3. Eliminating fast degrees of freedom

While the discussion in the last section addressed the identification of essential variables, we shall now explain how reduced models can be derived from the full set of equations of motion. The techniques that will be introduced in the course of this section range from thermodynamical free energy concepts to dynamical averaging techniques. In any event the physical idea behind the reduction process is that the fast degrees of freedom act as random forcing on the slowly evolving parts in the system. If the dynamics of the fast variables is well-posed in the sense that it admits a unique equilibrium distribution, we can simply average the random perturbations over their equilibrium distribution whereby the slow degrees of freedom are effectively driven by an averaged force. A generic slow-fast system has the form

$$\begin{aligned}\dot{x}_\epsilon(t) &= f(x_\epsilon(t), y_\epsilon(t), \epsilon)\\ \dot{y}_\epsilon(t) &= \frac{1}{\epsilon} g(x_\epsilon(t), y_\epsilon(t), \epsilon)\,,\end{aligned} \tag{3.1}$$

where x and y are the slow and fast coordinates, respectively. From the equations of motion it can be seen already that if both f and g are (globally) Lipschitz continuous, uniformly in ϵ and t, then the fast velocities will be of order $1/\epsilon$ faster than the slow ones if ϵ goes to zero. This situation becomes more intuitive if we switch to the slow timescale by scaling the free variable $t \mapsto \epsilon t$

$$\begin{aligned}\dot{x}_\epsilon(t) &= \epsilon f(x_\epsilon(t), y_\epsilon(t), \epsilon)\\ \dot{y}_\epsilon(t) &= g(x_\epsilon(t), y_\epsilon(t), \epsilon)\,,\end{aligned} \tag{3.2}$$

where we have labelled the scaled quantities again by $x_\epsilon(t), y_\epsilon(t)$. In the limit $\epsilon \to 0$ the slow variables are effectively frozen, for $\dot{x}_\epsilon(t) = \mathcal{O}(\epsilon)$, while the fast variables evolve conditional on the slow ones.[5] We assume that the conditional fast dynamics is well-posed for all values of the slow variables in a sense that will be specified below. This *slaving* mechanism is a common feature of molecular systems: for instance, it is a general phenomenon that the frequencies of the fast bond vibrations depend upon the slowly evolving conformations of the molecule; in turn, the varying bond vibrations couple back to the slow modes, usually torsion angles [29]. It may even happen that the back-coupling of the fast variables to the slow ones induces further timescales which may lie beyond the characteristic time of the slow degrees of freedom [35, 165].

The reader may wonder why timescales are an issue at all, besides the fact that systems with several different timescales are in some vague sense *complicated*. One difficulty in the context of molecular dynamics applications lies in the need for long-time simulations; in order to integrate the equations of motion any numerical scheme has to resolve the fastest modes on the order of femtoseconds which is a tedious task if the simulation ought to reveal the dynamics of the slowest modes that may take place on scales of milliseconds. Moreover the effect of the discretization error becomes more and more important for long trajectories, since for high-dimensional systems in a random environment (solvent) the discretized system departs from the exact trajectory very early during the integration.[6]

[5] We will make extended use of the Landau symbol $\mathcal{O}$ which we will, however, use in a very loose sense: here $h(\epsilon) = \mathcal{O}(\epsilon^\alpha)$ means that the limit $|h(\epsilon)\epsilon^{-\alpha}| \to c \geq 0$ exists for $\epsilon \to 0$.

[6] Yet this seems to be no problem whatsoever, since although single trajectories may be completely misdirected, the calculation of average quantities works surprisingly well; for a detailed discussion on the question *Why does molecular dynamics work?* the reader is referred to [166, 167, 168, 169].

3.1. Central paradigm in biophysics: free energy landscapes

There is a whole industry within the molecular dynamics community that is concerned with the calculation of free energy profiles. The free energy is arguably considered the most fundamental thermodynamical quantity in analyzing molecular systems, for there is a variety of phenomena as, for instance, molecular solvation, enzyme catalysis, or conformation dynamics, the analytical understanding of which is directly related to the corresponding free energy landscape [170]; see the review [1] and the references therein. Moreover it is a common believe that the *dynamics* of these phenomena is also driven by the free energy. For instance, it is often assumed that conformation dynamics is dynamics in the respective free energy landscape [171, 172]. We shall argue that this is not generally the case, even if there is a clear timescale separation between reaction coordinate and the remaining degrees of freedom. (The reason is that the free energy is not the potential of a force in the strict sense.) Before we come to this point let us briefly review the notion of free energy.

Speaking of *free energy* in the context of molecular applications, mostly means the Helmholtz or the Gibbs free energy. The Helmholtz free energy is the quantity of choice in order to describe the reversible work in a system at constant temperature in a fixed volume, whereas the Gibbs free energy describes reversible processes at constant temperature and pressure. In both cases the number of particles is kept constant. Here we are particularly interested in the Helmholtz free energy, which is most standard if no chemical reactions occur.

The statistical mechanics definition of the free energy is in terms of the partition function. Let us give an intuitive derivation: Recall the thermodynamical concept of Legendre transformations among thermodynamical potentials [173]. The Helmholtz free energy is given by $F = U - TS$, where U is the internal energy, T the temperature, and S is the entropy of the system. The partition function is simply the normalization constant of the respective probability density, say $\rho \propto \exp(-\beta H)$,

$$Z = \int_{T^*Q} \exp(-\beta H(z))\, dz\,, \quad z = (q,p)\,.$$

Now we can endeavour the Boltzmann definition of Shannon's information entropy,

$$S = -\int_{T^*Q} \rho(z) \ln \rho(z)\, dz\,,$$

which can be rewritten for a system in equilibrium, i.e., $\rho = Z^{-1}\exp(-\beta H)$:

$$S = \beta \mathbf{E} H(z) + \ln Z\,. \tag{3.3}$$

Noting that $\beta = 1/T$, it follows upon identifying $U = \mathbf{E}H(z)$ that

$$F = -\beta^{-1} \ln Z \tag{3.4}$$

which is the familiar expression that typically appears in molecular dynamics books. By replacing the Hamiltonian by the potential energy we can easily repeat the last few steps for the configurational Gibbs ensemble, but we could also consider a subensemble only, e.g., the distribution of the fast variables. This will be explained next.

Thermodynamic Integration We introduce the conditional free energy. Let $\Phi : \mathbf{R}^n \to \mathbf{R}^k$ denote a reaction coordinate. Unless otherwise stated we assume that Φ

is regular in the sense that its Jacobian $\mathbf{D}\Phi$ has full rank k almost everywhere.[7] The molecular Hamiltonian $H : T^*\mathbf{R}^n \to \mathbf{R}$ in mass-scaled coordinates reads

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

Following the relevant literature (e.g., [88]) we have:

Definition 3.1. *Consider the Hamiltonian H on the phase space $T^*\mathbf{R}^n \cong \mathbf{R}^n \times \mathbf{R}^n$ with the canonical coordinates (q,p), and let $\Phi : \mathbf{R}^n \to \mathbf{R}^k$ denote a smooth reaction coordinate. Then the free energy along the values of Φ is defined as*

$$F(\xi) = -\beta^{-1}\ln Z(\xi), \tag{3.5}$$

with the partition function

$$Z(\xi) = \int_{\mathbf{R}^n\times\mathbf{R}^n} \exp(-\beta H(q,p))\delta(\Phi(q)-\xi)\,dqdp\,, \tag{3.6}$$

where δ denotes the Dirac delta measure on $\mathbf{R}^k$.

The reader should bear in mind that (up to normalization) the integrand in (3.6) defines a conditional probability density. By application of the co-area formula we can write the partition function as the equivalent surface integral [70, 175]

$$Z(\xi) = \int_{\Sigma_\xi\times\mathbf{R}^n} \exp(-\beta H)\,(\mathrm{vol}J_\Phi)^{-1}d\mathcal{H}_\xi\,. \tag{3.7}$$

where $d\mathcal{H}_\xi$ is the Hausdorff measure (surface element) of $\Sigma_\xi \times \mathbf{R}^n$ considered as a submanifold of $\mathbf{R}^n \times \mathbf{R}^n$. Here $\Sigma_\xi \subset \mathbf{R}^n$ denotes the level set $\Phi^{-1}(\xi)$, but for the sake of simplicity we shall drop the subscript ξ and just write Σ for the level sets. The volume of the rectangular matrix J_Φ is defined as [176]

$$\mathrm{vol}J_\Phi(q) = \sqrt{\det J_\Phi^T(q)J_\Phi(q)}\,.$$

We believe that (3.5) together with (3.7) provides the appropriate mathematical representation of the free energy. For our purpose this form is more convenient than the one involving the Dirac delta, unless we want to dig into the depths of generalized functions and measure theory. For a formal derivation of the above identity using a simple change-of-variables argument the reader is referred to Appendix D.

From the definition it is clear that the free energy could be easily computed from the marginal probability distribution of the reaction coordinate. However the essential dynamics is typically slow, and so reliably sampling the marginal distribution is a rather tedious issue. Therefore a common approach is to constrain the system to fixed values of the reaction coordinate, and then sample the average force acting upon it. The free energy is recovered afterwards by numerical integration with respect to the reaction coordinate. This widely-used technique, which exploits the dichotomy of *free energy as the potential of mean force*, is known as Thermodynamic Integration and goes back to Kirkwood [8]. The hope is that, once one has successfully identified the reaction coordinate, sampling in the remaining variables is comparably fast.

We issue a warning: There is some ambiguity in the definition of free energy throughout the literature. Especially in the literature on transition state theory the term free energy is often used without the matrix volume; see, e.g., [3, 4]. We shall come back to that point at a later stage, and introduce yet another definition:

[7] According to Sard's Lemma [174] this can be guaranteed by choosing the $\Phi : \mathbf{R}^n \to \mathbf{R}^k$, such that it belongs to the class $\mathcal{C}^{n-k+1}(\mathbf{R}^n)$. Then the points, where $\mathbf{D}\Phi$ is rank-deficient, form a set of measure zero in $\mathbf{R}^{n-k}$, and the level sets $\Phi^{-1}(\xi)$ are regular submanifolds of codimension k in $\mathbf{R}^n$.

Definition 3.2. *The expectation for an integrable phase space function $f = f(q,p)$ conditional on the reaction coordinate $\Phi(q) = \xi$ is defined as*

$$\mathbf{E}_\xi f = \frac{1}{Z(\xi)} \int_{\Sigma\times\mathbf{R}^n} f \exp(-\beta H)\, (\mathrm{vol} J_\Phi)^{-1} d\mathcal{H}_\xi \,. \tag{3.8}$$

The following Lemma is standard, but we give the proof for the sake of illustration:

Lemma 3.3. *Let the free energy be defined as above. Then the derivative of the free energy takes the form of a conditional expectation*

$$\nabla F(\xi) = \mathbf{E}_\xi f_\xi \,, \tag{3.9}$$

where f_ξ is the generalized force along the reaction coordinate evaluated at $\Phi(\cdot) = \xi$,

$$f_\xi = \left.\frac{\partial H}{\partial \Phi}\right|_{\Phi=\xi} + \beta^{-1} \left(J_\Phi^T J_\Phi\right)^{-1} J_\Phi^T \nabla \ln \mathrm{vol} J_\Phi \,. \tag{3.10}$$

Proof. Differentiating the free energy (3.5) with respect to $\xi \in \mathbf{R}^k$ we obtain

$$\nabla F(\xi) = -\beta^{-1} \frac{1}{Z(\xi)} \frac{\partial Z}{\partial \xi} \,,$$

where $\partial/\partial\xi = (\partial/\partial\xi^1, \ldots, \partial/\partial\xi^k)$ is shorthand for the vector of partial derivatives with respect to ξ. Hence it remains to evaluate the integral

$$\frac{\partial Z}{\partial \xi} = \frac{\partial}{\partial \xi} \int_{\Sigma\times\mathbf{R}^n} \exp(-\beta H)\, (\mathrm{vol} J_\Phi)^{-1}\, d\mathcal{H}_\xi \tag{3.11}$$

The calculation is easily carried out in an adapted coordinate frame. To this end we let $\sigma : \mathbf{R}^d \to \Sigma$, $d = n - k$ be the embedding $\Sigma \subset \mathbf{R}^n$, and we let $\{n_1(\sigma), \ldots, n_k(\sigma)\}$ denote a set of orthonormal vectors that span the normal space over Σ. Further we denote by $N\Sigma_\varepsilon$ a sufficiently small tubular ε-neighbourhood of Σ, such that the map

$$\phi : \mathbf{R}^n \to N\Sigma_\varepsilon,\ (x,\eta) \mapsto \sigma(x) + \eta^i n_i(\sigma(x)) \,.$$

is a local embedding $N\Sigma_\varepsilon \subset \mathbf{R}^n$. By means of ϕ we can uniquely represent any point $q \in \mathbf{R}^n \cap N\Sigma_\epsilon$ in terms of the bundle coordinates as $q = \phi(x,\eta)$; for the details we refer to Appendix B. In particular the local coordinate expression for the potential is

$$V(x,\eta) = V(\sigma(x) + \eta^i n_i(\sigma(x))) \,.$$

Defining the conjugate momenta (u,ζ) in the standard way, we can easily extend ϕ to a symplectic transform $T^*\phi : T^*\mathbf{R}^n \to T^*N\Sigma_\varepsilon$. By construction, the transformation from (q,p) to the adapted coordinates (x,η,u,ζ) is symplectic, hence volume-preserving. Moreover the condition $\Phi(q) = \xi$, i.e., the restriction to $\Sigma \times \mathbf{R}^n$ amounts to setting $\eta = 0$. For convenience we define an augmented Hamiltonian by $H_\Phi = H + \beta^{-1} \ln \mathrm{vol} J_\Phi$. Using chain rule the derivative in (3.11) now becomes

$$\begin{aligned}\frac{\partial Z}{\partial \xi} &= -\int_{\mathbf{R}^d\times\mathbf{R}^n} B(x)^{-T} \frac{\partial}{\partial \eta} \exp(-\beta H_\Phi(x,\eta,u,\zeta))|_{\eta=0}\, dx du d\zeta \\ &= \beta \int_{\mathbf{R}^d\times\mathbf{R}^n} B(x)^{-T} \left.\frac{\partial H_\Phi}{\partial \eta}\right|_{\eta=0} \exp(-\beta H_\Phi(x,0,u,\zeta))\, dx du d\zeta \,,\end{aligned}$$

where $B(x)$ is the matrix $J_\Phi^T(\sigma(x))Q(\sigma(x))$ with $Q = (n_1, \ldots, n_k) \in \mathbf{R}^{n\times k}$. By definition of the augmented Hamiltonian this yields

$$\left.\frac{\partial H_\Phi}{\partial \eta}\right|_{\eta=0} = \left.\frac{\partial H}{\partial \eta}\right|_{\eta=0} + \beta^{-1}Q^T\nabla \ln \mathrm{vol}J_\Phi\,.$$

Upon multiplication with B^{-T} the last equation is equal to

$$\left.\frac{\partial H_\Phi}{\partial \Phi}\right|_{\Phi=\xi} = \left.\frac{\partial H}{\partial \Phi}\right|_{\Phi=\xi} + \beta^{-1}(Q^TJ_\Phi)^{-1}Q^T\nabla \ln \mathrm{vol}J_\Phi\,.$$

To complete the proof we show that the matrix $(Q^TJ_\Phi)^{-1}Q^T$ is the Moore-Penrose pseudoinverse of the Jacobian J_Φ. To this end consider a QR decomposition of the Jacobian J_Φ. That is, we consider $J_\Phi = QR$, where $Q \in \mathbf{R}^{n\times k}$ has orthonormal columns and $R \in \mathbf{R}^{k\times k}$ is upper triangular. Since R is invertible, the Moore-Penrose pseudoinverse of the Jacobian can be written as [177]

$$(J_\Phi^TJ_\Phi)^{-1}J_\Phi^T = (R^TQ^TJ_\Phi)^{-1}R^TQ^T = (Q^TJ_\Phi)^{-1}R^{-T}R^TQ^T\,,$$

by which the assertion immediately follows.[8] □

Remark 3.4. *The last result looks slightly different from what is typically found in the literature [71, 2]; see also [78, 10, 11, 12]; they are equivalent though. The difference can be explained by pointing out that these authors treat the partition function (3.6) as an ordinary surface integral (without the Jacobian), simultaneously considering considering Φ as if it were an independent coordinate [179]; cf. the results in [180].*

3.1.1. Contributions to the free energy Let us shortly comment on the last result. Apparently the derivative of the free energy is the conditional expectation of the mechanical force $\partial H/\partial\Phi$ in the direction of the reaction coordinate plus an additional term that is owed to the definition of the conditional probability density (pseudo force). Only in case that $\Sigma \subset \mathbf{R}^n$ is a linear subspace the matrix volume in (3.7) is constant, and the free energy is really the potential of the average mechanical force. We shall study the contributions to the mechanical force in more detail. From a geometrical viewpoint the Lagrangian formulation is more convenient, for the interpretation becomes more lucid. Taking advantage of the identity (2.8) we have

$$\left.\frac{\partial H}{\partial \eta}\right|_{\eta=0} = -\left.\frac{\partial L}{\partial \eta}\right|_{\eta=0}$$

along the integral curves of the Hamiltonian vector field. In coordinates L reads

$$L(x,\eta,\dot{x},\dot{\eta}) = \frac{1}{2}\left\langle (G+C)\dot{x},\dot{x}\right\rangle + \left\langle A^T\dot{x},\dot{\eta}\right\rangle + \frac{1}{2}\left\langle \dot{\eta},\dot{\eta}\right\rangle - V(x,\eta)\,,$$

with the submatrices of the metric tensor defined in (B.2); see the appendix for details. We can compute the derivative of the Langrangian with respect to the normal coordinate component-wise. This yields

$$\left.\frac{\partial L}{\partial \eta^i}\right|_{\eta=0} = \frac{1}{2}\left.\frac{\partial C_{\alpha\beta}}{\partial \eta^i}\right|_{\eta=0}\dot{x}^\alpha\dot{x}^\beta + \left.\frac{\partial A_{j\alpha}}{\partial \eta^i}\right|_{\eta=0}\dot{x}^\alpha\dot{\eta}^j - \left.\frac{\partial V}{\partial \eta^i}\right|_{\eta=0}.$$

[8] Intriguingly the last line would be true, even if Q were not orthogonal: in fact for arbitrary full-rank matrices $A, B \in \mathbf{R}^{n\times k}$ with $A = BS$ and S non-singular, it can be shown that $A^\sharp = (B^TA)^{-1}B^T$ is the uniquely defined Moore-Penrose pseudoinverse of A. It can be readily checked that the thus defined matrix meets the four Moore-Penrose conditions [178].

By chain rule it follows for the potential term

$$\left.\frac{\partial V}{\partial \eta^i}\right|_{\eta=0} = \langle n_i, \operatorname{grad} V\rangle$$

which is simply the directional derivative along the i-th normal direction. We can omit the potential in the following. The two other terms have a nice geometrical interpretation, too. Using the results from Appendix B we find

$$\left.\frac{\partial L}{\partial \eta^i}\right|_{\eta=0} = S^i_{\alpha\beta}(x)\dot{x}^\alpha\dot{x}^\beta + \omega^i_j(X_\alpha)\dot{x}^\alpha\dot{\eta}^j\,, \tag{3.12}$$

where

$$S^i_{\alpha\beta} = \langle dn_i(X_\alpha), X_\beta\rangle$$

are the matrix entries of the symmetric map that is associated with the second fundamental form of the embedding (extrinsic curvature of Σ) written in the basis of the local tangent vectors $X_\alpha = \partial\sigma/\partial x^\alpha$. The vectors $dn_i(X) = \nabla n_i \cdot X$ denote the directional derivatives of the normals n_i along a vector X. The coefficients ω^i_j are the normal fundamental forms that are associated with the normal frame $\{n_1, \ldots, n_k\}$:

$$\omega^i_j(X_\alpha) = \langle n_i, dn_j(X_\alpha)\rangle$$

Note that the term involving the normal connection in linear in both the normal and the tangential velocities. Hence it disappears upon taking the average over the velocities [15]. In particular if the codimension of Σ in $\mathbf{R}^n$ is one, then it is well-known that the connection term is identically zero; see Appendix B for details.

At first glance, the fact that the normal fundamental forms give no contribution to the free energy is quite remarkable. It says that the derivative of the mean force depends solely on points on $T\Sigma$, but not on the ambient space variables, in particular not on the normal velocities. At closer inspection, however, this is what we should expect, since the reaction coordinate does not depend on the velocities at all. Consequently we can disregard the connection term and compute the mean force by averaging over the remaining terms only. Reformulating the result from Lemma 3.3 accordingly, we thus have the expression for the derivative of the free energy

$$\nabla F(\xi) = \mathbf{E}_\xi \hat{f}_\xi\,,$$

where

$$\hat{f}_\xi = (Q^T(q)J_\Phi(q))^{-1}\left(Q^T(q)\,\nabla V_\Phi(q) - \langle \nabla n(q)\cdot v, v\rangle\right)\,, \tag{3.13}$$

with

$$V_\Phi(q) = V(q) + \beta^{-1}\ln \operatorname{vol} J_\Phi(q)\,.$$

The last quantity $\hat{f}_\xi$ in (3.13) is known as the *force of constraint* that is needed to constrain a natural mechanical system with potential V_Φ to the configuration submanifold $\Sigma = \Phi^{-1}(\xi)$; see the discussion in the Sections 4.1 and 4.2. Here the curvature term $\langle \nabla n \cdot v, v\rangle$ is understood as a k-vector with the single components $\langle \nabla n_i \cdot v, v\rangle$, and (q, v) are elements of the tangent bundle

$$T\Sigma = \{(q,v) \in \mathbf{R}^n \times \mathbf{R}^n \,|\, q \in \Sigma,\ J_\Phi(q)\cdot v = 0\}\,.$$

In order to reduce the computational effort it may convenient to recast (3.13) in a form that does not require to compute the orthonormal vectors n_i. Indeed $\hat{f}_\xi$ equals

$$\hat{f}_\xi = (J^T_\Phi(q)J_\Phi(q))^{-1}\left(J^T_\Phi(q)\nabla V_\Phi(q) - \left\langle \nabla^2\Phi(q)\cdot v, v\right\rangle\right)\,, \tag{3.14}$$

where again the rightmost term is explained component-wise for $\Phi = (\Phi_1, \ldots, \Phi_k)$.

Further notice that the reaction coordinate depends only on the configuration variables. Hence we can equally well integrate out the momenta in (3.7), which does not make a difference for the free energy. Modulo additive constants it becomes

$$F(\xi) = -\beta^{-1} \ln Q(\xi)\,, \quad Q(\xi) = \int_\Sigma \exp(-\beta V_\Phi) d\sigma_\xi\,,$$

where $d\sigma_\xi$ is the surface element of $\Sigma \subset \mathbf{R}^n$. Calculating the derivative yields

$$\nabla F(\xi) = \frac{1}{Q(\xi)} \int_\Sigma \bar{f}_\xi \exp(-\beta V_\Phi) d\sigma_\xi\,.$$

with

$$\bar{f}_\xi = (J_\Phi^T(q) J_\Phi(q))^{-1} \left(J_\Phi^T(q) \nabla V_\Phi(q) - \beta^{-1} \operatorname{tr}\left(P_T(q) \nabla^2 \Phi(q) \right) \right). \tag{3.15}$$

Here $P_T = \mathbf{1} - J_\Phi (J_\Phi^T J_\Phi)^{-1} J_\Phi^T$ denotes the point-wise projection onto the constrained tangent space $T_q\Sigma$. The last equation is in fact a velocity-averaged version of the generalized force (3.14) with respect to the Maxwellian velocity distribution [13, 16]. The trace term is known to be the extrinsic mean curvature of Σ in $\mathbf{R}^n$ with respect to the normal frame that is spanned by the gradient vectors $\operatorname{grad} \Phi_i$.

Remark 3.5. *Intriguingly equation (3.12) suggests a more general interpretation: Let X denote a generic vector field that is attached to a submanifold $\Sigma \subset \mathbf{R}^n$. For each $\sigma \in \Sigma$ consider the decomposition of tangent spaces $T_\sigma \mathbf{R}^n = T_\sigma\Sigma \oplus N_\sigma\Sigma$ with the respective projections P_T and P_N that are defined point-wise for $\sigma \in \Sigma$. Note that this is a decomposition of $\mathbf{R}^n$, since we can naturally identify $T_\sigma\mathbf{R}^n$ with $\mathbf{R}^n$. Then we can define two vector fields the first of which satisfies [181]*

$$P_N \nabla_X Y = I\!I(X, Y)\,, \tag{3.16}$$

where ∇ is the (covariant) differentiation in $\mathbf{R}^n$, and X, Y are both tangent along Σ. The second fundamental form $I\!I$ is defined by means of the Weingarten maps [182]

$$I\!I(X, Y) = \sum_i n_i \langle n_i, \nabla_X Y \rangle = \sum_i n_i \langle \mathfrak{S}_i X, Y \rangle\,.$$

The symmetric Weingarten maps $\mathfrak{S}_i : T_\sigma\Sigma \to T_\sigma\Sigma$ are given by $\mathfrak{S}_i = -P_T dn_i(\cdot)$ as can be readily checked by differentiating the relation $\langle n_i, Y \rangle = 0$ along X. (Here $dn_i(X)$ is just an alternative notation for $\nabla_X n_i$.) For a normal vector field ν, i.e., a vector field with $\nu(\sigma) \in N_\sigma\Sigma$ we have the following identity

$$P_N \nabla_X \nu = D_X \nu\,, \tag{3.17}$$

where $D_X\nu$ is the connection of the normal bundle. Given a normal frame $\{n_1, \ldots, n_k\}$ the connection can be written by means of the normal fundamental forms [183]:

$$\omega_j^i(X) = \langle D_X n_i, n_j \rangle = \langle dn_i(X), n_j \rangle\,.$$

The above identities (3.16) and (3.17) follow from the fundamental equations for submanifolds (Gauss formulae and Weingarten equations). But since any vector field on Σ can be represented in terms of $\nabla_X Y$ and $\nabla_X \nu$, the mechanical contribution in (3.10) can be regarded as the normal fraction of the Hamiltonian vector field [14].

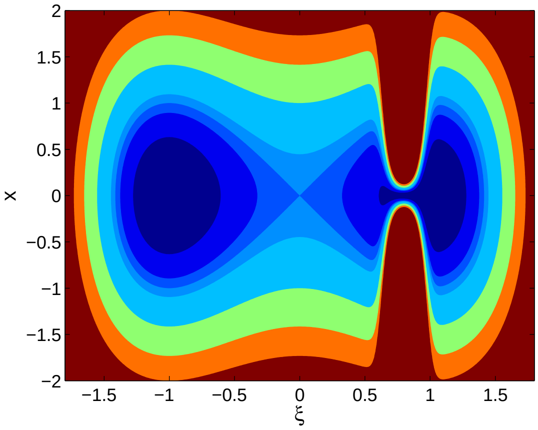


Figure 3. The plot shows the potential (3.19) with the dynamical barrier for the parameters $C = 15$, $\alpha = 200$, and $\xi_0 = 0.8$. The deep cut-away comes from the frequency peak of the harmonic oscillator.

