Chapter 1

Complementary Aspects

The described method was developed and constructed to reweight dynamical data in form of Markov State Models (MSM) between non-equilibrium steady states (NESS). The procedure in its current form is based on general ideas on non-equilibrium statistical mechanics and coarse-graining of trajectories by MSMs. We want to discuss in this chapter how these ideas can be expanded in other applications than reweighting simulated data.

The defined ensembles and the parallels to equilibrium statistical mechanics raises the question if we can push this analogy further. In particular we will discuss if there is an invariant measure, similar to the density of states (see section **??**) that depends on the interaction of the system but not on thermodynamic variables. Other questions raised is the role of partition functions and the relation to thermodynamic variables based on derivatives. These relations in connection to the Maximum Caliber were discussed by Dill et al. [1] for trajectory ensembles under global constraints. Equilibrium statistical ensembles were also expanded to quantum mechanical systems. In combination with the discrete nature of MSMs, it can be beneficial to analyse such systems in the context of the presented ensembles. The laser system in section **??** is such an example. It was approximated classically because simulation of quantum mechanical systems is beyond the scope of this thesis.

The definition of the NESS ensembles is based on local entropy productions between discretised states. We will examine how these discrete view on the dynamics is related to the underlying continuous trajectories. This will open up another way

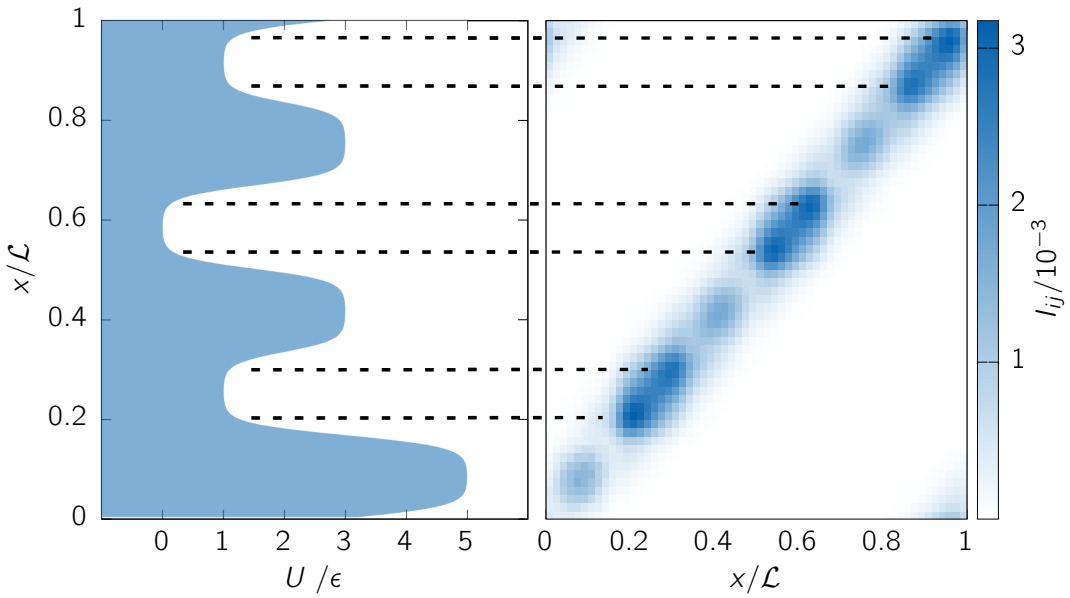

Figure 1.1: The invariant of the 1D system. The potential surface is shown to indicate the position of the maxima.

of estimating local entropy productions. More generally one could ask if the same constraints would apply in the non-discretised trajectory space.

