1 Markov State Modeling

Markov State Modeling (MSM) aims to map the slow dynamics of a complex system to an underlying discrete *markovian* process. This involves several steps, including phase space discretization, time discratization and dimensional reduction. There is numerous literature about the mentioned topics [**?**, **?**, **?**] and they are not part of this thesis. The analysed models are phenemological and constructed such that space discretization is performed manually and dimensional reduction is not needed.

1.1 Transfer Operator

All existing realisations of a process are combined in an ensemble of trajectories. Considering each trajectory occurs with equal probability, a probability distribution $P(x)$ can be constructed. This distribtion is evolved in time according to a generator, containing all dynamical information by knowledge of a all trejectories passing point x. The generator $\mathscr{L}$ is denoted by

$$P(x_{t+\tau}) = \mathscr{L} \circ P(x) \tag{1}$$

The initial probability distribution can be chosen according to physical conditions and evolves towards a steady state distribution π. The infinitesimal generator $\mathscr{L}$ is inconvenient for computational purposes and is replaced by a transfer operator $\mathscr{T}$ that evolves the system over a timerange τ. We assume the generator to be *Markovian*, as will be described in detail in section **??**. The generator is defined by

$$P_{t+\tau}(y) = \mathscr{T}(\tau) \circ P_t(y) = \int_\Omega \mathrm{d}x p(x, y; \tau) P_t(x) \tag{2}$$

where Ω represents the full phase space of the system and $p(x, y; \tau)$ is a transition probability in continuous space from state y to state x within time τ. The Operators are related by

$$\mathscr{T}(\tau) = \exp\left(\mathscr{L}\tau\right). \tag{3}$$

where the Eigenvectors Ψ_i of both are the same and the Eigenvalue λ_i of $\mathscr{T}$ and Λ_i of $\mathscr{L}$ are related by $\lambda = \exp(\tau\Lambda)$. This relations are needed to construct the infinitesimal generator from a transfer operator. The Eigenvalue decomposition has a maximal Eigenvalue 1 with the corresponding Eigenvector being the stationary distribtion. It is assumed to be singular such that the stationary distribution of a system is unique. The smaller Eigenvalues are related to the dynamics of the system as described in section **??**.

1.2 Markov Property

Time discretization needs special attention such that the resulting process fulfills the *Markov property*.

A Markov process is defined as a *stochastic process* (see section **??**) where for n successive timepoints ($t_1 < t_2 < .. < t_n$) the probabilty distribution $P(y_i, t_i)$ given the value y_i only depends on the information at the previous timepoint t_{i-1}. We denote

$$P(t_n, y_n|t_1, y_1; t_2, y_2; ...; t_{n-1}, y_{n-1}) = P(t_n, y_n|t_{n-1}, y_{n-1}) \tag{4}$$

as the time dependent markovian *transition probabilty* and note that the time evolution of a probabilty distribution can be expressed by the hierarchy

$$P(t_1, y_1; t_2, y_2; t_3, y_3) = P(t_1, y_1)P(t_2, y_2|t_1, y_1)P(t_3, y_3|t_2, y_2), \tag{5}$$

where $P(t_1, y_1)$ is the initial probabilty distribution. The Markov process is fully characterized by knowledge of the transition probabilties and the initial probabilty distribution. Integrating over y_2 and dividing by $P(t_1, y_1)$ gives

$$P(t_1, y_1|t_3, y_3) = \int \mathrm{d}y_2 P(t_2, y_2|t_1, y_1)P(t_3, y_3|t_2, y_2) \tag{6}$$

where the defintion of conditional probabilties was used on the left-hand side [**?**]. Equation **??** is known as the Chapman-Kolmogorov equation and will be used to test if the Markov property is fulfilled.

1.2.1 Validate Markovian Process

The time-dependence of the transition probabilty is dropped to suffice the focus on NESS and a constant timestep length called *lagtime* τ is assumed. The time-independent Chapman-Kolmogorov equation becomes

$$P(y_i|y_j, n\tau) = \int \mathrm{d}y_1 \int \mathrm{d}y_2 ... \int \mathrm{d}y_{j-1} P(y_1|y_i, \tau)P(y_2|y_1, \tau)...P(y_j|y_{j-1}, \tau) \tag{7}$$

for n time steps and the transition probabilties depend on the lagtime. The state space of a MSM is discretised so we denote $P(y_i|y_j, \tau) = p_{ij}(\tau)$ as the jump probabilty from state i to j within time τ.
In practice it is cumbersome to check the Chapman-Kolmogorov equation for each element so Prinz et al. [**?**] suggested to expand the irreducable transition probabilty matrix $p_{ij}(\tau)$ in its Eigenvalue decomposition

