Westfälische
Wilhelms-Universität
Münster

Andrea Winkler

Metabasins
or
a State Space Aggregation for Finite Markov Chains with exponentially small Transition Probabilities

2012

Mathematik

Metabasins
or
a State Space Aggregation for Finite Markov Chains with exponentially small Transition Probabilities

Inauguraldissertation
zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften
durch den Fachbereich Mathematik und Informatik
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Andrea Winkler

geboren in
Lüdinghausen

2012

Dekan:	Prof. Dr. Martin Stein
Erstgutachter:	Prof. Dr. Gerold Alsmeyer
Zweitgutachter:	Prof. Dr. Matthias Löwe
Tag der mündlichen Prüfung:	23. Januar 2013
Tag der Promotion:	23. Januar 2013

Zusammenfassung

Die vorliegende Arbeit behandelt das in Physik und Chemie eingeführte, aber bisher nur mittels Computersimulationen studierte und weder dort noch in der Mathematik rigoros definierte Phänomen der *Metabassins* - einer nach genauen Aggregationsvorschriften aus einem gegebenen endlichen Simulationspfad gebildeten Partition des Zustandsraumes gewisser physikalischer Systeme. Die zentrale Herausforderung dieser Arbeit liegt in der Herleitung und Analyse eines pfadunabhängigen Ansatzes, der gewisse, für Metabassins numerisch nachgewiesene Eigenschaften erfüllt (siehe Eigenschaften 1–5 in der Einleitung), etwa das Auftreten bestimmter Verweildauern, das Fehlen multipler Hin- und Rücksprünge, die Unabhängigkeit von Eintritts- und Austrittszustand oder die Ähnlichkeit der Energiebarrieren zwischen verschiedenen Bassins. Dies geschieht im Rahmen ergodischer, reversibler endlicher Markovketten mit exponentiell kleinen, durch eine Energiefunktion gegebenen Übergangswahrscheinlichkeiten.

Im ersten Teil der Arbeit wird eine Definition der Metabassins entwickelt, die auf die etablierte Theorie der Metastabilität zurückgeht und die Eigenschaften 1–5 erfüllt. Die Metabassins ergeben sich im Grunde als Täler und Vereinigungen von Tälern der Energielandschaft, wobei - anders als in anderen Arbeiten zu diesem Thema - Täler ganz verschiedener Ordnung oder Stabilität berücksichtigt werden. Nachdem diese Täler und der ihnen innewohnende Stabilitätsbegriff ausführlich eingeführt und studiert wurden, werden die gewünschten Eigenschaften hergeleitet, indem das Verhalten des Prozesses auf den einzelnen Tälern sowie die Übergänge zwischen diesen vollständig beschrieben werden. Dazu werden typische Trajektorien in den einzelnen Tälern bestimmt, durchschnittliche Verweildauern als exponentielle Funktionen der Tiefe der Täler identifiziert und insbesondere ein aggregierter Prozess hergeleitet, der jeweils nur das aktuelle Tal angibt, nicht aber den konkreten Zustand darin. Für diesen wird eine asymptotische (Semi-)Markoveigenschaft nachgewiesen und dessen Übergangswahrscheinlichkeiten werden ermittelt. Damit kann anschließend gezeigt werden, dass die Wahrscheinlichkeit für multiple Hin- und Rücksprünge genau dann klein ist, wenn die Energiebarrieren in etwa gleich hoch sind.

Der zweite Teil dieser Arbeit widmet sich der Güte der hergeleiteten Aggregation, indem die Übereinstimmung mit der pfadabhängigen Definition sowie die Auswirkung der Aggregation auf die Misch-, Überdeckungs- und Trefferzeiten studiert wird. Die Wahrscheinlichkeit für die Übereinstimmung beider Definitionen wird mit Hilfe gewisser Parameter beschrieben, die den Grad der Unordnung im System messen und für hochgradig ungeordnete Systeme eine große Wahrscheinlichkeit der Übereinstimmung liefern. Des Weiteren wird gezeigt, dass die Aggregation auf die oben genannten Zeiten einen nahezu vernachlässigbar kleinen Einfluss hat. Bezüglich der Mischzeiten wird dies im Rahmen der allgemeinen Fragestellung untersucht, welche Auswirkungen die Verlangsamung einer Markovkette durch Einführung zufälliger Verweildauern auf die Mischrate hat. Die naheliegende Vermutung, dass sich die Mischzeit bei Verlangsamung in etwa multiplikativ um die durchschnittliche Verweildauer erhöht, kann mit Hilfe von Spektraltheorie und Kopplungsargumenten für markovsche und auch semi-markovsche verlangsamte Ketten unter zusätzlichen Annahmen bewiesen werden. Die Notwendigkeit dieser zusätzlichen Annahmen wird ausführlich erläutert. Außerdem werden verschiedene Konvergenzraten explizit bestimmt, wie etwa die Mischraten der ursprünglichen Kette, der Einschränkung auf ein Metabassin oder der Konvergenz gegen die quasistationäre Verteilung.

Summary

The present work is devoted to a phenomenon known from physics and chemistry that is hitherto studied only by means of computer simulations and neither there nor in mathematics rigorously defined: the notion of *metabasins*. Metabasins are a partition of the state space of certain physical systems according to specific aggregation rules along a given finite simulation path. The main challenge of this thesis lies in the construction and analysis of a path-independent approach providing certain metabasin-intrinsic properties (see Properties 1–5 in the Introduction). These are for instance the occurrence of specific sojourn times, the absence of multiple forward-backward jumps, the independence of entrance and exit state, or the similarity of energy barriers between different basins. The study of this problem is done within the framework of ergodic, reversible finite Markov chains with exponentially small transition probabilities depending on some energy function.

In the first part of this thesis, a definition of metabasins is developed, which relies on the well established theory of metastability and complies with Properties 1–5. These metabasins basically emerge as valleys and unions of valleys of the energy landscape. Unlike similar works on this topic, valleys of completely different order are considered. Having introduced and analyzed those valleys and the notion of stability which is immanent to them in detail, the requested properties are derived. For this purpose, the process behavior on single valleys and the transitions between them are entirely specified. More specifically, typical trajectories on single valleys are determined, the average sojourn times are identified to depend on the depth of the valley in an exponential manner, and a certain aggregated process is defined that detects only the current valley and neglects the specific state therein. For this process, an asymptotic (semi-)Markov property is proved and its transition probabilities are determined. Using these probabilities, multiple forward-backward jumps are shown to be quite unlikely if and only if the energy barriers are approximately of the same height.

The second part of this work is addressed to the goodness of the aggregation and studies the accordance with the path-dependent definition as well as the impact of the aggregation on the mixing-, cover-, and hitting times. The probability of accordance of both definitions is described by means of certain parameters measuring the degree of disorder in the system. For highly disordered systems arises a high probability of accordance. Furthermore, it is proved that the impact of the aggregation on the mixing-, cover-, and hitting times is virtually negligible. For the mixing time, this is shown in the context of the general question about how the deceleration of a Markov chain via additional random sojourn times in every state affects the mixing time. The nearby conjecture is that a deceleration basically increases the mixing performance multiplicatively by the average sojourn time. With spectral theory and coupling arguments, this conjecture is proved to hold true for Markovian and semi-Markovian decelerated chains under some further assumptions. The need for these further assumptions is explained in detail. In addition, different rates of convergence are calculated explicitly, for instance the mixing rates of the original chain, of the restriction to a metabasin, or of the convergence against quasi-stationarity.

Contents

Introduction

Physical Background

Imagine a system of many (ball-like) particles moving in a given space or interacting with each other – a *many-particle system*. Examples are a cup of coffee or a piece of amber where the resin-particles (almost) stopped moving, but also a pattern of spins. Depending on the spatial arrangement of the particles, those systems exhibit different *energies*. For a pattern $\sigma \in \{\pm 1\}^{[1,\dots,N]\times[1,\dots,M]}$ of spins on an $N \times M$-lattice, the energy can be defined as the Hamiltonian Function

$$E(\sigma) \;:=\; -\sum_{i,j} \sigma_i \sigma_j,$$

where the sum ranges over all neighboring pairs $\langle i, j\rangle$ (see e.g. [39, Section 3.3.5]). For example,

$$E\left((+,+,-,+)\right) \;=\; -(1-1-1) \;=\; 1.$$

This model is known as the *Ising model* and extensively studied. For a cup of coffee or a piece of amber, the energy of a specific arrangement

$$\sigma \in \left\{(\sigma_1,\dots,\sigma_N) | \sigma_i \in [a,b]^3, 1 \le i \le N\right\}$$

of N particles in a box $[a,b]^3,\ 0 \le a < b < \infty$, can be defined as the Lennard-Jones-potential [4, Section 18.2.2]

$$E_{d,\varepsilon}(\sigma) \;:=\; \sum_{i\neq j} 4\varepsilon \left(\left(\frac{d}{\|\sigma_i - \sigma_j\|_2} \right)^{12} - \left(\frac{d}{\|\sigma_i - \sigma_j\|_2} \right)^{6} \right).$$

Here the sum ranges over all pairs of particles, d is a constant giving the distance at which the pair interaction changes from repellent to attractive, and ε equals the pair-energy at the distance $\sqrt[6]{2}d$. For example, the 3-particle configuration σ in Figure 0.1 has a Lennard-Jones potential $E_{2,1}(\sigma) = -0.28$. In both cases, the energy function as a function from the state space to $\mathbb{R}$ gives rise to a high-dimensional graph: the *(potential) energy landscape* (PEL).

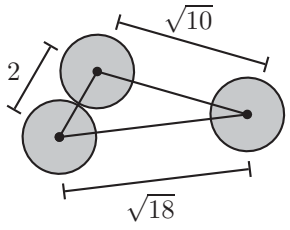

Figure 0.1.: Example of a 3-particle configuration

Now let us follow the trace of our system in the PEL when it traverses through the state space via movements of single particles or flippings of single spins. Every time it descends from an energetically higher state to a lower one, it loses energy, which may go as heat into the surroundings. When a movement increases the energy, the system seizes energy from the surroundings to afford this

ascent. In case it can not afford the ascent, the higher state is not feasible. Thus, by decreasing the temperature of the coffee, more and more states are unfeasible, until at 0℃ the highly ordered, crystalline icy state is the only possible arrangement of the particles in the cup, the *stable state* at that temperature. When increasing the temperature of the surroundings, the frozen coffee will melt again. During the melting, the temperature of the coffee stays at 0℃ and starts increasing not until it is completely liquid. Coffee (more precisely water) has a freezing/melting point at 0℃.

But there are liquids behaving differently, for example liquid amber. When heating up amber, it becomes more and more viscous but there is no accentuated melting point above which it is liquid and below which it is solid. When cooling it down, the particles are less and less mobile, the amber becomes more and more hardened, but there is no freezing point where the particles are well arranged in a specific structure with long-range order. This amorphous behavior is a consequence of the (in comparison for example to water) *highly disordered* energy landscape of amber: Independent of how we cool down water, before being too immobile, the particles always find a path of feasible states to reach the stable state. When cooling down liquid amber, it becomes stuck in some part of the PEL with no feasible transition to the stable crystalline state. The solid structure is a randomly frozen liquid arrangement, the particles are by no means well arranged. Hence, solid amber is an *unstable* particle-arrangement, a *supercooled liquid*.

Physical systems with this behavior are called *(structural) glasses* or *glass formers*, and should be kept in mind while reading this thesis. Examples are of course soda-lime glass known from windows or drinking vessels, but also coal tar, plastics, polymers, or metallic alloys [46, Section 11.8]. Spin patterns, though not belonging to the family of glasses, as well exhibit a high disorder, yet magnetic instead of positional. This disorder causes a phase transition and a spin-glass phase with a randomly frozen pattern [9]. Therefore, they are also called *spin-glasses* and another basic example.

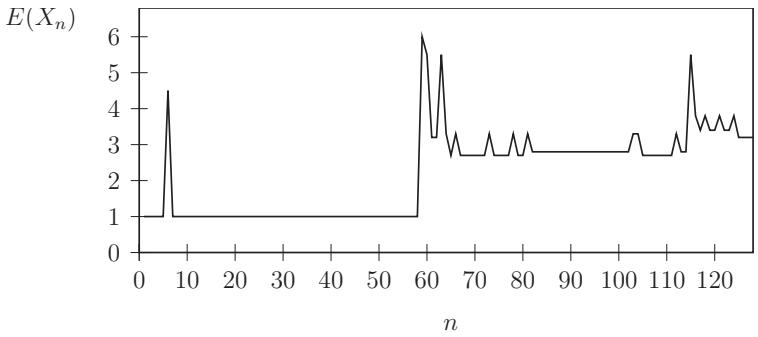

Figure 0.2.: Typical energy trajectory of a glass-forming system.

In modern natural sciences, real phenomenons are more and more studied with IT-techniques. Thus, systems of this kind are simulated on computers as random processes, sampling the state space resp. the energy landscape. Figure 0.2 shows a typical energy trajectory comprising *long sojourn times* in low-energy states with *forward-backward jumps* to neighbors and fast transitions between those very areas. To characterize glass-forming systems at low temperatures via their energy landscape was first done by Goldstein in 1969 [26] ("...when all is said and done, the existence of potential energy barriers large compared to thermal energy are intrinsic to the occurrence of the glassy state...") and has by now become a common method with the major goal to relate dynamics to properties of the PEL. Heuer [29] gives an exhaustive topical review on this issue until 2008. Typically, the trajectory of consecutively visited local minima is studied to determine the *partition function*

$$Z \;=\; \int e^{-\frac{E(\sigma_1,\ldots,\sigma_N)}{T}} d(\sigma_1,\ldots,\sigma_N),$$

or the *entropy* of an energy level e,

$$S(e) \;=\; \ln\left(|\{(\sigma_1,...,\sigma_N)|E(\sigma_1,...,\sigma_N)) = e\}|\right).$$

Heuer's review also includes his own results, namely that those computer simulations have shown important characteristics of such systems to be describable by energetic properties of the so-called *metabasins* (MB) [29, Chapter 6 and 7]. Based on Stillinger's concept of grouping states [60], those MB are formed in the following way by aggregation of suitable states of the describing process $(X_n)_{n\geq 0}$ along a simulated trajectory: Fixing a reasonable observation time T, define $\chi_0 \equiv 0$ and then recursively for $n \geq 1$

$$\chi_n \;:=\; \inf\left\{k > \chi_{n-1} \,|\, \{X_k,..., X_T\} \cap \{X_0,..., X_{k-1}\} = \emptyset\right\}.$$

MB up to $\upsilon := \sup\{n \geq 0 \,|\, \chi_n \leq T\}$ are chosen as

$$\mathcal{V}_n \;:=\; \{X_{\chi_n},..., X_{\chi_{n+1}-1}\}, \quad 0 \leq n \leq \upsilon,$$

each of which comprises typically a large number of minima of the PEL. There is an unprecedented information this aggregation provides for the goal to relate energy and mobility: The diffusion constant D (a system-inherent proportionality constant between the mean squared covered distance and the time [4, Equation 21-83]) and the relaxation time τ_α (the time needed for the supercooled liquid to approach the molecular state of a crystal) are proportional to fractions of moments of the MB-waiting time [29, Chapter 6 and 7]. This is in strong contrast to the unaggregated process, keeping track of every single visited local minimum: As there are very many forward-backward jumps between neighboring minima separated by an energy barrier lower than the barrier to other minima (see Figure 0.2), not only their average waiting time but also the number of these reciprocating jumps influences the time-scale of the long-range transport.

As developed in [29] and [53], the most important advantages of this model reduction, notably including the above one, are (referred to as Properties 1–5 hereafter):

1. Markov Property: The distribution of the successor of one MB is statistically independent of its entrance state. This justifies to lump them together and regard them as one single metastate.

2. Reciprocating Jumps: There are per definition no reciprocating jumps between MB. Hence, only the effective motion is displayed, giving the formula

$$D \;\sim\; \frac{1}{\langle\tau\rangle} \quad \text{and} \quad \tau_\alpha \;\sim\; \frac{\langle\tau^2\rangle}{\langle\tau\rangle},$$

   where $\langle\tau\rangle$ and $\langle\tau^2\rangle$ are the average resp. the average squared MB-waiting time.

3. Exit Time: The average MB-waiting time is proportional to its depth. Together with Property 2, this provides a strong and explicit relation between dynamics (D and τ_α) and thermodynamics (energies).

4. Trap Model: All energy barriers between different MB are approximately of the same height. Thus, for every MB-transition it requires to cross approximately the same energy threshold E_0, though via possibly different states. Such systems are called *trap models* (see [10]).

5. CTRW-Hypothesis: The sojourn times and jump distances between consecutively visited MB (measured in Euclidean distance) form sequences of statistically uncorrelated random variables, which are approximately mutually independent (*continuous time random walk-hypothesis* [53]). This simplifies the analysis in such a way that the diffusion constant equals the average spatial

MB-increment divided by the average MB-waiting time. Surprisingly, the average spatial MB-increment is almost independent of the temperature. Hence, the temperature dependence of the diffusion is completely characterized by the temperature dependence of the average MB-waiting time.

Despite these advantages, the suggested definition of MB has the obvious blemish that it, first, depends on the realization of the considered process and may thus vary from simulation to simulation, and second, proves the diffusion and relaxation to be determined by MB-coarse graining effects, without explaining how exactly the MB look like and why they look like that.

Outline

To provide and analyze a mathematically stringent definition of a *path-independent aggregation* of the state space which maintains the above properties is the principal concern here. We get a deeper understanding of the MB-concept along the way. In this endeavor, we will rely on ideas about *metastability* and *metastable states*, the mathematical framework in which systems such as glass-formers are studied because of their basic phenomenon: the sampling of the state space along trajectories comprising alternately *immobile phases of a pretended equilibrium* and very *mobile phases* sling-shooting the system to distant areas of the state space (see again Figure 0.2).

Over the years, metastability was studied in many different ways using various techniques and is still a field of great interest for complex physical systems. To mention only a few, as one of the first Lebowitz & Penrose [50] in 1971 identified metastable states via their long sojourn in a pretended equilibrium and the small return probability once this pretended equilibrium is left. In 1970, Freidlin & Wentzell started to use large deviation theory to study metastable states in deterministic dynamical systems with a stochastic perturbation tending to zero. We refer to their textbook [24] and the references therein. Influenced by that, for example Olivieri & Scoppola [49], Catoni & Cerf [16], or Beltrán & Landim [7] studied discrete (time and space) Markov chains with transition probabilities depending on a parameter β representing the inverse temperature such that the transition probabilities are either exponentially small in β or bounded away from zero. Cassandro, Galves, Olivieri & Vares [15], Bovier [11] and many more transferred this to stochastic mean field models and spin systems, where in contrast to the above works the number of states tends to infinity instead of the temperature tending to zero. Whereas in [15], following the ideas of [24], the focus is on typical trajectories of the process and their large deviations, in [11] the relation between exit times and capacities is emphasized. These two perceptions established themselves as the *pathwise approach* resp. *potential theoretic approach*. Furthermore, metastability has been linked to spectral methods (Bovier, Eckhoff, Gayrard & Klein [13], Matthieu [44]) and quasi-stationarity (Huisinga, Meyn & Schütte [34], Bianchi & Gaudillière [8]). We want to mention particularly the works of Scoppola ([56] and [55]) and Bovier, Eckhoff, Gayrard & Klein [12], as this thesis is highly influenced by and draws on many ideas and results from these. The former introduced a renormalization procedure that gives rise to an ordering of the states according to their (meta-) stability, and the latter links the metastable behavior of energy-driven systems to the PEL. For more literature and a general overview, we mention the monographs [11], [18] and [19] and the references therein.

This thesis is divided into two parts: the *construction of a definition of metabasins* (Chapters 1–4) and the study of the *goodness of the accompanying aggregation* (Chapters 5–7). Inspired by simulations of glass forming systems at very low temperatures with the Metropolis algorithm, we study reversible, ergodic finite Markov chains with exponentially small transition probabilities, which are determined by an energy function and a parameter $\beta > 0$ giving the inverse temperature. We are interested in the behavior of the process as $\beta \to \infty$ (low-temperature limit) and envisage a highly

disordered energy function. The precise assumptions are listed in Section 1.1. In the remainder of Chapter 1, the foundation towards an aggregation as outlined above is laid. Around certain metastable states (see also [55]), entailing an order from a kind of "weak" to a kind of "strong" metastability, we will define (Section 1.2) and study (Section 1.3) *valleys* characterized by minimal energy barriers and minimal paths. The notion of those energy barriers or saddles is very popular (see [12], [48] or [15] for a few examples), but to our best knowledge this closed definition of valleys, referring to barriers *and* paths, is new and so are the results on their shape. They will prove to be well behaving by being connected, adopting the hierarchical order of their ground states, and possessing a nice nesting structure. This definition of valleys allows a *microscopic view*, where the system behavior is studied when moving within a fixed valley, and a *macroscopic view*, which describes the transitions between valleys.

In Chapter 2, the process is studied from the *microscopic* perception. By slightly modifying results from [12] to comply with our situation, it is shown via variational techniques resp. a reduction to the one-dimensional case that in the limit of low temperatures the minimal energy barriers from the definition of valleys determine the speed resp. probability of a transition between the two states they separate. More precisely, in the limit $\beta \to \infty$, with probability tending to 1, the process reaches a state with lower barrier earlier than one with a higher barrier (Section 2.1). We infer that between the entrance and exit of a valley, with probability tending to 1 as $\beta \to \infty$, the ground state is reached. Section 2.2 regards the time needed to leave a given valley and connects this exit time with the PEL-parameters of the valley, namely the exponential dependence on its depth. This confirms Property 3. Similar forms of this result have been proved in different situations in many different ways, for example in [61, Chapter XI.2] for birth and death processes by solving linear equations of generating functions or in [24, Chapter 4, §4, Theorem 4.1] for continuous (time and space) processes by a geometric trials argument. We derive it with the help of the main theorem for the renormalization procedure in [55] and a discretization of the geometric trials argument in [24]. As a consequence, we obtain that there is no universal time scale since those depths are quite variable and the single time scales vary exponentially. We will also briefly touch on the phenomenon of *quasi-stationarity* in Section 2.3 by showing that the nesting structure of valleys induces an order on the absorption probability of the process with killing on the boundary of the valley.

On the *macroscopic* side in Chapter 3, we start by introducing two macroscopic processes which only detect the current valley and neglect specific states in it. The first one, denoted by $\overline{Y}$, considers the time spent in a valley, the second, denoted by Y, ignores it. With the use of the evolved properties of trajectories within a valley, we derive an asymptotic (semi-)Markovian structure (Section 3.2), confirming Property 1. The limiting (semi-)Markov chain is identified and studied in terms of recurrence and transience. Finally, Section 3.3 examines that, given an appropriate energy landscape, the nesting procedure of valleys annihilates (on the macroscopic scale) the accumulation of reciprocating jumps by merging valleys exhibiting such jumps into a single valley. By giving explicit bounds on macroscopic transition probabilities as consequences of the microscopic results, it will be shown that forward-backward jumps are more and more unlikely when the energy barriers are brought in line. Hence, valleys of sufficiently high order will satisfy Property 2. Finally, we can also identify conditions under which a diffusive behavior as described in Property 5 can be achieved, namely a sufficient homogeneity of the state space in such a way that MB-increments are centered and mutually uncorrelated.

From the first three chapters, we construct the aimed *definition of metabasins* in Chapter 4 which provides Property 4 inevitably. Though the clustering of states into basins of attraction or metastates is a common tool in metastability analysis, our approach differs from the literature by having a different thrust. As an example, Olivieri & Scoppola [49] fully describe the tube of exit from a domain in terms of which basins of attraction of increasing order are visited and for how long these basins are visited. In [7], Beltrán & Landim work with transition rates instead of energies, and

aim at finding a universal depth (and time scale) for all metastates. However, we rather aim at the *finest aggregation such that transitions back to an already visited metastate are very unlikely within a time frame used in simulations*. This finest aggregation will lead to valleys of very variable depth just as simulations do not exhibit a universal depth or timescale. Unlike the afore mentioned, we mix metastable states and valleys of different order to meet Properties 1–5. Most obviously, Property 1 loses its validity when assigning lower valleys to higher-order metastates against our aggregation rule, for which there is no natural way. Furthermore, we are interested in the aggregated processes, either time preserving or with an acceleration depending on the process itself, rather than a "blindly" accelerated version $(X_{nT})_{n\geq 0}$ for some $T \in \mathbb{R}_{>0}$ as in [7], or a version whose Markov property is artificially created as in [49].

Having established Properties 1–5 [Theorem 4.2.1], the second part of this work proceeds in Chapter 5 with a *comparison of the path-independent definition of MB with the path-dependent one*, so as to study the goodness of the above MB-aggregation. After illustrating that the trajectory-dependent procedure can not be valuably applied to all kinds of energy landscapes, we identify system parameters measuring the disorder (Section 5.1). Under some reasonable conditions on this disorder or connectivity, which, in essence, ensure the existence of reasonable path-dependent MB, Section 5.2 presents the accordance of both definitions.

The two remaining chapters deal with the *comparison of the mixing-, cover-, and hitting times* of the original and the aggregated processes. Using spectral-gap techniques as in [30], we will determine the asymptotic geometric mixing rate of the original process X, its restriction to a specific valley, the hit chain on the metastable states, a Markovian macroscopic process and the rate of mixing against the quasi-stationary distribution (Section 6.1). It turns out that all global processes (X, the hit chain and the Markovian macroscopic process) mix with geometric rates showing the same asymptotic behavior for $\beta \to \infty$ (on a logarithmic scale). The geometric rates of the local processes (the restricted and conditioned chains) are smaller but coincide asymptotically, too. This indicates that the aggregation procedure of the first part generates a macroscopic process with the same mixing behavior as X and faster mixing microscopic processes, confirming the metastability. The study of the mixing time of Y and $\overline{Y}$, one being the embedded jump chain of the other, motivates to analyze *how the deceleration of a Markov chain by introducing sojourn times in every state affects the mixing time*. Although results are known for the comparison of the mixing performance of some chains (see for example [21]), this precise problem is not investigated so far. We describe the problem in Section 6.2 and give various results for the three different cases where, first, the decelerated process is again Markovian (Section 6.3), second, the sojourn times are independent of the Markov chain (Subsection 6.4.1), and third, the decelerated process is semi-Markovian (Subsection 6.4.2). Broadly speaking, the deceleration increases the mixing time multiplicatively by the expected sojourn time. The proofs rely on spectral theory and coupling arguments. In Section 6.5, we come back to the originally raised problem and determine the mixing performance of the accelerated aggregated process by applying the theory just introduced. Here again, the mixing time of X is roughly at most the mixing time of Y times the average sojourn time. For the time preserving aggregated chain it is easy to see that it mixes at least as fast as the original process. Examples and preliminary results support the conjecture that they mix asymptotically with the same geometric rate, which remains as an open problem in this work.

In the final Chapter 7, we bound the relative difference between the *cover times* of the original and the aggregated process $\overline{Y}$ (Section 7.1) as well as the relative difference between the *hitting times* of the overall minimum of these two processes (Section 7.2). Though the according time for X is almost surely larger than the one for $\overline{Y}$, both results indicate that the difference is rather small compared to the dimension of those times.

In conclusion, we derive an aggregation procedure satisfying Properties 1–5 and coinciding to a certain extend with the definition established in physics so that the macroscopic process is consistent with the original one (in the studied aspects). Along the way, we develop results about the effect of deceleration of Markov chains on the mixing time.

Notation

Throughout this work, we use the following notation: In every chapter, there is an underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ which is assumed to be large enough to ensure every appearing random variable to be well defined. For stochastic processes on a state space $\mathcal{S}$, the initial distribution is specified via an index, that is $\mathbb{P}_\lambda$ for $\lambda \in \mathfrak{W}(\mathcal{S})$ resp. $\mathbb{P}_x$ for $\lambda = \delta_x, x \in \mathcal{S}$. The expectation with respect to $\mathbb{P}$ or $\mathbb{P}_\lambda$ is denoted by $\mathbb{E}$ or $\mathbb{E}_\lambda$. We use $\mathbb{P}^X$ and $\mathbb{P}_x^X$ as the law of a random variable X under $\mathbb{P}$ resp. under $\mathbb{P}_x$. For $A \subset \mathcal{S}$ and $x \in \mathcal{S}$, we define the entrance times

$$\tau_A := \inf\{n \geq 1 | X_n \in A\}, \quad \tau_A^0 := \inf\{n \geq 0 | X_n \in A\}, \quad \tau_x := \tau_{\{x\}}, \quad \tau_x^0 := \tau_{\{x\}}^0.$$

For two real numbers $a, b \in \mathbb{R}$ we denote with $a \wedge b$ their minimum and with $a \vee b$ their maximum. An empty sum $\sum_{i=n}^m a(i)$, $m < n$, is always defined to equal zero, whereas an empty product $\prod_{i=n}^m a(i)$, $m < n$, is understood as 1. Matrices are always assumed to be finite and real.

Acknowledgement

I would like to express my appreciation to my supervisor Prof. Dr. Gerold Alsmeyer for the trust he placed in me by giving me the opportunity of research on this topic, and for his encouragement and many instructive and fruitful discussions throughout this project. Moreover, my sincere thanks go to Prof. Dr. Andreas Heuer for his interest and many valuable hints, filling the theoretical concepts with physical live, and to Prof. Dr. Matthias Löwe, from whose funds my employment was payed. I am also much obliged to all members of the *Institut für Mathematische Statsitik* at the *WWU Münster* for a working environment on cordial terms, and to friends and family for a social environment exhilarant and supporting; first and foremost Sören Gröttrup, who had a careful look at a draft version of my thesis in the last weeks and at me in the last years.

Part I.

Construction of Metabasins

1. Valleys

Following the outline in the Introduction, we start this part with a precise model description and the analysis of *valleys*, using a definition of metastable states influenced by the one in [55] and the definition of essential saddles as in [12].

1.1. The Model

Let X be a Markov chain on a finite set $\mathcal{S}$ with transition matrix $\mathbf{P} = (p(r,s))_{r,s\in\mathcal{S}}$ and stationary distribution π, and let $E : \mathcal{S} \to \mathbb{R}$ be an energy function such that the following conditions hold:

Irreducibility: $\mathbf{P}$ is irreducible with $p(s,s) \geq \frac{1}{2}$ and $p(r,s) > 0$ if and only if $p(s,r) > 0$ for all $r, s \in \mathcal{S}$.