1.1 Invariant Measure

An invariant has the property of not changing under a set of transformations. The density of states in canonical ensemble is one example. It is defined under change of the Boltzmann measure. We expect a physical interpretation of the invariant based on the physical meaning of the Boltzmann factor. It is the number of microscopic states depending on chosen variables like energy and possibly other variables (see section **??**). Once we have shown that systems are recovered under a certain transformation, we expect an underlying invariant measure. For the set of transformations used here, the *dynamical invariant* was already identified in section **??** by

$$I_{ij}^2 = q_{ij}q_{ji}\exp\left(\frac{c_i + c_j}{2} + \zeta\right) \tag{1.1}$$

and is presented for the minimal 1D system (see section **??**) in figure 1.1. Knowing the invariant of a system, one can reweight to any ensemble that we are interested in. The reweighting formula becomes

$$p_{ij} = I_{ij} \exp\left(\frac{1}{2}(\tilde{\zeta} + \Delta S_{ij}) + \frac{1}{4}(\tilde{c}_i + \tilde{c}_j)\right), \tag{1.2}$$

where $\tilde{\mathbf{c}}$ is determined by enforcing $\sum_k p_{ik} = 1$ and applying the numerical solver used for the reweigthing formula. We want to understand the physical meaning of this quantity and show the technical use of it in enhanced sampling methods.

At first intuition the dynamical invariant is the density or number of trajectories connecting two microstates within lagtime τ, equivalent to the density of states. However, this quantity is difficult to grasp because time and length of trajectories have unknown relation without further information. We need thermodynamic properties like the diffusion coefficient $D = T\mu$, defined by the Einstein relation via temperature T multiplied with the mobility μ. For an ensemble of free-diffusive single-particles starting at position $X = 0$ and time $t_0 = 0$ with diffusion coefficient D, the probability distribution at time t is given by the solution to the Fokker-Planck equation [11]:

$$P(X, t) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{X^2}{4Dt}\right). \tag{1.3}$$

The probability distribution is interested as the density of trajectories of length t in free diffusion. The assumption still holds in NESS conditions because a generalised diffusion coefficient was shown to depend on an effective temperature [6]. In any case, the invariant in figure 1.1 does not agree with the expected invariant. In fact it shows marks that originate from the potential and peak on top of the potential barrier and on the sides of the potential wells. The middle of the potential wells are slightly lower populated than its sides. This shows that the invariant depends on the temperature as expected, but also in the potential surface. We believe that this dependence originates from the approximation used by the reweighting. A free diffusion invariant is expected for the full solution. An invariant for the full solution can be calculated if a numerical solution to the reweighting is found.

Nevertheless the approximated invariant holds some information. We note from the definition that the matrix is symmetric. All asymmetric information about dissipative effects are given by the local entropy productions. The invariant on the other hand contains information about symmetric local fluctuations, or non-dissipative effect. These contributions were distinguished by Maes [7] showing that symmetric

and asymmetric contributions to the dynamics are decoupled in equilibrium and coupled in off-equilibrium. The symmetric or *frenetic* contribution can be defined by $A_{ij} = \sqrt{p_{ij}p_{ji}}$ such that $p_{ij} = A_{ij} \exp(\Delta S_{ij}/2)$. The frenetic contribution changes with reweighting as predicted by Maes. The reweighting method recovers these frenetic contributions correctly so we assume the invariant to contain the necessary symmetric information. Accordingly we interprete the invariant as the non-dissipative local fluctuations of the system.

This point of view explains why the reweighting scheme does not apply to temperature-reweighting. The chosen constraints infer information about dissipative effects. The changes in the system under variation of temperature lie in the non-dissipative part of the dynamics so we cannot expect it work. It requires a different set of constraints that is unknown to us. In fact it might be difficult to find such constraint: We assume that the same trajectories can be sampled in any thermodynamic envirenment, only the probability of a trajectory changes. It was shown that a trajectory sampled at a certain temperature has unique fluctuation properties and cannot be sampled at another temperature, i.e. is has measure 0 []. A microscopic view on the trajectories does not allow temperature reweighting. In our mesoscopic view using the discrete MSM we might be saved by integrating out such microscopic properties.