Entropy, dynamical barriers It is about time coming to our first example which will guide us through the rest of this thesis: consider the Hamiltonian $H : T^*\mathbf{R}^n \to \mathbf{R}$

$$H(\xi, x, \zeta, u) = \frac{1}{2}\langle \zeta, \zeta\rangle + \frac{1}{2}\langle u, u\rangle + V_\epsilon(\xi, x)$$

with $\xi \in \mathbf{R}^k$, $x \in \mathbf{R}^d$, $d = n - k$ and the singularly perturbed interaction potential

$$V_\epsilon(\xi, x) = W(\xi) + \frac{1}{2\epsilon^2}\langle A(\xi)x, x\rangle \ ,$$

where $A \in \mathbf{R}^{d\times d}$ is an arbitrary symmetric, positive-definite (s.p.d.) matrix. Clearly the potential energy diverges as ϵ goes to zero. Observing that $V_\epsilon(\xi, x) = V_1(\xi, x/\epsilon)$, it is therefore convenient to introduce the scaled variables $x \mapsto \epsilon x$ in order to prevent the energy from blowing up. The scaling has a symplectic lift to the cotangent bundle that is given by $u \mapsto u/\epsilon$. The thus scaled Hamiltonian reads

$$H_\epsilon(\xi, x, \zeta, u) = \frac{1}{2}\langle \zeta, \zeta\rangle + \frac{1}{2\epsilon^2}\langle u, u\rangle + V_1(\xi, x)\,. \tag{3.18}$$

Physically speaking, the scaling has the effect that the second class of particles (with coordinates x) gets lighter as ϵ goes to zero. Therefore the particles get faster and faster, since the total energy remains finite. Accordingly we choose ξ as the reaction coordinate. The conditional density with respect to ξ is

$$\begin{aligned} Z(\xi) &= \int_{\mathbf{R}^d\times\mathbf{R}^n} \exp(-\beta H_\epsilon(\xi, x, \zeta, u))dxd\zeta du \\ &= \epsilon^d \left(\frac{2\pi}{\beta}\right)^{\frac{n+d}{2}} \left(\sqrt{\det A(\xi)}\right)^{-1} \exp(-\beta W(\xi))\,. \end{aligned}$$

Modulo constants the free energy becomes

$$F(\xi) = W(\xi) + \frac{1}{2\beta}\ln\det A(\xi)\,.$$

Now compare the free energy to the (conditional) internal energy of the system

$$U(\xi) = \mathbf{E}_\xi H_\epsilon(\xi, x, \zeta, u) = W(\xi) + \frac{d}{2\beta}\,,$$

where the conditional expectation is defined according to (3.8). From the last equality and equation (3.3) we directly obtain the Shannon entropy of the fast subsystem

$$S(\xi) = \frac{1}{2}(d - \ln\det A(\xi))\,.$$

Example 3.6. We shall exemplify the influence of the fast variables on the reaction coordinate in some more detail. Imagine the fast variables x represent the bond vibrations of a molecule, and ξ labels a conformational degree of freedom. Then it may happen that entropic effects from the bond vibrations alter the conformation dynamics. Let us carry the example above to the extremes, and set

$$V_1(\xi, x) = \frac{1}{4}(\xi^2 - 1)^2 + \frac{1}{2}\,\omega(\xi)^2 x^2 \tag{3.19}$$

with $\xi \in \mathbf{R}$, $x \in \mathbf{R}$ and a function $\omega(\xi) \geq c > 0$, which is defined as

$$\omega(\xi) = 1 + C\exp\left(-\alpha(\xi - \xi_0)^2\right)\,. \tag{3.20}$$

The potential function is shown in Figure 3. The frequency has a sharp peak at $\xi = \xi_0$ that induces a large force pointing towards the equilibrium manifold $x = 0$ (cf. Figure 4a). This has the effect that a particle which approaches ξ_0 with a large oscillation energy will bounce off the *dynamical barrier* that arises from the frequency peak, although the potential is almost flat this direction. In order to demonstrate the effect of the dynamical (or entropic) barrier we compute the free energy

$$F(\xi) = \frac{1}{4}(\xi^2 - 1)^2 + \beta^{-1}\ln\omega(\xi)\,. \tag{3.21}$$

which is depicted in Figure 4b. Apparently the entropic barrier in the full potential shows up as a potential barrier in the averaged potential. Nevertheless it is not a potential barrier in the usual sense, as it becomes harder and harder to cross it, if temperature $T = 1/\beta$ increases. In this sense the variation of bond frequencies results in entropic effects that may influence the conformational behaviour of a molecule.

3.1.2. Two distinct notions and the Fixman Theorem

We shall now come back to the problem of distinct notions of free energy. There is yet another quantity that circulates in the literature and which is often confused with the free energy (3.5):

$$G(\xi) = -\beta^{-1}\ln Z_\Sigma(\xi) \tag{3.22}$$

with

$$Z_\Sigma(\xi) = \int_{\Sigma\times\mathbf{R}^n} \exp(-\beta H)\, d\mathcal{H}_\xi\,. \tag{3.23}$$

This definition is quite important in the context of transition state theory [3, 4]. It has been shown [5] that the optimal dividing surface $\Sigma = \Phi^{-1}(\xi)$ that minimizes the transition rates between two sets over all hypersurfaces is a critical point of $G(\xi)$. Notice that the apparent difference to $F(\xi)$ lies in the matrix volume of the Jacobian J_Φ, which is not present here. The more subtle difference lies in the fact that G is intrinsically defined through the surface Σ, whereas F explicitly depends on the

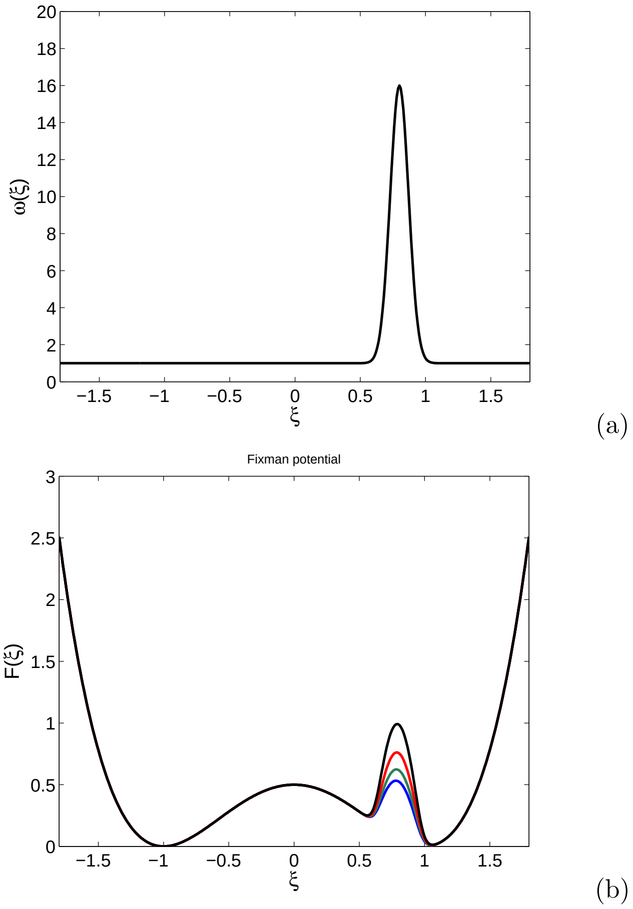


Figure 4. The oscillation frequency $\omega(\xi)$ and the free energy $F(\xi)$ are plotted — the latter for different inverse temperatures $\beta \in \{6.0, 5.0, 4.0, 3.0\}$ with the parameters $C = 15$, $\alpha = 200$, and $\xi_0 = 0.8$. Here $\beta = 3.0$ labels the highest peak at $\xi = \xi_0$, whereas the lowest one corresponds to $\beta = 6.0$, clearly indicating that the effect of the dynamical barrier becomes more and more important as temperature increases.

reaction coordinate Φ. This can be seen as follows: It is easy to recognize that we can switch between F and G by simply augmenting V with the Fixman potential W

$$V_\Phi(q) = V(q) + \underbrace{\beta^{-1} \ln \mathrm{vol} J_\Phi(q)}_{=:W(q)} . \tag{3.24}$$

Now suppose that we define a new reaction coordinate by $\Phi_g = g(\Phi)$ where g is a smooth, strictly monotonic function. Clearly $\Phi_g(q) = g(\xi)$ still defines the same surface Σ, so G is not altered. But since the Fixman potential

$$\beta^{-1} \ln \mathrm{vol} J_{\Phi_g} = \beta^{-1} \left(\ln \mathrm{vol} J_\Phi + \ln |\det \mathbf{D} g(\Phi)| \right)$$

depends on g, the free energy much depends on the reaction coordinate, viz.,

$$F(g(\xi)) = F(\xi) + \beta^{-1} \ln |\det \mathbf{D} g(\xi)| . \tag{3.25}$$

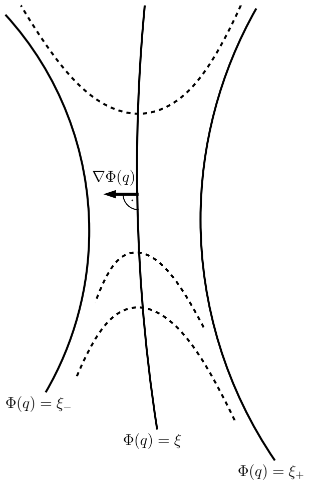


Figure 5. Giving some meaning to the Fixman potential $W=\beta^{-1}\ln\|\nabla\Phi\|$ for codimension-one submanifolds: The plot illustrates the squeezing of nearby level sets. The width of the harmonic confinement potential $W_\epsilon \propto (\Phi(q)-\xi)^2$ in the direction of the normal is of order $\|\nabla\Phi\|^2$ (see also Example 3.6).

We may call G the *geometric free energy* as it is invariant under transformations of the reaction coordinate. In contrast, we shall refer to F as the *standard free energy* or simply *free energy.* It can be readily checked that the corresponding Gibbs densities are related by a weighting factor in the way that

$$\exp(-\beta G(\xi)) = \mathbf{E}_\xi \mathrm{vol} J_\Phi(q)\,\exp(-\beta F(\xi))\,. \tag{3.26}$$

The Blue Moon relation The difference between F and G highlights another important aspect: In the seminal work [179] Fixman addressed the problem of how to compute unbiased averages for polymeric fluids that are subject to holonomic constraints. For instance, consider the objective of computing averages along certain prescribed reaction coordinates by Thermodynamic Integration methods. That is, the task is to compute the conditional expectation with respect to a reaction coordinate running constrained dynamics. This *bias problem* has been often understood in the sense that the bias were introduced by the integrability condition $\dot\Phi(q)=0$ (hidden constraint) that any system satisfies in addition to the reaction coordinate constraint $\Phi(q)=\xi$. This is certainly the case for a mechanical system if velocity- or momentum-dependent observables are considered. It is less known, however, that the bias problem remains if the dynamics is purely on configuration space, e.g., in case of Brownian motion. To understand this, recall the definition (3.8) of the conditional expectation. If we integrate out the momenta we have for an observable $f=f(q)$

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

where $d\sigma_\xi$ denotes the surface element of $\Sigma\subset\mathbf{R}^n$, and Q is the positional normalization constant. Suppose we want to compute the conditional expectation by imposing the constraint $\Phi(q)=\xi$ and averaging over the remaining variables. Of course, the constraint only specifies the submanifold $\Sigma=\Phi^{-1}(\xi)$ on which the system evolves; roughly speaking, the system knows its configuration manifold Σ but not the function Φ. The natural probability measure that is associated with the constrained system is therefore obtained by restricting the Gibbs measure to Σ. This defines another expectation that should be well distinguished from the conditional one:

$$\mathbf{E}_\Sigma f = \frac{1}{Q_\Sigma(\xi)} \int_\Sigma f \exp(-\beta V)\, d\sigma_\xi \,, \tag{3.27}$$

where Q_Σ is simply the configuration space version of (3.23). Equation (3.27) explains why averages that are computed subject to holonomic constraints differ from conditional expectations. In fact it is easy to see from the two definitions that

$$\mathbf{E}_\xi f = \frac{\mathbf{E}_\Sigma \left(f\, (\mathrm{vol} J_\Phi)^{-1}\right)}{\mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}} \,, \tag{3.28}$$

is the conditional expectation expressed by the constrained one. This identity which is known in the literature by the name of *Fixman theorem* or *Blue Moon ensemble method* holds true, no matter if the system involves momenta or not. Merging the Blue Moon relation together with equation (3.26) from above, we find a remarkably simple relation between F and G, namely

$$F(\xi) = G(\xi) - \underbrace{\beta^{-1} \ln \mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}}_{=:D(\xi)} \,. \tag{3.29}$$

This identity is remarkable, for we shall demonstrate in Section 4.3 below that the derivative of G can be written as an averaged force of constraint. That is, we can compute ∇G simply from quantities that are available anyway during the course of integration (Lagrange multipliers) with no need for extra reweighting. Once G is computed we obtain F by adding the term $D = -\beta^{-1} \ln \mathbf{E}_\Sigma (\mathrm{vol} J_\Phi)^{-1}$ which is also computed without any reweighting. For obvious reasons, the function D is called *Fixman potential*, too. As an additional treat the method does not involve second derivatives.

We can provide some physical interpretation of the Fixman potential W which is due to the work of van Kampen and Lodder on constraints [75]; see also [28, 17]. In some sense the Fixman potential mimics unconstrained dynamics, although the system is constrained. Consider a free dynamical system, either Brownian dynamics or stochastic Hamiltonian. Suppose we want to impose a constraint $\Phi(q) = \xi$ by adding a strong confining force that pushes a particle towards the surface $\Sigma = \Phi^{-1}(q)$ Imagine, this force is induced by the confinement potential

$$W_\epsilon(q) = \frac{1}{2\epsilon^2} \sum_{i=1}^{k} \left(\Phi_i(q) - \xi_i\right)^2 \,.$$

Letting ϵ become smaller and smaller while appropriately scaling the initial conditions (in order to prevent the energy from diverging) renders the particle to quickly oscillate around the constraint manifold Σ. The confinement potential has the property that its gully width orthogonal to Σ is of the order $(\mathrm{vol} J_\Phi)^2$; see Figure 5 for illustration. In case the dynamics is ergodic with respect to the canonical density the limit $\epsilon \to 0$ will result in: (i) confinement of the particle to the constraint manifold and (ii) an additional effective force, which is the derivative of the Fixman potential

$$W(q) = \beta^{-1} \ln \mathrm{vol} J_\Phi(q) \,.$$

In this sense adding the Fixman potential to a constrained system mimics unconstrained dynamics, by accounting for the influence of nearby level sets $\Phi(q) = \xi_\pm$

in Figure 5. This also explains why the standard free energy (which involves the Fixman potential) explicitly depends on the reaction coordinate $\Phi(q)$, whereas the geometric free energy depends only on the surface Σ. (This motivates the name *geometric free energy*.) Similar results for the microcanonical ensemble are available in the literature; see, e.g., [180, 27]. We refer to Section 3.4 for a detailed discussion of various confinement approaches.

We conclude by emphasizing that the two distinct notions of free energy, F and G, both have a configuration space analogue: since the reaction coordinate does not depend on the momenta at all, we have (modulo additive constants)

$$F(\xi) = -\beta^{-1} \ln Q(\xi) \tag{3.30}$$

with

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

for the standard free energy, and

$$G(\xi) = -\beta^{-1} \ln Q_\Sigma(\xi) \tag{3.31}$$

with

$$Q_\Sigma(\xi) = \int_\Sigma \exp(-\beta V) \, d\sigma_\xi$$

for the geometric free energy. Since the reaction coordinate is a purely configurational quantity, the thus defined free energies differ from the previously defined free energies (3.5) and (3.22) that were defined on phase space only by an additive constant.

Remark 3.7. *The traditional way in the literature to express the conditional expectation is in terms of the Dirac delta measure (e.g., see [71])*

$$\mathbf{E}_\xi f = \frac{1}{Q(\xi)} \int_{\mathbf{R}^n} f(q) \exp(-\beta V(q)) \delta(\Phi(q) - \xi) \, dq \,,$$

whereas the constrained average can be written as

$$\mathbf{E}_\Sigma f = \frac{1}{Q_\Sigma(\xi)} \int_{\mathbf{R}^n} f(q) \exp(-\beta V(q)) \delta(\Phi(q) - \xi) \, \mathrm{vol} J_\Phi(q) \, dq \,.$$

Accordingly the normalization constant Q_Σ reads

$$Q_\Sigma(\xi) = \int_{\mathbf{R}^n} \exp(-\beta V(q)) \delta(\Phi(q) - \xi) \, \mathrm{vol} J_\Phi(q) \, dq \,.$$

Comparing the last equations to each other, the assertion (3.28) follows as well. In an equal manner we could use the relation (3.24) to compute conditional expectations from constrained simulations by using the augmented potential V_Φ instead of V.

3.2. The Averaging Principle

Free energy profiles provide reduced statistical models for molecular system. A dynamical approach is the Method of Averaging which consists in replacing the full equations of motion by a reduced set of equations where certain degrees of freedom have been averaged out. The assertion that the trajectories of the reduced system are *close* to those of the original system is called the Averaging Principle. In its traditional formulation [184] it goes as follows: consider the initial value problem

$$\dot{z}_\epsilon(s) = \epsilon f(z_\epsilon(s), y(s)) \,, \quad z_\epsilon(0) = z$$

with uniformly Lipschitz continuous right hand side, where $y(t)$ is some forcing function. By continuity of the solution it follows that the limit solution for $\epsilon \to 0$ is constant on the interval $[0, T]$ for any fixed value of $T > 0$,

$$\lim_{\epsilon\to 0} z_\epsilon(s) = z \quad \forall s \in [0, T] .$$

Things change if we speed up time and consider the behaviour of the solution on an infinite time interval $[0, T/\epsilon]$. To this end we introduce the scaled variables $t = \epsilon s$ and $x_\epsilon(t) = z_\epsilon(t/\epsilon)$. Keeping in mind that $s \in [0, T/\epsilon]$ is equivalent to $t \in [0, T]$, we arrive at the classical averaging formulation

$$\dot{x}_\epsilon(t) = f(x_\epsilon(t), y(t/\epsilon)) , \quad x_\epsilon(0) = x , \tag{3.32}$$

where the initial value is independent of ϵ. This explains the idea of the fast dynamics as random perturbations, since $y(t/\epsilon)$ has now become a *fast* forcing function. Closing the last equation thus amounts to taking the limit $\epsilon \to 0$. Provided that $y(t)$ is ergodic with respect to probability measure μ, we can also close the equation by taking the ensemble average of the right hand side, i.e.,

$$\bar{f}(x) := \lim_{T\to\infty} \int_0^T f(x, y(t))\, dt = \int f(x, y)\mu(dy) .$$

If the integral exists, then $x_\epsilon(t) \to x_0(t)$ uniformly on compact time intervals $[0, T]$, and the limit solution $x_0(t)$ is governed by the averaged equation

$$\dot{x}_0(t) = \bar{f}(x_0(t)) , \quad x_0(0) = x .$$

For the convergence proof the reader is referred to the relevant literature [185, 24]. In the molecular dynamics case the forcing $y(t/\epsilon)$ in (3.32) is random and is the solution of the equations of motion for the fast variables. We can reformulate an analogous principle for the slow-fast system (3.1) from the last subsection,

$$\begin{aligned} \dot{x}_\epsilon(t) &= f(x_\epsilon(t), y_\epsilon(t), \epsilon) \\ \dot{y}_\epsilon(t) &= \frac{1}{\epsilon} g(x_\epsilon(t), y_\epsilon(t), \epsilon) . \end{aligned}$$

On the slow timescale the slow variables are effectively frozen, such that the fast dynamics (conditional on the slow variables) obeys the equation

$$\dot{y}_x(t) = g(x, y_x(t), 0) . \tag{3.33}$$

Let φ_t^x denote the respective *conditional* fast flow. That is, $y_x(t) = \varphi_t^x(y)$ is the solution of the last equation with initial value $y_x(0) = y$, where we use the subscript x to indicate the possible dependence on the slow variables. Assuming further that either φ_t^x is hyperbolic or mixing with unique invariant probability measure μ_x, then the conditional expectation of $f(x, \cdot)$ is uniquely defined [186, 187],

$$\bar{f}(x) = \lim_{T\to\infty} \int_0^T f(x, \varphi_t^x(y))\, dt = \int f(x, y)\mu_x(dy) , \tag{3.34}$$

provided the integral exists. In the molecular modelling case we face a very comfortable situation, since the equations of motion are either stochastic with non-degenerate noise matrix or Hamiltonian with randomized momenta. In any case the canonical invariant measure for the full system is unique, and so will be the conditional probability measure for the fast variables. Although the last statement may not be completely self-evident, we will show that the splitting into slow and fast variables can be carried out such as to maintain uniqueness of the invariant measure also for the fast dynamics.

The Averaging Principle is an assertion about the approximation properties of the averaged system on compact time intervals (observation time scale). If the right hand side of (3.32) averages to zero, then the dynamics of the accelerated system becomes trivial on the observation time scale. In this case the relevant dynamics happens on a longer time interval of order $1/\epsilon$ or even $\exp(-\epsilon)$, i.e., when fluctuations come into play. Averaging theorems for diverging time intervals can be found, e.g., in the work of Khas'minskii [33]. One such case is the high-friction limit of the Langevin equation. It has been claimed, however, that long-term corrections to the averaged equations (so-called diffusive limits) may become important even if the averaged dynamics is non-trivial on the observation time scale [34]. These authors notice that the *rareness* of the conformational transitions indicates that the relevant dynamics happens on time scales that lie beyond the observation time. There are two answers to this objection: First of all, we observe that the time scale of the transitions does not diverge as ϵ goes to zero (although, e.g., transition rates may change with ϵ). Hence conformation dynamics is essentially an $\mathcal{O}(1)$ effect. Moreover it seems that the methodology of diffusive limits is more targeted on systems with deterministic right hand side that is subject to random perturbations stemming from the fast variables. The problems considered in molecular dynamics are usually of a different type, but we will pick up this thread again in Section 6 below (see Remark 6.2).

Yet another open question up to now is whether the effective force $\bar{f}$ is somehow related to the free energy. In point of fact the free energy is also termed *potential of mean force*, and it is a common believe in the molecular dynamics community that the effective dynamics along a reaction coordinate is driven by the respective free energy.

Example 3.8. For the sake of illustration let us start with a simple (linear subspace) example: suppose the dynamics is given by a non-degenerate diffusion process,

$$\gamma\dot{q}(t) = -\nabla V(q(t)) + \sigma\dot{W}(t)$$

with $q = (x, y) \in \mathbf{R}^d \times \mathbf{R}^k$. Suppose further that the symmetric, positive-definite matrices γ, σ satisfy the fluctuation-dissipation relation $2\gamma = \beta\sigma\sigma^T$. In the Hamiltonian scenario timescale separation is often related to the mass ratio of fast and slow particles. For the Smoluchowski equation the situation is slightly different, since the equation of motion does not contain any masses. Now recall that in the elaboration upon covariant formulations of the Smoluchowski equation we have argued that $\gamma\dot{q}$ is an element of the cotangent space. That is, the friction matrix γ for diffusive motion takes over the role of the mass matrix for inertial motion. Let us assume for the moment that both friction and noise matrices are block diagonal,

$$\gamma = \begin{pmatrix} \gamma_1 & \mathbf{0} \\ \mathbf{0} & \gamma_2 \end{pmatrix}, \qquad \sigma = \begin{pmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \sigma_2 \end{pmatrix},$$

where each of the submatrices is proportional to the unit matrix (isotropy). In this case the equations of motion decay according to

$$\begin{aligned} \gamma_1\dot{x}(t) &= -\mathbf{D}_1 V(x(t), y(t)) + \sigma_1\dot{W}_1(t) \\ \gamma_2\dot{y}(t) &= -\mathbf{D}_2 V(x(t), y(t)) + \sigma_2\dot{W}_2(t) \end{aligned}$$

where $\mathbf{D}_1$, $\mathbf{D}_2$ denote the derivative with respect to the first and second slot. A simple comparison to (3.1) shows that we obtain the familiar slow-fast system by choosing $\gamma_2 = \epsilon\gamma_1$. Fluctuation-dissipation requires that $\sigma_2 = \sqrt{\epsilon}\sigma_1$, which yields for $\gamma_1 = \mathbf{1}$

$$\begin{aligned} \dot{x}_\epsilon(t) &= -\mathbf{D}_1 V(x_\epsilon(t), y_\epsilon(t)) + \sqrt{2\beta^{-1}}\dot{W}_1(t) \\ \dot{y}_\epsilon(t) &= -\frac{1}{\epsilon}\mathbf{D}_2 V(x_\epsilon(t), y_\epsilon(t)) + \sqrt{\frac{2\beta^{-1}}{\epsilon}}\dot{W}_2(t)\,. \end{aligned} \tag{3.35}$$

The invariant Gibbs measure $\mu \propto \exp(-\beta V)$ with $\beta = 2/\sigma_1^2$ is independent of ϵ as can be readily checked by substituting into the Kolmogorov forward equation. The conditional fast dynamics alone is obtained by switching to the slow timescale setting $t = \epsilon s$ and sending $\epsilon \to 0$. This yields the family of equations[9]

$$\dot{y}_x(s) = -\mathbf{D}_2 V(x, y_x(s)) + \sqrt{2\beta^{-1}}\dot{W}_2(s)\,.$$

This is a non-degenerate diffusion process. Hence it is certainly ergodic with respect to the conditional Gibbs measure, i.e., the Gibbs measure for fixed x,

$$\mu_x(dy) = \frac{1}{Q(x)}\exp(-\beta V(x,y))\,dy\,. \tag{3.36}$$

Letting ϵ in (3.35) going to zero, we obtain averaged equations of motion

$$\dot{x}_0(t) = -\nabla\bar{V}(x_0(t) + \sqrt{2\beta^{-1}}\dot{W}_1(t)\,,$$

where convergence in probability $x_\epsilon \to x_0$ is guaranteed by the Averaging Principle for stochastic processes [24, 184]. In our simple example the average force is

$$\nabla\bar{V}(x) = \int_{\mathbf{R}^k}\mathbf{D}_1 V(x,y)\mu_x(dy)$$

which turns out to be the derivative of both geometric or standard free energy. Here, the equivalence $F = G$ is owed to the fact that the reaction coordinate defines a linear subspace of the configuration space, such that the distinctive Jacobian term vanishes.

3.2.1. Averaging for linear reaction coordinates The following is basically a standard application of the Averaging Principle to molecular dynamics problems that involve a linear state space decomposition. In some sense it extends the ordinary Galerkin projection of first-order dynamical systems that is a popular reduction approach in the control community (e.g., [48, 188]). The crucial difference here is that the negligible degrees of freedom are averaged out rather than truncated. For an example of a Galerkin projection we refer to Example 3.23 below.

Let $\mathbf{R}^n$ be the Cartesian configuration space of our molecule with coordinates q, and assume we have applied any kind of spatial decomposition method (POD, PIP, ICA, ...) to $\mathbf{R}^n$. Let the k-dimensional (affine) dominant subspace found by any of these methods be denoted by $S \subset \mathbf{R}^n$, where S is characterized by a projection matrix $\mathscr{P} = PP^T$ (the k columns of P span the subspace S). The projection onto the orthogonal complement $S^\perp$ with respect to the Euclidean metric is denoted by $\mathscr{Q} = QQ^T$. Then $\mathscr{P} + \mathscr{Q} = \mathbf{1}$, and we have a unique decomposition of $\mathbf{R}^n$ due to

$$\mathscr{P}q \in S\,, \quad \mathscr{Q}q \in S^\perp\,.$$

Assume that the dynamics on S is slow as compared to the motion on $S^\perp$. We can define the respective slow and fast coordinates in the obvious way by $x = P^T q$ and $y = Q^T q$. Hence (x, y) form a complete set of new coordinates that are globally related to the Cartesian coordinates by $q = Px + Qy$, where x is the (linear) reaction coordinate. Since the slow-fast decomposition holds globally, we can easily get rid of the fast modes by simply averaging over the fast subspaces (fibres) $S_x^\perp \cong \mathbf{R}^{n-k}$ for

[9] The time scaling takes into account that the increments of the white noise are proportional to the square root of the time increments [117]. As a consequence the noise scales according to $\dot{W}(t) \mapsto \alpha\dot{W}(t/\alpha^2)$ under scaling transforms $t \mapsto t/\alpha$. Hence time scaling has the same effect as scaling the friction coefficient according to $\gamma \to \alpha\gamma$ subject to the condition $2\gamma = \beta\sigma\sigma^T$.

each value of the reaction coordinate. We assume that the dynamics is given by a diffusion process on the Euclidean configuration space $\mathbf{R}^n$,

$$\dot{q}(t) = -\mathrm{grad}\, V(q(t)) + \sqrt{2\beta^{-1}}\dot{W}(t)\,.$$

In terms of the new coordinates (x, y) we obtain the equations

$$\begin{aligned} \dot{x}_\epsilon(t) &= -\mathbf{D}_1 V(x_\epsilon(t), y_\epsilon(t)) + \sqrt{2\beta^{-1}}\dot{W}_1(t) \\ \dot{y}_\epsilon(t) &= -\frac{1}{\epsilon}\mathbf{D}_2 V(x_\epsilon(t), y_\epsilon(t)) + \sqrt{\frac{2\beta^{-1}}{\epsilon}}\dot{W}_2(t) \end{aligned} \tag{3.37}$$

with $\dot{W}_1 = P^T\dot{W}$ and $\dot{W}_2 = Q^T\dot{W}$, and $V(x, y) = V(Px + Qy)$. Note that we have already assigned the fast timescale to the second equation, where in contrast to the little example before the friction matrix is hidden in the scaled coordinates. Again the invariant Gibbs measure $\mu \propto \exp(-\beta V)$ is independent of ϵ, and for $\epsilon \to 0$ the fast process follows the conditional probability law

$$\mu_x(dy) = \frac{1}{Q(x)}\exp(-\beta V(x, y))\, dy$$

with the conditional partition function (normalization constant)

$$Q(x) = \int_{\mathbf{R}^k} \exp(-\beta V(x, y))\, dy\,.$$

The following averaging result is standard

Proposition 3.9 (Bogolyubov 1961)**.** *Assume that the integral*

$$\bar{f}(x) = -\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{D}_1 V(x, y_x(s))\, ds\,,$$

exists for all $x \in \mathbf{R}^k$, where $y_x(s)$ is the solution of the conditional fast flow

$$\dot{y}(s) = -\mathbf{D}_2 V(x, y(s)) + \sqrt{2\beta^{-1}}\dot{W}_2(s)\,.$$

Then as $\epsilon \to 0$ the solution $x_\epsilon(t)$ of the system of equations (3.37) converges in probability to a Markov process $x_0(t)$ that is governed by the equation

$$\dot{x}_0(t) = \bar{f}(x_0(t)) + \sqrt{2\beta^{-1}}\dot{W}_1(t)\,, \tag{3.38}$$

where for $T > 0$, $\delta > 0$

$$\lim_{\epsilon\to 0}\mathbf{P}\left[\sup_{0\le t\le T}|x_\epsilon(t) - x_0(t)| > \delta\right] = 0\,.$$

For the proof the reader is referred to the relevant literature, e.g., [24, 185]. In this simple case it is easy to recognize that the free energy is indeed directly related to the averaged equations of motion. Since the conditional fast process is ergodic with respect to the conditional probability measure $\mu_x(dy)$ as is defined above, we can express the averaged vector field $\bar{f}(x)$ as the conditional expectation

$$\bar{f}(x) = -\int_{\mathbf{R}^k}\mathbf{D}_1 V(x, y)\mu_x(dy)\,.$$

The last equation reveals that the mean force $\bar{f} = -\nabla\bar{V}$ has a potential

$$\bar{V}(x) = -\beta^{-1}\ln\int_{\mathbf{R}^k}\exp(-\beta V(x, y))\, dy\,, \tag{3.39}$$

that is formally equivalent to both of the two free energies F or G, respectively.

