

Model reduction in classical molecular dynamics

PhD dissertation
Fachbereich Mathematik und Informatik
Freie Universität Berlin

Carsten Hartmann

Revised version, as of July 2007

Für meinen Vater

Contents

1	Introduction	1
2	Modelling molecular motion	10
2.1	Lagrangian and Hamiltonian mechanics	10
2.1.1	Statistical mechanics: ensemble concepts	15
2.2	Stochastic Langevin dynamics	18
2.3	High-friction dynamics: Brownian motion	22
2.4	Essential degrees of freedom	26
2.4.1	Spatial decomposition methods	26
2.4.2	Reaction coordinates	31
2.5	Problems related to symmetry and further generalizations	31
3	Eliminating fast degrees of freedom	34
3.1	Central paradigm in biophysics: free energy landscapes	35
3.1.1	Contributions to the free energy	38
3.1.2	Two distinct notions and the Fixman Theorem	42
3.2	The Averaging Principle	46
3.2.1	Averaging for linear reaction coordinates	49
3.2.2	Nonlinear reaction coordinate dynamics	51
3.3	Projection operator techniques	58
3.3.1	Optimal prediction and the Mori-Zwanzig formalism	60
3.3.2	The generalized Langevin equation	65
3.4	Modelling fast degrees of freedom: adiabatic perturbation theory	70
3.4.1	Resonances in molecular systems	80
3.4.2	Relations to geometric singular perturbation theory	90
3.5	Summary and bibliographical remarks	93
4	Phase space of the fast variables	99
4.1	Excursus: constrained mechanical systems	99
4.1.1	Geometric considerations	100
4.1.2	Constrained Hamiltonian systems	101
4.1.3	Statistical mechanics of constrained molecular systems	104
4.2	Sampling constrained invariant measures	105
4.2.1	Blue Moon sampling	105
4.2.2	Constrained hybrid Monte-Carlo	107
4.2.3	Langevin and Brownian motion	113
4.3	Thermodynamic Integration	117
4.3.1	Free energy from constrained Langevin motion	118
4.3.2	Free energy from constrained hybrid Monte-Carlo	120
5	Algorithmic issues and numerical examples	123
5.1	The constrained hybrid Monte-Carlo algorithm	123
5.2	Ryckaert-Bellemans <i>n</i> -butane	126
5.3	Glycine dipeptide in vacuum	138

6	Deviations from reduced models: correcting Brownian motion	147
6.1	Moderate deviations from the Averaging Principle	147
6.1.1	Central Limit Theorem: fluctuations from equilibrium	147
6.2	Large deviations from the Averaging Principle	149
6.2.1	Diffusive limits	150
6.2.2	Metastability and conditional averaging	152
7	Summary	155
8	Zusammenfassung (deutsch)	157
A	Different free energy concepts and Fixman potentials	158
B	Coordinate expressions	159
C	More coordinate expressions and the mean curvature vector	161
D	A co-area formula for Dirac's delta function	163
E	Three-scale problems	164
F	Van Kampen's approximation	167

1. Introduction

This thesis explores different routes to model reduction in the context of classical molecular dynamics. Adopting a reductionist’s point of view, our main concern is the sound causal explanation of observable macroscopic properties, e.g., activation energies or dynamical stability of conformations in terms of the microscopic physical model. Notwithstanding computational aspects, the difficulty lies in the sheer complexity of the microscopic models with their vastly different spatial and temporal scales.

Roughly speaking, reduced modelling comes in two varieties: elimination of specific (e.g., fast) degrees of freedom (also: *modes*) from the original model, or parametrization of certain simplified models. The first approach is usually referred to as *mode reduction*, whereas the latter is often termed *remodelling*. In this thesis we mainly focus on mode reduction, since sticking to the original microscopic model means to keep as much of the problem’s physics as possible. (The microscopic models are based on profound physical and chemical knowledge, both theoretical considerations and experimental data, which constitutes their empirical adequacy and predictive power.) In doing so, we extend and refine available methods of mode reduction such as averaging or projection operator techniques and set them in context with one another. Nevertheless we also allude to aspects of parametrized models.

Model reduction, as it is understood here, relies on the knowledge of a possibly multidimensional *reaction coordinate*. Of course the problem of finding good (i.e., physically meaningful) reaction coordinates is highly sensitive to boundary conditions such as temperature or pressure and hence cannot be addressed without referring to a specific situation. However for medium-sized molecules such as polypeptides or Lennard-Jones clusters there are often few natural candidates for good reaction coordinates, e.g., torsion angles or radii of gyration. Difficulties arise, if solvent effects play a role, for then the molecules’ configuration space has to be extended to incorporate the solvent which may lead to systems with varying number of particles. However we do not address the reaction coordinate problem here.

Central paradigm in reduced modelling: free energy Free energy is arguably one of the most important and prevalent concepts in molecular dynamics (see the reviews [1, 2]). According to the general view, free energy describes the tendency of a molecular system to associate and react. By definition free energy encodes statistical information about such activated processes, provided the reaction coordinate is suitably chosen. Hence statistical equilibrium properties such as conformational weights can be expressed in terms of free energy. It is less obvious that also many dynamical properties such as transition rates are related to a particular variant of free energy, as has been pointed out on various occasions, e.g., [3, 4, 5]; for the original works on transition state theory we refer to the papers of Eyring [6] and Wigner [7]. Interestingly enough, we find that this specific free energy also appears as an effective molecular potential in most of the reduced models, which casts it a fundamental dynamical concept. For reasons that will become clear below, we term this second type of free energy *geometric free energy*, whereas the first one (which reflects the equilibrium statistics) will be referred to as *standard free energy*.

Literature and previous developments The reader may believe that giving a complete overview of the relevant free energy literature is hopeless. During the last few years progress has been made towards algorithms that sample standard free energy

profiles from their derivatives using constrained simulations, exploiting the dichotomy of *free energy as the potential of mean force*; the free energy is recovered afterwards by numerical integration. This method is known as *Thermodynamic Integration* and goes back to Kirkwood [8]. The idea of relating the derivative of the free energy to the averaged force of constraint appears in the work of Mülders *et al.* [9] for the first time; however these authors derive a wrong expression emanating from a wrong definition of the conditional expectation. Correct expressions have been established in, e.g., [10, 11, 12, 13, 14]. Most of these authors omit the problem of sampling the respective conditional expectation which is crucial for actual computations; recent work in that direction is [15, 16, 17]. Later articles by Vanden-Eijnden and co-workers [13, 5] address the problem of relations among different the free energy definitions, geometric and standard free energy. Current work that exploits this relations for the development of efficient algorithms has been done by the author [18].

Mode elimination techniques for ordinary or stochastic differential equations with multiple scales have a long lasting tradition in celestial mechanics, especially in the Russian literature, e.g., [19, 20], but also in the climate modelling community; for instance, see the proceedings [21]. In celestial mechanics typical problems boil down to finding an appropriate set of action-angle variables [22], whereas climate problems are often described by stochastic differential equations with slow and fast variables [23]. In either case the reduced models are obtained upon averaging over the random perturbations induced by the fast degrees of freedom; the relevant reference regarding the Averaging Principle is the textbook by Freidlin and Wentzell [24]. (A good and systematic overview of the current multiscale literature can be found in the review by Givon *et al.* [25]). Instances of the just mentioned averaging methods are rare in molecular dynamics, however. One such case that is studied by Bornemann & Schütte [26, 27] or Reich [28, 29] is the elimination of fast bond vibrations by introducing holonomic constraints (rigid bond approximations). Another current example that is treated in Yanao *et al.* [30] is the dynamics of gyration radii as collective variables in Hamiltonian system; yet these authors pursue a purely deterministic approach which is more in the spirit of De Leon *et al.* [31] or Uzer *et al.* [32]. The only approach known to the author that addresses stochastic dynamics is a projection operator type method in E & Vanden-Eijnden [13]. For such systems it may happen that the averaged dynamics is trivial on the typical observation time scale, whereas relevant effects appear on longer time scales only. In this case (which, however, does not appear in the examples considered by us) averaging theorems on diverging time intervals come into play which were originally stated by Khas'minskii [33]; see also the recent article [34]. The related problem of large deviations from the averaged equations (for example, if the unresolved system contains essential barriers) is considered in, e.g., [35].

A significant part of the model reduction literature deals with simplified parametrized models, typically linear differential equations. Examples involve rigid base or rigid base-pair models in DNA modelling [36], diffusion models for protein folding [37], or stochastic differential equations that are coupled to Hidden Markov Models [38, 39]. Related work on nonlinear Langevin equations which is based on a Maximum-Likelihood principle is [40]. We abstain from presenting an exhaustive list of references and instead refer to the bibliography in [41]. For the sake of completeness we also mention the equation-free approach that has been developed in [42].

Quite often model reduction is also understood in the sense of clustering or state space decomposition. Most of these methods aim at classifying an essential subspace onto which the full molecular time series is projected. By this, one obtains a dimension-

reduced time series which is easier to analyze. Methods of this kind are known, e.g., by the name of Principal Component Analysis [43, 44] or Proper Orthogonal Decomposition [45]. Yet another promising approach that allows for the identification of essential subspaces that are dynamically relevant is the transfer operator approach by Schütte *et al.* [46]; see also the book of Weber [47]. Essential subspace techniques can be easily linked with methods of mode reduction. For example, a popular approach in the optimal control community (e.g., see [48]) is to truncate the modes orthogonal to the essential subspace, a method known by the name of *Galerkin projection*. Applications to molecular problems are not known to us though.

Another important part of the literature is concerned with projection operator techniques that have been established by Mori [49] and Zwanzig [50] in the context of non-equilibrium statistical mechanics. (See also the review by Hynes [51] or the monograph by Evan & Morriss [52].) In molecular applications such methods amount to the derivation of the (non-Markovian) *generalized Langevin equation*. Examples can be found in [53, 54]; regarding a systematic study of the projection operator ansatz for molecular problems we refer to [55]. Contemporary mathematical works concern closure schemes for the generalized Langevin equation [56], Markovian approximations [57], or issues related to existence and uniqueness [58]. A good overview can be found in the new textbook by Chorin & Hald [59]. Related problems such as applications to Kac-Zwanzig heat bath models are discussed in [60]. A Markovian variant of the projection operator approach, which can be regarded as least-square approximation in some suitably defined function space, is called *optimal prediction*. Typically optimal prediction is applied to problems involving partial differential equations, for which the method works quite well (e.g., Burgers' equation [61], Korteweg-deVries-Burgers equation [62]). However the application to deterministic Hamiltonian systems yields rather poor results as has been repeatedly demonstrated, for example, by Hald & Kupferman [63] or Chorin *et al.* [64], and instances of stochastic Hamiltonian systems are not known to us. The only molecular dynamics application we are aware of is in the article of Seibold [65]; however therein the author mainly focuses on low temperature asymptotics and aspects of computational efficiency.

Issues addressed in this thesis This thesis deals with very different aspects of model reduction. The original models range from deterministic mechanical models on the one hand to stochastic differential equations such as Brownian motion or Langevin dynamics on the other hand. Each of these models comes along with its own formalism (covariant formulations in mechanics, Itô calculus for stochastic differential equations, etc.) which makes it difficult to handle all problems within a unifying framework. Moreover many problems in molecular dynamics are of genuinely thermodynamical nature which calls for an appropriate mathematical description of statistical concepts such as free energy. Here we adopt a more geometric language that is common to classical mechanics on manifolds. This may seem unusual, especially for readers that are familiar with stochastic differential equations. But in fact, many problems in molecular dynamics are problems on manifolds that have an interesting underlying geometric structure: constrained dynamics on a configuration submanifold, curvilinear reaction coordinates, and many more. Furthermore the covariant formalism of mechanics allows for straightforward generalization of statistical mechanics problems to curved spaces which is suitable, e.g., for sampling certain probability measures subject to holonomic constraints (cf. Section 4.2).

The general mathematical framework is established in Section 2. We basically

follow the relevant literature on geometric mechanics by Abraham & Marsden [66] and review ideas for Hamiltonian systems with randomized momenta that have been put forward in the work of Schütte [67]. It turns out that the covariant formalism of mechanics easily extends to stochastic differential equations which leads, among others, to a geometric version of Itô’s formula which proves that the Itô stochastic differential equation transforms like a second-order vector field (Lemma 2.11). The basic language of *stochastic differential geometry*, on which our considerations are based, is developed in the books by Stroock [68] and Hsu [69]. By expanding the ideas therein to Langevin processes, we reveal that the Langevin equation has some interesting transformation properties as compared to general hypo-elliptic diffusion processes which are due to its Hamiltonian origin (Lemma 2.10). In particular we find that the Itô-Stratonovich ambiguity vanishes, if we confine our attention to point transformations. The geometric viewpoint of molecular dynamics thus highlights that such different systems as second-order mechanical systems, first-order Brownian motion and stochastic Langevin equations exhibit common transformation properties. Additionally it gives rise to a physically intuitive and unifying picture for what is called *entropic effects* in conformation dynamics: in case of second-order mechanical systems, these stem from inertial contributions due to the kinetic energy, but they can be likewise explained by the interplay between ordinary diffusion and certain conformational degrees of freedom. In either case these effects are actuated by the underlying Riemannian structure (see Section 5 for some examples).

Section 3.1 tries to shed some light on the different free energy definitions that circulate in the literature; cf. the review [13], and see also the schematic overview in Appendix A. On a purely formal level, Federer’s co-area formula [70] links standard and geometric free energy by relating the underlying conditional probability densities. From a physical point of view, the standard free energy can be expressed as a sum of geometric free energy and an appropriately defined Fixman potential. Neither relation is actually new, but they both have useful practical implications for Thermodynamic Integration algorithms that, to the best of our knowledge, have not been taken into account so far: First of all, we explicate that the famous Blue Moon formula by Carter *et al.* [71] for the conditional expectation is an instance of the co-area formula (Section 3.1.2). The Blue Moon ensemble method is a popular and widely-used technique for the sampling of conditional expectations by means of constrained simulations with an appropriate reweighting strategy. Yet there has been (and still is) some confusion (e.g., see [16, 72, 73]) about whether the weight is affected by the presence of momenta or velocities in the system. But as we will argue below, reweighting is an issue for any type of constrained dynamics — no matter if the system involves momenta or not. Secondly, we demonstrate that geometric free energy can be viewed as the *potential of mean constraint force* (see Section 3.1.1 or Section 4.1 regarding holonomic constraints). By using Thermodynamic Integration it is hence possible to compute the derivative of the geometric free energy by simply averaging over the Lagrange multipliers (forces of constraint) that are explicitly available during the simulations without further function evaluations; cf. also the recent article [17]. Last but not least, the Fixman potential that marks the difference between geometric and standard free energy can be directly computed from constrained simulations without computing second derivatives of the reaction coordinate. This yields a remarkably simple formula for computing standard free energies that does neither require second derivatives nor reweighting à la Blue Moon (see Remark 4.14 or the recent article [18] by the author). To the best of the author’s knowledge all available algorithms do in fact require the calculation of the

reaction coordinate's Hessian; e.g., see [10, 11, 12, 16]. The various contributions to geometric and standard free energy have concise geometric and physical interpretations as is worked out in detail in Section 3.1.1. In Section 3.1.2 we also give an answer to the question in which sense free energy can be understood as a *potential of mean force*: if one takes up the position that force is understood as a differential 1-form, then only the derivative of the geometric free energy qualifies as a (mean) force (viz., the mean constraint force), while the derivative of the standard free energy exhibits additional gauge dependencies with regard to transformations of the reaction coordinate.

Section 3.2 is dedicated to the application of the Averaging Principle to diffusion models, i.e., stochastic differential equations with non-degenerate noise term. If the reaction coordinate is linear in the configuration variables, then the reduced equations describe simple diffusion in the free energy landscape; in this particular case, standard and geometric free energy coincide. If the reaction coordinate is nonlinear, the application of the Averaging Principle requires that we write the equation globally in terms of the resolved and the unresolved coordinates. Not only is this difficult (or even impossible), but it also makes the resulting equations in some respects intransparent. Therefore we take advantage of the fact that the reaction coordinate foliates configuration space and consider only local averages of the dynamics on each leaf, where each leaf is defined by a specific value of the reaction coordinate. The global picture can then be reconstructed by endowing the locally averaged equations with an appropriate Riemannian metric that is induced by the reaction coordinate and is defined for all of its possible values (Section 3.2.2). Although the just described approach is no longer covered by the Averaging Principle, it turns out that the reduced system is again an Itô equation which is covariant under transformations of the reaction coordinate and has a straightforward physical interpretation as a diffusion equation on a Riemannian manifold. In particular the effective potential energy is given by the geometric free energy. We briefly illustrate the method by suitable examples (see Examples 3.10 and 3.11) and discuss its relation to a related approach [13].

For mechanical systems the situation is more complicated, since the equations are essentially second-order, and thus the requirement that the fast dynamics exhibits a unique invariant measure for all values of the slow coordinates is difficult to handle analytically and numerically (see Example 3.12). Therefore we resort to projection operator techniques or least-square approximations such as optimal prediction [64]. Though similar to standard averaging, these methods account for the fact that the equations are second-order. As Chorin & Hald have proved in [56], optimal prediction for Hamiltonian systems leads to reduced models that are again Hamiltonian. Based on considerations therein, we derive a new and simple expression (3.63) for the effective total energy that allows for a lucid interpretation as a mechanical system on a Riemannian manifold which is spanned by the reaction coordinate. As in the Brownian dynamics case, we demonstrate that the effective potential is given by the geometric free energy. More sophisticated projection operator techniques like the Mori-Zwanzig procedure involve the derivation of a *generalized Langevin equation*, which is a suggestive way to rewrite Hamilton's equations as a Langevin-like equation that is formally equivalent [49, 50]. However we emphasize that the equivalence is only formal, for the derivation relies on the tacit (but wrong) assumption that the Hamiltonian system has a unique invariant measure. Although theoretically appealing, the ideas of Mori and Zwanzig have barely any practical relevance for studying complicated molecular processes; here we mention it only for the sake of completeness, while pointing out certain difficulties in connexion with the derivation of the generalized

Langevin equation (see Remark 3.17 and the preceding paragraph).

Finally in Section 3.4, we propose an *ad-hoc* alternative to averaging and optimal prediction that is based on the observation that the unresolved degrees of freedom often have small amplitude and can thus be approximated by harmonic motions. By averaging over these modes one ends up with semi-analytic reduced models (both diffusive and mechanical) that live only on the essential subspace, but still have few free parameters. Since the parametrization involves only the unresolved parts of a molecule, these models are easy to compute at the cost of restricted physical interpretability. Two interesting aspects emerge in connexion with the semi-analytic models: First, the reduced models are relevant in the context of stiff-bond approximations, since they explain how the dynamics is altered, when infinitely stiff bonds are replaced by rigid bonds (i.e., constraints). Work in this direction has been done by Hinch [74] for diffusive systems and by van Kampen & Lodder [75] and Reich [28] for mechanical systems. Although we do not contribute new results regarding rigid-bond approximations, knowledge about practical implications thereof are still not widespread in the molecular dynamics community (e.g., see [76]). Second, there are some interesting relations to adiabatic perturbation problems in mechanics [27]. For instance, it is well-known that averaging problems for small oscillations may suffer from resonances between the oscillators' frequencies [20]. It seems, however, that resonances do not play a role, if the system is appropriately thermalized. Since investigating resonance effects in stochastic Hamiltonian systems or Brownian motion in detail is far beyond the scope of this thesis, we provide only numerical evidence for this claim in Section 3.4.1. However more careful studies would be desirable.

To some extent reduced modelling results in the calculation of (geometric) free energy profiles. In fact there is a bunch of literature that addresses standard free energy calculation by means of Thermodynamic Integration, e.g., [71, 77]. However Thermodynamic Integration proceeds by constrained integration, and it is by no means clear how standard thermostating techniques fit constrained integration. It is striking that the question of how to sample the correct probability measure (constrained Gibbs measure) is typically ignored; e.g., see [10, 11, 78]. In particular there is a lot of confusion in the literature whether constrained Hamiltonian systems inherit fundamental thermodynamical properties from their unconstrained counterparts. For example, it is common sense in the molecular dynamics community that constrained Hamiltonian flows do not preserve phase space volume; e.g., see [73, 72, 79, 80]. Additionally there is an ongoing discussion [16] concerning the impact of so-called *hidden constraints* on the invariant distribution of constrained second-order systems and its relation to first-order systems. (Again this remark alludes to the co-area formula and the problem of Blue Moon reweighting for first-order systems.) We provide the theoretical background regarding constrained mechanical systems in Section 4.1. To this end, we basically review available results from the literature [66, 81]; in particular we adopt an argument in [82] that proves that constrained Hamiltonian systems are symplectic and therefore volume-preserving (Lemma 4.3). Taking advantage of this property, we then construct a novel hybrid Monte-Carlo (HMC) scheme that can be used together with the RATTLE symplectic integrator for constrained Hamiltonian systems. Following an idea in [83], we can prove that the corresponding discrete Markov chain is ergodic with respect to the constrained Gibbs measure on configuration space (Proposition 4.12). Related results for Brownian motion have recently become available in the work of Lelièvre *et al.* [17]. Therein, however, the authors prove ergodicity only for the time-continuous process, while

disregarding discretization issues. In Section 4.2.3 we generalize their results and construct a constrained Langevin dynamics, equation (4.27), that conserves a given holonomic constraint and that preserves the constrained Gibbs measure. (For this purpose we once more take advantage of the similarity between Hamilton and Langevin dynamics, borrowing ideas from index reduction techniques for differential-algebraic equations.) Furthermore we suggest a discretization scheme, equations (4.29)–(4.32), that can be regarded as a stochastic modification of the RATTLE algorithm. We should mention that an almost identical algorithm has been published by Vanden-Eijnden & Ciccotti [84] during the course of this thesis, where the authors could even prove that the algorithm is second-order accurate. However the article does not address issues of invariant measures and constrained probability distributions. We conclude with Section 4.3 by discussing the application of the various sampling schemes to the calculation of free energy profiles. We especially propose a novel algorithm that does neither require Blue Moon reweighting of the expectation values nor calculating second derivatives of the reaction coordinate (see Remark 4.14).

We illustrate the reduction schemes as well as the constrained hybrid Monte-Carlo sampling by means of several examples in Section 5. Both averaged Brownian motion and optimal prediction perform remarkably well in terms of dynamical observables such as transition rates or decay of correlations. (Especially for the latter approach this comes rather unexpected as optimal prediction for deterministic Hamiltonian system is known to yield fairly poor results; e.g., see Chorin *et al.* [64].) Moreover optimal prediction reveals an interesting (and yet unknown) physical mechanism that explains the backbone dynamics of a chain-like molecule: for *n*-butane, for example, we observe that the angular kinetic energy favours the *trans* conformation, which is characterized by a rather slim shape with respect to the principal axis of inertia and which should be contrasted with the bulky *cis* conformations. *Prima facie* this seems counter-intuitive, since one could expect that the mass distribution of a rotating molecule tends to spread out due to centrifugal forces. However here the situation is different, for the backbone rotation is an internal motion of the molecule. Since the kinetic energy tends to stabilize the more compact *trans* conformation by slightly increasing the total energy of the *cis* conformations, we term the induced force *internal centripetal force*. The same rotation mechanism explains the different conformational stabilities of the glycine dipeptide analogue, for which we study free energy landscapes and the optimal prediction Hamiltonian along the two central backbone angles: also here the kinetic energy stabilizes the extended C5 conformations by slightly lowering their total energy as compared to the bulky C7 conformations. Our calculations also reveal that the kinetic energy preserves the molecular potential’s symmetry under parity transformations in the Ramachandran plane, but exhibits an even higher symmetry itself: the matrix elements of the effective inverse metric are (approximately) invariant under reflections of the two backbone angles independently, where the slight perturbation of the symmetry reflects the non-uniform mass distribution along the peptide’s backbone. To the best of the author’s knowledge symmetry-breaking effects of the peptide backbones’ mass-distribution have not been studied so far, and more careful studies would be desirable.

Finally, we survey known results from the literature that deal with corrections to the Averaging Principle for non-degenerate stochastic differential equations. Problems involve moderate [85, 86] and large deviations [24, 35] or deviations on long time scales [33, 34]. We mention them for the sake of completeness and encourage their application to molecular dynamics problems in the future.

Summary of the main achievements

- The transformation properties of mechanical systems, non-degenerate diffusion equations (Brownian motion) and hypo-elliptic Langevin equations are studied. Restricting our attention to point transformations, it is a common feature of all such systems that they transform as second-order vector fields. In case of the diffusion equation, this property reflects the well-known Itô formula. For Langevin equations the Itô-Stratonovich ambiguity vanishes.
- The various contributions to standard and geometric free energy have concise geometric and physical interpretations. In particular we reveal that the geometric free energy is the potential of mean constraint force. As a by-product we provide an alternative version of the famous Blue Moon ensemble method that turns out to be an instance of Federer’s co-area formula. It therefore also applies to constrained first-order systems such as Brownian motion.
- For both averaging and optimal prediction we find that the geometric free energy appears in the reduced equations as an effective potential, which casts it a fundamental dynamical quantity. For all practical purposes the optimal prediction Hamiltonian can be approximated by a sum of kinetic and potential energy (i.e., geometric free energy), where the kinetic energy is defined with respect to an averaged Riemannian metric that is induced by the reaction coordinate.
- Hybrid Monte-Carlo (HMC) for constrained mechanical systems is a novel algorithm for sampling constrained Gibbs measures and free energy profiles. We prove a Law of Large Numbers for the time-discrete HMC Markov chain that holds for any stable step-size. Exploiting the close relationship between Hamilton and Langevin equations, we derive a constrained version of the Langevin dynamics that preserves the constrained canonical distribution and allows for calculating free energy profiles without reweighting or computing second derivatives.
- The performance of the different reduction schemes is demonstrated by means of two molecular examples: *n*-butane and the glycine dipeptide analogue. The reduced systems reproduce essential dynamical observables such as correlations or transition probabilities. Even more important, the models reveal a common rotation mechanism for the molecules’ conformational dynamics that can be explained by the interplay between geometric free energy and the (extrinsic) geometry of the reaction coordinate.

Some matter of notation We will make extended use of abstract index notations. Often we will use lowercase Greek indices α, β, γ and Latin indices i, j, k to distinguish between different types of coordinates (e.g., resolved and unresolved coordinates) with respect to an unspecified basis. Note, however, that this distinction is sometimes relaxed; then we use Latin indices h, l, m, n to label arbitrary coordinates. Moreover we use Einstein’s summation convention, that is, we sum over double upper and lower indices, where the range of the respective indices should be clear from the context. Using a particular coordinate system, for example, polar coordinates (r, φ, ϑ) we may also write g^{12} or $g^{r\varphi}$ to denote the $(1, 2)$ component of a contravariant tensor g (e.g., the inverse metric tensor). The reader should be aware of some other abusive notations that are common in the physical literature. For example, we will use the dot (time derivative) to denote tangent vectors. That is, if $q = (q^1, \dots, q^n)$ denotes coordinates on a configuration manifold Q , then we write (\dot{q}, \dot{q}) to denote the respective coordinates

on the tangent space TQ . We employ so-called mass-scaled coordinates which allows us to set the molecular mass to unity throughout this thesis. This is convenient, for it considerably simplifies the notation, and it allows us to identify tangent and cotangent space in the sense that $\dot{q} = p$. In doing so the velocity vector \dot{q} on the left is an element of the tangent space T_qQ , whereas the momentum on right hand side is from the dual space T_q^*Q . In general this identification will be procured by the metric tensor g on Q , but in the Euclidean case we will exploit this identification without further comment.

The various types of differential operators appearing in the text may seem a little confusing. Often it is important to distinguish between derivatives that are denoted by df or ∇f , and the gradient of a function f that is written as $\text{grad } f$ (the latter is a vector field, whereas the former denotes a 1-form). For vector fields X, Y we will sometimes use the symbols $\nabla_X Y$ for the covariant derivative between vector fields, or $dY(X) = \nabla Y \cdot X$ to denote the directional derivative of a vector field along a vector. Moreover the bold face symbol $\mathbf{D}f$ means the Jacobian of a vector-valued function f . We may also use the notation $\mathbf{D}_1 f(\cdot, \cdot)$ or $\mathbf{D}_2 f(\cdot, \cdot)$ to indicate derivatives with respect to the first or second slot (e.g., slow and fast coordinates) of a function.

Acknowledgments It is my pleasure to thank all those who supported and encouraged this thesis: First of all, my advisor, Christof Schütte, for directing my dissertation research, pointing out many interesting mathematical questions, giving constant encouragement and steady support. Eric Vanden-Eijnden, for providing many insightful remarks about free energy concepts that were invaluable for this work and for inviting me to the Courant Institute and UC Berkeley. Special thanks goes to Tony Lelièvre, Pete Kramer and John Maddocks for fruitful discussions on problems with constraints, and for persistently asking further questions. I also enjoyed the vibrant criticism by Giovanni Ciccotti (not to mention the excellent dinners).

I have appreciated the stimulating interdisciplinary research atmosphere at Freie Universität Berlin. In particular I very much thank my colleagues: Wilhelm Huisinga for introducing me to molecular dynamics, Jessika Walter and Burkhard Schmidt for stimulating discussions about my work (and their work, too), Tobias Jahnke, Oliver Sander and Heiko Berninger for any kind of sideline mathematics, and Illia Horenko for regularly joining me at the cafeteria. I am moreover grateful to Tim Conrad, Alexander Fischer and Eike Meerbach for helping me with the computer simulations.

Finally, this thesis would not be on hand without those persons, my family and friends, who accompanied me during the last few years. I am deeply indebted to my parents for backing me all the years, granting unrestricted freedom to follow my interests, and to Gebke for just being there.

The work was funded by the DFG Research Center MATHEON "Mathematics for Key Technologies" (FZT86) in Berlin.

2. Modelling molecular motion

2.1. Lagrangian and Hamiltonian mechanics

In classical molecular dynamics, a molecule is mostly modelled as a natural mechanical system [87, 88]. The two basic formulations are those of Lagrange and Hamilton: From a systematic point of view, Lagrangian mechanics seems more fundamental as it is based on a variational principle that directly leads to a coordinate-invariant formulation of mechanics. Intriguingly the *least action principle* was discovered by Hamilton itself decades after Lagrange stated the Euler-Lagrange equations, and which is thus also termed Hamilton’s principle [89, 90]. However the equivalent Hamiltonian formulation which should be distinguished from the variational principle, is better suited to establish statistical thermodynamics from mechanical principles or to point out the fundamental relationship with quantum mechanics. The related ensemble concepts are crucial ingredients of molecular models. We essentially follow the outline in the textbook [66]; see also [81] for some historical remarks.

Atomic interaction potential The key ingredient to all classical molecular models is the atomic interaction potential that contains all physically relevant interactions, where the word *physically* is to be understood within the scope of classical molecular dynamics. Here we consider only potentials that do not involve any velocity interactions and that are at least twice continuously differentiable.

Typically the molecular potential is modelled as a sum of contributions of different physical origin. For the sake of convenience we endeavour the multi-body notation with the bold face symbol $\mathbf{q}_i \in \mathbf{R}^3$, $i = 1, \dots, N$ for the position of the i -th atom. Roughly speaking, the molecular interaction can be divided into two parts: *local* interactions as induced by the bond structure of the molecule, and *long-range* interactions like electrostatic interactions, for instance. One rather simple but widely-used example is the Ryckaert-Bellemans potential for modelling alkane chains [91]. The local interactions are

- harmonic stretching of the covalent bond between the atoms i and $i + 1$ as described by $V_{\text{bd}}(\mathbf{q}_i, \mathbf{q}_{i+1}) \propto (r_i - r_{\text{eq}})^2$ with $r_i = \|\mathbf{q}_{i+1} - \mathbf{q}_i\|$,
- harmonic vibrations of the bond angle formed by the covalent bonds between three successive atoms which are modelled by $V_{\text{ba}}(\mathbf{q}_{i-1}, \mathbf{q}_i, \mathbf{q}_{i+1}) \propto (\psi_i - \psi_{\text{eq}})^2$ with the bond angle $\psi_i = \sphericalangle(\mathbf{q}_{i-1} - \mathbf{q}_i, \mathbf{q}_{i+1} - \mathbf{q}_i)$,
- motion of the torsion angle ω_i between two planes each of which is spanned by three atoms; this potential has a multi-well structure (see Figure 1), and it depends upon positions of four successive atoms $V_{\text{ts}} = V_{\text{ts}}(\mathbf{q}_{i-1}, \mathbf{q}_i, \mathbf{q}_{i+1}, \mathbf{q}_{i+2})$.

Local interactions that involve more than four atoms are not captured by this model. The corresponding non-bonded interactions are

- electrostatic interaction from the charges of the atoms j and k , that are described by a Coulomb potential $V_{\text{C}}(\mathbf{q}_j, \mathbf{q}_k) \propto 1/d_{jk}$ with $d_{jk} = \|\mathbf{q}_j - \mathbf{q}_k\|$,
- van der Waals interactions between polarizable atoms that are modelled by a Lennard-Jones potential $V_{\text{LJ}}(\mathbf{q}_j, \mathbf{q}_k) \propto 1/d_{jk}^{12} - 1/d_{jk}^6$ that contains both short-range and long-range interactions.

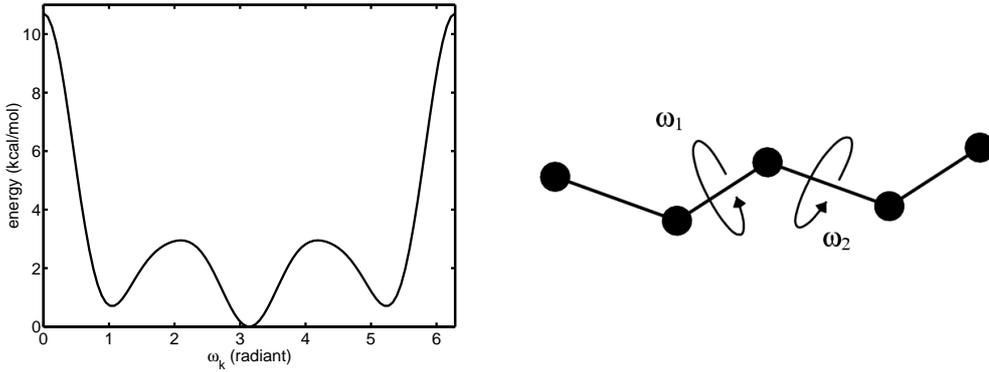


Figure 1. Ryckaert-Bellemans torsion angle potential as a function of the torsion angle ω_k ; see [91] for the details. The right panel shows the only two proper torsion angles (dihedral angles) of the *united atoms* pentane molecule.

This results in the following molecular potential

$$\begin{aligned}
 V(\mathbf{q}) = & \sum_i V_{\text{bd}}(\mathbf{q}_i, \mathbf{q}_{i+1}) + \sum_j V_{\text{ba}}(\mathbf{q}_{j-1}, \mathbf{q}_j, \mathbf{q}_{j+1}) \\
 & + \sum_k V_{\text{ts}}(\mathbf{q}_{k-1}, \mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{q}_{k+2}) \\
 & + \sum_{k,l} V_{\text{C}}(\mathbf{q}_k, \mathbf{q}_l) + \sum_{k,l} V_{\text{LJ}}(\mathbf{q}_k, \mathbf{q}_l),
 \end{aligned}$$

where $\mathbf{q} = (\mathbf{q}_1, \dots, \mathbf{q}_N)^T$ and the sums run over all the atoms that contribute to the respective interaction. Other possibly relevant interactions that are not directly contained in the potential are accounted for by adjusting the parameters in the potential in an appropriate way. In this sense the molecular potential is semi-empirical, and in fact using different potentials or *force fields* may produce vastly different results. See [92, 44] for the general force field methodology and [93] for a recent comparison of different force fields for the conformation dynamics of trialanine in water.

Lagrangian mechanics Consider a molecule with configuration space $Q \subseteq \mathbf{R}^n$ and molecular configurations $q = (q^1, \dots, q^n)^T$. Realistic molecular systems typically involve a large number N of atoms, therefore the spatial dimension $n = 3N$ is large. We denote by $\dot{q} = dq/dt$ the corresponding velocities, and by $TQ \cong Q \times \mathbf{R}^n$ the tangential bundle over the configuration space Q .

Let further $V : Q \rightarrow \mathbf{R}$ be a smooth molecular interaction potential. Throughout this thesis we shall assume that any of the two conditions is met: Either the system is *bounded* in the sense that $V \rightarrow \infty$ as $\|q\| \rightarrow \infty$, i.e., the configuration space is unbounded, but the potential is such that a particle cannot escape to infinity. Or the system is *periodic* in the sense that $Q \cong \mathbf{T}^n$. That is, Q is isomorphic to a flat n -dimensional torus. The latter is typically assumed for periodic boundary conditions or to bound the potential energy. In any case the Lagrange function $L : TQ \rightarrow \mathbf{R}$ is of the form *kinetic minus potential energy*,

$$L(q, \dot{q}) = \frac{1}{2} \langle M\dot{q}, \dot{q} \rangle - V(q), \quad (2.1)$$

where $M \in \mathbf{R}^{n \times n}$ is the diagonal, and positive-definite mass matrix, and $\langle \cdot, \cdot \rangle$ is the standard inner product between (tangent) vectors in \mathbf{R}^n . Consider a curve $q(t)$ in Q with $t \in [a, b]$ and fixed endpoints $q(a) = q_a$, $q(b) = q_b$. Hamilton's principle states that if a curve $q(t)$ minimizes the action integral

$$\int_a^b L(q(t), \dot{q}(t)) dt, \quad (2.2)$$

then it is a solution of the corresponding Euler-Lagrange equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0, \quad i = 1, \dots, n. \quad (2.3)$$

For a multi-body system in Cartesian space, the Euler-Lagrange equations reduce to the familiar Newton equations: Let again $\mathbf{q}_i \in \mathbf{R}^3$ denote the position vector of the i -th atom in Cartesian space $Q = \mathbf{R}^{3N}$, and let m_i be the corresponding atomic mass, such that $M = \text{diag}(m_1, m_1, m_1, \dots, m_N, m_N, m_N)$. Then (2.3) becomes

$$m_i \ddot{\mathbf{q}}_i = -\frac{\partial V}{\partial \mathbf{q}_i}, \quad i = 1, \dots, N.$$

A convenient property of the Euler-Lagrange equations (2.3) is that they can be easily put into a form which is invariant under coordinate transforms (covariant); clearly Newton's equations above are not covariant. The covariant formulation is particularly important, e.g., if a molecule is rigid or partially rigid [36].

Assume that Q is a n -dimensional Riemannian or pseudo-Riemannian manifold which is endowed with a coordinate-dependent metric $g(q) : T_q Q \times T_q Q \rightarrow \mathbf{R}$. Here $T_q Q$ is the tangent plane attached to $q \in Q$ which we identify with \mathbf{R}^n as usual. Alternatively we will denote the metric by $\langle \cdot, \cdot \rangle_g = \langle g(q) \cdot, \cdot \rangle$ and write g^b (or just g) for the associated linear map: the *metric tensor* with elements g_{ij} . The elements of the respective inverse map g^\sharp are labeled g^{ij} .

The proof of the basic statement below is standard and can be found in various textbooks on geometric mechanics [66]. We give it for the sake of illustration.

Lemma 2.1. *Let $L : TQ \rightarrow \mathbf{R}$ be a Lagrange function. Then in coordinates $(q, \dot{q}) \in TQ$ the Euler-Lagrange equations (2.3) have the covariant form*

$$\ddot{q}^i + \Gamma_{jk}^i \dot{q}^j \dot{q}^k + g^{ij} \frac{\partial V}{\partial q^j} = 0, \quad (2.4)$$

where

$$\Gamma_{jk}^i = \frac{1}{2} g^{il} \left(\frac{\partial g_{jl}}{\partial q^k} + \frac{\partial g_{kl}}{\partial q^j} - \frac{\partial g_{jk}}{\partial q^l} \right) \quad (2.5)$$

are the symmetric Christoffel symbols on TQ , and we have set in force the summation convention which means that we sum over double upper and lower indices.

Proof. Consider the Lagrange function on TQ

$$L(q, \dot{q}) = \frac{1}{2} g_{ij} \dot{q}^i \dot{q}^j - V(q),$$

which gives rise to the standard Euler-Lagrange equations (2.3), that now become

$$\frac{d}{dt} (g_{ij} \dot{q}^j) - \frac{1}{2} \frac{\partial g_{jk}}{\partial q^i} \dot{q}^j \dot{q}^k + \frac{\partial V}{\partial q^i} = 0.$$

Hence

$$g_{ij}\ddot{q}^j + \left(\frac{\partial g_{ij}}{\partial q^k} - \frac{1}{2} \frac{\partial g_{jk}}{\partial q^i} \right) \dot{q}^j \dot{q}^k + \frac{\partial V}{\partial q^i} = 0,$$

which upon multiplying by the inverse metric tensor g^\sharp and rearranging indices gives

$$\ddot{q}^i + \hat{\Gamma}_{jk}^i \dot{q}^j \dot{q}^k + g^{il} \frac{\partial V}{\partial q^l} = 0$$

with the non-symmetric Christoffel symbols

$$\hat{\Gamma}_{jk}^i = g^{il} \left(\frac{\partial g_{jl}}{\partial q^k} - \frac{1}{2} \frac{\partial g_{jk}}{\partial q^l} \right).$$

Taking into account the symmetry of the summation indices j and k in the Euler-Lagrange equations above, the quadratic expression $\hat{\Gamma}_{jk}^i \dot{q}^j \dot{q}^k$ can be rewritten as

$$g^{il} \left(\frac{\partial g_{jl}}{\partial q^k} - \frac{1}{2} \frac{\partial g_{jk}}{\partial q^l} \right) \dot{q}^j \dot{q}^k = \frac{1}{2} g^{il} \left(\frac{\partial g_{jl}}{\partial q^k} + \frac{\partial g_{kl}}{\partial q^j} - \frac{\partial g_{jk}}{\partial q^l} \right) \dot{q}^j \dot{q}^k$$

where the right hand side of the equation contains the symmetric Christoffel symbols Γ_{jk}^i just defined. From this the assertion follows. \square

Remark 2.2. Sometimes it is required to write the Euler-Lagrange equations in coordinate-free form. This is easily done by means of the lemma above using the language of covariant derivatives which is common in differential geometry [94, 95]. If $c(t)$ is a curve, and X is a vector field, the covariant derivative $\nabla_{\dot{c}(t)} X$ is declared as the directional derivative of X along $c(t)$ which reads

$$(\nabla_{\dot{c}(t)} X)^i = \frac{d}{dt} X^i(c(t)) + \Gamma_{jk}^i(c(t)) X^j(c(t)) \dot{c}^k(t).$$

See [66] for the details. If $c(t)$ is an integral curve of X , and X is the Lagrangian vector field, as defined by the Euler-Lagrange equations (2.4), we obtain the following coordinate-free representation

$$\nabla_{\dot{c}(t)} \dot{c}(t) + \text{grad} V(c(t)) = 0,$$

of the Euler-Lagrange equations, where the gradient of V is defined in coordinates as $(\text{grad} V)^i = g^{ij} \partial V / \partial q^j$. In fact, if we multiply the last equation by the metric tensor g (generalized mass matrix) we obtain Newton's second axiom, $F = ma$. For a modern survey of Riemannian geometry and related applications see [96].

Hamiltonian mechanics Given a molecular Lagrangian (2.1) we can proceed to the Hamiltonian formalism by introducing the conjugate momentum variable

$$p_i = \frac{\partial L}{\partial \dot{q}^i}, \quad i = 1, \dots, n. \quad (2.6)$$

We call the coordinate pair (q, p) a set of *conjugate variables*. We assume that the transformation $(q, \dot{q}) \mapsto (q, p)$ is invertible which requires the partial Hessian matrix $(\partial^2 L / \partial \dot{q}^i \partial \dot{q}^j)$ to be non-singular; so far this is the mass matrix M or the metric tensor g , respectively. We introduce the Hamiltonian as the Legendre transform of L

$$H(q, p) = \dot{q}^i p_i - L(q, \dot{q}),$$

where we have to substitute the velocities by the respective momentum variables. Keep in mind that the sum is taken over the double upper and lower index i . Once the Hamiltonian is defined we directly obtain Hamilton's equations of motion

$$\begin{aligned}\dot{q}^i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q^i}, \quad i = 1, \dots, n.\end{aligned}\tag{2.7}$$

It is a straight consequence of the above definition that

$$\frac{\partial H}{\partial q^i} = -\frac{\partial L}{\partial q^i}\tag{2.8}$$

which together with (2.6) shows the equivalence of Hamilton's equations and the Euler-Lagrange equations. Furthermore (2.6) assigns the phase space T^*Q with coordinates (q, p) to the configuration space Q with coordinates q ; the bundle $T^*Q \cong Q \times \mathbf{R}^n$ is called the cotangent bundle over Q . The mapping from the tangent space TQ to the cotangent space T^*Q is provided by the linear map g^\flat ,

$$g^\flat : T_q Q \rightarrow T_q^* Q, \quad p_i = g_{ij} \dot{q}^j, \quad i = 1, \dots, n$$

which is a slightly more sophisticated way to express the well-known relationship $p = M\dot{q}$ between Cartesian velocities and momenta. By the regularity assumption on the Lagrangian the matrix of g^\flat is invertible, so g^\sharp exists and takes T^*Q back to TQ . On the cotangent space the Hamiltonian has the generic form

$$H(q, p) = \frac{1}{2} g^{ij} p_i p_j + V(q)$$

and generates Hamilton's equations (2.7). A coordinate-free representation of Hamilton's equations can be obtained using Poisson brackets; see [81] for details.

Hamiltonian flows have some important conservation and symmetry properties: Let $\Psi_t : T^*Q \rightarrow T^*Q$ denote the one-parameter group of diffeomorphisms that is generated by the Hamiltonian. That is, $(q(t), p(t)) = \Psi_t(q, p)$ is the solution of the initial value problem (2.7) with $(q(0), p(0)) = (q, p)$.

First of all, the flow conserves the total energy, $H = H \circ \Psi_t$. Secondly, it is reversible in the sense that $\Psi_t(q, p) = \Psi_{-t}(q, -p)$. Last but not least, it conserves phase space volume, i.e., $\det \mathbf{D}\Psi_t = 1$. The last feature follows directly from the fact that the vector field (2.7) is divergence-free; an even stronger attribute is that Hamiltonian flows are *symplectic* which will be explained in the following.

Definition 2.3. *Let M be an even-dimensional smooth manifold. A symplectic structure on M is a closed non-degenerate differential 2-form $\Omega : T_m M \times T_m M \rightarrow \mathbf{R}$, i.e., the following conditions hold true*

$$d\Omega = 0 \quad \text{and} \quad \forall \xi \neq 0 \exists \eta : \Omega(\xi, \eta) \neq 0.$$

The pair (M, Ω) is called a symplectic manifold.