Transition Probabilities: There exists a parameter $\beta > 0$ and for every $\beta > 0$ some $\gamma_\beta > 0$ with $\lim_{\beta\to\infty} \gamma_\beta = 0$ and $\lim_{\beta\to\infty} \beta\gamma_\beta \in (0,\infty)$ such that

$$e^{-\beta((E(s)-E(r))^+ + \gamma_\beta)} \;\leq\; p(r,s) \;\leq\; e^{-\beta((E(s)-E(r))^+ - \gamma_\beta)}$$

for all distinct $r, s \in \mathcal{S}$ with $p(r,s) > 0$. Furthermore,

$$p^*(r,s) \;:=\; \lim_{\beta\to\infty} p(r,s)$$

exists for all $r, s \in \mathcal{S}$, is positive if $E(r) \geq E(s)$ and $= 0$ otherwise.

Reversibility: The pair $(\pi, \mathbf{P})$ satisfies the detailed balance equations, i.e.

$$\pi(r)p(r,s) \;=\; \pi(s)p(s,r)$$

for all $r, s \in \mathcal{S}$.

Non-Degeneracy: $E(r) \neq E(s)$ for all local minima $r, s \in \mathcal{S}, r \neq s$, of E, where $x \in \mathcal{S}$ is a local minimum of E if $E(x) \leq E(y)$ for every y with $p(x,y) > 0$.

Thus, we are dealing with a reversible Markov chain with exponentially small transition probabilities driven by an energy landscape. The non-degeneracy condition ensures the distinguishability of any two minima by their energy. This will later give us a way to determine the stable and the unstable one of any pair. Furthermore this condition provides that every local minimum is isolated. Results analog to those of this first part have been pre-published in [3] for a more restrictive non-degeneracy condition, namely the injectivity of the energy function.

An example of a Markov chain satisfying the above conditions, and also the main motivation behind this work, is a *Metropolis chain* with transition probabilities

$$p(r,s) \;=\; q(r,s)e^{-\beta(E(s)-E(r))^+}.$$

Here β is the inverse temperature and $(q(r,s))_{r,s\in\mathcal{S}}$ is an irreducible and symmetric transition matrix independent of β - the proposal chain. For $\gamma_\beta := -\min_{r,s} \ln(q(r,s))\beta^{-1}$, where the minimum is taken over all $r, s \in \mathcal{S}$ with $q(r,s) > 0$, the above conditions are fulfilled.

In the following, two states x, y with $p(x,y) > 0$ are called *neighbors* ($x \sim y$) and $\mathcal{N}(x) := \{y \in \mathcal{S} | p(x,y) > 0\}$ the *neighborhood* of x. Let us start with the following basic results for the stationary distribution.

Lemma 1.1.1. *For any two states $r, s \in \mathcal{S}$ we have*

$$e^{-\beta(E(r)-E(s)+2|\mathcal{S}|\gamma_\beta)} \;\leq\; \frac{\pi(r)}{\pi(s)} \;\leq\; e^{-\beta(E(r)-E(s)-2|\mathcal{S}|\gamma_\beta)}.$$

Proof: To start with, assume $r \sim s$. Reversibility and the assumptions on the transition probabilities imply

$$\frac{\pi(r)}{\pi(s)} \;=\; \frac{p(s,r)}{p(r,s)} \;\leq\; \frac{e^{-\beta((E(r)-E(s))^+-\gamma_\beta)}}{e^{-\beta((E(s)-E(r))^++\gamma_\beta)}} \;=\; e^{-\beta(E(r)-E(s)-2\gamma_\beta)}.$$

and

$$\frac{\pi(r)}{\pi(s)} \;=\; \frac{p(s,r)}{p(r,s)} \;\geq\; \frac{e^{-\beta((E(r)-E(s))^++\gamma_\beta)}}{e^{-\beta((E(s)-E(r))^+-\gamma_\beta)}} \;=\; e^{-\beta(E(r)-E(s)+2\gamma_\beta)}.$$

Now let r and s be arbitrary. By the irreducibility, there is a path $r = r_0, r_1, ..., r_n = s$ from r to s of neighboring states with $\pi(r_i)/\pi(r_{i+1}) \in [e^{-\beta(E(r_i)-E(r_{i+1})+2\gamma_\beta)}, e^{-\beta(E(r_i)-E(r_{i+1})-2\gamma_\beta)}]$, $0 \leq i \leq n-1$. Therefore,

$$\frac{\pi(r)}{\pi(s)} \;=\; \prod_{i=0}^{n-1} \frac{\pi(r_i)}{\pi(r_{i+1})} \quad \begin{cases} \leq e^{-\beta(E(r)-E(s)-2|\mathcal{S}|\gamma_\beta)} \\ \geq e^{-\beta(E(r)-E(s)+2|\mathcal{S}|\gamma_\beta)}. \end{cases}$$

□

Lemma 1.1.2. *Let $E_{min} = \min_s E(s)$ be the minimal energy on $\mathcal{S}$. For any state $s \in \mathcal{S}$ we have*

$$\frac{1}{|\mathcal{S}|} e^{-\beta(E(s)-E_{min}+2|\mathcal{S}|\gamma_\beta)} \;\leq\; \pi(s) \;\leq\; e^{-\beta(E(s)-E_{min}-2|\mathcal{S}|\gamma_\beta)}.$$

In particular, $\pi(s) \to 0$ as $\beta \to \infty$ if $E(s) > E_{min}$.

Proof: Using the previous lemma, we obtain

$$\sum_r \frac{\pi(r)}{\pi(s)} \;\geq\; \sum_r e^{-\beta(E(r)-E(s)+2|\mathcal{S}|\gamma_\beta)} \;\geq\; e^{\beta(E(s)-E_{min}-2|\mathcal{S}|\gamma_\beta)}$$

and

$$\sum_r \frac{\pi(r)}{\pi(s)} \;\leq\; \sum_r e^{-\beta(E(r)-E(s)-2|\mathcal{S}|\gamma_\beta)} \;\leq\; |\mathcal{S}| e^{\beta(E(s)-E_{min}+2|\mathcal{S}|\gamma_\beta)}.$$

Therefore,

$$\pi(s) \;=\; \frac{\pi(s)}{\sum_r \pi(r)} \;=\; \left(\sum_r \frac{\pi(r)}{\pi(s)}\right)^{-1} \quad \begin{cases} \leq\; e^{-\beta(E(s)-E_{min}-2|\mathcal{S}|\gamma_\beta)} \\ \geq\; \frac{1}{|\mathcal{S}|} e^{-\beta(E(s)-E_{min}+2|\mathcal{S}|\gamma_\beta)} \end{cases}.$$

□

1.2. Valleys: the Definition

Now we start to develop the notion of metastable states and valleys. In order to do so, we must first study minimal paths between two states and maximal energies along such paths.

1.2.1. Metastable States

Definition 1.2.1. (a) For any two distinct states $r, s \in \mathcal{S}$, let

$$\Gamma(r,s) \;:=\; \{(x_0,...,x_k)|\, k \in \mathbb{N},\, x_0 = r,\, x_k = s,\, x_i \neq x_j \text{ for } i \neq j,\, p(x_i, x_{i+1}) > 0 \text{ for } 0 \leq i \leq k-1\}$$

be the set of all finite self-avoiding paths from r to s having positive probability. For any such path $\gamma = (\gamma_0,...,\gamma_k) \in \Gamma(r,s)$, let $|\gamma| := k$ be its length. We further write $t \in \gamma$ if $t \in \{\gamma_1,...,\gamma_k\}$.

(b) A self-avoiding path $\gamma = (\gamma_1,...,\gamma_k)$ from r to s is called *minimal* if its maximal energy $\max_{1\leq i\leq k} E(\gamma_i)$ is minimal among all $\gamma' \in \Gamma(r,s)$. The set of these paths is denoted $\Gamma^*(r,s)$.

(c) The set of *essential saddles* $z^*(r,s)$ between r and s is defined as

$$z^*(r,s) \;:=\; \left\{ \operatorname*{argmax}_{t\in\gamma} E(t) \Big| \gamma \in \Gamma^*(r,s) \right\} \;\subset\; \mathcal{S}.$$

As for (c), it is to be noted that the essential saddle is not unique. Indeed, there may be several minimal paths with different maxima. But nevertheless, the energy of all essential saddles is the same, denoted by $E(z^*(r,s))$ hereafter.

Definition 1.2.2. Let $M^{(1)}$ be the set of local minima of E and $\mathfrak{n} := |M^{(1)}|$. For $2 \leq i \leq \mathfrak{n}$ let successively $\mathcal{M}^{(i-1)}$ be the set of states $m \in M^{(i-1)}$ with

$$\min_{s\in M^{(i-1)}\setminus\{m\}} E(z^*(m,s)) - E(m) \;=\; \min_{m'\in M^{(i-1)}} \min_{s\in M^{(i-1)}\setminus\{m'\}} E(z^*(m',s)) - E(m')$$

and $m^{(i-1)} := \operatorname{argmax}_{m\in\mathcal{M}^{(i-1)}} E(m)$. Then $M^{(i)} := M^{(i-1)}\setminus\{m^{(i-1)}\}$. The elements of $M^{(i)}$, $1 \leq i \leq \mathfrak{n}$, are called *metastable at level* i.

In words, given the minima of E, we denote the subset of minima with minimal *relative* barrier by $\mathcal{M}^{(1)}$, and the state with highest energy in this subset by $m^{(1)}$. We understand $m^{(1)}$ as the most unstable metastable state at level 1 so that the next-level metastable set $M^{(2)}$ no longer comprises $m^{(1)}$. We proceed inductively with identifying the most unstable state in $M^{(2)}$, remove it to obtain $M^{(3)}$ and so forth. Due to the non-degeneracy, the states $m^{(i)}$, $1 \leq i \leq \mathfrak{n}-1$, are well defined. With $\{m^{(\mathfrak{n})}\} := M^{(\mathfrak{n})}$ we yield for $1 \leq i \leq \mathfrak{n}$

$$M^{(i)} \;=\; \{m^{(i)},\ldots,m^{(\mathfrak{n})}\}.$$

Thus, the above definition gives a successive filtration of the state space into $\mathcal{S} =: M^{(0)} \supset M^{(1)} \supset ... \supset M^{(\mathfrak{n})} = \{m^{(\mathfrak{n})}\}$, where, starting with all local minima, in each step one local minimum is deleted. We want to emphasize that not the (absolute) minimal *energy* but the (relative) minimal *activation energy* or barrier (for a transition to another minimum of the same level) is the main criterion for being no longer metastable.

In [55] and [56], Scoppola introduced a similar filtration of the state space which gave rise to our definition in fact. From the sequence of sets defined above, take the subsequence given by those indices where all metastable states with the same relative barrier are deleted. This is as well that subsequence of the decreasing set-sequence in [55] which makes each inclusion proper. Calling the elements of $M^{(i)}$, $1 \leq i \leq \mathfrak{n}$, metastable arises for example from the fact that there exists a constant C such that

$$\mathbb{P}_m(X_n \notin M^{(i)}) \;\leq\; e^{-C\beta}$$

for all $m \in M^{(i-1)}, 1 \leq i \leq \mathfrak{n}$, and sufficiently large n [55, Theorem 2.2]. That is, the process is to be found most likely in those metastable states. Furthermore, as we will see in Section 2.2, the mean exit times of certain domains around those metastable states are very large with very many returns to the metastable state in between (metastability in the pathwise approach). A third reason for this appellation is that for a metastable state m of appropriate level the probability to reach any other metastable state without returning to m tends faster to zero than the corresponding probability for any non-metastable state (potential theoretic approach), see Corollary 2.1.4.

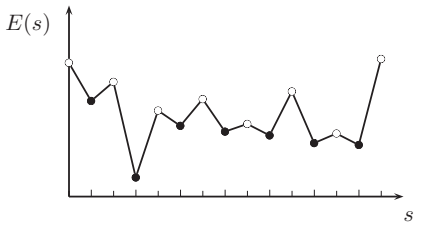

Figure 1.1.: Example of an energy landscape with minima shown as black dots (•)

Example 1.2.3. For the simple energy function depicted in Figure 1.1, a successive application of the algorithm from 1.2.2 as illustrated in Figure 1.2 leads to the following decomposition into subsets of metastable states:

$$M^{(1)} = \{2,4,6,8,10,12,14\}, \quad M^{(2)} = \{2,4,6,10,12,14\}, \quad M^{(3)} = \{2,4,6,10,14\},$$
$$M^{(4)} = \{2,4,10,14\}, \quad M^{(5)} = \{4,10,14\}, \quad M^{(6)} = \{4,14\},$$
$$M^{(7)} = \{4\}.$$

1.2.2. Valleys

Based on the filtration of $\mathcal{S}$ just described, we proceed to the definition of a sequence of *metastable sets* associated with the metastable states. Those sets will induce the MB.

Definition 1.2.4. For each $m \in M^{(i)}$, $1 \leq i \leq \mathfrak{n}$, let

$$V_<^{(i)}(m) \;:=\; \left\{ s \in \mathcal{S} \middle| E(z^*(s,m)) < E(z^*(s,m')) \text{ for all } m' \in M^{(i)} \backslash \{m\} \right\}.$$

We say that *state s is attracted by m at level i*, expressed as $s \rightsquigarrow m$ at level i, if

$$E(z^*(s,m)) \;=\; \min_{n \in M^{(i)}} E(z^*(s,n))$$

and every minimal path from s to a state $m' \in M^{(i)} \backslash \{m\}$ with $E(z^*(s,m')) = E(z^*(s,m))$ hits $V_<^{(i)}(m)$ at some time. Finally, let

$$l(i) \;:=\; \inf \left\{ i < j \leq \mathfrak{n} | m^{(i)} \rightsquigarrow m \text{ at level } j \text{ for some } m \in M^{(j)} \right\}$$

denote the minimal level at which the metastable state $m^{(i)}$ becomes attracted by a metastable state of superior level.

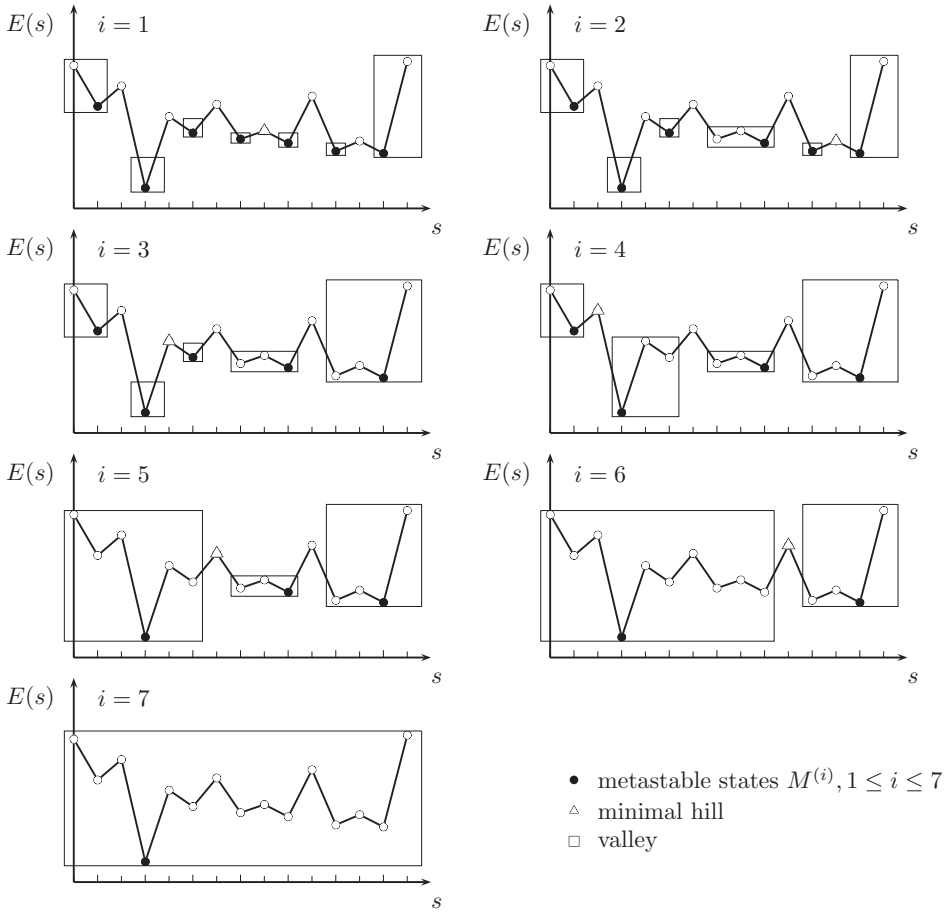

Figure 1.2.: Successive application of the algorithms in 1.2.2 and 1.2.5 to the energy landscape in Figure 1.1. For each step i, the metastable states and corresponding valleys are shown.

Definition 1.2.5. (a) Initialization: For each $m \in M^{(1)}$, define

$$V^{(1)}(m) \;:=\; \left\{ s \in \mathcal{S} \;\middle|\; s \rightsquigarrow m \text{ at level } 1 \right\}.$$

as the *valley of order* 1 *containing* m and let

$$N^{(1)} \;:=\; \left(\bigcup_{j=1}^{\mathfrak{n}} V^{(1)}(m^{(j)}) \right)^c$$

be the set of non-assigned states at level 1.

(b) Recursion: For $2 \le i \le \mathfrak{n}$ and $m \in M^{(i)}$, define

$$V^{(i)}(m) \;:=\; V^{(i-1)}(m) \cup \left\{ s \in N^{(i-1)} \;\middle|\; s \rightsquigarrow m \text{ at level } i \right\} \cup \bigcup_{j : l(j)=i, m^{(j)} \rightsquigarrow m \text{ at level } i} V^{(j)}(m^{(j)})$$

as the *valley of order* i *containing* m and let

$$N^{(i)} \;:=\; \left(\bigcup_{j=1}^{\mathfrak{n}} V^{(i \wedge j)}(m^{(j)}) \right)^c$$

be the set of non-assigned states at level i.

Here is a more intuitive description of what the previous two definitions render in a formal way: First, we define, for each level i and $m \in M^{(i)}$, the set $V_<^{(i)}(m)$ of those states s that are strongly attracted by m in the sense that $E(z^*(s,m))$ is strictly smaller than $E(z^*(s,m'))$ for any other $m' \in M^{(i)}$. Then, starting at level one, each valley $V^{(1)}(m)$, $m \in M^{(1)}$, is formed from $V_<^{(1)}(m)$ by adjoining all further states s attracted by m at this level. This leaves us with a set of non-assigned states, denoted $N^{(1)}$. In the next step (level 2), any $V^{(2)}(m)$ for $m \in M^{(2)}$ is obtained by adjoining to $V^{(1)}(m)$ all those $s \in N^{(1)}$ which are attracted by m at level 2. Moreover, if $m^{(1)}$ is attracted by m at level 2, then $V^{(1)}(m^{(1)})$ is merged into $V^{(2)}(m)$ as well. If no such m exists (thus $l(1) > 2$), it remains untouched until reaching level $l(1)$ where its bottom state $m^{(1)}$ becomes attracted by some $m' \in M^{(l(1))}$ causing its valley to be merged into $V^{(l(1))}(m')$. We will see that this ensures $V_<^{(2)}(m) \subset V^{(2)}(m)$. This procedure continues in the now obvious recursive manner until at level $\mathfrak{n}$ all states have been merged into one valley. Obviously, valleys of the same order are pairwise disjoint. Also, valleys once formed at some level can only be merged as a whole and will thus never be ripped apart during the recursive construction. For the energy function depicted in Figure 1.1, the successively derived valleys of order $i = 1,...,7$ are shown in Figure 1.2.

The reader should note that the valleys are defined in a completely deterministic way without any reference to dynamics, although they shall map a dynamical and random object. Hence, with these objects we can characterize the deterministic background of the random dynamical behavior.

1.3. Valleys: the Structure

1.3.1. Elements of a Valley

Before proceeding to results on the general shape of valleys, we collect some basic, mostly technical properties of essential saddles and first results on the structure of valleys. Both will be useful thereafter.

Proposition 1.3.1. *For any $r, s, u \in \mathcal{S}$, $1 \le i \le \mathfrak{n}$, $m_1, m_2 \in M^{(i)}, m_1 \neq m_2$, and $x_1, x_2 \in \mathcal{S}$ with $x_1 \in V_<^{(i)}(m_1)$ and $x_2 \in V^{(i)}(m_2)$, we have*

(a) $z^*(r,s) = z^*(s,r)$.

(b) $E(z^*(r,s)) \le E(z^*(r,u)) \vee E(z^*(u,s))$.

(c) $E(z^*(x_2,m_2)) \le E(z^*(x_2,m'))$ *for all* $m' \in M^{(i)}$.

(d) $E(z^*(x_1,m_2)) = E(z^*(m_1,m_2))$.

(e) $E(z^*(x_1,x_2)) \ge E(z^*(m_1,m_2))$.

(f) $x_1 \notin z^*(x_1,x_2)$.

Proof: Parts (a) and (b) are obvious.

For (c) we use an induction over i and note that there is nothing to show when $i = 1$. For general i, we must only verify that $E(z^*(x_2,m_2)) \le E(z^*(x_2,m'))$ for all $m' \in M^{(i)}$ if $x_2 \in V^{(j)}(m^{(j)})$ for some $j < i$ such that $l(j) = i$ and $m^{(j)} \rightsquigarrow m_2$ at level i (due to the recursive definition of $V^{(i)}(m_2)$). But the latter ensures that $E(z^*(x_2,m^{(j)})) \le E(z^*(x_2,n))$ for all $n \in M^{(j)} \supset M^{(i)}$ (inductive hypothesis) as well as $E(z^*(m^{(j)},m_2)) \le E(z^*(m^{(j)},m'))$ for all $m' \in M^{(i)}$. Consequently, for any such m',

$$\begin{aligned} E(z^*(x_2,m_2)) &\le E(z^*(x_2,m^{(j)})) \vee E(z^*(m^{(j)},m_2)) \\ &\le E(z^*(x_2,m^{(j)})) \vee E(z^*(m^{(j)},m')) \\ &\le E(z^*(x_2,m^{(j)})) \vee E(z^*(x_2,m'))) \\ &= E(z^*(x_2,m')). \end{aligned}$$

For assertion (d), note that $E(z^*(x_1,m_2)) > E(z^*(x_1,m_1))$, which in combination with (a) and (b) implies

$$E(z^*(m_1,m_2)) \;\le\; E(z^*(m_1,x_1)) \vee E(z^*(x_1,m_2)) \;=\; E(z^*(x_1,m_2))$$

and then further

$$E(z^*(x_1,m_2)) \;\le\; \underbrace{E(z^*(x_1,m_1))}_{<E(z^*(x_1,m_2))} \vee \underbrace{E(z^*(m_1,m_2))}_{\le E(z^*(x_1,m_2))} \;\le\; E(z^*(x_1,m_2)).$$

So the above must be an identity, i.e. $E(z^*(x_1,m_2)) = E(z^*(m_1,m_2))$.

Turning to part (e), we first infer with the help of (c) and (d) that

$$\begin{aligned} E(z^*(x_1,m_1)) \;&<\; E(z^*(x_1,m_2)) \;=\; E(z^*(m_1,m_2)) \\ &\le\; E(z^*(m_1,x_2)) \vee E(z^*(x_2,m_2)) \;=\; E(z^*(x_2,m_1)) \\ &\le\; E(z^*(x_2,x_1)) \vee E(z^*(x_1,m_1)), \end{aligned} \tag{1.1}$$

thus

$$E(z^*(x_1,m_1)) \;<\; E(z^*(x_1,x_2)). \tag{1.2}$$

Together with the just shown inequality $E(z^*(m_1,m_2)) \le E(z^*(x_2,m_1))$ (see (1.1)) and another use of (c), this yields

$$E(z^*(m_1,m_2)) \;\le\; E(z^*(x_2,m_1)) \;\le\; E(z^*(x_2,x_1)) \vee E(z^*(x_1,m_1)) \;=\; E(z^*(x_1,x_2)).$$

Finally, we infer with the help of (1.2) that

$$E(z^*(x_1,x_2)) \;>\; E(z^*(x_1,m_1)) \;\ge\; E(x_1)$$

and thus $x_1 \notin z^*(x_1,x_2)$, as claimed in (f). □

Remark 1.3.2. It is useful to point out the following consequence of the previous proposition. If, for an arbitrary state s and any two distinct metastable states $m,n \in M^{(i)}$, there exists a minimal path γ from s to n that hits a state r with $E(z^*(r,m)) < E(z^*(r,n))$, then there is also a minimal path from s to n that passes through m. Namely, if we replace the segment from r to n of the former path by the concatenation of two minimal paths from r to m and from m to n, then the maximal energy of this new path is

$$\begin{aligned} E(z^*(s,n)) \vee E(z^*(r,m)) \vee E(z^*(m,n)) \;&\le\; E(z^*(s,n)) \vee E(z^*(r,m)) \vee E(z^*(r,n)) \\ &=\; E(z^*(s,n)) \vee E(z^*(r,n)) \\ &=\; E(z^*(s,n)), \end{aligned}$$

by Proposition 1.3.1 (b), whence the new path has to be minimal from s to n as well. This yields two facts:

(a) A minimal path from s to n, where $s \rightsquigarrow n$ at level i, hits $V_<^{(i)}(n)$ before it hits any r with $E(z^*(r,m)) < E(z^*(r,n))$ for some $m \in M^{(i)}$. Otherwise, since the subpath from r to m can be chosen to stay in $\{t | E(z^*(t,m)) < E(z^*(t,n))\}$ and thus $E(z^*(s,m)) = E(z^*(s,n))$, there would be a path from s to m not hitting $V_<^{(i)}(n)$.

(b) If $s \rightsquigarrow n$ at level i and $m \in M^{(i)} \backslash \{n\}$ with $E(z^*(s,n)) = E(z^*(s,m))$, then a minimal path from s to m does not only hit $V_<^{(i)}(n)$ at some time, but in fact earlier than any other valley $V_<^{(i)}(m'), m' \in M^{(i)} \backslash \{n\}$.

Lemma 1.3.3. *Let $1 \le i < j \le \mathfrak{n}$, $m = m^{(i)}$ and $E(z^*(s,m)) \le E(z^*(s,m'))$ for every $m' \in M^{(i)}\backslash\{m\}$. Then $s \in V_<^{(j)}(m')$ for some $m' \in M^{(j)}$ implies $l(i) \le j$, $m \in V_<^{(j)}(m')$ and thus $V^{(i)}(m) \subset V^{(j)}(m')$.*

In other words, whenever $V^{(i)}(m^{(i)})$ contains an element s which at some higher level j belongs to some $V_<^{(j)}(m')$, $m' \in M^{(j)}$, the same must hold true for $m^{(i)}$ itself implying $V^{(i)}(m^{(i)}) \subset V^{(j)}(m')$. Conversely, this guarantees that $V^{(i)}(m^{(i)})$ will have no common elements with any $V_<^{(j)}(m')$ at levels $j < l(i)$ where it has not yet been merged into a valley of higher order. Furthermore, we conclude that $V_<^{(i)}(m) \subset V^{(i)}(m)$ for every $m \in M^{(i)}$, $1 \le i \le \mathfrak{n}$: By induction it suffices to show $x \in V^{(i)}(m)$ for $x \in V^{(i-1)}(m') \cap V_<^{(i)}(m)$, $m' \in M^{(i-1)}$, in which case by Proposition 1.3.1 (c) $m' \notin M^{(i)}$ and thus by the above Lemma $x \in V^{(i-1)}(m') \subset V^{(i)}(m)$.

Proof: Let us first note that, under the given assumptions,

$$E(z^*(s,m)) \;\le\; E(z^*(s,m')) \;<\; E(z^*(s,n))$$

for all $n \in M^{(j)}\backslash\{m'\}$, whence

$$E(z^*(m,n)) \;\le\; E(z^*(s,m)) \vee E(z^*(s,n)) \;=\; E(z^*(s,n)) \;\le\; E(z^*(s,m)) \vee E(z^*(m,n))$$

entails $E(z^*(m,n)) = E(z^*(s,n))$ for all such n. Using this fact, we find that

$$E(z^*(m,m')) \;\le\; E(z^*(s,m)) \vee E(z^*(s,m')) \;<\; E(z^*(s,n)) \;=\; E(z^*(m,n))$$

for all $n \in M^{(j)}\backslash\{m'\}$, which implies $m \rightsquigarrow m'$ at level j and thus $l(i) \le j$ as well as the other assertions. □

Lemma 1.3.4. *Given $1 \le i \le \mathfrak{n}$, $m \in M^{(i)}$ and $s \rightsquigarrow m$ at level i, let $\gamma = (\gamma_1,...,\gamma_k) \in \Gamma^*(s,m)$ be a path such that $E(z^*(\gamma_i,m)) \le E(z^*(\gamma_i,n))$ for all $n \in M^{(i)}\backslash\{m\}$, and which stays in $V_<^{(i)}(m)$ once hitting this set (such a γ exists by Remark 1.3.2 (a)). Then $\gamma_j \rightsquigarrow m$ at level i for each $j = 1,...,k$.*

Proof: There is nothing to prove for $\gamma_1 = s$ and any $\gamma_j \in V_<^{(i)}(m)$. So let r be any other state visited by γ, pick an arbitrary $n \in M^{(i)}\backslash\{m\}$ with $E(z^*(r,n)) = E(z^*(r,m))$ and then any minimal path τ from r to n. Let σ be the subpath of γ from s to r. We must show that τ hits $V_<^{(i)}(m)$. First, we point out that the maximal energy $E(z^*(s,r)) \vee E(z^*(r,n))$ of $\sigma\tau$, the concatenation of σ and τ, satisfies

$$\begin{aligned} E(z^*(s,n)) &\le E(z^*(s,r)) \vee E(z^*(r,n)) \\ &\le E(z^*(s,m)) \vee E(z^*(r,m)) \\ &= E(z^*(s,m)) \\ &\le E(z^*(s,n)), \end{aligned}$$

implying $\sigma\tau \in \Gamma^*(s,n)$ and, furthermore, $E(z^*(s,n)) = E(z^*(s,m))$. Thus $\sigma\tau$ must hit $V_<^{(i)}(m)$. But since σ does not hit $V_<^{(i)}(m)$ by assumption, we conclude that τ must hit $V_<^{(i)}(m)$. Since $\tau \in \Gamma^*(r,n)$ was arbitrary, we infer $r \rightsquigarrow m$ at level i. □

With the above preliminaries, we can give the following beautiful characterization of the states in a valley:

Lemma 1.3.5. *For each $1 \le i \le \mathfrak{n}$ and $m \in M^{(i)}$, we have that*

$$\left\{s \in \mathcal{S} \middle| s \rightsquigarrow m \text{ at level } i\right\} \;\subset\; V^{(i)}(m) \;\subset\; \left\{s \in \mathcal{S} \middle| E(z^*(s,m)) \le E(z^*(s,m')) \text{ for all } m' \in M^{(i)}\right\}.$$

Proof: For the second inclusion it suffices to refer to Proposition 1.3.1 (c). The first inclusion being obviously true for $s \in N^{(i-1)}$, we turn directly to the case when

$$s \rightsquigarrow n_1 \text{ at level } l_1, \quad n_1 \rightsquigarrow n_2 \text{ at level } l_2, \quad \ldots \quad n_{k-1} \rightsquigarrow n_k \text{ at level } l_k$$

with $k \geq 1$ and $1 \leq l_1 \leq \ldots \leq l_k \leq i-1$. Here, n_1 denotes the first minimum to which s is attracted (thus $s \in V^{(l_1)}(n_1)$), while n_k is the last minimum of this kind in the sequence. We may assume without loss of generality that $n_j \neq m$ for all j, for otherwise the assertion is clear.

