Zuse Institute Berlin

Takustrasse 7
D-14195 Berlin-Dahlem
Germany

PETER KOLTAI[1], GIOVANNI CICCOTTI[2], AND CHRISTOF SCHÜTTE[1,3]

[1] *Institute for Mathematics, Freie Universität Berlin, Germany*

[2] *Department of Physics, University of Roma "La Sapiensa", Italy*

[3] *Zuse Institute Berlin, Germany*

On metastability and Markov state models for non-stationary molecular dynamics

ZIB Report 16-11 (March 2016)

Zuse Institute Berlin
Takustrasse 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

On metastability and Markov state models for non-stationary molecular dynamics

Péter Koltai[*], Giovanni Ciccotti[†], and Christof Schütte[‡]

August 2, 2016

Abstract

Unlike for systems in equilibrium, a straight-forward definition of a metastable set in the non-stationary, non-equilibrium case may only be given case-by-case—and therefore it is not directly useful any more. We generalize the concept of metastability by relying on the theory of coherent sets. Based on this, we can derive finite-time non-stationary Markov state models. We illustrate this concept and its main differences to equilibrium Markov state modeling on simple, one-dimensional examples.

Keywords: coherent set, Markov state model, non-equilibrium molecular dynamics, metastability

MSC classification: 60J20 (primary), 60J35, 60J60 (secondary)

1 Introduction

Metastable molecular systems under non-equilibrium conditions caused by external fields have attracted increasing interest recently. For example, new experimental techniques like atomic force microscopy or simulation studies regarding the potential effects of electromagnetic radiation on the human body tissue have been extensively investigated in the literature. Specifically adapted molecular dynamics (MD) simulations have proved particularly useful for understanding the response of biomolecular conformations to external fields. Despite this significance, reliable tools for the quantitative description of non-equilibrium phenomena like the conformational dynamics of a molecular system under external forcing are still lacking.

For MD simulations in equilibrium such specific and reliable tools have been developed: Markov State Models (MSMs) allow for an accurate description of the transitions

[*] Institute of Mathematics, Freie Universität Berlin, Germany. E-mail: peter.koltai@fu-berlin.de.

[†] University of Roma "La Sapienza", Italy. E-mail: giovanni.ciccotti@roma1.infn.it.

[‡] Institute of Mathematics, Freie Universität Berlin, Germany; Zuse Institute Berlin, Germany. E-mail: Christof.Schuette@fu-berlin.de.

between the main conformations of the molecular system under investigation. MSMs for equilibrium MD have been well developed over the past decade in theory [SS13, PWS+11], applications (see the recent book [BPN14] for an overview), and software implementations [STSM+12, BaJLM+11]. They now form a set of standard tools. The principal idea of equilibrium MSMs is to approximate the MD system (in continuous state or phase space) by a reduced Markovian dynamics over a finite number of (macro-)states (i.e., in discrete state space). These (macro-)states represent the dominant metastable sets of the system, i.e., sets in which typical MD trajectories stay substantially longer than the system needs for a transition to another such set [SS13, SNL+11]. In equilibrium MD, these metastable sets are the main conformations of the molecular system under consideration which, often enough, are given by the main wells in its energy landscape. It has been shown that for many (bio)molecular systems the Markovian dynamics given by an MSM allows very close approximation of the longest relaxation processes of the underlying molecular system under equilibrium conditions [SNS10, DSS12].

However, in non-equilibrium settings with time-dependent external fields acting on the system the energy landscape depends on time, i.e., in principle the main wells of the energy landscape can move in time. That is, there may no longer be time-independent metastable sets in which the dynamics stays for long periods of time before exiting. Instead, the potentially metastable sets will move in state space. Generally, moving "metastable" sets cannot be anymore considered metastable. However, the so-called *coherent sets*, which have been studied for non-autonomous flow fields in fluid dynamics [FSM10], and have been theoretically discussed for systems described by ordinary differential equations in [Fro13], permit to get a meaning to the concept of metastability. This article will generalize the concept of metastability by utilizing coherent sets for diffusion processes in an energy landscape. Molecular dynamics is a possible application, and we will show how to build MSMs, for nonequilibrium MD, based on coherent sets.

2 Setting

We start with a diffusion process in a *time-dependent* potential $V : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$,

$$d\boldsymbol{x}_t = -\nabla V(t, \boldsymbol{x}_t)dt + \varepsilon d\boldsymbol{w}_t\,. \tag{1}$$

Here $\boldsymbol{w}_t$ is a standard Wiener process (Brownian motion), and $V(t,x) = V_{\text{int}}(x) + V_{\text{ext}}(t,x)$ with V_{int}, time-independent, which characterizes the inherent time-scales of the molecular system, and V_{ext}, the time-dependent external field. Here, and in the following, bold face symbols denote random variables. The noise intensity $\varepsilon = 2\beta^{-1}$ is a function of the inverse temperature β such that the dynamics would be ergodic with respect to the stationary density $\mu \propto \exp(-\beta V)$ if the energy landscape V were independent of t. The fact that the potential (or energy landscape) V explicitly depends on time means that in general there is no stationary probability distribution and no more meaning for ergodicity.

In principle, our diffusion model (1) could also be replaced by molecular dynamics equations of motion, e.g., by a mass-scaled Langevin equation,

$$\begin{aligned} d\boldsymbol{x}_t &= \boldsymbol{p}_t dt \\ d\boldsymbol{p}_t &= -\nabla V(t,\boldsymbol{x}_t)dt - \gamma \boldsymbol{p}_t dt + \varepsilon\sqrt{\gamma} d\boldsymbol{w}_t\,, \end{aligned} \tag{1a}$$

but we stick to (1) for simplicity of presentation.

By a sufficiently fine discretization (i.e. partition) of the phase space, we can approximate the Kolmogorov backward equation associated to (1),

$$\frac{\partial}{\partial t} f(t,x) = \frac{\varepsilon^2}{2}\Delta f(t,x) - \nabla_x V(t,x)\cdot\nabla_x f(t,x), \qquad f(0,\cdot) = f_0\,, \tag{2}$$

which describes the evolution of observables $f(t,x) = \mathbb{E}_{\boldsymbol{x}_0=x}[f_0(\boldsymbol{x}_t)]$. We obtain in the discretized state space

$$\dot{v}(t) = L(t)v(t), \tag{3}$$

where $L(t)\in\mathbb{R}^{N\times N}$ is the time-dependent generator. The discretization can be made such that this matrix is a generator indeed[1] [Kol10, LMHS11, FJK13]. The associated master equation, describing the propagation of probability distributions over the discrete state space, reads as

$$\dot{\mu}(t) = L(t)^T\mu(t)\,, \tag{4}$$

and approximates the Kolmogorov forward (or Fokker–Planck) equation associated to (1)

$$\frac{\partial}{\partial t} g(t,x) = \frac{\varepsilon^2}{2}\Delta g(t,x) + \nabla_x\cdot\big(V(t,x)g(t,x)\big), \qquad g(0,\cdot) = g_0\,, \tag{5}$$

describing the evolution of distributions over state space. Let the associated propagator, i.e. the evolution operator of (3), be given by $P(s,t)$, meaning that $t\mapsto P(s,t)v_s$ solves (3), given the initial condition $v(s) = v_s$. Then, of course, $P(s,t)^T$ is the solution operator of (4). Note that, due to the time-dependence of the dynamics, in general there will be no invariant distribution, and the system will *not* be *reversible*.

We wish to “extend” the notion and the treatment of metastability to this case. There is a subtle flavor to this, since there is no guarantee that the sets, which we would like to call metastable in this case, are fixed in time. To set the stage, let us start by looking at two asymptotic regimes.

3 Asymptotic regimes

In principle the dynamical behavior of (1) depends on the relation between the inherent timescales of the molecular system and the timescale on which the external fields change. We will have to distinguish at least the following three regimes.

[1] These are usually finite-volume type discretizations, which means that one partitions the part of the phase space of interest into disjoint sets, X_i, $i = 1,\dots,N$, with piecewise smooth boundary (e.g. hyperrectangles or Voronoi cells), and the v_i approximate the average value of f on X_i. The matrix entry $L(t)_{ij}$ is then the probability flux from X_i to X_j under (1), cf. [Kol10, Algorithm 5.15 and Section 5.4.1].

Very slow external field. Consider the *snapshot* systems (note that the snapshot t is fixed, i.e., the generator—and thus the external field—is frozen in time, and ϑ takes the role of time)

$$\frac{dv}{d\vartheta}(\vartheta) = L(t)v(\vartheta)\,. \tag{6}$$

If the implied time scales of these snapshot systems are much shorter than the time scale on which the change in the external field and the generator $L(t)$ takes place, then the original process (3) equilibrates *before* the external field can change a lot. Hence, on the time scale ϑ metastable behavior can be observed, where the metastable sets at time t are exactly those corresponding to $L(t)$. That is, the metastable sets move but slowly compared with the slowest internal relaxation timescales of the system. Accordingly, the system behaves in a quasi-stationary or adiabatic manner. If $\mu^*(t)$ denote the snapshot invariant distributions, i.e. $\mu^*(t)^T L(t) = 0^T$, then the distribution $\mu_t(\vartheta)$ of the system is drawn quickly to $\mu^*(t)$, and stays close for a sufficiently long while to this time-dependent equilibrium.

Quickly changing external field. The other extreme case, when the external field changes so quickly that the system can barely react, is more subtle. If the external forcing is sufficiently weak and faster than all other internal timescales of the system, then the system sees just a time-averaged ("blurred") potential. However, in general the situation is more complex and one has to be careful regarding the amplitude of the external field: if this is too large, the system will behave essentially as an ideal gas in an external field. Still, for a sufficiently weak field another situation can occur when the molecular system exhibits motion on a wide range of timescales, where the fastest ones could be much faster than the timescale on which the potential is changing, but the slowest ones might still be much slower than that.

In general, this situation does not lead to any simple solution, and should be treated as the general case. Instead, in the specific case of a fast but periodic external forcing, as e.g. discussed in [WSCDS14], using the periodic form of the forcing it is possible to build a MSM based on a quasi-stationary approach with a time-dependent family of metastable sets [SW15].

In between extremes. If the slowest internal timescales, i.e., the expected transition times between the main conformations, are comparable to the timescales on which the external field changes, *and* the external field is strong enough to alter the metastable behavior, then there is an interdependence between conformational switching and the motion of conformations induced by the external forcing. This case has not been studied up to now, at least not regarding MSM building, and will be the topic of the present work.

