4. Phase space of the fast variables

We have addressed the problem of the deriving simplified equations of motion for a given reaction coordinate in great detail. Yet the question of how to compute the coefficients and parameters of the reduced model (e.g., the free energy) has remained open. All of the reduced models depend on quantities that are averaged over the fast variables. Hence it is important to study the statistical properties of the fast variables, conditional on the particular value of the reaction coordinate. Especially we are going to explain how the conditional averages over the fast variables can be computed in practice.

4.1. Excursus: constrained mechanical systems

In this section we shall briefly discuss the properties of mechanical systems subject to holonomic constraints. In treating constraints it is most convenient to start within the framework of Lagrangian mechanics. For our purposes it suffices to define a holonomic constraint by specifying a submanifold $\Sigma \subset \mathbf{R}^n$ of the configuration space. Together with the natural inclusion $T\Sigma \subset T\mathbf{R}^n$ this determines the state space of the constrained system. Suppose that $\Sigma = \varphi^{-1}(0)$ is determined as the level set of a smooth function $\varphi : \mathbf{R}^n \to \mathbf{R}^s$. If the Jacobian $\mathbf{D}\varphi(q)$ has maximum rank on Σ, then Σ is a proper submanifold of codimension s in $\mathbf{R}^n$. The tangent space to $q \in \Sigma$ is then defined in the usual way considering the direction of curves in Σ which is equivalently expressed as [182, 243]

$$T_q\Sigma = \{v \in T_q\mathbf{R}^n \,|\, \mathbf{D}\varphi(q) \cdot v = 0\} \,.$$

For the sake of simplicity, we assume that Σ has codimension $s = 1$ in $\mathbf{R}^n$. For a much more general discussion of holonomic constraints the interested reader is referred to the textbook [81]; a good introduction into the geometry of submanifolds is [183].

We can now easily define a constrained Lagrangian by either restricting the original Lagrangian to the constrained tangent bundle $T\Sigma \subset T\mathbf{R}^n$, or to use the Lagrange Multiplier Theorem [97] to define an *augmented* Lagrangian,

$$\hat{L}(q, \dot{q}, \lambda) = L(q, \dot{q}) - \lambda\varphi(q) \,.$$

Note that the thus defined Lagrangian $\hat{L} : T\mathbf{R}^{n+1} \to \mathbf{R}$ is not regular as a function of q and λ, for it does not contain the velocity $d\lambda/dt$. Hence defining a Hamiltonian makes no sense at the moment. Nevertheless by minimizing the action integral for $\hat{L}$

$$\int_a^b \left(L(q(t), \dot{q}(t)) - \lambda(t)\varphi(q(t))\right)\, dt \,,$$

where the endpoints $q(a)$ and $q(b)$ both satisfy the constraint, we obtain the Euler-Lagrange equations in the unknowns q and λ,

$$\begin{aligned} \frac{d}{dt}\frac{\partial \hat{L}}{\partial \dot{q}^i} &= \frac{\partial \hat{L}}{\partial q^i} \\ 0 \qquad &= \frac{\partial \hat{L}}{\partial \lambda} \,. \end{aligned} \tag{4.1}$$

Evidently the second equation is simply the constraint $\varphi(q) = 0$. Accordingly the Euler-Lagrange equations form a differential-algebraic system which is of differential index three [244, 245]. The alternative method by restricting the original Lagrangian to $T\Sigma$ amounts to endowing Σ with an appropriate set of local coordinates $(x^1, \ldots, x^{n-1})$,

writing up the Lagrangian in these coordinates, and deriving local Euler-Lagrange equations. These will be then of the form (2.4). According to the theorem on Lagrange multipliers the local Euler-Lagrange equations are equivalent to the equations (4.1). We refer to the latter as *ambient space* formulation. It is by far the most common approach in molecular dynamics, for the equations can be discretized by standard numerical schemes [105, 246].

For related approaches the interested reader may consult the seminal work of Dirac [247], or constrained formulations using vakonomic mechanics [20]. A different method that is more in the spirit of index reduction techniques is treated in [248].

4.1.1. Geometric considerations Physically speaking, constraining a particle to a submanifold is achieved by (i) adding a constraining force $-\lambda\nabla\varphi$ to the original equations and (ii) imposing the condition $\varphi(q) = 0$. The more familiar constrained Newtonian equations read

$$M\ddot{q} + \nabla V(q) + \lambda\nabla\varphi(q) = 0\,, \quad \varphi(q) = 0\,.$$

Here the symbol ∇ is just a shorthand for $\nabla = (\partial/\partial q^1, \ldots, \partial/\partial q^n)^T$. For the sake of simplicity we set $M = \mathbf{1}$ and identify tangent and cotangent space in what follows.

We shall take a closer look at the origin of the constraining force. To this end we consider a curve $q(t)$ which is an integral curve of the constrained equations of motion. Let $n(q)$ be the unit normal to the constraint surface Σ. The tangent vectors $\dot{q}(t)$ then satisfy at all times t the orthogonality condition

$$\langle n(q), \dot{q}\rangle = 0\,,$$

where we have omitted the curve parameter t. Differentiation with respect to t yields

$$\langle \nabla n(q)\cdot\dot{q}, \dot{q}\rangle + \langle n(q), \ddot{q}\rangle = 0\,.$$

By assumption $q(t)$ is a solution of the constrained equations of motion. Hence we can insert the Newtonian equations into the last equation and solve for λ. This gives us the Lagrange multiplier $\lambda(t) = \lambda(q(t), \dot{q}(t))$, and thus the constraint force

$$-\lambda\nabla\varphi(q) = (\langle n(q), \nabla V(q)\rangle - \langle \nabla n(q)\cdot\dot{q}, \dot{q}\rangle)\, n(q)\,, \tag{4.2}$$

where $(q, \dot{q}) \in T\Sigma$. The last equation already reveals the mechanism of constraining a particle: Firstly, we define the point-wise projection onto the normal space to Σ,

$$P_N : (T\mathbf{R}^n)|\Sigma \to (T\Sigma)^\perp\,, \quad X \mapsto \langle n(q), X\rangle\, n(q)\,.$$

The contribution of the potential to the constraint force is easily identified as $P_N\nabla V$ which is the projection of the force field along the normal direction. This is physically intuitive, and accordingly the force that *intrinsically acts on the constrained particle* due to the potential is given by the tangential force $-P_T\nabla V$, where $P_T = \mathbf{1} - P_N$ denotes the projection onto $T\Sigma$. For the remaining part we shall prove:

Lemma 4.1. *Without loss of generality we set $V \equiv 0$. Then the constraint force $-\lambda\nabla\varphi$ is given by the second fundamental form of the embedding $\Sigma \subset \mathbf{R}^n$.*

Proof. Consider the unit normal $n \in \mathbf{R}^n$ as a map $n : \Sigma \to S^{n-1}$ which sends a point $q \in \Sigma$ to the unit sphere S^{n-1} (Gauss map). The second fundamental form is explained as the symmetric bilinear form $I\!I : T_q\Sigma \times T_q\Sigma \to \mathbf{R}$ that is defined by

$$I\!I(X, Y) = \langle \mathfrak{S}(q)\cdot X, Y\rangle\,, \quad \mathfrak{S}(q) = -P_T\nabla n(q)\,.$$

The map $\mathfrak{S} : T_q\Sigma \to T_q\Sigma$ is called the Weingarten map; in codimension one it is simply the negative derivative of the Gauss map, for $\nabla n \in T_q\Sigma$. Hence the assertion follows by comparing the last equation to (4.2) upon noting that $\dot{q} \in T_q\Sigma$. □

Remark 4.2. *The calculation of the constraint force for a scalar constraint is very instructive as it reveals the physical mechanism of constraining a particle to a submanifold of its configuration space. However we will also need an expression for the constraint force (and for the Lagrange multiplier) in the case when Σ has codimension $s > 1$. Since $\mathbf{D}\varphi$ has maximum rank s, we can construct an orthonormal frame $\{n_1(q), \ldots, n_s(q)\}$ for all $q \in \Sigma$ simply by orthonormalizing the columns of $\mathbf{D}\varphi$. By repeating the calculation above for each normal vector n_i we obtain*

$$\lambda = -(Q^T \mathbf{D}\varphi)^{-1} \left(\langle \mathfrak{S} \cdot \dot{q}, \dot{q} \rangle + Q^T \nabla V \right) , \tag{4.3}$$

where $(q, \dot{q}) \in T\Sigma$, and the matrix $Q = (n_1, \ldots, n_s) \in \mathbf{R}^{n \times s}$ contains the normal vectors as columns. The components of $\mathfrak{S}$ are the single Weingarten maps

$$\mathfrak{S}_i : T_q\Sigma \to T_q\Sigma , \quad \mathfrak{S}_i(q) = -P_T \nabla n_i(q) \quad (i = 1, \ldots, s) .$$

Here, in contrast to the scalar constraint, it is no longer true that $\nabla n_i \in T_q\Sigma$. But as $\dot{q} \in T_q\Sigma$ in the quadratic expression of (4.3), we can replace ∇n_i by its tangential projection $P_T \nabla n_i$ which then yields the second fundamental form of the embedding. Note that a common representation of λ that is frequently found in the literature is

$$\lambda = (\mathbf{D}\varphi^T \mathbf{D}\varphi)^{-1} \left(\langle \nabla^2 \varphi \cdot \dot{q}, \dot{q} \rangle - \mathbf{D}\varphi^T \nabla V \right) , \tag{4.4}$$

where $\nabla^2 \varphi$ is the Hessian matrix of $\varphi = (\varphi_1, \ldots, \varphi_s)$ that is understood componentwise. Both formulae for the Lagrange multipliers (4.3) and (4.4) are equivalent, which follows from considerations concerning pseudoinverses in the previous section and from the definition of the second fundamental form. In any event the constraint force $-\mathbf{D}\varphi\, \lambda$ is uniquely determined [66]. Comparing the last equations (4.3) and (4.4) to (3.13) and (3.14) suggests that we can compute the derivative of the free energy (3.9) by averaging over the Lagrange multiplier with the augmented potential

$$V_\varphi = V + \beta^{-1} \ln \mathrm{vol} J_\varphi .$$

4.1.2. Constrained Hamiltonian systems The transition from the Lagrangian to the Hamiltonian representation is not straightforward in the presence of constraints, at least in the ambient space formulation. In principle this would not be a problem, if we utilized local coordinates on the surface. Then the local Lagrangian would be regular, provided Σ were a regular hypersurface. Working with the augmented Lagrangian $\hat{L}$ we can formally define the conjugate momentum to q by

$$p^i = \frac{\partial \hat{L}}{\partial \dot{q}^i} .$$

This is the former momentum p, and we can derive a Hamiltonian $\hat{H}$ pretending that $\hat{L}$ is regular, while restricting the Legendre transform to the set defined by

$$0 = \frac{\partial \hat{L}}{\partial \dot{\lambda}} .$$

This yields the Hamiltonian

$$\hat{H}(q, p, \lambda) = \dot{q}^i p_i - \hat{L}(q, \dot{q}, \lambda) = H(q, p) + \lambda \varphi(q) .$$