The invariant is not completely understood, but it can be of computational purpose since it can be sampled from different simulations. This is the underlying idea of equilibrium enhanced sampling methods like multicanonical simulation [4], replica exchange [10] or metadynamics [5]. The general idea is to tweak the system such that regions of interest are sampled more frequently. This can be done by increasing the temperature to sample a high-energy states more frequently or adding artificial potentials to push or attract the system in a certain direction. The changes do not have to be of physical nature and are removed from the system after taking advantage of better sampling conditions. All gathered information are collected in the invariant as long as the corresponding relations are known by a reweighting formula. These relation are well-known for the equilibrium ensembles. Such methods can be applied to sample dynamics in NESS. This is a step to reach experimentally important timescales that are inaccessible by the limitation of computational power [8].

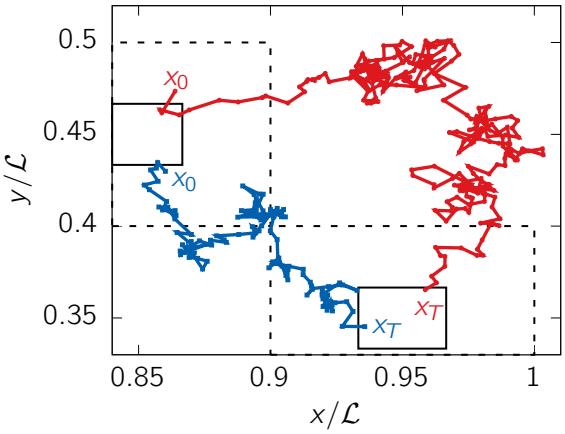

Figure 1.2: Two example trajectories from simulation of the 2D system between two microstates.The dashed boxes show the microstates, the solid boxes show the core of the microstates. A trajectory is counted when it starts and enters in a core-state.

1.2 Trajectory Analysis

MSMs are a coarse view on the dynamics of a system. The reweighting procedure acts on the MSM by enforcing the local entropy productions. We want to analyse how the single trajectories relate to the local entropy productions and how they can be calculated from the trajectories. The trajectory space allows us a detailed view on the system and one can judge if the microstate of the MSM and the assumptions for the analytic solution (see section **??**) are chosen correctly. The analytic solution ignores effects of fluctuations on the entropy production, the trajectory analysis takes them into account. Furthermore, the analytic solution for the entropy production only works when the new forces are applied along the collective variables (CV) chosen for the MSM. A trajectory analysis allows us to calculate entropy productions for forces depending on other system variables.

MSMs are sampled by counting all trajectories that start in a given microstate i and finish in another microstate j. The transition probabilities are defined by a trajectory ensemble average (see section **??**) connecting two microstates

$$\langle \chi_i(\mathbf{x}_0)\chi_j(\mathbf{x}_T)\rangle = \int\int \mathrm{d}\mathbf{x}_0\, \mathrm{d}[\mathbf{x}(t)]\, p[\mathbf{x}(t)|\mathbf{x}_0]p(\mathbf{x}_0)\chi_i(\mathbf{x}_0)\chi_j(\mathbf{x}_T), \tag{1.4}$$

where $\chi_i(\mathbf{x})$ is 1 if $\mathbf{x}$ is in state i and 0 else, $\mathbf{x}_0$ is the initial point of the trajectory $\mathbf{x}(t)$

and $\mathbf{x}_T$ the final point. Similarly we use this to calculate the local entropy production from trajectory space

$$\begin{aligned}\langle S\rangle_{ij} &= \langle \chi_i(\mathbf{x}_0)\chi_j(\mathbf{x}_T)\rangle \\ &= \int\int \mathrm{d}\mathbf{x}_0\, \mathrm{d}[\mathbf{x}(t)]\, p[\mathbf{x}(t)|\mathbf{x}_0]p(\mathbf{x}_0)\chi_i(\mathbf{x}_0)\chi_j(\mathbf{x}_T)\Delta S[\mathbf{x}(t)],\end{aligned} \tag{1.5}$$

where $\Delta S[\mathbf{x}(t)]$ is the entropy production of a trajectory that is calculated by