4 Finite-time coherent sets

4.1 Coherent pairs

We will introduce the concept of coherence for sets of our discrete state space system described by (4). The state space is now $S = \{1, \dots, N\}$, where we think of each $i \in S$ as corresponding to a small subset of $\mathbb{R}^d$, a "box". This way we may think very naturally of a stochastic process $(\boldsymbol{y}_t)_{t\geq 0}$ on S with law given by (4) as an approximation of the original process $(\boldsymbol{x}_t)_{t\geq 0}$ on $\mathbb{R}^d$. This approximation is "in distribution", meaning that the distribution of $(\boldsymbol{y}_t)_{t\geq 0}$ approximates that of $(\boldsymbol{x}_t)_{t\geq 0}$ given by (5).

With $t_0 = 0$ and $t_1 = 1$, defining a time-scale $\tau = t_1 - t_0$, we will consider the process $(\boldsymbol{y}_t)_{t\in[0,1]}$. Of course, everything applies for general t_0, t_1 as well. Let the process at time $t_0 = 0$, $\boldsymbol{y}_0$, be distributed according to μ_0 (denoted by $\boldsymbol{y}_0 \sim \mu_0$), and let at final time $\boldsymbol{y}_1 \sim \mu_1$. Then, if $P := P(0,1)$ is the associated propagator, we have $\mu_1 = P^T \mu_0$.

The definition of metastability for a time-independent system consists of finding sets A such that when starting an ensemble of realizations of the process in A, after a given *lag-time* $\tau = t_1 - t_0$ the majority of these realizations are still in A. Clearly, for time-dependent systems such a set A may not exist. Nevertheless, there could be a second set B, which has "size comparable to that of A", such that the majority of trajectories starting in A ends up in B. The notion of finite-time coherent pairs formalizes this idea. It originates from transport-based consideration of non-autonomous flow fields in fluid dynamics [FSM10, Fro13, FPG14].

Definition 1 (Coherent pairs)**:** We call a pair of sets $A, B \subset S$ *coherent* (on the chosen time interval $[t_0, t_1]$), if *both* of the following conditions are satisfied:

(i) *The forward condition:* If the process starts at initial time in A, then it ends up at final time with high probability in B; i.e.

$$\mathbb{P}[\boldsymbol{y}_1 \in B \,|\, \boldsymbol{y}_0 \in A] \approx 1 \,. \tag{7}$$

(ii) *The backward condition:* If the process ends up at final time in B, then it was at initial time with high probability in A; i.e.

$$\mathbb{P}[\boldsymbol{y}_0 \in A \,|\, \boldsymbol{y}_1 \in B] \approx 1 \,. \tag{8}$$

4.2 Characterization of coherence

Conditions (i) and (ii) of Definition 1 can be recast as

$$\mathbb{P}[\boldsymbol{y}_0 \in A,\, \boldsymbol{y}_1 \in B] \approx \mathbb{P}[\boldsymbol{y}_0 \in A] = \mu_0(A)\,, \tag{7*}$$

and

$$\mathbb{P}[\boldsymbol{y}_0 \in A,\, \boldsymbol{y}_1 \in B] \approx \mathbb{P}[\boldsymbol{y}_1 \in B] = \mu_1(B)\,. \tag{8*}$$

To simplify the notations to be used in the following development, we define the Euclidean and the ν-weighted (here, ν is some probability measure on S) scalar product of two vectors u, v by $\langle u,v\rangle := \sum_{i\in S} u_i v_i$, and $\langle u,v\rangle_\nu := \sum_{i\in S} u_i v_i \nu_i$, respectively. Further, we define the indicator vector $\mathbb{1}_A \in \mathbb{R}^N$ of a set $A \subset S$ by

$$\mathbb{1}_{A,i} := \begin{cases} 1, & i \in A, \\ 0, & i \notin A, \end{cases} \tag{9}$$

and the diagonal matrix $D_\nu := \mathrm{diag}(\nu)$. The joint probability can now be rewritten in different forms:

$$\begin{aligned} \mathbb{P}[\boldsymbol{y}_0 \in A,\, \boldsymbol{y}_1 \in B] &= \sum_{i\in A,\, j\in B} \mu_{0,i} P_{ij} && (10)\\ &= \langle \mathbb{1}_B, P^T D_{\mu_0} \mathbb{1}_A \rangle && (11)\\ &= \langle P\mathbb{1}_B, \mathbb{1}_A \rangle_{\mu_0} = \langle \mathbb{1}_A, P\mathbb{1}_B \rangle_{\mu_0} && (12)\\ &= \langle \mathbb{1}_B, D_{\mu_1}^{-1} P^T D_{\mu_0} \mathbb{1}_A \rangle_{\mu_1} && (13) \end{aligned}$$

In view of (13), $D_{\mu_1}^{-1} P^T D_{\mu_0}$ can be viewed as a *forward operator* (matrix) from a μ_0-weighted space into a μ_1-weighted space, since it transports the probability associated with $\mathbb{1}_A$ (i.e. the vector $D_{\mu_0}\mathbb{1}_A$) from initial time to final time, with respect to the μ_1-weighted scalar product.

Summarizing (7*) and (8*) with the new notation, a coherent pair satisfies the two conditions of coherence

$$\langle \mathbb{1}_A, \mathbb{1}_A \rangle_{\mu_0} \approx \langle \mathbb{1}_B, P^T D_{\mu_0} \mathbb{1}_A \rangle \approx \langle \mathbb{1}_B, \mathbb{1}_B \rangle_{\mu_1}\,. \tag{14}$$

Suppose A, B are a coherent pair. Then, we show in Appendix A that the two conditions of coherence imply

$$D_{\mu_1}^{-1} P^T D_{\mu_0} \mathbb{1}_A \approx \mathbb{1}_B\,, \tag{15}$$

where the approximate equality of to vectors is meant as "up to small error in the norm defined by $\langle\cdot,\cdot\rangle_\mu$", where $\mu = \mu_1$ in (15). In other words, if A, B is a coherent pair, then the forward operator maps the indicator of A approximately to the indicator of B. Note how this equation incorporates the forward and backward conditions of coherence: all that is in A, *goes* to B, and by

$$D_{\mu_1}^{-1} P^T D_{\mu_0} \mathbb{1}_{S\setminus A} = D_{\mu_1}^{-1} P^T D_{\mu_0} (\mathbb{1}_S - \mathbb{1}_A) \overset{(36)}{=} \mathbb{1}_S - D_{\mu_1}^{-1} P^T D_{\mu_0} \mathbb{1}_A \overset{(15)}{\approx} \mathbb{1}_S - \mathbb{1}_B = \mathbb{1}_{S\setminus B}, \tag{16}$$

what *does not come* from A, does not end up in B, or, equivalently, what ends up in B, *comes* from A.

The definition of coherence is based on pairs of coherent sets. It would be desirable to find a definition involving only one set. This requires the introduction of a backward process, as we see below.

4.3 The forward-backward process

Equation (15) allows us to reduce the characterization of coherence from a coherent *pair* of sets (one at initial and one at final time) to *only one* set (at initial time).

Substituting (15) into the equality of (12), and using the forward-backward condition (7*), we obtain

$$\langle \mathbb{1}_A, PD_{\mu_1}^{-1}P^T D_{\mu_0}\mathbb{1}_A\rangle_{\mu_0} \approx \mu_0(A)\,, \tag{17}$$

We are going to define a *forward-backward process.* Then we will see that the left-hand side of this expression is the probability that a forward-backward process starting in A, ends up in A.

To this end, let us consider the process $(\boldsymbol{y}_t)_{t\in[0,1]}$ and then define the *time-reversed* process $(\tilde{\boldsymbol{y}}_t)_{t\in[0,1]}$ by the transition probabilities (no Einstein convention is used)

$$\tilde{P}_{ij} \coloneqq \mathbb{P}[\tilde{\boldsymbol{y}}_1 = j \,|\, \tilde{\boldsymbol{y}}_0 = i] = \frac{\mu_{0,j}P_{ji}}{\mu_{1,i}}\,. \tag{18}$$

In matrix notation, $\tilde{P} = D_{\mu_1}^{-1}P^T D_{\mu_0}$.

Now, let us define the *forward-backward* process $(\boldsymbol{z}_t)$, by the transition matrix

$$C \coloneqq P\tilde{P} = PD_{\mu_1}^{-1}P^T D_{\mu_0}\,. \tag{19}$$

C leaves μ_0 invariant, i.e. $C^T\mu_0 = D_{\mu_0}PD_{\mu_1}^{-1}P^T\mu_0 = \mu_0$. Further, the forward-backward process is reversible with respect to μ_0, since C is self-adjoint with respect to the μ_0-weighted scalar product, $\langle Cu, v\rangle_{\mu_0} = \langle u, Cv\rangle_{\mu_0}$ for all $u, v \in \mathbb{R}^N$,[2] or, equivalently,

$$C^T D_{\mu_0} = D_{\mu_0}C\,. \tag{20}$$

C is also positive semidefinite with respect to this scalar product, since $\langle u, Cu\rangle_{\mu_0} = u^T D_{\mu_0}Cu = (u^T D_{\mu_0}PD_{\mu_1}^{-1/2})(D_{\mu_1}^{-1/2}P^T D_{\mu_0}u) \geq 0$.

Expression (17) is connected with the forward-backward process; in fact we have

$$\begin{aligned}\langle \mathbb{1}_A, PD_{\mu_1}^{-1}P^T D_{\mu_0}\mathbb{1}_A\rangle_{\mu_0} &= \langle \mathbb{1}_A, (D_{\mu_0}PD_{\mu_1}^{-1}P^T)D_{\mu_0}\mathbb{1}_A\rangle \\ &= \langle \mathbb{1}_A, D_{\mu_0}^{-1}C^T D_{\mu_0}\mathbb{1}_A\rangle_{\mu_0}\,.\end{aligned} \tag{21}$$

With this, analogous considerations to those made in Appendix A show that (17) implies

$$D_{\mu_0}^{-1}C^T D_{\mu_0}\mathbb{1}_A \approx \mathbb{1}_A\,, \tag{22}$$

which is the forward-backward analogue of (15). In other terms, the probability in the set A, i.e. $D_{\mu_0}\mathbb{1}_A$, is left almost invariant under transport by the forward-backward process $(\boldsymbol{z}_t)$, i.e. multiplication by C^T. With (20) and (22) we obtain

$$C\mathbb{1}_A \approx \mathbb{1}_A\,. \tag{23}$$

[2] This is equivalent to the detailed balance condition $\mu_{0,i}C_{ij} = \mu_{0,j}C_{ji}$.