A note about free energy as an averaging concept In the last example we could observe that the averaged dynamics was driven by the negative gradient of the free energy which explains why the (standard) free energy is sometimes termed *potential of mean force*. However we have to be careful, since according to equation (3.25) the derivative of the free energy neither transforms as a gradient field nor as a 1-form, i.e., a force. Moreover we have seen in Lemma 3.3 that the derivative of the free energy contains a pseudo force that has no straightforward dynamical interpretation, in case the essential variables do not span a linear subspace of the configuration space but rather a general Riemannian submanifold. Consequently we cannot expect that the free energy will provide the driving force of a general reaction coordinate dynamics.

A very simple argument convinces us that the standard free energy cannot be the right quantity to look at: consider the last example, where $x \in \mathbf{R}$ is one-dimensional. The reduced system in terms of the averaged force $\partial_x \bar{V}$ reads

$$\dot{x}(t) = -\partial_x \bar{V}(x(t)) + \sqrt{2\beta^{-1}}\dot{W}(t)\,.$$

Suppose we perform a change of coordinates, and we define a new coordinate z by $x = f(z)$. Expressing the equation of motion in terms of z using Lemma 2.11 yields

$$\dot{z} = -\frac{1}{f'(z)^2}\partial_z \bar{V}(f(z)) - \beta^{-1}\frac{f''(z)}{f'(z)^3} + \frac{1}{f'(z)}\sqrt{2\beta^{-1}}\dot{W}\,. \tag{3.40}$$

Now recall that the free energy carries some gauge dependence (3.25). That is,

$$F(f(z)) = F(z) + \beta^{-1}\ln f'(z)\,.$$

Hence for $\bar{V}(x) = F(x)$ we would obtain the transformed equation

$$\dot{z} = \, -\frac{1}{f'(z)^2}\partial_z F(z) - 2\beta^{-1}\frac{f''(z)}{f'(z)^3} + \frac{1}{f'(z)}\sqrt{2\beta^{-1}}\dot{W}\,, \tag{3.41}$$

which is different from (3.40) in general. Thus: although it may be that $\bar{V} = F$ holds true formally (and so does $G = F$ for the geometric free energy) the transformation properties of the standard free energy do not qualify its derivative as an averaged force. We leave it open to the reader to convince oneself that (3.41) is not an Itô equation (e.g., by choosing $\bar{V}(x) = x^2$ and $f(z) = z^2$).

3.2.2. Nonlinear reaction coordinate dynamics Presumably free energy landscapes do not appropriately describe the dynamics along arbitrary reaction coordinates, since their gradients do not transform like ordinary vector fields. Now consider a smooth reaction coordinate $\phi : \mathbf{R}^m \to \mathbf{R}^k$, and suppose we can globally decompose the system under consideration into a set of slow variables $\phi \in \mathbf{R}^k$ and another set of fast variables, say, $z \in \mathbf{R}^{m-k}$. This system will be of the form

$$\begin{aligned}\dot{\phi}_\epsilon(t) &= f(\phi_\epsilon(t), z_\epsilon(t), \epsilon)\\ \dot{z}_\epsilon(t) &= \frac{1}{\epsilon}g(\phi_\epsilon(t), z_\epsilon(t), \epsilon)\,.\end{aligned}$$

On condition that the fast dynamics for each value of the reaction coordinate $\phi = \xi$

$$\dot{z}_\xi(t) = g(\xi, z_\xi(t), 0)$$

is well-posed and admits a unique invariant measure, the Averaging Principle states that $\phi_\epsilon(t)$ converges in some appropriate sense to a limit process $\phi_0(t)$ as $\epsilon \to 0$.

The difficulty in setting up the slow-fast system is that it relies on a global change of coordinates which is hopeless for a general state space. However we observe that the equation for the fast dynamics and the conditional invariant measure are defined only locally for $\phi = \xi$. Noting that $\phi(\cdot) = \xi$ with ξ taking values in $\mathbf{R}^k$ defines a foliation of $\mathbf{R}^m$, we propose to decompose the full system into a family of slow-fast systems

$$\begin{aligned} \dot{y}_\epsilon(t) &= f_\xi(y_\epsilon(t), z_\epsilon(t), \epsilon) \\ \dot{z}_\epsilon(t) &= \frac{1}{\epsilon} g_\xi(y_\epsilon(t), z_\epsilon(t), \epsilon)\,, \end{aligned}$$

where the vector fields f_ξ, g_ξ are defined locally in a tubular neighbourhood of each fibre $\Sigma = \phi^{-1}(\xi)$. (This coordinate construction is explained in the appendix). The slow coordinates $y \in \mathbf{R}^k$ are intended to describe the dynamics orthogonal to each fibre. Averaging over over the fast variables then yields a family of vector fields

$$\bar{f}(\xi) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f_\xi(y_0 = 0, z_{\xi,0}(t), 0)\, dt\,,$$

that are defined fibre-wise for $\phi(\cdot) = \xi$, where $z_{\xi,0}(t)$ is the solution of the fast dynamics on each fibre. The effective dynamics of the reaction coordinate can be reconstructed by endowing the reaction coordinate space with an appropriate metric. To some extend the approach presented here can be considered a variant of the *accelerated dynamics* or *metadynamics* that is put forward in [13]; cf. also [189]. However, the local decomposition of state space here allows for a lucid physical and geometrical interpretation of the limit equation. This proves useful in designing algorithms that efficiently sample the coefficients of the reduced equation.

Unfortunately the standard Averaging Principle does not apply, since we can only study the local convergence to initial values on each fibre. Averaging over the initial values then gives the average vector field in the vicinity of the fibre but no dynamical information whatsoever, since the motion cannot leave the tubular neighbourhood. Therefore we warn the reader that the calculation is purely formal. Nevertheless we shall support the claims to be made by appropriate numerical examples later on.

Accelerating Brownian motion Let $V : \mathbf{R}^n \to \mathbf{R}$ be a smooth potential that is bounded from below, and let $\sigma > 0$ be scalar. The Smoluchowski equation reads

$$\dot{q}(t) = -\mathrm{grad}\, V(q(t)) + \sigma \dot{W}(t)\,.$$

Given a reaction coordinate $\Phi : \mathbf{R}^n \to \mathbf{R}^s$, the level sets of which define smooth configuration submanifolds of codimension s, we denote by $\sigma_\xi : \mathbf{R}^{n-s} \to \Sigma_\xi$ the embedding $\Sigma_\xi = \Phi^{-1}(\xi)$ into $\mathbf{R}^n$. To each $\sigma_\xi \in \Sigma_\xi$ we attach a set of normal vectors $(n_1(\sigma_\xi), \ldots, n_s(\sigma_\xi))$, and we introduce local coordinates z^α, $\alpha = 1, \ldots, n-s$ on Σ_ξ, and normal coordinates y^i, $i = 1, \ldots, s$ that measure the distance to Σ_ξ with respect to the normal frame $\{n_1, \ldots, n_s\}$. Fixing ξ, the original coordinates can be uniquely expressed in a sufficiently small tubular ε-neighbourhood $N\Sigma_{\xi,\varepsilon}$ of Σ_ξ by the map

$$q = \phi_\xi(z, y)\,, \quad \phi_\xi : (z, y) \mapsto \sigma_\xi(z) + y^i n_i(\sigma_\xi(z))\,.$$

According to (B.2) the Euclidean metric has the local coordinate expression

$$g_\xi(z, y) = \begin{pmatrix} G_\xi(z) + C_\xi(z, y) & A_\xi(z, y) \\ A_\xi(z, y)^T & \mathbf{1} \end{pmatrix}.$$

All local coordinate expressions, and the particular submatrices $G_\xi, C_\xi \in \mathbf{R}^{(n-s)\times(n-s)}$ or $A_\xi \in \mathbf{R}^{(n-s)\times s}$ are given in Appendix B. Note that all quantities depend parametrically on the value ξ of the reaction coordinate by virtue of the particular embedding of the normal bundle $N\Sigma_{\xi,\varepsilon}$ into $\mathbf{R}^n \times \mathbf{R}^n$. In local coordinates the Smoluchowski equation becomes (see Lemma 2.11)

$$\begin{aligned}
\dot{y}^i_\epsilon &= -g^{il}_\xi(z_\epsilon, y_\epsilon)\,\partial_l V_\xi(z_\epsilon, y_\epsilon) + b^i_\xi(z_\epsilon, y_\epsilon) + \sigma a^{il}_\xi(z_\epsilon, y_\epsilon)\dot{W}_l \\
\dot{z}^\alpha_\epsilon &= -\frac{1}{\epsilon} g^{\alpha l}_\xi(z_\epsilon, y_\epsilon)\,\partial_l V_\xi(z_\epsilon, y_\epsilon) + \frac{1}{\epsilon} b^\alpha_\xi(z_\epsilon, y_\epsilon) + \frac{\sigma}{\sqrt{\epsilon}} a^{\alpha l}_\xi(z_\epsilon, y_\epsilon)\dot{W}_l\,.
\end{aligned}$$

Note that the equations are only meaningful up to the first exit time from $N\Sigma_{\xi,\varepsilon}$. Moreover we have employed the following notation: $V_\xi = V \circ \phi_\xi$, and the function $b^h_\xi = -\beta^{-1} g^{kl}_\xi \Gamma^h_{\xi,kl}$ denotes the additional Itô drift term, whereas a^{kl}_ξ are the entries of the uniquely defined positive-definite matrix square root of g^{-1}_ξ. The symbol ∂_l is a shorthand for the partial derivatives with respect to z^α and y^i, respectively.[10]

By having assigned appropriate powers of ϵ to the equation of the fast variables, we force the dynamics tangential to the fibre Σ_ξ to be fast as compared to the orthogonal dynamics of the y^i (reaction coordinate dynamics); see 3.35 for comparison. For all $\epsilon > 0$ this system has an invariant Gibbs measure that is given by

$$\mu_\xi(dz, dy) = \frac{1}{Z_\xi} \exp(-\beta V_\xi(z, y)) \det g_\xi(z, y)\, dz dy\,. \tag{3.42}$$

The independence of ϵ can be easily verified by inserting the last expression into the Kolmogorov forward equation. Now we can repeat the time rescaling argument to see that on the microscopic timescale the equations read

$$\begin{aligned}
\dot{y}^i_\epsilon &= -\epsilon g^{il}_\xi(z_\epsilon, y_\epsilon)\,\partial_l V_\xi(z_\epsilon, y_\epsilon) + \epsilon b^i_\xi(z_\epsilon, y_\epsilon) + \sigma\sqrt{\epsilon} a^{il}_\xi(z_\epsilon, y_\epsilon)\dot{W}_l \\
\dot{z}^\alpha_\epsilon &= -g^{\alpha l}_\xi(z_\epsilon, y_\epsilon)\,\partial_l V_\xi(z_\epsilon, y_\epsilon) + b^\alpha_\xi(z_\epsilon, y_\epsilon) + \sigma a^{\alpha l}_\xi(z_\epsilon, y_\epsilon)\dot{W}_l\,.
\end{aligned}$$

Following [184] we obtain convergence to the initial value $y_\epsilon(t) \to y_0$ as $\epsilon \to 0$, where the restriction to the level set $\Phi^{-1}(\xi)$ clearly amounts to $y_0 = 0$. Using the formulae for the Christoffel symbols from Appendix B we obtain for the fast dynamics

$$\dot{z}^\alpha = -G^{\alpha\beta}_\xi(z)\,\partial_\beta V_\xi(z, 0) + b^\alpha_\xi(z, 0) + \sigma E^{\alpha\beta}_\xi(z)\dot{W}_\beta\,,$$

where

$$b^\alpha_\xi(z, 0) = -\beta^{-1} G^{\beta\gamma}_\xi(z)\Gamma^\alpha_{\xi,\beta\gamma}(z, 0)\,.$$

Here the $\Gamma^\alpha_{\xi,\beta\gamma}$ are the Christoffel symbols associated with the metric G_ξ on Σ_ξ, and E_ξ is the unique positive-definite matrix square root of G^{-1}_ξ. All other terms vanish at $y = 0$ since both $g^{\alpha i}_\xi = 0$ and $\Gamma^\alpha_{\xi,ij} = 0$. Hence the last equation is the local version for the intrinsic motion on Σ_ξ. Therefore, and according to Section 2.3, the invariant measure is the ordinary Gibbs measure (3.42) restricted to the fibre. That is,

$$\nu_\Sigma(dz) = \frac{1}{Q_\Sigma} \exp(-\beta V(\sigma_\xi(z)) \det G_\xi(z)\, dz\,. \tag{3.43}$$

Let us denote the right hand side of the slow equations of motion by

$$f^i_\xi(z, y) = -g^{il}_\xi(z, y)\,\partial_l V_\xi(z, y) + b^i_\xi(z, y) + \sigma a^{il}_\xi(z, y)\dot{W}_l\,.$$

Now averaging fibre-wise over the fast variables yields the static right hand side

$$\bar{f}^i(\xi) = \int \Big(b^i_\xi(z, 0) - g^{il}_\xi(z, 0)\partial_l V(z, 0) + \sigma a^{il}_\xi(z, 0)\dot{W}_l\Big)\, \nu_\Sigma(dz)\,.$$

[10] Note that there is some ambiguity in the use of the index i, as i is supposed to run from 1 to s whenever it indicates a normal coordinate as in y^i, but i also is considered as taking integer values from $n - s + 1$ to n, for instance, when labelling general vectors or matrices like $g^{\alpha i}$. Moreover the indices h, k, l run from 1 to n, whereas i, j only label the normal directions $1, \ldots, s$. We hope that their use will be clear from the particular context.

Employing the expressions in (B.6) for the Christoffel symbols and for the metric at $y = 0$, the mean vector field and the noise term get a considerably simpler form

$$\begin{aligned}\bar{f}^i(\xi) &= \int \Big(\beta^{-1} G_\xi^{\alpha\beta}(z) S^i_{\alpha\beta}(\sigma_\xi(z)) - \delta^{ij}\partial_j V_\xi(z,0)\Big)\,\nu_\Sigma(dz) + \sigma \dot{W}^i \\ &= \int \big(\beta^{-1}\kappa_{\xi,i}(z) - \delta^{ij}\partial_j V_\xi(z,0)\big)\,\nu_\Sigma(dz) + \sigma\dot{W}^i\,.\end{aligned} \tag{3.44}$$

The functions $\kappa_{\xi,i}(z)$ in the last row are the single components of the extrinsic mean curvature vector of Σ_ξ in $\mathbf{R}^n$ that is introduced in the following: Let $P_T : T_\sigma\mathbf{R}^n \to T_\sigma\Sigma_\xi$ denote the point-wise projection onto the tangent space to Σ_ξ, and recall the definition of the Weingarten maps $\mathfrak{S}_i = -P_T dn_i(\cdot)$ associated with the second fundamental form. The mean curvature vector H_ξ is defined as [190]

$$H_\xi(z) = \sum_{i=1}^{s} \kappa_{\xi,i}(z) n_i(\sigma_\xi(z))\,, \quad \kappa_{\xi,i} = -\operatorname{tr}\mathfrak{S}_i\,.$$

Reconstruction of the global dynamics We consider the deterministic part of $\bar{f}^i(\xi)$ as a force field on $\mathbf{R}^s$ by virtue of its parametric dependence on ξ and by identifying $T\mathbf{R}^s$ with $T^*\mathbf{R}^s$. Hence it remains to turn the stochastic force with respect to y into a force hat acts with respect to the reaction coordinate Φ. This is done so by endowing the limit system with an appropriate metric. To this end bear in mind that it follows from the Tubular Neighbourhood Theorem [191] that sufficiently close to the fibres $\Sigma = \Phi^{-1}(\xi)$ the uniquely invertible relation between the normal coordinate y and the reaction coordinate $r = \Phi$ and is given by

$$r = J_\Phi(\sigma_\xi(z))^T Q(\sigma_\xi(z))y + \xi\,,$$

where J_Φ denotes the Jacobian of Φ, and the columns of Q are the normal vectors $(n_1, \ldots, n_k)$. For each $\sigma \in \Sigma_\xi$ this transformation induces a metric on the normal space $N_{\sigma,0}\Sigma_\xi$, that is given by $m_\xi(z) = (J_\Phi^T J_\Phi)(\sigma_\xi(z))^{-1}$. By averaging over the fast variables with respect to their invariant distribution we can define an metric as follows

$$m(\xi) = \int m_\xi(z)\,\nu_\Sigma(dz)\,. \tag{3.45}$$

Notice that the deterministic part of (3.44) can be brought into the form

$$d^i(\xi) = \beta^{-1}\frac{\partial}{\partial y^i}\ln\int_{\Phi^{-1}(\xi)} \exp(-\beta V_\xi(z,y))\sqrt{\det g_\xi(z,y)}\,dz\Bigg|_{y=0}\,.$$

The averaged stochastic part is simply additive noise in the direction of the reaction coordinate. Hence we may write the *naked* reaction coordinate dynamics as

$$\dot{\xi}^i(t) = d^i(\xi(t)) + \sigma\dot{W}_i(t)\,,$$

which is ordinary diffusion in $\mathbf{R}^s$ with respect to the Euclidean metric. If we equip our configuration space $\mathbf{R}^s$ with the averaged metric $m(\xi)$ that comes along with the reaction coordinate, we obtain the global form of the averaged equations

$$\dot{\xi}^i(t) = -m^{ij}(\xi(t))\partial_j G(\xi(t)) + b^i(\xi(t)) + \sigma h^{ij}(\xi(t))\dot{W}_j(t)\,, \tag{3.46}$$

where h is the unique matrix square root of the inverse metric m^{-1}, and G is the geometric free energy (which should not be confused with the metric tensor G_ξ)

$$G(\xi) = -\beta^{-1}\ln Q_\Sigma(\xi) \quad \text{with} \quad Q_\Sigma(\xi) = \int_{\Phi^{-1}(\xi)} \exp(-\beta V)\,d\sigma_\xi$$

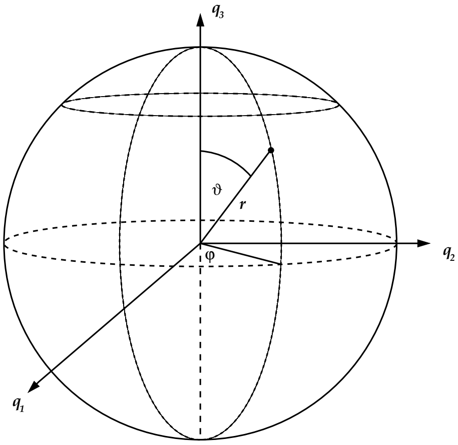


Figure 6. Spherical polar coordinates $(\varphi, \vartheta, r) \in S^2 \times \mathbf{R}_+$.

The additional term b is the usual Itô equation drift

$$b^i(\xi) = -\beta^{-1} m^{jk}(\xi) \bar{\Gamma}^i_{jk}(\xi)\,,$$

where $\bar{\Gamma}^i_{jk}$ are the Christoffel symbols associated with the metric m,

$$\bar{\Gamma}^i_{jk} = \frac{1}{2} m^{il} \left(\frac{\partial m_{jl}}{\partial \xi^k} + \frac{\partial m_{kl}}{\partial \xi^j} - \frac{\partial m_{jk}}{\partial \xi^l} \right) .$$

We emphasize that our approach is not unique, since it relies on an arbitrary manipulation of the equations of motion, speeding up the dynamics on the fibres. There is yet another possibility to accelerate the dynamics orthogonal to the reaction coordinate using a projection operator approach. For a single reaction coordinate the authors of [13] derive a representation that involves the free energy F

$$\dot{\xi}(t) = a(\xi(t)) F'(\xi(t)) + \beta^{-1} a'(\xi(t)) + \sigma \sqrt{a(\xi(t))}\dot{W}(t)\,, \tag{3.47}$$

where the metric factor a is defined as the conditional expectation

$$a(\xi) = \mathbf{E}_\xi \|\nabla \Phi(q)\|^2\,,$$

which should be distinguished from the expectation with respect to ν_Σ (compare equation (3.28)). It is not obvious that (3.47) really transforms like an Itô equation, as it does not have the standard covariant form (2.30). However it has been demonstrated that (3.47) is consistent with Itô formula under transformations of the reaction coordinate. Since this is also true for (3.46) one could expect that the two equations are equivalent. Intriguingly this is not the case, unless $\nabla\Phi$ is a function of ξ only, since then $a = m^{-1}$ (see the examples below). Presumably the difference in the result is owed to the fact that the authors of [13] organize the decomposition along the probability measures (gluing together different conditional measures), whereas we have endowed a decomposition of the state space (based on the foliation defined by Φ).

Example 3.10. Let us illustrate how the local averaging scheme works by means of an example. Consider the three-dimensional diffusion equation

$$\dot{q}(t) = -\mathrm{grad}\, V(q(t)) + \sigma \dot{W}(t)\,, \quad V(q) = V_0(\|q\|) + \delta(q)$$

where $\|\cdot\|$ is the Euclidean vector norm in $\mathbf{R}^3$. The potential V is bounded from below and is such that the first term defines the slow motion in the system, i.e., $|V_0| \ll |\delta|$. In this case there is a natural choice for the reaction coordinate

$$\Phi_1(q) = \|q\| = \sqrt{q_1^2 + q_2^2 + q_3^2}\,.$$

We first go through the reduction procedure using a global change of coordinates and then compare it to the local approach. The form of the problem suggests to use spherical polar coordinates. We introduce coordinates $(\varphi, \vartheta, r) \in S^2 \times \mathbf{R}_+$ by

$$\begin{aligned} q_1 &= r\cos\varphi\sin\vartheta\,, & r &\geq 0 \\ q_2 &= r\sin\varphi\sin\vartheta\,, & 0 &\leq \varphi < 2\pi \\ q_3 &= r\cos\vartheta\,, & 0 &\leq \vartheta \leq \pi\,, \end{aligned} \tag{3.48}$$

and therefore consider $NS^2 \cong S^2 \times \mathbf{R}_+$ as our new configuration space (see Figure 6). Pulling back the Euclidean metric to $S^2 \times \mathbf{R}_+$ induces the metric

$$h(\vartheta, r) = \begin{pmatrix} r^2\sin^2\vartheta & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} =: \begin{pmatrix} G(\vartheta, r) & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix},$$

where we have introduced the metric $G(\vartheta, r) = r^2 G_1(\vartheta)$ for the upper left 2×2 block of the full matrix, where $G_1(\vartheta)$ is the local metric on the unit 2-sphere S^2. Clearly $r = \Phi_1(q)$ is the reaction coordinate. The corresponding slow-fast system reads

$$\begin{aligned} \dot\omega_\epsilon^\alpha &= -\frac{1}{\epsilon}G^{\alpha\beta}(\vartheta_\epsilon, r_\epsilon)\partial_\beta V(\omega_\epsilon, r_\epsilon) + \frac{1}{\epsilon}b^\alpha(\vartheta_\epsilon, r_\epsilon) + \frac{\sigma}{\sqrt{\epsilon}}A^{\alpha\beta}(\vartheta_\epsilon, r_\epsilon)\dot W_\beta \\ \dot r_\epsilon &= -\partial_r V(\omega_\epsilon, r_\epsilon) + b^r(\vartheta_\epsilon, r_\epsilon) + \sigma\dot W\,, \end{aligned}$$

where $\omega = (\varphi, \vartheta)$ and $b^l = \beta^{-1}h^{jk}\Gamma^l_{jk}$. The noise amplitude $A = r^{-1}A_1$ is the positive-definite matrix square root of the inverse metric $G^{-1} = r^{-2}G_1^{-1}$. On the microscopic timescale $s = t/\epsilon$, we have convergence $r_\epsilon \to r$ for ϵ going to zero, such that the fast dynamics for frozen r is governed by the equation

$$\dot\omega_\epsilon^\alpha = -G^{\alpha\beta}(\vartheta_\epsilon, r_\epsilon)\partial_\beta V(\omega_\epsilon, r) + b^\alpha(\omega_\epsilon, r) + \sigma A^{\alpha\beta}(\vartheta_\epsilon, r_\epsilon)\dot W_\beta\,.$$

Notice that the fast dynamics is intrinsic to S_r^2 (the 2-sphere with radius r), since

$$\Gamma^\varphi_{rr} = \Gamma^\vartheta_{rr} = 0\,.$$

That is, the additional Itô drift $b^\alpha = -\beta^{-1}G^{\gamma\delta}\Gamma^\alpha_{\gamma\delta}$ depends only on the local metric G. Hence the conditional invariant measure of the fast process is simply given by the appropriately normalized Gibbs measure on the sphere S_r^2

$$\nu_r(d\omega) = \frac{1}{Q_{S_r^2}(r)}\exp(-\beta V(\omega, r))\sqrt{\det G(\omega, r)}\,d\omega\,.$$

The slow dynamics is governed by the equation

$$\dot r_\epsilon = -\partial_r V(\omega_\epsilon, r_\epsilon) + b^r(\omega_\epsilon, r_\epsilon) + \sigma\dot W$$

with

$$b^r = -\beta^{-1}h^{kl}\Gamma^r_{kl} = -\beta^{-1}\left(G_r^{\alpha\gamma}\Gamma^r_{\alpha\gamma} + \Gamma^r_{rr}\right)$$

and the Christoffel symbols

$$\Gamma^r_{\varphi\varphi} = -r\sin^2\vartheta\,,\quad \Gamma^r_{\vartheta\vartheta} = -r\,,\quad \Gamma^r_{rr} = 0\,.$$

By ergodicity of the fast process with respect to ν_r and application of the Averaging Principle we obtain convergence $r_\epsilon \to r_0$ as $\epsilon \to 0$. The limit process obeys

$$\dot{r}_0(t) = -\partial_r \bar{V}(r_0(t)) + \frac{2}{\beta r_0(t)} + \sigma \dot{W}(t)\,, \tag{3.49}$$

where the averaged potential is given by

$$\bar{V}(r) = V_0(r) + \int \delta(\omega, r)\nu_r(d\omega)\,. \tag{3.50}$$

We can obtain the same limit result by using the local embedding $N\Sigma \subset \mathbf{R}^3 \times \mathbf{R}^3$ with $\Sigma = S^2_\xi$. This can be seen as follows: As a first step consider the 2-sphere with radius ξ, that is defined by the reaction coordinate $\Phi_1(q) = \xi$. A local embedding $\sigma_\xi : S^2 \to S^2_\xi \subset \mathbf{R}^3$ is given by polar coordinates with fixed radius $r = \xi$

$$\begin{aligned} \sigma^1_\xi &= \xi \cos\varphi \sin\vartheta\,, && \xi \geq 0 \\ \sigma^2_\xi &= \xi \sin\varphi \sin\vartheta\,, && 0 \leq \varphi < 2\pi \\ \sigma^3_\xi &= \xi \cos\vartheta\,, && 0 \leq \vartheta \leq \pi\,. \end{aligned}$$

The next step is to construct a normal frame, for instance, by

$$n(\sigma_\xi(\varphi,\vartheta)) = \nabla\Phi_1(\sigma_\xi(\varphi,\vartheta)) = \sigma_1(\varphi,\vartheta)\,.$$

Since $\|\nabla\Phi_1(\sigma_\xi)\| = 1$ the normal coordinates that measure the distance to the surface S^2_ξ are simply given by $y = \Phi_1 - \xi$. In local coordinates (φ,ϑ,y) the metric tensor is

$$g_\xi(\varphi,\vartheta,y) = \begin{pmatrix} G_\xi(\vartheta) + C_\xi(\varphi,\vartheta,y) & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix},$$

where the local surface metric $G_\xi = \xi^2 G_1$ is defined as above, and

$$C_{\xi,\alpha\beta} = 2y\left\langle \partial_\alpha\sigma_\xi, dn(\partial_\beta\sigma_\xi)\right\rangle + y^2\left\langle dn(\partial_\alpha\sigma_\xi), dn(\partial_\beta\sigma_\xi)\right\rangle\,.$$

We can easily compute the matrix of the Weingarten map and the respective mean curvature. For the Weingarten map we have the expression

$$\mathfrak{S}_\xi(\varphi,\vartheta) = -dn(\cdot) = -\xi^{-1}P_T\,,$$

where $P_T : T_\sigma\mathbf{R}^n \to T_\sigma S^2$, $P_T = \mathbf{1} - n\left\langle n,\cdot\right\rangle$ is the point-wise projection onto the tangent plane to the unit sphere. Also the mean curvature is easily computed: Since all tangent spaces $T_\sigma S^2$ are two-dimensional, the projector P_T has rank 2. Thus

$$\kappa_\xi = -\operatorname{tr}\mathfrak{S}_\xi = \frac{2}{\xi}$$

which is the mean curvature of a 2-sphere in $\mathbf{R}^3$ with radius ξ. Using the result from the last subsection, the locally averaged equations take the form

$$\dot{\xi}(t) = -\partial_\xi \bar{V}(\xi(t)) + \frac{2}{\beta\xi(t)} + \sigma\dot{W}(t)\,, \tag{3.51}$$

where the averaged potential is given by

$$\bar{V}(\xi) = V_0(\xi) + \int \delta(\omega,\xi)\nu_\Sigma(d\omega)\,. \tag{3.52}$$

Since $r = y + \xi$ in this particular case, (3.51) equals already the global equation (3.49). In terms of the geometric free energy G the limit equation thus reads

$$\dot{r}(t) = -\partial_r G(r(t)) + \sigma\,\dot{W}(t)$$

which is full agreement with (3.46).

