Non-stationary Transition Path Theory
with applications to tipping and agent-based models
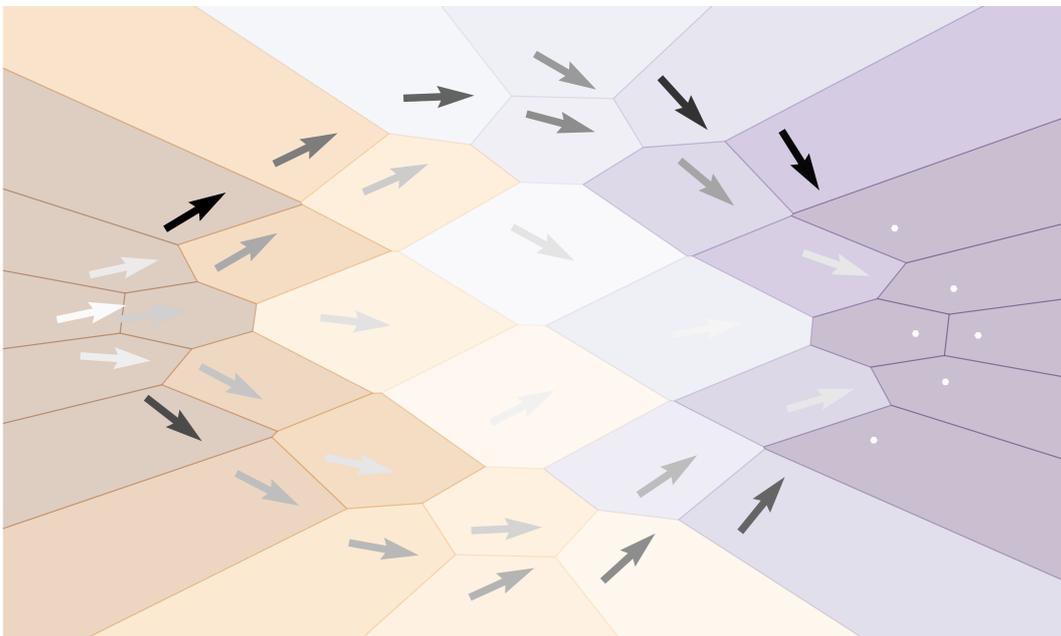
Dissertation

zur Erlangung des Grades eines Doktors der Naturwissenschaften
am Fachbereich Mathematik und Informatik der Freien Universität Berlin

vorgelegt von
Luzie Helfmann

Berlin, März 2022

Contents

Introduction

In a highly interconnected world with diverse actors, social interactions are becoming increasingly complex. Recent advances in data collection and growing computing power allow us to analyse the mechanisms of these social dynamics in new ways. Instead of modelling from the top down, we can use agent-based models (ABMs) to model the complex behaviours and interactions of many individual actors, called agents, from the bottom up [44]. This lets us assume only the behavioural rules of individual agents without knowledge of the collective dynamics of the whole population. The hope is that the local interactions of agents on the micro-scale will permit a more accurate description and better understanding of the dynamics and structures that emerge from them on the macro level of populations. Agents often interact with their local neighbours to pass on information, mimic or reinforce each other, resulting in formations or clusters of similar behaviour among agents. When the formed structures are meaningful on the larger scale, they constitute an example of an emergent structure or pattern. Examples of such structures are the spreading of a virus to peers resulting in local clusters of infected agents, the reinforcement of opinion bubbles on social media or the formation of schools of fish through alignment with neighbouring fish.

In agent-based models one is not so much interested in the stable states of the system, but rather the transient dynamics that lead to a stable state. When studying the spread of a virus one is not interested in the states where the virus or the population is extinct, but in the spreading dynamics leading to such a state. In an opinion model one is interested in the details of how either an opinion consensus or a polarized state is reached. More recently, there has been an increased interest in the tipping dynamics of social behaviours, that is, the transitions from a rather stable population state to another stable state triggered by external changes such as policies. These tipping dynamics can form a transition from a stable but socially undesirable state to a stable desirable state in the context of addressing global challenges such as climate change or biodiversity loss [51, 75]. For example when a city starts to expand its public transport system, it becomes more attractive to use buses and the subway instead of the car. If, in turn, more people use public transport, further expansion of the public transport system becomes necessary. This self-reinforcing feedback can cause a transition towards a socially favourable state. Another example is the rise of protests or social movements [26]. The more people take part in a protest the larger its impact, this feedback loop can turn an initially small protest into a large movement.

Research question and approach. Realistic social tipping processes display great complexity in terms of their drivers and mechanisms, leading to different possible transition paths [75]. In order to gain a better insight into these processes, specific methods for studying transition paths in forced systems are useful. The main focus of this thesis is to devise a method that allows a detailed analysis and understanding of tipping and other relevant transition paths in agent-based models.

Since there is not yet a unified mathematical formalisation of agent-based models and it is usually only assumed that the ABM can be written as a computer program, we will first introduce the class of agent-based models that we will consider.

1. We assume that the ABM is a Markov process. The inherent unpredictability and variability of human behaviour requires a stochastic model, and when this stochastic model is defined iteratively in terms of the most recent state of the system it is a Markov process. The assumption of a Markov process does not mean that agents cannot be influenced by their past states since the state space can be enlarged to include a memory of past states. In [30] it was indeed shown that several well-known agent-based models can be understood as Markov chains.

2. We assume the model to be non-stationary. ABMs are often not considered in their metastable or absorbing states but instead on a transient path leading to such states. Thus they are non-stationary, i.e., the distribution of the visited states does not correspond to the stationary distribution. Moreover, tipping is often triggered by external influences making the system dynamics even time-inhomogeneous.

3. We consider the model to be high-dimensional. Agent-based models usually describe large populations of autonomous agents with the size of the state space growing exponentially with the number of agents. For example when each agent can adopt one of two discrete opinions, the size of the state space is 2^N where N is the number of agents.

We need a method to study tipping for the above defined class of models. Due to the stochasticity and possibly high complexity of ABMs, there may be several very different and complicated transition paths leading from one set of stable or undesired states to another set of stable or desired states. We will call the set where the transition origins the source, and the set where the transition ends the sink.

To investigate these possibly complicated transition paths we consider the ensemble of all trajectories leading from the source to the sink and quantify them using Transition Path Theory (TPT) [21, 47, 71]. The main objects of TPT are the committor functions which give us information about the dynamical proximity to the source respectively the sink set. These committor functions allow us to find out about the most important transition channels, the bottlenecks during transitions or the frequency of transitions. But so far the theory of committors and TPT has only been developed for stationary ergodic Markov processes. To reach the goal of being able to study transition paths in ABMs, we generalize the committor equations and transition statistics of TPT to non-stationary Markov chains. We provide a general formulation of TPT and will highlight

and discuss the details of three special cases, namely, (i) time-dependent Markov chains on finite time intervals, (ii) periodically-varying dynamics, and (iii) dynamics with absorbing states. We expect the results to be generalizable to Markov jump processes and diffusion processes. Markov chains are a good starting point since they constitute the simplest form of a Markov process but can still describe rather rich dynamics. Many ABMs are already in the form of Markov chains or can be discretized in space and time to form one.

In the case that the dynamics of the ABM takes place in a very high-dimensional state space, we will need to reduce the model in order to solve for the committor functions. The appearance of emergent structures on the macro-scale and the presence of symmetries imply that the effective state space of the ABM dynamics is much smaller with fewer degrees of freedom. This effective state space can be parametrized by a few coordinates (e.g., the number of agents that have one opinion vs. those that have another opinion) and the effective dynamics in terms of these so-called collective variables can be estimated facilitating the application of Transition Path Theory to study tipping.

Guide through the thesis. We now give a short summary of each chapter of this thesis highlighting new results and findings.

Chapter 1 introduces the main background theory and tools for this thesis. We will introduce Markov chains, explain how continuous-space Markov processes can be discretized into Markov chains, and present Diffusion Maps, a dimension reduction algorithm that can be used for parametrizing lower dimensional structures in a data set.

In *Chapter 2* we provide a general formulation of the committor equations and Transition Path Theory for Markov chains on finite state spaces. We discuss several special cases of this formulation. As a first special case, we apply the formulation to stationary ergodic Markov chains. Therein we augment the existing stationary theory by new results and insights and also present a short application of TPT to study pollution paths of debris particles on the ocean surface that was published in [48]. As a second and third case, we cover the theory for periodically-varying Markov chains and Markov chains with time-dependent transition probabilities on a finite time interval allowing studies of transition paths in more realistic agent dynamics. The results of the last two cases were also published in [28].

In *Chapter 3* we describe two ways of how Markov chains with absorbing states can be modelled prior to their absorption into some terminal state. This allows us to compute the committor functions and study their transitions before being absorbed.

Chapter 4 elucidates how one can directly sample transitions leading from A to B. We extend the approach in [11] to construct a transition matrix that samples statistically exact transitions and depends on the TPT quantities. Additionally, we summarize two approaches [47, 6] for finding the most likely transition paths and most frequently taken cycles during transitions that will be needed in Chapter 6 for analysing tipping paths

in ABMs.

In *Chapter 5* we discuss several tipping mechanisms such as tipping induced by slow or by fast external parameter variations or due to noise. We demonstrate that all these forms of tipping can be studied with TPT.

In *Chapter 6* we finally study tipping in agent-based models after a model reduction of the high-dimensional ABM to a much smaller state space that still allows a description of the ABM's emergent patterns. More precisely, we employ Diffusion Maps to find suitable collective variables for the ABM and estimate a transition matrix on a discretization of the projected state space. This work resulted in a publication [29].

1 | Mathematical preliminaries

Stochastic processes can be used as models for the evolution of many real-world systems containing randomness, e.g., in finance, on the molecular scale, or in the social sciences. From a modeling perspective, randomness usually enters the dynamics by neglecting either certain external influences of the system or internal influences and substituting them by noise. The evolution of the model is therefore no longer deterministic, instead several different outcomes are possible.

In this chapter we will introduce the theoretical building blocks of this thesis. In the first part of this chapter we cover important aspects of Markov chains, i.e., of stochastic processes that are indexed by a discrete time, take values in a discrete state space and satisfy the Markov property. We closely follow and recommend the following books for further details [50, 10]. The second part is about the spatial and temporal discretization of continuous Markov processes into a Markov chain as well as about Diffusion Maps, a technique for finding a lower-dimensional representation of a data set by constructing a certain reversible Markov chain between the data points. Since continuous-time and -space Markov processes are of lesser importance in this thesis, we will not cover their basics and instead refer the reader to the following references on stochastic processes [34, 52].

1.1 Markov chains and asymptotics

Before we can define a Markov chain on a discrete and finite state space $\mathbb{X} = \{1, 2, \ldots, S\}$, we need to introduce the terms of a distribution and a stochastic matrix. In general, Markov chains can be defined on countable state spaces, but for our purposes finite state spaces are sufficient. A *distribution* on $\mathbb{X}$ can be represented by a vector $v = (v_i)_{i \in \mathbb{X}}$ of non-negative entries v_i summing to 1, i.e., $\sum_{i \in \mathbb{X}} v_i = 1$. A *(row-)stochastic matrix* is a matrix $A = (A_{ij})_{i,j \in \mathbb{X}}$ with non-negative entries and rows summing to 1.

With these objects at hand, let us continue with the definition of a general time-inhomogeneous Markov chain.

Definition 1.1. *A **Markov chain** with initial distribution $\mu = (\mu_i)_{i \in \mathbb{X}}$ on a discrete and finite state space $\mathbb{X} = \{1, 2, \ldots, S\}$ and with a collection of row-stochastic transition matrices $(P(n))_{n \in \mathbb{T}}$ with discrete index set $\mathbb{T} = \{0, 1, \ldots, N\}$ or $\mathbb{T} = \mathbb{N}_0$ is a stochastic process $(X_n)_{n \in \mathbb{T}}$ such that*

1.1. Markov chains and asymptotics

(i) the process at time $n = 0$ is distributed according to the initial distribution, i.e.,

$$\mathbb{P}(X_0 = i) = \mu_i, \tag{1.1}$$

*(ii) for any $j, i_1, \ldots, i_m \in \mathbb{X}$ and $n, n-1, \ldots, n-m \in \mathbb{T}$ the **Markov property** holds*

$$\mathbb{P}(X_n = j \mid X_{n-1} = i_1, \ldots, X_{n-m} = i_m) = \mathbb{P}(X_n = j \mid X_{n-1} = i_1) = P_{i_1 j}(n-1). \tag{1.2}$$

At each time n the *distribution of the Markov chain $\mu(n) = (\mu_i(n))_{i \in \mathbb{X}}$* is defined as the probability of finding the chain in a certain state at that time, i.e., as

$$\mu_i(n) := \mathbb{P}(X_n = i). \tag{1.3}$$

The initial distribution $\mu(0) = \mu$ together with the transition matrices $P(n)$ at times $n \in \mathbb{T}$, tell us about the distribution of the Markov chain at all later times and therefore about the likely whereabouts of the chain at all times. A simple application of the law of total probability

$$
\begin{aligned}
\sum_{i \in \mathbb{X}} \mu_i(n-1) P_{ij}(n-1) &= \sum_{i \in \mathbb{X}} \mathbb{P}(X_{n-1} = i) \mathbb{P}(X_n = j \mid X_{n-1} = i) \\
&= \sum_{i \in \mathbb{X}} \mathbb{P}(X_n = j, X_{n-1} = i) \\
&= \mathbb{P}(X_n = j) = \mu_j(n).
\end{aligned} \tag{1.4}
$$

shows that the distribution at time n can be found by multiplying the transition matrix from the right to the distribution at time $n-1$, i.e., that the iterative relation $\mu(n) = \mu(n-1)^\top P(n-1)$ holds.

The probability of the chain to be in states i and then j at two successive times n and $n+1$ is called the *current* or *flux* at time n and can be computed as follows

$$f_{ij}(n) := \mathbb{P}(X_n = i, X_{n+1} = j) = \mu_i(n) P_{ij}(n). \tag{1.5}$$

The current provides information about the average amount of trajectories that move from i at time n to j at time $n+1$.

The Markov property (ii) of Definition 1.1 states that the probability of the chain being in state j at time n depends only on the state at time $n-1$ and not on further information about the chain in the past. This property is the reason why Markov processes are also called *memory-less*. The following theorem allows the Markov property to be understood as the statement that *the future is independent of the past given the present*.

Proposition 1.2. *Let $(X_n)_{n \in \mathbb{T}}$ be a Markov chain, then conditional on $X_m = i$, the future and the past of the process are independent, i.e.,*

$$\mathbb{P}(A \cap B | X_m = i) = \mathbb{P}(A | X_m = i) \, \mathbb{P}(B | X_m = i) \tag{1.6}$$

where A is an event determined by the random variables X_n with $n \leq m$ and B is an event determined by X_n, $n \geq m$.

The Markov property also holds at so-called stopping times, and is then called the strong Markov property. A stopping time is a random time associated to a certain criterion. This criterion has to be such that at each time step of the Markov chain it is possible to evaluate whether the criterion has already been satisfied. More precisely, a stopping time is defined as follows.

Definition 1.3. *A random variable $\tau : \Omega \to \mathbb{T} \cup \{\infty\}$ is called a **stopping time**, if the event $\{\tau = n\}$ depends only on $X_0, X_1, \ldots, X_n$ for $n \in \mathbb{T}$.*

An important example of a stopping time is the first hitting time of a subset of $A \subset \mathbb{X}$ defined as[1]

$$\tau_A^+ := \min\{k \geq 0 \text{ s.t. } X_k \in A\} \tag{1.7}$$

with the convention $\min \varnothing = \infty$. It is a stopping time since

$$\{\tau_A^+ = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}$$

depends only on the random variables $X_0, \ldots, X_n$. Similarly, the first return time

$$T_A := \min\{k \geq 1 \text{ s.t. } X_k \in A\} \tag{1.8}$$

is a stopping time. Note that $T_A \geq 1$ while $\tau_A^+ \geq 0$, since the return time does not care about the state at time 0 and can be used to measure returns to a state. On the other hand, the last exit time of subset A,

$$\tau_A^- := \max\{k \geq 0 \text{ s.t. } X_k \in A\} \tag{1.9}$$

is not in general a stopping time since it depends on the future, in other words, at time n we cannot evaluate whether the stopping criterion has already been satisfied since it depends on whether the Markov chain will hit A again in the future. But the last exit time turns out to be a stopping time of the time-reversed Markov chain that we will learn about in the next section.

1.1.1 Time-independence and the time-reversed process

Since the study of Markov chains is simpler with time-independent transition probabilities or even time-independent distributions, we start by defining some necessary objects for this regime.

Definition 1.4. *A Markov chain is called **time-homogeneous** if the transition matrices do not depend on the time n, i.e., if $P(n) \equiv P$ for all $n \in \mathbb{T}$.*

---

[1]In Transition Path Theory, we need the more general first hitting times after time n, $\tau_A^+(n) := \min\{k \geq n \text{ s.t. } X_k \in A\}$. These only depend on the process after time n and are also stopping times.

1.1. Markov chains and asymptotics

Definition 1.5. *A distribution π is called **stationary or invariant** with respect to a time-homogeneous transition matrix P if $\pi^\top = \pi^\top P$.*

We will return to the question under which conditions a unique stationary distribution exists at a later time.

Definition 1.6. *A Markov chain $(X_n)_{n \in \mathbb{T}}$ is **stationary**, if all finite-dimensional distributions are invariant under time shifts, i.e., if*

$$\mathbb{P}(X_n = i_0, \ldots, X_{n+k} = i_k) = \mathbb{P}(X_{n+p} = i_0, \ldots, X_{n+k+p} = i_k) \tag{1.10}$$

holds for all times in $\mathbb{T}$ and all states $i_0, \ldots, i_k \in \mathbb{X}$.

In particular, if a time-homogeneous Markov chain has a stationary distribution π as its initial distribution, then the chain for all times stays in this distribution. It follows that the finite-dimensional distributions are invariant under time shifts and the chain is therefore stationary.

By considering a Markov chain backwards in time, we get another Markov chain, called the *time-reversed* or *backward* Markov chain as stated in the following theorem:

Theorem 1.7. *Let $(X_n)_{n \in \mathbb{T}}$ be a Markov chain on the time interval $\mathbb{T} = \{0, \ldots, N\}$ with transition matrix $P(n)$, initial distribution $\mu(0)$, and final distribution $\mu(N)$ at time N. Then the **time-reversed** or **backward** Markov chain $(X_n^-)_{n \in \mathbb{T}}$ is given by $X_n^- := X_{N-n}$. It is again a Markov chain but with initial distribution given by $\mu(N)$ and transition probabilities given by the **backward transition matrix***

$$P_{ij}^-(n) := \mathbb{P}(X_{n+1}^- = j \mid X_n^- = i) = \mathbb{P}(X_{N-n-1} = j \mid X_{N-n} = i)$$

$$= \begin{cases} P_{ji}(N - n - 1) \frac{\mu_j(N-n-1)}{\mu_i(N-n)} & \text{if } \mu_i(N - n) > 0 \\ S^{-1} & \text{else,} \end{cases} \tag{1.11}$$

where S is the size of the state space.

Proof. The proof follows by checking that the transition matrix is indeed row-stochastic and that the Markov property is satisfied. $\qquad\square$

An interesting question is how the time-dependence of the forward process affects the time-reversed chain. From the transition probabilities of the time-reversed chain, see Eq. (1.11), we can note that time-homogeneity of the forward process is not enough to guarantee that the time-reversed process has time-homogeneous transition probabilities. It is additionally needed that the distributions of the forward process are time-independent and hence that the chain is stationary. The reason why time-homogeneity of the forward process is not enough is the asymmetry of time. As time progresses, the distribution under a time-homogeneous transition matrix P equilibrates. The time-reversed Markov chain passes through the same distributions as the forward chain

but in backward time. The time-reversed transition matrix has to work against this equilibration progress by reversing all the individual fluxes which in general makes the matrix time-inhomogeneous.

In order to reverse a stationary Markov chain $(X_n)_{n\in\mathbb{N}_0}$ with stationary distribution π on the time index set $\mathbb{T} = \mathbb{N}_0$, we first have to extend the chain to a stationary chain $(\tilde{X}_n)_{n\in\mathbb{Z}}$ on the time set $\mathbb{T} = \mathbb{Z}$ by sending the initial condition given by a stationary distribution π to the infinite past. Then the time-reversal of $(\tilde{X}_n)_{n\in\mathbb{Z}}$ is simply given by $\tilde{X}_n^- := \tilde{X}_{-n}$, which is also stationary and admits the same stationary distribution π. Its backward transition probabilities are given by

$$P_{ij}^- = P_{ji}\frac{\pi_j}{\pi_i} \tag{1.12}$$

which follows from the stationarity of the chain

$$\mathbb{P}(\tilde{X}_{n+1}^- = j \mid \tilde{X}_n^- = i) = \frac{\mathbb{P}(\tilde{X}_{n+1}^- = j, \tilde{X}_n^- = i)}{\mathbb{P}(\tilde{X}_n^- = i)}$$

$$= \mathbb{P}(\tilde{X}_n^- = i \mid \tilde{X}_{n+1}^- = j)\frac{\mathbb{P}(\tilde{X}_{n+1}^- = j)}{\mathbb{P}(\tilde{X}_n^- = i)} = P_{ji}\frac{\pi_j}{\pi_i}.$$

1.1.2 Classification of states

We now come to some properties that tell us how the states of the Markov chain are connected through the transition probabilities. For the remainder of this chapter, we consider time-homogeneous Markov chains with a time-independent transition matrix P.

Definition 1.8. *We say that state i **leads to state** j, write $i \rightarrow j$, if there is a non-zero probability when starting in state i that the Markov chain eventually visits state j, i.e., if there exists some $n \in \mathbb{T}$ s.t. $\mathbb{P}(X_n = j|X_0 = i) > 0$. We say i **communicates with** j, write $i \leftrightarrow j$, whenever $i \rightarrow j$ and $j \rightarrow i$.*

Note that always $i \leftrightarrow i$ since $\mathbb{P}(X_0 = i|X_0 = i) = 1$. The communication relation $\leftrightarrow$ defines an equivalence relation on $\mathbb{X}$, therefore it partitions $\mathbb{X}$ into equivalence classes, so-called communication classes, in which all elements communicate with another. A communication class is called *closed* or *absorbing*, if the Markov chain cannot escape from it. If the Markov chain has only one single communication class $\mathbb{X}$, then the Markov chain is called *irreducible*, it cannot be further reduced and it is possible in a finite number of steps to reach any state from any other state.

On the other hand, if the matrix has several communication classes, then the Markov chain restricted to each closed communication class can be studied separately, has a row-stochastic transition matrix and is again irreducible. But what about the non-closed classes? As we will see soon, the non-closed classes are only visited during the transient phase of a Markov chain.

Definition 1.9. *A state i of a time-homogeneous Markov chain is called* **recurrent** *if*

$$\mathbb{P}(X_n = i \text{ for infinitely many } n \mid X_0 = i) = 1 \tag{1.13}$$

and **transient** *if*

$$\mathbb{P}(X_n = i \text{ for infinitely many } n \mid X_0 = i) = 0. \tag{1.14}$$

A transient state will eventually not be visited anymore whereas to a recurrent state the process will come back to again and again. Each state is either transient or recurrent [50, Theorem 1.5.3.], additionally the states of a communication class are either all transient or all recurrent [50, Theorem 1.5.4.]. If a communication class is recurrent then it is necessarily closed. Also, all finite closed classes are recurrent [50, Theorem 1.5.6.]. Therefore, in particular all irreducible Markov chains on finite state spaces are recurrent.

Another important property of Markov chains is their periodicity. Consider for example the Markov chain given by the transition matrix

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

When the process starts in state 1, it will switch deterministically from state 1 to 2 to 1 to 2 etc. For this Markov chain it is only possible to revisit a state every second time step. Therefore both states of this Markov chain, and actually the whole chain, is called periodic with period 2. On the other hand, when a Markov chain can return to a certain state i eventually at all times in the future, then it is not periodic and called aperiodic.

Let us define more formally what it means for a state to be aperiodic.

Definition 1.10. *A state $i \in \mathbb{X}$ is called* **aperiodic** *if for all sufficiently large times n, $(P^n)_{ii} > 0$. This condition can equivalently be also stated as $\gcd\{n \geq 0 \text{ s.t. } (P^n)_{ii} > 0\} = 1$, where $\gcd$ stands for the greatest common divisor.*

In an irreducible Markov chain, one state i is aperiodic if and only if, all states are aperiodic. Therefore the whole Markov chain is then said to be aperiodic. Whenever a state is not aperiodic, it is *periodic* and a period of

$$d_i = \gcd\{n \geq 0, (P^n)_{ii} > 0\} \tag{1.15}$$

can be assigned to it. Thus a state has period d_i if the chain can at most revisit state i at times that are multiples of d_i. Again in an irreducible Markov chain, all states share the same period.

1.1.3 Existence of a stationary distribution

Next we come to the conditions such that a unique stationary distribution exists and how it is reached.

Theorem 1.11. *Let P be irreducible and $(X_n)_{n \in \mathbb{T}}$ a finite Markov chain, then a unique stationary distribution π exists and it is given by the inverse of the expected return time of a state*

$$\pi_i = (\mathbb{E}(T_i \mid X_0 = i))^{-1}, \tag{1.16}$$

where $T_i = \min\{k \geq 1 \text{ s.t. } X_k = i\}$ is the first return time of state i.

Alternatively for finite state spaces the Perron-Frobenius Theorem [63] can be used to show that for irreducible transition matrices, a unique positive stationary distribution exists.

The next theorem states that if the chain is additionally aperiodic, then for any initial distribution, the chain converges to the stationary distribution.

Theorem 1.12. *Consider a Markov chain $(X_n)_{n \in \mathbb{T}}$ with irreducible and aperiodic transition matrix P. Then for all j and for all initial distributions,*

$$\mathbb{P}(X_n = j) \to \pi_j$$

as $n \to \infty$, where π is the unique stationary distribution of P.

Let us try to understand what may go wrong with the convergence in a periodic Markov chain by considering the following transition matrix $P = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right)$ of period 2. Due to the periodicity, the powers of the transition matrix P^n giving the $n-$step transition probabilities $\mathbb{P}(X_n = j \mid X_0 = i)$ will not converge for large n, here $P^n = P$ when n is odd and $P^n = I$ when n is even. Thus when the initial distribution is μ, the distribution at time n is given by μ when n is even and $\mu^\top P$ when n is odd. Unless the distribution μ is the stationary distribution, the distributions in time will switch periodically. It can be shown that for an irreducible, periodic Markov chain of period d, the limiting distribution will periodically vary with period d [50, Theorem 1.8.5].

1.1.4 Reversible Markov chains

Definition 1.13. *A Markov chain $(X_n)_{n \in \mathbb{T}}$ with initial distribution μ and irreducible transition matrix P is said to be reversible if for all $N \in \mathbb{T}$, the time-reversed chain $(X_{N-n})_{0 \leq n \leq N}$ is also a Markov chain with initial distribution μ and transition matrix P.*

So reversibility means that the forward process is indistinguishable from any backward process. From the above definition it immediately follows that a necessary condition for reversibility is that the initial distribution μ is stationary since the distribution has to remain unchanged in time. Hence the Markov chain has to be stationary. The following theorem gives a simple condition to check if a Markov chain is reversible.

Proposition 1.14. *A Markov chain $(X_n)_{n \in \mathbb{T}}$ with transition matrix P and initial distribution μ is reversible if and only if P and μ satisfy the so-called **detailed balance** relation:*

$$\mu_i P_{ij} = \mu_j P_{ji} \tag{1.17}$$

for all $i, j \in \mathbb{X}$.

A distribution that satisfies the detailed balance condition is stationary, since we can sum the relation $\sum_{i \in \mathbb{X}} \mu_i P_{ij} = \mu_j \sum_{i \in \mathbb{X}} P_{ji} = \mu_j$ using that the rows of the transition matrix sum to 1.

The detailed balance condition can be interpreted in two ways: either as a balance equation between the current from i to j and the current from j to i resulting in a net-zero change. Or, as a statement that the forward and backward transition probabilities agree. The backward transition probabilities of a stationary Markov chain with stationary distribution π and transition matrix P are given by $P_{ij}^- = P_{ji} \frac{\pi_j}{\pi_i}$. If the Markov chain fulfills the detailed balance relation, then $\pi_i P_{ij} = \pi_j P_{ji}$, hence in that case $P^- = P$.

1.1.5 Asymptotics

The next three theorems give conditions under which the average along a sequence of random variables converges to a mean. They provide the theoretical foundation for approximating the mean by an average of a finite number of random variables.

The first theorem, the strong law of large numbers, applies to random variables that are independent and identically distributed and with a finite mean.

Theorem 1.15 (Strong law of large numbers). *Let $(Y_n)_{n \in \mathbb{N}_0}$ be a sequence of independent, identically distributed random variables with mean $\mathbb{E}(Y_0) = \mu$. Then,*

$$\mathbb{P}\left(\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} Y_k = \mu \right) = 1. \tag{1.18}$$

The ergodic theorem gives the conditions for estimating the π-weighted average of some function by the average along a stationary trajectory exploring the whole state space. This if often referred to by the statement: *the space average equals the time average*.

Theorem 1.16 (Ergodic Theorem for Markov chains [50]). *Let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain on a finite state space $\mathbb{X}$ with irreducible, and time-homogeneous transition matrix P and with any initial distribution. Then for any bounded function $f : \mathbb{X} \to \mathbb{R}$,*

$$\mathbb{P}\left(\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \bar{f} \right) = 1, \tag{1.19}$$

where $\bar{f} = \sum_{i \in \mathbb{X}} \pi_i f_i$ is the average of f with respect to π, the unique stationary distribution of P.

We call a Markov chain *ergodic* when it satisfies the assumptions of the ergodic theorem.

And the last theorem in this respect shows how weighted cycles that are traversed by an ergodic Markov chain converge to the current of the Markov chain.

Theorem 1.17 (Cycle decomposition). *Let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with irreducible transition matrix P and any initial distribution. Denote by Γ the set of all cycles $\gamma = (i_1, \ldots, i_s, i_1)$ with distinct $i_1, \ldots, i_s \in \mathbb{X}$ that are associated to positive transition probabilities along all edges of the cycle. Here, a cycle is understood as an equivalence class containing all cyclic permutations of the cycle. By W_n^γ we denote the number of occurrences of the cycle γ in the Markov chain up to time n, i.e., the number of times the chain passes through the states $i_1, \ldots, i_s, i_1$ up to cyclic permutations. A cycle does not have to be visited in consecutive time steps but can be interrupted by visits to other cycles.*

Then the following almost sure convergence holds

$$\pi_i P_{ij} = \lim_{n \to \infty} \sum_{\gamma \in \Gamma} \frac{W_n^\gamma}{n} C_{ij}^\gamma \tag{1.20}$$

where π is the stationary distribution of P and C_{ij}^γ is 1 or 0 depending on whether (i, j) is an edge of γ or not.

Proof. Proofs can be found in [33, Theorem 3.3.1.] and [32, 1.3.5.]. □

The quantity $\pi_i P_{ij}$ gives the current between i and j of the stationary Markov chain. Thus the previous theorem gives a decomposition of the current between two states into fragments of the current carried by different cycles. For reversible Markov chains, it holds that the weight $w^\gamma := \lim_{n \to \infty} \frac{W_n^\gamma}{n}$ corresponding to a cycle $\gamma = (i_1, i_2, \ldots, i_s, i_1)$ agrees with the weight of the reversed cycle $\gamma^- = (i_s, i_{s-1}, \ldots, i_1, i_s)$.

1.2 Discretization and dimension reduction

After having recalled some fundamentals about Markov chains, we now continue by discussing methods to deal with continuous processes or high-dimensional data. We

start with the discretization of a continuous-time and space Markov process into a Markov chain simplifying its further analysis. We then introduce a method that allows the dimension reduction of a cloud of data points from a high-dimensional space to a low-dimensional space. These data points can be samples from some distribution, or form a realization of a stochastic process.

1.2.1 Discretization from trajectory data

In this first part we explain how to discretize a continuous-time Markov process $(X_t)_{t \in \mathbb{T}}$, $\mathbb{T} = [0, \infty)$ on a continuous state space $\mathbb{X} \subseteq \mathbb{R}^d$, $d \geq 1$ in terms of a Markov chain between subsets of $\mathbb{X}$. This coarse-graining approach is at the core of Markov State Modeling [62, 60] where high-dimensional molecular processes are approximated by Markov chains on metastable subsets. In other fields, the approach is also known under the name Ulam's method. We assume that the original, continuous-time and -space Markov process is time-homogeneous and stationary with measure π.

We start by defining a partition of $\mathbb{X}$ as a collection of non-overlapping sets $\{A_i\}_{i=1,\ldots,N}$ that cover the state space $\mathbb{X} = \cup_{i=1}^N A_i$. Assuming that each set is non-empty $\pi(A_i) > 0$, the transition probabilities

$$\mathbb{P}(X_{\Delta t} \in A_j \mid X_0 \in A_i) =: \mathbb{P}(\hat{X}_1 = j \mid \hat{X}_0 = i) = \hat{P}_{ij} \tag{1.21}$$

define a stationary Markov chain $(\hat{X}_n)_{n \in \mathbb{N}_0}$ on the discrete state space $\hat{\mathbb{X}} = \{1, \ldots, N\}$. The state $i \in \hat{\mathbb{X}}$ corresponds to the set A_i and one time step of the Markov chain represents a Δt−sized time step of the original Markov process.

The matrix $(\hat{P}_{ij})_{i,j \in \hat{\mathbb{X}}} = \hat{P}$ is stochastic since the disjoint sets $\{A_i\}_{i=1,\ldots,N}$ partition the state space. Moreover, the stationary distribution $(\hat{\pi}_i)_{i \in \hat{\mathbb{X}}}$ of the discrete Markov chain is given by the invariant measure of the partitioning sets $\hat{\pi}_i = \pi(A_i)$ [62, Lemma 1].

The transition matrix $\hat{P}$ also results from the orthogonal projection of the transfer operator, which is the continuous space analogue of the transition matrix evolving densities in time, onto the basis spanned by indicator functions on the partitioning sets $\{\mathbb{1}_{A_i}\}_{i=1,\ldots,N}$ [60, Theorem 1]. For reversible processes, the approximation error can be decreased by choosing sets that allow a good approximation of the dominant eigenfunctions of the original transfer operator [60]. In some situations better approximation results can be achieved by a projection of the transfer operator onto the space spanned by eigenfunctions [59], committors [60, Chapter 2.3.], or collective variables such as Diffusion Maps [67].

Usually, the probabilities (1.21) are difficult to compute or unknown. Computing trajectories of the original continuous stochastic process, on the other hand, is often easy. Therefore, quantities like (1.21) are usually approximated by a Monte-Carlo sum. We can rewrite the transition probabilities as conditional expectations

$$\hat{P}_{ij} = \mathbb{E}(\mathbb{1}_{A_j}(X_{\Delta t}) \mid X_0 \in A_i) \tag{1.22}$$

where $\mathbb{1}_A : \mathbb{X} \rightarrow \{0, 1\}$ is the indicator function of the set $A \subset \mathbb{X}$ that takes the value 1 when the argument lies in A and 0 else. Then, using i.i.d. samples $(x_k, y_k)_{k=1,\ldots,K}$ with x_k drawn from the stationary density on A_i and y_k being the evolution for time Δt of a trajectory started at x_k, we can approximate

$$\hat{P}_{ij} \approx \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{A_j}(y_k) =: \bar{P}_{ij}^K, \tag{1.23}$$

i.e., by the fraction of short trajectories started in A_i that end up in A_j. The matrix entries $\bar{P}_{ij}^K$ converge a.s. to $\hat{P}_{ij}$ by the strong law of large numbers in the limit of $K \rightarrow \infty$. Note that as required for a valid approximation, the matrix $\bar{P}^K$ is stochastic. Alternatively, on the basis of the ergodic theorem the matrix can also be estimated by the proportion of transitions between the sets in a long ergodic realization.

1.2.2 Example: The overdamped Langevin equation

As an illustrative example throughout this thesis, we often consider the overdamped Langevin process discretized as a Markov chain on rectangular boxes. The overdamped Langevin process is a homogeneous Markov process $(X_t)_{t \geq 0}$ on $\mathbb{R}^d$ satisfying the following stochastic differential equation (SDE)

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{1.24}$$

where $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is a potential function, $\sigma > 0$ determines the noise strength and $(W_t)_{t \geq 0}$ is a $d-$dimensional standard Wiener process (or Brownian motion). We will refer the reader to [52] for further details on the matter of SDE's.

The overdamped Langevin equation describes the random changes of $X_t \in \mathbb{R}^d$, which is usually interpreted as the position of a particle at time t. The first term of the SDE, $-\nabla V(X_t)$, gives a force that pushes the particle "downhill" in the direction of the gradient of the potential landscape. The second term adds normally distributed random kicks. The particle is therefore drawn towards the minima of the potential while simultaneously experiencing small random pushs in all directions and can therefore also move "uphill" and cross barriers of the potential. The process described by the SDE is reversible. When the potential V is confining, i.e., satisfying a sufficient growth condition, it is additionally ergodic and the invariant density is given by

$$\pi(x) = \tfrac{1}{Z} \exp\left(-\tfrac{2}{\sigma^2} V(x)\right), \tag{1.25}$$

where Z is a normalization constant.

The Euler-Maruyama scheme can be used to sample solutions of the SDE with a discrete time step Δt. Given an initial value, the time-discretization with fixed time

1.2. *Discretization and dimension reduction*

steps of size Δt reads [35]

$$X_{n+1} = X_n - \nabla V(X_n)\,\Delta t + \sigma\sqrt{\Delta t}\,\xi_n, \tag{1.26}$$

where ξ_n are $d-$dimensional vectors of i.i.d normally distributed random variables with mean 0 and variance 1.

Exemplarily, we will discuss the motion in the 1D potential $V(x) = \frac{1}{4}(1-x^2)^2$ which has two minima at ± 1. As a noise strength we choose $\sigma = 0.5$. In Fig. 1.1(a) the potential and the corresponding stationary density is shown, in Fig. 1.1(c) we show a realization that was sampled using the Euler-Maruyama scheme with $\Delta t = 0.1$.

(a) The potential and corresponding stationary density.

(b) The discretized stationary distribution on the grid cells.

(c) A metastable realization.
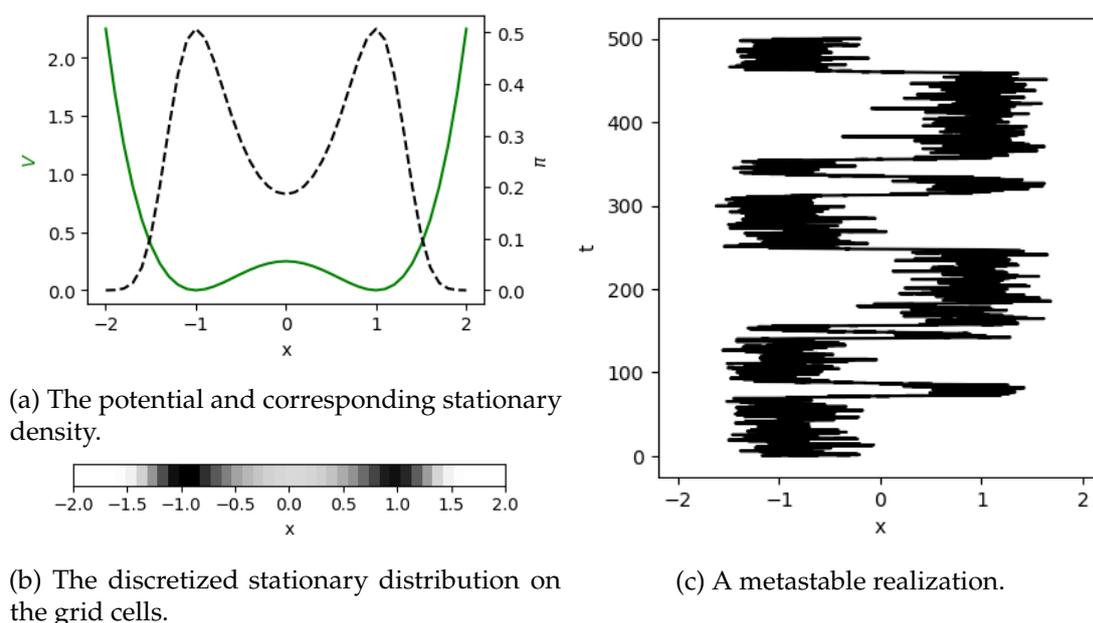
Figure 1.1: The overdamped Langevin process in a double well landscape.

The stationary density is concentrated around the two wells. The particle will spend most of its time near the minima of the well. These areas are *metastable* and attract the particle for a considerable amount of time. Metastability is a relaxed notion of stability for stochastic processes. But the realization also shows the sudden jumps from one well to the other that happen on a much slower time scale than the time scale of the process. These transitions are also called *rare events* or *noise-induced tipping* events.

We want to discretize this process on $[-2, 2]$ into a Markov chain on a regular grid of cell size 0.1 and time step size 0.1. The Euler-Maruyama scheme can be used for sampling i.i.d. short realizations of the SDE and estimating a transition matrix as in Eq. (1.23). The corresponding unique stationary distribution is displayed in Fig. 1.1(b) which quantitatively agrees with the stationary density.

1.2.3   Dimension reduction by Diffusion Maps

In many applications one encounters high-dimensional data sets that are difficult to analyse. Consider a set of M data points in a high-dimensional Euclidean space $\mathbb{D} = \{x_1, \ldots, x_M\} \subset \mathbb{R}^n$. The goal of dimension reduction is to find a meaningful mapping of the given sample into a low-dimensional Euclidean space

$$\xi : \mathbb{R}^n \to \mathbb{R}^d, \ d < n.$$

Dimension reduction methods can be split into those employing linear mappings such as Principal component analysis [53] which only work well if the data points are linearly related, and those using non-linear mappings such as Isomap [66] and Diffusion Maps [14, 36]. For these non-linear methods it is assumed that the data points are approximately sampled from a $d-$dimensional manifold embedded in the high-dimensional space $\mathbb{R}^n$. To unravel the structure of the manifold, the Euclidean distances between near-by data points can be used to approximate their geodesic distance on the manifold, while Euclidean distances between far-away data points are usually not a good approximation of their distance on the manifold. Here we will introduce the non-linear method Diffusion Maps which has the advantageous properties of being robust to noise and computationally inexpensive.

The general idea of Diffusion Maps is to define a random walk on the data points $\mathbb{D}$, where the transition probability between similar or near points is high and between far points is close to zero. The hope is that the random walk traverses the manifold and only follows its intrinsic structure. Then the dominant eigenpairs of the resulting transition matrix contain information about the different scales of the data and can be used for the mapping.

The transition matrix on the data points $\mathbb{D}$ is constructed as follows:

1. Choose a rotation-invariant kernel[2] $k^\epsilon(x, y) = h\left(\frac{\|x-y\|_2^2}{\epsilon}\right)$ that describes the similarity or closeness of two data points, for example the popular Gaussian kernel given by $h(z) = \exp(-z)$. Moreover one has to set the scale parameter $\epsilon > 0$, e.g., guided by the heuristic from [8, 36], that depends on the size of the data set.

2. Letting $q_j^\epsilon = \sum_{m=1}^M k^\epsilon(x_j, x_m)$, we evaluate the kernel at the data points and reweigh it to form the following kernel matrix

$$K_{ij}^\epsilon = \frac{k^\epsilon(x_i, x_j)}{q_i^\epsilon q_j^\epsilon}. \tag{1.27}$$

   The re-weighting cancels the bias of a nonuniform data sampling that is not connected to the geometry of the manifold. For more details and other possible re-weightings see [14].

---

[2]A kernel $k : X \times X \to \mathbb{R}$ is a non-negative and symmetric function.

3. Applying row-normalization of each row i by $d_i^\epsilon = \sum_{m=1}^{M} K_{im}^\epsilon$, we arrive at the row-stochastic matrix

$$P_{ij}^\epsilon = \frac{K_{ij}^\epsilon}{d_i^\epsilon}. \tag{1.28}$$

Since K^ϵ is symmetric, the Markov chain on the data described by P^ϵ is reversible with respect to the stationary distribution $\pi_i = \frac{d_i^\epsilon}{\sum_j d_j^\epsilon}$. The right eigenpairs (λ_j, ψ_j), $j = 0, \ldots, M - 1$ of P^ϵ contain information about the geometric structure of $\mathbb{D}$ at different scales and are real-valued due to P^ϵ being reversible. We order the eigenpairs by decreasing magnitude of their eigenvalues. Then the leading eigenvectors, i.e., with the largest eigenvalues in magnitude, scaled by their corresponding eigenvalue, are a good projection of the large-scale structures in the data

$$\xi(x_i) = (\lambda_1 (\psi_1)_i, \ldots, \lambda_d (\psi_d)_i) \in \mathbb{R}^d, \tag{1.29}$$

where $(\psi_j)_i$ is the ith component of the jth eigenvector. Since the eigenvector corresponding to the largest eigenvalue is just the 1-vector and contains no information, we excluded it from the mapping. As d we often choose the number of eigenvalues above the spectral gap. Whenever all eigenpairs are used for the mapping, we speak of a full mapping.

The reason why the projection using the eigenpairs is so valuable is that the Euclidean distances in these low-dimensional coordinates approximately correspond to the local diffusion distances on the manifold. It can be shown that [14]

$$\begin{aligned}
\|\xi(x_i) - \xi(x_j)\|_2^2 &= \sum_{l=1}^{d} \lambda_l^2 \left((\psi_l)_i - (\psi_l)_j \right)^2 \\
&\approx \sum_{l=0}^{M-1} \lambda_l^2 \left((\psi_l)_i - (\psi_l)_j \right)^2 \\
&= \sum_{l=0}^{M-1} \frac{|P_{il}^\epsilon - P_{jl}^\epsilon|^2}{\pi_l} =: D(x_i, x_j)^2
\end{aligned} \tag{1.30}$$

where it was used that the eigenvalues $\lambda_{d+1}, \ldots, \lambda_{M-1}$ are negligibly small. The diffusion distance D measures how the transition probabilities from two starting data points differ and thus reflects the connectivity structure of the data. The approximation becomes exact when the full embedding is used.