When (23) is satisfied, we say that the set A is coherent.

It seems from (23) as if almost invariance under the forward-backward process could be true without the existence of a set B such that A builds a coherent pair with B. Fortunately, this is not the case, and therefore (23) is equivalent to the pair of equations (7) and (8) defining coherence. To this end, let us consider C as the propagator of a "long" *forward process on the time interval* $[0,2]$. In fact, if we consider $C = P\tilde{P}$, the propagator of the "long" process, then P acts as propagator on $[0,1]$, and $\tilde{P}$ as propagator on $[1,2]$. Comparing (22) with (15), we see that (23) means that the pair A, A is coherent for the "long" forward process (with propagator C). Now, by Theorem 1 from Appendix B we find at time $t = 1$ a set B with $\mu_1(B) = \mu_0(A)$ such that (15) holds, and A, B are a coherent pair for our original process.

This notion of coherence allows for the extension of the concept of metastability to non-stationary systems. To do that, we need first to establish the result that if A is coherent for a time τ then it is also coherent for any $\tau' \leq \tau$ (monotony of coherence, see Appendix B). Even then, the time-scale τ will play a double role: on the one hand it is the length of a time window on which we have coherence, on the other hand it is now a time-scale for which these sets stay coherent. However, beyond τ the coherence will be, in general, lost, and only a new search can tell us if there are again metastabilities. In any event, we can now give the following definition.

Definition 2 (Metastability for non-stationary systems)**:** Given a time-scale τ, a set $A \subset S$ is called *metastable* in the time range $[t_0, t_0+\tau/2]$, if A is coherent on the time interval $[t_0, t_0+\tau]$.

Note, metastability now is not just the property of the set A at the time t_0, but involves the family of sets, $(A_t)_{t\in[t_0,t_0+\tau/2]}$, such that $A_t, A_{t'}$ are coherent pairs for every $t_0 \leq t < t' \leq t_0 + \tau/2$, and every A_t from this family stays coherent for at least a time $\tau/2$. In other words, metastability is not an instantaneous property. However, coherence on a time interval of length τ is needed to define metastability on a time window of length $\tau/2$.

4.4 Perturbed invariance and identification

Now, we are interested to identify not just one metastable set, but to partition the whole state space into metastable sets. Our intention is to try to construct—as done by MSM—an essentially rigorous coarse-grained dynamics. We restrict our attention to the case where the whole state space S can be partitioned into coherent sets.

By our definition, given a coherent set A, also the complement $S \setminus A$ is coherent. We are interested in the finest possible partition that is still coherent. This necessitates that the partition is disjoint. Otherwise, by taking intersections of non-disjoint partition elements as independent partition elements, we arrive at an equally coherent, finer partition. Condition (I2) below is due to this.

Let us assume that P has n *perfectly coherent* sets A_k, $k = 1, \ldots, n$, on some given fixed time interval. That means,

(I1) the forward-backward process returns to A_k with probability one, provided it started there, i.e., $\sum_{j\in A_k} C_{ij} = 1$ for every $i \in A_k$.

Furthermore,

(I2) we assume that no finer decomposition into perfectly coherent sets is possible.

Then, it can be seen easily that these conditions (I1)–(I2) imply that

- the A_k constitute a *disjoint partition* of the state space, i.e. $\dot\bigcup_{k=1}^n A_k = S$ (where $\dot\bigcup$ means that the sets we build union of are disjoint);
- there are sets B_k, $k = 1, \dots, n$, such that the A_k, B_k build perfectly coherent pairs of the forward process, i.e., if and only if the forward process starts in A_k, it will end up with probability one in the corresponding B_k at final time;[3]
- C from (19) is a block-diagonal matrix with blocks according to the A_k, meaning that $C_{ij} = 0$ whenever $i \in A_k$, $j \notin A_k$ for any k;[4] and
- all the diagonal blocks of C are irreducible (reducibility of a block would contradict (I2)).

Thus, C has an n-fold (degenerate) eigenvalue $\lambda = 1$. The corresponding right eigenvectors can be chosen as $u_k = \mathbb{1}_{A_k}$, and left eigenvectors as $v_k = D_{\mu_0}\mathbb{1}_{A_k}$, where μ_0 is an invariant measure of P (since C is reducible, μ_0 is not unique, in fact, it can be any linear combination of the invariant measures of the single blocks).

Standard perturbation arguments for self-adjoint matrices [Kat84, Chapter II, §6.2 and §6.3] show, that if C satisfies conditions (I1), (I2) just up to an ϵ error, and is an irreducible matrix, then $\lambda_1 = 1$ is a single eigenvalue of C with right and left eigenvectors $u_1 = \mathbb{1}$, $v_1 = D_{\mu_0}\mathbb{1}$ (now, μ_0 is unique due to irreducibility), and there are $n-1$ *real* eigenvalues satisfying $1 - \lambda_k = \mathcal{O}(\epsilon)$, $k = 2, \dots, n$. Moreover, also the subspace

$$\operatorname{span}(u_1, \dots, u_n) \approx \operatorname{span}(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}), \tag{24}$$

up to an error of order ϵ, i.e. the u_k can be expressed up to an error ϵ as linear combinations of the $\mathbb{1}_{A_k}$. Hence, for sufficiently small ϵ, the right eigenvectors u_k are linear combinations of "almost indicator vectors", which allows for the algorithmic identification of coherent sets [DW04]: for any fixed $k = 1, \dots, n$, we can find scalar constants $c_{k,1}, \dots, c_{k,n}$ such that, by components, $u_{1,i} \approx c_{k,1}, \dots, u_{n,i} \approx c_{k,n}$ for all $i \in A_k$.

[3] This is due to the monotony result Lemma 1. Perfect coherence means also that A_k, A_k is a perfectly coherent pair of the forward-backward process: without loss, we can think of the forward-backward process associated to the forward process on $[0,\tau]$, as just a forward process on $[0, 2\tau]$, which has the coherent pairs A_k, A_k (at initial time 0 and final time 2τ). Then, by Lemma 1 there has to be a B_k at any intermediate time—hence also at half-time of the forward-backward process, which is the final time τ of the forward process—such that A_k, B_k are a perfectly coherent pair.

[4] This can be seen from $C_{ij} = \sum_{\ell\in S} \mu_{0,i} P_{i\ell} \mu_{1,j}^{-1} P_{j\ell}$, where $\mu_0, \mu_1 > 0$, since we have by the second bullet '•' in the list that for all $\ell \in S$ either $P_{i\ell} = 0$ (if $\ell \notin B_k$), or $P_{j\ell} = 0$ (if $\ell \in B_k$), thus, in our hypothesis ($i \in A_k$, $j \notin B_k$), every term in the sum is zero.

Remark 1 (Metastability for time-dependent and time-independent dynamics)**:** Having characterized the relationship between metastability and spectral analysis, we can now point out the difference (or not) between the time-independent and time-dependent case. The latter requires the concept of coherence and is based on the spectral analysis of the matrix C. The former, instead, can be performed by simply referring to the spectral analysis of the matrix P. If the matrix P is time-reversible (i.e. satisfies detailed balance with invariant distribution μ_0), P and C share the same eigenvectors (because then, by reversibility, $\tilde{P} = P$, and thus $C = P^2$), so that also in the time-dependent case one could study metastability by spectral analysis of P [SS13]. However, for the non-reversible case, this is not true any more.

Remark 2: It could be inconvenient to elaborate the concept of coherence for the definition of metastability by using "crisp" sets (and so the associated concept of "crisp" macrostate). Alternatively, see Appendix C, one can develop the same idea using the formalism of fuzzy macrostates.

5 Towards Markov state models

Markov state models (MSMs) [SS13] as referred to in equilibrium molecular dynamics are Markov chains with a very small state space, designed to capture the dominant time scales of the original system. Note that if the system is in equilibrium, then its propagator $P(s,t)$ only depends on $(t-s)$, i.e. $P(s,t) = e^{(t-s)L}$ with a time-independent generator L, and hence its eigenvalues $\kappa_k = e^{(t-s)\Lambda_k}$ are exponentially decaying in $(t-s)$, with rates (eigenvalues of L) $\Lambda_k < 0$, $k \geq 2$. The *implied time scales* of the system are then the inverse rates, $|\Lambda_k|^{-1}$. The rates closest to zero are of the largest interest. A MSM is a (τ-dependent) stochastic matrix $\hat{P} \in \mathbb{R}^{n\times n}$, $n \ll N$, such that the eigenvalues of $\hat{P}$, $\hat{\kappa}_k$, approximate the dominant ones of $P(s,s+\tau)$; in other words $\hat{\kappa}_k \approx e^{\tau\Lambda_k}$, $k = 1,\dots,n$.

The "physical" connection to the original process is that one associates the states of the MSM with metastable sets of the original process. More precisely, we have the coarse-grained state space $\hat{S} = \{1,\dots,n\}$, such that the original state space S is partitioned into the sets A_k, i.e. $S = \dot\bigcup_{k=1}^{n} A_k = \dot\bigcup_{k\in\hat{S}} A_k$. Further, we have a coarse-graining function $\varphi : S \to \hat{S}$ such that $\varphi(i) = k$ if $i \in A_k$.

Our intention is to extend this concept to the scenario where metastable sets can only be understood in the sense of Definition 2, i.e. in form of coherent sets. Since we only use information about the system from a finite time-interval, we cannot have a MSM in the "classical" way, where one is able to infer something about the system for large τ. We are only able to infer statistical properties which show up on a time-scale $\tau' \leq \tau/2$.

Let us carry out the modeling for n coherent pairs, A_k, B_k, $k = 1,\dots,n$. As already mentioned in section 4, we restrict our attention to the case where the coherent sets constitute a full partition of the state space, namely $\dot\bigcup_{k=1}^{n} A_k = S = \dot\bigcup_{k=1}^{n} B_k$. In practice, it might not be feasible to satisfy this condition, however, one is interested in an "almost complete partition in measure", such that $\mu_0(\bigcup A_k) \approx 1 \approx \mu_1(\bigcup B_k)$. We will discuss below the consequences of this.

Given an initial time t_0, and a time-scale τ, yielding the terminal time $t_1 = t_0 + \tau$, we can define a (one-step) non-stationary MSM.