As a consequence of Darboux's Theorem [97] there is a natural symplectic form on the cotangent space T^*Q that is given by

$$\Omega = dq^i \wedge dp_i.\tag{2.9}$$

where the skew-symmetric wedge product of two differentials is defined as

$$\Omega(\xi, \eta) = dq^i(\xi) dp_i(\eta) - dq^i(\eta) dp_i(\xi).$$

with $dq^i(\zeta) = \zeta_1^i$, $dp_i(\zeta) = \zeta_2^i$ for $\zeta \in T_z T^*Q$ and $z = (q, p)$. If Q is one-dimensional, Ω is a volume form. For arbitrary dimension n , the *oriented Liouville volume* is obtained by taking n -fold exterior products of the symplectic form,

$$\Lambda = \frac{(-1)^{n(n-1)/2}}{n!} \underbrace{\Omega \wedge \dots \wedge \Omega}_{n \text{ times}} \quad (2.10)$$

which yields in coordinates

$$\Lambda = dq^1 \wedge \dots \wedge dq^n \wedge dp_1 \wedge \dots \wedge dp_n.$$

We will also need [66]

Definition 2.4. Let (M, Ω) and (N, Ξ) be symplectic manifolds. A smooth map $\phi : M \rightarrow N$ is called *symplectic*, if Ω is the pull-back of Ξ by ϕ . That is,

$$\phi^* : \Omega_z(\xi, \eta) = \Xi_{\phi(z)}(T_z \phi \cdot \xi, T_z \phi \cdot \eta), \quad \text{for } \xi, \eta \in T_z M,$$

where the tangent map

$$T_z \phi : T_z M \rightarrow T_{\phi(z)} N, \quad T_z \phi \cdot \zeta = \left. \frac{d}{ds} \phi(\gamma(s)) \right|_{s=0}$$

explains the derivative of ϕ at $z \in M$; here $\gamma(s)$ denotes a curve in M that satisfies $\gamma(0) = z$ and $\dot{\gamma}(0) = \zeta$, and the vector $T_z \phi \cdot \zeta$ only depends on ζ but not on the curve.

We consider the particular case of the Hamiltonian flow map $\Psi_t : T^*Q \rightarrow T^*Q$ with standard symplectic form $\Omega = dq^i \wedge dp_i$ for which the next statement holds [98].

Proposition 2.5 (Arnold 1989). *Hamiltonian flows are symplectic.*

Although it is common knowledge, we shall prove the *Liouville Theorem* as following from the symplecticness property for the readers convenience:

Corollary 2.6. *Hamiltonian flows preserve the Liouville volume $\Psi_t^* \Lambda = \Lambda$.*

Proof. The wedge product commutes with the pull-back, i.e., the pull-back of the wedge product is the wedge product of the pull-back. Hence (2.10) gives

$$\Psi_t^* \Lambda = \frac{(-1)^{n(n-1)/2}}{n!} (\Psi_t^* \Omega) \wedge \dots \wedge (\Psi_t^* \Omega),$$

but as $\Psi_t^* \Omega = \Omega$, according to Proposition 2.5, we conclude that $\Psi_t^* \Lambda = \Lambda$. \square

Assumption 2.7. *For the sake of convenience we shall introduce mass-scaled coordinates $q \mapsto M^{-1/2}q$ and $p \mapsto M^{-1/2}p$, where M is the symmetric positive-definite mass matrix. This is certainly a symplectic transform, i.e., it preserves the wedge product, and it allows us to set the molecular masses to one in what follows.*

2.1.1. Statistical mechanics: ensemble concepts So far we have considered a single molecule only; on physical grounds however this is often not the appropriate perspective: First of all, real experiments mostly deal with ensembles of molecules rather than with a single molecule. Secondly, it is doubtful that an experimentalist can precisely prepare or measure the initial state of a single molecule at room temperature.

Hence we study ensembles of molecules. The ensemble will be represented by a probability distribution of initial states, where the choice of the respective distribution will depend on the problem under consideration; situations where chemical reactions occur and the number of particles may change, require other ensemble concepts than

ensembles with a constant number of particles. We shall mainly treat the case of an ensemble at constant temperature with a constant number of particles and constant volume, the so-called *canonical ensemble*. The time evolution of the initial ensemble is then governed by the dynamics of the single molecules [67]. This can be understood in the following way: Let us abbreviate $z = (q, p)$, and let $f_0(z)$ be the initial preparation of the statistical ensemble. Moreover we denote by Ψ_t the Hamiltonian flow in phase space $E = T^*Q$. Then the density in time $f(z, t)$ will be the initial density that is transported along the orbits of the flow Ψ_t , that is,

$$f(z, t) = (f_0 \circ \Psi_{-t})(z), \quad \text{with} \quad f_0 = f(\cdot, 0). \quad (2.11)$$

Thus, the density f at time t is the *push-forward* of the initial density f_0 by the flow map Ψ_t . This is a consequence of Liouville's Theorem which implies conservation of probability. The last equation can be rephrased in terms of an evolution equation for f . If we denote by X_H the Hamiltonian vector field that is defined by (2.7), then f is governed by the Liouville equation

$$\partial_t f(z, t) = -\mathcal{L}f(z, t), \quad f(\cdot, 0) = f_0, \quad (2.12)$$

where the Liouville operator \mathcal{L} is defined as

$$\mathcal{L} = X_H(z) \cdot \nabla, \quad (2.13)$$

where ∇ is the derivative with respect to z^1, \dots, z^{2n} . The Liouvillian is skew-adjoint, if it is considered to act on an appropriate subspace of the Hilbert space $L^2(dz)$. For further reading the interested reader is referred to the original work of Koopman [99]. See also [100] and the references therein.

It is easy to check that the transported density (2.11) satisfies the Liouville equation, and in fact the Liouvillian is the generator of the corresponding *semigroup*,

$$f_0 \circ \Psi_{-t} = \exp(-t\mathcal{L})f_0. \quad (2.14)$$

Accordingly a stationary density is a function ρ for which $\rho(z) = (\rho \circ \Psi_{-t})(z)$ for all $z \in E$ at all instances of time t . Of course, *any* function of the Hamiltonian, $\rho(z) = f(H(z))$, is a stationary solution of the Liouville equation. One prominent representative of this class is the canonical density or Boltzmann density

$$\rho_{\text{can}}(z) = \frac{1}{Z} \exp(-\beta H(z)), \quad (2.15)$$

where

$$Z = \int_E \exp(-\beta H(z)) dz, \quad (2.16)$$

is the normalization constant (partition function), $\beta = 1/T$ is the inverse temperature, and dz denotes the Lebesgue measure associated with the Liouville form (2.10).

The semigroup notation above emphasizes the mathematical equivalence of the particle dynamics and the ensemble dynamics [101]. Indeed the relation (2.14) allows for likewise studying the statistical properties of the Hamiltonian system (2.7) in terms of the flow map Ψ_t : Consider the set \mathcal{M} of probability measures on E , that are absolutely continuous with respect to the Lebesgue measure dz (Liouville measure). According to the Liouville Theorem, dz is preserved by the flow and we can introduce the *transfer operator* or Frobenius-Perron operator P_t that acts on measures $\mu \in \mathcal{M}$,

$$(P_t \mu)(B) = (\mu \circ \Psi_{-t})(B) \quad \text{for all measurable sets } B \subset E$$

by means of the action on the associated smooth density,

$$(P_t f_0) = f_0 \circ \Psi_{-t}.$$

This amounts to interpreting the transfer operator in a probabilistic sense for a single particle, and we can rephrase the stationarity of the Gibbs density, stating that

$$\mu_{\text{can}}(dz) = \rho_{\text{can}}(z) dz$$

is an invariant measure of the Hamiltonian dynamics, that is, it satisfies $P_t \mu_{\text{can}} = \mu_{\text{can}}$. Consult [67, 102] for a detailed discussion in the context of metastability.

Stochastic Hamiltonian systems In general, statistical mechanics is about computing average quantities with respect to some probability measure. So far we have omitted the problem of how to compute such averages. In molecular dynamics the expectation value of an observable $A(z)$ is typically computed via time averages, exploiting the notion of ergodicity: a typical trajectory will eventually visit all possible states with $\mu > 0$. In a more formal language this means

$$\mathbf{E}A(z) = \int_E A(z) \mu(dz) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T A(\Psi_t z) dt.$$

The main difficulty here is that the deterministic flow Ψ_t has no unique invariant measure, since any function of the Hamiltonian yields a stationary solution of the Liouville equation. Even worse, the only candidates for ergodic invariant measures, like the equidistribution on the surface of constant energy, the microcanonical measure, are singular with respect to the Liouville measure. Conversely this means that we cannot compute averages with respect to the canonical measure μ_{can} from single particle trajectories [67]. In fact very few Hamiltonian systems are known to be ergodic: certain billiards and geodesic flows on surfaces of constant negative curvature [103, 104].

However it is possible to define a *stochastic* Hamiltonian system by averaging out the momenta from the deterministic flow, by which we obtain a dynamical system on configuration space only. For simplicity, we shall make two arrangements: Firstly, we consider only observables $A(q)$ that do not depend on the momenta $p \in T_q^*Q$. Secondly, we fix an observation time span $\tau > 0$ and define the *discrete* flow map

$$z_{k+1} = \Psi_\tau z_k, \quad z_k = (q(t_k), p(t_k)).$$

Notice that here the subscript denotes a time index that should not be confused with the respective coordinate label; anyway the distinction should be clear from the context. Further note that both restrictions are not too severe, as the restriction on the observables can mostly be circumvented by analytically doing the momentum averaging, and the discrete flow can be thought of stemming from a symplectic discretization of the Hamiltonian equations of motion [105, 82].

Let π denote the natural bundle projection $\pi : T^*Q \rightarrow Q$ from the cotangent bundle onto its base, and we let the parametric function

$$\varrho_q(p) = \frac{1}{Z(q)} \exp\left(-\frac{\beta}{2} g^{ij}(q) p_i p_j\right)$$

denote the probability density of the momentum variable p according to the Gibbs distribution ρ_{can} . The function $Z(q)$ normalizes the momentum density to one. Clearly the momentum distribution depends on the configuration q by means of the inverse metric tensor and the normalization constant; nevertheless it is locally Gaussian. We now introduce a *stochastic* Hamiltonian system as iterates of the map

$$q_{k+1} = (\pi \circ \Psi_\tau)(q_k, p_k) \tag{2.17}$$

with p_k randomly chosen according to the momentum distribution $\varrho_{q_k}(\cdot)$. The respective discrete *spatial* transfer operator S_τ that takes probability densities on Q forward in time, can then be defined as

$$S_\tau f(q) = \int (f \circ \pi \circ \Psi_\tau)(q, p) \varrho_q(p) dp.$$

Note the following important difference between the spatial and the full transfer operator P_t : The full propagator satisfies the semigroup property $P_{t+s} = P_s P_t$. However, the definition of the spatial propagator involves a momentum average, and in general averaging over the initial momenta and integrating up to time, say, 2τ gives a different result than averaging at the initial time and the intermediate time τ . In particular this means that S_τ has no generator.

In the following we will consider S_τ on the Hilbert space $L^2(\nu)$, that is weighted by the reduced probability measure $\nu(dq) = \varrho_0(q) dq$, where ϱ_0 is the marginal

$$\varrho_0(q) = \int_{T_q^* Q} \rho_{\text{can}}(q, p) dp.$$

The weighted Hilbert space is defined as

$$L^2(\nu) = \left\{ v : Q \rightarrow \mathbf{R} \mid \int_Q (v(q))^2 \nu(dq) < \infty \right\} \quad (2.18)$$

with the scalar product

$$\langle u, v \rangle_\nu = \int_Q u(q)v(q) \nu(dq). \quad (2.19)$$

The invariant probability measure that seems to be naturally associated with the stochastic Hamiltonian dynamical system (2.17) is ν . This means for the associated propagator that it must satisfy $S_\tau \mathbf{1} = \mathbf{1}$, as $\mathbf{1}$ denotes the invariant density with respect to ν . And indeed, it follows immediately from the fact that the canonical density is normalized that $S_\tau \mathbf{1} = \mathbf{1}$. Moreover the following is true [67]:

Proposition 2.8 (Schütte 1998). *Let $Q = \mathbf{T}^n$ be periodic. Then the transfer operator $S_\tau : L^2(\nu) \rightarrow L^2(\nu)$ has a simple eigenvalue $\lambda = 1$ that is bounded away from the remaining spectrum. Hence $\mathbf{1}$ is the unique invariant density in $L^2(\nu)$.*

2.2. Stochastic Langevin dynamics

Using deterministic dynamics has several advantages. First of all deterministic Hamiltonian systems are widely used in the molecular dynamics community and there is a variety of efficient numerical algorithms available [92, 106]. Moreover there is a well-established covariant formalism which is flexible enough so as to allow for the treatment of infinite dimensional problems, systems with symmetries or systems on manifolds [81]. However one severe drawback for molecular applications is the fact that there is no unique invariant phase space measure.

The *sampling problem* is mostly addressed by means of certain thermostating techniques like Nosé-Hoover, Berendsen or stochastic Andersen thermostats [88, 107]. Mostly, these algorithms modify the equations of motion in such a way that the dynamics samples the canonical density, provided the Hamiltonian flow is ergodic with respect to the microcanonical measure. This is a very strong assumption, and it is well-known that the ordinary Nosé-Hoover thermostat suffers from ergodicity problems for certain classes of Hamiltonians [108, 109]. This pathology can be

removed by employing extensions to the single-oscillator chain or by imposing constant temperature constraints [110, 111, 112]. But even then, the sampling works well *only if* the dynamics is ergodic, and conditions to guarantee ergodicity are still lacking. Additionally all these more sophisticated methods have in common that due to their complexity they are relatively hard to implement, and they require a careful adjustment of the parameters involved. See the recent survey article [113].

A promising alternative is stochastic Langevin dynamics or Brownian (Smoluchowski) dynamics [13]. These systems are proven to be ergodic under sufficiently weak assumptions which are similar to those from the last section, i.e., periodic or bounded configuration space [114, 115]. A Langevin system can be regarded as a mechanical system with additional noise and dissipation (friction). The noise can be thought of modelling the influence of a heat bath surrounding the molecule and the dissipation is chosen such as to counterbalance the energy fluctuations due to the noise. In its traditional form the equations of motion read [116]

$$\begin{aligned}\dot{q}(t) &= M^{-1}p(t) \\ \dot{p}(t) &= -\nabla V(q(t)) - \gamma M^{-1}p(t) + \sigma \dot{W}(t),\end{aligned}$$

where $\gamma, \sigma \in \mathbf{R}^{n \times n}$ are positive-definite matrices, and $W(t)$ is the standard n -dimensional Wiener process [117]. Here we use the expedient notation $\dot{W}(t)$ rather than $dW(t)$ for the increment of the Wiener process at the price of being less exact with regard to its correct interpretation [118]; so far it is sufficient to state that we understand the last equation in the sense of Itô, and therefore we will continue using this notation without further comments and omitting the time argument in most cases.

The fluctuation-dissipation theorem of Kubo [119] states that, if friction and noise satisfy the fluctuation-dissipation relation $2\gamma = \beta\sigma\sigma^T$, then the Langevin equation defines a Markov process that is ergodic with respect to the canonical measure μ_{can} . For our purposes it is convenient to rewrite the Langevin equation in a slightly different form: Let $H : T^*\mathbf{R}^n \rightarrow \mathbf{R}$ be the molecular Hamiltonian in mass-scaled coordinates

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

with conjugate coordinates $z = (q, p)$. Then the Langevin equation can be written as

$$\dot{z}(t) = (\mathbb{J} - \Gamma) \nabla H(z(t)) + \Sigma \cdot \dot{W}(t), \quad (2.20)$$

where ∇ denotes the derivative in \mathbf{R}^{2n} , and we have introduced the $2n \times 2n$ matrices

$$\mathbb{J} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \gamma \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma \end{pmatrix}.$$

Remark 2.9. *In principle the mass scaling $q \mapsto M^{-1/2}q$, $p \mapsto M^{-1/2}p$ affects both friction and noise terms in the Langevin equation in the way that the coefficients γ, σ scale according to $\gamma \mapsto M^{-1/2}\gamma M^{-1/2}$ and $\sigma \mapsto M^{-1/2}\sigma$. Hence the friction and noise coefficients become mass-dependent as well [39]. However it is easy to see that the mass scaling preserves the fluctuation-dissipation relation, and we refrain from making the scaling explicit. The reader should apologize this abuse of notation.*

Covariant form of the Langevin equation: Hamiltonian formulation

Unfortunately there is no canonical way that would lead to a covariant formulation of the Langevin equation in the spirit of the Euler-Lagrange equations or Hamilton's equations. As we shall see below the Langevin equation can be understood either as an Itô or as a Stratonovich equation, as long as we restrict our attention to

point transformations (i.e., symplectic lifts of configuration space maps to phase space); the distinction becomes irrelevant. The restriction to point transformations is not too severe, for the vast majority of physical problems that involve changes of variables are of that type. We will see below that equation (2.20) transforms like a dissipative Hamiltonian vector field, no matter whether it is interpreted as an Itô or a Stratonovich equation.¹ That is, the Langevin equation has some very specific properties as compared to general hypo-elliptic diffusion equations which are due to its Hamiltonian origin.

To the best of the author's knowledge the available approaches for covariant representations use projection operator techniques, where the system is coupled to a heat bath [121, 122]. These approaches involve very specific assumptions concerning timescale separation between system and bath variables or are based on Markov approximations of the friction kernel [123]. In principle one could simply apply Itô's formula to the Langevin equation, and recognize that it boils down to the familiar change-of-variables formula. Here we adopt an alternative approach by considering the associated evolution equation: The time evolution of a function $u \in L^1(dz)$ is governed by the Kolmogorov backward equation

$$\partial_t u(z, t) = \mathcal{A}_{\text{bw}} u(z, t), \quad u(\cdot, 0) = u_0, \quad (2.21)$$

where the backward generator is defined as

$$\mathcal{A}_{\text{bw}} = \frac{1}{2} \Delta_{\Sigma, \mathbf{R}^{2n}} + (\mathbb{J} - \Gamma) \nabla H(z) \cdot \nabla. \quad (2.22)$$

where $\Delta_{\Sigma, \mathbf{R}^{2n}}$ is the degenerate Laplacian with respect to the covariance matrix, i.e.,

$$\Delta_{\Sigma, \mathbf{R}^{2n}} = \text{tr}(\Sigma \Sigma^T \nabla^2).$$

Now we can easily study the transformation properties of the Langevin equation by means of the corresponding backward equation: Let $M = T^*Q$ be a symplectic manifold of dimension $2n$ that is endowed with a Riemannian metric h_0 , and understand ∇ as the covariant derivative in TM . Recall that the Laplace-Beltrami can be computed as the trace of the Hessian [68]. Further recall that the Hessian $\nabla^2 f : T_z M \rightarrow T_z M$ is defined as the matrix associated with the bilinear form $d^2 f(X, Y)$, where $X, Y \in TM$ are two vector fields that are evaluated point-wise at $z \in M$. It follows from Leibniz' rule that

$$\nabla_X df(Y) = d^2 f(X, Y) + df(\nabla_X Y),$$

and we obtain in coordinates:

$$d^2 f(X_i, X_j) = \frac{\partial^2 f}{\partial z^i \partial z^j} - \Gamma_{ij}^k \frac{\partial f}{\partial z^k}, \quad X_l = \frac{\partial}{\partial z^l}.$$

Here Γ_{ij}^k are the symmetric Christoffel symbols associated with the Riemannian metric h_0 on TM that contains the inverse covariance matrix. Now we define the following pseudo-Riemannian metric

$$h(z) = F(z) (\Sigma \Sigma^T)^{-1} F^T(z),$$

where F is obtained by a Cholesky decomposition of the phase space metric $h_0 = F^T F$. (The metric h is pseudo-Riemannian because only the momentum block is non-zero.) The corresponding symmetric Hessian matrix then reads in coordinates

$$(\nabla^2 f)_j^i = h^{il} \left(\frac{\partial^2 f}{\partial z^l \partial z^j} - \Gamma_{lj}^k \frac{\partial f}{\partial z^k} \right). \quad (2.23)$$

¹For the geometry of dissipative Hamiltonian vector fields without noise, see [120].

Double contraction (trace) of the Hessian yields the degenerate Laplace-Beltrami operator $\Delta_{\Sigma, M}$ on the phase space $M = T^*Q$, in other words: the diffusion generator of the Langevin equation on the phase space M . We state:

Lemma 2.10. *Let Q be a Riemannian manifold that is endowed with a metric g which is of the form $g = B^T B$, and let $H : T^*Q \rightarrow \mathbf{R}$ be the Hamiltonian*

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

Moreover we denote by $\tilde{\gamma}$ and $\tilde{\sigma}$ the usual friction and noise coefficients that satisfy the fluctuation-dissipation relation $2\tilde{\gamma} = \beta \tilde{\sigma} \tilde{\sigma}^T$. Abbreviating the canonical coordinates by $z = (q, p)$, the coordinate-invariant form of the Langevin equation is

$$\dot{z}(t) = (\mathbb{J} - \Gamma) \nabla H(z(t)) + \Sigma \cdot \dot{W}(t), \quad (2.24)$$

where the $2n \times 2n$ matrices are defined as

$$\mathbb{J} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B^T \tilde{\gamma} B \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B^T \tilde{\sigma} \end{pmatrix}.$$

Unwrapping the matrix-vector notation, (2.24) is equivalent to

$$\begin{aligned} \dot{q}^i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q^i} - \gamma_{ij} \frac{\partial H}{\partial p_j} + \sigma_{ij} \dot{W}^j, \quad i = 1, \dots, n \end{aligned} \quad (2.25)$$

where $\gamma = B^T \tilde{\gamma} B$, and $\sigma = B^T \tilde{\sigma}$ are q -dependent friction and noise matrices that satisfy the fluctuation-dissipation relation $2\gamma = \beta \sigma \sigma^T$.

Proof. The assertion concerning fluctuation-dissipation is trivial. For the first part it is sufficient to consider the evolution equation associated with the Langevin equation (2.20). To this end consider the molecular Hamiltonian

$$\tilde{H}(\eta, \xi) = \frac{1}{2} \langle \xi, \xi \rangle + \tilde{V}(\eta), \quad (\eta, \xi) \in T^*\mathbf{R}^n \cong \mathbf{R}^n \times \mathbf{R}^n,$$

which gives rise to the Langevin equation (2.20). We shall study the transformation properties of (2.20) by means of the corresponding backward equation

$$\partial_t u = \left(\frac{1}{2} \Delta_{\Sigma, \mathbf{R}^{2n}} + (\mathbb{J} - \tilde{\Gamma}) \tilde{\nabla} \tilde{H} \cdot \tilde{\nabla} \right) u.$$

Here $\tilde{\nabla} = (\partial/\partial\eta^1, \dots, \partial/\partial\xi_n)$ is the derivative in \mathbf{R}^{2n} , and $\tilde{\Gamma}, \tilde{\Sigma} \in \mathbf{R}^{2n \times 2n}$ are defined in an obvious way. Consider a diffeomorphism $\phi : Q \rightarrow \mathbf{R}^n$ with $\eta = \phi(q)$ which has a symplectic lift $T^*\phi : T^*Q \rightarrow T^*\mathbf{R}^n$ to phase space that is given by $(\eta, \xi) = (\phi(q), \mathbf{D}\phi(q)^{-T} p)$. In coordinates (q, p) the Hamiltonian reads

$$H(q, p) = \frac{1}{2} \langle g(q)^{-1} p, p \rangle + V(q) \quad (V = \tilde{V} \circ \phi),$$

with the induced metric tensor $g = B^T B$, $B = \mathbf{D}\phi$. Writing the backward equation in the new coordinates is straightforward: using (2.23) we obtain for the diffusion part

$$\frac{1}{2} \Delta_{\Sigma, T^*Q} = \frac{1}{2} \text{tr} (B^T \tilde{\sigma} \tilde{\sigma}^T B \mathbf{D}_2^2) =: \frac{1}{2} \text{tr} (\Sigma \Sigma^T \nabla^2),$$

where \mathbf{D}_2 denotes the derivatives with respect to the second (momentum) slot, and ∇ denotes the derivatives with respect to the new coordinates. Also the drift is

easily accounted for: since (q, p) are conjugate, the Hamiltonian vector field keeps its canonical form; moreover the dissipation part transforms like an ordinary vector field on $T^*Q \cong \mathbf{R}^{2n}$. The thus transformed drift part reads

$$(\mathbb{J} - \Gamma) \nabla H = \mathbb{J} \nabla H - \begin{pmatrix} \mathbf{0} \\ B^T \tilde{\gamma} B^{-T} p \end{pmatrix}$$

from which the assertion follows upon noting that $B^T \tilde{\gamma} B^{-T} = B^T \tilde{\gamma} B g^{-1} = \gamma g^{-1}$. \square

The proof reveals the origin of the Itô-Stratonovich equivalence in case of the Langevin equation without using the Itô formula: regarding the backward operator only the momentum derivatives contributes to the Laplace-Beltrami, and the momentum transforms linearly under point transformations. In the language of Itô calculus this means that the second order momentum derivatives in the Itô formula, which make up the additional terms, all vanish.

The time evolution of physical densities $f \in L^1(dz)$ is governed by the Kolmogorov forward equation, that is in some sense dual to the backward equation [102]

$$\partial_t f(z, t) = \mathcal{A}_{\text{fw}} f(z, t), \quad f(\cdot, 0) = f_0, \quad (2.26)$$

where the forward generator,

$$\mathcal{A}_{\text{fw}} = \frac{1}{2} \Delta_{\Sigma, M} - \nabla \cdot (\mathbb{J} - \Gamma) \nabla H. \quad (2.27)$$

is the formal adjoint of the backward generator on the Hilbert space $L^2(dz)$. (Note that the diffusion part of the operator is self-adjoint in $L^2(dz)$ since $\Delta_{\Sigma, M}$ involves only momentum derivatives, but the metric merely depends on the configuration variables.) The forward operator can then be rewritten according to

$$\mathcal{A}_{\text{fw}} = \frac{1}{2} \Delta_{\Sigma, M} - (\mathbb{J} - \Gamma) \nabla H \cdot \nabla + \nabla \cdot (\Gamma \nabla H),$$

where we have used that the Hamiltonian vector field is divergence-free. Alternatively the rightmost divergence can be written as $\text{tr}(\gamma g^{-1})$. Provided that the fluctuation-dissipation relation is met, the canonical (Boltzmann) density is a stationary solution of the forward equation, independently of the chosen coordinate system:

$$\mathcal{A}_{\text{fw}} \rho_{\text{can}} = 0, \quad \rho_{\text{can}} = \frac{1}{Z} \exp(-\beta H).$$

2.3. High-friction dynamics: Brownian motion

The Smoluchowski equation be understood as the high-friction limit of the Langevin equation. Unfortunately there is some ambiguity in the denotations throughout the literature: we shall use the term *Brownian motion* synonymously with *Smoluchowski equation* which describes a non-degenerate (elliptic) diffusion process in contrast to the degenerate (hypo-elliptic) Langevin equation. For the sake of illustration we shall briefly sketch the transition from Langevin to Smoluchowski for a one-dimensional system. To this end we write the Langevin equation in second-order form

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

We introduce a small parameter $\epsilon > 0$ by scaling the friction constant according to $\gamma \mapsto \gamma/\epsilon$; keeping the correct invariant density requires the scaling $\sigma \mapsto \sigma/\sqrt{\epsilon}$ for the

noise coefficient. In other words, the high-friction limit amounts to letting ϵ go to zero, so that we arrive at the singularly perturbed problem

$$M\ddot{q}_\epsilon = -\nabla V(q_\epsilon) - \frac{\gamma}{\epsilon}\dot{q}_\epsilon + \frac{\sigma}{\sqrt{\epsilon}}\dot{W},$$

So far the limit $\epsilon \rightarrow 0$ would yield the boring result $\dot{q} = 0$, as can be seen upon multiplying the whole equation by ϵ . However this changes if we consider the dynamics on the *diffusive timescale* $t \mapsto t/\epsilon$, on which the equations of motion take the form

$$\epsilon^2 M\ddot{q}_\epsilon = -\nabla V(q_\epsilon) - \gamma\dot{q}_\epsilon + \sigma\dot{W}.$$

Taking the limit $\epsilon \rightarrow 0$ formally leads to the Smoluchowski equation

$$\gamma\dot{q}_0 = -\nabla V(q_0) + \sigma\dot{W}. \quad (2.28)$$

The little word *formally* should be taken literally, because the white noise in the equation is unbounded, and therefore we cannot be sure that $\epsilon^2\ddot{q}_\epsilon$ goes to zero as ϵ goes to zero. Nevertheless it can be shown that the sample paths $q_\epsilon(t)$ converge to $q_0(t)$ with probability one [124, 125].

Covariant form of the Smoluchowski equation: geometric Itô formula

Loosely speaking the Smoluchowski equation can be understood as some sort of diffusive limit of the Langevin equation. However we have to take care that the Smoluchowski equation (SE) transforms in a way that is consistent with transformation properties of the Langevin equation (LE). Given a change of variables $\phi: \mathbf{R}^n \rightarrow Q$, this property is expressed in the following commutative diagram

$$\begin{array}{ccc} (\text{LE}, T^*\mathbf{R}^n) & \xrightarrow{T^*\phi: T^*Q \rightarrow T^*\mathbf{R}^n} & (\text{LE}, T^*Q) \\ \gamma \rightarrow \infty \downarrow & & \downarrow \gamma \rightarrow \infty \\ (\text{SE}, \mathbf{R}^n) & \xrightarrow{\phi: Q \rightarrow \mathbf{R}^n} & (\text{SE}, Q) \end{array}$$

Unlike the Langevin equations we understand the Smoluchowski equation strictly in the sense of Itô, because here the distinction really matters as we will see below. Without loss of generality we set $\gamma = \mathbf{1}$. Then the fluctuation-dissipation relation implies that $\sigma\sigma^T \propto \mathbf{1}$. In other words, the noise matrix σ is proportional to an orthogonal matrix. We may even choose $\sigma = \sqrt{2\beta^{-1}}$ scalar, since the white noise is invariant under orthogonal transformations [126]. Now the multidimensional Smoluchowski equation takes the form²

$$\dot{q} = -\text{grad} V(q) + \sqrt{2\beta^{-1}}\dot{W}, \quad q \in \mathbf{R}^n. \quad (2.29)$$

Note that we have tacitly replaced the derivative ∇ by the gradient; the first one is a one-form, whereas the latter defines a vector field on \mathbf{R}^n . This might seem somehow arbitrary but the replacement is owed to the fact that the friction matrix in the Langevin equation transforms according to $\gamma \mapsto B^T\gamma B$ under a change of coordinates. Hence $\gamma\dot{q}$ is an element of the cotangent space, and suppressing γ amounts to replacing ∇V by the gradient. For integrable functions $v \in L^1(dq)$ let us introduce the Kolmogorov backward equation that is associated with (2.29),

$$\partial_t v(q, t) = \mathcal{A}_{\text{bw}} v(q, t), \quad v(\cdot, 0) = v_0$$

²Setting $\gamma = \mathbf{1}$ is along the lines of the mass scaling, for we can always make γ vanish in the Smoluchowski equation by scaling the coordinates according to $q \mapsto \gamma^{-1/2}q$.

with the elliptic backward generator

$$\mathcal{A}_{\text{bw}} = \beta^{-1} \Delta - \text{grad} V(q) \cdot \nabla.$$

We shall prove the following statement which is due to [69]:

Lemma 2.11. *Let Q be a configuration manifold of dimension n that is equipped with a Riemannian metric $g = \langle \cdot, \cdot \rangle_g$ and coordinates $q = (q^1, \dots, q^n)$. Then the covariant Smoluchowski equation in Itô form reads*

$$\dot{q} = -\text{grad} V(q) + b(q) + a(q) \cdot \dot{W} \quad (2.30)$$

with the coefficients

$$b^i(q) = -\beta^{-1} g^{jk} \Gamma_{jk}^i \quad \text{and} \quad a(q) = \sqrt{2\beta^{-1} g(q)^{-1}} \quad (2.31)$$

where the Γ_{jk}^i are the matrices of the symmetric Christoffel symbols on TQ ,

$$\Gamma_{jk}^i = \frac{1}{2} g^{il} \left(\frac{\partial g_{kl}}{\partial q^j} + \frac{\partial g_{lj}}{\partial q^k} - \frac{\partial g_{jk}}{\partial q^l} \right).$$

The Stratonovich equivalent is obtained by omitting the additional drift vector $b(q)$.

Proof. Recall the definition of the diffusion backward generator

$$\mathcal{A}_{\text{bw}} = \beta^{-1} \Delta - \text{grad} V \cdot \nabla.$$

Now recall that the gradient of a function $V : Q \rightarrow \mathbf{R}$ is defined by the equation $dV(w) = \langle \text{grad} V, w \rangle_g$ for all vectors $w \in T_q Q$. In coordinates this means

$$(\text{grad} V)^i = g^{ij} \frac{\partial V}{\partial q^j}.$$

Accordingly, we know that the diffusion on Q is generated by the Laplace-Beltrami operator on Q which is obtained as the double contraction (trace) of the Hessian matrix (2.23) or, equivalently, $\Delta = \text{div} \circ \text{grad}$. In coordinates the Laplacian reads [127]

$$\Delta v = g^{jk} \left(\frac{\partial^2 v}{\partial q^j \partial q^k} - \Gamma_{jk}^i \frac{\partial v}{\partial q^i} \right),$$

where Γ_{jk}^i are the symmetric Christoffel symbols associated with the metric g . We observe that the first term in the Laplace-Beltrami appears as an ordinary diffusion, whereas the other one has the formal structure of a drift term. Thus the Laplacian can be split into two parts, and so the backward generator admits the representation

$$\mathcal{A}_{\text{bw}} = \mathcal{A}_{\text{diff}} + \mathcal{A}_{\text{drift}}$$

with the single operators

$$\begin{aligned} \mathcal{A}_{\text{diff}} &= \beta^{-1} g^{ij} \frac{\partial^2}{\partial q^i \partial q^j} \\ \mathcal{A}_{\text{drift}} &= - \left(\beta^{-1} g^{ij} \Gamma_{ij}^k + g^{kl} \frac{\partial V}{\partial q^l} \right) \frac{\partial}{\partial q^k} \end{aligned}$$

Writing down the associated Itô stochastic differential equation is straightforward. The Stratonovich form is obtained by simply considering the Smoluchowski equation as an ordinary first-order vector field which amounts to erasing all terms that involve Christoffel symbols. This proves the assertion. \square

Remark 2.12. We point out an important distinction: on the Euclidean configuration space $U \subseteq \mathbf{R}^n$ consider the Brownian motion of a free particle

$$\dot{u}(t) = \dot{W}(t).$$

Now let $\phi : Q \rightarrow U$ be a change of coordinates, such that $u = \phi(q)$. The map ϕ induces a Riemannian metric $g = J_\phi^T J_\phi$ on Q , where $J_\phi(q) = \mathbf{D}\phi(q)$ is the Jacobian of the transformation. According to Lemma 2.11 the respective Itô equation on Q becomes

$$\dot{q}^i(t) = -\frac{1}{2}g^{jk}\Gamma_{jk}^i + a^{ij}\dot{W}_j(t),$$

where a is the uniquely defined positive-definite matrix square root of g^{-1} . However note that a is not the only admissible choice for the diffusion matrix, that stems from reinterpreting the diffusion part $\mathcal{A}_{\text{diff}} = g^{ij}\partial^2/\partial q^i\partial q^j$ in the associated backward generator in terms of white noise in the differential equation. In particular if we demand that the two processes are related pathwise by the coordinate map, $u(t) = \phi(q(t))$, then the noise term is altered due to:

$$\dot{q}^i(t) = -\frac{1}{2}g^{jk}\Gamma_{jk}^i + J_\phi^{ij}\dot{W}_j(t),$$

where J_ϕ^{ij} are the entries of the inverse Jacobian J_ϕ^{-1} . (In many cases, J_ϕ is uniquely related to the metric $g = J_\phi^T J_\phi$ by a Cholesky decomposition. If ϕ is conformal, it may even happen that J_ϕ is nontrivial, although $g = \mathbf{1}$ and thus $\sigma = a$.)

As before, it is the Kolmogorov forward equation which governs the time evolution of densities. For functions $f \in L^1(dq)$ the forward equation reads [128]

$$\partial_t f(q, t) = \mathcal{A}_{\text{fw}} f(q, t), \quad f(\cdot, 0) = f_0.$$

The forward operator is defined as the formal adjoint of the backward operator in the Hilbert space $L^2(d\sigma)$, where $d\sigma = \sqrt{\det g(q)}dq$ denotes the volume element on Q . It is easy to see, using integration by parts, that the forward operator has the form³

$$\mathcal{A}_{\text{fw}} = \beta^{-1}\Delta + \text{grad } V(q) \cdot \nabla + \Delta V(q).$$

It is then straightforward to show that the Gibbs density

$$\rho(q) = \frac{1}{Z} \exp(-\beta V(q))$$

is a stationary solution of the forward equation, i.e., $\mathcal{A}_{\text{fw}}\rho = 0$. In turn it follows that

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

is the invariant measure (Gibbs measure) of the Smoluchowski equation (2.30). (Notice that the Gibbs density ρ is the density of the Gibbs measure ν with respect to the volume on Q .) There is yet another operator which we will often consider, and that is formally equivalent to the backward operator but acts on a different function space. Following [102], we consider the Fokker-Planck equation that evolves densities in time with respect to the invariant probability measure $\nu(dq)$

$$\partial_t w(q, t) = \mathcal{L} w(q, t), \quad w(\cdot, 0) = w_0$$

with the (backward) Fokker-Planck operator

$$\mathcal{L} = \beta^{-1}\Delta - \text{grad } V(q) \cdot \nabla,$$

which is regarded on a suitable subspace of $L^1(\nu)$. Clearly $\mathcal{L}\mathbf{1} = 0$, since the constant function $\mathbf{1}$ is the invariant Gibbs density with respect to ν . Moreover it can be shown that the Fokker-Planck generator is self-adjoint:

³It follows directly from two iterations of Stokes' Theorem (i.e., Green's formula) that the Laplace-Beltrami operator is essentially self-adjoint in the Hilbert space $L^2(d\sigma)$, but not in $L^2(dq)$.

Proposition 2.13. *Let the weighted Hilbert space be defined according to (2.18) with the respective weighted scalar product $\langle \cdot, \cdot \rangle_\nu$. Then $\langle \mathcal{L}u, v \rangle_\nu = \langle u, \mathcal{L}v \rangle_\nu$ for functions $u, v \in L^2(\nu)$, i.e., \mathcal{L} is self-adjoint in $L^2(\nu)$.*

Proof. The proof for the Euclidean case is standard can be found, e.g., in [102]. Exploiting that \mathcal{L} can be written as a Schrödinger operator in $L^2(d\sigma)$, the proof carries over to the manifold case considered here at almost no further expense. \square

2.4. Essential degrees of freedom

Most systems of nonlinear differential equations, either deterministic or stochastic, that model *real-world* problems are characterized by vastly different timescales, on which certain dynamical effects happen [129]. In many molecular systems such timescales are well-separated, since different physical interactions in the molecular potential induce different characteristic timescales. Roughly speaking, we can assign a natural timescale τ to a physical process that is inversely proportional to its average energy. Accordingly, the vibrations of the molecular bonds (about 1 femtosecond) are typically the fastest modes in the system. We introduce the ratio

$$\epsilon = \frac{\tau_{\text{fast}}}{\tau_{\text{slow}}} \ll 1$$

which serves as a measure for the separation of timescales in the system. Often things become more complicated as the degrees of freedom are coupled among each other. Then the fast variables may induce slow motions elsewhere in the system, such that the assignment of timescales to certain modes gets difficult. Dividing a molecular system into slow and fast degrees of freedom is motivated by the observation that the conformational dynamics of a molecule is typically slow, as compared to the remaining degrees of freedom, since transitions between conformations are rare events. However this statement is in some respects circular, since the distinguishability of conformations pretty much depends on the choice of the right essential (slow) coordinate.

We understand conformation dynamics as taking place in configuration space; in particular in the Hamiltonian context, the restriction to the configuration variables was owed to generating the Gibbs measure as the unique invariant measure of the dynamics. But even then one might imagine that the distinction between conformational degrees of freedom and those which are of minor interest is not along temporal scales but along spatial scales, in case of which we could be interested in those degrees of freedom that have the largest amplitude [43]. In fact conformational behaviour is a spatial property of a molecule, in the sense that it is related to the its geometric shape. Nevertheless the dynamics between conformations is often driven by either crossing of energy barriers or other temperature-dependent entropic mechanisms, both of which are inherent temporal effects [130, 131]. Clearly one might think of other possible subdivisions: the principal decision to be made is whether the velocities or momenta should be taken into account; this is possible in principle; standard references for such purely deterministic approaches that treat the full phase space are [31, 32]; see also [132, 133] for some recent developments.

2.4.1. Spatial decomposition methods It is not the aim of this thesis to discuss the problem of finding good reaction coordinates in great detail or even to resolve it; for our purposes it is sufficient to assume that we are given a set of slow coordinates with significant spatial amplitude, for instance, by sufficient physical insight a priori or by

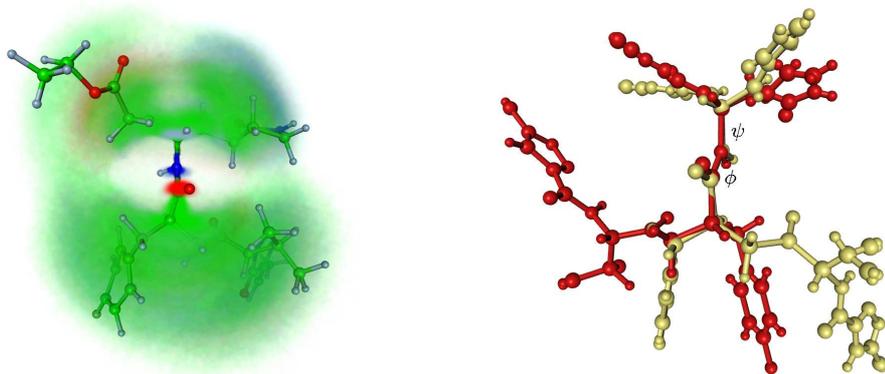


Figure 2. The left panel of the figure shows a volume rendering plot of the SARS-protease inhibitor at — allowedly unrealistic — temperature $T=1500\text{K}$. The clouds indicate the flexibility (fast oscillations) of the two main conformations which are depicted on the right, and which are well described by the two central torsion angles ϕ, ψ . See [135] for the simulation details.

statistical tools that operate on the produced time series *a posteriori*. Accordingly this subsection mainly surveys the available approaches for the identification of essential degrees in molecular systems, i.e., spatially extended motions that are slow in some appropriate sense. At the very beginning of this chapter we have seen that molecular interaction forces are defined in terms of internal degrees of freedom of a molecule. Hence it seems natural to suspect the essential coordinates among the internal coordinates. Indeed in many cases the conformation dynamics can be well described by specific dihedral (torsion) angles which link certain approximately rigid subunits of the molecule; see Figure 2 below for illustration. Other possible candidates for essential degrees of freedom are radii of gyration, certain intramolecular distances between selected endgroups of the molecule, or distances between parts of a molecule and some specific solvent molecules [134]. In any event the reader should keep in mind that essential degrees of freedom are not intrinsically defined by the molecule under consideration, but rather depend on the specific problem.

In case the essential degrees of freedom are not known a priori they can be possibly detected by statistical analysis of simulation data. There are plenty of such data-based methods and algorithms that all have in common that they sort out a linear subspace of the data space which is in some sense optimal, although the methods might be rather different in detail. This subspace is then assumed to be spanned by the essential variables [136]. Mostly these techniques work on Euclidean data only, and it would be nice to apply them in a more general framework to data that lies on a manifold; we shall discuss this issue below.

The most famous approach is certainly the method of Principal Component Analysis (PCA) which is also known as Karhunen-Loève expansion or Proper Orthogonal Decomposition (POD) to name just a few [137, 45].

The idea is to find a best-approximating k -dimensional subspace to a given set of data. We may characterize this subspace $S \subset \mathbf{R}^n$ by a projection operator \mathcal{P} ,

mapping \mathbf{R}^n onto S . We give a brief derivation of the method to provide some geometric intuition for the following discussion. Assume the data consist of sampled measurements $\{x(t_1), \dots, x(t_N)\}$ of a trajectory $x(t) \in \mathbf{R}^n$. Finding the optimal subspace S amounts to solving the least-square problem⁴

$$\min_{\mathcal{P}} \sum_{l=1}^N \|x(t_l) - \mathcal{P}x(t_l)\|_2^2$$

where the rank of \mathcal{P} is k . This problem can easily be reformulated as a related variational problem: Let S be a one dimensional subspace in \mathbf{R}^n spanned by the vector $w \in \mathbf{R}^n$, such that the projection \mathcal{P} becomes the linear map $\mathcal{P} = \langle \cdot, w \rangle w$. Since $\|x\|^2 = \|x - \mathcal{P}x\|^2 + \|\mathcal{P}x\|^2$ holds true for any projection, the least-square problem is equivalent to maximizing the *energy* of the projection,

$$E[w] = \sum_{l=1}^N \langle x(t_l), w \rangle^2 - \lambda(\|w\|_2^2 - 1),$$

where the Lagrange multiplier imposes normalization upon the vector w . Taking the variation of $w = w_\epsilon$ keeping x fixed it follows by chain rule $\delta E_\epsilon = \langle \mathbf{d}E, \delta w_\epsilon \rangle$ that the critical points of E have to satisfy the equation

$$2 \sum_{l=1}^N \langle x(t_l), w \rangle \langle x(t_l), \delta w_0 \rangle - 2\lambda \langle w, \delta w_0 \rangle = 0.$$

Since the last equation clearly must hold for arbitrary variations δw_ϵ we conclude that

$$\left(\sum_{l=1}^N \langle x(t_l), \cdot \rangle x(t_l) \right) w = \lambda w,$$

where (modulo normalization) the sum on the left hand side is an estimator of the symmetric covariance matrix of the data with accuracy $\mathcal{O}(N^{-1/2})$. Assuming that the covariance matrix has maximum rank n , the eigenvectors form a complete orthogonal set of vectors w_1, w_2, \dots, w_n . Consequently, we may repeat this whole procedure on the orthogonal complement to $w = w_1$, and find that then the best approximation is given by the second eigenvector w_2 , and so on. Finally the rank- k approximant S is spanned by the first k eigenvectors which is optimal in the sense that

$$\min_{\mathcal{P}} \sum_{l=1}^N \|x(t_l) - \mathcal{P}x(t_l)\|_2^2 = \sum_{j=k+1}^n \lambda_j$$

is the minimum achieved by any k -plane in \mathbf{R}^n , where the λ_j are the ordered eigenvalues of the covariance matrix; to put this differently, the chosen subspace contains the maximum variance of all k -dimensional linear subspaces [45]. By this procedure we find the optimal subspace; if we want to find the optimal *affine* subspace, we can exploit that it must pass through the mean of the data, and thus simply centre the data by subtracting the mean before computing the covariance.