Now we show that $n_1 \rightsquigarrow m$ at level i, which in turn implies $n_j \rightsquigarrow m$ at level i for all $1 \leq j \leq k$. As a consequence, $n_k \notin M^{(i)}$, $n_k \in V^{(i)}(m)$ and thus $s \in V^{(i)}(m)$. If $E(z^*(n_1,m)) < E(z^*(n_1,m'))$ for all $m' \in M^{(i)} \backslash \{m\}$, the assertion is proved. Hence suppose $E(z^*(n_1,m)) \geq E(z^*(n_1,m'))$ for some $m' \in M^{(i)} \backslash \{m\}$. Then

$$\begin{aligned} E(z^*(s,m')) &\leq E(z^*(s,n_1)) \vee E(z^*(n_1,m')) \leq E(z^*(s,n_1)) \vee E(z^*(n_1,m)) \\ &\leq E(z^*(s,n_1)) \vee E(z^*(s,m)) = E(z^*(s,m)) \\ &\leq E(z^*(s,m')) \end{aligned}$$

implies $E(z^*(s,m)) = E(z^*(s,m'))$ and also that the concatenation of any minimal path γ from s to n_1 and any minimal path τ from n_1 to m' (with maximal energy $E(z^*(s,n_1)) \vee E(z^*(n_1,m'))$) constitutes a minimal path from s to m' and must therefore hit $V_<^{(i)}(m)$. Note that by Lemma 1.3.4 we can choose γ to stay in the set of states attracted by n_1 at level l_1. Now, if τ hits $V_<^{(i)}(m)$, then $E(z^*(n_1,m)) = E(z^*(n_1,m'))$ and we are done. Otherwise, γ hits $V_<^{(i)}(m)$ implying the existence of some $r \in V_<^{(i)}(m)$ with $E(z^*(r,n_1)) \leq E(z^*(r,n'))$ for every $n' \in M^{(l_1)} \backslash \{n_1\}$. Now use Lemma 1.3.3 to conclude $n_1 \in V_<^{(i)}(m)$ and therefore $n_1 \rightsquigarrow m$ at level i. This completes the argument for the first inclusion. □

1.3.2. Connectivity and Nesting

The announced results on the shape of the valleys and their nested structure are provided by the next propositions.

Proposition 1.3.6. *For every $m \in M^{(i)}$ and $1 \leq i \leq \mathfrak{n}$, $V_<^{(i)}(m)$ is connected.*

Proof: Pick any $s \in V_<^{(i)}(m)$, any minimal path from s to m and finally any intermediate state r along this path for which $r \in V_<^{(i)}(m)$ must be verified. For every $m' \in M^{(i)} \backslash \{m\}$, we find

$$\begin{aligned} E(z^*(r,m)) &\leq E(z^*(r,s)) \vee E(z^*(s,m)) = E(z^*(s,m)) \\ &< E(z^*(s,m')) \\ &\leq \underbrace{E(z^*(s,r))}_{<E(z^*(s,m'))} \vee E(z^*(r,m')) \\ &= E(z^*(r,m')), \end{aligned}$$

which shows $r \in V_<^{(i)}(m)$ as required. □

Note that we have even shown that a minimal path from a state in $V_<^{(i)}(m)$ to m will never leave this set.

Proposition 1.3.7. *For every $m \in M^{(i)}$ and $1 \leq i \leq \mathfrak{n}$, $V^{(i)}(m)$ is connected.*

Proof: We use an inductive argument. If $i = 1$, the assertion follows directly from the definition of the level-one valleys because any $s \in V^{(1)}(m)$, $m \in M^{(1)}$, may be connected to m by a minimal path that eventually enters $V_<^{(1)}(m)$ without hitting any other $V_<^{(1)}(n)$ and is therefore completely contained in $V^{(1)}(m)$ by the previous lemma.

Turning to the inductive step, suppose the assertion holds true up to level $i-1$. Fix any $m \in M^{(i)}$ and notice that, by the inductive hypothesis, $V^{(i-1)}(m)$ as well as all $V^{(j)}(m^{(j)})$ with $l(j) = i$ and $m^{(j)} \rightsquigarrow m$ at level i are connected. Now, since these $m^{(j)}$ as well as all $s \in N^{(i-1)}$ attracted by m at level i may be connected to m by minimal paths as assumed in Lemma 1.3.4, we conclude that $V^{(i)}(m)$ is also connected. □

Proposition 1.3.8. *The following inclusions hold true:*

(a) $V^{(1)}(m) \subseteq ... \subseteq V^{(i)}(m)$ *for each* $m \in M^{(i)}$, $1 \leq i \leq \mathfrak{n}$.

(b) $V^{(i)}(m) \subseteq V^{(j)}(n)$ *for each* $1 \leq i < j \leq \mathfrak{n}$, $n \in M^{(j)}$ *and* $m \in M^{(i)} \cap V^{(j)}(n)$.

Proof: Since there is nothing to show for (a) we move directly to (b). But if $m \in M^{(i)} \cap V^{(j)}(n)$, then the definition of valleys ensures the existence of $1 \leq k \leq j - i$ and of $n_1,...,n_{k-1} \in M^{(j)} \backslash M^{(i)}$ such that $n_{p-1} \rightsquigarrow n_p$ at level l_p for each $p = 1,...,k$ and levels $i < l_1 < ... < l_k = j$, where $n_0 := m$ and $n_k := n$. As a consequence,

$$V^{(i)}(m) \subseteq V^{(l_1)}(n_1) \subseteq ... \subseteq V^{(l_{k-1})}(n_{k-1}) \subseteq V^{(j)}(n),$$

which proves the asserted inclusion. □

1.3.3. Leaving a Valley

To finish the analysis of the shape of the valleys, we study its boundary and exiting paths. In the next chapter we will need the following important property: A special class of minimal paths from the inside of any $V^{(i)}(m)$ to its outside must hit its interior $V_<^{(i)}(m)$.

Lemma 1.3.9. *Let* $m \in M^{(i)}, x \rightsquigarrow m$ *at level* i *and* $y \notin V^{(i)}(m)$. *Then either every minimal path from* x *to* y *hits the set* $V_<^{(i)}(m)$, *or* $E(z^*(x,y)) > E(z^*(x,m))$.

Proof: Suppose there is a minimal path γ from x to y avoiding $V_<^{(i)}(m)$. Since $y \notin V^{(i)}(m)$, it is not attracted by m at level i implying the existence of some $m' \in M^{(i)}$ with $E(z^*(y,m')) \leq E(z^*(y,m))$ and of some $\tau \in \Gamma^*(y,m')$ avoiding $V_<^{(i)}(m)$. Hence, the concatenation $\gamma\tau$ avoids $V_<^{(i)}(m)$ and must therefore have maximal energy larger than $E(z^*(x,m))$. Consequently,

$$\begin{aligned} E(z^*(x,m)) &< E(z^*(x,y)) \vee E(z^*(y,m')) \\ &\leq E(z^*(x,y)) \vee E(z^*(y,m)) \\ &\leq E(z^*(x,y)) \vee E(z^*(x,m)), \end{aligned}$$

and thus $E(z^*(x,y)) > E(z^*(x,m))$. □

Let us define the *outer boundary* $\partial^+ V$ of a valley V to be the set of those states outside of V which are adjacent to a state in V. In contrast, ∂V should denote the *inner boundary* of the valley V, that is, those states in V which are adjacent to a state in $\partial^+ V$. With the help of the previous result, we can easily show that $\partial^+ V$ contains only non-assigned states at any level where V has not yet been merged into a larger valley.

Lemma 1.3.10. *For any* $1 \leq i, j \leq \mathfrak{n}$ *and* $m = m^{(i)}$ *with* $l(i) > j$, *the outer part* $\partial^+ V$ *of the valley* $V := V^{(j \wedge i)}(m)$ *is a subset of* $N^{(j)}$ *and* $E(z^*(s,m)) = E(s)$ *for every* $s \in \partial^+ V$.

Proof: First, let $s \in \partial^+ V$ and suppose that $s \notin N^{(j)}$. Then $s \rightsquigarrow m'$ at level k, in particular $s \in V^{(k)}(m')$ for some $m' \in M^{(k)}$ and $k \leq j$. Pick any $r \in V$ with $r \sim s$ and note that $r \in \partial^+ V^{(k)}(m')$. W.l.o.g. we may assume that $r \rightsquigarrow m$ at level $j \wedge i$. Then Lemma 1.3.9 (with $x = r$ and $y = s$) ensures that either $E(z^*(r,s)) > E(z^*(r,m) \geq E(r)$, thus $z^*(r,s) = \operatorname{argmax}\{E(r), E(s)\} = s$ and $E(r) < E(s)$, or $r \in V_<(m)$ and, for some $n \in M^{(j \wedge i)}$,

$$\begin{aligned} E(z^*(r,m)) \;&<\; E(z^*(r,n)) \;\leq\; E(z^*(r,s)) \vee E(z^*(s,n)) \\ &\leq\; E(z^*(r,s)) \vee E(z^*(s,m)) \;\leq\; E(z^*(r,s)) \vee E(z^*(r,m)) \\ &=\; E(s) \vee E(z^*(r,m)), \end{aligned} \tag{1.3}$$

and thus again $E(r) < E(s)$. On the other hand, by the very same lemma (now with $x = s$ and $y = r$), we infer $E(r) > E(s)$ which is clearly impossible. Consequently, s must be non-assigned at level j as claimed.

For the second assertion take again $s \in \partial^+ V$ and a minimal path $\gamma = (s, r, ..., m)$ from s to m with $r \in V$. Again, by use of Lemma 1.3.9, we find either $E(z^*(r,m)) < E(z^*(r,s)) = E(s)$ or $r \in V_<(m)$, which leads analogously to equation (1.3) to $E(z^*(r,m)) < E(s) \vee E(z^*(r,m))$ and thus $E(z^*(r,m)) < E(s)$. In conclusion, both cases result in

$$E(z^*(s,m)) \;=\; E(s) \vee E(z^*(r,m)) \;=\; E(s),$$

finishing the proof. □

1.4. Remarks

In a nutshell, by going from $(V^{(i)}(m))_{m \in M^{(i)}}$ to $(V^{(i+1)}(m))_{m \in M^{(i+1)}}$, some valleys are merged into one (with only the smaller minimum retained as metastable state). Additionally those states from $N^{(i)}$ are added which at level i were attracted by metastable states now all belonging to the same valley. This induces the following tree-structure on the state space:

- Fix $\varnothing = s_0$.
- The first generation of the tree consists of all $m \in M^{(\mathfrak{n}-1)} \cup N^{(\mathfrak{n}-1)}$ and are thus connected to the root.
- The second generation of the tree consists of all $m \in M^{(\mathfrak{n}-2)} \cup N^{(\mathfrak{n}-2)}$, and m is connected to the node k of the first generation for which $E(z^*(m,k))$ is minimal or to itself (in the obvious sense).
- This continues until in the $\mathfrak{n}^{\text{th}}$ generation each state is listed and connected either with its unique point of attraction in the previous generation or with itself.

At each level i of such a tree, the nodes identify the metastable states that are not yet absorbed by higher order valleys as well as the saddles between them. The subtree rooted at any node $m \in M^{(i)}$ consists of the states in the valley $V^{(i)}(m)$.

Example 1.4.1. For the energy function of Example 1.2.3 and depicted in Figure 1.1, the described tree is shown in Figure 1.3. The sets of non-assigned states at the different levels are

$$\begin{aligned} &N^{(1)} = \{3,5,7,9,11,13\}, && N^{(2)} = \{3,5,7,11,13\}, && N^{(3)} = \{3,5,7,11\}, \\ &N^{(4)} = \{3,7,11\}, && N^{(5)} = \{7,11\}, && N^{(6)} = \{11\}, \\ &N^{(7)} = \emptyset. \end{aligned}$$

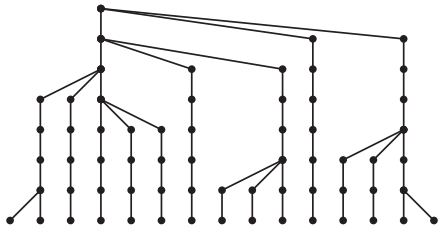

Figure 1.3.: The tree belonging to Figure 1.1

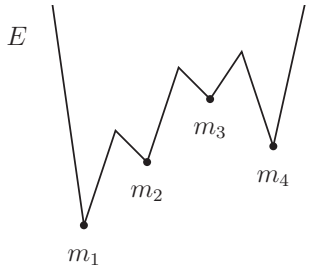

Figure 1.4.: Exemplary energy landscape

There are other authors who use similar graph-theoretical models in order to visualize high dimensional energy landscapes, for example, Okushima et al. in [47] or Bovier et al. in [12]. Both work with saddles of paths, just as we do. In contrast to our approach, in [47] every possible path, that is, every possible saddle, is represented as a node in the tree. But, as we will see, in the limit of low temperatures ($\beta \to \infty$) the essential saddle is all we need. In [12] the emphasis is on the separating saddles and absolute energies so that the order of metastability, given by the relative energies, is neglected.

The reader may wonder why valleys are defined via essential saddles and not via the at first glance more natural overall energy barriers, viz.

$$I(s,m) := \inf_{\gamma\in\Gamma(s,m)} I(\gamma_1,...,\gamma_{|\gamma|}) \quad \text{with} \quad I(s_1,...,s_n) := \sum_{i=1}^{n}(E(s_i)-E(s_{i-1}))^+$$

for a state s in a valley and the pertinent minimum m. This latter quantity, also called *cumulative activation energy*, is indeed an important parameter in [56]. The reason for our definition is, first, that the essential saddles are the critical parameters for the behavior of the aggregated chain (see Theorem 2.1.5), and second, that nice properties of the valleys fail when defining them through the cumulative activation energy instead of the energy of the essential saddle, first and foremost Lemma 1.3.5. This is visualized in Figure 1.4: If attraction and then valleys would be defined via the minimal cumulative activation energy, m_3 would be attracted by m_2 at level 2 which would be attracted by m_1 at level 3. But at level 3, m_3 would be attracted by m_4. That is, being attracted would no longer cause being in the corresponding valley and being in the valley would not cause the cumulative activation energy to be minimal.

Now there are two fundamental directions for further investigations:

Microscopic View: What happens while the process visits a fixed valley V?
In Chapter 2, we will show that during each visit of a valley V its minimum will be reached with probability tending to 1 as $\beta \to \infty$. We also calculate the expected residence time in V, establish Property 3 stated in the Introduction, and comment briefly on quasi-stationarity.

Macroscopic View: How does the process jump between the valleys?
In Chapter 3, by drawing on the results of Chapter 2, we will show that an appropriate aggregated chain is Markovian in the limit as $\beta \to \infty$ and calculate its transition probabilities. With this we will be able to establish Properties 1 and 2 listed in the Introduction and derive conditions under which Property 5 holds true.

The assembly of the results of these two chapters finally provides the envisaged definition of MB in Chapter 4.

2. Microscopic View: Fixing a Valley

Based on the provided definition of valleys of different orders, from the microscopic perception we are interested in the *system behavior when moving within a fixed valley*. We study typical trajectories of the process, determine the growth rate of the expected exit time when β tends to infinity (which is the low temperature limit in the Metropolis Algorithm) and introduce the topic of quasi-stationarity for the nesting procedure.

2.1. Low-Temperature Trajectories

The first goal in our study of the microscopic process and also the basic result for the subsequent analysis of the macroscopic process deals with the probabilities of reaching certain states earlier than others. We will conclude that in the limit $\beta \to \infty$ the process, when starting somewhere in a valley, will visit its minimum before leaving it. This will prove the metastability defined above to be conform with the potential theoretic approach to metastability defined by Bovier, Eckhoff, Gayrard & Klein in [13, Definition 1.1]. The deduction follows closely the ideas in [12], as this fundamental property of the valleys is a modification of Proposition 4.1. therein.

$\mathcal{S}$ may (and will) be viewed as a graph hereafter with edge set $\{(x,y)|x \sim y\}$. Given any subgraph Δ, we will write $\widetilde{\mathbf{P}}$ for the transition matrix of the chain restricted to Δ ($\tilde{p}(r,s) = p(r,s)$ for all distinct $r, s \in \Delta$, $\tilde{p}(r,r) = 1 - \sum_{s\in\Delta, s\neq r} \tilde{p}(r,s)$) and $\widetilde{\mathbb{P}}_x$ for probabilities when regarding this restricted chain starting at $x \in \Delta$.

2.1.1. Results on Hitting Probabilities

Theorem 2.1.1 (compare Proposition 4.1 in [12]). *Let $x, y, z \in \mathcal{S}$ be any pairwise distinct states with $x \notin z^*(x,z)$ and $E(z^*(x,z)) > E(z^*(x,y))$. Then*

$$\mathbb{P}_x(\tau_y > \tau_z) \;\leq\; |\mathcal{S}|^3 e^{-\beta(E(z^*(x,z))-E(z^*(x,y))-8|\mathcal{S}|\gamma_\beta)} \;=:\; \varepsilon(x,y,z,\beta) \overset{\beta\to\infty}{\longrightarrow} 0.$$

Thus, in the limit of low temperatures ($\beta \to \infty$), only the smallest of all possible energy barriers affects the speed of a transition. In particular, we have the following result which is preliminary to the subsequent one.

Theorem 2.1.2. *Given distinct $x, y \in \mathcal{S}$ and $m \in M^{(i)}$ such that $x \rightsquigarrow m$ at level i and $y \notin V^{(i)}(m)$, let $B := \{z \in \mathcal{S} | E(z^*(x,z)) > E(z^*(x,m))\}$. Then it holds true that*

$$\begin{aligned}
\mathbb{P}_x(\tau_m > \tau_y) \;&\leq\; \varepsilon(x,m,y,\beta)\,\mathbb{1}_B(y) \\
&\quad + \left(\sum_{z:E(z)>E(z^*(x,y))} \varepsilon(x,y,z,\beta) + \sum_{z\in V_<^{(i)}(m)} \varepsilon(z,m,y,\beta) \right) \mathbb{1}_{B^c}(y) \\
&=:\; \tilde{\varepsilon}(x,m,y,\beta) \\
&\overset{\beta\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Theorem 2.1.3. *Given $m \in M^{(i)}$, $x \in V^{(i)}(m)$ and $y \notin V^{(i)}(m)$, let $k \leq i$ be such that*

$$m_0 := x \rightsquigarrow m_1 \text{ at level } l_1, \quad m_1 \rightsquigarrow m_2 \text{ at level } l_2, \quad \ldots \quad m_{k-1} \rightsquigarrow m_k = m \text{ at level } l_k$$

for suitable $1 \leq l_1 < \ldots < l_k \leq i$, $m_j \in M^{(l_j)}$ for $j = 1,\ldots,k$ determined by the construction in Definition 1.2.5. Then

$$\mathbb{P}_x(\tau_m > \tau_y) \;\leq\; \sum_{j=1}^{k} \mathbb{P}_{m_{j-1}}(\tau_{m_j} > \tau_y) \;\leq\; \sum_{j=1}^{k} \tilde{\varepsilon}(m_{j-1}, m_j, y, \beta) \;\overset{\beta\to\infty}{\longrightarrow}\; 0.$$

Theorem 2.1.1 is the background for the definition of valleys in this work with the consequence in Theorem 2.1.3 that for β sufficiently large, with high probability the minimum of a valley is visited before this valley is left. In Section 3.2, this will be the basic ingredient to confirm Property 1 stated in the Introduction

As announced, we deduce another potential theoretic definition of metastability to hold true for our metastable states:

Corollary 2.1.4. *Let $1 \leq i \leq \mathfrak{n}$ be a level of aggregation such that*

$$\max_{m\in\mathcal{S}\setminus M^{(i)}} \left(\min_{m'\in M^{(i)}} E(z^*(m,m')) \right) - E(m) \;<\; \min_{m\in M^{(i)}} \left(\min_{m'\in M^{(i)}} E(z^*(m,m')) \right) - E(m).$$

Then

$$\frac{\max_{m\in M^{(i)}} \mathbb{P}_m(\tau_{M^{(i)}\setminus\{m\}} < \tau_m)}{\min_{m\in\mathcal{S}\setminus M^{(i)}} \mathbb{P}_m(\tau_{M^{(i)}} < \tau_m)} \;\to\; 0, \quad \beta \to \infty. \tag{2.1}$$

In [13], metastability is defined precisely via the convergence in Equation (2.1), having in mind that Bovier et al. let the number of states tend to infinity, not the temperature to zero. By the above corollary, this convergence holds true for every aggregation level for which all valleys of the same depth as the last erased minimum are erased as well since those level fulfill the above assumption. This characterization emphasizes that the probability to escape from a metastable state is much smaller than the corresponding probability for a non-metastable state, though both may tend to 0.

All three theorems and the corollary are proved at the end of the next subsection after a number of auxiliary results.

2.1.2. Auxiliary Results and Proofs

The proof of Theorem 2.1.1 will be accomplished by a combination of two propositions due to Bovier et al. [12] for a more special situation. We proceed with a reformulation of the first one in a weaker form and under weaker assumptions.

Proposition 2.1.5 (compare Theorem 1.8 in [12])**.** *Let $x, y, z \in \mathcal{S}$ be pairwise distinct such that $x \notin z^*(x,z)$. Then*

$$\begin{aligned} \mathbb{P}_x(\tau_x > \tau_z) \;&\leq\; |\mathcal{S}|^2\, e^{-\beta(E(z^*(x,z))-E(x)-3|\mathcal{S}|\gamma_\beta)}, \\ \mathbb{P}_x(\tau_x > \tau_y) \;&\geq\; |\mathcal{S}|^{-1}\, e^{-\beta(E(z^*(x,y))-E(x)+5|\mathcal{S}|\gamma_\beta)}. \end{aligned}$$

The proof requires two lemmata, the first of which may already be found in [40, Theorem 6.1] and is stated here in the notation of [12]. Its proof can be adopt to our situation without change.

Lemma 2.1.6 (see Theorem 2.1 in [12]). *Defining*

$$\mathcal{H}_z^x \;:=\; \{h : \mathcal{S} \to [0,1] \,|\, h(x) = 0, h(z) = 1\}$$

and the Dirichlet form

$$\mathcal{E}(h) \;:=\; \frac{1}{2} \sum_{r \sim s \in \mathcal{S}} \pi(r) p(r,s) (h(r) - h(s))^2,$$

we have

$$\mathbb{P}_x(\tau_x > \tau_z) \;=\; \frac{1}{\pi(x)} \inf_{h \in \mathcal{H}_z^x} \mathcal{E}(h).$$

Lemma 2.1.7 (see Lemma 2.2 and 2.5 in [12]). *For any one-dimensional subgraph $\Delta = (\omega_0,...,\omega_k)$ of $\mathcal{S}$ and corresponding transition matrix $\widetilde{\mathbf{P}}$, we have*

$$\mathbb{P}_{\omega_0}(\tau_{\omega_0} > \tau_{\omega_k}) \;\geq\; \widetilde{\mathbb{P}}_{\omega_0}(\tau_{\omega_0} > \tau_{\omega_k}) \;=\; \left(\sum_{i=1}^{k} \frac{\pi(\omega_0)}{\pi(\omega_i)} \frac{1}{p(\omega_i, \omega_{i-1})} \right)^{-1}.$$

Since the second part of this lemma is stated in [12] without proof, we include it in the Appendix (see Lemma A.1.1).

Proof of Proposition 2.1.5: In view of Lemma 2.1.6, we must find an appropriate function h for the upper bound. Let us define

$$A \;:=\; \{s \in \mathcal{S} | E(z^*(s,x)) < E(z^*(s,z))\}.$$

As $x \notin z^*(x,z)$,

$$E(z^*(x,x)) \;=\; E(x) \;<\; E(z^*(x,z)) \quad \text{and} \quad E(z^*(z,x)) \;\geq E(z) \;=\; E(z^*(z,z)),$$

and therefore $x \in A$ and $z \notin A$. Furthermore, for $r \notin A$, $s \in A$ with $p(r,s) > 0$,

$$E(z^*(s,x)) \;<\; E(z^*(s,z)) \;\leq\; E(z^*(s,x)) \vee E(z^*(x,z)) \;=\; E(z^*(x,z))$$

so that

$$\begin{aligned} E(z^*(x,z)) \;&\leq\; E(z^*(x,r)) \vee E(z^*(r,z)) \;=\; E(z^*(r,x)) \\ &\leq\; E(z^*(r,s)) \vee E(z^*(s,x)) \;=\; E(r) \vee E(s) \vee E(z^*(s,x)) \\ &=\; E(r) \vee E(s). \end{aligned}$$

Now define $h := \mathbb{1}_{A^c}$. By invoking Lemma 1.1.1, we obtain

$$\begin{aligned} \mathbb{P}_x(\tau_x > \tau_z) \;&=\; \frac{1}{\pi(x)} \inf_{h \in \mathcal{H}_z^x} \mathcal{E}(h) \\ &=\; \frac{1}{2\pi(x)} \sum_{r \neq s} \left(\mathbb{1}_{A^c}(r) - \mathbb{1}_{A^c}(s) \right)^2 \pi(r) p(r,s) \\ &=\; \sum_{r \notin A \ni s} \frac{\pi(r)}{\pi(x)} p(r,s) \\ &\leq\; \sum_{r \notin A \ni s} e^{-\beta(E(r) - E(x) + (E(s) - E(r))^+ - 3|\mathcal{S}|\gamma_\beta)} \\ &=\; \sum_{r \notin A \ni s} e^{-\beta(E(r) \vee E(s) - E(x) - 3|\mathcal{S}|\gamma_\beta)} \\ &\leq\; \sum_{r \notin A \ni s} e^{-\beta(E(z^*(x,z)) - E(x) - 3|\mathcal{S}|\gamma_\beta)} \\ &\leq\; |\mathcal{S}|^2 e^{-\beta(E(z^*(x,z)) - E(x) - 3|\mathcal{S}|\gamma_\beta)}, \end{aligned}$$

where of course the sum ranges over those $r \notin A \ni s$ with $p(r,s) > 0$.

Lemma 2.1.7 will enter in the proof of the lower bound. Consider the chain restricted to the one-dimensional subgraph given by a minimal path $\rho = (s_1,..., s_{|\rho|})$ from x to y. Let x^* be one essential saddle along this path. Then

$$\begin{aligned}
\mathbb{P}_x(\tau_x > \tau_y) &\geq \left(\sum_{i=1}^{|\rho|} \frac{\pi(x)}{\pi(s_i)} \frac{1}{p(s_i, s_{i-1})} \right)^{-1} \\
&\geq \frac{\pi(x^*)}{\pi(x)} \left(\sum_{i=1}^{|\rho|} \frac{\pi(x^*)}{\pi(s_i)} e^{\beta((E(s_{i-1})-E(s_i))^+ + \gamma_\beta)} \right)^{-1} \\
&\geq e^{-\beta(E(z^*(x,y))-E(x)+5|\mathcal{S}|\gamma_\beta)} \left(\sum_{i=1}^{|\rho|} e^{-\beta(E(z^*(x,y))-E(s_i)-(E(s_{i-1})-E(s_i))^+)} \right)^{-1} \\
&\geq e^{-\beta(E(z^*(x,y))-E(x)+5|\mathcal{S}|\gamma_\beta)} \frac{1}{|\mathcal{S}|}.
\end{aligned}$$

This completes the proof of Proposition 2.1.5. □

We proceed to the second proposition needed to prove Theorem 2.1.1.

Proposition 2.1.8 (see Corollary 1.6 in [12]). *Given $I \subset \mathcal{S}$ and distinct $x, z \in \mathcal{S} \backslash I$,*

$$\mathbb{P}_x(\tau_z < \tau_I) = \frac{\mathbb{P}_x(\tau_z < \tau_{I \cup \{x\}})}{\mathbb{P}_x(\tau_{I \cup \{z\}} < \tau_x)}.$$

Since no proof is provided in [12], we include it in the Appendix for completeness (see Proposition A.1.2). With the help of Propositions 2.1.5 and 2.1.8, the proof of Theorem 2.1.1 can now be given quite easily.

