

# Ambient space formulations and statistical mechanics of holonomically constrained Langevin systems

J. Walter<sup>1,a</sup>, C. Hartmann<sup>2,b</sup>, and J.H. Maddocks<sup>1,c</sup>

<sup>1</sup> LCVMM, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

<sup>2</sup> Institut für Mathematik, Freie Universität Berlin, 14195 Berlin, Germany

Received 09 September 2011 / Received in final form 10 October 2011

Published online 13 December 2011

**Abstract.** The most classic approach to the dynamics of an  $n$ -dimensional mechanical system constrained by  $d$  independent holonomic constraints is to pick explicitly a new set of  $(n - d)$  curvilinear coordinates parametrizing the manifold of configurations satisfying the constraints, and to compute the Lagrangian generating the unconstrained dynamics in these  $(n - d)$  configuration coordinates. Starting from this Lagrangian an unconstrained Hamiltonian  $H(q, p)$  on  $2(n - d)$  dimensional phase space can then typically be defined in the standard way via a Legendre transform. Furthermore, if the system is in contact with a heat bath, the associated Langevin and Fokker-Planck equations can be introduced. Provided that an appropriate fluctuation-dissipation condition is satisfied, there will be a canonical equilibrium distribution of the Gibbs form  $\exp(-\beta H)$  with respect to the flat measure  $dqdp$  in these  $2(n - d)$  dimensional curvilinear phase space coordinates. The existence of  $(n - d)$  coordinates satisfying the constraints is often guaranteed locally by an implicit function theorem. Nevertheless in many examples these coordinates cannot be constructed in any tractable form, even locally, so that other approaches are of interest. In ambient space formulations the dynamics are defined in the full original  $n$ -dimensional configuration space, and associated  $2n$ -dimensional phase space, with some version of Lagrange multipliers introduced so that the  $2(n - d)$  dimensional sub-manifold of phase space implied by the holonomic constraints and their time derivative, is invariant under the dynamics. In this article we review ambient space formulations, and explain that for constrained dynamics there is in fact considerable freedom in how a Hamiltonian form of the dynamics can be constructed. We then discuss and contrast the Langevin and Fokker-Planck equations and their equilibrium distributions for the different forms of ambient space dynamics.

<sup>a</sup> e-mail: [jessika.walter@gmx.net](mailto:jessika.walter@gmx.net)

<sup>b</sup> e-mail: [chartman@mi.fu-berlin.de](mailto:chartman@mi.fu-berlin.de)

<sup>c</sup> e-mail: [john.maddocks@epfl.ch](mailto:john.maddocks@epfl.ch)

## 1 Introduction

This article is primarily concerned with the statistical mechanics of finite dimensional Lagrangian dynamical systems that are subject to holonomic constraints, stochastic forcing, and viscous damping. One important class of physical systems of this type comprise coarse grain models of polymers, such as DNA, in a solvent heat bath. Such models involve rigid monomer sub-units, for example, in the case of DNA either rigid base or rigid base-pairs depending upon the level of coarse graining. These rigid sub-units interact with each other in a Lagrangian, conservative way, and with the heat bath in a nonconservative way, which, when the detailed motion of the solvent is not of interest, can be modelled in a Langevin fashion. In other words the time evolution can be described as a system of stochastic differential equations (SDEs) with a Hamiltonian part dependent on the polymer degrees of freedom, and the effects of the solvent bath are modelled by the addition of dissipative and stochastic forcing terms in the momentum balance equations. When the monomers are modelled as point masses, or beads on a string, the degrees of freedom are Cartesian and the mass matrix appearing in the Hamiltonian part is typically constant coefficient. In this case it is well known that under an appropriate fluctuation-dissipation hypothesis the Langevin system has an equilibrium distribution on phase space given by the usual canonical expression  $\exp(-\beta H)$  which also induces a measure on configuration space after the quadratic momentum variables are integrated away. Our interests lie primarily in examples where the mass matrix is not constant, but rather is a configuration dependent, generalized mass matrix, as is the case for polymers that are described in internal coordinates and are modelled as a system of rigid bodies whose configuration includes their orientation. To our knowledge the Langevin theory, in which the stochastic motion of both translational and rotational degrees of freedom can be strongly coupled, has only recently been fully described [32]; cf. also [11, 20, 21].

The motivation for the present article is the further special case in which the Langevin dynamics is subject to holonomic constraints. In the context of polymers these constraints can arise in at least two distinct ways. First, many polymer theories constrain the translational degrees of freedom between adjacent monomers, so as to freeze relative extension or shear. These constraints are similar in nature to bond length or distance constraints in molecular dynamics simulations. Second, in simulations of orientations it can be useful to parametrize the locally three dimensional group of rotations  $SO(3)$  with the four dimensional Euler parameter (or quaternion) description, in which the four parameters describing the orientation of one monomer are constrained to lie on the unit sphere in  $\mathbf{R}^4$ . The question of primary interest in this presentation is what effect a generalized mass matrix has on the constrained equilibrium distributions in both phase and configuration spaces. We answer the question by providing rather explicit expressions for the equilibrium distributions.

In this tutorial review article our objective is to describe the mathematical underpinnings of the statistical mechanics of holonomically constrained Langevin systems. In particular we point out that in the presence of holonomic constraints there is considerable freedom in how a Hamiltonian form of the stochastic dynamics can be introduced. We do not restrict ourselves to the case of stochastic polymer dynamics, but that motivation does set the three features of the class of systems that we do treat, namely, a) holonomically constrained Langevin systems, where b) the use of local coordinates satisfying the constraints is unduly cumbersome, and c) there is a configuration dependent kinetic energy. The first issue is therefore how to obtain an ambient space system of SDEs for which the holonomic constraints are (deterministically) satisfied. The first, comparatively standard, of two such systems of SDE is outlined in Section 2. Section 3 is then devoted to the analysis of the equilibrium distribution associated with this constrained dynamics; in particular the influence of

the momentum constraint on the distribution of the geometrically constrained configurations is analyzed (Section 3.2). In Section 4 we describe a second, much less standard, formulation of a Langevin ambient-space system where a different conjugate variable is introduced in the Legendre transformation to the Hamiltonian form (the impetus-striction formulation). In analogy with Section 3, the existence of a stationary probability distribution for the impetus-striction formulation is briefly discussed in Section 5, and the differences between the two approaches are highlighted. We summarize the results in Section 6.

The article has two appendices. In the first we record divers identities used in the analysis, while the second contains the lengthy, but straightforward, computations that form part of the proof of the invariance of the canonical distribution under the constrained Langevin dynamics.

## 1.1 Elementary Notation

We denote the entries of any matrix  $A \in \mathbf{R}^{n \times m}$  by  $A_{ij}$ , whereas when  $n = m$  the elements of the inverse matrix are labeled with upper indices:  $A_{ij}^{-1} = A^{ij}$ . The columns of the matrix  $A$  are denoted  $A_j$ . We do not explicitly distinguish between row and column vectors, with the meaning set by the context. We set the summation convention over repeated upper and lower indices. For example, for  $A \in \mathbf{R}^{n \times m}$ ,  $x \in \mathbf{R}^m$ ,

$$A^{ij}x_j = \sum_{j=1}^m A^{ij}x_j = \sum_{j=1}^m A_{ij}^{-1}x_j.$$

We use the notation  $x \cdot y = x^T y$  for the inner product between vectors and  $A : B = \text{tr}(A^T B)$  for the inner product between square matrices. We denote the tensor product by  $\otimes$ , i.e., for  $x, y \in \mathbf{R}^n$  we have  $(x \otimes y)_{ij} = x_i y_j \in \mathbf{R}^{N \times N}$ . The gradient of a function  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  is denoted

$$\nabla_x f = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right),$$

while

$$\nabla_x^2 f = \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{1 \leq i, j \leq n}$$

is the corresponding Hessian matrix. For a function  $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$  the nabla-operator denotes the corresponding Jacobian

$$\nabla_x f = \left( \frac{\partial f_i}{\partial x_j} \right)_{1 \leq i \leq m, 1 \leq j \leq n}.$$

When unambiguous, we omit the explicit differentiation parameter:  $\nabla f(x, y) = \nabla_{(x, y)} f$  for a function  $f = f(x, y)$  etc, while differentiation with respect to only the second argument is written  $\nabla_y f(x, y)$ . Similarly, we denote by  $\partial_i f$  the partial derivative  $\partial f / \partial x_i$  of a function  $f = f(x)$ , but use  $\partial_{x_i} f$  for  $f = f(x, y)$  depending on two vector arguments  $x, y$ . We also use the abbreviation

$$\frac{\partial}{\partial x} = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right).$$

## 2 Ambient-space formulation of constrained Langevin dynamics

We first record results for a finite dimensional system evolving on an  $n$ -dimensional configuration manifold  $\mathcal{Q} \subseteq \mathbf{R}^n$ . There is little loss of generality in assuming the case  $\mathcal{Q} = \mathbf{R}^n$ , but the more general case encompasses situations such as rigid body dynamics where some of the coordinates are Euler angles with a bounded range of definition. We further assume that there is a smooth potential energy  $V(q)$ ,  $V: \mathcal{Q} \rightarrow \mathbf{R}$ , where  $q = (q_1, \dots, q_n)$  denotes the configuration variable. The potential  $V$  is assumed to be bounded from below and to satisfy appropriate growth conditions so as to prevent escape to infinity. We further assume that there is a kinetic energy which is a quadratic function of the generalized velocities  $|\dot{q}|_g^2/2$  where  $|x|_g = \sqrt{x^T g x}$  is a Riemannian metric with respect to the  $n \times n$  generalized mass or metric tensor  $g = g^T > 0$ . In general when working in curvilinear coordinates the metric tensor  $g(q)$  will be position dependent. When the standard conjugate momenta  $p = (p_1, \dots, p_n)$  are introduced via  $p := g(q)\dot{q}$  then the total (i.e., kinetic plus potential) energy or Hamiltonian takes the form

$$H: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}, \quad H(q, p) = \frac{1}{2}|p|_{g^{-1}}^2 + V(q).$$

The Markovian model that we will assume for motion in the presence of a thermal bath is the Langevin or Hamiltonian-Langevin equation

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} - \gamma(q)\frac{\partial H}{\partial p} + \sigma(q)\dot{W} \end{aligned} \tag{1}$$

where  $W$  is standard  $n$ -dimensional Brownian motion (i.e.  $\dot{W}$  denotes Gaussian white noise),  $\gamma: \mathcal{Q} \rightarrow \mathbf{R}^{n \times n}$  is a field of symmetric positive definite friction coefficients and  $\sigma: \mathcal{Q} \rightarrow \mathbf{R}^{n \times n}$  is a field of full-rank noise coefficients. We assume throughout that the fluctuation-dissipation relation  $2\gamma(q) = \beta\sigma(q)\sigma(q)^T$  holds for almost all values of  $q \in \mathcal{Q}$ . As a consequence, the Gibbs-Boltzmann distribution

$$\rho(q, p) = \frac{1}{Z} \exp(-\beta H(q, p)), \quad Z = \int_{\mathcal{Q} \times \mathbf{R}^n} \exp(-\beta H(q, p)) dq dp$$

is a stationary distribution of (1). Furthermore, under suitable conditions on the potential  $V$  and the smallest eigenvalues of  $\gamma(\cdot)$  and  $g(\cdot)$ , it is the *unique* stationary distribution.

*Remark 1.* For later purposes it is important to note that  $\rho(q, p)$  is a probability density function with respect to the flat Liouville measure  $dq dp$  on phase space even when the configuration coordinates  $q$  are curvilinear.

*Remark 2.* The Langevin system (1) is a widely assumed model for a system interacting with a solvent bath. However it is rather difficult to derive it from a first principles analysis of the solvent for anything but very simple systems (e.g., see [19] and the references therein). Here we merely take (1) as the starting point of our analysis.

### 2.1 Constraints

We now discuss Langevin dynamics that are constrained to lie on a smooth (strict) submanifold  $\Sigma \subset \mathcal{Q}$  of admissible configurations. To begin, we discuss some general

aspects of holonomic constraints and suppose we can define  $\Sigma$  as the zero level set of a function  $f: \mathcal{Q} \rightarrow \mathbf{R}^d$ , i.e.,

$$\Sigma = \{q \in \mathcal{Q}: f(q) = 0\} \subset \mathcal{Q}.$$

We assume throughout that  $f$  is sufficiently differentiable and that its Jacobian  $\nabla f: \mathbf{R}^n \rightarrow \mathbf{R}^d$  has maximum rank  $d$  for almost all values of  $q \in \Sigma$ , in which case  $\Sigma$  is a smooth submanifold of  $\mathcal{Q}$  of dimension  $n - d$ . (We will say that  $\Sigma$  has *codimension*  $d$  in  $\mathcal{Q}$ .) In case of a scalar constraint the rank condition reduces to the requirement that  $\nabla f$  is not the null vector. We further assume that  $\Sigma$  is connected in order to exclude the situation of possibly having several invariant probability distributions with distinct topological support.