The computational cost of computing pair-wise distances and the eigenvectors of P^ϵ becomes very expensive if not impossible for very large data sets. To circumvent that, one can sub-sample the data set, compute the matrix and eigenpairs only for the sub-sample and interpolate the computed eigenvectors at the remaining data points with the help of the out-of-sample extension [15]. We refer the reader to [36] for an explanation of the extension.

2 | General Transition Path Theory

In this chapter we will introduce and generalize the main theory of this thesis, *Transition Path Theory* (TPT), which allows to study Markov processes regarding the properties of transitions between two predefined sets of the state space.

In stochastic dynamics it is possible for different trajectories started at the same initial condition to end up in very faraway places after a short amount of time. With Transition Path Theory we can filter out information only about the trajectories that started in some subset *A* of the state space, also called the *source set*, and end in some other set *B*, also called the *sink set*, without in-between returning back to *A*. These trajectory snippets that carry out a transition from *A* to *B* henceforth are also called *reactive trajectories*[1] and are the main focus of TPT, see Fig. 2.1.
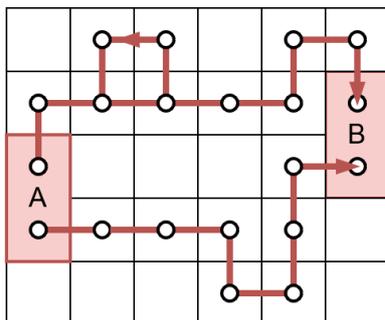
Figure 2.1: Two reactive trajectories that start in *A* and end in *B*. For visualization purposes, we will often employ Markov chains on a discrete grid.

Transition Path Theory exactly quantifies the rate and mean duration of the reactive trajectories, it determines their distribution thereby unveiling the bottlenecks during transitions from *A* to *B*, and gives their flux thus highlighting the importance of pathways. TPT is especially useful when only a small portion of trajectories that start in *A* end up in *B* instead of going back to *A* and when the reactive trajectories take qualitatively different routes from *A* to *B*. The results of TPT depend on the exact choice of the sets *A* and *B*. In some model contexts the set choice might be clear from the application. For example when studying transitions in a bistable system such as the conformational changes of a molecule from the unfolded to the folded state [49], a natural choice for *A* and *B* is given by two metastable regions of the state space.

---

[1] The name stems from the fact that TPT was originally used in the context of chemical reactions between the source (or reactant) set *A* and the sink (or product) set *B*.

TPT was originally developed in [21] for stationary Markov diffusion processes and later extended to stationary Markov jump processes [47] and chains [71]. The theory employs the information contained in the committor functions to quantify the transitions from A to B. The forward and backward committer function give the hitting probabilities of B in forward time resp. of A in backward time and are therefore crucial for quantifying the progress from A to B.

Our goal in this chapter is to provide a general formulation of the committor equations and TPT that also describes transitions in non-stationary dynamics. Non-stationary dynamics appear when a system is either not yet equilibrated or even externally influenced. The developed theory can be useful for investigating transition and tipping events in models of climate or social systems that are often externally influenced. For the underlying dynamics we assume a general time-inhomogeneous Markov chain $(X_n)_{n\in\mathbb{Z}}$ on a discrete and finite state space $\mathbb{X}$. We focus on Markov chains for two reasons: (i) They are the simplest Markov processes to start with, and (ii) by modeling or discretizing high-dimensional complex systems, such as agent-based models, we often arrive at Markov chains [30, 7, 62].

This chapter is based on a publication together with Enric Ribera Borrell [28]. Especially, a number of results from Section 2.4 are due to Enric Ribera Borrell in the context of his Master's thesis [56].

Example. Before we start with formulating the theory, we consider a simple example to demonstrate the advantage of TPT compared to simply analysing the most likely path. Assuming a stationary Markov chain $(X_n)_{n\in\mathbb{Z}}$ on the state space $\mathbb{X}$, the most likely path between source $A \subset \mathbb{X}$ and sink $B \subset \mathbb{X}$ is the path $(x_1, x_2, \ldots, x_{N-1}, x_N)$ with $x_1 \in A$, $x_2, \ldots, x_{N-1} \in (A \cup B)^c$, $x_N \in B$ that maximizes the probability of being observed

$$\mathbb{P}(X_1 = x_1, \ldots, X_N = x_N) = \pi_{x_1} P_{x_1 x_2} \ldots P_{x_{N-1} x_N}, \tag{2.1}$$

where π is the stationary distribution of the Markov chain with transition matrix P.[2] But for Markov chains individual paths often only carry little weight, i.e., have small path probabilities, and do not allow us to infer the general transition mechanisms. To observe this, let us consider the Markov chain in Fig. 2.2(a) with transition probabilities indicated on the edges and $0 < \epsilon < 1$. The probability of observing the lower transition path $A \to 2 \to B$ is

$$\mathbb{P}(X_1 = A, X_2 = 2, X_3 = B) = \pi_A \epsilon \tag{2.2}$$

while the probability of observing the upper path $A \to 1 \cdots \to 1 \to B$ with m self-transitions in state 1 is given by

$$\mathbb{P}(X_1 = A, X_2 = 1, \ldots, X_{2+m} = 1, X_{3+m} = B) = \pi_A (1 - \epsilon)^{m+1} \epsilon. \tag{2.3}$$

---

[2]For simplicity, we assume that a unique most likely path exists.
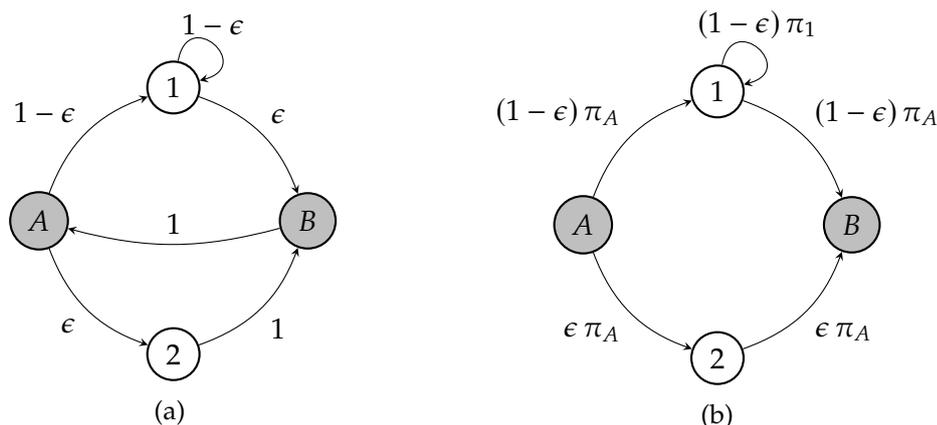
Figure 2.2: A Markov chain with 4 states, (a) transition probabilities, (b) the current of reactive trajectories.

Hence the probability of taking any of the upper paths is given by the following sum

$$\sum_{m=0}^{\infty} \mathbb{P}(X_1 = A, X_2 = 1, \ldots, X_{2+m} = 1, X_{3+m} = B) = \pi_A (1 - \epsilon) \epsilon \sum_{m=0}^{\infty} (1 - \epsilon)^m = (1 - \epsilon) \pi_A.$$

(2.4)

Therefore even though the lower transition path $A \rightarrow 2 \rightarrow B$ is more probable than a single upper transition path $A \rightarrow 1 \cdots \rightarrow 1 \rightarrow B$, it is more probable when ϵ is small ($\epsilon < 0.5$) that any one of the upper paths is taken, in particular, the probability is $(1 - \epsilon) \pi_A$.

Consequently, to get a better picture of the transition behaviour between A and B, it would be necessary to consider more than just the most likely path, for example the k most likely paths. For large state spaces, computing the k most likely paths can be very expensive and the paths can become difficult to analyse due to loops and excursions. TPT circumvents these problems and instead offers a global perspective by giving the transition current that concentrates around certain paths and the distribution of reactive trajectories. In this simple example, the transition current is given in Fig. 2.2(b) and the current along the two different pathways agrees with the above computed probabilities of taking any upper path or the lower path.[3]

2.1 General formulation

In this section we provide a general formulation of Transition Path Theory. We will derive the general form of committor equations followed by a presentation of the different transition statistics that we can obtain from them.

For now, we assume a Markov chain $(X_n)_{n \in \mathbb{Z}}$ with time-dependent distribution $\mu(n)$

---

[3] The current of reactive trajectories will be introduced in Section 2.1.2. By using the simple form of the forward committor: $q_A^+ = 0, q_i^+ = 1$ for $i = 1, 2, B$ and the backward committor: $q_B^- = 0, q_i^- = 1$ for $i = A, 1, 2$, and by additionally using that the amount of current into one node equals the amount of current out of that node, we arrive at the given result.

on the discrete and finite state space $\mathbb{X}$. The possibly time-dependent transition probabilities

$$P_{ij}(n) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

map the chain forward from time step n to $n + 1$ and the transition probabilities[4]

$$P_{ij}^-(n) = \mathbb{P}(X_{n-1} = j \mid X_n = i)$$

map from time step n backward to $n - 1$.

2.1.1 Committor probabilities

We start with the main objects of TPT, the forward and backward committor. We assume that the *source set A* and the *sink set B* are non-empty and disjoint subsets of $\mathbb{X}$ and that the *transition region C* $:= \mathbb{X} \setminus (A \cup B)$ is also non-empty. Then the forward committor tells us the probability of next hitting the sink B rather than the source A conditional on being in a certain state. The backward committor gives the same backward time, i.e., what is the probability of having last been in the source A rather than in the sink B? The backward and forward committor together give us information about how likely it is to have last come from A and to next go to B and therefore together they characterize the transitions from A to B.

For the exact definition of the committors, we need the following two stopping times, the *next hitting time of a set $S \subseteq \mathbb{X}$ after time n*,

$$\tau_S^+(n) := \min\{k \in \mathbb{Z} \text{ s.t. } k \geq n, X_k \in S\}, \ \min \varnothing := \infty \tag{2.5}$$

and the *last hitting time of set S before time n*,

$$\tau_S^-(n) := \max\{k \in \mathbb{Z} \text{ s.t. } k \leq n, X_k \in S\}, \ \max \varnothing := -\infty. \tag{2.6}$$

Note that we call it the next hitting time and last hitting time even though the definitions include the possibility of hitting the set S at the exact time point n.

Given the hitting times, we can define the committor functions. The *forward committor* function gives the probability of rather hitting B than A next when in state i at time n, i.e.,

$$q_i^+(n) := \mathbb{P}\left(\tau_A^+(n) > \tau_B^+(n) \mid X_n = i\right) \tag{2.7}$$

and the *backward committor* tells us the probability of having last come from A rather than B when in state i at time n,

$$q_i^-(n) := \mathbb{P}\left(\tau_A^-(n) > \tau_B^-(n) \mid X_n = i\right). \tag{2.8}$$

---

[4]Here we use a different notation compared to Eq. (1.11), where we let the backward transition matrix depend on the time of the time-reversed chain and not on the time of the forward chain.

For a given Markov chain, we can then find the forward and backward committor function by solving an iterative equation on the transition region $C = (A \cup B)^c$ with appropriate boundary conditions as well as initial and terminal conditions.

Theorem 2.1. *The forward committor $q^+(n) = (q_i^+(n))_{i \in \mathbb{X}}$ for a general Markov chain with forward transition matrix $P(n)$ and backward matrix $P^-(n)$ satisfies the following equation*

$$
\begin{cases}
q_i^+(n) & = \sum\limits_{j \in \mathbb{X}} P_{ij}(n) \, q_j^+(n+1) & i \in C \\
q_i^+(n) & = 0 & i \in A \\
q_i^+(n) & = 1 & i \in B
\end{cases}
\tag{2.9}
$$

for $n < N$ subject to a terminal condition $q^+(N)$ at time N. Analogously, the backward committor $q^-(n) = (q_i^-(n))_{i \in \mathbb{X}}$ solves the following equation

$$
\begin{cases}
q_i^-(n) & = \sum\limits_{j \in \mathbb{X}} P_{ij}^-(n) \, q_j^-(n-1) & i \in C \\
q_i^-(n) & = 0 & i \in B \\
q_i^-(n) & = 1 & i \in A
\end{cases}
\tag{2.10}
$$

for $n > 0$ with a given initial condition $q^-(0)$.

Proof. From the definition of the forward committor in Eq. (2.7), it immediately follows that we have $q_i^+(n) = 0$ when the process is in $i \in A$ at time n since we always have $\tau_A^+(n) = n$, while $\tau_B^+(n) > n$. Analogously we have $q_i^+(n) = 1$ for $i \in B$ since in that case $\tau_A^+(n) > n$ and $\tau_B^+(n) = n$. For the committor in state $i \in C$, we can sum the forward committor at all the other states j weighted with the transition probability to transition from i to j. This follows from

$$
\begin{aligned}
q_i^+(n) &= \mathbb{P}(\tau_B^+(n) < \tau_A^+(n) \mid X_n = i) = \sum_{j \in \mathbb{X}} \mathbb{P}(X_{n+1} = j, \tau_B^+(n) < \tau_A^+(n) \mid X_n = i) \\
&= \sum_{j \in \mathbb{X}} \mathbb{P}(X_{n+1} = j \mid X_n = i) \, \mathbb{P}(\tau_B^+(n) < \tau_A^+(n) \mid X_{n+1} = j, X_n = i) \\
&= \sum_{j \in \mathbb{X}} \mathbb{P}(X_{n+1} = j \mid X_n = i) \, \mathbb{P}(\tau_B^+(n+1) < \tau_A^+(n+1) \mid X_{n+1} = j, X_n = i) \\
&= \sum_{j \in \mathbb{X}} P_{ij}(n) \, q_j^+(n+1)
\end{aligned}
\tag{2.11}
$$

by first using the law of total probability, then conditioning on $\{X_{n+1} = j\}$, using that at time n the chain is in $i \in C$ and thus $\tau_A^+(n), \tau_B^+(n) \geq n + 1$, and last using the Markov property.

For the backward committor equations we can proceed in a similar way by additionally using the time-reversed process. □

Remark 2.2. *The committor equations can also be extended to source and sink sets that change*

in time, thereby enabling the study of transition between time-dependent sets. We might for example wonder, what are the transitions that leave a certain set in the beginning of some time interval and arrive in a certain set at the end of the interval? For this generalization we assume time-dependent sets $A(n)$ and $B(n)$ that are disjoint at each time step n and not empty for all n. Then a generalization of Theorem 2.1 follows by changing the hitting times and the boundary conditions to depend on $A(n)$ and $B(n)$ instead of A and B.

To fully define the committor equations, we need to equip them with appropriate terminal and initial conditions. Later in this chapter we will derive the following scenarios and provide more details on each case:

1. When the committors are time-independent they characterize a stationary Markov chain, compare with Section 2.2.

2. Periodic boundary conditions will allow us to study transitions of Markov chains where the law of the process is periodic, compare with Section 2.3.

3. Solving both equations over the finite time interval $\{0, \dots, N\}$ with conditions $q^+(N) = \mathbb{1}_B$ and $q^-(0) = \mathbb{1}_A$ will restrict the transitions of interest to take place during the time window, i.e., the reactive trajectories have to start in A at time 0 or later and to arrive in B at latest by time N, see Section 2.4.

4. By using as initial and terminal condition a stationary solution to the committor problem, we can study a system that was stationary but experiences a time-dependent shock during a finite time interval, compare with Section 2.4.5.

But before we will discuss these different cases, we will in the next section motivate the usage of committors for analysing transitions. We will define useful statistics of the transitions that are obtained from the committor functions.

2.1.2 Transition statistics

With TPT we learn the characteristics of *reactive trajectories*. As a reactive trajectory we consider a trajectory snippet $(X_n, X_{n+1}, \dots, X_{n+N})$

  (i) that starts in the source set: $X_n \in A$,

 (ii) ends in the sink set: $X_{n+N} \in B$,

(iii) and in-between stays in the transition region: $X_{n+1}, \dots, X_{n+N-1} \in C$.

When on this reactive trajectory it holds for times $k = n, \dots, n + N - 1$ both

$$\tau_A^-(k) > \tau_B^-(k) \text{ and } \tau_B^+(k+1) < \tau_A^+(k+1).$$

A few statistics that we will define below make statements only about reactive trajectories while they are in the transition region C, we will call the part of a reactive trajectory

without the start state in A and the end state in B, an *inner reactive trajectory* (Fig. 2.3). On an inner reactive trajectory the hitting times satisfy

$$\tau_A^-(k) > \tau_B^-(k) \text{ and } \tau_B^+(k) < \tau_A^+(k)$$

for the time points $k = n+1, \ldots, n+N-1$. After this section we will not always stress the difference between inner reactive trajectory and reactive trajectory and sometimes just say reactive trajectory.

Figure 2.3: Two reactive trajectories from A to B. The circles filled in grey highlight time points when the trajectory is in the transition region and thus on an inner reactive trajectory.

Let us start with the first statistical object telling us where the inner reactive trajectories spend most of their time.

Definition 2.3. *The **distribution of inner reactive trajectories** $\mu^{AB}(n) = (\mu_i^{AB}(n))_{i \in \mathbb{X}}$ gives the joint probability that the Markov chain is in a state i at time n while transitioning from A to B:*

$$\mu_i^{AB}(n) := \mathbb{P}\left(X_n = i, \tau_A^-(n) > \tau_B^-(n), \tau_B^+(n) < \tau_A^+(n)\right). \tag{2.12}$$

When the chain is stationary, we will write π^{AB} instead of $\mu^{AB}(n)$ to indicate the stationarity of the chain.

Note that $\mu_i^{AB}(n) = 0$ for $i \notin C$ since we only get information about the distribution of inner reactive trajectories.

Theorem 2.4. *For a Markov chain $(X_n)_{n \in \mathbb{T}}$ with committors $q^+(n), q^-(n)$, the distribution of inner reactive trajectories can be expressed as*

$$\mu_i^{AB}(n) = q_i^-(n)\, \mu_i(n)\, q_i^+(n). \tag{2.13}$$

Proof. We can compute from Eq. (2.12)

$$\mu_i^{AB}(n) = \mathbb{P}\left(\tau_A^-(n) > \tau_B^-(n), \tau_B^+(n) < \tau_A^+(n) \mid X_n = i\right) \mathbb{P}(X_n = i) = q_i^-\, q_i^+\, \pi_i$$

25

by first conditioning on $\{X_n = i\}$, and second by using the independence of the two events $\{\tau_A^-(n) > \tau_B^-(n)\}$, $\{\tau_B^+(n) < \tau_A^+(n)\}$ given $\{X_n = i\}$, which follows from the Markov property as stated in Proposition 1.2. $\qquad\qquad\square$

The distribution of inner reactive trajectories is not normalized but can easily be normalized by dividing $\mu^{AB}(n)$ by the *probability to be on an inner reactive trajectory*,

$$Z^{AB}(n) := \sum_{i \in C} \mu_i^{AB}(n) = \mathbb{P}(\tau_A^-(n) > \tau_B^-(n), \tau_B^+(n) < \tau_A^+(n)), \qquad (2.14)$$

to give a probability distribution on $\mathbb{X}$, therefore we define:

Definition 2.5. *Whenever $Z^{AB}(n) > 0$, we can define the **normalized distribution of inner reactive trajectories** at time n by*

$$\hat{\mu}_i^{AB}(n) := \mathbb{P}\left(X_n = i \mid \tau_A^-(n) > \tau_B^-(n), \tau_B^+(n) < \tau_A^+(n)\right) \qquad (2.15)$$

giving the distribution of states in which inner reactive trajectories spend their time.
    When the chain is stationary, we also write $\hat{\pi}^{AB}$.

The next object tells us about the amount of reactive trajectories that move from state i to j during one time step, i.e., the current of reactive trajectories.

Definition 2.6. *The **current (or flux) of reactive trajectories** $f^{AB}(n) = (f_{ij}^{AB}(n))_{i,j \in \mathbb{X}}$ is defined as*

$$f_{ij}^{AB}(n) := \mathbb{P}\left(X_n = i, X_{n+1} = j, \tau_A^-(n) > \tau_B^-(n), \tau_B^+(n+1) < \tau_A^+(n+1)\right). \qquad (2.16)$$

Theorem 2.7. *For a Markov chain $(X_n)_{n \in \mathbb{Z}}$ with transition matrix $P(n)$ and committors $q^+(n)$, $q^-(n)$, the current of reactive trajectories is given by*

$$f_{ij}^{AB}(n) = q_i^-(n)\, \mu_i(n)\, P_{ij}(n)\, q_j^+(n+1). \qquad (2.17)$$

Proof. The current of reactive trajectories can be computed as

$$
\begin{aligned}
f_{ij}^{AB}(n) &= \mathbb{P}(X_{n+1} = j, \tau_A^-(n) > \tau_B^-(n), \tau_B^+(n+1) < \tau_A^+(n+1) \mid X_n = i)\,\mathbb{P}(X_n = i) \\
&= \mathbb{P}(X_{n+1} = j, \tau_B^+(n+1) < \tau_A^+(n+1) \mid X_n = i)\,\mathbb{P}(\tau_A^-(n) > \tau_B^-(n) \mid X_n = i)\,\mu_i(n) \\
&= \mathbb{P}(\tau_B^+(n+1) < \tau_A^+(n+1) \mid X_{n+1} = j, X_n = i)\,\mathbb{P}(X_{n+1} = j \mid X_n = i)\,q_i^-(n)\,\mu_i(n) \\
&= q_i^-(n)\,\mu_i(n)\,P_{ij}(n)\,q_j^+(n+1),
\end{aligned}
$$

$$(2.18)$$

by first conditioning on $\{X_n = i\}$, then by independence of $\{X_{n+1} = j, \tau_B^+(n+1) < \tau_A^+(n+1)\}$ and $\{\tau_A^-(n) > \tau_B^-(n)\}$ given $\{X_n = i\}$, by conditioning on $\{X_{n+1} = j\}$, and last by the Markov property. $\qquad\qquad\square$

In addition to $Z^{AB}(n)$, we can also consider

$$H^{AB}(n) := \sum_{i,j} f_{ij}^{AB}(n) = \mathbb{P}(\tau_A^-(n) > \tau_B^-(n), \tau_A^+(n+1) > \tau_B^+(n+1)), \qquad (2.19)$$

giving the probability of being on a reactive trajectory at two consecutive time steps. Equivalently, it gives the probability of either being in A at time n and next transitioning to B or already having left A and being on an inner reactive trajectory.

The reactive trajectories only go out of A, not into A, moreover the reactive trajectories only enter B, and do not exit B. Therefore, A can be thought of as a source of reactive trajectories, whereas B acts like their sink. This motivates the next two quantities characterizing the amount of reactive trajectories that exit the source resp. enter the sink set in one time step. Since they quantify the amount that exits or enters per time step, we can understand them as discrete rates. We will find that by summing the current of reactive trajectories out of A we obtain the discrete rate of reactive trajectories that exit the set A, and by summing the current that goes into B, we get the rate of reactive trajectories flowing into B.

Definition 2.8. *The **discrete rate of reactive trajectories leaving** A, or in short **out-rate**, is given by*

$$k^{A\to}(n) := \mathbb{P}(X_n \in A, \tau_B^+(n+1) < \tau_A^+(n+1)), \qquad (2.20)$$

*and gives the probability of observing a reactive trajectory that leaves A at time n. The **discrete rate of reactive trajectories entering** B, or in short **in-rate**, is given by*

$$k^{\to B}(n) := \mathbb{P}(X_n \in B, \tau_A^-(n-1) > \tau_B^-(n-1)), \qquad (2.21)$$

and gives the probability of observing a reactive trajectory that enters B at time n.

Theorem 2.9. *For a Markov chain with current of reactive trajectories $f^{AB}(n)$, we find the discrete rates to be*

$$\begin{aligned} k^{A\to}(n) &= \sum_{\substack{i\in A \\ j\in \mathbb{X}}} f_{ij}^{AB}(n) \\ k^{\to B}(n) &= \sum_{\substack{i\in \mathbb{X} \\ j\in B}} f_{ij}^{AB}(n-1). \end{aligned} \qquad (2.22)$$

Proof. By using the law of total probability we find

$$\sum_{\substack{i\in A \\ j\in \mathbb{X}}} f_{ij}^{AB}(n) = \sum_{\substack{i\in A \\ j\in \mathbb{X}}} \mathbb{P}(X_n = i, X_{n+1} = j, \tau_A^-(n) > \tau_B^-(n), \tau_B^+(n+1) < \tau_A^+(n+1)) = k^{A\to}(n)$$

$$(2.23)$$

and

$$\sum_{\substack{i\in\mathbb{X}\\j\in B}} f_{ij}^{AB}(n-1) = \sum_{\substack{i\in\mathbb{X}\\j\in B}} \mathbb{P}(X_{n-1} = i, X_n = j, \tau_A^-(n-1) > \tau_B^-(n-1), \tau_B^+(n) < \tau_A^+(n)) = k^{\to B}(n).$$

(2.24)

□

2.2 Stationary Markov chains

In this section we consider the first useful case of Transition Path Theory. We study irreducible and stationary Markov chains $(X_n)_{n\in\mathbb{Z}}$ for which the committor probabilities and transition statistics are time-independent. Since the Markov chain is ergodic, we can obtain an interpretation of the transition statistics in terms of time-averages along a single infinitely long trajectory.

In this section we not only add a new perspective to the existing theory of TPT for stationary chains [47, 71] by considering it inside our more general framework from the previous section, but we will additionally extend the theory by some new details and results, e.g., the Proposition 2.15, the quantity (2.43) characterizing the mean duration of reactive trajectories, and the approaches for avoiding other sets during transitions or studying self-transitions in Section 2.2.6.

2.2.1 Stationary setting

We begin by describing the Markov chains of interest, often we simply abbreviate a Markov chain fulfilling the below assumptions as a stationary Markov chain.

Assumption 2.10. *We consider a Markov chain $(X_n)_{n\in\mathbb{Z}}$ taking values in a discrete and finite state space $\mathbb{X}$. A transition from state $i \in \mathbb{X}$ to state $j \in \mathbb{X}$ at the next time step occurs with time-homogeneous probability*

$$P_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) \tag{2.25}$$

stored in the row-stochastic transition matrix $P = (P_{ij})_{i,j\in\mathbb{X}}$. We assume that the process is irreducible and thus has a unique, strictly positive stationary distribution π, and that the chain is in stationarity.

For the setup of TPT we also need the time-reversed process $(X_n^-)_{n\in\mathbb{Z}}$ with $X_n^- := X_{-n}$. The time-reversed chain is also a Markov chain and stationary with respect to the same distribution. The transition matrix $P^- = (P_{ij}^-)_{i,j\in\mathbb{X}}$ of the time-reversed process is given by

$$P_{ij}^- = P_{ji}\frac{\pi_j}{\pi_i}. \tag{2.26}$$

Remark 2.11 (Irreducibility of the backward transition matrix)**.** *Note that due to the irreducibility of P, also P^- is irreducible. To show irreducibility of P^-, we need to show that for*

every i, j there exists $n > 0$ s.t. $((P^-)^n)_{ij} > 0$. We can rewrite the above Eq. (2.26) in matrix notation $P^- = D_\pi^{-1} P^\top D_\pi$, where D_π is the diagonal matrix with π on the diagonal, thus

$$(P^-)^n = D_\pi^{-1}(P^\top)^n D_\pi = D_\pi^{-1}(P^n)^\top D_\pi. \tag{2.27}$$

From P being irreducible and π being positive it follows that P^- is also irreducible.

2.2.2 Committor probabilities

Since the chain is stationary, the committors must be time-independent, i.e., for all n it must hold

$$q^+(n) = q^+(n+1) =: q^+, \tag{2.28}$$

$$q^-(n) = q^-(n+1) =: q^-. \tag{2.29}$$

Consequently, the committor equations from Theorem 2.1 reduce to

$$q_i^+ = \sum_{j \in \mathbb{X}} P_{ij} q_j^+, \tag{2.30}$$

$$q_i^- = \sum_{j \in \mathbb{X}} P_{ij}^- q_j^- \tag{2.31}$$

for $i \in C$ and with corresponding boundary conditions $q_i^+ = \mathbb{1}_B(i)$ and $q_i^- = \mathbb{1}_A(i)$ for $i \in A \cup B$.

The following proposition provides us with the necessary conditions such that the existence and uniqueness of the committors is guaranteed. In particular, due to the irreducibility of P and P^-, the assumptions are full-filled.

Proposition 2.12. *If P and P^- are such that eventually hitting $A \cup B$ from any $i \in C$ is certain, then the stationary committor equations in (2.30) and (2.31) each have a unique solution.*

Proof. The proposition covers a special case of $M-$periodic Markov chains of period $M = 1$, compare with Section 2.3. The proof therefore follows from Proposition 2.22 by setting $M = 1$. $\qquad \square$

Remark 2.13 (Reversible Markov chains). *Whenever the Markov chain is reversible, the analysis simplifies. A reversible Markov chain is indistinguishable from its time-reversal and it holds $P_{ij}\pi_i = P_{ji}\pi_j$. For reversible chains it follows from the committor equations that the forward and backward committor are related by $q_i^+ = 1 - q_i^-$, or in other words, that the probability of hitting B before A is the same as the probability of having last come from B not A.*

In the reversible case, we can additionally give a variational formulation of the forward committor [9, Theorem 7.33] as the unique minimizer $h = (h_i)_{i \in \mathbb{X}} \in [0,1]^{|\mathbb{X}|}$ of

$$\frac{1}{2} \sum_{i,j} \pi_i P_{ij}(h_i - h_j)^2 \tag{2.32}$$

subject to the constraints that $h|_A = 0$, $h|_B = 1$.

2.2.3 Statistics

The transition statistics follow from Section 2.1.2. Since the committors and the law of the chain are time-independent, the transition statistics are also time-independent and we denote them without their time-dependence, e.g., $\mu^{AB}(n) = \pi^{AB}$ and $f^{AB}(n) = f^{AB}$. Due to the stationarity, we can provide several additional results that relate the transition quantities to each other.

The first theorem shows that for a stationary Markov chain the current of reactive trajectories is conserved at states $i \in C$, i.e., the inflow of current into a state i equals the outflow of i. In the context of electrical circuits this property is known as Kirchhoff's law. The source and sink are characterized by only outflow resp. inflow of reactive trajectories, therefore the current conservation does not hold for them. Moreover we will find that for a stationary Markov chain the discrete rate of leaving A equals the discrete rate of entering B, thus we denote by $k^{AB} := k^{A\rightarrow} = k^{\rightarrow B}$ the global transition rate.

Theorem 2.14. *For a stationary Markov chain $(X_n)_{n\in\mathbb{Z}}$ the current of reactive trajectories out of a state $i \in C$ equals the current flowing into the state i, i.e.,*

$$\sum_{j\in\mathbb{X}} f_{ij}^{AB} = \sum_{j\in\mathbb{X}} f_{ji}^{AB}. \tag{2.33}$$

Further, the current of reactive trajectories flowing out of A into $\mathbb{X}$ (equivalently into $C \cup B$) equals the flow of reactive trajectories from $\mathbb{X}$ (equivalently from $C \cup A$) into B

$$\sum_{\substack{i\in A \\ j\in\mathbb{X}}} f_{ij}^{AB} = \sum_{\substack{i\in\mathbb{X} \\ j\in B}} f_{ij}^{AB}. \tag{2.34}$$

Proof. First, for any $i \in C$, we have

$$\sum_{j\in\mathbb{X}} \left(f_{ij}^{AB} - f_{ji}^{AB} \right) = q_i^- \pi_i \sum_{j\in\mathbb{X}} P_{ij}q_j^+ - q_i^+ \pi_i \sum_{j\in\mathbb{X}} q_j^- P_{ij}^- = q_i^- \pi_i q_i^+ - q_i^+ \pi_i q_i^- = 0 \tag{2.35}$$

using the definition of the time-reversed transition probabilities and the committor equations for $i \in C$.

Second, using that $f_{ij}^{AB} = 0$ if $i \in B$, $j \in \mathbb{X}$ and also if $i \in \mathbb{X}$, $j \in A$, we can compute

$$\sum_{\substack{i\in\mathbb{X} \\ j\in\mathbb{X}}} f_{ij}^{AB} = \sum_{\substack{i\in A \\ j\in\mathbb{X}}} f_{ij}^{AB} + \sum_{\substack{i\in C \\ j\in\mathbb{X}}} f_{ij}^{AB} + \underbrace{\sum_{\substack{i\in B \\ j\in\mathbb{X}}} f_{ij}^{AB}}_{=0} = \underbrace{\sum_{\substack{i\in\mathbb{X} \\ j\in A}} f_{ij}^{AB}}_{=0} + \sum_{\substack{i\in\mathbb{X} \\ j\in C}} f_{ij}^{AB} + \sum_{\substack{i\in\mathbb{X} \\ j\in B}} f_{ij}^{AB}. \tag{2.36}$$

By summing Eq. (2.35) over $i \in C$ and exchanging the indices i with j, we arrive at

$\sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ij}^{AB} = \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ji}^{AB} = \sum_{\substack{j \in C \\ i \in \mathbb{X}}} f_{ij}^{AB}$. This in turn allows us to rewrite Eq. (2.36) as

$$\sum_{\substack{i \in A \\ j \in \mathbb{X}}} f_{ij}^{AB} + \sum_{\substack{i \in \mathbb{X} \\ j \in C}} f_{ij}^{AB} = \sum_{\substack{i \in \mathbb{X} \\ j \in C}} f_{ij}^{AB} + \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f_{ij}^{AB} \tag{2.37}$$

implying that $\sum_{\substack{i \in A \\ j \in \mathbb{X}}} f_{ij}^{AB} = \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f_{ij}^{AB}$. $\qquad\qquad\square$

Moreover we can find a relation between the probabilities Z^{AB} and H^{AB}. The next proposition tells us that Z^{AB}, the probability to be on an inner reactive trajectory, and H^{AB}, the probability to be for two consecutive time steps on a reactive trajectory, differ exactly by k^{AB}.

Proposition 2.15. *For a stationary Markov chain as defined in Assumptions 2.10, the relation*

$$H^{AB} = Z^{AB} + k^{AB} \tag{2.38}$$

holds.

Proof. We know that whenever the event of being on an inner reactive trajectory $\{\tau_A^-(n) > \tau_B^-(n), \tau_A^+(n) > \tau_B^+(n)\} =: E_Z$ takes place, the definition of the hitting times, compare with Eq. (2.5) and (2.6), restricts the hitting times to be strictly smaller resp. greater than the current time n, i.e., $n > \tau_A^-(n) > \tau_B^-(n)$ and $\tau_A^+(n) > \tau_B^+(n) > n$ and thus it follows that also $\tau_A^+(n+1) > \tau_B^+(n+1)$ holds. Consequently

$$E_Z \subseteq \{\tau_A^-(n) > \tau_B^-(n), \tau_A^+(n+1) > \tau_B^+(n+1)\} =: E_H$$

and hence $Z^{AB} \leq H^{AB}$. Let the event E_D denote the difference between the events E_H and E_Z,

$$E_D := \{\tau_A^-(n) > \tau_B^-(n), \tau_A^+(n+1) > \tau_B^+(n+1), \tau_A^+(n) < \tau_B^+(n)\}$$

which can be rewritten as $E_D = \{\tau_A^-(n) = n, \tau_A^+(n+1) > \tau_B^+(n+1)\}$, thus

$$H^{AB} = \mathbb{P}(E_Z \cup E_D) = Z^{AB} + k^{AB}.$$

$\qquad\qquad\square$

Last, for stationary processes we often consider the following simplified current:

Definition 2.16. *The **effective current of reactive trajectories** $f^+ = (f_{ij}^+)_{i,j \in \mathbb{X}}$ gives the net amount of the current of reactive trajectories from i to j,*

$$f_{ij}^+ := \max\{f_{ij}^{AB} - f_{ji}^{AB}, 0\}. \tag{2.39}$$

When the dynamics are additionally reversible, the effective current simplifies to

$$f_{ij}^+ = \begin{cases} \pi_i P_{ij} \left(q_j^+ - q_i^+ \right), & \text{if } q_j^+ > q_i^+ \\ 0, & \text{else} \end{cases} \tag{2.40}$$

and thus only flows along edges where the current increases. Due to this form, the effective current of reversible dynamics is free of cycles, i.e., there cannot be a cyclic path with positive effective current between all successively visited states. If there would exist such a cyclic path, then the committor would have to strictly increase along the cycle for f^+ to be positive, but this is impossible. For non-reversible processes, the effective current is however not guaranteed to be cycle-free and a different approach can be taken to extract a cycle-free current from the current of reactive trajectories. We will come to this in Section 4.2.

2.2.4 Interpretation of the statistics as time-averages

The transition statistics give us dynamical information about the ensemble of reactive trajectories. Due to the Markov chains' ergodicity, the Markov chain will visit all states infinitely many times. Therefore by Birkhoff's ergodic theorem the ensemble space average of a quantity equals the time average of this quantity along a single infinitely long trajectory. Building on this, the transition statistics characterizing the ensemble of reactive trajectories can also be found by considering the reactive pieces along a single infinitely long trajectory and by averaging over them.

> **Theorem 2.17.** *For a Markov chain $(X_n)_{n \in \mathbb{Z}}$ satisfying Assumption 2.10, we have the following $\mathbb{P}-$almost sure convergence results:*
>
> $$\pi_i^{AB} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{1}_{\{i\}}(X_n) \, \mathbb{1}_A \left(X_{\tau_{A \cup B}^-(n)} \right) \, \mathbb{1}_B \left(X_{\tau_{A \cup B}^+(n)} \right)$$
>
> $$f_{ij}^{AB} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{1}_{\{i\}}(X_n) \, \mathbb{1}_A \left(X_{\tau_{A \cup B}^-(n)} \right) \, \mathbb{1}_{\{j\}}(X_{n+1}) \, \mathbb{1}_B \left(X_{\tau_{A \cup B}^+(n+1)} \right)$$
>
> $$k^{AB} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{1}_A(X_n) \, \mathbb{1}_B \left(X_{\tau_{A \cup B}^+(n+1)} \right) \tag{2.41}$$
>
> $$= \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{1}_A \left(X_{\tau_{A \cup B}^-(n-1)} \right) \, \mathbb{1}_B(X_n)$$
>
> *where $i, j \in \mathbb{X}$ and $\mathbb{1}_S(x)$ is the indicator function on the set S.*

The proof of Theorem 2.17 can be found in [56, Thm 3.3.2, Thm 3.3.7, Thm 3.3.11]. Since in the above theorem the functions that are evaluated at each time point n along the realization not only depends on the process at time n but also on the process at

hitting times, one cannot simply apply the ergodic theorem for Markov chains as in Theorem 1.16. Instead one has to directly rely on Birkhoff's ergodic theorem [72] for the canonical representation of the process as a Markov shift.

Theorem 2.17 not only offers an approach to approximate the transition statistics by averaging along a sufficiently long, stationary trajectory but also gives interpretability of the statistics. While π^{AB} as the relative frequency of visits of reactive trajectories to a certain state and f^{AB} as the relative frequency of consecutive visits of reactive trajectories to two states are still straightforward to understand, we can also give meaning to the rate k^{AB} and to H^{AB}. By Theorem 2.17, we can understand k^{AB} as the total number of transitions that are started in A within the time interval $\{-N, \dots, N\}$ divided by the size of the interval $2N + 1$ in the limit of $N \to \infty$. Or equivalently, as the total number of transitions that are ended in B during $\{-N, \dots, N\}$ divided by the size of the interval $2N + 1$ in the limit of $N \to \infty$. This clarifies the interpretation of k^{AB} as a discrete rate. Similarly we can give meaning to

$$H^{AB} = \sum_{i,j \in \mathbb{X}} f_{ij}^{AB} = \lim_{N \to \infty} \frac{1}{2N + 1} \sum_{n=-N}^{N} \mathbb{1}_A \left(X_{\tau_{A\cup B}^-(n)} \right) \mathbb{1}_B \left(X_{\tau_{A\cup B}^+(n+1)} \right) \tag{2.42}$$

as the fraction of time of either being in A and in the next time step on an inner reactive trajectory or being on an inner reactive trajectory in the limit of an infinitely large time interval. Equivalently, we can understand H^{AB} as the fraction of time of being on a reactive trajectory during two consecutive time step.

Due to the above interpretations, we note that the ratio between H^{AB} and k^{AB} provides us with a further transition characteristic for stationary Markov chains,

$$t^{AB} := \frac{H^{AB}}{k^{AB}}, \tag{2.43}$$

telling us the *expected duration of a reactive trajectory*. To clarify, the duration of a single reactive trajectory $(X_n, X_{n+1}, \dots, X_{n+N})$ is not the number of time points while being on a reactive trajectory, i.e., $N + 1$ here, but has to be understood as the number of passages when reactive, i.e., N here. Also note that due to the relation $H^{AB} = Z^{AB} + k^{AB}$ from Proposition 2.15, $\frac{Z^{AB}}{k^{AB}} = t^{AB} - 1$ gives the expected number of time steps while being on an inner reactive trajectory.

There is a different way to get the same result. We can find that

$$\frac{H^{AB}}{k^{AB}} = \left(\mathbb{P}(X_n \in A \mid \tau_A^-(n) > \tau_B^-(n), \tau_A^+(n+1) > \tau_B^+(n+1)) \right)^{-1}. \tag{2.44}$$

For an irreducible Markov chain on a finite state space, the inverse of the stationary distribution in a state i equals the expected return time to i when starting in that same state, see Theorem 1.11. Here, we have the inverse of the stationary distribution in A conditional on being reactive. Therefore it gives the expected return time to A when only considering the reactive parts, and thus the expected duration of a reactive

trajectory. Note that the process conditional on being on a reactive trajectory is still irreducible on all states that are ever visited by a reactive trajectory.

2.2.5 Example: Transitions in a triplewell potential

We can give a first example in which we exemplarily study the transition behaviour of a particle diffusing in a triplewell potential landscape. In particular, we consider the diffusive motion of a particle at position $X_t \in \mathbb{R}^2$ at time t according to the overdamped Langevin equation (compare with Section 1.2.2)

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{2.45}$$

where $\sigma > 0$ is the noise strength, W_t is a 2$-$dimensional standard Brownian motion and $V : \mathbb{R}^2 \to \mathbb{R}$ is the triple well potential from [47, 62],

$$
\begin{aligned}
V(x,y) = & \tfrac{3}{4}\exp\left(-x^2 - \left(y - \tfrac{1}{3}\right)^2\right) - \tfrac{3}{4}\exp\left(-x^2 - \left(y - \tfrac{5}{3}\right)^2\right) \\
& - \tfrac{5}{4}\exp\left(-(x-1)^2 - y^2\right) - \tfrac{5}{4}\exp\left(-(x+1)^2 - y^2\right) + \tfrac{1}{20}x^4 + \tfrac{1}{20}\left(y - \tfrac{1}{3}\right)^4.
\end{aligned}
\tag{2.46}
$$

This potential landscape (Fig. 2.4) has three minima, the two deeper wells are approximately centered at $(\pm 1, 0)$ and the shallow well at $(0, 1.5)$. The minima are separated by saddle points approximately at $(0, -0.25)$ and $(\pm 0.6, 1.1)$. At approximately $(0, 0.5)$ there is a local maximum.
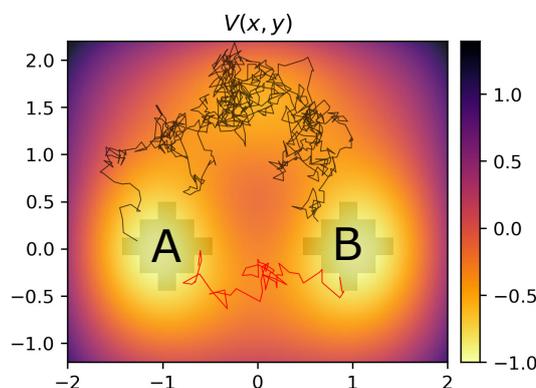
Figure 2.4: Triple well potential landscape and two possible reactive trajectories from set A (around the left deep well) to set B (around the right deep well).

We are now interested in the transition behaviour between the deep wells of V at $(\pm 1, 0)$ when the dynamics is stationary. In Fig. 2.4, we show the potential landscape and two possible reactive trajectories from one of the deep wells (marked by set A) to the other deep well (marked by set B). With TPT we now hope to answer the following question: Which pathway is more likely, the pathway via the two lower barriers and the third metastable shallow well at $(0, 1.5)$, or the direct crossing over the higher barrier

between the two deep wells?