Definition 3 (Non-stationary MSM)**:** We call the $\hat{S}$-valued *one-step* process $(\hat{\boldsymbol{y}}_{t_0}, \hat{\boldsymbol{y}}_{t_1})$, where $\hat{S} = \{1, \dots, n\}$, a MSM of the original process $\boldsymbol{y}_t$, if

(i) $\hat{\mu}_{0,i} := \mathbb{P}[\hat{\boldsymbol{y}}_{t_0} = i] = \mathbb{P}[\boldsymbol{y}_{t_0} \in A_i]$, for $i \in \hat{S}$; and

(ii) the propagator of $\hat{\boldsymbol{y}}_t$, the matrix $\hat{P} \in \mathbb{R}^{n\times n}$, satisfies

$$\hat{P}_{ij} := \mathbb{P}[\hat{\boldsymbol{y}}_{t_1} = j \,|\, \hat{\boldsymbol{y}}_{t_0} = i] = \mathbb{P}[\boldsymbol{y}_{t_1} \in B_j \,|\, \boldsymbol{y}_{t_0} \in A_i]\,. \tag{25}$$

A few remarks are in order:

- Any one state $i \in \hat{S}$ at different times may correspond to a different set in phase space. For instance, $\hat{\boldsymbol{y}}_{t_0} = i$ corresponds to $\boldsymbol{y}_{t_0} \in A_i$, but $\hat{\boldsymbol{y}}_{t_1} = i$ corresponds to $\boldsymbol{y}_{t_1} \in B_i$. Thus, in our two-time scenario, rows of $\hat{P}$ correspond to sets at initial time, and columns of $\hat{P}$ correspond to sets at final time.

- We have from (25) and (12) that

$$\hat{P}_{ij} = \frac{\langle P\mathbb{1}_{B_j}, \mathbb{1}_{A_i}\rangle_{\mu_0}}{\langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_{\mu_0}}\,. \tag{26}$$

- Moreover, $\hat{\mu}_1 := \hat{P}^T\hat{\mu}_0$ componentwise is $\hat{\mu}_{1,j} = \mu_1(B_j)$, since we have by $\hat{\mu}_{0,j} = \mathbb{P}[\boldsymbol{y}_{t_0} \in A_j] = \langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_{\mu_0}$ and (26), that

$$\hat{\mu}_{1,j} = \sum_{i=1}^n \hat{\mu}_{0,i}\hat{P}_{ij} = \sum_{i=1}^n \langle P\mathbb{1}_{B_j}, \mathbb{1}_{A_i}\rangle_{\mu_0} = \langle P\mathbb{1}_{B_j}, \mathbb{1}\rangle_{\mu_0} = \langle \mathbb{1}_{B_j}, \underbrace{P^T\mu_0}_{=\mu_1}\rangle = \mu_1(B_j)\,. \tag{27}$$

  Hence, $\hat{\boldsymbol{y}}_t$ is indeed a coarse-grained dynamics of the original process.

In equilibrium MSM-building, a MSM corresponding to a full partition (what we assume to have here) allows to identify the MSM propagator $\hat{P}$ as a (Galerkin) projection of P onto the space spanned by the indicator functions $\mathbb{1}_{A_k}$; for a comprehensive treatment see [SS13, Theorem 5.5].

In our non-stationary case, we have

$$\hat{P}_{ij} \overset{(26)}{=} \frac{\langle P\mathbb{1}_{B_j}, \mathbb{1}_{A_i}\rangle_{\mu_0}}{\langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_{\mu_0}} \overset{(13)}{=} \frac{\langle \mathbb{1}_{B_j}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_{A_i}\rangle_{\mu_1}}{\langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_{\mu_0}}\,. \tag{28}$$

This shows, that a non-stationary MSM is in fact a projection of the transition matrix onto the *basis functions* $\mathbb{1}_{A_k}$, $k = 1, \dots, n$, with respect to the *test functions* $\mathbb{1}_{B_k}$, $k = 1, \dots, n$. Thus, the set of test and basis functions do not coincide, and (26) describes a generalized Galerkin projection, a so-called *Petrov–Galerkin* projection.

Remark 3 (Coarse-grained master equation)**:** Since the original process is given by the time-continuous master equation (3), it would be desirable to have a coarse-grained, MSM master equation

$$\dot{\hat{\mu}}(t) = \hat{L}(t)^T \hat{\mu}(t)\,, \quad t \in [t_0, t_1]\,. \tag{29}$$

Naively, $\hat{L}(t)$ could be found by setting up $\hat{P} = \hat{P}(t,t')$ for all times $t,t' \in [t_0,t_1]$, and differentiating $\hat{P}(t,t')$ with respect to t' at $t' = t$ to obtain $\hat{L}(t)$. By (26), we see that $\hat{P}(t,t')$ depends on t' through P, $\mathbb{1}_A$, and $\mathbb{1}_B$. Now, $\mathbb{1}_A = \mathbb{1}_A(t,t')$, $\mathbb{1}_B = \mathbb{1}_B(t,t')$ are indicator vectors for every t', thus they cannot be changing continuously in t', unless they are constant in time. Thus, their derivative with respect to t' does not exist in general (note that this problem is just as severe in a continuous state space as in a discrete one). To go around this problem one can relax the requirement for *crisp* coherent sets, and crisp indicator vectors, and approximate the coherent sets by so-called fuzzy *membership* or *affiliation* vectors, which take values between 0 and 1, instead of only 0 or 1. A theoretical framework for coherence of fuzzy macrostates is given in Appendix C. We leave to elaborate on the derivation of the master equation (29) to future work.

6 Examples

6.1 Shifting potential

As our first example, we consider a one-dimensional double-well potential shifting to the right as time passes,

$$V_1(t,x) = ((x-t)^2 - 1)^2\,. \tag{30}$$

We choose the time window $[t_0,t_1] = [0,3]$, and restrict ourselves to the state space $X = [-2,5]$. Figure 1 depicts the potential V_1 at initial and final times.

We consider the diffusion process (1) with $\varepsilon = 0.3$. Note that the temporal change rate of the potential is chosen to approximately match the time scale of the (slow) dynamics, such that we can encounter coherent sets, and that we are truly situated between the two asymptotic regimes described in section 3.

As a rule of thumb, this will be the case whenever the metastable time scale τ of any of the snapshot systems (if we would freeze the movement of the potential, and run the process in this stationary situation) match with the time scale of the (relative) change of V in time; say, $\|\frac{1}{V}\frac{d}{dt}V\| = \mathcal{O}(\tau^{-1})$ for every $t \in [t_0,t_1]$, where $\|\cdot\|$ is some norm for functions on X.

To proceed, we discretize the continuous state space $[-2,5]$ into $N = 256$ uniform subintervals, i.e. $S = \{1,\dots,256\}$. The fineness of the discretization is chosen such that further refinement essentially does not alter the results on the level of accuracy we consider. The discrete generators $L(t) \in \mathbb{R}^{N\times N}$ are obtained by discretizating [Kol10, LMHS11, FJK13] the continuous Fokker–Planck equation (5). This yields the propagator $P = P(0,3)$ by solving

$$\frac{d}{dt}P(s,t) = L(t)P(s,t), \quad P(s,s) = I\,, \tag{31}$$

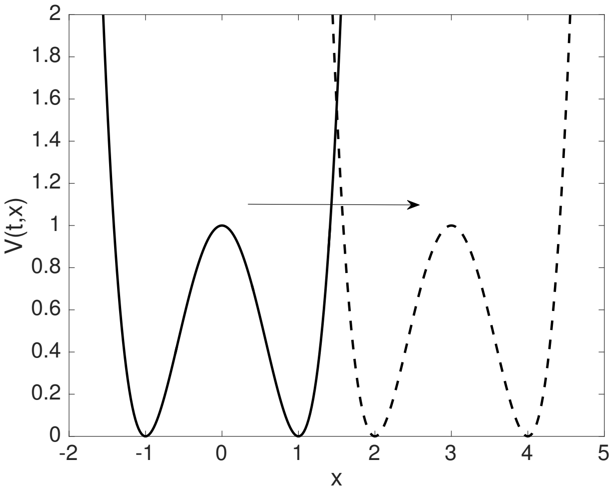

Figure 1: Potential V_1 at initial time $t_0 = 0$ (solid line), and at final time $t_1 = 3$ (dashed line). The potential shifts with constant speed to the right.

on the chosen time interval, hence obtaining the evolution operator which solves (4). The computation has been done in Matlab with a standard solver. We take the uniform distribution $\mu_0 = N^{-1}\mathbb{1}$ as the initial distribution, and compute the final distribution $\mu_1 = P^T\mu_0$. Both are shown in Figure 2.

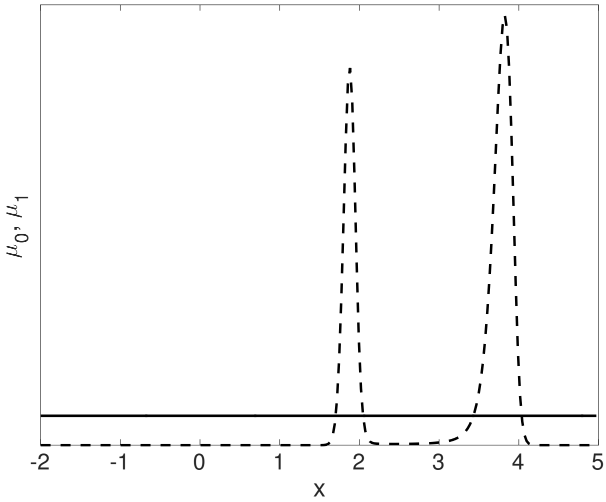

Figure 2: Distributions μ_0 at initial time $t_0 = 0$ (solid line), and μ_1 at final time $t_1 = 3$ (dashed line).

Note that the distribution at final time is concentrated in the wells at that time, but their peaks' positions do not coincide with the well minima, they are a bit more to the left. This is explained by the movement of the potential wells: they move with comparable speed to the dynamical motion of trajectories, hence the distribution has to "catch up" with their motion.

Also, although the wells are equally deep, μ_1 gives more weight to the right well. This is due to the initial configuration: the potential barrier at initial time was at position $x = 0$, hence there was more "mass" right to the barrier (recall the non-symmetry of the state space $[-2, 5]$ with respect to 0, and, especially, that μ_0 is chosen as a uniform distribution, and not as a Boltzmann–Gibbs measure). After starting the process each potential well "collects" the part of the distribution which started on its side of the barrier, and this is then slowly equilibrated by diffusion. This is reflected by μ_1.