Although POD is optimal at approximating a given data set it is not necessarily so for describing the dynamics that generates the particular data, for features of low

⁴There is some freedom in choosing a vector norm with respect to which the minimization is carried out. Frequent choices, for instance, are energy-based norms that are induced by an inner product $\langle u, v \rangle_A = \langle Au, v \rangle$, where $A = M$ may be the molecular mass matrix or the matrix of a quadratic Lyapunov function (notice that the former is equivalent to using mass-scaled coordinates). The specific choice of an inner product should be suited to the problem under consideration.

variance (energy) may induce important effects. In the molecular dynamics community a large amount of variance in the subspace is often taken as equivalent to the statement that the conformational dynamics takes place in this subspace. This is often wrong, as has been pointed out on various occasions; see [138, 139], for instance. Although in the derivation above we have nowhere assumed that the data follows an unimodal or even Gaussian distribution, it is self-evident that the method is reliable only for unimodal data. Hence a one-dimensional subspace containing, say, 99% of the variance is not necessarily the essential subspace with regard to the conformation dynamics.

Apparently, choosing the essential degrees of freedom merely according to their variance is not a good idea in general. Nevertheless we can consider the full set of eigenvectors w_1, w_2, \dots, w_n as a new basis for \mathbf{R}^n that gives an ordering according to variance. This may serve as a hint what the essential subspace could be, and we know that the slow modes are good candidates for the conformational degrees of freedom. Hence we introduce characteristic timescales for the thus rotated modes,

$$z(t) = R \cdot x(t), \quad R = (w_1, \dots, w_n) \in O(n),$$

by the respective *decorrelation time* τ which is defined as the integral

$$\tau = \int_0^\infty |\rho_i(s)| ds, \quad \rho_i(s) = \text{cor}(z_i(0), z_i(s)).$$

Loosely speaking, the decorrelation time indicates when $z(t)$ and $z(t+s)$ are effectively independent. We stress that the thus defined characteristic timescale is a technical notion rather than a sound and uniquely defined mathematical term; there are alternative definitions: in particular in the climate modelling community the integral is sometimes computed using the square of the autocorrelation function or the function itself [140]. However we claim that a slowly decaying autocorrelation function with $\tau \gg 1$ together with a large variance indicates possibly relevant dynamical behaviour. Of course this approach is not completely tight, but it is based on the fundamental observation that conformational changes in a molecule are related to spatially large and relatively slow rearrangements of the configuration variables.

Diagonalizing the data covariance matrix is not the only way to proceed; in particular the assignment of timescales can be linked with any other method as well. Another technique takes advantage of the insight that *correlated motions* in molecules, in particular proteins, are ubiquitous and often essential for biomolecular function. In principle, the covariance analysis can be extended to using the correlation matrix instead which means using an appropriately *normalized* covariance matrix. This approach, however, heavily relies on a quasi-harmonic treatment of the configurational ensemble, as it detects only linearly correlated motions [141, 142].

A quite promising approach is the Full Correlation Analysis which is based on the information theoretical concept of mutual information. It allows to detect and quantify any correlated motion from the Cartesian molecular dynamics trajectories. As in the previous cases the methods singles out a linear subspace of \mathbf{R}^n which is spanned by maximally uncorrelated basis vectors which are the solution of a generalized nonlinear eigenvalue problem [54]. This method has proven superior to the classical correlation analysis on some occasions [143]. Related approaches are Independent Component Analysis [144] and Non-Gaussian Component Analysis [145].

Yet another dynamics-based technique is the method of Principal Interaction Patterns (PIP), as introduced by Hasselmann [146]. It is predominantly used in the climate modelling community, and it takes into account the dynamical system upon

which a reduced dynamical model is built: the aim is to minimize the difference

$$J[\Theta] = \int_0^T \|x(t) - x_\Theta(t)\| dt,$$

where $x_\Theta(t)$ is the solution of a reduced dynamical system that is described by a set of functions or parameters Θ ; for instance, one may think of Θ as the linear projection \mathcal{P} from above, such that the reduced system is simply the Galerkin projection of the original equations. Now PIP involves two steps, the first of which is an ensemble average over the initial values $x(0), x_\Theta(0)$, followed by minimization with respect to Θ . Eventually, finding the essential degrees of freedom x_Θ boils down to solving a nonlinear optimization problem in \mathcal{P} . Alternatively one might imagine that Θ parametrizes a reduced model [147]. Although theoretically appealing we believe that this method does not lead to computationally tractable problems for most molecular systems; we quote it for the sake of completeness; see also [148].

We finally mention that the problem of maximizing the decorrelation time can be addressed directly, which is known by the name of Optimal Persistence Patterns. This technique aims at identifying the linear subspace that has the slowest decay of correlation. There are several possible ways to measure persistence (e.g., by means of the decorrelation time τ), and it is shown in [149] that finding the optimally persistent subspace results in solving a generalized eigenvalue problem. However since it is our strong belief that conformation dynamics manifests itself as macroscopic behaviour in both space and time, we prefer the combined approach of least-square approximations like POD together with the analysis of decorrelation times.

Transfer operator approach The transfer operator approach to metastability rests upon the observation that conformation dynamics can be understood as flipping dynamics between distinct subsets of configuration space that are *almost invariant* under the dynamics [67, 150]. Hence the problem of identifying conformations amounts to the identification of almost invariant sets in configuration space, where *almost* is understood in a way that transitions between those sets are rare. The method exploits the close relationship between the flow of a dynamical system, either Langevin, Smoluchowski or stochastic Hamiltonian, with its associated Frobenius-Perron operator S_t on configuration space.

The key idea is that the natural *invariant* sets and measures are given by the eigenvectors or eigenfunctions to the eigenvalue $\lambda_0 = 1$, whereas the eigenfunctions to $\lambda_k < 1$ that are close to $\lambda_0 = 1$ correspond to the *almost invariant* sets [151]. Suppose that the probability of the system to be in configuration $q \in Q$ at time $t = 0$ is given by the unique invariant density $f_0 = \rho_{\text{can}}$. Then the transition probability $p(\tau, A, B)$ from $A \subset Q$ to $B \subset Q$ within the observation time τ is given by the fraction of the ensemble that has started in A at $t = 0$, and which has ended up in B at time $t = \tau$. Hence $p(\tau, A, B)$ is the conditional probability

$$\mathbf{P}_\rho [q(\tau) \in B \mid q(0) \in A] = \frac{\mathbf{P}_\rho [q(\tau) \in B \cap q(0) \in A]}{\mathbf{P}_\rho [q(0) \in A]},$$

where subscript indicates that the initial preparation is given by ρ_{can} . Finally, a metastable set A is characterized by the requirement that $p(\tau, A, A) \approx 1$. This is to say that during a fixed observation time τ the system is likely to stay within the set. The algorithmic strategy that is put forward in [67] is to identify metastable subsets from the eigenfunctions of the Frobenius-Perron operator S_τ that correspond

to eigenvalues $\lambda < 1$ close to the Perron root $\lambda_0 = 1$. The number of metastable sets is then equal to the number of eigenvalues close to one, including $\lambda_0 = 1$ and counting multiplicity. The justification of this strategy is given in [102, 152].

2.4.2. Reaction coordinates The term *reaction coordinate* is typically used in a very loose sense and synonymous with *essential degrees of freedom*. More precise, a reaction coordinate can be understood as defining a family of isocommittor surfaces, such that all trajectory launched from the same surface have equal probability to first reach one metastable set before the other [153, 154]. The isocommittor surfaces are the codimension-one level sets of the reaction coordinates which, within the framework of Brownian motion, can be systematically computed from the Kolmogorov backward equation. For simplicity assume that there are only two conformations, or metastable sets $A \subset Q$ and $B \subset Q$. If \mathcal{A}_{bw} is the backward generator associated with the molecular diffusion process, then the solution of the boundary value problem

$$\mathcal{A}_{\text{bw}}\Phi = 0, \quad \Phi|_{\partial A} = 0, \Phi|_{\partial B} = 1$$

yields the reaction coordinate in the sense that the probability that a process starting from $q \in Q$ reaches B before A is $\Phi(q)$; so to speak, the level sets $\Phi(q) = \xi$ with ξ between 0 and 1 are the isocommittor surfaces of the process [13]. This clearly presumes that the metastable sets have been identified in advance, e.g., by the transfer operator algorithm as presented in the last paragraph.

In principle the procedure works in the same way for processes that are generated by Langevin equations, with the only difference that the reaction coordinate becomes a function on phase space [155]. But notice that *reaction coordinate* with its probabilistic meaning lacks a physical interpretation in terms of the intramolecular motion, i.e., it does not *explain* the transition mechanism. This issue is addressed by algorithms like the String Method [156] or the Nudged Elastic Band method [157]. Further note that although the direct identification of reaction coordinates is theoretically appealing, it requires solving a high-dimensional PDE which is part of a current PhD thesis [158].

2.5. Problems related to symmetry and further generalizations

As we have illustrated any of the introduced methods can provide a set of essential variables, and it depends on the specific problem which method to take. However we have concealed that all of the methods require some sort of preprocessing, before they can be applied to the raw Cartesian data. The reason is that in the absence of external forces, the intramolecular force field is translationally and rotationally invariant; therefore also the equations of motion are equivariant with respect to the Euclidean group $SE(3) \cong \mathbf{R}^3 \times SO(3)$ which consists of translations and proper rotations in \mathbf{R}^3 (rigid body symmetry). But since the rigid body symmetry carries no interesting information for the conformation analysis, the symmetry group is factored out before the data analysis starts; otherwise the motion that is associated with the symmetry deteriorates the statistical analysis, in particular the covariance analysis.

Typically the symmetry reduction is done by means of a molecular alignment of the simulation data, also sometimes called *fitting*, which removes the overall translations and rotations from the Cartesian data. This has several drawbacks: First of all, molecular alignment algorithms operate with respect to an arbitrarily chosen reference configuration of the molecule, such that all rotations and translations are carried out such as to minimize some distance to the reference state [159]. Often only a

specific group of atoms is chosen as reference, with respect to which the data is aligned, and, as the reader can imagine, the subspace identification depends on the choice of the reference configuration. Secondly and even worse, the preprocessing destroys the correspondence between data and equations of motion, for the symmetry reduction is performed on the data indeed, but the equations of motion are left untouched. This inconsistency has the effect that, e.g., the dichotomy *free energy as the potential of mean force* for a reaction coordinate that is not invariant under translations and rotation is annihilated. As a consequence the free energy as computed from the distribution of the symmetry-reduced data is different from the free energy that is computed by means of Thermodynamic Integration.

Therefore we want to apply the techniques to general non-Euclidean data which could, e.g., be obtained from nonlinear transformations of the original data. On the one hand these transformations can be chosen such as to respect the symmetry, like the transformation to internal coordinates. On the other hand this requires that we know how to compute the covariance matrix in a meaningful way, say, for data lying on a torus. We shall explain this idea in more detail: Consider a set of curvilinear coordinates that are represented by the following map

$$\varphi : \mathbf{R}^n \rightarrow M \subset \mathbf{R}^m \quad \text{with} \quad \varphi(t) := \varphi(x(t)),$$

where in most cases $m \leq n$, but also the reverse case can be dealt with [160]. Given a series of observation data $x(t)$, we define a centered data set by simply subtracting the mean $\varphi(t) \mapsto \varphi(t) - \bar{\varphi}$, where the term *mean* has to be specified in a way that fits the problem; see [161] and the references therein. It seems appealing to decompose M directly. However neither do we believe that there is a straightforward generalization of the above mentioned concepts to manifolds nor do we assume that such a procedure would lead to computationally tractable problems. Instead we favour the following approach going back to considerations from nonlinear elasticity [162]: the idea is to embed M into a linear space V of higher dimension, and apply a linear subspace decomposition to V ; once a subspace $U \subset V$ is chosen, we can construct the approximant $S \subset M$ as the intersection $M \cap U$. The last step is to be understood as follows: Let v be a local coordinate map on V , and let $\mathcal{P}v \in U$, where \mathcal{P} is the usual projection from V to $U \subset V$. Moreover let the map $\tau : M \rightarrow V$ denote the embedding $M \subset V$. Then $\tau^{-1} \circ \mathcal{P} \circ \tau$ maps $\varphi \in M$ to points on the approximant.

The particular embedding τ is open to choice, and should be motivated by the physics of the problem. For an example using a polar decomposition of torsion space, see [163]; this particular embedding is quite problematic, for all points lie on a hypersphere S^{2m-1} , where m is the number of torsion angles. There is yet another difficulty: Notice that the approximant is a regular submanifold of codimension s in M ; but then, considered as a submanifold of \mathbf{R}^n it has codimension s , too. Hence the approximant has dimension $n - s$ which is still high-dimensional, since mostly $m \ll n$.

It is a good idea to change the view slightly: consider the non-approximating subspaces which are defined as the fibres over the essential variables or reaction coordinates, and which are given by a family of configurational submanifolds

$$\Sigma_\xi = \{x \in \mathbf{R}^n \mid \Phi : \mathbf{R}^n \rightarrow \mathbf{R}^{m-s}, \Phi(x) = \xi\}, \quad (2.32)$$

with the reaction coordinate

$$\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^{m-s}, \quad \Phi = \chi \circ \tau^{-1} \circ \mathcal{P} \circ \tau \circ \varphi,$$

where $\chi : \mathbf{R}^m \rightarrow \mathbf{R}^{m-s}$ is a local coordinate map on S . By construction, $\mathbf{D}\Phi$ has rank $m - s$ which is equivalent to stating that the fibres $\Phi^{-1}(\xi)$ are smooth submanifolds

of \mathbf{R}^n , and so $\Phi(x) = \xi$ defines a foliation of the Euclidean configuration space. The reaction coordinates can then be thought of as *spanning* the essential configuration space [42]. Nevertheless it is important to bear in mind that the object of interest is the approximating subspace rather than the respective reaction coordinates. Moreover it is no longer the case that Σ_ξ is the orthogonal complement of S in \mathbf{R}^n as was true in the linear scenario. However proceeding this way has the advantage that a clever choice of both φ and τ can lead to relatively low-dimensional essential variables, e.g., linear combinations of a few torsion angles or radii of gyration (cf. Figure 2).

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 $\epsilon \rightarrow 0$ 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, 164].

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

⁵We 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 $|h(\epsilon)\epsilon^{-\alpha}| \rightarrow c \geq 0$ exists for $\epsilon \rightarrow 0$.

⁶Yet 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 [165, 166, 167, 168].

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 [169]; see the review [1] and the references therein. Moreover it is a common belief 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 [170, 171]. 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 [172]. 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., $\rho = Z^{-1} \exp(-\beta H)$:

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

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

$$F = -\beta^{-1} \ln Z \quad (3.4)$$

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 $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ 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 $H : T^*\mathbf{R}^n \rightarrow \mathbf{R}$ 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 $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ denote a smooth reaction coordinate. Then the free energy along the values of Φ is defined as

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

with the partition function

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

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, 174]

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

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 [175]

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

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:

⁷According to Sard's Lemma [173] this can be guaranteed by choosing the $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$, 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*

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

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*

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

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

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

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

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

The calculation is easily carried out in an adapted coordinate frame. To this end we let $\sigma : \mathbf{R}^d \rightarrow \Sigma$, $d = n - k$ be the embedding $\Sigma \subset \mathbf{R}^n$, and we let $\{n_1(\sigma), \dots, n_k(\sigma)\}$ 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

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

is a local embedding $N\Sigma_\varepsilon \subset \mathbf{R}^n$. By means of ϕ we can uniquely represent any point $q \in \mathbf{R}^n \cap N\Sigma_\varepsilon$ 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 $T^*\phi : T^*\mathbf{R}^n \rightarrow T^*N\Sigma_\varepsilon$. 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 $H_\Phi = H + \beta^{-1} \ln \text{vol} J_\Phi$. Using chain rule the derivative in (3.11) now becomes

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

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

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

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

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

To complete the proof we show that the matrix $(Q^T J_{\Phi})^{-1} Q^T$ 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 [176]

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

by which the assertion immediately follows.⁸ \square

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 [178]; cf. the results in [179].*

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

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

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

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

⁸Intriguingly 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 $A^{\sharp} = (B^T A)^{-1} B^T$ 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 [177].

By chain rule it follows for the potential term

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

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

$$\left. \frac{\partial L}{\partial \eta^i} \right|_{\eta=0} = S_{\alpha\beta}^i(x) \dot{x}^\alpha \dot{x}^\beta + \omega_j^i(X_\alpha) \dot{x}^\alpha \dot{\eta}^j, \quad (3.12)$$

where

$$S_{\alpha\beta}^i = \langle dn_i(X_\alpha), X_\beta \rangle$$

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_j^i are the normal fundamental forms that are associated with the normal frame $\{n_1, \dots, n_k\}$:

$$\omega_j^i(X_\alpha) = \langle n_i, dn_j(X_\alpha) \rangle$$

Note that the term involving the normal connection is 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

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

with

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

The last quantity \hat{f}_ξ 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

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

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}_ξ equals

$$\hat{f}_\xi = (J_\Phi^T(q)J_\Phi(q))^{-1} (J_\Phi^T(q) \nabla V_\Phi(q) - \langle \nabla^2 \Phi(q) \cdot v, v \rangle), \quad (3.14)$$

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

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

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

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 $\text{grad } \Phi_i$.

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 [180]*

$$P_N \nabla_X Y = \mathbb{I}(X, Y), \quad (3.16)$$

where ∇ is the (covariant) differentiation in \mathbf{R}^n , and X, Y are both tangent along Σ . The second fundamental form \mathbb{I} is defined by means of the Weingarten maps [181]

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

The symmetric Weingarten maps $\mathfrak{S}_i : T_{\sigma} \Sigma \rightarrow T_{\sigma} \Sigma$ 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

$$P_N \nabla_X \nu = D_X \nu, \quad (3.17)$$

where $D_X \nu$ is the connection of the normal bundle. Given a normal frame $\{n_1, \dots, n_k\}$ the connection can be written by means of the normal fundamental forms [182]:

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

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].

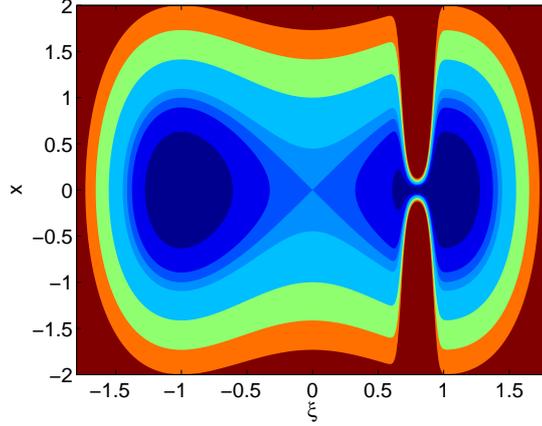


Figure 3. The plot shows the potential (3.19) with the dynamical barrier for the parameters $C = 15$, $\alpha = 200$, and $\xi_0 = 0.8$. The deep cut-away comes from the frequency peak of the harmonic oscillator.

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 $H : T^*\mathbf{R}^n \rightarrow \mathbf{R}$

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

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

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

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

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

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 conformational dynamics. Let us carry the example above to the extremes, and set

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

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

$$\omega(\xi) = 1 + C \exp(-\alpha(\xi - \xi_0)^2). \quad (3.20)$$

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

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

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):

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

with

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

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

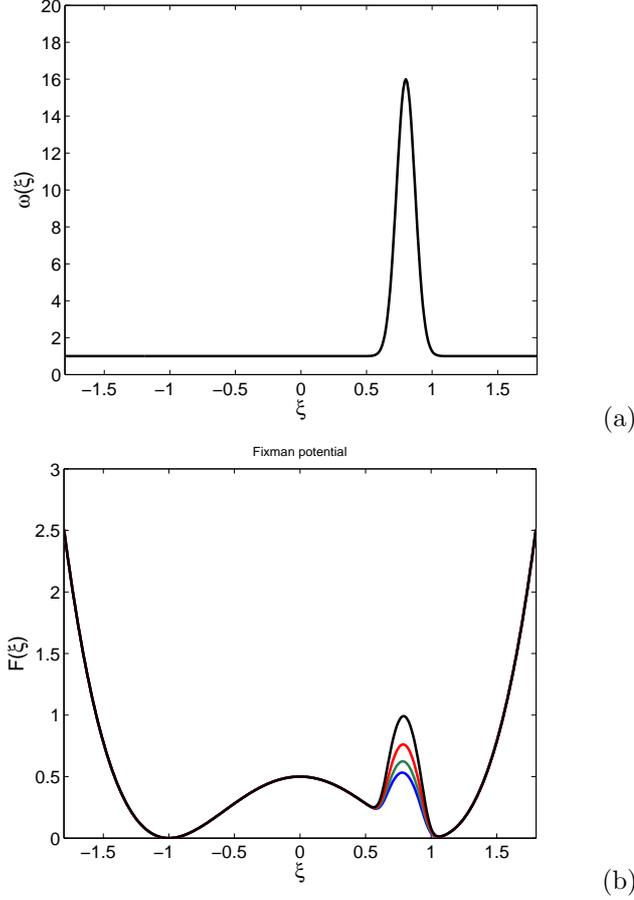


Figure 4. The oscillation frequency $\omega(\xi)$ and the free energy $F(\xi)$ are plotted — the latter for different inverse temperatures $\beta \in \{6.0, 5.0, 4.0, 3.0\}$ with the parameters $C = 15$, $\alpha = 200$, and $\xi_0 = 0.8$. Here $\beta = 3.0$ labels the highest peak at $\xi = \xi_0$, whereas 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

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

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

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

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

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

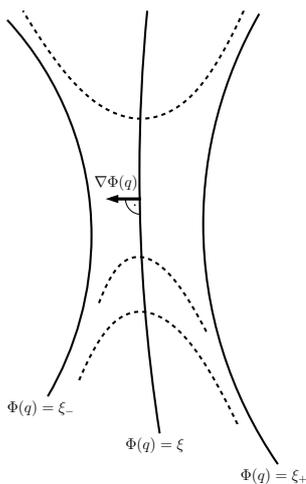


Figure 5. Giving some meaning to the Fixman potential $W = \beta^{-1} \ln \|\nabla\Phi\|$ for codimension-one submanifolds: The plot illustrates the squeezing of nearby level sets. The width of the harmonic confinement potential $W_\epsilon \propto (\Phi(q) - \xi)^2$ in the direction of the normal is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

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

The Blue Moon relation The difference between F and G highlights another important aspect: In the seminal work [178] 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 $f = f(q)$

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

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:

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

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

$$\mathbf{E}_\xi f = \frac{\mathbf{E}_\Sigma (f (\text{vol} J_\Phi)^{-1})}{\mathbf{E}_\Sigma (\text{vol} J_\Phi)^{-1}}, \quad (3.28)$$

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

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

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 $D = -\beta^{-1} \ln \mathbf{E}_\Sigma (\text{vol} J_\Phi)^{-1}$ 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

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

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 $(\text{vol} J_\Phi)^2$; see Figure 5 for illustration. In case the dynamics is ergodic with respect to the canonical density the limit $\epsilon \rightarrow 0$ 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

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

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., [179, 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)

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

with

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

for the standard free energy, and

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

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

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

Accordingly the normalization constant Q_{Σ} reads

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

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 [183] 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 $\epsilon \rightarrow 0$ is constant on the interval $[0, T]$ for any fixed value of $T > 0$,

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

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

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

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 $\epsilon \rightarrow 0$. 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.,

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

If the integral exists, then $x_\epsilon(t) \rightarrow x_0(t)$ 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 [184, 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

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

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 [185, 186],

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

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,

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

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 $\epsilon \rightarrow 0$. This yields the family of equations⁹

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

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 ,

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

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 $x_\epsilon \rightarrow x_0$ is guaranteed by the Averaging Principle for stochastic processes [24, 183]. 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, 187]). 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 $\mathcal{P} = PP^T$ (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 $\mathcal{Q} = QQ^T$. Then $\mathcal{P} + \mathcal{Q} = \mathbf{1}$, and we have a unique decomposition of \mathbf{R}^n due to

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

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

⁹The 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 $\gamma \rightarrow \alpha\gamma$ 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 ,

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

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 $\epsilon \rightarrow 0$ 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*

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

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 $\epsilon \rightarrow 0$ 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 $T > 0, \delta > 0$

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

For the proof the reader is referred to the relevant literature, e.g., [24, 184]. 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

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

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

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

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 $\phi : \mathbf{R}^m \rightarrow \mathbf{R}^k$, 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 $\epsilon \rightarrow 0$.

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

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

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 extent the approach presented here can be considered a variant of the *accelerated dynamics* or *metadynamics* that is put forward in [13]; cf. also [188]. 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 $V : \mathbf{R}^n \rightarrow \mathbf{R}$ be a smooth potential that is bounded from below, and let $\sigma > 0$ be scalar. The Smoluchowski equation reads

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

Given a reaction coordinate $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^s$, the level sets of which define smooth configuration submanifolds of codimension s , we denote by $\sigma_\xi : \mathbf{R}^{n-s} \rightarrow \Sigma_\xi$ 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 $(n_1(\sigma_\xi), \dots, n_s(\sigma_\xi))$, and we introduce local coordinates z^α , $\alpha = 1, \dots, n-s$ on Σ_ξ , and normal coordinates y^i , $i = 1, \dots, s$ that measure the distance to Σ_ξ with respect to the normal frame $\{n_1, \dots, n_s\}$. 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 $N\Sigma_{\xi,\epsilon}$ into $\mathbf{R}^n \times \mathbf{R}^n$. In local coordinates the Smoluchowski equation becomes (see Lemma 2.11)

$$\begin{aligned}\dot{y}_\epsilon^i &= -g_\xi^{il}(z_\epsilon, y_\epsilon) \partial_l V_\xi(z_\epsilon, y_\epsilon) + b_\xi^i(z_\epsilon, y_\epsilon) + \sigma a_\xi^{il}(z_\epsilon, y_\epsilon) \dot{W}_l \\ \dot{z}_\epsilon^\alpha &= -\frac{1}{\epsilon} g_\xi^{\alpha l}(z_\epsilon, y_\epsilon) \partial_l V_\xi(z_\epsilon, y_\epsilon) + \frac{1}{\epsilon} b_\xi^\alpha(z_\epsilon, y_\epsilon) + \frac{\sigma}{\sqrt{\epsilon}} a_\xi^{\alpha l}(z_\epsilon, y_\epsilon) \dot{W}_l.\end{aligned}$$

Note that the equations are only meaningful up to the first exit time from $N\Sigma_{\xi,\epsilon}$. Moreover we have employed the following notation: $V_\xi = V \circ \phi_\xi$, and the function $b_\xi^h = -\beta^{-1} g_\xi^{kl} \Gamma_{\xi,kl}^h$ denotes the additional Itô drift term, whereas a_ξ^{kl} are the entries of the uniquely defined positive-definite matrix square root of g_ξ^{-1} . 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

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

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

$$\begin{aligned}\dot{y}_\epsilon^i &= -\epsilon g_\xi^{il}(z_\epsilon, y_\epsilon) \partial_l V_\xi(z_\epsilon, y_\epsilon) + \epsilon b_\xi^i(z_\epsilon, y_\epsilon) + \sigma \sqrt{\epsilon} a_\xi^{il}(z_\epsilon, y_\epsilon) \dot{W}_l \\ \dot{z}_\epsilon^\alpha &= -g_\xi^{\alpha l}(z_\epsilon, y_\epsilon) \partial_l V_\xi(z_\epsilon, y_\epsilon) + b_\xi^\alpha(z_\epsilon, y_\epsilon) + \sigma a_\xi^{\alpha l}(z_\epsilon, y_\epsilon) \dot{W}_l.\end{aligned}$$

Following [183] we obtain convergence to the initial value $y_\epsilon(t) \rightarrow y_0$ as $\epsilon \rightarrow 0$, 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

$$\dot{z}^\alpha = -G_\xi^{\alpha\beta}(z) \partial_\beta V_\xi(z, 0) + b_\xi^\alpha(z, 0) + \sigma E_\xi^{\alpha\beta}(z) \dot{W}_\beta,$$

where

$$b_\xi^\alpha(z, 0) = -\beta^{-1} G_\xi^{\beta\gamma}(z) \Gamma_{\xi,\beta\gamma}^\alpha(z, 0).$$

Here the $\Gamma_{\xi,\beta\gamma}^\alpha$ 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_ξ^{-1} . All other terms vanish at $y = 0$ since both $g_\xi^{\alpha i} = 0$ and $\Gamma_{\xi,ij}^\alpha = 0$. 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,

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

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

$$f_\xi^i(z, y) = -g_\xi^{il}(z, y) \partial_l V_\xi(z, y) + b_\xi^i(z, y) + \sigma a_\xi^{il}(z, y) \dot{W}_l.$$

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

$$\bar{f}^i(\xi) = \int \left(b_\xi^i(z, 0) - g_\xi^{il}(z, 0) \partial_l V(z, 0) + \sigma a_\xi^{il}(z, 0) \dot{W}_l \right) \nu_\Sigma(dz).$$

¹⁰Note 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 $1, \dots, s$. 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

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

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 $P_T : T_\sigma \mathbf{R}^n \rightarrow T_\sigma \Sigma_\xi$ 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 [189]

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

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 that 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 [190] 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, \dots, 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

$$m(\xi) = \int m_\xi(z) \nu_\Sigma(dz). \quad (3.45)$$

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

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

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

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

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)

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

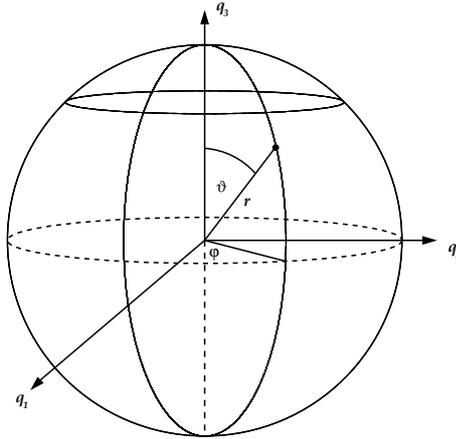


Figure 6. Spherical polar coordinates $(\varphi, \vartheta, r) \in S^2 \times \mathbf{R}_+$.

The additional term b is the usual Itô equation drift

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

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

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

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

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

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

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

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

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

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 $b^l = \beta^{-1} h^{jk} \Gamma_{jk}^l$. 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 $r_\epsilon \rightarrow r$ 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

$$\Gamma_{rr}^\varphi = \Gamma_{rr}^\vartheta = 0.$$

That is, the additional Itô drift $b^\alpha = -\beta^{-1} G^{\gamma\delta} \Gamma_{\gamma\delta}^\alpha$ 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

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

and the Christoffel symbols

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

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

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

where the averaged potential is given by

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

We can obtain the same limit result by using the local embedding $N\Sigma \subset \mathbf{R}^3 \times \mathbf{R}^3$ with $\Sigma = S_\xi^2$. 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 $\sigma_\xi : S^2 \rightarrow S_\xi^2 \subset \mathbf{R}^3$ is given by polar coordinates with fixed radius $r = \xi$

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

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_ξ^2 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

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

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 $P_T : T_\sigma \mathbf{R}^n \rightarrow T_\sigma S^2$, $P_T = \mathbf{1} - n \langle n, \cdot \rangle$ 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

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

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

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

where the averaged potential is given by

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

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

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

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

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

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 [191, 192]. 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 $L : TNS^2 \rightarrow \mathbf{R}$ 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

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

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 :

$$\begin{aligned}\dot{\omega}_r^\alpha(t) &= \zeta_r^\alpha(t) \\ \dot{\zeta}_r^\alpha(t) &= -\Gamma_{\beta\gamma}^\alpha \zeta_r^\beta(t) \zeta_r^\gamma(t) - 2\Gamma_{\beta r}^\alpha \zeta_r^\beta(t) p.\end{aligned}\tag{3.54}$$

Note that since $\Gamma_{\beta r}^\alpha \neq 0$, 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.,

$$\ddot{\omega}_r^\alpha(t) = -\Gamma_{\beta\gamma}^\alpha \dot{\omega}_r^\beta(t) \dot{\omega}_r^\gamma(t).$$

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

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

with the abbreviations

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

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,

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

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 $\Gamma_{\varphi\varphi}^r = -r \sin^2 \vartheta$ and $\Gamma_{\vartheta\vartheta}^r = -r$. 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 $\Gamma_{\varphi r}^\varphi = \Gamma_{\vartheta r}^\vartheta = 1/r$ 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 $\mu_{\text{can}}(dz)$ denote the Gibbs measure on the phase space $E = T^*\mathbf{R}^n$. We introduce the weighted Hilbert space

$$L^2(\mu) = \left\{ v : E \rightarrow \mathbf{R} \mid \int_E v(z)^2 \mu_{\text{can}}(dz) < \infty \right\}$$

that is endowed with an appropriately weighted scalar product

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

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

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

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]

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

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,

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

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

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

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

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

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 [194, 195].

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

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

¹¹We 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, 193]). 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]

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

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, \dots, z^n

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

This transformation is non-singular, for $\det \mathbf{D}\psi = \text{vol} J_{\Phi}$ 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 $d\mathcal{L}_\xi(r, s) = \sqrt{\det G_\xi(r)} dr ds$ 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

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

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

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

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

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

where

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

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

¹²Note 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 $r = q^1, \dots, q^{n-k}$ as local coordinates on Σ .

¹³This 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 $r = q^1, \dots, q^{n-k}$ 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

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

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

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

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

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

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

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

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, 196, 52].

Let us consider the problem how phase space functions evolve in time. To this end consider a Hamiltonian $H : T^*\mathbf{R}^n \rightarrow \mathbf{R}$ 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., $\Psi_t : E \rightarrow E$, $E = T^*\mathbf{R}^n$ is the Hamiltonian flow map). For our purposes it is convenient to cast Hamilton's equations in the form

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

Given a function $f_0 : E \rightarrow \mathbf{R}$, 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

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

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,

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

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 $\Pi : L^2(\mu) \rightarrow L^2(\mu)$ 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, 197] and define S_t^\perp as the propagator of

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

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 [198], which results in a Volterra integral equation for the orthogonal dynamics $w(z, t)$,

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

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

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

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

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

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., [199]. 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

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

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,

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

which is known by the name of Schrödinger representation. According to [200] the adjoint semigroup is generated by the adjoint Liouvillian $\mathcal{L}^* = -\mathcal{L}$, i.e., $T_t^* = T_{-t}$. Noting that $\mathbf{E}_S f = \langle f_0, T_{-t} \rho_0 \rangle$ 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 $\mathbf{E}_H = \mathbf{E}_S$

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

$$\begin{aligned} \partial_t \varphi_1(z, t) &= \mathbf{E}_{\xi, \eta} \varphi_3(z, t) \\ \partial_t \varphi_3(z, t) &= -\nabla G(\varphi_1(z, t)) + \int_0^s K(t-s) w(z, s) ds + w(z, t), \end{aligned} \quad (3.71)$$

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,

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

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 $\eta(t) = (\mathbf{E}_{\xi,\eta}\varphi_3)(\xi, \eta, t)$, 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:

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

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 [201] 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

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

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 [202] the formal solution of (3.68) is¹⁴

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

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

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

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

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

¹⁴Of 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., $u_k(z, h) \rightarrow w(z, h)$ for $h \in [0, \tau]$ as $k \rightarrow \infty$. 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 [203, 204]. 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

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

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. [205, 206]. 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

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

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 [207])

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

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, 208, 209]; see also [209]. Regarding stability of the solutions with respect to perturbations of the (unresolved) initial conditions we refer to the excellent survey article [202] 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 [210]; 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, 211]. 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 [212].

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 [213] 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, \dots, x^{n-s} , and assume that the dynamics stays close to this subspace over a finite time interval. Given a local orthonormal frame $\{n_1(\sigma(x)), \dots, n_s(\sigma(x))\}$ 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

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

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

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

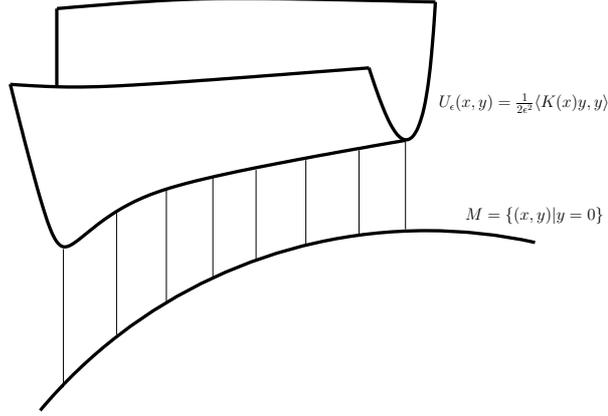


Figure 7. Schematic plot of the confining potential.

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 be 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 $\epsilon \rightarrow 0$ for the Smoluchowski equation. Let $V_\epsilon : \mathbf{R}^n \rightarrow \mathbf{R}$ be the modelling potential. Then for $\beta > 0$ the Smoluchowski equation on \mathbf{R}^n reads

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

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 $\sigma : \mathbf{R}^{n-s} \rightarrow M \subset \mathbf{R}^n$. As before we introduce local coordinates x^α , $\alpha = 1, \dots, n-s$ on M , and normal coordinates y^i , $i = 1, \dots, s$ that measure the distance to M with respect to the normal frame $\{n_1, \dots, n_s\}$. In terms of the local coordinates the Smoluchowski equation becomes according to Lemma 2.11

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

where $b^h = -\beta^{-1} g^{kl} \Gamma_{kl}^h$ denotes the additional Itô drift term with the symmetric Christoffel symbols Γ_{kl}^h , 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 $\epsilon \rightarrow 0$ 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*

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

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

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

with the Christoffel symbols $\Gamma_{\gamma\delta}^\alpha(x) = \Gamma_{\gamma\delta}^\alpha(x, 0)$ 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

$$\begin{aligned} \dot{x}_\epsilon^\alpha &= -\frac{1}{\epsilon} g_\epsilon^{\alpha j} \partial_j V_1 - g_\epsilon^{\alpha\beta} \partial_\beta V_1 + b_\epsilon^\alpha + a_\epsilon^{\alpha l} \dot{W}_l \\ \dot{z}_\epsilon^i &= -\frac{1}{\epsilon^2} g_\epsilon^{ij} \partial_j V_1 - \frac{1}{\epsilon} g_\epsilon^{i\beta} \partial_\beta V_1 + \frac{1}{\epsilon} b_\epsilon^i + \frac{1}{\epsilon} a_\epsilon^{il} \dot{W}_l, \end{aligned} \quad (3.79)$$

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

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

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 $\dot{W}(t) \mapsto \epsilon^{-1} \dot{W}(\epsilon^2 t)$:

$$\begin{aligned} \dot{x}^\alpha &= -\epsilon^2 g_\epsilon^{\alpha j} \partial_j V_1 - \epsilon^2 g_\epsilon^{\alpha\beta} \partial_\beta V_1 + \epsilon^2 b_\epsilon^\alpha + \epsilon a_\epsilon^{\alpha l} \dot{W}_l \\ \dot{z}_\epsilon^i &= -g_\epsilon^{ij} \partial_j V_1 - \epsilon g_\epsilon^{i\beta} \partial_\beta V_1 + \epsilon b_\epsilon^i + a_\epsilon^{il} \dot{W}_l. \end{aligned}$$

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

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

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

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

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

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

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

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

Since $\mu_x(dz)$ does not depend on ϵ , and the integrand is uniformly continuous in z we may interchange the limit $\epsilon \rightarrow 0$ 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

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

and into a remainder that gives

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

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 $\text{grad } \bar{V} = G^{-1} \nabla \bar{V}$ 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 $x_\epsilon(t) \rightarrow x(t)$ is a straight consequence of the Averaging Principle for non-degenerate diffusion processes [24]. (See also the recent paper [17] for a convergence proof.) \square

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

¹⁶Moreover 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

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

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, 214]). We support our conjectures by suitable numerical examples below.