To every point  $q \in \Sigma$ , we assign the tangent space

$$T_q \Sigma = \{v \in \mathbf{R}^n: \nabla f(q) v = 0\}$$

as the set of vectors tangential to  $\Sigma$ . Note that the requirement that the admissible generalized velocities  $\dot{q}$  are in the kernel of  $\nabla f$  simply expresses the fact that

$$\left. \frac{d}{dt} f(c(t)) \right|_{t=0} = \nabla f(q) \dot{c}(0)$$

must be identically zero for any differentiable curve  $c: (-\epsilon, \epsilon) \rightarrow \mathcal{Q}$  going through  $c(0) = q$ . We call the “dual” to the tangent space,

$$T_q^* \Sigma = \{p \in \mathbf{R}^n: \nabla f(q) g^{-1}(q) p = 0\}$$

the cotangent space at  $q \in \Sigma$ , which characterizes the admissible momenta  $p$  for each  $q \in \Sigma$ . In the usual manner we identify both tangent and cotangent spaces with the respective  $(n - d)$ -dimensional linear subspaces of  $\mathbf{R}^n$  and emphasize that for each  $q \in \Sigma$  there is a natural linear transformation  $p = g(q)v$  between the tangent and cotangent subspaces.

Later we argue that imposing constraints on the Langevin system (1) can be usefully interpreted as a projection of the dynamics onto the sets of admissible states. To this end we abbreviate  $A = \nabla f(q)$ ,  $A \in \mathbf{R}^{d \times n}$  and introduce the orthogonal projection onto the tangent space as the linear map

$$P: \mathbf{R}^n \rightarrow T_q \Sigma, \quad Pu = u - A^T(AA^T)^{-1}Au. \quad (2)$$

As can be readily checked,  $P = P^T$  and  $P^2 = P$ , so that  $P$  is the orthogonal projection on to the tangent space  $T_q \Sigma$  with the normal space  $(T_q \Sigma)^\perp := \text{span}\{\nabla f_1, \dots, \nabla f_d\}$  as nullspace. (For any  $c \in \mathbf{R}^d$ ,  $Pm = 0$  for any vector  $m = A^T c \in \text{span}\{\nabla f_1, \dots, \nabla f_d\}$ , i.e.,  $Pm = 0$  for any vector  $m$  normal to  $\Sigma$ .) Similarly, using the isomorphism between  $T_q \Sigma$  and  $T_q^* \Sigma$ , it follows that

$$P^*: \mathbf{R}^n \rightarrow T_q^* \Sigma, \quad P^*w = w - g^{-1}A^T(Ag^{-2}A^T)^{-1}Ag^{-1}w \quad (3)$$

is the orthogonal projection onto the corresponding cotangent space  $T_q^* \Sigma$  (in other words, a projection of the momentum vectors) with the normal space  $(T_q^* \Sigma)^\perp := \text{span}\{g^{-1}\nabla f_1, \dots, g^{-1}\nabla f_d\}$  as nullspace. A third, oblique, projection will be key in our discussion of the constrained equations of motion, namely

$$\mathcal{P}: \mathbf{R}^n \rightarrow \mathbf{R}^n, \quad \mathcal{P}v = v - A^T(Ag^{-1}A^T)^{-1}Ag^{-1}v, \quad (4)$$

which satisfies  $\mathcal{P}^2 = \mathcal{P}$  and  $\mathcal{P}^T g^{-1} = g^{-1} \mathcal{P}$ . The first property implies that  $\mathcal{P}$  is indeed a projection, while the second implies

$$(\mathcal{P}x)^T g^{-1}(y - \mathcal{P}y) = 0,$$

which is to say that  $\mathcal{P}$  and  $I - \mathcal{P}$  are orthogonal with respect to the inner product weighted with the inverse generalized mass matrix

$$\langle x, y \rangle_{g^{-1}} = x^T g^{-1} y.$$

Note that  $Ag^{-1}\mathcal{P}w = 0$  for all  $w \in \mathbf{R}^n$ . Accordingly we may regard  $\mathcal{P}$  as a mapping  $\mathcal{P}: \mathbf{R}^n \rightarrow T_q^*\Sigma$ , with nullspace  $(T_q\Sigma)^\perp$ . Nonetheless  $P^*w \neq \mathcal{P}w$  for arbitrary vectors  $w \in \mathbf{R}^n$ , with equality being attained only if  $w \in T_q^*\Sigma$ .

## 2.2 Constrained dynamics

With the above preliminaries in hand, we now turn to our main subject and generalize the stochastic dynamics (1) to the case including holonomic constraints  $f(q) = 0$ . We will verify that the constrained dynamics can be written in the form

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} - \gamma(q) \frac{\partial H}{\partial p} + \sigma(q) \dot{W} + \Lambda \end{aligned} \quad (5)$$

with  $\Lambda = \Lambda(t)$  being a suitable stochastic process adapted to  $W$ , such that

$$P(q(t))\Lambda(t) = 0 \quad \text{and} \quad f(q(t)) = 0. \quad (6)$$

In complete analogy to the deterministic case (cf. [10, 11, 20, 30, 31] and Section 4 of the present article), the modifying term  $\Lambda$  can be interpreted as a Lagrange multiplier (or constraint reaction force) which enforces that the dynamics respect the constraint. The requirement that  $\Lambda$  be normal to the constraint manifold (i.e.,  $P\Lambda = 0$ ) is known as d'Alembert's principle; it implies that the multiplier is uniquely determined.

In the present stochastic case, the constraint reaction force  $\Lambda(t)$  has both a deterministic and stochastic part, depending on the realization of the Brownian motion  $W(s)$  up to time  $s \leq t$  (which is what is meant by "adapted to  $W$ "). The reaction force can be explicitly computed to satisfy (6) by a method directly analogous to that used in the classic case of a deterministic Hamiltonian system subject to holonomic constraints, but with Itô's formula replacing the chain rule for differentiation. To this end we define the "time-differentiated" constraint

$$h(q, p) = A(q)g(q)^{-1}p, \quad A = \nabla f,$$

which must vanish along the solutions  $(q(t), p(t))$  of the constrained Langevin equation (5) due to the requirement that  $f(q(t)) = 0$  for all  $t > 0$ . We seek a representation of the constraint reaction force that is of the form  $\Lambda(t) = F(q(t), p(t), \dot{W}(t))$ , with a suitable function  $F$  satisfying  $PF = 0$ . Without loss of generality we can decompose the multiplier according to

$$\Lambda(t) = C(t) + D(t)\dot{W}(t)$$

where  $C$  is a deterministic process of bounded variation, whereas  $D\dot{W}$  is generally unbounded. Using Itô's formula for  $h$  yields (see [28, Sec. 4.2])

$$\begin{aligned} \frac{d}{dt}h(q(t), p(t)) &= \frac{\partial}{\partial q^k} \left( g^{ij} \frac{\partial f}{\partial q^i} p_j \right) \dot{q}^k + \frac{\partial f}{\partial q^i} g^{ik} \dot{p}_k \\ &= \frac{\partial}{\partial q^k} \left( g^{ij} \frac{\partial f}{\partial q^i} p_j \right) g^{kl} p_l - \frac{\partial f}{\partial q^i} g^{ik} \left( \frac{\partial H}{\partial q^k} + \gamma_{kl} \frac{\partial H}{\partial p_l} - C_k \right) \\ &\quad + \frac{\partial f}{\partial q^i} g^{ik} (\sigma_{kj} + D_{kj}) \dot{W}^j. \end{aligned}$$

In order that  $dh/dt = 0$  we must have that the deterministic drift and the white noise term vanish independently. Recalling that, by d'Alembert's principle,  $\Lambda$ , and therefore  $C$  and  $D\dot{W}$  are aligned with the normal directions  $\nabla f_1, \dots, \nabla f_d$ , we can use the ansatz  $C = A^T c$  and  $D\dot{W} = A^T d$ , and solve for  $c$  and  $d$ . We easily find

$$\begin{aligned} Kc &= Ag^{-1} \left( \frac{\partial H}{\partial q} + \gamma \frac{\partial H}{\partial p} \right) - S \\ Kd &= -Ag^{-1} \sigma \dot{W} \end{aligned}$$

where  $K = Ag^{-1}A^T$ , and the vector  $S \in \mathbf{R}^d$  is a curvature term that is quadratic in  $p$  and has the entries

$$S_k(p, p) = p^T \nabla (g^{-1} \nabla f_k) g^{-1} p, \quad k = 1, \dots, d.$$

By the assumptions on  $f$  and  $A = \nabla f$ , respectively, the  $d \times d$  matrices  $K(q)$  are invertible for almost all  $q \in \Sigma$ , so that

$$\begin{aligned} C &= A^T K^{-1} \left( Ag^{-1} \left( \frac{\partial H}{\partial q} + \gamma \frac{\partial H}{\partial p} \right) - S \right) \\ D\dot{W} &= -A^T K^{-1} Ag^{-1} \sigma \dot{W}. \end{aligned}$$

Upon noting that  $\mathcal{P}$  in (4) can be written as  $\mathcal{P} = I - A^T K^{-1} Ag^{-1}$ , the last equations can be recast as

$$\begin{aligned} C &= (I - \mathcal{P}) \left( \frac{\partial H}{\partial q} + \gamma \frac{\partial H}{\partial p} \right) - A^T K^{-1} S \\ D &= -(I - \mathcal{P}) \sigma, \end{aligned}$$

which uniquely determine the reactive force  $\Lambda(t) = C(t) + D(t)\dot{W}(t)$  as a function of the phase variables  $(q(t), p(t))$  and the noise  $\dot{W}(t)$  at time  $t > 0$ .

The terms in the constraint force have suggestive interpretations: The two terms that involve the projection  $(I - \mathcal{P})$  are the negative of the driving and random forces in the Langevin equation perpendicular to the constraint surface. The quadratic form  $\mathcal{H}(p, p) = A^T K^{-1} S(p, p)$  is a curvature term known as the *second fundamental form* of the submanifold  $\Sigma \subset \mathcal{Q}$ . Roughly speaking, it generates accelerations according to the curvatures of the constraint surface in the direction of the momenta [14].

By substitution for  $\Lambda = C + D\dot{W}$  in equation (5), we find the following ambient-space formulation of the constrained Langevin system

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}, \\ \dot{p} &= -\mathcal{P} \left( \frac{\partial H}{\partial q} + \gamma \frac{\partial H}{\partial p} - \sigma \dot{W} \right) - \mathcal{H}(p, p). \end{aligned} \tag{7}$$

Properties of the ambient-space formulation and its relation to (5)–(6) are summarized in the following result.

**Proposition 1.** *Let  $h: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}^d$  be defined by  $h(q, p) = A(q)g(q)^{-1}p$ . The Langevin system (7) has the following properties:*

1. *The function  $h(q, p)$  is an integral for (7).*
2. *For any regular value  $c$  of  $f: \mathcal{Q} \rightarrow \mathbf{R}^d$ , the set*

$$\mathcal{A}(c) = \{(q, p) \in \mathcal{Q} \times \mathbf{R}^n : f(q) = c \text{ and } h(q, p) = 0\}$$

*is invariant under the dynamics (7).*

3. *If  $(q(t), p(t))$  is a solution of (7), then it is also a solution of (5)–(6) where the multiplier  $\Lambda = \Lambda(t)$  is given by*

$$\Lambda = (I - \mathcal{P}) \left( \frac{\partial H}{\partial q} + \gamma \frac{\partial H}{\partial p} - \sigma \dot{W} \right) - \Pi(p, p).$$

### 3 Equilibrium distribution

We now discuss some properties of the invariant probability distribution for the dynamics (7). The corresponding phase space is defined as the set

$$\mathcal{A} = \{(q, p) \in \mathcal{Q} \times \mathbf{R}^n : f(q) = 0 \text{ and } \nabla f(q)g(q)^{-1}p = 0\}$$

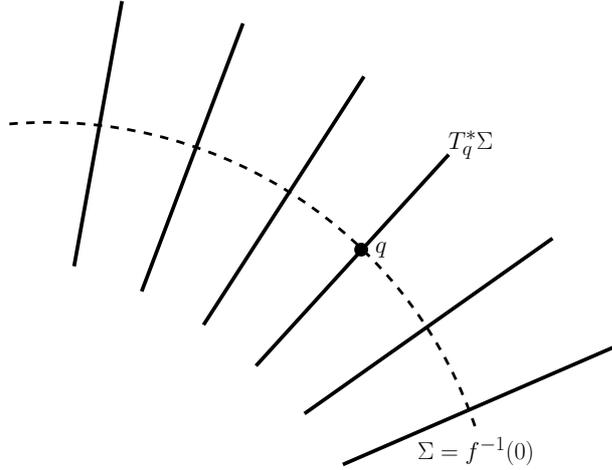
that is isomorphic to the cotangent bundle  $T^*\Sigma = \cup_{q \in \Sigma} T_q^*\Sigma$ . As the energy of the constrained system (7) is the Hamiltonian  $H$  restricted to  $\mathcal{A}$ , a natural candidate for the stationary distribution of the constrained dynamics (7) seems to be the restriction of the unconstrained canonical density

$$\rho(q, p) = \frac{1}{Z} \exp(-\beta H(q, p)), \quad Z = \int_{\mathcal{Q} \times \mathbf{R}^n} \exp(-\beta H(q, p)) dq dp$$

to the set  $\mathcal{A}$ . However, care must be taken with the precise meaning of this restriction; while we can unambiguously restrict the canonical density function  $\rho \propto \exp(-\beta H)$  to  $\mathcal{A}$ , it is less obvious how to obtain an explicit expression for the appropriate restricted measure defined on  $\mathcal{A}$  corresponding to the restriction of the Liouville measure  $d\Gamma = dq dp$ . For deterministic Hamiltonian systems, i.e., with  $\gamma, \sigma = 0$ , the restricted Liouville measure is typically defined in terms of the associated *symplectic form*, and so is the Liouville measure of the constrained system (see, e.g., [14, 23] and the discussion in the Abstract). But this measure should not be confused with the surface measure on the embedded manifold  $\mathcal{A} \subset \mathcal{Q} \times \mathbf{R}$  that is obtained from restricting the Lebesgue measure  $d\lambda = dq dp$ . We next propose an explicit “ambient” expression for the appropriate restricted measure on  $\mathcal{A}$ .