Before applying TPT on this example, we have to discretize the process into a Markov chain. For estimating the transition matrix as in Eq. (1.23), we sampled $10,000$ short trajectories of time step size $\Delta t = 0.3$ and with noise strength $\sigma = 1$ and discretized the interval $[-2, 2] \times [-1, 2]$ into regular square grid cells of size 0.2×0.2. Then we choose sets A and B as centered at $(-1, 0)$ and $(1, 0)$ respectively.
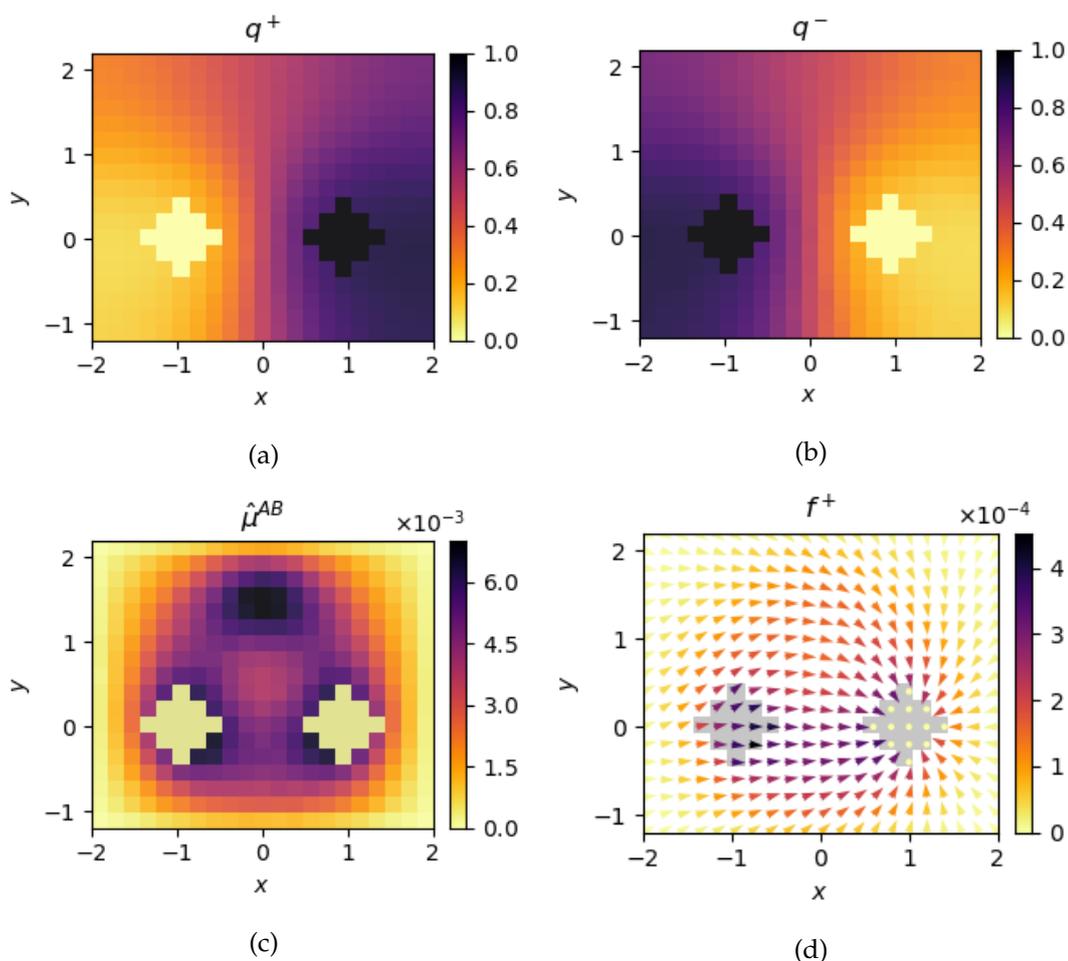
Figure 2.5: (a) Forward committor, (b) Backward committor, (c) normalized distribution of reactive trajectories, (d) the accumulated f^+ out of each state.

The computed committor functions and statistics of the reactive trajectories are shown in Fig. 2.5. Since the dynamics is reversible, the backward committor is just $1 - q^+$. Additionally, due to the symmetry of the potential and the sets A and B about the line $\{x = 0\}$, the probabilities of reaching A and B forward or backward in time are balanced for $x = 0$. The computed committor functions confirm this behaviour and take values of approximately 0.5 on the grid cells around $x = 0$. Also, we can note that the committors are close to constant inside the metastable sets given by the wells due to the fast mixing inside the wells, but vary across the barriers. The normalized distribution of reactive trajectories is spread around the two pathways: the pathway via the upper

shallow well and the pathway via the direct barrier between A and B. The distribution is highest in the shallow well, indicating that reactive trajectories spend a lot of time there if they reach this set. The computed effective current[5] is much higher along the pathway over the direct barrier between A and B. It therefore seems that most reactive trajectories transition from A to B directly via the higher barrier instead of visiting the upper well and crossing the two lower barriers. The rate of transitions is $k^{AB} = 0.0142$. Thus a reactive trajectory is started on average every $(k^{AB})^{-1} = 70$th time step and with a mean transition time of $t^{AB} = 11.04$.

2.2.6 Avoiding sets, self-transitions and transitions between several sets

Our theory so far always tells us about transitions from one set, A, to another different set, B. In this section we will demonstrate that we can use the same tools to consider variations of that basic setting. For example we can answer the following questions: What are the pathways of trajectories that last came from A but without hitting B return back to A? Or, what are the pathways of trajectories that transition from A avoiding the set D to set B? Or, what are the direct transition pathways between a larger number of disjoint subsets?

(a)   (b)

Figure 2.6: (a) Sketch showing self-returning trajectories, (b) trajectories from A to B that avoid the set D.

We will call the trajectories that last came from A (not B) and next go to A (not B), *self-returning trajectories* (Fig. 2.6(a)). These self-returning trajectories can also be considered as being resilient since they come back to their original purpose. They can stay in A or leave A but only to later come back to A. For these self-returning trajectories it holds that the process last came from A, i.e., $\tau_A^+(n) < \tau_B^+(n)$, and next visits A, i.e., $\tau_A^-(n) > \tau_B^-(n)$, so by computing the forward committor to set A rather than B

$$q_i^+(A, B) := \mathbb{P}(\tau_A^+(n) < \tau_B^+(n) \mid X_n = i) \tag{2.47}$$

---

[5]Regarding plotting f^+ in Fig. 2.5(d): If the underlying process is a diffusion process in $\mathbb{R}^d$, we can estimate for each grid cell i (corresponding to a state of the Markov chain) the vector of the average effective current out of that grid cell, i.e., to each i we can attach the vector $\sum_{j \neq i} f_{ij}^+ v_{ij}$, where v_{ij} is the unit vector pointing from the center of the grid cell i to the center of the grid cell j.

and the backward committor of coming from set A rather than B

$$q_i^-(A,B) := \mathbb{P}(\tau_A^-(n) > \tau_B^-(n) \mid X_n = i), \tag{2.48}$$

we can characterize them. We we can multiply the committors with the stationary distribution analogously as in Theorem 2.4 to get the distribution of self-returning trajectories

$$\mathbb{P}(X_n = i, \tau_A^-(n) > \tau_B^-(n), \tau_A^+(n) < \tau_B^+(n)) = q_i^-(A,B)\,\pi_i\,q_i^+(A,B). \tag{2.49}$$

Analogously, we can compute the current of the self-returning trajectories.

When we are on the other hand interested in the transitions from set A to set B that avoid the third set D (Fig. 2.6(b)), which is non-empty and not intersecting A or B, then the transitions are such that they last came from A and not from B or D, i.e., it holds $\tau_A^-(n) > \tau_{B\cup D}^-(n)$, and next go to B not $A \cup D$, i.e., $\tau_B^+(n) < \tau_{A\cup D}^+(n)$. The respective committors $q_i^-(A, B \cup D)$ and $q_i^+(B, A \cup D)$, provide us with the important information about these transitions such as their distribution (by multiplying the committors with the stationary distribution) and their current. In a similar way we can study the transitions between several disjoint, non-empty sets $S_k, k = 1, \ldots, K$. Let us assume that we want to better understand the possible direct transitions from each S_l to each S_j that avoid all other sets $S_k, k \neq l, j$. This can for example be helpful to coarse-grain the dynamics between several sets into a transition network (as e.g. later in Fig. 5.9(b)). By computing the backward committors $q_i^-(S_l, \cup_{k\neq l}S_k)$ and forward committors $q_i^+(S_j, \cup_{k\neq j}S_k)$ for each l, j, the distribution of direct transitions between S_l and S_j is given by $q_i^-(S_l, \cup_{k\neq l}S_k)\,\pi_i\,q_i^+(S_j, \cup_{k\neq j}S_k)$.

In this section we have seen that by slight changes of the sets with respect to which we compute the committors, we can study the transitions between all kinds of sets. We will need these results in the next example where we study debris paths and later for studying canards in Chapter 5.2.

2.2.7 Real-world example: Studying debris pathways in the ocean

We will give a first real-world application of TPT to the dynamics of debris particles on the ocean surface. But TPT can be useful in diverse field, e.g., for studying rare molecular folding events [49], tipping paths in climate models [40, 23, 41, 24] and later in this thesis, we will apply TPT to study social tipping.

In this application we are interested in analysing the possible routes of floating plastic and debris particles on the surface of the ocean. In particular, we will build a Markov chain model of their motion fitted to data and use it to study the pollution paths from their insertion at the coastlines to garbage patches in the ocean, i.e., to the distinct areas that accumulate plastic and debris over time. The results and figures where originally published in [48] and we herein only give a short summary of the main approach and results. The data processing, numerical studies and figures where

37

done by Philippe Miron, while the author helped with the modeling and designed the TPT analysis.

Markov chain model.   We want to model the motion of a single floating debris particle on the ocean surface by a stationary Markov chain $(X_n)_{n \in \mathbb{Z}}$ fitted to data. The motion of floating debris on the ocean surface can be estimated from trajectory data of undrogued drifter buoys from the last 30 years [43, 42] that follow the motion of wind and ocean currents in a similar way as debris particles. We can construct a transition matrix on a box covering of the ocean surface by counting the number of transitions from one box to another in the trajectory data and row-normalizing this count matrix, compare with Section 1.2.1. In this application, the resulting matrix has several absorbing states and communication classes and is not irreducible. Since we are only interested in the main global dynamics, we restrict the transition matrix to its largest communication class resulting in a substochastic but irreducible transition matrix. We denote the state space of this transition matrix by O. By adding an artificial state ω to the Markov chain, we can close the transition matrix on the state space $\mathbb{X} = O \cup \omega$ by letting all the outflow of O go to ω. To achieve a physically more realistic model, we strengthen the beaching of debris particles at the coasts, i.e., at boxes $L \subset O$ that are partially filled by land or ice. Specifically we let an additional portion of the mass from $i \in L$ flow into ω in proportion to the size of land or ice in this cell.

Since it is unrealistic that all the debris particles ultimately end in ω, we re-inject them back into O, thereby making the transition matrix irreducible again. In particular, we insert it into one of the land boxes $i \in L$ with probability $\frac{W_i}{\sum_{j \in L} W_j}$ where W_i gives the estimated amount of mismanaged waste in the neighbourhood of box i from 2010 [31], see Fig. 2.7(a). The resulting transition matrix on $\mathbb{X} = O \cup \omega$ has the following form

$$P = \begin{pmatrix} P^O & a \\ r & 0 \end{pmatrix} \tag{2.50}$$

where P^O describes the transition probabilities between boxes on the ocean surface O, while

$$a = \left(1 - \sum_{j \in O} P^O_{ij} \right)_{i \in O} \tag{2.51}$$

gives the probabilities of outflow from O to ω, and $r = (r_i)_{i \in O}$ gives the probability of inflow from ω into O with

$$r_i = \begin{cases} \frac{W_i}{\sum_{j \in L} W_j} & \text{for } i \in L \\ 0 & \text{for } i \in O \setminus L. \end{cases} \tag{2.52}$$

We consider the Markov chain in stationarity. Its stationary distribution $\pi = (\pi^O, \pi^\omega)$

(a)

(b)
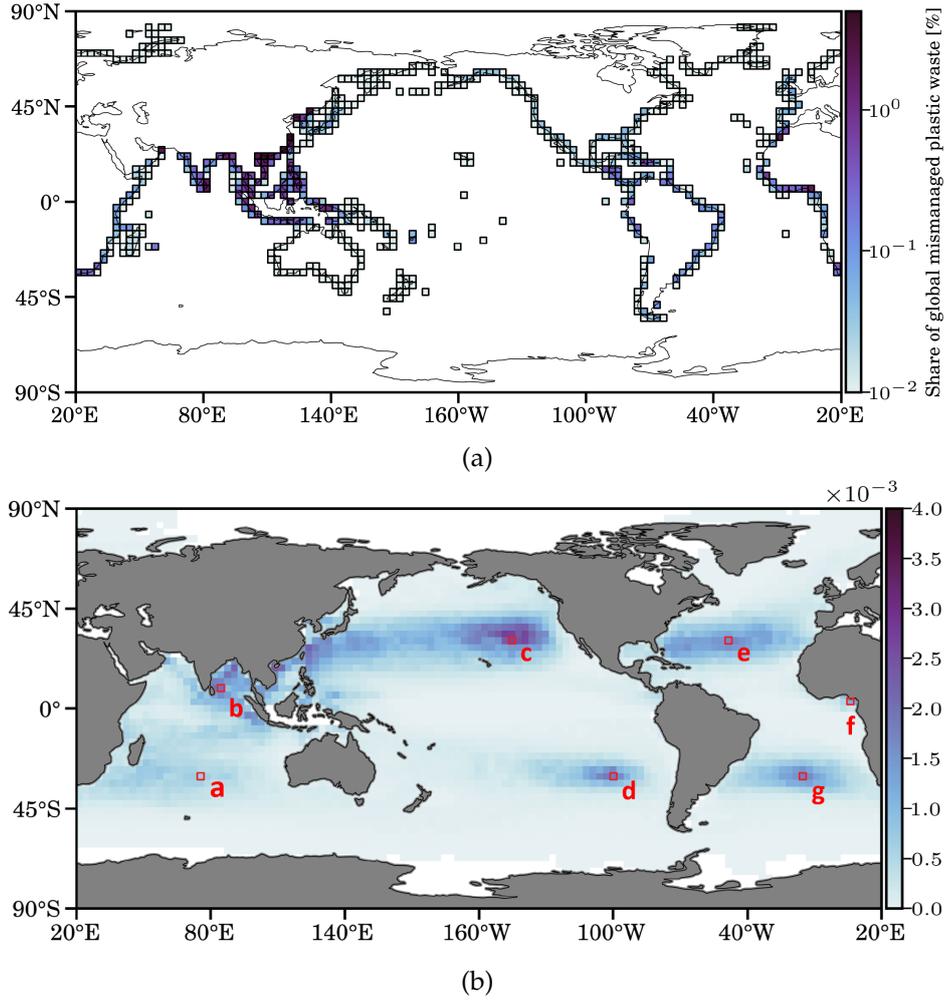
Figure 2.7: (a) Estimated share of mismanaged waste in the neighbourhood of each box from 2010 [31] giving the insertion distribution r. (b) The stationary distribution π^O on boxes covering the ocean O. The boxes where the distribution peaks locally are assumed to be garbage patches and highlighted by the red boxes. We denote the garbage patches as follows, a: the Indian Ocean patch, b: Bay of Bengal Patch, c: the North Pacific patch (also known as the Great Pacific Garbage Patch), d: South Pacific patch, e: North Atlantic patch, f: Gulf of Guinea Patch, g: South Atlantic patch.

on the combined domain $\mathbb{X} = O \cup \omega$ fulfills

$$\begin{pmatrix} \pi^O & \pi^\omega \end{pmatrix} \begin{pmatrix} P^O & a \\ r & 0 \end{pmatrix} = \begin{pmatrix} \pi^O & \pi^\omega \end{pmatrix}, \tag{2.53}$$

consequently, the stationary distribution on O satisfies $\pi^O P^O + \pi^\omega r = \pi^O$, which can be rearranged to give

$$\pi^O = \pi^\omega r (I - P^O)^{-1} \tag{2.54}$$

where the inverse of $(I - P^O)$ exists since the matrix is weakly chained diagonally dominant (Prop. A.2). The stationary distribution is shown in Fig. 2.7(b). The boxes

where the distribution is peaked are assumed to show the locations of garbage patches that in the long-run accumulate debris particles. The patch in the North Pacific ocean appears as the strongest while the patch in the Indian ocean is the weakest.

We will now show that the stationary distribution on the ocean surface allows a sensible physical interpretation. If at each time point we insert a debris particle into the ocean according to the distribution r, then in the long-time limit the accumulated mass distribution on the ocean boxes is given by

$$\sum_{k=0}^{\infty} r\,(P^O)^k = r\,(I - P^O)^{-1}.$$ (2.55)

This distribution is not a probability distribution, instead it carries the amount of debris particles that are on average on the ocean surface if one debris particle enters per time step with distribution r. Moreover, this mass distribution is proportional to π^O, thus the stationary distribution of our Markov chain model of one debris particle reflects the long-time limit distribution of debris particles that are injected at a constant pace according to the distribution r.

The resulting Markov chain model describes the infinite cycle of one debris particle from its insertion at the coast, its passage to garbage patches and eventually its beaching at a coast, thereafter it starts another journey at some coast. The model is time-homogeneous and therefore does not account for seasonal and climatic effects as well as changes in the amount of ocean pollution.

Remark 2.18. *In Chapter 3 we will discuss how systems with outflow into an absorbing state can be modeled and studied with Transition Path Theory prior to the absorption. But the setting here is rather different, even though we have outflow to ω, it is unrealistic to assume that in the long-time limit debris beaches and no debris remains in the ocean. Therefore we chose to recirculate it back into the domain O. In the TPT computations we have to ensure that the reactive trajectories avoid the state ω since the closure introduces artificial mixing in ω that would otherwise nonphysically change the results by allowing transitions via ω.*

Transition paths of debris.   With the Markov chain model of debris particles, we can start to uncover the pollution paths of debris particles with the help of TPT. In particular we are interested in the transitions from their insertion, i.e., from source set $A = \omega$, to the different garbage patches in the ocean. We therefore choose the union of boxes covering garbage patches (compare with Fig. 2.7(b)) as the sink set B. Since the artificially added ω state is the starting state of transitions, the considered transitions cannot again pass ω once they started there and the well-mixing property of the state ω poses no problem.

The results of computing the reactive currents with TPT are shown in Fig. 2.8. The North Pacific Garbage Patch noticeably has the strongest inflow of debris according to our model. The debris takes on average 2.6 years from its insertion to a garbage
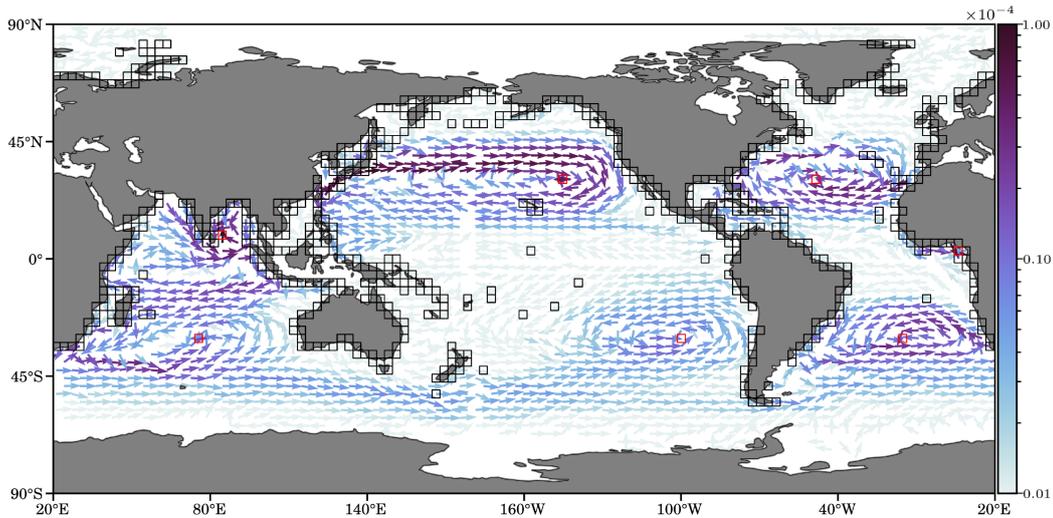
Figure 2.8: Reactive current from source set $A = \omega$ to B, the union of garbage patches (highlighted by red boxes). Coastal boxes L that are filled by land or ice and through which debris enters the ocean are highlighted as black boxes.

patch. By doing separate TPT computations towards each garbage patch while avoiding the other patches, compare with Section 2.2.6 for avoiding sets, we can compute the expected duration to each garbage patch individually. This shows that there is a big difference in the duration of debris routes towards the near-coastal garbage patches (an expected duration of 0.2 resp. 0.6 years towards the patch in the Gulf of Guinea resp. Bay of Bengal) and towards garbage patches in the center of the ocean (a duration of 7.3, 8.6, 4.3, 4.0, and 4.2 years to the North Pacific, South Pacific, North Atlantic, South Atlantic, and Indian Ocean patches).
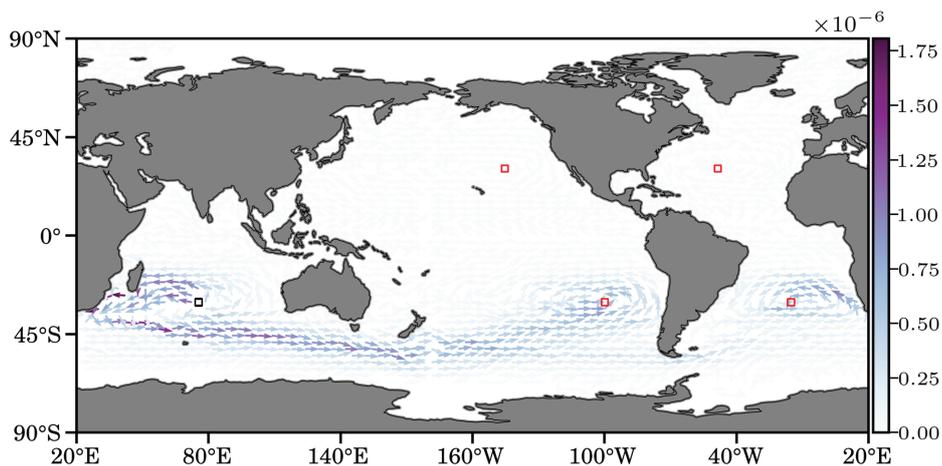
Figure 2.9: Outflow of debris particles (shown by the reactive current) from the garbage patch in the Indian Ocean to the remainder of garbage patches while avoiding ω.

We are also interested in the exchange of debris between garbage patches. To understand this better, we can study the transitions from one garbage patch to the remainder of garbage patches while avoiding the state ω. The garbage patch in the

Indian Ocean does not accumulate as much debris as the other patches (compare with the stationary distribution). Therefore we exemplarily study the emission of debris from the patch in the Indian Ocean to the other patches (Fig. 2.9). We can note that most of the outflow of debris flows to the patch in the South Atlantic and the South Pacific ocean.

2.3   Periodic dynamics

Many real-world systems showcase periodicity, for example ecological or climate systems subject to seasonal driving. Before considering the transitions in these systems, we begin by describing Markov chains with periodically-varying transition probabilities and equilibrated distributions, i.e., the chain cycles through the same distributions each period. If the period is only one time step long, this case reduces to the case of stationary Markov chains from the previous section. We will then come to the committor equations with periodic boundary conditions and the transition statistics that consequently also vary periodically.

2.3.1   Setting

We consider a Markov chain $(X_n)_{n \in \mathbb{Z}}$ on a finite and discrete state space $\mathbb{X} = \{1, \ldots, S\}$ with transition probabilities $P_{ij}(n)$ that are periodically varying in time with period length M, i.e., the transition matrices fulfill

$$P(n) = P(n + M) \; \forall n \in \mathbb{Z}. \tag{2.56}$$

As a consequence, the transition matrices within one period $\mathbb{M} := \{0, \ldots, M-1\}$ are sufficient to describe all the dynamics and we represent them by $P_m := P(m)$ for $m \in \mathbb{M}$.

   We are interested in describing Markov chains equilibrated to the periodically varying transition matrices and therefore ask: What are the conditions under which the distribution of the chain is periodic too, i.e., such that $\mu(n) = \mu(n + M)$ holds for all n? When we want the Markov chain to have a periodically varying distribution, we will need that for each m the more coarsely resolved chain $(X_{m+nM})_{n \in \mathbb{Z}}$ has a stationary distribution. The transition matrix that maps exactly M time points forward in time when currently in a time congruent to $m \in \mathbb{M}$ modulo M is given by the following product of M matrices

$$\bar{P}_m := P_m P_{m+1} \cdots P_{m-1} \tag{2.57}$$

and exactly describes the transition probabilities of the chain $(X_{m+nM})_{n \in \mathbb{Z}}$.

   In the following proposition we will use these ideas further to find the conditions such that a periodically-varying distribution exists.

Proposition 2.19. *If $\bar{P}_0$ is irreducible and assuming the setting as described above, then for all $m \in \mathbb{M}$ there exists a unique stationary distribution π_m of the transition matrix $\bar{P}_m$.*

The distribution is such that it fulfills $\pi_m^\top = \pi_{m-1}^\top P_{m-1}$.[6] Further, for $m = 0$ the stationary distribution is positive, i.e, $\pi_{0,i} > 0$ for all $i \in \mathbb{X}$.

Proof. Since $\bar{P}_0$ is irreducible and the state space is finite, the Markov chain induced by $\bar{P}_0$ has a unique and positive stationary distribution $\pi_0 = (\pi_{0,i})_{i\in\mathbb{X}}$ such that $\pi_0^\top \bar{P}_0 = \pi_0^\top$. It follows that also $\bar{P}_1$ has a stationary distribution given by $\pi_1^\top := \pi_0^\top P_0$, since by the invariance of π_0 under $\bar{P}_0$ it holds

$$\pi_1^\top \bar{P}_1 = \pi_0^\top P_0 \bar{P}_1 = \pi_0^\top P_0 \cdots P_{M-1} P_0 = \pi_0^\top \bar{P}_0 P_0 = \pi_0^\top P_0 = \pi_1^\top. \tag{2.58}$$

Analogously, each $\bar{P}_m$ for $m = 2, \ldots, M-1$, has a stationary distribution $\pi_m^\top := \pi_{m-1}^\top P_{m-1}$ fulfilling $\pi_m^\top \bar{P}_m = \pi_m^\top$.

To show uniqueness of the family $(\pi_m)_{m=0,\ldots,M-1}$, we employ a proof by contradiction. Assume there exists a $v \in \mathbb{R}^S$ with $v \neq \pi_m$ such that $v^\top \bar{P}_m = v^\top$ (for $m \neq 0$). Since v is an invariant vector of $\bar{P}_m$, it follows

$$v^\top P_m \cdots P_{M-1} = v^\top \bar{P}_m P_m \cdots P_{M-1} = v^\top P_m \cdots P_{M-1} \bar{P}_0, \tag{2.59}$$

i.e., that $v^\top P_m \cdots P_{M-1}$ is an invariant vector of $\bar{P}_0$. But the unique invariant vector of $\bar{P}_0$ is given by π_0, so it must be that $v^\top P_m \cdots P_{M-1} = \pi_0^\top$. This in turn implies that

$$v^\top = v^\top P_m \cdots P_{M-1} P_0 \cdots P_{m-1} = \pi_0^\top P_0 \cdots P_{m-1}, \tag{2.60}$$

thereby contradicting our assumption that v is not π_m. □

For the above proposition we need that there is one time step that fulfills the irreducibility assumption of the transition matrix mapping M time steps forward. Without loss of generality we assume that this happens at the time step 0.

Just because $\bar{P}_0$ is irreducible does not mean that all other $\bar{P}_m$ are so too. It is an easy task to construct a counter example, e.g., by forcing all the trajectories to move to the same state at a certain time step. But a weaker statement can be made, namely, it is possible for the Markov chain $(X_n)_{n\in\mathbb{Z}}$ to reach any state from any other state in a finite number of time steps. Irreducibility of $\bar{P}_0$ means that whenever the chain is in some state, say i, at a time point congruent to 0 modulo M, then for any other state j, a $p \in \mathbb{N}_1$ exists such that the process can reach that other state at pM time steps in the future. And since from any state k at any time point during the period, some state i is reached at a time congruent to 0 modulo M, and from there is is possible to reach any j, it is possible from every state k to reach every state j in a finite number of steps.

We will call a Markov chain *M−stationary* when it admits the unique periodic family of distributions from Proposition 2.19. The distribution is then given by $\mu(n) = \pi_m$ whenever time n is congruent to m modulo M. Equivalently we can phrase *M−stationarity* as the property that the joint probability $\mathbb{P}(X_n = i_n, .., X_{n+k} = i_{n+k})$ does not change

---

[6]Sometimes such a family $(\pi_m)_{m=0,\ldots,M-1}$ of invariant densities is called equivariant.

when shifting time by the period length M, i.e., that

$$\mathbb{P}(X_n = i_n, .., X_{n+k} = i_{n+k}) = \mathbb{P}(X_{n+M} = i_n, .., X_{n+k+M} = i_{n+k}) \qquad (2.61)$$

for all n, k.

Having the long-time behavior of chains in mind in this section, we will assume:

Assumption 2.20. *We assume that $\bar{P}_0$ is irreducible and that the chain is M-stationary.*

For computing the backward committors, we will also need the transition matrix of the time-reversed chain $(X_n^-)_{n\in\mathbb{Z}}$ with $X_n^- = X_{-n}$. Due to $M-$stationarity, the transition probabilities of the time-reversed chain are also $M-$periodic, and it is enough to give the transition probabilities backward in time for each time point during the period $m \in \mathbb{M}$

$$\begin{aligned}
P_{m,ij}^- := \mathbb{P}(X_{m-1} = j \mid X_m = i) &= \mathbb{P}(X_{-m+1}^- = j \mid X_{-m}^- = i) \\
&= \mathbb{P}(X_m = i \mid X_{m-1} = j) \frac{\mathbb{P}(X_{m-1} = j)}{\mathbb{P}(X_m = i)} = P_{m-1,ji} \frac{\pi_{m-1,j}}{\pi_{m,i}}
\end{aligned} \qquad (2.62)$$

whenever $\pi_{m,i} > 0$, else for $\pi_{m,i} = 0$ we set $P_{m,ij}^- := 0$.

2.3.2   The committor equations

For $M-$stationary Markov chains the law of the process is indistinguishable from the law pM times later for any $p \in \mathbb{N}_1$. Therefore the committors are identical every M time steps. We denote

$$\begin{aligned}
q_m^+ &:= q^+(n) \\
q_m^- &:= q^-(n)
\end{aligned} \qquad (2.63)$$

whenever n is congruent to $m \in \mathbb{M}$ modulo M.

The committor equations for $M-$periodic dynamics can therefore be found from the general committor equations (2.9) and (2.10) by using the periodicity conditions (2.63). The $M-$periodic forward committor fulfills the following iterative system

$$\begin{cases}
q_{m,i}^+ &= \sum\limits_{j \in \mathbb{X}} P_{m,ij} \, q_{m+1,j}^+ \quad & i \in C \\
q_{m,i}^+ &= 0 & i \in A \\
q_{m,i}^+ &= 1 & i \in B
\end{cases} \qquad (2.64)$$

with the periodicity condition $q_M^+ = q_0^+$, whereas the $M-$periodic backward committor satisfies

$$\begin{cases}
q_{m,i}^- &= \sum\limits_{j \in \mathbb{X}} P_{m,ij}^- \, q_{m-1,j}^- \quad & i \in C \\
q_{m,i}^- &= 0 & i \in B \\
q_{m,i}^- &= 1 & i \in A
\end{cases} \qquad (2.65)$$

subject to $q_M^- = q_0^-$.

The following characterization of the forward and backward committors in terms of the path probabilities over one period $\mathbb{M}$ will be useful for showing the existence of unique solutions to the committor problems.

Proposition 2.21. *For any time n congruent to m modulo M and $i \in C$ the committor functions* (2.63) *satisfy the following equalities*

$$q_{m,i}^+ = \sum_{i_1 \dots i_M \in C} P_{m,ii_1} \cdots P_{m+M-1,i_{M-1}i_M}\, q_{m,i_M}^+ + \sum_{k=1}^{M} \sum_{\substack{i_1 \dots i_{k-1} \in C \\ i_k \in B}} P_{m,ii_1} \cdots P_{m+k-1,i_{k-1}i_k} \quad (2.66)$$

$$q_{m,i}^- = \sum_{i_1 \dots i_M \in C} P_{m,ii_1}^- \cdots P_{m-M+1,i_{M-1}i_M}^-\, q_{m,i_M}^- + \sum_{k=1}^{M} \sum_{\substack{i_1 \dots i_{k-1} \in C \\ i_k \in A}} P_{m,ii_1}^- \cdots P_{m-k+1,i_{k-1}i_k}^- \quad (2.67)$$

Proof. First it follows from (2.64) for $i \in C$ that

$$q_{m,i}^+ = \sum_{i_1 \in C} P_{m,ii_1} q_{m+1,i_1}^+ + \sum_{i_1 \in B} P_{m,ii_1} \quad (2.68)$$

since $q_{m+1,i_1}^+ = 1$ if $i_1 \in B$, $q_{m+1,i_1}^+ = 0$ if $i_1 \in A$. By inserting the committor equations at the following times iteratively and by using that $q_0^+ = q_M^+$, we get (2.66). We can proceed analogously for the backward committor, starting from (2.65) and re-inserting committor equations. $\qquad\square$

This proposition tells us that the forward committor at the period start is made up of the probability of all possible paths that stay in C during the period multiplied with the forward committor M time steps later and the probability of all paths that stay in C for some time but hit B within the period. The time-resolution of the Markov chain during the period is important for the committors since we can resolve hitting events at time points within the period. The committors one would compute for a more coarsely resolved, but stationary chain as described by any $\bar{P}_m$ (time-homogeneous, but mapping one period in time forward) we would not notice the chain hitting A or B during the period.

The following proposition provides us with the conditions such that unique solutions to the forward and backward committor problems exist. In particular, when $\bar{P}_0$ is irreducible and thus also the backward matrix $\bar{P}_0^- := P_0^- \cdot P_{M-1}^- \cdots P_1^-$ is irreducible, the conditions are satisfied.

Proposition 2.22. *We assume that for every state $i \in C$ at a time congruent to 0 modulo M it is possible to reach $A \cup B$ and also possible to have come from $A \cup B$. Then the solutions to* (2.64) *and* (2.65) *exist and are unique.*

Proof. We start with the case of the forward committor problem. We can rewrite (2.66)

2.3. Periodic dynamics

from Proposition 2.21 with $m = 0$ as the matrix equation

$$\left(q_{0,i_0}^+\right)_{i_0 \in C} = \underbrace{P_0|_{C \to C} \cdots P_{M-1}|_{C \to C}}_{=:D} \left(q_{0,i_M}^+\right)_{i_M \in C} + \sum_{k=1}^{M} P_0|_{C \to C} \cdots P_{k-1}|_{C \to B} \, \mathbb{1}, \quad (2.69)$$

where $A|_{I \to J}$ denotes the restriction of the matrix A to entries from $i \in I$ to $j \in J$ and $\mathbb{1}$ denotes a vector of ones of size $|B|$. By rearranging the Eq. (2.69), we arrive at

$$(I - D)\left(q_{0,i_0}^+\right)_{i_0 \in C} = \sum_{k=1}^{M} P_0|_{C \to C} \cdots P_{k-1}|_{C \to B} \, \mathbb{1}. \quad (2.70)$$

To prove that a unique solution $(q_{0,i_0}^+)_{i_0 \in C}$ to Eq. (2.70) exists, we need to show that $W := (I - D)$ is invertible. If q_0^+ uniquely exists, also $q_1^+, \ldots, q_{M-1}^+$ uniquely exist due to the iterative form of the committor problem. In the following we will show that W is a weakly chained diagonally dominant matrix (Definition A.1) and therefore invertible by Proposition A.2.

We know that D is a matrix with non-negative entries and row sum $\sum_j D_{ij} \leq 1$ since it is a product of substochastic matrices. It follows that $\sum_{j \neq i} D_{ij} \leq 1 - D_{ii}$ for all $i \in C$. Hence W is weakly diagonally dominant. It remains to be shown that all rows i in W that are not strictly diagonally dominant, lead via a path $i \to i_1 \cdots \to i_k$ with $W_{ii_1}, \ldots, W_{i_{k-1}i_k} \neq 0$ to a strictly diagonally dominant row i_k. The matrix D contains the following entries:

$$D_{ij} = \mathbb{P}(X_1, \ldots, X_{M-1} \in C, X_M = j \mid X_0 = i) \quad (2.71)$$

for $i, j \in C$. Therefore the not strictly diagonally dominant rows of W correspond to the rows i of D such that

$$\mathbb{P}(X_1, \ldots, X_{M-1}, X_M \in C \mid X_0 = i) = \sum_j D_{ij} = 1, \quad (2.72)$$

while the strictly diagonally dominant rows corresponds to rows i with

$$\mathbb{P}(X_1, \ldots, X_{M-1}, X_M \in C \mid X_0 = i) = \sum_j D_{ij} < 1. \quad (2.73)$$

Let us now assume that there does not exist such a path for the not strictly diagonally dominant rows, then starting from i the process will forever stay in C since there is no path that leads to a state i_k with a possibility to enter $A \cup B$. This is contrary to our assumption, and therefore such a path must always exist. Hence W is weakly chained diagonally dominant and therefore invertible.

For the case of the backward committor, we can proceed analogously by starting from Eq. (2.67) and using the properties of the backward process. □

We have shown that unique solutions to the two committor problems exist as long as it is possible from every state in the transition region at the fixed time 0 of the period to reach $A \cup B$ in forward and backward time. In order to solve the committor equation with periodic boundary conditions, one can solve the Eqs. (2.66) and (2.67) with $m = 0$ which contains only one unknown: the committor at time 0. The committors for the remaining times can then iteratively be computed thereof by using (2.64) and (2.65). An alternative approach is to solve the equations on the time-augmented space using a time-augmented transition matrix that pushes the dynamics deterministically forward in time:

$$P_{\text{Aug}} = \begin{pmatrix} 0 & P_0 & & 0 \\ & \ddots & \ddots & \\ & & 0 & P_{M-2} \\ P_{M-1} & & & 0 \end{pmatrix}. \tag{2.74}$$

The corresponding committor equations are derived in [28].

2.3.3 Transition statistics

We have seen that the forward and backward committor in the case of periodically driven dynamics are also $M-$periodic and can be computed from the iterative Eqs. (2.64), (2.65) with periodic conditions in time. Since committors, densities and transition matrices are $M-$periodic, the statistics computed from them are so too. It follows for the distribution of (inner) reactive trajectories

$$\mu_i^{AB}(n) = \pi_{m,i} q_{m,i}^- q_{m,i}^+ =: \pi_{m,i}^{AB} \tag{2.75}$$

and for the current of reactive trajectories

$$f_{ij}^{AB}(n) = \pi_{m,i} q_{m,i}^- P_{m,ij} q_{m+1,j}^+ =: f_{m,ij}^{AB} \tag{2.76}$$

whenever time n is congruent to m modulo M.

Compared to the case of stationary, infinite-time Markov chains, the discrete rate of reactive trajectories leaving A at time m, $k_m^{A\rightarrow} = \sum_{i \in A, j \in \mathbb{X}} f_{m,ij}^{AB}$, does not anymore equal the discrete rate of reactive trajectories arriving in B at time m, $k_m^{\rightarrow B} = \sum_{i \in \mathbb{X}, j \in B} f_{m-1,ij}^{AB}$.

The next theorem provides us with the conservation laws of the current of reactive trajectories in the case of periodic dynamics and will allow us to find the relation between $k_m^{A\rightarrow}$ and $k_m^{\rightarrow B}$.

Theorem 2.23. *Consider a Markov chain satisfying Assumption 2.20. Then, for each state $i \in C$ and time $m \in \mathbb{M}$ we have the following current conservation law*

$$\sum_{j \in \mathbb{X}} f^{AB}_{m,ij} = \sum_{j \in \mathbb{X}} f^{AB}_{m-1,ji}, \tag{2.77}$$

i.e., all the reactive trajectories that flow out of a state i or stay in i at a time congruent to m, have flown into i or been in i at a time congruent to $m-1$.

Further, over one period the amount of reactive flux leaving A is the same as the amount of flux entering B, i.e.,

$$\sum_{m \in \mathbb{M}} \sum_{\substack{i \in A \\ j \in \mathbb{X}}} f^{AB}_{m,ij} = \sum_{m \in \mathbb{M}} \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f^{AB}_{m,ij}. \tag{2.78}$$

Proof. To show that the flux conservation in state $i \in C$ holds, we compute

$$\sum_{j \in \mathbb{X}} \left(f^{AB}_{m,ij} - f^{AB}_{m-1,ji} \right) \stackrel{(1)}{=} \pi_{m,i} q^-_{m,i} \sum_{j \in \mathbb{X}} P_{m,ij} q^+_{m+1,j} - \pi_{m,i} q^+_{m,i} \sum_{j \in \mathbb{X}} q^-_{m-1,j} P^-_{m,ij} \stackrel{(2)}{=} 0 \quad (2.79)$$

using (1) $P^-_{m,ij}\pi_{m,i} = P_{m-1,ji}\pi_{m-1,j}$ and (2) the backward and forward committor equations for $i \in C$.

Next we want to show that the current of reactive trajectories leaving A during one period equals the current entering B during one period. We calculate

$$
\begin{aligned}
\sum_{m \in \mathbb{M}} \sum_{\substack{i \in \mathbb{X} \\ j \in \mathbb{X}}} f^{AB}_{m,ij} &= \sum_{m \in \mathbb{M}} \left(\sum_{\substack{i \in A \\ j \in \mathbb{X}}} f^{AB}_{m,ij} + \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f^{AB}_{m,ij} + \underbrace{\sum_{\substack{i \in B \\ j \in \mathbb{X}}} f^{AB}_{m,ij}}_{=0} \right) \\
&= \sum_{m \in \mathbb{M}} \left(\underbrace{\sum_{\substack{i \in \mathbb{X} \\ j \in A}} f^{AB}_{m,ij}}_{=0} + \sum_{\substack{i \in \mathbb{X} \\ j \in C}} f^{AB}_{m,ij} + \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f^{AB}_{m,ij} \right)
\end{aligned} \tag{2.80}
$$

using that $f^{AB}_{m,ij} = 0$ if $i \in B$, $j \in \mathbb{X}$ and if $i \in \mathbb{X}$, $j \in A$. And by the current conservation for $i \in C$, $m \in \mathbb{M}$ and by relabeling i, j, m,

$$\sum_{m \in \mathbb{M}} \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f^{AB}_{m,ij} = \sum_{m \in \mathbb{M}} \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f^{AB}_{m-1,ji} = \sum_{m \in \mathbb{M}} \sum_{\substack{j \in C \\ i \in \mathbb{X}}} f^{AB}_{m,ij} \tag{2.81}$$

we arrive at

$$\sum_{m \in \mathbb{M}} \left(\sum_{\substack{i \in A \\ j \in \mathbb{X}}} f^{AB}_{m,ij} + \sum_{\substack{i \in \mathbb{X} \\ j \in C}} f^{AB}_{m,ij} \right) = \sum_{m \in \mathbb{M}} \left(\sum_{\substack{i \in \mathbb{X} \\ j \in C}} f^{AB}_{m,ij} + \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f^{AB}_{m,ij} \right) \tag{2.82}$$

implying that $\sum_{m \in \mathbb{M}} \sum_{\substack{i \in A \\ j \in \mathbb{X}}} f^{AB}_{m,ij} = \sum_{m \in \mathbb{M}} \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f^{AB}_{m,ij}$.

$\square$

As a result of (2.78), the discrete out-rate averaged over one period equals the average discrete in-rate, which we define to be $\bar{k}_M^{AB}$, i.e.,

$$\bar{k}_M^{AB} := \frac{1}{M} \sum_{m \in \mathbb{M}} k_m^{A\rightarrow} = \frac{1}{M} \sum_{m \in \mathbb{M}} k_m^{\rightarrow B}. \tag{2.83}$$

This period-averaged discrete rate tells us the average probability of a reactive trajectory to depart in A per time step, or in other words, the expected number of reactive trajectories leaving A per time step.

2.3.4 Example: A twisted triple well potential

We are now illustrating the introduced theory on the example of a particle in the triplewell landscape V from Eq. (2.46) when additionally a periodically varying force is applied. In particular, we add the force

$$F(x, y, t) = 1.4 \, \cos\left(\frac{2\pi t}{1.8}\right)(-y, x) \tag{2.84}$$

to the gradient dynamics in the triple well. Due to the cosine modulation, the force alternatingly exhibits an anti-clockwise circulation and a clockwise circulation. The resulting diffusion process with 1.8-periodic forcing is given by

$$dX_t = (-\nabla V(X_t) + F(X_t, t)) \, dt + \sigma \, dW_t. \tag{2.85}$$

We discretize the dynamics into a Markov chain and estimate transition matrices $P_0, P_1, \ldots, P_{M-1}$ ($M = 6$) for Δt−spaced time points during the period, each transition matrix is mapping $\Delta t = 0.3$ into the future. In Fig. 2.10 the force coming from the potential plus circulation is shown for time points corresponding to $m = 0$ and $m = 3$.
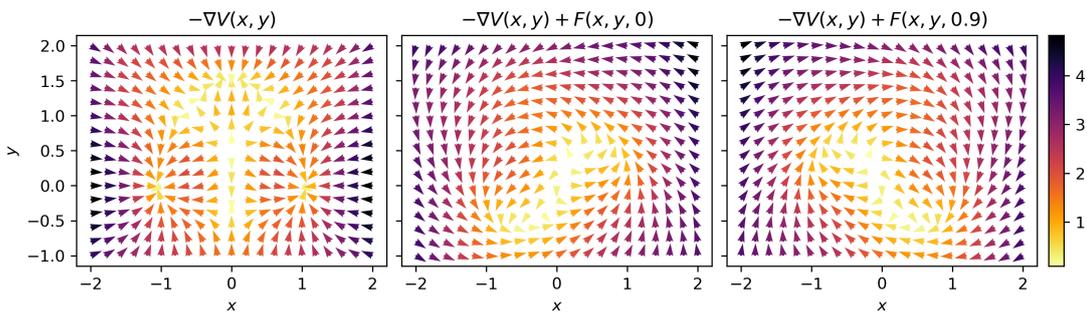
Figure 2.10: The negative gradient of the potential $-\nabla V(x, y)$ gives the force in the unperturbed system (first panel). For this example we add a periodically varying force inducing a circulation. The forces at times $m = 0$ and $m = 3$ are the two extremes of the cosine-modulated force and shown in the middle resp. right panel.