We consider an ϵ -family of Lagrangians $L_\epsilon : TNM \rightarrow \mathbf{R}$ 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

$$\begin{aligned} \dot{x}_\epsilon^\alpha &= u_\epsilon^\alpha \\ \dot{u}_\epsilon^\alpha &= -\Gamma_{kl}^\alpha(x_\epsilon, y_\epsilon)w_\epsilon^k w_\epsilon^l - g^{\alpha l}(x_\epsilon, y_\epsilon)\partial_l V_\epsilon(x_\epsilon, y_\epsilon) \\ \dot{y}_\epsilon^i &= v_\epsilon^i \\ \dot{v}_\epsilon^i &= -\Gamma_{kl}^i(x_\epsilon, y_\epsilon)w_\epsilon^k w_\epsilon^l - g^{il}(x_\epsilon, y_\epsilon)\partial_l V_\epsilon(x_\epsilon, y_\epsilon) \end{aligned} \tag{3.81}$$

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 $\epsilon \rightarrow 0$. The thus scaled equations of motion are

$$\begin{aligned} \dot{x}_\epsilon^\alpha &= u_\epsilon^\alpha \\ \dot{u}_\epsilon^\alpha &= -\Gamma_{\epsilon,kl}^\alpha w_\epsilon^k w_\epsilon^l - g_\epsilon^{\alpha\gamma}\partial_\gamma V_1 - \frac{1}{\epsilon}g_\epsilon^{\alpha j}\partial_j V_1 \\ \dot{y}_\epsilon^i &= \frac{1}{\epsilon}v_\epsilon^i \\ \dot{v}_\epsilon^i &= -\Gamma_{\epsilon,kl}^i w_\epsilon^k w_\epsilon^l - g_\epsilon^{i\alpha}\partial_\alpha V_1 - \frac{1}{\epsilon}g_\epsilon^{ij}\partial_j V_1 \end{aligned} \tag{3.82}$$

with the same abbreviation as before: $g_\epsilon = g(x, \epsilon z)$ and $\Gamma_{\epsilon,kl}^h = \Gamma_{kl}^h(x, \epsilon z)$. 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

$$\nu^\epsilon(dr, ds) = Z_\epsilon^{-1} \exp(-\beta E_\epsilon(r, s)) \det g_\epsilon(r) dr ds.$$

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

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

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

$$\begin{aligned} \dot{x}_\epsilon^\alpha &= \epsilon u_\epsilon^\alpha \\ \dot{u}_\epsilon^\alpha &= -\epsilon \Gamma_{\epsilon,kl}^\alpha w_\epsilon^k w_\epsilon^l - \epsilon g_\epsilon^{\alpha\gamma}\partial_\gamma V_1 - g_\epsilon^{\alpha j}\partial_j V_1 \\ \dot{y}_\epsilon^i &= v_\epsilon^i \\ \dot{v}_\epsilon^i &= -\epsilon \Gamma_{\epsilon,kl}^i w_\epsilon^k w_\epsilon^l - \epsilon g_\epsilon^{i\alpha}\partial_\alpha V_1 - g_\epsilon^{ij}\partial_j V_1. \end{aligned}$$

Sending $\epsilon \rightarrow 0$ and exploiting that $g_\epsilon^{\alpha j} \rightarrow 0$ in the equation for u^α , since the off-diagonal entries of the inverse metric g_ϵ^{-1} vanish, we have $\dot{x}(t) \rightarrow 0$ and $\dot{u}(t) \rightarrow 0$. 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

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

with the conditional normal energy

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

Observing that $\Gamma_{ij, \epsilon}^\alpha \rightarrow 0$ as $\epsilon \rightarrow 0$, 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

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

with the Christoffel symbols $\Gamma_{\gamma\delta}^\alpha$ 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

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

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*

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

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 $\sigma : \mathbf{R}^{n-s} \rightarrow M \subset \mathbf{R}^n$, 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 [215] for hypo-elliptic diffusion processes; see also [216, 217]. 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 $H : T^*\mathbf{R}^2 \rightarrow \mathbf{R}$

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

with the potential

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

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

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

As $\epsilon \rightarrow 0$ 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

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

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

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

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. [218]).) On the slow timescale $t \mapsto \epsilon t$ we find that the fast dynamics alone is given by

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

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

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

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

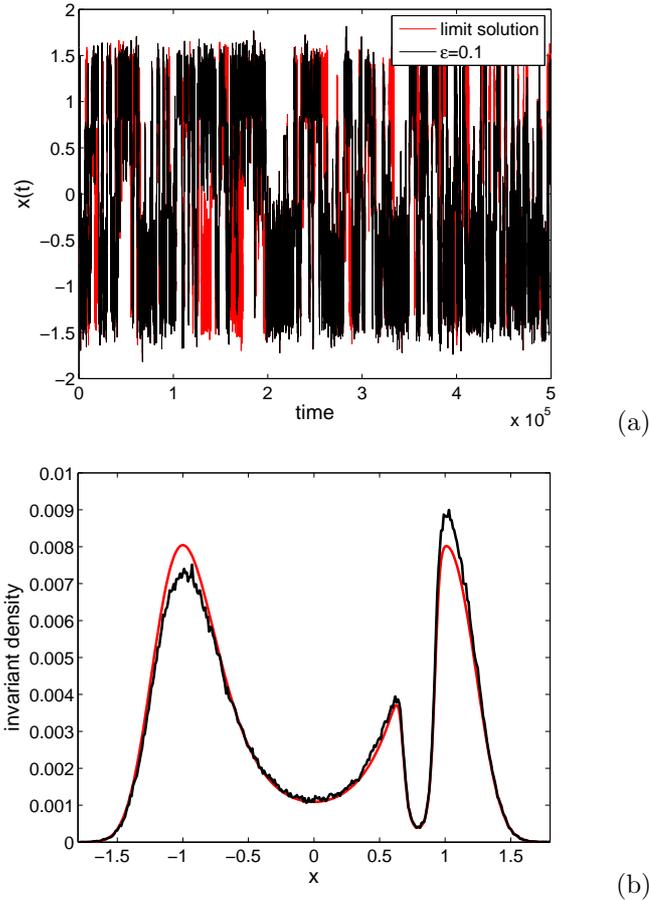


Figure 8. Long-term behaviour of the solution of (3.85) for $\epsilon = 0.1$ versus the limit solution. Upper panel: Typical hybrid Monte-Carlo (HMC) realization for $\beta = 4.0$ and 100 integration steps between the HMC points. Lower panel: invariant density of the slow dynamics computed from 500 000 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 [178]. Note that the result is similar to the results in classical mechanics [179, 75, 27], and it is well-known [218] 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

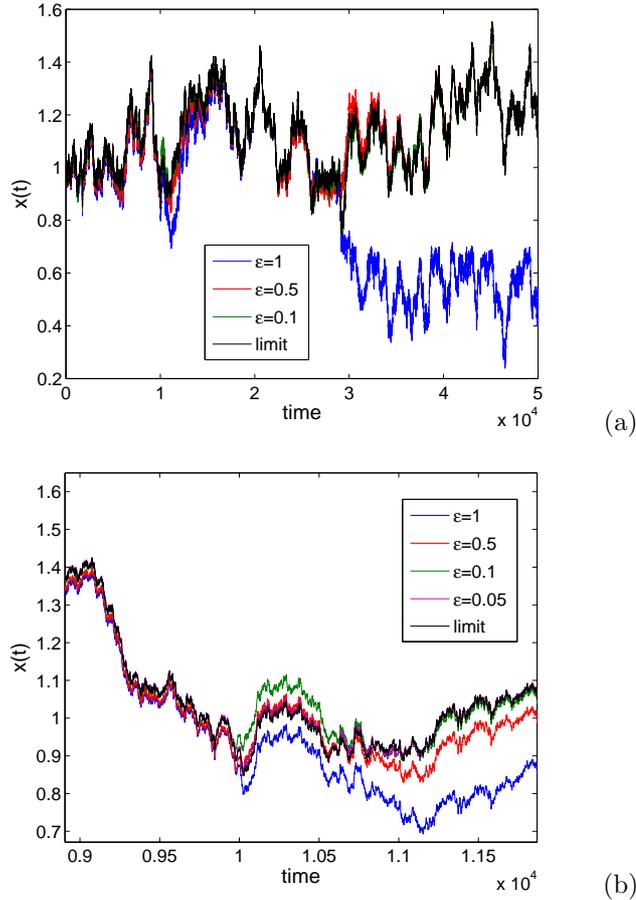


Figure 9. The two plots illustrate convergence of the full system of equations (3.85) towards the limit system (3.86). It can be seen that the error for a typical realization of a HMC trajectory is maximum at the dynamical barrier $x = x_0$; for $\epsilon = 1$ it even happens that the full dynamics makes 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 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., $W = \beta^{-1} \ln \text{vol} J_\Phi$. To this end we remind the reader that $\text{vol} J_\Phi = \sqrt{\det \mathbf{D}\Phi^T \mathbf{D}\Phi}$ for a function $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^s$, and we consider a free Hamiltonian system onto which a

constraint $\Phi(q) = \xi$ is imposed by adding the confining potential

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

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 $\epsilon \rightarrow 0$. 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 $\omega_i^{\dot{}}(X)$ 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 $\omega_i^{\dot{}}(X)$ 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 $\epsilon \rightarrow 0$ the $I = (I_1, \dots, I_m)$ become first integrals of the resulting vector field, where the condition $I = I_0$ singles out an invariant torus \mathbf{T}^m with coordinates $\varphi = (\varphi_1, \dots, \varphi_m)$. 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

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

and is independent of φ_0 . The classical Averaging Principle of Neishtadt [219] 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 $t \rightarrow \infty$ 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

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

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 [220]. Enforcing the non-resonance condition (3.89) and letting ϵ tend to zero, one obtains an averaged system that is Hamiltonian with the energy [221]

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

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

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

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 [222, 223, 224]). 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 $s = 2$:

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

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 $c \rightarrow 0$ 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

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

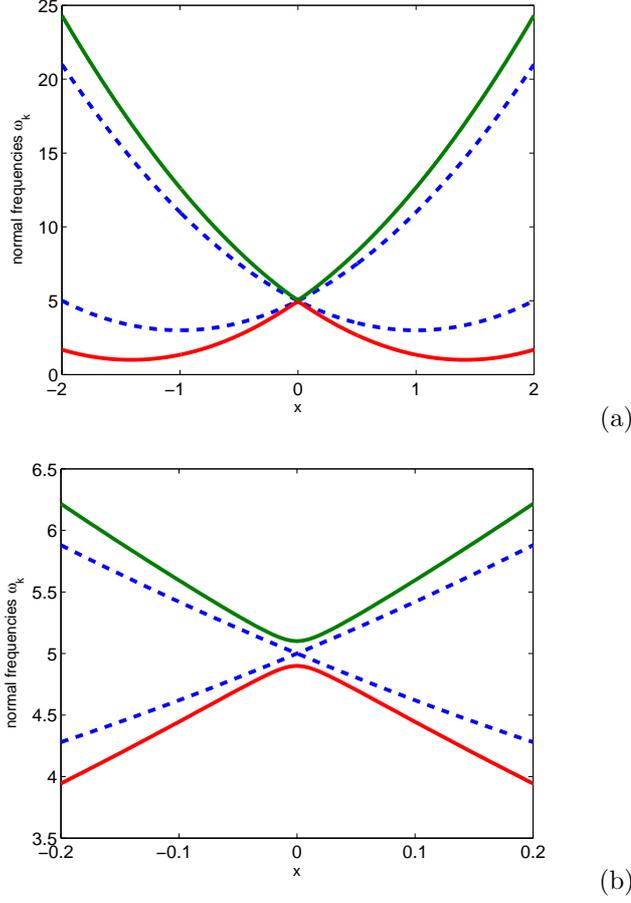


Figure 10. Eigenvalues of the matrix A : the dotted blue lines show a_1 and a_2), whereas the red and green curves show the eigenvalues $\lambda_1 = \omega_1^2$ and $\lambda_2 = \omega_2^2$. As the zoom in the lower panel illustrates the eigenvalues exhibit an *avoided crossing* at $x = 0$ with frequency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

with the frequencies $\omega_i(x) = \sqrt{\lambda_i(x)}$ from above and the action variables [221]

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

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 $\epsilon \rightarrow 0$ we expect that the action variables uniformly converge to the adiabatic invariants, $I_k^\epsilon \rightarrow J_k$. 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

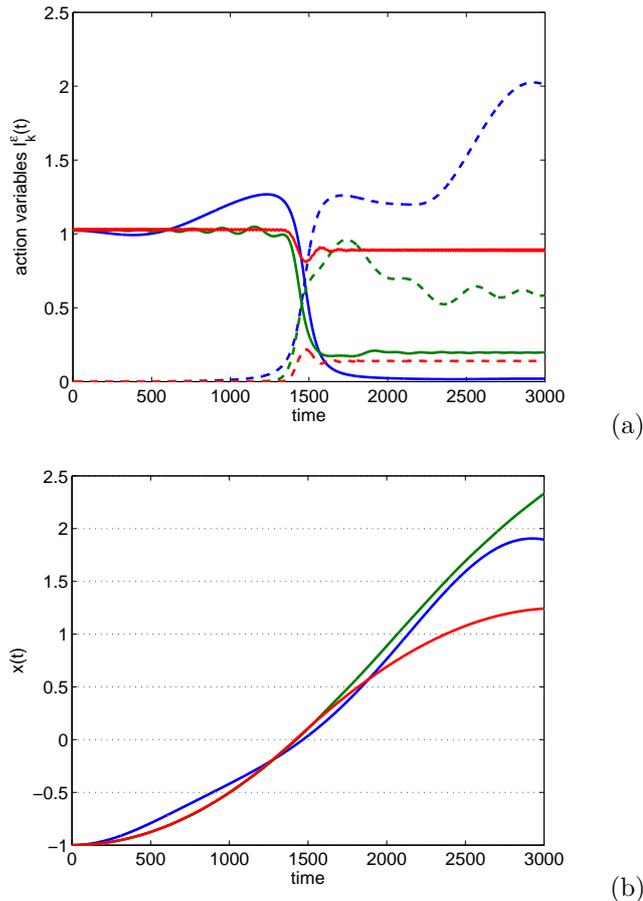


Figure 11. Jump of the action variables $I_k^\epsilon(t)$ at the avoided crossing. It can be seen that jumps occur exactly when the dynamics reaches $x = 0$. The plots show the dynamics of the $I_k^\epsilon(t)$ for $\epsilon = 1$ (blue), $\epsilon = 0.1$ (green), and $\epsilon = 0.01$ (red). All numerical simulations were carried out at constant step-size $h = 0.0002$ 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 [225, 226].

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

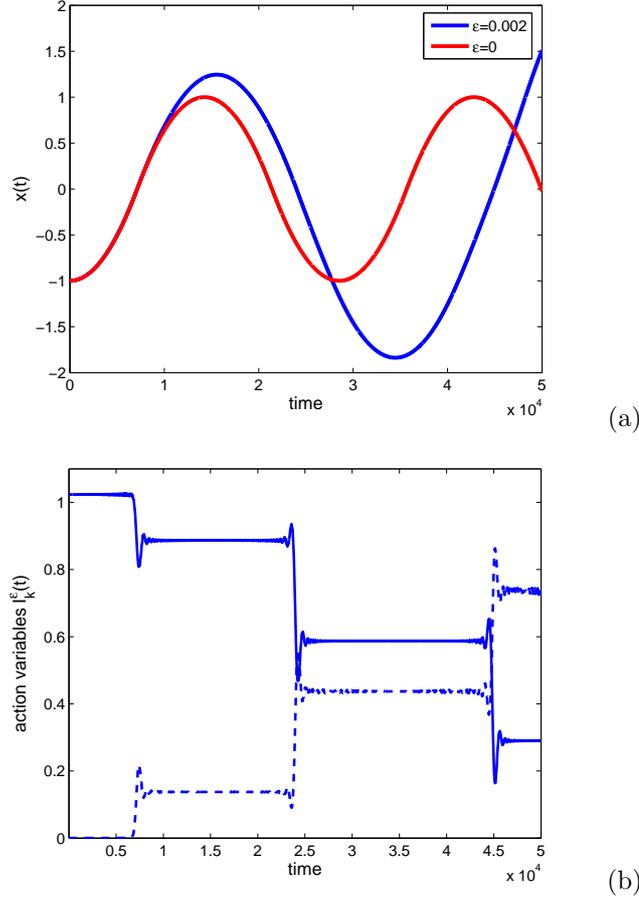


Figure 12. The approximation of the full solution with $\epsilon = 0.002$ by the limit solution $\epsilon = 0$ becomes worse each time the system passes through a resonance (crossing). The integration was carried out with step-size $h = 10^{-5}$ and initial values $(x(0), y^1(0), y^2(0), u(0), v^1(0),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

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 [227] 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

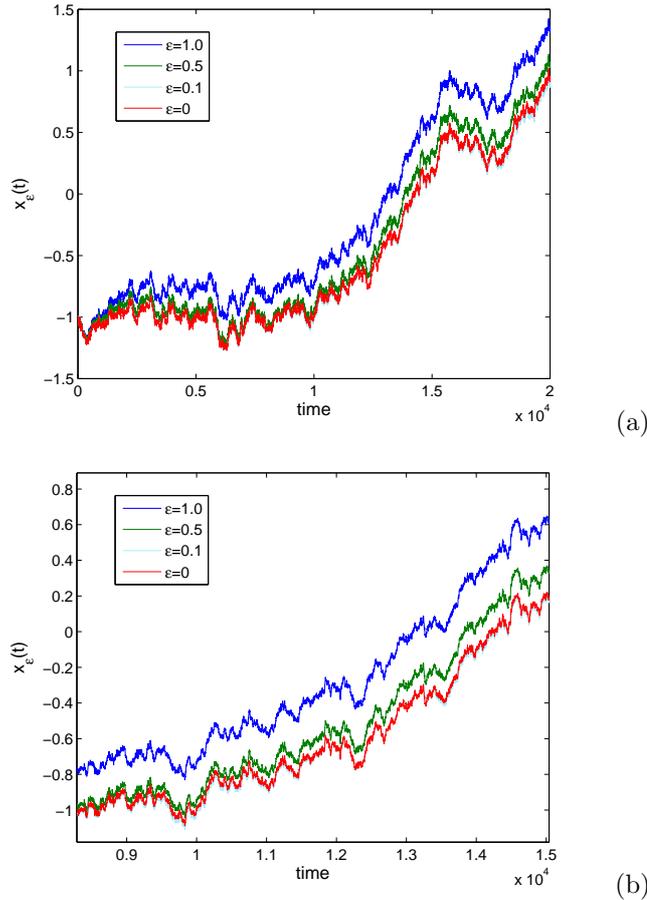


Figure 13. Typical realization for the stochastic Hamiltonian system associated with (3.90) for moderate coupling $c = 0.1$. The simulations were performed using a hybrid Monte-Carlo (HMC) scheme at temperature $\beta = 3.0$ with step-size $h = 0.0005$ for the Leapfrog integrator choosing new momenta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

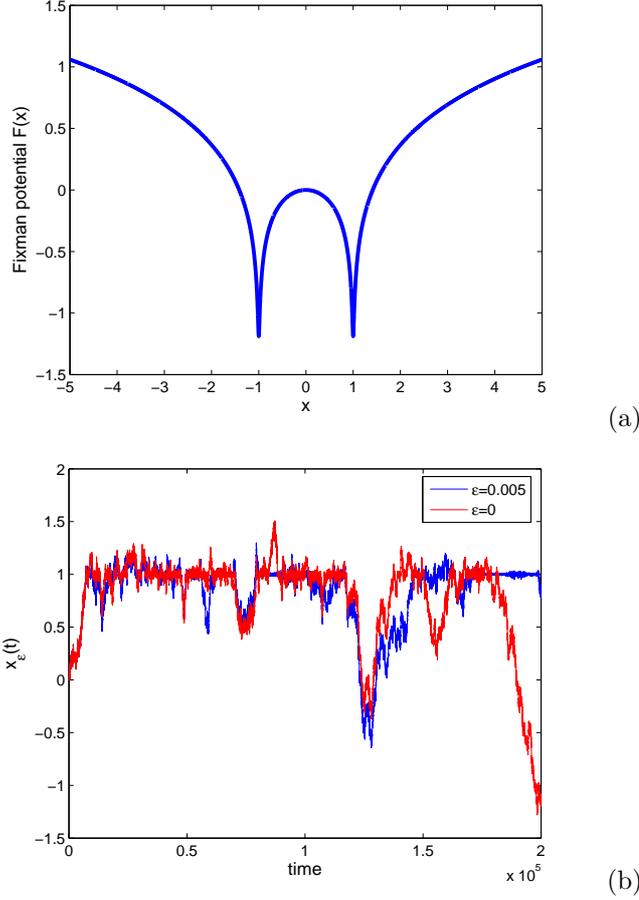


Figure 14. HMC simulation for weak coupling $c = 0.0001$ and $\epsilon = 0.05$. The upper panel shows the corresponding Fixman potential for $\beta = 3.0$.

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

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

We are interested in the quasi-deterministic limit $\gamma, \sigma \rightarrow 0$ 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).

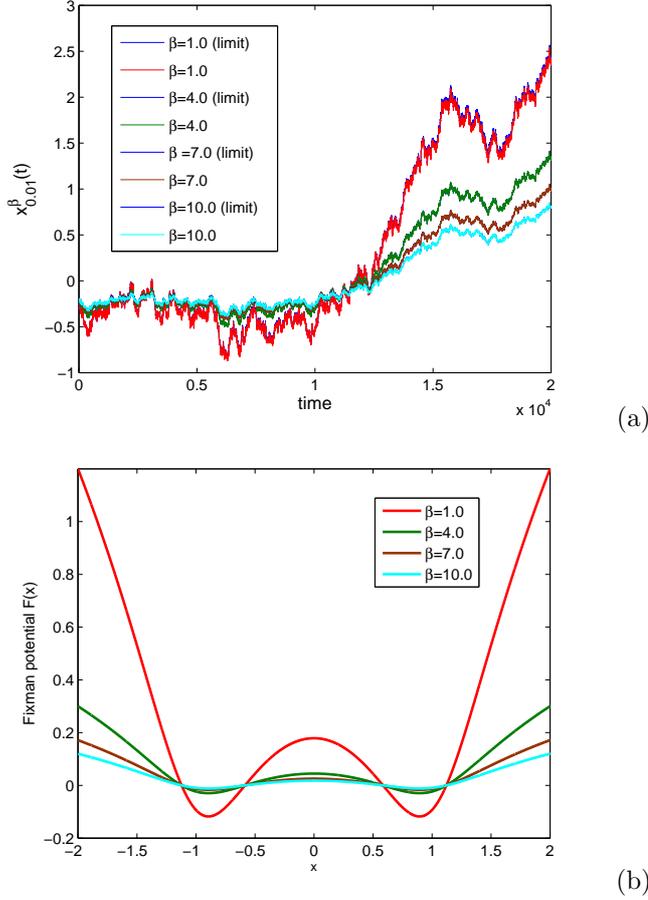


Figure 15. Typical HMC simulations for $c = 0.1$ and $\epsilon = 0.01$ at various temperatures. Note that the limit and the full trajectories are virtually indistinguishable. The lower panel shows the respective Fixman potentials.

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

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

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, $\delta \rightarrow 0$ 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 $\epsilon \sim \delta$.

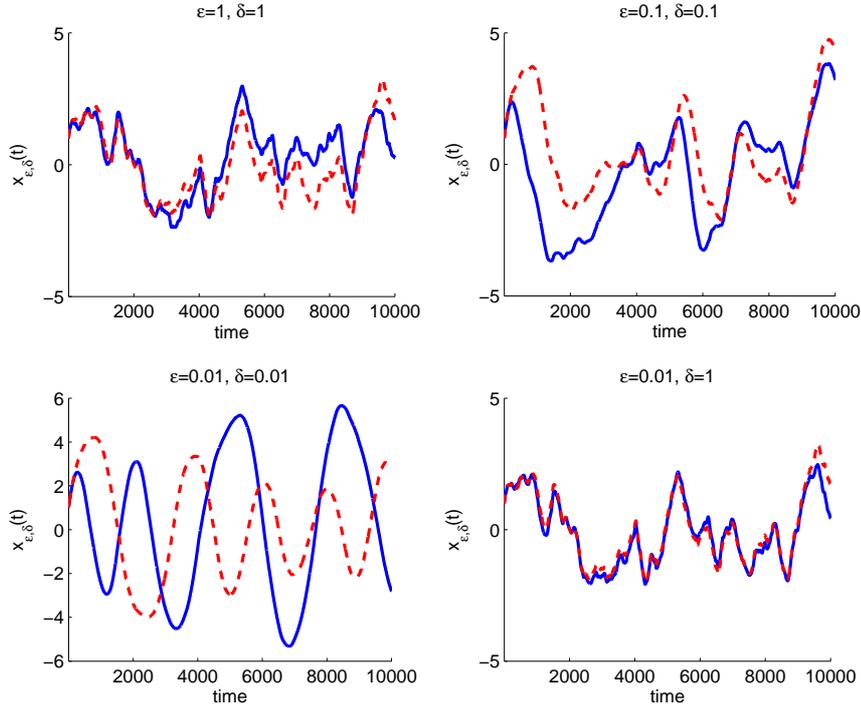


Figure 16. Typical realizations of the slow variable $x_{\epsilon, \delta}$ for the two-parameter Langevin equation (3.92) with coupled parameters $\epsilon \sim \delta$ (blue curves: full system, red curves: limit dynamics for $\epsilon = 0$). The realizations indicate that for $\epsilon, \delta \rightarrow 0$ the averaged dynamics with the Fixman potential does no longer approximate the full (quasi-deterministic) system. In contrast, taking the limit $\epsilon \rightarrow 0$ 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 $\epsilon, \delta \rightarrow 0$, even far away from the avoided crossing.

Indeed the realizations shown in Figure 16 indicate that for $\epsilon, \delta \rightarrow 0$ 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 $\epsilon \rightarrow 0$ 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 [228] 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 $\epsilon \rightarrow 0$, 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

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

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 $\mathbf{D}_2g(x, y, 0) \neq 0$ 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 system¹⁷

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

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 [229, 230] 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 [231], 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 [232] 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

¹⁷We call M uniformly (hyperbolic) asymptotically stable, if and only if all eigenvalues of the Jacobian $\mathbf{D}_2g(x, h(x), 0)$ have negative real parts and are uniformly bounded away from zero.

covariance matrix aa^T of the noise it has been shown recently [129] 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

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

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 $\epsilon \rightarrow 0$, and therefore the estimation for the approximation error becomes useless. For the technical intricacies we refer to [129, 233].

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

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

with the sharply-peaked frequency (see Figure 4)

$$\omega(x) = 1 + C \exp(-\alpha(x - x_0)^2). \quad (3.98)$$

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

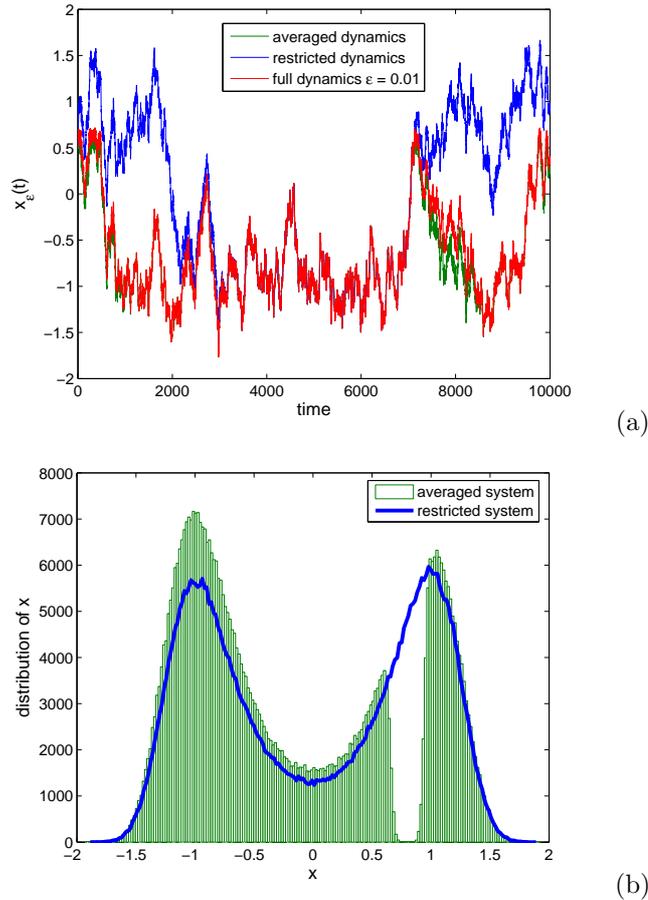


Figure 17. The upper panel shows typical realizations of the slow-fast Smoluchowski equation (3.97) versus the averaged and its restricted limit equation. The integration was performed using an Euler-Maruyama scheme with step-size $h = 10^{-4}$ and initial values $(x(0), z(0)) = (x_0, 0)$ as is consistent 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

¹⁸In 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 [234].

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 $PP^T z \in M$, and we can introduce local coordinates $x = P^T z$ on M . The Galerkin projection then consists in the projection of the full system

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

onto the tangent space of M . That is,

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

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 [162]. 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 $f = -\text{grad} V$ 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 $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ 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,

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

or

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

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, 235].

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 [178, 71],

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

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

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

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

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

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

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

where $\text{grad} G = m^{-1} \nabla G$ 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}_ξ 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

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

with Γ_{jk}^i denoting the symmetric Christoffel symbols associated with the Riemannian metric m . We emphasize that the derivation of the reduced 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):

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

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 $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ denote again a reaction coordinate with (yet unknown) conjugate momentum $\Theta : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^k$. Then the conditional expectation

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

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

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

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, 197, 236]) 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 $\Pi = \mathbf{E}_{\xi, \eta}$, $\Pi : L^2(\mu) \rightarrow L^2(\mu)$. 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 [201]. 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. [237, 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 $\epsilon \rightarrow 0$ 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 $\epsilon \rightarrow 0$, 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 [179] and [238]. For a detailed discussion we refer to the textbook [239] 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 $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}^k$, i.e., $M = \varphi^{-1}(0)$. If we impose the constraint $\varphi(q) = 0$ by adding a strong potential,

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

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

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

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 [218] for mechanical systems and in [240, 241] 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].

4. Phase space of the fast variables

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

4.1. Excursus: constrained mechanical systems

In this section we shall briefly discuss the properties of mechanical systems subject to holonomic constraints. In treating constraints it is most convenient to start within the framework of Lagrangian mechanics. For our purposes it suffices to define a holonomic constraint by specifying a submanifold $\Sigma \subset \mathbf{R}^n$ of the configuration space. Together with the natural inclusion $T\Sigma \subset T\mathbf{R}^n$ this determines the state space of the constrained system. Suppose that $\Sigma = \varphi^{-1}(0)$ is determined as the level set of a smooth function $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}^s$. If the Jacobian $\mathbf{D}\varphi(q)$ has maximum rank on Σ , then Σ is a proper submanifold of codimension s in \mathbf{R}^n . The tangent space to $q \in \Sigma$ is then defined in the usual way considering the direction of curves in Σ which is equivalently expressed as [181, 242]

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

For the sake of simplicity, we assume that Σ has codimension $s = 1$ in \mathbf{R}^n . For a much more general discussion of holonomic constraints the interested reader is referred to the textbook [81]; a good introduction into the geometry of submanifolds is [182].

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

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

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

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

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

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

Evidently the second equation is simply the constraint $\varphi(q) = 0$. Accordingly the Euler-Lagrange equations form a differential-algebraic system which is of differential index three [243, 244]. The alternative method by restricting the original Lagrangian to $T\Sigma$ amounts to endowing Σ with an appropriate set of local coordinates (x^1, \dots, x^{n-1}) ,

writing up the Lagrangian in these coordinates, and deriving local Euler-Lagrange equations. These will be then of the form (2.4). According to the theorem on Lagrange multipliers the local Euler-Lagrange equations are equivalent to the equations (4.1). We refer to the latter as *ambient space* formulation. It is by far the most common approach in molecular dynamics, for the equations can be discretized by standard numerical schemes [105, 245].

For related approaches the interested reader may consult the seminal work of Dirac [246], or constrained formulations using vakonomic mechanics [20]. A different method that is more in the spirit of index reduction techniques is treated in [247].

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

This yields the Hamiltonian

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

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

equations of motion for the variables q and p ,

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

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

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

which is the image of the Legendre transform of $(T\mathbf{R}^n)|_{T\Sigma}$ which we can identify with $T^*\Sigma$. Here H is the original (i.e., unconstrained) Hamiltonian, and \mathbf{D}_2 denotes the derivative with respect to the second slot. The condition on the momentum is exactly the condition $\dot{\varphi}(q) = 0$, and is referred to as *hidden* constraint; it is hidden because it does not appear explicitly in the equations of motion. Notice that the identification of \mathcal{B} with $T^*\Sigma$ is a rather subtle issue which is related to the non-regularity of the augmented Lagrangian; in general this identification is valid only up to a symplectic diffeomorphism $p \mapsto p + \alpha \nabla \varphi$, where α is chosen such that p satisfies the hidden constraint; see [248, 249] regarding this discussion.

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

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

which is zero, since $(q(t), p(t))$ is a curve in \mathcal{B} , and hence satisfies the hidden constraint. The last equation is quite important from the viewpoint of numerics, since it states that a numerical discretization scheme of the differential-algebraic system (4.5) should take care of the hidden constraint in order to preserve the energy conservation property of the continuous flow [82, 250]. Furthermore it is obvious from the equations of motion that the constrained system is still reversible in time.

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

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

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

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

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

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

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

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

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

From this we immediately conclude:

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

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

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

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

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

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

$$\nu_{\text{can}} = (\rho_{\text{can}}|_{\mathcal{B}}) d\lambda_{\Sigma}. \quad (4.6)$$

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

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

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

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

with

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

and the partition function

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

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

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

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

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

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

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

Let

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

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

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

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

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

4.2. Sampling constrained invariant measures

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

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

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

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

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

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

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

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

and

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

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

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

and

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

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

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

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

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

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

4.2.2. Constrained hybrid Monte-Carlo The goal of this section is to introduce an alternative to the usual microcanonical sampling methods (Nosé-Hoover, isokinetic ensemble) that may not be ergodic, or standard Monte-Carlo which may be poorly mixing. We adopt the hybrid Monte-Carlo (HMC) technique, which emulates the general Metropolis Monte-Carlo strategy of proposal and acceptance steps, where, however, the proposal is generated by short runs of the Hamiltonian system with randomly chosen initial conditions. This method circumvents the common Monte-Carlo problem, namely, that the acceptance probability of an arbitrary move to an energetically unfavourable state becomes incredibly small. As ordinary Metropolis Monte-Carlo, HMC is conceptually very simple, and is designed to handle symplectic integration, i.e., one can use standard integrators for constrained Hamiltonian systems. Moreover it can be proved that the dynamics is ergodic with respect to the positional density under rather mild conditions which are met for our purposes [83, 251, 252]. As an additional treat the acceptance procedure also controls the numerical error, because HMC rejects those moves that have too large energy fluctuations.

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

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

where $\sigma(x)$ denotes the embedding of Σ into \mathbf{R}^n , and $x = (x^1, \dots, x^d)$ are local coordinates on Σ . Now consider the symplectic and reversible discrete flow map $\Psi_\tau : \mathcal{B} \rightarrow \mathcal{B}$ on the constrained phase space $\mathcal{B} = T^*\Sigma$, and consider iterates of Ψ_τ with initial momenta that are randomly chosen according to the constrained Maxwell distribution $\varrho_x(\cdot)$ in (4.8). This generates a sequence $\{x_0, \dots, x_{N-1}\} \subset \mathbf{R}^d$ in configuration space. Note that if the flow Ψ_τ were exactly energy-preserving, then the x_k would be distributed according to ν_Σ . However it is impossible to find a numerical discretization scheme that is symplectic, reversible, and exactly energy-conserving at the same time as follows from backward error analysis [105]. The best we can achieve is that the energy error remains uniformly bounded on compact time intervals and oscillates around its exact value [253].

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

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

where

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

is the energy error. Accordingly we reject the proposal (i.e., $x_{k+1} = x_k$) with probability $1 - p_\tau$. In this form, HMC yields a configuration sampling, and by repeating the procedure of generating proposals, the resulting HMC Markov chain $\{x_1, \dots, x_N\}$ allows for approximating the conditional expectation, if the system is ergodic [254]. In order to prove ergodicity for the constrained HMC Markov chain we make use of an idea in [83] that rests upon the following strong Law of Large Numbers [255, 256].

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

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

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

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

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

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

Invariance of the constrained Gibbs measure Invariance of the constrained Gibbs measure can be shown following the outline of the proof in [252] for separable Hamiltonians. Here we cannot separate the canonical density into merely momentum and position dependent parts, and so we write

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since the HMC acceptance probability (4.11) is strictly positive, it does not alter the accessibility properties of the Markov chain. Hence, and for the sake of notational convenience, we shall omit it in what follows. Proving the accessibility condition then requires two steps: In a first step we follow the approach in [83] and construct ambient space sample paths that satisfy the accessibility condition in $\Sigma \subset Q$. In doing so, it turns out that the problem boils down to a standard symplectic discretization of constrained systems. In a second step we demonstrate that the ambient space discretization has an equivalent formulation in local coordinates, hence satisfying the accessibility condition (4.10). Regarding the former problem we endeavour a discrete variant of Hamilton's principle of least action. Following [257], we introduce a discrete Lagrangian as a map $L_h : Q \times Q \rightarrow \mathbf{R}$. The discrete counterpart of the classical action is a mapping $S_h : Q^{N+1} \rightarrow \mathbf{R}$, that is defined as the sum

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

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

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

where $\mathbf{D}_1, \mathbf{D}_2$ denote the derivatives with respect to the first and second slot. If $\mathbf{D}_2 L_h$ (the generalized discrete momentum) is invertible, then (4.15) implicitly defines a discrete flow by means of the map $(q_{k+1}, q_k) = \Phi_h(q_k, q_{k-1})$. The particular discretization scheme that leads to (4.14) is open to choice and should depend on the problem; for the details we refer to the seminal work of Marsden and West [257].

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

$$\mathbf{P}[q(\tau) \in B \mid q(0) = q_0] > 0.$$

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

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

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

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

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

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

the discrete Euler-Lagrange equations turn out to be [258]

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

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

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

which are known as the SHAKE algorithm [259]. The Lagrange multiplier λ_k is chosen such as to enforce the constraint at time $k+1$. The discrete conjugate momenta is defined by the discrete Legendre transform of $\hat{L}_h = L_h - \langle \lambda_k, \varphi(q_k) \rangle$, that is,

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

Hence we can consider the SHAKE algorithm as a mapping $\mathcal{B} \rightarrow \mathcal{B}$ (or $T^*\Sigma \rightarrow T^*\Sigma$). It is symplectic by virtue of its variational character (cf. the related work [245, 82]). By choosing initial conditions $q(0) = q_0$ and $p(0) = -\mathbf{D}_1 \hat{L}_h(q_0, q_1, \lambda_0)$ the discrete flow generates a discrete trajectory that connects q_0 and q_τ . Finally, it follows by continuity of Φ_τ on the initial conditions that the endpoints of trajectories with perturbed initial momenta $p_\epsilon(0) = p(0) + \epsilon$ remain in $B \subset \Sigma$ whenever ϵ is sufficiently small. \square

A frequently used variant of the SHAKE algorithm is called RATTLE and goes back to [260]. It can be considered as a constrained version of the ordinary velocity Verlet scheme. SHAKE and RATTLE are equivalent to each other by dint of (4.18). Moreover they are variational with the discrete Lagrangian L_h defined above, and therefore both SHAKE and RATTLE are symplectic.

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

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

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

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

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

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

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

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

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

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

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

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

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

which gives rise to the following discrete Euler-Lagrange equations

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

Solving the equation for x_{k+1} given x_k, x_{k-1} defines a map $\Theta_h : \mathbf{R}^d \rightarrow \mathbf{R}^d$. By computing the conjugate momenta $u_k = -\mathbf{D}_1 L_{\Sigma, h}(x_k, x_{k+1})$ we can augment this map to a symplectic map $\Psi_h : T^*\mathbf{R}^d \rightarrow T^*\mathbf{R}^d$. The following statement is true [261]:

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

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

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

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

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

The last assertion does not say anything about the speed of convergence, which remains an open problem; see [83, 262] for some numerical studies. In particular the speed of convergence depends on the choice of the HMC integration time $\tau = Nh$. Exploring state space becomes certainly faster if τ is increased. However increasing τ while keeping the step-size h constant decreases the acceptance probability, since energy fluctuations become an issue.

Before we conclude the Monte-Carlo section, we shortly mention that the HMC algorithm with lag time $\tau = h$ and without the acceptance-rejection procedure is equivalent to an Euler discretization of the Smoluchowski equation [263]. However letting the acceptance step account for the discretization error, HMC can be regarded as an exact discretization of the Smoluchowski equation at step-size $\tau = h$, i.e., HMC generates a diffusion-like flow. Therefore the algorithm converges for any stable step-size without introducing a bias.

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

Unlike for deterministic dynamics, there are many situations in which stochastic dynamics is proved to be ergodic [115]. This requires that the coefficients in the equations are globally Lipschitz, a condition which is typically not satisfied; in practice, this seems to be no problem whatsoever [264].

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

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

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

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

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

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

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

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

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

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

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

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

with $\Delta W_n = W_{n+1} - W_n$ denoting the increment of the Brownian motion. We can enforce the constraint by introducing an appropriate projection onto the tangent space of Σ which gives rise to the implicit Euler-Maruyama scheme [265]

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

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

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

We emphasize that both discretization schemes are consistent with the constrained Itô equation (4.20). Certainly the implicit scheme will allow for larger step-sizes, but it requires to solve the implicit and nonlinear equation. In turn, the choice of the nonlinear solver will affect the stability of the numerical solution (cf. [82, 266]).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

we have $\alpha = \delta = 0$ and therefore $\varphi(t) = 0$ at all times $t > 0$. Borrowing a denomination from the theory of differential algebraic equations [250], we term (4.27) the *underlying stochastic differential equation* to (4.25).²⁰ It remains to check whether the constrained canonical distribution μ_Σ is invariant under the constrained Langevin dynamics.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

We can perform the momentum integration by application of the divergence theorem [267] for submanifolds of arbitrary codimension; since $T_q^*\Sigma$ is a linear subspace of $T_q^*\mathbf{R}^n \cong \mathbf{R}^n$ and linear subspaces have zero mean curvature, it follows that the remaining forcing/dissipation integral above is zero, i.e.,

$$\int_{\mathcal{B}} \text{div}_\Sigma (\mathbf{D}_2 (u \exp(-\beta H))) d\lambda_\Sigma = 0.$$

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

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

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

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

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

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

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

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

The final momentum step is

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

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

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

The integrator is quasi-symplectic in the sense of [268], and we expect that it is strongly convergent of order two. Preliminary numerical simulations seem to support this claim, but we refrain from detailed numerical studies for the sake of brevity. In fact, a very similar result has appeared recently during the writing of this thesis. Elaborating upon the RATTLE integrator, the authors of [84] obtain an integrator almost identical to (4.29)–(4.32), but with a different implementation of the white noise term; in addition, they prove that the integrator is second-order accurate.

4.3. Thermodynamic Integration

We have presented methods for sampling constrained Gibbs measures in either configuration or phase space. By these means we can now to sample, for example, the derivative of the free energy or other quantities that appear in any of the reduced models. In order to access configuration space regions which correspond to improbable values of the reaction coordinate, we have to resort to methods like Thermodynamic Integration [8, 269] or the closely related Thermodynamic Perturbation [270, 271]. Accordingly this section explains Thermodynamic Integration from the point of view of the different types of constrained dynamics and gives an overview of different methods of free energy calculation. In particular we will explain how both geometric and standard free energy can be computed rather efficiently from the force of constraint.

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

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

where the effective mass m is given by

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

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

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

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

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

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

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

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

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

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

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

where $w_{n,l}$ are the weights of the particular quadrature rule (see, e.g., [272, 273]).

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

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

with the generalized partition function

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

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

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

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

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

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

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

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

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

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

$$\nabla U_\alpha = \int_{\mathcal{B}} \lambda_\alpha \mu_{\Sigma, \alpha}, \quad (4.34)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

which is in perfect agreement with the formulae that have been derived in various instances, e.g., [12, 15, 11, 274].

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

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

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

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

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

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

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

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

with

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

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

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

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

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

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

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

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

with

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

Bibliographical remarks For the sake of completeness we mention just a few other methods that are available in the literature. A Brownian dynamics approach that exploits the relation between the derivative of a generalized free energy and the respective mean constraint force is given in [17]. Another widely-used Monte-Carlo-based algorithm for free energy calculations is Umbrella Sampling [275], where the system is forced to sample a certain range of the reaction coordinate by adding a confining potential. Though easy to implement (and thus popular), Umbrella Sampling involves unphysical manipulations of the original system and uncontrolled sources of error due to the choice of the confining potential [235]. Another class of approaches can be subsumed under the name of *Adaptive Biasing Forces*. These approaches, like conformational flooding [276], scaled force [77], or metadynamics [277], estimate the mean force during the course of integration. While sampling of phase space proceeds, the estimate is progressively refined, and introduced in the equations of motion as

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

The formerly mentioned methods exploit that the free energy is equal to the reversible work that a system performs while undergoing an adiabatic change of state. This requires that the system always stays in its thermodynamic equilibrium conditional on the (frozen) reaction coordinate. But if the reaction coordinate is controlled in such a way that the remaining system cannot relax to its thermodynamic equilibrium, then the amount of performed work typically exceeds the free energy (second law of thermodynamics). Hence the above mentioned algorithms suffer from a systematic overestimation of free energy differences due to finite sampling times. However it is possible to compute free energy differences by averaging the irreversible work using an appropriate exponential weighting which is due to [278, 279]. For applications of the *Jarzynski equality* we refer to the recent preprint [280].

5. Algorithmic issues and numerical examples

Based on the considerations of the Sections 2.4, 3, and 4.2 we shall qualitatively study some of the introduced reduction schemes for two model systems: Ryckaert-Bellemans n -butane and the glycine dipeptide analogue with a GROMOS96 vacuum force field. The latter is a small peptide that contains a central amino acid and which is a popular benchmark system for spectroscopy and conformation dynamics [281, 282]. On the other hand Ryckaert-Bellemans' n -butane molecule is particularly convenient for our purposes, since many properties are known on analytical grounds (e.g., reaction coordinate, torsional free energy).

We have argued (and this is confirmed by numerical simulations) that hybrid Monte-Carlo (HMC) generates a diffusion-like flow, and, in point of fact, we can draw similar conclusion from HMC simulations or simulations of the corresponding Smoluchowski equation. Langevin dynamics is not used in the model studies. The reason is that (i) the dynamics of a Langevin equation can be vastly different for different friction and noise coefficients, and (ii) there is an ongoing discussion about the choice of good integrators that preserve the thermodynamical properties of the systems (e.g., temperature, equipartition of energy, invariant measure). We refer to the monograph [283] and the references therein.

5.1. The constrained hybrid Monte-Carlo algorithm

For the evaluation of constrained and conditional expectation values we confine our attention to the hybrid Monte-Carlo scheme as has been introduced in Section 4.2.2. The reason is twofold: first of all, all quantities of interest can be computed as positional averages, and, secondly, we know for sure that the constrained expectations eventually converge to the correct values (strong law of large numbers).

To this end we briefly explain how the constrained hybrid Monte-Carlo algorithm actually works. It is convenient to switch back to a representation of the equations of motion and the invariant measure in terms of the ambient space variable (q, p) . We shall also drop the mass scaling assumption. Given a symmetric, positive-definite molecular mass matrix $M \in \mathbf{R}^{n \times n}$, and an interaction potential $V : \mathbf{R}^n \rightarrow \mathbf{R}$ that is bounded from below, the unconstrained Lagrangian is defined as

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

The respective unconstrained Hamiltonian thus reads

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

Introducing the reaction coordinate constraint $\Phi(q) = \xi$, the constrained equations of motion (4.1) are then generated by the augmented Lagrangian $\hat{L} = L - \lambda^i (\Phi_i(q) - \xi_i)$. The SHAKE discretization of the equations of motion for a time step $h > 0$ is

$$\begin{aligned} q_{n+1} - 2q_n + q_{n-1} &= -h^2 M^{-1} (\nabla V(q_n) + \mathbf{D}\Phi(q_n)\lambda_n) \\ \xi &= \Phi(q_{n+1}). \end{aligned} \quad (5.1)$$

In the classical formulation of Ryckaert *et al.*, the momentum is approximated by [259]

$$p_n = M \left(\frac{q_{n+1} - q_{n-1}}{h} \right). \quad (5.2)$$

This method has two major drawbacks: First of all, the mapping $(q_n, p_n) \mapsto (q_{n+1}, p_{n+1})$ defined by (5.1)–(5.2) is not symplectic.²² Secondly, the three-term recursion in (5.1) may lead to an accumulation of round-off errors. Therefore the scheme may become unstable, as has been pointed out in [105]. A remedy of both problems is to make the SHAKE algorithm a variational integrator: following [257] we replace (5.2) by the correct discrete conjugate momentum (4.18). This amounts to formulating SHAKE as a one-step method which yields the RATTLE algorithm [260]

$$\begin{aligned}
p_{n+1/2} &= p_n - \frac{h}{2} (\nabla V(q_n) + \mathbf{D}\Phi(q_n)\lambda_n) \\
q_{n+1} &= q_n + hM^{-1}p_{n+1} \\
\xi &= \Phi(q_{n+1}) \\
p_{n+1} &= p_{n+1/2} - \frac{h}{2} (\nabla V(q_{n+1}) + \mathbf{D}\Phi(q_{n+1})\mu_n) \\
0 &= \mathbf{D}\Phi(q_{n+1})M^{-1}p_{n+1},
\end{aligned} \tag{5.3}$$

The Lagrange multipliers λ_n, μ_n are chosen, such that the two constraints are satisfied. The RATTLE integrator (or SHAKE considered as a mapping $T^*\Sigma \rightarrow T^*\Sigma$) is symplectic as following from its variational nature; cf. the related articles [245, 82].²³

Recall that hybrid Monte-Carlo (HMC) requires that we draw a initial momentum from the constrained Maxwell distribution at each Monte-Carlo step, where the constrained Maxwellian depends point-wise on the constrained position variables. This can be understood as follows: consider the unconstrained expression for the kinetic energy in terms of the velocity variables,

$$T(v) = \frac{1}{2} \langle Mv, v \rangle := \frac{1}{2} \langle v, v \rangle_M,$$

where $\langle \cdot, \cdot \rangle_M$ denotes the metric with respect to the mass matrix. We have shown in Section 4.1.3 that the constrained canonical probability distribution is simply the restriction of the unconstrained distribution. In order to restrict the Maxwell density to the constrained tangent space $T_q\Sigma$, $q \in \Sigma$, we define the M -orthogonal projection $P_{M,T} : T_q\mathbf{R}^n \rightarrow T_q\Sigma$ that is defined point-wise for each $q \in \Sigma$:

$$P_{M,T} = \mathbf{1} - M^{-1}J_\Phi(J_\Phi^T M^{-1}J_\Phi)^{-1}J_\Phi^T, \quad J_\Phi = \mathbf{D}\Phi(q).$$

Strictly speaking, $P_{M,T}$ sends vectors $v \in \mathbf{R}^n$ to vectors in $\tilde{v} \in \mathbf{R}^n$, such that \tilde{v} satisfies the hidden constraint $\mathbf{D}\Phi \cdot \tilde{v} = 0$. It can be readily checked that (i) the matrix $P_{M,T}$ meets the idempotency property $P_{M,T}^2 = P_{M,T}$, and that (ii) it is symmetric with respect to the mass-weighted scalar product $\langle \cdot, \cdot \rangle_M$. That is,

$$\langle P_{M,T}u, v \rangle_M = \langle u, P_{M,T}v \rangle_M$$

for any two vectors $u, v \in \mathbf{R}^n$. Hence $P_{M,T}$ is an orthogonal projection with respect to the metric defined by the mass matrix M . Consequently, we shall refer to $P_{M,T}$ as M -orthogonal projection. Since $P_{M,T}$ maps to the constrained velocity space, we obtain the restricted Maxwell density $\exp(-\beta T_\Sigma)$ by restricting the kinetic energy,

$$T_\Sigma(q, v) := T(P_{M,T}v) = \frac{1}{2} \langle P_{M,T}v, v \rangle_M.$$

²²The mapping preserves the wedge product though. However the thus defined flow is not a map $T^*\Sigma \rightarrow T^*\Sigma$, since the momenta do not satisfy the hidden constraint $\mathbf{D}\Phi^T M^{-1}p = 0$.

²³A convenient numerical scheme for solving the nonlinear constraint equation $\Phi(q_{n+1}) = \xi$ is provided by original SHAKE iteration; see [259]. The SHAKE iteration can be considered a nonlinear one-step Gauss-Seidel-Newton iteration as has been argued in [266]. It is incredibly stable at rather large step-size, e.g., 5fs with a torsion angle constraint and the Ryckaert-Bellemans force field.