#### 3.1 Integration over fibres

We will show that the constrained Liouville measure on  $\mathcal{A}$  can be represented as a product of the surface element on  $\Sigma$  (the so-called *base manifold*) together with the  $(n - d)$ -dimensional Lebesgue measure on the cotangent space at  $q \in \Sigma$  (the so-called *fibre*), and a  $q$ -dependent multiplicative correction that depends upon the angle between the tangent and cotangent spaces at  $q$ . It is the presence of the angle correction factor that is perhaps somewhat surprising.



**Fig. 1.** Integration over the fibres: integration against the constrained phase space Liouville measure can be decomposed into an integration along the cotangent space fibre for each  $q$ , then multiplication by a  $q$ -dependent angle correction term, finally followed by integration over the base manifold  $\Sigma$  regarded as a subset of  $\mathcal{Q} \subseteq \mathbf{R}^n$ .

While our final formula does not depend upon local coordinates, our derivation of it does, so to give precise statements, we first introduce local coordinates (which we tacitly assume to exist) by the continuously differentiable and locally invertible map

$$\phi: \begin{cases} \mathbf{R}^{n-d} \rightarrow \Sigma, & x \mapsto q(x) \\ \mathbf{R}^{n-d} \rightarrow T_q^*\Sigma, & v \mapsto [\mathbf{t}_1^*(q), \dots, \mathbf{t}_{n-d}^*(q)]v \end{cases} \quad (8)$$

where  $\{\mathbf{t}_1^*(q), \dots, \mathbf{t}_{n-d}^*(q)\}$  is an orthonormal basis of  $T_q^*\Sigma$  for any fixed  $q \in \Sigma$ .<sup>1</sup>

It will be convenient to state our result for a single scalar constraint  $d = 1$  first. Then the aforementioned surface elements can be written as

$$d\sigma_\Sigma(q) = \sqrt{\det(\nabla q \nabla q^T)} dx, \quad d\lambda_q^*(p) = dv. \quad (9)$$

We denote the constrained Liouville measure by  $\Gamma_{\mathcal{A}}$ , and introduce the notation  $\angle(x, y)$  for the angle between two vectors, i.e.,

$$\angle(x, y) = \arccos \left( \frac{\langle x, y \rangle}{\|x\| \|y\|} \right).$$

Integration against  $d\Gamma_{\mathcal{A}}$  can now be understood in the following way: for any  $\Gamma_{\mathcal{A}}$ -integrable function  $w : \mathcal{A} \rightarrow \mathbf{R}$  and subsets  $A \subset \Sigma$  and  $B_q \subset T_q^*\Sigma$ ,  $q \in A$ , we have

$$\begin{aligned} & \int_{\{q \in A, p \in B_q\} \subset \mathcal{A}} w(q, p) d\Gamma_{\mathcal{A}}(q, p) \\ &= \int_A \cos \angle \{ \nabla f(q), g(q)^{-1} \nabla f(q) \} \left( \int_{B_q} w(q, p) d\lambda_q^*(p) \right) d\sigma_\Sigma(q), \end{aligned}$$

<sup>1</sup> Recall that we identify all tangent and cotangent spaces with the respective subspaces of  $\mathbf{R}^n$ ; finding an orthonormal basis for  $T_q^*\Sigma$  therefore merely means finding an orthonormal basis for a linear subspace of  $\mathbf{R}^n$ .

The integral identity can be readily verified by doing a suitable change of coordinates (see the Appendices for details). Since the integration splits into an integral over the momentum in the cotangent space  $p \in B_q \subset T_q^* \Sigma$ , followed by an integration over the admissible configurations  $q \in A \subset \Sigma$ , we will describe it as “integration over the fibres”. The cosine correction term involves the angle between the normal vectors to the tangent and cotangent spaces at  $q$ , which, as a matter of definition, is also the angle between the codimension one tangent and cotangent spaces.

The only complication in the case of multiple constraints is that we must introduce the notion of angle between subspaces, rather than between vectors. While this notion of angle is less well known, it is nevertheless entirely standard (see, e.g., [17, 26, 33]).

**Definition 1.** Let  $X, Y \subset \mathbf{R}^n$  be subspaces of equal dimension  $\dim X = \dim Y = k$ , and let  $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$  and  $\{\mathbf{y}_1, \dots, \mathbf{y}_k\}$  be arbitrary bases of  $X$  and  $Y$ . The angle  $\angle\{X, Y\}$  between  $X$  and  $Y$  is defined by

$$\angle\{X, Y\} := \arccos \frac{\det([\mathbf{x}_1, \dots, \mathbf{x}_k]^T [\mathbf{y}_1, \dots, \mathbf{y}_k])}{\sqrt{\det([\mathbf{x}_1, \dots, \mathbf{x}_k]^T [\mathbf{x}_1, \dots, \mathbf{x}_k])} \sqrt{\det([\mathbf{y}_1, \dots, \mathbf{y}_k]^T [\mathbf{y}_1, \dots, \mathbf{y}_k])}}$$

Note that  $\angle\{X, Y\}$  is well-defined as the right hand side does not depend on the specific choice of the bases, and

$$\begin{aligned} & \det([\mathbf{x}_1, \dots, \mathbf{x}_k]^T [\mathbf{y}_1, \dots, \mathbf{y}_k]) \\ & \leq \sqrt{\det([\mathbf{x}_1, \dots, \mathbf{x}_k]^T [\mathbf{x}_1, \dots, \mathbf{x}_k])} \sqrt{\det([\mathbf{y}_1, \dots, \mathbf{y}_k]^T [\mathbf{y}_1, \dots, \mathbf{y}_k])}. \end{aligned}$$

Based on Definition 1 we can now state our main result, the proof of which can be found in Appendix 6.

**Proposition 2.** Let  $N_q \Sigma = (T_q \Sigma)^\perp$  and  $N_q^* \Sigma = (T_q^* \Sigma)^\perp$  be the orthogonal complements of  $T_q \Sigma$  and  $T_q^* \Sigma$  in  $\mathbf{R}^n$  and assume that the fluctuation-dissipation relation  $2\gamma = \beta\sigma\sigma^T$  holds. Further let  $H_{\mathcal{A}} = H|_{\mathcal{A}}$  denote the restriction of the total energy to the constrained phase space  $\mathcal{A}$ . Then

$$\rho_{\mathcal{A}} = \frac{1}{Z_{\mathcal{A}}} \cos \angle\{N_q \Sigma, N_q^* \Sigma\} \exp(-\beta H_{\mathcal{A}}(q, p)),$$

with  $Z_{\mathcal{A}}$  being the normalization constant

$$Z_{\mathcal{A}} = \int_{\Sigma} \int_{T_q^* \Sigma} \cos \angle\{N_q \Sigma, N_q^* \Sigma\} \exp(-\beta H_{\mathcal{A}}(q, p)) d\lambda_q^*(p) d\sigma_{\Sigma}(q),$$

is the density function for the canonical equilibrium distribution of the Langevin dynamics (7) when the measure is the product  $d\lambda_q^*(p) d\sigma_{\Sigma}(q)$  of Lebesgue measure over the cotangent fibres, followed by integration with respect to the surface measure in the configuration constraint manifold  $\Sigma(q) \in \mathbf{R}^n$ .

*Remark 3.* Even if the generalized mass matrix  $g$  is independent of  $q$ , the angle correction term will in general not be constant unless the normal subspace  $N_q \Sigma$  is also independent of  $q$ .

*Remark 4.* If  $X$  and  $Y$  are two subspaces of  $\mathbf{R}^n$  with equal dimension  $k$ , and  $X^\perp, Y^\perp$  are their orthogonal complements in  $\mathbf{R}^n$ , then it is a standard result that

$$\angle\{X, Y\} = \angle\{X^\perp, Y^\perp\}.$$

Hence

$$\cos \angle \{N_q \Sigma, N_q^* \Sigma\} = \cos \angle \{T_q \Sigma, T_q^* \Sigma\}. \quad (10)$$

In other words, the correction factor involves the angle between velocity and momentum (i.e., tangent and cotangent) spaces to the constraint set  $\Sigma(q)$ , just as in the codimension one case.

*Remark 5.* There is a considerable literature on how to compute the angle between two subspaces (see, e.g., [17, 26, 33]). If we write  $\{\nabla f_1, \dots, \nabla f_d\}$  and  $\{g^{-1}\nabla f_1, \dots, g^{-1}\nabla f_d\}$  as bases of  $N_q \Sigma$  and  $N_q^* \Sigma$ , and write  $K = Ag^{-1}A^T$  with  $A = \nabla f$ ,  $G = AA^T$ , and  $J = Ag^{-2}A^T$ , the angle formula becomes

$$\angle \{N_q \Sigma, N_q^* \Sigma\} = \arccos \left( \frac{\det K(q)}{\sqrt{\det G(q) \det J(q)}} \right). \quad (11)$$

The angle can also be computed from the eigenvalues of the matrix  $KG^{-1}K^TJ^{-1}$  [16]. Specifically

$$\cos \angle \{N_q \Sigma, N_q^* \Sigma\} = \prod_{i=1}^d \cos(\phi_i). \quad (12)$$

where the cosines of the so-called principal angles  $\phi_1, \dots, \phi_d$  are the square roots of the eigenvalues of the matrix  $KG^{-1}K^TJ^{-1}$ .

### 3.2 The configuration space marginal distribution

In many situations the observables of interest are independent of the momenta  $p$ . In such cases the expectations can be evaluated with respect to the marginal of the canonical distribution in the configuration variables  $q$ . The latter can be obtained from the phase space distribution by integrating out the momenta, i.e.,

$$d\nu_\Sigma(\cdot) = \int_{T_q^* \Sigma} d\rho_{\mathcal{A}}(\cdot, p),$$

and because the kinetic energy is assumed to be pure quadratic we expect this marginal to be explicitly computable. Using (8) we conclude that

$$\int_{T_q^* \Sigma} \exp(-\beta H_{\mathcal{A}}(q, p)) d\lambda_q^*(p) = (\det((\mathbf{t}^*)^T g^{-1} \mathbf{t}^*))^{-1/2} \exp(-\beta V)$$

with the columns of the  $n \times (n-d)$ -matrix  $\mathbf{t}^* = [\mathbf{t}_1^*, \dots, \mathbf{t}_{n-d}^*]$  being a set of orthonormal basis vectors for the cotangent space  $T_q^* \Sigma$ . Such an orthonormal basis for  $T_q^* \Sigma$  may exist only locally in a neighbourhood of  $q \in \Sigma$ , but this is all we need here. Further using the identity (11), it follows that

$$\frac{\cos \angle \{N_q \Sigma, N_q^* \Sigma\}}{\sqrt{\det((\mathbf{t}^*)^T g^{-1} \mathbf{t}^*)}} = \frac{\sqrt{\det((\mathbf{t}^*)^T g^{-1} \mathbf{t}^*)}}{\sqrt{\det((\mathbf{t}^*)^T g^{-2} \mathbf{t}^*)}},$$

which implies

$$d\nu_\Sigma(q) = \frac{1}{Z_\Sigma} \frac{\sqrt{\det((\mathbf{t}^*)^T g^{-1} \mathbf{t}^*)}}{\sqrt{\det((\mathbf{t}^*)^T g^{-2} \mathbf{t}^*)}} \exp(-\beta V(q)) d\sigma_\Sigma(q).$$

Here  $Z_\Sigma$  is the constant that normalizes the total probability to one. The appearance of the orthonormal vectors  $\mathbf{t}_i^*$  in the ratio of determinants seems rather inconvenient and we may further simplify the expression by instead using an arbitrary basis of the tangent space. Exploiting the fact that the metric tensor  $g \in \mathbf{R}^{n \times n}$  can be regarded as a linear map  $g: T_q \Sigma \rightarrow T_q^* \Sigma$  that transforms velocity vectors into momentum vectors, we arrive at:

**Lemma 1.** *Let  $\{\mathbf{t}_1, \dots, \mathbf{t}_{n-d}\}$  be an arbitrary basis of the tangent space  $T_q \Sigma$  and set  $\mathbf{t} = [\mathbf{t}_1, \dots, \mathbf{t}_{n-d}] \in \mathbf{R}^{n \times (n-d)}$ . Then the configuration marginal distribution  $\nu_\Sigma$  for the Langevin dynamics (7) has the form*

$$d\nu_\Sigma(q) = \frac{1}{Z_\Sigma} \frac{\sqrt{\det(\mathbf{t}^T g \mathbf{t})}}{\sqrt{\det(\mathbf{t}^T \mathbf{t})}} \exp(-\beta V(q)) d\sigma_\Sigma(q).$$

*Remark 6.* When there is no constraint, i.e.  $\Sigma = \mathcal{Q}$  the matrix  $\mathbf{t}$  is square and the measure reduces to

$$d\nu = \frac{1}{Z} \sqrt{\det(g(q))} \exp(-\beta V(q)) dq,$$

in which  $\sqrt{\det(g(q))}$  is the classic metric correction factor, and the distribution is pure Boltzmann with respect to  $dq$  if and only if the generalized mass matrix is constant. In contrast the constrained marginal distribution may not be Boltzmann with respect to  $d\sigma_\Sigma(q)$  even if the generalized mass matrix is constant.