Example 3.11. One might imagine that the reaction coordinate is defined by

$$\Phi_2(q) = \|q\|^2 = q_1^2 + q_2^2 + q_3^2\,,$$

which is also a frequently used reaction coordinate for distance-based problems. Let us denote $\rho = \Phi_2$. Transforming the averaged equation (3.49) to an equation for $\rho = r^2$ is straightforward: we find with (3.49) and Lemma 2.11

$$\dot\rho(t) = -4\rho(t)\,\partial_\rho \bar V\left(\sqrt{\rho(t)}\right) + \frac{6}{\beta} + 2\sigma\sqrt{\rho(t)}\,\dot W(t)\,, \tag{3.53}$$

where we have used that the Christoffel symbol Γ^r_{rr} transforms like [81]

$$\Gamma^\rho_{\rho\rho} = \left(\frac{\partial r}{\partial \rho}\right)^2 \Gamma^r_{rr}\frac{\partial \rho}{\partial r} + \frac{\partial \rho}{\partial r}\frac{\partial^2 r}{\partial \rho^2} = -\frac{1}{2\rho}$$

Other than in the equation for r, we have $\Gamma^\rho_{\rho\rho} \neq 0$ which, in fact, yields the correct limit equation as would be obtained by using modified polar coordinates from the outset (replacing r by $\sqrt{\rho}$), and then stepping through the averaging procedure. The same equation is obtained by endowing the local limit equation (3.51) with the metric $m(\rho) = (4\rho)^{-1}$ that is induced by the reaction coordinate Φ_2 due to (3.45).

3.3. Projection operator techniques

It remains to address the reaction coordinate dynamics for a second-order mechanical system. For second-order systems we encounter the problem that the conditional expectation over the fast degrees of freedom involves position and velocity (momentum) variables. Now recall that in the Hamiltonian picture both positions and momenta were treated as independent variables. However fixing the reaction coordinate at a certain value amounts to imposing a holonomic constraint which inevitably determines the conjugate momenta. In turn, by varying the slow position and momentum variable independent of each other we obtain a fast subsystem that is dissipative and no longer Hamiltonian. The natural invariant probability measures for dissipative systems of this kind, so-called Axiom A flows, are Sinai-Ruelle-Bowen (SRB) measures [192, 193]. Although SRB measures are special cases of Gibbs measures (for example, they can be written in the form $\exp(-S)$, where S is a suitably defined pseudo-potential), they are difficult to handle both analytically and numerically; for example, if the flow of the fast subsystem is unbounded and expanding, there is no way of sampling the invariant measure by numerical long-term simulations. Moreover it is by no means clear whether the averaged system preserves the structure of the original mechanical equations. We shall illustrate the problem:

Example 3.12. Let us again adopt the Lagrangian viewpoint for a second, and consider the Lagrangian $L : TNS^2 \to \mathbf{R}$ in polar coordinates $(\varphi, \vartheta, r) \in S^2 \times \mathbf{R}_+$

$$L = \frac{1}{2}\left\langle G(\vartheta, r)(\dot\varphi, \dot\vartheta)^T, (\dot\varphi, \dot\vartheta)\right\rangle + \frac{1}{2}\dot r^2 - V(r)\,,$$

where V is a smooth, spherically-symmetric potential, and $G(\vartheta, r) = r^2 G(\vartheta, 1)$ is the metric of the 2-sphere with radius r. See Example 3.10 for details. Speeding up the angle variables by scaling the respective velocities according to

$$L_\epsilon(\varphi, \vartheta, r, \dot\varphi, \dot\vartheta, \dot r) = L(\varphi, \vartheta, r, \epsilon\dot\varphi, \epsilon\dot\vartheta, \dot r)\,,$$

we obtain Euler-Lagrange equations in first-order form with slow and fast variables

$$\begin{aligned}
\dot{r}_\epsilon(t) &= p_\epsilon(t)\\
\dot{p}_\epsilon(t) &= -\Gamma^r_{\alpha\beta}\zeta^\beta_\epsilon(t)\zeta^\gamma_\epsilon(t) - \partial_r V(r_\epsilon(t))\\
\dot{\omega}^\alpha_\epsilon(t) &= \frac{1}{\epsilon}\zeta^\alpha_\epsilon(t)\\
\dot{\zeta}^\alpha_\epsilon(t) &= -\frac{1}{\epsilon}\Gamma^\alpha_{\beta\gamma}\zeta^\beta_\epsilon(t)\zeta^\gamma_\epsilon(t) - \frac{2}{\epsilon}\Gamma^\alpha_{\beta r}\zeta^\beta_\epsilon(t)p_\epsilon(t)
\end{aligned}$$

subject to appropriate initial conditions. We have abbreviated $\omega = (\varphi, \vartheta)$. On the microscopic timescale $s = t/\epsilon$ we find the fast dynamics for frozen slow variables r, p:

$$\begin{aligned}
\dot{\omega}^\alpha_r(t) &= \zeta^\alpha_r(t)\\
\dot{\zeta}^\alpha_r(t) &= -\Gamma^\alpha_{\beta\gamma}\zeta^\beta_r(t)\zeta^\gamma_r(t) - 2\Gamma^\alpha_{\beta r}\zeta^\beta_r(t)\, p\,.
\end{aligned} \tag{3.54}$$

Note that since $\Gamma^\alpha_{\beta r} \neq 0$, the system is dissipative unless $p = 0$. In this particular case the fast equations of motion describe geodesics on the 2-sphere of radius r, i.e.,

$$\ddot{\omega}^\alpha_r(t) = -\Gamma^\alpha_{\beta\gamma}\dot{\omega}^\beta_r(t)\dot{\omega}^\gamma_r(t)\,.$$

The associated Gibbs measure is the ordinary Gibbs measure for the full system restricted to the 2-sphere with radius r. That is,

$$\mu_r(d\omega, d\dot{\omega}) = \frac{1}{Z_{S^2_r}(r)}\exp(-\beta T_r(\omega,\dot{\omega}))\det G(\omega, r)\, d\omega d\dot{\omega}$$

with the abbreviations

$$Z_{S^2_r}(r) = 4\pi r^2\left(\frac{\beta}{2\pi}\right)^{-3/2}$$

for the normalization constant, and

$$T_r(\omega,\dot{\omega}) = \frac{1}{2}\langle G(\omega, r)\dot{\omega}, \dot{\omega}\rangle$$

for the kinetic energy. We can write the slow equations again in second-order form,

$$\ddot{r}_\epsilon(t) = -\Gamma^r_{\alpha\beta}\dot{\omega}^\beta_\epsilon(t)\dot{\omega}^\gamma_\epsilon(t) - \partial_r V(r_\epsilon(t))\,,$$

and average the quadratic part in the slow equation with respect to μ_r. This yields

$$\ddot{r}(t) = -\frac{2}{\beta}\frac{1}{r(t)} - \partial_r V(r(t))\,, \tag{3.55}$$

where we have used that $\Gamma^r_{\varphi\varphi} = -r\sin^2\vartheta$ and $\Gamma^r_{\vartheta\vartheta} = -r$. We easily recognize that equation (3.55) is just the mechanical analogue of the stochastic limit equation (3.49). Now let us revisit equation (3.54) assuming that $p < 0$. The Christoffel symbols are $\Gamma^\varphi_{\varphi r} = \Gamma^\vartheta_{\vartheta r} = 1/r$ and zero else. Therefore the system is strictly hyperbolic, whenever $p < 0$ is sufficiently large in modulus. If the system were purely deterministic, the damping would dominate the dynamics, but its stationary points, and therefore its invariant measures, would clearly depend on the initial values. Consequently, the averaged equations would depend on which invariant measure we choose. For the stochastic system with randomized velocities anything can happen. Strictly speaking, the stochastic Hamiltonian system was defined only with regard to the symplectic, time-reversible and energy-preserving Hamiltonian flow (which we no longer have). But, having in mind the fluctuation-dissipation relation from the Langevin equation, we can imagine that the dynamics will depend on how friction and velocity perturbations counterbalance each other. And so will the invariant measure.

3.3.1. Optimal prediction and the Mori-Zwanzig formalism Originally, the idea of averaging stems from celestial mechanics [20]. Although the models considered there were purely mechanical, i.e., second-order, the problems are slightly different from ours. Indeed, the above considerations reveal that the application of the Averaging Principle is beyond the scope of this thesis. A central paradigm in molecular dynamics which comes from nonequilibrium thermodynamics is the method of Mori [49] and Zwanzig [50]. It is a formal procedure to rewrite the equations of motion in a specified set of essential variables (resolved variables). Unlike the Averaging Principle the Mori-Zwanzig proceeds without eliminating degrees of freedom, but rather incorporates them as some sort of heat bath, involving memory and noise. What is called *noise* here actually results from the unresolved variables and is the solution of an auxiliary equation which describes the dynamics orthogonal to the subspace of the resolved (essential) variables. The key element of this procedure is a projection operator, that projects the full set of equations onto the set of essential degrees of freedom. The projection is orthogonal in the Hilbert space L^2; thus it projects onto a space of functions that depend on the essential variables only. However this projection is not unique, and there is some freedom of choice. For instance, for first-order systems the conditional expectation (3.8) provides such a projection, but likewise the expectation (3.27) with respect to the constrained Gibbs measure. There is a subtle point concerning the relation between projection and the orthogonal dynamics as has been pointed out recently in [58]: the validity of the Mori-Zwanzig procedure relies on the well-posedness of the equations for the unresolved variables; this issue is similar to the closure problem for the fast dynamics in the averaging scheme, whereby the projection must account for positions and the momenta (velocities) in an appropriate manner to obtain well-posed equations of motion.

Before we proceed with the Mori-Zwanzig formalism, let us first consider the problem of optimally projecting the equations of motion onto the (function) subspace that is spanned by the reaction coordinate. This gives rise to a method called *optimal prediction*: Suppose we want to approximate the dynamics of an unresolved variable in some function space norm, say, in the Hilbert space L^2. Basically, this is to say that we want to study the best-approximation of an observable with regard to its expectation value. To this end let $\mu_{\mathrm{can}}(dz)$ denote the Gibbs measure on the phase space $E = T^*\mathbf{R}^n$. We introduce the weighted Hilbert space

$$L^2(\mu) = \left\{ v : E \to \mathbf{R} \,\Big|\, \int_E v(z)^2 \, \mu_{\mathrm{can}}(dz) < \infty \right\}$$

that is endowed with an appropriately weighted scalar product

$$\langle u, v \rangle_\mu = \int_E u(z)v(z) \, \mu_{\mathrm{can}}(dz) \, .$$

Recall the problem of optimal subspace projection, e.g., by the method of Principal Component Analysis (PCA) in Section 2.4. In some sense, optimal prediction can be considered the function space analogue of optimally projecting onto a dominant subspace. For example, consider the conditional expectation $\mathbf{E}_\xi(\cdot) = \mathbf{E}(\cdot|\Phi = \xi)$ as defined in (3.8) for a reaction coordinate Φ. It is easy to check that the conditional expectation defines an orthogonal projection

$$\Pi : L^2(\mu) \to L^2(\bar{\mu}) \subset L^2(\mu) \, , \quad (\Pi f)(\xi) = \mathbf{E}_\xi f \, ,$$

where $\bar{\mu}(d\xi) \propto Z(\xi)\, d\xi$ is the marginal probability of the reaction coordinate. In other words, the conditional expectation is an orthogonal projection onto the space of functions that depend only on the reaction coordinate. Given an arbitrary function $\phi \in L^2(\mu)$, this projection has the following useful property [64]

$$\|\phi - \Pi\phi\|_\mu^2 \le \|\phi - \psi\|_\mu^2 \quad \forall \psi \in L^2(\bar{\mu})\,.$$

where $\|\cdot\|_\mu$ denotes the norm in $L^2(\mu)$. Labelling by $\mathbf{E}(\cdot)$ the expectation with respect to μ_{can}, then the last inequality can be expressed in terms of expectation values,

$$\mathbf{E}|\phi - \mathbf{E}_\xi\phi|^2 \le \mathbf{E}|\phi - \psi|^2 \quad \forall \psi \in L^2(\bar{\mu})\,.$$

For the sake of illustration consider a reaction coordinate $\Phi(t) = \Phi(q(t))$. Since

$$\frac{d}{dt}\Phi(q(t)) = \mathbf{D}\Phi(q(t))^T p(t)$$

is linear in the momenta, the best-approximation of the reaction coordinate with respect to the conditional expectation $\mathbf{E}_\xi(\cdot)$ becomes trivial, viz.,

$$\dot{\xi}(t) = 0\,, \quad \xi(0) = \xi\,.$$

The approach is clearly not unique, and the optimal prediction equation very much depends on the choice of the conditional expectation. For example, one could project onto functions that depend on both Φ and $\dot{\Phi}$ or other relevant quantities. For our purpose it is more convenient to define a conditional expectation, that involves the reaction coordinate Φ and its conjugate momentum Θ.

Definition 3.13. *Let the function $\Phi : \mathbf{R}^n \to \mathbf{R}^k$ denote a smooth reaction coordinate, and let $\Theta : T^*\mathbf{R}^n \to \mathbf{R}^k$ be its conjugate momentum map.*[11] *We define the marginal probability density of Φ, Θ in the canonical ensemble by*

$$R(\xi, \eta) = \int_{\mathbf{R}^n \times \mathbf{R}^n} \delta(\Phi(q) - \xi)\delta(\Theta(q,p) - \eta)\mu_{\text{can}}(dq, dp)\,. \tag{3.56}$$

The conditional probability measure is denoted $\mu_{\xi,\eta} = \delta(\Phi - \xi)\delta(\Theta - \eta)\mu_{\text{can}}$. Then for an integrable function $f = f(q,p)$, we define the conditional expectation by

$$\mathbf{E}_{\xi,\eta} f = \frac{1}{R(\xi,\eta)} \int_{\mathbf{R}^n \times \mathbf{R}^n} f(q,p)\mu_{\xi,\eta}(dq, dp) \tag{3.57}$$

Quite remarkably, $\mathbf{E}_{\xi,\eta}(\cdot)$ comprises the expectation with respect to the constrained canonical ensemble as the special case $\mathbf{E}_{\xi,0}(\cdot)$. Hence the expectation $\mathbf{E}_{\xi,0}(\cdot) \neq \mathbf{E}_\xi(\cdot)$ is intrinsic to the constrained phase space $T^*\Sigma$, where $\Sigma = \Phi^{-1}(\xi)$. That is, it does not depend on the function Φ but only on the surface Σ. For the details the interested reader is referred to the relevant literature [195, 196].

Now optimal prediction proceeds as follows: Suppose we are given the molecular Hamiltonian H explicitly in terms of the reaction coordinate Φ, its conjugate momentum Θ, and a bunch of unresolved coordinates and momenta. This gives rise to equations for the reaction coordinate and its conjugate momentum

$$\begin{aligned}\dot{\Phi}^i &= \frac{\partial H}{\partial \Theta_i}\\ \dot{\Theta}_i &= -\frac{\partial H}{\partial \Phi^i}\,, \quad i = 1, \ldots, k\,.\end{aligned}$$

[11] We understand the term *momentum map* in a rather loose sense and not in accordance with the definition that is conventionally used in geometric mechanics (e.g., see [81, 194]). Nevertheless we regard the conjugate momentum Θ as a function of q and p, thus a momentum *map*.

The equations are not closed; they depend on both resolved and unresolved variables. Replacing the right hand side of the equations by its best-approximation by taking the conditional expectation yields the optimal prediction equations due to Hald [56]

$$\begin{aligned}\dot{\xi}^i &= \mathbf{E}_{\xi,\eta}\left(\frac{\partial H}{\partial \Theta_i}\right)\\ \dot{\eta}_i &= -\mathbf{E}_{\xi,\eta}\left(\frac{\partial H}{\partial \Phi^i}\right), \quad i = 1, \ldots, k\,.\end{aligned} \tag{3.58}$$

Proposition 3.14 (Hald 2000)**.** *The system (3.58) is Hamiltonian*

$$\begin{aligned}\dot{\xi}^i &= \frac{\partial E}{\partial \eta_i}\\ \dot{\eta}_i &= -\frac{\partial E}{\partial \xi^i}\,.\end{aligned}$$

with total energy

$$E(\xi,\eta) = -\beta^{-1}\ln R(\xi,\eta)\,.$$

Formally the optimal prediction Hamiltonian resembles the free energy expressions from the previous subsections. In fact it is some sort of free energy (in phase space though) which is related to the geometric free energy. For better distinguishability we shall speak of E as the *optimal prediction free energy.*

Optimal prediction equations In many relevant cases the representation of the reduced equations (3.58) in terms of the optimal prediction free energy E is not convenient, since E may not be accessible so easily (cf. Section 3.5). Even worse, in general the conjugate momentum Θ is not known explicitly. Nevertheless it is possible to recast (3.58) in a form that contains only quantities that are either already known or that can be sampled by means of Thermodynamic Integration. Assume that J_Φ has maximum rank. For convenience we introduce new coordinates $z^1, \ldots, z^n$

$$\psi:\ z^l = \begin{cases} \Phi^l(q) & \text{for } l = 1, \ldots, k\\ q^l & \text{for } l = k+1, \ldots, n\,. \end{cases} \tag{3.59}$$

This transformation is non-singular, for $\det \mathbf{D}\psi = \mathrm{vol} J_\Phi$ does not vanish by assuming that J_Φ has maximum rank. Hence we can write the molecular Lagrangian as

$$L(z,\dot{z}) = \frac{1}{2}a_{kl}(z)\dot{z}^k\dot{z}^l - V(z)\,, \tag{3.60}$$

where a_{kl} are the entries of the metric $(\mathbf{D}\psi^T\mathbf{D}\psi)^{-1}\circ\psi^{-1}$ that is induced by the change of coordinates. Due to (2.6) the conjugate momenta are given by $w_j = \partial L/\partial \dot{z}^j$. The Hamiltonian is then obtained as the Legendre transform $H(z,w) = w_j\dot{z}^j - L(z,\dot{z})$. We may split the new coordinates according to $z = (\xi, r)$ and $w = (\eta, s)$, such that

$$H(\xi,r,\eta,s) = \frac{1}{2}a^{ij}\eta_i\eta_j + \frac{1}{2}a^{i\alpha}\eta_i s_\alpha + \frac{1}{2}\delta^{\alpha\gamma}s_\alpha s_\gamma + V(\xi,r)\,, \tag{3.61}$$

where the a^{kl} are the matrix elements of

$$(\mathbf{D}\psi^T\mathbf{D}\psi)\circ\psi^{-1} = \begin{pmatrix} J_\Phi^T J_\Phi & M_\Phi^T\\ M_\Phi & \mathbf{1}\end{pmatrix}. \tag{3.62}$$

Here we employ Latin indices to enumerate the reaction coordinate (upper left matrix block), whereas the Greek indices label the unresolved modes (lower right unit block).

The off-diagonal submatrix $M_\Phi \in \mathbf{R}^{(n-k)\times k}$ is the minor of J_Φ that is made out of the first $n-k$ rows. Modulo normalization the marginal density thus becomes

$$R(\xi,\eta) = \int_{T^*\Sigma} \exp(-\beta H(\xi,r,\eta,s))\, d\mathcal{L}_\xi(r,s)\,,$$

where $d\mathcal{L}_\xi(r,s) = \sqrt{\det G_\xi(r)}drds$ is the Hausdorff measure of $\Sigma \times \mathbf{R}^{n-k} \subset \mathbf{R}^n \times \mathbf{R}^n$, and G_ξ is the induced metric on $\Sigma = \Phi^{-1}(\xi)$, that is obtained as the restriction of the Euclidean metric to Σ.[12] We need to compute the partial derivatives of

$$E(\xi,\eta) = -\beta^{-1}\ln R(\xi,\zeta)\,.$$

We have

$$\frac{\partial E}{\partial \eta^i} = \frac{1}{R}\int_{T^*\Sigma} \frac{\partial H}{\partial \eta^i} \exp(-\beta H)\, d\mathcal{L}_\xi\,.$$

Since the off-diagonal terms are linear in s, they vanish on average, and it remains

$$\frac{\partial E}{\partial \eta^i} = A^{ij}(\xi,\eta)\eta_j \quad \text{with} \quad A^{ij} = \mathbf{E}_{\xi,\eta}\left\langle \nabla\Phi^i, \nabla\Phi^j \right\rangle\,.$$

Other than the constrained expectation $\mathbf{E}_{\xi,0} = \mathbf{E}_\Sigma$ which is intrinsic to the surface Σ, the conditional expectation $\mathbf{E}_{\xi,\eta}$ does not give rise to a proper dynamical system that has $\mu_{\xi,\eta}$ as its invariant distribution. However we can sample $\mathbf{E}_{\xi,0}$ and the fact that Φ is only a function of the configurational variables suggests to do a Taylor expansion of the conditional expectation in powers of η. If the temperature is low as compared to the atomic masses (i.e., $\beta \gg 1$), the Maxwellian momentum distribution will be sharply peaked at $\eta = 0$. It is therefore convenient to replace $\mathbf{E}_{\xi,\eta}$ by $\mathbf{E}_{\xi,0}$ while neglecting higher order terms, in which case the last expression becomes

$$A^{ij} = \mathbf{E}_\Sigma\left\langle \nabla\Phi^i, \nabla\Phi^j \right\rangle + \mathcal{O}(\|\eta\|^2)\,.$$

Accounting for the dependence of the Hausdorff measure $d\mathcal{L}_\xi$ (surface element) on the foliation parameter ξ by appropriately extending G_ξ to the ambient space of $\Sigma = \Phi^{-1}(\xi)$, we can compute the derivative with respect to ξ^i. This yields

$$\frac{\partial E}{\partial \xi^i} = \frac{1}{R}\int_{T^*\Sigma} \left(\frac{\partial H}{\partial \xi^i} + \frac{1}{2}\operatorname{tr}\left(G_\xi^{-1}\frac{\partial G_\xi}{\partial \xi^i}\right)\right) \exp(-\beta H)\, d\mathcal{L}_\xi\,,$$

where

$$\frac{\partial}{\partial \xi^i}\sqrt{\det G_\xi} = \frac{1}{2}\operatorname{tr}\left(G_\xi^{-1}\frac{\partial G_\xi}{\partial \xi^i}\right)\sqrt{\det G_\xi}$$

is basically the i-th component of the mean curvature of Σ in $\mathbf{R}^n$; see Appendix C for the calculation.[13] Omitting again all terms that are linear in s, expanding all other terms around $\eta = 0$, what remains decays into two parts: The first part is the derivative of V with respect to ξ^i which, together with the mean curvature, can be summarized to yield the derivative of the familiar geometric free energy (3.31). The

[12] Note that we still have to integrate over a manifold, and that G_ξ is simply the metric of Σ from the preceding sections, where we have explicitly chosen $r = q^1, \ldots, q^{n-k}$ as local coordinates on Σ.

[13] This looks like a contradiction to Hald's Theorem, since we have an extra term in addition to the derivative of the Hamiltonian. However one should bear in mind that the coordinates $r = q^1, \ldots, q^{n-k}$ in the Hamiltonian (3.61) are *not* the unresolved variables, unless q is restricted to the fibre $\Phi^{-1}(\xi)$. But this means nothing but shifting the metric G_ξ from the Hausdorff measure to the (unresolved) kinetic energy part in the Hamiltonian. Yet this does not affect the integral.

other term is the derivative of the (average) kinetic energy of the reaction coordinate, such that we finally obtain

$$\frac{\partial E}{\partial \xi^i} = \frac{\partial}{\partial \xi^i}\left(K(\xi) + A^{jk}(\xi)\eta_j\eta_k\right) + \mathcal{O}(\|\eta\|^4)\,.$$

As before the A^{jk} are the lowest-order components of the effective inverse mass, and K is the geometric free energy (which we have labelled by K in order to avoid confusion with the metric tensor G_ξ)

$$K(\xi) = -\beta^{-1}\ln\int_\Sigma \exp(-\beta V)\,d\sigma_\xi$$

with $d\sigma_\xi$ denoting the surface element of $\Sigma \subset \mathbf{R}^n$. Conclusively, the optimal prediction free energy or effective Hamiltonian splits into kinetic and potential energy in the way that is easily interpretable, and probably more handy for practical applications

$$E(\xi,\eta) \approx \frac{1}{2}A^{ij}(\xi)\eta_i\eta_j + K(\xi)\,, \tag{3.63}$$

where both the inverse mass A^{-1} and the geometric free energy K (more precisely: the mean force $-\nabla K$) can be directly sampled by means of Thermodynamic Integration using constrained molecular dynamics; see the detailed discussion in Section 4.2.

The reader may wonder whether one could recover the standard free energy by integrating $\exp(-\beta E)$ over the momenta. In fact, integrating out the momenta yields

$$\int \exp(-\beta E)\,d\eta \propto \left(\sqrt{\det \mathbf{E}_\Sigma J_\Phi^T J_\Phi}\right)^{-1}\exp(-\beta K)\,.$$

But this is different from (3.26) which states the relation between geometric and standard free energy, and which — upon using (3.28) — can be recast in the form

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

Example 3.15. Let us reconsider the three-dimensional toy problem with radial potential. Choosing coordinates (φ,ϑ,ρ) on $N\Sigma \cong S^2 \times \mathbf{R}^+$, where $\rho = \|q\|^2$ denotes the resolved coordinate (reaction coordinate), the Hamiltonian takes the form

$$H = \frac{1}{2}\left\langle G(\vartheta,\rho)^{-1}u, u\right\rangle + 2\rho\zeta^2 + W(\rho)\,.$$

Again, $G(\vartheta,\rho) = \rho G_1(\vartheta)$ is the metric on the 2-sphere with radius $\sqrt{\rho}$, and $W(\rho) = V(\sqrt{\rho})$ is the radial potential. In this particular case the expression for the optimal prediction free energy (3.63) is exact and reads

$$\begin{aligned} E(\rho,\zeta) &= 2\rho\zeta^2 - \beta^{-1}\ln\int_{S^2}\exp(-\beta W(\rho))\sqrt{\det G(\vartheta,\rho)}\,d\varphi d\vartheta \\ &= 2\rho\zeta^2 + W(\rho) - \beta^{-1}\ln\rho \end{aligned}$$

plus additional constants which we have omitted. This puts forward the equations

$$\begin{aligned} \dot\rho(t) &= 4\rho(t)\zeta(t) \\ \dot\zeta(t) &= -\,\partial_\rho W(\rho(t)) - 2\zeta(t)^2 + \frac{1}{\beta\rho(t)}\,. \end{aligned}$$

3.3.2. The generalized Langevin equation Optimal prediction can be considered a lowest-order approximation of the equations of motion, similar to the averaging procedure. However it is possible to derive an exact evolution equation for the essential variables which is very intuitive, and from which we can derive non-Markovian corrections to optimal prediction. For this purpose we briefly review the projection operator approach of Mori and Zwanzig as can be found in, e.g., [51, 197, 52].

Let us consider the problem how phase space functions evolve in time. To this end consider a Hamiltonian $H : T^*\mathbf{R}^n \to \mathbf{R}$ with coordinates $z = (q, p)$. Let X_H be the Hamiltonian vector field generated by H, and denote by $z(t) = \Psi_t(z)$ with $z = z(0)$ the integral curves of X_H (i.e., $\Psi_t : E \to E$, $E = T^*\mathbf{R}^n$ is the Hamiltonian flow map). For our purposes it is convenient to cast Hamilton's equations in the form

$$\frac{d}{dt}\Psi_t^i(z) = X_H(\Psi_t^i(z))\,, \quad \Psi_0^i(z) = z^i \tag{3.64}$$

Given a function $f_0 : E \to \mathbf{R}$, we define $f(z,t) = (f_0 \circ \Psi_t)(z)$ as the pull-back of f_0 by the flow map. It follows by (3.64) and chain rule that f obeys the differential equation

$$\frac{d}{dt}(f_0 \circ \Psi_t)(z) = \nabla f(\Psi_t(z)) \cdot X_H(\Psi_t(z))\,. \tag{3.65}$$

Clearly the last equation is not closed in the sense that it does not give rise to the time evolution of f without solving Hamilton's equations for $z(t) = \Psi_t(z)$. Recall that

$$\tilde{X}_H(\Psi_t(z)) = \mathbf{D}\Psi_t(z) \cdot X_H(z)$$

is the transformation rule (chain rule) for a generic vector field. But since Ψ_t is symplectic and therefore preserves Hamilton's equations, the identity $\tilde{X}_H = X_H$ holds true for the push-forward of a Hamiltonian vector field by its flow. Now recall the definition of the Liouville equation (2.12). Using chain rule again and the definition (2.13) of the Liouville operator, we can rewrite the ordinary differential equation (3.65) as a partial differential equation in z and t. That is,

$$\partial_t f(z,t) = \mathcal{L} f(z,t)\,, \quad f(z,0) = f_0(z)\,, \tag{3.66}$$

where now the symbol ∇ in $\mathcal{L} = X_H(z) \cdot \nabla$ denotes the derivative with respect to z. (For the relation to the adjoint Liouville equation that governs the time evolution of probability densities see the remark below.) We may endeavour the semigroup notation from Section 2.1.1 and write the solution of the Liouville equation as

$$f = f_0 \circ \Psi_t = \exp(t\mathcal{L}) f_0\,.$$

In particular we can choose $f_0 = z_0^i$, such that $\exp(t\mathcal{L}) z_0^i = \Psi_t^i(z_0)$ describes the time evolution of the i-th coordinate. The aim is to split the transfer operator $T_t = \exp(t\mathcal{L})$ into a part S_t that acts only on the subspace of the essential (resolved) variables, and a part $S_t^\perp$ that operates on the orthogonal subspace.