We are now interested in finding out how the transition pathways are changed when the dynamics are in equilibrium to the periodically varying forcing. As sets A

2.3. Periodic dynamics
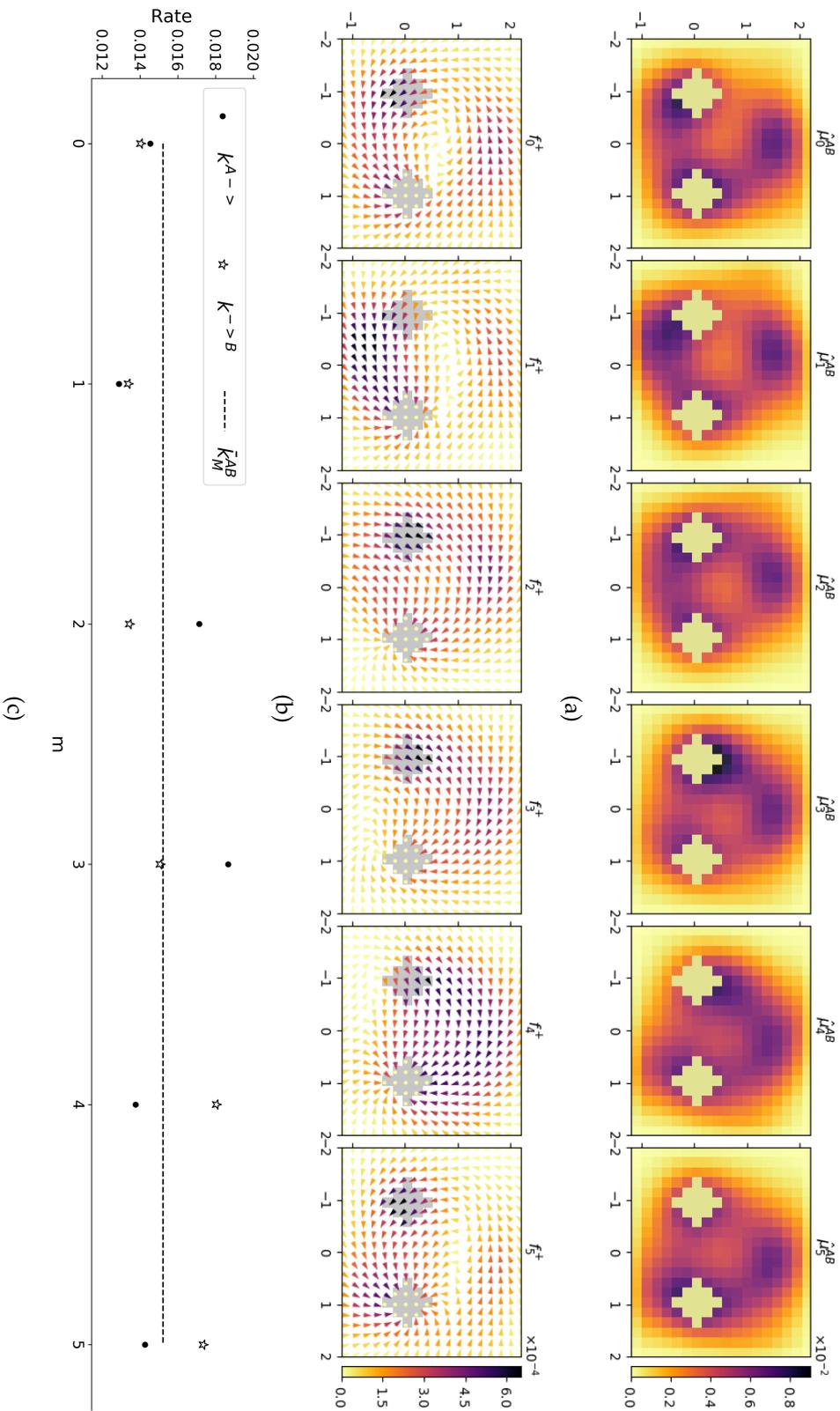
Figure 2.11: (a) The distribution and (b) effective current of reactive trajectories for the triple well dynamics with periodic circulative forcing that is sometimes clockwise, sometimes anti-clockwise. (c) The rates $k_m^{A\rightarrow}$ and $k_m^{\rightarrow B}$ indicate the probability of a reactive trajectory to leave A respectively enter B per time step.

and B we again consider the sets around the two deep minima of the triple well. The results of TPT are shown in Fig. 2.11. The time-dependent effective current indicates that at the beginning of the period the majority of transitions traverses the lower channel via the direct barrier. This agrees with the fact that the forcing gives the particle an additional push in the anti-clockwise direction during the beginning of the period. Towards the end of the period transitions more likely pass through the upper channel via the shallow well which can be explained by the clockwise forcing during the end of the period. Interestingly, the effective current in the upper channel points alternatingly in the positive and the negative $x-$direction. We can therefore assume that some reactive trajectories in the upper channel are moving back and forth near the shallow well for some time without being able to reach B. The rate of reactive trajectories leaving A is highest at times of the period $m = 2, 3$ and the rate of entering B is highest towards $m = 4, 5$, suggesting that most reactive trajectories start and arrive during this part of the period. But transitions are not necessarily completed within the same period. The averaged rate of transitions is given by 0.0152, thus the transitions are more frequent compared to the stationary setting.

2.4 Time-dependent dynamics on a finite time interval

In this section we will obtain transition statistics for Markov chains defined on a finite time interval. The dynamics are defined through an initial distribution at the beginning of the time interval and a possibly time-dependent transition matrix that gives information about how the distribution changes in time. A wide range of processes can be modelled in such a way, for example, processes that are not in stationarity or externally forced.

We are interested in transitions from source A to sink B that are restricted to start and end during the finite time interval. This restriction can be obtained by choosing certain terminal resp. initial conditions for the time-dependent committor equations. In particular, by setting $q_i^+ = \mathbb{1}_B(i)$ as a terminal condition for the forward committor, hitting B after or at the final time N is only possible when the process is already in B. By choosing $q_i^- = \mathbb{1}_A(i)$ as the initial condition of the backward committor, having come from A at the initial time is only possible when being already in A. In this section we will work this out in more detail, discuss the resulting statistics and explain how this can be changed to also account for transitions that start or end outside of the fixed time frame.

2.4.1 Setting

We start by describing the dynamics of interest in this section.

Assumption 2.24. *We consider a Markov chain* $(X_n)_{0 \leq n \leq N}$ *on a finite time interval* $\{0. \ldots, N\}$ *taking values in a discrete and finite state space* $\mathbb{X} = \{1, \ldots, S\}$. *The probability of moving*

51

from the state i at time n to the state j at the next time point n + 1 is given by the (i, j)−entry of the row-stochastic transition matrix P(n):

$$P_{ij}(n) := \mathbb{P}(X_{n+1} = j \mid X_n = i). \tag{2.86}$$

Given an initial distribution $\mu(0)$, the distributions at later times are given by $\mu(n + 1)^\top = \mu(n)^\top P(n)$.

By these assumptions, the chain can have time-inhomogeneous transition probabilities; and even if $P(n) \equiv P$ for all n, the distributions $\mu(n)$ need not be constant in time. We are also not requiring the chain to be irreducible anymore.

We also need the time-reversed process $(X_n^-)_{0 \le n \le N}$ defined by $X_n^- := X_{N-n}$ which is also a Markov chain (Theorem 1.7). Its transition probabilities are given for any $n \in \{1, \ldots, N\}$ by

$$P_{ij}^-(n) := \mathbb{P}(X_{n-1} = j \mid X_n = i) = \frac{\mu_j(n-1)}{\mu_i(n)} P_{ji}(n-1) \tag{2.87}$$

whenever $\mu_i(n) > 0$, and otherwise by $P_{ij}^-(n) = 1/S$.

2.4.2   Committor probabilities

We are only interested in reactive trajectories that start in A at or after time 0 and reach B until time N, therefore we have to adjust the hitting times and restrict them to the interval $\{0, \ldots, N\}$ as follows:

$$\tau_S^+(n) := \min\{k \in \{0, \ldots, N\} \text{ s.t. } k \ge n, X_k \in S\}, \min \varnothing := +\infty. \tag{2.88}$$

and

$$\tau_S^-(n) := \max\{k \in \{0, \ldots, N\} \text{ s.t. } k \le n, X_k \in S\}, \max \varnothing := -\infty. \tag{2.89}$$

In the case of transitions only during a finite time interval, it is no longer ensured that $\tau_{A\cup B}^+(n) \ne +\infty$ and $\tau_{A\cup B}^-(n) \ne -\infty$.[7] For example when the chain is in an absorbing state it can no longer reach A or B, hence $\tau_{A\cup B}^+(n) = +\infty$. To have a well-defined committor problem, we define $\mathbb{P}(\infty < \infty) := 0$ implying that for states from which neither A nor B can be reached, the committor is zero. This definition is sensible from the standpoint of reactive trajectories which cannot pass states from which neither A nor B is reachable and results in a zero distribution and current of reactive trajectories in these states. Additionally, this definition gives a consistent committor problem. Whenever we consider a state $i \in C$ from which neither A nor B can be reached, then from this state i one can at most reach a set of states $J \subseteq C$ that shares this property. The committor for all these states is defined to be 0 making the committor problem consistent.

---

[7]In the stationary case this was ensured by the irreducibility of P. Also in the M−stationary case this was provided by the irreducibility of $\bar{P}_0$.

So let us set up the committor problem. We only need to clarify the terminal and initial condition since the rest follows from Theorem 2.1. The terminal condition for the forward committor on $i \in B$ can be rewritten as

$$q_i^+(N) = \mathbb{P}(\tau_B^+(N) < \tau_A^+(N) \mid X_N = i) = \mathbb{P}(N < \infty \mid X_N = i) = 1 \qquad (2.90)$$

since $\tau_B^+(N) = N$ and $\tau_A^+(N) = \infty$ when $X_N \in B$. For $i \in A$ on the other hand we arrive at

$$q_i^+(N) = \mathbb{P}(\infty < N \mid X_N = i) = 0. \qquad (2.91)$$

On $i \in C$ we get

$$q_i^+(N) = \mathbb{P}(\tau_B^+(N) < \tau_A^+(N) \mid X_N = i) = \mathbb{P}(\infty < \infty \mid X_N = i) = 0 \qquad (2.92)$$

by the above definition of $\mathbb{P}(\infty < \infty) := 0$. Analogously, the initial condition of the backward committor turns out to be $q_i^-(0) = 1$ on $i \in A$, $q_i^-(0) = 0$ on $i \in B$, and $q_i^-(0) = 0$ on $i \in C$.

Then the forward committor problem for transitions restricted to $\{0, \ldots, N\}$ is given by

$$\begin{cases} q_i^+(n) & = & \sum\limits_{j \in \mathbb{X}} P_{ij}(n)\, q_j^+(n+1) & i \in C \\ q_i^+(n) & = & 0 & i \in A \\ q_i^+(n) & = & 1 & i \in B \end{cases} \qquad (2.93)$$

for $n \in \{0, \ldots, N-1\}$ and with terminal condition $q_i^+(N) = \mathbb{1}_B(i)$. Analogously, the backward committor satisfies for $n \in \{1, \ldots, N\}$

$$\begin{cases} q_i^-(n) & = & \sum\limits_{j \in \mathbb{X}} P_{ij}^-(n)\, q_j^-(n-1) & i \in C \\ q_i^-(n) & = & 1 & i \in A \\ q_i^-(n) & = & 0 & i \in B \end{cases} \qquad (2.94)$$

with initial condition $q_i^-(0) = \mathbb{1}_A(i)$. Because the committor equations (2.93) and (2.94) are defined iteratively, the solutions exist and are unique.

The following theorem provides us with a finite-time analogue result to Proposition 2.21 for the path probabilities of the paths that start in A and end in B within the restricted time frame $\{0, \ldots, N\}$.

Proposition 2.25. *The forward committor at time $n \in \{0, \ldots, N-1\}$ and the backward committor at time $n \in \{1, \ldots, N\}$, respectively, satisfy for $i \in C$ the following equalities:*

$$q_i^+(n) = \sum_{k=n+1}^{N-1} \sum_{\substack{i_k \in B \\ i_{n+1}, \ldots, i_{k-1} \in C}} P_{i i_{n+1}}(n) \cdots P_{i_{k-1} i_k}(k-1), \qquad (2.95)$$

$$q_i^-(n) = \sum_{k=0}^{n-1} \sum_{\substack{i_k \in A \\ i_{k+1}, \ldots, i_{n-1} \in C}} P_{ii_{n-1}}^-(n) \cdots P_{i_{k+1}i_k}^-(k+1). \tag{2.96}$$

Proof. The proof follows from iteratively substituting the forward committor equation of $q^+(n+1)$ into the equation of $q^+(n)$ until reaching the terminal condition. Similarly for the backward equation. □

2.4.3   Transition statistics and interpretations

We have seen that the forward and backward committors for a finite-time Markov chain can be computed from the iterative Eqs. (2.93) and (2.94) with terminal respectively initial conditions. Based on these, we will next introduce the corresponding transition statistics.

The distribution of inner reactive trajectories is given for any time $n \in \{0, \ldots, N\}$ by

$$\mu_i^{AB}(n) = q_i^-(n)\, \mu_i(n)\, q_i^+(n). \tag{2.97}$$

Observe that $\mu_i^{AB}(0) = \mu_i^{AB}(N) = 0$ for all i because there are yet no inner reactive trajectories at these times. Therefore the distribution of inner reactive trajectories cannot be normalized at times 0 and N, consequently the normalized distribution of reactive trajectories $\hat{\mu}^{AB}(n)$ is just defined for times $n \in \{1, \ldots, N-1\}$. The current of reactive trajectories for times $n \in \{0, \ldots, N-1\}$ is given by

$$f_{ij}^{AB}(n) = q_i^-(n)\, \mu_i(n)\, P_{ij}(n)\, q_j^+(n+1). \tag{2.98}$$

Due to the special form of the committors at the interval boundaries, we know that the current $f_{ij}^{AB}(0)$ can only be non-zero for $i \in A$ and the current $f_{ij}^{AB}(N-1)$ can only be non-zero for $j \in B$. The discrete rate of transitions leaving A is defined for times $n \in \{0, \ldots, N-1\}$ by

$$k^{A\rightarrow}(n) = \sum_{\substack{i \in A \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) \tag{2.99}$$

whereas the discrete rate of transitions entering B is given for times $n \in \{1, \ldots, N\}$ by

$$k^{\rightarrow B}(n) = \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f_{ij}^{AB}(n-1). \tag{2.100}$$

Given all these statistical properties, we can show that the current of reactive trajectories into a state equals the amount of current going out of that state one time step later. Further, the discrete out-rate averaged over the whole time interval equals the time-averaged in-rate.

Theorem 2.26. *For a finite-time Markov chain $(X_n)_{0 \leq n \leq N}$ satisfying Assumption 2.24, the current of reactive trajectories flowing into a node $i \in C$ at time $n - 1$ equals the current flowing out of the node $i \in C$ at time n, i.e.,*

$$\sum_{j \in \mathbb{X}} f_{ji}^{AB}(n-1) = \sum_{j \in \mathbb{X}} f_{ij}^{AB}(n) \qquad (2.101)$$

where $n \in \{1, \ldots, N-1\}$. Further, the rate of transitions flowing out of A over the whole time interval $\{0, \ldots, N\}$ equals the rate of reactive trajectories into B over the interval

$$\sum_{n=0}^{N-1} k^{A \to}(n) = \sum_{n=1}^{N} k^{\to B}(n). \qquad (2.102)$$

Proof. First, for any $i \in C$ and $n \in \{1, \ldots, N-1\}$ we have on the one hand that

$$\sum_{j \in \mathbb{X}} f_{ij}^{AB}(n) = q_i^-(n)\,\mu_i(n) \left(\sum_{j \in \mathbb{X}} P_{ij}(n)\, q_j^+(n+1) \right) \overset{(1)}{=} q_i^-(n)\,\mu_i(n)\, q_i^+(n), \qquad (2.103)$$

where (1) follows from the committor equation (2.93). On the other hand we find that

$$
\begin{aligned}
\sum_{j \in \mathbb{X}} f_{ji}^{AB}(n-1) &= \sum_{j \in \mathbb{X}} q_j^-(n-1)\,\mu_j(n-1)\, P_{ji}(n-1)\, q_i^+(n) \\
&\overset{(2)}{=} q_i^+(n)\,\mu_i(n) \left(\sum_{j \in \mathbb{X}} P_{ij}^-(n)\, q_j^-(n-1) \right) \overset{(3)}{=} q_i^+(n)\,\mu_i(n)\, q_i^-(n),
\end{aligned}
\qquad (2.104)
$$

where (2) follows from the backward transition probabilities (2.87) and (3) from the backward committor equation in (2.94). Thus the inflow and successive outflow are equal to another and given by the distribution of inner reactive trajectories.

Second, by using that $f_{ij}^{AB}(n) = 0$ for any $n \in \{0, \ldots, N-1\}$ if $i \in B, j \in \mathbb{X}$ or if $i \in \mathbb{X}, j \in A$ we arrive at the following equality

$$\sum_{n=0}^{N-1} \left(\sum_{\substack{i \in \mathbb{X} \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) \right) = \sum_{n=0}^{N-1} \left(\sum_{\substack{i \in A \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) + \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) \right) = \sum_{n=0}^{N-1} \left(\sum_{\substack{i \in \mathbb{X} \\ j \in C}} f_{ij}^{AB}(n) + \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f_{ij}^{AB}(n) \right).$$

$$(2.105)$$

Then, we can show that

$$
\begin{aligned}
\sum_{n=0}^{N-1} \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) &= \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ij}^{AB}(0) + \sum_{n=1}^{N-1} \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) \overset{(4)}{=} \sum_{n=1}^{N-1} \sum_{\substack{i \in C \\ j \in \mathbb{X}}} f_{ji}^{AB}(n-1) \\
&\overset{(5)}{=} \sum_{n=0}^{N-2} \sum_{\substack{j \in C \\ i \in \mathbb{X}}} f_{ij}^{AB}(n) + \sum_{\substack{j \in C \\ i \in \mathbb{X}}} f_{ij}^{AB}(N-1) = \sum_{n=0}^{N-1} \sum_{\substack{j \in C \\ i \in \mathbb{X}}} f_{ij}^{AB}(n),
\end{aligned}
\qquad (2.106)
$$

2.4. *Time-dependent dynamics on a finite time interval*

where in (4) we have applied the time-dependent current conservation for $i \in C$, $n \in \{1, \dots, N-1\}$ and we have used that $f_{ij}^{AB}(0) = 0$ for $i \in C$. And in (5) we have relabeled i, j and used that $f_{ij}^{AB}(N-1) = 0$. As a consequence

$$\sum_{n=0}^{N-1} \sum_{\substack{i \in A \\ j \in \mathbb{X}}} f_{ij}^{AB}(n) = \sum_{n=0}^{N-1} \sum_{\substack{i \in \mathbb{X} \\ j \in B}} f_{ij}^{AB}(n) \qquad (2.107)$$

and thus by the definitions of the discrete rates

$$\sum_{n=0}^{N-1} k^{A \to}(n) = \sum_{n=1}^{N} k^{\to B}(n). \qquad (2.108)$$

$\square$

For stationary infinite-time Markov chains the out-rate equals the in-rate. For finite-time dynamics this only holds when averaging over the time interval, as shown by result (2.102). Analogously to the stationary case, we denote the time-averaged rate by $\bar{k}_N^{AB}$, i.e.,

$$\bar{k}_N^{AB} := \frac{1}{N+1} \sum_{n=0}^{N-1} k^{A \to}(n) = \frac{1}{N+1} \sum_{n=1}^{N} k^{\to B}(n). \qquad (2.109)$$

In the infinite-time stationary case, Theorem 2.17 tells us that k^{AB} equals the time average of the number of reactive trajectories departing A per time step along a single infinitely long trajectory. Here, we cannot apply the ergodic theorem to turn $\bar{k}_N^{AB}$ into an average along a single trajectory. Instead, we can use the law of large numbers to write the time-averaged rate in terms of an ensemble average to get a better understanding. For this, we take an infinite number of i.i.d. copies of the finite-time chain, i.e., for each $i = 1, 2, \dots$ the chain $(X_n^i)_{0 \le n \le N}$ is distributed according to the law of the finite-time dynamics with $X_0^i \sim \mu(0)$, $X_1^i \sim P_{X_0^i}(0)$ etc. Then by the law of large numbers in Theorem 1.15, we have the following $\mathbb{P}-$a.s. convergence

$$\lim_{K \to \infty} \frac{1}{K} \sum_{i=1}^{K} \frac{1}{N+1} \sum_{n=0}^{N-1} \mathbb{1}_A(X_n^i) \mathbb{1}_B \left(X_{\tau_{A \cup B}^+(n+1)}^i \right)$$

$$= \frac{1}{N+1} \sum_{n=0}^{N-1} \mathbb{E} \left[\mathbb{1}_A(X_n) \mathbb{1}_B \left(X_{\tau_{A \cup B}^+(n+1)} \right) \right] \qquad (2.110)$$

$$= \frac{1}{N+1} \sum_{n=0}^{N-1} \mathbb{P}(X_n \in A, \tau_B^+(n+1) < \tau_A^+(n+1)) = \bar{k}_N^{AB}.$$

Hence we can interpret the average rate $\bar{k}_N^{AB}$ as the total expected number of reactive trajectories within $\{0, \dots, N\}$ divided by the number of time steps. Analogously, we

can apply the same argument for the time-averaged probability of being on a transition

$$\bar{H}_N^{AB} := \frac{1}{N+1} \sum_{n=0}^{N} H^{AB}(n) = \frac{1}{N+1} \mathbb{E}\left[\sum_{n=0}^{N} \mathbb{1}_A \left(X_{\tau_{A\cup B}^-(n)} \right) \mathbb{1}_B \left(X_{\tau_{A\cup B}^+(n+1)} \right) \right], \qquad (2.111)$$

such that $\mathbb{P}-$a.s.

$$\lim_{K\to\infty} \frac{1}{K} \sum_{i=1}^{K} \frac{1}{N+1} \sum_{n=0}^{N} \mathbb{1}_A \left(X_{\tau_{A\cup B}^-(n)}^i \right) \mathbb{1}_B \left(X_{\tau_{A\cup B}^+(n+1)}^i \right) = \bar{H}_N^{AB} \qquad (2.112)$$

and $\bar{H}_N^{AB}$ can be understood as the expected number of time steps the Markov chain is on an inner reactive trajectory or being in A just before beginning a transition during $\{0, \ldots, N\}$ divided by $N + 1$. Last, we define the ratio

$$\bar{t}_N^{AB} := \frac{\bar{H}_N^{AB}}{\bar{k}_N^{AB}} \qquad (2.113)$$

and observe that it provides us with an approximation of the expected duration of a reactive trajectory over $n \in \{0, \ldots, N\}$. It is only an approximation since the expected value of a quotient of random variables is in general not the expected value of the numerator divided by the expected value of the denominator.

In conclusion, some of the properties of the stationary case also hold in the finite-time case after averaging.

2.4.4 Example: Restricted transitions in the triplewell

To demonstrate the effect of the finite-time restriction on the transition behaviour between A and B, we now study the stationary dynamics in a triple well (Section 2.2.5) restricted to a finite time window and initiated in the stationary distribution. We use the same discretization as before and consider the interval $\mathbb{T} = \{0, \ldots, N-1\}$, $N = 6$.

Even though we study the same underlying dynamics as in the case of stationary dynamics with infinite time, the possible transition paths between A and B are limited to the pathway that is fast to traverse, i.e., the lower channel via the direct barrier, see Fig. 2.12. Since only a small portion of the reactive trajectories from the infinite-time example has a short enough transition time to be considered in this case, the average rate of transitions $\bar{k}_6^{AB} = 0.0017$ is much lower than the corresponding rate $k^{AB} = 0.0142$ in the infinite-time case, and the average time a transition takes $\bar{t}_6^{AB} = 3.07$ is much shorter than in the infinite-time case.

In a second example, Fig. 2.13, we want to highlight the usage of finite-time TPT to study large qualitative changes in the transition behavior of a system. We consider the stationary triple well example over a finite interval $\mathbb{T} = \{0, \ldots, N-1\}$, but this time with a smaller noise strength of $\sigma = 0.26$ compared to the previous example. As we increase the interval length N from short $N = 20$ to long $N = 500$, we allow

2.4. *Time-dependent dynamics on a finite time interval*

(a)

(b)
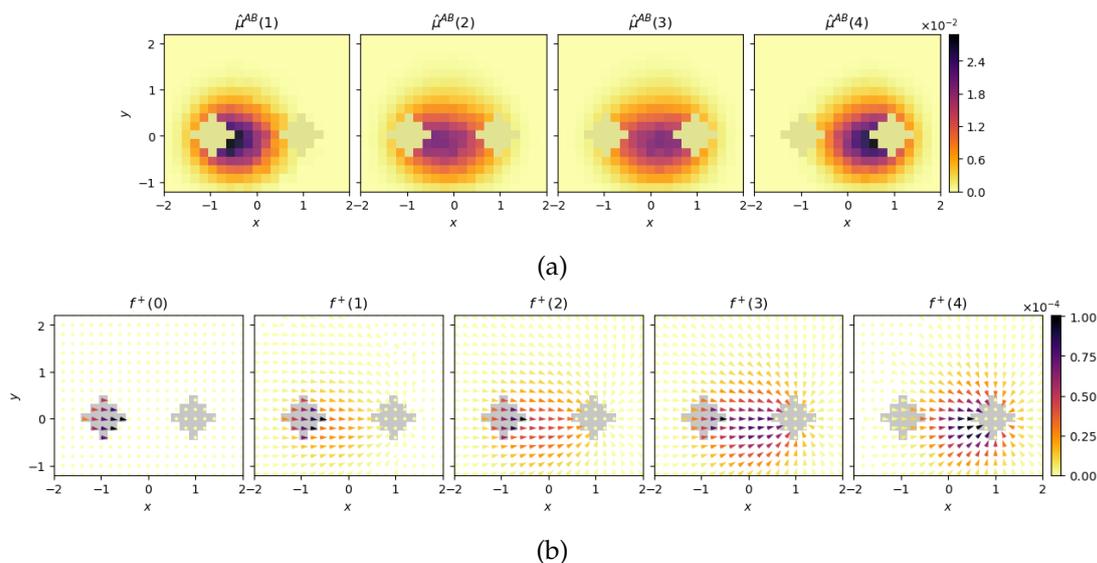
Figure 2.12: (a) The distribution and (b) the effective current of reactive trajectories are shown for the stationary triple well dynamics restricted to a finite-time window $\{0, \dots, 5\}$.

reactive trajectories to be longer and longer, and thus the average reactive trajectory changes. Whereas for the short interval most of the distribution and current is around the lower transition channel, see Fig. 2.13, for the long interval most of the distribution and current is around the upper transition channel. The transition behaviour for the long interval is already not distinguishable from the infinite-time transition dynamics (not shown here, see [28]).

2.4.5 Transitions that extend the interval boundaries

Whenever we know something about the dynamics outside the time interval $\{0, \dots, N\}$, we can also employ other initial and terminal conditions for the committors and allow reactive trajectories to extend over the boundaries of the finite time interval.

Consider the following dynamics of a mostly stationary Markov chain that experiences a shock, i.e., the transition matrix changes only during a small time interval. Let the Markov chain $(X_n)_{n \in \mathbb{Z}}$

1. be stationary and time-homogeneous for $n < 0$ with respect to the irreducible transition matrix $P_{n<0}$ and with stationary distribution $\pi_{n<0}$

2. for times $n \in \{0, .., N-1\}$ the transition probabilities change and are given by $P(n)$, thus the distribution at time $n = 1, \dots, N$ becomes $\mu^\top(n) = \pi_{n<0}^\top P(0) \cdots P(n-1)$

3. after time $n = N - 1$ the dynamics is again given by a time-independent and irreducible transition matrix $P_{n \geq N}$ and the distribution approaches the stationary distribution of $P_{n \geq N}$ for large $n \geq N$

58

(a)

(b)
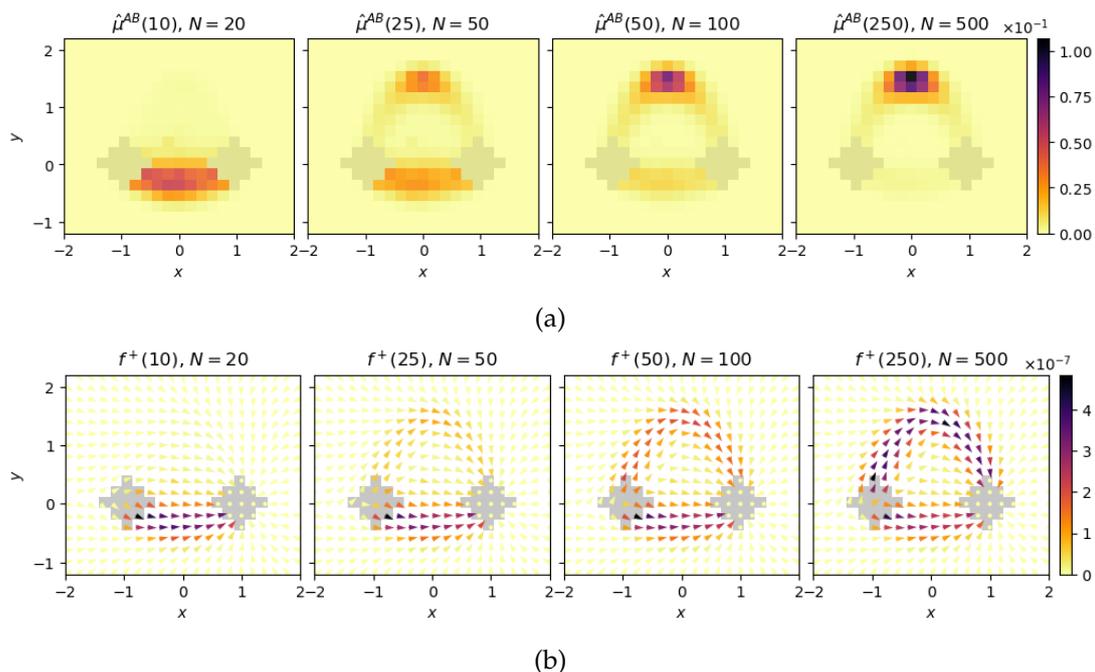
Figure 2.13: Qualitative changes in the transition behaviour for finite-time, stationary TPT and increasing time-interval lengths N. The plots show (a) the distribution of reactive trajectories (normalized to being reactive) and (b) the effective current at the time point $n = \frac{N}{2}$.

Then the backward committor for $n \leq 0$ is constant since the backward transition matrix $P^-_{n<0}$ (the reversal of $P_{n<0}$ with respect to $\pi_{n<0}$) is time-independent and the backward committor cannot be influenced by the shock that lies in the future. The backward committor for $n \leq 0$ solves the stationary committor equation with the time-reversed transition matrix $P^-_{n<0}$. Analogously, the forward committor for times $n \geq N$ is also time-independent and solves the stationary forward committor equation with transition matrix $P_{n \geq N}$. The committors on the finite time interval can then be computed with initial resp. terminal conditions given by the stationary backward committor before time 1 resp. the forward committor from after $N - 1$.

In many scenarios these initial and terminal conditions might be more natural and do not introduce singularities at the time interval boundaries by forbidding transitions that extend the interval. In that way we consider all reactive trajectories that pass the time interval.

3 | Transition Path Theory for absorbing chains

In this chapter we are interested in systems that eventually leave the domain of interest and cannot come back, i.e., are absorbed outside of the domain of interest. An example for such a dynamic is a population process where the population will eventually become extinct but the interesting regime is the transient phase where the population is still alive. But many other agent-based models have one or more of such absorbing states [30]. The question we will try to answer in this chapter is, what can we say about the transitions from A to B before the Markov chain leaves the states of interest? Ultimately the chain will be absorbed, therefore the stationary distribution of such a chain is concentrated on the absorbing states. Consequently, we cannot rely on the stationary TPT setting from Section 2.2 to study the interesting transitions. Instead, we will start by describing possible ways for modeling such dynamics and finding the distribution before the chain has left the domain. Then we can discuss the application of Transition Path Theory to study the transitions prior to the escape from the domain of interest.

3.1 Dynamics prior to ultimate absorption

We consider a time-homogeneous Markov chain $(X_n)_{n \in \mathbb{Z}}$ on the finite state space $\mathbb{X} \cup \omega$. We assume that the process is irreducible and aperiodic on $\mathbb{X}$, the *domain of interest*, and that from at least one state in $\mathbb{X}$ the Markov chain can escape the domain $\mathbb{X}$ and enter the *absorbing state* given by ω. The transition matrix Q on $\mathbb{X}$ is *substochastic*, i.e., there is at least one row whose sum is less than one due to a positive transition probability to ω. The state ω is absorbing in the sense that once the process has reached it, it will stay there with probability one.

We denote by $\tau_\omega := \min\{n \in \mathbb{Z} \text{ s.t. } X_n = \omega\}$ the first hitting time of the absorbing state ω. Due to the above assumption of an irreducible process on $\mathbb{X}$, absorption to ω is certain. The transition matrix on the joint state space $\mathbb{X} \cup \omega$ takes the following form:

$$P = \begin{pmatrix} (Q_{ij})_{i,j \in \mathbb{X}} & (1 - \sum_{k \in \mathbb{X}} Q_{ik})_{i \in \mathbb{X}} \\ 0 & 1 \end{pmatrix}. \tag{3.1}$$

3.1. *Dynamics prior to ultimate absorption*

There are two viewpoints of the process: either as a process on $\mathbb{X}$ with substochastic transition matrix Q and which eventually leaves the domain $\mathbb{X}$, or as a process on $\mathbb{X} \cup \omega$ where ω is absorbing and the states of $\mathbb{X}$ are transient states.

Remark 3.1. *The theory of this chapter is also applicable to Markov chains with several absorbing states or classes by simply aggregating all the absorbing states and classes into one state ω and considering the remainder as the domain of interest $\mathbb{X}$ on which the dynamics remains unaltered by the aggregation.*

Following [18, 16], we will next discuss the dominant eigenvalue and corresponding eigenvectors of Q and explain their significance. For notational simplicity we denote $\mathbb{P}_f(\cdot) := \mathbb{P}(\cdot \mid X_0 \sim f)$ and $\mathbb{P}_i(\cdot) := \mathbb{P}(\cdot \mid X_0 = i)$. From the Perron-Frobenius Theorem [63, Theorem 1.5.] we know that Q has a positive left eigenvector ν and positive right eigenvector w corresponding to the eigenvalue that is largest in absolute value and positive, $\lambda < 1$.[1] We assume in the following that both eigenvectors are normalized such that $\sum_i \nu_i = 1$ and $\sum_i \nu_i w_i = 1$.

Then by the following computation,

$$\lambda^n = \sum_{i \in \mathbb{X}} (\nu Q^n)_i = \sum_{i \in \mathbb{X}} \mathbb{P}_\nu(X_n = i) = \mathbb{P}_\nu(X_n \in \mathbb{X}) = \mathbb{P}_\nu(\tau_\omega > n) \tag{3.2}$$

the dominant eigenvalue λ gives the probability to remain at least one more time step in $\mathbb{X}$. Further, the left eigenvector $\nu = (\nu_i)_{i \in \mathbb{X}}$ gives the following distribution

$$\nu_i = \frac{1}{\lambda^n} (\nu Q^n)_i = \mathbb{P}_\nu(X_n = i \mid \tau_\omega > n) \tag{3.3}$$

which is also called the *quasi-stationary distribution*. When the process is initialized in the quasi-stationary distribution, then by conditioning on not yet having been absorbed, this distribution is stationary. Moreover from [63, Theorem 1.2] by using the aperiodicity of the chain, it follows that

$$(Q^n)_{ij} = \lambda^n w_i \nu_j + O(n^{m_2-1} |\lambda_2|^n), \tag{3.4}$$

where λ_2 is the second largest eigenvalue in absolute value of Q, i.e., $|\lambda_2| < \lambda$, and assumed non-zero (otherwise a similar result can be established). The algebraic multiplicity of λ_2 is m_2, and it is assumed that if there exists another eigenvalue λ_j with $|\lambda_j| = |\lambda_2|$, then its multiplicity $m_j \leq m_2$. Eq. (3.4) allows us to give meaning to the right eigenvector w of Q. We can first sum Eq. (3.4) over $j \in \mathbb{X}$ to give

$$\mathbb{P}_i(\tau_\omega > n) = \sum_{j \in \mathbb{X}} (Q^n)_{ij} = \lambda^n w_i + O(n^{m_2-1} |\lambda_2|^n) \tag{3.5}$$

---

[1] The Perron-Frobenius Theorem only tells us that $\lambda \leq 1$. But Prop. A.2 implies that the weakly chained diagonally dominant matrix $(I - Q)$ is invertible and therefore does not have an eigenvalue 0, hence Q cannot have an eigenvalue 1.

and by dividing by $\lambda^n = \mathbb{P}_\nu(\tau_\omega > n)$ and taking the limit of $n \to \infty$, we arrive at

$$w_i = \lim_{n \to \infty} \frac{\mathbb{P}_i(\tau_\omega > n)}{\mathbb{P}_\nu(\tau_\omega > n)}. \tag{3.6}$$

Thus the right eigenvector is proportional to the probability of never exiting $\mathbb{X}$ when starting in state i. The constant factor $(\mathbb{P}_\nu(\tau_\omega > n))^{-1}$ ensures that the probability does not approach zero in the limit. Moreover using Eq. (3.4), we can give ν the interpretation as a limiting distribution, namely in the limit ν looses its dependence on the starting state:

$$\nu_i = \lim_{n \to \infty} \mathbb{P}_j(X_n = i \mid \tau_\omega > n) \text{ for every } j \in \mathbb{X}. \tag{3.7}$$

In the next two sections we will introduce two possibilities for modeling the dynamics before absorption.

3.1.1 The survival process that never leaves $\mathbb{X}$

We denote by $(Z_n)_{n \in \mathbb{Z}}$ the original process $(X_n)_{n \in \mathbb{Z}}$ conditioned on never leaving $\mathbb{X}$, and we call $(Z_n)_{n \in \mathbb{Z}}$ the *survival process*. The law of the survival process conditional on starting in $\mathbb{X}$ is therefore given by the following limit of the absorption time in the infinitely far away future

$$\mathbb{P}\left(Z_{n_1} = i_1, \ldots, Z_{n_k} = i_k \mid Z_0 = i_0\right) := \lim_{m \to \infty} \mathbb{P}\left(X_{n_1} = i_1, \ldots, X_{n_k} = i_k \mid \tau_\omega > m, X_0 = i_0\right) \tag{3.8}$$

with time points $0 = n_0 < n_1 < \ldots < n_k < m$ and states $i_0, \ldots, i_k \in \mathbb{X}$. We follow the derivation in [16, Chapter 3.2] to obtain an expression of the law in terms of transition probabilities. We can rewrite

$$
\begin{aligned}
&\lim_{m \to \infty} \mathbb{P}\left(X_{n_1} = i_1, \ldots, X_{n_k} = i_k \mid \tau_\omega > m, X_0 = i_0\right) \\
&= \lim_{m \to \infty} \frac{\mathbb{P}\left(X_{n_1} = i_1, \ldots, X_{n_k} = i_k, \tau_\omega > m \mid X_0 = i_0\right)}{\mathbb{P}\left(\tau_\omega > m \mid X_0 = i_0\right)} \\
&= \mathbb{P}\left(X_{n_1} = i_1, \ldots, X_{n_k} = i_k \mid X_0 = i_0\right) \lim_{m \to \infty} \frac{\mathbb{P}\left(\tau_\omega > m - n_k \mid X_0 = i_k\right)}{\mathbb{P}\left(\tau_\omega > m \mid X_0 = i_0\right)},
\end{aligned} \tag{3.9}
$$

where in the last line we used the Markov property and time-homogeneity of the original chain. Further, from Eq. (3.5) we get

$$\lim_{m \to \infty} \frac{\mathbb{P}_j(\tau_\omega > m - n)}{\mathbb{P}_i(\tau_\omega > m)} = \lambda^{-n} \frac{w_j}{w_i} \tag{3.10}$$

and can therefore rewrite the conditional distribution (3.9) as a product of transition probabilities

$$
\begin{aligned}
(3.9) = & \mathbb{P}\left(X_{n_1} = i_1, \ldots, X_{n_k} = i_k \mid X_0 = i_0\right) \lambda^{-n_k} \frac{w_{i_k}}{w_{i_0}} \\
= & \prod_{l=1}^{k} \lambda^{n_{l-1}-n_l} \frac{w_{i_l}}{w_{i_{l-1}}} \mathbb{P}\left(X_{s_l} = i_l \mid X_{s_{l-1}} = i_{l-1}\right).
\end{aligned}
\tag{3.11}
$$

The forward transition probabilities $P_{Z,ij}$ of the survival process are therefore given by

$$
P_{Z,ij} = \mathbb{P}(Z_{n+1} = j \mid Z_n = i) = \lim_{m \to \infty} \mathbb{P}(X_{n+1} = j \mid X_n = i, \tau_\omega > m) = \lambda^{-1} Q_{ij} \frac{w_j}{w_i} \tag{3.12}
$$

for $i, j \in \mathbb{X}$, while its stationary distribution can be shown to be $(v_i w_i)_{i \in \mathbb{X}} =: vw$. The difference between the distribution v and vw is the following: The quasi-stationary distribution v gives the equilibrated distribution at time n conditioned on not having left by time n, see Eq. (3.7). The stationary survival distribution on the other hand gives the limiting distribution [18]

$$
w_i v_i = \lim_{\substack{m-n \to \infty \\ n \to \infty}} \mathbb{P}(X_n = i \mid \tau_\omega > m), \tag{3.13}
$$

i.e., the equilibrated distribution at time n conditional on not leaving $\mathbb{X}$ for an infinitely long time in the future. The backward transition matrix of the survival process when in stationarity amounts to

$$
P_{Z,ij}^{-} = \lambda^{-1} Q_{ji} \frac{v_j}{v_i}. \tag{3.14}
$$

The survival process is a stationary process but it conditions on the infinitely far away future. For many applications it is more natural to condition on not having left $\mathbb{X}$ by a certain time, therefore at next we derive another process that does not condition on the future but instead on a certain time step. This is also a more natural setting for Transition Path Theory, where we usually observe the system at a certain time point at which absorption has not yet taken place. We will see that the conditioning on absorption after a certain time step comes at the price of making the chain time-inhomogeneous.

3.1.2 The mortal process that stays in $\mathbb{X}$ at least until time 0

We condition the original process on not having left $\mathbb{X}$ by a certain time. Without loss of generality we choose this time step to be 0 and we condition on $\tau_\omega > 0$. This new process will be denoted by $(Y_n)_{n \in \mathbb{Z}}$ and called the *mortal process*. With this conditioning, transitions to ω after time 0 are allowed and therefore gradually all the probability mass will leave $\mathbb{X}$.

By Eq. (3.7), the equilibrated distribution at time $n = 0$ conditional on not yet having left $\mathbb{X}$ is given by the quasi-stationary distribution v on $\mathbb{X}$. Consequently the distribution

on the joint state space $\mathbb{X} \cup \omega$ is given by

$$(\mathbb{P}(Y_0 = i))_{i \in \mathbb{X} \cup \omega} = (\nu, 0). \tag{3.15}$$

After time 0, mass can enter the absorbing state ω and the dynamics for $n \geq 0$ is naturally described by the transition matrix

$$
\begin{aligned}
P_{Y,ij}(n) &:= \mathbb{P}(Y_{n+1} = j \mid Y_n = i) \\
&= \mathbb{P}(X_{n+1} = j \mid X_n = i, \tau_\omega > 0) \\
&= \mathbb{P}(X_{n+1} = j \mid X_n = i, X_0 \in \mathbb{X}) = P_{ij}
\end{aligned} \tag{3.16}
$$

using the Markov property.[2] Thus at time $n = 1$, the distribution is given by

$$(\mathbb{P}(Y_1 = i))_{i \in \mathbb{X} \cup \omega} = (\mathbb{P}(X_1 = i \mid \tau_\omega > 0))_{i \in \mathbb{X} \cup \omega} = (\lambda \nu, 1 - \lambda), \tag{3.17}$$

which in turn by iteratively applying $P_Y(n) = P$ is mapped to

$$(\mathbb{P}(Y_n = i))_{i \in \mathbb{X} \cup \omega} = (\lambda^n \nu, 1 - \lambda^n) \tag{3.18}$$

at time n. As $n \to \infty$, the distribution in $\mathbb{X}$ will approach zero.