$$\Delta S[\mathbf{x}(t)] \approx \frac{1}{2T}\sum_d \sum_{t=1}^{t=T} \left(x_t^{(d)} - x_{t-1}^{(d)}\right)\left(F^{(d)}(\mathbf{x}_t) + F^{(d)}(\mathbf{x}_{t-1})\right) \tag{1.6}$$

in discrete form as produced by the simulation, discussed in section **??**. The relevant trajectories are chosen as illustrated in figure 1.2. The dashed lines represent a microstate, the box in the middle is the core state. A trajectory starts in a core state and terminates when the final core state is reached. The center state ensures that fluctuations on the boundaries are not taken into account. A maximum length of trajectory is set for computational efficiency and does not influence the distribution of entropy productions. The distribution of the entropy productions are shown for chosen transitions on the 1D and 2D model in figure 1.3. The average of the distribution is shown by the dashed black line, the solid yellow line represents the average as calculated from the analytic model. The two values agree well, indicating that the analytic approximation works well on the coarse-grained level. Larger distances reduce the number of available trajectories and produce more noise. The full difference of analytic and measured value is shown in figure 1.4 for the 1D system and locally for the 2D system. We consider the trajectory analysis to be exact, because it takes all information into account and is sufficiently sampled for the given systems. The absolute deviations are small compared to the absolute entropy productions in the range $[-10\frac{\epsilon}{K}, 10\frac{\epsilon}{K}]$. Yet the 1D systems shows a pattern: When a trajectory start or ends in a state dominated by large local forces the deviation of the analytic solution is larger but still in an acceptable range. The deviation might origin in a non-symmetric distribution of entropy productions. This indicates that the fluctuations do not each other out and the analytic assumptions is flawed. A third way to estimate the entropy production is use the transition probabilities of the MSM by $\ln \Delta S_{ij} = \frac{p_{ij}}{p_{ji}}$. This method agrees with the others but suffers from sampling issues for larger distances and needs longer simulation times to gather sufficient data, so it is not of further interest. The analytic method gives a good estimate and can thus be used for estimation of entropy productions without additional data used. The trajectory method can estimate the entropy production well but samples the

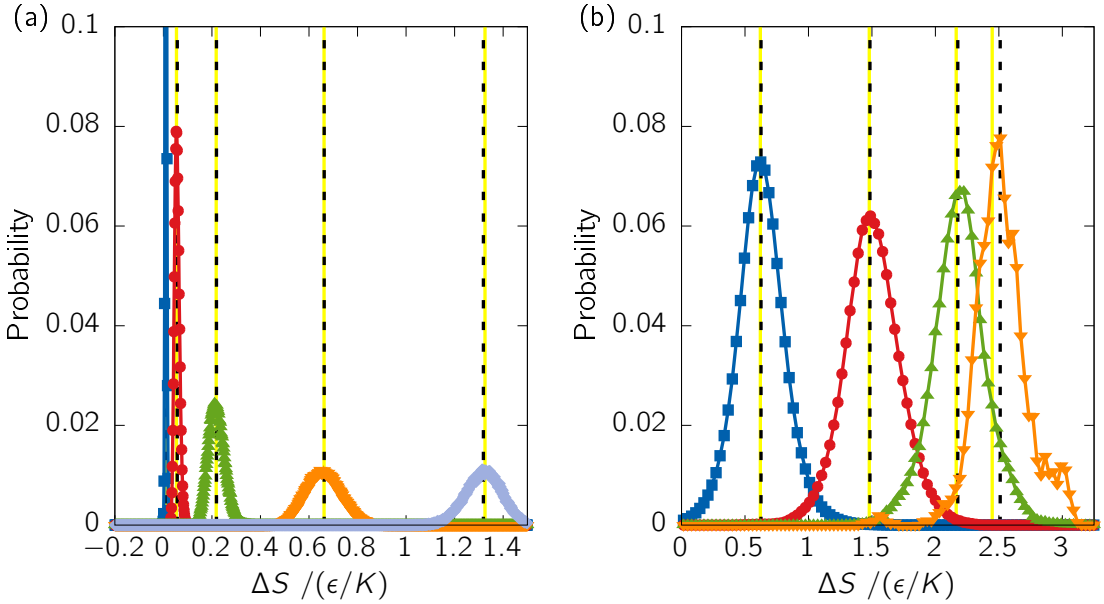

Figure 1.3: Distribution of entropy productions for (a) the single particle in 1D potential and (b) single particle in 2D potential for chosen transitions. The average of the distribution is shown by a black dashed line, the analytic solution is shown by a solid yellow line.

whole distribution. It gathers more information than needed for the reweighting and requires simulation of the target system.