*Remark 7.* If we introduce a local parametrization  $x \mapsto q(x)$  of the embedding  $\Sigma \subset \mathcal{Q}$  we may take  $\mathbf{t} = \nabla q(x)$  and  $d\sigma_\Sigma(q) = \sqrt{\det(\nabla q^T \nabla q)} dx$ . Then we obtain

$$d\nu_\Sigma(q) = \frac{1}{Z_\Sigma} \sqrt{\det(\nabla q^T g(q) \nabla q)} \exp(-\beta V(q)) dx,$$

which again may not be Boltzmann with respect to  $dx$  even if the generalized mass matrix is constant.

*Remark 8.* Finally we note that when  $\mathcal{Q}$  is considered as a Riemannian manifold that is endowed with the Riemannian metric  $\langle \cdot, \cdot \rangle_g$ , then the Riemannian surface element that is induced by  $g$  and the embedding  $\Sigma \subset \mathcal{Q}$  is

$$d\sigma_{\Sigma,g}(q) = \sqrt{\det(\nabla q^T g(q) \nabla q)} dx$$

so that we arrive at

$$d\nu_\Sigma(q) = \frac{1}{Z_\Sigma} \exp(-\beta V(q)) d\sigma_{\Sigma,g}(q).$$

which is pure Boltzmann with respect to the Riemannian surface element  $d\sigma_{\Sigma,g}(q)$ .

*Remark 9.* An ambient configuration space representation of  $d\nu_\Sigma$  in terms of Dirac's delta function that is common in the molecular dynamics literature is (see, e.g., [1, 2, 7])

$$d\nu_\Sigma(q) = \frac{1}{Z_\Sigma} \exp(-\beta V(q)) \sqrt{\det K(q)} \delta(f(q)) dq. \quad (13)$$

The appearance of the weight  $\sqrt{\det K}$  with  $K = \nabla f g^{-1} \nabla f^T$  can be understood as follows: Letting the level sets  $\Sigma(c) = \{q \in \mathcal{Q}: f(q) = c\}$  define a complete foliation of  $\mathcal{Q} \subseteq \mathbf{R}^n$ , Federer's coarea formula asserts [8, 9, 12]

$$\int_{\mathcal{Q}} h d\lambda_{\mathcal{Q}} = \int_{\mathbf{R}^d} \left( \int_{\Sigma(c)} h (\det K)^{-1/2} d\sigma_{\Sigma(c),g} \right) dc,$$

for any integrable function  $h \in L^1(d\lambda_{\mathcal{Q}})$ . Here  $d\lambda_{\mathcal{Q}}(q) = \sqrt{\det g(q)}dq$  denotes the volume element on  $\mathcal{Q}$  with respect to the Riemannian metric  $\langle \cdot, \cdot \rangle_g$ . Noting that

$$\int_{\mathcal{Q}} h d\lambda_{\mathcal{Q}} = \int_{\mathbf{R}^d} \left( \int_{\mathcal{Q}} h(q) \delta(f(q) - c) d\lambda_{\mathcal{Q}}(q) \right) dc,$$

by definition of the delta function, it follows that

$$d\sigma_{\Sigma, g}(q) = \sqrt{\det K(q)} \delta(f(q) - c) dq$$

which implies (13). Note that the delta function  $\delta(f(q) - c)$  depends on the specific choice of the level set function  $f$ , even though the constraint surface  $\Sigma(c)$  does not. The weight  $\sqrt{\det K} = \sqrt{\det(\nabla f g^{-1} \nabla f^T)}$  absorbs this gauge dependence.

## 4 The Impetus-Striction ambient space formulation of constrained dynamics

In this section we will present a second ambient space formulation of holonomically constrained Langevin dynamics, which has the feature that the constraints are by construction integrals of the resulting Hamiltonian-Langevin system. The central idea is that in the presence of holonomic constraints there is a freedom in the choice of the conjugate variable leading to the Hamiltonian form of the dynamics. No matter what choice of conjugate variable is made, the evolution in configuration space is identical, but the phase space dynamics differ, and this freedom can be used to some advantage. The impetus-striction approach has previously been exploited in the analysis of various deterministic infinite dimensional Lagrangian systems where the dynamics are governed by partial differential equations [3, 4, 10, 22], but the approach applies equally well to the finite dimensional case governed by ordinary differential equations where the introduction of the stochastic Langevin terms is more straightforward (e.g., see [13]). As the impetus-striction formulation is not entirely standard we first introduce the basic ideas for deterministic systems. More details can be found in [10].

### 4.1 The deterministic case of holonomically constrained dynamics

The most natural way to introduce the impetus-striction formulation is via Hamilton's least action principle. In the unconstrained case the action integral takes the form

$$S[q] = \int_a^b L(q(t), \dot{q}(t)) dt$$

where, consistent with the earlier presentation, we are primarily interested in Lagrangians of the special form

$$L: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}, \quad L(q, v) = \frac{1}{2}|v|_g^2 - V(q).$$

Then the Lagrangian dynamics are the Euler-Lagrange equations of the action integral, and their Hamiltonian form is the system

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} \end{aligned} \tag{14}$$

(corresponding to the unconstrained Hamiltonian-Langevin equation (1) with  $\gamma = \sigma = 0$ ) which arises via the Legendre transform, with the conjugate variable unambiguously defined as  $p := L_{\dot{q}}(q, \dot{q})$ , and

$$H: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}, \quad H(q, p) = \frac{1}{2}|p|_{g^{-1}}^2 + V(q).$$

If  $d$  holonomic constraints of the form  $f(q) = 0$  are now introduced, then the action principle is typically modified by the addition of a term of the form  $-\lambda \cdot f(q)$  where  $\lambda(t)$  is a  $d$ -vector of Lagrange multipliers. (The Lagrange multiplier  $\lambda$  should not be confused with the Lebesgue measure from the previous section.) One then arrives at the augmented Lagrangian

$$\bar{L}: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}, \quad \bar{L}(q, v) = \frac{1}{2}|v|_g^2 - V(q) - \lambda \cdot f(q),$$

and the dynamics is given by the differential-algebraic system comprising the Euler-Lagrange equations of the augmented Lagrangian  $\bar{L}$  with respect to  $q$ , and the additional  $d$  algebraic equations  $f(q(t)) = 0$  corresponding to the additional  $d$  unknowns  $\lambda(t)$  whose time derivative appears nowhere.

It is then also standard to write the Hamiltonian form of the Euler-Lagrange equations with respect to  $q$  for the augmented Lagrangian  $\bar{L}$ . The significant point is that the additional augmented terms are effectively a modification of the potential part of the action because the generalized velocity  $\dot{q}$  does not appear in them. Consequently the conjugate momentum for this augmented Lagrangian is the same as for the unconstrained case  $p := \bar{L}_{\dot{q}}(q, \dot{q}) = L_{\dot{q}}(q, \dot{q})$ , and one arrives at the differential-algebraic Hamiltonian system:

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} - \nabla f(q)^T \lambda \\ f(q) &= 0, \end{aligned} \tag{15}$$

with precisely the same Hamiltonian as before.

It is then a standard computation to eliminate the algebraic equations in (15) via the (much simpler) deterministic ( $\sigma = \gamma = 0$ ) version of the stochastic computation that we carried out in Sect. 2.2 to arrive at an explicit functional form  $\lambda(q, p)$  which guarantees that the constraint set  $f(q) = 0 = \nabla f(q)g^{-1}(q)p$  is an invariant of the dynamics

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} - \nabla f(q)^T \lambda(q, p). \end{aligned} \tag{16}$$

Equations (16) are an ambient space formulation of the holonomically constrained dynamics for which the constraint set is invariant. The dynamics are perfectly well-defined off the constraint set, as they must be, but there are two unattractive features. First the dynamics off the constraint set are in general not Hamiltonian. Second the constraint set is itself repelling in the sense that trajectories that start close by move away, which is a feature that can be problematic for numerics.

Now we are ready to introduce the impetus-striction formulation. The starting point is that from the calculus of variations point of view it is equally valid to augment the Lagrangian with the differentiated constraint  $h(q, \dot{q}) = 0$  (where

$h(q, v) = \nabla f(q)v$ , and a multiplier vector  $\mu(t) \in \mathbf{R}^d$  associated with the time differentiated constraint, which we call the striction for brevity, to arrive at a different augmented Lagrangian

$$\hat{L}(q, v, \mu) = L(q, v) - \mu \cdot h(p, v),$$

which imposes the constraint at the velocity level rather than on the configuration level. By construction, the velocity constraint is integrable so that  $h(q, \dot{q}) = 0$  implies  $f(q) = c$  and, by properly choosing the initial conditions, we can make the integration constant  $c = 0$ . An integration by parts in time in the action integral reveals that the striction  $\mu$  and multiplier  $\lambda$  are related through the identity  $-\dot{\mu} = \lambda$ . Then it is simple to compute that the Euler-Lagrange equations of the two augmented Lagrangians with respect to  $q$  are identical.<sup>2</sup>

The interesting features of the impetus-striction formulation arise when we compute the Hamiltonian form of the dynamics. Now there is a different conjugate momentum, which we call the impetus:

$$\xi := \partial \hat{L} / \partial \dot{q} = \partial L / \partial \dot{q} - \nabla f(q)^T \mu = p - \nabla f(q)^T \mu.$$

In particular the striction and constraint Jacobian  $\nabla f(q)$  enter in the definition of the impetus  $\xi$ .

Then the impetus-striction formulation of the constrained dynamics is as follows. The constraint  $h(q, \dot{q}) = 0$  is enforced by choosing the striction  $\mu(q, \xi)$  via

$$\mu(q, \xi) := \underset{\mu}{\operatorname{argmin}} H(q, \xi + \nabla f(q)^T \mu),$$

where as before  $H$  is the Hamiltonian associated with the unconstrained Lagrangian  $L$ . This prescription defines the function  $\mu(q, \xi)$  uniquely for any Hamiltonian that is strictly convex in the momenta  $p$ , but for the simple class of Lagrangians and Hamiltonians assumed here the minimization step is particularly straightforward and has linear first-order (necessary *and* sufficient) conditions

$$0 = \nabla f g^{-1} (\xi + \nabla f^T \mu) = \nabla f g^{-1} p = \nabla f \dot{q} = h(q, \dot{q}),$$

so that the minimization step enforces the constraint, and yields the explicit formula

$$\mu(q, \xi) = - (A g^{-1} A^T)^{-1} A g^{-1} \xi,$$

where we have again introduced the abbreviation  $A = \nabla f$ . Recalling the definition (4) of the oblique projection  $\mathcal{P}$ , it follows that

$$\xi + \nabla f^T \mu = \mathcal{P} \xi = p.$$

Now we introduce the impetus-striction Hamiltonian

$$\mathcal{H}: \mathcal{Q} \times \mathbf{R}^n, \quad \mathcal{H}(q, \xi) = H(q, \mathcal{P} \xi) = \min_{\mu} H(q, \xi + A(q)^T \mu). \quad (17)$$

---

<sup>2</sup> For readers interested in nonholonomic constraints we remark that the Euler-Lagrange equations of an augmented Lagrangian with a general velocity level constraint of the form  $\eta(q)\dot{q} = 0$  for  $\eta$  some  $d \times n$  matrix do not in general correspond to the dynamics given by d'Alembert's principle for constrained systems. Analogously our equivalence of the two systems of Euler-Lagrange equations for the two different forms of augmented Lagrangian depends crucially upon the particular structure that for the holonomic case  $\eta(q) = \nabla f(q)$  is a Jacobian matrix.