Following [56] we denote by $\Pi : L^2(\mu) \to L^2(\mu)$ and $Q = \mathbf{1} - \Pi$ a pair of orthogonal projections (e.g., the conditional expectation). Modulo some technical assumptions we require that $Q\mathcal{L}Q$ is the infinitesimal generator of a strongly continuous semigroup. In other words, we demand that $Q\mathcal{L}$ generates a flow on the Q subspace. For the details we refer to [58, 198] and define $S_t^\perp$ as the propagator of

$$\begin{aligned} \partial_t w(z,t) &= Q\mathcal{L} w(z,t) \\ w(z,0) &= w_0(z) \in \ker \Pi \end{aligned} \tag{3.67}$$

which can be equivalently written as an inhomogeneous equation for $w = \exp(tQ\mathcal{L}) w_0$:

$$\begin{aligned} \partial_t w(z,t) &= \mathcal{L} w(z,t) - \Pi \mathcal{L} w(z,t) \\ w(z,0) &= w_0(z) \in \ker \Pi\,. \end{aligned}$$

The solution of the last equation is easily obtained by Variation of Constants [199], which results in a Volterra integral equation for the orthogonal dynamics $w(z,t)$,

$$w(z,t) = T_t w_0(z) - \int_0^t T_{t-s}\Pi\mathcal{L}w(z,s)\,ds\,. \tag{3.68}$$

Using that $T_t\mathcal{L} = \mathcal{L}T_t$ we may write the Liouville equation (3.66) in the form

$$\partial_t f(z,t) = \partial_t T_t f_0(z) = T_t\Pi\mathcal{L}f_0(z) + T_t Q\mathcal{L}f_0(z)\,.$$

In the second term the transfer operator T_t acts on a function that lies in the nullspace of Π. Hence we can insert the solution (3.68) of the orthogonal dynamics with initial condition $w_0 = Q\mathcal{L}f_0$. Omitting the argument z from now on this gives

$$\partial_t f(t) = T_t\Pi\mathcal{L}f_0 + S_t^{\perp}Q\mathcal{L}f_0 + \int_0^t T_{t-s}\Pi\mathcal{L}S_s^{\perp}Q\mathcal{L}f_0\,ds\,. \tag{3.69}$$

The last equation is often referred to as *generalized Langevin equation.* By no means this equation is simpler than the original problem. In point of fact, the complexity of the full-dimensional evolution problem has been transferred to the solution of the Volterra integral equation of the second kind for the orthogonal dynamics.

The various terms in the generalized Langevin equation have suggestive physical interpretations: The first term on the right hand side is Markovian. Indeed,

$$T_t\Pi\mathcal{L}f_0 = \Pi\mathcal{L}f_0 \circ \Psi_t = \Pi\mathcal{L}f(t)\,.$$

The second term in (3.69), which is usually interpreted as noise evolves the unresolved variables according to the orthogonal dynamics' equation. It remains in the orthogonal subspace for all times, for $S_t^{\perp}Q$ commutes with $Q = Q^2$. Finally, the third term depends on the value of the observable f at times $s \in [0,t]$, i.e., it depends on the past evolution up to time t. Accordingly it embodies memory effects that stem from dynamical interaction between the two subspaces.

Introducing the abbreviations $w(t) = S_t^{\perp}Q\mathcal{L}f_0$ and $K(t-s) = T_{t-s}\Pi\mathcal{L}$ we can cast the generalized Langevin equation in the slightly more compact form

$$\partial_t f(t) = \Pi\mathcal{L}f(t) + \int_0^t K(t-s)w(s)\,ds + w(t)\,, \tag{3.70}$$

where $w(t)$ is the solution of the Volterra integral equation (3.68) for the orthogonal dynamics with $w_0 = Q\mathcal{L}f_0$. So far, the last equation is completely equivalent to the Liouville equation (3.66), but in practice it can only be solved approximately.

Remark 3.16. *Note the different signs in the Liouville equation (2.12) for densities and the Liouville equation (3.66), and remember that the Liouvillian is skew-adjoint in the Hilbert space $L^2(dz)$ (and so is in $L^2(\mu)$ for any smooth probability measure μ preserved by the Hamiltonian flow). Accordingly the Liouville equation (3.66) for phase space functions can be regarded as the formal adjoint of (2.12).*

This duality is the classical analogue of the famous dichotomy of Schrödinger and Heisenberg picture in quantum mechanics; see, e.g., [200]. Recall that the time evolution of a probability density ρ is the push-forward of an initial density ρ_0 by the Hamiltonian flow, i.e., $\rho = \rho_0 \circ \Psi_{-t}$, whereas the time-dependence of an observable f is induced by the pull-back, $f = f_0 \circ \Psi_t$ of an initial value f_0. We can make the Schrödinger-Heisenberg duality more specific: Suppose we are interested in the time-dependent expectation value of an observable f. As we have seen in (3.65) we can calculate $f(z,t)$ by following an initial preparation $f(z,0) = f_0(z)$ along a trajectory

$z(t) = \Psi_t(z)$. *If the initial values z are distributed according to some probability distribution $\rho_0(z)$, then*

$$\mathbf{E}_{\mathrm{H}} f(z,t) = \int_E \rho_0(z) T_t f_0(z)\, dz\,,$$

where we have employed the semigroup notation $T_t = \exp(t\mathcal{L})$. This representation of time-dependent expectation values is called Heisenberg picture *(or* Lagrangian picture *in fluid dynamics, respectively). Changing our point of view slightly we may consider the observable at a fixed point in phase space, while weighting the observed quantity with the current value of the initial ensemble,*

$$\mathbf{E}_{\mathrm{S}} f(z,t) = \int_E f_0(z) T_{-t}\rho_0(z)\, dz\,,$$

which is known by the name of Schrödinger representation. *According to [201] the adjoint semigroup is generated by the adjoint Liouvillian $\mathcal{L}^* = -\mathcal{L}$, i.e., $T_t^* = T_{-t}$. Noting that $\mathbf{E}_{\mathrm{S}} f = \langle f_0, T_{-t}\rho_0\rangle$ we see immediately that $\langle f_0, T_{-t}\rho_0\rangle = \langle f_0, T_t^*\rho_0\rangle = \langle T_t f_0, \rho_0\rangle$. Hence both representations are equivalent in the sense that $\mathbf{E}_{\mathrm{H}} = \mathbf{E}_{\mathrm{S}}$*

Approximations and closures Although it seems appealing to make further assertions, e.g., concerning a generalized fluctuation-dissipation relation, (3.70) is the best we can achieve, unless we reinforce further assumptions. In particular we choose Π to be the conditional expectation. We briefly review the most common approximation schemes that are available in the relevant literature. To this end, we restrict our attention to the case of a separable Hamiltonian that is of the form

$$H(x,y,u,v) = \frac{1}{2}\langle u,u\rangle + \frac{1}{2}\langle v,v\rangle + V(x,y)\,,$$

where $(x,u) \in \mathbf{R}^k \times \mathbf{R}^k$ denotes the reaction coordinate with its conjugate momentum, whereas $(y,v) \in \mathbf{R}^{n-k} \times \mathbf{R}^{n-k}$ labels a set of unresolved conjugate variables.

The Mori-Zwanzig approach is very elegant on the formal level of deriving the generalized Langevin equation, but it becomes a bit messy when it comes to specific the equations of motion. Therefore, and for the sake of clarity, we shall be very explicit regarding notation: we let $z = (x,y,u,v)$ abbreviate the state vector, and we write $\varphi(z,t) = \Psi_t(z)$ for the solution curves that are generated by the Hamiltonian H. Moreover let the projection Π be the conditional expectation $\mathbf{E}_{\xi,\eta} = \mathbf{E}(\cdot | z_1 = \xi, z_3 = \eta)$ that is understood with respect to the initial conditions, where the corresponding probability density is given by (3.56). Note that this point of view is different from the optimal prediction viewpoint, where simply the right hand side of Hamilton's equations was *replaced* by its optimal L^2-projection given the current value of the reaction coordinate. (Consult the recent textbook [59] for some clarifying remarks.) It can readily checked that the generalized Langevin equation takes the form

$$\begin{aligned} \partial_t \varphi_1(z,t) &= \mathbf{E}_{\xi,\eta}\varphi_2(z,t) \\ \partial_t \varphi_2(z,t) &= -\nabla G(\varphi_1(z,t)) + \int_0^s K(t-s) w(z,s)\, ds + w(z,t)\,, \end{aligned} \tag{3.71}$$

where the integral kernel $K(t-s) = T_{t-s}\mathbf{E}_{\xi,\eta}\mathcal{L}$ is defined as above, and $\nabla G = \mathbf{E}_{\xi,\eta}\mathbf{D}_1 V(\cdot,\cdot)$. The fluctuation term stems from the orthogonal dynamics equation,

$$w(z,t) = -S_t^{\perp}\nabla\left(V(\varphi_1(z,t),\cdot) - G(\varphi_1(z,t))\right)\,.$$

So far the generalized Langevin equation involves no approximations, notwithstanding the separability assumption on the Hamiltonian. But obviously the equations are not closed, for they still depend on the initial values of the unresolved variables. A commonly used simplification is obtained by taking the conditional expectation on either sides of the equation which, by definition of the orthogonal dynamics, annihilates the fluctuation term. Defining $\xi(t) = (\mathbf{E}_{\xi,\eta}\varphi_1)(\xi,\eta,t)$ and $\eta(t) = (\mathbf{E}_{\xi,\eta}\varphi_2)(\xi,\eta,t)$, the generalized Langevin equation (3.71) becomes upon projecting from the left

$$\begin{aligned}\dot{\xi}(t) &= \eta(t)\\ \dot{\eta}(t) &= -\mathbf{E}_{\xi,\eta}\nabla G(\varphi_1(z,t)) + \int_0^s \mathbf{E}_{\xi,\eta}K(t-s)w(z,s)\,ds\,.\end{aligned}$$

Still the equations are not closed, since the conditional expectation does not commute with the evaluation of the nonlinear force term, i.e., $\mathbf{E}_{\xi,\eta}\nabla G(\varphi_1(z,t)) \neq \nabla G(\xi(t))$. In order to obtain an equation for (ξ,η) we follow [64] and interchange the evaluation of the effective force and the conditional expectation:

$$\mathbf{E}_{\xi,\eta}\nabla G(\varphi_1(z,t)) \approx \nabla G(\mathbf{E}_{\xi,\eta}\varphi_1(z,t)) = \nabla G(\xi(t))\,. \tag{3.72}$$

We refer to this step as *mean-field approximation*. The reader should not be bothered by this step, since the sole alternative would be to neglect the spreading of $\varphi_1(z,t)$ due to different initial conditions in z. However it has turned out [202] that one is better off preserving the distributed initial conditions, while mistreating them slightly, than completely ignoring them. This yields a non-Markovian optimal prediction equation

$$\begin{aligned}\dot{\xi}(t) &= \eta(t)\\ \dot{\eta}(t) &= -\nabla G(\xi(t)) + \int_0^s \mathbf{E}_{\xi,\eta}K(t-s)w(z,s)\,ds\,.\end{aligned} \tag{3.73}$$

Note that the memory integral contains information about the unresolved modes, and so we still have to solve the orthogonal dynamics equation. Suppose the Volterra equation (3.68) is well-posed. Following [203] the formal solution of (3.68) is[14]

$$w(z,t) = \zeta(z,t) - \int_0^t R(t,s)\zeta(z,s)\,ds\,,$$

where $R(t,s)$ is the resolvent kernel

$$R(t,s) = \sum_{i=1}^{\infty}(-1)^{i-1}\kappa_i(t,s)\,, \quad \kappa_i(t,s) = \int_0^t K(t-\varsigma)\kappa_{i-1}(\varsigma,s)\,d\varsigma$$

with $\kappa_1(t,s) = K(t-s)$. The smoothness of $w(z,\cdot)$ depends on the smoothness of the memory kernel. Clearly solving the equations numerically is not necessarily easier than directly solving the Liouville equation (3.67) for the orthogonal dynamics. Nevertheless the Neumann series above is related to an iterative scheme that is useful once an approximate solution is known. For a sufficiently small time step h we consider

$$w(z,h) = \zeta(z,h) - \int_0^h K(h-s)w(z,s)\,ds\,, \tag{3.74}$$

where $\zeta(z,h) = T_h w_0(z)$, and $\mathbf{E}_{\xi,\eta}w(z,s) = 0$, i.e., $w(\cdot,s)$ lies in the nullspace of the projection $\Pi = \mathbf{E}_{\xi,\eta}$. We shall apply the method of successive approximations to the integral equation (3.74). This method consists in constructing a sequence

$$u_{k+1}(z,h) = \zeta(z,h) - \int_0^h K(h-s)u_k(z,s)\,ds$$

[14] Of course, well-posedness depends upon the choice of the underlying function space. In particular the existence of weak L^2-solutions has been proved recently in the article [58]

with $\mathbf{E}_{\xi,\eta}u_k(z,s) = 0$ and initialization $u_0(z,h) = \zeta(z,h)$. It can be regarded as a Picard iteration for the differential equation (3.65). Pushing the iteration to the next order u_1, exploiting the semigroup property $T_h = T_{h-s} \circ T_s$, we find

$$u_1(z,h) = (1 - h\mathbf{E}_{\xi,\eta}\mathcal{L})\,\zeta(z,h)$$

and so forth. It is known that for a sufficiently smooth integral kernel $K(h-s)$ that satisfies a local Lipschitz condition the sequence $\{u_k\}$ eventually converges to the orthogonal dynamics solution in some interval $[0,\tau]$, i.e., $u_k(z,h) \to w(z,h)$ for $h \in [0,\tau]$ as $k \to \infty$. However existence and uniqueness is guaranteed only locally; basically the maximally achievable τ up to which the solution can be continued depends on boundedness and decay of the integral kernel. For details the reader may consult the references [204, 205]. It is interesting to note that extending the lowest order approximation $w(z,h) \approx u_0(z,h)$ to $h = t$ and substituting it into (3.73) yields what circulates in the literature as *t-damping equation*

$$\begin{aligned}\dot{\xi}(t) &= \eta(t)\\ \dot{\eta}(t) &= -\nabla G(\xi(t)) - t\,\gamma(\xi(t))\cdot\eta(t)\,,\end{aligned}\tag{3.75}$$

where the positive semi-definite friction matrix γ is given by

$$\gamma(\xi) = \mathbf{E}_{\xi,\eta}\left(\nabla\left(V(z_1,\cdot) - G(z_1)\right)\nabla\left(V(z_1,\cdot) - G(z_1)\right)^T\right)\,.$$

In the last step we have once more interchanged the conditional expectation with the function evaluation (mean-field approximation). Roughly speaking the t-damping equation amounts to the approximation $S_t^{\perp} \approx T_t$; see [64] and the references therein. However we note that neither u_0 nor u_1 ought to be considered a systematic asymptotic expansion for the orthogonal dynamics that is valid beyond the characteristic decay time h of the orthogonal dynamics. In particular the energy in the t-damping equation will quickly decay to zero. This seems rather unphysical, and we therefore suggest to approximate the memory kernel not until the level of numerical discretization.

A related approximation which is popular in the nonequilibrium statistical mechanics community consists in introducing a characteristic time τ that indicates the support of the memory integral backwards in time; see, e.g. [206, 207]. The basic idea is to replace (3.74) by a modified Volterra equation

$$w(t) = \zeta(z,t) - \int_0^t K(t-s)\hat{w}(z,s)\,ds\,,\quad w,\hat{w} \in \ker\mathbf{E}_{\xi,\eta}\,,$$

where $\hat{w}(s) = w_0(z)k(s/\tau)$, and $k(s/\tau)$ is an arbitrary function satisfying

$$k(0) = 1 \quad \text{and} \quad \int_0^\infty k(s/\tau) = \tau\,.$$

For $k(s/\tau) = \exp(-s/\tau)$ we can easily expand the integral in powers of τ and obtain a t-damping-like equation which reads to lowest order in τ (see [208])

$$\begin{aligned}\dot{\xi}(t) &= \eta(t)\\ \dot{\eta}(t) &= -\nabla G(\xi(t)) - \tau\,\gamma(\xi(t))\cdot\eta(t)\end{aligned}$$

with the previously defined friction matrix. Unlike (3.75) the friction term in the last equation does not increase as time evolves, provided γ stays bounded. Nevertheless the system is dissipative in the sense that the total energy of the system is decreasing along the solution curves and eventually goes to zero. A further *ad-hoc* modification that has been suggested recently in the PhD thesis [55] consists in adding an extra stochastic term to the equations with (yet unknown) statistics. This leads to

$$\begin{aligned}\dot{\xi}(t) &= \eta(t)\\ \dot{\eta}(t) &= -\nabla G(\xi(t)) - \tau\,\gamma(\xi(t))\cdot\eta(t) + F(\xi(t),t)\,,\end{aligned}$$

which is a linear Langevin equation but should not be confused with the covariant Langevin equation (2.25) with configuration-dependent friction and noise coefficients. If $F(\xi,t)$ is an uncorrelated, zero-mean stochastic process that satisfies the generalized fluctuation-dissipation relation,

$$\mathbf{E}F(\xi,s)F(\xi,t)^T = 2\tau\beta^{-1}\gamma(\xi)\delta(s-t)\,,$$

then the linear Langevin equation has the invariant probability density

$$\rho(\xi,\eta) \propto \exp(-\beta E(\xi,\eta)) \quad \text{with} \quad E(\xi,\eta) = \frac{1}{2}\langle\eta,\eta\rangle + G(\xi)\,.$$

Remark 3.17. *We mention that there is an ongoing discussion about whether the Volterra equation or approximations thereof are well-posed and numerical solutions exist [58, 209, 210]; see also [210]. Regarding stability of the solutions with respect to perturbations of the (unresolved) initial conditions we refer to the excellent survey article [203] and the references given there.*

Many authors study a special case of a Volterra integro-differential equation that relates the velocity autocorrelation function of the reaction coordinate to the memory kernel, in case the system consists of harmonic oscillators only [211]; however these authors rarely take into account the specific assumptions under which the equations have been derived (e.g., linear projections rather than conditional expectations); see, e.g., [53, 212]. Moreover this type of Volterra equation suffers from various degrees of ill-posedness, and the numerical integration is notoriously unstable. Therefore many authors resort to regularization techniques, e.g., (sequential) Tikhonov regularization, or choosing local ansatz functions for the memory kernel [54]..

To the best of the author's knowledge there are no statements regarding the numerical efficiency of the Mori-Zwanzig method as compared to simulations of the full model, and detailed numerical studies of the generalized Langevin equation are desirable. Moreover, systematic studies of Markov approximations are rare, e.g., [57]. But addressing the computational aspects in an adequate way is far beyond the scope of this thesis, and we leave it at the few remarks given above. For related approaches using a moment expansion of the Liouville equation we refer to [213].

3.4. Modelling fast degrees of freedom: adiabatic perturbation theory

In this subsection we put forward another approach to get rid of certain irrelevant (unresolved) degrees of freedom. The name *adiabatic perturbation theory* is borrowed from the theory of adiabatic invariants of integrable systems which is a common topic in celestial mechanics. The theory of adiabatic invariants relies on the formalism of canonical transformations: an oscillatory system is recast into an equivalent one with action-angle coordinates (I,φ), such that I is invariant under the Hamiltonian flow, and φ is an angular coordinate on a torus [20]. If the action variables I are not preserved but slowly varying (*slow* is meant in comparison with the angle variables), we arrive at the classical averaging problem; see [214] and the references therein.

The method which is proposed in this section can be considered a thermodynamical variant of the action-angle problem, which is better suited to problem involving a heat bath. It leads to a simplification of the former averaging problem, and it relies on the basic insight that certain degrees of freedom are fast and have comparably small amplitude, such that we can treat them as harmonic oscillations. Not only does this considerably simplify the analysis of the models and their numerical simulation, but most of the unresolved variables are harmonic anyway, e.g., bond and bond angle vibrations, or solvent motion to mention just a few.

By no means the averaging results that we present are new. However the current approach places emphasis on two different aspects: First of all it gives rise to a alternative view on fast motions from which semi-analytic, reduced models can be developed that have few free parameters. Secondly, it explains once more the relation between stiff harmonic modes, e.g., bonds, and constrained variables. In other words, it points out the (in principle well-known but often ignored) difference between a constrained system, where certain modes are held fixed at equilibrium values, and very stiff systems, where the system is allowed to oscillate around these values. The last remark concerns the difference between conditional and constrained expectations (Fixman Theorem or Blue Moon formula), and it provides a physical understanding of techniques like the widely-used umbrella sampling; cf. [76].

A modelling potential Suppose that any of the subspace reduction methods from Section 2.4 has given us an approximating subspace M that is spanned by a few slow variables, say, $x^1, \ldots, x^{n-s}$, and assume that the dynamics stays close to this subspace over a finite time interval. Given a local orthonormal frame $\{n_1(\sigma(x)), \ldots, n_s(\sigma(x))\}$ over M with normal coordinates $y \in \mathbf{R}^s$ we define a confining potential by

$$U_\epsilon(\sigma, n) = \frac{1}{2\epsilon^2} \langle B(\sigma)n, n\rangle \ ,$$

where $n \in N_\sigma M$ with $n = y^j n_j(\sigma(x))$, and $\epsilon \ll 1$ is an empirical scaling parameter, that might be chosen, for instance, as the autocorrelation time ratio of the slowest and the first truncated dominant degree of freedom. Suppose that for each $\sigma \in M$ the matrix $B(\sigma) \in \mathbf{R}^{n\times n}$ is positive-semidefinite of rank s. In bundle coordinates (x, y) the confinement potential then takes the form

$$U_\epsilon(x, y) = \frac{1}{2\epsilon^2} \langle K(x)y), y\rangle \ . \tag{3.76}$$

Note that if we assume that the matrix $B(\sigma)$ above has maximum rank s, then the symmetric, and positive-definite matrix $K(x) \in \mathbf{R}^{s\times s}$ is simply $B(\sigma)$ written in the basis of the normal frame. In fact, it is recommendable to construct the normal frame from the eigenvectors of $B(\sigma)$ corresponding to non-zero eigenvalues.

The confinement potential U_ϵ is designed in such a way that it achieves its minimum exactly on the approximant [27]. This is always possible if the matrix K has s strictly positive eigenvalues (recall that the codimension of $M \subset \mathbf{R}^n$ is s). If ϵ tends to zero, it generates a force in the neighbourhood of M that pushes the moving particle to the manifold. Clearly in the limit the particle must remain on M, and we obtain a reduced system that lives only on the approximant.

By construction, U captures the influence of the normal modes which have small variance.[15] This offers a reliable description of the motion close to the approximant M, provided the matrix family $B(\sigma)$ is appropriately chosen. For example, one may think of $B(\sigma)$ as the covariance or correlation matrix of the system conditional on x. This

[15] The term *normal mode* is not to be confused with what is typically called *Normal Mode Analysis.*

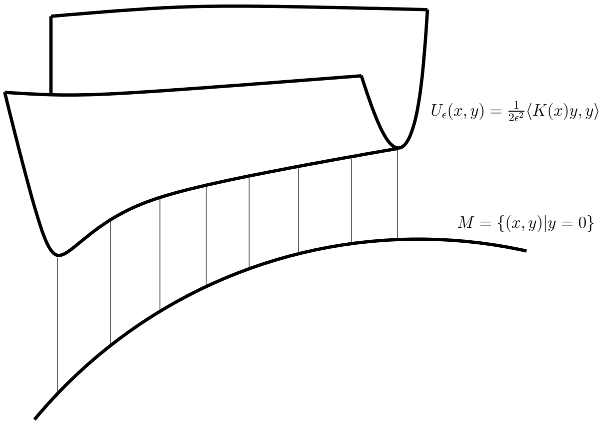

$U_\epsilon(x,y) = \frac{1}{2\epsilon^2}\langle K(x)y, y\rangle$
$M = \{(x,y)|y=0\}$

Figure 7. Schematic plot of the confining potential.

would guarantee that the normal modes reproduce the statistics of the unresolved motion in the vicinity of the approximant. The idea now is to replace the original potential V by a modelling potential

$$V_\epsilon(x,y) = V_M(x) + U_\epsilon(x,y)$$

in a tubular neighbourhood of M. For example, one might think of $V_M(x) = V(\sigma(x))$ as the restriction of the molecular to the approximant, or $V_M(x) = F(x)$ could be some kind of free energy in the essential variables x. This can by rephrased saying that the fast variables are modelled by appropriate Ornstein-Uhlenbeck processes (Brownian motion) or harmonic oscillators, respectively (second-order equations).

Strong confinement limit: diffusive motion To formulate our idea precisely we start studying the limit $\epsilon \to 0$ for the Smoluchowski equation. Let $V_\epsilon : \mathbf{R}^n \to \mathbf{R}$ be the modelling potential. Then for $\beta > 0$ the Smoluchowski equation on $\mathbf{R}^n$ reads

$$\dot{q}_\epsilon(t) = -\mathrm{grad}\, V_\epsilon(q_\epsilon(t)) + \sqrt{2\beta^{-1}}\dot{W}(t)\,.$$

We assume that the approximant M that is spanned by the essential variables is a smoothly embedded submanifold of codimension s in $\mathbf{R}^n$, and we denote this embedding by $\sigma : \mathbf{R}^{n-s} \to M \subset \mathbf{R}^n$. As before we introduce local coordinates x^α, $\alpha = 1, \ldots, n-s$ on M, and normal coordinates y^i, $i = 1, \ldots, s$ that measure the distance to M with respect to the normal frame $\{n_1, \ldots, n_s\}$. In terms of the local coordinates the Smoluchowski equation becomes according to Lemma 2.11

$$\begin{aligned} \dot{x}^\alpha_\epsilon &= -g^{\alpha l}(x_\epsilon, y_\epsilon)\, \partial_l V_\epsilon(x_\epsilon, y_\epsilon) + b^\alpha(x_\epsilon, y_\epsilon) + a^{\alpha l}(x_\epsilon, y_\epsilon)\, \dot{W}_l \\ \dot{y}^i_\epsilon &= -g^{il}(x_\epsilon, y_\epsilon)\, \partial_l V_\epsilon(x_\epsilon, y_\epsilon) + b^i(x_\epsilon, y_\epsilon) + a^{il}(x_\epsilon, y_\epsilon)\, \dot{W}_l\,, \end{aligned} \tag{3.77}$$

where $b^h = -\beta^{-1} g^{kl}\Gamma^h_{kl}$ denotes the additional Itô drift term with the symmetric Christoffel symbols Γ^h_{kl}, and a^{kl} are the entries of the uniquely defined positive-definite matrix square root of g^{-1} multiplied by the noise amplitude $\sqrt{2\beta^{-1}}$ (see Appendix B for the definition of the metric tensor g). The effect of confining a Brownian particle to the submanifold M is expressed in the next statement following an idea due to [74].

Proposition 3.18. *For all $\epsilon > 0$ let the process $(x_\epsilon(t), y_\epsilon(t)) \in \mathbf{R}^n$ defined by (3.77) with ϵ-dependent initial values $(x_\epsilon(0), y_\epsilon(0)) = (x, \epsilon y)$ be a continuous Markov process. Furthermore let the processes admit a family of unique invariant measures $\mu^\epsilon(dx, dy)$. Then as $\epsilon \to 0$ the process $x_\epsilon(t) \in \mathbf{R}^{n-s}$ converges in probability to a stochastic process $x(t) \in \mathbf{R}^{n-s}$ satisfying the following differential equation*

$$\dot{x}(t) = \bar{b}(x(t)) - \operatorname{grad} \bar{V}(x(t)) + \bar{a}(x(t))\, \dot{W}(t)\,, \tag{3.78}$$

where the effective potential is given by

$$\bar{V}(x) = V_M(x) + \frac{1}{2\beta} \ln \det K(x)\,.$$

The rightmost term is a Fixman potential. The remaining coefficients are

$$\bar{b}^\alpha(x) = \beta^{-1} G^{\gamma\delta}(x) \Gamma^\alpha_{\gamma\delta}(x)\,, \quad \bar{a}^{\alpha\gamma}(x) = \sqrt{2\beta^{-1}} \left(\sqrt{G^{-1}(x)}\right)_{\alpha\gamma}$$

with the Christoffel symbols $\Gamma^\alpha_{\gamma\delta}(x) = \Gamma^\alpha_{\gamma\delta}(x, 0)$ of the metric $G(x)$ on M.