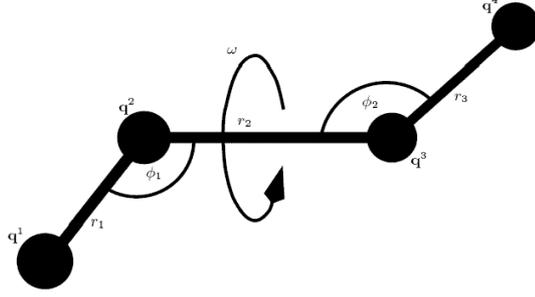


Figure 18. Ryckaert-Bellemans united-atoms butane molecule [91].

Defining $K(p) = T(M^{-1}p)$, the phase space analogue of T_Σ is found to be

$$K_\Sigma(q, p) := \frac{1}{2} \langle P_{M,T}^* p, p \rangle_{M^{-1}}, \quad P_{M,T}^* = M P_{M,T} M^{-1}.$$

It is easy to see that $P_{M,T}^*$ is idempotent and symmetric with respect to $\langle \cdot, \cdot \rangle_{M^{-1}}$. Hence $P_{M,T}^*$ is the M^{-1} -orthogonal projection onto the constrained momentum space $T_q^* \Sigma$. In other words, $P_{M,T}^*$ sends $p \in \mathbf{R}^n$ to $\tilde{p} \in \mathbf{R}^n$, such that \tilde{p} satisfies the hidden constraint $\mathbf{D}\Phi^T M^{-1} \tilde{p} = 0$. Omitting normalization, the constrained Maxwellian is

$$\varrho_\Sigma(q, p) \propto \exp(-\beta K_\Sigma(q, p)) \quad (5.4)$$

which is exactly the ambient space analogue of the constrained density (4.8) in local coordinates. The easiest way to draw momenta from the constrained distribution (5.4) is to generate a vector p due to the unconstrained distribution $\exp(-\beta K(p))$, and then apply the projection onto the constrained cotangent plane $T_q^* \Sigma$, $q \in \Sigma$. This then yields a vector $\tilde{p} = P_{M,T}^* p$ that is properly distributed according to ϱ_Σ . In this way the projection maintains the full dimensionality for the HMC algorithm, and we can completely work in the ambient space coordinates q and p .

The algorithm We summarize the considerations from Section 4.2.2 and the last few paragraphs. Given an initial position q_0 that satisfy the constraint $\Phi(q_0) = \xi$, the constrained hybrid Monte-Carlo algorithm proceeds as follows.

- (i) Draw a random vector due to the unconstrained momentum distribution

$$p \sim \exp(-\beta K(p)), \quad K(p) = \frac{1}{2} \langle M^{-1} p, p \rangle.$$

- (ii) Project p so as to satisfy the hidden constraint, i.e., $p_0 = P_{M,T}^* p$ with

$$P_{M,T}^* = \mathbf{1} - J_\Phi (J_\Phi^T M^{-1} J_\Phi)^{-1} J_\Phi^T M^{-1}, \quad J_\Phi = \mathbf{D}\Phi(q_0).$$

- (iii) Propagate $(\tilde{q}_1, \tilde{p}_1) = \Psi_\tau(q_0, p_0)$, where Ψ_τ is the numerical flow up to time $\tau > 0$, that is defined by the RATTLE discretization (5.3).

- (iv) Accept $q_1 = \tilde{q}_1$ with probability

$$r = \min \left(1, \frac{\exp(-\beta H(\tilde{q}_1, \tilde{p}_1))}{\exp(-\beta H(q_0, p_0))} \right),$$

or reject, i.e., set $q_1 = q_0$. (Here $H = K + V$ is the unconstrained Hamiltonian.)

- (v) Repeat.

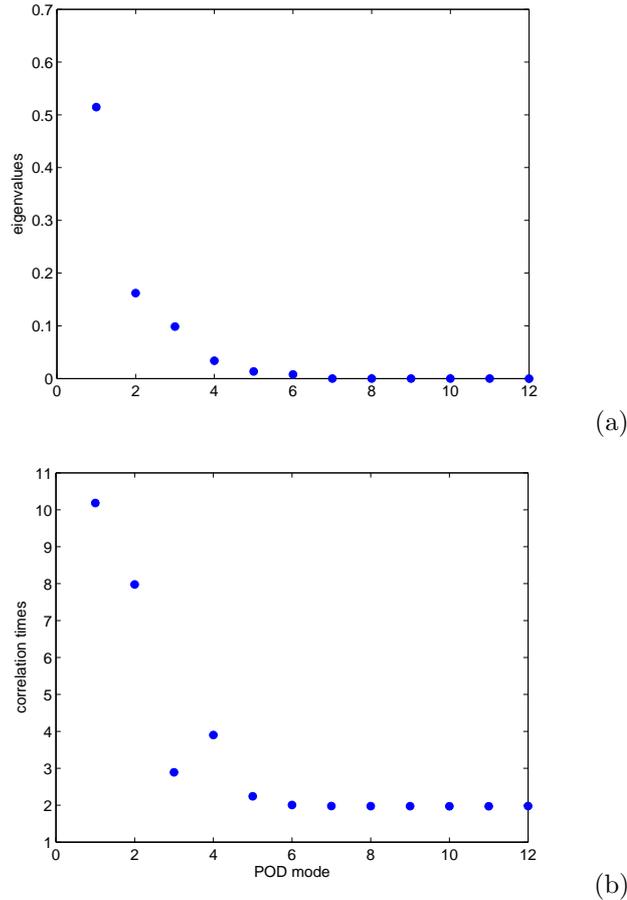


Figure 19. The figures show the POD analysis of the Cartesian configuration data of a butane molecule. The data stems from a HMC run at $T = 300\text{K}$ and $\tau = 50\text{fs}$ observation time span (step-size $h = 1\text{fs}$). The total number of steps is $N = 200\,000$ Upper panel: eigenvalues of the covariance matrix. Lower panel: characteristic time scales of the rotated modes (the scale is arbitrary)

5.2. Ryckaert-Bellemans *n*-butane

Proper orthogonal decomposition We study the spatio-temporal decomposition of the 12-dimensional Cartesian configuration space of a united atoms butane molecule. To this end we generate a HMC time series at $T = 300\text{K}$ with observation interval $\tau = 50\text{fs}$ between the HMC points. The integration is carried out with an ordinary Leapfrog/Verlet integrator with time step $h = 1\text{fs}$. For the chosen parameters h, τ the HMC acceptance probability is nearly one. Denoting by $\{q_1, \dots, q_N\}$ the HMC Markov chain of length $N = 200\,000$ we estimate the covariance matrix by

$$\hat{C} = \frac{1}{N-1} \sum_{k=1}^N (q_k - \bar{q})(q_k - \bar{q})^T .$$

where \bar{q} denotes the arithmetic mean of the data

$$\bar{q} = \frac{1}{N} \sum_{k=1}^N q_k.$$

Moreover the data has been aligned in order to remove overall translations and rotations (i.e., the rigid body symmetry which lowers the rank \hat{C} by six).

Let us write the singular value decomposition of the symmetric covariance matrix as $\hat{C} = U\Lambda V^T$ with $\Lambda = \text{diag}(\lambda_k)$ with $\lambda_6, \dots, \lambda_{12} = 0$ and orthogonal matrices $U = V$. The POD modes are defined as

$$z_k = V^T (q_k - \bar{q}).$$

For the butane data we observe that the eigenvalues decay almost exponentially with one dominant eigenvalue that explains about 50% of the total variance (see Figure 19a). The first two modes cover about 81% of the total variance. To determine the number of important modes, we take a look at the corresponding decorrelation times

$$\tau_{d,i} = \frac{1}{N} \sum_{k=1}^N |c_k^i|,$$

where the discrete autocorrelation function c_k^i of the i -th POD mode at time lag k is estimated via fast Fourier transform of the data [284]. It can be seen in Figure 19b that the first two POD modes are slow in the sense that their autocorrelation functions decay slowly. The characteristic timescales of the remaining modes are comparably shorter. Note the following two features of the characteristic timescales: First of all the last six modes do not show any interesting behaviour which reflects the symmetry-reduction by means of the molecular alignment. (The same is true for the eigenvalues of the covariance matrix.) Furthermore the decay time τ_d is not a monotonic function of the number of modes, i.e., there may always be slow modes that have small variance.

Let us study the approximation capabilities of POD by means of the single torsion angle ω which is the observable of interest for the conformation dynamics. If we denote by $P_k \in \mathbf{R}^{n \times k}$ the matrix that contains the first k eigenvectors of the covariance matrix, we can define the rank- k approximant of the original data q_k as

$$\hat{q}_l = P_k P_k^T (q_l - \bar{q}) + \bar{q} = P_k (z_l^1, \dots, z_l^k)^T + \bar{q}.$$

Accordingly we obtain a reconstruction of the torsion angle by $\hat{\omega}_l = \omega(\hat{q}_l)$. We find in Figure 20a that even the single mode approximation yields the correct qualitative conformation behaviour between the three metastable sets (one *trans* and two *cis* conformations) although the variance is reduced as expected. This is contrasted with the two-mode approximation in Figure 20b which captures the main dynamical features of the conformation dynamics. Yet the invariant distribution of the torsion angle is not fully captured by the two-mode approximation, since small errors can accumulate over time and thus lead to the wrong distribution (see Figure 21 below).

Free energy calculation and optimal prediction We shall restrict our attention to the single torsion angle of the butane molecule; it will serve as the reaction coordinate from now on. Let us start with the free energy along the torsion angle which equals the torsion potential independently of the temperature [285]. (Note that this is not true in general, but only for the n -butane potential that does not involve Lennard-Jones interactions.) We compare the two quantities: Helmholtz free energy

$$F(\omega) = -\beta^{-1} \ln \int_{\Sigma} \exp(-\beta V) \|\nabla \Phi\|^{-1} d\sigma_{\omega},$$

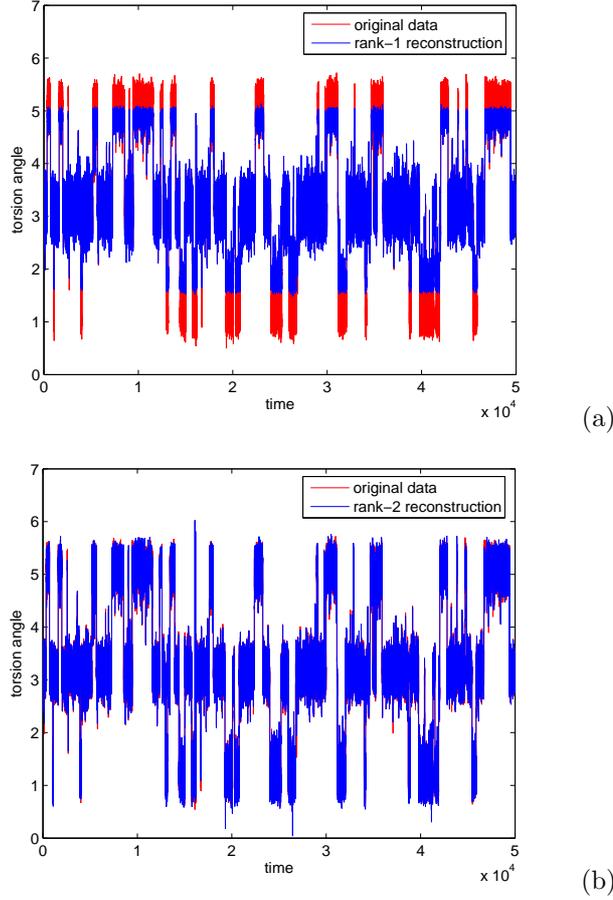


Figure 20. The plot compares the single mode approximation of the central torsion angle with the two-mode approximation for the first 50 000 HMC steps.

and geometric free energy

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

where $\Sigma = \Phi^{-1}(\omega)$ in either case. Both energies are computed by Thermodynamic Integration using the constrained hybrid Monte-Carlo algorithm at temperature $T = 300\text{K}$ with $\tau = 50\text{fs}$ between the HMC points and 5 000 sample points per constrained run. The interval $[-\pi, \pi[$ was subdivided using a uniform grid of 60 values between 0.14 and 6.14 (see Section 5.1 for the simulation details).

The standard and the geometric free energy are shown in Figure 22a. One observes that the geometric free energy barrier from the *cis* to the *trans* conformation is about 1.0kJ/mol higher than the standard free energy barrier, which confirms that free energy barriers give a only lower bound for the (reversible) work that is needed to switch between different conformations. Simultaneously we have sampled the effective inverse mass, that appears in the optimal prediction Hamiltonian (3.63) and that is depicted in Figure 22b. The plot clearly indicated that the inverse mass or

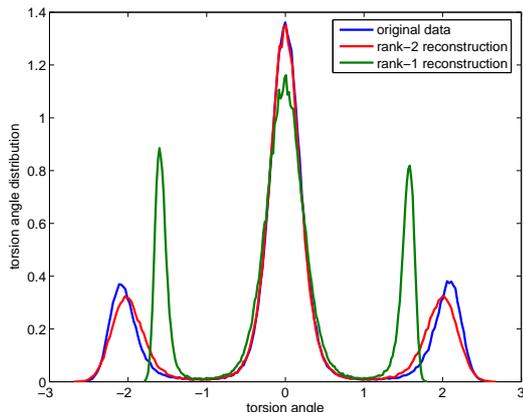


Figure 21. The plot shows histograms the torsion angle distribution at $T = 300\text{K}$ that were computed over the full time series and approximations thereof. The variance in both reconstructions is lowered as compared to the original data. The red histogram moreover illustrates that the correct distribution may not be reached although the dynamics seems well-captured. The reason is that small errors can accumulate over time and thus lead to the wrong distribution.

inverse metric, respectively, depends on the reaction coordinate, even in this simple case of a single torsion angle living on the unit circle S^1 . We observe that the kinetic energy favours the *trans* conformation, which is characterized by a rather slim shape with respect to the principal axis of inertia and which should be contrasted with to the more clustered *cis* conformations. This seems somehow counter-intuitive since one could expect that the mass distribution of a rotating molecule tends to spread out due to centrifugal forces. However here the situation is different since the rotation we are dealing with is an internal motion of the molecule. Likewise the respective kinetic energy is *internal* and should not be confused with the rotational energy of a rigid-body. Physically speaking, the effective mass indicates the redistribution of atomic masses for different conformations. Since the kinetic energy tends to stabilize the more compact *trans* conformation by slightly increasing the total energy of the *cis* conformations, we shall speak of an *internal centripetal force*.²⁴

In order to study the dynamical properties of the reduced model we compare the torsion dynamics of the full HMC simulation to a HMC simulation of the optimal prediction Hamiltonian as has been defined in Section 3.3.1:

$$E(\omega, \eta) = \frac{1}{2g(\omega)}\eta^2 + G(\omega).$$

Here

$$\frac{1}{g(\omega)} = \mathbf{E}_\Sigma \|\nabla\Phi\|_{M^{-1}}^2 \quad \text{with} \quad \|x\|_{M^{-1}}^2 = \langle M^{-1}x, x \rangle$$

is the mass-weighted effective inverse mass. A phase plot of the effective Hamiltonian is presented in Figure 23. For the sake of comparison we also do a HMC simulation in

²⁴Although the conjugate momentum to the torsion angle is certainly angular momentum, the reader should not be tempted to interpret the effective mass as moment of inertia. Both *cis* and *trans* have the same moments of inertia with respect to the rotation axis. In fact it is constant.

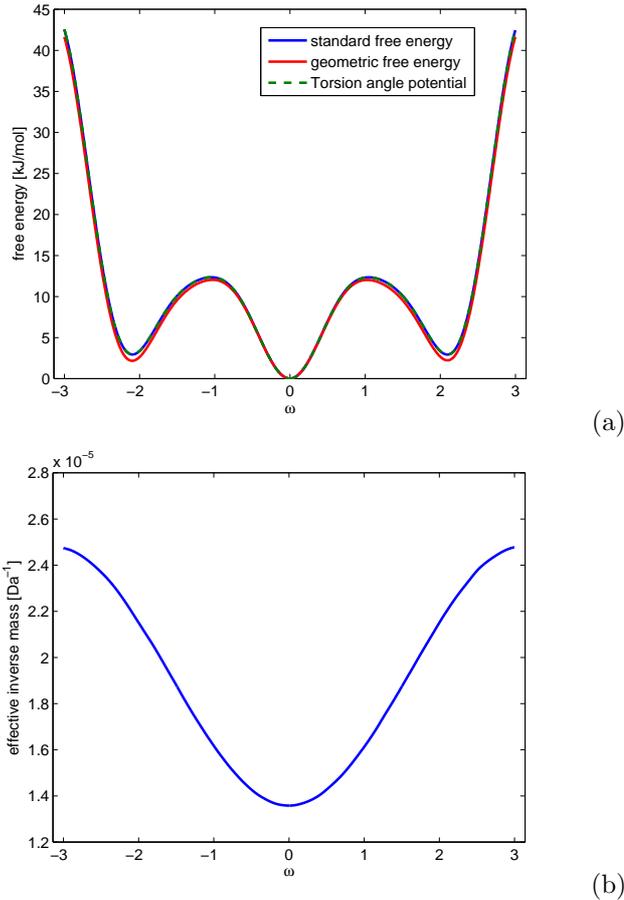


Figure 22. Comparison of free energy $F(\omega)$ and geometric free energy $G(\omega)$ at $T = 300\text{K}$ along the single torsion angle. For the Ryckaert-Bellemans n -butane the free energy does not depend on T . The energy barrier in the geometric free energy between *cis* and *trans* conformation is about 1.0kJ/mol higher than for the standard free energy. The lower panel shows the angle-dependent effective inverse mass that appears in the optimal prediction Hamiltonian.

the free energy landscape. The corresponding free energy Hamiltonian is defined by

$$H_{\text{free}}(\omega, \eta) = \frac{1}{2\mu}\eta^2 + F(\omega),$$

where μ is the effective mass

$$\mu = \beta^{-1} \text{cov}(\dot{\Phi})^{-1}.$$

Here $\dot{\Phi}$ means the torsion angle velocity. The definition of an effective mass by means of the covariance is based on the assumption that the kinetic energy in the simulated ensemble is equally partitioned among all degrees of freedom (equipartition of energy).²⁵ We run HMC simulations for E , H_{free} and for the full system, each at

²⁵Keep in mind that the hybrid Monte-Carlo trajectory does not contain momenta or velocities. Therefore we have computed the torsion angle velocity taking finite differences along the trajectory.

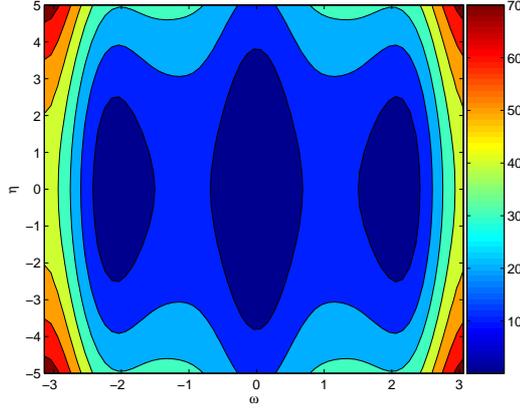


Figure 23. Optimal prediction Hamiltonian E at $T = 300\text{K}$ (arbitrary units).

room temperature $T = 300\text{K}$ with $\tau = 50\text{fs}$ and total length $N = 500\,000$ steps. The results in terms of torsion angle distribution and decay of correlations is presented in Figure 24 below. By definition the separable free energy Hamiltonian reproduces the correct marginal distribution of the torsion angle. However this is not true for the optimal prediction Hamiltonian: the marginal distribution here is²⁶

$$\int_{\mathbf{R}} \exp(-\beta E(\omega, \eta)) d\eta \propto \exp(-\beta G(\omega)) \left(\sqrt{\mathbf{E}_{\Sigma} \|\nabla \Phi\|_{M^{-1}}^2} \right)^{-1},$$

which should not be confused with the respective Fixman relation (3.26),

$$\exp(-\beta F(\omega)) = \exp(-\beta G(\omega)) \mathbf{E}_{\Sigma} \|\nabla \Phi\|_{M^{-1}}^{-1}.$$

Nevertheless it seems that the deviation is not too severe, since the optimal prediction simulation almost reaches the correct torsion angle distributions (see Figure 24a). As a dynamical observable we have computed the autocorrelation functions for each torsion angles. It can be seen in Figure 24b that the decay of correlations of the optimal prediction simulation and the free energy simulation is close to the behaviour of the full HMC Markov chain. Unfortunately the correct decay of correlations is no indicator for the correct transition behaviour between the conformations. Indeed if we define metastable sets by subdividing the interval $[-\pi, \pi[$ into three boxes $M_1 = [-\pi, -1[$, $M_2 = [-1, 1[$, and $M_3 = [1, \pi[$ and compute the transition matrix of the respective Markov chains by simply counting transitions within the HMC time step τ , we find the following three transition matrices:

$$P_{\text{full}} = \begin{pmatrix} 0.9927 & 0.0073 & 0 \\ 0.0022 & 0.9956 & 0.0022 \\ 0 & 0.0075 & 0.9925 \end{pmatrix} \quad (5.5)$$

$$P_{\text{op}} = \begin{pmatrix} 0.9927 & 0.0073 & 0 \\ 0.0022 & 0.9956 & 0.0022 \\ 0 & 0.0075 & 0.9925 \end{pmatrix} \quad (5.6)$$

²⁶The occurrence of mass-weighted metric is owed to the fact that the constrained expectation is understood as the configuration space marginal of a constrained Hamiltonian system. Therefore the respective Gibbs density involves the molecular masses by virtue of the Riemannian surface element.

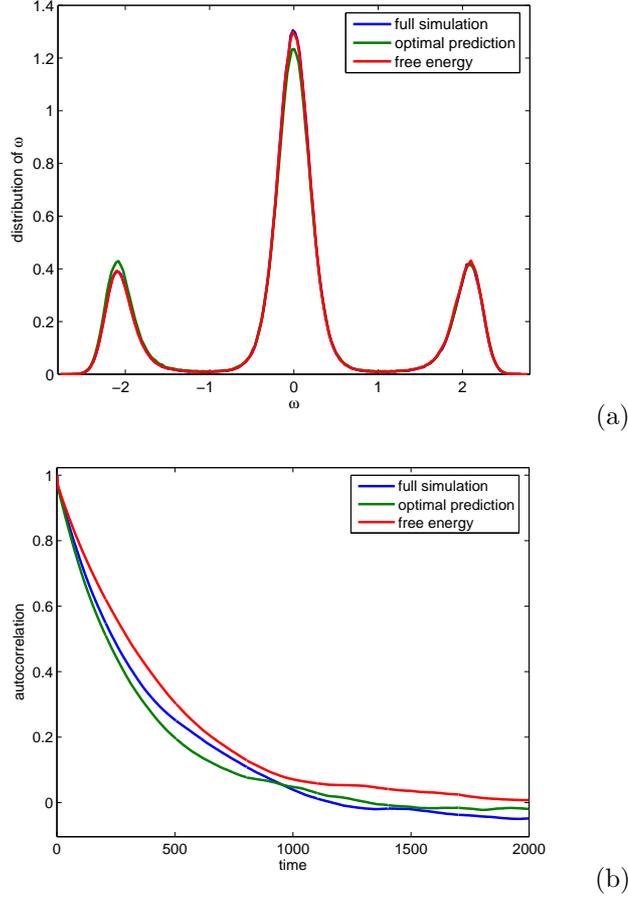


Figure 24. Comparison of reduced models for conformation dynamics. Upper panel: distribution of the torsion angles at $T = 300\text{K}$ over a 25ns HMC simulation. Lower panel: autocorrelation functions of the torsion angle (arbitrary time units).

$$P_{\text{free}} = \begin{pmatrix} 0.9941 & 0.0059 & 0 \\ 0.0016 & 0.9968 & 0.0016 \\ 0 & 0.0056 & 0.9944 \end{pmatrix}. \quad (5.7)$$

The three states M_1, M_2, M_3 are chosen such that the respective conformations are separated by the saddle points of the free energy landscape. Notice that $P_{\text{op}} = P_{\text{full}}$. In particular the probability to stay inside the *cis* transformations is lower as compared to the free energy system. Provided the coarse-grained dynamics is still Markovian, then the three-state Markov chains are completely characterized by the respective transition matrices (e.g., exit times, transition rates). Of course the deviations between the transition matrices are not very drastic, but we have observed that the identity $P_{\text{op}} = P_{\text{full}}$ is quite robust with regard to different discretizations. Moreover we should keep in mind that the matrices we all computed from finite samples.

Recapitulating, the optimal prediction Hamiltonian provides a reasonable reduced model for studying the dynamics between conformations on short time intervals (the

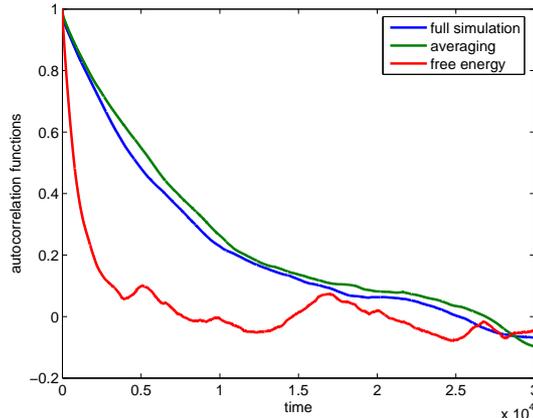


Figure 25. Decay of correlations for the torsion angle using Brownian dynamics at temperature $T = 300\text{K}$. (The time axis is arbitrarily scaled.)

correct invariant distribution is clearly given by the free energy). The main feature of optimal prediction however is that it allows for a physical interpretation of the different dynamical contributions in terms of inertial forces and forces coming from the reduced potential (i.e., the geometric free energy).

Brownian motion Just for illustration we repeat the simulation using the reduced diffusion model (3.46) from Section 3.2. The equation of motion is

$$\dot{\omega} = -\frac{1}{m(\omega)}G'(\omega) - \frac{1}{2\beta} \frac{m'(\omega)}{m(\omega)^2} + \sqrt{\frac{2}{\beta} \frac{1}{m(\omega)}}\dot{W}$$

where

$$m(\omega) = \mathbf{E}_{\Sigma} \|\nabla\Phi\|^{-2}$$

should not be confused with the effective mass g in the optimal prediction Hamiltonian. In particular the constrained expectation is understood with respect to the mass-free constrained Brownian dynamics. We have run a full simulation at $T = 300\text{K}$ and step-size $h = 1\text{fs}$, and also simulated diffusion in the respective free energy landscape using the Euler-Maruyama scheme. Then we compared both realizations to the reduced diffusion model. The full length of the simulation was $N = 1\,000\,000$ steps. (Note that the total length of the HMC trajectory was 25ns as compared to only 1ns here.)

It turns out that the naïve free energy model gives a remarkably wrong decay of correlations, whereas the averaged model does a fairly good job (see Figure 25). Apparently, it is necessary to introduce additional friction and noise parameters $\gamma, \sigma > 0$, similar to the effective mass μ in the free energy Hamiltonian, that controls the decay of correlations. Following [286, 287, 288] we can estimate the diffusion coefficient σ from the decay of the velocity autocorrelation function (Einstein-Green-Kubo formula). That is, we consider the reduced equation

$$\gamma\dot{\bar{\omega}} = -F'(\bar{\omega}) + \sigma\dot{W},$$

where

$$2\gamma = \beta\sigma^2 \quad \text{and} \quad \sigma^2 = \int_0^\infty c_\omega(t) dt$$

with $c_\omega(t)$ denoting the velocity autocorrelation function of the torsion angle. The result is shown in Figure 26, where it turns out that the parametrized free energy diffusion model now reproduces the correct decay of correlations. (It better be, for we have chosen friction and noise coefficients in such a way as to reflect the correct correlations.) Moreover the model captures the correct marginal distribution of the torsion angle by definition of the standard free energy — at least for a sufficiently long simulation. As we have argued, the decay of correlations does not tell us anything about other dynamical observables such as transition rates between metastable conformations. For further comparison we compute the Markov transition matrices between the sets $M_1 = [-\pi, -1[$, $M_2 = [-1, 1[$, and $M_3 = [1, \pi[$:

$$P_{\text{full}} = \begin{pmatrix} 0.9988 & 0.0012 & 0 \\ 0.0005 & 0.9990 & 0.0005 \\ 0 & 0.0011 & 0.9989 \end{pmatrix} \quad (5.8)$$

$$P_{\text{avg}} = \begin{pmatrix} 0.9988 & 0.0012 & 0 \\ 0.0005 & 0.9990 & 0.0005 \\ 0 & 0.0011 & 0.9989 \end{pmatrix} \quad (5.9)$$

$$P_{\text{free}} = \begin{pmatrix} 0.9982 & 0.0018 & 0 \\ 0.0004 & 0.9991 & 0.0005 \\ 0 & 0.0012 & 0.9988 \end{pmatrix}. \quad (5.10)$$

As before in case of optimal prediction, we observe that the transfer operator of the averaged model coincides exactly with the full propagator, $P_{\text{full}} = P_{\text{avg}}$. Conclusively, and in accordance with the considerations of the Hamiltonian system, we expect that the diffusion model with the geometric free energy is able to capture the correct transition rates between the *cis* and the *trans* conformations. The standard free energy model gives the correct statistics of the conformations though, and the deviation between the free energy propagator and the full one is not very drastic anyway.

Fixman potential: replacing fast degrees of freedom In Section 3.4 we have introduced an *ad-hoc* method to derive a semi-analytic reduced model. Since the calculation of effective models or free energy profiles, may be numerically expensive, we have proposed to replace the energy of the non-approximating modes in the system by an appropriately parametrized harmonic modelling potential. Averaging over the harmonic degrees of freedom then leads to a reduced model in terms of the reaction coordinate and an additional Fixman potential, the parametrization of which is open to choice. Moreover there is some freedom in choosing an appropriate metric on the approximant. For the butane molecule we suggest the following: Choosing the torsion angle as reaction coordinate the configuration space is essentially $S^1 \subset \mathbf{R}^{12}$. Given a metric $h(\omega)$ on S^1 , we consider the modelling Hamiltonian

$$H_{\text{fix}}(\omega, \eta) = \frac{1}{2h(\omega)} + V_{\text{tor}}(\omega) + \frac{1}{2\beta} \ln \det K(\omega).$$

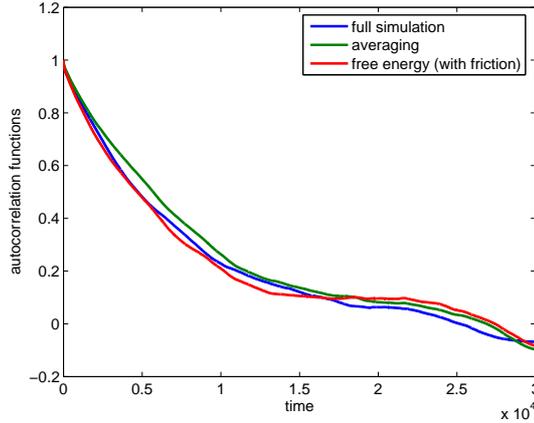


Figure 26. Decay of correlations for the torsion angle dynamics for Brownian dynamics with friction, where noise and friction coefficients σ, γ are computed from the Einstein-Green-Kubo relations (cf. Figure 25 above).

Here $K \in \mathbf{R}^{5 \times 5}$ is a yet unspecified, symmetric, positive-definite matrix.²⁷ As a potential V_{tor} , we choose the torsion potential; for the metric we choose either the constant effective mass $h(\omega) = \mu$ or $h(\omega) = g(\omega)$ from the last subsection. We decide to employ the covariance matrix of the symmetry-reduced data to feed the Fixman potential. There are essentially two possibilities

$$K_1(\omega) = \mathbf{E}_\xi(q_r(t) - \bar{q}_r)(q_r(t) - \bar{q}_r)^T,$$

or

$$K_2(\omega) = \mathbf{E}_\Sigma(q_r(t) - \bar{q}_r)(q_r(t) - \bar{q}_r)^T,$$

where \mathbf{E}_ξ and \mathbf{E}_Σ mean conditional or constrained averages, respectively, and the subscript r indicates that the data is aligned in order to account for the overall rotations and translations. Strictly speaking, the K_i denote only the irreducibly part of the covariance matrices. Running a sufficiently long unconstrained simulation the conditional covariance matrix can be computed from sorting the data according to the different values of the torsion angle. Alternatively, one could run short constrained simulations to get a local estimate of K_1 or K_2 . Here we have run constrained simulations on a 30 points grid each with about 5 000 HMC sampling points in order to obtain rather accurate estimates. In fact, it turns out that $\det K_1$ and $\det K_2$ are quite different (see Figure 27): the modelling potential that involves the constrained covariance matrix K_2 is right between the free energy and the geometric free energy $G(\omega)$, whereas the one involving K_1 is significantly different.

Hence we shall only consider two instances of the latter: a separable system with constant effective mass $h(\omega) = \mu$ and a non-separable system with $h(\omega) = g(\omega)$. For each system we run a HMC simulation at $T = 300\text{K}$ with $\tau = 50\text{fs}$. The total length is $T = 25\text{ns}$. The results are shown in Figure 28 below: it seems that the

²⁷Clearly, the normal space over S^1 is 11-dimensional. But if we further take into account the $SE(3)$ -symmetry of rigid motions, then the quotient space is effectively only 5-dimensional. Therefore the matrix in the Fixman potential has rank $s = 5$, and it suffices to consider its irreducible part.

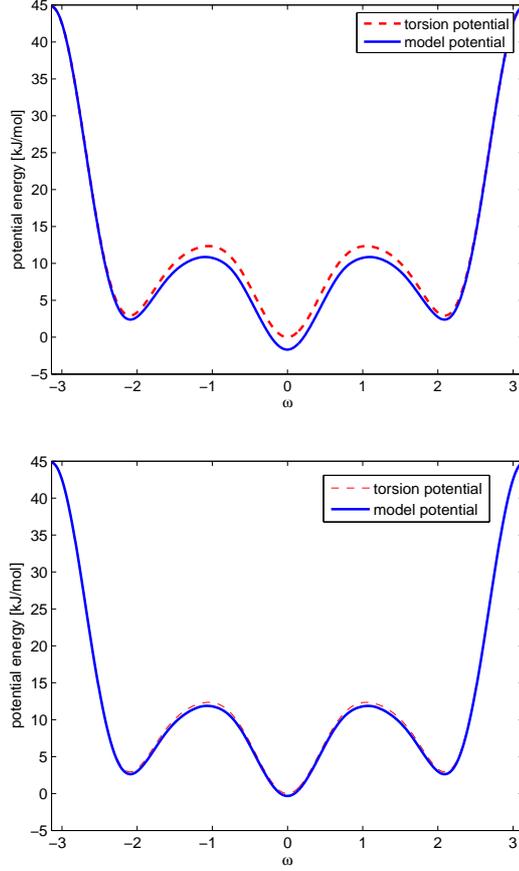


Figure 27. Different modelling potentials at $T = 300\text{K}$. Upper panel: potential energy for a Fixman potential computed from a conditional covariance matrix. Lower panel: modelling potential that has been computed from the constrained covariance matrix.

separable system with the constant effective mass performs slightly better in terms of torsion angle distribution and autocorrelation function, whereas the torsion-dependent mass overemphasizes the influence of the *cis* conformations' inertia which is already accounted for by the Fixman potential.

For further comparison we compute again the Markov transition matrices between the sets $M_1 = [-\pi, -1[$, $M_2 = [-1, 1[$, and $M_3 = [1, \pi[$. We find

$$P_{\text{full}} = \begin{pmatrix} 0.9936 & 0.0064 & 0 \\ 0.0017 & 0.9965 & 0.0018 \\ 0 & 0.0063 & 0.9937 \end{pmatrix} \quad (5.11)$$

$$P_{\text{non}} = \begin{pmatrix} 0.9915 & 0.0085 & 0 \\ 0.0012 & 0.9976 & 0.0012 \\ 0 & 0.0090 & 0.9910 \end{pmatrix} \quad (5.12)$$

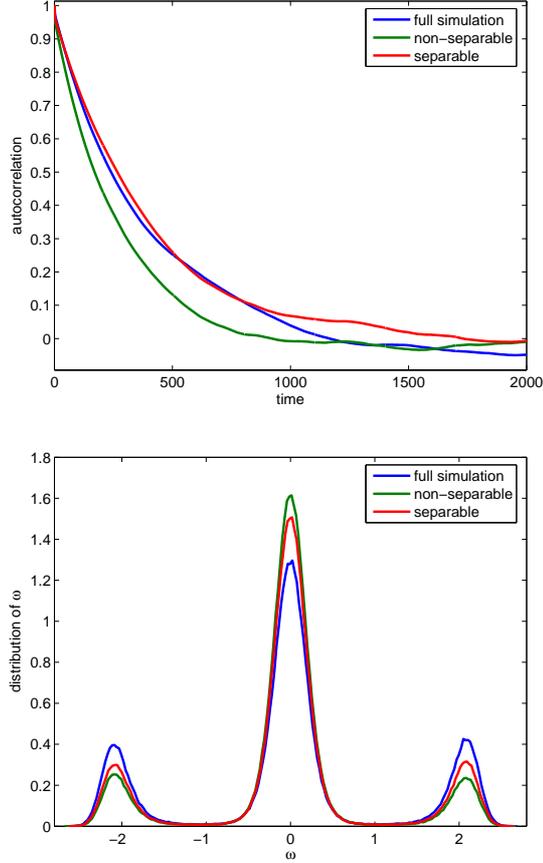


Figure 28. Simulation results for the modelling Hamiltonians involving the conditionally averaged covariance matrix at $T = 300\text{K}$. Upper panel: autocorrelation functions. Lower panel: distributions of the central torsion angles

$$P_{\text{sep}} = \begin{pmatrix} 0.9935 & 0.0065 & 0 \\ 0.0012 & 0.9976 & 0.0012 \\ 0 & 0.0066 & 0.9934 \end{pmatrix}. \quad (5.13)$$

Notice that the two (1,1) and (3,3) entries in the transition matrix P_{non} reflect the effect of the rotational energy of the *cis* transformation. To confirm that the separable model scores well, we compute the matrix norm of the differences $\Delta P_{\text{non}} = P_{\text{full}} - P_{\text{non}}$ and $\Delta P_{\text{sep}} = P_{\text{full}} - P_{\text{sep}}$. Indeed, we have

$$\|\Delta P_{\text{non}}\|_2 = 0.0046 \quad \text{and} \quad \|\Delta P_{\text{sep}}\|_2 = 0.0014,$$

where $\|\cdot\|_2$ denotes the spectral norm that is induced by the Euclidean vector norm. One should be careful with drawing conclusions from this simple example, as there are many different routes leading towards reduced models, and there is no *a priori* justification for preferring the separable model to the non-separable one. Setting up a good parametrization already requires some physical insight into the system; cf. [41, 38]. In fact it is not even clear whether an accurate sampling of the Fixman

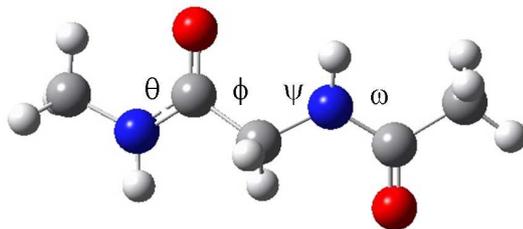


Figure 29. Glycine dipeptide in its extended C5 conformation.

potential improves the result, as we have carried out a substitution of the unresolved modes only in a small tubular neighbourhood of the approximant (here: S^1). However the simulations seem encouraging to promote further systematic studies.

5.3. Glycine dipeptide in vacuum

We consider the glycine dipeptide analogue as a paradigm for small biomolecules that exhibit interesting dynamical features, e.g., macroscopically distinct conformations and transitions between these conformations. The data is obtained from a 100ns hybrid Monte-Carlo simulation with a 100fs integration between the Monte-Carlo points. For the simulation we employ the GROMOS96 vacuum force field of GROMACS [289, 290] together with the native Java interface METAMACS [291]. It is important to note that we do not want to discuss issues whether the peptide model is physically meaningful or even realistic. It literally serves as a paradigm by means of which we can illustrate certain effects or compare different reduction schemes.

Proper orthogonal decomposition of torsion space We apply proper orthogonal decomposition (POD) to the torsion space of the glycine dipeptide analogue (GLDA). Regarding the macroscopic conformations we may confine our attention to the four central dihedral angles (see Figure 29); the leftmost and the rightmost torsions only rotate the methyl (CH_3) endgroups, which typically does not give rise to interesting physical effects. The four angles $(\theta, \phi, \psi, \omega) \in S^1 \times S^1 \times S^1 \times S^1$ span the 4-torus \mathbf{T}^4 which we shall consider as the essential configuration space. (Doing a decomposition of the Cartesian configuration space does not seem to make sense here, since the conformation dynamics of GLDA in vacuum can be completely described in terms of the torsion angles.)

In principle, we could embed the torus into a linear space, e.g., $\mathbf{T}^4 \subset \mathbf{R}^5$. But then we have to take into account that the embedding induces a nontrivial metric on \mathbf{R}^5 , which complicates both the POD and the following reduction steps. Instead we favour the idea of considering a flat torus, i.e., we regard \mathbf{T}^4 as the rectangle $[-\pi, \pi]^4$ (or $[-180, 180]^4$, respectively) upon identifying opposite edges. In this case it may happen that edges separate points lying in a single conformation which then show up as several clusters. However we are free to place the edges wherever we like. In particular we can cut the torus in a way that no clusters get separated. This amounts to a shift of the data as in Figure 30 which is optimal in the sense that the number of transitions over the edges within one Monte-Carlo step is minimized. As POD is affine

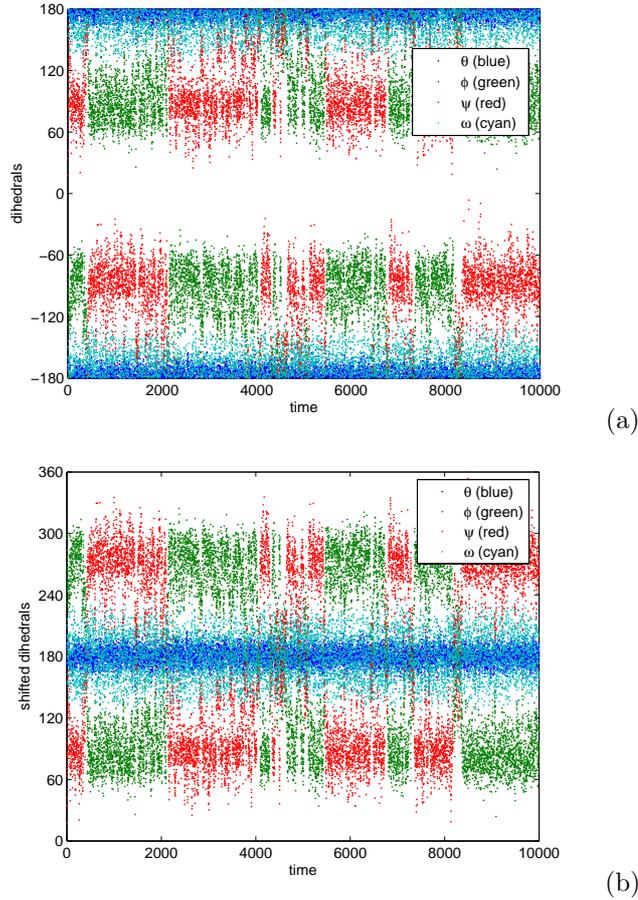


Figure 30. Optimal cut of the torus \mathbf{T}^4 . Notice that unlike before in Section 5.2, all angles are measured in degrees.

invariant it is clearly not affected by the shifting, and we can apply standard POD.

Diagonalizing the covariance matrix of the centered torsion data, it turns out that there is a single POD mode that is separated from the remaining modes by a considerable eigenvalue gap. This explains why the torsion data can be reconstructed from a single POD mode as can be seen from Figure 31. Moreover the dominant mode is much slower mixing than the other modes as is indicated by the decay of autocorrelations, which qualifies it for the systems reaction coordinate. (The eigenvalues of the covariance matrix and the characteristic time scales can be found in the caption of Figure 31.) Intriguingly the first mode depends almost completely on the two central backbone angles (ϕ, ψ) , viz.,

$$z_1 \approx \frac{1}{\sqrt{2}} (\psi - \phi) .$$

Furthermore we see from the time series that $\psi(t)$ equals $-\phi(t)$ (to all appearances). This behaviour reflects the fact that the glycine dipeptide analogue is symmetric with respect to the plane that is spanned by the peptide bond in the extended C5

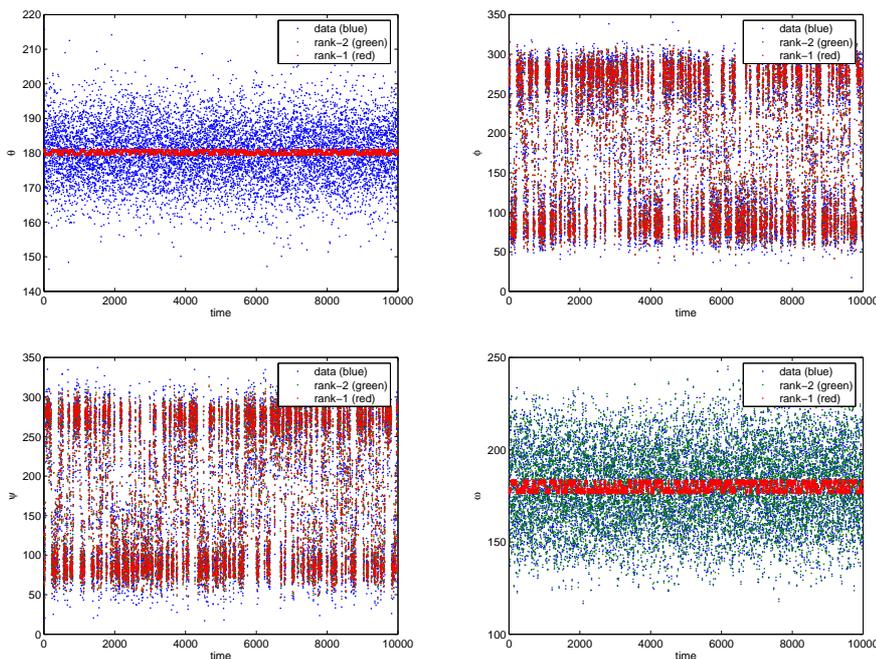


Figure 31. POD of the shifted dihedrals. The eigenvalues of the data covariance matrix are $\sigma = \{1.4448, 0.0416, 0.0371, 0.0066\}$. The corresponding autocorrelation times are $\tau = \{238.2034, 2.1211, 3.4585, 2.4682\}$.

conformation. Accordingly we have $z_1 \approx \sqrt{2}\psi$. The agreement is indeed surprising, as Figure 32 indicates. However it should be kept in mind that the approximation is likely not to capture the correct long-term behaviour (e.g., invariant distribution), since small deviations can accumulate over time.