Proof of Theorem 2.1.1: By first using the previous result and then Proposition 2.1.5, we find

$$\mathbb{P}_x(\tau_z < \tau_y) = \frac{\mathbb{P}_x(\tau_z < \tau_{\{x,y\}})}{\mathbb{P}_x(\tau_{\{z,y\}} < \tau_x)} \leq \frac{\mathbb{P}_x(\tau_z < \tau_x)}{\mathbb{P}_x(\tau_y < \tau_x)} \leq |\mathcal{S}|^3 e^{-\beta(E(z^*(x,z))-E(z^*(x,y))-8|\mathcal{S}|\gamma_\beta)}.$$

The argument is completed by noting that $E(z^*(x,z)) > E(z^*(x,y))$. □

Proof of Theorem 2.1.2: If B occurs, the asserted bound follows directly from Theorem 2.1.1. Proceeding to the case when B^c occurs, i.e. $E(z^*(x,y)) \leq E(z^*(x,m))$, we first point out that

$$\begin{aligned}
\mathbb{P}_x(\tau_m > \tau_y) &= \mathbb{P}_x(\tau_m > \tau_y, E(X_n) > E(z^*(x,y)) \text{ for some } n \leq \tau_y) \\
&\quad + \mathbb{P}_x(\tau_m > \tau_y, E(X_n) \leq E(z^*(x,y)) \text{ for all } n \leq \tau_y) \\
&=: P_1 + P_2.
\end{aligned}$$

For all $z \in \mathcal{S}$ with $E(z) > E(z^*(x,y))$, we have $x \notin z^*(x,z)$ and $E(z^*(x,z)) > E(z^*(x,y))$, for

$$E(z^*(x,z)) \geq E(z) > E(z^*(x,y)) \geq E(x).$$

Therefore, by an appeal to Theorem 2.1.1,

$$\begin{aligned}
P_1 &\leq \mathbb{P}_x(\tau_y > \tau_z \text{ for some } z \text{ with } E(z) > E(z^*(x,y))) \\
&\leq \sum_{z: E(z) > E(z^*(x,y))} \mathbb{P}_x(\tau_y > \tau_z) \\
&\leq \sum_{z: E(z) > E(z^*(x,y))} \varepsilon(x,y,z,\beta).
\end{aligned}$$

To get an estimate for P_2, note that every minimal path from x to y must pass through $V_<^{(i)}(m)$ (Lemma 1.3.9). With this observation and by another appeal to Theorem 2.1.1, we infer

$$P_2 \;\leq\; \sum_{z\in V_<^{(i)}(m)} \mathbb{P}_z(\tau_m > \tau_y) \;\leq\; \sum_{z\in V_<^{(i)}(m)} \varepsilon(z,m,y,\beta),$$

having further utilized that (by Proposition 1.3.1 (f) and (e)) $z \notin z^*(z,y)$ and

$$\begin{aligned} E(z^*(z,m)) \;&<\; E(z^*(z,m')) \;\leq\; E(z^*(z,y)) \vee E(z^*(y,m')) \\ &\leq\; E(z^*(z,y)) \vee E(z^*(y,m)) \;\leq\; E(z^*(z,y)) \vee E(z^*(z,m)) \\ &=\; E(z^*(z,y)) \end{aligned}$$

for some $m' \in M^{(i)}\backslash\{m\}$ with $E(z^*(y,m)) \geq E(z^*(y,m'))$, which must exist since $y \notin V^{(i)}(m)$. □

Proof of Theorem 2.1.3: We first note that $y \notin V^{(l_j)}(m_j)$ for all $1 \leq j \leq k$. With $m_0,...,m_k$ as stated in the theorem (recall $m_0 = x$ and $m_k = m$), we obtain

$$\begin{aligned} \mathbb{P}_x(\tau_m > \tau_y) \;&=\; \mathbb{P}_x(\tau_m > \tau_y > \tau_{m_1}) + \mathbb{P}_x(\tau_{m_1} \wedge \tau_m > \tau_y) \\ &\leq\; \mathbb{P}_{m_1}(\tau_m > \tau_y) + \mathbb{P}_x(\tau_{m_1} > \tau_y) \\ &\leq\; \mathbb{P}_{m_2}(\tau_m > \tau_y) + \mathbb{P}_{m_1}(\tau_{m_2} > \tau_y) + \mathbb{P}_x(\tau_{m_1} > \tau_y) \\ &\;\;\vdots \\ &\leq\; \sum_{j=1}^{k} \mathbb{P}_{m_{j-1}}(\tau_{m_j} > \tau_y). \end{aligned}$$

Finally, use Theorem 2.1.2 to infer

$$\mathbb{P}_{m_{j-1}}(\tau_{m_j} > \tau_y) \;\leq\; \tilde{\varepsilon}(m_{j-1}, m_j, y, \beta)$$

for each $j = 1,...,k$. □

Proof of Corollary 2.1.4: For every $m \in M^{(i)}$, it holds true that

$$\begin{aligned} \mathbb{P}_m(\tau_{M^{(i)}\backslash\{m\}} < \tau_m) \;&\leq\; \sum_{m'\in M^{(i)}\backslash\{m\}} \mathbb{P}_m(\tau_{m'} < \tau_m) \\ &\leq\; |\mathcal{S}|^2 e^{-\beta(\min_{m\in M^{(i)}}(\min_{m'\in M^{(i)}} E(z^*(m,m'))-E(m))-3|\mathcal{S}|\gamma_\beta)}, \end{aligned}$$

while for every $m \in \mathcal{S}\backslash M^{(i)}$,

$$\mathbb{P}_m(\tau_{M^{(i)}} < \tau_m) \;\geq\; \mathbb{P}_m(\tau_{m'} < \tau_m) \geq |\mathcal{S}|^{-1}\, e^{-\beta(E(z^*(m,m'))-E(m)+5|\mathcal{S}|\gamma_\beta)},$$

for some $m' \in M^{(i)}$ maximizing $E(z^*(m,m'))$. Thus, under the given assumptions,

$$\begin{aligned} \frac{\max_{m\in M^{(i)}} \mathbb{P}_m(\tau_{M^{(i)}\backslash\{m\}}) < \tau_m)}{\min_{m\in\mathcal{S}\backslash M^{(i)}} \mathbb{P}_m(\tau_{M^{(i)}} < \tau_m)} \;&\leq\; \frac{|\mathcal{S}|^2 e^{-\beta(\min_{m\in M^{(i)}}(\min_{m'\in M^{(i)}} E(z^*(m,m'))-E(m))-3|\mathcal{S}|\gamma_\beta)}}{|\mathcal{S}|^{-1}\, e^{-\beta(\max_{m\in\mathcal{S}\backslash M^{(i)}}(\min_{m'\in M^{(i)}} E(z^*(m,m'))-E(m)+5|\mathcal{S}|\gamma_\beta)}} \\ &\to\; 0 \end{aligned}$$

as $\beta \to \infty$. □

2.2. Exit Time

In view of Property 3, in this section we want to analyze how long a given valley is visited, that is, we study its *exit time*. There is an extensive literature on exit problems for different kinds of stochastic processes. We mention [61, Ch. XI.2] and [24, Ch. 4, §4, Theorem 4.1] as two related to our work. The latter one studies perturbed systems on a continuous space. We can discretize their argument to get, with use of a result from [55], a nice result on the time needed to leave a valley $V^{(i)}(m)$ for any fixed $1 \leq i \leq \mathfrak{n}$ and $m \in M^{(i)}$. This result is more explicit than the one in [49, Proposition 4.6], as it only uses system parameters of the specific valley $V^{(i)}(m)$ and not of the whole energy landscape.

Definition 2.2.1. For $1 \leq i \leq \mathfrak{n}$, $N := N^{(i)}$, we define the following stopping (entrance/exit) times:

$$\begin{aligned} \xi_0^{(i)} &:= \tau_{N^c}^0 \\ \zeta_n^{(i)} &:= \inf\left\{k \geq \xi_n^{(i)} | X_k \in N\right\} \\ \xi_{n+1}^{(i)} &:= \inf\left\{k \geq \zeta_n^{(i)} | X_k \in N^c\right\},\ n \geq 0. \end{aligned}$$

The entrance times $\xi_n^{(i)}$ mark the epochs when a new valley is visited, while the exit times $\zeta_n^{(i)}$ are the epochs at which a valley is left. The reader should notice that we do not restrict ourselves to valleys of order i but include those valleys which up to order i have not yet been absorbed by some larger valley. Exit and entrance times never coincide, for there is no way to go from one valley to another one without hitting a non-assigned state - crests are always non-assigned (see Lemma 1.3.10).

In this section, we will focus on $\zeta_0^{(i)}$ for any fixed i, thus writing $\zeta_0 := \zeta_0^{(i)}$ hereafter, but later for the macroscopic process the other times will be needed as well.

For each valley $V^{(i)}(m)$, $m \in M^{(i)}$, let us define

$$s_m \;=\; s_m^{(i)} \;:=\; \left\{s \in \partial^+ V^{(i)}(m) | E(s) \leq E(s') \text{ for any } s' \in \partial^+ V^{(i)}(m)\right\}. \tag{2.2}$$

Due to Lemma 1.3.10, s_m is as well characterized by

$$s_m \;=\; \left\{s \in \partial^+ V^{(i)}(m) | E(z^*(m,s)) \leq E(z^*(m,s')) \text{ for any } s' \in \partial^+ V^{(i)}(m)\right\}.$$

Like the essential saddle between to states, the (in terms of energy) minimal state on the outer boundary of a valley is not unique - s_m is a whole set of states. But, as well similar to the essential saddle, every $s \in s_m$ has the same energy, which we denote by $E(s_m)$.

2.2.1. Mean Exit Time

Theorem 2.2.2. *Let $m \in M^{(i)}$, $1 \leq i \leq \mathfrak{n}$. Then*

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln \mathbb{E}_r \zeta_0 \;=\; E(s_m) - E(m)$$

for any $r \in V^{(i)}(m)$.

That the mean exit time tends to infinity with increasing β is not surprising, neither is the exponential growth since the escape probabilities are exponentially small. That the normalized mean (by taking logarithms and dividing by β) is independent of the starting point is up to the diminutiveness

(in comparison to the exit time) of the time to reach the minimum. In [29], our physical encouragement, Heuer refers to $E(s_m) - E(m)$, $m \in M^{(i)}$, as the *apparent activation energy* (page 22) or *barrier height* (page 7), which is nothing else than the *depth* of the valley $V(m)$. Therefore, Property 3 from the Introduction holds true and we can relate thermodynamics of the system (energies) to dynamics of the chain (holding times) in a very precise way.

For the upper bound, we need a result from [55], which in our notation is:

Proposition 2.2.3 (see Theorem 2.2 in [55])**.** *For any $1 \leq i \leq \mathfrak{n}$, β sufficiently large and $t > 2^{i-1}\exp(\beta(E(s_{m^{(i-1)}}) - E(m^{(i-1)}) + 2i|\mathcal{S}|\gamma_\beta))$,*

$$\sup_{x \in M^{(i-1)}} \mathbb{P}_x(\tau_{M^{(i)}} > t) \;\leq\; \exp(-\Delta\beta)$$

holds true with a positive constant Δ, where $M^{(0)} = \mathcal{S}$ should be recalled.

Scoppola proved it inductively, using the renormalization procedure defined in that work. We conclude the following result as a corollary.

Corollary 2.2.4. *Fix $1 \leq i \leq \mathfrak{n}$, $m \in M^{(i)}$ and $r \in V^{(i)}(m)$. Then, for any β sufficiently large and $t > 2^i \exp(\beta(E(s_m) - E(m) + 2(i+1)|\mathcal{S}|\gamma_\beta))$, it holds true that*

$$\mathbb{P}_r(\zeta_0 < (i+1)t) \;\geq\; \frac{1}{4}.$$

Proof: Let us first note that we can always arrange for m being equal to $m^{(i)}$ by sufficiently decreasing the energy function at any $m' \in M^{(i)} \backslash \{m\}$ so as to make $E(s_m) - E(m)$ minimal among all states in $M^{(i)}$. This affects neither the valley $V^{(i)}(m)$ and its outer boundary nor the distribution of ζ_0 when starting in m, for this distribution does not depend on the energy landscape outside of $V^{(i)}(m) \cup \partial^+ V^{(i)}(m)$. When applying the previous proposition, the constant Δ may have changed but is still positive which suffices for our purposes. So let $m = m^{(i)}$ hereafter.

Fix $t > 2^i \exp(\beta(E(s_m) - E(m) + 2(i+1)|\mathcal{S}|\gamma_\beta))$ and $T := it$. Since

$$E(s_m) - E(m) \;\geq\; E(s_{m^{(j)}}) - E(m^{(j)})$$

for every $1 \leq j \leq i$, we infer

$$\begin{aligned}
\mathbb{P}_r(\tau_{M^{(i)}} \leq T) \;&\geq\; \mathbb{P}_r(\tau_{M^{(i)}} \leq T, \tau_{M^{(1)}} \leq t) \\
&\geq\; \mathbb{P}_r(\tau_{M^{(1)}} \leq t) \inf_{x \in M^{(1)}} \mathbb{P}_x(\tau_{M^{(i)}} \leq (i-1)t) \\
&\geq\; \mathbb{P}_r(\tau_{M^{(1)}} \leq t) \inf_{x \in M^{(1)}} \mathbb{P}_x(\tau_{M^{(2)}} \leq t) \inf_{x \in M^{(2)}} \mathbb{P}_x(\tau_{M^{(i)}} \leq (i-2)t) \\
&\geq\; \prod_{j=1}^{i} \inf_{x \in M^{(j-1)}} \mathbb{P}_x(\tau_{M^{(j)}} \leq t) \\
&\geq\; \big(1 - \exp(-\Delta\beta)\big)^i \\
&\geq\; \frac{3}{4}
\end{aligned}$$

for β sufficiently large. Furthermore, for β so large that $\mathbb{P}_r(\tau_{M^{(i)}} < \tau_m) \leq 1/4$, we find that

$$\begin{aligned}
\mathbb{P}_r(\tau_{M^{(i)}} \leq T) \;&=\; \mathbb{P}_r(\tau_{M^{(i)}} = \tau_m \leq T) + \mathbb{P}_r(\tau_{M^{(i)}} \leq T, \tau_{M^{(i)}} < \tau_m) \\
&\leq\; \mathbb{P}_r(\tau_m \leq T) + \mathbb{P}_r(\tau_{M^{(i)}} < \tau_m) \\
&\leq\; \mathbb{P}_r(\tau_m \leq T) + \frac{1}{4}.
\end{aligned}$$

By combining both estimates, we obtain

$$\mathbb{P}_r(\tau_m \le T) \;\ge\; \mathbb{P}_r(\tau_{M^{(i)}} \le T) - \frac{1}{4} \;\ge\; \frac{1}{2}.$$

Hence, state m is hit in time T with at least probability $1/2$ when starting in r. Since $m = m^{(i)}$, we further have

$$\mathbb{P}_m(\zeta_0 \le t) \;\ge\; \mathbb{P}_m(\tau_{M^{(i+1)}} \le t) \;\ge\; 1 - \exp(-\Delta\beta) \;\ge\; \frac{1}{2}$$

for β sufficiently large. Hence, the valley is left in time t with at least probability $1/2$ when starting in m. By combining the estimates, we finally obtain

$$\begin{aligned}\mathbb{P}_r(\zeta_0 \le (i+1)t) &\ge \mathbb{P}_r(\zeta_0 \le T + t | \tau_m \le T)\, \mathbb{P}_r(\tau_m \le T) \\ &\ge \mathbb{P}_r(\tau_m \le T)\, \mathbb{P}_m(\zeta_0 \le t) \\ &\ge \frac{1}{4},\end{aligned}$$

which proves our claim. □

Proof of Theorem 2.2.2: Using the lemma just shown, we infer

$$\begin{aligned}\mathbb{E}_r(\zeta_0) &\le (i+1)t \sum_{n\ge 0} (n+1)\, \mathbb{P}_r\left(n(i+1)t \le \zeta_0 < (n+1)(i+1)t\right) \\ &= (i+1)t \sum_{n\ge 0} (n+1) \Big(\mathbb{P}_r\left(\zeta_0 \ge n(i+1)t\right) - \mathbb{P}_r\left(\zeta_0 \ge (n+1)(i+1)t\right) \Big) \\ &= (i+1)t \sum_{n\ge 0} \mathbb{P}_r\left(\zeta_0 \ge n(i+1)t\right) \\ &\le (i+1)t \sum_{n\ge 0} \left(\max_{x\in V} \mathbb{P}_x\left(\zeta_0 \ge (i+1)t\right) \right)^n \\ &\le (i+1)t \sum_{n\ge 0} \left(\frac{3}{4}\right)^n \\ &= 4(i+1)t,\end{aligned}$$

where $t := 2^i \exp(\beta(E(s_m) - E(m) + 2(i+1)|\mathcal{S}|\gamma_\beta)) + 1$. Since $\gamma_\beta \to 0$, we get in the limit

$$\limsup_{\beta\to\infty} \frac{1}{\beta} \ln \mathbb{E}_r \zeta_0 \;\le\; E(s_m) - E(m)$$

for all $r \in V^{(i)}(m)$.

Turning to the lower bound, define a sequence of stopping times, viz. $\rho_0 := 0$ and

$$\rho_n \;:=\; \inf\{k > \rho_{n-1} | X_k = m \text{ or } X_k \in \partial^+ V\}$$

for $n \ge 1$. Then $Z_n := X_{\rho_n}$, $n \ge 0$, forms a Markov chain, the transition probabilities of which can be estimated with the help of Proposition 2.1.5, namely

$$\begin{aligned}\mathbb{P}(Z_1 \in \partial^+ V^{(i)}(m) | Z_0 = m) &= \mathbb{P}_m(\rho_1 = \zeta_0) \\ &= \mathbb{P}_m(\zeta_0 < \tau_m) \\ &\le \sum_{s\in\partial^+ V} \mathbb{P}_m(\tau_s < \tau_m) \\ &\le |\mathcal{S}|^3 e^{-\beta(\min_{s\in\partial^+ V} E(z^*(m,s)) - E(m) - 3|\mathcal{S}|\gamma_\beta)} \\ &= |\mathcal{S}|^3 e^{-\beta(E(s_m) - E(m) - 3|\mathcal{S}|\gamma_\beta)}.\end{aligned}$$

Further defining $\nu := \inf\{k \geq 0 | Z_k \in \partial^+ V^{(i)}(m)\}$, this implies in combination with a geometric trials argument that

$$\mathbb{P}_m(\nu > n) \;\geq\; \Big(1 - |\mathcal{S}|^3 e^{-\beta(E(s_m)-E(m)-3|\mathcal{S}|\gamma_\beta)}\Big)^{n-1}.$$

As a consequence,

$$\mathbb{E}_m\zeta_0 \;=\; \sum_{n\geq 1} \mathbb{E}_m(\underbrace{\rho_n - \rho_{n-1}}_{\geq 1})\mathbb{1}_{\{\nu\geq n\}} \;\geq\; \sum_{n\geq 1}\mathbb{P}_m(\nu \geq n) \;\geq\; |\mathcal{S}|^{-3} e^{\beta(E(s_m)-E(m)-3|\mathcal{S}|\gamma_\beta)}.$$

For arbitrary $r \in V$, we now infer

$$\begin{aligned}
\mathbb{E}_r\zeta_0 \;&=\; \mathbb{E}_r\zeta_0\mathbb{1}_{\{\zeta_0\leq\rho_1\}} + \mathbb{E}_r\zeta_0\mathbb{1}_{\{\zeta_0>\rho_1\}} \\
&\geq\; \mathbb{E}_r\big(\mathbb{E}_r(\zeta_0\mathbb{1}_{\{\zeta_0>\rho_1\}}|X_{\rho_1} = m)\big) \\
&\geq\; \mathbb{E}_r\mathbb{1}_{\{\zeta_0>\rho_1\}}\mathbb{E}_m\zeta_0 \\
&\geq\; \mathbb{P}_r(\zeta_0 > \rho_1)|\mathcal{S}|^{-3}e^{\beta(E(s_m)-E(m)-3|\mathcal{S}|\gamma_\beta)} \\
&\geq\; \frac{1}{2}\,|\mathcal{S}|^{-3}e^{\beta(E(s_m)-E(m)-3|\mathcal{S}|\gamma_\beta)}
\end{aligned}$$

for all sufficiently large β, because $\lim_{\beta\to\infty}\mathbb{P}_r(\zeta_0 > \rho_1) = 1$ (Theorem 2.1.3). Finally, by taking logarithms and letting β tend to ∞, we arrive at the inequality

$$\liminf_{\beta\to\infty}\frac{1}{\beta}\ln\mathbb{E}_r\zeta_0 \;\geq\; E(s_m) - E(m),$$

which completes the proof. □

2.2.2. Universal Time Scale

The next result distinguishes our approach from others mentioned in the Introduction which aim at universal time scales or universal stability. As the exit times depend on the depth and those depths differ exponentially from valley to valley, in our situation it is impossible to identify a universal scale for the times spent in different valleys.

Corollary 2.2.5. *Let $T = e^{\beta t}$ for some $t > 0$.*

(a) For every $r \in V^{(i)}(m)$ with $m \in M^{(i)}$, $1 \leq i \leq \mathfrak{n}$, satisfying $E(s_m) - E(m) > t$,

$$\lim_{\beta\to\infty}\mathbb{P}_r(\zeta_0 \leq T) \;=\; 0.$$

(b) For $t > \max_{m\in M^{(i)}}\left(E(s_m) - E(m)\right)$, $1 \leq i \leq \mathfrak{n}$, and every $r \in \mathcal{S}$,

$$\lim_{\beta\to\infty}\mathbb{P}_r(T < \zeta_1) \;=\; 0.$$

Thus, if we choose $T = e^{\beta t}$, $t > 0$, as a typical time scale, that is, study $(X_{nT})_{n\geq 0}$, once a valley deeper than t is reached, it will a.a.s. not be left within K steps for every $K \in \mathbb{N}$. But if every valley is shallower than t, the process $(X_{nT})_{n\geq 0}$ overlooks the valleys.

Proof: (a) We need the bound

$$\mathbb{P}(X > a) \geq \frac{\mathbb{E}(X-a)^2}{\mathbb{E}(X^2)}$$

for every integrable random variable X and $a \leq \mathbb{E}(X)$, which is proved in [51, Equation (10)]. Thus,

$$\begin{aligned}
\mathbb{P}_r(\zeta_0 \leq T) &= 1 - \mathbb{P}_r(\zeta_0 \geq T) \\
&\leq 1 - \frac{\mathbb{E}_r(\zeta_0 - T)^2}{\mathbb{E}_r(\zeta_0^2)} \\
&= \frac{2T\mathbb{E}_r(\zeta_0) - T^2}{\mathbb{E}_r(\zeta_0^2)} \\
&\leq \left(2e^{\beta(t+E(s_m)-E(m)+o(1))} - e^{2\beta t}\right) e^{-2\beta(E(s_m)-E(m)+o(1))} \\
&= 2e^{\beta(t-E(s_m)+E(m)+o(1))} - e^{2\beta(t-E(s_m)+E(m)+o(1))} \\
&= e^{\beta(t-E(s_m)+E(m)+o(1))} \\
&\to 0
\end{aligned}$$

as $\beta \to \infty$, because $t < E(s_m) - E(m)$.

(b) In this case,

$$\mathbb{P}_s(\zeta_0 > T/2) \leq 2\frac{\mathbb{E}_s(\zeta_0)}{T} = 2e^{\beta(E(s_m)-E(m)-t+o(1))} \to 0$$

for every $s \in V^{(i)}(m)$, $m \in M^{(i)}$. Furthermore,

$$\begin{aligned}
\mathbb{P}_r(\xi_1 > T/2) &\leq \mathbb{P}_r(\zeta_0 > T/4) + \mathbb{P}_r(\xi_1 > T/2, \zeta_0 \leq T/4) \\
&\leq 4e^{\beta(E(s_m)-E(m)-t+o(1))} + \mathbb{P}_r(\xi_1 - \zeta_0 > T/4) \\
&\to 0
\end{aligned}$$

for every $r \in V^{(i)}(m)$, $m \in M^{(i)}$, since $\xi_1 - \zeta_0$ is a sum of geometrically distributed random variables with parameters $p = p(s,s)$ for some $s \in N^{(i)}$. These parameters converge to positive constants as β tends to infinity, because for every $s \in N^{(i)}$ there is at least one $r \sim s$ with $E(r) \leq E(s)$. Thus,

$$\begin{aligned}
\mathbb{P}_r(T < \zeta_1) &\leq \mathbb{P}_r(\xi_1 > T/2) + \mathbb{P}_r(\xi_1 \leq T/2, \zeta_1 > T) \\
&= \mathbb{P}_r(\xi_1 > T/2) + \sum_{s \in \mathcal{S}} \mathbb{P}_r(X_{\xi_1} = s, \xi_1 \leq T/2, T < \zeta_1) \\
&\leq \mathbb{P}_r(\xi_1 > T/2) + \sum_{s \in \mathcal{S}} \mathbb{P}_r(X_{\xi_1} = s)\mathbb{P}_s(\zeta_0 > T/2) \\
&\to 0.
\end{aligned}$$

□

2.3. Quasi-Stationarity

Naturally, several other questions concerning the behavior of the process when moving in a fixed valley are of interest, and *quasi-stationarity* may appear as one to come up with first. Quasi-stationarity is strongly linked to metastability, see for example the new results of Bianchi & Gaudillière [8]. In [52], Pollett gives a bibliography on the immense literature on the topic of quasi-stationarity.

For a given valley V (of any level), a quasi-stationary distribution $\nu = (\nu(j))_{j \in V}$ is characterized by the *quasi-invariance*, viz.

$$\mathbb{P}_\nu(X_n = j | \tau_{\mathcal{S} \setminus V} > n) = \nu(j) \quad \text{for all } j \in V,\, n \geq 0. \tag{2.3}$$

It also satisfies

$$\lim_{n\to\infty} \mathbb{P}_\mu(X_n = j | \tau_{\mathcal{S}\setminus V} > n) \;=\; \nu(j) \quad \text{for all } j \in V, \tag{2.4}$$

if μ is an arbitrary distribution with $\mu(V) = 1$, for X is irreducible. The latter property renders uniqueness of ν. Since $\mathcal{S}$ is finite, the existence of ν follows by an old result due to Darroch & Seneta [17, Section 4]. It is obtained as the normalized eigenvector of the Perron-Frobenius eigenvalue $\lambda = \lambda(V)$ of the submatrix $(p(x,y))_{x,y\in V}$. This eigenvalue λ is also the probability for the chain to stay in V at least one step when started with ν, thus $\mathbb{P}_\nu(\tau_{V^c} > 1) = \lambda$. As an immediate consequence, one finds that the exit time τ_{V^c} has a geometric distribution with parameter $1-\lambda$ under $\mathbb{P}_\nu$. In the present context, this naturally raises the question about how the parameter λ relates to the transition probabilities or the energies of the valley V. A simple probabilistic argument shows the following basic and intuitively obvious result concerning the eigenvalues associated with the nesting $V^{(1)}(m) \subset ... \subset V^{(i)}(m)$ (Proposition 1.3.8) for any $1 \le i \le \mathfrak{n}$ and $m \in M^{(i)}$.

Proposition 2.3.1. *Fixing any $1 \le i \le \mathfrak{n}$ and $m \in M^{(i)}$, let $\lambda^{(j)} := \lambda(V^{(j)}(m))$ for $j = 1,...,i$. Then $\lambda^{(1)} \le ... \le \lambda^{(i)}$.*

Proof: Write ν_j as shorthand for the quasi-stationary distribution on $V^{(j)}(m)$ and T_j for $\tau_{\mathcal{S}\setminus V^{(j)}(m)}$. Plainly, $T_j \le T_{j+1}$ and

$$\begin{aligned}(\lambda^{(j)})^n \;&=\; \mathbb{P}_{\nu_j}(T_j > n) \;\le\; \mathbb{P}_{\nu_j}(T_{j+1} > n) \\ &=\; \int_{\{T_{j+1}>k\}} \mathbb{P}_{X_k}(T_{j+1} > n-k)\; d\mathbb{P}_{\nu_j} \\ &=\; \mathbb{P}_{\nu_j}(T_{j+1} > k)\, \mathbb{P}_{\mu_k}(T_{j+1} > n-k),\end{aligned} \tag{2.5}$$

where $\mu_k(x) := \mathbb{P}_{\nu_j}(X_k = x | T_{j+1} > k)$ for $x \in V^{(j+1)}$. Since $\mathcal{S}$ is finite and by virtue of (2.4), we have that $\mu_k \le 2\nu_{j+1}$ when choosing k sufficiently large. For any such k, we find that (2.5) has upper bound

$$2\,\mathbb{P}_{\nu_j}(T_{j+1} > k)\, \mathbb{P}_{\nu_{j+1}}(T_{j+1} > n-k) \;=\; 2\,\mathbb{P}_{\nu_j}(T_{j+1} > k)\, (\lambda^{(j+1)})^{n-k}.$$

Hence, we finally conclude

$$\lambda^{(j)} \;\le\; \left(2\,\mathbb{P}_{\nu_j}(T_{j+1} > k)\, (\lambda^{(j+1)})^{-k}\right)^{1/n} \lambda^{(j+1)}$$

and thereby the assertion upon letting $n \to \infty$. □

An alternative matrix-analytic proof draws on an old result by Frobenius [25], here cited from Seneta [57, Theorem 1.1], which is a very useful monograph on Matrix theory needed for time-homogeneous and -inhomogeneous Markov chains on finite and countable state spaces.

Lemma 2.3.2 (see Theorem 1.1 in [57])**.** *If $A = (a_{ij})$ and $C = (c_{ij})$ denote two real $k \times k$-matrices such that A is nonnegative and irreducible with maximal eigenvalue λ_A^* and $|c_{ij}| \le a_{ij}$ for all $1 \le i,j \le k$, then $|\lambda| \le \lambda_A^*$ for all eigenvalues λ of C.*

Second proof of Proposition 2.3.1: For any fixed valley V, collapse all states $s \notin V$ into an absorbing state (grave) Δ. This leaves transition probabilities between states in V unchanged. A proper rearrangement of states allows us to assume that the new transition matrix has the form

$$\mathbf{P} \;=\; \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{p} & \mathbf{Q} \end{pmatrix}$$

for a $|V| \times 1$-column vector $\mathbf{p} \neq 0$ and a nonnegative, substochastic and irreducible $|V| \times |V|$-matrix $\mathbf{Q}$. Now, for any $2 \leq j \leq i$, let A be this matrix $\mathbf{Q}$ when $V = V^{(j)}(m)$, and D be this matrix when $V = V^{(j-1)}(m)$. Then, obviously,

$$A := \begin{pmatrix} A_1 & A_2 \\ A_3 & D \end{pmatrix}$$

and A is irreducible and nonnegative with maximal eigenvalue $\lambda^{(j)}$. Defining furthermore

$$C := \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix},$$

The largest eigenvalue of C equals the largest eigenvalue of D, thus $\lambda^{(j-1)}$. Finally, the desired conclusion follows from the previous lemma since $|c_{ij}| = c_{ij} \leq a_{ij}$ for all $i \leq i, j \leq k$. □

In our special case, we can say more about the eigenvalues $\lambda(V(m))$:

Proposition 2.3.3. *Fixing any $1 \leq i \leq \mathfrak{n}$ and $m \in M^{(i)}$, $V := V^{(i)}(m)$, then*

$$\lim_{\beta \to \infty} \frac{1}{\beta} \ln\left(1 - \lambda(V)\right) = -(E(s_m) - E(m)).$$

Proof: Since $\mathbb{E}_\nu(\tau_{V^c}) = \frac{1}{1-\lambda(V)}$, we have

$$\lim_{\beta \to \infty} \frac{1}{\beta} \ln\left(1 - \lambda(V)\right) = \lim_{\beta \to \infty} \frac{1}{\beta} \ln\left(\frac{1}{\mathbb{E}_\nu(\tau_{V^c})}\right) = -\lim_{\beta \to \infty} \frac{1}{\beta} \ln\left(\mathbb{E}_\nu(\tau_{V^c})\right) = -(E(s_m) - E(m)).$$

□

The above result is also known for continuous time Markov chains driven by a Brownian Motion with absorption outside of a smooth bounded domain of $\mathbb{R}^n$, $n \geq 2$, see Mathieu [44, Theorem 1]. He even determined the whole spectrum of the sub-stochastic operator. In Section 6.1 we will come back to the spectrum of $(p(x,y))_{x,y \in V(m)}$ when determining the speed of convergence against the quasi-stationary distribution.