Proof. For the relation between the various free energies and the Fixman potential see the paragraph above Remark 3.21 below. First of all observe that $V_\epsilon(x, y) = V_1(x, \epsilon^{-1} y)$. Hence we suggest to introduce scaled variables $y = \epsilon z$, in order to circumvent a blow up of the normal energy in the confinement limit. Moreover we assume that all realizations will stay in the tubular neighbourhood of M. In the scaled coordinates (x, z) the equations of motion read

$$\begin{aligned} \dot{x}^\alpha_\epsilon &= -\frac{1}{\epsilon} g^{\alpha j}_\epsilon\, \partial_j V_1 - g^{\alpha\beta}_\epsilon\, \partial_\beta V_1 + b^\alpha_\epsilon + a^{\alpha l}_\epsilon\, \dot{W}_l \\ \dot{z}^i_\epsilon &= -\frac{1}{\epsilon^2} g^{ij}_\epsilon\, \partial_j V_1 - \frac{1}{\epsilon} g^{i\beta}_\epsilon\, \partial_\beta V_1 + \frac{1}{\epsilon} b^i_\epsilon + \frac{1}{\epsilon} a^{il}_\epsilon\, \dot{W}_l\,, \end{aligned} \tag{3.79}$$

where we have introduced the scaled quantities $g_\epsilon = g(x, \epsilon z)$, $b_\epsilon = b(x, \epsilon z)$ and $a_\epsilon = a_\epsilon(x, \epsilon z)$. Now the normal energy remains finite as ϵ goes to zero, and the equations have the standard form to which the Averaging Principle applies. It can be readily checked that the ϵ-family of invariant measures is given by

$$\mu^\epsilon(dx, dz) = \frac{1}{Z_\epsilon} \exp\left(-\beta V_1(x, z)\right) \sqrt{\det g(x, \epsilon z)}\, dxdz\,.$$

In order to compute the conditional invariant measure of the fast process we make a time scaling $t \mapsto \epsilon^2 t$, taking into account that the noise scales like $\dot{W}(t) \mapsto \epsilon^{-1} \dot{W}(\epsilon^2 t)$:

$$\begin{aligned} \dot{x}^\alpha &= -\epsilon^2 g^{\alpha j}_\epsilon\, \partial_j V_1 - \epsilon^2 g^{\alpha\beta}_\epsilon\, \partial_\beta V_1 + \epsilon^2 b^\alpha_\epsilon + \epsilon a^{\alpha l}_\epsilon\, \dot{W}_l \\ \dot{z}^i_\epsilon &= -g^{ij}_\epsilon\, \partial_j V_1 - \epsilon g^{i\beta}_\epsilon\, \partial_\beta V_1 + \epsilon b^i_\epsilon + a^{il}_\epsilon\, \dot{W}_l\,. \end{aligned}$$

Letting ϵ go to zero yields the fast process conditioned on the frozen slow variables x

$$\dot{z}^i_x = -\,\delta^{ij}\, \partial_j V_1(x, z) + \sqrt{2\beta^{-1}} \delta^{il}\, \dot{W}_l\,, \tag{3.80}$$

where we have taken advantage of the identity $g^{il}(x, 0) = \delta^{il}$. The conditional invariant measure then is independent of ϵ and has the remarkably simple form

$$\mu_x(dz) = \frac{1}{Q(x)} \exp\left(-\beta U_1(x, z)\right)\, dz\,,$$

which is owed to the fact that the fibres $N_\sigma M$ locally look like $\mathbf{R}^s$, since we have dilated the normal direction in the just described way; no functional determinant is involved.[16] Endeavouring the Averaging Principle we have to compute the integral

$$\bar{f}^\alpha(x) = \lim_{\epsilon \to 0} \lim_{T \to \infty} \frac{1}{T} \int_0^T f_\epsilon^\alpha(x, z_x(t))\, dt\,,$$

where f_ϵ^α denotes the right hand side of the x-equation in (3.79). Note that the conditional fast process is a non-degenerate Ornstein-Uhlenbeck process, and therefore $z_x(t)$ is exponentially mixing, i.e., ergodic. Hence we can replace the time average by

$$\bar{f}^\alpha(x) = \lim_{\epsilon \to 0} \int f_\epsilon^\alpha(x, z)\, \mu_x(dz)\,.$$

Since $\mu_x(dz)$ does not depend on ϵ, and the integrand is uniformly continuous in z we may interchange the limit $\epsilon \to 0$ with the integration. We can split $f_\epsilon^\alpha = h_\epsilon^\alpha + k_\epsilon^\alpha$ into one part that becomes independent of z as ϵ goes to zero

$$\lim_{\epsilon \to 0} h_\epsilon^\alpha(x, z) = b^\alpha(x, 0) + a^{\alpha l}(x, 0)\dot{W}_l$$

and into a remainder that gives

$$\begin{aligned} \lim_{\epsilon \to 0} k_\epsilon^\alpha(x, z) &= \lim_{\epsilon \to 0} g^{\alpha l}(x, \epsilon z)\, \partial_l V_1(x, z) \\ &= G^{\alpha\gamma}(x) \left(\partial_\gamma V_1(x, z) - \omega_j^i(X_\beta) z^j \partial_i V_1(x, z)\right) \end{aligned}$$

The second term which contains the 1-form coefficients $\omega_j^i(\cdot)$ of the normal connection is determined by those off-diagonal terms of the inverse metric tensor which are linear in z, as follows upon Taylor expanding the inverse of g in powers of z; since $g^{\alpha i}(x, 0) = 0$, the singular term vanishes completely (cf. Appendix B). Clearly only terms that are quadratic in z will survive the averaging procedure, since z has zero mean; therefore all terms $\omega_j^i(\cdot) z^j z^i$ with $i \neq j$ are averaged out, where the additional z^i comes from the partial derivative of the quadratic potential. However $\omega_j^i(\cdot)$ is a skew-symmetric form and thus $\omega_i^i(X) = 0$ (see the remark below).

In order to complete the proof it remains to evaluate $\bar{f}^\alpha = \bar{h}^\alpha + \bar{k}^\alpha$ with $\bar{h} = h_0$. Since $g^{\alpha l}(x, 0) = \delta_\beta^l G^{\alpha\beta}$ we have $\bar{b}^\alpha = b^\alpha(x, 0)$ and therefore

$$\begin{aligned} \bar{f}^\alpha &= -G^{\alpha\gamma} \int \partial_\gamma V_1(x, z) \mu_x(dz) + \bar{b}^\alpha + \bar{a}^{\alpha\gamma}\dot{W}_\gamma \\ &= -G^{\alpha\gamma}\partial_\gamma \left(V_M(x) + \frac{1}{2\beta} \ln \det K(x)\right) + \bar{b}^\alpha + \bar{a}^{\alpha\gamma}\dot{W}_\gamma\,. \end{aligned}$$

Noting that $\operatorname{grad} \bar{V} = G^{-1}\nabla\bar{V}$ and $\bar{a} = \sqrt{2\beta^{-1}G^{-1}}$ we see that $\bar{f}$ is the right hand side of (3.78). Finally, convergence in probability $x_\epsilon(t) \to x(t)$ is a straight consequence of the Averaging Principle for non-degenerate diffusion processes [24]. (See also the recent paper [17] for a convergence proof.) □

Note that the degree of complexity in the reduced equations is of course a matter of how the approximant M is embedded into the $\mathbf{R}^n$, since the metric G is induced

[16] Moreover the dilation has the consequence that we can extend the average of the slow process over the full fibres in the normal bundle (i.e., without the restriction to the tubular neighbourhood), as effects of the extrinsic geometry vanish anyway as ϵ goes to zero.

by the embedding $M \subset \mathbf{R}^n$ which is open to choice. In point of fact, M will often be a linear subspace of $\mathbf{R}^n$, such that the reduced equations simply become

$$\dot{x}(t) = -\mathrm{grad}\,\bar{V}(x(t)) + \sqrt{2\beta^{-1}}\dot{W}(t)\,, \quad (\mathrm{grad}\,\bar{V} = \nabla\bar{V})\,.$$

Strong confinement limit: mechanical system We have to be careful with regard to naïve application of the Averaging Principle: the situation is less clear here than in the diffusion case, since in general the equations do not admit a unique invariant measure. Therefore we shall restrict our attention to the stochastic version of the equations of motion (i.e., with randomized momenta and distributed initial conditions) and give only informal statements concerning convergence (cf. [24, 215]). We support our conjectures by suitable numerical examples below.

We consider an ϵ-family of Lagrangians $L_\epsilon : TNM \to \mathbf{R}$ with the modelling potential V_ϵ that has been substituted for the molecular potential. Using bundle coordinates (x, y) the Euler-Lagrange equations can be written in first-order form

$$\begin{aligned}
\dot{x}^\alpha_\epsilon &= u^\alpha_\epsilon \\
\dot{u}^\alpha_\epsilon &= -\Gamma^\alpha_{kl}(x_\epsilon, y_\epsilon)w^k_\epsilon w^l_\epsilon - g^{\alpha l}(x_\epsilon, y_\epsilon)\partial_l V_\epsilon(x_\epsilon, y_\epsilon) \\
\dot{y}^i_\epsilon &= v^i_\epsilon \\
\dot{v}^i_\epsilon &= -\Gamma^i_{kl}(x_\epsilon, y_\epsilon)w^k_\epsilon w^l_\epsilon - g^{il}(x_\epsilon, y_\epsilon)\partial_l V_\epsilon(x_\epsilon, y_\epsilon)
\end{aligned} \tag{3.81}$$

with the shorthand $w = (u, v)$ for the tangent space coordinates. As before we introduce scaled coordinates $z = y/\epsilon$ in order to prevent the normal energy from diverging for $\epsilon \to 0$. The thus scaled equations of motion are

$$\begin{aligned}
\dot{x}^\alpha_\epsilon &= u^\alpha_\epsilon \\
\dot{u}^\alpha_\epsilon &= -\Gamma^\alpha_{\epsilon,kl}w^k_\epsilon w^l_\epsilon - g^{\alpha\gamma}_\epsilon\partial_\gamma V_1 - \frac{1}{\epsilon}g^{\alpha j}_\epsilon\partial_j V_1 \\
\dot{y}^i_\epsilon &= \frac{1}{\epsilon}v^i_\epsilon \\
\dot{v}^i_\epsilon &= -\Gamma^i_{\epsilon,kl}w^k_\epsilon w^l_\epsilon - g^{i\alpha}_\epsilon\partial_\alpha V_1 - \frac{1}{\epsilon}g^{ij}_\epsilon\partial_j V_1
\end{aligned} \tag{3.82}$$

with the same abbreviation as before: $g_\epsilon = g(x, \epsilon z)$ and $\Gamma^h_{\epsilon,kl} = \Gamma^h_{kl}(x, \epsilon z)$. The Lagrangian that corresponds to the scaled Euler-Lagrange equations then is $K_\epsilon(x, z, \dot{x}, \dot{z}) = L_\epsilon(x, \epsilon z, \dot{x}, \epsilon\dot{z})$. If we let $E_\epsilon(r, s)$ with $r = (x, z)$ and $r = \dot{s}$ denote the total energy of the Lagrangian K_ϵ, then the corresponding invariant Gibbs measure can be written in terms of a smooth density. That is, for each value of ϵ we have

$$\nu^\epsilon(dr, ds) = Z^{-1}_\epsilon \exp\left(-\beta E_\epsilon(r, s)\right)\det g_\epsilon(r)\, drds\,.$$

The finite energy scaling has the effect that the Gibbs measure ν^ϵ will contract to the Gibbs measure on TM as ϵ goes to zero with an additional term that comes from the scaled constraining potential $U_1(x, z)$. Indeed

$$\nu^0 \propto \exp\left(-\beta(E_1(x, 0, \dot{x}, 0) + U_1(x, z))\right)\det G(x)\,.$$

Scaling the free variable according to $t \mapsto \epsilon t$ (microscopic timescale), we find

$$\begin{aligned}
\dot{x}^\alpha_\epsilon &= \epsilon u^\alpha_\epsilon \\
\dot{u}^\alpha_\epsilon &= -\epsilon\Gamma^\alpha_{\epsilon,kl}w^k_\epsilon w^l_\epsilon - \epsilon g^{\alpha\gamma}_\epsilon\partial_\gamma V_1 - g^{\alpha j}_\epsilon\partial_j V_1 \\
\dot{y}^i_\epsilon &= v^i_\epsilon \\
\dot{v}^i_\epsilon &= -\epsilon\Gamma^i_{\epsilon,kl}w^k_\epsilon w^l_\epsilon - \epsilon g^{i\alpha}_\epsilon\partial_\alpha V_1 - g^{ij}_\epsilon\partial_j V_1\,.
\end{aligned}$$

Sending $\epsilon \to 0$ and exploiting that $g_\epsilon^{\alpha j} \to 0$ in the equation for u^α, since the off-diagonal entries of the inverse metric g_ϵ^{-1} vanish, we have $\dot{x}(t) \to 0$ and $\dot{u}(t) \to 0$. This yields equations of motion for $z(t)$ conditioned on the frozen slow variable x

$$\ddot{z}_x^i = -g^{ij}(x,0)\,\partial_j V_1(x,z) = -\partial_i U_1(x,z)\,,$$

such that the conditional invariant measure becomes

$$\nu_x(dz,dv) = \frac{1}{Q(x)}\exp\left(-\beta E_x(z,v)\right)\,dzdv\,.$$

with the conditional normal energy

$$E_x(z,v) = \frac{1}{2}\langle v,v\rangle + U_1(x,z)\,.$$

Observing that $\Gamma^\alpha_{ij,\epsilon} \to 0$ as $\epsilon \to 0$, computing the average of the slow dynamics is no different than in the diffusion case. Since all terms which are linear in v vanish, it remains the average of the potential terms; the mechanical analogue of (3.78) is

$$\begin{aligned}\dot{x}^\alpha &= u^\alpha\\ \dot{u}^\alpha &= -\Gamma^\alpha_{\gamma\delta}(x)u^\gamma u^\delta - G^{\alpha\gamma}(x)\partial_\gamma\bar{V}(x)\end{aligned} \tag{3.83}$$

with the Christoffel symbols $\Gamma^\alpha_{\gamma\delta}$ of the metric G on M and the averaged potential

$$\bar{V}(x) = V_M(x) + \frac{1}{2\beta}\ln\det K(x)\,.$$

Clearly the confined system is Hamiltonian with energy

$$H_0(x,p) = \frac{1}{2}\left\langle G(x)^{-1}p,p\right\rangle + U(x)\,,$$

and we claim that the original model system (3.81) (appropriately randomized) with initial values that are distributed according to $(x,y,u,v) \sim \exp(-\beta E_\epsilon(x,\epsilon y,u,\epsilon v))$ converges in distribution to the (randomized) confined system given by (3.83) with initial conditions that are distributed according to $(x,u) \sim \exp(-\beta H_0(x,Gu))$.

Remark 3.19. *The Langevin equation that is associated to (3.83) reads*

$$\begin{aligned}\dot{x}^\alpha &= \frac{\partial H_0}{\partial p^\alpha}\\ \dot{p}^\alpha &= -\frac{\partial H_0}{\partial x^\alpha} - \hat{\gamma}_{\alpha\delta}\frac{\partial H_0}{\partial p^\delta} + \hat{\varsigma}_{\alpha\delta}\dot{W}^\delta\,.\end{aligned} \tag{3.84}$$

In accordance with Lemma 2.10, friction $\hat{\gamma} = J_\sigma^T\gamma J_\sigma$ and noise coefficients $\hat{\varsigma} = J_\sigma^T\varsigma$ satisfy the fluctuation-dissipation relation, where $J_\sigma = \mathbf{D}\sigma$ is the Jacobian of the embedding $\sigma : \mathbf{R}^{n-s} \to M \subset \mathbf{R}^n$, and $\dot{W}$ denotes the Wiener process in $\mathbf{R}^{n-s}$. Basically the derivation of (3.84) is along the lines of the last paragraph, applying the L^2-convergence result of Kifer [216] for hypo-elliptic diffusion processes; see also [217, 218]. We omit this lengthy calculation, that involves some subtleties (non-resonance and exponential mixing conditions) and refer to the next subsection where a numerical illustration for a Langevin system with an eigenvalue resonance is given.

Example 3.20. Consider the Hamiltonian function $H : T^*\mathbf{R}^2 \to \mathbf{R}$

$$H(x,y,u,v) = \frac{1}{2}u^2 + \frac{1}{2}v^2 + V_\epsilon(x,y)$$

with the potential

$$V_\epsilon(x,y) = \frac{1}{4}(x^2-1)^2 + \frac{1}{2\epsilon^2}\,\omega(x)^2y^2\,,$$

and $x \in \mathbf{R}$, $y \in \mathbf{R}$. The function $\omega(x) \geq c > 0$ is defined as before:

$$\omega(x) = 1 + C \exp\left(-\alpha(x - x_0)^2\right) .$$

As $\epsilon \to 0$ the potential V_ϵ induces a large force pushing a particle towards the equilibrium manifold $y = 0$. Choosing initial values $y = \mathcal{O}(\epsilon)$ the confinement to the x-axis then results in additional force on the particle that is given by the derivative of the Fixman potential. In order to let the energy remain finite we apply a scaling transform to the fast variables, $(y, v) \mapsto (\epsilon y, \epsilon^{-1} v)$. This yields a scaled Hamiltonian H_ϵ to which the following Lagrangian is associated

$$L_\epsilon(x, y, \dot{x}, \dot{y}) = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\left(\epsilon \dot{y}\right)^2 - V_1(x, y) .$$

The corresponding Euler-Lagrange equations can be written as a first-order system

$$\begin{aligned} \dot{x}_\epsilon(t) &= r_\epsilon(t) \\ \dot{r}_\epsilon(t) &= -x_\epsilon(t)(x_\epsilon(t)^2 - 1) - \omega'(x_\epsilon(t))\, \omega(x_\epsilon(t)) y_\epsilon(t)^2 \\ \dot{y}_\epsilon(t) &= \frac{1}{\epsilon} s_\epsilon(t) \\ \dot{s}_\epsilon(t) &= -\frac{1}{\epsilon}\omega(x_\epsilon(t))^2 y_\epsilon(t) . \end{aligned} \tag{3.85}$$

with initial values that are distributed according to $(x, y, r, s) \sim \exp(-\beta E_1(x, y, r, s))$ independently of ϵ. (Here E_1 is the total energy of the Lagrangian L_ϵ for $\epsilon = 1$. Note that without scaling the initial values, the total energy diverges. As a consequence the limit orbits may not lie on the x-axis at all (cf. [219]).) On the slow timescale $t \mapsto \epsilon t$ we find that the fast dynamics alone is given by

$$\begin{aligned} \dot{y}_x(t) &= s_x(t) \\ \dot{s}_x(t) &= -\,\omega(x)^2 y_x(t) , \end{aligned}$$

which can be regarded as a Hamiltonian system with the oscillation energy

$$E_x(y, s) = \frac{1}{2}s^2 + \omega(x)^2 y^2$$

and the conditional invariant measure

$$\mu_x(dy, ds) = \frac{1}{Q(x)} \exp(-\beta E_x(y, s))\, dy ds .$$

Application of the Averaging principle yields the limit equation

$$\begin{aligned} \dot{x}_0(t) &= r_0(t) \\ \dot{r}_0(t) &= -x_0(t)(x_0(t)^2 - 1) - \beta^{-1} \ln \omega(x_0(t)) . \end{aligned} \tag{3.86}$$

Notice that the rightmost term is again the derivative of the Fixman potential. It is furthermore easy to see that in our particular example the mean force is the derivative of the free energy. A comparison of the limit solution and the full solution for various values of ϵ is shown in Figure 9. Apparently, the averaged solution is always pretty close to the limit solution, except at the dynamical barrier. The reason is that the frequency of the fast oscillator is almost constant away from the barrier, such that the two degrees of freedom are virtually decoupled, and averaging trivially gives good approximations (see Section 6 for a detailed discussion of the deviations from the averaged dynamics). For values below $\epsilon = 0.1$ the two solutions are almost indistinguishable; notice that the convergence is even pathwise. The long-term dynamics of the slow variable is depicted in Figure 8. Here we have integrated both the limit solution and the full equation

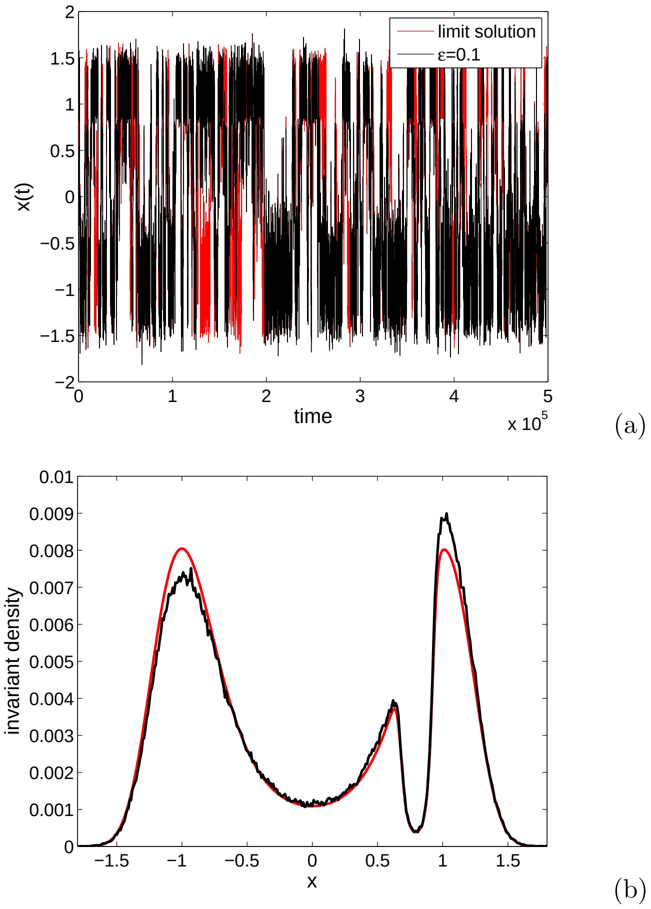


Figure 8. Long-term behaviour of the solution of (3.85) for $\epsilon = 0.1$ versus the limit solution. Upper panel: Typical hybrid Monte-Carlo (HMC) realization for $\beta = 4.0$ and 100 integration steps between the HMC points. Lower panel: invariant density of the slow dynamics computed from 500 000 sample points.

using a hybrid Monte-Carlo scheme with internal step-size $h = 10^{-3}$ and an step-size $\tau = 10^{-2}$ between the Monte-Carlo points, i.e., new momenta were drawn every 10 integration steps. The long-term simulation has been carried out with $\tau = 10^{-1}$ and 500 000 Monte-Carlo points.

Fixman potential reloaded Summarizing, the confinement (also: strong molecular restraint) has the effect that a correction potential, the Fixman potential

$$U = \beta^{-1} \ln \sqrt{\det K}\,,$$

has to be added to the restricted dynamics on M in order to capture the influence of the fast modes [179]. Note that the result is similar to the results in classical mechanics [180, 75, 27], and it is well-known [219] that in case the normal energy is finite the correction potential does not depend on the embedding of M into $\mathbf{R}^n$. Here keeping the

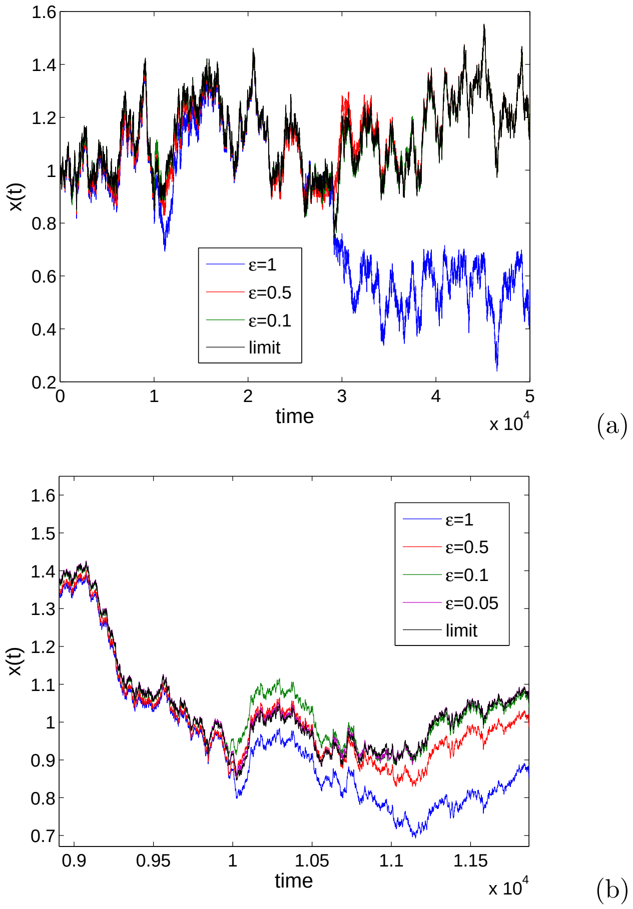


Figure 9. The two plots illustrate convergence of the full system of equations (3.85) towards the limit system (3.86). It can be seen that the error for a typical realization of a HMC trajectory is maximum at the dynamical barrier $x = x_0$; for $\epsilon = 1$ it even happens that the full dynamics makes a transition to a neighbouring metastable set and deviates completely. All simulations have been carried out at the temperature $\beta = 4.0$, and we have have chosen the parameters $A = 15$, $\alpha = 200$, $x_0 = 0.8$ for the frequency function $\omega(x)$. The lower panel gives a zoom into the upper graphics around $x = 1$.

normal energy bounded is achieved by the dilatation $y \mapsto \epsilon y$ of the fibres in the normal bundle. In the example above it turned out that the mean force could be expressed as the derivative of the (geometric) free energy. However in general the Fixman potential is different from the free energy which very well depends on the extrinsic geometry as we have seen in the section on free energy (cf. the discussion about dynamical barriers in the specific case of a flat geometry).

Before we conclude let us let us briefly clarify the relation between the Fixman potential here and the quantity formerly denominated the Fixman potential, viz., $W = \beta^{-1} \ln \mathrm{vol} J_\Phi$. To this end we remind the reader that $\mathrm{vol} J_\Phi = \sqrt{\det \mathbf{D}\Phi^T \mathbf{D}\Phi}$ for a function $\Phi : \mathbf{R}^n \to \mathbf{R}^s$, and we consider a free Hamiltonian system onto which a constraint $\Phi(q) = \xi$ is imposed by adding the confining potential

$$W_\epsilon(q) = \frac{1}{2\epsilon^2}\left(\Phi(q) - \xi\right)^2 .$$

As before the spatial initial conditions $q_0 = q_\epsilon(0)$ are located in a tubular ϵ-neighbourhood of $\Sigma = \Phi^{-1}(\xi)$. That is, we require $\Phi(q_0) - \xi = \mathcal{O}(\epsilon)$ in order to prevent the energy from diverging in the limit $\epsilon \to 0$. By expanding W_ϵ in terms of the normal coordinates around the constraint manifold and repeating the calculation from above, it is then straightforward to show that the Fixman potential W becomes the potential of the limiting confining force perpendicular to Σ.

In molecular simulations, the Fixman potential is sometimes added to a constrained Hamiltonian (e.g., with frozen molecular bonds) in order to mimic unconstrained dynamics and to reproduce the correct statistics of an unconstrained system [28]. By the way, the same can be done for Brownian dynamics [17]. However as we have argued in the proof of Lemma 3.18 and in the last example (cf. Figure 9), the convergence of the confined system to the limit system is often pathwise. That is, by adding the Fixman potential to a constrained system do even approximate single trajectories of the stiff, unconstrained system.

Remark 3.21. *Let us shortly comment on the relevance of the connection 1-forms $\omega_i^i(X)$ associated with the normal frame. There is one possible scenario where the connection gives contributions to the average force, namely, if the embedded manifold has singular points σ_* where $n_i(\sigma_*) = 0$ for some of the normal vectors. In this case $\omega_i^i(X)$ is different from zero, and in some $\sqrt{\epsilon}$-scale neighbourhood of these points the averaged dynamics will differ from the full solution. However it follows from Sard's Theorem that such points form a set of measure zero, and therefore the confinement result holds whenever the reaction coordinate is sufficiently smooth.*

3.4.1. Resonances in molecular systems

For purely deterministic systems it is well-known that eigenvalue crossings in the matrix K of the confining potential may have large impact on the limit equation. It is an open question whether degeneracies of the matrix K can affect the approximation capabilities of the stochastic limit system as well. To address this question, let us briefly review the Averaging Principle for almost integrable system as it appears in celestial mechanics. To this end we follow the outline in [98] and consider the Hamiltonian $H_\epsilon = H_\epsilon(I, \varphi)$ that is assumed to give rise to the following weakly perturbed system

$$\dot{I} = \epsilon f(I, \varphi, \epsilon) \tag{3.87}$$

$$\dot{\varphi} = -\omega(I) + \epsilon g(I, \varphi, \epsilon)\,, \tag{3.88}$$

where $I \in \mathbf{R}^m$ and $\varphi \in \mathbf{T}^m$ (cf. equation (3.2)). In the limit $\epsilon \to 0$ the $I = (I_1, \ldots, I_m)$ become first integrals of the resulting vector field, where the condition $I = I_0$ singles out an invariant torus $\mathbf{T}^m$ with coordinates $\varphi = (\varphi_1, \ldots, \varphi_m)$. For $\epsilon = 0$ the equation $\dot{\varphi} = -\omega(I_0)$ defines a conditional flow on the torus, which can be easily solved,

$$\varphi(t) = \varphi_0 - \omega t\,, \quad \omega = \omega(I_0)\,.$$

Now assume that the right hand side of the slow equation is periodic for $\epsilon = 0$, i.e., $f(I, \varphi + 2\pi, 0) = f(I, \varphi, 0)$. The time average of the slow equation is simply

$$\bar{f}(I_0) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(I_0, \varphi_0 - \omega t, 0)\, dt\,,$$

and is independent of φ_0. The classical Averaging Principle of Neishtadt [220] consists in replacing the full system above by the spatially averaged system

$$\dot{J} = \epsilon \bar{f}(J)\,, \quad \bar{f}(J) = \frac{1}{(2\pi)^m}\int_{\mathbf{T}^m} f(J,\varphi,0)\,d\varphi\,.$$

The last equality states that the conditional flow $\varphi(t)$ is such that in the limit $t \to \infty$ the torus is uniformly sampled which excludes periodic orbits, for example. Basically, replacing the time average by the spatial average requires that the components of the frequency are *non-resonant*. That is, for all $J \in \mathbf{R}^m$ we require (at least) that there are no integer coefficients $k_i \in \mathbf{Z}$, such that

$$k_1\omega_1(J) + \ldots + k_m\omega_m(J) = 0\,, \quad \sum_{i=1}^{s} |k_i| \neq 0. \tag{3.89}$$

If, for instance, the two frequencies of a two-dimensional harmonic oscillator are related by $\omega_1 = k\omega_2$ with $k \in \mathbf{N}$, then the system admits a periodic orbit with $\omega_* = \min(\omega_1, \omega_2)$. Hence the conditional fast flow covers only a one-dimensional submanifold (namely, the periodic orbit) of the two-dimensional torus $\mathbf{T}^2$.