Free energy and the optimal prediction Hamiltonian We should not be tempted to think that ψ is the slowest mode in the system, for equal reasoning applies for the other backbone angle ϕ . If we assume that dihedral angles are typically the slowest degrees of freedom in the system as compared to bond and bond angle vibrations, then it follows that only ϕ and ψ together can be considered as slow variables. Nevertheless it is instructive to first take a look at standard and geometric free energy in terms of a single backbone angle, e.g., the ψ angle.

To this end we constrain the reaction coordinate ψ , and do Thermodynamic Integration at $T = 300\text{K}$ along 36 different values ψ between -180° and $+180^\circ$. The simulation is carried out using the constrained hybrid Monte-Carlo scheme from Section 5.1, where each constrained run involves 50 000 sample points with $\tau = 100\text{fs}$ spacing between the Monte-Carlo points and internal step-size $h = 1\text{fs}$, such that the acceptance rate varies between 98% and 99%. Each integration starts from a configuration for the respective angle after minimizing the potential energy. Again we compare standard free energy

$$F(\psi) = -\beta^{-1} \ln \int_{\Sigma} \exp(-\beta V) \|\nabla \Phi\|^{-1} d\sigma_{\psi},$$

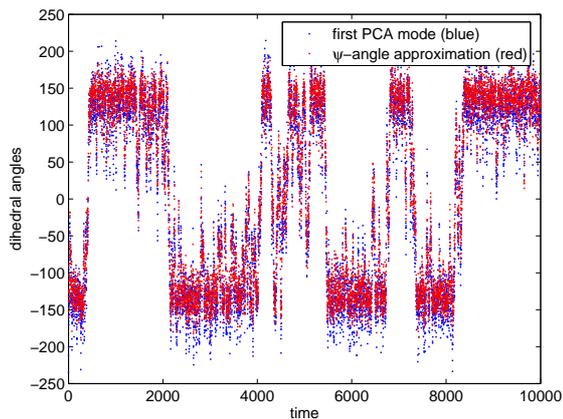


Figure 32. Approximation of the dominant POD mode by the angle ψ .

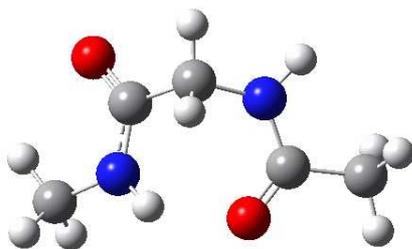


Figure 33. Glycine dipeptide in its C7 conformation.

and geometric free energy

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

where $\Sigma = \Phi^{-1}(\psi)$ is of codimension one in the configuration space \mathbf{R}^{57} . Both standard and geometric free energy are shown in Figure 34a below. The C7 conformations at $\psi \approx \pm 85^\circ$ and the stretched C5 conformations at $\psi \approx \pm 150^\circ$ are clearly distinguished (see Figure 33). Recalling the considerations from the butane example we face the same effect, namely, that the geometric free energy favours the cluster-like C7 conformation as compared to the profile of the standard free energy. This becomes more lucid if we look at the corresponding effective inverse mass $1/g$ as it appears in the optimal prediction Hamiltonian, and which is defined by the constrained expectation

$$g(\psi) = (\mathbf{E}_{\Sigma} \|\nabla \Phi\|_{M^{-1}}^2)^{-1}.$$

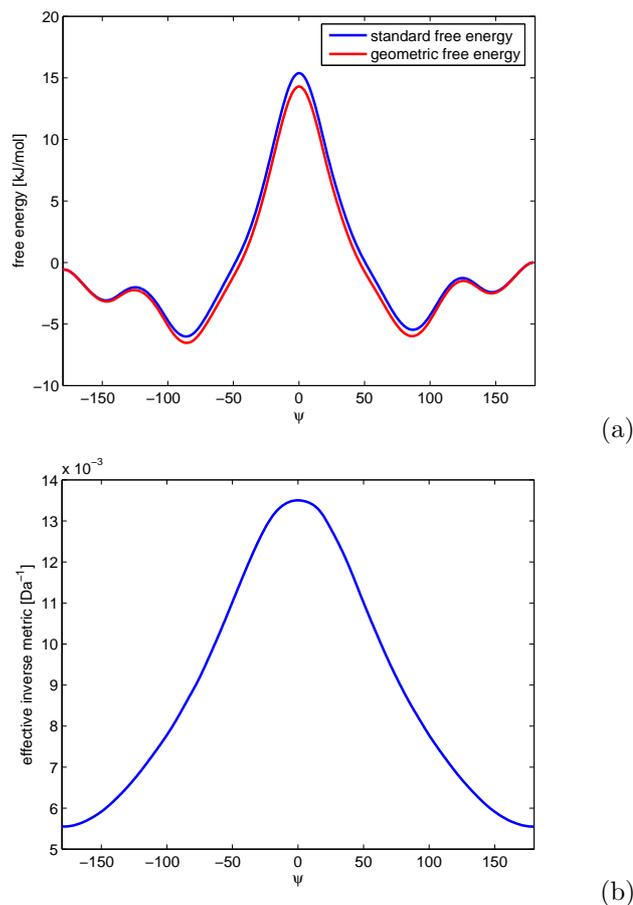


Figure 34. Free energy and effective inverse mass of the ψ angle at $T = 300\text{K}$.

The effective inverse mass is depicted in Figure 34b. It again reflects the tendency of the kinetic energy to stabilize those conformations which are slim with respect to the principal axis of inertia. (The destabilizing effect is even more evident for the knobby 0° conformation which, however, is extremely unfavourable anyway.)

Note that the effective inverse mass has almost the same shape as in the butane example before (up to a 180° shift). This suggests that the inverse effective mass is a genuine property of the chosen reaction coordinate (here: a torsion angle). This is quite remarkable, since the quantity g is computed as a configurational average which depends very well on the molecular potential and thus on the specific molecule.

Effective models in the Ramachandran plane The one-dimensional free energy profiles and the effective inverse mass of the optimal prediction Hamiltonian could already provide some insight into the conformation dynamics of GLDA. As in the butane example it has turned out that the extrinsic geometry of the reaction coordinate can have significant dynamical effects on the conformation dynamics that compete with the effects induced by the potential energy. We shall complete the picture by

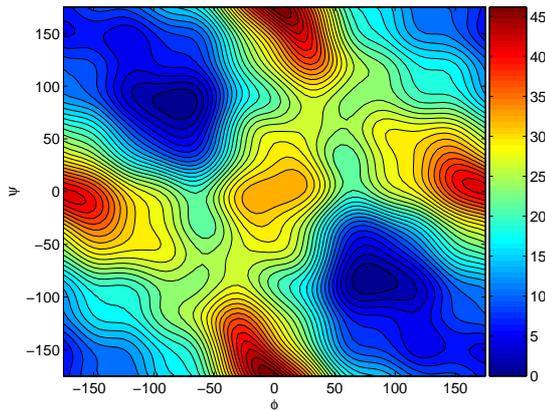


Figure 35. Standard free energy $F(\phi, \psi)$ at $T = 300\text{K}$.

looking at free energy landscapes and effective kinetic energy along the two central backbone angles ϕ and ψ .

In order to compute free energy and effective inverse mass, we perform Thermodynamic Integration in the Ramachandran plane. Such calculations are rare (e.g., [11]), although easy-to-use Thermodynamic Integration formulae in more than one dimension have been put forward during the last few years (see also [292], using a simplified force expression). We cover the Ramachandran plane with a two-dimensional, uniform 36×36 grid, and run constrained hybrid Monte-Carlo simulations at $T = 300\text{K}$ on each grid point (ϕ_i, ψ_j) . The step-size was chosen to be $h = 1\text{fs}$ with 100 integration steps between the Monte-Carlo points. Starting from a energy-minimized configuration, each simulation involves $N = 10\,000$ sample points, hence equivalently 1ns of total integration time for each ϕ, ψ combination. Using the expressions from the Corollaries 4.15 and 4.16, we compare both standard free energy

$$F(\phi, \psi) = -\beta^{-1} \ln \int_{\Sigma} \exp(-\beta V) (\text{vol} J_{\Phi})^{-1} d\sigma_{\phi, \psi},$$

and geometric free energy

$$G(\phi, \psi) = -\beta^{-1} \ln \int_{\Sigma} \exp(-\beta V) d\sigma_{\phi, \psi}$$

with $\Sigma = \Phi^{-1}(\phi, \psi)$. Taking advantage of the reaction coordinate's periodicity, we reconstruct the smooth free energy surfaces from the mean forces by expanding F and G into truncated, two-dimensional Fourier series [293]. The respective Fourier coefficients are then determined from the sampled derivatives, ∇G and ∇F , in a least square sense which amounts to solving an underdetermined linear system of equations.

The results are shown in the Figures 35 and 36. Both plots clearly reveal the C7 conformation at about $(\phi, \psi) = (\pm 80^\circ, \mp 80^\circ)$. Moreover, but less clearly, we can see the extended C5 conformation around $(\phi, \psi) = (\pm 180^\circ, \pm 150^\circ)$ which is about 5 – 10kJ/mol higher than the C7 conformation. Again the C5 energy in the geometric free energy landscape is raised as compared to the standard free energy (circa 3.5kJ/mol). For illustration the minimized potential energy function projected

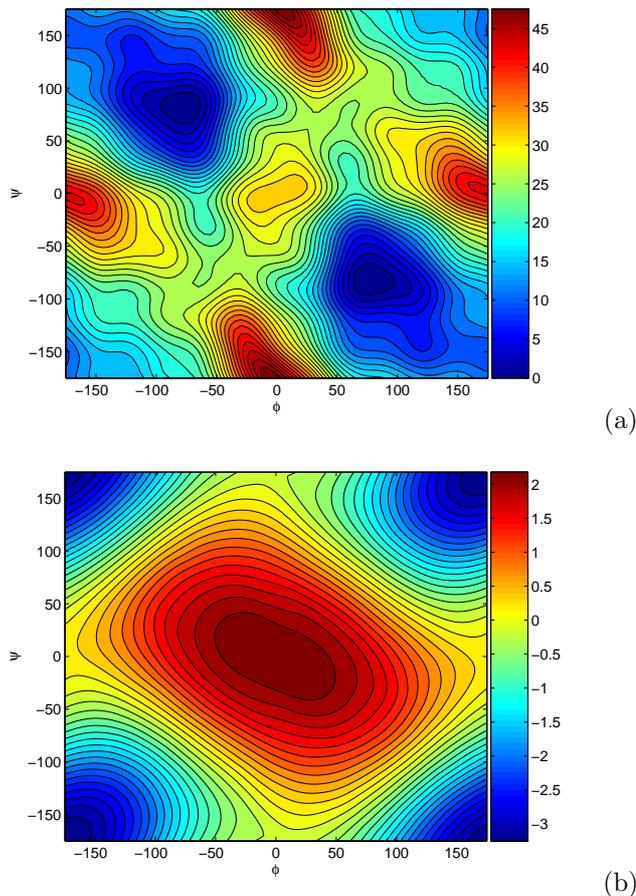


Figure 36. Geometric free energy $G(\phi, \psi)$ at $T = 300\text{K}$. The upper panel shows the geometric free energy, whereas the lower panel depicts the difference to the standard free energy. See Figure 35 and the Fixman potential in Figure 38.

onto the Ramachandran plane is shown in Figure 37 below. The most noticeable difference is that the energy barriers of the strongly repulsive O-O ring-like state at $\phi = 0^\circ$ and the H-H ring-like state at $\psi = 0^\circ$ are far more pronounced than in the free energy landscapes, which, however, does not belong to any admissible transition path anyway (cf. Figure 30). The plots indicate that both geometric and free energy are mainly affected by the potential energy landscape of the system.

For further comparison we have plotted the difference $F - G$ in Figure 36b. Letting Φ denote the reaction coordinate, the following is known on analytical grounds

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

where the mass-weighted matrix volume is defined as

$$\text{vol}J_\Phi = \sqrt{\det(\mathbf{D}\Phi M^{-1} \mathbf{D}\Phi)},$$

This should be contrasted with the Fixman potential $A = -\beta^{-1} \ln \sqrt{\det g}$ that is

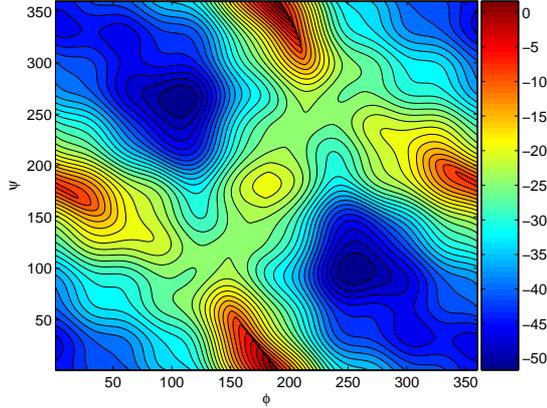


Figure 37. Minimized potential energy landscape of the backbone angles.

defined by the relation

$$\int \exp(-\beta E) d\eta \propto \exp(-\beta A) \exp(-\beta G),$$

where E is the optimal prediction Hamiltonian

$$E(\phi, \psi, \eta_1, \eta_2) = \frac{1}{2} g^{ij}(\phi, \psi) \eta_i \eta_j + G(\phi, \psi)$$

with the effective inverse mass

$$g^{-1} = \mathbf{E}_\Sigma \mathbf{D}\Phi^T M^{-1} \mathbf{D}\Phi \in \mathbf{R}^{2 \times 2}.$$

The Fixman potential A is shown in Figure 38 below. Note that clearly $A \neq F - G$, since the constrained expectation does not commute with the operations performed on the mass-weighted Gramian. (They are quite similar though.) We abstain from doing numerical simulations of the reduced system, and focus on qualitative features instead. To this end we shall study the specific form of the effective kinetic energy in more detail. Roughly, Figure 38 reflects the familiar property of the free energy to favour the extended C5 conformation: the Fixman potential achieves its global minimum in the $\pm 180^\circ$ corners of the Ramachandran plane. Moreover we see that the effective inverse mass is not diagonal, i.e., the two angles are coupled by the effective kinetic energy, where the off-diagonal terms' order of magnitude in Figure 38 is about one tenth of that of the diagonal entries. The reader should compare this to the calculations of the kinetic energy for a system with constrained bonds in [294] which reveal quite similar features. Additionally we see that the reduced model inherits the systems' symmetry, since also the kinetic energy is invariant under parity $(\phi, \psi) \mapsto -(\phi, \psi)$. Intriguingly we find that, other than the free energy, the kinetic energy carries a higher (approximate) symmetry that stems from the (almost) invariance with respect to the transformations $\phi \mapsto -\phi$ or $\psi \mapsto -\psi$. The slight perturbation of the symmetry stems from the non-uniform mass distribution along the peptide's backbone.

The kinetic energy plots reveal an interesting feature of the system: Let us assume that no extra potential were present in the optimal prediction Hamiltonian, i.e., $G = 0$, while keeping the shape of the kinetic energy as is (which is not so far off as the

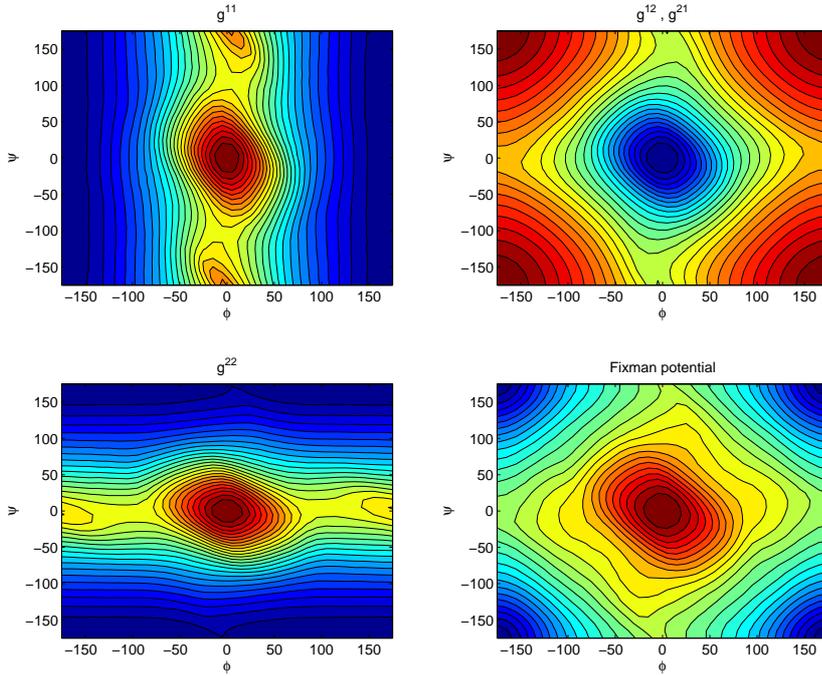


Figure 38. Effective inverse mass $g^{-1} = \mathbf{E}_{\Sigma}(J_{\Phi}^T M^{-1} J_{\Phi})$ of the two central backbone angles at $T = 300\text{K}$. The Fixman potential $A = -\beta^{-1} \ln \sqrt{\det g}$ is shown in the lower right corner (cf. also Figure 36).

comparison with the Ryckaert-Bellemans example shows). Then the system is (in a loose sense) geodesic, but is still likely to remain in the C5 conformation whenever the momentum is nonzero; only for $\eta = 0$ (inertia-free motion) the total energy is identically zero in the whole Ramachandran plane. Note that this comes up as a property of the effective Riemannian metric in the expression of the kinetic energy which therefore acts as a dynamical force that is mainly induced by the structure of the backbone (cf. the discussion in Example 3.20). In this sense the extrinsic geometry of the reaction coordinate renders conformation *dynamics* of a molecule.

6. Deviations from reduced models: correcting Brownian motion

Recall the idea of the Averaging Principle as introduced in Section 3.2. Remember moreover that we have obtained averaged diffusion equations for the reaction coordinates by artificially accelerating the dynamics of the unresolved modes. The averaged equations can then be considered the asymptotic result of the (singular) limit of infinite time scale separation between the reaction coordinate and the remaining degrees of freedom.

Correspondingly, we shall briefly sketch possible scenarios where this asymptotic strategy fails to capture the effective dynamics, e.g., due to a lack time scale separation or due to metastability in the unresolved modes. Anyway in realistic examples there is no control over the scale separation; in fact there is no small parameter at all.

6.1. Moderate deviations from the Averaging Principle

For the sake of clarity we restrict our attention to the case of a linear reaction coordinate. (The generalization to curvilinear reaction coordinates is straightforward using the results of Section 2.3.) Accordingly, we consider a Smoluchowski equation with separated slow and fast modes (x, y) and a potential $V : \mathbf{R}^s \times \mathbf{R}^k \rightarrow \mathbf{R}$, viz.,

$$\begin{aligned}\dot{x}_\epsilon(t) &= -\mathbf{D}_1 V(x_\epsilon(t), y_\epsilon(t)) + \sigma \dot{W}_1(t) \\ \dot{y}_\epsilon(t) &= -\frac{1}{\epsilon} \mathbf{D}_2 V(x_\epsilon(t), y_\epsilon(t)) + \frac{\sigma}{\sqrt{\epsilon}} \dot{W}_2(t).\end{aligned}\tag{6.1}$$

Here $\sigma^2 = 2/\beta$. As we know from Proposition 3.9, for $\epsilon \rightarrow 0$ the slow process $x_\epsilon(t)$ converges pathwise to a Markov process $x(t)$ for all $t \in [0, T]$ which is governed by

$$\dot{x}(t) = -\nabla G(x(t)) + \sigma \dot{W}_1(t),\tag{6.2}$$

where the averaged potential G is the (geometric) free energy,

$$G(x) = -\beta^{-1} \ln \int \exp(-\beta V(x, y)) dy.$$

6.1.1. Central Limit Theorem: fluctuations from equilibrium We want to study the error of the averaged motion, $x_\epsilon(t) - x(t)$, on a fixed time interval $[0, T]$. It was discovered by Khas'minskii [85] that the normalized error

$$\xi_\epsilon(t) = \frac{x_\epsilon(t) - x(t)}{\sqrt{\epsilon}}$$

has a limiting distribution for $\epsilon \rightarrow 0$ that is Gaussian.²⁸ What is interesting to note is the difference to deterministic systems, for which it is possible to asymptotically expand the error $x_\epsilon(t) - x(t) = \epsilon \xi_1(t) + \epsilon^2 \xi_2(t) + \dots$ in powers of the small parameter. For diffusive systems it can be shown however, that the error is of order $\sqrt{\epsilon}$, where all higher-order error terms are exponentially small [24]. Hence no further terms of the asymptotic expansion can be written down.

Suppose that $\mathbf{D}_1 V(x, y)$ has bounded first and second derivatives in x . We continue the analysis of the error by setting $x_\epsilon(t) = x(t) + \sqrt{\epsilon} \xi_\epsilon(t)$. Then

$$\dot{\xi}_\epsilon(t) = \frac{1}{\sqrt{\epsilon}} (\mathbf{D}_1 V(x(t) + \sqrt{\epsilon} \xi_\epsilon(t), y_\epsilon(t)) - \nabla G(x(t))).\tag{6.3}$$

²⁸We write $\xi_\epsilon(t)$ rather than $\xi_\epsilon(t, \omega)$, where $\omega \in \Omega$ is an element of some set of elementary events. However the reader should keep in mind that $\xi_\epsilon(t)$ is a stochastic process that depends on the realizations of the white noise, i.e., $\xi_\epsilon(t, \cdot)$ is a random variable for each $t \in [0, T]$.

Expanding the right hand side formally in terms of ϵ , we obtain [184]

$$\begin{aligned}\dot{\xi}_\epsilon(t) &= \frac{1}{\sqrt{\epsilon}} (\mathbf{D}_1 V(x(t), y_\epsilon(t)) - \nabla G(x(t))) \\ &\quad + \mathbf{D}_1^2 V(x(t), y_\epsilon(t)) \cdot \xi_\epsilon(t) + \mathcal{O}(\epsilon^\infty).\end{aligned}$$

The remainder $\mathcal{O}(\epsilon^\infty)$ is far from obvious, but we refer the reader to the article [86] and the references given there. Now recall that $y_\epsilon(t) = y(t/\epsilon)$. Integrating yields

$$\begin{aligned}\xi_\epsilon(t) &= \frac{1}{\sqrt{\epsilon}} \int_0^t (\mathbf{D}_1 V(x(s), y(s/\epsilon)) - \nabla G(x(s))) ds \\ &\quad + \int_0^t \mathbf{D}_1^2 V(x(s), y(s/\epsilon)) \cdot \xi_\epsilon(s) ds + \mathcal{O}(\epsilon^\infty).\end{aligned}$$

Obviously the second term on the right hand side converges in the way that

$$\int_0^t \mathbf{D}_1^2 V(x(s), y(s/\epsilon)) \cdot \xi_\epsilon(s) ds \rightarrow \int_0^t \nabla^2 G(x(s)) \cdot \xi(s) ds \quad (6.4)$$

as $\epsilon \rightarrow 0$. Upon rescaling time according to $s \mapsto \epsilon s$, the first term becomes

$$\frac{1}{\sqrt{\epsilon}} \int_0^{t/\epsilon} (\mathbf{D}_1 V(x(\epsilon s), y(s)) - \nabla G(x(\epsilon s))) ds.$$

Letting ϵ going to zero, the last expression would be simply an instance of the ordinary Central Limit Theorem, if there were no time dependence in x . But it has been proved that the integral (6.4) converges weakly to a Gaussian Markov process [85]. Abbreviating $f(x, y) = \mathbf{D}_1 V(x, y)$, this Gaussian process has the covariance matrix

$$a(x)a(x)^T = \lim_{T \rightarrow \infty} \int_0^T \int_0^T \text{cov}(f(x, y_x(s)), f(x, y_x(t))) ds dt, \quad (6.5)$$

where $y_x(t)$ is the solution of the fast dynamics for a fixed value of x , and $\text{cov}(f, g)$ denotes the covariance of two random vector fields f, g . (Note that $y_x(t)$ for fixed t is a random variable by virtue of the different realizations of the Brownian motion.) The limit of the normalized deviation satisfies the following family of equations

$$\dot{\xi}(t) = \nabla^2 G(x) \cdot \xi(t) + a(x) \cdot \dot{W}_\xi(t)$$

with $\dot{W}_\xi(t)$ denoting standard Brownian motion in \mathbf{R}^k . As a result the averaged equation (6.2) is replaced by the following pair of equations

$$\begin{aligned}\dot{x}(t) &= -\nabla G(x(t)) + \sigma \dot{W}_1(t) \\ \dot{\xi}(t) &= \nabla^2 G(x(t)) \cdot \xi(t) + a(x(t)) \cdot \dot{W}_\xi(t).\end{aligned} \quad (6.6)$$

Equation (6.6) has an interesting structure, a so-called *skew product structure*. That is, the slow equation for x is decoupled from the equation for ξ , while the deviations depend on the solution of the averaged equation. Therefore it is by no means obvious how to recover the full solution from the averaged one, including the deviations. One possibility to do so is by employing *van Kampen's approximation* [295]

$$x_\epsilon(t) \approx x(t) + \sqrt{\epsilon} \xi(t), \quad (6.7)$$

where the " \approx " symbol should not be taken literally, for convergence $\xi_\epsilon(t) \rightarrow \xi(t)$ was only in the sense of probability distributions, whereas we had convergence in probability with regard to $x_\epsilon(t) \rightarrow x(t)$. Another frequently used approach, e.g., [296], is to add some extra noise to the averaged solution, while omitting the drift:

$$\dot{x}_\epsilon(t) = -\nabla G(x_\epsilon(t)) + (\sigma + \sqrt{\epsilon} a(x_\epsilon(t))) \dot{W}_1(t) \quad (6.8)$$

All these approximations are in some sense obvious, but *ad-hoc*. Therefore we refer to the steps (6.7) and (6.8) as *remodelling* of the full dynamics. Nevertheless the method offers interesting opportunities for the practical implementation, since the covariance matrix (6.5) can be computed numerically running constrained simulations of the fast dynamics at fixed values of the slow coordinate.

Example 6.1. For the sake of illustration we reconsider our guiding example 3.20:

$$\begin{aligned}\dot{x}_\epsilon(t) &= -\partial_x V(x_\epsilon(t), y_\epsilon(t)) + \sigma \dot{W}_1(t) \\ \dot{y}_\epsilon(t) &= -\frac{1}{\epsilon} \partial_y V(x_\epsilon(t), y_\epsilon(t)) + \frac{\sigma}{\sqrt{\epsilon}} \dot{W}_2(t).\end{aligned}$$

Here $(x, y) \in \mathbf{R} \times \mathbf{R}$, and the potential is given by

$$V(x, y) = \frac{1}{4} (x^2 - 1)^2 + \frac{1}{2} \lambda(x)^2 y^2$$

with the function $\lambda(x) \geq c > 0$ defined by

$$\lambda(x) = 1 + C \exp(-\alpha(x - x_b)^2).$$

Note that the ϵ -scaling is slightly different from the previous occurrences (see Example 3.20, for instance). The skew-structured limit equation then reads

$$\begin{aligned}\dot{x}(t) &= -\partial_x G(x(t)) + \sigma \dot{W}_1(t) \\ \dot{\xi}(t) &= \partial_x^2 G(x(t)) \xi(t) + a(x(t)) \dot{W}_\xi(t),\end{aligned}\tag{6.9}$$

where G is the free energy exhibiting the entropic (dynamical) barrier,

$$G(x) = \frac{1}{4} (x^2 - 1)^2 + \beta^{-1} \ln \lambda(x),$$

and

$$a(x) = \frac{1}{\beta} \left| \frac{\lambda'(x)}{\lambda(x)^2} \right|.$$

is the standard deviation of the Gaussian error which can be computed analytically, if we assume that the fast process is ergodic (see Appendix F). Note that the second derivative of $F(x)$, which appears in the limit equation of the error is not negative definite (see Figure 39). In fact the error is completely governed by the dynamical barrier at $x = x_b$, where both $|\lambda'(x)|$ and $|\lambda''(x)|$ attain their maximum values. This is in good agreement with the numerical results in Example 3.20.

We emphasize that the averaged dynamics $x(t)$ already provides the *best-approximation* to the full dynamics $x_\epsilon(t)$ for reasonably small ϵ . Therefore, including the error does by no means improve the approximation quality of the limit dynamics. It rather serves as a vehicle to increase the variability of the averaged model, in order to incorporate certain fluctuations, that would not be present otherwise. Moreover it turns out that incorporating moderate deviations can account for situations where the scale separation is not good, i.e., ϵ is not small.

6.2. Large deviations from the Averaging Principle

The method of averaging and the Central Limit Theorem make assertions for the dynamics on the finite time intervals. Typically they both may become invalid on diverging time intervals of order $1/\epsilon$ or even $\exp(1/\epsilon)$. However there are many interesting phenomena that occur on such time scales (rare events), and which are not always captured by the Averaging Principle. One such case is the *hopping* of the fast dynamics between local attractors (metastable conformations), which may happen on time scales that are beyond the time scale of the slowest modes in the system.

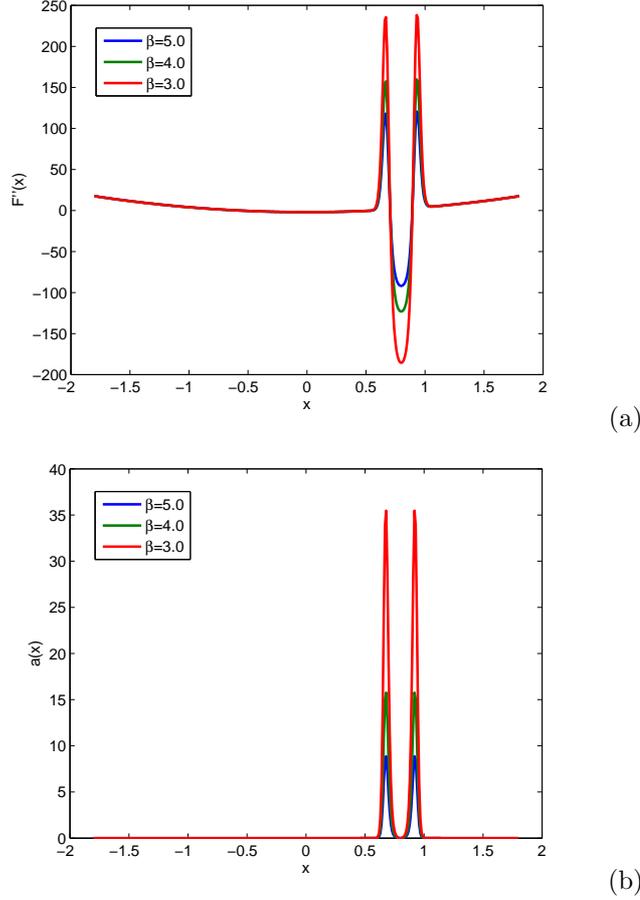


Figure 39. These plots illustrate the coefficients in the equations of motion (6.9) for the error for various inverse temperatures $\beta \in \{5.0, 4.0, 3.0\}$. Again $\beta = 3.0$ labels the most noticeable peak at $x = x_b$, whereas the little one corresponds to $\beta = 5.0$. Note first of all the error achieves its maximum in the vicinity of the dynamical barrier, and furthermore that the second derivative of $F(x)$ that drives the averaging error is not negative definite.

6.2.1. Diffusive limits Reconsider the generic slow-fast system (3.1). Here we seek to derive a reduced equation on the longer, diffusive, time scale of order $1/\epsilon$. This may be of interest, if the averaged system is deterministic on the $\mathcal{O}(1)$ time scale with vanishing drift, whereas the fast dynamics is stochastic. (One such case is the high-friction limit of the Langevin equation on the observation time scale of order 1.)

Changing the free variable in (6.1) according to $t \mapsto t/\epsilon$ while omitting the noise in the slow equation, we obtain the following system on the diffusive time scale

$$\begin{aligned}
 \dot{x}_\epsilon(t) &= -\frac{1}{\epsilon} \mathbf{D}_1 V(x_\epsilon(t), y_\epsilon(t)) \\
 \dot{y}_\epsilon(t) &= -\frac{1}{\epsilon^2} \mathbf{D}_2 V(x_\epsilon(t), y_\epsilon(t)) + \frac{\sigma}{\epsilon} \dot{W}_2(t).
 \end{aligned}
 \tag{6.10}$$

We assume that the slow dynamics averages to zero under the fast process, i.e.,

$$\int_{\mathbf{R}^k} \mathbf{D}_1 V(x, y) \mu_x(dy) = 0, \quad (6.11)$$

where $\mu_x(dy) \propto \exp(-\beta V(x, y)) dy$ is the invariant Gibbs measure of the fast process $y_x(t)$. It can then be shown that the slow drift gives significant contributions on the diffusive time scale by coupling to the noise of the fast equation as $\epsilon \rightarrow 0$, hence the term *diffusive limit*; see the rich literature, e.g., [33, 297, 23]. Associated with the system (6.10) is the Kolmogorov backward generator

$$\mathcal{A}^\epsilon = \frac{1}{\epsilon^2} \mathcal{A}_1 + \frac{1}{\epsilon} \mathcal{A}_2$$

with

$$\begin{aligned} \mathcal{A}_1 &= \frac{\sigma^2}{2} \operatorname{tr} \mathbf{D}_2^2 - \mathbf{D}_2 V(x, y) \cdot \mathbf{D}_2 \\ \mathcal{A}_2 &= -\mathbf{D}_1 V(x, y) \cdot \mathbf{D}_1. \end{aligned}$$

By adopting arguments from semigroup perturbation theory, e.g., [298], we can expand the solution of the backward equation into a power series in ϵ . Equating coefficients of equal powers gives rise to reduced equations in terms of the slow variable x , namely

$$\dot{x}(t) = b(x(t)) + a(x(t)) \dot{W}(t). \quad (6.12)$$

Here $\dot{W}(t)$ denotes standard Brownian motion in \mathbf{R}^s . As is shown in Appendix E, the drift and the noise coefficients are given by the expressions

$$\begin{aligned} a(x)a(x)^T &= 2 \int \mathbf{D}_1 V_1(x, y)^T \int_0^\infty \mathbf{E}_y \mathbf{D}_1 V(x, y_x(t)) dt \mu_x(dy) \\ b(x) &= \int \mathbf{D}_1 V(x, y) \int_0^\infty \mathbf{E}_y \mathbf{D}_1^2 V(x, y_x(t)) dt \mu_x(dy), \end{aligned}$$

where $y_x(t)$ denotes the fast process at time t starting at $y_x(0) = y$, and \mathbf{E}_y labels the average over all realizations up to time t conditional on the initial value y .

A few important bibliographical remarks are in order: Firstly, giving a rigorous proof of the diffusive limit equation is far beyond the scope of the present thesis, and refer to the appendix for a rough sketch of derivation using a perturbative expansion of the backward equation. Yet another issue is convergence: the vast majority of the results in the literature deals with weak convergence $x_\epsilon(t) \rightarrow x(t)$ on condition that the centering condition (6.11) is satisfied. Stronger results are available though; see, e.g., the original paper by Khas'minskii [33] or the textbook [124]. Other papers, like [34], relax the centering condition demanding for a two-scale expansion of the backward equation (which is much more challenging). See the paper [299] for a numerical scheme, which is based upon a multiple time stepping discretization of the original system (6.10) that can be used to simulate the limit system (6.12) quite effectively.

Remark 6.2. *The reader may wonder what happens if the slow equation already contains some noise on the $\mathcal{O}(1)$ time scale, i.e., if we face a situation like*

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

Apparently the system admits an invariant measure, the Gibbs measure

$$\mu(dx, dy) = \frac{1}{Z} \exp(-\beta V(x, y)) dx dy,$$

where the normalization constant Z is defined by

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

Existence of the Gibbs measure, however, contradicts the centering condition (6.11), as can be easily seen by interchanging the order of integration in the last equation,

$$Z = \int_{\mathbf{R}^s} \underbrace{\left(\int_{\mathbf{R}^k} \exp(-\beta V(x, y)) dy \right)}_{\text{constant due to (6.11)}} dx.$$

The centering condition is crucial for the derivation of the diffusive limit equation in the present form (see Remark F.1 below). In fact it seems that diffusive limits are more targeted on systems with deterministic right hand side, subject to random perturbations stemming from the fast variables. This is slightly different from problems usually considered in molecular dynamics. However the method is useful for studying hypo-elliptic diffusion processes, such as the high-friction limit of the Langevin equation.

6.2.2. Metastability and conditional averaging Roughly speaking the Averaging Principle relies on the idea that the fast dynamics explores its state space, sampling its invariant measure, while the slow dynamics is at rest. Therefore averaging fails, if there is some subset of state space, in which the fast dynamics is likely to get trapped. Consider the slow-fast system (6.1), and assume that the fast subsystem

$$\dot{y}_x(t) = -\mathbf{D}_2 V(x, y_x(t)) + \sigma \dot{W}_2(t)$$

is ergodic with respect to the conditional Gibbs measure $\mu_x(dy)$. Assume further that for some values of $x \in \mathbf{R}^s$ the fast dynamics shows metastability. For example, we can assume that for a particular value x_{crit} the potential $V(x, y)$, considered along the fast direction, has a significant potential barrier $\Delta V(x_{\text{crit}})$ separating two wells. If $2\Delta V(x_{\text{crit}}) \gg -\sigma^2 \ln \epsilon$ then Large Deviation Theory explains that the exit time from the potential wells induces an additional time scale that is of the order

$$\tau_{\epsilon, \sigma} \sim \epsilon \exp(\beta \Delta V(x_{\text{crit}})) \quad (\beta = 2/\sigma^2 \gg 1). \quad (6.13)$$

The last equation can be considered as some sort of Arrhenius law; for a mathematical justification we refer to the standard textbook on Large Deviation Theory by Freidlin and Wentzell [24]. It is easy to see that, at low temperature, $\tau_{\epsilon, \sigma}$ exceeds any other time scale in the system. In turn, rapid mixing of the fast variable then requires ϵ to be exponentially small as compared to the noise level σ . We would like to study the system (6.1) in the case that $\tau_{\epsilon, \sigma} = \mathcal{O}(1)$. Fixing the order of exit times in that way amounts to a scaling relation between ϵ and σ by virtue of (6.13).

We shall briefly explain the basic idea of the *conditional averaging* approach that has been put forward in [35], and which has been refined recently in [164, 300]. To this end assume that for each x , we can identify two more or less metastable sets $B_1(x), B_2(x) \subset \mathbf{R}^k$. (Here \mathbf{R}^k denotes the state space of the fast variables.) As one observes that the fast process $y_x(t)$ is rapidly mixing inside each metastable set $B_1(x)$ or $B_2(x)$, respectively, it makes sense to average the slow dynamics with respect to the (almost invariant) probability measures on each metastable set. This results in two locally averaged models, one for $B_1(x)$ and another one for $B_2(x)$, which are coupled by means of a rate matrix that governs the transitions between the metastable sets. The reduced model is then a time inhomogeneous model of the form

$$\dot{x}(t) = -\nabla G_{i(t)}(x(t)) + \sigma \dot{W}_1(t), \quad (6.14)$$

where G_i is the local free energy

$$G_i(x) = -\beta^{-1} \ln \int_{\mathbf{R}^k} \mu_{x,i}(dy) \quad (6.15)$$

with $\mu_{x,i}(dy) = (\mu_x|B_i(x))(dy)$, appropriately normalized. The switching $i(t) = i(t, x)$ is a two-state Markov jump process, that mimics the transition between $B_1(x)$ and $B_2(x)$. Under certain conditions the rates of the jump process are determined by the second dominant eigenvalue $\lambda_2(x)$ of the infinitesimal generator of the fast process $y_x(t)$, as has been demonstrated in [35] (however in a non-rigorous fashion).

A more detailed multiscale analysis is carried out in the PhD thesis [164]. There the author also exposes how the metastable fast dynamics can be approximated in some L^1 -sense by a family of Ornstein-Uhlenbeck processes that are coupled by appropriately designed Markov jump processes; cf. also the results in [301, 302].

Realization as a stochastic particle method The method of conditional averaging allows for an elegant numerical realization as a stochastic particle method which makes it accessible for practical applications. The discretization is based on a Trotter splitting of the generator associated with the conditionally averaged system (6.14), an idea borrowed from so-called *surface hopping* algorithms in quantum-classical dynamics [303, 304]. To derive the numerical scheme, consider the Fokker-Planck equation associated with the original system (6.1)

$$\partial_t \rho(x, y, t) = \mathcal{L} \rho(x, y, t), \quad u(x, y, 0) = g(x),$$

where \mathcal{L} is the backward generator in the semi-weighted Hilbert space $L^2(\mu_x)$. Given two families of metastable sets $B_1(x)$ and $B_2(x)$, we seek for a Galerkin decomposition of the full solution of the Fokker-Planck equation in the form

$$\rho(x, y, t) = c_1(x, t) \chi_1(x, y) + c_2(x, t) \chi_2(x, y).$$

Here χ_1, χ_2 span the two-dimensional dominant subspace of the fast dynamics' generator. Provided certain conditions are met (e.g., regularity of the boundary between the metastable sets), then it has been shown in [35] that the Fokker-Planck equation associated with the reduced system (6.14)–(6.15) has the following intuitive representation in terms of the coefficient vector $c = (c_1, c_2)^T$, namely

$$\partial c(x, t) = (\bar{\mathcal{L}} + \bar{\mathcal{R}}) c(x, t)$$

with $\bar{\mathcal{L}}$ containing the generators \mathcal{L}_i of the locally averaged systems (6.14)

$$\bar{\mathcal{L}} = \begin{pmatrix} \mathcal{L}_1 & 0 \\ 0 & \mathcal{L}_2 \end{pmatrix},$$

and a rate matrix $\bar{\mathcal{R}} \in \mathbf{R}^{2 \times 2}$ that provides the switching between the states $i \in \{1, 2\}$. Here the rate depends on the second eigenvalue $\lambda_2(x)$ of the generator of the fast dynamics $y_x(t)$. A time discretization at time step $h = \mathcal{O}(\epsilon^2)$ is obtained by a splitting,

$$\exp(h(\bar{\mathcal{L}} + \bar{\mathcal{R}})) = \exp(h\bar{\mathcal{L}}) \exp(h\bar{\mathcal{R}}) + \mathcal{O}(h^2),$$

of the propagator. The thus defined propagator has a nice pathwise interpretation: apparently the first exponential simply gives the propagation according to the locally averaged equations (6.14) up to time h , where the second one represents the exchange between the two states $i = 1$ and $i = 2$. In point of fact, $\exp(h\bar{\mathcal{R}})$ is a stochastic matrix for all $h > 0$, i.e., it is the transition matrix of the Markov chain $\{i(0), i(h), i(2h), \dots\}$.

Considering an ensemble of N particles $x_k(t)$, $k = 1, \dots, N$ in the respective states $i_k \in \{1, 2\}$, the system (6.14)–(6.15) has the following realization as a stochastic particle method: Propagate each particle, $x_k(t) \mapsto x_k(t+h)$, by solving

$$\dot{x}_k(t) = -\nabla G_{i_k(t)}(x_k(t)) + \sigma \dot{W}_1(t).$$

Then let each particle make a transition $i_k \mapsto j_k$ according to the transition probabilities contained in the stochastic matrix $\exp(h\bar{\mathcal{R}})$. If the $x_k(t)$ represent the ensemble $c(x, t)$ at time t , then the resulting ensemble at time $t+h$ yields a representation of $c(x, t+h)$. For the details we refer to [300].

Finally, we claim that conditional averaging provides a useful extension of the proposed reduction scheme for diffusion at low temperature. This presupposes that it is possible to estimate the transition rates between possible metastable sets along the fast variables which is typically difficult, whenever the state space of the unresolved variables is high-dimensional. One possible way then may be to resort to iterative schemes like [47], and confine the attention to very few fast variables, e.g., certain torsion angles, to get a rough estimate of metastabilities or exit times, respectively.

Unlike the extensions based on Central Limit Theorem or diffusive limits, the generalization of conditional averaging to Riemannian manifolds and curvilinear reaction coordinates is not straightforward; it requires a careful study of the boundary between the metastable sets and the restriction of the fast generator to these sets. This problem is under current investigation by the author.

7. Summary

Model reduction for molecular problems is mainly understood as the identification of suitable reaction coordinates and the calculation of free energy profiles along these. This thesis addresses model reduction for both mechanical models and stochastic differential equation from the unifying viewpoint of geometric mechanics, thereby reviewing and extending available techniques (e.g., from celestial mechanics or climate modelling) to dynamical molecular problems.

We present a systematic elaboration of the transformation properties of such different molecular models as mechanical models, Brownian motion or hypo-elliptic Langevin equations. Regarding the latter we take advantage of the close relationship with Hamiltonian systems and demonstrate that the Itô-Stratonovich ambiguity vanishes, if we confine our attention to point transformations (i.e., symplectic lifts of transformations of the configuration variables). A central paradigm in reduced modelling is the (thermodynamical) free energy. We give a detailed and concise analysis of different notions of free energy (standard and geometric free energy) and develop a generalized version of the famous Blue Moon Ensemble method which does not make any reference to the underlying dynamical system. Most notably, we gain precise understanding of the notion *free energy as the potential of mean (constraint) force* which allows for designing novel and efficient algorithms for the calculation of free energy profiles; a schematic overview of the various concepts can be found in Appendix A. For both diffusion and mechanical models we derive reduced models that are structure-preserving and covariant with regard to transformations of the reaction coordinate. As a common feature, we recognize that the reduced models contain the geometric free energy as an effective potential, which casts geometric free energy a fundamental dynamical quantity. In particular we show that the optimal prediction Hamiltonian can be written as a sum of kinetic energy and geometric free energy, where the kinetic energy is defined with respect to an averaged Riemannian metric that is induced by the reaction coordinate and which is easily computed numerically. (A brief survey over the various reduction schemes is given in Section 3.5.)