Next, we compute a non-stationary MSM based on coherent sets, as described in the theoretical part of this paper. Recall, that in section 4.4 we identify coherent sets that are slightly perturbed versions of perfectly coherent sets. We use (Lipschitz) continuity of the eigenvalues and corresponding eigenspaces of the forward-backward transition matrix C in the perturbation parameter ϵ. The noise level ε from the examples we show here can be identified with the perturbation parameter ϵ, however note, that the concept is not restricted to cases where ϵ has a direct relation to a process parameter. All we assume is that the transition matrix P of the perturbed process is such that there is some transition matrix P_0, in some sense close to P, which admits perfectly coherent sets. The P_0 does not have to be unique, or let alone have *any* physical meaning or connection to our system. Bearing this in mind, it can nevertheless support the understanding to consider a case where a physical connection is present. Thus, we discuss this briefly.

If $\varepsilon = 0$, the stochastic process (1) reduces to a deterministic gradient flow, and the saddle of the potential at $x = 0$ divides[5] the state space into two perfectly coherent sets $A_1 = [-2, 0]$ and $A_2 = S \setminus A_1 = (0, 5]$ (at time $t = 0$). They can be identified from the two-dimensional eigenspace of C at the eigenvalue $\lambda = 1$ (which has multiplicity 2). Recall, that then the corresponding eigenvectors u_1, u_2 (we take an arbitrary basis of the two-dimensional eigenspace) are exact linear combinations of the indicator vectors $\mathbb{1}_{A_1}, \mathbb{1}_{A_2}$, i.e. there are constant scalars $c_{1,1}, c_{2,1}, c_{1,2}, c_{2,2}$ such that $u_1 = c_{1,1}\mathbb{1}_{A_1} + c_{2,1}\mathbb{1}_{A_2}$ and $u_2 = c_{1,2}\mathbb{1}_{A_1} + c_{2,2}\mathbb{1}_{A_2}$ (due to discretization of the continuous process, these crisp identities may only hold up to a negligible error). Hence, the components of both u_1 and u_2 take only two values, respectively, which can easily be determined and used to identify $\mathbb{1}_{A_1}$ and $\mathbb{1}_{A_2}$.

Now, the solution of the problem relies on the assumption that the perturbation is sufficiently small, such that we can carry out the identification of the coherent sets as just described, now *approximately*. To this end we consider the eigenspectrum (computed by the Matlab routine `eigs`, which uses a Lánczos-type algorithm) of the forward-backward transition matrix $C = PD_{\mu_1}^{-1}P^TD_{\mu_0}$; the first few eigenvalues are given by

$$\lambda_1 = 1.000, \quad \lambda_2 = 0.871, \quad \lambda_3 = 3.24 \cdot 10^{-6}\,.$$

Viewing this as perturbation of the perfectly invariant case, the theory in section 4.4 tells us to expect as many eigenvalues close or equal to one, as coherent sets we have; in this case two.

[5] Actually, the saddle of the potential would be the boundary between the coherent sets only if the potential is changing infinitely slowly, and hence we would be dealing with a quasi-stationary case. In the finite-speed case, the process starting just right of the saddle will still converge to the bottom of the left well, because initially it moves so slowly that the saddle "overtakes" it. Thus, the boundary of the coherent sets is slightly to the right from the saddle, at the point where $-\frac{d}{dx}V_1(0,x) = 1$.

We note the large spectral gap after the second dominant eigenvalue, thus we expect to find two coherent pairs A_1, B_1 and A_2, B_2 (which are "close" to the pairs from the unperturbed case), where $A_2 = S \setminus A_1$ and $B_2 = S \setminus B_1$. Recalling the perturbation argument (24), we expect the two dominant eigenvectors of C, namely u_1 and u_2, to be approximate linear combinations of $\mathbb{1}_{A_1}$ and $\mathbb{1}_{A_2}$. That is, there are scalars $c_{1,1}, c_{1,2}, c_{2,1}, c_{2,2}$ such that

$$u_1 \approx c_{1,1}\mathbb{1}_{A_1} + c_{2,1}\mathbb{1}_{A_2}, \quad u_2 \approx c_{1,2}\mathbb{1}_{A_1} + c_{2,2}\mathbb{1}_{A_2} \,. \tag{32}$$

This should provide $\mathbb{1}_{A_1}, \mathbb{1}_{A_2}$. However, (32) is merely an approximate identity, hence we can identify $\mathbb{1}_{A_1}, \mathbb{1}_{A_2}$ also only approximately. Since $\lambda_1 = 1$, we have $u_1 = \mathbb{1}$, hence $c_{1,1} = c_{2,1} = 1$. Thus, we get by inverting (32) that

$$I_{A_1} := -\frac{c_{2,2}}{c_{1,2} - c_{2,2}} u_1 + \frac{1}{c_{1,2} - c_{2,2}} u_2 \,, \quad I_{A_2} := -\frac{c_{1,2}}{c_{2,2} - c_{1,2}} u_1 + \frac{1}{c_{2,2} - c_{1,2}} u_2 \,, \tag{33}$$

so that

$$I_{A_1} \approx \mathbb{1}_{A_1}, \quad I_{A_2} \approx \mathbb{1}_{A_2} \,. \tag{34}$$

Given we know the form of u_2 from (32) we can infer that its components form two groups of values, one concentrated around $c_{1,2}$ and the other concentrated around $c_{2,2}$. Hence, any element of the particular group should be a good approximation to the corresponding $c_{i,j}$, and for simplicity we set $c_{1,2} = \max_{i \in S} u_{2,i}$ and $c_{2,2} = \min_{i \in S} u_{2,i}$. Now that we have $c_{1,2}, c_{2,2}$, Equation (33) gives us I_{A_1}, I_{A_2}, approximations to $\mathbb{1}_{A_1}$ and $\mathbb{1}_{A_2}$, respectively.

Further, $D_{\mu_1}^{-1} P^T D_{\mu_0} u_k$, $k = 1, 2$, should be approximate linear combinations of $\mathbb{1}_{B_1}$ and $\mathbb{1}_{B_2}$. Now, since $u_1 = \mathbb{1}$, it is sufficient to consider u_2. Figure 3 (left) shows u_2 and its push-forward, $D_{\mu_1}^{-1} P^T D_{\mu_0} u_2$.

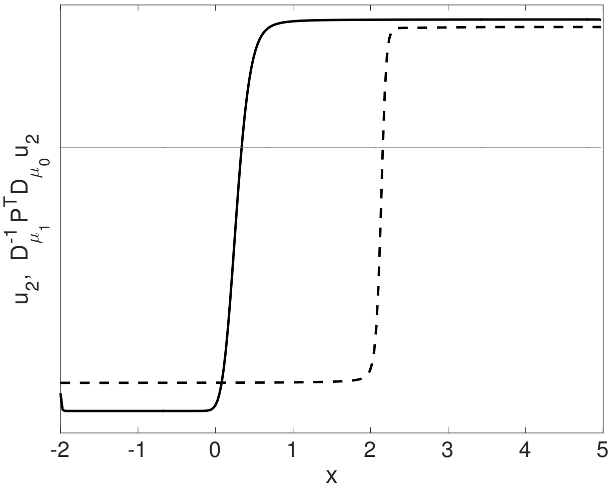

Figure 3: Second eigenvector u_2 of C, showing coherent sets at initial time (solid line), and $D_{\mu_1}^{-1} P^T D_{\mu_0} u_2$, showing coherent sets at final time (dashed line). The thin horizontal line indicates the value zero.

Both the eigenvector u_2 and its push-forward are approximate step functions. In fact, the smaller the diffusivity constant ε is, the closer these objects get to being step functions. This suggests to take $A_1 \approx [-2,0]$, $A_2 \approx (0,5]$, and $B_1 \approx [-2,2]$, $B_2 \approx (2,5]$. Comparing these sets with the position of the potential barrier at initial and final times, the initial separation is aligned with the barrier at $x = 0$ for $t_0 = 0$, but one might have expected B_1 and B_2 to be separated by $x = 3$. The reason, just as in Figure 2, lies in the non-negligible shift speed of the potential: due to this, it is more likely to find the system in the left half of the respective potential well, than in the right half. If we would slow down the potential shift, say, it would shift the same amount, but uniformly on the time interval, e.g., $[t_0,t_1] = [0,30]$, the boundary of the coherent sets would align with the positions of the potential barrier at the respective times.

Instead of relying on our visual identification of coherent sets, we approximate the $\mathbb{1}_{A_i}$ and $\mathbb{1}_{B_i}$, $i = 1,2$, by I_{A_i} from (33) and their push-forwards $I_{B_i} := D_{\mu_1}^{-1} P^T D_{\mu_0} I_{A_i}$. A more sophisticated way of extracting the sets of interest from spectral analysis, especially for more than two sets, can be found e.g. in [DW04]. The focus of this work is however conceptual, thus we will use this simplified construction of the indicators.

Finally, the MSM transition matrix $\hat{P}$ is computed by (26), by substituting I_{A_i}, I_{B_i} from (33) instead of $\mathbb{1}_{A_i}, \mathbb{1}_{B_i}$, which yields

$$\hat{P} = \begin{pmatrix} 0.9683 & 0.0295 \\ 0.0150 & 0.9599 \end{pmatrix}.$$

Note that $\hat{P}$ is only approximately stochastic. This is due to the fact that (i) our approximate indicators I_{A_1}, I_{A_2}, I_{B_1}, and I_{B_2} as obtained from (33) both at initial and final time do not form a partition of unity, and (ii) the system is not closed, and trajectories are lost at the state space boundaries. Still, $\hat{P}_{ij}$ approximates the probabilities from moving from A_i to B_j very well.

6.2 Merging potential wells

In the equilibrium case one can (vaguely) identify *one* potential well with *one* metastable set. Our next example shows that in the time-variant case this intuition does not have to hold even in the simplest examples.

Let us consider the time-dependent potential

$$V_2(t,x) = 5\left((x+1)\left(x - \frac{t}{2\tau}\right)\left(x - 1 + \frac{t}{2\tau}\right)\right)^2 \tag{35}$$

on the time interval $[t_0,t_1] = [0,\tau]$ for $\tau = 10$, and state space $[-2,2]$. It is depicted in Figure 4 (left) at different times.