Clearly this Hamiltonian does not give an equation for λ in the usual way. Therefore the evolution of the Lagrange multiplier is undetermined. Nevertheless, we obtain

equations of motion for the variables q and p,

$$\begin{aligned} \dot{q}^i &= \frac{\partial \hat{H}}{\partial p_i} \\ \dot{p}^i &= -\frac{\partial \hat{H}}{\partial q^i} \\ 0 &= -\frac{\partial \hat{H}}{\partial \lambda}\,, \end{aligned} \tag{4.5}$$

that are equivalent to the Euler-Lagrange equations (4.1) modulo the restriction $\partial \hat{L}/\partial \dot{\lambda} = 0$. Similar to the former Lagrangian formulation on the tangent bundle the dynamics now takes place on the constrained phase space bundle

$$\mathscr{B} = \left\{ (q,p) \in T^*\mathbf{R}^n \,\middle|\, q \in \Sigma \text{ and } \langle \nabla\varphi(q), \mathbf{D}_2 H(q,p) \rangle = 0 \right\}$$

which is the image of the Legendre transform of $(T\mathbf{R}^n)|T\Sigma$ which we can identify with $T^*\Sigma$. Here H is the original (i.e., unconstrained) Hamiltonian, and $\mathbf{D}_2$ denotes the derivative with respect to the second slot. The condition on the momentum is exactly the condition $\dot{\varphi}(q) = 0$, and is referred to as *hidden* constraint; it is hidden because it does not appear explicitly in the equations of motion. Notice that the identification of $\mathscr{B}$ with $T^*\Sigma$ is a rather subtle issue which is related to the non-regularity of the augmented Lagrangian; in general this identification is valid only up to a symplectic diffeomorphism $p \mapsto p + \alpha\nabla\varphi$, where α is chosen such that p satisfies the hidden constraint; see [249, 250] regarding this discussion.

Let $\Phi_t : \mathscr{B} \to \mathscr{B}$ with $\mathscr{B} \cong T^*\Sigma$ be the flow of the equations of motion (4.5). Then it is easy to show that the total energy remains a first integral, $H|_{\mathscr{B}} = H|_{\mathscr{B}} \circ \Phi_t$, where $H|_{\mathscr{B}}$ is the unconstrained Hamiltonian, restricted to $\mathscr{B}$. In fact, for a solution $(q(t), p(t))$ of the constrained equations of motion (4.5), the variation of the total energy along that curve is equal to

$$\frac{d}{dt} H(q(t), p(t)) = -\lambda \frac{\partial \varphi}{\partial q^l} \frac{\partial H}{\partial p_l}$$

which is zero, since $(q(t), p(t))$ is a curve in $\mathscr{B}$, and hence satisfies the hidden constraint. The last equation is quite important from the viewpoint of numerics, since it states that a numerical discretization scheme of the differential-algebraic system (4.5) should take care of the hidden constraint in order to preserve the energy conservation property of the continuous flow [82, 251]. Furthermore it is obvious from the equations of motion that the constrained system is still reversible in time.

Concerning the volume-preservation property or symplecticness there is some disagreement in the molecular dynamics community, for it is often stated that constrained Hamiltonian flows were not volume-preserving [73]. Although agreement on this issue is immediately obtained, if the Hamiltonian is considered in local coordinates on $T^*\Sigma$ which is no different from the standard case in $\mathbf{R}^n$, people disagree upon the ambient space formulation; see, for instance, [79, 72, 80]. Since both approaches are equivalent in the sense that the trajectories coincide, we expect that the ambient space Hamiltonian has the same structural properties as its local counterpart. Indeed, the following can be shown [82].

Lemma 4.3 (Leimkuhler & Reich 2004)**.** *Let the flow $\Phi_t : \mathscr{B} \to \mathscr{B}$, $\mathscr{B} \cong T^*\Sigma$ be the solution of the ambient space Hamiltonian system (4.5), and let $\omega = \Omega|_{\mathscr{B}}$ denote the restriction of the standard symplectic form $\Omega = dq^i \wedge dp_i$ on $T^*\mathbf{R}^n$ to the constrained phase space $\mathscr{B}$. Then Φ_t is symplectic, i.e., $\Phi_t^*\omega = \omega$.*

Proof. We give the proof for the sake of illustration. We start by introducing the differential one-forms dq and dp on full phase space, and then specify the restriction to $\mathscr{B}$ by considering the symplectic form along integral curves of the constrained equations of motion. From the equations of motion (4.5) we have

$$\begin{aligned} d\dot{q}^i &= \frac{\partial^2 H}{\partial p_i \partial q^l} dq^l + \frac{\partial^2 H}{\partial p_i \partial p_l} dp_l \\ d\dot{p}_i &= -\left(\frac{\partial^2 H}{\partial q^i \partial q^l} + \lambda \frac{\partial^2 \varphi}{\partial q^i \partial q^l} \right) dq^l - \frac{\partial^2 H}{\partial q^i \partial p_l} dp_l \\ 0 &= \frac{\partial \varphi}{\partial q^l} dq^l \,, \end{aligned}$$

where the last equation is the differential version of the constraint $\varphi(q) = 0$. Now consider a solution $(q(t), p(t))$ of the system (4.5). We have to show that $d\omega/dt = 0$. By definition, $\omega = \Omega|_{\mathscr{B}}$; therefore invariance of ω under the flow Φ_t is equivalent to state that the time derivative of the unconstrained symplectic form,

$$\frac{d\Omega}{dt} = \frac{d}{dt} \left(dq^i \wedge dp_i \right) = d\dot{q}^i \wedge dp_i + dq^i \wedge d\dot{p}_i \,,$$

vanishes along a constrained curve $(q(t), p(t)) \in \mathscr{B}$. Plugging the differentials from the equations of motion into the rightmost terms in the last equation we arrive at

$$\begin{aligned} \frac{d\Omega}{dt} &= \frac{\partial^2 H}{\partial p_i \partial q^l} dq^l \wedge dp_i + \frac{\partial^2 H}{\partial p_i \partial p_l} dp_l \wedge dp_i \\ &\quad + \left(\frac{\partial^2 H}{\partial q^i \partial q^l} + \lambda \frac{\partial^2 \varphi}{\partial q^i \partial q^l} \right) dq^l \wedge dq^i + \frac{\partial^2 H}{\partial q^i \partial p_l} dp_l \wedge dq^i \\ &= \frac{\partial^2 H}{\partial p_i \partial q^l} dq^l \wedge dp_i + \frac{\partial^2 H}{\partial q^i \partial p_l} dp_l \wedge dq^i \\ &= \frac{\partial^2 H}{\partial p_i \partial q^l} dq^l \wedge dp_i - \frac{\partial^2 H}{\partial p_l \partial q^i} dq^i \wedge dp_l \,, \end{aligned}$$

where we have taken advantage of the skew-symmetry of the wedge product: all terms of the form $A_{ij} dz^i \wedge dz^j$ cancel with the respective $-A_{ji} dz^j \wedge dz^i$, for $A_{ij} = A_{ji}$ is symmetric due to interchangeability of second order partial derivatives; by the skewness property the diagonal terms are zero, too. Finally, notice that all terms in the double sum appear twice with alternating signs. Hence all terms in the last line cancel, and the assertion follows. □

From this we immediately conclude:

Corollary 4.4. *Let λ_Σ be the Liouville form corresponding to $\omega = \Omega|_{\mathscr{B}}$ with $\mathscr{B} \cong T^*\Sigma$. Then the constrained flow $\Phi_t : \mathscr{B} \to \mathscr{B}$ preserves the Liouville volume, $\Phi_t^* \lambda_\Sigma = \lambda_\Sigma$.*

Proof. The assertion directly follows from Lemma 4.3 and the definition of the Liouville form (2.12) with the restricted symplectic form $\omega = \Omega|_{\mathscr{B}}$. □

4.1.3. Statistical mechanics of constrained molecular systems Let us shortly revisit the problem of evolving phase space densities in time. The line of discussion is similar to section 2.1.1: we abbreviate $z = (q, p)$ and consider an initial preparation $f_0(z)$. As the only difference we require $z \in \mathscr{B}$.

Since the constrained flow $\Phi_t : \mathscr{B} \to \mathscr{B}$ preserves the Liouville measure, i.e, the Hausdorff measure on $\mathscr{B}$ considered as a submanifold of $T^*\mathbf{R}^n \cong \mathbf{R}^n \times \mathbf{R}^n$, the Frobenius-Perron operator is simply defined as the push-forward of f_0 by the flow,

$$P_t f_0 = f_0 \circ \Phi_{-t} \,.$$

The energy of the constrained system is the Hamiltonian H restricted to $\mathscr{B}$. Hence the Gibbs measure ν_{can} naturally associated with the constrained system is the restriction of the full measure $\mu_{\mathrm{can}}(dz) = \rho_{\mathrm{can}}(z)dz$ to the constraint subspace, i.e.,

$$\nu_{\mathrm{can}} = (\rho_{\mathrm{can}}|_{\mathscr{B}})\, d\lambda_\Sigma \,. \tag{4.6}$$

Here $d\lambda_\Sigma$ is the Hausdorff measure (Liouville measure) of $\mathscr{B} \subset \mathbf{R}^n \times \mathbf{R}^n$. It is helpful to write down the local coordinate expression of ν_{can}: introducing again bundle coordinates (x, y) on $N\Sigma$, and defining the conjugate momenta (u, v) in the usual way (see Appendix B), the unconstrained symplectic form becomes

$$\Omega = dx^\alpha \wedge du_\alpha + dy \wedge dv \,,$$

where we used the index α to label the local coordinates x^α, u_α, $\alpha = 1, \ldots, n-1$ on the constrained phase space $\mathscr{B}$. The constrained symplectic form is obtained by restricting the standard symplectic form according to $\omega = \Omega|_{\mathscr{B}}$ which amounts to erasing the last term $dy \wedge dv$ in the sum. Using the local coordinate expression (B.4) of the unconstrained Hamiltonian, the constrained Gibbs measure reads

$$\nu_{\mathrm{can}}(dx, du) = \frac{1}{Z_\Sigma} \exp(-\beta H_\Sigma(x, u))\, dx^1 \ldots du_{n-1}$$

with

$$H_\Sigma(x, u) = \frac{1}{2} \left\langle G(x)^{-1} u, u \right\rangle + V(x, 0) \,,$$

and the partition function

$$Z_\Sigma = \int_{\mathscr{B}} \exp(-\beta H_\Sigma(x, u))\, dx^1 \ldots du_{n-1} \,.$$

Here we encounter the same problem as without constraints: the invariant measure of the system (4.5) is not unique and, in particular, the only candidate for an ergodic measure, namely the microcanonical measure, is singular with respect to $d\lambda_\Sigma$. However Section 2.1.1 has already set the stage for the constrained case: we introduce a discrete stochastic constrained Hamiltonian system as iterates of the map

$$x_{k+1} = (\pi \circ \Phi_\tau)(x_k, u_k) \,, \quad \pi : T^*\Sigma \to \Sigma \,. \tag{4.7}$$

Now let u_k be chosen randomly according to the constrained momentum distribution

$$\varrho_x(u) \propto \exp\left(-\beta T(x, u)\right) \,, \quad T(x, u) = \frac{1}{2} G^{\alpha\beta}(x) u_\alpha u_\beta \,, \tag{4.8}$$

where the $G^{\alpha\beta}$ are the elements of the inverse metric of $\Sigma \subset \mathbf{R}^n$. Then the discrete spatial transfer operator S_τ that takes probability densities on Σ forward in time is

$$S_\tau f(x) = \int (f \circ \pi \circ \Phi_\tau)(x, u)\, \varrho_x(u)\, du \,.$$

Let

$$\nu_\Sigma(dx) = \frac{1}{Q_\Sigma} \exp(-\beta V(x,0)) \sqrt{\det G(x)}\, dx\,, \tag{4.9}$$

be the Gibbs measure on Σ that is obtained from ν_{can} by integrating out the momenta. (The constant Q_Σ simply normalizes the total probability to one.) According to Section 2.1.1 we consider S_τ on the weighted Hilbert space $L^2(\nu_\Sigma)$ with the respective scalar product defined in (2.19). Consulting Proposition 2.8, we immediately obtain that $\nu_\Sigma(dx)$ is the unique invariant measure of the constrained stochastic Hamiltonian system (4.7). The algorithmic realization will be exposed in the following section.