Since reduced modelling essentially boils down to the calculation of geometric free energy, sampling free energy landscapes becomes a problem of its own. We solve this problem in the context of Thermodynamic Integration by introducing two novel sampling algorithms that can handle reaction coordinate constraints: First of all, we introduce a robust hybrid Monte-Carlo scheme for constrained mechanical systems, for which a strong Law of Large Numbers is proved. Additionally, we derive a constrained Langevin equation that preserves the canonical distribution. Both algorithms prove useful for the calculation of free energy profiles, and we propose a very simple Thermodynamic Integration scheme for the Langevin equation that does without Blue Moon reweighting and without the evaluation of second derivatives. We illustrate the performance of the reduction strategies by means of two paradigmatic model systems: *n*-butane and glycine dipeptide analogue. With regard to the former we do rather detailed numerical simulations of the reduced models comparing them against the full system. We observe that both averaged Brownian motion and optimal prediction perform remarkably well in terms of dynamical observables such as decay of correlations or transition rates between *cis* and *trans* conformations (which was rather unexpected in the case of optimal prediction). Both systems exhibit the common feature that the (extrinsic) geometry of the reaction coordinate has significant dynamical effects on the conformation dynamics that compete with the effects

induced by the potential energy (geometric free energy). Moreover optimal prediction reveals an interesting physical mechanism: the kinetic energy tends to stabilize the extended *trans* conformation by slightly increasing the total energy of the bulky *cis* conformations. Since the conformational change is actuated by an internal rotation of the molecule we have termed this effect *internal centripetal force*. For glycine the calculations confirm the former observation, namely, that the kinetic energy stabilizes the extended C5 conformations by slightly lowering their energy as compared to the C7 conformations. We moreover recognize that the kinetic energy preserves the molecular potential's symmetry under parity transformations in the Ramachandran plane of the two central backbone angles, while exhibiting even a higher symmetry. Both the configuration dependence of the effective mass and possible symmetry-breaking may lead to interesting physical effects and demand for systematic studies in the future.

8. Zusammenfassung (deutsch)

In der Moleküldynamik wird Modellreduktion überwiegend im Sinne der Identifikation von Reaktionskoordinaten verstanden; eher selten wird dabei auf das zugrunde liegende dynamische System abgehoben. Diese Dissertation schließt nun ebendiese Lücke, indem sie vorhandene Verfahren zur Modellreduktion, vor allem Mittelungs- und Bestapproximationsverfahren (optimale Vorhersage) aus der Himmelsmechanik und der Klimamodellierung, auf Molekülmodelle überträgt, wobei wir besonderes Augenmerk auf Strukturhaltung der jeweiligen dynamischen Systeme legen.

Die geometrische Mechanik bietet für unsere Zwecke die einheitliche mathematische Beschreibung der unterschiedlichen Modellierungsansätze, die zudem die Begriffe der statistischen Thermodynamik (Ensembles, Entropie etc.) einschließt. So können wir unter anderem beweisen, dass so verschiedenartige Systeme wie Hamiltongleichungen oder stochastische Differentialgleichungen gemeinsame Transformationseigenschaften besitzen. Die geometrische Sichtweise liefert uns auch ein sehr präzises Verständnis der einzelnen Beiträge zur freien Energie, die ein zentrales Paradigma der reduzierenden Modellierung ist. Insbesondere zeigen wir, dass eine spezifische Form der freien Energie, die *geometrische freie Energie*, als effektives Potential in der reduzierenden dynamischen Beschreibung molekularer Systeme auftaucht. Diese Eigenschaft der geometrischen freien Energie ist allein deshalb bemerkenswert, weil vergleichbare Zusammenhänge mit Reaktionsraten bisher nur aus der Theorie der Übergangszustände bekannt waren.

Zur numerischen Berechnung der freien Energie entwickeln wir zunächst eine verallgemeinerte *Blue-Moon-Umgewichtungsmethode*, die unabhängig von dem zugrunde liegenden dynamischen System funktioniert. (Das bisherige Blue-Moon-Verfahren beschränkt sich auf Differentialgleichungen zweiter Ordnung, das heißt mechanische Systeme.) Darauf aufbauend geben wir zwei neuartige Algorithmen an, die zur Klasse thermodynamischer Integrationsmethoden gehören: Ein hybrides Monte-Carlo-Verfahren und eine Langevindynamik, die es beide erlauben, bedingte Wahrscheinlichkeitsdichten entlang von Reaktionskoordinaten abzutasten. Mit Hilfe des zweiten Verfahrens ist es insbesondere möglich, mehrdimensionale Profile freier Energie ohne Umgewichtung und ohne zweite Ableitungen der Reaktionskoordinaten zu berechnen, was numerisch einen erheblichen Effizienzgewinn darstellt und unseres Wissens nach keines der bekannten Verfahren vermag.

Um die Güte der reduzierenden Modelle zu testen, betrachten wir zunächst das Molekül *n*-Butan. Dabei zeigt sich, dass sowohl die eindimensionale, gemittelte stochastische Differentialgleichung wie auch die Bestapproximation des mechanischen Systems die Dynamik zwischen den drei wesentlichen Konformationen exakt reproduziert. Zusätzlich liefert die reduzierende Beschreibung der Bestapproximationsmethode eine plausible physikalische Erklärung für die unterschiedlichen Metastabilitäten der verschiedenen Konformationen, die sich aus dem Wechselspiel von kinetischer Energie und geometrischer freier Energie ergeben: Die kinetische Energie (Rotationsenergie) stabilisiert die gestreckte *trans*-Konformation gegenüber der etwas "massigeren" *cis*-Transformation. Mathematisch lässt sich dieser Effekt durch die Riemannsche Struktur beschreiben, die von der Geometrie der Reaktionskoordinate (Torsionswinkel) induziert wird. Darüber hinaus untersuchen wir qualitativ die zweidimensionale Bestapproximation des Biomoleküls Glyzin-Dipeptidanalogen durch seine beiden zentralen Torsionswinkel. Auch bei diesem System erklärt derselbe Rotationsmechanismus die verschiedenen Metastabilitäten von C5 und C7 Konformationen.

A. Different free energy concepts and Fixman potentials

The table below shows the various notions of free energy in this thesis and surveys their related Fixman potentials (for a more detailed overview, we refer to Section 3.5). As before $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ is a smooth reaction coordinate, where *smooth* is meant such that the level sets $\Sigma_\xi = \Phi^{-1}(\xi)$ are \mathcal{C}^2 -submanifolds of codimension k in \mathbf{R}^n for all regular values $\xi \in \mathbf{R}^k$. The Jacobian is abbreviated as $J_\Phi = \mathbf{D}\Phi$. Following the previous nomenclature, $d\sigma$ labels the surface element (or Hausdorff measure) of $\Sigma_\xi \subset \mathbf{R}^n$, whereas $d\mathcal{H}$ denotes the surface element of the phase space submanifold $\Sigma_\xi \times \mathbf{R}^n \subset \mathbf{R}^n \times \mathbf{R}^n$. Note that standard and geometric free energies coincide, if the reaction coordinate is linear in the configurations. In this case also $A = D = F - G$ below.

free energy	second-order system	first-order system	Fixman potential
standard	$F = -\beta^{-1} \ln Z$ with $Z = \int \exp(-\beta H) (\text{vol} J_\Phi)^{-1} d\mathcal{H}$	$F = -\beta^{-1} \ln Q$ with $Q = \int \exp(-\beta V) (\text{vol} J_\Phi)^{-1} d\sigma$	$D = -\beta^{-1} \ln M (= F - G)$ with $M = \frac{1}{Q_\Sigma} \int (\text{vol} J_\Phi)^{-1} \exp(-\beta V) d\sigma$
geometric	$G = -\beta^{-1} \ln Z_\Sigma$ with $Z_\Sigma = \int \exp(-\beta H) d\mathcal{H}$	$G = -\beta^{-1} \ln Q_\Sigma$ with $Q_\Sigma = \int \exp(-\beta V) d\sigma$	$W = \beta^{-1} \ln \text{vol} J_\Phi$, e.g., $V \mapsto V + W$ for Blue Moon reweighting
optimal prediction	$E = \frac{1}{2} \langle I\eta, \eta \rangle + G$ with $I = \frac{1}{Q_\Sigma} \int J_\Phi^T J_\Phi \exp(-\beta V) d\sigma$		$A = -\beta^{-1} \ln \int \exp(-\beta E) d\eta - G$, where typically $A \neq F - G$
confinement			$U = \beta^{-1} \ln \sqrt{\det K}$, (K s.p.d.) if $K = J_\Phi^T J_\Phi$, then $U = W$

B. Coordinate expressions

We introduce the local coordinate expressions and expressions for the metric tensor that are used throughout this thesis. Let $\Sigma \subset \mathbf{R}^n$ be a smooth submanifold of codimension k in \mathbf{R}^n . Recall the definition of the normal bundle over Σ

$$N\Sigma = \{(\sigma, n) \mid \sigma \in \Sigma, n \in N_\sigma\Sigma\} \subset \mathbf{R}^n \times \mathbf{R}^n$$

with the natural diffeomorphism of $N\Sigma$ into \mathbf{R}^n given by $\iota : (\sigma, n) \mapsto \sigma + n$. In a sufficiently small tubular neighbourhood $N\Sigma_\varepsilon$ of Σ with $\|n\| < \varepsilon$ we can pull back the Euclidean metric, considering $N\Sigma_\varepsilon$ as our configuration space. Then, given an orthonormal frame $\{n_1(\sigma), \dots, n_k(\sigma)\}$, we can introduce local coordinates on $N\Sigma_\varepsilon$ by

$$\phi_1 : \mathbf{R}^n \rightarrow N\Sigma_\varepsilon, (x, y) \mapsto (\sigma(x), y^i n_i(\sigma(x))), \quad (\text{B.1})$$

By means of ϕ_1 we can represent any point $(\sigma, n) \in N\Sigma_\varepsilon$ in terms of the bundle coordinates (x, y) , hence any point $q \in \mathbf{R}^n$ close to the submanifold Σ . We shall make the arrangement that all coordinates that belong to Σ are indexed by Greek letters $\alpha, \beta, \gamma, \dots$, whereas the normal coordinates are indexed by Latin letters i, j, k, \dots . Whenever it is necessary, we will use Latin indices \dots, l, m, n that run over all coordinates which, however, should become clear from the context.

We endow the tangent space $TN\Sigma_\varepsilon$ with the standard bases $\partial/\partial x^\alpha \in T_{\sigma, n}N\Sigma_\varepsilon$ and $\partial/\partial y^i \in T_{\sigma, n}N\Sigma_\varepsilon$ which give rise to local coordinates in the usual way. Let us abbreviate $z = (x, y)$. Then the local coordinate expression of the metric tensor is obtained by pulling back the Euclidean metric by the map $\phi = \iota \circ \phi_1$

$$g_{ij}(z) = \delta_{kl} \frac{\partial \phi^k}{\partial z^i} \frac{\partial \phi^l}{\partial z^j}, \quad i, j, k, l = 1, \dots, n$$

Hence the metric tensor takes the form

$$g(x, y) = \begin{pmatrix} G(x) + C(x, y) & A(x, y) \\ A(x, y)^T & \mathbf{1} \end{pmatrix}. \quad (\text{B.2})$$

where the matrix $G(x) \in \mathbf{R}^{d \times d}$, $d = n - k$ is the metric induced on Σ by restricting the Euclidean metric. Introducing the shorthand $X_\alpha = \partial\sigma/\partial x^\alpha$ for the vectors tangent to Σ , we have $G_{\alpha\beta} = \langle X_\alpha, X_\beta \rangle$. The matrix $C(x, y) \in \mathbf{R}^{d \times d}$ has the entries

$$C_{\alpha\beta} = 2y^k \langle dn_k(X_\alpha), X_\beta \rangle + y^k y^l \langle dn_k(X_\alpha), dn_l(X_\beta) \rangle,$$

where $dn(X) = \nabla n \cdot X$ denotes the directional derivative of n along X .²⁹ Note that we have exploited the symmetry $\langle dn_k(X_\alpha), X_\beta \rangle = \langle X_\alpha, dn_k(X_\beta) \rangle$ in the last equation. Finally the elements of the off-diagonal matrix $A(x, y) \in \mathbf{R}^{d \times k}$ are given by

$$A_{i\beta} = y^j \langle n_i, dn_j(X_\beta) \rangle.$$

If the codimension of Σ in \mathbf{R}^n is one, then the metric tensor takes a particularly simple form, and the matrices G, C can be given a nice geometrical meaning:

$$g(x, y) = \begin{pmatrix} G(x)(\mathbf{1} - M(x, y))^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$$

where M denotes the matrix of the Weingarten map that is associated with the second fundamental form of the embedding $\Sigma \subset \mathbf{R}^n$ (cf. [81, 218]). It is defined by

$$M_\gamma^\alpha = y G^{\alpha\beta} \langle dn_1(X_\beta), X_\gamma \rangle.$$

²⁹We will also sometimes use the common notation $dn(X) = \nabla_X n$.

The vanishing of the off-diagonal matrix A is related to the fact that the normal connection is identically zero for submanifolds of codimension one. This can be seen by differentiating the expression $\|n_1\|^2 = 1$ along Σ

$$0 = \left. \frac{d}{dt} \|n_1(\sigma(t))\|^2 \right|_{t=0} = 2 \langle dn_1(X), n_1 \rangle ,$$

where $\sigma(t)$ is a curve in Σ with tangent X at $t = 0$. More generally, the coefficients $\omega_j^i(X) = \langle dn_i(X), n_j \rangle$ are 1-forms which are called the *normal fundamental forms*. By the same differentiation argument it is easy to check that these 1-forms are skew-symmetric, $\omega_j^i = -\omega_i^j$. For the details the reader is referred to [180, 182].

Hamiltonian and Lagrange function We now state the local expressions of the molecular Lagrangian and the corresponding Hamiltonian. Without loss of generality we set the atomic masses to unity. The Lagrangian $L : TN\Sigma_\varepsilon \rightarrow \mathbf{R}$ is considered first:

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

Note that this is the ordinary Lagrangian (2.1) with $M = \mathbf{1}$, where the inner product of tangent vectors in $TN\Sigma_\varepsilon$ is defined by

$$\langle (X, Y), (X', Y') \rangle = \langle X + Y, X' + Y' \rangle ,$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product in \mathbf{R}^n . The local coordinate expression of L is

$$L(x, y, \dot{x}, \dot{y}) = \frac{1}{2} \langle g(x, y) \cdot (\dot{x}, \dot{y})^T, (\dot{x}, \dot{y}) \rangle - V(x, y) , \quad (\text{B.3})$$

where $V(x, y) = V(\sigma(x) + y^i n_i(\sigma(x)))$, and the metric tensor (B.2) can be written as

$$g = \begin{pmatrix} G + C & A \\ A^T & \mathbf{1} \end{pmatrix} = P \cdot \begin{pmatrix} G + C - AA^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \cdot P^T$$

with the matrix

$$P = \begin{pmatrix} \mathbf{1} & A \\ \mathbf{0} & \mathbf{1} \end{pmatrix} .$$

The inverse metric tensor then takes the form

$$g^{-1} = P^{-T} \cdot \begin{pmatrix} (G + C - AA^T)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \cdot P^{-1}$$

with the inverse of P ,

$$P^{-1} = \begin{pmatrix} \mathbf{1} & -A \\ \mathbf{0} & \mathbf{1} \end{pmatrix} .$$

Defining the conjugate momenta to (x, y) by

$$u_\alpha = \frac{\partial L}{\partial \dot{x}^\alpha}, \quad \alpha = 1, \dots, d$$

$$v_i = \frac{\partial L}{\partial \dot{y}^i}, \quad i = 1, \dots, k .$$

we obtain the Hamiltonian as the Legendre transform of L

$$H(x, y, u, v) = \frac{1}{2} \langle g(x, y)^{-1} \cdot (u, v)^T, (u, v) \rangle + V(x, y) . \quad (\text{B.4})$$

Calculation of the Christoffel symbols Our averaging results rely on local coordinates. Hence we need to compute the (symmetric) Christoffel symbols

$$\Gamma_{jk}^i = \frac{1}{2}g^{il} \left(\frac{\partial g_{jl}}{\partial z^k} + \frac{\partial g_{kl}}{\partial z^j} - \frac{\partial g_{jk}}{\partial z^l} \right)$$

with $z = (x, y)$. Since we can assume that all curves $(\sigma(t), n(t))$ stay close to Σ it makes sense to consider only terms up to zeroth order in y (linear terms have mean zero anyway). At $y = 0$ the first $\alpha \leq n - k$ Christoffel symbols read

$$\begin{aligned} \Gamma_{\beta\gamma}^\alpha &= \frac{1}{2}G^{\alpha\delta} \left(\frac{\partial G_{\beta\delta}}{\partial x^\gamma} + \frac{\partial G_{\gamma\delta}}{\partial x^\beta} - \frac{\partial G_{\beta\gamma}}{\partial x^\delta} \right) \\ \Gamma_{i\beta}^\alpha &= G^{\alpha\gamma} S_{\gamma\beta}^i \\ \Gamma_{ij}^\alpha &= 0, \end{aligned} \tag{B.5}$$

where $\Gamma_{\beta\gamma}^\alpha$ are simply the Christoffel symbols that are associated with the metric G on the surface Σ . The symmetric matrix that is associated with the second fundamental form in the local coordinate basis $X_\alpha = \partial\sigma/\partial x^\alpha$ has the entries

$$S_{\gamma\beta}^i = \langle dn_i(X_\gamma), X_\beta \rangle.$$

The remaining Christoffel symbols for the normal coordinates ($i \leq k$) are given by

$$\begin{aligned} \Gamma_{\alpha\beta}^i &= -S_{\alpha\beta}^i \\ \Gamma_{j\alpha}^i &= \frac{1}{2} \left(\omega_{i\alpha}^j - \omega_{j\alpha}^i \right) \\ \Gamma_{jk}^i &= 0. \end{aligned} \tag{B.6}$$

Notice that $\Gamma_{j\alpha}^i$ is skew-symmetric in the upper and lower indices, as follows from the definition of the skew-symmetric coefficients of the normal connection

$$\omega_{j\alpha}^i = \langle dn_i(X_\alpha), n_j \rangle.$$

C. More coordinate expressions and the mean curvature vector

We first address the problem how to parametrize the constraint manifold. A submanifold Σ of \mathbf{R}^n defined by the vector-valued equation $\Phi(q) = 0$ is properly immersed, if $\mathbf{D}\Phi$ has maximum rank, i.e., is non-singular almost everywhere on the surface [181, 242]. According to Sard's Lemma [173] this can be guaranteed by choosing $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$, such that it belongs to the class $\mathcal{C}^{n-k+1}(\mathbf{R}^n)$. Then the points, at which $\mathbf{D}\Phi$ 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 . The following Lemma holds:

Lemma C.1. *Let $q^* \in \Sigma$ be any non-singular point, and let $U_\delta(q^*)$ denote a sufficiently small tubular δ -neighbourhood including that point. Then there is a parametrization of Σ in $U_\delta(q^*)$ given by $\{q^1, \dots, q^{n-k}; f^1, \dots, f^k\}$ that is an embedding, where $f : \mathbf{R}^{n-k} \rightarrow \mathbf{R}^k$ is the local inverse of Φ as defined by $q^{n-k+l} = f^l(q^1, \dots, q^{n-k})$.*

Proof. Let q^* be a non-singular point of Σ , and consider the square $k \times k$ minor K of $\mathbf{D}\Phi$ which is made by, say, cropping the Jacobian's first $n - k$ columns. Suppose $\det K \neq 0$ at $q^* \in \Sigma$. Then the *Implicit Function Theorem* guarantees that we can locally solve the equation $\Phi(q) = 0$ for the vector (q^{n-k+1}, \dots, q^n) , obtaining smooth functions of the remaining coordinates. Let these function be denoted by

$q^{n-k+l} = f^l(q^1, \dots, q^{n-k})$, such that $\Phi(q^1, \dots, q^{n-k}; f^1, \dots, f^k) = 0$. Consequently $\sigma = (q^1, \dots, q^{n-k}; f^1, \dots, f^k)^T$ is an immersion and moreover by the Inverse Function Theorem and smoothness of f an embedding of Σ into \mathbf{R}^n . \square

Note that the specific choice of K does not affect our considerations, for we can always choose a different parametrization $\tilde{\sigma}$, with any k coordinates q^l being functions of the remaining $n - k$ coordinates. For instance if Φ is of class \mathcal{C}^∞ , then so are the transition functions $\psi = \sigma \circ \tilde{\sigma}^{-1}$. Thus Σ will be globally smooth.

Hence we can define local coordinates $\{x^1, \dots, x^{n-k}\}$ with $x^\alpha = q^\alpha$ for $\alpha = 1, \dots, n - k$, such that $\sigma = \sigma(x)$ is an embedding $\Sigma \subset \mathbf{R}^n$. Then we obtain from implicit differentiation of the equality $\Phi(x^1, \dots, x^{n-k}; f^1, \dots, f^k) = 0$

$$d\Phi = \left(\frac{\partial\Phi}{\partial x^\alpha} + \frac{\partial\Phi}{\partial f^l} \frac{\partial f^l}{\partial x^\alpha} \right) dx^\alpha = 0, \quad (f^l = q^{n-k+l})$$

where the sum is taken over $\alpha = 1, \dots, n - k$ and $l = 1, \dots, k$. Since the 1-forms dx^α are linearly independent, we demand that each of the brackets vanishes. For convenience we may bring the last equation into matrix vector form. We have

$$\mathbf{D}f = -(\mathbf{D}_2\Phi)^{-1}\mathbf{D}_1\Phi.$$

Here we used the symbol \mathbf{D}_1 to denote the derivative with respect to the first $n - k$ coordinates and \mathbf{D}_2 for the remaining slot. Clearly $K = \mathbf{D}_2\Phi$ is the invertible $k \times k$ minor of $\mathbf{D}\Phi$. In particular in the codimension $k = 1$ case, we can explicitly assert $\mathbf{D}f = -(\partial\Phi/\partial q^n)^{-1}\mathbf{D}_1\Phi$. Finally we obtain the restriction of the Euclidean metric to Σ as the metric that is induced by the embedding of Σ into \mathbf{R}^n ,

$$G_{\alpha\gamma} = \langle \partial\sigma/\partial x^\alpha, \partial\sigma/\partial x^\gamma \rangle = \delta_{\alpha\gamma} + \sum_l \frac{\partial f^l}{\partial x^\alpha} \frac{\partial f^l}{\partial x^\gamma}$$

which can be equivalently written in the form

$$G = \mathbf{1} + (\mathbf{D}_1\Phi)^T (\mathbf{D}_2\Phi)^{-T} (\mathbf{D}_2\Phi) (\mathbf{D}_1\Phi).$$

Mean curvature vector The map $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ with the rank of $\mathbf{D}\Phi$ equal to k defines a foliation of \mathbf{R}^n of codimension k by the collection of all connected components $\Sigma_\xi = \Phi^{-1}(\xi)$, where ξ varies throughout \mathbf{R}^k . In the calculation of the optimal prediction equations in Section 3.3.1 we have utilized a relation between the variation of the surface element $d\sigma_\xi$ with ξ and the components of the mean curvature vector of the leaf Σ_ξ . The justification is given now:

For each regular ξ value consider the normal bundle over $\Sigma = \Sigma_\xi$. We have seen in Appendix B that in a sufficiently small tubular neighbourhood we can pull back the Euclidean metric to the normal bundle $N\Sigma$. Using bundle coordinates (x, y) the pulled-back metric takes the form (B.2), viz.,

$$g(x, y) = \begin{pmatrix} G(x) + C(x, y) & A(x, y) \\ A(x, y)^T & \mathbf{1} \end{pmatrix}.$$

where the matrix G is the metric induced on Σ by the embedding. By

$$dV = \sqrt{\det g(x, y)} dx dy$$

we define the volume element on $N\Sigma$. Since $g(x, 0) = G(x) \otimes \mathbf{1}$ the surface element $d\sigma_\xi = \sqrt{\det G_\xi(x)} dx$ of Σ in can be expressed accordingly as

$$d\sigma_\xi = \sqrt{\det g(x, 0)} dx$$

where the subscript ξ is used to indicate the implicit dependence of the surface metric on the foliation parameter ξ (on the other hand, $g(x, 0) = G(x) \otimes \mathbf{1}$ without subscript, since the normal coordinate $y = 0$ takes over the role of the parameter ξ).

Given an orthonormal frame $\{n_1(\sigma), \dots, n_k(\sigma)\}$ for the normal bundle we can write the map Φ in terms of the bundle coordinates $\Phi(x, y) - \xi = (J_{\Phi}^T Q)(\sigma(x))y$, where $Q \in \mathbf{R}^{n \times k}$ is the matrix (n_1, \dots, n_k) , and $J_{\Phi} = \mathbf{D}\Phi$ denotes the Jacobian. By chain rule we can evaluate the derivative of the surface metric G_{ξ}

$$\frac{\partial}{\partial \xi^i} \sqrt{\det G_{\xi}(x)} = (J_{\Phi}^T Q)^{ij} \frac{\partial}{\partial y^j} \sqrt{\det g(x, 0)}.$$

Here $(J_{\Phi}^T Q)^{ij}$ are the elements of the inverse matrix $(J_{\Phi}^T Q)^{-1}$. Taking advantage of the identity $(\det g)' = \det g \cdot \text{tr}(g^{-1}g')$ we find

$$\begin{aligned} \frac{\partial}{\partial \xi^i} \sqrt{\det G_{\xi}} &= \frac{1}{2} (J_{\Phi}^T Q)^{ij} \text{tr} \left(g(x, 0)^{-1} \frac{\partial g}{\partial y^j} \Big|_{y=0} \right) \sqrt{\det g(x, 0)} \\ &= -(J_{\Phi}^T Q)^{ij} \text{tr} (G^{-1} \mathfrak{S}_j) \sqrt{\det G}. \end{aligned}$$

From the particular form of the metric $g(x, y)$ we can conclude that $G^{-1} \mathfrak{S}_j$ with $\mathfrak{S}_j = -P_T dn_j(\cdot)$ are the matrices of the Weingarten maps with respect to the local basis of the tangent vectors $\partial\sigma/\partial x^{\alpha} \in T_{\sigma}\Sigma$, where $P_T : T_{\sigma}\mathbf{R}^n \rightarrow T_{\sigma}\Sigma$ is the tangential projection. The trace gives the negative components of the mean curvature vector

$$H(\sigma(x)) = \sum_{i=1}^s \kappa_i(x) n_i(x), \quad \kappa_i = -\text{tr}(G^{-1} \mathfrak{S}_i),$$

with respect to the normal coordinates y^1, \dots, y^k (or the respective normal frame). Accordingly $\partial_i \sqrt{\det G_{\xi}} = (J_{\Phi} Q)^{ij} \kappa_j$ is the mean curvature with respect to the foliation $\Phi^{-1}(\xi)$. The dependence on Φ via the Jacobian does not come as a surprise, since, as is known, the mean curvature is an extrinsic curvature measure.

D. A co-area formula for Dirac's delta function

We briefly outline how to write the conditional probability (3.6) as an ordinary surface integral (3.7). Some definitions first: a function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is *quickly decaying* if

$$\lim_{\|z\| \rightarrow \infty} z^{\alpha} f(z) = 0, \quad \forall \alpha \in \mathbf{N}_0,$$

where z^{α} is declared component-wise [305]. Then the space of quickly decaying functions $f \in \mathcal{C}^{\infty}(\mathbf{R}^n)$ with quickly decaying derivatives is called Schwartz space and is denoted by $\mathcal{S}(\mathbf{R}^n)$. Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ be a smooth function, such that the fibres $\Sigma_{\xi} = \Phi^{-1}(\xi)$ are smooth submanifolds of codimension k in \mathbf{R}^n . Then for any function $f \in \mathcal{S}(\mathbf{R}^n)$ we define the Dirac measure $\delta(\Phi(z) - \xi)$ by

$$\int_{\mathbf{R}^n} f(z) \delta(\Phi(z) - \xi) dz = \int_{\Sigma_{\xi}} f(\text{vol} J_{\Phi})^{-1} d\sigma_{\xi}. \quad (\text{D.1})$$

Here $J_{\Phi} = \mathbf{D}\Phi$ denotes the Jacobian of Φ , and $d\sigma_{\xi}$ is the Hausdorff measure (surface element) of $\Sigma_{\xi} \subset \mathbf{R}^n$. The matrix volume for the rectangular matrix J_{Φ} is given by

$$\text{vol} J_{\Phi}(z) = \sqrt{\det J_{\Phi}^T(z) J_{\Phi}(z)}$$

Without loss of generality we set $\xi = 0$ and omit the argument ξ from now on. In order to show that the definition (D.1) makes sense let us introduce a non-negative function

$\varphi : \mathbf{R}^k \rightarrow \mathbf{R}$ that has compact support and which satisfies $\varphi(0) > 0$. Moreover for $\varepsilon > 0$ we define the family of functions $\delta_\varepsilon(y) = \varepsilon^{-1}\varphi(\varepsilon^{-1}y)$. The following is standard: for a test function $h \in \mathcal{S}(\mathbf{R}^k)$ we introduce the Dirac distribution $\delta(y)$ by

$$\lim_{\varepsilon \rightarrow 0} \int_{\mathbf{R}^k} \delta_\varepsilon(y)h(y) dy = \int_{\mathbf{R}^k} h(y)\delta(y) dy,$$

where the rightmost integral is defined as the point evaluation [306]

$$h(0) = \int_{\mathbf{R}^k} h(y)\delta(y) dy.$$

Here we face a slightly different problem: Using (D.1) we have to show that

$$\lim_{\varepsilon \rightarrow 0} \int_{\mathbf{R}^n} \delta_\varepsilon(\Phi(z))f(z) dz = \int_{\mathbf{R}^n} f(z)\delta(\Phi(z)) dz. \quad (\text{D.2})$$

By definition, the support of δ_ε shrinks as ε goes to zero. Therefore we can restrict the integration domain to a tubular neighbourhood $N\Sigma_\varepsilon$ of Σ with local coordinates given by the map $\phi(x, y) = \sigma(x) + y^i n_i(\sigma(x))$. Hence we have

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \int_{N\Sigma_\varepsilon \cap \mathbf{R}^n} \delta_\varepsilon(\Phi(z))f(z) dz \\ = \lim_{\varepsilon \rightarrow 0} \int_{\mathbf{R}^n} f(x, y)\delta_\varepsilon(B(x)y) \sqrt{g(x, y)} dx dy \end{aligned}$$

with the abbreviation $\sqrt{g} = \sqrt{\det g}$. Note that we have used the somehow abusive notation $f(x, y)$ for the pull-back ($f \circ \phi$)(x, y). The matrix $B = Q^T J_\Phi \in \mathbf{R}^{k \times k}$ with $Q = (n_1, \dots, n_k)$ stems from the local representation of $(\Phi \circ \phi)(x, y) = B(x)y$. We introduce a new variable ζ by setting $\zeta = B(x)y$. Thus by the above definition of the Dirac distribution $\delta(\zeta)$ the last equation becomes

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \int_{\mathbf{R}^n} f(x, B(x)^{-1}\zeta)\delta_\varepsilon(\zeta)(\det B(x))^{-1} \sqrt{g(x, B(x)^{-1}\zeta)} dx d\zeta \\ = \int_{\mathbf{R}^n} f(x, B(x)^{-1}\zeta)\delta(\zeta)(\det B(x))^{-1} \sqrt{g(x, B(x)^{-1}\zeta)} dx d\zeta \\ = \int_{\mathbf{R}^{n-k}} f(x, 0)(\det B(x))^{-1} \sqrt{g(x, 0)} dx, \end{aligned}$$

It follows from the particular form of the metric (B.2) that $\sqrt{g(x, 0)} = \sqrt{G(x)}$, where G is the metric of Σ . Observing further that $\det B(x) = \text{vol} J_\Phi(\sigma(x))$ we obtain

$$\lim_{\varepsilon \rightarrow 0} \int_{N\Sigma_\varepsilon \cap \mathbf{R}^n} \delta_\varepsilon(\Phi(z))f(z) dx = \int_{\Sigma} f(\text{vol} J_\Phi)^{-1} d\sigma,$$

which gives the assertion (D.2). We conclude by noting that the definition (D.1) is independent of the choice of any compactly supported function φ .

E. Three-scale problems

The common structure of the systems considered in Section 6 is that they involve three time scales rather than two time scales as in typical slow-fast systems. Hence we consider a generic three-scale system (for simplicity we assume that $(x, y) \in \mathbf{R} \times \mathbf{R}$)

$$\begin{aligned} \dot{x}_\varepsilon(t) &= \frac{1}{\varepsilon} f(x_\varepsilon(t), y_\varepsilon(t)) + g(x, y) \\ \dot{y}_\varepsilon(t) &= -\frac{1}{\varepsilon^2} h(x_\varepsilon(t), y_\varepsilon(t)) + \frac{\sigma}{\varepsilon} \dot{W}_2(t) \end{aligned} \quad (\text{E.1})$$

which is basically the former slow-fast system after a rescaling of time according to $t \mapsto t/\epsilon$, such that the right hand side of the slow equation has the same order of magnitude as the diffusion term in the fast equation [23, 34]; see also [307, 297].

In order to derive the limit equation for $\epsilon \rightarrow 0$, we shall employ a perturbation-like argument. To this end consider the backward equation associated with (E.1):

$$\partial_t v^\epsilon(x, y, t) = \mathcal{A}^\epsilon v^\epsilon(x, y, t) \quad (\text{E.2})$$

with

$$\mathcal{A}^\epsilon = \epsilon^{-2} \mathcal{A}_1 + \epsilon^{-1} \mathcal{A}_2 + \mathcal{A}_3,$$

and the three generators

$$\mathcal{A}_1 = \frac{\sigma^2}{2} \partial_y^2 + g(x, y) \partial_y$$

$$\mathcal{A}_2 = f(x, y) \partial_x$$

$$\mathcal{A}_3 = g(x, y) \partial_x.$$

Suppose that the fast process, generated by \mathcal{A}_1 , has a unique invariant measure $\mu_x(dy) = \rho(x, y)dy$, where the density ρ satisfies $\mathcal{A}_1 \rho = 0$.

We expand the solution of the backward equation into a perturbation series according to $v^\epsilon = v_0 + \epsilon v_1 + \epsilon^2 v_2 + \dots$ choosing an initial density $v^\epsilon(x, y, 0) = v^\epsilon(x, 0)$ that only depends on the slow variable [35]. Plugging v^ϵ into (E.2) equating powers of ϵ yields a hierarchy of equations, the first three of which are

$$\epsilon^{-2} : \mathcal{A}_1 v_0 = 0, \quad (\text{E.3})$$

$$\epsilon^{-1} : \mathcal{A}_1 v_1 = -\mathcal{A}_2 v_0, \quad (\text{E.4})$$

$$\epsilon^0 : \mathcal{A}_1 v_2 = \partial_t v_0 - \mathcal{A}_2 v_1 - \mathcal{A}_3 v_0. \quad (\text{E.5})$$

As the operator \mathcal{A}_1 acts on function in the fast variable only, and its kernel is one-dimensional, we can conclude that v_0 depends on x only. In order to unveil the lowest order time evolution (E.5) we define the projection $\Pi : L^2(\mu_x) \rightarrow \ker(\mathcal{A}_1) \subset L^2(\mu_x)$ onto the nullspace of \mathcal{A}_1 as the map

$$(\Pi u)(x) = \int u(x, y) \mu_x(dy). \quad (\text{E.6})$$

In fact Π is the conditional expectation with respect to $\mu_x(dy)$. We address the next equation (E.4). For it to be uniquely solvable in $L^2(\mu_x)$, it is helpful to see that [308]

$$u \in \text{ran } \mathcal{A}_1 \iff u \in (\ker \mathcal{A}_1)^\perp \iff \Pi u = 0.$$

Hence orthogonality to the kernel amounts to averaging of $\mathcal{A}_2 v_0$ to zero under the fast dynamics, which can be equivalently expressed by $\Pi \mathcal{A}_2 \Pi = 0$, because the kernel of \mathcal{A}_1 is one-dimensional. If the centering condition (6.11) holds, this condition is clearly satisfied, such that we can invert \mathcal{A}_1 on the second equation:

$$v_1 = -\mathcal{A}_1^{-1} \mathcal{A}_2 v_0.$$

If we insert this expression into the evolution equation (E.5), and apply the orthogonal projection onto the nullspace of \mathcal{A}_1 , we obtain the diffusive limit equation

$$\partial_t v_0(x, t) = \bar{\mathcal{A}} v_0(x, t) \quad \text{with} \quad \bar{\mathcal{A}} = \Pi \mathcal{A}_3 \Pi - \Pi \mathcal{A}_2 \mathcal{A}_1^{-1} \mathcal{A}_2 \Pi. \quad (\text{E.7})$$

In view of the fact that \mathcal{A}_2 and \mathcal{A}_3 are first-order differential operator, and \mathcal{A}_1 does not involve any x -derivatives at all, we can interpret the last equation again as a backward equation in $L^1(dx)$ with a generator that can be cast into standard form

$$\bar{\mathcal{A}} = A(x) \partial_x^2 + B(x) \partial_x,$$

and to which the following Itô stochastic differential equation is associated [33]

$$\dot{x}(t) = B(x(t)) + \sqrt{2A(x(t))}\dot{W}(t).$$

We will show below how the coefficients A, B are to be computed. Generally speaking, the procedure works by solving (E.4) which is an ordinary differential equation:

$$\mathcal{A}_1^{-1}\mathcal{A}_2v_0 = \mathcal{A}_1^{-1}f(x, y)\partial_x v_0 =: w(x, y)\partial_x v_0,$$

where the function $w(x, y)$ solves the *cell problem*

$$\mathcal{A}_1w(x, y) = f(x, y) \quad \text{with} \quad w(x, \cdot) \in (\ker \mathcal{A}_1)^\perp. \quad (\text{E.8})$$

Note that the initial conditions and the respective integration constants of the cell problem are chosen such that $w(x, \cdot)$ does not lie in the nullspace of \mathcal{A}_1 . Solving the equation subject to consistent initial conditions, $\bar{\mathcal{A}}$ can be written as

$$\begin{aligned} \bar{\mathcal{A}}v_0 = & - \int f(x, y)w(x, y)\mu_x(dy) \partial_x^2 v_0 \\ & + \int (g(x, y) - f(x, y)\partial_x w(x, y)) \mu_x(dy) \partial_x v_0 \end{aligned} \quad (\text{E.9})$$

using the definition of the conditional expectation with respect to μ_x . It remains to show that the covariance matrix of the diffusion is positive definite. Indeed by means of (E.8) the first term under the integral can be rewritten as the quadratic expression

$$f(x, y)w(x, y) = w(x, y)\mathcal{A}_1w(x, y),$$

which is strictly negative, since the spectrum of \mathcal{A}_1 , considered on functions that are orthogonal to the kernel of \mathcal{A}_1 , lies entirely on the negative real axis.

Integral representation of the averaged generator Extracting the coefficients a, b by solving the cell problem (E.8) may not be possible in general. An alternative approach [25] uses an explicit integral representation \mathcal{A}^{-1} . Thus let $g(x, \cdot)$ be orthogonal to the kernel of $L^2(\mu_x)$, i.e., $\Pi g = 0$. Then the function

$$G(x, y) = - \int_0^\infty \exp(t\mathcal{A}_1)g(x, y) dt$$

is an integral representation of $\mathcal{A}_1^{-1}g$, for

$$\begin{aligned} \mathcal{A}_1G &= - \int_0^\infty \mathcal{A}_1 \exp(t\mathcal{A}_1)g dt \\ &= - \int_0^\infty \frac{d}{dt} \exp(t\mathcal{A}_1)g dt \\ &= (1 - \lim_{t \rightarrow \infty} \exp(t\mathcal{A}_1))g, \end{aligned}$$

and \mathcal{A}_1 is negative-definite for all functions $g \in (\ker \mathcal{A}_1)^\perp$. Hence $\exp(t\mathcal{A}_1) \rightarrow 0$ and

$$w(x, y) = - \int_0^\infty \exp(t\mathcal{A}_1)f(x, y) dt,$$

which gives upon substitution into (E.9)

$$\begin{aligned} \bar{\mathcal{A}}v_0 &= \int f(x, y) \int_0^\infty \exp(t\mathcal{A}_1)f(x, y) dt \mu_x(dy) \partial_x^2 v_0 \\ &+ \int f(x, y) \partial_x \int_0^\infty \exp(t\mathcal{A}_1)f(x, y) dt \mu_x(dy) \partial_x v_0 \\ &+ \int g(x, y) \mu_x(dy) \partial_x v_0. \end{aligned}$$

Exploiting the semigroup property of $\exp(t\mathcal{A}_1)$ the coefficients become

$$\begin{aligned} A(x) &= \int f(x, y) \int_0^\infty \mathbf{E}_y f(x, y_x(t)) dt \mu_x(dy) \\ B(x) &= \int \left(g(x, y) + f(x, y) \int_0^\infty \mathbf{E}_y \partial_x f(x, y_x(t)) dt \right) \mu_x(dy), \end{aligned}$$

where $y_x(t)$ denotes the fast process at time t starting at $y_x(0) = y$, and \mathbf{E}_y labels the average over all realizations up to time t conditional on the initial value y .

F. Van Kampen's approximation

We shall demonstrate how studying the normalized deviations leads to a three-scale problem of the type (E.1). To this end consider the scaled deviation

$$\xi_\epsilon(t) = \frac{x_\epsilon(t) - x(t)}{\sqrt{\epsilon}}. \quad (\text{F.1})$$

For the sake of convenience we restrict our attention to the case $\xi_\epsilon \in \mathbf{R}$. We augment the system (6.1) by the (redundant) differential equation (6.3) for ξ_ϵ . In other words, we replace (6.1) by the joint system for $(x, \xi_\epsilon, y_\epsilon) \in \mathbf{R} \times \mathbf{R} \times \mathbf{R}$:

$$\begin{aligned} \dot{x} &= -\partial_x G(x) + \sigma \dot{W}_1 \\ \dot{\xi}_\epsilon &= -\frac{1}{\sqrt{\epsilon}} \partial_x (V(x, y_\epsilon) - G(x)) - \partial_x^2 V(x, y_\epsilon) \xi_\epsilon + \mathcal{O}(\epsilon^\infty) \\ \dot{y}_\epsilon &= -\frac{1}{\epsilon} \partial_y V(x, y_\epsilon) + \frac{\sigma}{\sqrt{\epsilon}} \dot{W}_2. \end{aligned}$$

Clearly the averaged equation for x is decoupled from the rest, but we keep it, since otherwise the system would become time inhomogeneous. The associated backward equation then has the form

$$\partial_t u^\epsilon(x, \xi, y, t) = \mathcal{A}^\epsilon u^\epsilon(x, \xi, y, t)$$

with

$$\mathcal{A}^\epsilon = \epsilon^{-1} \mathcal{A}_1 + \epsilon^{-1/2} \mathcal{A}_2 + \mathcal{A}_3,$$

where the single generators are given by

$$\begin{aligned} \mathcal{A}_1 &= \frac{\sigma^2}{2} \partial_y^2 - \partial_y V(x, y) \partial_y \\ \mathcal{A}_2 &= -\partial_x (V(x, y) - G(x)) \partial_\xi \\ \mathcal{A}_3 &= \frac{\sigma^2}{2} \partial_x^2 - (\partial_x V(x, y) + \partial_x^2 V(x, y) \xi) \partial_x. \end{aligned}$$

In contrast to the previous section the nullspace of \mathcal{A}_1 now consists of function that depend on x as well as ξ (the slow coordinates). Accordingly the projection $\Pi : L^2(\mu_x) \rightarrow \ker \mathcal{A}_1$ maps to functions $g(x, \xi)$. Quite remarkably the operator \mathcal{A}_2 meets the solvability condition $\Pi \mathcal{A}_2 \Pi = 0$. The powers of $\sqrt{\epsilon}$ in the backward equation suggests that we shall expand its solution as follows

$$u^\epsilon = u_0 + \sqrt{\epsilon} u_{1/2} + \epsilon u_1 + \dots$$

Equating powers of $\sqrt{\epsilon}$ yields a hierarchy of equations, the first three of which are

$$\begin{aligned} \epsilon^{-1} &: \mathcal{A}_1 u_0 = 0, \\ \epsilon^{-1/2} &: \mathcal{A}_1 u_{1/2} = -\mathcal{A}_2 u_0, \\ \epsilon^0 &: \mathcal{A}_1 u_1 = \partial_t u_0 - \mathcal{A}_2 u_{1/2} - \mathcal{A}_3 u_0. \end{aligned}$$

Repeating the procedure from the last section taking into account the solvability condition and the fact that the projection Π commutes with the Ornstein-Uhlenbeck generator \mathcal{A}_3 gives the familiar limit equation

$$\partial_t u_0 = \bar{\mathcal{A}}u_0 \quad \text{with} \quad \bar{\mathcal{A}} = -\Pi\mathcal{A}_2\mathcal{A}_1^{-1}\mathcal{A}_2\Pi + \Pi\mathcal{A}_3\Pi.$$

By running through the calculation from the last section, Appendix E, it can be shown that the term containing \mathcal{A}_1^{-1} yields the diffusion expression for the normalized deviation ξ without further drift (see below), whereas the rightmost term yields the averaged equation for the slow variable x and the drift $G''(x)\xi$ of the error. That is, $\bar{\mathcal{A}}$ turns out to be the generator of the skew system (6.6).

Calculation of the diffusion coefficient Consider the family of cell problems (E.8) for functions $w(x, \xi, \cdot) \in (\ker \mathcal{A}_1)^\perp$. In particular we take a look at

$$f(x, \xi, y) = -\frac{\partial}{\partial x} \left(\frac{\lambda(x)^2}{2} y^2 - \frac{\sigma^2}{2} \ln \lambda(x) \right)$$

from Example 6.1 above. The cell problem is then independent of ξ ,

$$\frac{\sigma^2}{2} \frac{d^2 w}{dy^2} - y\lambda(x)^2 \frac{dw}{dy} + \left(y^2 \lambda'(x) \lambda(x) - \frac{\sigma^2}{2} \frac{\lambda'(x)}{\lambda(x)} \right) = 0.$$

Hence $w(x, \xi, y) = w(x, y)$. The solution of the homogeneous problem is easily found:

$$w_0(x, y) = C_2(x) + \frac{\sigma\sqrt{\pi}}{2} \frac{C_1(x)}{\lambda(x)} \operatorname{Erfi} \left[\frac{\lambda(x)y}{\sigma} \right],$$

where C_1, C_2 are integration constants, that may or may not depend on the slow variable x , and $\operatorname{erfi}[z]$ is the *complex error function* that is defined by

$$\operatorname{erfi}[z] = -i \operatorname{erf}[iz] \quad \text{with} \quad \operatorname{erf}[z] = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-\zeta^2) d\zeta.$$

Variation of constants including the solvability condition (6.11) finally leads to

$$w(x, y) = C_2(x) + \frac{\sigma\sqrt{\pi}}{2} \frac{C_1(x)}{\lambda(x)} \operatorname{Erfi} \left[\frac{\lambda(x)y}{\sigma} \right] + \frac{1}{2} \frac{\lambda'(x)}{\lambda(x)} y^2, \quad (\text{F.2})$$

where the integration constant $C_1(x)$ is arbitrary, and $C_2(x)$ is determined by the requirement $\Pi w = 0$. That is, $C_2(x)$ is found to be

$$C_2(x) = -\frac{\sigma^2}{4} \frac{\lambda'(x)}{\lambda(x)^3}.$$

Intriguingly the solvability condition does not rely on $C_1(x)$ at all. For this reason we may fix $C_1 \equiv 0$ without any loss of generality. In fact, the computed coefficients do not depend on $C_1(x)$ anyway. By this we obtain the diffusion coefficient in (6.9),

$$A(x) = -\int f(x, y)w(x, y)\mu_x(dy) = \frac{\sigma^4}{4} \left(\frac{\lambda'(x)}{\lambda(x)^2} \right)^2. \quad (\text{F.3})$$

Remark F.1. Consider again the problem of inverting \mathcal{A}_1 . Clearly the result of this operation is defined only up to functions that vanish under the action of \mathcal{A}_1 ,

$$v_1 = \mathcal{A}_1^{-1}\mathcal{A}_2v_0 + \zeta \quad \text{with} \quad \zeta \in \ker \mathcal{A}_1,$$

where the additional function ζ accounts for the indeterminacy of inverting \mathcal{A}_1 . Gladly this does not change the diffusive limit equation as long as the solvability condition

$\Pi\mathcal{A}_2\Pi = 0$ is met, for then also $\Pi\mathcal{A}_2\zeta = 0$, and so the indeterminacy disappears from the effective equation. See also [85].