It can be computed that the Euler-Lagrange equations of motion of the augmented action  $\hat{L}$  with respect to  $q$  can be recast as the Hamiltonian system

$$\begin{aligned}\dot{q} &= \frac{\partial \mathcal{H}}{\partial \xi} \\ \dot{\xi} &= -\frac{\partial \mathcal{H}}{\partial q}.\end{aligned}\tag{18}$$

For our special class of Lagrangians we arrive at the concrete form

$$\begin{aligned}\dot{q} &= g^{-1}(\xi + \nabla f^T \mu) \\ \dot{\xi} &= -\nabla V(q) + \frac{1}{2}(\xi + \nabla f^T \mu)^T [\nabla g^{-1}] (\xi + \nabla f^T \mu) + \mu^T \nabla^2 f g^{-1} (\xi + \nabla f^T \mu),\end{aligned}$$

where it is understood that  $\mu(q, \xi)$  is the known function defined through the minimization step above.<sup>3</sup> Here  $\nabla^2 f$  and  $\nabla g^{-1}$  are order three tensors, so that (with the summation convention)  $\mu^T \nabla^2 f = \mu_i \nabla^2 f_i$  and  $\eta^T [\nabla g^{-1}] \eta$  is understood as a vector with the entries

$$(\eta^T [\nabla g^{-1}] \eta)_k = \frac{\partial}{\partial q^k} g^{ij} \eta_i \eta_j.$$

*Remark 10.* The system (18) is an ambient space formulation of the constrained Lagrangian dynamics that is a standard canonical, unconstrained Hamiltonian in all of phase space  $\mathcal{Q} \times \mathbf{R}^n$  with a completely explicit formula for the Hamiltonian  $\mathcal{H}$ . The only unusual feature is that while  $\mathcal{H}$  is a convex function of the conjugate momentum (or impetus)  $\xi$ , it is *not strictly* convex. This follows because while the Hamiltonian  $H(q, p)$  is a strictly convex function of  $p$ , we have  $\mathcal{H}(q, \xi) = H(q, \mathcal{P}\xi)$  and the intervening projection  $\mathcal{P}$  has a nullspace, specifically the column space of  $\nabla f(q)^T$ , or equivalently the normal space to the tangent space  $T_q \Sigma$ . Thus for each  $q$ ,  $\mathcal{H}(q, \xi)$  is exactly constant on a  $d$ -dimensional subspace. This corresponds to a  $d$ -dimensional infinitesimal (variational) symmetry of  $\mathcal{H}$ ,  $\xi \mapsto \xi + \nabla f(q)^T \epsilon$  and so, via Noethers Theorem, generates  $d$  integrals of motion, which are precisely the quantities  $f(q)$ . Thus the impetus-striction formulation starts from a constrained Lagrangian that is strictly convex in the generalized velocities  $\dot{q}$  and constructs an unconstrained Hamiltonian formulation which is only convex in the impetus  $\xi$ , and for which the Lagrangian constraints are recast as Hamiltonian integrals. In this sense the impetus-striction construction is dual to Dirac's theory of constraints, which starts from a Lagrangian that is a convex, but not strictly convex, function of the generalized velocities, and constructs an equivalent strictly convex Hamiltonian system with constraints [5, 6].

*Remark 11.* The component of the impetus  $\xi$  in the normal space to the tangent space  $T_q \Sigma$  is in some sense not a physical quantity, and it corresponds to a gauge freedom. If the right hand side of the impetus equation is modified by the addition of a vector field in the constraint normal space, but is otherwise arbitrary, then the evolution of  $\xi$  is of course changed, but the associated evolution of the configuration variable  $q$  is unaltered.

*Remark 12.* Because the  $f(q)$  are integrals, the level sets form (at least locally) a foliation of configuration space. As a consequence any particular constraint set, e.g.,  $f(q) = 0$  is neutrally stable to the full ambient space dynamics, in contrast to the instability in the more standard ambient space formulation described previously.

<sup>3</sup> When computing partial derivatives it should be noted that the minimizing property of the function  $\mu(q, \xi)$  means that its partial derivatives do not contribute any terms.

*Remark 13.* The fact that the Hamiltonian is only defined via an auxiliary minimization, which for the class of Lagrangians considered here leads to an explicit projection, is reminiscent of Pontryagin's maximum principle in optimal control. However the resemblance is to our understanding merely superficial.

## 4.2 Impetus-striction Langevin dynamics

As the Hamiltonian system (18) is now a standard canonical one it is natural to modify it to be a Langevin equation by adding friction and noise in the standard way, which then yields

$$\begin{aligned}\dot{q} &= \frac{\partial \mathcal{H}}{\partial \xi} \\ \dot{\xi} &= -\frac{\partial \mathcal{H}}{\partial q} - \gamma(q) \frac{\partial \mathcal{H}}{\partial \xi} + \sigma(q) \dot{W}\end{aligned}\tag{19}$$

which is our second ambient space Langevin formulation. Here we assume that friction and noise coefficients are the same as in (1) or (7). We have:

**Proposition 3.** *Define the impetus-striction Hamiltonian as*

$$\mathcal{H}: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}, \quad \mathcal{H}(q, \xi) = H(q, \mathcal{P}\xi)$$

where  $\mathcal{P} = I - A^T(Ag^{-1}A^T)^{-1}Ag^{-1}$  and  $H$  is the standard Hamiltonian dual to the given Lagrangian

$$H(q, p) = \frac{1}{2}|p|_{g^{-1}}^2 + V(q).$$

Then the Langevin system (19) has the following properties

1. For any regular value  $c$  of  $f: \mathcal{Q} \rightarrow \mathbf{R}^d$ , the set

$$\Sigma(c) \times \mathbf{R}^n = \{(q, p) \in \mathcal{Q} \times \mathbf{R}^n : f(q) = c\}$$

is invariant under the dynamics (19). In other words the  $f(q)$  are integrals of the dynamics, so that a constraint of the form  $f(q) = c$  is automatically satisfied along trajectories of the dynamics for all  $t$ , with the value of  $c$  set by the initial conditions on the configuration  $q(0)$ .

2. If  $(q(t), \xi(t)) \subset \Sigma(c) \times \mathbf{R}^n$  is the solution of (19), then, with probability one,  $(q(t), p(t)) \subset \mathcal{A}(c)$  is the solution of (7) where  $\xi(t)$  and  $p(t)$  are related by

$$p(t) = \mathcal{P}(q(t))\xi(t),$$

which holds almost surely (i.e., with probability one).

## 5 The Impetus-Striction-Langevin equilibrium distributions

It remains to discuss the properties of the equilibrium distribution associated with the dynamics (19). Because (19) are a standard Hamiltonian system, it is well-known that provided the fluctuation-dissipation relation  $2\gamma = \beta\sigma\sigma^T$  holds, the Gibbs-Boltzmann density  $\exp(-\beta\mathcal{H})$  is a stationary solution of the Fokker-Planck or Kolmogorov forward equation

$$\left(\frac{\partial}{\partial t} - L^*\right)\psi = 0, \quad \psi(\cdot, \cdot, 0) = \psi_0$$

that propagates the evolution of probability densities forward in time. Here  $L^*$  is the corresponding Fokker-Planck or forward operator

$$L^* = \frac{1}{2}\sigma\sigma^T : \nabla_\xi^2 - \nabla_\xi \mathcal{H} \cdot \nabla_q + (\nabla_q \mathcal{H} + \gamma \nabla_\xi \mathcal{H}) \cdot \nabla_\xi + (\gamma : \nabla_\xi^2 \mathcal{H}).$$

To verify that  $\psi_0 = \exp(-\beta\mathcal{H})$  is a stationary solution with  $\partial\psi/\partial t = 0$  one merely computes the various expressions on the right hand side of the Fokker-Planck equation in the standard way. Because the  $d$  constraints  $f(q)$  are by construction integrals of the impetus-striction dynamics, both with and without the fluctuation-dissipation terms, an analogous computation reveals that any expression of the form  $\exp(-\beta\{\mathcal{H} + \Phi(f(q))\})$  is another stationary solution of the Fokker-Planck equation, where  $\Phi : \mathbf{R}^d \rightarrow \mathbf{R}$  is arbitrary (cf. [29]). Thus it can be seen that the only equilibrium measures of interest are marginals on  $f(q) = 0$ .

In the case of an impetus-striction Hamiltonian there remains a problem of normalizability. Specifically we cannot call  $\exp(-\beta\mathcal{H})$  a stationary *probability* distribution because

$$\int_{\mathbf{R}^n} \exp(-\beta\mathcal{H}(\cdot, \xi)) d\xi = \infty,$$

i.e.,  $\exp(-\beta\mathcal{H})$  is not normalizable over all of  $\mathbf{R}^{2n}$ . And the difficulty remains even with the restriction  $f(q) = 0$ . The reason for the divergence of the integral are the flat directions of the Hamiltonian under the impetus-striction symmetry  $\xi \rightarrow \xi + \nabla f(q)^T \epsilon$ . The saving grace is that, as previously remarked, these flat directions of  $\xi$  are not true physical unknowns, so that any expectation of physical interest should involve only the tangential component of  $\xi$ . We will show that the appropriate marginal distribution is normalizable. Specifically we restrict to the tangent bundle  $T\Sigma$  on which  $f(q) = 0$  and the degenerate directions  $\xi$  are suppressed.

## 5.1 Integration over fibres revisited

**Proposition 4.** *Let  $T_q\Sigma$  be the constraint tangent space in  $\mathbf{R}^n$  and assume that the fluctuation-dissipation relation  $2\gamma = \beta\sigma\sigma^T$  holds. Let  $d\lambda_q(\xi)$  denote the Lebesgue measure on the tangent space  $T_q\Sigma$ . Then the probability distribution*

$$d\rho_{\mathcal{A}} = \frac{1}{Z_{\mathcal{A}}} \exp(-\beta\mathcal{H}(q, \xi)) d\lambda_q(\xi) d\sigma_{\Sigma}(q), \quad \xi \in T_q\Sigma,$$

with the partition function given by

$$Z_{\mathcal{A}} = \int_{\Sigma} \left( \int_{T_q\Sigma} \exp(-\beta\mathcal{H}(q, \xi)) d\lambda_q(\xi) \right) d\sigma_{\Sigma}(q),$$

is invariant under the dynamics (19).

*Proof.* Using that  $\angle\{N_q\Sigma, N_q^*\Sigma\} = \angle\{T_q\Sigma, T_q^*\Sigma\}$ , it is sufficient to show that

$$\int_{U \subset T_q\Sigma} \exp(-\beta\mathcal{H}) d\lambda_q = \cos \angle\{T_q\Sigma, T_q^*\Sigma\} \int_{U^* \subset T_q^*\Sigma} \exp(-\beta H) d\lambda_q^*.$$

Let the vectors  $\{\mathbf{t}_1, \dots, \mathbf{t}_{n-d}\}$  form a locally orthonormal basis of  $T_q\Sigma$  and define the matrix  $\mathbf{t} = [\mathbf{t}_1, \dots, \mathbf{t}_{n-d}] \in \mathbf{R}^{n \times (n-d)}$ . On  $T_q\Sigma$ , we introduce local coordinates by

$$\varphi : \mathbf{R}^{n-d} \rightarrow T_q\Sigma, \quad u \mapsto \mathbf{t}u.$$

Then, for any fixed  $q \in \Sigma$ ,

$$\begin{aligned} \int_{U \subset T_q \Sigma} \exp(-\beta \mathcal{H}(q, \xi)) d\lambda_q(\xi) &= \int_{U \subset T_q \Sigma} \exp(-\beta H(q, \mathcal{P}\xi)) d\lambda_q(\xi) \\ &= \int_{U \subset T_q \Sigma} \exp(-\beta H(q, \mathcal{P}t u)) du. \end{aligned}$$

But the column space of the matrix  $\mathcal{P}t$  equals the cotangent space  $T_q^* \Sigma$ , where a parametrization of  $T_q^* \Sigma$  is given by  $p = \varphi^*(w)$  with

$$\varphi^*: \mathbf{R}^{n-d} \rightarrow T_q^* \Sigma, \quad w \mapsto g t \hat{g}^{-1} w.$$

Hence

$$\begin{aligned} \int_{U \subset T_q \Sigma} \exp(-\beta \mathcal{H}) d\lambda_q &= \int_{U \subset T_q \Sigma} \exp(-\beta H(\cdot, \mathcal{P}t u)) du \\ &= \frac{1}{\sqrt{\det([g t \hat{g}^{-1}]^T [g t \hat{g}^{-1}])}} \int_{U^* \subset T_q^* \Sigma} \exp(-\beta H) d\lambda_q^*. \end{aligned}$$

The assertion follows upon noting that

$$\frac{1}{\sqrt{\det([g t \hat{g}^{-1}]^T [g t \hat{g}^{-1}])}} = \frac{\det(t^T g t)}{\sqrt{\det(t^T g^2 t)}} = \cos \angle \{T_q \Sigma, T_q^* \Sigma\}.$$

*Remark 14.* Because it is orthogonal projection of the impetus  $\xi$  onto the tangent (and not cotangent) space that encapsulates the physical variables, there is now no angle correction term in the expression for stationary measure.