Remark 4.5. *A frequently used (symbolic) formula for the constrained canonical measure in the ambient space variables (q,p) that involves Dirac's delta function is*

$$\nu_\Sigma \propto \exp(-\beta H(q,p))\delta(\varphi(q))\delta(\dot\varphi(q))(\mathrm{vol}J_\varphi(q))^2\,,$$

where $\mathrm{vol}J_\varphi = \|\nabla\varphi\|$ denotes the matrix volume of $\nabla\varphi$. This representation is intrinsic to the constrained phase space $T^*\Sigma$ since the matrix volume annihilates the explicit dependence on φ stemming from the delta function (compare equation (3.7)).

4.2. Sampling constrained invariant measures

We are aiming at algorithms that allow for sampling the (invariant) Gibbs measure of a constrained systems. The algorithms should be easy to implement on a computer and offer control over the numerical discretization error. Without constraints, sampling the Gibbs measure can be accomplished using any of the standard thermostatting techniques. Here the task is more involved, for two major requirements have to be met: firstly the thermostat must be consistent with constrained dynamics (fixed reaction coordinate), and secondly the dynamics has to be ergodic with respect to the constrained Gibbs measure. It is well-known that the ordinary Nosé-Hoover thermostat suffers from ergodicity problems for certain classes of Hamiltonians [108, 109]. This pathology can be removed by using extensions to the single-oscillator chain or by imposing constant temperature constraints [110, 111]. But even then, expectation values converge *only if* the dynamics is ergodic, and conditions to guarantee ergodicity are still lacking (notice the circularity in the argument). Additionally all these more sophisticated methods have in common that due to their complexity they are relatively hard to implement, and they require a careful adjustment of the parameters involved. Even worse, it is not clear a priori how these methods fit constrained symplectic integration; see [112] for a discussion on that topic. In particular in the Nosé-Hoover method the constraint force, which is the relevant quantity in free energy calculations, becomes dependent on the thermostat variables, which means that it can no longer be interpreted as the constraint force of the molecular subsystem. A promising alternative is stochastic Langevin dynamics or Brownian (Smoluchowski) dynamics [13]. These systems are proven to be ergodic under sufficiently weak assumptions like periodic or bounded configuration space [114, 115]. Since the noise term is usually unbounded constraining such systems to submanifolds of its state space is a challenging problem that has been recently addressed for the high-friction case [17].

4.2.1. Blue Moon sampling Recall the discussion of the Fixman Theorem in Section 3.1.2: we have to distinguish between the conditional and the constrained probability measure and the respective conditional expectations. Let $\Phi : \mathbf{R}^n \to \mathbf{R}^k$ be a smooth reaction coordinate, and denote by $\Sigma = \Phi^{-1}(\xi)$ its smooth fibre given a

regular value ξ of the reaction coordinate. The conditional probability measure of a Hamiltonian system reads

$$\mu_\xi(A) = \frac{1}{Z(\xi)} \int_A \exp(-\beta H)(\mathrm{vol} J_\Phi)^{-1} d\mathcal{H}_\xi \,,$$

where $A \subseteq \Sigma \times \mathbf{R}^n$ is a measurable Borel set, and $d\mathcal{H}_\xi$ is the surface measure of $\Sigma \times \mathbf{R}^n$ considered as a submanifold of $T^*\mathbf{R}^n \cong \mathbf{R}^n \times \mathbf{R}^n$. In contrast, the Gibbs measure generated by the constrained flow is defined as

$$\nu_{\mathrm{can}}(B) = \frac{1}{Z_\Sigma} \int_B \exp(-\beta H) d\lambda_\Sigma$$

with $B \subseteq T^*\Sigma$ and $d\lambda_\Sigma$ denoting the constrained Liouville measure on $T^*\Sigma \cong \Sigma \times \mathbf{R}^d$, where $d = n - k$ is the dimension of Σ. We define the respective expectation values

$$\mathbf{E}_\xi f = \frac{1}{Z(\xi)} \int_{\Sigma \times \mathbf{R}^n} f \exp(-\beta H)(\mathrm{vol} J_\Phi)^{-1} d\mathcal{H}_\xi \,,$$

and

$$\mathbf{E}_\Sigma f = \frac{1}{Z_\Sigma} \int_{\Sigma \times \mathbf{R}^d} f \exp(-\beta H) d\lambda_\Sigma \,.$$

If we restrict our attention to configuration observables f the relation between the two expectation values is easier to comprehend. First of all observe that $\Sigma = \Phi^{-1}(\xi)$ does not involve any momenta, from which the identity $d\mathcal{H}_\xi = d\sigma_\xi dp$ follows, where $d\sigma_\xi$ is the surface element of $\Sigma \subset \mathbf{R}^n$, and p denotes the original momenta. Hence we can integrate out the momenta and find that

$$\mathbf{E}_\xi f = \frac{1}{Q(\xi)} \int_\Sigma f \exp(-\beta V)(\mathrm{vol} J_\Phi)^{-1} d\sigma_\xi \,,$$

and

$$\mathbf{E}_\Sigma f = \frac{1}{Q_\Sigma} \int_\Sigma f \exp(-\beta V) d\sigma_\xi \,,$$

where the reduced normalization constants $Q(\xi)$ and Q_Σ are related by

$$Q(\xi) = Q_\Sigma \, \mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1} \,.$$

That is, as long as we consider only position-dependent observables we can compute averages with respect to either probability measure just by altering the potential function according to $V \mapsto V \pm \beta^{-1} \ln \mathrm{vol} J_\Phi$; compare the discussion of the Fixman Theorem in Section 3.1.2. In particular we can compute conditional expectations by running constrained simulations with the augmented potential $V_\Phi = V + \beta^{-1} \ln \mathrm{vol} J_\Phi$. Recall that this was just another way to read the Blue Moon relation (3.28),

$$\mathbf{E}_\xi f = \frac{\mathbf{E}_\Sigma \left(f \, (\mathrm{vol} J_\Phi)^{-1} \right)}{\mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}} \,,$$

which expresses the conditional expectation of a configurational observable f by the constrained expectation $\mathbf{E}_\Sigma(\cdot) = \mathbf{E}(\cdot | q \in \Sigma)$. It can be computed either with respect to ν_{can} as defined above or likewise with respect to ν_Σ as given by (4.9).

4.2.2. Constrained hybrid Monte-Carlo The goal of this section is to introduce an alternative to the usual microcanonical sampling methods (Nosé-Hoover, isokinetic ensemble) that may not be ergodic, or standard Monte-Carlo which may be poorly mixing. We adopt the hybrid Monte-Carlo (HMC) technique, which emulates the general Metropolis Monte-Carlo strategy of proposal and acceptance steps, where, however, the proposal is generated by short runs of the Hamiltonian system with randomly chosen initial conditions. This method circumvents the common Monte-Carlo problem, namely, that the acceptance probability of an arbitrary move to an energetically unfavourable state becomes incredibly small. As ordinary Metropolis Monte-Carlo, HMC is conceptually very simple, and is designed to handle symplectic integration, i.e., one can use standard integrators for constrained Hamiltonian systems. Moreover it can be proved that the dynamics is ergodic with respect to the positional density under rather mild conditions which are met for our purposes [83, 252, 253]. As an additional treat the acceptance procedure also controls the numerical error, because HMC rejects those moves that have too large energy fluctuations.

In order to explain how HMC works recall the concept of the discrete spatial transfer operator S_τ that evolves spatial densities forward in time, and which is associated with a stochastic Hamiltonian system with random momenta. According to Proposition 2.8 and the considerations from the last section, the randomized flow preserves the spatial probability measure (4.9) that we may write as

$$\nu_\Sigma(dx) = \frac{1}{Q_\Sigma} \exp(-\beta V(\sigma(x)))\sqrt{\det G(x)}dx\,,$$

where $\sigma(x)$ denotes the embedding of Σ into $\mathbf{R}^n$, and $x = (x^1 \dots, x^d)$ are local coordinates on Σ. Now consider the symplectic and reversible discrete flow map $\Psi_\tau : \mathscr{B} \to \mathscr{B}$ on the constrained phase space $\mathscr{B} = T^*\Sigma$, and consider iterates of Ψ_τ with initial momenta that are randomly chosen according to the constrained Maxwell distribution $\varrho_x(\cdot)$ in (4.8). This generates a sequence $\{x_0, \dots, x_{N-1}\} \subset \mathbf{R}^d$ in configuration space. Note that if the flow Ψ_τ were exactly energy-preserving, then the x_k would be distributed according to ν_Σ. However it is impossible to find a numerical discretization scheme that is symplectic, reversible, and exactly energy-conserving at the same time as follows from backward error analysis [105]. The best we can achieve is that the energy error remains uniformly bounded on compact time intervals and oscillates around its exact value [254].

The hybrid Monte-Carlo (HMC) method accounts for this drawback by accepting or rejecting points with a certain probability that depends on the energy error. We start the integration from $x_k \in \mathbf{R}^d$ with initial momentum $u_k \sim \varrho_x(u)$. Integrating the underlying Hamiltonian system for a time τ then generates a proposal $\tilde{x}_k = (\pi \circ \Psi_\tau)(x_k, u_k)$, which is accepted (i.e., $x_{k+1} = \tilde{x}_k$) with probability

$$p_\tau(x_k, u_k) = \min\left(1, \exp(-\beta \Delta H_\Sigma(x_k, u_k; \tau))\right)\,,$$

where

$$\Delta H_\Sigma(x_k, u_k; \tau) = (H_\Sigma \circ \Psi_\tau)(x_k, u_k) - H_\Sigma(x_k, u_k)$$

is the energy error. Accordingly we reject the proposal (i.e., $x_{k+1} = x_k$) with probability $1 - p_\tau$. In this form, HMC yields a configuration sampling, and by repeating the procedure of generating proposals, the resulting HMC Markov chain $\{x_1, \dots, x_N\}$ allows for approximating the conditional expectation, if the system is ergodic [255]. In order to prove ergodicity for the constrained HMC Markov chain we make use of an idea in [83] that rests upon the following strong Law of Large Numbers [256, 257].