$$p\Phi_k = \lambda_k \Phi_k. \tag{8}$$

The time evolution of the probabilty distribution can then be described by

$$P(t) = \sum_k c_k \Phi_k \exp\left(-\lambda_k t\right), \tag{9}$$

where c_k are constants that are determined by the initial state of the system. If the dynamics of the system fulfil detailed balance (i.e. system is in equilibrium), the Eigenvalues and Eigenvectors are real and one can construct a hierarchy

starting with the slowest process $\lambda_0 = 1 > \lambda_1 > ... > 0$. Typically the slow dynamical processes are of interested and the fast processes with low λ are cut off from the expansion. Faster vibrations were already deleted by state space discretisation. The characteristic timescales are identified from equation **??** by $t_i = \frac{-1}{\log \lambda_i}$ and are calculated for different lagtimes. Figure **??** shows an example of the slowest timescales for varying lagtimes. The region where $t_i < \tau$ is forbidden because a observed timescale cannot be smaller than the minimal timestep of the Markov process. The timescales reach a plateau for large enough lagtimes, indicating that the Chapman-Kolmogorov equation **??** is valid in this region, i.e. the slowest dynamics of the system are equally described by all choices of lagtimes. A test of consistency of the Eigenvectors is needed for full comformation of Markovianity, but is as inefficient as comparing single transition probabilty matrix entries. It is merely a tool to choose a lagtime for further analysis. It is suggested to identify metastable states and compare detailed relaxation probabilties from this state to the trajectory data. The relaxation profile of the MSM should be within the errorbars of the trajectory data as shown in figure **??** for full conformation, that the model defined by space discretisation and lagtime is markovian for the slowest processes. The Eigenvalue decomposition allows detailed analysis of a MSM by isolating each process and showing detailed probabilty fluxes involved via the corresponding Eigenvector. The described methods rely on the transition probabilty matrix being symmetric and the Eigenvalue decomposition being real-valued. In NESS this condition is not met and the Eigenvalues may become complex. A timescale seperation with a hierarchy to delete fast processes with small real part of the Eigenvalues is not possible [**?**]. A similarly powerfull tool for analysis of NESS is not known yet, however the Schur-decomposition might be a good candidate for timescale seperation. In this thesis, a reference equilibrium system is used to choose a lagtime for further analysis. First-passage-time-distributions (FPTD) are used to confirm markovianity for systems in and out of equilibrium and for analysis of the processes involved.

1.3 First-Passage-Time Distribution

First-Passage-Times Distributions are widely used to characterize processes in biology, chemistry and physics and are often associated with a free-energy barrier a system has to overcome. The FPTD contains detailed transition information by collecting numerous realisations of a process. In experiment and simulation of rare events, only the mean of the process is given due to limited observered process realisations [**?**, **?**]. Given a MSM, the metastable states are identified by the steady state distribution (see chapter **??**) and the FPTD between all metastable states are calculated. A metastable state can consist of several microstates, the collection of intitial states is denoted by I , of final states by F. For the purpose of calculating the FPTS from I to F the MSM is modified such that all final microstates fF become a sink, i.e. all jumps out of the metastable state have probability 0 and staying in the state has probability

1

$$\begin{aligned} \tilde{p}_{fj} &= 0 \forall j, \forall f F \\ \tilde{p}_{ff} &= 1 f F. \end{aligned} \tag{10}$$

An initial state is defined, where full probability is in one microstate iI of the starting metastable state

$$\begin{aligned} \rho_i^{(0)} &= 1 \\ \rho_j^{(0)} &= 0 \forall j \neq i, \end{aligned} \tag{11}$$

where the superscript denotes the number of iterations. The probability distributions is iterated with the modified Markov model $\hat{p}$ until all probability is trapped in the sink.

$$\vec{\rho}^{(t+1)} = \hat{p}\vec{\rho}^{(t)}. \tag{12}$$

The first-passage probability $\mathrm{FPT}(t)$ at step t is then the probability flow in the sink

$$\mathrm{FPT} i \to F(t) = \sum_{lF} \rho_l^{(t)} - \rho_l^{(t-1)}. \tag{13}$$

This calcultion is repeated for all initial microstates i. The FPTD is then calculated by weighting each distribution with the probability of a trajectory starting in each initial microstate

$$\mathrm{FPT} I \to F(t) = \sum i \in I \frac{\sum_{k \notin I} \pi_k p_{ki}}{\sum i \in I \left(\sum_{k \notin I} \pi_k p_{ki} \right)} \mathrm{FPT} i \to F(t). \tag{14}$$

Knowing the FPTD, all momnents of the distrubtion can be calculated

$$M^n_{I \to F} = \sum_t \mathrm{FPT}_{I \to F}(t) * t^n \tag{15}$$

1.4 Entropy Production