecular dynamics and free energy*, Commun. Math. Sci. **3** (2005), no. 1, 1–20. MR2132822 (2005k:70036)
26. ———, *Comment on two distinct notions of free energy*, Physica D **228** (2007), no. 1, 59–63. MR2334508 (2008c:82005)
27. L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math. **119** (1967), 147–171. MR0222474 (36:5526)
28. A. M. Horowitz, *A generalized guided Monte Carlo algorithm*, Phys. Lett. B **268** (1991), 247–252.
29. C. Jarzynski, *Nonequilibrium equality for free energy differences*, Phys. Rev. Lett. **78** (1997), no. 14, 2690–2693.
30. W. Kliemann, *Recurrence and invariant measures for degenerate diffusions*, Ann. Probab. **15** (1987), no. 2, 690–707. MR885138 (88d:58134)
31. J. Latorre, C. Hartmann, and Ch. Schütte, *Free energy computation by controlled Langevin processes*, Procedia Computer Science **1** (2010), 1591–1600.
32. B. J. Leimkuhler and S. Reich, *Simulating Hamiltonian dynamics*, Cambridge Monographs on Applied and Computational Mathematics, vol. 14, Cambridge University Press, 2005.
33. B. J. Leimkuhler and R. D. Skeel, *Symplectic numerical integrators in constrained Hamiltonian systems*, J. Comput. Phys. **112** (1994), no. 1, 117–125. MR1277499 (95h:58053)
34. T. Lelièvre, M. Rousset, and G. Stoltz, *Computation of free energy differences through nonequilibrium stochastic dynamics: The reaction coordinate case*, J. Comput. Phys. **222** (2007), no. 2, 624–643. MR2313418 (2008a:82052)
35. ———, *Free energy computations. A mathematical perspective*, Imperial College Press, 2010. MR2681239

36. T. Lelièvre, M. Rousset, and G. Stoltz, *Langevin dynamics with constraints and computation of free energy differences*, arXiv preprint **1006.4914** (2010).
37. P. B. Mackenzie, *An improved hybrid Monte Carlo method*, Phys. Lett. B **226** (1989), no. 3-4, 369–371.
38. J. Marsden and T. Ratiu, *Introduction to mechanics and symmetry*, Texts in Applied Mathematics, vol. 17, Springer, 2003. MR1723696 (2000i:70002)
39. G. N. Milstein and M. V. Tretyakov, *Quasi-symplectic methods for Langevin-type equations*, IMA J. Numer. Anal. **23** (2003), 593–626. MR2011342 (2004h:37131)
40. ———, *Stochastic numerics for mathematical physics*, Scientific Computation, Springer, 2004. MR2069903 (2005f:60004)
41. D. D. L. Minh and A. B. Adib, *Optimized free energies from bidirectional single-molecule force spectroscopy*, Phys. Rev. Lett. **100** (2008), 180602.
42. S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, *Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality*, J. Chem. Phys. **119** (2003), no. 6, 3559–3566.
43. D. C. Rapaport, *The art of molecular dynamics simulations*, Cambridge University Press, 1995.
44. S. Reich, *Smoothed Langevin dynamics of highly oscillatory systems*, Physica D **138** (2000), 210–224. MR1744627 (2001k:37137)
45. J. Schlitter and M. Klähn, *A new concise expression for the free energy of a reaction coordinate*, J. Chem. Phys. **118** (2003), no. 5, 2057–2060.
46. J. E. Straub, M. Borkovec, and B. J. Berne, *Molecular-dynamics study of an isomerizing diatomic in a Lennard-Jones fluid*, J. Chem. Phys. **89** (1988), no. 8, 4833–4847.
47. E. Vanden-Eijnden and G. Ciccotti, *Second-order integrators for Langevin equations with holonomic constraints*, Chem. Phys. Lett. **429** (2006), no. 1-3, 310–316.

Université Paris Est, CERMICS and INRIA, MICMAC project-team Ecole des Ponts ParisTech, 6 & 8 Av. Pascal, 77455 Marne-la-Vallée, France
E-mail address: lelievre@cermics.enpc.fr

INRIA Lille, Nord Europe, Parc Scientifique de la Haute Borne, 40 avenue Halley, Bt. A Park Plaza, 59650 Villeneuve d'Ascq, France
E-mail address: mathias.rousset@inria.fr

Université Paris Est, CERMICS and INRIA, MICMAC project-team Ecole des Ponts ParisTech, 6 & 8 Av. Pascal, 77455 Marne-la-Vallée, France
E-mail address: stoltz@cermics.enpc.fr