Proposition 4.6 (Meyn & Tweedie 1993, Tierney 1994)**.** *Let* $\{x_t \in \mathbf{R}^d, t = 0, \tau, 2\tau, \ldots\}$ *be a Markov chain with invariant probability measure* ν_Σ *that satisfies*

$$\mathbf{P}\left[x_{k+1} \in B \,|\, x_k = x\right] > 0 \quad \forall x \in U \subseteq \mathbf{R}^d, \forall B \in \mathcal{B}(U)\,, \tag{4.10}$$

where $\mathcal{B}(U)$ *is the Borel* σ*-algebra of* $U \subset \mathbf{R}^d$*, and* $B \in \mathcal{B}(U)$ *has positive Lebesgue measure. Then* $\{x_t \in \mathbf{R}^d, t = 0, \tau, 2\tau, \ldots\}$ *satisfies the strong Law of Large Numbers,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(\sigma(x_i)) = \int_{\mathbf{R}^d} f(\sigma(x))\nu_\Sigma(dx) \quad \textit{(almost surely)}$$

for almost all $x_0 \in \mathbf{R}^d$*, where* $f \circ \sigma \in L^1(\nu_\Sigma)$ *is a measurable function.*

It is convenient to understand f as an observable that is defined on the original n-dimensional configuration space, such that $f \circ \sigma$ denotes the restriction to Σ. For example, the reader may think of the system's potential energy $f = V(q)$. We proceed step by step, checking (i) invariance of the constrained Gibbs measure ν_Σ, and (ii) the phase space accessibility condition (4.10) for the HMC algorithm.

Invariance of the constrained Gibbs measure Invariance of the constrained Gibbs measure can be shown following the outline of the proof in [253] for separable Hamiltonians. Here we cannot separate the canonical density into merely momentum and position dependent parts, and so we write

$$\nu_{\mathrm{can}}(dx, du) = \frac{1}{Z_\Sigma} \underbrace{\exp(-\beta T(x,u))}_{\varrho_x(u)} \underbrace{\exp(-\beta V(\sigma(x)))}_{\eta(x)}\, dxdu$$

indicating that the momentum density depends on the position coordinates as well. We introduce the HMC acceptance probability for a τ-step by $(\tilde{x}, \tilde{u}) = \Psi_\tau(x, u)$

$$p_\tau(x,u) = \min\left(1,\, \frac{\varrho_{\tilde{x}}(\tilde{u})\eta(\tilde{x})}{\varrho_x(u)\eta(x)}\right). \tag{4.11}$$

The definition of $p_\tau(x,u)$ is the standard Metropolis-Hastings acceptance probability for symplectic and reversible flow maps, and it can be readily checked that it coincides with the acceptance probability defined above. Clearly we have $p_\tau = 1$ for an exactly energy-conserving flow. We prove the following statement.

Lemma 4.7. *The constrained Gibbs measure* ν_Σ *is invariant under the HMC flow that is generated by the symplectic and reversible flow map* Ψ_τ *together with the Metropolis acceptance-rejection procedure with acceptance probability* p_τ*.*

Proof. It is sufficient to show that the HMC preserves expectation values with respect to ν_Σ. Let $\zeta \in \mathbf{R}^d$ be an accepted position value after a single integration and acceptance step. We assume that the initial momentum u is distributed according to $\varrho_x(u)$. Furthermore, let $\vartheta(d\zeta)$ denote the marginal distribution of the position variables after one HMC step. Hence we have to show that

$$\int_{\mathbf{R}^d} f(\sigma(x))\nu_\Sigma(dx) = \int_{\mathbf{R}^d} f(\sigma(\zeta))\vartheta(d\zeta)\,.$$

Suppose the initial position x follows the canonical distribution ν_{can}. Then for a given x we draw a momentum vector from $\varrho_x(u)$, and propagate a time step τ according to $(\tilde{x}, \tilde{u}) = \Psi_\tau(x, u)$. We can perform the acceptance-rejection procedure

for the rightmost expectation using a change-of-variables argument. Exploiting that the constrained Liouville measure $d\lambda_\Sigma$ is preserved under the flow Ψ_τ, we obtain

$$\begin{aligned}
&\int_{\mathbf{R}^d} f(\sigma(\zeta))\vartheta(d\zeta) \\
&= \int_{\mathbf{R}^d} f(\sigma(\zeta))\, p_\tau(\Psi_{-\tau}(\zeta,\tilde{u}))\, \rho(\Psi_{-\tau}(\zeta,\tilde{u}))\, d\lambda_\Sigma \\
&\quad + \int_{\mathbf{R}^d} f(\sigma(\zeta))\ (1 - p_\tau(\zeta,-\tilde{u}))\, \rho(\zeta,-\tilde{u})\, d\lambda_\Sigma\,,
\end{aligned}$$

where $\rho(x,u) = \varrho_x(u)\eta(x)$ denotes the smooth density of $\nu_\Sigma(dx,du) = \rho(x,u)dxdu$. Note that the first integral on the right hand side originates from the acceptance, the second one stems from the rejection step. Taking advantage of the identity

$$p_\tau(\Psi_{-\tau}(\zeta,\tilde{u}))\, \rho(\Psi_{-\tau}(\zeta,\tilde{u})) = p_\tau(\zeta,-\tilde{u})\, \rho(\zeta,-\tilde{u})\,, \tag{4.12}$$

using the reversibility $\Psi_{-\tau}(x,u) = \Psi_\tau(x,-u)$ of the flow and that $\rho(x,-u) = \rho(x,u)$ is even in its second argument, we find upon integrating out the momenta

$$\begin{aligned}
&\int_{\mathbf{R}^d} f(\sigma(\zeta))\vartheta(d\zeta) \\
&= \int_{\mathbf{R}^d} f(\sigma(\zeta))\ (1 + A_\tau(\zeta,\tilde{u}) - A_\tau(\zeta,\tilde{u}))\, \rho(\zeta,\tilde{u})\, d\lambda_\Sigma \\
&= \int_{\mathbf{R}^d} f(\sigma(\zeta))\, \rho(\zeta,\tilde{u})\, d\lambda_\Sigma \\
&= \frac{1}{Z_\Sigma} \int_{\mathbf{R}^d} f(\sigma(\zeta))\, \sqrt{\det G(\zeta)}\, d\zeta\,.
\end{aligned}$$

In the second line we have introduced the abbreviation $A_\tau = p_\tau\rho$ for the two terms in the identity (4.12) above. The assertion follows, observing that the last equation is simply the expectation with respect to the constrained Gibbs measure ν_Σ. □

Remark 4.8. *HMC gives a time-reversible mapping, as can be verified directly by checking detailed balance for* $(\tilde{x},\tilde{u}) = \Psi_\tau(x,u)$*:*

$$\begin{aligned}
\rho(x,u)p_\tau(x,\tilde{x}) &= \rho(x,u)\min\left(1,\, \frac{\rho(\tilde{x},\tilde{u})}{\rho(x,u)}\right) \\
&= \min\left(\rho(\tilde{x},\tilde{u}), \rho(x,u)\right) \\
&= \rho(\tilde{x},\tilde{u})\min\left(1,\, \frac{\rho(x,u)}{\rho(\tilde{x},\tilde{u})}\right) \\
&= \rho(\tilde{x},\tilde{u})p_{-\tau}(\tilde{x},x)\,.
\end{aligned} \tag{4.13}$$

The assertion follows from the symmetry with respect to the initial and propagated variables after the second line. Hence HMC generates a reversible flow.

Configuration space accessibility To verify the accessibility condition (4.10) we basically have to show that there is a discrete flow map that connects any two points $x(0) \in U \subseteq \mathbf{R}^d$ and $x(\tau) \in B$, where $B \in \mathcal{B}(U)$. To this end we borrow an argument from [83], where the accessibility condition in case of an unconstrained, separable system has been proved. Therein the authors use a discrete version of Hamilton's assuming that the system is bounded, i.e., either $U \cong \mathbf{T}^d$ (compact), or $U \cong \mathbf{R}^d$ with $V \circ \sigma$ uniformly bounded from above.

Since the HMC acceptance probability (4.11) is strictly positive, it does not alter the accessibility properties of the Markov chain. Hence, and for the sake of notational convenience, we shall omit it in what follows. Proving the accessibility condition then requires two steps: In a first step we follow the approach in [83] and construct ambient space sample paths that satisfy the accessibility condition in $\Sigma \subset Q$. In doing so, it turns out that the problem boils down to a standard symplectic discretization of constrained systems. In a second step we demonstrate that the ambient space discretization has an equivalent formulation in local coordinates, hence satisfying the accessibility condition (4.10). Regarding the former problem we endeavour a discrete variant of Hamilton's principle of least action. Following [258], we introduce a discrete Langrangian as a map $L_h : Q \times Q \to \mathbf{R}$. The discrete counterpart of the classical action is a mapping $S_h : Q^{N+1} \to \mathbf{R}$, that is defined as the sum

$$S_h = \sum_{k=0}^{N-1} L_h(q_k, q_{k+1}) \tag{4.14}$$

where $q_k \in Q$ and k labels the discrete time. Given fixed endpoints $q_0, q_N \in Q$ the discrete variational principle states that the discretized equations of motion minimize the action sum. The discretized equations are obtained by variation over the $q_1, \ldots, q_{N-1}$ which yields the *discrete Euler-Lagrange* equations

$$\mathbf{D}_2 L_h(q_{k-1}, q_k) + \mathbf{D}_1 L_h(q_k, q_{k+1}) = 0\,, \quad \forall k \in \{1, \ldots, N-1\}\,, \tag{4.15}$$

where $\mathbf{D}_1, \mathbf{D}_2$ denote the derivatives with respect to the first and second slot. If $\mathbf{D}_2 L_h$ (the generalized discrete momentum) is invertible, then (4.15) implicitly defines a discrete flow by means of the map $(q_{k+1}, q_k) = \Phi_h(q_k, q_{k-1})$. The particular discretization scheme that leads to (4.14) is open to choice and should depend on the problem; for the details we refer to the seminal work of Marsden and West [258].