3. Macroscopic View: Transitions between Valleys

With the help of the nested state space decomposition into valleys of different orders and around bottom states of different stability, we will now be able to define and study *macroscopic versions of the original process* $X = (X_n)_{n\geq 0}$. Those coarse grained processes keep track only of the valleys that are visited by X, neglecting the specific state in them. As already explained, we do not fix the order of stability of the considered valleys, rather we consider every valley which is at a given level not yet absorbed by another valley, though its ground state may be of lower stability. Aggregated processes like these, as a special kind of hidden Markov models, have received quite some attention in the literature and are studied for example as canonical models of systems with observability restrictions (e.g. [58] and [62]), in terms of lumpability (e.g.. [37, Chapter VI] and [38]), or in terms of transition probabilities under different conditions (e.g. [54], [35] and [59]). Here we will show an asymptotically (semi-)Markovian behavior and study the trajectories of the aggregated process in order to identify the reasons for forward-backward jumps.

3.1. Macroscopic Processes

In the subsequent definition of aggregated versions of X, we will distinguish between two variants:

A time-scale preserving aggregation that, for a fixed level and each n, keeps track of the valley the original chain visits at time n and thus only blinds its exact location within a valley.

An accelerated version that, while also keeping track of the visited valleys, further blinds the sojourn times within a valley by counting a visit just once.

Actually, the definition of these aggregations at a chosen level i is a little more complicated because their state space, denoted $\mathcal{S}^{(i)}$ below and the elements of which we call *level i metastates*, also comprises the non-assigned states at level i as well as the minima of those valleys that were formed at an earlier level and whose merger is pending at level i because their minima are not attracted at this level.

Definition 3.1.1. Fix $1 \leq i \leq \mathfrak{n}$, let $\mathcal{S}^{(i)} := \{m^{(j)} \in M^{(1)} | \, l(j) > i\} \cup N^{(i)}$ and

$$V^{(i)}(s) \;:=\; \begin{cases} V^{(i)}(m^{(j)}), & \text{if } s = m^{(j)} \text{ for some } j \geq i \\ V^{(j)}(m^{(j)}), & \text{if } s = m^{(j)} \text{ for some } j < i \\ \{s\}, & \text{if } s \in N^{(i)} \end{cases}$$

for $s \in \mathcal{S}^{(i)}$. Then define

$$\overline{Y}_n^{(i)} \;:=\; \sum_{s\in\mathcal{S}^{(i)}} s\, \mathbb{1}_{\{X_n \in V^{(i)}(s)\}} \quad \text{and} \quad Y_n^{(i)} \;:=\; \overline{Y}_{\sigma_n}^{(i)},$$

$n \geq 0$, where $\sigma_0 = \sigma_0^{(i)} :\equiv 0$ and

$$\sigma_n \;=\; \sigma_n^{(i)} \;:=\; \inf\left\{k > \sigma_{n-1} \middle| \overline{Y}_k^{(i)} \neq \overline{Y}_{k-1}^{(i)}\right\}$$

for $n \geq 1$. We call $\overline{Y}^{(i)} = (\overline{Y}_n^{(i)})_{n\geq 0}$ and $Y^{(i)} = (Y_n^{(i)})_{n\geq 0}$ the *aggregated chain (AC)* and the *accelerated aggregated chain (AAC) (at level i)* associated with $X = (X_n)_{n\geq 0}$.

Hence, starting in an arbitrary valley, the original chain stays there for a time $\sigma_1 = \zeta_0 = \zeta_0^{(i)}$ (as defined in Definition 2.2.1) before it jumps via some non-assigned states $k_1,...,k_l$, $l \geq 1$, (staying a geometric time in each of these states) to another valley at time $\sigma_{l+1} = \xi_1 = \xi_1^{(i)}$. There it stays for $\sigma_{l+2} - \sigma_{l+1} = \zeta_1 - \xi_1$ time units, before it moves on in a similar manner. By going from X to its aggregation $\overline{Y}^{(i)}$ at level i, we regard the whole valley $V^{(i)}(s)$ for $s \in \mathcal{S}^{(i)}$ as one single metastate and therefore give up information about the exact location of X within a valley. $\overline{Y}^{(i)}$ is a jump process on $\mathcal{S}^{(i)}$ with successive sojourn times $\sigma_{n+1} - \sigma_n$, $n \geq 0$, which do not only depend on the valley but also on the states of entrance and exit. The AAC is the embedded jump process, viz.

$$\overline{Y}_n^{(i)} \;=\; \sum_{j\geq 0} Y_j^{(i)} \mathbb{1}_{\{\sigma_j \leq n < \sigma_{j+1}\}},$$

giving the states only at jump epochs. Starting from the minimum of the first valley, it moves to the states $k_1,..., k_l \in N^{(i)}$, proceeds to the minimum of a second valley and so on. Its sojourn times are always 1.

Of course, at small temperatures the time spent in a non-assigned state or in a valley around a low order metastable state is very small compared to the time spent in a valley around a metastable state of higher order. Thus, such states can be seen as instantaneous but not are necessarily of little importance for the evolution of the process. We account for them for several reasons: First, in the path-dependent definition mentioned in the Introduction and used in physics, they build small MB of great transitional activity of the process and are thus relevant in view of our goal to provide a definition of MB that conforms as much as possible with the path-dependent one. Additionally, there are very many low-order metastable states, a good approximation for the energies of the local minima is to be normally distributed. Though a single "instantaneous" state has only little effect, they become importance due to their mere number. Second, a complete partitioning of the state space, that is, an assignment of every $s \in \mathcal{S}$ to a metastate $m \in \mathcal{S}^{(i)}$ via a global algorithm, fails when merely focusing on $\{V^{(i)}(m), m \in M^{(i)}\}$. There is neither an obvious nor natural way to assign non-assigned states or non-attracted minima to them. Any such assignment would cause Property 1 to fail (as we will see below). However, we see no reason to ignore low order metastates completely, as they possess the same features as higher order metastates.

As energetically lower states are always more likely than higher ones under the stationary distribution π, X is most likely to be found in a minimum and $\mathbb{P}_\pi(X_n \neq \overline{Y}_n^{(i)}) \leq e^{-\Delta(i)\beta}$, $n \geq 0$, $1 \leq i \leq \mathfrak{n}$, for an appropriate constant $\Delta(i)$. Therefore, it is not far-fetched to study the AC $\overline{Y}^{(i)}$ in order to obtain results concerning the underlying system. We dedicate this chapter to the identification of the most convenient level i. In the second part of the present work, most notably Chapters 6 and 7, we come back to the question about which information of the original process can be obtained from the AAC and AC.

3.2. (Semi-)Markov Property

In general, both aggregated chains are no longer Markovian. Transition probabilities of the AAC not only depend on the current state, i.e. the current valley, but also on the entrance state into that valley, whereas transition probabilities of the AC depend on the current sojourn times, which in turn depend on the previous, the present and the next valley. On the other hand, since valleys are defined in such a way that asymptotically almost surely (a.a.s.), i.e., with probability tending to one as $\beta \to \infty$, the minimum will be reached from anywhere inside the valley before the valley is left, the

AAC will be shown below to converge to a certain Markov chain on $\mathcal{S}^{(i)}$. Also, when passing over to the bivariate AC $(\overline{Y}_n^{(i)}, \overline{Y}_{n+1}^{(i)})_{n\geq 0}$, it converges to a semi-Markov chain.

Definition 3.2.1. Given any nonempty countable set $\mathcal{S}$, let $(M_n, T_n)_{n\geq 0}$ be a bivariate temporally homogeneous Markov chain on $\mathcal{S} \times \mathbb{N}$, with transition kernel $Q(s, \cdot)$ only depending on the first component, viz., for all $n \geq 0, s \in S$ and $t \geq 0$,

$$\mathbb{P}(M_{n+1} = s, T_{n+1} \leq t | M_n, T_n) \;=\; Q(M_n, \{s\} \times [0, t]) \tag{3.1}$$

holds. Put $S_n := \sum_{i=0}^{n} T_i$ for $n \geq 0$ and $\nu(t) := \max\{n \geq 0 | S_n \leq t\}$ ($\max \emptyset := 0$) for $t \geq 0$. Then $Z_n := M_{\nu(n)}, n \geq 0$, is called *semi-Markov chain* with *embedded Markov chain* $(M_n)_{n\geq 0}$ and *sojourn* or *holding times* T_0, T_1,....

Note that equation (3.1) holds if and only if $M = (M_n)_{n\geq 0}$ forms a temporally homogeneous Markov chain and the $(T_n)_{n\geq 0}$ are conditionally independent given M such that the distribution of T_n only depends on M_{n-1}, M_n for $n \geq 1$ (in a temporally homogeneous manner), and on M_0 for $n = 0$. Note further that we have specialized to the case where holding times take values in $\mathbb{N}$ only (instead of $(0, \infty)$). For a very vivid introduction to discrete time semi-Markov chains on countable state spaces see for example [5, Chapter 3] or [1, Chapter 9] for more general results.

3.2.1. Markov Property

Recall from (2.2) the definition of s_m for $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and notice that the second equality there entails $E(z^*(m, s)) < E(z^*(m, s'))$ for any $s \in s_m$ and any $s' \in \partial^+ V^{(i)}(m) \backslash s_m$. Further recall from our basic assumptions that for any $r, s \in \mathcal{S}$ the limit $p^*(r, s) = \lim_{\beta\to\infty} p(r, s)$ exists and is positive if $r \sim s$ and $E(r) \geq E(s)$. The following result, revealing the announced convergence for the AAC, confirms in particular that a valley $V^{(i)}(m)$, $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$, is a.a.s. to be left via s_m.

We are interested in $\mathbb{P}_r(Y_1^{(i)} = s)$, $r, s \in \mathcal{S}^{(i)}$ when β tends to infinity. The model assumptions ensure the existence of this limit if, either $r \in N^{(i)}$, or $r \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $s \notin s_r$, or $r \in \mathcal{S}^{(i)} \backslash N^{(i)}$, $s \in s_r$ and $|s_r| = 1$. We will see this in the proof of the next proposition. However, the assumptions do not ensure the existence for $r \in \mathcal{S}^{(i)} \backslash N^{(i)}$, $s \in s_r$ and $|s_r| \geq 2$. Regard as an example a Markov chain on $\mathcal{S} = \{1, 2, 3\}$ with $p(2, 1) = \exp(-\beta - (-1)^{\lfloor \beta \rfloor})$ and $p(2, 3) = \exp(-\beta)$, and observe that $\mathbb{P}_2(\tau_1 < \tau_3) = 1/(1+\exp((-1)^{\lfloor \beta \rfloor}))$ does not converge. Therefore, we have to pass on to an appropriate subsequence along which β tends to infinity. That is, we restrict to a specific *annealing schedule*.

Definition 3.2.2. We call a subsequence $(\beta_n)_{n\geq 0}$ with $\beta_n \to \infty$ as $n \to \infty$ a *proper annealing schedule*, if $\lim_{\beta_n\to\infty} \mathbb{P}_r(Y_1^{(i)} = s)$ exists for every $r, s \in \mathcal{S}^{(i)}$.

Such a proper annealing schedule exists since the probabilities are bounded and $\mathcal{S}^{(i)}$ is finite. In the following, whenever it is necessary to pass on to such a subsequence, we write $\beta_n \to \infty$ instead of $\beta \to \infty$ and assume that we have fixed some proper annealing schedule $(\beta_n)_{n\geq 0}$. The limit of course depends on this schedule.

Proposition 3.2.3. *For each $1 \leq i \leq \mathfrak{n}$ and as $\beta_n \to \infty$ along a proper annealing schedule, the level i AAC $Y^{(i)}$ converges to a Markov chain $\widehat{Y}^{(i)} = (\widehat{Y}_n^{(i)})_{n\geq 0}$ on $\mathcal{S}^{(i)}$ with transition probabilities $\widehat{p}(r, s) = \widehat{p}_i(r, s)$ stated below, that is,*

$$\lim_{\beta_n\to\infty} \mathbb{P}(Y_{n+1}^{(i)} = s | Y_n^{(i)} = r, Y_{n-1}^{(i)} = m_{n-1}, ..., Y_0^{(i)} = m_0) \;=\; \widehat{p}(r, s)$$

for all $m_0,...,m_{n-1},r,s\in\mathcal{S}^{(i)}$ *and* $n\geq 0$. *For* $r\in\mathcal{S}^{(i)}\backslash N^{(i)}$ *we have* $\widehat{p}(r,s):=\lim_{\beta_n\to\infty}\mathbb{P}_r(X_{\zeta_0}=s)$, *which is positive if and only if* $s\in s_r$, *and for* $r\in N^{(i)}$

$$\widehat{p}(r,s) \;:=\; \begin{cases} \dfrac{p^*(r,s)}{1-p^*(r,r)}, & s\in\mathcal{N}(r)\cap N^{(i)} \\ \displaystyle\sum_{r'\in\mathcal{N}(r)\cap V^{(i)}(s)} \frac{p^*(r,r')}{1-p^*(r,r)}, & s\in\mathcal{S}^{(i)}\backslash N^{(i)}. \end{cases}$$

Proof: Fix $1\leq i\leq\mathfrak{n}$ and write Y_n for $Y_n^{(i)}$. The first step is to verify that, as $\beta\to\infty$,

$$\mathbb{P}(Y_{n+1}=s|Y_n=r,Y_{n-1}=m_{n-1},...,Y_0=m_0) \;=\; \mathbb{P}_r(Y_1=s)+o(1)$$

for all $m_0,...,m_{n-1},r,s\in\mathcal{S}^{(i)}$ and $n\geq 0$. If $r\in N^{(i)}$, then $Y_n=X_{\sigma_n}$ and the Markov property of X provide us with the even stronger result

$$\mathbb{P}(Y_{n+1}=s|Y_n=r,Y_{n-1}=m_{n-1},...,Y_0=m_0) \;=\; \mathbb{P}_r(Y_1=s).$$

A little more care is needed if $r\in\mathcal{S}^{(i)}\backslash N^{(i)}$. For any $s\in\mathcal{S}^{(i)}$, $x\in V^{(i)}(r)$ and $n\geq 0$, we have

$$\begin{aligned}\mathbb{P}(Y_{n+1}=s|Y_n=r,X_{\sigma_n}=x) \;&=\; \mathbb{P}_x(Y_1=s,\tau_r<\sigma_1)+\mathbb{P}_x(Y_1=s,\tau_r>\sigma_1) \\ &=\; \mathbb{P}_r(Y_1=s)\,\mathbb{P}_x(\tau_r<\sigma_1)+\mathbb{P}_x(Y_1=s,\tau_r>\sigma_1).\end{aligned}$$

The last two summands can further be bounded by

$$\mathbb{P}_r(Y_1=s)\,\mathbb{P}_x(\tau_r<\sigma_1) \;\leq\; \mathbb{P}_r(Y_1=s) \quad\text{and}\quad \mathbb{P}_x(Y_1=s,\tau_r>\sigma_1) \;\leq\; \mathbb{P}_x(\sigma_1<\tau_r).$$

For the last probability, Theorem 2.1.3 ensures

$$\mathbb{P}_x(\sigma_1<\tau_r) \;\leq \sum_{z\in\partial^+V^{(i)}(r)}\mathbb{P}_x(\tau_z<\tau_r) \;\overset{\beta\to\infty}{\longrightarrow}\; 0.$$

Consequently, as $\beta\to\infty$,

$$\begin{aligned}\mathbb{P}_r(Y_1=s) \;&=\; \left(1-\sum_{z\in\partial^+V^{(i)}(r)}\tilde{\varepsilon}(x,r,z,\beta)\right)\mathbb{P}_r(Y_1=s)+o(1) \\ &\leq\; \big(1-\mathbb{P}_x(\sigma_1<\tau_r)\big)\,\mathbb{P}_r(Y_1=s)+o(1) \\ &\leq\; \mathbb{P}(Y_{n+1}=s|X_{\sigma_n}=x,Y_n=r)+o(1) \\ &\leq\; \mathbb{P}_r(Y_1=s)+o(1),\end{aligned}$$

and therefore

$$\begin{aligned}&\mathbb{P}(Y_{n+1}=s|Y_n=r,Y_{n-1}=m_{n-1},...,Y_0=m_0) \\ &\qquad=\sum_{x\in V^{(i)}(r)}\mathbb{P}(Y_{n+1}=s|X_{\sigma_n}=x,Y_n=r)\,\mathbb{P}(X_{\sigma_n}=x|Y_n=r,Y_{n-1}=m_{n-1},...,Y_0=m_0) \\ &\qquad=\;\mathbb{P}_r(Y_1=s)+o(1).\end{aligned}$$

It remains to verify that $\mathbb{P}_r(Y_1 = s) = \widehat{p}(r,s) + o(1)$ for any $r, s \in \mathcal{S}^{(i)}$. First, if $r \in \mathcal{S}^{(i)} \backslash N^{(i)}$, then $\sigma_1 = \tau_{N^{(i)}} = \zeta_0$ and $Y_1 = X_{\zeta_0}$. Since $E(z^*(r,s)) < E(z^*(r,s'))$ for any $s \in s_r$ and any $s' \in \partial^+ V^{(i)}(r) \backslash s_r$, we now infer with the help of Theorem 2.1.1

$$\begin{aligned}
\mathbb{P}_r(Y_1 \notin s_r) &= \mathbb{P}_r\big(\tau_{s_r} > \tau_{s'} \text{ for some } s' \in \partial^+ V^{(i)}(r) \backslash s_r\big) \\
&\leq \sum_{s' \in \partial^+ V^{(i)}(r) \backslash s_r} \mathbb{P}_r\big(\tau_s > \tau_{s'}\big) \\
&\leq \sum_{s' \in \partial^+ V^{(i)}(r)} \varepsilon(r,s,s',\beta) \\
&= o(1),
\end{aligned}$$

as $\beta \to \infty$, where s should denote an arbitrary state in s_r. This yields $\mathbb{P}_r(Y_1 = s) = \mathbb{P}_r(X_{\zeta_0} = s) + o(1)$ and $\mathbb{P}_r(X_{\zeta_0} = s) \to 0$ if $s \notin s_r$. If otherwise $s \in s_r$, then let $s' \sim s$ be a neighboring state of s in $V^{(i)}(r)$ with $E(z^*(r,s')) < E(s_r)$. Thus,

$$\mathbb{P}_r(\tau_{s'} > \zeta_0) \;\leq \sum_{s \in \partial^+ V^{(i)}(r)} \mathbb{P}_r(\tau_{s'} > \tau_s) \;\leq \sum_{s \in \partial^+ V^{(i)}(r)} \varepsilon(r,s',s,\beta) \;\leq\; \frac{1}{2}$$

for β large enough. This yields

$$\begin{aligned}
\mathbb{P}_r(Y_1 = s) &\geq \mathbb{P}_r(\tau_{s'} < \zeta_0, Y_1 = s) \\
&\geq \mathbb{P}_r(\tau_{s'} < \zeta_0) \mathbb{P}_{s'}(Y_1 = s) \\
&= \mathbb{P}_r(\tau_{s'} < \zeta_0) \frac{\mathbb{P}_{s'}\big(\tau_s < \tau_{\partial^+ V^{(i)}(r) \cup \{s'\} \backslash \{s\}}\big)}{\mathbb{P}_{s'}(\tau_{s_r} < \tau_{s'})} \\
&\geq \mathbb{P}_r(\tau_{s'} < \zeta_0) \frac{p(s',s)}{\sum_{s \in s_r} \mathbb{P}_{s'}(\tau_s < \tau_{s'})} \\
&\geq \mathbb{P}_r(\tau_{s'} < \zeta_0) \frac{e^{-\beta(E(s_r) - E(s') + \gamma_\beta)}}{\sum_{s \in s_r} e^{-\beta(E(s_r) - E(s') - 2|\mathcal{S}|\gamma_\beta)}} \\
&\geq \mathbb{P}_r(\tau_{s'} < \zeta_0) \frac{e^{-3\beta|\mathcal{S}|\gamma_\beta}}{|s_r|},
\end{aligned}$$

by use of Proposition 2.1.8 and Theorem 2.1.1. We can finally conclude that

$$\lim_{\beta_n \to \infty} \mathbb{P}_r(Y_1 = s) \;\geq\; \lim_{\beta \to \infty} \frac{1}{2|\mathcal{S}|} e^{-3\beta|\mathcal{S}|\gamma_\beta} \;>\; 0.$$

Here we need to pass on to the proper annealing schedule to ensure the existence of the limit. Second, if $r \in N^{(i)}$, then either $Y_1 = s \in \mathcal{N}(r) \cap N^{(i)}$, or otherwise $Y_1 = s \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $X_{\sigma_1} = r'$ for some $r' \in \mathcal{N}(r) \cap V^{(i)}(s)$. It thus follows that

$$\mathbb{P}_r(Y_1 = s) \;=\; \mathbb{P}_r(X_{\sigma_1} = s) \;=\; \frac{p(r,s)}{1 - p(r,r)} \;=\; \frac{p^*(r,s)}{1 - p^*(r,r)} + o(1)$$

if $s \in \mathcal{N}(r) \cap N^{(i)}$, while

$$\mathbb{P}_r(Y_1 = s) \;=\; \sum_{r' \in \mathcal{N}(r) \cap V^{(i)}(s)} \mathbb{P}_r(X_{\sigma_1} = r') \;=\; \sum_{r' \in \mathcal{N}(r) \cap V^{(i)}(s)} \frac{p^*(r,r')}{1 - p^*(r,r)} + o(1)$$

in the second case. □

3.2.2. The Asymptotic Jump Chain

The limiting chain $\widehat{Y}^{(i)} = (\widehat{Y}_n^{(i)})_{n\geq 0}$ from above is called the *asymptotic jump chain at level i* hereafter. Note that, typically, it is not irreducible. It may have transient states, not necessarily non-assigned, and several irreducibility classes. For the energy landscape given in Figure 1.1 (see also Figure 1.2), the states $2, 3$ and 11 are transient states of $\widehat{Y}^{(4)}$ and $\{4, 7, 10\}$ is one recurrence class.

Proposition 3.2.4. *Let $\widehat{Y} := \widehat{Y}^{(i)}$ be the asymptotic jump chain at level i (according to a proper annealing schedule) and write $s_m := m$ for $m \in N^{(i)}$.*

(a) A state $m \in \mathcal{S}^{(i)}$ is recurrent for $\widehat{Y}$ if and only if $E(s_{m'}) = E(s_m)$ for every $m' \in \mathcal{S}^{(i)}$ with $E(z^*(m, m')) = E(s_m)$. In particular, if for $m \in \mathcal{S}^{(i)}\backslash N^{(i)}$ one $s \in s_m$ is transient, then already every $s \in s_m$ and m itself is transient.

(b) Any two recurrent states $m, n \in \mathcal{S}^{(i)}$ are in the same irreducibility class of $\widehat{Y}$ if and only if $E(z^*(m, n)) = E(s_m) = E(s_n)$. In particular, if $E(z^*(r, s)) \equiv E$ for every $r, s \in \mathcal{S}^{(i)}\backslash N^{(i)}$ and some $E \in \mathbb{R}^+$, then every $s \in \mathcal{S}^{(i)}\backslash N^{(i)}$ is recurrent and they are all the the same irreducibility class of $\widehat{Y}$.

Proof: For part (a), first note that m is not reachable by $\widehat{Y}$ from any state m' with $E(s_{m'}) < E(s_m)$. Therefore, if there is some m' with $E(s_{m'}) < E(s_m)$ and $E(z^*(m, m')) = E(s_m)$, then there is a path γ from m to m' of states in $\mathcal{S}^{(i)}$ which does not exceed $E(s_m)$. Let γ_j be the first state n along this path with $E(s_n) < E(s_m)$. Then $\widehat{p}(\gamma_i, \gamma_{i+1}) > 0$ for every $0 \leq i \leq j - 1$. This means that γ_j is reachable from m but m is not reachable from γ_j. Thus, m is transitive. On the other hand, if $E(s_{m'}) = E(s_m)$ for every $m' \in \mathcal{S}^{(i)}$ with $E(z^*(m, m')) = E(s_m)$, then m is reachable from any state s which is reachable from m. Since, furthermore, for any $r, r' \in s_m$,

$$E(z^*(r, s)) \;\leq\; E(z^*(r, m)) \vee E(z^*(m, r')) \vee E(z^*(r', s)) \;=\; E(s_m) \vee E(z^*(r', s)) \;=\; E(z^*(r', s)),$$

and in the same way $E(z^*(r', s)) \leq E(z^*(r, s))$ for any $s \in \mathcal{S}^{(i)}$, the second claim in (a) follows.

(b) It suffices to note that in the proof of (a) we have seen that a recurrent state m is reachable by and reaches every m' with $E(z^*(m, m')) = E(s_m) = E(s_{m'})$. □

3.2.3. Semi-Markov Property

Having shown that $Y^{(i)}$ behaves asymptotically as a Markov chain, viz. the jump chain $\widehat{Y}^{(i)}$, it is fairly easy to verify with the help of the next simple lemma that the augmented bivariate AC $\big(\overline{Y}_n^{(i)}, \overline{Y}_{n+1}^{(i)}\big)_{n\geq 0}$ is asymptotically semi-Markovian.

Lemma 3.2.5. *For each $\beta > 0$, the sojourn times $\sigma_{n+1} - \sigma_n$, $n \geq 0$, of the AC $\overline{Y}^{(i)}$ are conditionally independent given $Y^{(i)}$. The conditional law of $\sigma_{n+1} - \sigma_n$ depends only on $(Y_{n-1}^{(i)}, Y_n^{(i)}, Y_{n+1}^{(i)})$ and satisfies*

$$\begin{aligned}\mathbb{P}\big(\sigma_{n+1} - \sigma_n \in \cdot \,|Y_{n-1}^{(i)} = x, Y_n^{(i)} = y, Y_{n+1}^{(i)} = z\big) \;&=\; Q((x, y, z), \cdot\,) \\ &:= \begin{cases} \mathit{Geom}(1 - p(y, y)), & \text{if } y \in N^{(i)} \\ \sum_{s\in V^{(i)}(y), s\sim x} \mathbb{P}_s(\sigma_1 \in \cdot \,|Y_1^{(i)} = z)\, \mathbb{P}_x(X_{\sigma_1} = s), & \text{if } y \notin N^{(i)} \end{cases}\end{aligned} \tag{3.2}$$

for all $x, y, z \in \mathcal{S}^{(i)}$ with $\mathbb{P}(Y_{n-1}^{(i)} = x, Y_n^{(i)} = y, Y_{n+1}^{(i)} = z) > 0$ and $n \geq 1$.

Proof: The assertions follow easily when observing that, on the one hand, at least one state $y \in N^{(i)}$ must be visited between two states $x, z \in \mathcal{S}^{(i)}\backslash N^{(i)}$ (Lemma 1.3.10) and that, on the other hand, the original chain X and its aggregation $\overline{Y}^{(i)}$ coincide at any epoch where a non-assigned state is hit, which renders the Markov property of $\overline{Y}^{(i)}$ at these epochs. Further details are omitted. □

In order to formulate the next result, let $0 = \widehat{\sigma}_0 < \widehat{\sigma}_1 < ...$ be an increasing sequence of random variables such that its increments $\widehat{\sigma}_{n+1} - \widehat{\sigma}_n$, $n \geq 0$, are conditionally independent given the asymptotic jump chain $\widehat{Y}^{(i)}$. Moreover, let the conditional law of $\widehat{\sigma}_{n+1} - \widehat{\sigma}_n$ depend only on $\big(\widehat{Y}^{(i)}_{n-1}, \widehat{Y}^{(i)}_n, \widehat{Y}^{(i)}_{n+1}\big)$ and be equal to $Q\big(\big(\widehat{Y}^{(i)}_{n-1}, \widehat{Y}^{(i)}_n, \widehat{Y}^{(i)}_{n+1}\big), \cdot\,\big)$, with Q as defined in (3.2). Then $((\widehat{Y}^{(i)}_n, \widehat{Y}^{(i)}_{n+1}), \widehat{\sigma}_{n+1})_{n\geq 0}$ forms a Markov renewal process and $\big(\widehat{Y}^{(i)}_{\widehat{\nu}(n)}, \widehat{Y}^{(i)}_{\widehat{\nu}(n+1)}\big)_{n\geq 0}$ a semi-Markov chain, where $\widehat{\nu}(n) := \sup\{k \geq 0 | \widehat{\sigma}_k \leq n\}$.

Proposition 3.2.6. *For each $1 \leq i \leq \mathfrak{n}$, $((Y^{(i)}_n, Y^{(i)}_{n+1}), \sigma_{n+1})_{n\geq 0}$ converges to the Markov renewal process $((\widehat{Y}^{(i)}_n, \widehat{Y}^{(i)}_{n+1}), \widehat{\sigma}_{n+1})_{n\geq 0}$ as $\beta_n \to \infty$ along a proper annealing schedule, in the sense that*

$$\lim_{\beta_n\to\infty} \frac{\mathbb{P}_{y_0}\big(\big(Y^{(i)}_k, Y^{(i)}_{k+1}\big) = (y_k, y_{k+1}),\, \sigma_{k+1} = i_{k+1},\, 0 \leq k \leq n\big)}{\mathbb{P}_{y_0}\big(\big(\widehat{Y}^{(i)}_k, \widehat{Y}^{(i)}_{k+1}\big) = (y_k, y_{k+1}),\, \widehat{\sigma}_{k+1} = i_{k+1},\, 0 \leq k \leq n\big)} \;=\; 1$$

for all $y_0,...,y_{n+1} \in \mathcal{S}^{(i)}$, $0 < i_1 < ... < i_{n+1}$ and $n \geq 0$ such that the denominator is positive. Furthermore, $(\overline{Y}^{(i)}_n, \overline{Y}^{(i)}_{n+1})_{n\geq 0}$ is asymptotically semi-Markovian in the sense that

$$\lim_{\beta_n\to\infty} \frac{\mathbb{P}_{y_0}\big(\big(\overline{Y}^{(i)}_k, \overline{Y}^{(i)}_{k+1}\big) = (y_k, y_{k+1}), 0 \leq k \leq n\big)}{\mathbb{P}_{y_0}\big(\big(\widehat{Y}^{(i)}_{\widehat{\nu}(k)}, \widehat{Y}^{(i)}_{\widehat{\nu}(k+1)}\big) = (y_k, y_{k+1}), 0 \leq k \leq n\big)} \;=\; 1$$

for all $y_0,...,y_{n+1} \in \mathcal{S}^{(i)}$ and $n \geq 0$ such that the denominator is positive.