To see how the above problem is related to ours, consider the family Hamiltonians H_ϵ with confinement potential as is obtained as the Legendre transform of the Lagrangian L_ϵ in the last section. We shall restrict our attention to initial value problems at constant energy (i.e., the microcanonical setting). The Hamiltonian reads

$$H_\epsilon(x,y,u,v) = \frac{1}{2}\langle u,u\rangle + \frac{1}{2}\langle v,v\rangle + V_M(x) + \frac{1}{2\epsilon^2}\langle K(x)y,y\rangle\,.$$

By construction, the conditional system of equations for frozen x is integrable. Hence coordinates (I,φ) exist, such that there is a (x,ϵ)-parameter family of canonical (i.e., symplectic) transformations. The corresponding family of Hamiltonians is

$$H_{x,\epsilon}(I,\varphi) = \sum_{k=1}^{s} I_k(x,\epsilon)\omega_k(x)\,,$$

where the $\omega_k(x)$ are square roots of the eigenvalues of $K(x)$, and $I_k = I_k(y,v;x,\epsilon)$. Although $(z,w)_{x,\epsilon} \mapsto (I,\varphi)_{x,\epsilon}$ is a symplectic transformation when x is fixed, the full transformation $S_\epsilon : (x,z,p,w) \mapsto (x,\varphi,p,I)$ is not unless we set $\epsilon = 0$ (note that $\omega = \partial H_0/\partial I$ in (3.87) above). However we can compute the equations of motion with respect to the pulled-back (non-standard) symplectic form, which of course becomes ϵ-dependent [221]. Enforcing the non-resonance condition (3.89) and letting ϵ tend to zero, one obtains an averaged system that is Hamiltonian with the energy [222]

$$H_J(x,p) = \frac{1}{2}\langle p,p\rangle + V_M(x) + \sum_{k=1}^{s} J_k\omega_k(x)\,.$$

Here the averaged action variables $J_k = \bar{I}_k$ are constant and depend solely on the initial conditions $(x(0), y(0), v(0))$ of the original system. Hence also in microcanonical setting the confinement has the effect that an additional potential is added to the constrained dynamics on T^*M. This should be compared to the Fixman potential,

$$W_0(x) = \sum_{k=1}^{s} J_k\omega_k(x) \quad \text{vs.} \quad U_0(x) = \beta^{-1}\sum_{k=1}^{s}\ln\omega_k(x)\,,$$

noting that U_0 depends on the temperature $1/\beta$, whereas W_0 only depends on the scaled initial energy of fast system via the initial values $(x(0), y(0), v(0))$ which is easily explained by the different underlying ensemble concepts, i.e., canonical vs. microcanonical; see the monograph [27] for a detailed discussion.

An interesting question is how resonances could affect the confinement result in the molecular dynamics case. For the classical situation it is well-known [20] that the approximation capability of the limit system is related to the exponent $\gamma > 0$ that appears in so-called Diophantine conditions

$$|\langle k, \omega(J)\rangle| > c\|k\|^{-\gamma}, \quad J \in \mathbf{R}^s, \forall k \in \mathbf{Z}^s \backslash \{0\}.$$

That is, if for given γ the measure of frequencies $\omega_k(J)$ that violate the Diophantine condition is large (almost resonant regimes), the averaged system is likely to be a bad approximation to the original dynamics. However the effect of the resonance also depends on how long the system stays in the vicinity of an almost resonant set. If the normal motion is generated by non-degenerate Ornstein-Uhlenbeck processes the system is mixing and we expect no problems. However for a stochastic Hamiltonian system or Langevin dynamics at low friction and noise the situation is less clear.

To determine the measure of the frequency set that violates the Diophantine condition is a tedious and challenging mathematical task that goes far beyond the scope of the present thesis (cf. the articles [223, 224, 225]). Therefore we will not take up this discussion here, but we shall study the problem by means of an illustrative model system instead. To this end consider a singularly perturbed potential which constrains to a submanifold of codimension $s = 2$:

$$U_\epsilon(x, y) = \frac{1}{2\epsilon^2}\langle A(x)y, y\rangle \qquad A(x) = \begin{pmatrix} a_1(x) & c \\ c & a_2(x) \end{pmatrix}$$

with $a_i(x) = (x \pm 1)^2 + \Delta$, and a coupling constant $0 < c \ll 1$. The additive constant $\Delta > 0$ is chosen such that A is a positive matrix (e.g., $\Delta = 2c$). The frequencies ω_k are the eigenvalues of A which are shown in Figure 10. The eigenvalues of A are $\lambda_i = \omega_i^2$

$$\lambda_i(x) = \frac{a_1(x) + a_2(x)}{2} \pm \sqrt{\frac{(a_1(x) - a_2(x))^2}{4} + c^2}.$$

Note that at $x = 0$ the eigenvalues are separated by a gap of width $\Delta\lambda = 2c$ (avoided crossing). As $c \to 0$ the gap closes, and the system has a resonance $\omega_1 = \omega_2$.

We compare the classical singularly perturbed Hamiltonian initial value problem and compare it to the stochastic Hamiltonian system with randomized momenta. To this end consider the three-dimensional model Hamiltonian

$$H_\epsilon(x, y, u, v) = \frac{1}{2}u^2 + \frac{1}{2}\langle v, v\rangle + \frac{1}{2\epsilon^2}\langle A(x)y, y\rangle,$$

putting forward the equations of motion

$$\begin{aligned} \dot{x}_\epsilon &= u_\epsilon \\ \dot{u}_\epsilon &= -\frac{1}{2\epsilon^2}\langle A(x_\epsilon)y_\epsilon, y_\epsilon\rangle \\ \dot{y}_\epsilon &= v_\epsilon \\ \dot{v}_\epsilon &= -\frac{1}{\epsilon^2}A(x_\epsilon)y_\epsilon. \end{aligned} \tag{3.90}$$

The system is integrated subject to the initial conditions $(x(0), y(0), u(0), v(0)) = (x_*, \epsilon y_*, u_*, v_*)$. The associated limit Hamiltonian has the form

$$H_J(x, u) = \frac{1}{2}u^2 + J_1\omega_1(x) + J_2\omega_2(x)$$

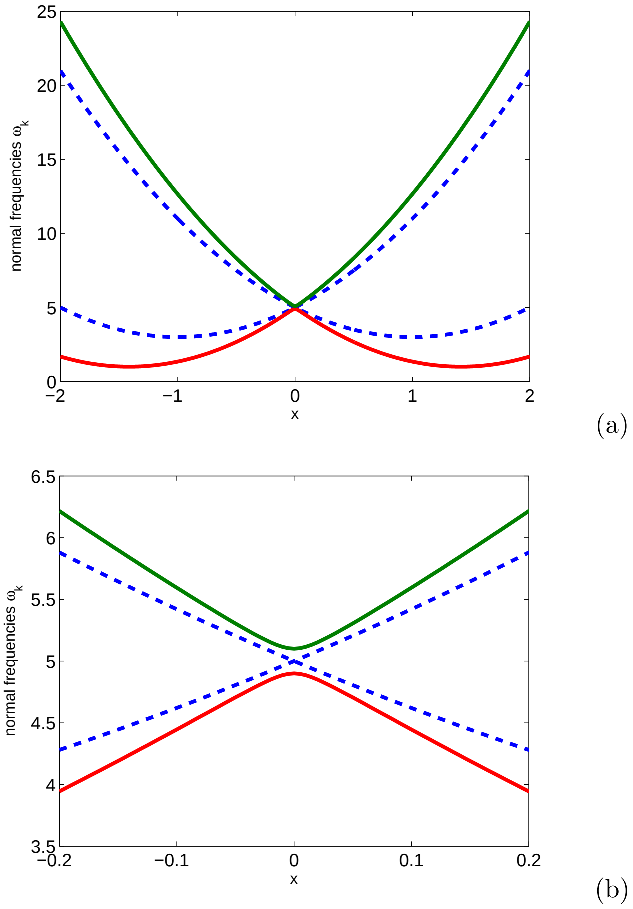


Figure 10. Eigenvalues of the matrix A: the dotted blue lines show a_1 and a_2), whereas the red and green curves show the eigenvalues $\lambda_1 = \omega_1^2$ and $\lambda_2 = \omega_2^2$. As the zoom in the lower panel illustrates the eigenvalues exhibit an *avoided crossing* at $x = 0$ with frequency gap $\Delta\lambda = 2c$ (right panel).

with the frequencies $\omega_i(x) = \sqrt{\lambda_i(x)}$ from above and the action variables [222]

$$J_i = \frac{1}{\omega_i(x_*)}\left(\frac{1}{2}w_i^2 + \frac{1}{2}\omega_i^2(x_*)z_i^2\right)$$

Here $z = C(x_*)y_*$, and $w = \dot{z}$, where $C(x) \in O(2)$ is the orthogonal matrix that point-wise diagonalizes $A(x) = C^T(x)\Lambda(x)C(x)$. We start the integration of the full Hamiltonian system (3.90) with initial values that are chosen such that the action variables $I_k^\epsilon(t) = I_k(y(t), v(t); x(t), \epsilon)$ satisfy $I_1^\epsilon(0) \approx 1$ and $I_2^\epsilon(0) \approx 0$. Then as $\epsilon \to 0$ we expect that the action variables uniformly converge to the adiabatic invariants, $I_k^\epsilon \to J_k$. As can be seen from Figure 11 the action variables remain almost constant unless the system reaches the resonant regime around $x = 0$, where energy is suddenly transferred from one normal mode (oscillation) to the other, such that the action variables vary significantly. For fixed coupling constant $c > 0$ between the oscillators these *non-adiabatic transitions* become weaker as ϵ decreases. In fact it is known that

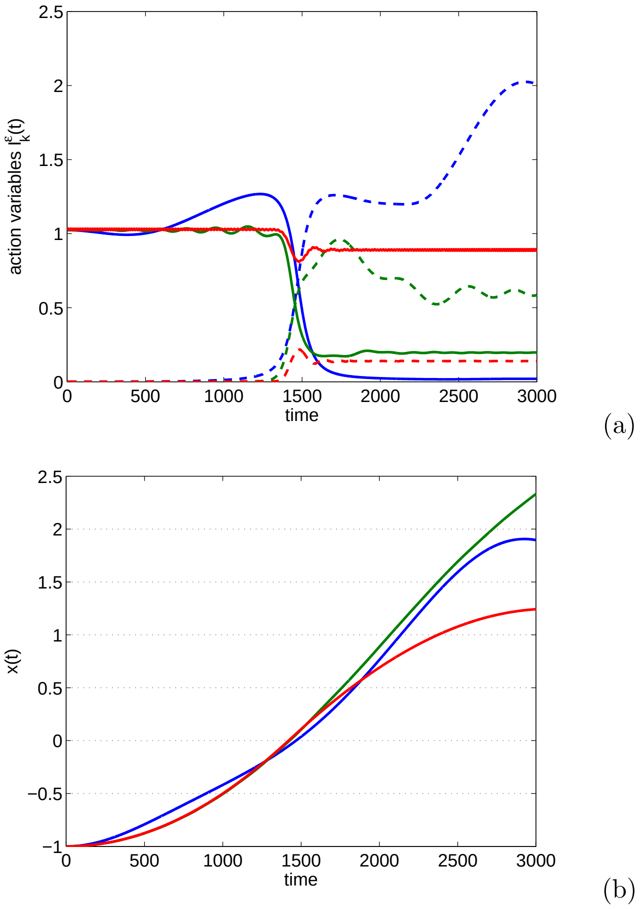
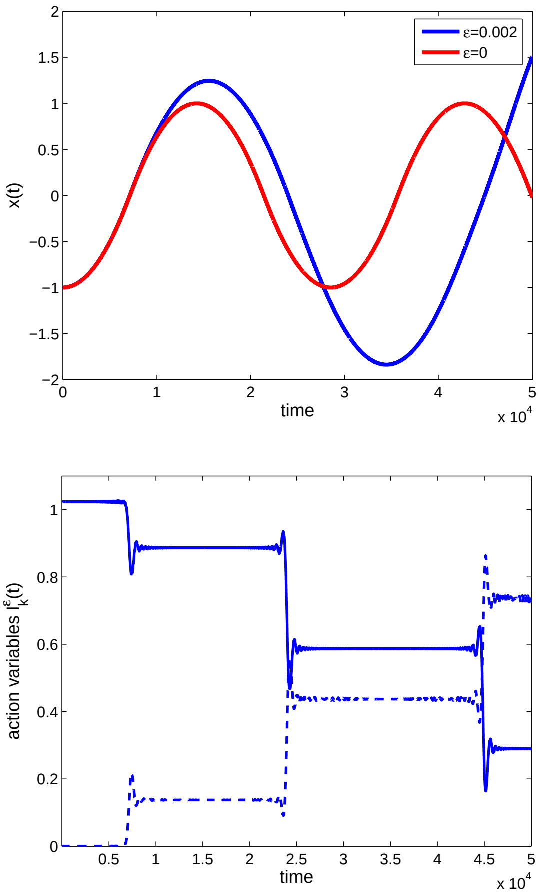


Figure 11. Jump of the action variables $I_k^\epsilon(t)$ at the avoided crossing. It can be seen that jumps occur exactly when the dynamics reaches $x = 0$. The plots show the dynamics of the $I_k^\epsilon(t)$ for $\epsilon = 1$ (blue), $\epsilon = 0.1$ (green), and $\epsilon = 0.01$ (red). All numerical simulations were carried out at constant stepsize $h = 0.0002$ using a symplectic Leapfrog/Verlet scheme with initial values $(x(0), y^1(0), y^2(0), u(0), v^1(0), v^2(0)) = (-1, 0, \epsilon, 0, 0, 0)$.

non-adiabatic transitions occur in a $\sqrt{\epsilon}$-neighbourhood of a resonance [226, 227].

Of course we have to keep in mind that we are not interested in tracing the I_k^ϵ but rather in approximating the slow variable $x_\epsilon(t)$ by the effective motion $x(t)$ which is generated by H_J. Here the situation is even worse, since once the system has passed through the (avoided) crossing, though constant again, the values of the I_k^ϵ have been altered. Yet the limit Hamiltonian H_J is still the same with $J_k = I_k^0(0)$ which is likely not to capture the true dynamics after a non-adiabatic transition has occurred. Hence the limit solution and full solution deviate more and more whenever the system passes through the crossing (see Figure 12).

Now let us repeat the experiment for a stochastic Hamiltonian system with randomized momenta. Of course it does not make sense to look at action variables which are anyway stochastic variables, since they depend on the random momenta



(a)

(b)

Figure 12. The approximation of the full solution with $\epsilon = 0.002$ by the limit solution $\epsilon = 0$ becomes worse each time the system passes through a resonance (crossing). The integration was carried out with step-size $h = 10^{-5}$ and initial values $(x(0), y^1(0), y^2(0), u(0), v^1(0), v^2(0)) = (-1, 0, \epsilon, 0, 0, 0)$.

of the fast dynamics. Anyway there is no limit result which states that they should become constant as ϵ goes to zero. Nonetheless we may compare the slow motion $x_\epsilon(t)$ to the limit motion $x(t)$. A typical realization of the Hamiltonian system (3.90) at the temperature $\beta = 4.0$ is shown in Figure 13. Apparently for relatively large ϵ the avoided crossing does not affect the dynamics at all. Even if we close the eigenvalue gap by letting the coupling constant c go to zero, the limit dynamics still approximates the full dynamics (a typical realization and the corresponding Fixman potential for $c = 0.0001$ is shown in Figure 14 below). Observe that the nascent resonance at $x = 0$ induces an additional potential barrier that renders the system to be (though weakly) metastable. Last but not least we illustrate the dynamics at various temperatures while keeping ϵ, c fixed. We choose $c = \mathcal{O}(\sqrt{\epsilon})$, which is typically considered the worst case (e.g., see [228] and the references therein). For $\epsilon = 0.01$ we observe that for our test problem the system full dynamics and the limit dynamics are almost indistinguishable

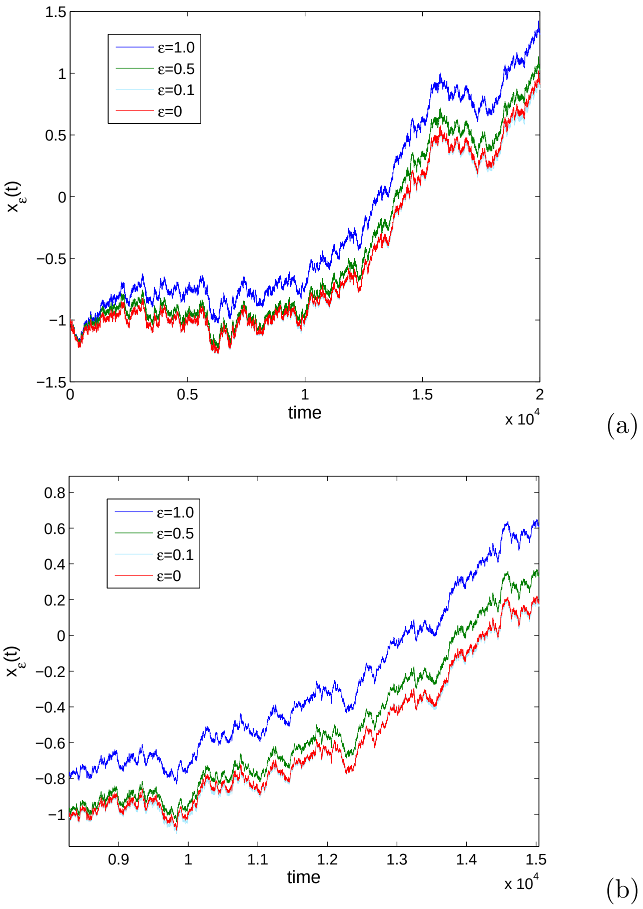


Figure 13. Typical realization for the stochastic Hamiltonian system associated with (3.90) for moderate coupling $c = 0.1$. The simulations were performed using a hybrid Monte-Carlo (HMC) scheme at temperature $\beta = 3.0$ with step-size $h = 0.0005$ for the Leapfrog integrator choosing new momenta every 100 steps. The lower panel shows a zoom into the upper one.

for various values of β (see Figure 15).

Certainly these short simulations are nothing more than illustrations of what can happen in the presence of resonances or almost-resonances (avoided crossings). However they should get the impression to the reader that the impact of resonances on the limiting behaviour of appropriately "thermalized" systems does not seem as severe as for purely deterministic systems.

The reader may wonder if the Fixman potential U_0 is just the average of the deterministic potential W_0 over all initial values with respect to the Gibbs distribution. It is easy to see that this is not the case, for

$$\bar{W}_0(x, x_*) = \sum_{k=1}^{s} \omega_k(x) \int J_k(x_*, y, v)\nu_{x_*}(dy, dv) \neq U_0(x)\,,$$

where the average is with respect to the Gibbs measure ν_x of the normal modes.

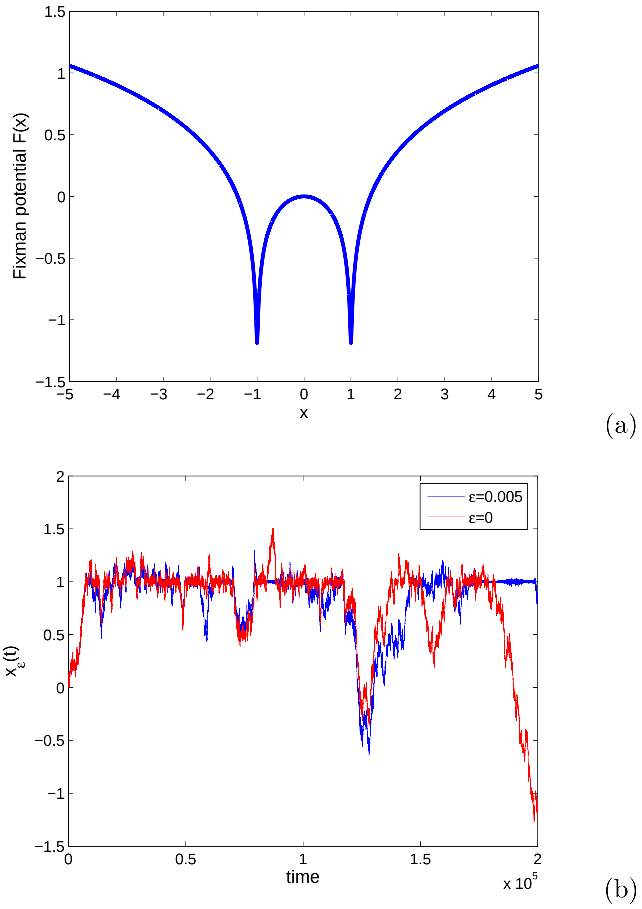


Figure 14. HMC simulation for weak coupling $c = 0.0001$ and $\epsilon = 0.05$. The upper panel shows the corresponding Fixman potential for $\beta = 3.0$.

Remark 3.22. *We take a brief look at Langevin dynamics in the limit of low friction and noise which represents a particular case — even in the absence of resonances: Consider the Langevin equation for the confinement problem. For σ, γ scalar satisfying the fluctuation-dissipation relation $2\gamma = \beta\sigma^2$ we have the equations of motion*

$$\begin{aligned}
\dot{x}_\epsilon &= u_\epsilon \\
\dot{u}_\epsilon &= -\frac{1}{2\epsilon^2}\left\langle A(x_\epsilon)y_\epsilon, y_\epsilon\right\rangle - \gamma u_\epsilon + \sigma \dot{W}_1 \\
\dot{y}_\epsilon &= v_\epsilon \\
\dot{v}_\epsilon &= -\frac{1}{\epsilon^2}A(x_\epsilon)y_\epsilon - \gamma v_\epsilon + \sigma \dot{W}_2\,.
\end{aligned} \tag{3.91}$$

We are interested in the quasi-deterministic limit $\gamma, \sigma \to 0$ with $\gamma \sim \sigma^2$ (constant temperature). For this purpose we introduce a scaling parameter $\delta \ll 1$ and we set $\gamma = \delta\gamma_0$ and $\sigma = \sqrt{\delta}\sigma_0$. As before we dilate the normal coordinates according to $(y, v) \mapsto (\epsilon y, v)$, defining $z = y/\epsilon$ (note that z and v are no longer conjugate variables).

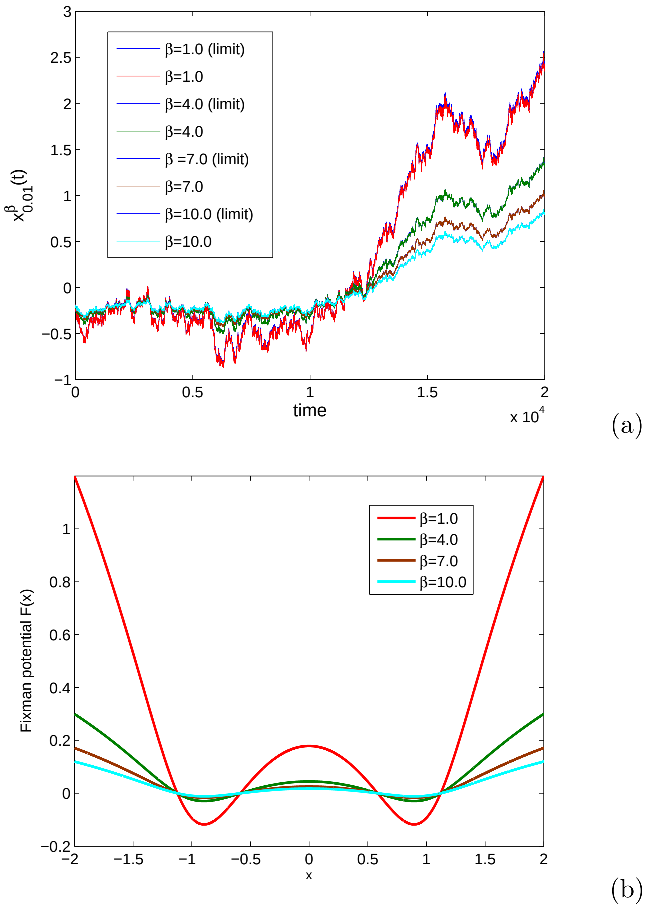


Figure 15. Typical HMC simulations for $c = 0.1$ and $\epsilon = 0.01$ at various temperatures. Note that the limit and the full trajectories are virtually indistinguishable. The lower panel shows the respective Fixman potentials.

On the microscopic (i.e., slow) timescale the Langevin equation now becomes

$$\begin{aligned}
\dot{x}_{\epsilon,\delta} &= \epsilon u_{\epsilon,\delta} \\
\dot{u}_{\epsilon,\delta} &= -\frac{\epsilon}{2}\left\langle A(x_{\epsilon,\delta})z_{\epsilon,\delta}, z_{\epsilon,\delta}\right\rangle - \epsilon\delta\gamma_0 u_{\epsilon,\delta} + \sqrt{\epsilon\delta}\,\sigma_0\dot{W}_1 \\
\dot{z}_{\epsilon,\delta} &= v_{\epsilon,\delta} \\
\dot{v}_{\epsilon,\delta} &= -A(x_{\epsilon,\delta})z_{\epsilon,\delta} - \epsilon\delta\gamma_0 v_{\epsilon,\delta} + \sqrt{\epsilon\delta}\,\sigma_0\dot{W}_2\,.
\end{aligned} \tag{3.92}$$

Suppose the coupling constant $c > 0$ is kept fixed. Even then we are caught in a complicated situation since there are two distinct scaling parameters, where the limiting behaviour very much depends on the order of letting ϵ, δ tend to zero, and we have to consider certain distinguished limits. *Roughly speaking, $\delta \to 0$ brings us straight to the deterministic world, and the description using the Fixman potential becomes inappropriate, whereas letting ϵ go to zero first amounts to the fully stochastic situation. Therefore it is recommendable to couple the two scales in a way that $\epsilon \sim \delta$.*

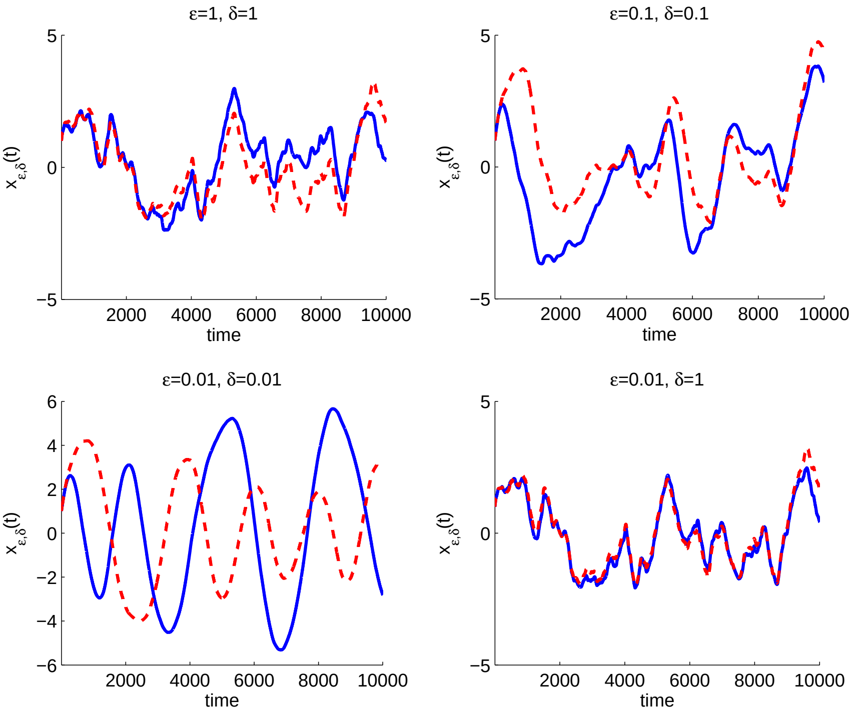


Figure 16. Typical realizations of the slow variable $x_{\epsilon,\delta}$ for the two-parameter Langevin equation (3.92) with coupled parameters $\epsilon \sim \delta$ (blue curves: full system, red curves: limit dynamics for $\epsilon = 0$). The realizations indicate that for $\epsilon, \delta \to 0$ the averaged dynamics with the Fixman potential does no longer approximate the full (quasi-deterministic) system. In contrast, taking the limit $\epsilon \to 0$ while keeping $\delta = 1$ fixed leads to the usual (stochastic) limiting behaviour which is also robust in the vicinity of the avoided crossing.