Lemma 4.9. *Suppose the potential $V : Q \to \mathbf{R}$ is sufficiently smooth and bounded from above. Given $q_0, q_\tau \in \Sigma$, there is a symplectic mapping $(q(\tau), p(\tau)) = \Phi_\tau(q(0), p(0))$ and an open neighbourhood $B \subset \Sigma$ of q_τ, such that*

$$\mathbf{P}\left[q(\tau) \in B \,|\, q(0) = q_0\right] > 0\,.$$

Proof. We define the constraint manifold Σ as the level set (fibre) of the smooth function $\varphi : Q \to \mathbf{R}$. That is, we set $\Sigma = \varphi^{-1}(0)$ for a regular value 0 of φ. For simplicity we assume that V is uniformly bounded on Σ (otherwise we may restrict our attention to a subset $M \subset \Sigma$ which can be done at the price of further notation). We let $L : TQ \to \mathbf{R}$ denote the continuous Lagrangian

$$L(q, \dot{q}) = \frac{1}{2}\,\langle \dot{q}, \dot{q} \rangle - V(q)\,,$$

and introduce the discrete Lagrangian $L_h : Q \times Q \to \mathbf{R}$ for a time step $h > 0$:

$$L_h(q_k, q_{k+1}) = \frac{1}{2}\left(L\left(q_{k+1}, \frac{q_{k+1} - q_k}{h}\right) + L\left(q_k, \frac{q_{k+1} - q_k}{h}\right)\right)$$

We fix endpoints $q_0, q_N \in \Sigma$ and set $q_N = q_\tau$. Since V is bounded, the action sum is bounded from below, and the limit of the unconstrained problem exists. Extremizing the unconstrained action sum subject to the constraint $q_k \in \Sigma$ for $k \in \{1, \ldots, N-1\}$,

$$\min_{q_k \in \Sigma,\, \lambda_k} \sum_{k=0}^{N-1} \left(L_h(q_{k+1}, q_k) - \langle \lambda_k, \varphi(q_k) \rangle\right)\,,$$

the discrete Euler-Lagrange equations turn out to be [259]

$$\begin{aligned} 0 &= \mathbf{D}_2 L_h(q_{k-1}, q_k) + \mathbf{D}_1 L_h(q_k, q_{k+1}) + \lambda_k^T \mathbf{D}\varphi(q_k) \\ 0 &= \varphi(q_k) \end{aligned} \tag{4.16}$$

for all $k \in \{1, \ldots, N-1\}$. Given $q_{k-1}, q_k \in \Sigma$, i.e., $\varphi(q_k) = \varphi(q_{k-1}) = 0$, we can evaluate the derivatives of the discrete Lagrangian L_h and solve the last equation for q_{k+1} subject to the condition that $q_{k+1} \in \Sigma$. This yields the equations of motion

$$\begin{aligned} q_{k+1} - 2q_k + q_{k-1} &= -h^2(\nabla V(q_k) + \mathbf{D}\varphi(q_k))\lambda_k \\ 0 &= \varphi(q_{k+1})\,, \end{aligned} \tag{4.17}$$

which are known as the SHAKE algorithm [260]. The Lagrange multiplier λ_k is chosen such as to enforce the constraint at time $k+1$. The discrete conjugate momenta is defined by the discrete Legendre transform of $\hat{L}_h = L_h - \langle \lambda_k, \varphi(q_k) \rangle$, that is,

$$p_k = -\mathbf{D}_1 L_h(q_k, q_{k+1}) + \mathbf{D}\varphi(q_k)\lambda_k\,. \tag{4.18}$$

Hence we can consider the SHAKE algorithm as a mapping $\mathscr{B} \to \mathscr{B}$ (or $T^*\Sigma \to T^*\Sigma$). It is symplectic by virtue of its variational character (cf. the related work [246, 82]). By choosing initial conditions $q(0) = q_0$ and $p(0) = -\mathbf{D}_1 \hat{L}_h(q_0, q_1, \lambda_0)$ the discrete flow generates a discrete trajectory that connects q_0 and q_τ. Finally, it follows by continuity of Φ_τ on the initial conditions that the endpoints of trajectories with perturbed initial momenta $p_\epsilon(0) = p(0) + \epsilon$ remain in $B \subset \Sigma$ whenever ϵ is sufficiently small. □

A frequently used variant of the SHAKE algorithm is called RATTLE and goes back to [261]. It can be considered as a constrained version of the ordinary velocity Verlet scheme. SHAKE and RATTLE are equivalent to each other by dint of (4.18). Moreover they are variational with the discrete Lagrangian L_h defined above, and therefore both SHAKE and RATTLE are symplectic.

Lemma 4.9 guarantees accessibility from any point $q \in \Sigma$ to any open set. However condition (4.10) requires accessibility of any Borel set of positive Hausdorff measure (irreducibility), which excludes certain pathologies that otherwise might occur in the HMC transition probabilities; see [83]. This is expressed in:

Lemma 4.10. *Let $\Psi_\tau : T^*\Sigma \to T^*\Sigma$ denote the symplectic numerical flow map that is defined by the RATTLE algorithm. Then the HMC transition probabilities satisfy*

$$\mathbf{P}\left[q(\tau) \in B \,|\, q(0) = q_0\right] > 0 \quad \forall q \in \Sigma \subset Q$$

for all $B \in \mathcal{B}(\Sigma)$ with positive Hausdorff measure $\mathcal{H}^d$ on Σ.

Proof. Given an initial point $q \in \Sigma$, we have to show that any Borel set B of positive measure can be reached from a set of momenta with positive measure.

To this end consider the subset $M_B(q) \subset T_q^*\Sigma$ that is determined by all initial momenta p for which $(\pi \circ \Psi_\tau)(q, p) \in B$. Omitting the acceptance step, the HMC transition probabilities $p(q, B, \tau) = \mathbf{P}\left[q(\tau) \in B \,|\, q(0) = q\right]$ can be written as

$$p(q, B, \tau) = \int_{M_B(q)} \varrho_q(q)\, dp\,.$$

Since the constrained Maxwell density $\varrho_q(p)$ is strictly positive, it is enough to show that $M_B(q)$ has positive measure. Since we can naturally identify all cotangent spaces $T_q^*\Sigma$ with the d-dimensional subspaces of $\mathbf{R}^n$ that are determined by the hidden constraint $\langle \nabla(q), p \rangle = 0$, we have to show that $M_B(q)$ has positive Hausdorff measure

$\mathcal{H}^d$. Now suppose the contrary, i.e., assume $\mathcal{H}^d(M_B(q)) = 0$, and consider the map $F_q : M_B(q) \to B$, $p \mapsto (\pi \circ \Psi_\tau)(q,p)$. By definition F_q is onto. Therefore we have [70]

$$\mathcal{H}^d(B) = \mathcal{H}^d(F_q(M_B(q))) \leq L\mathcal{H}^d(M_B(q)) = 0$$

which contradicts $\mathcal{H}^d(B) > 0$. Here $0 < L < \infty$ is the Lipschitz constant of F_q (since Ψ_τ is volume-preserving, such a constant obviously exists). □

It remains to show that the flow $(q_k, p_k) \to (q_{k+1}, p_{k+1})$ has an equivalent counterpart $(x_{k+1}, u_{k+1}) = \Psi_h(x_k, u_k)$ in local coordinates (which inherits all its structural properties). As we know from the continuous world, the local coordinate version of the Euler-Lagrange equations can be derived from the restricted Lagrangian $L_\Sigma = L|_{T\Sigma}$. Accordingly we define the constrained discrete Lagrangian as $L_{\Sigma,h} = (L|_{T\Sigma})_h$. Given an embedding $\sigma : \mathbf{R}^d \to \Sigma \subset Q$ we can define the constrained discrete Lagrangian $L_{\Sigma,h} : \Sigma \times \Sigma \to \mathbf{R}$ as the map

$$L_{\Sigma,h}(x_k, x_{k+1}) = L_h\left(\sigma(x_k), \sigma(x_{k+1})\right) ,$$

which gives rise to the following discrete Euler-Lagrange equations

$$0 = \mathbf{D}_2 L_{\Sigma,h}(x_{k-1}, x_k) + \mathbf{D}_1 L_{\Sigma,h}(x_k, x_{k+1}) . \tag{4.19}$$

Solving the equation for x_{k+1} given x_k, x_{k-1} defines a map $\Theta_h : \mathbf{R}^d \to \mathbf{R}^d$. By computing the conjugate momenta $u_k = -\mathbf{D}_1 L_{\Sigma,h}(x_k, x_k + 1)$ we can augment this map to a symplectic map $\Psi_h : T^*\mathbf{R}^d \to T^*\mathbf{R}^d$. The following statement is true [262]:

Lemma 4.11 (Wendlandt & Marsden 1997)**.** *Equation (4.16) has a solution $(q_{k+1}, q_k) = \Phi_h(q_k, q_{k-1})$ if and only if $(x_{k+1}, x_k) = \Theta_h(x_k, x_{k-1})$ is a solution of (4.19). Furthermore Φ_h and Θ_h are equivalent in the sense that $\Phi_h = \sigma \circ \Theta_h$.*

This completes the proof that the accessibility condition (4.10) holds true for the HMC Markov chain together with the SHAKE or RATTLE iteration. Together with the invariance of the constrained Gibbs measure ν_Σ we therefore conclude

Proposition 4.12. *Let $V : Q \to \mathbf{R}$ be sufficiently smooth and bounded from above. Then, for measurable $f \circ \sigma \in L^1(\nu_\Sigma)$, the strong Law of Large Numbers,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(q_i) = \int_{\mathbf{R}^d} f(\sigma(x))\nu_\Sigma(dx) \quad \textit{(almost surely)} ,$$

holds true for almost all initial values $q_0 \in \Sigma$, where $\{q_0, q_1, q_2, \ldots\}$ with $q_i \in \Sigma$ stems from the RATTLE symplectic integrator (4.17)–(4.18).

The last assertion does not say anything about the speed of convergence, which remains an open problem; see [83, 263] for some numerical studies. In particular the speed of convergence depends on the choice of the HMC integration time $\tau = Nh$. Exploring state space becomes certainly faster if τ is increased. However increasing τ while keeping the step-size h constant decreases the acceptance probability, since energy fluctuations become an issue.

Before we conclude the Monte-Carlo section, we shortly mention that the HMC algorithm with lag time $\tau = h$ and without the acceptance-rejection procedure is equivalent to an Euler discretization of the Smoluchowski equation [264]. However letting the acceptance step account for the discretization error, HMC can be regarded as an exact discretization of the Smoluchowski equation at step-size $\tau = h$, i.e., HMC generates a diffusion-like flow. Therefore the algorithm converges for any stable step-size without introducing a bias.

4.2.3. Langevin and Brownian motion As this section does not address dynamics but rather sampling of probability distributions in order to compute certain expectation values we may accept any sampling scheme that does the job. Popular sampling method in molecular dynamics are Brownian motion and Langevin dynamics, and we shall explain how they fit into the framework of constrained integration.

Unlike for deterministic dynamics, there are many situations in which stochastic dynamics is proved to be ergodic [115]. This requires that the coefficients in the equations are globally Lipschitz, a condition which is typically not satisfied; in practice, this seems to be no problem whatsoever [265].