Proof: The first assertion being obvious by Proposition 3.2.3, note that it implies, with $\nu(n) := \sup\{k \geq 0 | \sigma_k \leq n\}$,

$$\lim_{\beta_n\to\infty} \frac{\mathbb{P}_{y_0}\big(\big(Y^{(i)}_{\nu(k)}, Y^{(i)}_{\nu(k+1)}\big) = (y_k, y_{k+1}), 0 \leq k \leq n\big)}{\mathbb{P}_{y_0}\big(\big(\widehat{Y}^{(i)}_{\widehat{\nu}(k)}, \widehat{Y}^{(i)}_{\widehat{\nu}(k+1)}\big) = (y_k, y_{k+1}), 0 \leq k \leq n\big)} \;=\; 1$$

for all $y_0,...,y_{n+1} \in \mathcal{S}^{(i)}$ and $n \geq 0$ such that the denominator is positive. Therefore, the second assertion follows when finally noting that

$$Y^{(i)}_{\nu(n)} \;=\; \sum_{j\geq 0} Y^{(i)}_j \mathbb{1}_{\{\sigma_j \leq n < \sigma_{j+1}\}} \;=\; \overline{Y}^{(i)}_n$$

for each $n \geq 0$. □

So we have shown that, although aggregation generally entails the loss of the Markov property, here it leads back to processes of this kind (Markov or semi-Markov chains) in an asymptotic sense at low temperature regimes. This confirms Property 1.

3.3. Reciprocating Jumps

As discussed to some extent in the Introduction, we want to find an aggregation level at which *reciprocating jumps* appear to be very unlikely so as to obtain a better picture of essential features of the observed process. To render precision to this informal statement requires to further specify the term "reciprocating jump" and to provide a measure of likelihood for its occurrence. It is useful to point out first that the original chain X exhibits two types of forward-backward jumps (compare Figure 3.1):

Intra-valley jumps which occur between states inside a valley (starting in a minimum the process falls back to it many times before leaving the valley).

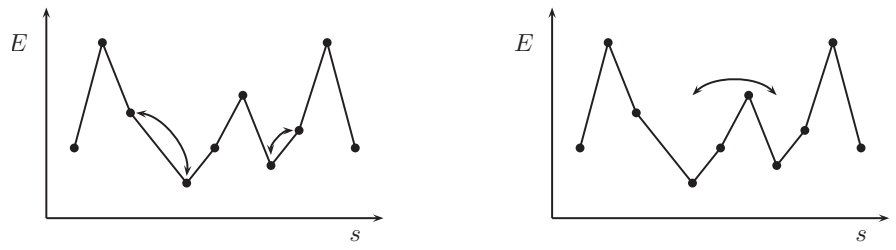

Figure 3.1.: Illustration of intra-valley jumps (left panel) versus inter-valley jumps (right panel).

Inter-valley jumps which occur between two valleys (typically, when the energy barrier between these valleys is much lower than the barrier to any other valley).

Clearly, intra-valley jumps disappear by aggregating valleys into metastates, while inter-valley jumps may also be viewed as intra-valley jumps for higher order valleys and do occur when transitions between any two of them are much more likely than those to other valleys, in which case they should be aggregated into one valley. This motivates the following definition.

Definition 3.3.1. We say the process $(Y_n^{(i)})_{n\in\mathbb{N}}$ exhibits *reciprocating jumps of order* $\varepsilon \geq 0$ if there exists a nonempty subset $A \subsetneq \mathcal{S}^{(i)}\backslash N^{(i)}$ with the following property: For each $m_1 \in A$, there exists $m_2 \in A$ such that

$$\liminf_{\beta\to\infty} \frac{1}{\beta}\left(\ln\left(\mathbb{P}_{m_1}\left(X_{\xi_1} \in V^{(i)}(m_2)\right)\right) - \ln\left(\mathbb{P}_{m_1}\left(X_{\xi_1} \in V^{(i)}(m)\right)\right)\right) \;\geq\; \varepsilon$$

for all $m \in \mathcal{S}^{(i)}\backslash(N^{(i)} \cup A)$. In other words, it is exponentially (with rate ε) more likely to stay in A than to leave it (ignoring intermediate visits to non-assigned states).

In view of our principal goal to give a path-independent definition of MBs, we must point out that, by irreducibility, forward-backward jumps *always* occur with positive probability at any nontrivial level of aggregation and can therefore never be ruled out completely. This is in contrast to the path-dependent version by Heuer [29] in which the non-occurrence of forward-backward jumps appears to be the crucial requirement. As a consequence, Definition 3.3.1 provides an alternative, probabilistic and verifiable criterion for reciprocating jumps to be sufficiently unlikely in a chosen aggregation.

3.3.1. Macroscopic Transition Probabilities

The following proposition contains further information on which valleys are visited consecutively by providing the probabilities of a transition from $V^{(i)}(m)$ to $V^{(i)}(m')$ for any $m, m' \in \mathcal{S}^{(i)}\backslash N^{(i)}$. It is a direct consequence of the asymptotic results in the previous section, notably Proposition 3.2.3.

Proposition 3.3.2. *Let $m \in \mathcal{S}^{(i)}\backslash N^{(i)}, s_m \subset \partial^+ V^{(i)}(m)$ be as defined in* (2.2)*. Then*

$$\lim_{\beta_n\to\infty} \mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m'))$$
$$= \sum_{s\in s_m} \widehat{p}(m,s)\left(\widehat{p}(s,m') + \sum_{n\geq 1}\sum_{r_1,\ldots,r_n\in N^{(i)}} \widehat{p}(s,r_1)\cdot\ldots\cdot\widehat{p}(r_{n-1},r_n)\,\widehat{p}(r_n,m')\right)$$

for any $m' \in \mathcal{S}^{(i)}\backslash N^{(i)}$.

The reader should notice that, as $\widehat{p}(s,r) = 0$ whenever $E(s) < E(r)$, the last sum actually ranges only over those non-assigned $r_1,...,r_n$ with $E(s_m) \geq E(r_1) \geq ... \geq E(r_n) \geq E(s_{m'})$. In particular for $m' = m$ it has to be $E(r_i) = E(s_m)$ for every $1 \leq i \leq n$.

Proof: Let us first point out that $\mathbb{P}_r(X_{\xi_0} \in V^{(i)}(m)) = o(1)$ as $\beta \to \infty$ for any $r \in N^{(i)}$ such that $E(r) < E(s_m)$. Namely, since the last property implies $r \notin \partial^+ V^{(i)}(m)$, any path from r into $V^{(i)}(m)$ must traverse a state $s \in \partial^+ V^{(i)}(m)$ with $E(s) \geq E(s_m) > E(r)$, whence the probability for such a path goes to zero as $\beta \to \infty$. Noting further that $\mathbb{P}_m(Y_1^{(i)} \notin s_m) = o(1)$ as $\beta \to \infty$ by Proposition 3.2.3, we now infer (with $\xi_n = \xi_n^{(i)}$)

$$\mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m')) \;=\; \sum_{s \in s_m} \mathbb{P}_m(Y_1^{(i)} = s) \mathbb{P}_s(X_{\xi_0} \in V^{(i)}(m')) + o(1)$$

with

$$\begin{aligned} &\mathbb{P}_s(X_{\xi_0} \in V^{(i)}(m')) \\ &\quad= \; \mathbb{P}_s(Y_1^{(i)} = m') \;+\; \sum_{n \geq 1} \sum_{r_1,...,r_n \in N^{(i)}} \mathbb{P}_s(Y_1^{(i)} = r_1,..., Y_n^{(i)} = r_n, Y_{n+1}^{(i)} = m') \\ &\quad= \; \widehat{p}(s,m') + \sum_{n \geq 1} \sum_{r_1,...,r_n \in N^{(i)}} \widehat{p}(s,r_1) \cdot ... \cdot \widehat{p}(r_{n-1},r_n)\, \widehat{p}(r_n,m') + o(1) \end{aligned}$$

as $\beta_n \to \infty$. □

In essence, the previous result tells us that a valley $V^{(i)}(m')$ is neighboring to $V^{(i)}(m)$, that is, reachable with positive probability by the asymptotic jump chain $\widehat{Y}^{(i)}$ (and thus by $Y^{(i)}$ at any temperature level β) without intermediately hitting any other valley, if and only if there exists at least one (in terms of energies) decreasing path in $N^{(i)}$ from some $s \in s_m$ to $V^{(i)}(m')$. For any other such pair of valleys, connected by a non-decreasing path through states in $N^{(i)}$, the transition probability decreases to zero exponentially in β. If this path can chosen to be unimodal, here called *uphill-downhill-path*, this can be stated in a very precise way. In order to do so, we need the following lemma about the decay of the probability to leave $V^{(i)}(m)$ via a specific $s \in \partial^+ V^{(i)}(m)$.

Lemma 3.3.3. *Let $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $s \in \partial^+ V^{(i)}(m)$. Then*

$$\lim_{\beta \to \infty} \frac{1}{\beta} \ln \mathbb{P}_m(Y_1 = s) \;=\; -(E(s) - E(s_m)).$$

Proof: First, note that for $s \in s_m$ the right-hand side is zero and the probability on the left-hand side is bounded above by one and below by $C \exp(-3\beta |\mathcal{S}| \gamma_\beta)$, which converges to some strictly positive constant. Therefore, the whole term on the left-hand side converges to zero as well, in particular independently of the annealing schedule. We proceed to $s \in \partial^+ V^{(i)}(m) \backslash s_m$. Decompose the event $\{Y_1 = s\}$ with respect to the number of visits to m before $V(m) := V^{(i)}(m)$ is left (or use Proposition 2.1.8), giving

$$\mathbb{P}_m(Y_1 = s) \;=\; \frac{\mathbb{P}_m(\sigma_1 = \tau_s < \tau_m)}{\mathbb{P}_m(\sigma_1 < \tau_m)}.$$

For the upper bound, we obtain by another appeal to Proposition 2.1.5 that

$$\mathbb{P}_m(Y_1 = s) \;=\; \frac{\mathbb{P}_m(\sigma_1 = \tau_s < \tau_m)}{\mathbb{P}_m(\sigma_1 < \tau_m)} \;\leq\; \frac{\mathbb{P}_m(\tau_s < \tau_m)}{\mathbb{P}_m(\tau_{s_m} < \tau_m)} \;\leq\; e^{-\beta(E(s) - E(s_m) + o(1))}.$$

The proof of the lower bound is much more technical. Let $\gamma = (\gamma_1,...,\gamma_n) \in \Gamma^*(m,s)$ be a minimal path which leaves $V(m)$ only in the last step and such that for any $\gamma_i, \gamma_j \in \gamma$ both, the subpath from γ_i to γ_j, and the inversed path from γ_j to γ_i, are minimal. Define

$$r_0 := m \quad \text{and} \quad r_1 := \gamma_{i_0} \quad \text{with} \quad i_0 := \inf\{0 \le i \le n-1 | E(\gamma_{i+1}) \ge E(s_m)\}.$$

In particular, $E(r_1) < E(s_m)$ and $E(z^*(r_1,r_0)) < E(s_m)$. Define furthermore the first record by $s_1 := \gamma_{i_1}$ with

$$i_1 \;:=\; \inf\Big\{ i_0 < i \le n \,\Big|\, E(\gamma_i) \ge E(s_m),\, \inf\{j \ge i | E(\gamma_j) < E(\gamma_i)\} < \inf\{j \ge i | E(\gamma_j) > E(\gamma_i)\} \Big\},$$

and then successively for $k \ge 1$ with $s_k = \gamma_{i_k} \ne s$ the records $s_{k+1} := \gamma_{i_{k+1}}$ with

$$\begin{aligned} i_{k+1} \;:=\; \inf\Big\{ n \ge i \ge \inf\{j \ge i_k | E(\gamma_j) < E(s_k)\} \,\Big|\, & E(\gamma_i) \ge E(s_k), \\ & \inf\{j \ge i | E(\gamma_j) < E(\gamma_i)\} < \inf\{j \ge i | E(\gamma_j) > E(\gamma_i)\} \Big\}. \end{aligned}$$

Note that the energy of these records is increasing, though not strictly increasing. Let s_{k-1} be the last record defined in this way and $s_k := s$. Since $E(z^*(m,s)) = E(s)$, s_k is as well a record. Given the records $s_1,...,s_k$, for $1 \le i \le k-1$ let r_{2i} be the first (not necessarily isolated) minimum along γ after s_i and r_{2i+1} the last (not necessarily isolated) minimum along γ before s_{i+1}. Here a minimum along γ is some $\gamma_i \in \gamma$ such that it is a minimum of E restricted to γ. Finally, let $r_{2k} := s_k = s$. In the following we will proof that

(a) $\mathbb{P}_{r_{2j}}(\tau_{r_{2j+1}} < \zeta_0) \to 1$ as $\beta \to \infty$ for any $0 \le j \le k-1$,

(b) $\mathbb{P}_{r_1}(\tau_{r_2} < \zeta_0) \ge e^{-\beta(E(z^*(r_1,r_2)) - E(s_m) + o(1))}$,

(c) $\mathbb{P}_{r_{2j+1}}(\tau_{r_{2j+2}} < \zeta_0) \ge e^{-\beta(E(z^*(r_{2j+1},r_{2j+2})) - E(z^*(r_{2j-1},r_{2j})) + o(1))}$ for any $1 \le j \le k-2$,

(d) $\mathbb{P}_{r_{2k-1}}(\tau_{r_{2k}} = \zeta_0) \ge e^{-\beta(E(z^*(r_{2k-1},r_{2k})) - E(z^*(r_{2k-3},r_{2k-2})) + o(1))}$.

This gives for β large enough

$$\begin{aligned} \mathbb{P}_m(Y_1 = s) \;&\ge\; \left(\prod_{j=0}^{2k-2} \mathbb{P}_{r_j}(\tau_{r_{j+1}} < \zeta_0) \right) \mathbb{P}_{r_{2k-1}}(\tau_{r_{2k}} = \zeta_0) \\ &\ge\; \frac{1}{2} \left(\prod_{j=0}^{k-2} \mathbb{P}_{r_{2j+1}}(\tau_{r_{2j+2}} < \zeta_0) \right) \mathbb{P}_{r_{2k-1}}(\tau_{r_{2k}} = \zeta_0) \\ &\ge\; \frac{1}{2} e^{-\beta(E(z^*(r_1,r_2)) - E(s_m) + o(1))} \cdot \prod_{j=1}^{k-1} e^{-\beta(E(z^*(r_{2j+1},r_{2j+2})) - E(z^*(r_{2j-1},r_{2j})) + o(1))} \\ &=\; \frac{1}{2} e^{-\beta(E(z^*(r_{2k-1},r_{2k})) - E(s_m) + o(1))} \\ &=\; e^{-\beta(E(s) - E(s_m) + o(1))}, \end{aligned}$$

and thus the assertion.

(a) For $j = 0$ this is obvious since $E(z^*(r_0,r_1)) < E(s_m)$. For $1 \le j \le k-1$ and any $r' \in \partial^+ V(m)$ it holds true that

$$E(z^*(r_{2j}, r_{2j+1})) \;<\; E(z^*(r_{2j}, r_{2j-1})) \;\le\; E(z^*(r_{2j}, m)) \;\le\; E(z^*(r_{2j}, r')), \tag{3.3}$$

where we make use of the fact that between r_{2j} and r_{2j+1} the energy stays below the last record, and that all subpaths of γ are minimal as well.

(b) By the definition of r_1 and r_2, there is a unimodal path between them so that the cumulative activation energy along this path equals $E(z^*(r_1, r_2)) - E(r_1)$. Therefore

$$\mathbb{P}_{r_1}(\tau_{r_2} < (\zeta_0 \wedge \tau_{r_1})) \;\geq\; e^{-\beta(E(z^*(r_1,r_2))-E(r_1)+o(1))}.$$

Furthermore, for any $r' \in \partial^+ V(m)$ we have

$$E(z^*(r_1, r')) \;\geq\; E(s_m) > E(r_1) \quad \text{and} \quad E(z^*(r_1, r_2)) \;\geq\; E(s_m) > E(r_1).$$

Therefore,

$$\begin{aligned}
\mathbb{P}_{r_1}((\tau_{r_2} \wedge \zeta_0) < \tau_{r_1}) \;&\leq \sum_{r' \in \partial^+ V(m) \cup \{r_2\}} \mathbb{P}_{r_1}(\tau_{r'} < \tau_{r_1}) \\
&\leq \sum_{r' \in \partial^+ V(m) \cup \{r_2\}} e^{-\beta(E(z^*(r_1,r'))-E(r_1)+o(1))} \\
&\leq\; e^{-\beta(E(s_m)-E(r_1)+o(1))}.
\end{aligned}$$

Combining the two estimates, we get

$$\mathbb{P}_{r_1}(\tau_{r_2} < \zeta_0) \;=\; \frac{\mathbb{P}_{r_1}(\tau_{r_2} < (\zeta_0 \wedge \tau_{r_1}))}{\mathbb{P}_{r_1}((\tau_{r_2} \wedge \zeta_0) < \tau_{r_1})} \;\geq\; e^{-\beta(E(z^*(r_1,r_2))-E(s_m)+o(1))}.$$

(c) Let $1 \leq j \leq k-2$. We use the same strategy as in the proof of (b). So, again, by the definition of r_{2j+1} and r_{2j+2}, there is a unimodal path between them with cumulative activation energy $E(z^*(r_{2j+1}, r_{2j+2})) - E(r_{2j+1})$ along this path, and

$$\mathbb{P}_{r_{2j+1}}(\tau_{r_{2j+2}} < (\zeta_0 \wedge \tau_{r_{2j+1}})) \;\geq\; e^{-\beta(E(z^*(r_{2j+1},r_{2j+2}))-E(r_{2j+1})+o(1))}.$$

Furthermore,

$$E(z^*(r_{2j-1}, r_{2j})) \;\leq\; E(z^*(r_{2j+1}, r_{2j+2})) \quad \text{and} \quad E(r_{2j+1}) \;<\; E(z^*(r_{2j+1}, r_{2j+2})).$$

Finally, for any $r' \in \partial^+ V(m)$, by use of Equation (3.3),

$$\begin{aligned}
E(r_{2j+1}) \;&\leq E(z^*(r_{2j+1}, r_{2j})) \\
&<\; E(z^*(r_{2j}, r')) \\
&\leq\; E(z^*(r_{2j}, r_{2j+1})) \vee E(z^*(r_{2j+1}, r')) \\
&=\; E(z^*(r_{2j+1}, r')).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{P}_{r_{2j+1}}((\tau_{r_{2j+2}} \wedge \zeta_0) < \tau_{r_{2j+1}}) \;&\leq \sum_{r' \in \partial^+ V(m) \cup \{r_{2j+2}\}} \mathbb{P}_{r_{2j+1}}(\tau_{r'} < \tau_{r_{2j+1}}) \\
&\leq \sum_{r' \in \partial^+ V(m) \cup \{r_{2j+2}\}} e^{-\beta(E(z^*(r_{2j+1},r'))-E(r_{2j+1})+o(1))} \\
&\leq\; e^{-\beta(E(r_{2j-1},r_{2j})-E(r_{2j+1})+o(1))}.
\end{aligned}$$

Combining the two estimates, we get

$$\mathbb{P}_{r_{2j+1}}(\tau_{r_{2j+2}} < \zeta_0) \;=\; \frac{\mathbb{P}_{r_{2j+1}}(\tau_{r_{2j+2}} < (\zeta_0 \wedge \tau_{r_{2j+1}}))}{\mathbb{P}_{r_{2j+1}}((\tau_{r_{2j+2}} \wedge \zeta_0) < \tau_{r_{2j+1}})} \;\geq\; e^{-\beta(E(z^*(r_{2j+1},r_{2j+2}))-E(r_{2j-1},r_{2j})+o(1))}.$$

(d) All bounds for energies in (c) can be proved in the very same way (here $r' \in \partial^+ V(m)\backslash\{s\}$), so that

$$\mathbb{P}_{r_{2k-1}}(\tau_{r_{2k}} = \zeta_0) \;=\; \frac{\mathbb{P}_{r_{2k-1}}(\tau_{r_{2k}} = (\zeta_0 \wedge \tau_{r_{2k-1}}))}{\mathbb{P}_{r_{2k-1}}((\tau_{r_{2k}} \wedge \zeta_0) < \tau_{r_{2k-1}})} \;\geq\; e^{-\beta(E(z^*(r_{2k-1},r_{2k}))-E(r_{2k-3},r_{2k-2})+o(1))}.$$

□

Lemma 3.3.4. *Let $m_0, m_1 \in \mathcal{S}^{(i)}\backslash N^{(i)}$ be two distinct local minima for some $1 \leq i \leq \mathfrak{n}$.*

(a) It holds true that

$$\limsup_{\beta\to\infty} \frac{1}{\beta} \ln \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \;\leq\; -\left(E(z^*(m_0,m_1)) - E(s_{m_0})\right).$$

(b) Suppose there exists a path $\gamma = (\gamma_0,..., \gamma_k)$ from some $s \in \partial^+ V^{(i)}(m_0)$ with $E(z^*(s,m_1)) = E(z^*(m_0,m_1))$ to $V^{(i)}(m_1)$, avoiding every other valley. Then

$$\liminf_{\beta\to\infty} \frac{1}{\beta} \ln \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \;\geq\; -\left(I(\gamma_0,..., \gamma_k) + E(s) - E(s_{m_0})\right).$$

Note that the equality $I(\gamma_0,..., \gamma_k) = E(z^*(s,m_1)) - E(s)$, in which case

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \;=\; -\left(E(z^*(m_0,m_1)) - E(s_{m_0})\right),$$

does indeed imply the already mentioned property

$$E(\gamma_i) \;\geq\; E(\gamma_{i-1}) \quad \text{for } 1 \leq i \leq j \quad \text{and} \quad E(\gamma_i) \;\leq\; E(\gamma_{i-1}) \quad \text{for } j+1 \leq i \leq k$$

with $\gamma_j \in z^*(s,m_1)$. We call such a path an *uphill-downhill-path* because it first straddles the energy barrier $E(z^*(s,m_1))$ and then falls down to the valley $V^{(i)}(m_1)$. The existence of such a path can be found in most 2- or higher dimensional energy landscapes. Furthermore, we write $m_0 \upharpoonleft\downharpoonright m_1$ if there exists an uphill-downhill-path from some $s \in \partial^+ V^{(i)}(m_0)$ with $E(z^*(s,m_1)) = E(z^*(m_0,m_1))$ to $V^{(i)}(m_1)$ which avoids every other valley.

Proof: With γ as stated, the lower bound for $\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1))$ in (b) is easily obtained as follows:

$$\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \;\geq\; \mathbb{P}_{m_0}(X_{\zeta_0+i} = \gamma_i,\, 0 \leq i \leq k) \;\geq\; \mathbb{P}_{m_0}(X_{\zeta_0} = s)\, e^{-\beta I(\gamma_0,\ldots,\gamma_k) - \gamma_\beta \beta |\mathcal{S}|}.$$

Since $\mathbb{P}_{m_0}(X_{\zeta_0} = s) = e^{-\beta(E(s)-E(s_{m_0})+o(1))}$ by Lemma 3.3.3, the exponential decay in β is with an exponent of at most $I(\gamma_0,...,\gamma_k) + E(s) - E(s_{m_0})$ as claimed.

For the upper bound in (a), we decompose the event into disjoint sets depending on the number of visits N, say, to m_0 between 1 and $\zeta_0 = \zeta_0^{(i)}$ (as $\zeta_0 \neq \xi_1$, we can not apply Proposition 2.1.8). This leads to

$$\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N=0) \;=\; \mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})$$

and, for $k \geq 1$,

$$\begin{aligned} &\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N=k) \\ &\qquad = \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1),\, |\{\tau_{m_0} < n \leq \zeta_0 | X_n = m_0\}| = k-1,\, \tau_{m_0} < \zeta_0) \\ &\qquad = \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N=k-1)\, \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0) \\ &\qquad \;\vdots \\ &\qquad = \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N=0)\, \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0)^k \\ &\qquad = \mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})\, \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0)^k. \end{aligned}$$

Consequently,

$$\begin{aligned}\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) &= \sum_{k\geq 0} \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N = k) \\ &= \sum_{k\geq 0} \mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})\, \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0)^k \\ &= \frac{\mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})}{\mathbb{P}_{m_0}(\zeta_0 < \tau_{m_0})}.\end{aligned}$$

By invoking Proposition 2.1.5, we infer

$$\begin{aligned}\frac{\mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})}{\mathbb{P}_{m_0}(\zeta_0 < \tau_{m_0})} &\leq \frac{\sum_{r\in V^{(i)}(m_1)} \mathbb{P}_{m_0}(\tau_r < \tau_{m_0})}{\mathbb{P}_{m_0}(\tau_x < \tau_{m_0})} \\ &\leq |\mathcal{S}|^3 \sum_{r\in V^{(i)}(m_1)} e^{-\beta(E(z^*(m_0,r))-E(z^*(m_0,x))-8|\mathcal{S}|\gamma_\beta)} \qquad (3.4)\end{aligned}$$

for all $x \in V^{(i)}(m_0)^c$. For any $r \in V^{(i)}(m_1)$, we have $E(z^*(m_0, r)) \geq E(z^*(m_1, r))$, and thereby

$$E(z^*(m_0, m_1)) \leq E(z^*(m_0, r)) \vee E(z^*(r, m_1)) = E(z^*(m_0, r)).$$

Using this in (3.4), we obtain

$$\frac{\mathbb{P}_{m_0}(\xi_1^{(i)} = \tau_{V^{(i)}(m_1)} < \tau_{m_0})}{\mathbb{P}_{m_0}(\zeta_0 < \tau_{m_0})} \leq |\mathcal{S}|^4 e^{-\beta(E(z^*(m_0,m_1))-E(z^*(m_0,x))-8|\mathcal{S}|\gamma_\beta)}$$

and then, upon choosing $x \in s_{m_0}$ and noting that $E(z^*(m_0, x)) = E(s_{m_0})$,

$$\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \leq |\mathcal{S}|^4 e^{-\beta\left(E(z^*(m_0,m_1))-E(s_{m_0})-8|\mathcal{S}|\gamma_\beta\right)}.$$

We finally conclude

$$\limsup_{\beta\to\infty} \frac{1}{\beta} \ln \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \leq -\left(E(z^*(m_0, m_1)) - E(s_{m_0})\right)$$

as asserted. □

To summarize, which valleys are visited consecutively depends on (a) their spatial arrangement and (b) the energy barriers between them: A transition from one valley $V^{(i)}(m_0)$ to another valley $V^{(i)}(m_1)$ is only possible if there exists a path from $s \in \partial^+ V^{(i)}(m_0)$ to $V^{(i)}(m_1)$, not hitting any other valley. This transition is made at small temperatures (i.e. large β) if the additional energy barrier $E(z^*(s_{m_0}, m_1)) - E(s_{m_0})$ is sufficiently small or in other words the energy barrier $E(z^*(s_{m_0}, m_1))$ is approximately of the same height as all other energy barriers, including the barrier $E(z^*(s_{m_0}, m_0)) = E(s_{m_0})$. Results similar to the two lemmata above can be found in [16], where they used the cycle-approach due to Freidlin and Wentzell [24]. The obtained bounds are more general but the derivation of this special case from those results would be more work-intensive than proving them as done above.

Whereas the upper bound is proved in Lemma 3.3.4 and Lemma 3.3.3 in the same way by decomposing the event according to the number of visits of the ground state before the valley is left, the proofs of the lower bound differ. For Lemma 3.3.4 we simply determine the probability of one specific path by its cumulative activation energy. For Lemma 3.3.3 we identify the relevant minima along such a path and the probabilities to reach one minimum from the other one earlier than $\partial^+ V(m)$. The crucial point is that, first, for any such minimum r_{2j+1} both, the essential saddle to the next minimum r_{2j+2}, and the essential saddle to $\partial^+ V(m)$, are at least as high as the essential saddle

between the previous pair (r_{2j-1}, r_{2j}), and, second, there is an uphill-downhill path between r_{2j+1} and r_{2j+2}. Thus, the rate for the exponential decay of the probability to reach the next minimum before $\partial^+V(m)$ is at most the *additional* energy barrier $E(z^*(r_{2j+1}, r_{2j+2})) - E(z^*(r_{2j-1}, r_{2j}))$ and not the energy barrier $E(z^*(r_{2j+1}, r_{2j+1})) - E(r_{2j+1})$. When calculating the product in the proof, this leads to the additional energy barrier $E(z^*(s,m)) - E(s_m)$ instead o the cumulative activation energy.

This strategy fails in Lemma 3.3.4: If there is a minimum r_1 along the path from s_{m_0} to $V(m_1)$, then by definition of this path, r_1 is non-assigned. Therefore, although a minimum along the path, it is no local minimum of E. In fact, there is a decreasing path to some other valley, the essential saddle being r_1 itself. Thus, the additional energy barrier equals the energy barrier $E(z^*(r_1, r_2)) - E(r_1)$, which therefore is the rate for the exponential decay of the probability to reach r_2 earlier than the union of all valleys. The sum of all these energy barriers leads to the cumulative activation energy.

Now we come back to the reciprocating jumps in the accelerated chain and see:

Proposition 3.3.5. *Fix $1 \le i \le \mathfrak{n}$ and $\varepsilon \ge 0$. For $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$, define*

$$\overline{\mathcal{N}}(m,\varepsilon) \;:=\; \left\{ m' \in \mathcal{S}^{(i)} \backslash N^{(i)} \,\middle|\, m \,\mathrel{\uparrow\!\!\downarrow}\, m',\, E(z^*(m,m')) - E(s_m) \le \varepsilon \right\}.$$

Then the AAC at level i exhibits no reciprocating jumps of order ε if for every $m, m' \in \mathcal{S}(i) \backslash N^{(i)}$ there exist a sequence $m = m_0, m_1, ..., m_n = m' \in \mathcal{S}^{(i)} \backslash N^{(i)}$ such that $m_{j+1} \in \overline{\mathcal{N}}(m_j, \varepsilon)$ for every $0 \le j \le n-1$.