To repeat the analysis from our previous example, we proceed identically, also by choosing the same parameters $\varepsilon = 0.3$, $N = 256$. Again, the initial distribution is chosen to be uniform. Figure 4 (right) shows the initial and final distributions.

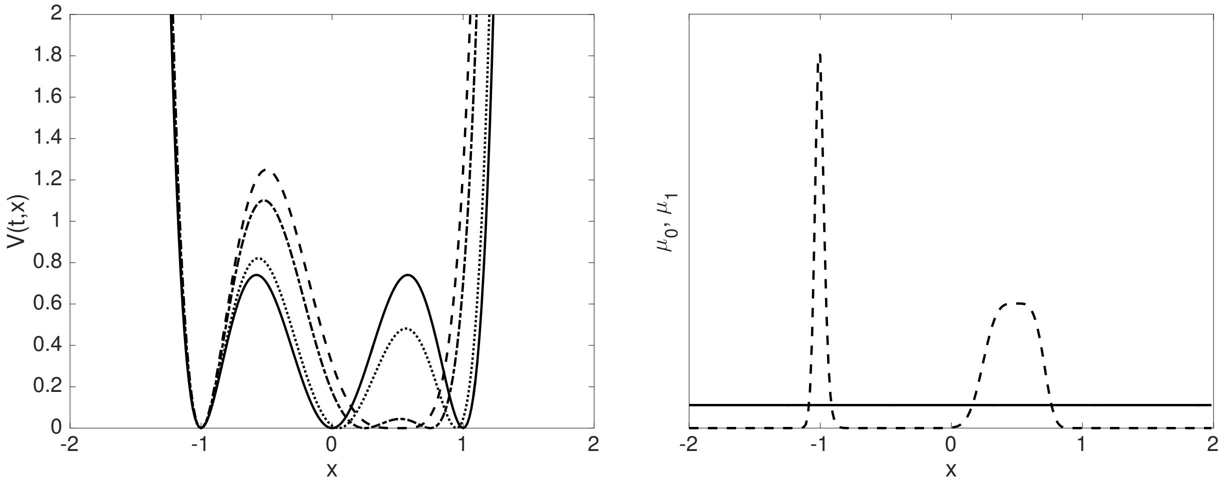

Figure 4: Left: Potential V_2 at times $t = 0, 1, 5, 10$, shown by solid, dotted, dash-dotted, and dashed lines, respectively. The middle and the right wells merge (the barrier between them vanishes), while the barrier separating the left well from the others rises. Right: initial (solid) and final (dashed) distributions, μ_0 and μ_1, respectively.

What is going to be a coherent pair now? Clearly, the left well stays coherent. However, is the middle well at initial time with the right well at final time coherent? Or is it the right well at initial time and the right one at final time?

Since our analysis is analogous to that in the previous example, we will not repeat it step-by-step, but show directly the second eigenvector of the forward-backward transition matrix C in Figure 5, which has dominant spectrum

$$\lambda_1 = 1.0000, \quad \lambda_2 = 0.9614, \quad \lambda_3 = 1.3138 \cdot 10^{-6}\,.$$

The eigenvalues tell us that there are only two coherent pairs. The second eigenvector shows which sets form these coherent pairs. The first coherent pair is not a surprise: $A_1 \approx B_1 \approx [-2, -0.5]$. Regarding the other pair(s), neither of our guesses from above were right. The other coherent pair has to involve *both* the middle and right wells at initial time, i.e. $A_2 \approx B_2 \approx (-0.5, 2]$. Neither the middle, nor the right well at initial time can *alone* be coherent. This is in line with the forward-backward characterization: the initial set of a coherent pair has to be metastable under the forward-backward process. Now, the forward process starting in the two rightmost wells at initial time ends up in the large right well at final time, then starting a backward process from there will lead with essentially equal probabilities to one of the two rightmost wells at initial time. Hence, there *union* has to form the coherent set A_2.

7 Conclusion

We have shown that the theory of coherent sets allows us to extend the concept of metastability and that of Markov state models (MSMs) from stationary reversible processes to those

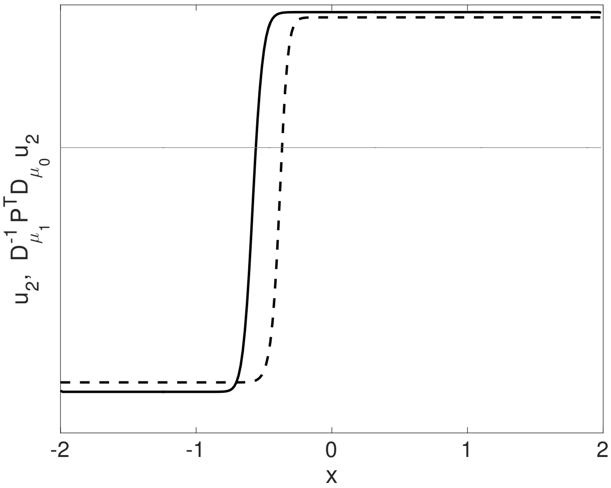

Figure 5: Second eigenvector u_2 of C, showing coherent sets at initial time (solid line), and $D_{\mu_1}^{-1}P^T D_{\mu_0}u_2$, showing coherent sets at final time (dashed line). The thin horizontal line indicates the value zero.

which are non-stationary (and, hence, in general also non-reversible). If the process under consideration is stationary and reversible, our construction reduces to the standard identification of metastability and spectral theory based MSM analysis (Remark 1). The presented examples show that the intuition and tools taken from the stationary case can fail to understand and describe the essential dynamical properties for non-stationary processes, whereas our approach is capable of doing so.

We consider this work as a possible conceptual foundation for building MSMs of non-stationary processes. In order to achieve this goal, there are more features that have to be incorporated, which make MSMs a practicable tool. For instance, is there an "optimal" choice for the time scale τ, which yields the "best" MSMs? Are there conditions on the non-stationary forcing (e.g. periodicity in time), which allow for inferring the desired information about the system for arbitrary times, just by setting up *one* MSM for *one* single time τ? Also, equilibrium MSMs are often built with respect to metastable sets which do not build a full partition of state space (so-called "core set MSMs"), and this shall be carried over to the non-stationary case as well. There are also behaviors not present in stationary reversible systems, which one would possibly like to capture in MSMs for non-stationary systems, such as cyclicity. Considering the minimal (discrete time) example

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

every state turns out to be perfectly coherent for every time interval, but an analysis based on finite time coherent sets does not explicitly indicate that a cycle of period 3 is present (unless we are lucky enough to consider the system on a time interval of length 3). Last, it has to be investigated whether for a forced system of physical interest the concept of coherent sets can still remain useful to define metastability and build MSMs. These issues will be addressed elsewhere.

A Functional characterization of finite-time coherent pairs

In this appendix we derive Equation (15).

Note that both $D_{\mu_1}^{-1}P^TD_{\mu_0}$ and P

- are positive operators (An operator Q is positive, if $Qv \geq 0$ for every $v \geq 0$, here and in the following the inequality is meant componentwise. Positivity directly implies monotony: if $v \leq w$, then $Qv \leq Qw$ follows.); and
- leave $\mathbb{1}$ invariant, i.e.,
$$D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1} = \mathbb{1} \quad \text{and} \quad P\mathbb{1} = \mathbb{1}\,, \tag{36}$$
where the former equation follows from $\mu_1 = P^T\mu_0$, which can be rewritten as $(D_{\mu_1}\mathbb{1}) = P^T(D_{\mu_0}\mathbb{1})$, and the latter one follows from P being a stochastic transition matrix.

Additionally, we have $\langle \mathbb{1}_A, \mathbb{1}_A\rangle_{\mu_0} = \mu_0(A)$, and $\langle \mathbb{1}_B, \mathbb{1}_B\rangle_{\mu_1} = \mu_1(B)$. Recall (7*), (8*), and (11), yielding the *forward-backward condition*

$$\langle \mathbb{1}_A, \mathbb{1}_A\rangle_{\mu_0} \approx \langle \mathbb{1}_B, P^TD_{\mu_0}\mathbb{1}_A\rangle \approx \langle \mathbb{1}_B, \mathbb{1}_B\rangle_{\mu_1}\,. \tag{37}$$

With A, B a coherent pair, the forward-backward condition (37) gives with (13) that

$$\langle \mathbb{1}_B, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} \approx \langle \mathbb{1}_B, \mathbb{1}_B\rangle_{\mu_1}\,. \tag{38}$$

Monotony of $D_{\mu_1}^{-1}P^TD_{\mu_0}$ with $\mathbb{1}_A \leq \mathbb{1}$, and the invariance of $\mathbb{1}$ imply that $D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A \leq D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1} = \mathbb{1}$, thus, together with (38) we get that $(D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A)_i \approx 1$ for $i \in B$. This is merely a condition for the subset of components $i \in B$. It remains to show that $(D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A)_i \approx 0$ for $i \in S \smallsetminus B$. This we can achieve by estimating the "total probability" that ends up in $S \smallsetminus B$. We have that

$$\langle \mathbb{1}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} \approx \langle \mathbb{1}_B, \mathbb{1}_B\rangle_{\mu_1}\,. \tag{39}$$

In fact,

$$\langle \mathbb{1}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} = \langle P\mathbb{1}, D_{\mu_0}\mathbb{1}_A\rangle = \langle \mathbb{1}, D_{\mu_0}\mathbb{1}_A\rangle = \langle \mathbb{1}_A, \mathbb{1}_A\rangle_{\mu_0} \approx \langle \mathbb{1}_B, \mathbb{1}_B\rangle_{\mu_1}\,, \tag{40}$$

where the first equality follows by changing from a weighted scalar product to a nonweighted one, and the second from the invariance of $\mathbb{1}$ under P. Now, with $(D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A)_i \approx 1$ for $i \in B$, and with $\mathbb{1} = \mathbb{1}_B + \mathbb{1}_{S\smallsetminus B}$, we have for the left-hand side of (39) that

$$\begin{aligned}
\langle \mathbb{1}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} &= \langle \mathbb{1}_B, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} + \langle \mathbb{1}_{S\smallsetminus B}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} & (41)\\
&\approx \langle \mathbb{1}_B, \mathbb{1}_B\rangle_{\mu_1} + \langle \mathbb{1}_{S\smallsetminus B}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} & (42)
\end{aligned}$$

Together with (39) we obtain $\langle \mathbb{1}_{S\smallsetminus B}, D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A\rangle_{\mu_1} \approx 0$, and with $D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A \geq 0$, that $(D_{\mu_1}^{-1}P^TD_{\mu_0}\mathbb{1}_A)_i \approx 0$ for $i \in S \smallsetminus B$. In summary, we have shown (15).