Constrained Brownian motion We briefly review the work in [17], where an ergodicity proof for constrained Brownian motion is given. For this purpose we let again $\Sigma = \varphi^{-1}(0)$ denote a smooth submanifold of codimension k in $\mathbf{R}^n$, where $\varphi : \mathbf{R}^n \to \mathbf{R}^k$ with regular value $0 \in \mathbf{R}^k$. For each $\sigma \in \Sigma$ let $(n_1(\sigma), \ldots, n_k(\sigma))$ be the normal frame attached to Σ. If $Q \in \mathbf{R}^{n\times k}$ is the matrix the columns of which are the normal vectors n_k, then

$$P_T(\sigma) = \mathbf{1} - Q(\sigma)Q^T(\sigma)$$

is the point-wise orthogonal projection $P_T : T\mathbf{R}^n|\Sigma \to T\Sigma$ of vectors onto the tangent space of Σ. Here $\sigma : \mathbf{R}^k \to \Sigma$ labels again the embedding $\Sigma \subset \mathbf{R}^n$. It is convenient to use the ambient space notation $q = \sigma(x)$ for $q \in \mathbf{R}^n$ lying on Σ. Assuming the usual boundedness conditions on the potential V we have [17]

Proposition 4.13 (Lelièvre 2006)**.** *Let ν_Σ be the constrained Gibbs measure (4.9). Then ν_Σ is the unique invariant measure of the following Itô equation*

$$\dot q = -P_T(q)\left(\operatorname{grad} V(q) - \sqrt{2\beta^{-1}}\dot W\right) + \beta^{-1}\sum_{i=1}^{k}\kappa_i(q)n_i(q) \tag{4.20}$$

with initial value $q(0) \in \Sigma$, and the components κ_i of the mean curvature vector

$$H(q) = \sum_{i=1}^{s}\kappa_i(q)n_i(q)\,, \quad \kappa_i = -\operatorname{tr}(P_T\nabla n_i)\,.$$

Moreover the solutions $q(t)$ of (4.20) satisfy a Law of Large Numbers

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(q(t))\,dt \to \int_\Sigma f(\sigma(x))\nu_\Sigma(dx) \quad \textit{(almost surely)}$$

where $f \in L^1(\nu_\Sigma)$, and convergence holds for almost all initial values $q(0) \in \Sigma$.

Regarding the conditional measure we encounter the same situation as in the HMC case: simply changing the molecular potential to $V_\varphi = V + \beta^{-1}\ln \mathrm{vol} J_\varphi$ the constrained diffusion process samples the conditional probability measure. Equation (4.20) can be considered the ambient space formulation for diffusion on a submanifold of $\mathbf{R}^n$, similar to the constrained Euler-Lagrange equations (4.1). This representation is especially convenient for numerical discretization. Itô-Taylor expansion of (4.20) with step-size $h > 0$ leads to the following variational formulation [17]

$$\begin{aligned} q_* &= q_n - h\operatorname{grad} V(q_n) + \sqrt{2\beta^{-1}}\Delta W_n \\ q_{n+1} &= \operatorname*{argmin}_{z\in\mathbf{R}^n}\left(\|z - q_*\|^2 \mid \varphi(z) = 0\right) \end{aligned} \tag{4.21}$$

with $\Delta W_n = W_{n+1} - W_n$ denoting the increment of the Brownian motion. We can enforce the constraint by introducing an appropriate projection onto the tangent space of Σ which gives rise to the implicit Euler-Maruyama scheme [266]

$$\begin{aligned} q_{n+1} &= q_n - h\,(\operatorname{grad} V(q_n) + \mathbf{D}\varphi(q_{n+1})\lambda_n) + \sqrt{2\beta^{-1}}\Delta W_n \\ \varphi(q_{n+1}) &= 0\,, \end{aligned} \tag{4.22}$$

where the Lagrange multiplier $\lambda_n \in \mathbf{R}^k$ is chosen, such that $\varphi(q_{n+1}) = 0$. It is further possible to simplify the above scheme by attaching the constraint force $-\lambda^T \mathbf{D}\varphi$ at q_n, from which we obtain a semi-explicit discretization scheme [13, 16]

$$\begin{aligned} q_{n+1} &= q_n - h\,(\operatorname{grad} V(q_n) + \mathbf{D}\varphi(q_n)\lambda_n) + \sqrt{2\beta^{-1}}\Delta W_n \\ \varphi(q_{n+1}) &= 0\,, \end{aligned} \tag{4.23}$$

We emphasize that both discretization schemes are consistent with the constrained Itô equation (4.20). Certainly the implicit scheme will allow for larger step-sizes, but it requires to solve the implicit and nonlinear equation. In turn, the choice of the nonlinear solver will affect the stability of the numerical solution (cf. [82, 267]).

Constrained Langevin dynamics We address the problem of constraining Langevin dynamics to a configuration submanifold $\Sigma \subset \mathbf{R}^n$. Of course it is possible to treat the Langevin equation as an ordinary hypo-elliptic diffusion by applying Proposition 4.13. This would, however, completely ignore the underlying (symplectic) geometry of the phase space in the Langevin equation. Therefore we propose an approach that comes close to common index reduction techniques for mechanical systems with constraints.[19]

For the sake of simplicity we assume that φ be real-valued. Now consider the Langevin equation for a constrained natural mechanical system (4.5),

$$\begin{aligned} \dot{q}^i &= \frac{\partial \hat{H}}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial \hat{H}}{\partial q^i} - \gamma_{ij}\frac{\partial \hat{H}}{\partial p_j} + \sigma_{ij}\dot{W}^j\,, \quad i = 1, \ldots, n \\ 0 &= \frac{\partial \hat{H}}{\partial \lambda}\,, \end{aligned} \tag{4.24}$$

with the constrained Hamiltonian $\hat{H} = H + \lambda\varphi$,

$$\hat{H}(q,p) = \frac{1}{2}\langle p, p\rangle + V(q) + \lambda\varphi(q)\,.$$

Moreover, let us assume that γ, σ are scalar satisfying $2\gamma = \beta\sigma^2$. Then, more concretely, the constrained Langevin equation for a separable Hamiltonian reads

$$\begin{aligned} \dot{q} &= p \\ \dot{p} &= -\nabla V(q) - \lambda\nabla\varphi(q) - \gamma p + \sigma\dot{W} \\ 0 &= \varphi(q)\,, \end{aligned} \tag{4.25}$$

Unlike in the mechanical case considered earlier, the Lagrange multiplier has now become a random variable that depends on the particular realization of the Brownian

[19] A related approach has been put forward recently during the writing of this thesis [84].

motion. Recall from the discussion of the constrained Hamiltonian system that the dynamics takes place on the constrained phase space bundle that by

$$\mathscr{B} = \left\{ (q,p) \in T^*\mathbf{R}^n \,\middle|\, q \in \Sigma \text{ and } \left\langle \nabla\varphi(q), \mathbf{D}_2\hat{H}(q,p) \right\rangle = 0 \right\} .$$

Let $n(q)$ be the unit normal to Σ. Since the gradient $\operatorname{grad}\varphi = \nabla\varphi$ is normal to the fibre $\varphi^{-1}(0)$ and $p = \mathbf{D}_2\hat{H}$, we have the orthogonality condition

$$\langle n(q(t)), p(t) \rangle = 0$$

for the solutions $(q(t), p(t))$ of (4.25). Therefore (by differentiation with respect to t)

$$\langle n(q(t)), \dot{p}(t) \rangle + \langle \nabla n(q(t)) \cdot p(t), p(t) \rangle = 0$$

By inserting the equation of motion for p, and solving for $-\lambda\nabla\varphi$, we find

$$-\lambda\nabla\varphi(q) = P_N^*(q) \left(\nabla V(q) + \gamma p - \sigma\dot{W} \right) + S_q(p,p) , \tag{4.26}$$

where $P_N^* : (T^*\mathbf{R}^n)|\Sigma \to (T^*\Sigma)^\perp$, $P_N^* = nn^T$ is the point-wise projection onto the orthogonal complement of $T^*\Sigma$ and $S_q(p,p)$ is the second fundamental form of the embedding $\Sigma \subset \mathbf{R}^n$ (compare Lemma 4.1 and keep in mind that the mass scaling allows us to identify $T\mathbf{R}^n$ with $T^*\mathbf{R}^n$), viz.,

$$S_q(p,p) = -n(q) \langle \nabla n(q) \cdot p, p \rangle .$$

Plugging the constraint force back into the Langevin equation (4.25) eliminates the constraint, and we end up with the phase space equivalent of (4.20):

$$\begin{aligned} \dot{q} &= p \\ \dot{p} &= -P_T^*(q) \left(\nabla V(q) + \gamma p - \sigma\dot{W} \right) + S_q(p,p) . \end{aligned} \tag{4.27}$$

Here $P_T^* = \mathbf{1} - P_N^*$ denotes the orthogonal projection onto the constrained phase space $T^*\Sigma$. The function $\varphi(q)$ is a conserved quantity of the constrained Langevin equation which can be seen as follows: By construction of the constraint force, we have $\ddot{\varphi}(t) = 0$ along the solutions $(q(t), p(t))$ of (4.27). Integrating with respect to time we conclude that $\varphi(t) = \alpha t + \delta$. Choosing suitable initial conditions $(q(0), p(0)) = (q_0, p_0)$, such that

$$\varphi(q_0) = 0 \quad \& \quad \langle \nabla\varphi(q_0), p_0 \rangle = 0 ,$$

we have $\alpha = \delta = 0$ and therefore $\varphi(t) = 0$ at all times $t > 0$. Borrowing a denomination from the theory of differential algebraic equations [251], we term (4.27) the *underlying stochastic differential equation* to (4.25).[20] It remains to check whether the constrained canonical distribution μ_Σ is invariant under the constrained Langevin dynamics.

As before, let $H_\Sigma = H|_{\mathscr{B}}$ denote the restriction of the original Hamiltonian to the constrained phase space $\mathscr{B} \cong T^*\Sigma$. We employ the notation $d\lambda_\Sigma(q,p)$ for the constrained Liouville measure expressed in the ambient space coordinates.[21] Then, abbreviating $z = (q,p)$, the invariant measure can be written as

$$\mu_\Sigma(dz) = \frac{1}{Z_\Sigma} \exp(-\beta H_\Sigma(z)) d\lambda_\Sigma(z) .$$

[20] Exactly the same result would be obtained by applying Itô's formula to the orthogonality condition above, for the orthogonality condition is linear in the momenta, and the noise comes solely from the momentum equation. Therefore there are no extra second-order contributions from the noise.

[21] The notation $d\lambda_\Sigma(q,p)$ becomes clear if one bears in mind that the constrained Liouville volume form λ_Σ is defined by exterior products of the constrained symplectic form ω_Σ which is simply the restriction of the unconstrained symplectic form, $\omega_\Sigma = (dq^i \wedge dp_i)|_{\mathscr{B}}$.