*Remark 15.* Addressing the question of whether or not the dynamics (7) or (19) are ergodic with respect to the marginal distribution  $\rho_A$  would involve verification of various technical conditions such as weak controllability (i.e., Hörmander's condition) and growth conditions on the functions  $V$  and  $f$ . However such analyses lie beyond the scope of this article [24, 25].

*Remark 16.* The situation is similar to that of deterministic Hamiltonian dynamics, in which any function of the Hamiltonian is a stationary solution of the Liouville equation. Moreover, the only candidate for an ergodic invariant measure, namely, the microcanonical measure  $\delta(H - E)$ , is singular with respect to the Liouville measure. In point of fact rather few concrete, physically motivated, Hamiltonian systems are actually known to be ergodic. Examples that are known to be ergodic include certain billiards and geodesic flows on surfaces of constant negative mean curvature.

## 5.2 The configuration space marginal

We can of course continue and compute further marginals. For example, the configuration space marginal should not depend on which Hamiltonian formulation is used in phase space. In fact the configuration space marginal from the impetus-striction Langevin dynamics coincides with any of the forms described in Section 3.2. For example, using the Riemannian surface measure we have:

**Corollary 1.** *The configuration space marginal of the stationary measure for (19) can be expressed as*

$$d\nu_\Sigma(q) = \frac{1}{Z_\Sigma} \exp(-\beta V(q)) d\sigma_{\Sigma, g}(q).$$

## 6 Summary and conclusions

We have developed various expressions for the stationary or equilibrium measures of holonomically constrained Langevin dynamics, both in phase and configuration spaces. Here by holonomically constrained Langevin dynamics we mean  $n$ -dimensional Lagrangian dynamics to which are appended both  $d$  independent holonomic constraints  $f(q) = 0$ ,  $f = (f_1, \dots, f_d)$ , and dissipative and stochastic terms, which are assumed to be related through the standard fluctuation-dissipation relation. We assume that it is not feasible to construct explicitly  $n-d$  generalized coordinates leading to an unconstrained Langevin formulation, for otherwise the ambient space approach we follow is moot.

The natural route to consider the statistical mechanics of such systems is via a Hamiltonian form of the Lagrangian dynamics, and we have explained that in the presence of holonomic constraints, and whether or not the Langevin fluctuation-dissipation terms are present or not, there is freedom in how the Hamiltonian part of the dynamics can be constructed. The first possibility is that a Legendre transform may be performed in the standard way, specifically independent of the holonomic constraints and their associated Lagrange multipliers, by the introduction of the standard momentum as conjugate variable. A second, less standard approach, the impetus-striction formulation, is to introduce an alternative conjugate variable, the impetus, which does involve the time-differentiated constraints and their Lagrange multipliers, or strictions. (As discussed in [10] there is actually a continuous family of possible impetus-striction formulations, but we only consider the simplest case here.) It is explained that in either case the appropriate multipliers or strictions can be evaluated as functions of the phase variables, which leads to an ambient space dynamics in  $2n$ -dimensional phase space. In the standard formulation there is a  $2(n-d)$ -dimensional submanifold formed by the zero level set of the constraints and their time-derivatives that is invariant under the dynamics. However the ambient dynamics is non-Hamiltonian off this sub-manifold, and the sub-manifold is itself a repelling set. In contrast in the impetus-striction formulation the ambient space dynamics is Hamiltonian in the full  $2n$ -dimensional phase space, and the  $d$ -dimensional constraints are by construction  $d$  (commuting) integrals of both the conservative Hamiltonian version, where the dissipation and fluctuation terms are dropped, and the full Hamiltonian-Langevin system. Of course integrals are by definition neutrally stable invariant sets.

We have not investigated the properties of numerical implementations of the impetus-striction formulation of ambient space Langevin dynamics. However a numerical impetus-striction simulation of a deterministic, infinite-dimensional (partial differential equation), holonomically constrained, Lagrangian system was described in [3], and the associated numerics were found to be rather stable and robust. On the other hand it remains unclear whether or not the impetus-striction formulation can lead to particularly efficient numerical implementations. To our knowledge very few of the existing numerical schemes for Langevin equations can handle holonomic constraints [18, 21, 31] or general (possibly degenerate) Hamiltonians [27].

With the ambient space dynamics in hand we then turned to consideration of various expressions for the associated equilibrium measures of the constrained Langevin dynamics that allow expectations to be computed. In the standard Hamiltonian formulation this stationary measure reduces to the (anticipated) Gibbs-Boltzmann density  $\exp(-\beta H(q, p))$  evaluated on the invariant set where integration is understood to be decomposed into a (flat) integration along the  $(n-d)$ -dimensional cotangent space fibre for each  $q$ , followed by (an unanticipated) multiplication by the cosine of the angle between the tangent and cotangent spaces at  $q$ , and finally by integration over the base configuration constraint set  $\Sigma$  regarded as a subset of  $\mathbf{R}^n$ . In contrast in

the impetus-striction formulation, the natural stationary measure reduces to the (anticipated) Gibbs-Boltzmann density  $\exp(-\beta\mathcal{H}(q, \xi))$  evaluated on the tangent bundle of  $\Sigma$  where integration is understood to be decomposed into a (flat) integration of the impetus  $\xi$  along the  $(n-d)$ -dimensional tangent space at each  $q$ , followed (with no angle correction term) by integration over the base configuration constraint set  $\Sigma$  regarded as a subset of  $\mathbf{R}^n$ .

Finally we showed that the appropriate configuration space marginal equilibrium distribution implied by (necessarily either of) the Hamilton-Langevin ambient phase space equilibrium distributions involves the Boltzmann density  $\exp(-\beta V(q))$  integrated against the Riemannian surface measure on  $\Sigma$ , where the Riemannian inner product is weighted by the generalized mass matrix  $g$ . We presented various explicit forms of this measure, with the conclusion that except for the very particular case in which  $\det(\nabla f(q)g(q)\nabla f(q)^T) = a \det(\nabla f(q)\nabla f(q)^T)$  for some constant  $a > 0$ , the constrained configuration space stationary measure does not in fact coincide with the simplest possible conjecture, namely the Boltzmann weight  $\exp(-\beta V(q))$  integrated over the constraint set  $\Sigma$  regarded as a subset of  $\mathbf{R}^n$ .

It is a pleasure to be able to thank the Editors, and the anonymous referees for their constructive suggestions for revisions to this article. The work of JW and JHM was partially supported by the Swiss National Science Foundation, and that of CH by the DFG Research Center MATHEON “Mathematics for Key Technologies” in Berlin.

## Appendix A: Some useful equalities

We provide a few useful identities that are used throughout the text. First of all, recall that the oblique projection  $\mathcal{P}$  that was introduced in Section 2.1 is orthogonal with respect to the metric  $\langle \cdot, \cdot \rangle_{g^{-1}}$  and projects onto the constrained cotangent space  $T_q^*\Sigma$ . It will be convenient to have a more explicit representation for the projection matrix in terms of a basis of  $T_q^*\Sigma$ . To this end let  $\{\mathbf{t}_1, \dots, \mathbf{t}_{n-d}\}$  be an arbitrary basis of the tangent space  $T_q\Sigma$  and define the  $(n-d) \times (n-d)$ -matrix  $\hat{g}$  by

$$\hat{g}(q) := [\mathbf{t}_1, \dots, \mathbf{t}_{n-d}]^T g [\mathbf{t}_1, \dots, \mathbf{t}_{n-d}] \quad (\text{A.1})$$

with entries  $\hat{g}_{ij} = \mathbf{t}_i^T g \mathbf{t}_j$ . For brevity, we use the notation  $\mathbf{t} = [\mathbf{t}_1, \dots, \mathbf{t}_{n-d}]$ . The projector  $\mathcal{P}$  now can be expressed as

$$\mathcal{P} = g \mathbf{t} \hat{g}^{-1} \mathbf{t}^T. \quad (\text{A.2})$$

The representation (A.2) shows that  $\ker(\mathcal{P}) = N_q\Sigma$  and that we can regard the projector  $\mathcal{P}$  as a compound mapping: first, a vector  $x \in \mathbf{R}^n$  is orthogonally projected onto the tangent space  $T_q\Sigma$ , then the tangent vector  $Px$  is mapped one-to-one onto the cotangent space  $T_q^*\Sigma$ ; cf. also (2)–(3). Therefore, considering the restriction of  $\mathcal{P}$  to the tangent space  $T_q\Sigma$  we may define its inverse  $\mathcal{P}^{-1} : \mathbf{R}^n \rightarrow T_q\Sigma$  as the  $n \times n$ -matrix

$$\mathcal{P}^{-1} = \mathbf{t}(\mathbf{t}^T \mathbf{t})^{-1} \hat{g}(\mathbf{t}^T \mathbf{t})^{-1} \mathbf{t}^T g^{-1}, \quad (\text{A.3})$$

which, in the case that  $\mathbf{t}$  is orthonormal, simplifies to

$$\mathcal{P}^{-1} = \mathbf{t} \hat{g} \mathbf{t}^T g^{-1}. \quad (\text{A.4})$$

The projection matrix  $P$  that maps onto the tangent space and which was defined in (2) has the following property.

**Lemma 2.** *Let  $P: \mathbf{R}^n \rightarrow T_q \Sigma$  be as defined in (2). Then, we have for  $k = 1, \dots, n$*

$$P_{jk} \frac{\partial P_{ij}}{\partial q_i} = P_{kj} \frac{\partial \ln \sqrt{\det G}}{\partial q_j}.$$

The proof follows directly from the Jacobi identity

$$(\det A)' = \det A \cdot \text{tr}(A^{-1}A'),$$

for an invertible matrix  $A = A(s)$  where the  $'$  denotes differentiation with respect to the parameter  $s \in \mathbf{R}$ .

We also make use of a simplified version of the divergence theorem on manifolds: Define a Riemannian submanifold  $\mathcal{M} \subset \mathbf{R}^n$  that is compact (i.e., has no boundary) and that is equipped with surface element  $d\sigma_{\mathcal{M}}$ . Further let  $P_{\mathcal{M}}$  denote the orthogonal projection onto corresponding tangent spaces, and let  $X: \mathcal{M} \rightarrow T\mathcal{M}$  be a smooth vector field tangential to  $\mathcal{M}$ , i.e., we have  $P_{\mathcal{M}}X = X$ . Then

$$\int_{\mathcal{M}} \text{div}_{\mathcal{M}}(X) d\sigma_{\mathcal{M}} = 0, \quad (\text{A.5})$$

where  $\text{div}_{\mathcal{M}}(X) = \text{tr}(P_{\mathcal{M}}\nabla X)$  denotes the surface divergence. In case of non-compact submanifolds, the divergence theorem can be generalized to fast decaying functions, e.g., the  $X_i$  being Schwartz functions. In particular we will use the following two versions for  $\mathcal{M} = \Sigma$  and cotangent spaces  $\mathcal{M} = T_q^*\Sigma$  for fixed  $q \in \Sigma$ :

$$\forall X: T_q^*\Sigma \rightarrow TT_q^*\Sigma \text{ with } P^*X = X : \quad \int_{T_q^*\Sigma} \text{div}_{T_q^*\Sigma}(X) d\lambda_q^* = 0, \quad (\text{A.6})$$

and

$$\forall X: \Sigma \rightarrow T\Sigma \text{ with } PX = X : \quad \int_{\Sigma} \text{div}_{\Sigma}(X) d\sigma_{\Sigma} = 0. \quad (\text{A.7})$$

Here the surfaces divergences are given by

$$\text{div}_{T_q^*\Sigma}(X) = \text{tr}(P^*(q)\nabla X(p)), \quad \text{div}_{\Sigma}(X) = \text{tr}(P(q)\nabla X(q)).$$

## Appendix B: Stationarity of the constrained canonical distribution

This section is devoted to the proof of Proposition 2 using the Kolmogorov backward equation. To this end let the function  $w \in C^{2,1}(\mathcal{Q} \times \mathbf{R}^n, \mathbf{R})$  be the solution of the backward Kolmogorov equation

$$\left( \frac{\partial}{\partial t} - L \right) w = 0, \quad w(\cdot, \cdot, 0) = \phi$$

where

$$L = \frac{1}{2} \mathcal{P} \sigma \sigma^T \mathcal{P}^T : \nabla_p^2 + \nabla_p H \cdot \nabla_q - (\mathcal{P} \nabla_q H + \mathcal{P} \gamma \nabla_p H + \mathbb{I}) \cdot \nabla_p$$

is the infinitesimal generator of the Langevin dynamics (7). The solution of the backward equation is given by

$$w(q, p, t) = \mathbf{E}[\phi(q(t), p(t)) | (q(0), p(0)) = (q, p)]$$

with  $(q(t), p(t))$  being the solution to (7) for initial conditions  $(q, p) \in \mathcal{A}(0) \cong T^*\Sigma$ . We will call a probability measure  $\mu$  *invariant* under  $(q(t), p(t))$  if

$$\int_{\mathcal{A}} w(q, p, t) d\mu(q, p) = \int_{\mathcal{A}} w(q, p, 0) d\mu(q, p).$$