The trajectory analysis is useful when one does not want to reweight the CVs of the MSM. The analytic method can only be used when the forces are defined on the same CV. Otherwise sampling the trajectories of the target system is the only way to estimate the target system for reweighting. One can gather data by sampling the system at different thermodynamic states and gather the information at any of these states afterwards. Continuous reweighting is not possible in this case. The second use is testing for double peaks in the distribution of entropy productions. It indicates that two microstates are connected by more than one set of pathways. The MSM should be changed to resolve these pathways by a different choice of microstates. Otherwise the MSM is not able to distinguish all pathways and it would cause errors in the reweighting procedure.

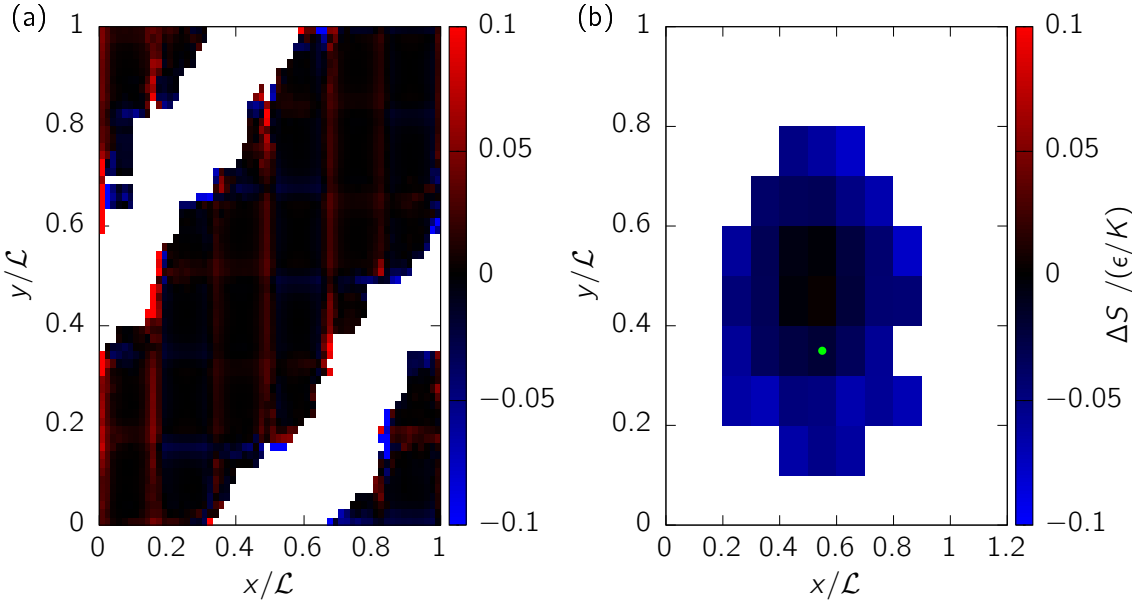

Figure 1.4: Deviation of local entropy production between analytic solution and averaging over simulated trajectories for (a) a single particle on a 1D potential surface and (b) a single particle on a 2D potential surface for a starting point at (0.55,0.35), marked by the green dot.

1.3 Off-Equilibrium Reweighting in Literature

Reweighting of dynamical data in and out of equilibrium has attracted some attention lately. There are three works focussing on the problem for off-equilibrium dynamics. We want to discuss the assumptions of these methods and compare them to our presented method.