In order to show that μ_Σ is indeed invariant we consider the Kolmogorov backward equation associated with the Langevin equation (4.27) and study its solution

$$u(z,t) = \mathbf{E}_z f(z(t))\,, \quad u(z,0) = f(z)\,,$$

where $z(t) = (q(t), p(t))$ is the solution of (4.27), and $\mathbf{E}_z(\cdot)$ is the expectation conditional on the initial value $z = (q_0, p_0)$. The measure μ_Σ is invariant, if

$$\int_{\mathscr{B}} u(z,t)\mu_\Sigma(dz) = \int_{\mathscr{B}} u(z,0)\mu_\Sigma(dz) \quad \forall t > 0\,. \tag{4.28}$$

The backward generator (2.22) for the constrained Langevin equation (4.27) reads

$$\mathcal{A}_{\mathrm{bw}} = \frac{\sigma^2}{2} P_T^* : \mathbf{D}_2^2 + p \cdot \mathbf{D}_1 - P_T^* \left(\nabla V + \gamma p\right) \cdot \mathbf{D}_2 + S_q \cdot \mathbf{D}_2\,.$$

The double contraction $A : B = \mathrm{tr}(AB)$ denotes the matrix inner product, whereas the simple dot is the pairing between tangent and cotangent vectors in $\mathbf{R}^n$. Taking the time derivative of (4.28), omitting the normalization constant Z_Σ, we obtain

$$\begin{aligned}
&\frac{\partial}{\partial t} \int_{\mathscr{B}} u \exp(-\beta H_\Sigma) d\lambda_\Sigma \\
&\quad = \int_{\mathscr{B}} (\mathcal{A}_{\mathrm{bw}} u) \exp(-\beta H_\Sigma) d\lambda_\Sigma \\
&\quad = \underbrace{\int_{\mathscr{B}} \left(\frac{\sigma^2}{2} P_T^* : \mathbf{D}_2^2 - \gamma P_T^* p \cdot \mathbf{D}_2 \right) u \exp(-\beta H) d\lambda_\Sigma}_{\text{forcing and dissipation}} \\
&\qquad + \underbrace{\int_{\mathscr{B}} \left(p \cdot \mathbf{D}_1 - (P_T^* \nabla V(q) - S_q(p,p)) \cdot \mathbf{D}_2\right) u \exp(-\beta H) d\lambda_\Sigma}_{\text{constrained Hamiltonian}}\,,
\end{aligned}$$

where we have replaced H_Σ by H under the integral. We can address the two terms separately: Regarding the latter, we observe that the integral contains the Liouvillian of the index-reduced deterministic system. As μ_Σ is invariant under the constrained deterministic flow, and the index-reduced system generates the same flow on $T^*\Sigma$ as the constrained Hamiltonian vector field (4.5), it follows that the integral vanishes identically. The integrand in the first integral can be written as

$$\begin{aligned}
&\int_{\mathscr{B}} \left(\frac{\sigma^2}{2} P_T^* : \mathbf{D}_2^2 - \gamma P_T^* p \cdot \mathbf{D}_2 \right) u \exp(-\beta H) d\lambda_\Sigma \\
&\qquad = \frac{\sigma^2}{2} \int_{\mathscr{B}} \mathrm{div}_\Sigma \left(\mathbf{D}_2 \left(u \exp(-\beta H)\right)\right) d\lambda_\Sigma
\end{aligned}$$

with div_Σ labelling the divergence on the linear momentum subspace $T_q^*\Sigma \subset T_q^*\mathbf{R}^n$

$$\mathrm{div}_\Sigma X(q,p) = \mathrm{tr}\left(P_T^* \mathbf{D}_2 X(q,p)\right)\,.$$

We can perform the momentum integration by application of the divergence theorem [268] for submanifolds of arbitrary codimension; since $T_q^*\Sigma$ is a linear subspace of $T_q^*\mathbf{R}^n \cong \mathbf{R}^n$ and linear subspaces have zero mean curvature, it follows that the remaining forcing/dissipation integral above is zero, i.e.,

$$\int_{\mathscr{B}} \mathrm{div}_\Sigma \left(\mathbf{D}_2 \left(u \exp(-\beta H)\right)\right) d\lambda_\Sigma = 0\,.$$

Hence we conclude that μ_Σ is invariant under the constrained Langevin motion (4.27). Assuming that the associated Markov process $z(t) = (q(t), p(t))$ with $(q(0), p(0)) =$

(q_0, p_0) has a strictly positive transition function (accessibility condition) yields the Law of Large Numbers

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(z(t))dt \to \int f(z)\mu_\Sigma(dz) \quad \text{(almost surely)}$$

for all functions $f \in L^1(\mu_\Sigma)$ and consistent initial conditions $(q_0, p_0) \in T^*\Sigma$. We omit the generalization of (4.27) to vector-valued constraints and refer the reader to the section on constrained Hamiltonian systems.

An *ad-hoc* numerical discretization of (4.27) can be built upon integrators for constrained deterministic systems. Modifying the SHAKE algorithm (4.17) with the discrete conjugate momentum (4.18) accordingly, we propose the scheme

$$\begin{aligned} p_{n+1/2} &= p_n - \frac{h}{2}\left(\nabla V(q_n) + \gamma p_n + \lambda_n \nabla\varphi(q_n)\right) + \sigma\Delta W_{n+1/2} \\ q_{n+1} &= q_n + h p_{n+1}\,, \end{aligned} \tag{4.29}$$

where $\Delta W_{n+1/2} = W_{n+1/2} - W_n$, and the Lagrange multiplier λ_n is chosen, such that

$$\varphi(q_{n+1}) = 0\,. \tag{4.30}$$

The final momentum step is

$$\begin{aligned} p_{n+1} = p_{n+1/2} &- \frac{h}{2}(\nabla V(q_{n+1}) + \gamma p_{n+1/2} \\ &+ \mu_n \nabla\varphi(q_{n+1})) + \sigma\Delta W_{n+1}\,, \end{aligned} \tag{4.31}$$

where $\Delta W_{n+1} = W_{n+1} - W_{n+1/2}$, and μ_n is determined by the hidden constraint

$$\langle \nabla\varphi(q_{n+1}), p_{n+1}\rangle = 0\,. \tag{4.32}$$

The integrator is quasi-symplectic in the sense of [269], and we expect that it is strongly convergent of order two. Preliminary numerical simulations seem to support this claim, but we refrain from detailed numerical studies for the sake of brevity. In fact, a very similar result has appeared recently during the writing of this thesis. Elaborating upon the RATTLE integrator, the authors of [84] obtain an integrator almost identical to (4.29)–(4.32), but with a different implementation of the white noise term; in addition, they prove that the integrator is second-order accurate.

4.3. Thermodynamic Integration

We have presented methods for sampling constrained Gibbs measures in either configuration or phase space. By these means we can now to sample, for example, the derivative of the free energy or other quantities that appear in any of the reduced models. In order to access configuration space regions which correspond to improbable values of the reaction coordinate, we have to resort to methods like Thermodynamic Integration [8, 270] or the closely related Thermodynamic Perturbation [271, 272]. Accordingly this section explains Thermodynamic Integration from the point of view of the different types of constrained dynamics and gives an overview of different methods of free energy calculation. In particular we will explain how both geometric and standard free energy can be computed rather efficiently from the force of constraint.

We shall exemplify the basic approach by means of the optimal prediction equations. For the sake of simplicity we consider a scalar reaction coordinate $\Phi : \mathbf{R}^n \to \mathbf{R}$. In this particular case the optimal prediction Hamiltonian (3.63) reads

$$E(\xi, \eta) = \frac{1}{2} m(\xi)^{-1}\eta^2 + G(\xi)\,,$$

where the effective mass m is given by

$$m(\xi) = \left(\frac{1}{Q_\Sigma} \int_{\Sigma_\xi} \|\nabla \Phi\|^2 \exp(-\beta V) d\sigma_\xi \right)^{-1} .$$

Here Σ_ξ denotes the level sets $\Phi^{-1}(\xi)$ for all regular values ξ of Φ, and $d\sigma_\xi$ is the corresponding surface element (no conditional expectation here, Q_Σ is the normalization constant). Recall further that the geometric free energy is defined as

$$G(\xi) = -\beta^{-1} \ln \int_{\Sigma_\xi} \exp(-\beta V) d\sigma_\xi \, .$$

In order to sample the effective mass, we can simply use any method that samples ν_Σ, like HMC or even Langevin dynamics. If we denote by $\{q_0, \ldots, q_{N-1}\} \subset \Sigma_\xi$ the respective (constrained) Markov chain we can approximate m by

$$m(\xi) \approx \left(\frac{1}{N} \sum_{i=0}^{N-1} \|\nabla \Phi(q_i)\|^2 \right)^{-1} .$$

Recall that it is theoretically possible to compute the standard free energy by running brute force simulations, sampling the marginal distribution of the reaction coordinate. In principle the geometric free energy could also be directly computed from unconstrained simulation data building histograms of the reaction coordinate: upon backwards application of the Blue Moon reweighting formula (3.28), we have

$$\exp(-\beta G(\xi)) \approx \left(\sum_{i=0}^{N-1} \|\nabla \Phi(q_i)\| \right)^{-1} \sum_{i=0}^{N-1} \chi_\xi(\Phi(q_i)) \, \|\nabla \Phi(q_i)\| \, ,$$

where χ_ξ denotes the indicator function of the set $[\xi, \xi + \Delta\xi[$ for sufficiently small increment $\Delta\xi$. The last formula makes the geometric free energy directly observable. Of course for all reaction coordinates of actual interest, the sampling along the reaction coordinate will be rather poor due to slow mixing and metastability. Resorting to Thermodynamic Integration instead, we can estimate G from its derivative,

$$G'(\xi) \approx \frac{1}{N} \sum_{i=0}^{N-1} \frac{\langle n(q_i), \nabla V(q_i) \rangle - \beta^{-1} \mathrm{div}\, n(q_i)}{\|\nabla \Phi(q_i)\|} \, , \quad n = \frac{\nabla \Phi}{\|\nabla \Phi\|} \, .$$

Basically, the formula for G' is obtained by disregarding the Fixman potential in the expression (3.15). In principle, the optimal prediction equations (3.58) would require only that the mean force be given, however it might be desirable to have its potential at hand. Given n samplings at various values ξ_l we can recover G by numerical integration (i.e., Thermodynamic Integration) over ξ using any suitable quadrature rule

$$G_n(\xi) = \sum_{l=1}^{n} w_{n,l} \, G'(\xi_l) \, , \tag{4.33}$$

where $w_{n,l}$ are the weights of the particular quadrature rule (see, e.g., [273, 274]).