Taking the time derivative on both sides of the equation and exploiting the fact that  $w$  solves the backward equation, it follows that  $\mu$  is invariant if

$$\int_{\mathcal{A}} Lw d\mu = 0.$$

Thus to prove Proposition 2, it remains to show that

$$\int_{\Sigma} \int_{T_q^*\Sigma} (Lw) \cos \angle\{N_q\Sigma, N_q^*\Sigma\} \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma} = 0.$$

To abbreviate notation we set

$$\rho(q) = \angle\{N_q\Sigma, N_q^*\Sigma\}.$$

We split the contributions in the integrand as follows:

$$\begin{aligned} & \int_{\Sigma} \int_{T_q^*\Sigma} (Lw) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma} \\ &= \underbrace{\int_{\Sigma} \int_{T_q^*\Sigma} \left( \frac{1}{2} \mathcal{P} \sigma \sigma^T \mathcal{P}^T : \nabla_p^2 w - \mathcal{P} \gamma \nabla_p H \cdot \nabla_p w \right) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma}}_{\text{forcing and dissipation}} \\ &+ \underbrace{\int_{\Sigma} \int_{T_q^*\Sigma} (\nabla_p H \cdot \nabla_q w - (\mathcal{P} \nabla_q H + \mathbb{I}) \cdot \nabla_p w) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma}}_{\text{constrained Hamiltonian dynamics}}. \end{aligned}$$

We now address the two terms separately. The first term represents the stochastic part of the Langevin equation for  $\gamma, \sigma \neq 0$ . Keeping  $q$  fixed, we can regard it as a degenerate Ornstein-Uhlenbeck process in  $p$  that is known to have a stationary distribution that is Gaussian with a possibly singular covariance matrix [28]. The second integrand represents the constrained Liouville operator corresponding to Hamiltonian dynamics that is subject to a holonomic constraint  $f(q) = 0$ .

### B.1 Forcing and dissipation part

We first consider the forcing and dissipation integral. Using  $\nabla f \nabla_p H = 0$  for all  $p \in T_q^*\Sigma$ , we observe that  $\mathcal{P}^T \nabla_p H = \nabla_p H$  when  $p \in T_q^*\Sigma$ . Hence the second term in the integrand becomes

$$\begin{aligned} & - \int_{\Sigma} \int_{T_q^*\Sigma} (\mathcal{P} \gamma \nabla_p H \cdot \nabla_p w) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma} \\ &= \frac{1}{\beta} \int_{\Sigma} \int_{T_q^*\Sigma} (\mathcal{P} \gamma \mathcal{P}^T [\nabla_p \exp(-\beta H)] \cdot \nabla_p w) \cos \rho d\lambda_q^* d\sigma_{\Sigma} \\ &= \frac{1}{\beta} \int_{\Sigma} \int_{T_q^*\Sigma} \text{tr} (P^* \nabla_p \cdot [\mathcal{P} \gamma \mathcal{P}^T (\nabla_p w) \exp(-\beta H)]) \cos \rho d\lambda_q^* d\sigma_{\Sigma} \\ & - \int_{\Sigma} \int_{T_q^*\Sigma} \left( \frac{1}{2} \mathcal{P} \sigma \sigma^T \mathcal{P}^T : \nabla_p^2 w \right) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma} \end{aligned}$$

where, in the second equality, we used  $P^* \mathcal{P} = \mathcal{P}$  and the fluctuation-dissipation relation  $2\gamma = \beta \sigma \sigma^T$ . But the second term in the last equality annihilates the fluctuation terms in the forcing and dissipation integral, while the first integral vanishes by application of the divergence theorem (A.6). Hence it follows that

$$\int_{\Sigma} \int_{T_q^* \Sigma} \left( \frac{1}{2} \mathcal{P} \sigma \sigma^T \mathcal{P}^T : \nabla_p^2 w - \mathcal{P} \gamma \nabla_p H \cdot \nabla_p w \right) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma} = 0.$$

## B.2 Constrained Hamiltonian part

This second part is slightly more involved (cf. [15]). We demonstrate that

$$\int_{\Sigma} \int_{T_q^* \Sigma} (\nabla_p H \cdot \nabla_q w - (\mathcal{P} \nabla_q H + \mathbb{I}) \cdot \nabla_p w) \cos \rho \exp(-\beta H) d\lambda_q^* d\sigma_{\Sigma} = 0. \quad (\text{B.8})$$

For this purpose we switch to a locally adapted coordinate system on  $T_q^* \Sigma \oplus N_q^* \Sigma$  for any fixed  $q \in \Sigma$ . Although the integration over the fibres in (B.8) does not involve the normal part of  $p$ , including it is necessary, for the integrand involves derivatives with respect to  $p \in \mathbf{R}^n$ . We introduce orthonormal bases  $\{\mathbf{t}_1^*, \dots, \mathbf{t}_{n-d}^*\}$  and  $\{\mathbf{n}_1^*, \dots, \mathbf{n}_d^*\}$  of  $T_q^* \Sigma$  and  $N_q^* \Sigma$ , respectively, to define the orthogonal  $n \times n$  matrix  $R = R(q)$  by

$$R = [\mathbf{t}^*, \mathbf{n}^*], \quad \mathbf{t}^* = [\mathbf{t}_1^*, \dots, \mathbf{t}_{n-d}^*], \quad \mathbf{n}^* = [\mathbf{n}_1^*, \dots, \mathbf{n}_d^*].$$

An adapted coordinate chart is now defined by the map

$$\theta: \mathbf{R}^{n-d} \times \mathbf{R}^d \rightarrow T_q^* \Sigma \oplus N_q^* \Sigma, \quad (u, \zeta) \mapsto \mathbf{t}^* u + \mathbf{n}^* \zeta.$$

We use the shorthand  $\mathbf{R}_0^{n-d} := \mathbf{R}^{n-d} \times \{0\}$  with  $\{0\} \subset \mathbf{R}^d$  for vectors

$$Rv \in T_q^* \Sigma \iff v = (u, 0) \in \mathbf{R}_0^{n-d}.$$

In order to avoid confusion in taking derivatives, we introduce the following notation: for a function  $a: \mathbf{R}^n \rightarrow \mathbf{R}$ ,  $p \mapsto a(p)$  we write  $\nabla a(p) = R \nabla_v a(Rv)$ , whereas for functions  $b: \mathcal{Q} \times \mathbf{R}^n \rightarrow \mathbf{R}$ ,  $(q, p) \mapsto b(q, p)$  we write

$$\nabla_q b(q, p) = \nabla_1 b(q, p) + [\nabla_q R(q)v]^T R(q) \nabla_v b(q, p) \quad (\text{B.9})$$

with  $\nabla_1$  denoting the derivative with respect to the first argument.

We now prove (B.8). We start with the first term in the integrand and use (B.9):

$$\begin{aligned} & \int_{\Sigma} \int_{T_q^* \Sigma} (\nabla_p H \cdot \nabla_q w) \cos \rho e^{-\beta H(q,p)} d\lambda_q^*(p) d\sigma_{\Sigma}(q) \quad (\text{here } \nabla_q = \nabla_1) \\ &= \underbrace{\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R \nabla_v H \cdot \nabla_q w) \cos \rho e^{-\beta H} dv d\sigma_{\Sigma}}_{=: \mathbf{A}_1} \\ & \quad - \underbrace{\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R \nabla_v H \cdot (\nabla_q [Rv])^T R \nabla_v w) \cos \rho e^{-\beta H} dv d\sigma_{\Sigma}}_{=: \mathbf{A}_2}, \end{aligned}$$

and, addressing these two terms separately,

$$\begin{aligned} \mathbf{A}_1 = & \underbrace{\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} ((R\nabla_v H \cdot \nabla_q [w \cos \rho e^{-\beta H}]) - (R\nabla_v H \cdot \nabla_q \cos \rho) w e^{-\beta H}) dv d\sigma_{\Sigma}}_{=: \mathbf{C}_1} \\ & + \beta \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R\nabla_v H \cdot \nabla_q H) w \cos \rho e^{-\beta H} dv d\sigma_{\Sigma}, [-3pt] \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}_2 = & - \underbrace{\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P}\nabla_q [Rv] R\nabla_v H \cdot R\nabla_v w) \cos \rho e^{-\beta H} dv d\sigma_{\Sigma}}_{=: \mathbf{C}_2} \\ & + \int_{\Sigma} \int_{T_q^* \Sigma} (\lambda^n \nabla f_n \cdot \nabla_p w) \cos \rho e^{-\beta H(q,p)} d\lambda_q^*(p) d\sigma_{\Sigma} \end{aligned}$$

where  $\lambda^n = \lambda^n(q, p)$  denotes the Lagrange multiplier from the constraint force (not to be confused with the Lebesgue measure  $\lambda_q^*$ ). Substitution of the constraint force in last integral can be readily verified by using (A.2) with  $p = Rv$ :

$$(I - \mathcal{P})\nabla_q [Rv] R\nabla_v H = K^{km} \langle \nabla_q [Rv] g^{-1} Rv, g^{-1} \nabla f_k \rangle \nabla f_m = -\lambda^n(q, p) \nabla f_n.$$

We proceed by calculating the second integral in (B.8). Again we use (B.9) and get

$$\begin{aligned} & - \int_{\Sigma} \int_{T_q^* \Sigma} (\mathcal{P}\nabla_q H \cdot \nabla_p w) \cos \rho e^{-\beta H(q,p)} d\lambda_q^*(p) d\sigma_{\Sigma}(q) \quad (\text{here } \nabla_q = \nabla_1) \\ & = - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} ((\mathcal{P}\nabla_q H - (\nabla_q [Rv])^T R\nabla_v H) \cdot R\nabla_v w) \cos \rho e^{-\beta H} dv d\sigma_{\Sigma} \\ & = - \underbrace{\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P}\nabla_q H \cdot R\nabla_v (w e^{-\beta H})) \cos \rho dv d\sigma_{\Sigma}}_{=: \mathbf{B}_1} \\ & \quad - \beta \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P}\nabla_q H \cdot R\nabla_v H) w \cos \rho e^{-\beta H} dv d\sigma_{\Sigma} \\ & \quad + \underbrace{\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P}(\nabla_q [Rv])^T R\nabla_v H \cdot R\nabla_v w) \cos \rho e^{-\beta H} dv d\sigma_{\Sigma}}_{=: \mathbf{B}_2}. \end{aligned}$$

We observe that the second integral on the r.h.s. of the above equation cancels with the second integral of the term  $\mathbf{A}_1$ . Moreover the third integrand countervails the second integral of the term  $\mathbf{A}_2$ . Reassembling the results so far, (B.8) reads

$$\mathbf{B}_1 + \mathbf{B}_2 + \mathbf{C}_1 + \mathbf{C}_2 = 0. \quad (\text{B.10})$$