From the distribution and forward transition matrix for times $n \geq 0$, we can compute the backward transition matrix. For $n \geq 0$ we can express the flux between times n and $n + 1$ and states $i, j \in \mathbb{X}$ as follows:

$$
\begin{aligned}
\mathbb{P}(Y_n = j, Y_{n+1} = i) &= \mathbb{P}(Y_{n+1} = i \mid Y_n = j)\,\mathbb{P}(Y_n = j) = Q_{ji}\,\lambda^n\,\nu_j \\
&= \mathbb{P}(Y_n = j \mid Y_{n+1} = i)\,\mathbb{P}(Y_{n+1} = i) = \mathbb{P}(Y_n = j \mid Y_{n+1} = i)\,\lambda^{n+1}\,\nu_i.
\end{aligned}
$$

This gives the backward transition matrix

$$P_{Y,ij}^-(n) := \mathbb{P}(Y_{n-1} = j \mid Y_n = i) = \lambda^{-1} \frac{\nu_j}{\nu_i} Q_{ji} \tag{3.19}$$

for $i, j \in \mathbb{X}$ and $n > 0$. On $\mathbb{X}$ the backward transition matrix is thus time-independent and equals the backward transition matrix of the survival process, compare with Eq. (3.14). By studying also the flux of state ω, we arrive at the complete backward transition matrix on $\mathbb{X} \cup \omega$ and for $n > 0$

$$P_{Y,ij}^-(n) = \begin{pmatrix} \lambda^{-1} \frac{\nu_j}{\nu_i} Q_{ji} & 0 \\ (1 - \sum_{k \in \mathbb{X}} Q_{jk})\,\nu_j \frac{\lambda^{n-1}}{1-\lambda^n} & \frac{1-\lambda^{n-1}}{1-\lambda^n} \end{pmatrix}. \tag{3.20}$$

Can we also say something about the mortal process before time $n = 0$? Looking at Eq. (3.13) we can expect that the distribution should converge to $(w\nu, 0)$ for $n \to -\infty$ since absorption to ω lies infinitely far in the future from the point of view of the

---

[2]For $n = 0$, $i = \omega$, the transition probability $P_{Y,ij}(n)$ is not defined. But since the distribution in ω is zero until $n = 0$, this does not pose any problems.

3.1. *Dynamics prior to ultimate absorption*

infinite past. Also the forward and backward transition matrix should in the limit that $n \to -\infty$ be given by the forward and backward matrix of the survival process. A simple computation shows, that the forward matrix for $n < 0$ and $i, j \in \mathbb{X}$ is given by:

$$
\begin{aligned}
P_{Y,ij}(n) &:= \mathbb{P}(X_{n+1} = j \mid X_n = i, \tau_\omega > 0) \\
&= \frac{\mathbb{P}(X_{n+1} = j, \tau_\omega > 0 \mid X_n = i)}{\mathbb{P}(\tau_\omega > 0 \mid X_n = i)} \\
&= \mathbb{P}(X_{n+1} = j \mid X_n = i) \frac{\mathbb{P}(\tau_\omega > 0 \mid X_{n+1} = j)}{\mathbb{P}(\tau_\omega > 0 \mid X_n = i)} \\
&= \mathbb{P}(X_{n+1} = j \mid X_n = i) \frac{\mathbb{P}(X_0 \in \mathbb{X} \mid X_{n+1} = j)}{\mathbb{P}(X_0 \in \mathbb{X} \mid X_n = i)} \\
&= Q_{ij} \frac{\sum_{k \in \mathbb{X}} \left(Q^{|n|-1}\right)_{jk}}{\sum_{k \in \mathbb{X}} \left(Q^{|n|}\right)_{ik}}
\end{aligned}
\tag{3.21}
$$

which is time-dependent due to conditioning on the future and indeed converges to the forward matrix of the survival process for $n \to -\infty$. For $i = \omega$, the transition probabilities are undefined.

The computation of the distribution for times $n < 0$ is more complicated. Below we will show that for $i \in \mathbb{X}$, $n < 0$

$$
\mathbb{P}(Y_n = i) = v_i \lambda^n \sum_{k \in \mathbb{X}} \left(Q^{|n|}\right)_{ik} = v_i \lambda^{-|n|} \mathbb{P}_i(\tau_\omega > |n|),
\tag{3.22}
$$

which indeed converges to $w_i v_i$ as $n \to -\infty$ by (3.6). We start by rewriting the distribution for $n < 0$ and any starting state $j \in \mathbb{X}$ at time $N < n$ as

$$
\begin{aligned}
\mathbb{P}(Y_n = i) &= \lim_{N \to -\infty} \mathbb{P}(X_n = i \mid \tau_\omega > 0, X_N = j) \\
&= \lim_{N \to -\infty} \frac{\mathbb{P}(\tau_\omega > 0 \mid X_n = i)\, \mathbb{P}(X_n = i \mid X_N = j)}{\mathbb{P}(\tau_\omega > 0 \mid X_N = j)} \\
&= \left(\sum_{k \in \mathbb{X}} \left(Q^{|n|}\right)_{ik}\right) \lim_{N \to -\infty} \frac{(Q^{n-N})_{ji}}{\sum_{k \in \mathbb{X}} (Q^{|N|})_{jk}}.
\end{aligned}
\tag{3.23}
$$

Then by using Eq. (3.4) and (3.5), we can rewrite the limit as

$$
\begin{aligned}
\lim_{N \to -\infty} \frac{(Q^{n-N})_{ji}}{\sum_{k \in \mathbb{X}} (Q^{|N|})_{jk}} &= \lim_{N \to -\infty} \frac{\lambda^{n-N} w_j v_i + \mathcal{O}\left((n-N)^{m_2-1} |\lambda_2|^{n-N}\right)}{\lambda^{|N|} w_j + \mathcal{O}\left(|N|^{m_2-1} |\lambda_2|^{|N|}\right)} \\
&= \lim_{N \to -\infty} \frac{\lambda^n w_j v_i + \mathcal{O}\left((n-N)^{m_2-1} |\lambda_2|^{n-N} \lambda^{-|N|}\right)}{w_j + \mathcal{O}\left(|N|^{m_2-1} |\lambda_2|^{|N|} \lambda^{-|N|}\right)} \\
&= \lambda^n v_i
\end{aligned}
\tag{3.24}
$$

where the convergence follows since exponential growth is stronger than polynomial

growth.

By computing the flux between states $i, j \in \mathbb{X}$ for $n \leq 0$, we can find the backward transition matrix to be given by the survival backward matrix, i.e., $P^-_{Y,ij}(n) := \lambda^{-1} \frac{v_j}{v_i} Q_{ji}$ for $n \leq 0$. This agrees with the observation that in the infinite past the process becomes the survival process.

3.1.3 Example

The following example (Fig. 3.1) shows how the stationary distribution of the survival process $(Z_n)_{n \in \mathbb{Z}}$ differs from the distribution of the mortal process $(Y_n)_{n \in \mathbb{Z}}$ which stays in $\mathbb{X}$ at least until time 0.
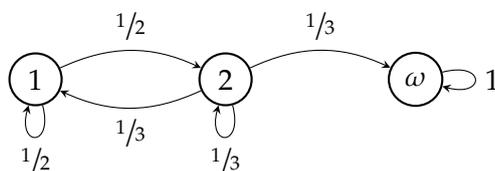
Figure 3.1: Transition probabilities of a simple process that gets absorbed in ω.

While the stationary distribution of the survival process is given by

$$\mathbb{P}(Z_n = 1) = 0.6, \ \mathbb{P}(Z_n = 2) = 0.4,$$

the time-dependent distribution of the mortal process for $n \geq 0$ is given by

$$\mathbb{P}(Y_n = 1) = \mathbb{P}(Y_n = 2) = \lambda^n \, 0.5$$

with $\lambda = \frac{5}{6}$. The distribution of the survival process is biased towards state 1 since from state 1 it is less likely to get absorbed in ω in the near future. The distribution of the mortal process on the other hand is balanced between the two states 1 and 2. The leading eigenvalue is given by 5/6, thus in the $Y-$process for $n \geq 1$ per time step 1/6th of the probability mass moves to ω.

3.2 Transitions before absorption

We are now interested in the transitions from $A \subset \mathbb{X}$ to $B \subset \mathbb{X}$. In order to study the transition behaviour of the Markov chain before being absorbed, we need a description of the dynamics prior to absorption. In the previous two sections we introduced two possibilities for modeling the process by making minimal assumptions. We will next explain how these results can be combined with TPT in order to study the reactive trajectories while the chain is still in $\mathbb{X}$.

3.2. *Transitions before absorption*

3.2.1 TPT for the survival process

By applying the TPT approach for stationary Markov chains from Chapter 2.2 to the stationary survival chain, we gain knowledge about the transitions from A to B conditional on never escaping $\mathbb{X}$. Since the survival process is only defined on the domain of interest $\mathbb{X}$, the committors are also defined on $\mathbb{X}$.

The forward committor of the survival process is given by

$$q_i^+ = \mathbb{P}(\tau_A^+(n) > \tau_B^+(n) \mid Z_n = i) = \lim_{m \to \infty} \mathbb{P}(\tau_A^+(n) > \tau_B^+(n) \mid X_n = i, \tau_\omega > m) \qquad (3.25)$$

and solves

$$\begin{cases} q_i^+ &= \sum\limits_{j \in \mathbb{X}} P_{Z,ij}\, q_j^+ = \sum\limits_{j \in \mathbb{X}} \lambda^{-1} Q_{ij} \frac{w_j}{w_i}\, q_j^+ & i \in \mathbb{X} \setminus (A \cup B) \\ q_i^+ &= 0 & i \in A \\ q_i^+ &= 1 & i \in B. \end{cases} \qquad (3.26)$$

The backward committor becomes

$$q_i^- = \mathbb{P}(\tau_A^-(n) > \tau_B^-(n) \mid Z_n = i) = \lim_{m \to \infty} \mathbb{P}(\tau_A^-(n) > \tau_B^-(n) \mid X_n = i, \tau_\omega > m) \qquad (3.27)$$

and solves

$$\begin{cases} q_i^- &= \sum\limits_{j \in \mathbb{X}} P_{Z,ij}^-\, q_j^- = \sum\limits_{j \in \mathbb{X}} \lambda^{-1} \frac{v_j}{v_i} Q_{ji}\, q_j^- & i \in \mathbb{X} \setminus (A \cup B) \\ q_i^- &= 0 & i \in B \\ q_i^- &= 1 & i \in A. \end{cases} \qquad (3.28)$$

The equations for the transition statistics follow using the stationary distribution $(w_i v_i)$ of the survival chain, e.g., the distribution of reactive trajectories is given by $w_i v_i q_i^- q_i^+$.

3.2.2 TPT for the mortal process

The approach we have to take for the Markov chain that has not been absorbed by time $n = 0$ is a bit more advanced due to being time-inhomogeneous. Since the mortal chain is conditioned on no absorption until time 0 it is more natural to ask questions about what happens thereafter. We focus on the reactive trajectories that happen for times $n \geq 0$ but will still discuss the committors for the whole time frame $\mathbb{Z}$.

The equations for the forward and backward committor for time-dependent Markov chains are given by (2.9) resp. (2.10). We are interested in the time frame $\mathbb{Z}$, similar as in Section 2.4.5. The committors of $(Y_n)_{n \in \mathbb{Z}}$,

$$q_i^+(n) = \mathbb{P}(\tau_A^+(n) > \tau_B^+(n) \mid Y_n = i) = \mathbb{P}(\tau_A^+(n) > \tau_B^+(n) \mid X_n = i, \tau_\omega > 0), \qquad (3.29)$$

$$q_i^-(n) = \mathbb{P}(\tau_A^-(n) > \tau_B^-(n) \mid Y_n = i) = \mathbb{P}(\tau_A^-(n) > \tau_B^-(n) \mid X_n = i, \tau_\omega > 0), \qquad (3.30)$$

are defined on $\mathbb{X} \cup \omega$. But we can reduce the committor problem to $\mathbb{X}$ since transitions via ω are impossible. First of all, we know that $q_\omega^+(n) = 0$ since from ω it is no longer possible to hit A or B and therefore $\mathbb{P}(\tau_A^+(n) > \tau_B^+(n) \mid Y_n = \omega) = \mathbb{P}(\infty > \infty \mid Y_n = \omega) = 0$. Consequently, the state ω does not affect any other states regarding hitting A or B in the future and the forward committor equation can be reduced to an equation on $\mathbb{X}$. Secondly, even though $q_\omega^-(n) > 0$ since it is possible in backward time to reach A or B from ω, the remaining committor values $q_i^-(n), i \in \mathbb{X}$ are not influenced by $q_\omega^-(n)$ since there is a zero probability of being mapped into ω backwards in time due to no escape from ω. Thus setting $q_\omega^-(n) = 0$ leaves the committor problem unaffected and additionally also the density and flux of reactive trajectories. By setting it to zero, also the backward committor equation is reduced to $\mathbb{X}$.

By plugging the forward transition matrices into the time-dependent committor problem, the equation for the forward committor becomes

$$
\begin{cases}
q_i^+(n) & = \; \sum\limits_{j \in \mathbb{X}} Q_{ij}\, q_j^+(n) & i \in \mathbb{X} \setminus (A \cup B), n \geq 0, \\
q_i^+(n) & = \; \sum\limits_{j \in \mathbb{X}} Q_{ij} \frac{\sum_{k \in \mathbb{X}}(Q^{|n|-1})_{jk}}{\sum_{k \in \mathbb{X}}(Q^{|n|})_{ik}}\, q_j^+(n+1) & i \in \mathbb{X} \setminus (A \cup B), n < 0, \\
q_i^+(n) & = \; 0 & i \in A, \\
q_i^+(n) & = \; 1 & i \in B.
\end{cases}
\tag{3.31}
$$

For $n \geq 0$ the equation is essentially stationary due to the time-independent forward matrix and since the terminal condition lies at infinity. Since the forward transition matrix becomes time-dependent for $n < 0$, the forward committor becomes time-dependent for $n < 0$ with terminal condition given by the stationary solution of $n \geq 0$. We denote the stationary committor for $n \geq 0$ by q^+ due to the time-independence.

The equation for the backward committor is stationary on the whole time frame $\mathbb{Z}$, we have

$$
\begin{cases}
q_i^- & = \; \sum\limits_{j \in \mathbb{X}} \lambda^{-1} \frac{v_j}{v_i} Q_{ji}\, q_j^- & i \in \mathbb{X} \setminus (A \cup B) \\
q_i^- & = \; 0 & i \in B \\
q_i^- & = \; 1 & i \in A.
\end{cases}
\tag{3.32}
$$

We now want to understand the transitions for $n \geq 0$. The interesting observation is that for $n \geq 0$ the distribution of reactive trajectories as well as the current of reactive trajectories only decrease by a constant factor for larger and larger times. The distribution of reactive trajectories for $i \in \mathbb{X}, n \geq 0$ is given by

$$
\mu_i^{AB}(n) = \lambda^n v_i q_i^+ q_i^-
\tag{3.33}
$$

while the current for $i, j \in \mathbb{X}, n \geq 0$ is given by

$$
f_{ij}^{AB}(n) = \lambda^n v_i q_i^- Q_{ij} q_j^+.
\tag{3.34}
$$

3.2. *Transitions before absorption*

Therefore,

$$\frac{\mu_i^{AB}(n+1)}{\mu_i^{AB}(n)} = \lambda < 1 \text{ and } \frac{f_{ij}^{AB}(n+1)}{f_{ij}^{AB}(n)} = \lambda < 1 \tag{3.35}$$

and the factor determining the reduction of reactive trajectories is given by λ. For large n it becomes more and more unlikely that the Markov chain has not been absorbed in ω and can thus still transition from A to B, therefore eventually both the distribution and the current of reactive trajectories will converge to 0. But by normalizing the distribution and current of reactive trajectories, both quantities remain constant for all $n \geq 0$.

The *expected number of transitions* that are on average started and completed for $n \geq 0$ is given by

$$\sum_{n \geq 0} k^{A \rightarrow}(n) = \frac{1}{1-\lambda} \sum_{\substack{i \in A \\ j \in \mathbb{X}}} \nu_i Q_{ij} q_j^+. \tag{3.36}$$

This quantity can be compared to the *expected number of time steps before absorption* for $n \geq 0$ which is the following sum of the probabilities that at time step n the chain is still not absorbed

$$\mathbb{E}(\tau_\omega - 1 \mid Y_0 \sim \nu) = \sum_{n \geq 0} \mathbb{P}(Y_n \in \mathbb{X} \mid Y_0 \sim \nu) = \sum_{\substack{i,j \in \mathbb{X} \\ n \geq 0}} \nu_i \, (Q^n)_{ij}$$

$$= \sum_{\substack{j \in \mathbb{X} \\ n \geq 0}} \lambda^n \nu_j = \sum_{n \geq 0} \lambda^n = \frac{1}{1-\lambda}. \tag{3.37}$$

In conclusion we have shown two approaches for studying transitions prior to absorption. Depending on the application, one or the other might give a more suitable description. Both approaches result in stationary committor functions for the interesting time frame. While the backward committors are the same for both approaches, the forward committors differ since they allow for absorption in the future or not.

3.2.3 Example

We conclude this chapter with a simple example. We consider a Markov chain (Fig. 3.2) which on $\mathbb{X} = \{A, 1, 2, B\}$ is aperiodic and irreducible and which is ultimately absorbed in ω. The states A and B are completely symmetric. From states 1 and 2 the probabilities of going to A to B are balanced. But from state 2 it is additionally possible with probability $1/4$ to reach the absorbing state ω and with probability $1/4$ to stay in the state for one more time step.

First, we consider the transition behaviour of the survival process $(Z_n)_{n \in \mathbb{Z}}$. The survival process has the following stationary distribution

$$(\nu_i w_i)_{i \in \mathbb{X}} = (0.23, 0.28, 0.26, 0.23)$$

(states are ordered as $A, 1, 2, B$). The probability of finding the process in A and B is
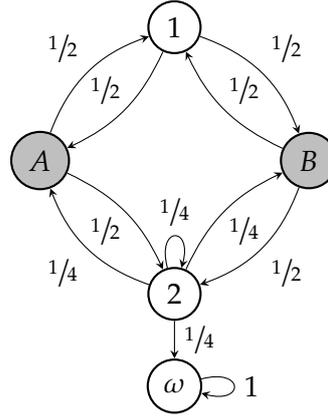
Figure 3.2: Transition probabilities of a Markov chain with unbalanced transition paths due to an absorbing state.

the same, which was to be expected due to the symmetry. The probability of finding the survival process in 1 is slightly larger than in 2. The process is thus biased to the state 1 from which it is less likely that the original process will soon transition to ω. The committors are given by

$$q^+ = (0, 1/2, 1/2, 1) \text{ and } q^- = (1, 1/2, 1/2, 0).$$

The committor probabilities agree on the states 1 and 2 since the probability of reaching A or B forward or backward in time is just $1/2$ when neglecting the possibility of ever hitting ω. Consequently, the normalized distribution of reactive trajectories is given by

$$\hat{\mu}^{AB} = (0, 0.51, 0.49, 0)$$

which is proportional to the stationary distribution in C. It is more likely to find a reactive trajectory in the state 1 than 2. When in A, the probability of a reactive trajectory to take the route via 1 is given by $f_{A1}^{AB}/\left(\sum_j f_{Aj}^{AB}\right) = 0.59$ compared to $f_{A2}^{AB}/\left(\sum_j f_{Aj}^{AB}\right) = 0.41$ via state 2. Last, the rate of transitions is given by $k^{AB} = 0.12$ implying that a transition is started on average every 8.61th time step with a mean duration of $t^{AB} = 2.15$.

Now we come to the mortal process conditioned on staying in $\mathbb{X}$ until time 0. Its distribution when $n \geq 0$ is given by

$$(\lambda^n v_i)_{i \in \mathbb{X}} = \lambda^n (0.22, 0.24, 0.32, 0.22)$$

where $\lambda = 0.92$. Thus contrary to the survival process, the probability of finding the process in 2 is higher than finding the process in state 1. This process considers all trajectories but those that have left $\mathbb{X}$ before time 0. It is more likely to find the process in state 2 than 1 which agrees with the intuition that the distribution should be higher in state 2 due to the possibility of self-transitions. Next, we consider the transitioning behaviour from A to B for times $n \geq 0$. On average the mortal process stays in $\mathbb{X}$ for

$\frac{1}{1-\lambda}$ = 12.25 time steps before being absorbed. The expected number of transitions that are started and completed during $n \geq 0$ is given by 1.11. The backward committor is the same as before

$$q^- = (1, 0.5, 0.5, 0)$$

but the forward committor is now given by

$$q^+ = (0, 0.5, 1/3, 1)$$

due to the possibility of reaching ω from state 2. Therefore the resulting normalized distribution of reactive trajectories is again higher in state 1 than 2,

$$\hat{\mu}^{AB} = (0, 0.52, 0.48, 0).$$

When in A, the probability that a reactive trajectory takes the route along 1 is given by 0.6, while a reactive trajectory along 2 is taken with probability 0.4. This agrees with the intuition that along state 1 it is more likely that a trajectory reaches B, in particular, when in state 1 the probability to reach B rather than A is 1/2, while from state 2 the probability to reach B rather than $A \cup \omega$ is given by $\sum_{n>0} (1/4)^n = 1/3$.

This example demonstrates that even though the two modelled processes at first have very different distributions, the properties of their reactive trajectories (when normalized) can be rather similar.

4 | Transition sampling and pathway analysis

We have seen that Transition Path Theory provides us with useful statistical properties of the transitions from a chosen source to a sink. In this chapter we are concerned with getting a more detailed understanding of the transition pathways.

To advance this goal, we will first in Section 4.1 explain how we can directly sample reactive trajectories of stationary Markov chains when the forward committor probabilities are known. A large ensemble of sampled reactive trajectories can form the basis of further analysis, e.g., for extracting pathways or deriving more application-specific transition statistics.

Secondly, in Section 4.2, we will explain and illustrate two methods from the literature for extracting the dominant transition paths connecting the source with the sink. Especially in high-dimensional systems it can be difficult to comprehend the global information about the pathways of reactive trajectories as provided by the current of reactive trajectories. Individual paths, on the other hand, can often still be understood. It is important though to consider more than just the most likely path to get a representative understanding of the dynamics. We will present two methods with a different focus. The first method [47] studies the dominant cycle-free, and thus less complicated, transition paths of reversible, stationary dynamics while the second approach [6] is especially targeted at non-reversible dynamics and separately considers the cycle-free paths and the cyclic structures. But this happens at the expense of the computational efficiency since a large batch of sampled reactive trajectories has to be analysed. In Chapter 6 we will need the latter method to study the dominant tipping paths in agent-based models and to gain a better understanding of the cyclic structures during transitions.

4.1 The Sisyphus chain

We consider the original Markov chain $(X_n)_{n \in \mathbb{Z}}$ on the finite state space $\mathbb{X}$ to be of the type described in 2.10, in particular the chain is stationary with distribution π and transition matrix P. Our interest lies in sampling statistically exact reactive trajectories from a chosen source $A \subset \mathbb{X}$ to a sink $B \subset \mathbb{X}$. Related constructions that sample inner reactive trajectories can be found in the literature for stationary, reversible Markov

jump processes [11] and chains [71]. We will present a construction that additionally considers the last visited state in A before entering the transition region and the first state in B and is valid for non-reversible chains.

We will achieve our goal by constructing a transition matrix P^S that recurrently samples reactive trajectories. This means in particular that the stationary distribution and the current of P^S should coincide with the reactive distribution resp. reactive current of P on the subset of states visited by reactive trajectories. The transition matrix P^S will map the chain from A via C to B and thereby correctly sample reactive trajectories from A via C to B. The modelled process then stays in B for a long time to account for the non-reactive part of the original trajectory, before being mapped back to a state in A and going on another transition. Mapping all the non-reactive parts of the trajectory onto B is needed to get a correct weighting of the reactive and the non-reactive parts of the trajectory. Due to the recurrent motion of the constructed chain from A to B and again from A to B, the Markov chain will be called *Sisyphus chain* and the transition matrix P^S will be named *Sisyphus matrix*. As an illustrative example we show a realization of a random walk on a chain of four boxes, as well as the corresponding Sisyphus trajectory in Fig. 4.1. During the reactive phase both trajectories agree, but during the non-reactive phase the Sisyphus chain stays in B while the original trajectory can wander around.
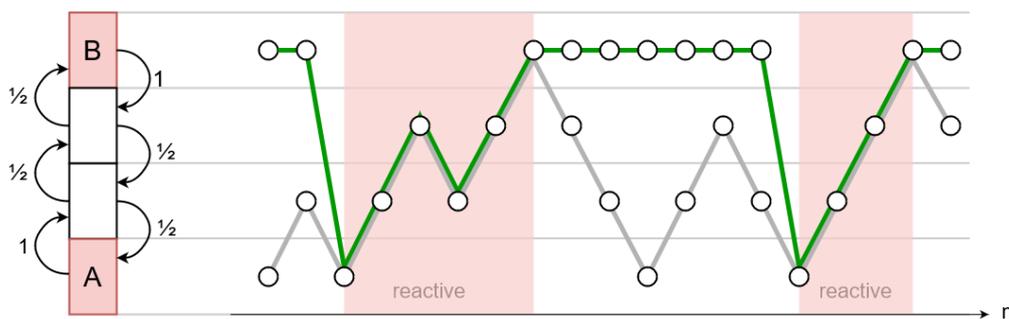
Figure 4.1: We give a simple example of a trajectory (in grey) and the extracted Sisyphus trajectory (in green) that stays in B while the original trajectory is non-reactive.

To construct the Sisyphus matrix, we first have to restrict the state space $\mathbb{X}$ to states that are visited by reactive trajectories. Since not all states in A are ever the starting state of a reactive trajectory and not all states in B are potential last states, we only consider the subset of A with positive outflow of reactive trajectories

$$\tilde{A} := \{i \in A \text{ s.t. } k_i^{A\rightarrow} > 0\} \subseteq A, \tag{4.1}$$

where $k_i^{A\rightarrow} := \sum_{j\in\mathbb{X}} f_{ij}^{AB}$ is the current of reactive trajectories out of a state $i \in A$, and similarly we consider the subset of B with positive inflow of reactive trajectories

$$\tilde{B} := \{i \in B \text{ s.t. } k_i^{\rightarrow B} > 0\} \subseteq B, \tag{4.2}$$

where $k_i^{\to B} := \sum_{j \in \mathbb{X}} f_{ji}^{AB}$ gives the current of reactive trajectories into $i \in B$. Further we reduce the set C to the set of states that can be visited by a reactive trajectory, i.e.,

$$\tilde{C} := \{i \in C \text{ s.t. } q_i^+ q_i^- > 0\} \subseteq C. \tag{4.3}$$

Then we can state the main result of this section.

Theorem 4.1. *We assume a stationary Markov chain as stated in Assumption 2.10 with transition quantities q^+, k^{AB}, H^{AB}, π^{AB} and f^{AB} for chosen sets A and B. Then the transition matrix*

$$P_{ij}^S = \begin{cases} \dfrac{P_{ij} q_j^+}{\sum_{k \in \tilde{\mathbb{X}}} P_{ik} q_k^+}, & i \in \tilde{A}, j \in \tilde{C} \cup \tilde{B} \\[2.5ex] P_{ij} \dfrac{q_j^+}{q_i^+}, & i \in \tilde{C}, j \in \tilde{C} \cup \tilde{B} \\[2.5ex] 1 - \dfrac{k^{AB}}{1-H^{AB}}, & i = j \in \tilde{B} \\[2.5ex] \dfrac{k_j^{A\to}}{1-H^{AB}}, & i \in \tilde{B}, j \in \tilde{A} \\[2.5ex] 0, & else \end{cases} \tag{4.4}$$

on the state space $\tilde{\mathbb{X}} = \tilde{C} \cup \tilde{A} \cup \tilde{B}$ recurrently samples statistically exact reactive trajectories from $\tilde{A}$ to $\tilde{B}$. In particular the stationary distribution of P^S coincides on $\tilde{C}$ with π^{AB} and its current coincides on pairs of states $i \in \tilde{A} \cup \tilde{C}, j \in \tilde{C} \cup \tilde{B}$ with f_{ij}^{AB}. Additionally, the process stays in $\tilde{B}$ to account for the non-reactive parts of the original Markov chain and is then mapped back to $\tilde{A}$.

Proof. It is straightforward to check that P^S is a stochastic matrix with non-negative entries and rows summing to 1 by using the committor equations from Theorem 2.1, the relation from Proposition 2.15 and that $k^{AB} = \sum_{i \in \tilde{A}} k_i^{A\to} = \sum_{i \in \tilde{B}} k_i^{\to B}$.

We proceed by confirming that the stationary distribution of P^S is given by the following probability distribution

$$\pi_i^S = \begin{cases} \pi_i^{AB} & i \in \tilde{C} \\ k_i^{A\to} & i \in \tilde{A} \\ k_i^{\to B}(1 - H^{AB})(k^{AB})^{-1} & i \in \tilde{B}, \end{cases} \tag{4.5}$$

i.e., that the distribution is invariant under the action of P^S: $(\pi^S)^T P^S = (\pi^S)^T$. We can check the different entries separately: For $j \in \tilde{C}$, it holds that

$$\sum_{i \in \tilde{\mathbb{X}}} \pi_i^S P_{ij}^S = \sum_{i \in \tilde{A} \cup \tilde{C}} \pi_i q_i^- P_{ij} q_j^+ = \pi_j q_j^- q_j^+ = \pi_j^{AB} \tag{4.6}$$

by using that $\pi_i P_{ij} = \pi_j P_{ji}^-$ and the backward committor equation. The only inflow into

$j \in \tilde{A}$ comes from states $i \in \tilde{B}$, therefore:

$$\sum_{i \in \tilde{\mathbb{X}}} \pi_i^S P_{ij}^S = k_j^{A \to} \left(1 - H^{AB}\right)^{-1} \sum_{i \in \tilde{B}} k_i^{\to B}(1 - H^{AB})(k^{AB})^{-1} = k_j^{A \to}. \tag{4.7}$$

Last, the probability mass mapped into $j \in \tilde{B}$ can be rewritten as the inflow from the same state $i = j$ and from states $i \in \tilde{A} \cup \tilde{C}$

$$
\begin{aligned}
\sum_{i \in \tilde{\mathbb{X}}} \pi_i^S P_{ij}^S &= \left(1 - H^{AB} - k^{AB}\right)\left(1 - H^{AB}\right)^{-1} k_j^{\to B}(1 - H^{AB})(k^{AB})^{-1} \\
&\quad + \sum_{i \in \tilde{C}} \pi_i^{AB} P_{ij} \frac{q_j^+}{q_i^+} + \sum_{i \in \tilde{A}} k_i^{A \to} P_{ij} q_j^+ \left(\sum_{k \in \tilde{\mathbb{X}}} P_{ik} q_k^+\right)^{-1} \\
&= \left(1 - H^{AB} - k^{AB}\right) k_j^{\to B}(k^{AB})^{-1} + \sum_{i \in \tilde{A} \cup \tilde{C}} q_i^- \pi_i P_{ij} q_j^+ \\
&= \left(1 - H^{AB} - k^{AB}\right) k_j^{\to B}(k^{AB})^{-1} + k_j^{\to B},
\end{aligned}
\tag{4.8}
$$

which on $\tilde{B}$ is equal to π^S.

We are left to confirm that the current of P^S is given by

$$f_{ij}^S = \begin{cases} f_{ij}^{AB} & i \in \tilde{A} \cup \tilde{C}, j \in \tilde{C} \cup \tilde{B} \\ k_i^{\to B}\left(1 - H^{AB} - k^{AB}\right)(k^{AB})^{-1} & i = j \in \tilde{B} \\ k_i^{\to B} k_j^{A \to}(k^{AB})^{-1} & i \in \tilde{B}, j \in \tilde{A}. \end{cases} \tag{4.9}$$

A simple calculation shows that f^S sums to 1, f^S is non-negative since $1 - H^{AB} - k^{AB}$ gives the probability of not being on a reactive trajectory. To verify the current, we simply have to plug in the formulas (4.4), (4.5) and (4.9) into $f_{ij}^S = \pi_i^S P_{ij}^S$ to see that it holds. $\qquad\square$

In the remainder of this section, we will discuss three questions: (i) What is the structure of the Sisyphus matrix? (ii) What are the properties of the Sisyphus matrix? (iii) In which ways can we extend the construction?

What is the structure of the Sisyphus matrix? Let us assume that we start the Sisyphus chain in $\tilde{A}$, then the next state lies in $\tilde{B} \cup \tilde{C}$ and is taken with a probability that is proportional to the forward committor of the next state. When in the transition region $\tilde{C}$, the probability to transition to another state in $\tilde{B} \cup \tilde{C}$ is proportional to the ratio of the forward committors of the next versus the current state. Thus the Sisyphus chain will move along a path with mostly increasing forward committor values until it reaches $\tilde{B}$.

When the Sisyphus chain has arrived in $\tilde{B}$, the probability of going back to any state in $\tilde{A}$ is given by the ratio $k^{AB}/(1 - H^{AB})$. This ratio indeed gives the probability

of becoming reactive since its inverse $(1 - H^{AB})/k^{AB}$ gives the *expected duration of the non-reactive parts* of the original trajectory.[1] Consequently the transition probability from $\tilde{B}$ to a specific state in $\tilde{A}$ is the probability of become reactive but weighted with the probability to start a transition from this particular state in $\tilde{A}$.

What are the properties of the Sisyphus matrix? The constructed matrix is non-reversible due to the directed flow from $\tilde{A}$ to $\tilde{B}$ and back to $\tilde{A}$. In particular the fluxes over the boundaries of $\tilde{A}$ and $\tilde{B}$ are not balanced.

Due to the following three properties one can reach any state $j \in \tilde{\mathbb{X}}$ from any state $i \in \tilde{\mathbb{X}}$ in a finite number of steps, hence the matrix P^S is irreducible:

(i) every state in $\tilde{C}$ is reachable from some state in $\tilde{A}$, and every state in $\tilde{A}$ leads to some state in $\tilde{C} \cup \tilde{B}$,

(ii) from every state in $\tilde{C}$ we can reach some state in $\tilde{B}$, and each state of $\tilde{B}$ is reachable from some state in $\tilde{C}$,

(iii) from any state in $\tilde{B}$ we can reach any state in $\tilde{A}$ (there is mixing).

In which ways can we extend the construction? Usually one is interested in the reactive parts of the Sisyphus chain only. By slightly changing P^S such that a new reactive trajectory is immediately restarted upon reaching $\tilde{B}$, i.e., by setting

$$
\begin{aligned}
P^S_{ii} &= 0 && \text{for} && i \in \tilde{B}, \\
P^S_{ij} &= \frac{k^{A\to}_j}{k^{AB}} && \text{for} && i \in \tilde{B}, j \in \tilde{A},
\end{aligned}
\tag{4.10}
$$

one can sample reactive trajectories without a long pause in $\tilde{B}$. It can easily be shown that the stationary distribution of this adapted Sisyphus matrix is given by

$$
\pi^S_i = \begin{cases}
\pi^{AB}_i \, (H^{AB} + k^{AB})^{-1}, & i \in \tilde{C} \\
k^{A\to}_i \, (H^{AB} + k^{AB})^{-1}, & i \in \tilde{A} \\
k^{\to B}_i \, (H^{AB} + k^{AB})^{-1}, & i \in \tilde{B}.
\end{cases}
\tag{4.11}
$$

In the case that the original Markov chain is reversible, one can also construct a transition matrix that samples cycle-free reactive trajectories [11, 71], i.e., reactive trajectories that do not self-intersect on their way from A to B. The trick is to construct a transition matrix whose current is given by the effective current which is free of cycles.

Last, what can we say about the construction for non-stationary Markov chains? It is possible but technical to write down the Sisyphus matrix for time-dependent dynamics, therefore we left it out here. The construction goes along the same lines by replacing stationary quantities by their time-dependent counter parts. Additionally the sets $\tilde{A}, \tilde{B}, \tilde{C}$ become time-dependent, which further complicates the construction.

---

[1]This follows from a similar reasoning as for the expected duration of reactive trajectories in Eq. (2.44).

In the case of finite-time dynamics, an initial distribution of reactive trajectories has to be derived and used for sampling reactive trajectories in combination with the time-dependent Sisyphus matrix.

4.2 Algorithms for pathway-finding

In this second part of the chapter we are reviewing two methods for extracting information about the dominant paths from A to B in stationary chains. We are again assuming a Markov chain as in Assumption 2.10. The first approach by Metzner et al. [46, 47] searches for the dominant cycle-free paths from A to B. The second approach by Banisch et al. [6] additionally studies the most important cycles that are visited during transitions.

We already used the word *cycle* and also *cycle-free path*, but what exactly do we mean by the terms? As a cycle-free path we consider a path $\gamma = (i_1, \ldots, i_s)$ with $i_k \in \mathbb{X}$ that is non-intersecting, i.e., all traversed states i_k are pairwise different. A cycle on the other hand is a path $\gamma = (i_1, \ldots, i_s, i_1)$ with all i_k being pairwise different and that in the end comes back to its starting state. We consider a cycle as an equivalence class containing all cyclic permutations of the cycle. Note that self-cycles $\gamma = (i, i)$, i.e., paths that stay in i, are also considered as cycles.

4.2.1 Path decomposition for reversible, stationary chains

Metzner et al. [46, 47] has introduced an algorithm that decomposes the effective current of a reversible Markov chain to find the dominant cycle-free paths that start in A and end in B. Even though reactive trajectories in reversible Markov chains can contain cycles, the cycles do not lead to progress along the transition. The effective current f^+, which gives the net amount of current of reactive trajectories along an edge, is free of cycles for reversible Markov chains and therefore a good starting point for the search for simple and meaningful transition paths that are freed of cycles.

The decomposition builds on an iterative search for the most dominant reactive paths and the associated amount of current that each can carry. The crucial assumption of this algorithm is that the current along a path γ is restricted by the minimal current value of an edge (i, j) along the path, i.e., by $c(\gamma) := \min_{(i,j)\in\gamma}\{f_{ij}^+\}$. Then, paths γ are considered as more dominant, if they can potentially carry more current of reactive trajectories, equivalently, if their $c(\gamma)$ value is larger.

The method can be summarized as follows: Given the effective current f^+, the algorithm starts by searching for the paths γ from A to B with the largest minimal current value along any edge of that path, i.e., the largest $c(\gamma)$. These paths will be called *dominant*. For simplicity it is assumed that all all entries of f^+ are pairwise different. Then if there are several dominant paths, they will contain the same edge with maximal minimal current value, this edge will be called the *bottleneck*. To define a unique dominant path, one searches for the path that is also dominant from A to the

bottleneck edge and from the bottleneck edge to B. If these are also not unique, then this procedure can be continued recursively until a unique representative of the dominant paths is found. This path will be denoted by γ_D^1 and the amount of current $c(\gamma_D^1)$ is assigned to it. The second most dominant path can be found by searching for a representative of the dominant paths in the residuum effective current

$$f_{ij}^{+,1} := \begin{cases} f_{ij}^+ - c(\gamma_D^1) & (i,j) \in \gamma_D^1 \\ f_{ij}^+ & (i,j) \notin \gamma_D^1, \end{cases} \tag{4.12}$$

resulting from subtracting $c(\gamma_D^1)$ from the effective current along edges of the path γ_D^1. This procedure can be iterated to fully decompose f^+ into representative dominant paths carrying portions of the current.

Pseudo-code and more details on the algorithm can be found in [46, 47].

4.2.2 Decomposition for non-reversible, stationary chains

For stationary Markov chains that are non-reversible, cyclic structures are an important feature. Therefore in [6] it is proposed to decompose the current of reactive trajectories f^{AB} into the current carried by cycle-free paths that start in A and end in B and are therefore called *productive* for the transition and into the current carried by cycles that are per se *unproductive* for the undertaking of the transition. For the decomposition a large ensemble of reactive trajectories is first split into cycle-free paths from A to B and into unproductive cycles, compare with Fig. 4.2, and then analysed with regard to their frequency.
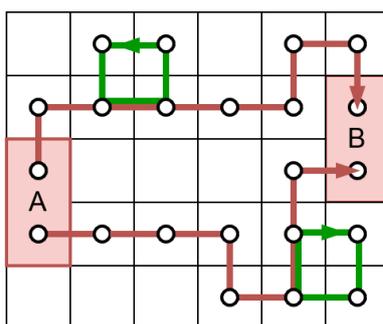
Figure 4.2: Two reactive trajectories transitioning from A to B. Each trajectory is split into a productive path (in red) and unproductive cycles (in green).

We denote by Γ^P the set of all non-intersecting paths that start in A, end in B, and in-between pass through the transition region C. By Γ^U we denote the set of cycles through the transition region C. Now we are equipped to state the result. The current of reactive trajectories can be split as follows into the current coming from all cycle-free

productive paths Γ^P and the current induced by the set of unproductive cycles Γ^U [6]

$$f^{AB} = \underbrace{\sum_{\gamma \in \Gamma^P} w^\gamma \, C^\gamma}_{=:f^P} + \underbrace{\sum_{\gamma \in \Gamma^U} w^\gamma \, C^\gamma}_{=:f^U}, \tag{4.13}$$

where C^γ is the incidence matrix of the path γ

$$C_{ij}^\gamma = \begin{cases} 1, & \text{if } \gamma = (\dots, i, j, \dots) \\ 0, & \text{else} \end{cases} \tag{4.14}$$

and w^γ encodes the associated path weight, i.e., the relative frequency of visits of a reactive trajectory to the path or cycle γ. Thereby the edges of γ have to be passed in the right order but excursions to other cycles or paths in-between are allowed. Given an infinitely long ergodic trajectory $(X_n)_{n \in \mathbb{N}}$ of the Markov chain, we have the following almost sure convergence

$$w^\gamma = \lim_{N \to \infty} \frac{W_N^\gamma}{N}, \tag{4.15}$$

where W_N^γ counts the number of times that $(X_n)_{n=0,\dots,N-1}$ passes through γ while also being reactive and allowing for interrupting visits to other paths or cycles.

This decomposition not only allows an evaluation of the dominant productive paths and most important unproductive cycles by considering the paths resp. cycles with the largest weights w^γ but also enables a separate consideration of the productive and the unproductive current.

Remark 4.2 (Derivation). *The original derivation in [6] proceeds as follows. They consider the current of the Sisyphus matrix, f^S, which from $\tilde{A}$ to $\tilde{B}$ agrees with the current of reactive trajectories, compare with Theorem 4.1. The **cycle decomposition** (Theorem 1.17) can be applied to uniquely decompose the current f^S into weighted cycles solely in $\tilde{C}$, weighted self-cycles in $\tilde{B}$ and weighted cycles that contain edges from $\tilde{B}$ to $\tilde{A}$. The cycles solely in C correspond to Γ^U while the cycles containing edges from $\tilde{B}$ to $\tilde{A}$ can be identified with the productive paths Γ^P upon deleting the edges from $\tilde{B}$ to $\tilde{A}$. The result in Eq. (4.13) follows.*

Since the Sisyphus chain from Theorem 4.1 agrees with the original chain while being reactive, the convergence in (4.15) also holds by counting the passages through γ in a realization of the Sisyphus chain. Therefore, the easiest way to numerically estimate the decomposition is as follows [5]:

1. Sample a long ergodic trajectory of the Sisyphus chain $(x_n^S)_{n=0,\dots,N-1}$ according to Theorem 4.1 that contains sufficiently many transitions from A to B. Since we only need the reactive trajectory pieces for the computation of (4.15) but correctly weighted compared to the non-reactive pieces, we can adapt P^S as in Eq. (4.10) and immediately start a new reactive trajectory once $\tilde{B}$ is reached but reweigh the found reactive trajectories correctly with $H^{AB} + k^{AB}$.

2. Estimate $w(\gamma)$ by averaging along this sampled trajectory as follows: First prune out all the reactive pieces. Then for each reactive snippet iteratively cut out all the cycles by going through the trajectory until for the first time a state is revisited, i.e., until we find r such that $x_r^S = x_m^S$, $m < r$. Take out the cycle $(x_m^S, \ldots x_{r-1}^S) = \gamma$ and increment W_N^γ by 1. If the Sisyphus chain is sampled with immediate restarting once $\tilde{B}$ is reached, we instead have to increment W_N^γ by $H^{AB} + k^{AB}$. Repeat until from the reactive snippet only a cycle-free transition path γ is left, increment W_N^γ accordingly by 1 resp. by $H^{AB} + k^{AB}$. Then move on to the next reactive trajectory piece.

Remark 4.3 (Pathways in time-dependent dynamics). *Finding important transition pathways is much more complicated in the case of underlying time-dependent dynamics as described in Sections 2.3 and 2.4. The reactive trajectories that start in A at different time points differ statistically and ideally should be analysed separately. The most general framework would consider space-time paths from A to B, which are naturally free of cycles since they only flow forward in time but can be of a rather complicated form. Also the number of different space-time paths from A to B is much larger than the number of spatial paths in a stationary chain. In the case that the timing of a reactive trajectory is not of interest, one can also study the space-time paths from A to B projected onto space, i.e., neglecting information about their starting time.*

Can we adapt the proposed algorithms for finding pathways in time-dependent dynamics? The space-time current $f^{AB}(n)$ is cycle-free and obeys Kirchhoff's law except at the source and sink states. Therefore Metzner's algorithm can be applied to iteratively find the dominant space-time pathways that have the largest minimum current value. But the approach might become expensive due to the large size of space-time. We already mentioned that space-time paths from A to B do not contain cycles, thus a decomposition into cycles and cycle-free paths similar as in Eq. (4.13) is not very useful. Still, one can sample reactive trajectories by extending Theorem 4.1 and analyse them regarding the paths that are dominantly taken.

4.2.3 Illustration of the two algorithms on simple examples

We will illustrate the two decomposition algorithms on a reversible and on a non-reversible Markov chain.

A reversible Markov chain We consider the reversible Markov chain in Fig. 4.3(a) where states A and B and states 1 and 2 are symmetric with respect to the dynamics. The current of reactive trajectories from state A to state B is shown in Fig. 4.3(b). The effective current agrees with the current of reactive trajectories except along the edges between states 1 and 2 where the currents effectively cancel.