However we have to be very careful in relaxing the centering condition (6.11). To see what can happen consider the cell problem, and do not assume that \mathcal{A}_2v_0 be orthogonal to the kernel of \mathcal{A}_1 ; but then projecting equation (E.4) to the nullspace of \mathcal{A}_1 yields a contradiction, for $\Pi\mathcal{A}_1 = \mathcal{A}_1\Pi$ and therefore

$$0 = \Pi\mathcal{A}_1v_1 = \Pi\mathcal{A}_2v_0 \neq 0.$$

Of course the solvability condition $\Pi\mathcal{A}_2\Pi = 0$ is somehow weaker than the centering condition (6.11). Nevertheless the last equation clearly shows that the perturbation series breaks down, if the right hand side of equation (E.4) has a component in the nullspace of \mathcal{A}_1 . In fact $\Pi\mathcal{A}_2\Pi = 0$ can be considered a consistency condition for the whole ansatz to make sense.

References

- [1] P. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.*, 93:2395–2417, 1993.
- [2] C. Chipot and D.A. Pearlman. Free energy calculations. The long and winding gilded road. *Mol. Sim.*, 28(1-2):1–12, 2002.
- [3] C.H. Bennett. Molecular dynamics and transition state theory: the simulation of infrequent events. In R.E. Christoffersen, editor, *Algorithms for Chemical Computations*, pages 63–97. Amer. Chem. Soc., Washington D.C., 1977.
- [4] D. Chandler. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem Phys.*, 68:2959–2970, 1978.
- [5] E. Vanden-Eijnden and F.A. Tal. Transition state theory: Variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.*, 123:184103, 2005.
- [6] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3:107–115, 1935.
- [7] E.P. Wigner. Calculation of the rate of elementary association reactions. *J. Chem. Phys.*, 5:720–725, 1937.
- [8] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–313, 1935.
- [9] T. Mülders, P. Krüger, W. Swegat, and L. Schlitter. Free energy as the potential of mean force. *J. Chem. Phys.*, 104(12):4869–4870, 1996.
- [10] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *J. Chem. Phys.*, 109(18):7737–7744, 1998.
- [11] W.K. den Otter. Free energy from molecular dynamics with multiple constraints. *Mol. Phys.*, 98(12):773–781, 2000.
- [12] E. Darve, M.A. Wilson, and A. Pohorille. Calculating free energies using a scaled-force molecular dynamics algorithm. *Mol. Sim.*, 28(1-2):113–144, 2002.
- [13] W. E and E. Vanden-Eijnden. Metastability, conformation dynamics, and transition pathways in complex systems. In S. Attinger and P. Koumoutsakos, editors, *Multiscale, Modelling, and Simulation*, pages 35–68. Springer, Berlin, 2004.
- [14] C. Hartmann and Ch. Schütte. A geometric approach to free energy calculations. *Commun. Math. Sci.*, 3(1):1–20, 2005.
- [15] C. Hartmann and Ch. Schütte. A constrained hybrid Monte-Carlo algorithm and the problem of calculating the free energy in several variables. *ZAMM*, 85(10):700–710, 2005.
- [16] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem*, 6(9):1809–1814, 2005.
- [17] G. Ciccotti, T. Lelièvre, and E. Vanden-Eijnden. Sampling Boltzmann-Gibbs distributions restricted on a manifold with diffusions: Application to free energy calculations. *Rapport de recherche du CERMICS*, 2006-309, 2006.
- [18] C. Hartmann and Ch. Schütte. Comment on two distinct notions of free energy. *Physica D*, 2007. accepted.
- [19] E.A. Grebenikov. Methods of averaging equations in celestial mechanics. *Soviet Astronomy*, 9:146–148, 1965.
- [20] V.I. Arnold, V.V. Kozlov, and A.I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics*. Springer, New York, 1997.
- [21] P. Imkeller and J.-S. von Storch, editors. *Stochastic Climate Models*. Birkhäuser, Cambridge, 2001.
- [22] V.I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Russ. Math. Surv.*, 18:85–192, 1963.
- [23] E. Vanden-Eijnden A.J. Majda, I. Timofeyev. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, 54:891–974, 2001.
- [24] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, New York, 1984.
- [25] D. Givon, R. Kupferman, and A.M. Stuart. Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity*, 17:R55–R127, 2004.
- [26] F.A. Bornemann and Ch. Schütte. A mathematical approach to smoothed molecular dynamics: Correcting potentials for freezing bond angles. *Physica D*, 102:52–77, 1997.
- [27] F.A. Bornemann. *Homogenization of Singularly Perturbed Mechanical Systems*, volume 1687 of *Lecture Notes in Mathematics*. Springer, Berlin, 1998.
- [28] S. Reich. Smoothed dynamics of highly oscillatory Hamiltonian systems. *Physica D*, 89:28–42, 1995.
- [29] S. Reich. Torsion dynamics of molecular systems. *Phys. Rev. E*, 53:4176–4181, 1996.
- [30] T. Yanao, W.S. Koon, J.E. Marsden, and I.G. Kevrekidis. Gyration-radius dynamics in

- structural transitions of atomic clusters. *preprint*, 2006.
- [31] N De Leon, M.A. Mehta, and R.Q. Topper. Cylindrical manifolds in phase space as mediators of chemical reaction dynamics and kinetics. i. theory. *J. Chem. Phys.*, 94(12):8310–8328, 1991.
- [32] T. Uzer, C. Jaffé, J. Palacián, P. Yanguas, and S. Wiggins. The geometry of reaction dynamics. *Nonlinearity*, 15:957–992, 2002.
- [33] R.Z. Khas'minskii. A limit theorem for the solution of differential equations with random right-hand sides. *Theory Prob. Applications*, 11(3):390–406, 1966.
- [34] W. E, D. Liu, and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Commun. Pure Appl. Math.*, 58:1–42, 2005.
- [35] Ch. Schütte, J. Walter, C. Hartmann, and W. Huisinga. An averaging principle for fast degrees of freedom exhibiting long-term correlations. *SIAM Mult. Mod. Sim.*, 2(3):501–526, 2004.
- [36] O. Gonzalez and J.H. Maddocks. Extracting parameters for base-pair level models of DNA from molecular dynamics simulations. *Theoretical Chemistry Accounts*, 106:76–82, 2001.
- [37] K. Moritsugu and J.C. Smith. Langevin model of the temperature and hydration dependence of protein vibrational dynamics. *J. Phys. Chem. B*, 109:12182–12194, 2005.
- [38] I. Horenko, E. Dittmer, and A. Fischer Ch. Schütte. Automated model reduction for complex systems exhibiting metastability. *SIAM Multiscale Modeling and Simulation*, 5(3):802–827, 2006.
- [39] I. Horenko and C. Hartmann. Blind model reduction for high-dimensional time-dependent data. *Physica D*, 2006. submitted.
- [40] Y. Pokern, A. Stuart, and P. Wiberg. Parameter estimation for partially observed hypoelliptic diffusions. *preprint*, 2006.
- [41] I. Horenko, E. Dittmer, and Ch. Schütte. Reduced stochastic models for complex molecular systems. *Comp. Vis. Sci.*, 9(2):89–102, 2005.
- [42] G. Hummer and I.G. Kevrekidis. Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations. *J. Chem. Phys.*, 118:10762–10773, 2003.
- [43] A. Amadei, A.B.M. Linssen, and H.J.C Berendsen. Essential dynamics on proteins. *Proteins*, 17:412–425, 1993.
- [44] T. Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer, New York, 2002.
- [45] P. Holmes, J.L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 1996.
- [46] Ch. Schütte, W. Huisinga, and P. Deuffhard. Transfer operator approach to conformational dynamics in biomolecular systems. In B. Fiedler, editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 191–223. Springer, Berlin, 2001.
- [47] M. Weber. *Meshless Methods in Conformation Dynamics*. Hut, München, 2006.
- [48] S. Lall, J.E. Marsden, and S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Int. J. Robust Nonlin. Control*, 12:519–535, 2002.
- [49] H. Mori. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.*, 33:423–450, 1965.
- [50] R.W. Zwanzig. Nonlinear generalized Langevin equations. *J. Stat. Phys.*, 9:215–220, 1975.
- [51] J.T. Hynes and J.M. Deutch. Non-equilibrium problems - projection operator techniques. In H. Eyring, D. Henderson, and W. Jost, editors, *Physical Chemistry - An Advanced Treatise*, pages 729–836. Academic Press, New York, 1975.
- [52] D.J. Evans and G.P. Morriss. *Statistical Mechanics of Nonequilibrium Liquids*. Academic Press, London, 1990.
- [53] M.E. Tuckerman and B.J. Berne. Stochastic molecular dynamics in systems with multiple time scales and memory friction. *J. Chem. Phys.*, 95(6):4389–4396, 1991.
- [54] O.F. Lange and H. Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins*, 62(4):1053–1061, 2006.
- [55] J.L. Barber. *Application of Optimal Prediction to Molecular Dynamics*. PhD Thesis, UC Berkeley, University of California, Berkeley, 2004.
- [56] A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Natl. Acad. Sci. USA*, 97(7):2969–2973, 2000.
- [57] R. Kupferman and A.M. Stuart. Fitting SDE models to nonlinear Kac- Zwanzig heat bath models. *Physica D*, 199:279–316, 2004.
- [58] D. Givon, O.H. Hald, and R. Kupferman. Existence proof for orthogonal dynamics and the Mori-Zwanzig formalism. *Isr. J. Math.*, 145:221–332, 2005.
- [59] A.J. Chorin and O.H. Hald. *Stochastic Tools in Mathematics and Science*. Springer, New

- York, 2006.
- [60] R. Kupferman. Fractional kinetics in Kac-Zwanzig heat bath models. *J. Stat. Phys.*, 114:291–326, 2004.
- [61] D. Bernstein. Optimal prediction of burgers’s equation. *SIAM Mult. Mod. Sim.*, 6(1):27–52, 2007.
- [62] A.J. Chorin. Averaging and renormalization for the KortewegdeVriesBurgers equation. *Proc. Natl. Acad. Sci. USA*, 100, 2003.
- [63] O.H. Hald and R. Kupferman. Convergence of optimal prediction for nonlinear Hamiltonian systems. *SIAM J. Numer. Anal.*, 39(3):983–1000, 2001.
- [64] A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D*, 166:239–257, 2002.
- [65] B. Seibold. Optimal prediction in molecular dynamics. *Monte Carlo Methods and Applications*, 10(1):25–50, 2004.
- [66] R. Abraham and J.E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Massachusetts, 1978.
- [67] C. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1998.
- [68] D.W. Stroock. *An Introduction to the Analysis of Paths on a Riemannian Manifold*. AMS, Providence, 2000.
- [69] E.P. Hsu. *Stochastic Calculus on Manifolds*. AMS, Providence, 2002.
- [70] H. Federer. *Geometric Measure Theory*. Springer, Berlin, 1969.
- [71] E.A. Carter, G. Ciccotti, J.T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156(5):472–477, 1989.
- [72] S. Melchionna. Constrained systems and statistical distribution. *Phys Rev. E*, 61(6):6165–6170, 2000.
- [73] M.E. Tuckerman, Y. Liu, G. Ciccotti, and G.J. Martyna. Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian dynamics. *J. Chem. Phys.*, 115(4):1678–1702, 2001.
- [74] E.J. Hinch. Brownian motion with stiff bonds and rigid constraints. *J. Fluid. Mech.*, 271:219–234, 1994.
- [75] N.G. van Kampen and J.J. Lodder. Constraints. *Am. J. Phys.*, 52(5):419–424, 1984.
- [76] J. Kästner and W. Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *J. Chem. Phys.*, 123(14):144104, 2005.
- [77] E. Darve and A. Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, 115(20):9169–9183, 2001.
- [78] K. Kuczera. One- and multidimensional conformational free energy simulations. *J. Comp. Chem.*, 17(15):1726–1749, 1996.
- [79] G. Kalibaeva, M. Ferrario, and G. Ciccotti. Constant pressure-constant temperature molecular dynamics: a correct constrained NPT ensemble using the molecular virial. *Mol. Phys.*, 101(6):765–778, 2003.
- [80] A. Sergi. Generalized bracket formulation of constrained dynamics in phase space. *Phys. Rev. E*, 69:021109, 2004.
- [81] J.E. Marsden and T.S. Ratiu. *Introduction to Mechanics und Symmetry*. Springer, New York, 1999.
- [82] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004.
- [83] E. Cancès, F. Legoll, and G. Stoltz. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *IMA Preprint Series*, 2040:1–35, 2005.
- [84] E. Vanden-Eijnden and G. Ciccotti. Second-order integrators for Langevin equations with holonomic constraints. *Chem. Phys. Lett.*, 429:310–316, 2006.
- [85] R.Z. Khas’minskii. On stochastic processes defined by differential equations with a small parameter. *Theory Prob. Applications*, 11(2):211–228, 1966.
- [86] Yu. Kifer. Averaging and climate models. In P. Imkeller and J.-S. von Storch, editors, *Stochastic Climate Models*. Birkhäuser, Cambridge, 2001.
- [87] M. Allen and D. Tildesley. *Computer Simulations of Liquids*. Clarendon Press, Oxford, 1990.
- [88] D. Frenkel and B Smit. *Understanding Molecular Dynamics: From Algorithms to Applications*. Academic Press, London, 2002.
- [89] W.R. Hamilton. On a general method in dynamics. *Phil. Trans. Roy. Soc. Lon.*, part II:247–308, 1834.

- [90] W.R. Hamilton. Second essay on a general method in dynamics. *Phil. Trans. Roy. Soc. Lon.*, part I:95–144, 1835.
- [91] J.-P. Ryckaert and A. Bellemans. Molecular dynamics of liquid n-Butane near its boiling point. *Chem. Phys. Lett.*, 30(1):123–125, 1975.
- [92] A.R. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, Dorchester, 2001.
- [93] Y.G. Mu, D.S. Kosov, and G. Stock. Conformational dynamics of Trialanine in water. 2. comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. *Journal of Physical Chemistry B*, 107(21):5064–5073, 2003.
- [94] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry I*. Wiley, New York, 1963.
- [95] M. Spivak. *Differential Geometry*, volume 2. Publish or Perish, Boston, 1974.
- [96] M. Nakahara. *Geometry, Topology and Physics*. IOP, Bristol, 2003.
- [97] R. Abraham, J.E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer, New York, 1988.
- [98] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, New York, 1989.
- [99] B.O. Koopman. Hamiltonian dynamics and transformations in Hilbert space. *Proc. Natl. Acad. Sci. USA*, 17:315–318, 1931.
- [100] A. Lasota and M.C. Mackey. *Chaos, Fractals and Noise: Stochastic Aspects of Dynamics*. Springer, New York, 1994.
- [101] E.B. Davies. *One-parameter semigroups*. Academic Press, London, 1980.
- [102] W. Huisinga. *Metastability of Markovian systems: A transfer operator based approach in applications to molecular dynamics*. PhD Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2001.
- [103] Ya.G. Sinai. Dynamical systems with elastic reflections. *Russ. Math. Surv.*, 25(2):137–189, 1970.
- [104] D.V. Anosov. *Geodesic Flows on Closed Riemannian Submanifolds with Negative Curvature*. Proceedings of the Steklov Institute of Mathematics, No. 90 (1967). Amer. Math. Soc., 1969.
- [105] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer, Berlin, 2002.
- [106] M. Griebel, S. Knapek, G. Zumbusch, and A. Caglar. *Numerische Simulation in der Moleküldynamik: Numerik, Algorithmen, Parallelisierung, Anwendungen*. Springer, Berlin, 2003.
- [107] H.C. Andersen. Molecular dynamics simulations at constant temperature and/or pressure. *J. Chem. Phys.*, 71(4):2384–2393, 1980.
- [108] W.G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.
- [109] B.L. Holian and W.G. Hoover. Numerical test of the Liouville equation. *Phys. Rev. A*, 34(5):4229–4239, 1986.
- [110] G.J. Martyna, M.L. Klein, and M.E. Tuckerman. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(4):2635–2643, 1992.
- [111] D.J. Evans and G.P. Morriss. The isothermal/isobaric molecular dynamics ensemble. *Phys. Lett. A*, 8-9:433–436, 1983.
- [112] S.D. Bond, B.J. Leimkuhler, and B.B. Laird. The Nosé-Poincaré method for constant temperature molecular dynamics. *J. Comput. Phys.*, 151:114–134, 1999.
- [113] P.H. Hünenberger. Thermostat algorithms for molecular dynamics simulations. In C. Holm and K. Kremer, editors, *Advanced Computer Simulation: Approaches for Soft Matter Sciences I*, volume 173 of *Advances in Polymer Science*, pages 105–149. Springer, Berlin, 2005.
- [114] J.C. Mattingly and A.M. Stuart. Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Process. Related Fields.*, 8(2):199–214, 2001.
- [115] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.
- [116] H. Risken. *The Fokker-Planck equation : methods of solution and applications*. Springer, Berlin, 1996.
- [117] B. Øksendal. *Stochastic differential equations : an introduction with applications*. Springer, Berlin, 2003.
- [118] L. Arnold. *Stochastic differential equations : theory and applications*. Wiley, New York, 1974.
- [119] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29:255–284, 1966.
- [120] V.E. Tarasov. Phase-space metric of non-Hamiltonian systems. *J. Phys. A*, 38:2145–2155, 2005.
- [121] O. Cepas and J. Kurchan. Canonically invariant formulations of Langevin equations and

- Fokker-Planck equations. *Euro. Phys. J. B*, 2:221–223, 1998.
- [122] W.C. Kerr and A.J. Graham. Generalized phase space version of Langevin equations and associated Fokker-Planck equations. *Euro. Phys. J. B*, 15:305–311, 2000.
- [123] R.L.C. Akkermans. The Langevin equation for generalized coordinates. In B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, and Ch. Schütte R. Skeel, editors, *New Algorithms for Macromolecular Simulation*, volume 49 of *Lecture Notes in Computational Science and Engineering*, pages 155–164. Springer, Berlin, 2006.
- [124] E. Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, 1967.
- [125] C. Hartmann, Ch. Schütte, and J.H. Maddocks. Diffusive limit of the Langevin equation. in *preparation*, 2006.
- [126] T. Hida, K.S. Lee, and S.S. Lee. Conformal invariance of white noise. *Nagoya Math. J.*, 98:87–98, 1985.
- [127] T. Frankel. *The Geometry of Physics: An Introduction*. Cambridge University Press, 1997.
- [128] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 1991.
- [129] N. Berglund and B. Gentz. *Noise-Induced Phenomena in Slow-Fast Dynamical Systems: A Sample-Paths Approach*. Springer, London, 2006.
- [130] W. Huisinga, S. Meyn, and Ch. Schütte. Phase transitions & metastability in markovian and molecular systems. *Ann. Appl. Probab.*, 14:419–458, 2004.
- [131] V. Knecht and H. Grubmüller. Mechanical coupling via the membrane fusion SNARE protein Syntaxin 1A: A molecular dynamics study. *Biophys J.*, 84(3):1527–1547, 2003.
- [132] M. Dellnitz, K.A. Grubits, J.E. Marsden, K. Padberg, and B. Thieme. Set oriented computation of transport rates in 3-degree of freedom systems: the Rydberg atom in crossed fields. *Regular and Chaotic Dynamics*, 10(2):173–192, 2005.
- [133] F. Gabern, W.S. Koon, J.E. Marsden, S.D. Ross, and T. Yanao. Application of tube dynamics to non-statistical reaction processes. *Few-Body Systems*, 38:167–172, 2006.
- [134] P.G. Bolhuis. Kinetic pathways of β -hairpin (un)folding in explicit solvent. *Biophys J.*, 88:50–61, 2005.
- [135] P. Deuffhard and Ch. Schütte. Molecular conformation dynamics and computational drug design. In J.M. Hill and R. Moore, editors, *Applied Mathematics Entering the 21st Century: Invited Talks from the ICIAM 2003 Congress*, pages 91–119. SIAM, 2004.
- [136] M. Eichinger, B. Heymann, H. Heller, H. Grubmüller, and P. Tavan. Conformational dynamics simulation of proteins. In P. Deuffhard, J. Hermans, B. Leimkuhler, A. Mark, S. Reich, and R.D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 74–93. Springer, Berlin, 1999.
- [137] K. Karhunen. Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fenicae*, 36:1–7, 1946.
- [138] M.A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Principal Component Analysis and long time protein dynamics. *J. Phys. Chem.*, 100:2567–2572, 1996.
- [139] W. Huisinga, C. Best, F. Cordes, R. Roitzsch, and Ch. Schütte. From simulation data to conformational ensembles: Structure and dynamics-based methods. *J. Comp. Chem.*, 20:1760–1774, 1999.
- [140] D. Crommelin and A.J. Majda. Strategies for model reduction: Comparing different optimal bases. *J. Atmos. Sci.*, 61(17):2206–2217, 2004.
- [141] T. Ichiye and M. Karplus. Collective motions in proteins – a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, 11:205–217, 1991.
- [142] P.H. Hünenberger, A.E. Mark, and W.F. van Gunsteren. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.*, 252:492–503, 1995.
- [143] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, 2004.
- [144] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, New York, 2001.
- [145] G. Blanchard, M. Sugiyama, M. Kawanabe, V. Spokoiny, and K.-R. Müller. Non-Gaussian component analysis: a semi-parametric framework for linear dimension reduction. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 131–138. MIT Press, Cambridge, 2006.
- [146] K. Hasselmann. PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, 93:11015–11021, 1988.

- [147] F. Kwasniok. Empirical low-order models of barotropic flow. *J. Atmos. Sci.*, 61:235–245, 2004.
- [148] F. Kwasniok. The reduction of complex dynamical systems using principal interaction patterns. *Physica D*, 92:28–60, 1996.
- [149] T. DelSole. Optimally persistent patterns in time-varying fields. *J. Atmos. Sci.*, 58:1341–1356, 2001.
- [150] P. Deuffhard, M. Dellnitz, O. Junge, and Ch. Schütte. Computation of essential molecular dynamics by subdivision techniques. In P. Deuffhard, J. Hermans, B. Leimkuhler, A. Mark, S. Reich, and R.D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 94–111. Springer, Berlin, 1999.
- [151] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behaviour. *SIAM J. Numer. Anal.*, 36(2):491–515, 1999.
- [152] W. Huisinga and B. Schmidt. Metastability and dominant eigenvalues of transfer operators. In C. Chipot, R. Elber, A. Laaksonen, B. Leimkuhler, A. Mark, T. Schlick, Ch. Schütte, and R. Skeel, editors, *New Algorithms for Macromolecular Simulation*, volume 49 of *Lecture Notes in Computational Science and Engineering*, pages 167–182. Springer, Berlin, 2005.
- [153] P.G. Bolhuis, D. Chandler, C. Dellago, and P.L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Ann Rev. Phys. Chem.*, 59:291–318, 2002.
- [154] G. Hummer. From transition paths to transition states and rate coefficients. *J. Chem. Phys.*, 120:516–523, 2004.
- [155] W. E, W. Ren, and E. Vanden-Eijnden. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces and transition tubes. *Chem. Phys. Lett.*, 413:242–247, 2005.
- [156] W. E, W. Ren, and E. Vanden-Eijnden. Finite temperature string method for the study of rare events. *J. Phys. Chem. B*, 109(14):6688–6693, 2005.
- [157] H. Jónsson, G. Mills, and K.W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In B.J. Berne, G. Ciccotti, and D.F. Coker, editors, *Classical and Quantum Dynamics in Condensed Phase Simulations*, pages 385–404. World Scientific, 1998.
- [158] P. Metzner. *Algorithmic Identification of Transition Channels in Molecular Systems*. PhD Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2006.
- [159] J. Schmidt-Ehrenberg, D. Baum, and H.-C. Hege. Visualizing dynamic molecular conformations. *Proc. IEEE Visualization*, VIS '02:235–242, 2002.
- [160] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comp.*, 10:1299–1319, 1998.
- [161] N. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, 1993.
- [162] S. Lall, P. Krysl, and S. Glavaski. Structure-preserving model reduction for mechanical systems. *Physica D*, 184:304–318, 2003.
- [163] Y. Mu, P.H. Ngyuen, and G. Stock. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins*, 58(1):45–52, 2004.
- [164] J. Walter. *Averaging for Diffusive Fast-Slow Systems with Metastability in the Fast Variable*. PhD Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2005.
- [165] A. Crisanti, M. Falcioni, and A. Vulpiani. On the effects of an uncertainty on the evolution law in dynamical systems. *Physica A*, 160:482–502, 1989.
- [166] R.D. Skeel, G. Zhang, and T. Schlick. A family of symplectic integrators: stability, accuracy, and molecular dynamics applications. *SIAM J. Sci. Comput.*, 18:203–222, 1997.
- [167] R.D. Skeel and J. Izaguirre. The five femtosecond time step barrier. In P. Deuffhard, J. Hermans, B. Leimkuhler, A. Mark, S. Reich, and R.D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 303–318. Springer, Berlin, 1999.
- [168] S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36(5):1549–1570, 1999.
- [169] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, 125:024106, 2006.
- [170] A.J. Golumbfskie, V.S. Pande, and A.K. Chakraborty. Simulation of biomimetic recognition between polymers and surfaces. *Proc. Natl. Acad. Sci. USA*, 96(12):11707–11712, 1999.
- [171] I. Tavernelli, S. Cotesta, and E.E. Di Iorio. Protein dynamics, thermal stability, and free-energy landscapes: A molecular dynamics investigation. *Biophys J.*, 85(4):2641–2649, 2003.
- [172] V.P. Carey. *Statistical Thermodynamics and Microscale Thermophysics*. Cambridge University Press, 1999.

- [173] V. Guillemin and A. Pollack. *Differential Topology*. Prentice Hall, 1973.
- [174] H. Airault and P. Malliavin. Intégration géométrique sur l'espace de Wiener. *Bull. Sci. Math*, 112:3–52, 1988.
- [175] R.A. Horn. *Matrix analysis*. Cambridge University Press, 1987.
- [176] A. Ben-Israel and T.N.E. Greville. *Generalized Inverses: Theory and Applications*. Wiley, New York, 1974.
- [177] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1989.
- [178] M. Fixman. Classical statistical mechanics of constraints: a theorem and applications to polymers. *Proc. Natl. Acad. Sci. USA*, 71:3050–3053, 1974.
- [179] H. Rubin and P. Ungar. Motion under a strong constraining force. *Comm. Pure Appl. Math.*, 10:65–87, 1957.
- [180] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry II*. Wiley, New York, 1969.
- [181] M.P. do Carmo. *Riemannian Geometry*. Birkhäuser, Boston, 1992.
- [182] M. Spivak. *Differential Geometry*, volume 4. Publish or Perish, Boston, 1975.
- [183] N.N. Bogolyubov and Y.A. Mitropol'skii. *Asymptotic Methods in the Theory of Nonlinear Oscillations*. Gordon & Breach, New York, 1961.
- [184] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structured*. North-Holland, Amsterdam, 1978.
- [185] Yu. Kifer. Averaging in dynamical systems and large deviations. *Invent. math.*, 110:337–370, 1992.
- [186] M. Viana. *Stochastic Dynamics of Deterministic Systems*. IMPA, Rio de Janeiro, 1997.
- [187] T. Stykel. Gramian-based model reduction for descriptor systems. *Math. Control Signals Systems*, 16:297–319, 2004.
- [188] M. Ianuzzi, A. Laio, and M. Parinello. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.*, 90(23):238302, 2003.
- [189] B.-Y. Chen. *Geometry of submanifolds*. M. Dekker, New York, 1973.
- [190] T. Bröcker and K. Jänich. *Einführung in die Differentialtopologie*. Springer, Berlin, 1990.
- [191] Ya.G. Sinai. Markov partitions and C-diffeomorphisms. *Funct. Anal. Appl.*, 2:64–89, 1968.
- [192] R. Bowen and D. Ruelle. The ergodic theory of axiom A flows. *Invent. Math.*, 29(3):181–202, 1975.
- [193] J.-P. Ortega and T.S. Ratiu. *Momentum Maps and Hamiltonian Reduction*, volume 222 of *Progress in Mathematics*. Birkhäuser, Boston, 2004.
- [194] J.P. Ryckaert and G. Ciccotti. Introduction of Andersen's demon in the molecular dynamics of systems with constraints. *J. Chem Phys.*, 78(12):7368–7374, 1983.
- [195] G. Ciccotti and J.P. Ryckaert. Molecular dynamics simulation of rigid molecules. *Comput. Phys. Rep.*, 4:345–392, 1986.
- [196] R.W. Zwanzig. Problems in nonlinear transport theory. In J. Ehlers, K. Hepp, R. Kippenhahn, H.A. Weidenmüller, and J. Zittarz, editors, *Systems Far From Equilibrium*. Springer, Berlin, 1980.
- [197] O.F. Bandtlow and P.V. Coveney. On the existence of dynamical systems with exponentially decaying collision operators. *J. Phys. A*, 34:4585–4599, 2001.
- [198] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer, New York, 1999.
- [199] L.E. Ballentine. *Quantum Mechanics: A Modern Development*. World Scientific, Singapore, 1998.
- [200] K. de Leeuw. On the adjoint semigroup and some problems in the theory of approximation. *Mathematische Zeitschrift*, 73(3):219–234, 1960.
- [201] A.J. Chorin. Conditional expectations and renormalization. *SIAM Mult. Mod. Sim.*, 1:105–118, 2003.
- [202] C.T.H. Baker. A perspective on the numerical treatment of volterra equations. *J. Comp. Appl. Math.*, 125:217–249, 2000.
- [203] J. Steinberg. Numerical solution of Volterra integral equation. *Numer. Math.*, 19(3):212–217, 1972.
- [204] C. Corduneanu. *Integral Equations and Applications*. Cambridge University Press, 1991.
- [205] S. Lovesey. Single-particle motion in classical monatomic liquids. *J. Phys. C*, 6:1856–1862, 1973.
- [206] L. Sjörgen and A. Sjölander. Kinetic theory of self-motion in monatomic liquids. *J. Phys. C*, 12:4369–4392, 1979.
- [207] D. Reichman and E. Rabani. Self-consistent mode coupling theory for self-diffusion in quantum

- liquids. *Phys. Rev. Lett.*, 87(26):265702, 2001.
- [208] G. Gripenberg, S.-O. Londen, and O. Staffans. *Volterra Integral and Functional Equations*. Cambridge University Press, 1990.
- [209] N.J. Ford, J.T. Edwards, and K. Frischmuth. Qualitative behaviour of approximate solutions of some Volterra integral equations with non-Lipschitz nonlinearity. *Numerical Analysis Report, University of Manchester*, 310:1–15, 1997.
- [210] G.D. Harp and B.J. Berne. Time-correlation functions, memory functions and molecular dynamics. *Phys. Rev A*, 2(3):975–996, 1970.
- [211] T. Srokowski. Stochastic processes with finite correlation time: Modeling and application to the generalized Langevin equation. *Phys. Rev. E*, 64(3):031102, 2001.
- [212] W. Just, H. Kantz, C. Rödenbeck, and M. Helm. Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A*, 34:3199–3213, 2001.
- [213] V.I. Arnold. *Geometrische Methoden in der Theorie gewöhnlicher Differentialgleichungen*. Birkhäuser, Basel, 1987.
- [214] J.A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems*, volume 59 of *Applied Mathematical Sciences*. Springer, New York, 1985.
- [215] Yu. Kifer. Stochastic versions of Anosov’s and Neistadt’s theorems on averaging. *Stoch. Dyn.*, 1:1–21, 2001.
- [216] R.Z. Khas’minskii. On the averaging principle for Itô stochastic differential equations. *Kybernetika (Prague)*, 4:260–279, 1968.
- [217] A. Yu. Veretennikov. On rate of mixing and the averaging principle for hypoelliptic stochastic differential equations. *Math. USSR IZV*, 33(2):221–231, 1989.
- [218] R. Herbst and I. Froese. Realizing holonomic constraints in classical and quantum mechanics. *Comm. Math. Phys.*, 220:489–535, 2001.
- [219] A.I. Neishtadt. Averaging in multifrequency systems, ii. *Sov. Phys. Dokl.*, 21(2):80–82, 1976.
- [220] J. Koiller. A note on classical motion under strong constraints. *J. Phys. A*, 23:L521–L527, 1990.
- [221] F. Takens. Motion under the influence of a strong constraining force. In C. Robinson Z. Nitecki, editor, *Global Theory of Dynamical Systems*, volume 99. Springer, Berlin, 1980.
- [222] A.I. Neishtadt. Passage through a separatrix in a resonance problem with slowly varying parameter. *J. Appl. Math. Mech.*, 39(4):594–605, 1975.
- [223] A.I. Neishtadt. Passage through resonances in two-frequency problem. *Sov. Phys. Dokl.*, 20(3):189–191, 1975.
- [224] M.M. Dodson, B.P. Rynne, and J.A.G. Vickers. Averaging in multifrequency systems. *Nonlinearity*, 2:137–148, 1989.
- [225] P. Lochak and C. Meunier. *Multiphase Averaging for Classical Systems: With Applications to Adiabatic Theorems*. Springer, New York, 1988.
- [226] V.I. Arnold. Conditions for the applicability and estimate of the error of an averaging method for systems which pass through states of resonance in the course of their evolution. *Sov. Math. Dokl.*, 6:331–334, 1956.
- [227] D. Cohen, T. Jahnke, K. Lorenz, and Ch. Lubich. Numerical integrators for highly oscillatory Hamiltonian systems: a review. In A. Mielke, editor, *Analysis, Modeling and Simulation of Multiscale Problems*, pages 553–576. Springer, Berlin, 2006.
- [228] R. Klein. An applied mathematical view of meteorological modelling. In J.M. Hill and R. Moore, editors, *Applied Mathematics Entering the 21st Century: Invited Talks from the ICIAM 2003 Congress*. SIAM, Philadelphia, 2004.
- [229] A.N. Tikhonov. Systems of differential equations containing small parameters in the derivatives. *Mat. Sbornik N.S.*, 31:575–586, 1952.
- [230] I.S. Gradstein. Application of A.M. Lyapunov’s theory of stability to the theory of differential equations with small coefficients in the derivatives. *Mat. Sbornik N.S.*, 32:263–286, 1953.
- [231] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *J. Differential Equations*, 31:53–98, 1979.
- [232] C.K.R.T. Jones. Geometric singular perturbation theory. In L. Arnold, C. Jones, K. Mischaikow, G. Raugel, and R. Johnson, editors, *Dynamical Systems: Lectures Given at the 2nd Session of the Centro Internazionale Matematico Estivo (C.I.M.E.) Held in Montecatini Terme, Italy, June 13–22, 1994*, volume 1609 of *Lecture Notes in Mathematics*, pages 44–118. Springer, Berlin, 1995.
- [233] N. Berglund and B. Gentz. Geometric singular perturbation theory for stochastic differential equations. *J. Differential Equations*, 191:1–54, 2003.
- [234] H. Grubmüller. Proteins as molecular machines: Force probe simulations. In N. Attig, K. Binder, H. Grubmüller, and K. Kremer, editors, *Computational Soft Matter: From*

- Synthetic Polymers to Proteins*, volume 23 of *NIC Series*, pages 401–421. Gustav-Stresemann-Institut, Bonn, 2004.
- [235] B. Roux. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91:275–282, 1995.
- [236] A.J. Chorin, O.H. Hald, and R. Kupferman. Prediction from partial data, renormalization, and averaging. *J. Sci. Comp.*, 28(2-3):245–261, 2006.
- [237] O.H. Hald and R. Kupferman. Asymptotic and numerical analyses for mechanical models of heat baths. *J. Stat. Phys.*, 106:1121–1184, 2002.
- [238] H. Koppe and H. Jensen. Das Prinzip von Lagrange-d’Alembert in der klassischen Mechanik und in der der Quantenmechanik. *Sitzber. Heidelberg. Akad. Wiss., Math.-Naturw. Kl.*, Abh. 5, 1971.
- [239] G. Gallavotti. *The Elements of Mechanics*. Springer, New York, 1983.
- [240] O.G. Smolyanov, H. v. Weizsaecker, and O. Wittich. Brownian motion on a manifold as limit of stepwise conditioned standard Brownian motions. In F. Gesztesy, H. Holden, and J. Jost, editors, *Stochastic processes, Physics and Geometry: New Interplays. II: A Volume in Honour of S. Albeverio*, volume 29 of *Canadian Mathematical Society Conference Proceedings*, pages 589–602. AMS, Providence, 2000.
- [241] H. v. Weizsäcker, O.G. Smolyanov, and O. Wittich. Diffusions on compact riemannian manifolds and surface measures. *Dokl. Math.*, 61(2):230–234, 2000.
- [242] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov. *Modern Geometry – Methods and Applications: Part I. The Geometry of Surfaces, Transformation Groups, and Fields*. Springer, Berlin, 1984.
- [243] W.C Rheinboldt. Differential-algebraic systems as differential equations on manifolds. *Math. Comp.*, 43:473–482, 1984.
- [244] K.E. Brenan, S.L.L. Campbell, and L.R. Petzold. *Numerical solution of initial value problems in differential algebraic equations*. North-Holland, 1989.
- [245] B. Leimkuhler and R.D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comput. Phys.*, 112:117–125, 1994.
- [246] P.A.M. Dirac. Generalized Hamiltonian dynamics. *Can. J. Math*, 2:129–148, 1950.
- [247] O. Gonzalez, J.H. Maddocks, and R.L Pego. Multi-multiplier ambient-space formulations of constrained dynamical systems, with an application to elastodynamics. *Arch. Rational Mech. Anal.*, 157:285–323, 2001.
- [248] R. I. McLachlan and C. Scovel. Equivariant constrained symplectic integration. *J. Nonlinear Sci.*, 5:233–256, 1995.
- [249] J.H. Maddocks and R.L. Pego. An unconstrained Hamiltonian formulation for incompressible flows. *Commun. Math. Phys.*, 170:207–217, 1995.
- [250] B. Leimkuhler and S. Reich. Symplectic integration in constrained Hamiltonian systems. *Math. Comp.*, 63:589–605, 1994.
- [251] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York, 1999.
- [252] J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.
- [253] G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms. *J. Stat. Phys.*, 74:1117–1143, 1994.
- [254] L. Tienrey. Introduction to general state-space Markov chains. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov chain Monte-Carlo in practice*, pages 59–74. Chapman & Hall, London, 1997.
- [255] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
- [256] L. Tienrey. Markov chains for exploring posterior distributions. with discussion and a rejoinder by the author. *Ann. Statist.*, 22(4):1701–1762, 1994.
- [257] J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numer.*, 9:357–514, 2001.
- [258] R. MacKay. Some aspects of the dynamics of Hamiltonian systems. In D.S. Broomhead and A. Iserles, editors, *The Dynamics of numerics and the numerics of dynamics*, pages 137–193. Clarendon Press, Oxford, 1992.
- [259] J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.
- [260] H.C. Andersen. RATTLE: A “velocity” version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52:24–34, 1983.

- [261] J.M. Wendlandt and J.E. Marsden. Mechanical integrators derived from a discrete variational principle. *Physica D*, 106:223–246, 1997.
- [262] B. Mehlig, D.W. Heermann, and B.M. Forrest. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B*, 45(2):679–685, 1992.
- [263] B. Mehlig, D.W. Heermann, and B.M. Forrest. Exact Langevin algorithms. *Mol. Phys.*, 76:1347–1357, 1992.
- [264] G.N. Milstein and M.V. Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Numer. Anal.*, 43(3):1139–1154, 2005.
- [265] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, Berlin, 1992.
- [266] E. Barth, K. Kuczera, B. Leimkuhler, and R.D. Skeel. Algorithms for constrained molecular dynamics. *J. Comput. Chem.*, 16:1192–1209, 1995.
- [267] L. Ambrosio and H.M. Soner. Level set approach to mean curvature flow in arbitrary codimension. *J. Differential Geom.*, 43:693–737, 1996.
- [268] G.N. Milstein and M.V. Tretyakov. Quasi-symplectic methods for Langevin-type equations. *IMA Journal of Numerical Analysis*, 23(4):593–626, 2003.
- [269] T.P. Straatsma and H.J.C. Berendsen. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.*, 89:5876–5886, 1988.
- [270] R.W. Zwanzig. High-temperature equation of state by a perturbation method i. Nonpolar gases. *J. Chem. Phys.*, 22:1420–1426, 1954.
- [271] A.E. Mark. Free energy perturbation calculations. In P. v. Ragué Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, and P.R. Schreiner, editors, *Encyclopedia of Computational Chemistry*, pages 1070–1083. Wiley & Sons, Chichester, 1998.
- [272] P. Deuffhard and A. Hohmann. *Numerische Mathematik I*. de Gruyter, Berlin, 2002.
- [273] H.R. Schwarz and N. Köckler. *Numerische Mathematik*. Teubner, Wiesbaden, 2004.
- [274] I. Coluzza, M. Sprik, and G. Ciccotti. Constrained reaction coordinate dynamics for systems with constraints. *Mol. Phys.*, 101(18):2885–2894, 2003.
- [275] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte-Carlo free energy estimation: Umbrella sampling. *J. Chem. Phys.*, 23:187–199, 1977.
- [276] H. Grubmüller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E*, 52:2893–2906, 1995.
- [277] A. Laio and M. Parinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99(20):12562–12566, 2002.
- [278] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14):2690–2693, 1997.
- [279] G.E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *J. Stat. Phys.*, 90(5/6):1481–1487, 1998.
- [280] T. Lelièvre, M. Rousset, and G. Stoltz. Computation of free energy differences through nonequilibrium stochastic dynamics: the reaction coordinate case. *Rapport de recherche du CERMICS*, 2006-306, 2006.
- [281] W.F. Lau and B.M. Pettitt. Conformations of the glycine dipeptide. *Biopolymers*, 26(11):1817–1831, 1987.
- [282] P. Hamm and S. Woutersen. Coupling of the Amide I Modes of the Glycine Dipeptide. *Bull. Chem. Soc. Jpn.*, 75(5):985–988, 2001.
- [283] G.N. Milstein and M.V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, Berlin, 2004.
- [284] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in Fortran 90*. Cambridge University Press, 1996.
- [285] A. Fischer. *Die Hybride Monte-Carlo Methode in der Molekülphysik*. Diplomarbeit, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1997.
- [286] A. Einstein. Zur Theorie der Brownschen Bewegung. *Ann. Phys.*, 19:371–381, 1906.
- [287] M.S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II. irreversible processes in fluids. *J. Chem. Phys.*, 22(3):398–413, 1954.
- [288] R. Kubo, M. Yokota, and S. Nakajima. Statistical-mechanical theory of irreversible processes. ii. response to thermal disturbance. *J. Phys. Soc. Japan*, 12(11):1203–1211, 1957.
- [289] E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, 7:306–317, 2001.
- [290] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible and free. *J. Comp. Chem.*, 26:1701–1718, 2005.

- [291] METAMACS – a Java package for simulation and analysis of metastable Markov chains. Available from: <http://biocomputing.mi.fu-berlin.de/Metamacs/>.
- [292] Y. Wang and K. Kuczera. Multidimensional conformational free energy surface exploration: Helical states of ala_n and aib_n peptides. *J. Phys. Chem. B*, 101:5205–5213, 2000.
- [293] T. Conrad. *Berechnung von freie-Energie-Flächen am Beispiel des Alanin-Dipeptides*. Bachelor’s Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2003.
- [294] J. Antony, B. Schmidt, and Ch. Schütte. Nonadiabatic effects on peptide vibrational dynamics induced by conformational changes. *J. Chem. Phys.*, 122:014309, 2005.
- [295] N.G. van Kampen. Elimination of fast variables. *Phys. Rep.*, 124(2):69–160, 1985.
- [296] C. Rödenbeck, C. Beck, and H. Kantz. Dynamical systems with timescale separation: Averaging, stochastic modelling, and central limit theorems. In P. Imkeller and J.-S. von Storch, editors, *Stochastic Climate Models*. Birkhäuser, Cambridge, 2001.
- [297] G.C. Papanicolaou. Some probabilistic problems and methods in singular perturbation. *Rocky Mtn. J. Math.*, 6(4):653–674, 1976.
- [298] T.G Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Analysis*, 12:55–67, 1973.
- [299] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Comm. Math. Sci.*, 1(2):385–391, 2003.
- [300] J. Walter and Ch. Schütte. Analysis, modelling and simulation of multiscale problems. In A. Mielke, editor, *Conditional Averaging for Diffusive Fast-Slow Systems: A Sketch for Derivation*. Springer, Berlin, 2006.
- [301] I. Pavlyukevich. *Stochastic Resonance*. Logos, Berlin, 2002.
- [302] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes ii: Precise estimates for small eigenvalues. *J. Eur. Math. Soc.*, 7:69–99, 2005.
- [303] J.C. Tully. Molecular dynamics with electronic transitions. *J. Chem. Phys.*, 93(2):1061–1071, 1990.
- [304] I. Horenko, Ch. Salzmann, B. Schmidt, and Ch. Schütte. Quantum-classical liouville approach to molecular dynamics: Surface hopping gaussian phase-space packets. *J. Chem. Phys.*, 117(24):11075–11088, 2002.
- [305] D. Werner. *Funktionalanalysis*. Springer, Berlin, 1997.
- [306] L. Schwartz. *Théorie des Distributions*. Hermann, 1997.
- [307] G.C. Papanicolaou and W. Kohler. Asymptotic theory of mixing stochastic ordinary differential equations. *Commun. Pure Appl. Math.*, XXVII:641–668, 1974.
- [308] P.R. Kramer and A.J. Majda. Stochastic mode reduction for the immersed boundary method. *SIAM J. Appl. Math.*, 64(2):368–400, 2003.