4.3.1. Free energy from constrained Langevin motion We study Thermodynamic Integration in case the constrained dynamics is on phase space. For this purpose consider a generalized free energy along a vectorial reaction coordinate $\Phi : \mathbf{R}^n \to \mathbf{R}^k$

$$U_\alpha(\xi) = -\beta^{-1} \ln Z_\alpha(\xi)$$

with the generalized partition function

$$Z_\alpha(\xi) = \int_{\Sigma_\xi \times \mathbf{R}^n} \exp(-\beta H_\alpha)\, d\mathcal{H}_\xi \,,$$

where $d\mathcal{H}_\xi = d\sigma_\xi dp$ is the surface measure of $\Sigma \times \mathbf{R}^n \subset \mathbf{R}^n \times \mathbf{R}^n$, and H_α is the Hamiltonian that is augmented by the Fixman potential with weight α,

$$H_\alpha = H + \beta^{-1} \ln \mathrm{vol} J_\alpha \,, \quad \mathrm{vol} J_\alpha = \sqrt{\det \mathbf{D}\alpha^T \mathbf{D}\alpha} \,.$$

Choosing $\alpha = q$ the generalized free energy turns into the geometric free energy $G = U_q$, whereas for $\alpha = \Phi$ we recover the standard free energy $F = U_\Phi$. Now recall that according to Lemma 3.3 the derivative of the free energy can be written as

$$\nabla U_\alpha = \frac{1}{Z_\alpha} \int_{\Sigma \times \mathbf{R}^n} \left. \frac{\partial H_\alpha}{\partial \Phi} \right|_{\Phi = \xi} \exp(-\beta H_\alpha) d\mathcal{H}_\xi \,,$$

Moreover we know from the discussion in Section 3.1.1 that the derivative of the free energy with respect to the reaction coordinate is independent of the normal momenta (or velocities). Then, upon comparing equation (3.13) to the expression (4.26) for the Langevin constraint force (note that both the noise term and the linear friction term have zero mean), it turns out that ∇U_α can be equivalently written as

$$\nabla U_\alpha = \frac{1}{Z_\alpha} \int_{\Sigma \times \mathbf{R}^n} \lambda_\alpha \, \exp(-\beta H_\alpha) d\mathcal{H}_\xi \,,$$

where λ_α is the Lagrange multiplier in (4.25) that is necessary to constrain a Langevin system with Hamiltonian H_α to the constraint phase space

$$\mathscr{B} = \{(q,p) \in \mathbf{R}^n \times \mathbf{R}^n \,|\, q \in \Sigma,\, \mathbf{D}\Phi(q) \cdot \mathbf{D}_2 H_\alpha(q,p) = 0\} \,,$$

which is clearly independent of the weight α, since $\mathbf{D}_2 H_\alpha = \mathbf{D}_2 H$. Notice that by definition of the constraint force, λ_α depends only on the constrained momenta. Hence we can replace the expectation above by the constrained average. This yields

$$\nabla U_\alpha = \int_{\mathscr{B}} \lambda_\alpha \, \mu_{\Sigma,\alpha} \,, \tag{4.34}$$

where $\mu_{\Sigma,\alpha}$ is the constrained canonical probability measure with Hamiltonian H_α,

$$\mu_{\Sigma,\alpha} = \frac{1}{Z_{\Sigma,\alpha}} \exp(-\beta H_{\Sigma,\alpha}) d\lambda_\Sigma \,, \tag{4.35}$$

that is preserved by the constrained Langevin system with Hamiltonian H_α. Clearly, $\alpha = q$ simply amounts to the constrained canonical probability measure, whereas the invariant measure of the Langevin system with $\alpha = \Phi$ is the conditional canonical measure. The respective Lagrange multipliers are related by

$$\lambda_\alpha = \lambda_q - \beta^{-1} (J_\alpha^T J_\alpha)^{-1} J_\alpha^T \nabla \ln \mathrm{vol} J_\alpha \,, \tag{4.36}$$

which, upon choosing $\alpha = \Phi$, becomes the correct expression (3.13) for computing the derivative of the standard free energy. Hence formulae (4.34)–(4.36) reveals both ∇F and ∇G by appropriately adapting the weight function α. Assuming ergodicity for the discretization of the constrained Langevin equation (4.27) we claim that the following is true: Let (q_k, p_k), $k = 0, \ldots N-1$ be a discretized solution of (4.27) with initial values $(q_0, p_0) \in \mathscr{B}$. Then we have for the geometric free energy

$$\nabla G(\xi) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} \lambda_{q,k}(q_k, p_k) \,, \tag{4.37}$$

where $\lambda_{q,k}$ is the Lagrange multiplier of the stochastic RATTLE algorithm (4.29)–(4.32). If we replace the potential V in (4.27) by the augmented potential $V_\Phi = V + \beta^{-1} \ln \mathrm{vol} J_\Phi$, generating a realization $\{(\tilde{q}_0, \tilde{p}_0), \ldots, (\tilde{q}_{N-1}, \tilde{p}_{N-1})\} \subset \mathscr{B}$, then we obtain the same relation for the derivative of the standard free energy

$$\nabla F(\xi) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \lambda_{\Phi,k}(\tilde{q}_k, \tilde{p}_k)\,, \tag{4.38}$$

where $\lambda_{\Phi,k}$ is now the discrete RATTLE Lagrange multiplier for the augmented potential. At this point the reader may wonder, whether we can replace the continuous Lagrange multiplier in (4.34) by its discrete counterpart. For the deterministic RATTLE algorithm, equivalence between the Lagrange multipliers has been established in [12]. Indeed, by simply repeating the argument given there, it follows that the same is true for the stochastic RATTLE algorithm.

The evaluation of ∇V_Φ may be a tedious task, since it requires to compute the Hessians $\nabla^2 \Phi_i$. If the mean force is not updated at each integration step it may be more efficient to use the original potential instead of the augmented one. Then we can explicitly augment the Lagrange multiplier according to (4.36) and reweight the average employing the Blue Moon relation (3.28). If $\{(q_0, p_0), \ldots, (q_{N-1}, p_{N-1})\} \subset \mathscr{B}$ is a realization of the constrained Langevin equation (without the additional Fixman potential), then we can compute the standard free energy by means of

$$\nabla F(\xi) = \lim_{N\to\infty} \left(\sum_{k=0}^{N-1} w(q_k) \right)^{-1} \sum_{k=0}^{N-1} w(q_k) \lambda_{\Phi,k}(q_k, p_k)\,, \tag{4.39}$$

with the Blue moon weight $w = (\mathrm{vol} J_\Phi)^{-1}$ and

$$\lambda_{\Phi,i} = \lambda_{q,i} - \beta^{-1} (J_\Phi^T J_\Phi)^{-1} J_\Phi^T \nabla \ln \mathrm{vol} J_\Phi\,,$$

which is in perfect agreement with the formulae that have been derived in various instances, e.g., [12, 15, 11, 275].

Remark 4.14. *An even simpler way to compute F directly goes via equation (3.29). Recall the considerations concerning the co-area formula that have led to the Blue Moon reweighting relation in Section 3.1.2. In particular we have found that standard and geometric free energy are simply related by*

$$F(\xi) = G(\xi) - \beta^{-1} \ln \mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}\,.$$

Since we can obtain the components of ∇G by just averaging over the ordinary Lagrange multipliers, it is evident that the most efficient way to compute F is by first computing G and then adding the Fixman potential $D = -\beta^{-1} \ln \mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}$. This method is certainly the most efficient one, since it only requires the evaluation of $\mathrm{vol} J_\Phi$, where the Jacobians are available anyway during the constrained integration without extra reweighting or the calculation of second derivatives.

4.3.2. Free energy from constrained hybrid Monte-Carlo Free energy calculation with HMC trajectories is slightly different from the Langevin case, since HMC samples only the configurational Gibbs density. The Lagrange multipliers, however, are functions of both positions and momenta.

We can easily compute the momentum average analytically: averaging the quadratic curvature term in the constraint force over the constrained Gaussian momentum density gives the mean curvature as can be seen from equation (3.15). To a

certain extend this is obvious, as the momentum average of a quadratic form gives the trace of the matrix inside the quadratic form, and the Lagrange multipliers contain the second fundamental form that involves the matrices of the Weingarten maps; the trace of the Weingarten maps then yields the coefficients of the mean curvature vector. The next statement is a consequence of (3.15) and the Law of Large Numbers (4.12):

Corollary 4.15. *Let $\{q_0, \dots, q_{N-1}\} \subset \Sigma$ denote a HMC Markov chain with the integrator (4.17)–(4.18) and the acceptance probability (4.11). Then*

$$\nabla G(\xi) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} g(q_k)$$

with

$$g = (J_\Phi^T J_\Phi)^{-1} \left(J_\Phi^T \nabla V - \beta^{-1} \operatorname{tr}\left(P_T \nabla^2 \Phi \right) \right) ,$$

where $P_T = \mathbf{1} - J_\Phi (J_\Phi^T J_\Phi)^{-1} J_\Phi^T$ is the point-wise projection onto $T_q\Sigma$, and the rightmost term is understood component-wise for $\Phi = (\Phi_1, \dots, \Phi_k)^T$. Accordingly, the derivative of the standard free energy takes the obvious form

$$\nabla F(\xi) = \lim_{N\to\infty} \left(\sum_{k=0}^{N-1} w(q_k) \right)^{-1} \sum_{k=0}^{N-1} w(q_k) f(q_k) ,$$

with $w = (\mathrm{vol} J_\Phi)^{-1}$ and

$$f = g - \beta^{-1} (J_\Phi^T J_\Phi)^{-1} J_\Phi^T \nabla \ln \mathrm{vol} J_\Phi .$$

Note that the reasoning of Remark 4.14 applies as well: we can directly compute the standard free energy F by Thermodynamic Integration of ∇G and adding the Fixman term $D = -\beta^{-1} \ln \mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}$ to G. The alternative way to compute the ∇F without extra reweighting à la Blue Moon is expressed in the following statement:

Corollary 4.16. *Let $\{\tilde{q}_0, \dots, \tilde{q}_{N-1}\} \subset \Sigma$ denote a constrained HMC Markov chain with the augmented Hamiltonian $H_\Phi = T + V_\Phi$, where $V_\Phi = V + \beta^{-1} \ln \mathrm{vol} J_\Phi$. Then*

$$\nabla F(\xi) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(\tilde{q}_k)$$

with

$$f = (J_\Phi^T J_\Phi)^{-1} \left(J_\Phi^T \nabla V_\Phi - \beta^{-1} \operatorname{tr}\left(P_T \nabla^2 \Phi \right) \right) .$$

Bibliographical remarks For the sake of completeness we mention just a few other methods that are available in the literature. A Brownian dynamics approach that exploits the relation between the derivative of a generalized free energy and the respective mean constraint force is given in [17]. Another widely-used Monte-Carlo-based algorithm for free energy calculations is Umbrella Sampling [276], where the system is forced to sample a certain range of the reaction coordinate by adding a confining potential. Though easy to implement (and thus popular), Umbrella Sampling involves unphysical manipulations of the original system and uncontrolled sources of error due to the choice of the confining potential [236]. Another class of approaches can be subsumed under the name of *Adaptive Biasing Forces*. These approaches, like conformational flooding [277], scaled force [77], or metadynamics [278], estimate the mean force during the course of integration. While sampling of phase space proceeds, the estimate is progressively refined, and introduced in the equations of motion as

a biasing force, which guarantees that the force acting along the reaction coordinate averages to zero over time. Eventually the free energy is recovered from the added force. For an overview of various kinds of methods we refer to the reviews [13, 2] and the references therein.

The formerly mentioned methods exploit that the free energy is equal to the reversible work that a system performs while undergoing an adiabatic change of state. This requires that the system always stays in its thermodynamic equilibrium conditional on the (frozen) reaction coordinate. But if the reaction coordinate is controlled in such a way that the remaining system cannot relax to its thermodynamic equilibrium, then the amount of performed work typically exceeds the free energy (second law of thermodynamics). Hence the above mentioned algorithms suffer from a systematic overestimation of free energy differences due to finite sampling times. However it is possible to compute free energy differences by averaging the irreversible work using an appropriate exponential weighting which is due to [279, 280]. For applications of the *Jarzynski equality* we refer to the recent preprint [281].