The algorithm by Metzner et al. decomposes the effective current into two paths: $(A, 1, B)$ and $(A, 2, B)$ that each carry a portion of 0.05 of the effective current. Here the algorithm gives all the paths that are possible from the effective current. The decomposition by Banisch et al., on the other hand, finds additional paths and also
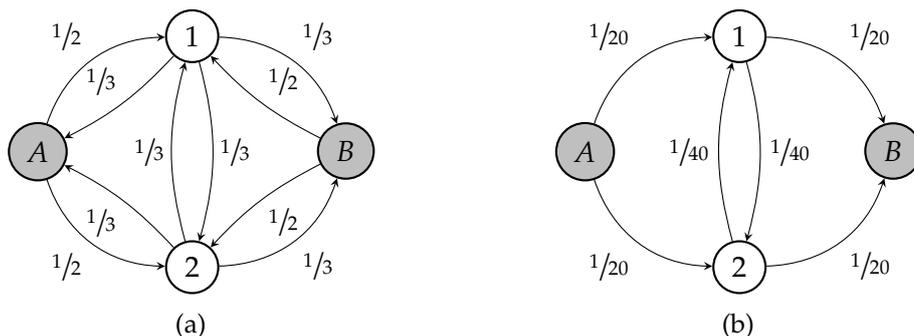
Figure 4.3: Reversible Markov chain with (a) transition probabilities, and (b) f^{AB} given along the edges.

cycles since it considers the current of reactive trajectories instead. The algorithm finds the following productive paths from A to B:

$(A, 1, B)$ and $(A, 2, B)$ each with weight 0.0375

$(A, 1, 2, B)$ and $(A, 2, 1, B)$ each with weight 0.0125

as well as the cycle

$(1, 2, 1)$ with weight 0.0125.

Even though the algorithm by Metzner et al. misses the cycle $(1, 2, 1)$ as well as the two cycle-free paths $(A, 1, 2, B)$ and $(A, 2, 1, B)$, in the case of larger Markov chains this approach might be very useful. The found paths already might help in understanding the mechanism of transitions from A to B better. The decomposition by Banisch et al. gives us a complete picture of the different paths and cycles during transitions.

A non-reversible Markov chain   At next we study a non-reversible Markov chain, where Metzner's decomposition approach is no longer applicable. Cycles play an important role in non-reversible processes. We consider a Markov chain that consists of 5 states. The transition probabilities are shown in the Fig. 4.4(a) and the current of reactive trajectories is given in Fig. 4.4(b).
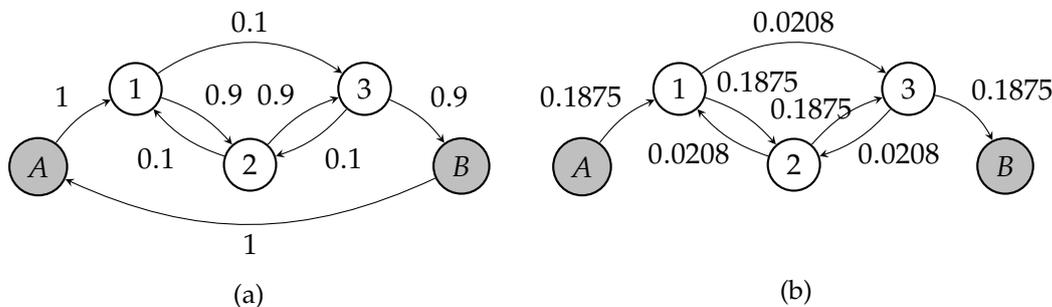
Figure 4.4: Non-reversible Markov chain with (a) transition probabilities, and (b) f^{AB} given along the edges.

Applying the decomposition into productive paths and unproductive cycles, we arrive at the following paths from A to B:

$(A, 1, 2, 3, B)$ with weight 0.1669

$(A, 1, 3, B)$ with weight 0.0206

and unproductive cycles

$(1, 3, 2, 1)$ with weight 0.0002

$(1, 2, 1)$ and $(2, 3, 2)$ each with weight 0.0206.

In the case of reversible processes, each cycle has a reversed cycle with the same weight [6] and consequently f^U is symmetric. This is no longer ensured for non-reversible processes, here for example the cycle $(1, 3, 2, 1)$ is only traversed in one direction. This is the reason behind saying that cyclic structures are characteristic of non-reversible processes.

5 | Different tipping mechanisms

In the previous chapters we developed the tools for studying transitions in Markov chains. In this chapter we will review different tipping mechanisms and illustrate how Transition Path Theory can be employed to study them, thereby demonstrating its broad applicability.

Tipping of a dynamics is herein understood as a sudden change of the system state from one attractor to another due to some small quantitative change of a parameter or noise. Building on the large tipping literature, in Ashwin et al [3] they distinguish between three causes of tipping: bifurcation-induced, rate-induced, and noise-induced. A deterministic ordinary differential equation (ODE)

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, \lambda(t))$$

with time-dependent parameter $\lambda(t)$ can be considered as an open system subject to the external variation of λ in time. A constant parameter would in this picture correspond to a closed system.

In this setting, *bifurcation-induced tipping* happens when λ varies infinitesimally slowly and at a certain critical value of λ, *the bifurcation point*, the attractor loses its stability or disappears and the trajectory tips to a different attractor. Not all bifurcations necessarily correspond to tipping, some bifurcations entail a rather smooth qualitative change of the dynamics such as a Hopf bifurcation. In [68] they distinguish between safe, explosive and dangerous bifurcations. Only the dangerous bifurcations correspond to tipping, they are characterized by a sudden and fast jump to a distant attractor and often exhibit hysteresis, i.e., the system does not immediately tip back to the original attractor upon shifting λ back past the bifurcation point.

Rate-induced tipping occurs when the external parameter λ changes at such a high rate that the dynamics can no longer follow the changing attractor and therefore shifts away.

Last, if the system is forced randomly for example by Brownian motion as in the following stochastic differential equation (SDE)

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

then *noise-induced tipping* can take place whereby the noisy fluctuations cause the

system to escape from an attractor. Intuitively, we can also understand this SDE as an open system subject to some external, additive parameter that varies in a random way. Of course, combinations of the above tipping mechanisms are also possible.

Recently other related tipping mechanisms have been classified. In shock-tipping [27] an external shock pushes the system out of the basin of attraction[1], while phase-tipping [1] describes a situation where tipping can only occur during certain phases of a limit cycle or a strange attractor. Additionally, the concept of save overshoots has been defined [57, 58]. When a control parameter only temporarily exceeds the critical bifurcation point, a trajectory that has started to tip might be able to return back to its original attractor. A phenomenon related to tipping is given by the delayed tipping events of canards [20, 37]. Canards are trajectories in a slow-fast system that after reaching a certain critical point follow an unstable slow manifold for a substantial time and only delayed either tip or safely go back to their original attractor. Due to the real-world consequences of tipping such as in the climate system, tipping cascades, i.e., interacting subsystems that can individually tip and influence each other with respect to the onset and likelihood of tipping, have become a focus of research [19, 76].

Since we are interested in the transition dynamics of Markov processes, in the following we will illustrate the three main tipping mechanisms as well as canards under the additional effect of small noise. We will discretize the resulting Markov process in space and time to get a Markov chain and apply Transition Path Theory. Of course it would also be an interesting question to think about how the above concepts of tipping can be defined and extended to Markov chains. Here we follow the more straightforward approach of studying the discretized dynamics of an SDE with respect to tipping.

5.1 Tipping due to noise or external parameter variations

In this first part, we will illustrate the three main types of tipping: noise-induced, bifurcation-induced and rate-induced on the example of a particle whose position $(X_t)_{t \in \mathbb{R}}$ in time is changed according to the following overdamped Langevin equation in 1D

$$dX_t = -\frac{dV}{dx} dt + \sigma dW_t \qquad (5.1)$$

with noise strength $\sigma = 0.5$, see also Section 1.2.2 for an introduction to the overdamped Langevin equation. The described motion of a particle in a potential $V : \mathbb{R} \to \mathbb{R}$ allows a simple physical interpretation of a particle that moves towards the minima of the potential while also being subject to small random forcings in all directions.

In a simple symmetric double well landscape such as given by $V(x) = \frac{1}{4}\left(x^2 - 1\right)^2$ noise-induced tipping can take place between the two wells. Without the noise $\sigma = 0$, the dynamics would simply settle in one of the two wells. But the small fluctuations

---

[1]The basin of attraction of a certain attractor is the set of initial conditions that eventually lead to the attractor.

given by the Brownian motion term enable rare crossings over the potential barrier and thus tipping from one well to the other.

By additionally linearly tilting the potential $V(x) = \frac{1}{4}\left(x^2 - 1\right)^2 + \lambda x$ with an ever increasing slope parameter λ, a particle without noise in the right well will tip into the left well at the bifurcation point when the right well disappears. The added term λx can also be interpreted as causing a force of strength λ pushing from right to left (when $\lambda > 0$) and thus enforcing bifurcation-induced tipping when the parameter λ is large enough. We will see that by adding noise to the dynamics tipping can happen before the well has fully disappeared.

On the other hand, by moving the double well in the positive direction along the $x-$axis at a high enough rate, a particle in the right well can no longer follow the stable minimum as the potential moves away and thus rate-induced tipping happens from the right well to the left well in the deterministic gradient dynamics. Adding noise will blur the threshold behaviour that rate-induced tipping only happens when the rate is larger than a certain critical value.

We will now illustrate all three cases in more depth.

5.1.1 Noise-induced tipping in the double well

We start by considering the simplest case of only noise-induced tipping in the constant double well landscape

$$V(x) = \frac{1}{4}\left(x^2 - 1\right)^2. \tag{5.2}$$

In this system noise-induced tipping can happen from one basin of attraction across the barrier to the other basin of attraction. The two basins of attraction are centered at the minima of the landscape at $x = \pm 1$.

To study the tipping dynamics more quantitatively with the stationary TPT framework, we discretize the dynamics on the spatial interval $[-2, 2]$ with $\Delta x = 0.05$ and $\Delta t = 0.1$ to get a Markov chain with the transition matrix estimated by counting transitions as in Eq. (1.23). We consider the boxes belonging to the right well with $x \geq 0.8$ as set A and the boxes corresponding to the left well with $x \leq -0.8$ as set B. Then the committors (Fig. 5.1) are roughly linear in the transition region. On top of the barrier at $x = 0$, both committors are $1/2$. Due to the symmetry of the potential landscape around $x = 0$ and the reversibility of the process, the probability of next moving towards A from the barrier top agrees with the probability of next going to B both in forward and backward time. The computed rate of transitions amounts to $k^{AB} = 0.0013$, i.e., in stationarity a transition can be observed to start in A on average every $1/0.0013 = 769^{\text{th}}$ time step, and the mean duration of such a transition amounts to 27 time steps. By considering the rate of transitions conditional on being in A,

$$\mathbb{P}(\tau_B^+(n + 1) < \tau_A^+(n + 1) \mid X_n \in A) = (\mathbb{P}(X_n \in A))^{-1} k^{AB} \tag{5.3}$$

we learn about the frequency of starting a transition when the process is currently in A. Here the conditional transition rate when in A is given by 0.004, thus it takes on average 250 time steps when in A until a transition is started. Transition pathways are not yet interesting in this 1D example, they only become meaningful in higher dimensions where paths can be qualitatively different.
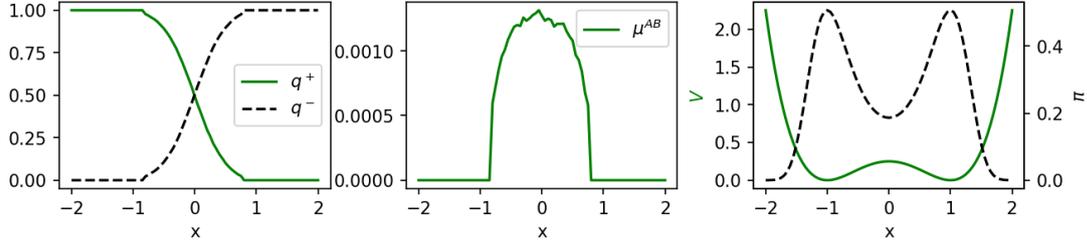
Figure 5.1: TPT analysis for noise-induced tipping in a symmetric double well potential.

5.1.2 Bifurcation-induced tipping in a tilting double well

Next we will come to bifurcation-induced tipping in combination with noise in the potential landscape

$$V(x, \lambda) = \frac{1}{4}\left(x^2 - 1\right)^2 + \lambda x \tag{5.4}$$

as the external parameter λ is varied. At $\lambda = 0$, both wells of the potential have the same depth and this parameter corresponds to the case from the above paragraph. When the external parameter λ becomes negative, the landscape is tilted such that the left well becomes more shallow and eventually disappears, while the right well becomes deeper and thus more metastable. When the parameter becomes positive, the right well loses depth and eventually disappears while the left well gains depth.

Bifurcations are usually considered in deterministic systems ($\sigma = 0$) under infinitesimally small parameter changes such that the system tracks the attractor until the bifurcation point where the qualitative dynamical change is immediate compared to the time scale of the parameter variation. In the first panel of Fig. 5.2, we plot the locations of the stable fixed points corresponding to the local minima of the potential (solid line) and the unstable fixed points corresponding to local maxima (dashed line) of the deterministic dynamics in the potential

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\mathrm{d}V}{\mathrm{d}x}(x, \lambda) \tag{5.5}$$

for different values of the control parameter λ. This gives the so-called bifurcation diagram showing the asymptotic states of the system as a function of the external parameter. In this example there are two saddle-node (or fold) bifurcations at the bifurcation points $\lambda = \mp\frac{2}{3\sqrt{3}}$ where a stable and an unstable fixed point annihilate each other upon varying λ and where the dynamics without noise will tip from the just

disappearing well to the other. The deterministic dynamics will exhibit hysteresis: at $\lambda = \frac{2}{3\sqrt{3}}$ the system can tip from the right well to the left well, but when the parameter variation is reversed, back-tipping happens at a different value of the parameter, namely at $\lambda = -\frac{2}{3\sqrt{3}}$.

Remark 5.1. *This S−shaped bifurcation diagram of two saddle-node bifurcations and the associated bistability is paradigmatic for tipping in climate [38] and ecological systems [61]. Once the critical bifurcation point is reached, intrinsic self-enforcing feedbacks lead to a shift towards another stable state [70]. Also safe overshoots [57, 58] have been studied in the context of the S−shaped bifurcation curve. When the control parameter exceeds the bifurcation point but is quickly reversed and settles at a value below the bifurcation point and when the response of the system is very slow, then it is possible for a trajectory that has started to shift to safely come back to its initial stable state.*

When we consider the tilting landscape with added noise, then the object corresponding to the bifurcation diagram is given by the stationary distribution of the discretized Markov chain for each parameter value λ (first panel of Fig. 5.2). Since the parameter is assumed to vary infinitesimally slowly, the stochastic dynamics will settle in the stationary distribution rather quickly compared to the infinitesimally slow parameter variations. While the bifurcation diagram highlights two points of qualitative change, the stationary distribution changes only near $\lambda = 0$. The stationary distribution changes from being concentrated in the right well when $\lambda < 0$ to being peaked in the left well when $\lambda > 0$. A particle in this slowly changing landscape will therefore change its predominant state as the parameter is varied.

We are now interested in studying the tipping dynamics in more detail using TPT. We will first assume that λ is changed infinitesimally slowly. Since the change is infinitesimally slow, it can be assumed that the dynamics is always equilibrated to the current parameter value and we are therefore interested in the stationary TPT quantities under change of λ. Since more realistically λ changes at a finite speed, we will study this next. With the extension of TPT to time-dependent dynamics in Chapter 2.4, we have the tool to also study tipping dynamics under time-dependent external parameters.

In Figs. 5.2, 5.3, we study how the tipping from the right to the left well depends on λ. The forward committor tells us whether the particle starting in some state and for some fixed value of λ will more likely tip to B than to A thereby making it the most important quantifier of the tipping likelihood. The potential barrier (curve of unstable fixed points) still indicates the positions where tipping to B is as likely as safely going back to A. Moreover we can note that the likelihood of starting a transition to B from a point close to the minimum of the right well only becomes large when the parameter λ comes close to the bifurcation point. This is confirmed by the second panel of Fig. 5.3 showing the transition rate conditional on being in A for different values of λ. Only when λ comes close to the bifurcation point the conditional rate starts to increase.

Due to the stochasticity, a trajectory experiencing the infinitesimally slow parameter variations will not only tip once but can go back and forth between A and B several
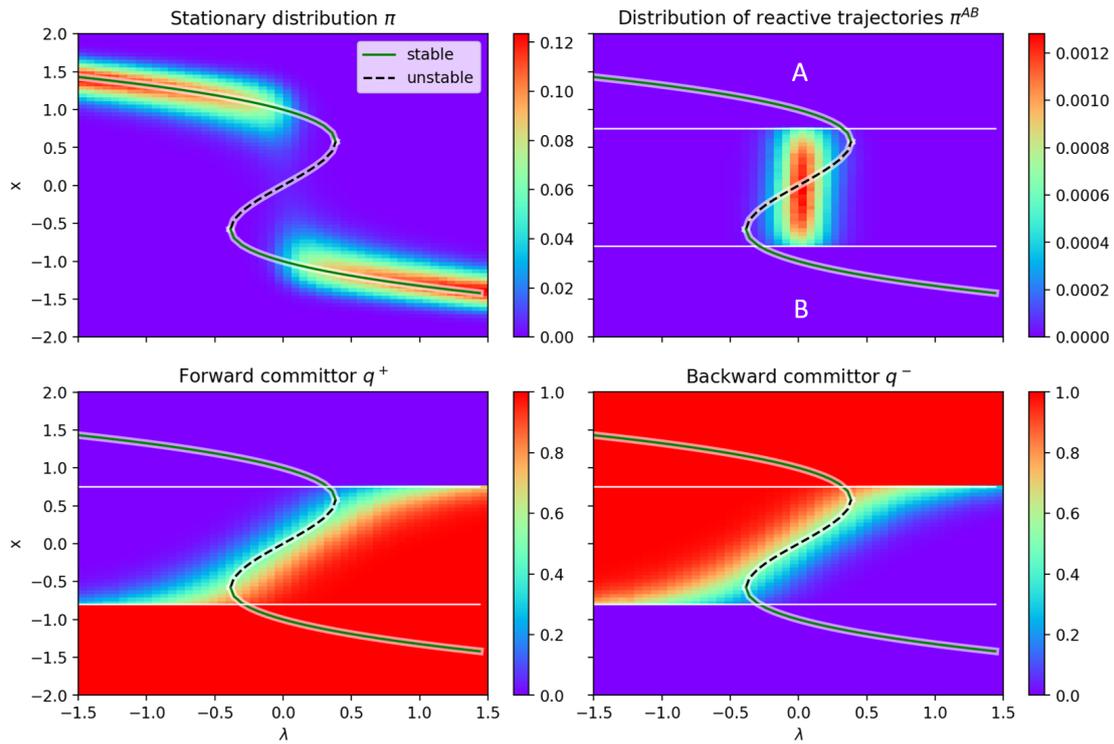
Figure 5.2: Infinitesimally slow bifurcation-induced tipping in combination with noise in a tilting double well.
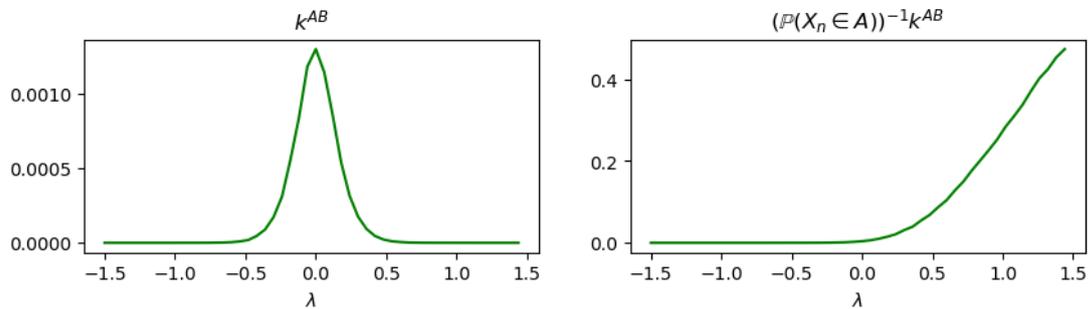
Figure 5.3: The dependence of the parameter on the transition rates in a slowly tilting double well.

times. The distribution π^{AB} and the rate of observing transitions k^{AB} characterize the ensemble of all these transitions. Both quantities are peaked at $\lambda = 0$. This means that most transitions happen near $\lambda = 0$ where the potential is symmetric. When $\lambda < 0$, the trajectory only tips rarely since the barrier to B is very high, while for $\lambda > 0$, the trajectory is most likely already in B and therefore less transitions are observed. We can conclude that tipping with noise can happen several times while the parameter is increased. Most transitions happen near $\lambda = 0$ and thus much earlier than the bifurcation point. Note that the hysteresis has disappeared and the attained state no longer depends on the history of the process but rather on the likelihood of the attained state which changes near $\lambda = 0$.
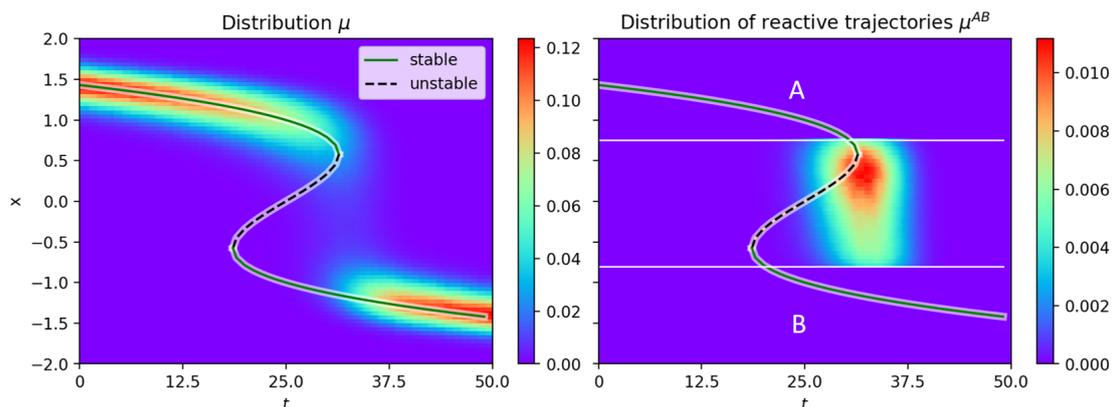
Figure 5.4: Bifurcation-induced tipping when the parameter is varied at a finite, constant rate. Tipping happens later and the parameter continues to vary while the system tips.

In realistic situations where bifurcation-induced tipping appears, the parameter is not varied infinitesimally slowly but just rather slowly compared to the time scale at which the dynamics evolves. Then the tipping is not immediate compared to the parameter variation but takes some time while the external parameter continues to vary. In Fig. 5.4, we show the resulting TPT quantities when the parameter is varied at constant rate $\frac{d\lambda}{dt} = 0.06$. This was realized by applying TPT to the time-evolving transition matrix on a finite time interval, see Section 2.4. We assumed as an initial distribution the stationary distribution of the initial transition matrix and as terminal and initial conditions for the committors the indicator functions on the sets B and A, thereby restricting the transitions to happen inside the time interval of interest. Since the size of the time interval where tipping is likely is rather short, most likely only one tipping event takes place within this interval. Indeed, the expected number of transitions is $\sum_{n=0}^{N-1} k^{A \to}(n) = 1.004$. The reactive distribution is peaked near the time point where the right well disappears. This implies that most reactive trajectories only start to tip away from the right well when the parameter crosses the bifurcation point at $\lambda = \frac{2}{3\sqrt{3}}$ and the right well disappears. If we would study reversed tipping in this example from the left to the right well while also reversing the parameter variation, tipping would happen near the other bifurcation point at $\lambda = -\frac{2}{3\sqrt{3}}$. Thus the dynamics again displays hysteresis. We can also expect that for slower rates of parameter variation, the tipping behaviour is a mixture of the two studied cases and that the distribution of reactive trajectories is peaked at an intermediate value of the external parameter.

5.1.3 Rate-induced tipping in a shifting double well

We have seen that the added noise in bifurcation-induced tipping blurs the critical parameter threshold that determines the onset of tipping. We are now interested in rate-induced tipping, i.e., in deterministic dynamics that tip because the rate of change of the external parameter has crossed a certain threshold. We will illustrate this phenomenon

5.1. *Tipping due to noise or external parameter variations*

and study the effect of added noise.

Inspired by [3], we consider the example of a particle in a shifting double well landscape. In particular, we let the potential shift in the positive $x-$direction

$$V(x, \lambda) = \frac{1}{4} \left((x - \lambda)^2 - 1 \right)^2 \tag{5.6}$$

as the parameter λ is varied at constant rate $\frac{d\lambda}{dt} = r > 0$. We will see below that particles will likely tip from the right to the left well when the rate r is large.
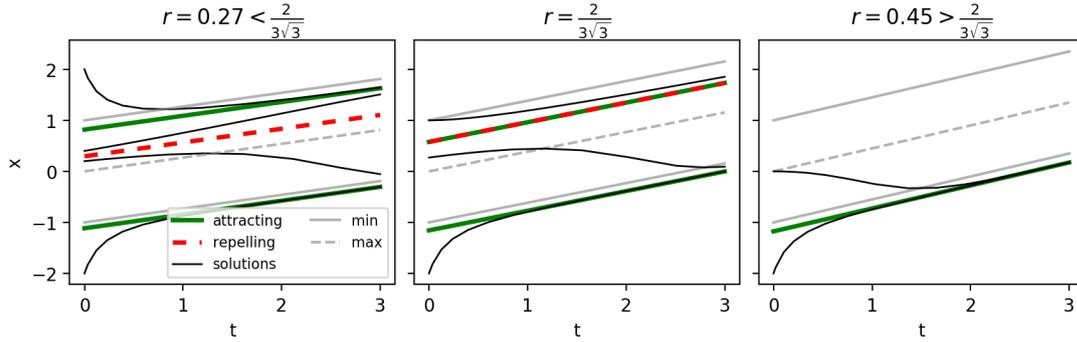
Figure 5.5: The $r-$dependence of the attracting (green) and repelling orbits (red) of ODE (5.7) as well as near-by trajectories (black). In the background the minima and maxima of the moving doublewell are shown (grey). Trajectories that start out in one of the wells and end up in the other are considered as tipped.

To get a better understanding of the dynamics and the tipping from the right to the left well, we will first consider the deterministic system without noise

$$\begin{cases} \dfrac{dx}{dt} &= -\dfrac{dV}{dx}(x, \lambda) \\[2mm] \dfrac{d\lambda}{dt} &= r \end{cases} \tag{5.7}$$

with $\lambda(0) = 0$. Depending on the rate r, the deterministic system has $1 - 3$ attracting or repelling orbits $x^o(t) = \lambda(t) + c$ where c solves $r = -c^3 + c$. These different cases can be summarized as follows, see also Fig. 5.5 for an illustration.

1. When $r = 0$, the system has two stable fixed points at the minima of the wells and one unstable fixed point at the top of the potential barrier, there is no rate-induced tipping in the deterministic dynamics.

2. For small rates $0 < r < \frac{2}{3\sqrt{3}}$, three orbits emerge from the three fixed points of $r = 0$, two that attract trajectories and that are close to the minima of the wells and one repelling orbit that is close to the barrier of the potential. When a trajectory is started inside the right well at a position between the repelling orbit and the moving barrier top, then it is drawn to the attracting orbit in the left well and thus there is rate-induced tipping from inside the right well to inside the left.
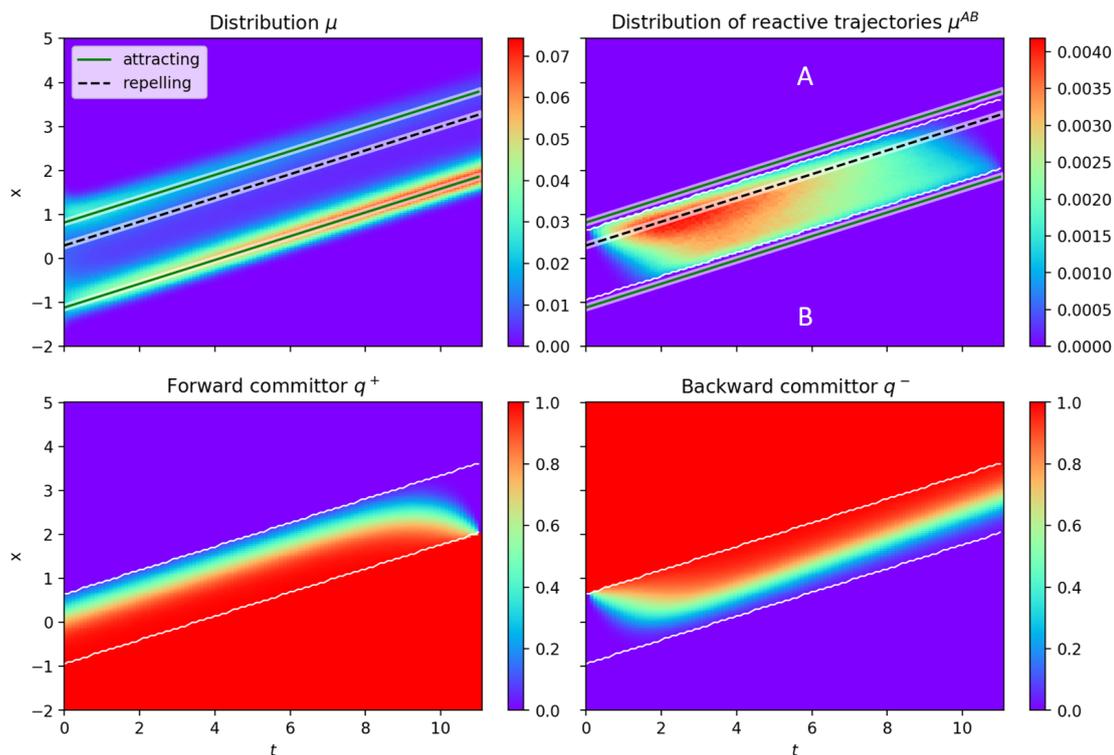
Figure 5.6: TPT analysis for rate-induced tipping (with noise) in a double well shifting at rate r along the positive $x-$axis.

3. At $r = \frac{2}{3\sqrt{3}}$, one repelling and one attracting orbit merge and there are only two solutions to $r = -c^3 + c$. Again, trajectories that start out to the left of the merged orbit can tip into the left well.

4. For $r > \frac{2}{3\sqrt{3}}$, the merged orbits have disappeared and there is only one attracting orbit left. For all starting conditions in the right well the trajectory will end up in the left well and thus tip.

Now that we understand the conditions such that rate-induced tipping happens in the deterministic system, we can start investigating the dynamics with added noise. We are interested in the regime $0 < r < \frac{2}{3\sqrt{3}}$ where the dynamics are still bistable, in particular we choose $r = 0.27$. Adding noise enables tipping in this parameter regime even for trajectories that start out to the right of the repelling orbit.

Since the dynamics is time-dependent, we will make use of TPT for dynamics on a finite-time window with the transitions restricted to the time frame of interest, compare Section 2.4. Due to the shifting of the potential landscape our sets A and B will be time-dependent. In particular we choose as A the set that just covers the right attracting orbit, while set B covers the left attracting orbit. Note that in the deterministic case there would be no tipping with this choice of sets since A lies to the right of the repelling line. As an initial distribution we consider the stationary distribution of the initial transition matrix.

In Fig. 5.6 we show the computed results. From the first panel, it is clearly visible that the distribution in the right well decreases and in the left well increases upon shifting the potential, thus some trajectories are tipping. Further, the forward committor and backward committor are shifted compared to Fig. 5.1. In particular, when a particle coming from A reaches the barrier top, it will almost always move towards B since $q^+ \approx 1$ near the barrier top. Also the backward committor is close to 1 near the barrier top, thus most trajectories near the barrier top have come from A. The distribution of reactive trajectories is peaked during the first half of the time frame, suggesting that most rate-induced tipping events from the right to the left well happen then. Note that the distribution only considers the tipping events that will arrive in B within the considered time interval. The expected number of transitions turns out to be $\sum_{n=0}^{N-1} k^{A\to}(n) = 0.29$. Since a trajectory at time 0 is with an equal probability in the left or in the right well, approximately $0.29/0.5 = 0.58$ of trajectories that are started in the right well make it to the center of the other well within the time interval. The mean rate of transitions conditional on being in A is given by 0.0147, thus when in A a transition departs 3.7 times faster as in the case of a constant double well. Further, the mean duration of transitions amounts to $\bar{t}_N^{AB} = 25.69$ time steps implying that reactive trajectories take slightly less time for the transition than in the case of a constant double well. In conclusion, the additional forcing by noise blurs the threshold behaviour. Tipping with added noise can happen for all rates of change r and all initial conditions inside the right well. The rate of change of the parameter only amplifies the tipping.

5.2   Canards with noise

Canards are a phenomenon that can occur in slow-fast systems. The dynamics in slow-fast systems alternates between parts of slow motion and parts of fast motion. Canards are trajectories that can for a substantial time track a repelling part of the dynamics before gaining speed and being pushed away, sometimes they can therefore be considered as delayed tipping events.

To illustrate this phenomenon, we consider the well-known slow-fast equation of the van der Pol oscillator which in slow time t can be written as [37]:

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} &= \dfrac{1}{\epsilon}\left(y - \dfrac{1}{3}x^3 + x\right) \\[2mm] \dfrac{\mathrm{d}y}{\mathrm{d}t} &= a - x. \end{cases} \tag{5.8}$$

The fast variable is given by x and evolves on the fast time scale $s = \frac{t}{\epsilon}$, while the slow variable y evolves on the slow time scale t. The separation between the two time scales is defined through the parameter $0 < \epsilon \ll 1$. The external parameter $a > 0$ determines the position of the fixed point at $(x, y) = (a, -a + \frac{1}{3}a^3)$ and its stability. The fixed point is unstable for $a < 1$ and stable for $a > 1$.

By considering the equation in terms of t and in terms of s separately and letting $\epsilon \to 0$ we can consider the slow and the fast dynamics independently in the limit of infinite time-scale separation. Letting $\epsilon \to 0$ in the ODE in slow time (5.8), we get the so-called *reduced* or *slow problem*

$$\begin{cases} 0 &= y - \frac{1}{3}x^3 + x \\ \frac{dy}{dt} &= a - x \end{cases} \tag{5.9}$$

which is an ODE describing only the slow motion with an algebraic constraint restricting the dynamics to

$$C := \{(x,y) \in \mathbb{R}^2, y = \frac{1}{3}x^3 - x\}, \tag{5.10}$$

the so-called *critical manifold*.

We can also re-scale time to the fast dynamics and let $\epsilon \to 0$ to get the *fast problem* or *layer equation*:

$$\begin{cases} \frac{dx}{ds} &= y - \frac{1}{3}x^3 + x \\ \frac{dy}{ds} &= 0. \end{cases} \tag{5.11}$$

The fast dynamics in the limit only takes place horizontally in the $x-$direction with $1-3$ fixed points depending on the current $y-$layer. The union of all fixed points of the layer equation corresponds exactly to the $S-$shaped critical manifold. We can find that the branch $C \cap \{-1 < x < 1\} =: C^r$ consists of repelling points of the fast system, while the branches $C \cap \{x < -1\} =: C^{a-}$ and $C \cap \{1 < x\} =: C^{a+}$ consist of attracting points. The points $C \cap \{x = -1\} =: p^-$ and $C \cap \{x = 1\} =: p^+$ on the critical manifold are change points where the repelling and attracting branch meet. They can be considered as points where the layer equation goes through a saddle-node bifurcation. The limiting dynamic of the fast and the slow problem is sketched in Fig. 5.7.
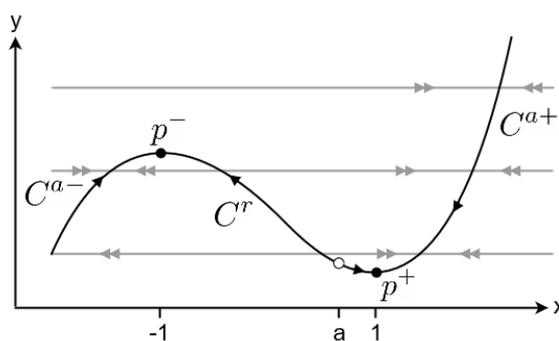
Figure 5.7: The limiting behaviour of the van der Pol equation for infinite time-scale separation. When the unstable fixed point (white dot) shifts to the right onto the branch C^{a+}, it becomes stable.

Now that we treated the limit $\epsilon \to 0$, let us consider the case of small $0 < \epsilon \ll 1$ [20,

37]. The slow and the fast problem already tell us something about the dynamics for finite ϵ. When a trajectory comes close to an attracting branch of the critical manifold, the fast x−motion becomes slow and thus the x−and y−motion co-evolve on the same time scale and follow close to the critical manifold in the direction of the slow y−motion. Away from the critical manifold, the x−motion is much faster than the y−motion such that the trajectory moves nearly horizontally.

The dynamics of the system depends crucially on the value of the external parameter $a > 0$. When $a > 1$, there is a stable fixed point situated on the attracting branch that attracts all trajectories. When a is far below 1, the fixed point is unstable and placed on the repelling branch. The overall motion is a stable limit cycle called *relaxation oscillation* that alternates between parts of slow motion near the attracting branches and parts of fast motion when the system switches to the other attracting branch near a saddle-node point.

Canards only appear when the parameter a is very close to 1 from below.[2] Canards are trajectories that for a substantial time stay in the neighbourhood of the repelling branch. When a trajectory slowly follows close to the attracting branch until it is near the saddle-node point p^+, then the y−dynamics can bring the particle close to the repelling branch which it then tracks for some time since the x−dynamics is slow near the repelling branch. It takes some time for the trajectory to gain speed in the horizontal direction and move away from the repelling branch. Depending on the precise value of a, the trajectory will either return to the attracting branch it came from (small canard cycle) or visit the other attracting branch (large canard cycle).

In terms of the external parameter a, the slow-fast system undergoes a Hopf bifurcation at $a = 1$. For $a < 1$ the system is oscillatory with a stable limit cycle around an unstable fixed point, while for $a > 1$ a stable fixed point exists. This bifurcation is characterized by stark variations of the size of oscillations prior to the bifurcation point where canards appear, this phenomenon is also called *canard explosion*.

Now that we have understood the general dynamics of the van der Pol oscillator and the configurations where canards appear, we are interested in studying the dynamics under the additional influence of small noise as in the SDE

$$\begin{cases} \mathrm{d}X_t &= \frac{1}{\epsilon}\left(Y_t - \frac{1}{3}X_t^3 + X_t\right)\mathrm{d}t + \sigma^X \sqrt{\frac{1}{\epsilon}}\mathrm{d}W_t^X \\ \mathrm{d}Y_t &= (a - X_t)\,\mathrm{d}t + \sigma^Y\mathrm{d}W_t^Y \end{cases} \tag{5.12}$$

where $\sigma^X, \sigma^Y > 0$ give the noise strength. The noise scaling in the fast variable is such that the Brownian motion in the fast variable is also fast. We choose $\sigma^X = \sigma^Y = 0.01$ and a time-scale separation of $\epsilon = 0.15$.

In the first panel of Fig. 5.8 we show a long realization where the external parameter a is close to 1 such that canards appear, more precisely $a = 0.9999$. Because of the noise, the system can pass through different small and large canard cycles which in the

---

[2]It can be shown that canards in the van der Pol oscillator appear near $a = 1 - \frac{\epsilon}{8} + O(\epsilon^{3/2})$ [37, Theorem 8.1.3.].
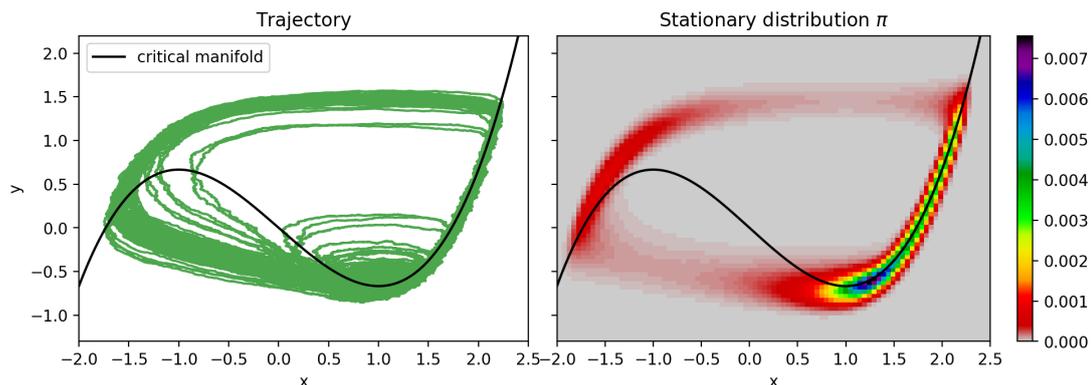
Figure 5.8: For the stochastic van der Pol equation, we show a long equilibrium trajectory of the stochastic system (left panel) and the corresponding stationary distribution (right panel).

deterministic system would only appear when a is varied. Due to the two time scales of the process, a continuous-time discretization in terms of a rate matrix is more appropriate than in terms of a transition matrix with a fixed time step Δt. For simplicity in this thesis, where the formalism is in terms of Markov chains, we still work with a transition matrix with a very small time step $\Delta t = 0.005$ such that the fast dynamics is also well resolved. The stationary distribution of the resulting Markov chain (Fig. 5.8) is spread over the limit cycle with most of its mass near the right attracting branch where motion is slow and which the trajectory will visit most often.

Since the motion on the attracting branches is slow, we can consider the fast transitions between the attracting branches as tipping events. The difference to the previous tipping mechanisms is that tipping here is not caused by parameter variations or noise. The most interesting question for us is how often the delayed tipping events from the right to the left attracting branch, i.e., the large canard cycles, happen compared to the small canard cycles corresponding to safe transitions back to the right attracting branch. To answer this question, we separately study the direct paths between each pair of branches of the critical manifold with TPT. We define three disjoint sets that cover C^{a-}, C^r and C^{a+} respectively. Then we use TPT to separately study all direct transitions paths from any of the source sets C^{a-}, C^r or C^{a+} to any of the sink sets C^{a-}, C^r or C^{a+} that avoid the remainder of sets as described in Section 2.2.6. The distributions of these direct paths and corresponding transition rates are shown in Fig. 5.9(a). We can also abstract the transition behaviour into a transition network between the three sets. By dividing the individual rates by the probability of being in the given source set, we get the conditional transition rates between the three sets. These are indicated on the edges of the transition network in Fig. 5.9(b).

We can conclude that for the chosen parameters, it is slightly less likely to directly tip from the attracting branch C^{a+} to the other attracting branch C^{a-} than to transition from the attracting branch C^{a+} first to the repelling branch C^r. When on the repelling branch, it is 4 times as likely to return back to C^{a+} (small canard cycle) instead of a
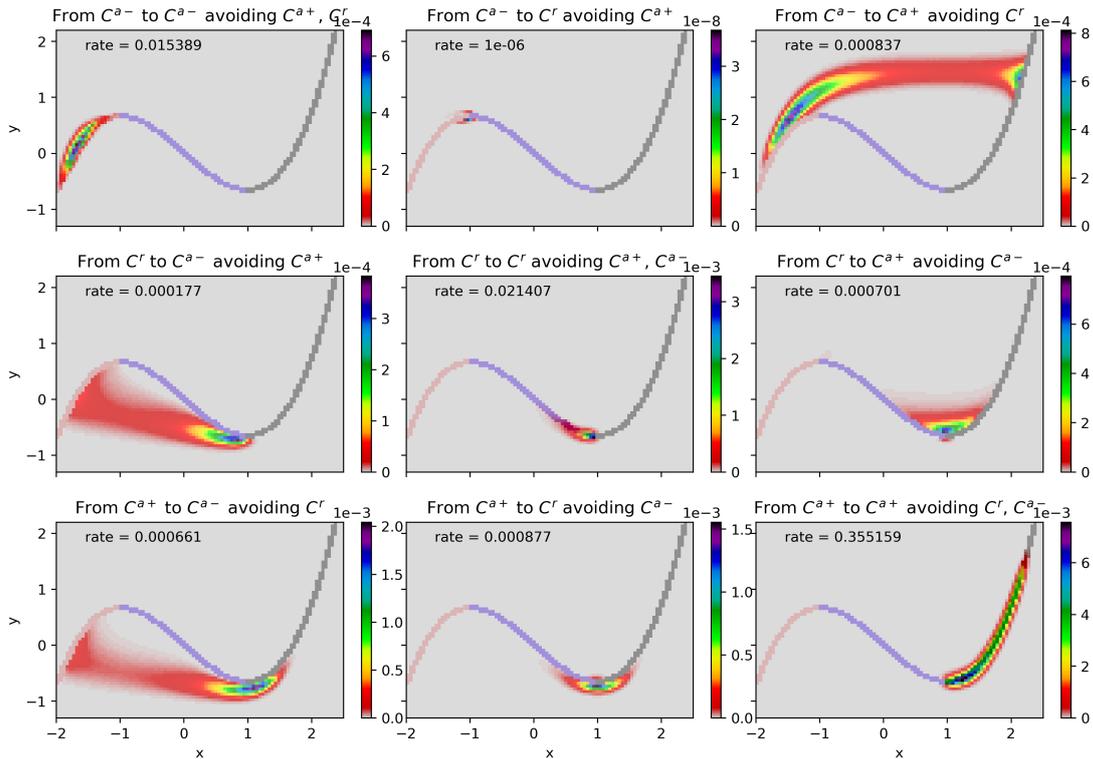
(a)
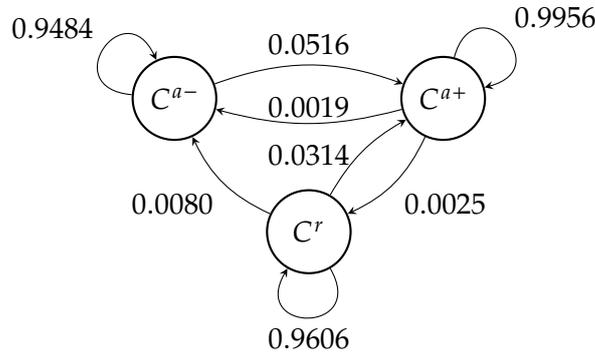
(b)

Figure 5.9: For the stochastic van der Pol equation we study the direct paths between the different branches of the critical manifold. In (a) we show the densities of the direct transitions and the corresponding transition rates. In (b) we give the coarse-grained transition network between the different parts of the critical manifold and the associated transition probabilities.

delayed tipping to C^{a-} (large canard cycle). Thus for our choice of parameters, small canard cycles are 4 times as likely as the large canard cycles.