Letting now ϵ, δ go to zero we see that friction and noise vanish at a higher rate than the slow variable $x_{\epsilon,\delta}$ freezes. Hence the assumptions underlying the Averaging Principle fail, for the fast dynamics does no longer admit a unique invariant measure. Accordingly we expect that the Fixman potential does not provide the correct limit description for $\epsilon, \delta \to 0$, even far away from the avoided crossing.

Indeed the realizations shown in Figure 16 indicate that for $\epsilon, \delta \to 0$ the averaged system of equations (3.84) with the Fixman potential does no longer approximate the full (quasi-deterministic) system. In contrast, taking the limit $\epsilon \to 0$ while keeping $\delta \gg \epsilon$ fixed leads to the usual (stochastic) limiting behaviour which is also robust in the vicinity of the avoided crossing. We emphasize that these hand-waving arguments can only provide restricted insight; a rigorous study of the two-parameter system (3.92) requires profound knowledge of the system itself and careful analysis of the distinguished limits which cannot be given here. For the method of distinguished limits and perturbative multiscale expansions we refer to [229] and the references therein.

3.4.2. Relations to geometric singular perturbation theory This whole section has surveyed different techniques for the elimination of fast degrees of freedom. All these techniques have in common that the fast degrees of freedom are averaged out with respect to some particular probability distribution that is either the invariant measure of the fast dynamics (Averaging Principle) or a prescribed probability measure (optimal prediction). Here we shall briefly mention yet another approach which proceeds by discarding (and hence disregarding) the fast variables, which is reasonable under certain conditions. Let us consider a deterministic slow-fast system

$$\begin{aligned} \dot{x}(t) &= f(x(t), y(t), \epsilon) \\ \dot{y}(t) &= \frac{1}{\epsilon} g(x(t), y(t), \epsilon)\,, \end{aligned} \tag{3.93}$$

where $\epsilon \ll 1$, and $(x, y) \in \mathbf{R}^d \times \mathbf{R}^s$ are slow and fast coordinates, respectively. So far we have considered the limit $\epsilon \to 0$, but the limiting equation clearly depends on how the limit is reached. In fact by simply setting $\epsilon = 0$, the system degenerates to a differential-algebraic equation of the form

$$\begin{aligned} \dot{x}(t) &= f(x(t), y(t), 0) \\ 0 \quad &= g(x(t), y(t), 0)\,. \end{aligned}$$

Suppose that g is sufficiently smooth, such that the equation $g(x, y, 0) = 0$ defines a differentiable manifold $M = g^{-1}(0)$. Further assuming that $\mathbf{D}_2 g(x, y, 0) \neq 0$ on M, the Implicit Function Theorem states that we can locally solve for $y = h(x)$. Upon reinserting h into the slow equation we obtain the reduced system[17]

$$\dot{x}(t) = F(x(t))\,, \quad F(x) = f(x, h(x), 0)\,. \tag{3.94}$$

In some sense this restriction can be understood as averaging over the fast variables, where the corresponding conditional invariant measure is singular with support on M, i.e., $\mu_x(dy) = \delta_M(x, y)$. It has been shown [230, 231] that, if M is uniformly asymptotically stable, then the full system (3.93) stays in a tubular ϵ-neighbourhood of M, such that it can be approximated by solving the reduced system (3.94).

The proper geometric description of the dynamics in the vicinity of the invariant manifold M is due to Fenichel [232], who has shown that for sufficiently small ϵ an invariant manifold M_ϵ exist that can be parametrized by a formal series

$$\xi = \xi(x, \epsilon) \quad \text{with} \quad \xi(x, \epsilon) = h(x) + \epsilon h_1(x) + \epsilon^2 h_2(x) + \dots\,.$$

The corresponding reduced equations of motion for $0 < \epsilon \ll 1$ then are

$$\dot{x}(t) = F_\epsilon(x(t))\,, \quad F_\epsilon(x) = f(x, \xi(x, \epsilon), \epsilon)\,.$$

For the general theory and conditions that guarantee convergence of the formal power series we refer to the review [233] and the references given there. Nicely, the above considerations can be generalized to stochastic systems of Smoluchowski type

$$\begin{aligned} \dot{x}(t) &= f(x(t), y(t), \epsilon) + \sigma a(x(t), y(t), \epsilon)\dot{W}(t) \\ \dot{y}(t) &= \frac{1}{\epsilon} g(x(t), y(t), \epsilon) + \frac{\sigma}{\sqrt{\epsilon}} b(x(t), y(t), \epsilon)\dot{W}(t) \end{aligned} \tag{3.95}$$

with $\sigma^2 = 2/\beta$. By applying the above arguments to the deterministic part in the stochastic equations of motion, and imposing some non-degeneracy condition on the

[17] We call M uniformly (hyperbolic) asymptotically stable, if and only if all eigenvalues of the Jacobian $\mathbf{D}_2 g(x, h(x), 0)$ have negative real parts and are uniformly bounded away from zero.

covariance matrix aa^T of the noise it has been shown recently [130] that the sample paths remain concentrated inside a tubular σ-neighbourhood of M_ϵ. Under certain conditions it is then possible to approximate (3.95) by the reduced stochastic system

$$\dot{x}(t) = F_\epsilon(x(t)) + \sigma A_\epsilon(x(t))\dot{W}(t) \tag{3.96}$$

with

$$F_\epsilon(x) = f(x, \xi(x,\epsilon), \epsilon) \quad \text{and} \quad A_\epsilon(x) = a(x, \xi(x,\epsilon), \epsilon)$$

The reduced equation provides an approximation up to the first exit time τ_ϵ from M_ϵ. The approximation is of order $\sigma\sqrt{\epsilon(1+\chi(t))}$, where $\chi(t)$ depends on the associated deterministic system and is bounded whenever the deterministic system admits a uniformly hyperbolic, asymptotically stable invariant manifold. In particular for $\epsilon = 0$ the reduced system gives simply the slow diffusion restricted to the invariant manifold $M = M_0$ that is defined by the algebraic equation $g(x, y, 0) = 0$.

Replacing the full system (3.95) by the reduced system (3.96) in a controlled manner involves many subtleties; in particular the first exit time τ_ϵ from the invariant manifold goes to zero as $\epsilon \to 0$, and therefore the estimation for the approximation error becomes useless. For the technical intricacies we refer to [130, 234].

Example 3.23. Reconsider our familiar confinement problem for a diffusion process in $\mathbf{R}^2$. Using the scaling $y = \epsilon z$ of the fast coordinate we have the system of equations

$$\begin{aligned} \dot{x}_\epsilon &= -\partial_x V(x_\epsilon) - \partial_x \omega(x_\epsilon)\omega(x_\epsilon) z_\epsilon^2 + \sigma \dot{W}_1 \\ \dot{z}_\epsilon &= -\frac{1}{\epsilon^2}\omega^2(x_\epsilon) z_\epsilon + \frac{1}{\epsilon}\sigma \dot{W}_2 \,. \end{aligned} \tag{3.97}$$

with the sharply-peaked frequency (see Figure 4)

$$\omega(x) = 1 + C \exp\left(-\alpha(x - x_0)^2\right) \,. \tag{3.98}$$

The invariant manifold of the deterministic equation that is defined by the condition $z = 0$ is clearly uniformly hyperbolic and asymptotically stable, for $\omega(x) \geq c > 0$. For fixed $\epsilon > 0$ the diameter of the invariant manifold M_ϵ is determined by the second derivative of the constraining potential, and it becomes wider, if $\omega(x)$ is large, i.e., the potential is stiff, and it becomes narrower, if $\omega(x)$ is small. This accounts for the fact that for a stiff potential there is less spreading of trajectories. For $\epsilon = 0$ the reduced system turns out to be the confined system (3.78), but without the additional Fixman potential,

$$\dot{x}_0 = -\partial_x V(x_0) + \sigma \dot{W}_1 \,.$$

As we have seen throughout several examples, the confined system including the Fixman potential $U = \beta^{-1} \ln \omega$ approximates the full dynamics rather well, and the reader may wonder, if solutions of the last equation can do better. Figure 17 shows a typical realization of the Smoluchowski equation above for small ϵ versus the averaged and the restricted dynamics. The plot clearly indicates that the averaged dynamics yields the better approximation. Especially the long-term behaviour (the invariant distribution) is not captured by the restricted dynamics at all.

Of course even for $\epsilon > 0$ the reduced equations on the invariant manifold M_ϵ were never meant to approximate the long-term behaviour of the full system, since the system is likely to leave M_ϵ after some time. Nevertheless we mention this approach, as discarding fast harmonic and quasi-harmonic motions is quite common in molecular applications; for instance, almost every popular molecular dynamics code imposes

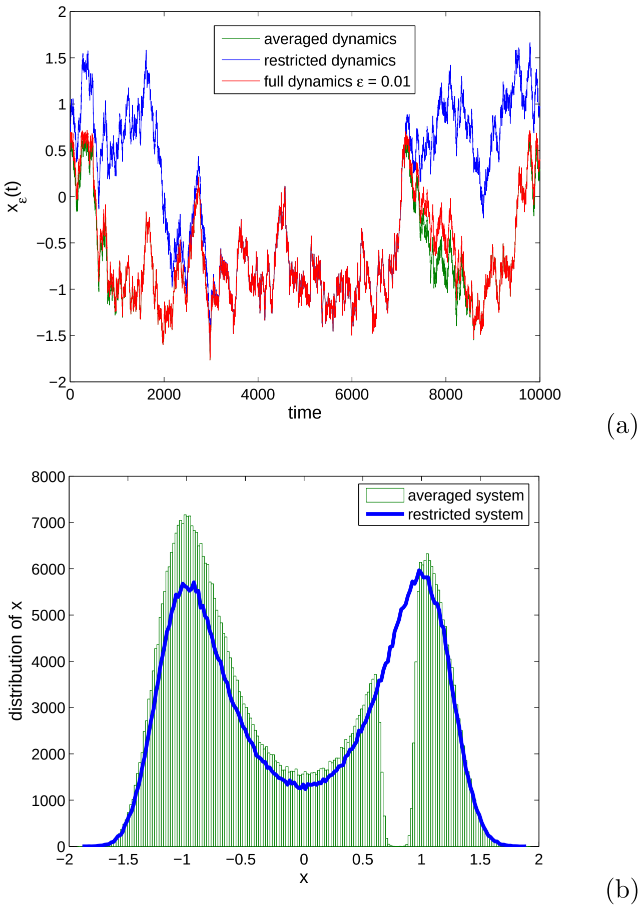


Figure 17. The upper panel shows typical realizations of the slow-fast Smoluchowski equation (3.97) versus the averaged and its restricted limit equation. The integration was performed using an Euler-Maruyama scheme with step-size $h = 10^{-4}$ and initial values $(x(0), z(0)) = (x_0, 0)$ as is consistent with the restriction to the invariant manifold M. The lower panel shows unnormalized histograms of the slow coordinates. Notice that only the averaged system reproduces the three metastable sets correctly, since the additional barrier at $x = 0.8$ stems from the entropy contribution of the fast modes (cf. the discussion regarding the entropy contribution of fast bond vibrations in Section 3.1.1).

constraints on the fast bond vibrations without accounting for their contribution (given by the Fixman potential) to the remaining system.[18]

Spatial decomposition methods reconsidered If the invariant manifold M is known from the outset there are plenty of methods to restrict a system to it. For first-order systems and invariant manifolds that are linear subspaces of the systems' configuration space a convenient route is the Galerkin projection: Recall the discussion

[18] In fact, this is not quite correct, since many molecular force fields are parametrized such as to reproduce certain physical effects subject to frozen bond lengths or even bond angles [235].

from Section 2.4, and let $z \in \mathbf{R}^n$ denote the original configuration variable. Denote further by P the $n \times d$ matrix the rows of which span the d-dimensional subspace M. Then $PP^Tz \in M$, and we can introduce local coordinates $x = P^Tz$ on M. The Galerkin projection then consists in the projection of the full system

$$\dot{z}(t) = f(z(t), t),\ z \in \mathbf{R}^n$$

onto the tangent space of M. That is,

$$\dot{x}(t) = P^T f(Px(t), t),\ x \in \mathbf{R}^d .$$

For mechanical systems one has to be more careful, since the Galerkin projection does not preserve the Hamiltonian property of the system, even if it is written as a first-order system. The canonical way to restrict a mechanical system to a submanifold of its configuration space is by means of holonomic constraints [163]. That is, the restriction of the equations of motion is obtained by, firstly, restricting the original Lagrangian to TM and then, secondly, computing the corresponding Euler-Lagrange equations. It clearly depends on the particular system whether the reduced equations are really simpler to evaluate than the original ones. For example, if $f = -\operatorname{grad} V$ in the equations above, where V is the molecular potential, then the right hand side of the reduced equations still requires the gradient evaluation of the full molecular force field which is typically the most expensive operation in numerical simulations.

3.5. Summary and bibliographical remarks

This section briefly revisits the variety of different strategies that have been introduced to systematically deduce reduced models for conformation dynamics of molecules provided a suitable reaction coordinate is known.

Distinct notions of free energy Consider a molecule with configurations $q \in \mathbf{R}^n$ and conjugate momenta $p \in T_q^*\mathbf{R}^n \cong \mathbf{R}^n$. Let further $\Phi : \mathbf{R}^n \to \mathbf{R}^k$ be a smooth reaction coordinate. If the molecular Hamiltonian is denoted by $H = T + V$, then the standard free energy is defined by the marginal density of the reaction coordinate,

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

or

$$F(\xi) = -\beta^{-1} \ln \int_{\Sigma} \exp(-\beta V)(\mathrm{vol} J_\Phi)^{-1} d\sigma_\xi$$

which differs from the former only by an additive constant (recall that $J_\Phi = \mathbf{D}\Phi$). Here $\Sigma = \Phi^{-1}(\xi)$ is the level set of the function Φ that is defined by the equation $\Phi(q) = \xi$, where $d\sigma_\xi$ denotes its surface element. In contrast to that, $d\mathcal{H}_\xi$ is the Hausdorff measure of $\Sigma \times \mathbf{R}^n$ considered as a submanifold of $\mathbf{R}^n \times \mathbf{R}^n$. By construction, F captures the correct statistical weights between different conformations [1, 236].

There is yet another definition that is important in the context of transition state theory [3, 4] which is based on the probability density of the surface $\Sigma \subset \mathbf{R}^n$,

$$G(\xi) = -\beta^{-1} \ln \int_{\Sigma \times \mathbf{R}^n} \exp(-\beta H) d\mathcal{H}_\xi ,$$

or

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

We have termed this second type of free energy the *geometric free energy*, since it depends only on the surface Σ but not on the reaction coordinate Φ. The difference between the two free energies and implications thereof have been clearly stated for the first time in the review [13]. The authors of [5] insist on calling only F a proper free energy, since G is not a function of the reaction coordinate. However we think that only G deserves the name *potential of mean force*, for only the derivative of G can be written as an average generalized force as has been pointed out in Section 3.1.1. Moreover unlike ∇F, only ∇G transform like a 1-form (i.e., a force).

By analyzing the different probability densities underlying the two free energies, we recover the famous Fixman Theorem or the Blue Moon reweighting formula, that allows for computing conditional expectations from constrained simulations [179, 71],

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

The leftmost expectation is a conditional expectation $\mathbf{E}_{\xi}(\cdot) = \mathbf{E}(\cdot\,|\,\Phi(q) = \xi)$, whereas the one on the right denotes the expectation with respect to the Gibbs measure restricted to the fibre $\Sigma = \Phi^{-1}(\xi)$, i.e., $\mathbf{E}_{\Sigma}(\cdot) = \mathbf{E}(\cdot\,|\,q \in \Sigma)$. The formula marks the important difference between a function Φ and a surface Σ that is defined as its level set: there are many functions that have identical level sets. Basically, the Blue Moon formula can be considered an instance of Federer's co-area formula [70]. Accordingly, the reasoning that leads to Blue Moon does not involve any reference to an underlying dynamical system. Therefore, and in contrast to what is commonly asserted, the formula holds whether or not the system involves momenta. Moreover the relation is true for any configurational probability measure. As a straight consequence F and G are related by the simple formula

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

Averaging for stochastic differential equations Consider the diffusion of a molecule with configurations $q \in \mathbf{R}^n$ in the potential energy landscape $V : \mathbf{R}^n \to \mathbf{R}$,

$$\dot{q}(t) = -\mathrm{grad}\, V(q(t)) + \sqrt{2\beta^{-1}}\dot{W}(t) .$$

Suppose we can arbitrarily speed up all variables except the reaction coordinate. Basically this amounts to speeding up the dynamics along the fibres $\Phi^{-1}(\xi)$ for all regular values ξ of the reaction coordinate. Of course it is not possible to find a global coordinate transformation so as to rewrite the above equation in terms of the reaction coordinate and the remaining coordinates. However we can *locally* consider the accelerated dynamics on each fibre $\Sigma = \Phi^{-1}(\xi)$ and average the right hand side of the equations of motion over the invariant measure $\nu_{\Sigma} \propto \exp(-\beta V) d\sigma_{\xi}$ of the thus accelerated dynamics. This yields an effective drift and noise orthogonal to each fibre. In order to recover the *global* picture, we endow the state space that is spanned by the reaction coordinate with an appropriate averaged metric

$$m(\xi) = \mathbf{E}_{\Sigma}(\mathbf{D}\Phi^T \mathbf{D}\Phi)^{-1} .$$

By this we obtain a reduced model for the dynamics of the reaction coordinate

$$\dot{\xi} = -\mathrm{grad}\, G(\xi) + b(\xi) + \sqrt{2\beta^{-1}} a(\xi) \dot{W}_{\xi} ,$$

where $\mathrm{grad}\, G = m^{-1}\nabla G$ is the gradient of the geometric free energy, a is the positive-definite square root of the inverse metric tensor m^{-1}, and $\dot{W}_{\xi}$ denotes standard

Brownian motion in $\mathbf{R}^k$ (here, k is the dimension of the reaction coordinate). The additional drift comes from interpreting the equation in the sense of Itô; it is given by

$$b^i(\xi) = \beta^{-1} m^{jk} \Gamma^i_{jk}$$

with Γ^i_{jk} denoting the symmetric Christoffel symbols associated with the Riemannian metric m. We emphasize that the derivation of the reduces system is based on an arbitrary manipulation of the original model which is not unique.

In point of fact, there is yet another possibility to accelerate the dynamics orthogonal to the reaction coordinate using a projection operator approach. This amounts to a decomposition along the lines of the invariant measure of the system (gluing together different conditional measures). For a single reaction coordinate the authors of [13] derive a reduced equation that involves the free energy F

$$\dot{\xi} = h(\xi)\partial_\xi F(\xi) + \beta^{-1}\partial_\xi h(\xi) + \sqrt{2\beta^{-1}h(\xi)}\dot{W}_\xi \,,$$

where the metric factor h is defined as the conditional expectation

$$h(\xi) = \mathbf{E}_\xi \|\nabla\Phi(q)\|^2 \,,$$

which should be distinguished from the (constrained) expectation with respect to ν_Σ. It is not obvious that the second equation really transforms like an Itô equation, as it does not have the standard covariant form. However it has been demonstrated that it is consistent with Itô formula under transformations of the reaction coordinate. Since this is also true for the other reduced equation one could expect that the two equations are equivalent. Intriguingly this is not the case, unless $\nabla\Phi$ is a function of ξ only. Then $h = m^{-1}$. The difference can be explained by drawing upon to the different decompositions into fast and slow variables (probabilistic versus geometric).

Optimal prediction and the Mori-Zwanzig procedure If the original system is Hamiltonian the methods of choice can be subsumed under the name of *projection operator techniques*. Unlike the ordinary averaging techniques these methods do not explicitly rely on the assumption of time scale separation, and they take into account that the configurational variables and their conjugate momenta are independent variables (i.e., the equations is effectively second-order):

$$\begin{aligned} q^i &= \frac{\partial H}{\partial p_i} \\ p_i &= -\frac{\partial H}{\partial q^i} \,, \quad i = 1, \ldots, n \,. \end{aligned}$$

Let us assume the system is appropriately thermalized, i.e., we consider a stochastic perturbations of the original deterministic system, such that the system at temperature $T = 1/\beta$ is ergodic with respect to the canonical probability measure $\mu \propto \exp(-\beta H)$. Let $\Phi : \mathbf{R}^n \to \mathbf{R}^k$ denote again a reaction coordinate with (yet unknown) conjugate momentum $\Theta : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^k$. Then the conditional expectation

$$\mathbf{E}_{\xi,\eta}(\cdot) = \mathbf{E}\left(\,\cdot \mid \Phi(q) = \xi,\, \Theta(q,p) = \eta\right)$$

defines an orthogonal projection in the Hilbert space $L^2(\mu)$, where $\mathbf{E}(\cdot)$ is meant with respect to μ. Exploiting the best-approximation property of orthogonal projections, one can show that the optimal approximation of Hamilton's equations in $L^2(\mu)$ in terms of ξ and η solely is obtained by the projected equations of motion

$$\begin{aligned} \xi^j &= \frac{\partial E}{\partial \eta_j} \\ \eta_j &= -\frac{\partial E}{\partial \xi^j} \,, \quad j = 1, \ldots, k \,, \end{aligned}$$

where the optimal prediction free energy E (effective Hamiltonian) is defined by

$$E(\xi, \eta) = -\beta^{-1} \ln \int_{T^*\Sigma} \exp(-\beta H) d\mathcal{L}_{\xi,\eta} \,.$$

Here $d\mathcal{L}_{\xi,\eta}$ is the Hausdorff measure of the submanifold $\Sigma \times \mathbf{R}^{n-k} \subset \mathbf{R}^n \times \mathbf{R}^n$ that is defined as the level set of the reaction coordinate and its conjugate momentum.

The *optimal prediction equations* in Hamiltonian form are due to Hald and were stated in [56]. We could show that the effective Hamiltonian E relates to known quantities as the geometric free energy G in the following intuitive way

$$E(\xi, \eta) = \frac{1}{2} \langle I(\xi)\eta, \eta \rangle + G(\xi) + \mathcal{O}(\|\eta\|^4) \,.$$

The effective inverse mass is given by

$$I(\xi) = \mathbf{E}_\Sigma J_\Phi^T J_\Phi \,,$$

where the expectation is understood with respect to the constrained Gibbs measure $\nu_\Sigma \propto \exp(-\beta V) d\sigma_\xi$. Neither G nor I depend on the momentum variables. If the temperature is low as compared to the atomic masses (i.e., $\beta \gg 1$) the Maxwellian momentum distribution is sharply peaked around $\eta = 0$, such that we can neglect all higher-order contributions and interpret the effective Hamiltonian in the usual way as a sum of kinetic and potential energy. Doing so, the reader may wonder whether one could recover the standard free energy by integrating $\exp(-\beta E)$ over the momenta. In fact, integrating out the momenta yields

$$\int \exp(-\beta E) \, d\eta \neq C \exp(-\beta F) \,.$$

That is, the reaction coordinate distribution generated by the optimal prediction system is not given by $\exp(-\beta F)$ which is no surprise whatsoever, as we have neglected all terms that are at least $\mathcal{O}(\|\eta\|^4)$.

The Mori-Zwanzig procedure (e.g., [51, 198, 237] consists in decomposing the Liouville equation that is associated with the Hamiltonian system into a part that acts only in the direction of the reaction coordinate plus a remainder. To this end we define the projection $\Pi = \mathbf{E}_{\xi,\eta}$, $\Pi : L^2(\mu) \to L^2(\mu)$. If $(q(t), p(t))$ denotes the solution of Hamilton's equations depending on initial values $q = q(0)$ and $p = p(0)$, then the generalized Langevin equation for a function $f(t) := f(q(t), p(t))$ reads

$$\partial_t f(t) = \Pi \mathcal{L} f(t) + \int_0^t K(s-t) w(s) \, ds + w(t) \,.$$

Here K is a friction kernel that makes the equation non-Markovian, and w is the solution of an Volterra integral equation that is defined on the subspace orthogonal to the reaction coordinate. The operator $\mathcal{L}$ is the usual Liouvillian that is generated by the Hamiltonian vector field. Although the various terms in the last equation have appealing physical interpretations (drift, friction and noise) the equation is useless without further assumptions and approximations. For example, if the Hamiltonian is separable, explicitly containing the reaction coordinate and its conjugate momentum, a (rather bold) approximation to the generalized Langevin equation is the so-called *t-damping equation*, proposed by the authors of [202]. It reads

$$\begin{aligned} &\dot{\xi}(t) = \eta(t) \\ &\dot{\eta}(t) - \nabla G(\xi(t)) + t \, \gamma(\xi(t)) \cdot \eta(t) \,, \end{aligned}$$

and it is the formerly introduced optimal prediction equation with a Markovian friction term that increases with time. The symmetric and positive semi-definite matrix γ describes configuration-dependent friction, and G is the geometric free energy (which coincides with the standard free energy F in this particular case). An alternative equation, where t in the friction term is replaced by a constant characteristic time scale τ is suggested in [55]. However in either case the system is dissipative and the energy of the system quickly decays to zero (which is not true for the original system).

Systematic studies of the Mori-Zwanzig procedure are extremely rare; see, e.g. [238, 57, 58]. Even worse, they rely on rather restrictive assumptions (e.g., separable, quadratic Kac-Zwanzig Hamiltonian as in [60]) which considerably limits the usability of the Mori-Zwanzig procedure.

Modelling fast degrees of freedom: Fixman potential A basic insight of conformation dynamics is that once a reaction coordinate is well chosen, then the remaining degrees of freedom are fast and have comparably small amplitude. This leads to the idea to treat all unresolved variables as being harmonic, with a stiffness matrix which may depend on the reaction coordinate. Consequently, we replace the original molecular potential V by a *modelling potential*

$$V_\epsilon(x,y) = V_M(x) + \frac{1}{2\epsilon^2}\,\langle C(x)y, y\rangle \ ,$$

where M is the configuration manifold that is spanned by the reaction coordinate, (x,y) are local coordinates on the normal bundle NM, and C is a symmetric and positive-definite matrix. The particular form of the V_M is open to choice; for example, one can choose it as the restriction of the molecular potential to M. We have studied the singular limit $\epsilon \to 0$ of both the diffusion system or the Hamiltonian system, while keeping the total energy finite. In either case the model potential constrains the motion to the dominant subspace M giving pathwise convergence in most cases. The averaged drift in the limit system stems from the effective potential,

$$\bar{V}(x) = V_M(x) + (2\beta)^{-1} \ln\det C(x)\,.$$

The rightmost term is the *Fixman potential.* It pops up when taking the limit $\epsilon \to 0$, and it describes the influence of the coupling between the (fast) oscillations normal to M and the motion along M. Physically speaking, it accounts for the difference between a constrained system and a very stiff (but unconstrained) system. This connection has been established in [75] from the viewpoint of statistical mechanics; see also [28]. The equivalent problem in the microcanonical ensemble goes back to [180] and [239]. For a detailed discussion we refer to the textbook [240] or [98].

Furthermore the confinement mechanism provides a physical explanation of the Fixman Theorem and the Blue Moon formula. Imagine, the dominant subspace $M \subset \mathbf{R}^n$ is determined as the level set of some function $\varphi : \mathbf{R}^n \to \mathbf{R}^k$, i.e., $M = \varphi^{-1}(0)$. If we impose the constraint $\varphi(q) = 0$ by adding a strong potential,

$$V_\epsilon(q) = V_M(q) + \frac{1}{2\epsilon^2}\sum_{i=1}^{k}\left(\varphi_i(q)\right)^2 \ ,$$

then the corresponding limit potential for $\epsilon \to 0$ has the familiar form

$$\bar{V}(q) = V_M(q) + \beta^{-1} \ln \mathrm{vol} J_\varphi(q)\,.$$

Hence it turns out that the Fixman Theorem describes the difference between an ideal constraint, i.e., a configuration submanifold $M \subset \mathbf{R}^n$, and a penalty function

φ that is added to confine the system to the fibre $M = \varphi^{-1}(0)$. The analogous relationship between the invariant constrained and conditional probability measures has been exposed in the recent paper [17], where also a strong convergence proof for the confinement of diffusion processes is given. (The infinite energy scenario is discussed in [219] for mechanical systems and in [241, 242] for diffusion processes.)

The confinement method can be viewed as a simplification of the former reduction schemes that works for both stochastic differential equation models and (stochastic) Hamiltonian systems. Especially the limit potential can be interpreted as a free energy in a flat geometry, where the influence of the extrinsic geometry of M has vanished due to the finite energy scaling (see Section 3.4). Moreover the stiffness matrix can be freely chosen (modulo the condition that it be symmetric and positive-definite). Hence the modelling potential offers some flexibility in setting up a reduced model by means of parametrization. For alternative approaches that are built on fully parametrized reduced models we refer to the recent preprints [41, 39].