In the next step we evaluate the terms  $\mathbf{B}_2$  and  $\mathbf{C}_2$  both of which contain the Jacobian  $\nabla_q [R(q)v]$  as a part of their integrands. This gives

$$\mathbf{B}_2 + \mathbf{C}_2 = \mathbf{B}'_2 + \mathbf{C}'_2 \quad (\text{B.11})$$

with

$$\begin{aligned} \mathbf{B}'_2 & = \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P}(\nabla_q [Rv])^T R\nabla_v H \cdot R\nabla_v (w e^{-\beta H})) \cos \rho dv d\sigma_{\Sigma} \\ \mathbf{C}'_2 & = - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P}\nabla_q [Rv]^T R\nabla_v H \cdot R\nabla_v (w e^{-\beta H})) \cos \rho dv d\sigma_{\Sigma}, \end{aligned}$$

so that (B.10) can be equivalently expressed as

$$\mathbf{B}_1 + \mathbf{B}'_2 + \mathbf{C}_1 + \mathbf{C}'_2 = 0.$$

We turn to the computation of  $B'_2$  and  $C'_2$  and get after some simple manipulations

$$\begin{aligned} \mathbf{B}'_2 &= - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} R_{ij} [\mathcal{P}]_{im} \partial_{v_j} \left[ \frac{\partial R_{sl}}{\partial q_m} v_l g^{so} R_{on} v_n \right] w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &= - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} R_{ij} [\mathcal{P}]_{im} \frac{\partial R_{sj}}{\partial q_m} R_{sn} \partial_{v_n} H w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} R_{ij} [\mathcal{P}]_{im} \frac{\partial R_{sl}}{\partial q_m} v_l g^{so} R_{oj} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &= - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} [\mathcal{P}]_{im} R_{ij} \frac{\partial^2 H}{\partial q_m \partial v_j} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad + \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} [\mathcal{P}]_{im} \frac{\partial (g^{in} R_{no} v_o)}{\partial q_m} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} g^{si} [\mathcal{P}]_{im} \frac{\partial R_{sl}}{\partial q_m} v_l w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \end{aligned} \tag{B.12}$$

while the other integral becomes

$$\begin{aligned} \mathbf{C}'_2 &= - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\mathcal{P} \nabla_q [Rv] R \nabla_v H \cdot R \nabla_v (w e^{-\beta H})) \cos \rho \, dv d\sigma_{\Sigma} \\ &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} R_{ij} [\mathcal{P}]_{im} \partial_{v_j} \left[ \frac{\partial R_{ml}}{\partial q_s} v_l g^{so} R_{on} v_n \right] w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} R_{ij} [\mathcal{P}]_{im} \frac{\partial R_{mj}}{\partial q_s} g^{so} R_{on} v_n w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad + \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} [\mathcal{P}]_{im} \frac{\partial R_{ml}}{\partial q_s} v_l g^{si} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &= - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} R_{ij} \frac{\partial R_{mj}}{\partial q_s} g^{so} R_{on} v_n w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad + \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} g^{si} [\mathcal{P}]_{im} \frac{\partial R_{ml}}{\partial q_s} v_l w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}. \end{aligned} \tag{B.13}$$

Exploiting that  $g^{-1} \mathcal{P} = \mathcal{P}^T g^{-1}$ , we observe that the last integral cancels with the last integral in (B.12). Moreover the first integral of (B.12) cancels with  $\mathbf{B}_1$ . Hence

$$\begin{aligned} \mathbf{B}_1 + \mathbf{B}'_2 + \mathbf{C}'_2 + \mathbf{C}_1 &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} [\mathcal{P}]_{im} \frac{\partial (g^{in} R_{no} v_o)}{\partial q_m} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} R_{ij} \frac{\partial R_{mj}}{\partial q_s} g^{so} R_{on} v_n w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ &\quad + \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R \nabla_v H \cdot \nabla_q [w \cos \rho e^{-\beta H}]) \, dv \, d\sigma_{\Sigma} \\ &\quad - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R \nabla_v H \cdot \nabla_q \cos \rho) w e^{-\beta H} \, dv d\sigma_{\Sigma}. \end{aligned} \tag{B.14}$$

Now, the first integral on the r.h.s. can be expressed as follows:

$$\begin{aligned}
 & 7 \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} [\mathcal{P}]_{im} \frac{\partial(g^{in} R_{no} v_o)}{\partial q_m} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} P_{im} \frac{\partial(R_{mn} \partial_{v_n} H)}{\partial q_i} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &+ \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (I - P)_{im} \frac{\partial(R_{mn} \partial_{v_n} H)}{\partial q_i} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &- \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} \frac{\partial(g^{is} R_{so} v_o)}{\partial q_m} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}
 \end{aligned}$$

where the first integral is further manipulated according to

$$\begin{aligned}
 & \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} P_{im} \frac{\partial(R_{mn} \partial_{v_n} H)}{\partial q_i} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} \text{tr} (P \nabla_q [R \nabla_v H \cdot w \cos \rho e^{-\beta H}]) \, dv d\sigma_{\Sigma} \\
 &- \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (P R \nabla_v H \cdot \nabla_q [w \cos \rho e^{-\beta H}]) \, dv d\sigma_{\Sigma}.
 \end{aligned}$$

The first integral after the equality is seen to be zero by applying the divergence theorem, equation (A.7); the second one cancels with the second one from below in (B.14). Collecting all the terms, in order that (B.10) be fulfilled, we need that

$$\begin{aligned}
 & \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R \nabla_v H \cdot \nabla_q [w e^{-\beta H}]) \cos \rho \, dv d\sigma_{\Sigma} \\
 &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (I - P)_{im} \frac{\partial(R_{mn} \partial_{v_n} H)}{\partial q_i} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &- \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} \frac{\partial(g^{is} R_{so} v_o)}{\partial q_m} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &- \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} R_{ij} \frac{\partial R_{mj}}{\partial q_s} g^{so} R_{on} v_n w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}
 \end{aligned} \tag{B.15}$$

We address each term on the r.h.s. separately. For the first integral, we use Lemma 2 together with  $R \nabla_v H = P R \nabla_v H$  to get

$$\begin{aligned}
 & \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (I - P)_{im} \frac{\partial(R_{mn} \partial_{v_n} H)}{\partial q_i} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (\nabla_q \ln((\det G)^{-1/2}) \cdot P R \nabla_v H) w \cos \rho \, dv d\sigma_{\Sigma}.
 \end{aligned} \tag{B.16}$$

The second integral of (B.15) is changed with  $[\nabla f_1, \dots, \nabla f_d]^T g^{-1} R v = 0$  to

$$\begin{aligned}
 & - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} \frac{\partial(g^{is} R_{so} v_o)}{\partial q_m} w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\
 &= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} g^{ms} \partial_s f_n \partial_{im}^2 f_k g^{io} R_{ol} v_l w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}.
 \end{aligned} \tag{B.17}$$

Further manipulating the third integral on the r.h.s. of (B.15) yields

$$\begin{aligned} & - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} R_{ij} \frac{\partial R_{mj}}{\partial q_s} g^{so} R_{on} v_n w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ & = - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} \frac{\partial [K^{kn} (\nabla f_k \otimes g^{-1} \nabla f_n) R]_{ij}}{\partial q_s} R_{ij} g^{so} R_{ol} v_l w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}. \end{aligned} \quad (\text{B.18})$$

To further simplify (B.18) we recall that  $R$  is an orthogonal matrix given by  $R = [\mathbf{t}^*, \mathbf{n}^*]$  and note that for some invertible matrix  $\zeta \in \mathbf{R}^{d \times d}$

$$\mathbf{n}^* := [\mathbf{n}_1^*, \dots, \mathbf{n}_d^*] = g^{-1} [\nabla f_1, \dots, \nabla f_d] \zeta$$

such that

$$\zeta^{-1} = [\mathbf{n}_1^*, \dots, \mathbf{n}_d^*]^T g^{-1} [\nabla f_1, \dots, \nabla f_d],$$

by which (B.18) becomes

$$\begin{aligned} & - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} [\nabla f_k \otimes g^{-1} \nabla f_n]_{im} R_{ij} \frac{\partial R_{mj}}{\partial q_s} g^{so} R_{on} v_n w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ & = \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} \frac{\partial (K^{kn} \zeta^{jn})}{\partial q_s} \langle \nabla f_k, \mathbf{n}_j^* \rangle g^{so} R_{ol} v_l w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma} \\ & - \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} K^{kn} \zeta^{jn} \partial_{is}^2 f_k g^{ir} \nabla_r f_m \zeta_{mj} g^{so} R_{ol} v_l w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}. \end{aligned} \quad (\text{B.19})$$

A careful inspection now reveals that the last integral cancels with (B.17). The remaining terms can be further simplified by using that

$$\langle \nabla f_k, \mathbf{n}_j^* \rangle = \zeta_{mj} \langle \nabla f_k, g^{-1} \nabla f_m \rangle = \zeta_{mj} K_{mk}$$

which, upon collecting all the remaining terms in (B.16) and (B.19), and applying Lemma 2 to the second line of (B.19), shows that (B.15) reduces to

$$\int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} (R \nabla_v H \cdot \nabla_q \cos \rho(q)) w e^{-\beta H} \, dv d\sigma_{\Sigma} \quad (\text{B.20})$$

$$= \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} \left( \nabla_q \ln((\det G)^{-1/2}) \cdot PR \nabla_v H \right) w \cos \rho \, dv d\sigma_{\Sigma} \quad (\text{B.21})$$

$$+ \int_{\Sigma} \int_{\mathbf{R}_0^{n-d}} \left( \nabla_q \ln(\det(K\zeta)) \cdot PR \nabla_v H \right) w e^{-\beta H} \cos \rho \, dv d\sigma_{\Sigma}. \quad (\text{B.22})$$

The equation obviously is satisfied if

$$\nabla_q \ln(\cos \rho) = \nabla_q \ln \left( \frac{\det(K\zeta)}{\sqrt{\det G}} \right).$$

But (11) implies that

$$\cos \rho = \left( \frac{\det K}{\sqrt{\det G \det J}} \right),$$

so that the above assertion follows upon noting that

$$\zeta^{-T} \zeta^{-1} = [\nabla f_1, \dots, \nabla f_d] g^{-2} [\nabla f_1, \dots, \nabla f_d] \Rightarrow \det(\zeta) = \frac{1}{\sqrt{\det(J)}}.$$

Hence Proposition 2 is proved.

## References

1. E.A. Carter, G. Ciccotti, J.T. Hynes, R. Kapral, *Chem. Phys. Lett.* **156**, 472 (1989)
2. G. Ciccotti, R. Kapral, E. Vanden-Eijnden, *Chem. Phys. Chem.* **6**, 1809 (2005)
3. D. Dichmann, J.H. Maddocks, *Nonlinear Science* **6**, 271 (1996)
4. D. Dichmann, J.H. Maddocks, R. Pego, *Arch. Rational Mech. Anal.* **135**, 357 (1996)
5. P. Dirac, Generalized Hamiltonian dynamics, *Can. J. Math.* **2**, 129 (1950)
6. P. Dirac, Generalized Hamiltonian dynamics, *Proc. R. Soc. A* **246**, 326 (1958)
7. W. E, E. Vanden-Eijnden, in *Multiscale, Modelling, and Simulation*, edited by S. Attinger, P. Koumoutsakos (Springer, Berlin, 2004), p. 35
8. L. Evans, R. Gariepy, *Measure Theory and Fine Properties of Functions* (CRC Press, 1992)
9. H. Federer, *Geometric Measure Theory* (Springer, 1969)
10. O. Gonzalez, J.H. Maddocks, R. Pego, *Arch. Rational Mech. Anal.* **157**, 285 (2001)
11. C. Hartmann, Ph.D. thesis, Free University Berlin, 2007
12. C. Hartmann, J. Latorre, G. Ciccotti, *Eur. Phys. J. Special Topics* **200**, 79 (2011)
13. C. Hartmann, J. Latorre, C. Schütte, *Proc. Comp. Sci.* **1**, 1591 (2010)
14. C. Hartmann, C. Schütte, *ZAMM* **85**, 700 (2005)
15. C. Hartmann, C. Schütte, G. Ciccotti, *J. Chem. Phys.* **132**, 111103 (2010)
16. H. Hotelling, *Biometrika* **28**, 321 (1936)
17. C. Ipsen, C. Meyer, *Am. Math. Mon.* **102**, 904 (1995)
18. B. Kallemov, G. H. Miller, *SIAM J. Sci. Comput.* **33**, 653 (2011)
19. R. Kupferman, *J. Stat. Phys.* **114**, 291 (2004)
20. T. Lelièvre, M. Rousset, G. Stoltz., *Free Energy Computations: A Mathematical Perspective* (Imperial College Press, 2010)
21. T. Lelièvre, M. Rousset, G. Stoltz, *Math. Comp.* (to appear) (2011)
22. J.H. Maddocks, R. Pego, *Commun. Math. Phys.* **170**, 207 (1995)
23. J. Marsden, T. Ratiu, *Introduction to Mechanics and Symmetry* (Springer, New York, 1999)
24. J. Mattingly, A. Stuart, *Markov Process. Related Fields* **8**, 199 (2001)
25. J. Mattingly, A. Stuart, D. Higham, *Stochastic Process. Appl.* **101**, 185 (2002)
26. J. Miao, A. Ben-Israel, *Linear Algebra Appl.* **171**, 81 (1992)
27. G. Milstein, M. Tretyakov, *IMA J. Numer. Anal.* **23**, 593 (2003)
28. B. Øksendal, *Stochastic Differential Equations*, 5th edn. (Springer, Berlin, Heidelberg, 1998)
29. M. Tuckerman, Y. Liu, G. Ciccotti, G. Martyna, *J. Chem. Phys.* **115**, 1678 (2001)
30. W.F. van Gunsteren, H.J.C. Berendsen, J.A.C. Rullmann, *Molec. Phys.* **44**, 69 (1981)
31. E. Vanden-Eijnden, G. Ciccotti, *Chem. Phys. Lett.* **429**, 310 (2006)
32. J. Walter, O. Gonzalez, J.H. Maddocks, *Multisc. Model. Simul.* **8**, 1018 (2010)
33. P. Wedin, in *Matrix Pencils*, Vol. 973 of *Lecture Notes in Mathematics*, edited by B. Kågström, A. Ruhe (Springer, Berlin Heidelberg, 1983), p. 263