Obviously, if all metastable states are in the same recurrence class of $\widehat{Y}$, that is, if $E(z^*(m,m')) \equiv E$ for every $m, m' \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and some $E \in \mathbb{R}^+$, then there are no reciprocating jumps of order $\varepsilon = 0$. Or, the other way around, if there are reciprocating jumps to some order $\varepsilon > 0$, then $\widehat{Y}$ is not irreducible (referring to $\mathcal{S}^{(i)} \backslash N^{(i)}$). But for finite β and sufficiently small ε, transitions between the different recurrence classes of $\widehat{Y}$ still are not too unlikely.

3.3.2. Diffusive Behavior

The origin of our endeavor to define aggregations with no reciprocating jumps of an order larger than a small ε is to obtain an associated process with (almost) *decorrelated increments* (in Euclidean state space), for this and a proper centering causes the mean squared displacement up to the n-th jump to grow with n instead of n^2. This is known as *normal diffusive behavior* in physics, described by *Fick's second law*, see e.g. [4, Section 21.3.3], most notably Equation (21-83) therein which says:

Definition 3.3.6. A stochastic process $(M_n)_{n\ge 0}$ on $\mathbb{R}$ is *diffusive under* $\mathbb{P}_x$, $x \in \mathbb{R}$, if there exists $a \le b \in \mathbb{R}_{>0}$ such that

$$an \;\le\; \mathbb{E}_x(M_n - M_0)^2 \;\le\; bn$$

for every $n \ge 0$.

In physical diffusive systems it is observed that $\mathbb{E}_x(M_n - M_0)^2 n^{-1}$ is almost constant in n on a moderate time scale, which depends on the system itself. This constant is called *diffusion coefficient*.

The above definition differs from the traditional, continuous time one where "[a] continuous time parameter stochastic process which possesses the (strong) Markov property and for which the sample paths $X(t)$ are (almost always) continuous functions of t is called a diffusion process" [36, Section 15.1, page 157], for example a Brownian motion. In most situations, it is additionally assumed that

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E}\left((X(t+h) - X(t))^2 \,|X(t) = x \right) \;=\; \sigma^2(x,t),$$

where $\sigma^2(x,t)$ is called *diffusion parameter*. Thus, whereas in the continuous case the diffusion parameter is defined via infinitesimal motion, we define it in accordance to the physical meaning via long-range motion. As the normalized mean squared displacement of the Brownian motion is 1, both definitions of the diffusion parameter coincide for this diffusion.

Lemma 3.3.7. *A stochastic process $(M_n)_{n\geq 0}$ on $\mathbb{R}$ whose increments $(M_n - M_{n-1})_{n\geq 1}$ are centered and uncorrelated under $\mathbb{P}_x$ (in particular a martingale in $\mathfrak{L}_2$) is diffusive under $\mathbb{P}_x$.*

Proof: The assertion can be deduced by the following easy calculation: For every $n \geq 0$ we have

$$\begin{aligned}\mathbb{E}_x\left((M_n - M_0)^2\right) &= \mathbb{E}_x\left(\left(\sum_{i=1}^{n} M_i - M_{i-1}\right)^2\right)\\ &= \mathbb{E}_x\left(\sum_{i=1}^{n}(M_i - M_{i-1})^2\right) + 2\sum_{1\leq i<j\leq n}\mathbb{E}_x\left((M_i - M_{i-1})(M_j - M_{j-1})\right)\\ &= \sum_{i=1}^{n}\mathbb{E}_x\left((M_i - M_{i-1})^2\right)\end{aligned}$$

with

$$n \cdot \min_{x\sim y}(x-y)^2 \;\leq\; \sum_{i=1}^{n}\mathbb{E}_x\left((M_i - M_{i-1})^2\right) \;\leq\; n \cdot \max_{x\sim y}(x-y)^2.$$

□

In our situation, identifying $\mathcal{S}$ with a subspace of $\mathbb{R}^d$ for some $d \in \mathbb{N}$, without aggregation the increments are highly correlated due to the following argument: At any given time, the process is with high probability in a minimum and when leaving it, say by making a positive jump, the next increment is most likely negative because of the drift back to the minimum. When regarding the asymptotic jump chain, correlations like the above vanish: On an irreducibility class there are no forward-backward jumps, that is, no reciprocating jumps of any order $\varepsilon \geq 0$, when the energy landscape is well behaving. However, we still have to grapple with the finiteness of the state space preventing the increments from being strictly uncorrelated and having mean zero. But if the irreducibility class is quite large in comparison to the observation time n, and the energy landscape is sufficiently homogeneous, there is a diffusive behavior up to the n-th jump. We obtain the largest irreducibility class if we assume $E(s_m) \equiv E$ for every $m \in \mathcal{S}^{(i)}\backslash N^{(i)}$, that is, $E(z^*(m_0,m_1)) - E(s_{m_0}) = 0$ for all distinct $m_0, m_1 \in \mathcal{S}^{(i)}\backslash N^{(i)}$. Then all minima $m \in \mathcal{S}^{(i)}\backslash N^{(i)}$ are in the same irreducibility class of $\widehat{Y}^{(i)}$.

Lemma 3.3.8. *Identify every state $m \in \mathcal{S}^{(i)}$ with a point $m \in \mathbb{R}^d$ for some $d \in \mathbb{N}$. Let $1 \leq i \leq \mathfrak{n}$ be an aggregation level with $E(z^*(m,m')) = E(s_m)$ for every two $m, m' \in \mathcal{S}^{(i)}\backslash N^{(i)}$ and let $\mathcal{R}$ be the irreducibility class of $\widehat{Y}^{(i)}$ comprising the minima $\mathcal{S}^{(i)}\backslash N^{(i)}$. Let the energy landscape be homogeneous enough to ensure that there is some $m_0 \in \mathcal{R}$ from which $\widehat{Y}^{(i)}$ needs at least n steps to leave the set*

$$\left\{m \in \mathcal{R} | \mathbb{E}_m(\widehat{Y}^{(i)}_{1,j} - \widehat{Y}^{(i)}_{0,j}) = 0,\, 1 \leq j \leq d\right\}.$$

Then the asymptotic jump chain shows a diffusive behavior up to time n when starting in m_0, that is,

$$ak \;\leq\; \mathbb{E}_{m_0}\left(\left(\widehat{Y}^{(i)}_{k,j} - \widehat{Y}^{(i)}_{0,j}\right)^2\right) \;\leq\; bk$$

for every $1 \leq j \leq d$, some $a \leq b \in \mathbb{R}_{>0}$ and every $0 \leq k < n$.

Proof: Since

$$\begin{aligned}
&\mathbb{E}_{m_0}\left(\left(\widehat{Y}^{(i)}_{l,j}-\widehat{Y}^{(i)}_{l-1,j}\right)\left(\widehat{Y}^{(i)}_{l',j}-\widehat{Y}^{(i)}_{l'-1,j}\right)\right)\\
&\quad=\mathbb{E}_{m_0}\left(\mathbb{E}_{m_0}\left(\left(\widehat{Y}^{(i)}_{l,j}-\widehat{Y}^{(i)}_{l-1,j}\right)\middle|\widehat{Y}^{(i)}_0,\dots,\widehat{Y}^{(i)}_{l-1}\right)\left(\widehat{Y}^{(i)}_{l',j}-\widehat{Y}^{(i)}_{l'-1,j}\right)\right)\\
&\quad=\mathbb{E}_{m_0}\left(\mathbb{E}_{m_0}\left(\widehat{Y}^{(i)}_{l,j}-\widehat{Y}^{(i)}_{l-1,j}\middle|\widehat{Y}^{(i)}_{l-1}\right)\left(\widehat{Y}^{(i)}_{l',j}-\widehat{Y}^{(i)}_{l'-1,j}\right)\right)\\
&\quad=\mathbb{E}_{m_0}\left(\mathbb{E}_{\widehat{Y}^{(i)}_{l-1}}\left(\widehat{Y}^{(i)}_{1,j}-\widehat{Y}^{(i)}_{0,j}\right)\left(\widehat{Y}^{(i)}_{l',j}-\widehat{Y}^{(i)}_{l'-1,j}\right)\right)\\
&\quad=0
\end{aligned}$$

for every $1\le l'<l\le k$, the assertion is a direct consequence of the previous lemma. □

We want to conclude the diffusion-studies for this case of a constant energy threshold between the basins by the following informal observation: Recall the definition of $\widehat{\nu}(k)$ before Definition 3.2.1 as the time needed by the semi-Markov process associated with $\widehat{Y}^{(i)}$ to undertake k transitions. Under more restrictive conditions, namely

$$\mathbb{E}_{m_0}\left(\left(\widehat{Y}^{(i)}_{l,j}-\widehat{Y}^{(i)}_{l-1,j}\right)\left(\widehat{Y}^{(i)}_{l',j}-\widehat{Y}^{(i)}_{l'-1,j}\right)\middle|\widehat{\nu}(k)\right)=0\quad\mathbb{P}_{m_0}\text{-a.s.}$$

for every $1\le l'<l\le\widehat{\nu}(k)$, it can be shown in the same way that

$$\mathbb{E}_{m_0}\left(\left(\widehat{Y}^{(i)}_{\widehat{\nu}(k),j}-\widehat{Y}^{(i)}_{0,j}\right)^2\right)=\mathbb{E}_{m_0}\left(\widehat{\nu}(k)\cdot\underbrace{\sum_{i=1}^{\widehat{\nu}(k)}\frac{\mathbb{E}_{m_0}\left(\left(\widehat{Y}_{i,j}-\widehat{Y}_{i-1,j}\right)^2\right)}{\widehat{\nu}(k)}}_{=:A}\right)$$

for every $1\le j\le d$ and every $0\le k$ with $\widehat{\nu}(k)<n$ a.s. Due to the fact that $\widehat{Y}^{(i)}$ is irreducible on $\mathcal{R}$ with stationary distribution π^* it holds true that

$$\frac{k}{\widehat{\nu}(k)}\begin{cases}\le\ \frac{(\widehat{\sigma}_1-\widehat{\sigma}_0)+\ldots+(\widehat{\sigma}_{\widehat{\nu}(k)+1}-\widehat{\sigma}_{\widehat{\nu}(k)})}{\widehat{\nu}(k)}\ \to\ \mathbb{E}_{\pi^*}(\widehat{\sigma}_1)\\ \ge\ \frac{(\widehat{\sigma}_1-\widehat{\sigma}_0)+\ldots+(\widehat{\sigma}_{\widehat{\nu}(k)}-\widehat{\sigma}_{\widehat{\nu}(k)-1})}{\widehat{\nu}(k)}\ \to\ \mathbb{E}_{\pi^*}(\widehat{\sigma}_1)\end{cases}\quad\text{a.s.}$$

as $k\to\infty$. Therefore,

$$\frac{\mathbb{E}_{m_0}(\widehat{\nu}(k))}{k}\ \to\ \frac{1}{\mathbb{E}_{\pi^*}(\widehat{\sigma}_1)}$$

as $k\to\infty$. Given that the average A of the squared increments converges, we finally observe for large but not too large k ($\widehat{\nu}(k)<n$) that $\frac{1}{k}\mathbb{E}_{m_0}\left(\left(\widehat{Y}^{(i)}_{\widehat{\nu}(k),j}-\widehat{Y}^{(i)}_{0,j}\right)^2\right)$, $1\le j\le d$, the mean squared displacement of the limiting semi-Markov chain $(\widehat{Y}^{(i)}_{\widehat{\nu}(k)})_{k\ge0}$, is close to the fraction of the average squared increment of $\widehat{Y}$ and the expected sojourn time in the stationary regime. For sufficiently large β_n, the same holds true for $\overline{Y}^{(i)}=(Y^{(i)}_{\nu(k)})_{k\ge0}$ because it is close to $(\widehat{Y}^{(i)}_{\widehat{\nu}(k)})_{k\ge0}$. It is interesting to note that for large β_n the temperature dependence of the mean squared displacement of $\overline{Y}^{(i)}$ is solely given by the temperature dependence of the sojourn times. As the mean squared displacement equals the diffusion coefficient, this is in good accordance to the results of computer simulations of glassy systems with small but positive temperature. As already pointed out in the Introduction, there as well the diffusion coefficient is roughly the fraction of an almost temperature independent constant and the average sojourn time, the last depending highly on the temperature.

Now regard the situation with different barriers between different basins, that is, some level $1\le i\le\mathfrak{n}$ where $\varepsilon>0$ is the smallest order for which there are no reciprocating jumps. Thus, $E(z^*(m_0,m_1))-$

$E(s_{m_0}) \leq \varepsilon$ for all distinct $m_0, m_1 \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and the bound ε in the inequality is reached for some $m_0, m_1 \in \mathcal{S}^{(i)} \backslash N^{(i)}$. Then $\mathcal{S}^{(i)} \backslash N^{(i)}$ disintegrates into several irreducibility classes of $\widehat{Y}^{(i)}$, each of which may consist only of very few metastates of order i. Consequently, due to the smallness of these irreducibility classes, the increments of $\widehat{Y}^{(i)}$ are no longer decorrelated, as we can not assume the process to stay for a long time in a set analog to the one in the above lemma. Thus, for the increments of $Y^{(i)}$ to be decorrelated it is not appropriate to study the limit $\beta_n \to \infty$. Rather β has to be small enough to ensure the probability of a jump with additional energy ε to be likely enough. Or, the other way around, for a given finite β, the bound ε has to be small enough.

Example 3.3.9. Let E be a one dimensional energy landscape, identify the state space with a subset of $\mathbb{R}$, and let m be a metastable state for some level $1 \leq i \leq \mathfrak{n}$ such that for its neighboring metastable states m_1 and m_2 we have

$$E(z^*(m, m_1)) \;=\; E(s_m) \quad \text{and} \quad E(z^*(m, m_2)) \;>\; E(s_m).$$

Then the expected increment is given by

$$\mathbb{E}_m(Y_1 - Y_0) \;=\; (s_m - m)\mathbb{P}_m(Y_1 = s_m) + (z^*(m, m_2) - m)\mathbb{P}_m(Y_1 = z^*(m, m_2))$$

with

$$\mathbb{P}_m(Y_1 = s_m) \;\to\; 1 \quad \text{and} \quad \mathbb{P}_m(Y_1 = z^*(m, m_2)) \;\to\; 0$$

as $\beta \to \infty$. Thus, as $\beta \to \infty$, the increment is not centered. Therefore, fix some large $\beta < \infty$ with

$$\mathbb{P}_m(Y_1 = z^*(m, m_2)) \;=\; \mathbb{P}_m(Y_1 = s_m) e^{-\beta(E(z^*(m,m_2)) - E(s_m) + R(\beta))},$$

where $R(\beta)$ is the remainder and converges to 0. To obtain a centered increment in m, the difference $E(z^*(m, m_2)) - E(s_m)$ has to be small enough to ensure

$$s_m - m \;=\; -(z^*(m, m_2) - m) e^{-\beta(E(z^*(m,m_2)) - E(s_m) + R(\beta))},$$

that is

$$\frac{1}{\beta} \ln\left(\frac{s_m - m}{m - z^*(m, m_2)}\right) \;=\; -\left(E(z^*(m, m_2)) - E(s_m) + R(\beta)\right).$$

When increasing β (that is decreasing the temperature) the additional energy barrier $E(z^*(m, m_2)) - E(s_m)$ has to be smaller and smaller to allow a centered increment.

4. Metabasins

4.1. A Path-Independent Definition...

A path-independent definition of metabasins, announced and to some extent discussed in the Introduction, can now be given on the basis of the previous considerations as the valleys of an appropriate order.

Definition 4.1.1. A finite Markov chain X driven by an energy function E satisfying the assumptions stated at the beginning of Chapter 1 has *metabasins of order* $\varepsilon \geq 0$ if there exists an aggregation level $i < \mathfrak{n} - 1$ such that the following conditions are fulfilled for each $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$:

(MB1) $\sup_{m' \in \mathcal{S}^{(i)} \backslash (N^{(i)} \cup \{m\})} E(z^*(m, m')) - E(s_m) \leq \varepsilon$.

(MB2) For every $m' \in \mathcal{S}^{(i)} \backslash N^{(i)}$ there exist states $m = m_0, m_1, ..., m_n = m' \in \mathcal{S}^{(i)} \backslash N^{(i)}$, $n \in \mathbb{N}$, such that $m_j \not\parallel m_{j+1}$ for every $0 \leq j \leq n - 1$.

In this case, the valleys $(V^{(i)}(m))_{m \in \mathcal{S}^{(i)}}$ are called *metabasins (MB) of order* ε.

The reader should notice that each singleton set $\{s\}$ consisting of a non-assigned state $s \in N^{(i)}$ forms a MB. The conditions (MB1) and (MB2) ensure the good nature of (a) the energy barriers and (b) the spatial arrangement of minima. As already pointed out, this determines which valleys are visited consecutively. Furthermore, we note again that the MB are defined purely deterministic on the basis of the energy function, but nevertheless satisfy all the random dynamical properties, as we will see in the next section. Therefore, unlike with the path-dependent definition, we are able to *explain* those properties and behavior, and not only *observe* it: All Properties 1–5 result from necessary transitions of energy thresholds and a high-dimensional energy landscape.

Before concluding this chapter and the first part of this thesis with a numerical example and the final theorem which summarizes the results obtained so far, we want to compare our approach and definition with the literature. Traditionally, when studying metastability, one is interested in the sequence of metastates visited by the system and their sojourn times. Those metastates can only be defined via metastable sets around metastable states because the system will not strictly stay in the metastable state very long and traverse to another metastable state. It will stay in the basin of attraction for a long time and then traverse to another basin of attraction. Now, in contrast to others (for example Olivieri & Scoppola [49] or Beltrán & Landim [7], already discussed in the Introduction), we are not interested in the *aging effect* which is to find metastates of increasing or at least monotone stability. This perception would neglect all less stable areas, though they behave in the very same way as the more stable ones, only on a smaller time scale. A description of the effective motion in state space can only be given when regarding all the different time scales in the different areas and therefore valleys of different stability. There is a second distinguishing mark by which our approach goes beyond previous works: We do not look at a "blindly" accelerated version $(X_{nT})_{n \geq 0}$ for some appropriate time-scale T. We rather identify the embedded jump process of the aggregated chain, that is, the accelerated aggregated chain, to be in the limit Markovian (along a proper annealing schedule). This AAC and its decelerated version, the AC, are the processes of interest in this work.

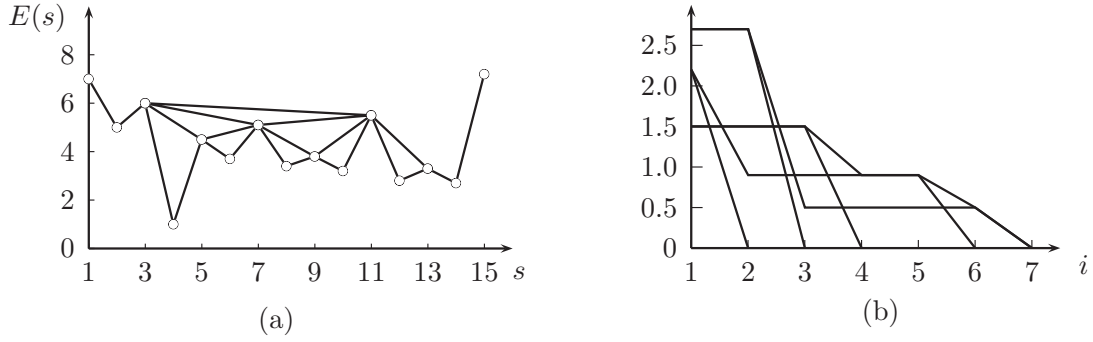

Figure 4.1.: (a) 2-dimensional modification of the energy landscape from Example 1.2.3.
(b) $\sup_{m' \in \mathcal{S}^{(i)} \setminus N^{(i)}} |E(z^*(m, m')) - E(s_m)|$ for the various metastable states in $\mathcal{S}^{(i)} \setminus N^{(i)}$ in dependence of the level $1 \leq i \leq \mathfrak{n}$.

4.2. ...Possessed of the Requested Properties

Properties of MB which can be concluded from the results of the previous chapters are summarized in the next theorem. The reader is reminded of Properties 1–5 stated in the Introduction.

Theorem 4.2.1. *For MB as defined in Definition 4.1.1 we have:*

(1) The transition probabilities for jumps between MB do not depend on the point of entrance as $\beta \to \infty$ and the AAC converges towards a limiting Markov chain along a proper annealing schedule (Property 1).

(2) There are no reciprocating jumps of order ε (Property 2).

(3) The expected residence time in a MB depends on E exponentially via the depth of the MB as $\beta \to \infty$ (Property 3).

(4) Regarding only MB pertaining to local minima, the system is a trap model (Property 4).

(5) For $\varepsilon = 0$, all metastable states $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ are in the same irreducibility class of the asymptotic jump chain which is diffusive up to a specific time, given the homogeneity assumptions of E in Lemma 3.3.8 (compare Property 5).

Proof: (1) follows from Proposition 3.2.3, (2) from Proposition 3.3.5, (3) from Theorem 2.2.2, (4) directly from the definition, and (5) from Lemma 3.3.8. □

We defined MB as valleys of an appropriate order since we know that, first, the AAC exhibits the Markov property in the limit $\beta_n \to \infty$, and, second, the expected residence times behave like in the simulations. To determine the appropriate level, we noticed that the additional energy barrier has to be small to arrange for no reciprocating jumps resp. has to be zero to arrange for the irreducibility of $\widehat{Y}$ (with respect to $\mathcal{S}^{(i)} \setminus N^{(i)}$). That we end up with a trap model is an immediate consequence of this procedure.

Example 4.2.2. We return to Example 1.2.3 given in Chapter 1, but modify the energy landscape by allowing direct transitions between some saddles (see Figure 4.1 (a)) because (MB2) can clearly not be fulfilled in a one-dimensional model. While having no effect on the metastable states $m \in M^{(i)}$, valleys change in the way that, for levels $i \in \{5, 6\}$, the states $\{1, 2, 3\}$ do no longer belong to the valley around state 4 and $\{1, 2\}$ forms its own valley. The energy-differences $\sup_{m' \in \mathcal{S}^{(i)} \setminus N^{(i)}} E(z^*(m, m')) -$

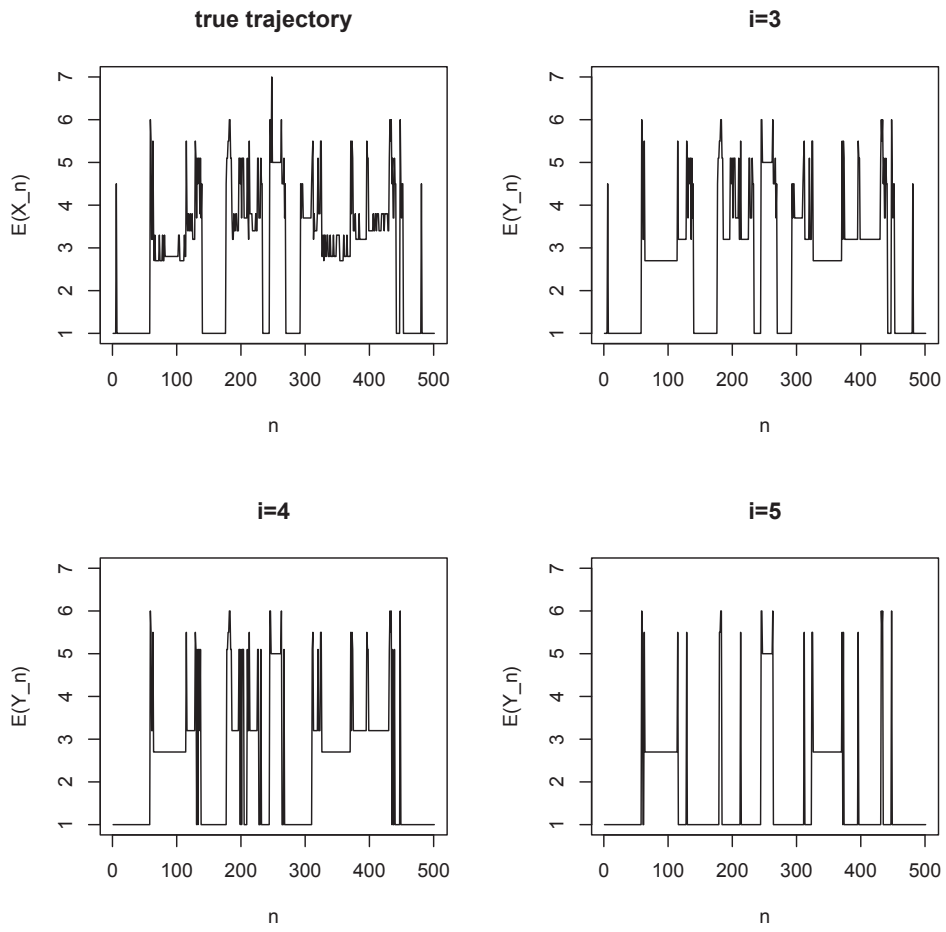

Figure 4.2.: Energies of the true trajectory and of the trajectories of the aggregated chains at levels $i = 3, 4, 5$.

$E(s_m)$ of the various metastable states m at each level $1 \leq i \leq \mathfrak{n} = 7$ are shown in Figure 4.1 (b). The supremum of these energy differences decreases in i, and we obtain MB of order 1 for $i \geq 4$, and of order 0.5 for $i \geq 6$.

To illustrate the behavior, we have run a Metropolis Algorithm on this energy landscape. For initial state $s = 4$ and $\beta = 0.75$, the energies of the trajectories of the original and the aggregated process at levels $i = 3, 4, 5$ are shown in Figure 4.2. The following observations are worth to be pointed out:

Reciprocating Jumps: In the original trajectory there are correlated forward-backward jumps between the times $n = 60$ and $n = 110$, between 320 and 380, and between 410 and 440. They result from reciprocating jumps between (12,13) and (8,10). At aggregation level 3, where those pairs belong to the same level, these reciprocating jumps are no longer observed. When further increasing the level of aggregation, remaining forward-backward jumps vanish as well, for example the unsuccessful trials to escape at the beginning and at the end of the trajectory. But since in this example the energy landscape is not at all high dimensional, re-visits of valleys occur due to the comparatively long simulation time.

Exit Time: In accordance to our results, the sojourn times increase with increasing depth of the valley. Furthermore, those exits really mark epochs after which a return takes not place within

some reasonable time. In the real trajectory the exit times as well depend on the depth of the 1-state valleys, but those "exits" are most often improper, as they are followed by a return within a rather short time (see the trajectory between $n = 320$ and $n = 380$). Therefore, those exit times without aggregation can not describe the time spent in some region of the state space.

Trap Model: Whereas in the real trajectory there is no particular energy threshold which has to be crossed by a transition, this is the fact most obviously in aggregation level $i = 5$. This energy threshold is between 5.75 ± 0.25. As explained, this gives rise to a trap model, which can clearly be identified in the trajectory for $i = 5$.

Long-Range Motion: Since with the cumulated forward-backward jumps there is no real transport within one valley, the motion in state space is well described by the aggregated process.

Part II.

Goodness of Metabasin-Aggregation

5. Comparison with Path-Dependent Definition

In Part I, we constructed a definition of MB which satisfies Properties 1 to 5 of the Introduction. Now we want to study its goodness, here measured by, first, the *consistence with the MB from computer simulations*, and, second, the *similarity of the mixing-, cover-, and hitting times of X and $\overline{Y}$*. This will prove our coarse graining procedure to be in fact rewarding, as it is coarse enough to provide benefits like the nice structure of the aggregated processes, but also subtle enough to map important characteristics of the original system. The consistence with the path-dependent definition, topic of this chapter, will finally clarify what the simulated MB are made of in most trajectories. As the MB trajectory determines the long-range transport, their deterministic description identifies the mechanisms behind.

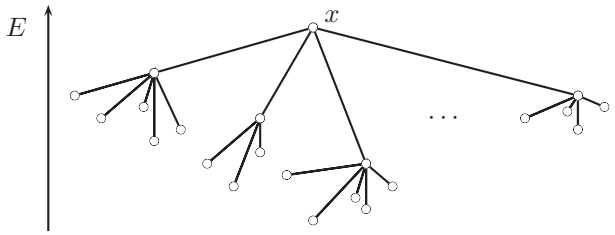

Figure 5.1.: Example of an energy landscape with a tree-like structure and a marked root x.

It should not be surprising that the path-dependent definition of MB by Heuer [29] and stated in the Introduction differs from our path-independent one. For example, the energy landscape in Figure 5.1 has no reasonable path-dependent MB because every transition between two branches of the shown tree must pass through the state x. For a typical trajectory, there will be at most three MB: the states visited before the first occurrence of x, the states visited between the first and the last occurrence of x, and the states visited after the last occurrence of x. The reason for this poor performance is the tree-like structure of the energy landscape or, more generally, the fact that the connectivity between the branches is too small to allow a selfavoiding walk through more than two branches. This results in a small recurrence time for x (compared to the number of states visited in between). However, every branch constitutes a MB when using the path-independent definition for sufficiently small ε, in which case the AAC forms a Markov chain and, in case of the Metropolis algorithm, even a random walk on the graph.

Having thus exemplified that the two definitions of MB do not necessarily coincide, where the path-independent approach applies to a wider class of energy landscapes, we turn to the question about conditions for them to coincide with at least a given probability. As already pointed out, we have to assume a sufficient connectivity between the metastates to ensure the existence of reasonable path-dependent MB. In terms of this connectivity (for a precise definition see Definition 5.1.1) and the parameters β and ε, the main result of this chapter, Theorem 5.2.3 below, provides lower bounds for the probability that both definitions yield the same partition of the state space.

5.1. Connectivity Parameters

The first step towards bounding the probability for the definitions to coincide is to identify and count, for a fixed level $1 \leq i \leq \mathfrak{n}$, each $m \in \mathcal{S}^{(i)}$ and a given β, the states $s \in \mathcal{S}^{(i)}$ for which a transition of $Y^{(i)}$ from m to s is likely. This leads to the announced connectivity parameters, where of course it is to specify what "likely" is.