B Temporal monotony of coherence

We would like to show that if A, B constitute a coherent pair on $[t_0, t_1]$, then for all $t \in (t_0, t_1)$ there is an $E \subset S$, such that A, E are a coherent pair on $[t_0, t]$, with coherence not smaller than that of A, B. To get an idea for the problem, we start with a specific case.

Lemma 1: If A, B constitute a *perfectly* coherent pair on $[t_0, t_1]$, i.e. $\mu_0(A) = \mu_1(B)$, and

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \in A] = 1 = \mathbb{P}[\boldsymbol{y}_{t_0} \in A \,|\, \boldsymbol{y}_{t_1} \in B], \tag{43}$$

then for all $t \in (t_0, t_1)$ there is an $E \subset S$, such that A, E are a perfectly coherent pair on $[t_0, t]$.

Proof. Fix $t \in (t_0, t_1)$, and let μ_t denote the distribution of the process at time t. To start, note that the following two statements are equivalent:

(a) There is no set $E \subset S$ satisfying $\mu_t(E) = \mu_0(A)$ with

$$\mathbb{P}[\boldsymbol{y}_t \in E \,|\, \boldsymbol{y}_{t_0} \in A] = 1 = \mathbb{P}[\boldsymbol{y}_{t_0} \in A \,|\, \boldsymbol{y}_t \in E]\,. \tag{44}$$

(b) There is an $i \in S$ such that

$$\mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \in A] > 0 \quad \text{and} \quad \mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \notin A] > 0\,. \tag{45}$$

To see this, note that if (b) is true, then $i \notin E$ has to hold, because $\mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \notin A] > 0$. However, also $i \notin S \setminus E$ has to hold, because $\mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \in A] > 0$, hence (a) follows. If (b) is false, then for every $i \in S$ either $\mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \in A] = 0$ or $\mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \notin A] = 0$. Thus,

$$E = \{i \in S \,|\, \mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \notin A] = 0\} \tag{46}$$

would be a set perfectly coherent with A, because $\mathbb{P}[\boldsymbol{y}_t \in S \setminus E \,|\, \boldsymbol{y}_0 \in A] = 0$ follows from $\mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \in A] = 0$ for every $i \notin E$. We have by perfect coherence also $\mu_t(E) = \mu_0(A)$, falsifying (a).

To show the claim of the lemma, we assume that (a) holds, and show that this leads to a contradiction. Let $i \in S$ be the state from (b). Then we have two cases to consider for the conditional probability $\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = i]$:

1) If $\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = i] > 0$, then

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \notin A] \geq \mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = i] \cdot \mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \notin A] > 0\,, \tag{47}$$

contradicting the perfect coherence of the pair A, B.

2) If $\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = i] = 0$, then

$$\mathbb{P}[\boldsymbol{y}_{t_1} \notin B \,|\, \boldsymbol{y}_{t_0} \in A] \geq \mathbb{P}[\boldsymbol{y}_{t_1} \notin B \,|\, \boldsymbol{y}_t = i] \cdot \mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \in A] > 0\,, \tag{48}$$

contradicting the perfect coherence of the pair A, B.

□

For the general claim, where the coherent pair is not perfectly coherent, we need a quantifier for coherence. Let us assume from now on that $\mu_0(A) = \mu_1(B)$, then we have

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \in A] = \frac{\mathbb{P}[\boldsymbol{y}_{t_1} \in B,\, \boldsymbol{y}_{t_0} \in A]}{\mu_0(A)} = \frac{\mathbb{P}[\boldsymbol{y}_{t_1} \in B,\, \boldsymbol{y}_{t_0} \in A]}{\mu_1(B)} = \mathbb{P}[\boldsymbol{y}_{t_0} \in A \,|\, \boldsymbol{y}_{t_1} \in B]\,. \tag{49}$$

Thus, both the "forward" probability (we will also call it the forward rate) from (7), and the "backward" probability from (8) identically quantify coherence: this quantity is between 0 and 1, and it is equal to 1 if and only if we have perfect coherence.

We are going to show that the forward rate is monotonically decreasing in time, i.e. that for every time $t \in (t_0, t_1)$ there is a set E, with $\mu_t(E) = \mu_0(A)$, such that the forward rate from A to E is at least the forward rate from A to B. We shall do this in a constructive manner: if we want to find the set E at time t which is most coherent with A, we have to collect those states $i \in S$ for which it is the most likely that the process being in i at time t came from A. To this end we will look at the ratios $\frac{a_i}{\eta_i}$, where a_i denotes the probability that the process is in i at time t, given it started in A at time t_0, i.e. $a_i = \mathbb{P}[\boldsymbol{y}_t = i \,|\, \boldsymbol{y}_{t_0} \in A]$, and η_i is the probability that the process is in i at time t (i.e. η is the distribution of $\boldsymbol{y}_t$, i.e. $\eta_i = \mathbb{P}[\boldsymbol{y}_t = i]$). We will construct E by adding those states $i \in S$ to it for which $\frac{a_i}{\eta_i}$ is largest, until $\mu_t(E) = \mu_0(A)$. More precisely, let $i_1, i_2, \ldots, i_N \in S$ be an ordering of the states (i.e. $i_\alpha \neq i_\beta$ if $\alpha \neq \beta$) such that

$$\frac{a_{i_1}}{\eta_{i_1}} \geq \frac{a_{i_2}}{\eta_{i_2}} \geq \ldots \geq \frac{a_{i_N}}{\eta_{i_N}}\,, \tag{50}$$

and let $E = \{i_1, i_2, \ldots, i_{k^*}\}$ be a set, where k^* is chosen such that[6]

$$\sum_{k=1}^{k^*} \eta_{i_k} = \mu_0(A)\,, \tag{51}$$

then we can show the following.

Theorem 1: Let $\mu_0(A) = \mu_1(B)$, and $t \in (t_0, t_1)$. Then, assuming the existence of a k^* such that (51) holds, the set $E \subset S$ defined above is such that $\mu_t(E) = \mu_0(A)$, and

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \in A] \leq \mathbb{P}[\boldsymbol{y}_t \in E \,|\, \boldsymbol{y}_{t_0} \in A]\,. \tag{52}$$

Proof. Note that by conservation of total probability we have

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \in A] = \sum_{j \in S} \underbrace{\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = j]}_{=b_j} \cdot \underbrace{\mathbb{P}[\boldsymbol{y}_t = j \,|\, \boldsymbol{y}_{t_0} \in A]}_{=a_j}\,. \tag{53}$$

[6] For a general process on the discrete state space S there is no guarantee that an $E \subset S$ exists with $\mu_t(E) = \mu_0(A)$. However, if we think of $(\boldsymbol{y}_t)$ as a discretization of a continuous-space process $(\boldsymbol{x}_t)$ with a probability density function, then we can find a sufficiently fine discretization, such that (51) and (52) hold with an arbitrarily small error. If $S \subset \mathbb{R}^d$ is a continuous state space, Theorem 1 even holds without the additional assumption on the existence of such a k^*. The set E is then chosen as a suitable superlevel set of the function $\frac{a(x)}{\eta(x)}$. We refrain from spelling out the technical details here.

Additionally to $a_j = \mathbb{P}[\boldsymbol{y}_t = j \,|\, \boldsymbol{y}_{t_0} \in A], \eta_j = \mathbb{P}[\boldsymbol{y}_t = j] = \mu_{t,j}$, already defined, we define $b_j = \mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = j]$ as in (53). We see that

$$\sum_{j\in S} \eta_j b_j = \sum_{j\in S} \mathbb{P}[\boldsymbol{y}_t = j] \cdot \mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_t = j] = \mathbb{P}[\boldsymbol{y}_{t_1} \in B] = \mu_1(B) = \mu_0(A)\,. \tag{54}$$

Our objective is to prove that $E \subset S$, defined above (51), with

$$\sum_{j\in E} \eta_j = \mu_0(A) = \sum_{j\in S} \eta_j b_j\,, \tag{55}$$

satisfies $\sum_{j\in S} a_j b_j \leq \sum_{j\in E} a_j$, which is equivalent to (52).

To this end, note first that if E satisfies (55), then

$$\sum_{j\in E} \eta_j (1 - b_j) = \sum_{j\in E} \eta_j - \sum_{j\in E} \eta_j b_j \overset{(55)}{=} \sum_{j\in S} \eta_j b_j - \sum_{j\in E} \eta_j b_j = \sum_{j\notin E} \eta_j b_j\,. \tag{56}$$

This equation can be read as the probability that the process starts in E at time t, and goes to $S \setminus B$ (left-hand side) is being equal to the probability that the process starts in $S \setminus E$ at time t, and goes to B (right-hand side). The idea of the proof is now to use E, and show that the probability that comes from A, through E, going to $S \setminus B$ is greater than the probability coming from A, through $S \setminus E$, going to B. This will imply that A is more coherent with E than with B.