6 | Social tipping in agent-based models

Now that we have the tools at hand to study general transition dynamics in Markov chains and have learned about the possible tipping mechanisms, we come back to the main problem of studying tipping in agent-based models. Transition Path Theory allows us to gain a complete statistical understanding of the diverse tipping paths that can occur in ABMs. The theory in Chapters 2 and 3 can be utilized to study tipping caused by noise or external parameter variations as long as the high dimensionality of the model does not pose a problem. In this chapter we will focus on high-dimensional agent-based models that exhibit noise-induced tipping. By parametrizing the most important degrees of freedom in the ABM by collective variables, we can reduce the size of the state space substantially and carry out a TPT analysis.

The results and figures of this chapter were originally published in [29].

6.1   Reduction of agent-based models

Before we start with the model reduction, we want to understand the complexity behind social systems and agent-based models

6.1.1   Social systems and agent-based models

Social systems are complex systems and as such characterized by rich, nonlinear and usually local (i.e., only between neighbours) interactions among a large number of individual constituents [13, 12]. The individual entities are ignorant of the behaviour of the system as a whole and only respond to local information. Social systems are usually open, i.e., continuously in interaction with their environment, and therefore often not in a simple equilibrium. In addition, social systems are affected by their history and can display hierarchies and multi-scale structures [64].

When modeling the dynamics of social systems (or, parts thereof), agent-based models (ABMs) are a natural choice [44]. One defines the characteristics of a large but finite number of discrete entities (e.g., people, households, companies, . . .), the so-called agents, and a set of possible actions and local interactions rules for the agents. Often the behavioural rules are stochastic, thus reflecting the unpredictability and individuality of the agents. By iteratively changing the states of the agents according to the local behavioural rules of each agent, the dynamics are simulated. From the interplay of

local interactions, patterns on the global scale can emerge. These global patterns are usually the main interest not the individual agent states.

In theory, one can describe the full complexity of social systems with these models, but ABMs quickly become difficult to parametrize, understand and analyse. Simple ABMs are therefore usually employed to test hypotheses and to learn about the laws of interactions, while more complicated ABMs are used to make predictions about the real-world.

6.1.2   Model assumptions

For our purposes it is useful to view agent-based models as Markov processes. Indeed, when an ABM is stochastic and defined iteratively as a computer program, it can be considered as a Markov processes but on a possibly very high-dimensional or large state space $\mathbb{X}$ [7, 30]. The Markovianity assumption means that the next state of the system only depends on the current state and not the history. This does not necessarily mean that agents have no memory or cannot be influenced by their past. By enlarging the state space formally to include a memory of past states, Markovian dynamics can be retained.

Let us denote the state of agent i at time t by X_t^i and, for simplicity, we assume that all agents can adopt the same states $X_t^i \in \mathbb{A}$. The individual agent state can for example be a binary opinion such as when $\mathbb{A} = \{0, 1\}$, a continuous position, e.g., $\mathbb{A} = \mathbb{R}^2$, or a combination of both. The state of the whole population of N_A agents is then given by the collection of all individual agent states $\mathbf{X}_t = \left(X_t^i\right)_{i=1}^{N_A}$ and thus the state space is given by $\mathbb{X} := \mathbb{A}^{N_A}$. Finally, the Markov process $(\mathbf{X}_t)_{t \in \mathbb{T}}$ describes how these population states evolve in time.

Since the high dimensionality of the model is already difficult enough, we consider in this chapter only ABMs in stationarity and assume the models to be ergodic such that the whole state space can be explored by a single realization.

6.1.3   Model reduction

The main difficulty when analysing stationary agent-based models lies in their high dimensionality. The size of the state space grows exponentially with the number of agents, and usually one is interested in studying a rather large population of agents with each of them having many internal states.

But most high-dimensional complex systems can be described by a few meaningful coordinates that parametrize the most important dynamical changes and that allow a substantial reduction of the state space. These so-called *collective variables* (or *reaction coordinates*, *order parameters*)

$$\xi : \mathbb{X} \to \mathbb{R}^d$$

allow an approximate description of the actual system's dynamics in a "reduced" or "simpler" state space of a smaller, so-called intrinsic dimension d.

There are two reasons that give rise to the existence of these coordinates for ABMs. First, the dynamics of most ABMs is concentrated on a subset of the state space with fewer degrees of freedom. When the state space is continuous this might mean that the ABM resides most of the time in the vicinity of some low-dimensional manifold. For example, many ABMs are concentrated on states in which neighbouring agents display coherent behaviours due to the local interaction rules (e.g., imitation), resulting in a much smaller effective state space.

Secondly, many ABMs have symmetries with respect to the dynamics that allow a further reduction of the state space. For example when all agents are identical (or exchangeable), then the dynamics does not depend on which of these agents is in a certain state but only on how many of them are in a certain state. Even when not all agents are identical, there can be groups of agents that are identical for example due to a similar position in the interaction network. Many real-world social networks are highly modular [25, 69], i.e., contain many communities, and agents in these communities can be considered as approximately identical when all their other defining features also agree.

From now on, we assume agent-based models for which collective variables exist. Only in certain situations, e.g., when agents behave rather homogeneous or identical, one can guess appropriate collective variables based on an intuition about the system's dominant feedbacks. On the other hand, when the agents are rather heterogeneous and the dynamics are complicated, then it is difficult to identify meaningful collective variables. We therefore seek an automated way of finding collective variables ξ, which should allow us to represent the dominant model behaviour, as well as all the important transition pathways between the metastable regions in state space.

Not only is a long simulation of an ergodic agent-based model comparably cheap to simulate, even in high-dimensional or large state spaces, but most of the visited states will be concentrated in a small part of the state space, i.e., they lie near the low-dimensional manifold or the discrete analogue we wish to parametrize. Methods such as Diffusion Maps, introduced in Section 1.2.3, can then be employed to parametrize this object from the sampled trajectory. The sampled data, together with a suitable distance function representing the dynamical closeness between them, yield a weighted graph that can be parametrized by the dominant eigenfunctions of a random walk on the graph. These eigenfunctions will be used as collective variables. It is worth noting that the eigenfunctions reflect the symmetries of the dynamics and thus the symmetries are also taken into account in the collective variables. In the literature [45, 39] Diffusion Maps have already been used to determine collective variables of ABMs.

Given the collective variables, we still need to find a reduced description of the model evolution in terms of these coordinates. The simplest approach is the construction of a Markov chain model on a discretization of the projected state space $\xi[\mathbb{X}]$.

All in all, we propose to reduce the ABM as follows:

1. Sample a long realization $\mathbb{D} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_M\} \subset \mathbb{X}$ of the ABM.

2. Apply the Diffusion Maps algorithm to the whole realization or a large enough sub-sample of it to find suitable collective variables $\xi : \mathbb{X} \to \mathbb{R}^d$. The found collective variables can be interpolated at the remaining states such that we can write the reduced realization as $\xi[\mathbb{D}] = \{\xi(\mathbf{x}_0), \xi(\mathbf{x}_1), \ldots, \xi(\mathbf{x}_M)\} \subset \mathbb{R}^d$.

   Remark 6.1. *Note that the state space $\mathbb{X}$ is not necessarily an Euclidean space. We assume that it is a metric space and that the important results from the theory of Diffusion Maps, for example, regarding the implications of re-normalizations of the kernels, can be transferred to metric spaces. In particular, we will later consider the discrete state space $\{0, 1\}^{N_A}$ with the Hamming distance.*

3. To find a suitable partition of the reduced space into L cells, we apply the K-means clustering algorithm to $\xi[\mathbb{D}]$ and use the found L centers as centers of Voronoi grid cells denoted by $\{A_l\}_{l=1,\ldots,L}$. Thus we finally arrive at the *discrete collective variables*

$$\hat{\xi} : \mathbb{X} \to \{1, \ldots, L\}$$

   with $\hat{\xi}(\mathbf{x}) = l$ whenever $\xi(\mathbf{x}) \in A_l$.

4. Estimate a transition matrix $\hat{P}$ on $\hat{\mathbb{X}} = \{1, \ldots, L\}$ by counting the transitions between the grid cells in the reduced realization $\xi[\mathbb{D}]$, compare with Section 1.2.1.

6.2 Tipping analysis of two agent-based models

In this section, we present two agent-based models and examine their tipping paths after a suitable model reduction. The two presented models are part of the large class of models of opinion and behavioural change due to social interactions [65]. Often in these models one is interested in understanding the emergence of a stable macro-state of either opinion consensus or synchronous behaviour on the one hand, or opinion polarization or asynchronous behaviour on the other. In contrast, we are interested in the transitions between states of locally converged agents and have thus chosen models where the stochasticity enables tipping.

To be more specific, in our models agents are making binary behavioural decisions and change their binary opinions in response to the social influence of their network neighbours, potentially mediated by an additional macroscopic interaction. Apart from their fixed position in the network, agents are identical. We will assume interaction networks consisting of several groups of nodes which are densely linked among themselves but with only few connections to the other groups. The densely linked agents in each block are nearly identical because they are connected to very similar sets of other agents, and thus behave rather similarly due to the local interaction rules. Both ABMs can be viewed as ergodic Markov chains on finite but large state spaces and have many metastable states, where agents behave collectively in each of the blocks. Due to these properties, both ABMs exhibit noise-induced tipping, i.e., transitions between metastable sets, that can be studied with Transition Path Theory.

6.2.1 A threshold model of social contagion or activation

We will start by introducing and analysing a very simple ABM describing phenomena such as the spreading of cultural fads, hypes or consumption behaviours, or the activation for some collective action such as rioting. The model is ultimately based on Granovetter's threshold model [26] and similar to several network-based versions of Granovetter's model [73, 74].

Let us define our threshold model in more detail:

Setting. We consider a population of N_A agents with social connections among them given by the edges of a static network $\mathcal{G}$ of N_A vertices. Each agent i can be in one of two discrete states: being *inactive* or *active* in the collective action. We denote the binary state of agent i at the discrete time point n by $X_n^i \in \{0,1\}$ with $X_n^i = 0$ corresponding to inactivity and $X_n^i = 1$ meaning activity.

Interaction network. We assume that the interaction network $\mathcal{G}$ has two scales: it consists of *blocks*, sometimes also referred to as *communities*, in which the nodes are densely connected, whereas nodes of different blocks are sparsely connected. One popular approach to randomly generate such a network is by the stochastic block model. Each node i is uniquely assigned to a *block* $\mathcal{B}_k$, $k = 1, \ldots, K$. When node i belongs to $\mathcal{B}_k$, we also write $i \in \mathcal{B}_k$. After defining a symmetric matrix $W = (W_{kl})$ of size $K \times K$ that contains the edge wiring probabilities between a node of block $\mathcal{B}_k$ and one of block $\mathcal{B}_l$, we go through all pairs of nodes independently and with probability W_{kl} place an edge between them when they belong to blocks $\mathcal{B}_k$ and $\mathcal{B}_l$. The diagonal entries of W determine the edge wiring probabilities for agents from the same block. In the case of only one block, this is equivalent to the Erdős–Rényi random graph model.

Interaction rules. When an agent makes a binary decision its network neighbours exert a a threshold-like influence: if more than a certain fraction of neighbours are in the opposite state to that of the agent, the agent will switch its state with a high probability. Thus each agent aligns its state with the state of the majority of its social neighbours. In addition, there is a small probability for the agent to switch its state when less than a certain fraction of neighbours are in the opposite state, which can either be interpreted as a form of exploration or as representing otherwise unmodelled additional causes for switching one's state.

More precisely, at each time n, each agent i in state $X_n^i = 0$ (resp. = 1) will change their state to $X_{n+1}^i = 1$ (resp. = 0)

- with probability p, if more than or exactly a fraction θ of neighbouring agents at time n are in the opposite state 1 (resp. 0),

- or with the exploration probability e, if less than a fraction θ of neighbours is in the opposite state,

where we assume $1 > p \gg e > 0$ such that social influence is stronger than exploration.

Markov chain.   We can also view the system as a Markov chain $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ on the state space $\mathbb{X} = \{0, 1\}^{N_A}$, where we denote the *population state* at time n by $\mathbf{X}_n = (X_n^i)_{i=1,\dots,N_A}$. Since agents in every time step change their state synchronously and independently of each other, the transition matrix on $\mathbb{X}$ decomposes into the product of the "transition probabilities" for each individual agent

$$P_{\mathbf{xy}} := \mathbb{P}(\mathbf{X}_{n+1} = \mathbf{y} \mid \mathbf{X}_n = \mathbf{x}) = \prod_{i=1}^{N_A} \mathbb{P}(X_{n+1}^i = y^i \mid \mathbf{X}_n = \mathbf{x}). \tag{6.1}$$

The exploration probability ensures that agents are never stuck in a state. In every time step an agent has a positive probability to remain in the same state as well as to change the state, i.e., $\mathbb{P}(X_{n+1}^i = 0 \mid \mathbf{X}_n = \mathbf{x}) > 0$ and $\mathbb{P}(X_{n+1}^i = 1 \mid \mathbf{X}_n = \mathbf{x}) > 0$ respectively. Thus by (6.1) there is a positive probability to go from any population state to any other within one time step, implying that the Markov chain is irreducible and also aperiodic.

Remark 6.2. *Note that the form of this transition matrix (6.1) demonstrates two important properties of ABMs. First, each agent is autonomous, therefore the matrix decomposes into the transition rules of each individual agent. Second, since every agent only takes their neighbours' activities into account, the transition probabilities of each individual agent do not depend on the state of the whole population but can be reduced to depend only on the state of their neighbours*

$$\mathbb{P}(X_{n+1}^i = y^i \mid \mathbf{X}_n = \mathbf{x}) = \mathbb{P}(X_{n+1}^i = y^i \mid X_n^j = x^j, j \in \mathcal{N}_i) \tag{6.2}$$

where $\mathcal{N}_i$ is the set containing agent i as well as its neighbours. This is a manifestation of the property that agents in most ABMs only use local and limited information for determining their actions.

Resulting dynamics.   If we consider a population where every agent is interacting with every other agent, i.e., they are interacting on a complete network, then in the interesting parameter regime the system switches between two metastable regions: (i) where a majority of agents is inactive and (ii) where a majority is active (Fig. 6.1). These transitions where agents are changing their activities are fast and drastic and can be considered as *noise-induced* events. If we now study a population consisting of several complete or mostly complete blocks, with some connections between the blocks, then this dynamic is multiplied. The majority in each block can switch between inactivity and activity, but depending on the number of connections between the blocks, all blocks are either synchronized, only weakly influencing each other, or behaving mostly independently. By *tipping* we here understand a transition from one metastable region to another, i.e., one block of agents drastically changes its state from majority of agents active to majority inactive (or vice versa). Sometimes we might refer to the tipping of the whole population, i.e., when all blocks one after the other change between majority
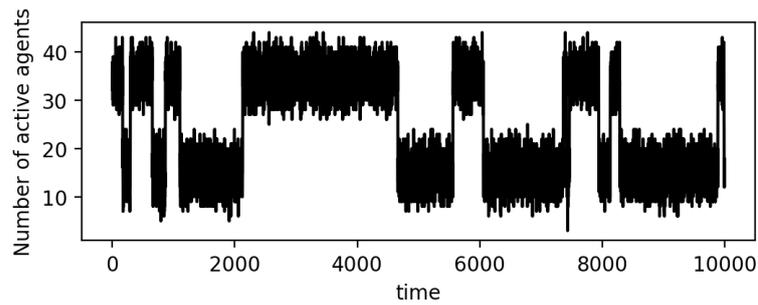
Figure 6.1: Realization of the threshold model with 50 agents that are interacting on a complete network. The dynamics switches between two metastable macrostates. The model parameters are $e = 0.3$, $p = 0.7$, $\theta = 0.5$.

of agents being active and the majority being inactive. This happens via the individual blocks' successive tipping, i.e., via a *tipping cascade*.

We will now study two examples in more detail.

Example 1

In the first example we will consider a small population of just 10 agents that are evenly split into two blocks. We set parameters to the case where tipping occurs and the blocks are weakly influencing each other in order to avoid a trivial behaviour and to focus on the most interesting dynamical regime, see Fig. 6.2(b) for the network. In particular, we set the change probability as $p = 0.3$ and the exploration probability as $e = 0.03$. As long as $p \gg e$, the actual scale of the probabilities determines mostly how fast agents are changing their behaviour. The threshold was set to the most focal value of $\theta = 0.5$, meaning that agents are influenced by the majority behaviour in their neighbourhood. With a size of $|\mathbb{X}| = 2^{10}$, the state space is already nontrivial, but still small enough to be able to do direct computations on the state space.

The realization in Fig. 6.2(a) indicates that the system remains in four metastable regions most of the time: where (i-ii) a majority in block 1 (resp. 2) but not in the other block is active, (iii) a majority in both blocks is inactive, and (iv) a majority in both blocks is active. It seems that those states (i-ii) where the two blocks show a differing majority activity are less metastable than those states (iii-iv) where agents in both blocks are conforming. Moreover, the realization suggests that the tipping of one block induces the other block to also tip.

Several reasons suggest that we can find meaningful collective variables for the dynamics. First, the interaction rules of this model encourage coherent behaviour among neighbouring agents, thus the dynamics should concentrate only on a subset of the state space where agents in the different blocks behave conform. Additionally the dynamics has several symmetries. The model is symmetric under swapping active with inactive. Further, the agents in each block are approximately identical and thus the dynamics are approximately symmetric under swapping agents in the same block.

105
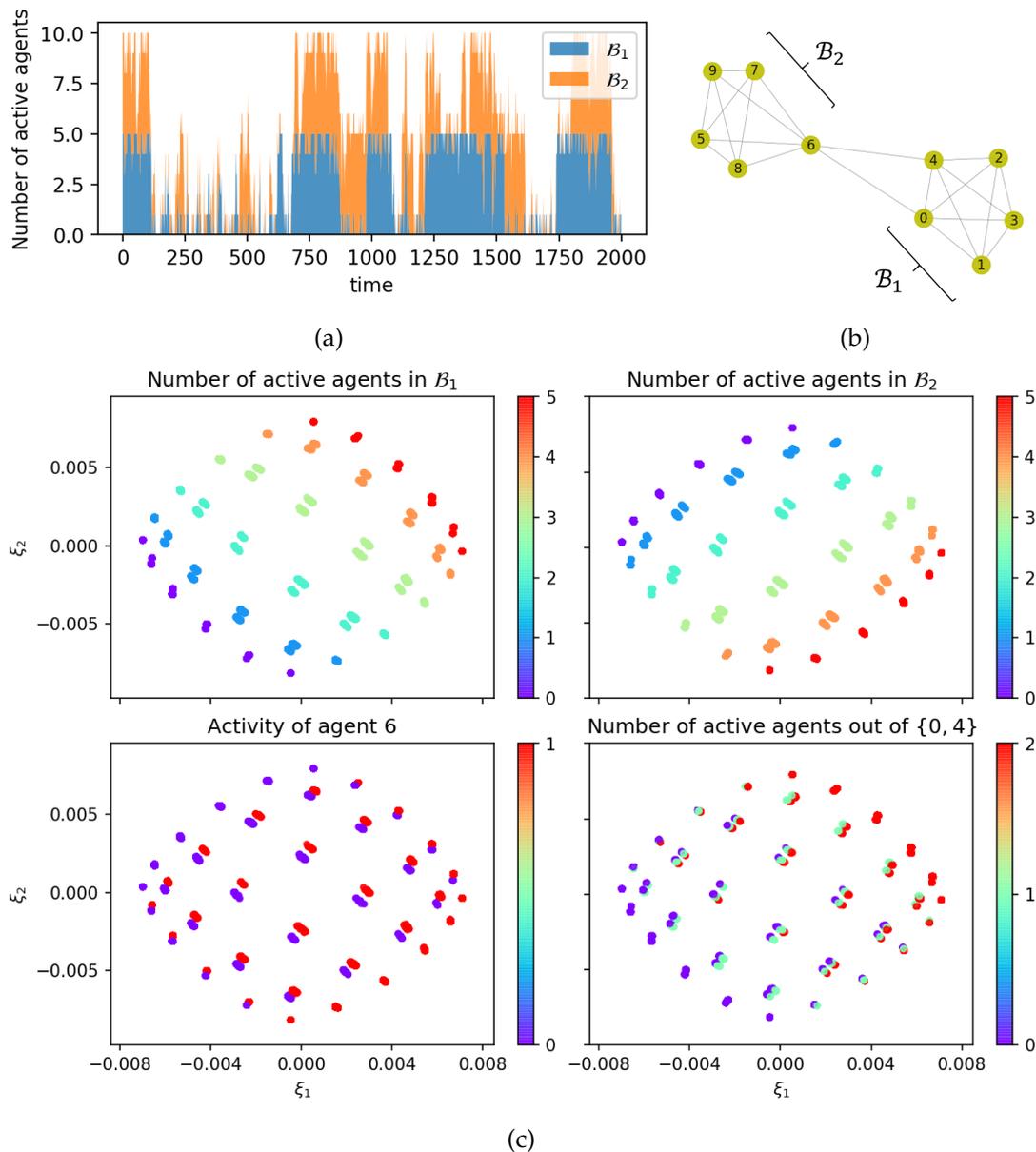
Figure 6.2: Threshold model with a network of two small blocks, see Example 1: (a) The realization is shown using a stackplot, i.e., the number of active agents in $\mathcal{B}_2$ is plotted vertically on top of the number of active agents in $\mathcal{B}_1$. Several tipping events can be seen. (b) Interaction network of two blocks. (c) Projection of population states into the dominant two Diffusion Maps coordinates, the Diffusion Maps scale parameter turned out to be $\epsilon = 0.25$. To better understand the projection, the data points are colored according to the number of active agents in each block and the activity of agents $0, 4$ and 6.

Last the dynamics is approximately symmetric under swapping all agents in one block with the agents in the other block.

So how meaningful are the coordinates from the Diffusion Maps algorithm when using a realization of $M = 20,000$ population states $\mathbb{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{20,000}\}$? Due to the length of the realization, many states occur multiple times and we use these repeated data points also for the Diffusion Maps projection. As a kernel we use the Gaussian kernel $k^{\epsilon}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{d(\mathbf{x},\mathbf{y})^2}{\epsilon}\right)$ and compute the distance $d(\mathbf{x}, \mathbf{y})$ between two data points via the Hamming distance, which measures the distance between two binary vectors as the number of entries where they differ. The Hamming distance between two population vectors reflects also their dynamical distance since population vectors where few agents have changed their states are much closer in Hamming distance than population vectors where many agents have switched. Further, we estimate an appropriate scale parameter ϵ using the heuristic from [8]. The coordinates from the Diffusion Maps algorithm are the dominant eigenfunctions of a random walk on a weighted network between the data points. The weighted network contains information about the relative occurrence of states of the model, their estimated dynamical closeness, but also the symmetries of the dynamics are reflected in the network.

The projection into the dominant two coordinates can be found in Fig. 6.2(c). Judging from the location of the spectral gap of the Diffusion Maps spectrum (not shown here), the intrinsic dimension of the dynamics seems to coincide with the number of blocks, i.e., $d = 2$. The data set of population states are embedded into a square. The coloring of the data points indicates that the two orthogonal directions encode the number of active agents in each block and thus these directions constitute the most important degrees of freedom in the data. Note that these coordinates should remain approximately the same under swapping the states of agents in the same block and thus the symmetry of agents in the same block is reflected in the collective variables. The Diffusion Map coordinates are refining the structure of the essential dynamics with each additional coordinate and are ordered by the scales they encode. The first Diffusion Map coordinate encodes the total number of active agents in the population and the second refines this by splitting them into two blocks. Looking more closely, we can see that the projected groups of points (corresponding to a certain number of active agents in each block) consist of some substructures on a smaller scale. These substructures encode whether agent 6 is active or not, and how many of agent 0 and 4 are active. Higher-order Diffusion Maps coordinates, in this case the coordinate ξ_4 (not shown in the figure), also decode the information about the activity of agents $0, 4$ and 6. We will later investigate the importance of agents $0, 4$ and 6 with respect to the dynamics.

After discretizing the space spanned by the collective variables into $L = 36$ Voronoi cells using the K-Means algorithm, and estimating the transition matrix on this discrete state space by using the projected realization, we now come to studying the tipping paths with TPT. For the threshold model we are interested in studying how the activity in collective behaviour spreads through the population. We will therefore choose A as

107

the set of states where less than 3 agents are active, and B as the set where more than 7 agents are active.[1] The results of the tipping analysis are presented in Fig. 6.3.

In the panels (a) and (b) we show the committors q^+, q^- on the reduced state space. The original state space $\mathbb{X} = \{0,1\}^{10}$ is small enough such that we can solve the system of linear equations for the exact forward resp. backward committors of the ABM, denoted by q^+_{ABM} resp. q^-_{ABM}. We can therefore study the relative error in the π-weighted l^2-norm between the committors q^+, q^- on the reduced space and the exact committors of the ABM

$$\frac{\|q^+_{\text{ABM}} - q^+ \circ \hat{\xi}\|_{l^2(\pi)}}{\|q^+_{\text{ABM}}\|_{l^2(\pi)}} = \left(\frac{\sum_{\mathbf{x} \in \mathbb{X}} \left(q^+_{\text{ABM},\mathbf{x}} - q^+_{\hat{\xi}(\mathbf{x})} \right)^2 \pi_{\mathbf{x}}}{\sum_{\mathbf{x} \in \mathbb{X}} \left(q^+_{\text{ABM},\mathbf{x}} \right)^2 \pi_{\mathbf{x}}} \right)^{\frac{1}{2}} \tag{6.3}$$

by approximating the π-weighted sums by evaluations along the stationary realization $\mathbb{D}$. The relative error of the forward and the backward committors is 0.03 and confirms that the collective variables allow a good approximation of the tipping dynamics.

From Fig. 6.3 (a) we can see that the forward committor is not perfectly symmetric with respect to the two blocks: when block 1 has completely tipped but block 2 has not (these are the states around $\xi_1 \approx 0$ and $\xi_2 > 0.005$, compare with Fig. 6.2 (c)), the forward committor is much higher than in the opposite scenario, when block 2 has tipped but block 1 not (the states around $\xi_1 \approx 0$ and $\xi_2 < -0.005$). Also the reactive distribution is higher when block 1 has tipped and 2 has not. The effective current[2] in Fig. 6.3 (e) indicates two dominant transition channels from A to B:

(I) $A \rightarrow$ agents in block 1 get active $\rightarrow$ agents in block 2 get active $\rightarrow B$

(II) $A \rightarrow$ agents in block 2 get active $\rightarrow$ agents in block 1 get active $\rightarrow B$,

with slightly more flow through channel (I). In order to better compare the likelihood of both transition channels, we group similar states of the discrete state space $\hat{\mathbb{X}} = \{1, \dots, L\}$ together to form a coarser partition of the discrete state space and compute the aggregated currents between these groups of states. We denote the groups by $\{G_r\}_{r=1,\dots,R}$, they have to be disjoint, respect the boundaries of A and B and partition the state space $\cup_{r=1}^{R} G_r = \hat{\mathbb{X}}$. Here we split the transition region C into three channels and join the states of A and of B, compare the coloring in Fig. 6.3 (f). We can then compute the *reactive macro-current* F^{AB} from one group of states G_r to another group of states G_s by summing all the individual reactive currents from each state of the first group to each state of the final group, i.e., $F^{AB}_{rs} = \sum_{l \in G_r, m \in G_s} f^{AB}_{lm}$. From F^{AB}, we can also compute the effective macro-current F^+. With these macro-currents (Fig. 6.3 (f))

---

[1] Note that we have to be able to express A and B via the collective variables.

[2] The threshold model is only very slightly non-reversible, therefore we are not doing a decomposition into cycles and productive parts. Instead the approximately cycle-free effective current highlights the transition pathways.
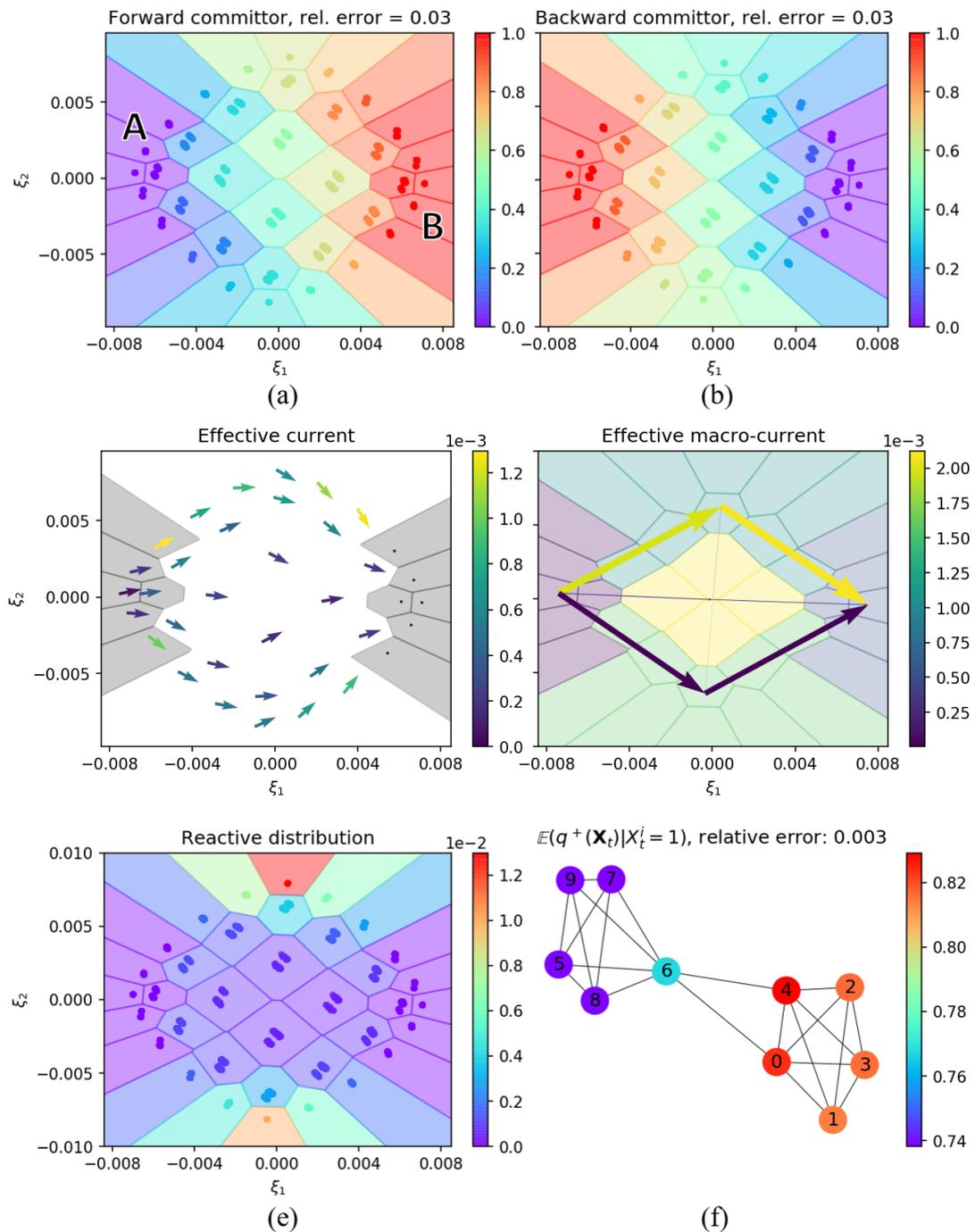
Figure 6.3: Tipping analysis for Example 1: (a), (b) Committors on the discretized space. (c) Effective current, *A* and *B* are indicated by the two shaded areas. (d) Effective macro-current through the three channels indicated in shaded blue, yellow and green. (e) Reactive distribution. (f) Agents as indicators of the overall tipping.

we can confirm that there is more effective current going through channel (I). The transition rate from A to B amounts to $k^{AB} = 0.0039$ meaning that in a stationary trajectory, a transition from A to B of mean duration $t^{AB} = 18.85$ is started on average every $1/k^{AB} \approx 256$th time step. Transitions along channel (I) contribute 53% to k^{AB}, while channel (II) only contributes 41% to the rate. The reason should lie in the slight asymmetry of the network between block 1 and 2: Agents 0 and 4 of block 1 are both connected to agent 6, see the network in Fig. 6.2(b). From the interaction rules of the ABM it follows that the likelihood of agent 0 and 4 to become active when agent 6 is active is smaller than the likelihood of agent 6 to become active after agents 0 and 4. These results also fit with the asymmetry in the committor: as soon as block 1 has tipped, it is very likely that block 2 also tips.

To further study the role of each individual agent with respect to the overall tipping between A and B, we consider the expected forward committor conditional on agent i being active

$$\mathbb{E}\left(q^+ \circ \hat{\xi} \mid X_n^i = 1\right) =: I_i^{AB}. \tag{6.4}$$

The agents with the largest I_i^{AB} constitute the best *(individual-agent) indicators* that the overall tipping of the population will soon happen. When these agents are active, the system is the most likely to tip to B, thus one should especially consider these agents to access the tipping likelihood. We can estimate I_i^{AB} by a Monte-Carlo approximation with a sufficiently long stationary ABM trajectory $\mathbb{D} = \{\mathbf{x}_0, \ldots, \mathbf{x}_M\}$:

$$
\begin{aligned}
I_i^{AB} &= \frac{\mathbb{E}\left((q^+ \circ \hat{\xi}) \, \mathbb{1}_{\{X_n^i=1\}}\right)}{\mathbb{P}\left(X_n^i = 1\right)} = \frac{\sum_{\mathbf{x}\in\mathbb{X}} q^+_{\hat{\xi}(\mathbf{x})} \, \mathbb{1}_{\{x^i=1\}}(\mathbf{x}) \, \pi_{\mathbf{x}}}{\sum_{\mathbf{x}\in\mathbb{X}} \mathbb{1}_{\{x^i=1\}}(\mathbf{x}) \, \pi_{\mathbf{x}}} \\
&\approx \frac{\sum_{m=1}^{M} q^+_{\hat{\xi}(\mathbf{x}_m)} \, \mathbb{1}_{\{x_m^i=1\}}(\mathbf{x}_m)}{\sum_{m=1}^{M} \mathbb{1}_{\{x_m^i=1\}}(\mathbf{x}_m)}.
\end{aligned}
\tag{6.5}
$$

Fig. 6.3 (f) shows that generally the agents from block 1 are the better tipping indicators. Moreover, agents 0 and 4 are the overall best indicators of tipping, while agent 6 is the best indicator from block 2. This is probably due to them being connected to the other block, thus increasing the tipping likelihood when they are active. One has to be careful in the interpretation of I^{AB}, it only shows us the correlations of the state of agent i and the forward committor and not a causation, i.e., which agent has the largest individual impact on the overall tipping.

Example 2

As a second example we consider a large population structured into four blocks of different sizes. Block 1 contains 20 agents, all other blocks consist of 25 agents, see Fig. 6.5 (d) for the network. The four blocks are circularly connected, and the network is generated by the stochastic block model where each agent has a wiring probability of 0.9

to agents in the same block and of 0.04 to agents from circularly neighboured blocks. Further we set $e = 0.23$, $p = 0.66$, $\theta = 0.5$. The realization in Fig. 6.4(a) indicates that the dynamics are metastable and shows several tipping events where the majority of agents in one block switch their state. Since the states where the majority in each block is either active or inactive should be rather stable, there are potentially 16 metastable regions. One can assume that the tipping of one block induces neighbouring blocks to also tip.

With a state space size of 2^{95}, the state space of this model is too large for computations and we will therefore search for collective variables. From the results of the previous example, we can expect that the Diffusion Maps coordinates tell the number of active agents in each block. Indeed, the spectral gap of the Diffusion Maps eigenvalues indicates that the intrinsic dimension of the model is $d = 4$ and the projection into the two most dominant collective variables (Fig. 6.4) forms a four-dimensional hypercube, a *tesseract*. The coloring of the states in Fig. 6.4 indicates that the four orthogonal directions encode the number of active agents in each of the four blocks. Thus the corners of the hypercube correspond to the metastable regions where the majority of agents in a certain set of blocks are active and the others not. The edges of the hypercube are much less visible but also present. They are not visited that frequently, since they correspond to the rare transitions where the majority in one block is changing their activity. The faces and the inside of the tesseract are empty since they would correspond to states that are unlikely to visit. For our computations later we will use all four dominant Diffusion Map coordinates.

After discretizing the projected state space into $L = 150$ cells using K-Means and estimating a transition matrix using a trajectory of total length $M = 40,000,000$, we can study the tipping events with TPT (Fig. 6.5). As A we consider states where $\leq 25\%$ of agents are active, and as B the states where $\geq 75\%$ are active. The dominant Diffusion Maps coordinate encodes the number of agents that are active, and from Fig. 6.5 (a) we can see that along this coordinate the forward committor increases in distinct steps from 0 to 1. Due to the faster decorrelation inside each metastable set, i.e., in the regions where agents in the same block are behaving conform, the committor is approximately constant in each metastable set. From the committors we can compute transition statistics, such as the average duration of reactive trajectories $t^{AB} = 79.3$ and their frequency which tells us that a transition from A to B is completed on average every $1/k^{AB} \approx 627$th time step.

Visualizing the TPT results for this example is difficult since the space spanned by the collective variables is 4–dimensional. Therefore, we are again interested in grouping similar states of $\hat{\mathbb{X}}$ together and computing the aggregate currents between these macrostates. In this way we can form a transition network from A via different macrostates to B. Since the system is much larger this time, we want to group cells together which are dynamically close by means of a clustering algorithm such as the

111
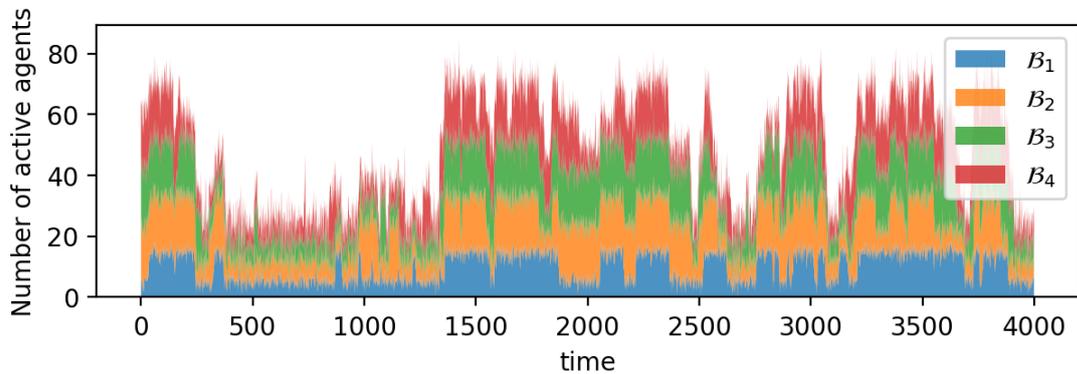
(a)
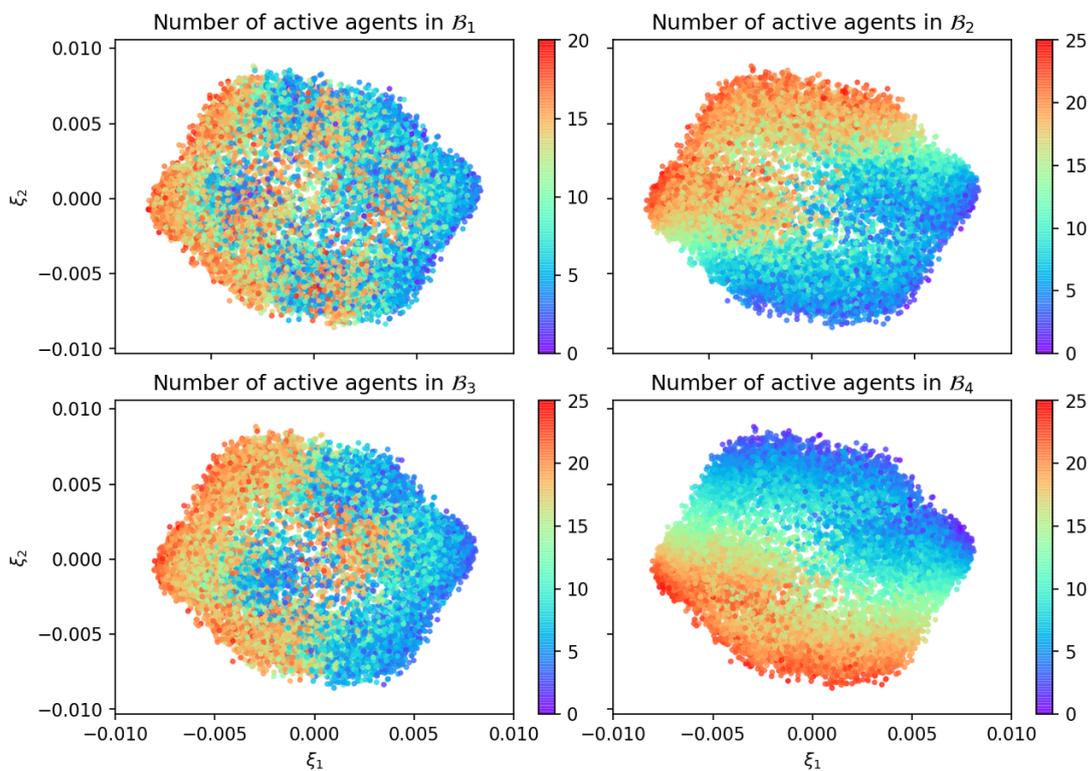
(b)

Figure 6.4: Threshold model on a network of four incomplete blocks as in Example 2: (a) The realization is shown as a stackplot. Several small tipping events (where agents in just one block switch their state) as well as tipping cascades (where nearly all agents change their activity) are apparent. (b) The Diffusion Maps projection into the first two coordinates is colored according to the number of active agents in each block which suggests the tesseract structure. The Diffusion Maps scale parameter came out as $\epsilon = 0.15$.
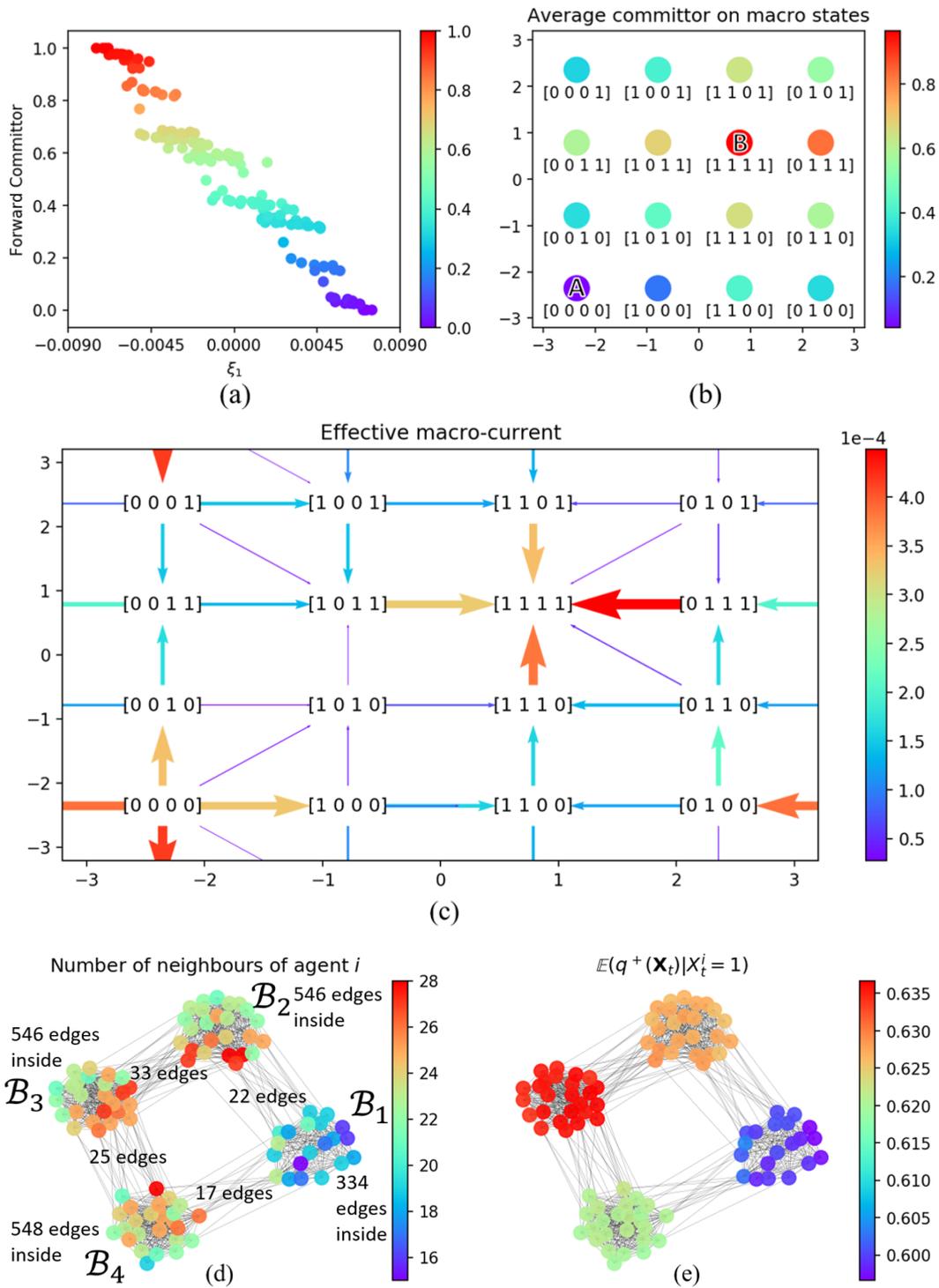
Figure 6.5: Tipping analysis of Example 2: (a) Forward committor against the dominant Diffusion Maps coordinate. (b) Mean forward committor on the macrostates that are placed on a torus. We denoted macrostates as a 4–D vector of 0's and 1's encoding the majority activity in each of the four blocks, e.g., $[0, 0, 1, 0]$ reads as majority of agents in block 1,2 and 4 are inactive and majority in block 3 is active. (c) Effective macro-current on the torus, the color and width of the arrow indicates the magnitude of the current. (d) Number of neighbours of each agent as well as the total number of connections inside and between blocks. (e) Agents as indicators of overall tipping.

well-known clustering method PCCA+ [59].[3] Since we expect the macrostates to be of the form where the majority of agents in each block is either inactive or active, we cluster the state space into 2^4 macrostates. These macrostates also correspond to the corners of the tesseract. Since the net of a tesseract can be visualized nicely on a 2−D torus, we placed the macrostates and resulting transition network on a 2−D torus.