Two methods relying on the Girsanov transformation [2, 12] reweight each trajectory in conformational coordinates from the reference data individually. The first work by Warren and Allen relies on long trajectories and is tested successfully on a birth-death process in NESS. The authors point out two practical problems of the method: First they note that long trajectories are needed for unbiased reweighting and the length is set by the unknown target system. Second, the distribution of trajectory weights becomes broader with increasing trajetory length. These problems of long trajectories were avoided by Donati and Keller by employing short trajectories with Markov State Models. The method requires full data on the trajectory and

the random numbers generated during the simulation for the Langevin thermostat. Storing all this information would require large computational effort. In practice the lagtime of the MSM and the target state is defined before running the reference simulation and the trajectories are reweighted on runtime. It is shown that the equilibrium MSM of a Hexapeptide can be recovered. The method is combined with metadynamics [3] for efficient sampling of dynamics and construction of MSM. The reweighting method is in theory applicable for off-equilibrium MSMs, albeit it was not shown yet. In contrast, our method relying on the Maximum Caliber is applied a posteriori on a MSM. The target system can be chosen freely after the simulation data are gathered. The two methods mentioned above rely on reweighting every trajectory individually, forcing them to reweight during simulation runtime. Furthermore our method requires minimal computational effort by only solving a set of non-linear convex equations. The problem of long trajectories discussed by Warren and Allen was adressed by Markov State Modeling for both, the method of Donati and Keller and our method.

Another new reweighting method applicable for NESS by Zuckermann et al. [9] was published by submission of this thesis. A number of short markov-like trajectories are collected from a reference simulation. The target system for reweighting is defined by its probabilities distribution. The weights of the reference trajectories are updated individually until they recover the target probability distribution and are stationary. This algorithm can be extended to NESS by adding wells and sinks of probability at chosen positions. In this case the trajectories have to be arranged such that the probability is transported from well to sink while meeting the requirements on probabilities distribution. We refer to the paper for details of the algorithm. The method uses space discretisation like our method but does not require markovianity unlike our method and the Girsanov reweighting. The trajectories are treated in their continuous form, similar to the Girsanov reweighting. This method was not yet tested on NESS systems.

The sparse number of works on reweighting dynamics in off-equilibrium shows that it is relatively new problem. Most of the method are only expected to work in off-equilibrium but are not tested yet. The presented methods rely on different theoretical bases and assumptions and operate on different spaces(long or short trajectories, MSM). We hope to see continuous work on this problem in the future.

Bibliography

[1] Purushottam D Dixit, Jason Wagoner, Corey Weistuch, Steve Pressé, Kingshuk Ghosh, and Ken A Dill. Perspective: Maximum caliber is a general variational principle for dynamical systems. *The Journal of chemical physics*, 148(1):010901, 2018.

[2] Lorenzo Donati, Carsten Hartmann, and Bettina G Keller. Girsanov reweighting for path ensembles and markov state models. *The Journal of chemical physics*, 146(24):244112, 2017.

[3] Luca Donati and Bettina G Keller. Girsanov reweighting for metadynamics simulations. *The Journal of Chemical Physics*, 149(7):072335, 2018.

[4] Wolfhard Janke. Multicanonical monte carlo simulations. *Physica A: Statistical Mechanics and its Applications*, 254(1-2):164–178, 1998.

[5] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[6] Boris Lander, Udo Seifert, and Thomas Speck. Mobility and diffusion of a tagged particle in a driven colloidal suspension. *EPL (Europhysics Letters)*, 92(5):58001, 2011.

[7] Christian Maes. *Non-dissipative effects in nonequilibrium systems*. Springer, 2018.

[8] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Current opinion in structural biology*, 31:64–74, 2015.

[9] John D Russo, Jeremy Copperman, and Daniel M Zuckerman. Iterative trajectory reweighting for estimation of equilibrium and non-equilibrium observables. *arXiv preprint arXiv:2006.09451*, 2020.

[10] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[11] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[12] Patrick B Warren and Rosalind J Allen. Trajectory reweighting for non-equilibrium steady states. *Molecular Physics*, 116(21-22):3104–3113, 2018.