Elementary manipulation yields for an arbitrary $E \subset S$:

$$\sum_{j\in S} a_j b_j = \sum_{j\in E} a_j b_j + \sum_{j\notin E} a_j b_j \tag{57}$$

$$= \sum_{j\in E} a_j - \underbrace{\sum_{j\in E} a_j (1 - b_j)}_{=p_{out}} + \underbrace{\sum_{j\notin E} a_j b_j}_{=p_{in}}\,, \tag{58}$$

where we can see that $p_{out} = \mathbb{P}[\boldsymbol{y}_{t_1} \notin B,\, \boldsymbol{y}_t \in E \,|\, \boldsymbol{y}_{t_0} \in A]$, and $p_{in} = \mathbb{P}[\boldsymbol{y}_{t_1} \in B,\, \boldsymbol{y}_t \notin E \,|\, \boldsymbol{y}_{t_0} \in A]$. Now, invoking (50) and (51), and the set $E = \{i_1, i_2, \ldots, i_{k^*}\}$, we can guarantee that there is some $c > 0$ such that

$$\frac{a_j}{\eta_j} \geq c \text{ for } j \in E, \quad \text{and} \quad \frac{a_j}{\eta_j} \leq c \text{ for } j \notin E\,, \tag{59}$$

with $\mu_t(E) = \mu_0(A)$. Thus, we have in (58):

$$p_{in} = \sum_{j\notin E} \frac{a_j}{\eta_j} \eta_j b_j \leq c \sum_{j\notin E} \eta_j b_j \overset{(56)}{=} c \sum_{j\in E} \eta_j (1 - b_j) \leq \sum_{j\in E} \frac{a_j}{\eta_j} \eta_j (1 - b_j) = p_{out}\,. \tag{60}$$

This means that $p_{in} \leq p_{out}$, thus, substituting into (58) gives

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \in A] = \sum_{j\in S} a_j b_j \leq \sum_{j\in E} a_j = \mathbb{P}[\boldsymbol{y}_t \in E \,|\, \boldsymbol{y}_{t_0} \in A]\,. \tag{61}$$

This was to show. □

Remark 4 (The knapsack problem)**:** The construction of the set E is strongly related to the solution of the *knapsack problem* from mathematical optimization:

$$\text{maximize} \sum_{j\in E} a_j \text{ w.r.t. } E \subset S, \quad \text{subject to } \sum_{j\in E} \eta_j = \sum_{j\in S} \eta_j b_j = \text{const}\,. \tag{62}$$

Viewing the a_j as rewards and the η_j as costs, the optimization problem is to maximize the reward by choosing items j from S, given the allowed total cost of these items is given by a constant $\sum_{j\in S} \eta_j b_j$. It can be shown that assuming the existence of k^* in (51) basically guarantees that the *greedy strategy* (i.e., choose the items with the highest reward-to-cost ratio a_j/η_j first) for solving (62) is optimal. Then, Theorem 1 states that the optimal value of this problem is greater or equal $\sum_{j\in S} a_j b_j$, cf. (61).

C Formulation in terms of fuzzy macrostates

So far, the affiliation of the respective states to a set A has been indicated by the vector $\mathbb{1}_A$, such that $\mathbb{1}_{A,j} \in \{0,1\}$. Now, let $f_A \in [0,1]^N$ be an *affiliation vector*, $f_{A,j} \in [0,1]$ indicating to which extent $j \in S$ belongs to a *macrostate* A. Note that thus A in general ceases to have the interpretation of a crisp set, although it can be thought of probabilistically: any state $j \in S$ belongs to the macrostate A with probability $f_{A,j}$, so that

$$\mathbb{P}[\boldsymbol{y}_{t_0} \in A] = \sum_{j\in S} \mu_{0,j} f_{A,j} = \langle \mathbb{1}, f_A \rangle_{\mu_0}\,, \tag{63}$$

where we vaguely abuse the notation $\boldsymbol{y}_{t_0} \in A$, meaning that $\boldsymbol{y}_{t_0}$ contributes to the macrostate A.

Let thus $\mu_{0,A} := \langle \mathbb{1}, f_A \rangle_{\mu_0}$ denote the total probability in macrostate A. Let us analogously define a macrostate B at final time, f_B, such that $\mu_{1,B} = \langle \mathbb{1}, f_B \rangle_{\mu_1} = \mu_{0,A}$. Hence, the two macrostates carry the same probability. In analogy with (13), the joint probability of being in macrostate A at initial time and in macrostate B at final time can be expressed by

$$\mathbb{P}[\boldsymbol{y}_{t_0} \in A,\, \boldsymbol{y}_{t_1} \in B] = \sum_{i,j\in S} \mu_{0,i} f_{A,i} P_{ij} f_{B,j} = \langle P f_B,\, D_{\mu_0} f_A \rangle = \langle f_B,\, D_{\mu_1}^{-1} P^T D_{\mu_0} f_A \rangle_{\mu_1}\,. \tag{64}$$

We are ready to state the monotony result for coherence of fuzzy macrostates.

Theorem 2: Let A, B be fuzzy macrostates defined by the affiliation vectors f_A, f_B, such that $\mu_{0,A} = \mu_{1,B}$. Then, for any fixed $t \in (t_0, t_1)$, the vector $f_E := P_{t,1} f_B$, where $P_{t,1}$ is the propagator from t to t_1, is an affiliation vector, and thus defines a macrostate E. Further, this macrostate satisfies $\mu_{t,E} = \mu_{0,A}$, and

$$\mathbb{P}[\boldsymbol{y}_{t_1} \in B \,|\, \boldsymbol{y}_{t_0} \in A] = \mathbb{P}[\boldsymbol{y}_t \in E \,|\, \boldsymbol{y}_{t_0} \in A]\,. \tag{65}$$

Proof. We start by showing that f_E is an affiliation vector (i.e. $f_{E,i} \in [0,1]$ for all $i \in S$) with $\langle \mathbb{1}, f_E \rangle_{\mu_t} = \mu_{0,A} = \mu_{1,B}$.

To this end, we have, componentwise, $P_{t,1} f_B \geq 0$, because $f_B \geq 0$ and $P_{t,1}$ is a stochastic matrix. Further,

$$(P_{t,1} f_B)_i = \sum_{j \in S} (P_{t,1})_{ij} f_{B,j} \leq \max_{i \in S} |f_{B,i}| \underbrace{\sum_{j \in S} (P_{t,1})_{ij}}_{=1} \leq 1\,. \tag{66}$$

Last, we have

$$\langle \mathbb{1}, f_E \rangle_{\mu_t} = \langle \mathbb{1}, P_{t,1} f_B \rangle_{\mu_t} = \langle \mu_t, P_{t,1} f_B \rangle = \langle \underbrace{P_{t,1}^T \mu_t}_{=\mu_1}, f_B \rangle = \langle \mathbb{1}, f_B \rangle_{\mu_1} = \mu_{1,B} = \mu_{0,A}\,. \tag{67}$$

In summary, f_E defines a macrostate with $\mu_{t,E} = \mu_{0,A}$.

Note that By the Markov property, we have $P = P_{0,t} P_{t,1}$, where $P_{0,t}$ is the propagator from t_0 to t, and we have by elementary manipulations, and re-weighting of the scalar products:

$$\begin{aligned} \mathbb{P}[\boldsymbol{y}_{t_0} \in A,\, \boldsymbol{y}_{t_1} \in B] &= \langle f_B,\, D_{\mu_1}^{-1} P^T D_{\mu_0} f_A \rangle_{\mu_1} \\ &= \langle f_B,\, D_{\mu_1}^{-1} P_{t,1}^T P_{0,t}^T D_{\mu_0} f_A \rangle_{\mu_1} \\ &= \langle f_B,\, P_{t,1}^T P_{0,t}^T D_{\mu_0} f_A \rangle \\ &= \langle P_{t,1} f_B,\, D_{\mu_t}^{-1} P_{0,t}^T D_{\mu_0} f_A \rangle_{\mu_t} \\ &= \langle f_E,\, D_{\mu_t}^{-1} P_{0,t}^T D_{\mu_0} f_A \rangle_{\mu_t} \\ &= \mathbb{P}[\boldsymbol{y}_{t_0} \in A,\, \boldsymbol{y}_t \in E] \end{aligned} \tag{68}$$

Thus, dividing by $\mu_{0,A}$, (65) follows. □

Theorem 2 shows, that, as time progresses, the level of *maximal* coherence can only decrease. In fact, we have demonstrated that it certainly exists a macrostate E at any intermediate time t, which has the same level of coherence as A and B.

References

[BaJLM+11] Kyle A Beauchamp, Gregory R Bowman andThomas J Lane, Lutz Maibaum, Imran S Haque, and Vijay S Pande. MSMBuilder2: Modeling conformational dynamics at the picosecond to millisecond scale. *J Chem Theor Comput*, 2011.

[BPN14] G. R. Bowman, V. S. Pande, and F. Noé, editors. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, volume 797 of *Advances in Experimental Medicine and Biology*. Springer, 2014.

[DSS12] N. Djurdjevac, M. Sarich, and Ch Schütte. Estimating the eigenvalue error of Markov state models. *Multiscale Modeling & Simulation*, 10(1):61–81, 2012.

[DW04] Peter Deuflhard and Marcus Weber. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.*, 398:161–184, 2004. Special Issue on Matrices and Mathematical Biology.

[FJK13] Gary Froyland, Oliver Junge, and Péter Koltai. Estimating long term behavior of flows without trajectory integration: the infinitesimal generator approach. *SIAM J. Numer. Anal.*, 51(1):223–247, 2013.

[FPG14] Gary Froyland and Kathrin Padberg-Gehle. Almost-invariant and finite-time coherent sets: directionality, duration, and diffusion. In *Ergodic Theory, Open Dynamics, and Coherent Structures*, pages 171–216. Springer, 2014.

[Fro13] Gary Froyland. An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. *Physica D: Nonlinear Phenomena*, 250:1–19, 2013.

[FSM10] Gary Froyland, Naratip Santitissadeekorn, and Adam Monahan. Transport in time-dependent dynamical systems: Finite-time coherent sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):043116, 2010.

[Kat84] Tosio Kato. *Perturbation Theory for Linear Operators.* Springer-Verl., 2. edition, 1984.

[Kol10] Péter Koltai. *Efficient approximation methods for the global long-term behavior of dynamical systems – Theory, algorithms and examples.* PhD thesis, Technische Universität München, 2010.

[LMHS11] Juan C Latorre, Philipp Metzner, Carsten Hartmann, and Christof Schütte. A structure-preserving numerical discretization of reversible diffusions. *Commun. Math. Sci*, 9(4):1051–1072, 2011.

[PWS+11] J.H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schütte, and F. Noé. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.*, 134:174105, 2011.

[SNL+11] C. Schütte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden. Markov State Models based on Milestoning. *J. Chem. Phys.*, 134:204105, 2011.

[SNS10] Marco Sarich, Frank Noé, and Christof Schütte. On the approximation quality of Markov state models. *Multiscale Modeling & Simulation*, 8(4):1154–1177, 2010.

[SS13] Ch. Schütte and M. Sarich. *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, volume 24 of *Courant Lecture Notes*. American Mathematical Society, December 2013.

[STSM+12] M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, Ch. Schütte, and F. Noé. Emma - a software package for Markov model building and analysis. *J. Chem. Theory Comput.*, 8:2223–2238, 2012.

[SW15] Christof Schütte and Han Wang. Building Markov state models for periodically driven non-equilibrium systems. *Journal of Chemical Theory and Computation*, 11(4):18191831, 2015.

[WSCDS14] Han Wang, Christof Schütte, Giovanni Ciccotti, and Luigi Delle Site. Exploring the conformational dynamics of alanine dipeptide in solution subjected to an external electric field: A nonequilibrium molecular dynamics simulation. *Journal of Chemical Theory and Computation*, 10(4):1376–1386, 2014.