In Figs. 6.5 (b) and (c) we show the mean forward committor of the macrostates as well as the resulting transition network given by the effective macro-current. The macro-current is larger between macrostates where a neighbouring block tips than when a non-neighbouring block tips. Thus the dominant pathways from *A* to *B* are of the form of a tipping cascade from one block to its neighbours and then to their neighbours etc. The macro-current also indicates that it is most likely for block 4 to tip first and for block 1 to tip last. This can be explained as follows: Every agent in block 4 has on average 21.92 neighbours from the same block and 1.68 neighbours from the other blocks. Compared to the other blocks, agents from block 4 have the highest proportion of neighbours from the same block. Thus block 4 is the most independent block and therefore can change its activity most freely. The role of block 1 is also special. It is the smallest block with only 20 agents and also the block where each agent has the largest proportion of extraneous neighbours. The role of block 1 is also reflected in the mean forward committor values: Out of all the macrostates, where only one block has tipped, the committor is the smallest when only block 1 has tipped. This indicates that when block 1 has tipped, it easily tips back due to the strong influence from its neighbouring blocks. Moreover, out of all the macrostates where three blocks have tipped, the forward committor is the highest when block 1 is the still inactive block.

For the network of four blocks we can study which agents are the best indicators of the overall tipping (Fig. 6.5 (e)). We can immediately see that the values of I_i^{AB} do not differ that much for the different agents, possibly due to the four blocks being of a rather similar size and similarly connected. Still, block 3 seems to result in the highest expected forward committor when an agent of that block is active. Block 3 has the most connections to other blocks, and can therefore possibly exert the most influence on neighbouring blocks. This might explain why the expected forward committor is the largest when an agent from block 3 is active.

6.2.2   An oscillating, bivariate complex contagion model

We now come to a second agent-based model where we are considering the changes of binary opinions and separately of binary actual behavioural choices with respect to a certain practice, such as a climate-friendly lifestyle or a certain preventive measure

---

[3]The advantage of the fuzzy method PCCA+ compared to other fuzzy clustering algorithms is that it takes the dynamical information into account and results in a clustering that tries to preserve the slow time scales of the dynamical process. We will use PCCA+ for non-reversible processes [17, 22, 55], which takes the dominant real Schur vectors of the transition matrix and by a linear transformation maps them into a set of non-negative membership vectors that form a partition of unity and are as crisp as possible. Other optimization criteria also exist [59]. By assigning each state to the cluster with the highest membership value, we can make the clustering crisp.

against an epidemic. For illustrative purposes, we use the context of climate-friendly lifestyles and the metaphor of "green" behaviour. We hence say that each agent has a *non-green* or *green* opinion, and also displays a *non-green* or *green* actual behaviour.

The model again considers a complex contagion process, where the social reinforcement from multiple agents at the same time is needed for an agent to change its state. But this time an agent's state has two components: opinion and actual behavioural. The model describes that opinions and actual behaviours do not always have to be aligned. There might be a time lag between holding a certain opinion and behaving accordingly. Additionally, the incentive to hold a certain opinion can drop when many agents in the neighbourhood are behaving in that way. The resulting dynamics is oscillatory.

In detail the model is formulated as follows:

Setting. We consider a system of N_A agents, each agent i with a binary opinion $O_n^i \in \{0, 1\}$ and a binary behaviour $B_n^i \in \{0, 1\}$ at time n. For illustration we consider 0 as *non-green* and 1 as *green*.

We again assume a static interaction network $\mathcal{G}$ with many distinct communities or blocks, e.g, generated by the stochastic block model. Since each agent i is influenced by the average state of agents within the same block, we define for each agent $i \in \mathcal{B}_k$ the following block fraction:

$$\bar{O}_n^i := \frac{|\{j \in \mathcal{B}_k : O_n^j = 1\}|}{|\mathcal{B}_k|},$$ (6.6)

i.e., the fraction of agents with a *green* opinion in the same block as i, and

$$\bar{B}_n^i := \frac{|\{j \in \mathcal{B}_k : B_n^j = 1\}|}{|\mathcal{B}_k|},$$ (6.7)

the fraction of agents with a *green* behaviour in the same block. Note that these quantities, viewed as functions of the agents' index i, are constant on each block.

Interaction rules. At each discrete time point n, each agent i independently chooses two distinct neighbours j, k uniformly at random.[4]

A change in the actual behaviour of agent i is triggered when the two chosen neighbours both display the opposite behaviour and the change is made more likely by the respective opinion in the agent's block. In particular,

  – if $B_n^i = 0$ and $B_n^j = B_n^k = 1$: agent i changes its behaviour to $B_{n+1}^i = 1$
    with probability $\tau \, (b \, \bar{O}_n^i + (1 - b))$,

  – if $B_n^i = 1$ and $B_n^j = B_n^k = 0$: agent i changes its behaviour to $B_{n+1}^i = 0$
    with probability $\tau \, (c \, (1 - \bar{O}_n^i) + (1 - c))$,

---

[4]We assume that each agent has at least two neighbours.

where the parameters $b, c \in [0,1]$ determine how strongly a *green* resp. *non-green* change in behaviour is influenced by the opinions in the block, while the general rate parameter $\tau \in (0,1)$ is for scaling the amount of change per time step.

As a result, an agent has a chance of changing its behaviour to *green* when interacting with two neighbours of *green* behaviour and the more likely the more agents in his block have a *green* opinion.[5] An agent can change its behaviour to *non-green*, when interacting with two neighbours of *non-green* behaviour and the more likely the more agents in his block show a *non-green* behaviour.

Now we come to the opinion changes. An agent i is triggered to switch its opinion when the two chosen neighbours j, k both hold the opposite opinion and the change of opinion is amplified if the block displays the contrary behaviour to the neighbouring agents, in particular

– if $O_n^i = 0$ and $O_n^j = O_n^k = 1$: agent i changes its opinion to $O_{n+1}^i = 1$
$\qquad$ with probability $\tau\,(f\,(1 - \bar{B}_n^i) + (1 - f))$,

– if $O_n^i = 1$ and $O_n^j = O_n^k = 0$: agent i changes its state $O_{n+1}^i = 0$
$\qquad$ with probability $\tau\,(g\,\bar{B}_n^i + (1 - g))$,

where the parameters $f, g \in [0,1]$ determine how strongly a *green* resp. *non-green* change in opinion is influenced by the actual behaviour in the block.

Thus the opinion changes are the other way around, when an agent with a certain opinion (e.g., *green*) meets two neighbours of a different opinion (e.g., *non-green*) the change probability is higher the more agents in his block do not show this behaviour (i.e., the more show a *green* behaviour).

Last, when the chosen neighbours j, k do not both display the opposite behaviour resp. opinion, agent i changes its behaviour resp. opinion only with a small exploration probability e. The exploration probability e should be small compared to τ such that the dominant dynamics is given by the interactions and not by exploration.

Markov chain. The dynamics of the whole population can again be viewed as a Markov chain $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ on the state space $\mathbb{X} = \{0,1\}^{2 \times N_A}$, where we denote the population state at time n by $\mathbf{X}_n = (\mathbf{B}_n, \mathbf{O}_n) = (B_n^i, O_n^i)_{i=1,\dots,N_A}$. Requiring $0 < e, \tau < 1$ ensures that the Markov chain is irreducible and aperiodic.

Resulting dynamics. Let us consider a fully-connected population, in other words a complete network, and choose a large block influence strength $b, c, f, g = 0.7$. Since an agent first has to interact with two agents of a different state in order to have a high chance for switching its state, it is hard for the dynamics to escape from a situation where agents in a block have converged. Consequently, the Markov chain is metastable

---

[5]Note that if we disregard the rate τ, the first behaviour-change probability is a convex combination with factor b between the probabilities $\bar{O}_n^i$ ("fraction in block with *green* opinion") and 1 ("change with certainty to the behaviour of the two chosen neighbors").

with four metastable regions where the large majority of agents share the same opinion and display the same behaviour (either the one aligned with the shared opinion or the opposite one). The realization in Fig. 6.6(a) shows that the dynamics cycles in one direction through the four possible metastable regions. Due to these cyclic dynamics the Markov chain is strongly non-reversible. Starting from a majority in the population with a *non-green* opinion and behaviour, first the majority changes their opinion to green, then after some time they switch their behaviour also to green, followed by a change to a *non-green* opinion, and then also a change to a *non-green* behaviour. Since the angle of rotation in the coordinate plane, also called *phase*, contains all the information about the dynamics, we can essentially reduce the plot to 6.6(b), where we show how the *phase* $\theta_n \in [0, 2\pi)$ varies in time. Whenever the phase remains approximately constant for some time, the system is in a metastable region.

(a)   (b)
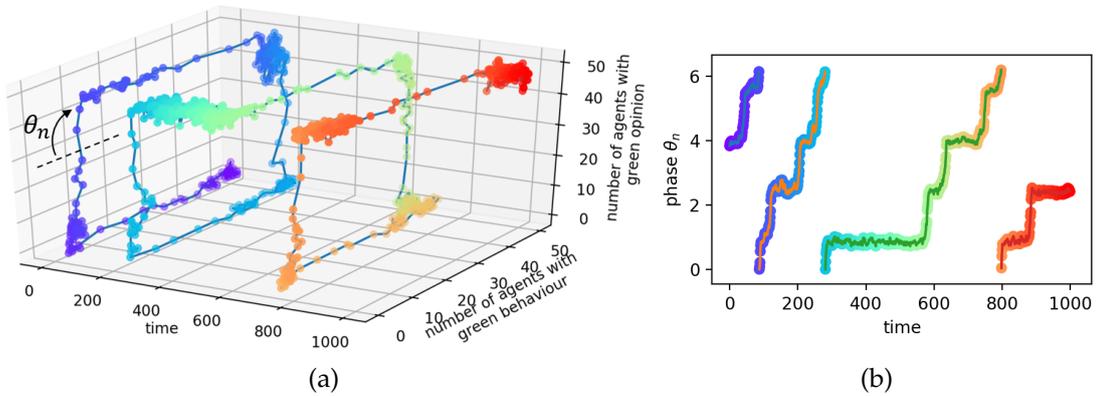
Figure 6.6: Realization of the complex contagion dynamics for one complete network of 50 interacting agents. (a) The dynamics is strongly cyclic in the plane spanned by the two coordinates "number of agents with *green* opinion" and "number of agents with *green* behaviour". (b) Therefore we can also visualize the dynamics by plotting the clockwise-angle in the coordinate plane, i.e., the phase θ_n. The model parameters are $b, c, f, g = 0.7$, $e = 0.02$, $\tau = 0.99$.

We will next study these oscillatory dynamics and tipping pathways on a clustered network of agents.

Example 3

We consider a population of 40 agents split into two blocks that are each complete, i.e., $W_{11} = W_{22} = 1$. The two blocks resemble two coupled oscillators with a coupling strength given by the edge wiring probability between the blocks. An edge wiring probability of $W_{12} = W_{21} = 0.055$ between the two blocks ensures that the two blocks are synchronized only most of the time. The interaction network is shown in Fig. 6.7 (b). The interaction parameters are set to $b = c = f = g = 0.7$, $e = 0.02$ and $\tau = 0.99$. In Fig. 6.7 (a) we show how the phase varies in time for each block $\mathcal{B}_j$ in a short simulation. The block-wise phase θ_n^j at time n is the angle of rotation of the system state in the coordinate plane of "number of agents with *green* opinion in block $\mathcal{B}_j$" versus "number

117

of agents with *green* behaviour in block $\mathcal{B}_j$" as measured from the center point $(10, 10)$. We see that most of the time the phases of the two blocks are synchronized or mimicking each other, but they can also be completely out of phase.

To analyse the transition dynamics, we first need to reduce the state space. For this model one would guess that the number of agents with a *green* opinion in each block and the number of agents with a *green* behaviour in each block will constitute good collective variables. Or, since the blocks behave like coupled oscillators, one could use the phase of each oscillator as a coordinate. In Fig. 6.7 we show the Diffusion Maps projection into the dominant three coordinates using again a Gaussian kernel and the Hamming distance. The number of dominant eigenvalues of P^ϵ indicates that the intrinsic dimension is $d = 4$. Still, the three dominant coordinates already visually indicate that the projected data lie on a 4–D hypercube, i.e., a tesseract. Diffusion Maps cannot embed the tesseract onto a 2−D torus only into a 4−D Euclidean space. For ease of visualization, we will further post-process the embedding and project the net of the tesseract onto a 2-D torus. By choosing two two-dimensional planes in $\mathbb{R}^4$ that are meeting orthogonally in the center of the projected tesseract, we can measure the angles in these planes and thereby untangle the net of the tesseract without edges crossing each other, see Fig. 6.7(d) for the resulting net of the tesseract on the 2−D torus. The coloring indicates that indeed the four orthogonal directions encode the number of agents with a *green* behaviour or opinion in each of the blocks. All the computations will still be done using the four Diffusion Maps coordinates, the projection onto the 2-D torus is only for visualization purposes.

To study the dynamics and especially the tipping pathways, we discretize the projected space into $L = 150$ Voronoi cells and estimate a transition matrix on this space using a trajectory of length $M = 100,000$. We are interested in the tipping dynamics between states A, where a majority of more than 80% has a *non-green* opinion and behaviour to states B, where more than 80% of the population have a *green* opinion and behaviour. The results of the TPT analysis are shown in Fig. 6.8.

The forward and backward committor take values close to 0 or 1 for many states, thus they are rather deterministic, see Figs. 6.8 (a), (b). The reason lies in the strong directedness and non-reversibility of the dynamics. When the states with committors close to 1 are visited the system will most likely come from A and directly go to B due to the directedness in the dynamics. The states with a committor near 0 correspond to states where the system most likely comes from B and is send to A. But there are also some states with a committor of around $\frac{1}{2}$ and thus the future states thereafter are less predictable.

We will next look at the transition current in order to understand the possible transition pathways from A to B. Due to the high non-reversibility, the effective current is no longer cycle-free. Instead we can decompose the reactive current into a productive, cycle-free and an unproductive cyclic current as described in Section 4.2.2. From the decomposition in Figs. 6.8 (c), (d) we see that the dominant productive pathways are

(a)

(b)

(c)

(d)
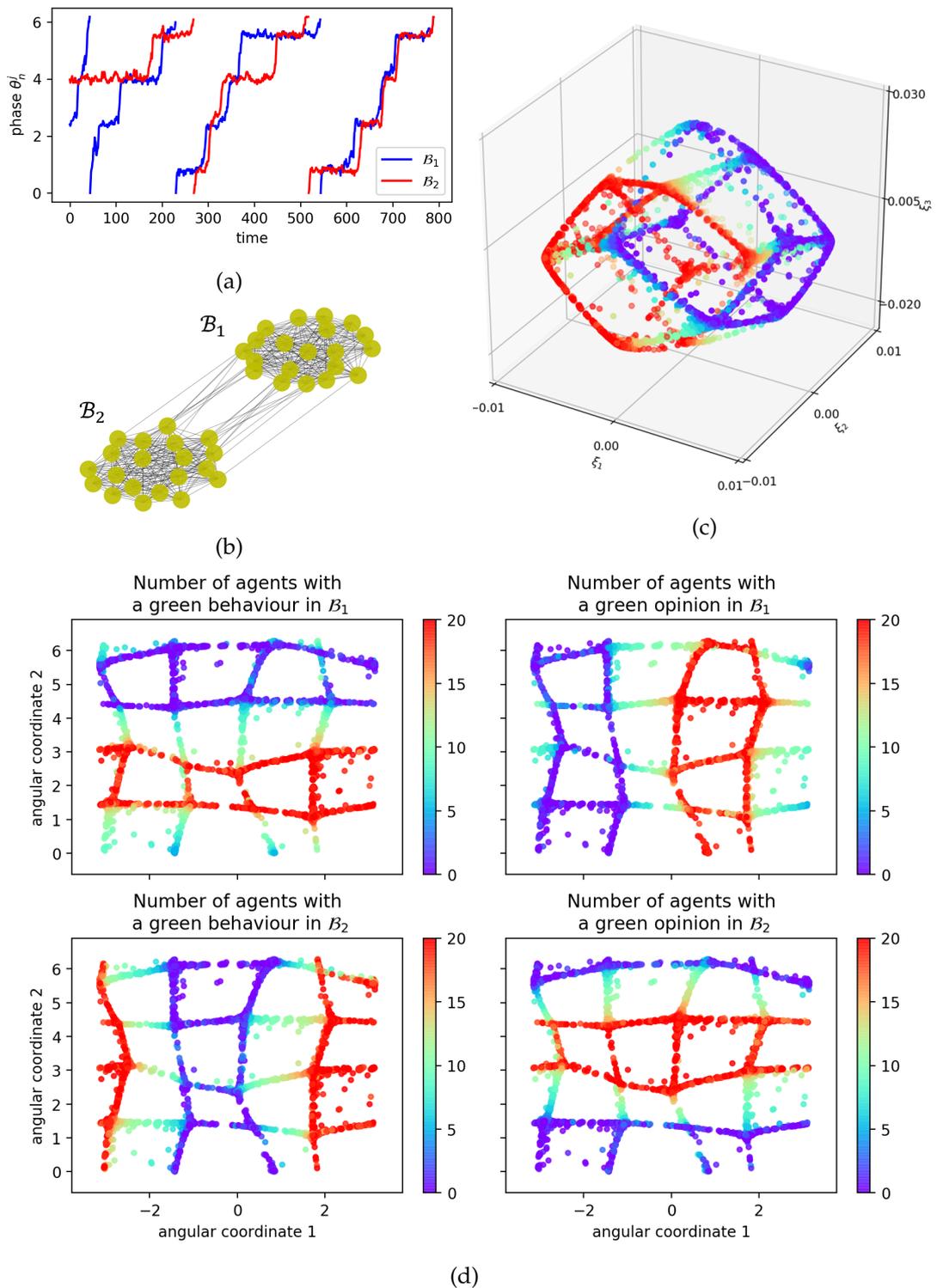
Figure 6.7: Dynamics of the complex contagion model with two blocks as in Example 3: (a) For both blocks we plot how the phase θ_n^j varies in time, mostly they are synchronized. (b) Interaction network. (c) Projection of the data set along the first three Diffusion Maps coordinates with scale parameter $\epsilon = 0.1$ forming a tesseract. (d) Post-processing of the 4D Diffusion Maps projection by visualizing the data points on a torus (this means that the opposite sides of the plot are identified with each other) and thereby unfolding the net of the tesseract. Data points are colored according to the number of agents of a certain opinion and behaviour in each block.

119
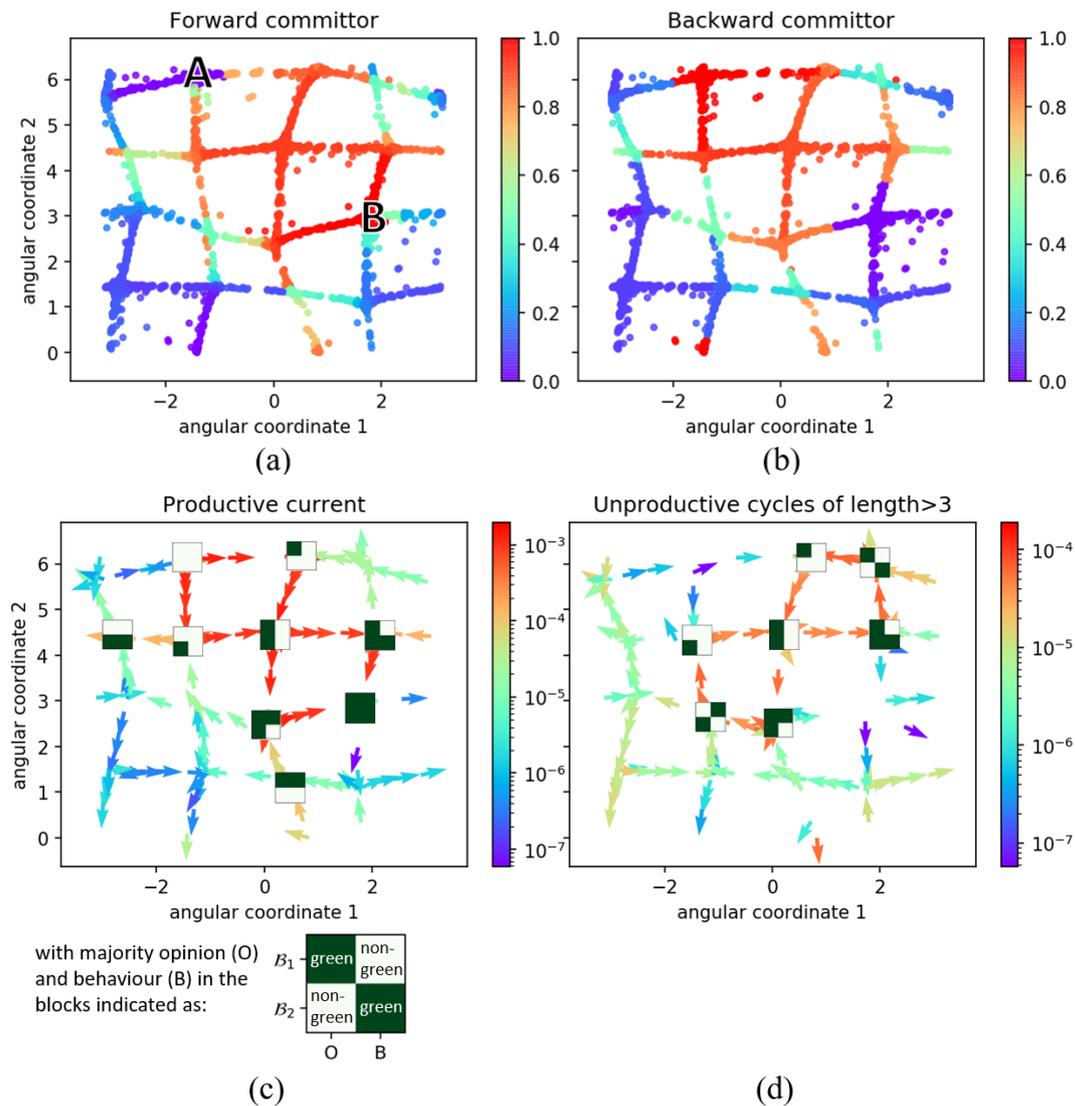
Figure 6.8: Tipping analysis of Example 3: (a) Forward and (b) backward committor. To understand the projected states, one can compare with the coloring in Fig. 6.7 (d). (c) Productive, cycle-free current from A to B (note the logarithmic colour scale). (d) Unproductive current of cycles whose length is larger than 3. For (c) and (d) we labeled the important macrostates by a table that indicates whether for that macrostate the majority in a block has a *non-green* or *green* opinion (O) or behaviour (B).

of the form:

(I) *A* → agents in one of the blocks change their opinion to *green* → agents in the other block change their opinion to *green* → agents in one of the blocks change their behaviour to *green* → agents in the other block change their behaviour to *green* → *B*,

while there are also some less likely productive pathways:

(II) *A* → agents in one of the blocks change their opinion to *green* → agents in the same block change their behaviour to *green* → agents in the other block change their opinion to *green* → agents in the other block change their behaviour to *green* → *B*.

The dominant unproductive cycles are of the general form:

(III) Both blocks display a *non-green* behaviour, majority of agents in block 1 (resp. 2) has a *green* opinion → agents in block 2 (resp. 1) change their opinion also to *green* → agents in block 2 (resp. 1) further switch their behaviour to *green* → agents in block 2 (resp. 1) change their opinion back to *non-green* → then agents in block 2 (resp. 1) change their behaviour back to *non-green*.

In these unproductive cycles, one block does a solo-cycle through the behaviour and opinion space. These are also common in coupled oscillators and called "2π phase jumps" [54, 2].

By comparing the strength of the current along the dominant productive paths (I) (around $9 \times 10^{-4} - 10^{-3}$) with the values of the current along the dominant unproductive cycles (III) (around $4 \times 10^{-5} - 6 \times 10^{-5}$) in Figs. 6.8 (c) and (d), we can deduce that the pathways (I) are visited $15 - 25$ times as much as the dominant cyclic structure (III). Beyond the dominant pathways, we can give some general quantitative statements: Conditional on being on a reactive trajectory, the probability to be on a productive path is $(H^{AB})^{-1} \sum_{i,j} f_{ij}^{P} = 0.175$, while the probability to be on a cycle of length > 3 is $(H^{AB})^{-1} \sum_{i,j} f_{ij}^{U,>3} = 0.014$. The remaining conditional probability is attributed to cycles of length ≤ 3.

6.3   Conclusion

For dynamics of large agent populations and hence large state spaces the committor equations become impossible to solve and a TPT analysis therefore infeasible. A meaningful reduction of the state space is necessary to investigate the ABM's tipping behaviour with Transition Path Theory.

Using two stationary ABMs as examples, we have shown in this chapter how ABMs can be reduced using collective variables and how their tipping behaviour can be studied. The dynamics of most ABMs is predominantly concentrated on a subset of

the state space. In the two considered models, agents are strongly affiliated with a subpopulation (also called block). Due to the local interaction rules, the population states where the individual agents in the same subpopulation agree in their actions or attitudes are predominantly visited and metastable. Rarely it can happen that one agent after the other in a single subpopulation changes their state. This chain reaction represents a transition from one metastable state to another of the ABM. We relied on a long ergodic simulation of the model to cluster around the subset of predominantly occupied population states and of the rare transition paths. Then we employed Diffusion Maps to parametrize the graph induced by this data set. Diffusion maps do not not use the temporal ordering of the population states but simply the Hamming distance as a local distance measure between two population states to construct a weighted graph between the data points. Since individual agents in the ABM can autonomously flip their state according to the rules of the model, the Hamming distance is a suitable distance measure reflecting also the dynamical closeness of two population vectors. The found collective variables showed that the population states concentrate around a structure resembling a hypercube. The corners of the hypercube represent the metastable states while the edges make up the transition paths between metastable states. Remarkably the coordinates only provide information about the number of agents of a certain block that are in some state and not which agents exactly, and thus they also reflect the approximate symmetry between agents of the same block. By discretizing the space spanned by the collective variables and estimating the effective dynamics, we arrived at a meaningful, much smaller Markov chain.

We applied Transition Path Theory to the reduced Markov chain to quantify the tipping dynamics. We could for instance reveal the agents that are most indicative of an impending tipping event, the dominant cascading pathways as well as the likely cyclic dynamics on the paths from A to B. In the two examples, noise-induced tipping between the two extreme metastable regimes happens as a tipping cascade among connected subpopulations. A subpopulation that tips triggers connected subpopulations to also tip. Note that by considering each subpopulation as a separate system, the remainder of subpopulations plays the role of an external influence to that subpopulation. In that way the tipping of a subpopulation can also be considered as triggered by external parameter variations.

Summary

The interesting dynamical regimes in agent-based models (ABMs) of social dynamics are the transient dynamics leading to metastable or absorbing states, and the transition paths between metastable states possibly caused by external influences. In this thesis, we are particularly interested in the pathways of rare and critical transitions such as the tipping of the public opinion in a population or the forming of social movements. For a detailed quantitative analysis of these transition paths, we consider the agent-based models as Markov chains and employ Transition Path Theory. Since ABMs are usually not considered in stationarity and possibly even forced, we generalize Transition Path Theory to time-dependent dynamics, for example on finite-time intervals or with periodically varying transition probabilities. We also specifically consider the case of dynamics with absorbing states and show how the transitions prior to absorption can be studied. These generalizations can also be useful in other application domains such as for studying tipping in climate models or transitions in molecular models with external stimuli.

Another obstacle when analysing the dynamics of agent-based models is the large number of agents resulting in a high-dimensional state space for the model. However, the emergent dynamics of the ABM usually has significantly fewer degrees of freedom and many symmetries enabling a model reduction. On the example of two stationary ABMs we demonstrate how a long model simulation can be employed to find a lower-dimensional parametrization of the state space using a manifold learning algorithm called Diffusion Maps. In the considered models, agents adapt their binary behaviour to the local neighbourhood. When the interaction network consists of several densely connected communities, the dynamics result in a largely coherent behaviour in each community. The low-dimensional structure of the state space is therefore a hypercube. The corners represent metastable states with coherent agent behaviour in each group and the edges correspond to transition paths where agents in a community change their behaviour through a chain reaction. Finally, we can apply Transition Path Theory to the effective dynamics in the reduced space to reveal, for example, the dominant transition paths or the agents that are most indicative of an impending tipping event.

Zusammenfassung

Die interessanten dynamischen Regime in agentenbasierten Modellen sind einerseits die Transienten die zu metastabilen oder absorbierenden Zuständen führen, und andererseits die Übergangsdynamiken zwischen metastabilen Zuständen. Wir interessieren uns in dieser Arbeit insbesondere für die kritischen und seltenen Übergänge wie zum Beispiel das Umkippen der öffentlichen Meinung oder die Bildung von sozialen Bewegungen. Diese Übergangsdynamiken wollen wir in agentenbasierten Modellen, die wir hier als Markov Ketten betrachten, mithilfe von Transition Path Theory analysieren. Da agentenbasierte Modelle in der Regel als nicht-stationär und möglicherweise sogar unter externer Beinflussung betrachtet werden, verallgemeinern wir Transition Path Theory auf zeitabhängige Dynamiken, zum Beispiel auf endlichen Zeitintervallen oder mit periodisch variierenden Übergangswahrscheinlichkeiten. Wir betrachten auch speziell den Fall einer Dynamik mit absorbierenden Zuständen und zeigen, wie die Übergänge vor der Absorption untersucht werden können. Diese Verallgemeinerungen können auch in anderen Anwendungsbereichen nützlich sein, z.B. bei der Untersuchung von Kippvorgängen in Klimamodellen oder in Molekülmodellen unter externer Beeinflussung.

Ein weiteres Hindernis bei der Analyse von agentenbasierten Modellen ist der hochdimensionale Zustandsraum aufgrund der vielen Agenten. Die emergente Dynamik hat jedoch oftmals deutlich weniger Freiheitsgrade und zusätzlich viele Symmetrien, was eine Modellreduktion ermöglicht. Am Beispiel von zwei Modellen zeigen wir, wie eine lange Modellsimulation eingesetzt werden kann, um mithilfe von Diffusion Maps (einem Manifold Learning Algorithmus) eine niedrigdimensionale Parametrisierung des Zustandsraumes zu finden. In den betrachteten Modellen passen Agenten ihr binäres Verhalten der lokalen Nachbarschaft an. Wenn die Interaktionsnetzwerke aus mehreren dicht verbundenen Gemeinschaften bestehen, ist das Resultat ein weitgehend kohärentes Verhalten in jeder Gemeinschaft. Die niedrigdimensionale Struktur der Zustandsraumes ist daher ein Hyperwürfel. Die Ecken stellen metastabile Zustände mit kohärentem Agentenverhalten in jeder Gruppe dar und die Kanten entsprechen Übergangspfaden, bei denen Agenten in einer Gemeinschaft durch eine Kettenreaktion ihr Verhalten ändern. Auf die effektive Dynamik in dem reduzierten Raum können wir schlussendlich Transition Path Theory anwenden um z.B. die dominanten Übergangspfade aufzudecken oder die Agenten zu finden, die am meisten auf ein bevorstehendes Kippereignis hinweisen.

Appendix

For convenience we give the definition of a weakly chained diagonally dominant matrix. This matrix property is used several times in this thesis to show that a matrix is invertible.

Definition A.1 (Weakly chained diagonally dominant matrices [4])**.** *A complex square matrix $A = (a_{ij})$ is termed **weakly chained diagonally dominant (w.c.d.d.)** if the following two properties hold:*

(i) *each row is **weakly diagonally dominant**, i.e., $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$ holds for all i*

(ii) *whenever $|a_{ii}| = \sum_{j \neq i} |a_{ij}|$ for a row i, there exists a path in the matrix $i \to i_1 \to i_2 \cdots \to i_k$ with $a_{ii_1}, a_{i_1 i_2} \ldots, a_{i_{k-1} i_k} \neq 0$ and such that the path ends at a row i_k that is **strictly diagonally dominant**, i.e., such that $|a_{i_k i_k}| > \sum_{j \neq i_k} |a_{i_k j}|$.*

Proposition A.2. *A weakly chained diagonally dominant matrix is invertible.*

The proof can be found in [4, Lemma 3.2.].

Bibliography

[1] Hassan Alkhayuon, Rebecca C Tyson, and Sebastian Wieczorek. Phase tipping: how cyclic ecosystems respond to contemporary climate. *Proceedings of the Royal Society A*, 477(2254):20210059, 2021.

[2] Alex Arenas, Albert Díaz-Guilera, Jürgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics reports*, 469(3):93–153, 2008.

[3] Peter Ashwin, Sebastian Wieczorek, Renato Vitolo, and Peter Cox. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1962):1166–1184, 2012.

[4] Parsiad Azimzadeh and Peter A Forsyth. Weakly chained matrices, policy iteration, and impulse control. *SIAM Journal on Numerical Analysis*, 54(3):1341–1364, 2016.

[5] Ralf Banisch and Nataša Djurdjevac Conrad. Cycle-flow–based module detection in directed recurrence networks. *EPL (Europhysics Letters)*, 108(6):68008, 2015.

[6] Ralf Banisch, Natasa Djurdjevac Conrad, and Christof Schütte. Reactive flows and unproductive cycles for random walks on complex networks. *The European Physical Journal Special Topics*, 224(12):2369–2387, 2015.

[7] Sven Banisch. *Markov chain aggregation for agent-based models*. Springer, 2015.

[8] Tyrus Berry and John Harlim. Variable bandwidth diffusion kernels. *Applied and Computational Harmonic Analysis*, 40(1):68–96, 2016.

[9] Anton Bovier and Frank Den Hollander. *Metastability: a potential-theoretic approach*, volume 351. Springer, 2016.

[10] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media, 2013.

[11] Maria Cameron and Eric Vanden-Eijnden. Flows in complex networks: theory, algorithms, and application to Lennard–Jones cluster rearrangement. *Journal of Statistical Physics*, 156(3):427–454, 2014.

[12] Paul Cilliers. Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management*, 5(02):135–147, 2001.

[13] Paul Cilliers and David Spurrett. Complexity and post-modernism: Understanding complex systems. *South African Journal of Philosophy*, 18(2):258–274, 1999.

[14] Ronald R Coifman and Stéphane Lafon. Diffusion Maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[15] Ronald R Coifman and Stéphane Lafon. Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions. *Applied and Computational Harmonic Analysis*, 21(1):31–52, 2006.

[16] Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-stationary distributions: Markov chains, diffusions and dynamical systems*. Springer Science & Business Media, 2012.

[17] Natasa Djurdjevac Conrad, Marcus Weber, and Christof Schütte. Finding dominant structures of nonreversible markov processes. *Multiscale Modeling & Simulation*, 14(4):1319–1340, 2016.

[18] John N Darroch and Eugene Seneta. On quasi-stationary distributions in absorbing discrete-time finite markov chains. *Journal of Applied Probability*, 2(1):88–100, 1965.

[19] Mark M Dekker, Anna S Von Der Heydt, and Henk A Dijkstra. Cascading transitions in the climate system. *Earth System Dynamics*, 9(4):1243–1260, 2018.

[20] Marc Diener. The canard unchained or how fast/slow dynamical systems bifurcate. *The Mathematical Intelligencer*, 6(3):38–49, 1984.

[21] Weinan E and Eric Vanden-Eijnden. Towards a theory of transition paths. *Journal of statistical physics*, 123(3):503, 2006.

[22] Konstantin Fackeldey, Alexander Sikorski, and Marcus Weber. Spectral clustering for non-reversible Markov chains. *Computational and Applied Mathematics*, 37(5):6376–6391, 2018.

[23] Justin Finkel, Dorian S Abbot, and Jonathan Weare. Path properties of atmospheric transitions: illustration with a low-order sudden stratospheric warming model. *Journal of the Atmospheric Sciences*, 77(7):2327–2347, 2020.

[24] Justin Finkel, Robert J Webber, Edwin P Gerber, Dorian S Abbot, and Jonathan Weare. Learning forecasts of rare stratospheric transitions from short simulations. *Monthly Weather Review*, 149(11):3647–3669, 2021.

[25] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[26] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[27] Lukas Halekotte and Ulrike Feudel. Minimal fatal shocks in multistable complex networks. *Scientific reports*, 10(1):1–13, 2020.

[28] Luzie Helfmann, Enric Ribera Borrell, Christof Schütte, and Péter Koltai. Extending transition path theory: Periodically driven and finite-time dynamics. *Journal of nonlinear science*, 30(6):3321–3366, 2020.

[29] Luzie Helfmann, Jobst Heitzig, Péter Koltai, Jürgen Kurths, and Christof Schütte. Statistical analysis of tipping pathways in agent-based models. *The European Physical Journal Special Topics*, pages 1–23, 2021.

[30] Luis R Izquierdo, Segismundo S Izquierdo, Jose Manuel Galan, and José Ignacio Santos. Techniques to understand computer simulations: Markov chain analysis. *Journal of Artificial Societies and Social Simulation*, 12(1):6, 2009.

[31] Jenna R. Jambeck, Roland Geyer, Chris Wilcox, Theodore R. Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. Plastic waste inputs from land into the ocean. *Science*, 347:768–771, 2015.

[32] Da-Quan Jiang and Donghua Jiang. Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems, 2004.

[33] Sophia L Kalpazidou. *Cycle representations of Markov processes*, volume 28. Springer Science & Business Media, 2007.

[34] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[35] Peter E Kloeden and Eckhard Platen. *Numerical Solution of Stochastic differential Equations*. Springer, 1992.

[36] Péter Koltai and Stephan Weiss. Diffusion Maps embedding and transition matrix analysis of the large-scale flow structure in turbulent Rayleigh–Bénard convection. *Nonlinearity*, 33(4):1723, 2020.

[37] Christian Kuehn. *Multiple time scale dynamics*, volume 191. Springer, 2015.

[38] Timothy M Lenton, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping elements in the earth's climate system. *Proceedings of the national Academy of Sciences*, 105(6):1786–1793, 2008.

[39] Ping Liu, Hannah R Safford, Iain D Couzin, and Ioannis G Kevrekidis. Coarse-grained variables for particle-based models: Diffusion Maps and animal swarming simulations. *Computational Particle Mechanics*, 1(4):425–440, 2014.

[40] Dario Lucente, Stefan Duffner, Corentin Herbert, Joran Rolland, and Freddy Bouchet. Machine learning of committor functions for predicting high impact climate events. *arXiv preprint arXiv:1910.11736*, 2019.

[41] Dario Lucente, Corentin Herbert, and Freddy Bouchet. Committor functions for climate phenomena at the predictability margin: The example of el nino southern oscillation in the jin and timmerman model. *arXiv preprint arXiv:2106.14990*, 2021.

[42] R. Lumpkin, S. A. Grodsky, L. Centurioni, M.-H. Rio, J. A. Carton, and D. Lee. Removing spurious low-frequency variability in drifter velocities. *J. Atm. Oce. Tech.*, 30:353–360, 2012.

[43] R. Lumpkin and M. Pazos. Measuring surface currents with Surface Velocity Program drifters: the instrument, its data and some recent results. In A. Griffa, A. D. Kirwan, A. Mariano, T. Özgökmen, and T Rossby, editors, *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, chapter 2, pages 39–67. Cambridge University Press, 2007.

[44] Michael W Macy and Robert Willer. From factors to actors: Computational sociology and agent-based modeling. *Annual review of sociology*, 28(1):143–166, 2002.

[45] Christian Marschler, Jens Starke, Ping Liu, and Ioannis G Kevrekidis. Coarse-grained particle model for pedestrian flow using Diffusion Maps. *Physical Review E*, 89(1):013304, 2014.

[46] Philipp Metzner. *Transition path theory for Markov processes*. PhD thesis, Freie Universität Berlin, 2008.

[47] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Transition path theory for Markov jump processes. *Multiscale Modeling & Simulation*, 7(3):1192–1219, 2009.

[48] Philippe Miron, FJ Beron-Vera, Luzie Helfmann, and Péter Koltai. Transition paths of marine debris and the stability of the garbage patches. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(3):033101, 2021.

[49] Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, 106(45):19011–19016, 2009.

[50] James R Norris. *Markov chains*. Number 2 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[51] Karine Nyborg, John M Anderies, Astrid Dannenberg, Therese Lindahl, Caroline Schill, Maja Schlüter, W Neil Adger, Kenneth J Arrow, Scott Barrett, Stephen Carpenter, et al. Social norms as solutions. *Science*, 354(6308):42–43, 2016.

[52] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[53] Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

[54] Arkady Pikovsky, Jürgen Kurths, and Michael Rosenblum. Synchronization: a universal concept in nonlinear sciences, 2003.

[55] Bernhard Reuter, Konstantin Fackeldey, and Marcus Weber. Generalized Markov modeling of nonreversible molecular kinetics. *The Journal of chemical physics*, 150(17):174103, 2019.

[56] Enric Ribera Borrell. From ergodic infinite-time to finite-time transition path theory. Master's thesis, Freie Universität Berlin, 2019.

[57] Paul Ritchie, Özkan Karabacak, and Jan Sieber. Inverse-square law between time and amplitude for crossing tipping thresholds. *Proceedings of the Royal Society A*, 475(2222):20180504, 2019.

[58] Paul DL Ritchie, Joseph J Clarke, Peter M Cox, and Chris Huntingford. Overshooting tipping point thresholds in a changing climate. *Nature*, 592(7855):517–523, 2021.

[59] Susanna Röblitz and Marcus Weber. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification*, 7(2):147–179, 2013.

[60] Marco Sarich. *Projected transfer operators*. PhD thesis, Freie Universität Berlin, 2011.

[61] Marten Scheffer, Steve Carpenter, Jonathan A Foley, Carl Folke, and Brian Walker. Catastrophic shifts in ecosystems. *Nature*, 413(6856):591–596, 2001.

[62] Christof Schütte and Marco Sarich. *Metastability and Markov State Models in Molecular Dynamics*, volume 24. American Mathematical Soc., 2013.

[63] Eugene Seneta. *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.

[64] Herbert A Simon. The architecture of complexity. In *Facets of systems science*, pages 457–476. Springer, 1991.

[65] Alina Sîrbu, Vittorio Loreto, Vito DP Servedio, and Francesca Tria. Opinion dynamics: models, extensions and external effects. In *Participatory sensing, opinions and collective awareness*, pages 363–401. Springer, 2017.

[66] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[67] Erik H Thiede, Dimitrios Giannakis, Aaron R Dinner, and Jonathan Weare. Galerkin approximation of dynamical quantities using trajectory data. *The Journal of chemical physics*, 150(24):244111, 2019.

[68] J Michael T Thompson, HB Stewart, and Y Ueda. Safe, explosive, and dangerous bifurcations in dissipative dynamical systems. *Physical Review E*, 49(2):1019, 1994.

[69] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*, 2011.

[70] Egbert H van Nes, Babak MS Arani, Arie Staal, Bregje van der Bolt, Bernardo M Flores, Sebastian Bathiany, and Marten Scheffer. What do you mean,'tipping point'? *Trends in ecology & evolution*, 31(12):902–904, 2016.

[71] Eric Vanden-Eijnden. Transition path theory. In *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, pages 91–100. Springer, 2014.

[72] Peter Walters. *An introduction to ergodic theory*, volume 79. Springer Science & Business Media, 2000.

[73] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[74] Marc Wiedermann, E Keith Smith, Jobst Heitzig, and Jonathan F Donges. A network-based microfoundation of Granovetter's threshold model for social tipping. *Scientific Reports*, 10(1):1–10, 2020.

[75] Ricarda Winkelmann, Jonathan F Donges, E Keith Smith, Manjana Milkoreit, Christina Eder, Jobst Heitzig, Alexia Katsanidou, Marc Wiedermann, Nico Wunderling, and Timothy M Lenton. Social tipping processes for sustainability: An analytical framework. *arXiv preprint arXiv:2010.04488*, 2020.

[76] Nico Wunderling, Jonathan F Donges, Jürgen Kurths, and Ricarda Winkelmann. Interacting tipping elements increase risk of climate domino effects under global warming. *Earth System Dynamics*, 12(2):601–619, 2021.

Selbstständigkeitserklärung

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht.
Mit einer Prüfung meiner Arbeit durch ein Plagiatsprüfungsprogramm erkläre ich mich einverstanden.

Datum:

Unterschrift:

      Luzie Helfmann