Definition 5.1.1. Let $\varepsilon \geq 0$ and $1 \leq i \leq \mathfrak{n}$. Define the *connectivity parameters*

$$\begin{aligned}
\eta_1 &= \eta_{1,\varepsilon} := \min_{n \in N^{(i)}, r \in \mathcal{S}^{(i)} : V^{(i)}(r) \cap \mathcal{N}(n) \neq \emptyset} \left| \left\{ s \in \mathcal{N}(n) \backslash V^{(i)}(r) \middle| \, E(s) \leq E(n) + \varepsilon \right\} \right|, \\
\eta_2 &= \eta_{2,\varepsilon} := \min_{m \in \mathcal{S}^{(i)} \backslash N^{(i)}} \left| \left\{ s \in \partial^+ V^{(i)}(m) | E(s) \leq E(s_m) + \varepsilon \right\} \right|, \\
\eta_3 &= \eta_{3,\varepsilon} := \min_{n \in N^{(i)}} \left| \left\{ s \in \mathcal{S}^{(i)} | E(x) \leq E(n) + \varepsilon \text{ for some } x \in V^{(i)}(s) \cap \mathcal{N}(n) \right\} \right|.
\end{aligned} \tag{5.1}$$

η_1 is the minimal number of neighboring sites of a non-assigned state n which do not belong to a particular neighboring valley and whose energy is at most ε plus the energy of n. η_2 is the minimal number of neighboring sites/valleys of a valley $V^{(i)}(m)$ whose energy is at most ε plus the energy of s_m. Finally, η_3 is the minimal number of neighboring valleys of a non-assigned state n which comprise a state with energy of at most ε plus the energy of n. Independent from ε, η_1 and η_3 are quite large in the high dimensional and complex energy landscape of a structural glass, where non-assigned states are neighbored to very many different valleys. For very small ε, η_2 may be 1, but if X has MB of order ε in a high dimensional energy landscape, then η_2 can be assumed to be quite large as well.

For $\varepsilon = 0$, the connectivity parameters η_2 and η_3 count the neighboring sites according to the asymptotic jump chain $\widehat{Y}^{(i)}$, that is, transitions of $\widehat{Y}^{(i)}$ to states counted above have positive probability. That for $\varepsilon > 0$ transitions to states counted above have reasonable large probabilities is content of the next lemma, revealing that the defined parameters in fact measure the connectivity of the MB.

Lemma 5.1.2. *Let $\varepsilon > 0$ and $1 \leq i \leq \mathfrak{n}$ with connectivity parameters defined in* (5.1)*. Writing Y_k for $Y_k^{(i)}$ and $V(m)$ for $V^{(i)}(m)$, $m \in \mathcal{S}^{(i)}$, the following assertions hold true for β sufficiently large:*

(a) If $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $s \in \partial^+ V(m) \cap \{x | E(x) - E(s_m) \leq \varepsilon\}$, or $m \in N^{(i)}$ and $s \in \mathcal{S}^{(i)}$ satisfies $V(s) \cap \{x \in \mathcal{N}(m) | E(x) - E(m) \leq \varepsilon\} \neq \emptyset$, then

$$\mathbb{P}_m(Y_1 = s) \geq e^{-2\beta\varepsilon}.$$

(b) $\mathbb{P}_m(Y_1 \neq s) \geq \eta_1 \, e^{-2\beta\varepsilon}$ for any distinct $m \in N^{(i)}$ and $s \in \mathcal{S}^{(i)}$.

(c) $\mathbb{P}_m(Y_1 \neq s) \geq (\eta_2 - 1) \, e^{-2\beta\varepsilon}$ for any distinct $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $s \in \mathcal{S}^{(i)}$.

We see that, for ε small enough according to β, transitions with an energy barrier of at most ε are still quite likely and thus a jump to a particular valley quite unlikely in the case of high connectivity.

Proof: (a) Choose $\beta_0 > 0$ so large that, for $\beta \geq \beta_0$, $\gamma_\beta \leq \varepsilon$ and $\mathbb{P}_m(Y_1 = s) \geq e^{-2\beta(E(s) - E(s_m))}$ for any $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $s \in \partial^+ V(m)$, the latter being possible by Lemma 3.3.3. Then for any such m and s, we infer $\mathbb{P}_m(Y_1 = s) \geq e^{-2\varepsilon\beta}$ provided that additionally $E(s) \leq E(s_m) + \varepsilon$ holds true. If $m \in N^{(i)}$, then $\mathbb{P}_m(Y_1 = s) \geq e^{-2\varepsilon\beta}$ for any $s \in \mathcal{S}^{(i)}$ such that $E(x) \leq E(s) + \varepsilon$ for some $x \in V(s) \cap \mathcal{N}(m)$, for

$$\mathbb{P}_m(Y_1 = s) \geq \mathbb{P}_m(X_{\sigma_1} = x) = p(m, x) \geq e^{-\beta((E(x) - E(m))^+ + \gamma_\beta)}.$$

(b) Pick again β_0 so large that $\gamma_\beta \le \varepsilon$ for all $\beta \ge \beta_0$. Then, by definition of η_1,

$$\begin{aligned}\mathbb{P}_m(Y_1 \neq s) &\ge \sum_{x\in\mathcal{N}(m),x\notin V(s)} p(m,x)\\ &\ge \sum_{x\in\mathcal{N}(m),x\notin V(s),E(x)\le E(m)+\varepsilon} \exp\big(-\beta((E(x)-E(m))^+ + \gamma_\beta)\big)\\ &\ge \eta_1 \exp(-2\beta\varepsilon).\end{aligned}$$

(c) Fix β_0 so large that $\mathbb{P}_m(Y_1 = x) \ge e^{-2\beta(E(x)-E(s_m))}$ for any $x \in \partial^+V(m)$ and $\beta \ge \beta_0$. In the very same way as in part (b), we get for all $\beta \ge \beta_0$, by definition of η_2,

$$\begin{aligned}\mathbb{P}_m(Y_1 \neq s) &\ge \sum_{x\in\partial^+V(m),x\neq s} \mathbb{P}_m(Y_1 = x)\\ &\ge \sum_{x\in\partial^+V(m),x\neq s,E(x)\le E(s_m)+\varepsilon} \exp\big(-2\beta(E(x)-E(s_m))\big)\\ &\ge (\eta_2 - 1)\exp(-2\beta\varepsilon).\end{aligned}$$

□

5.2. Accordance of Path-Dependent and Path-Independent MB

The above result motivates that in the case of high connectivity the probability to revisit a particular valley within a fixed time T is quite small, or in other words, the probability for the AAC to jump along a selfavoiding path is quite high. This is the main step towards the announced theorem and stated below. The observation time T of course has to be small compared to the cover time of the process.

Lemma 5.2.1. *Let $\varepsilon > 0$ and $1 \le i \le \mathfrak{n}$ with connectivity parameters defined in* (5.1). *Writing Y_k for $Y_k^{(i)}$ and $V(m)$ for $V^{(i)}(m)$, $m \in \mathcal{S}^{(i)}$, define*

$$\tau^{(i)}_{V(m)} := \inf\{k \ge 1 | Y_k = m\}.$$

Then the following assertions hold true for all sufficiently large β:

(a) For any $0 < \delta < 1 - \mathbb{P}_m(Y_2 = m)$ and $1 \le T \le T(m,\beta) + 1$,

$$\mathbb{P}_m\left(\tau^{(i)}_{V(m)} > T\right) \ge \delta,$$

where

$$T(m,\beta) := \frac{\ln\delta}{\ln\big(\min_{m'\neq m}\mathbb{P}_{m'}(Y_1 \neq m)(1 - \mathbb{1}_{\{m'\notin N^{(i)}\}}\delta(m',\beta))\big)}$$

and

$$\delta(m',\beta) := \max_{x\in V(m')} \sum_{z\in\partial^+V(m')} \tilde{\varepsilon}(x,m',z,\beta).$$

In particular, if $\delta \le \big((\eta_1 \wedge (\eta_2 - 2))e^{-2\beta\varepsilon}\big)^T$ for some $T > 0$, then $T(m,\beta) \ge T$.

(b) For each $k \geq 1$ and $m_0 \in \mathcal{S}^{(i)}$,

$$\sum_{m_1,\ldots,m_k} \prod_{j=0}^{k-1} \mathbb{P}_{m_j}(Y_1 = m_{j+1}) \;\geq\; [\eta_2 \wedge \eta_3]_k \, e^{-2k\varepsilon\beta}$$

where summation ranges over all pairwise distinct $m_1,\ldots,m_k \in \mathcal{S}^{(i)}\backslash\{m_0\}$ and for $N \in \mathbb{N}$ we write $[N]_k := N(N-1)\cdot\ldots\cdot(N-k+1)$.

It should be noticed that $\mathbb{P}_m(\tau^{(i)}_{V(m)} > 1) = 1$ (the AAC never stays put) and

$$\mathbb{P}_m(\tau^{(i)}_{V(m)} > T) \;\leq\; \mathbb{P}_m(\tau^{(i)}_{V(m)} > 2) \;=\; 1 - \mathbb{P}_m(Y_2^{(i)} = m)$$

for every $T \geq 2$ with equality holding only if $T = 2$. We thus see that $\mathbb{P}_m(\tau^{(i)}_{V(m)} > T) \geq \delta$ entails $\delta < 1 - \mathbb{P}_m(Y_2 = m)$, the latter being typically large. Furthermore, the bound on the number of self-avoiding path of length k is very crude and can be improved when knowing more about the spatial arrangement of the metastable states.

Proof: (a) Recall from the first part of the proof of Proposition 3.2.3 that

$$\begin{aligned}
\mathbb{P}_m(Y_{n+1} \neq z | Y_n = y, X_{\sigma_n} = x) \;&\geq\; \mathbb{P}_y(Y_1 \neq z)\left(1 - \mathbb{1}_{\{y \notin N^{(i)}\}} \sum_{r \in \partial^+ V(y)} \mathbb{P}_x(\tau_r < \tau_y)\right) \\
&\geq\; \mathbb{P}_y(Y_1 \neq z)\left(1 - \mathbb{1}_{\{y \notin N^{(i)}\}} \sum_{r \in \partial^+ V(y)} \tilde{\varepsilon}(x,y,r,\beta)\right) \\
&\geq\; \mathbb{P}_y(Y_1 \neq z)\Big(1 - \mathbb{1}_{\{y \notin N^{(i)}\}}\, \delta(y,\beta)\Big)
\end{aligned}$$

holds true for all $y, z \in \mathcal{S}^{(i)}$, $x \in V(y)$ and $\beta > 0$. Putting $\mathfrak{m}(x) := m'$ if $x \in V(m')$ for $m' \in \mathcal{S}^{(i)}$, we obtain for $T \geq 2$, by using the above bound repeatedly,

$$\begin{aligned}
\mathbb{P}_m\Big(\tau^{(i)}_{V(m)} > T\Big) \;&=\; \mathbb{P}_m(Y_1 \neq m,\ldots,Y_T \neq m) \\
&=\; \sum_{x_1,\ldots,x_{T-1} \notin V^{(i)}(m)} \mathbb{P}_m(X_{\sigma_1} = x_1) \prod_{k=1}^{T-2} \mathbb{P}_m\big(X_{\sigma_{k+1}} = x_{k+1} | X_{\sigma_k} = x_k, Y_k = \mathfrak{m}(x_k)\big) \\
&\qquad\qquad \times \mathbb{P}_m(Y_T \neq m | X_{\sigma_{T-1}} = x_{T-1}, Y_{T-1} = \mathfrak{m}(x_{T-1})) \\
&\geq\; \sum_{x_1,\ldots,x_{T-1} \notin V^{(i)}(m)} \mathbb{P}_m(X_{\sigma_1} = x_1) \prod_{k=1}^{T-2} \mathbb{P}_m\big(X_{\sigma_{k+1}} = x_{k+1} | X_{\sigma_k} = x_k, Y_k = \mathfrak{m}(x_k)\big) \\
&\qquad\qquad \times \left(\min_{m' \neq m} \Big(\mathbb{P}_{m'}(Y_1 \neq m)(1 - \mathbb{1}_{\{m' \notin N^{(i)}\}}\, \delta(m',\beta))\Big)\right) \\
&\;\;\vdots \\
&\geq\; \min_{m' \neq m} \Big(\mathbb{P}_{m'}(Y_1 \neq m)(1 - \mathbb{1}_{\{m' \notin N^{(i)}\}}\, \delta(m',\beta))\Big)^{T-1}.
\end{aligned}$$

But this establishes the asserted inequality when finally observing that the last expression is $\geq \delta$ if and only if $T \leq T(m,\beta) + 1$.

Having just said that $T(m,\beta) \geq T$ holds if and only if

$$\min_{m' \neq m} \Big(\mathbb{P}_{m'}(Y_1 \neq m)(1 - \mathbb{1}_{\{m' \notin N^{(i)}\}} \delta(m',\beta)) \Big)^T \;\geq\; \delta,$$

it suffices to note that, as $\beta \to \infty$, $\delta(m',\beta) \to 0$ holds true if $m' \in \mathcal{S}^{(i)} \backslash N^{(i)}$, giving

$$1 - \delta(m',\beta) \;\geq\; \frac{\eta_2 - 2}{\eta_2 - 1}$$

for sufficiently large β. Together with Lemma 5.1.2 (b), this further yields

$$\min_{m' \neq m} \Big(\mathbb{P}_{m'}(Y_1 \neq m)(1 - \mathbb{1}_{\{m' \notin N^{(i)}\}} \delta(m',\beta)) \Big)^T \;\geq\; \Big((\eta_1 \wedge (\eta_2 - 2)) e^{-2\beta\varepsilon} \Big)^T$$

and then the assertion.

(b) Here it suffices to notice that, by Lemma 5.1.2 (a), $[\eta_2 \wedge \eta_3]_k$ forms a lower bound for the number of self-avoiding paths $(m_0,...,m_k)$ such that $\mathbb{P}_{m_j}(Y_1 = m_{j+1}) \geq e^{-2\beta\varepsilon}$ for each $j = 0,...,k-1$. □

For $\varepsilon = 0$, we have a similar result for the asymptotic jump chain.

Lemma 5.2.2. *Let $\varepsilon = 0$ and $1 \leq i \leq \mathfrak{n}$ with connectivity parameters defined in* (5.1). *Writing $\widehat{Y}_k$ for $\widehat{Y}_k^{(i)}$ and $V(m)$ for $V^{(i)}(m)$, $m \in \mathcal{S}^{(i)}$, define*

$$\widehat{\tau}^{(i)} \;:=\; \inf\{k \geq 1 | \widehat{Y}_k = m\}.$$

Then the following assertions hold true:

(a) For any $0 < \delta < 1 - \mathbb{P}_m(\widehat{Y}_2 = m)$ and $1 \leq T \leq T(m) + 1$,

$$\mathbb{P}_m \Big(\widehat{\tau}^{(i)} > T \Big) \;\geq\; \delta,$$

where

$$T(m) \;:=\; T(m,\infty) \;:=\; \frac{\ln \delta}{\ln\big(\min_{m' \neq m} \mathbb{P}_{m'}(\widehat{Y}_1 \neq m) \big)}.$$

In particular, if $\delta \leq \big((\eta_1 \wedge (\eta_2 - 2)) \min_{r,s:\widehat{p}(r,s)>0} \widehat{p}(r,s) \big)^T$ for some $T > 0$, then $T(m) \geq T$.

(b) For each $k \geq 1$ and $m_0 \in \mathcal{S}^{(i)}$,

$$\sum_{m_1,...,m_k} \prod_{j=0}^{k-1} \mathbb{P}_{m_j}(\widehat{Y}_1 = m_{j+1}) \;\geq\; [\eta_2 \wedge \eta_3]_k \min_{r,s:\widehat{p}(r,s)>0} \widehat{p}(r,s)^k$$

where summation ranges over all pairwise distinct $m_1,...,m_k \in \mathcal{S}^{(i)} \backslash \{m_0\}$.

Proof: (a) In perfect analogy to the previous proof we obtain

$$\mathbb{P}_m\Big(\widehat{\tau}^{(i)} > T \Big) \;=\; \mathbb{P}_m(\widehat{Y}_1 \neq m,..., \widehat{Y}_T \neq m) \;\geq\; \min_{m' \neq m} \Big(\mathbb{P}_{m'}(\widehat{Y}_1 \neq m) \Big)^{T-1},$$

which is bounded below by δ if and only if $T \leq T(m) + 1$. Furthermore,

$$\min_{m' \neq m} \Big(\mathbb{P}_{m'}(\widehat{Y}_1 \neq m) \Big)^T \;\geq\; \Big((\eta_1 \wedge (\eta_2 - 1)) \min_{r,s:\widehat{p}(r,s)>0} \widehat{p}(r,s) \Big)^T,$$

which yields the assertions.

(b) Again, it suffices to notice that $[\eta_2 \wedge \eta_3]_k$ forms a lower bound for the number of self-avoiding paths $(m_0,...,m_k)$ such that $\widehat{p}(m_j, m_{j+1}) > 0$ for each $j = 0,...,k-1$. □

We proceed with the announced result about the relation between path-dependent and path-independent MB. To this end we fix $T = \sigma_K$ for some $K \in \mathbb{N}$. Let $\mathcal{V}_k$ for $k = 1,...,\upsilon$ denote the random MB obtained from $X_0,...,X_T$ as defined in the Introduction. For $x \in \mathcal{S}$, we further let $\mathcal{V}(x)$ denote the MB $\mathcal{V}_k$ containing x and put $\mathcal{V}(x) := \emptyset$ if no such MB exists which is the case if and only if $x \notin \{X_0,...,X_T\}$.

Theorem 5.2.3. *Let $\varepsilon > 0$ and suppose that X has MB of order ε at level i with connectivity parameters defined in (5.1). Fix $K \in \mathbb{N}$, $T = \sigma_K$ and $0 < \delta \leq \big((\eta_1 \wedge (\eta_2 - 1) - 1)e^{-2\beta\varepsilon}\big)^K$. Then, for each $0 \leq k < K$ and $m_0 \in \mathcal{S}^{(i)}$, there exists $\beta_0 > 0$ such that for all $\beta \geq \beta_0$*

(a) with $V_<^{(i)}(s) := \{s\}$ if $s \in N^{(i)}$

$$\mathbb{P}_{m_0}\Big(V_<^{(i)}(Y_k) \subseteq \mathcal{V}(Y_k)\Big) \;\geq\; 1 - \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m,\beta) \left(\max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} |V_<(m)| + 2 \right),$$

and the right-hand side goes to 1 as $\beta \to \infty$.

(b) $\mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subseteq V^{(i)}(Y_j),\, 0 \leq j < k) \geq 1 - k(\max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m,\beta) + (1 - \delta))$.

(c) If $\eta_2 \wedge \eta_3 > K - 1$, then

$$\mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subseteq V^{(i)}(Y_j), 0 \leq j \leq K-1) \;\geq\; [\eta_2 \wedge \eta_3]_K \left(1 - \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m,\beta)\right)^{K-1} e^{-2K\varepsilon\beta}.$$

We conclude that the inner part of the path-independent MB is a.a.s. part of the path-dependent one and the probability of the path-dependent MB to be part of the path-independent ones is high whenever the connectivity is high and ε is small according to β. Namely, for the occurring bounds to be significant, two requirements must be met. First, K must be small compared to the cover time of the AAC and ε must be small compared to β_0 to ensure $\exp(-2\beta\varepsilon) \gg 0$. Second, the connectivity must be high to ensure $1 - \delta \ll 1$ and $[\eta_2 \wedge \eta_3]_K e^{-2K\varepsilon\beta} \gg 0$.

Typically, the inclusions in parts (b) and (c) are strict because of high energy states within a valley that will probably be missed during one simulation run and therefore not belong to any path-dependent MB. On the other hand, since our approach strives to cover the state space as completely as possible by valleys, the latter comprise such high energy states whenever they are assignable in the sense described in Chapter 1.

Further note that there is obviously an analog statement for $\varepsilon = 0$ where only $e^{-2\beta\varepsilon}$ has to be replaced by $\min_{r,s:\widehat{p}(r,s)>0} \widehat{p}(r,s)$ and $e^{-2K\beta\varepsilon}$ by $\big(\min_{r,s:\widehat{p}(r,s)>0} \widehat{p}(r,s)\big)^K$. We omit stating it here.

Proof: With i being fixed, let us write as earlier $V(m)$ for $V^{(i)}(m)$, and also $V_<(m)$ for $V_<^{(i)}(m)$.

(a) For a given $0 \leq k < K$, define

$$\begin{aligned} A_k &:= \{\sigma_k \leq \tau_{Y_k} < \sigma_{k+1}\}, \\ B_k &:= \{ \text{ for every } x \in V_<(m) \text{ there exists } \tau_{Y_k} \leq l_x < \sigma_k \text{ such that } X_l = x\}, \\ C_k &:= \left\{ X_l = Y_k \text{ for some } \max_{x \in V_<(Y_k)} \tau_x \leq l < \sigma_k \right\}, \end{aligned}$$

as the events that, first, during a visit of a valley its ground state is visited, second, every element of $V_<(m)$ is visited after visiting the ground state and before leaving the valley, and third, the ground state is again visited after the whole inner part $V_<(m)$ is visited and before the valley is left. With $\delta(m,\beta)$ as defined in Lemma 5.2.1 and using

$$\mathbb{P}_x(\sigma_1 < \tau_m) \;\leq\; \sum_{y \in \partial^+ V(m)} \tilde{\varepsilon}(x,m,y,\beta) \;\leq\; \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m,\beta) \;=:\; \delta_{\max} \tag{5.2}$$

for $x \in V(m)$, $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$, we obtain

$$\begin{aligned}
&\mathbb{P}_{m_0}\Big(V_<^{(i)}(Y_k) \subseteq \mathcal{V}(Y_k)\Big) \\
&\quad \geq \ \mathbb{P}_{m_0}(A_k \cap B_k \cap C_k) \\
&\quad = \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(\{X_{\sigma_k} = r\} \cap A_k \cap B_k \cap C_k) \\
&\quad = \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(X_{\sigma_k} = r) \mathbb{P}_{m_0}(A_k | X_{\sigma_k} = r) \mathbb{P}_{m_0}(B_k \cap C_k | \{X_{\sigma_k} = r\} \cap A_k) \\
&\quad = \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(X_{\sigma_k} = r) \mathbb{P}_r(\tau_m < \sigma_1) \\
&\qquad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_<(m),\, X_l = m \text{ for some } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1) \\
&\quad \geq \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(X_{\sigma_k} = r)(1 - \delta_{\max}) \\
&\qquad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_<(m),\, X_l = m \text{ for some } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1) \\
&\quad = \ (1 - \delta_{\max}) \sum_{m \in \mathcal{S}^{(i)}} \mathbb{P}_{m_0}(Y_k = m) \qquad (5.3) \\
&\qquad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_<(m),\, X_l = m \text{ for some } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1).
\end{aligned}$$

Thus, in order to show that with high probability a path-dependent MB comprises the inner part of a valley, we show that with high probability, when starting in its minimum, the whole inner part will be visited and the process will return to the minimum once more before the valley is left. This is trivial if $m \in N^{(i)}$ and thus $V_<^{(i)}(m) = \{m\}$, for then

$$\mathbb{P}_m\left(\tau_x < \sigma_1 \text{ for every } x \in V_<(m),\, X_l = m \text{ for some } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1\right) \ = \ 1.$$

More needs to be done if $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$, where

$$\begin{aligned}
&\mathbb{P}_m\left(\tau_x < \sigma_1 \text{ for every } x \in V_<(m),\, X_l = m \text{ for some } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1\right) \\
&\qquad \geq \ 1 - \mathbb{P}_m\left(\tau_x > \sigma_1 \text{ for some } x \in V_<(m)\right) - \mathbb{P}_m\left(X_l \neq m \text{ for each } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1\right).
\end{aligned}$$

The second probability in the preceding line can further be bounded with the help of (5.2), viz.

$$\begin{aligned}
&\mathbb{P}_m(X_l \neq m \text{ for each } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1) \\
&\quad = \sum_{y \in V_<(m)} \mathbb{P}_m(\max_{x \in V_<(m)} \tau_x = \tau_y,\ X_l \neq m \text{ for every } \max_{x \in V_<(m)} \tau_x \leq l < \sigma_1) \\
&\quad \leq \sum_{y \in V_<(m)} \mathbb{P}_m(\max_{x \in V_<(m)} \tau_x = \tau_y) \mathbb{P}_y(\tau_m > \sigma_1) \\
&\quad \leq \ \delta_{\max}.
\end{aligned}$$

For the first probability, we obtain with the help of Theorem 2.1.1

$$\begin{aligned}\mathbb{P}_m\left(\tau_x > \sigma_1 \text{ for some } x \in V_<(m)\right) &\leq \sum_{x\in V_<(m)} \mathbb{P}_m(\sigma_1 < \tau_x)\\ &\leq \sum_{x\in V_<(m)} \sum_{y\in\partial^+V(m)} \mathbb{P}_m(\tau_y < \tau_x)\\ &\leq \sum_{x\in V_<(m)} \sum_{y\in\partial^+V(m)} \varepsilon(m,x,y,\beta), \end{aligned} \tag{5.4}$$

because $E(z^*(m,y)) > E(z^*(m,x))$ for $x \in V_<(m)$ and $y \in \partial^+V(m)$. The latter can be seen as follows: It has been shown in the proof of Theorem 2.1.2 that $E(z^*(x,m)) < E(z^*(x,y))$. Hence,

$$E(z^*(x,m)) \;<\; E(z^*(x,y)) \;\leq\; E(z^*(x,m)) \vee E(z^*(y,m)) \;=\; E(z^*(y,m)),$$

as asserted. Next, we infer

$$E(z^*(x,y)) \;\leq\; E(z^*(x,m)) \vee E(z^*(m,y)) \;\leq\; E(z^*(x,m)) \vee E(z^*(x,y)) \;=\; E(z^*(x,y)),$$

thus $E(z^*(x,y)) = E(z^*(m,y))$. Recalling the definition of $\varepsilon(m,x,y,\beta)$, the last equality implies $\varepsilon(m,x,y,\beta) = \varepsilon(x,m,y,\beta)$, which will now be used to further bound the expression in (5.4), namely

$$\begin{aligned}\sum_{x\in V_<(m)} \sum_{y\in\partial^+V(m)} \varepsilon(m,x,y,\beta) &= \sum_{x\in V_<(m)} \sum_{y\in\partial^+V(m)} \varepsilon(x,m,y,\beta)\\ &= \sum_{x\in V_<(m)} \sum_{y\in\partial^+V(m)} \tilde{\varepsilon}(x,m,y,\beta)\\ &\leq |V_<(m)|\,\delta_{\max}\\ &\leq \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} |V_<(m)|\,\delta_{\max}.\end{aligned}$$

Together with (5.3), this yields, as asserted,

$$\begin{aligned}&\mathbb{P}_{m_0}\Big(V_<^{(i)}(Y_k) \subseteq \mathcal{V}(Y_k)\Big)\\ &\quad\geq\; (1-\delta_{\max}) \sum_{m\in\mathcal{S}^{(i)}} \mathbb{P}_{m_0}(Y_k = m)\\ &\qquad\qquad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_<(m),\, X_l = m \text{ for some } \max_{x\in V_<(m)} \tau_x \leq l < \sigma_1)\\ &\quad\geq\; (1-\delta_{\max}) \left(1 - \left(\max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} |V_<(m)| + 1\right)\delta_{\max}\right)\\ &\quad\geq\; 1 - \left(\max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} |V_<(m)| + 2\right)\delta_{\max}.\end{aligned}$$

(b) According to Lemma 5.2.1, choose $\beta_0 > 0$ such that

$$\max_{m\in\mathcal{S}^{(i)}} \mathbb{P}_m\left(\tau_{V(m)}^{(i)} \leq K\right) \;\leq\; 1-\delta \tag{5.5}$$

for each $\beta \geq \beta_0$. By using (5.5) and (5.2), we now infer

$$\begin{aligned}
&\mathbb{P}_{m_0}(Y_l = Y_k \text{ for some } k+1 \leq l \leq K) \\
&\quad = \sum_{s\in\mathcal{S}} \mathbb{P}_{m_0}(Y_l = Y_k \text{ for some } k+1 \leq l \leq K, X_{\sigma_k} = s) \\
&\quad \leq \sum_{s\in\mathcal{S}} \mathbb{P}_{m_0}(X_{\sigma_k} = s)\Big(\mathbb{P}_s(\tau^{(i)}_{V(Y_0)} \leq K-k,\, X_j = \mathfrak{m}(s) \text{ for some } 0 \leq j < \sigma_1) \\
&\qquad\qquad + \mathbb{1}_{\{s\notin N^{(i)}\}}\, \mathbb{P}_s(\tau^{(i)}_{V(Y_0)} \leq K-k, X_j \neq \mathfrak{m}(s) \text{ for all } 0 \leq j < \sigma_1)\Big) \\
&\quad \leq \sum_{s\in\mathcal{S}} \mathbb{P}_{m_0}(X_{\sigma_k} = s)\, \mathbb{P}_{\mathfrak{m}(s)}(\tau^{(i)}_{V(Y_0)} \leq K-k) + \sum_{s\notin N^{(i)}} \mathbb{P}_{m_0}(X_{\sigma_k} = s)\, \mathbb{P}_s(\sigma_1 < \tau_{\mathfrak{m}(s)}) \\
&\quad \leq 1 - \delta + \delta_{\max},
\end{aligned}$$

and finally

$$\begin{aligned}
\mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subset V^{(i)}(Y_j),\, 0 \leq j < k) &\geq \mathbb{P}_{m_0}\left(\bigcap_{j=0}^{k-1} \{Y_l \neq Y_j, j+1 \leq l \leq K\}\right) \\
&\geq 1 - \sum_{j=0}^{k-1} \mathbb{P}_{m_0}(Y_l = Y_j \text{ for some } j+1 \leq l \leq K) \\
&\geq 1 - k(\delta_{\max} + (1-\delta)).
\end{aligned}$$

(c) In the following calculation, let $r_0 = m_0$, $\sum_{m_j}$ range over all K-vectors $(m_1,...,m_K)$ with pairwise distinct components in $\mathcal{S}^{(i)}\backslash\{m_0\}$ and, for each $k < K$, let $\sum_{r_1,...,r_k}$ range over all k-vectors $(r_1,...,r_k)$ such that $r_j \in V(m_j)$ for each $j = 1,...,k$. As in part (b), use (5.2) repeatedly to infer

$$\begin{aligned}
&\mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subset V^{(i)}(Y_j),\, 0 \leq j \leq K-1) \\
&\qquad \geq \sum_{m_j} \sum_{r_1,...,r_{K-1}} \mathbb{P}_{m_0}\left(\bigcap_{j=0}^{K-1} \{Y_j = m_j, X_{\sigma_j} = r_j,\, \tau_{m_j} < \sigma_{j+1}\} \cap \{Y_K = m_K\}\right) \\
&\qquad = \sum_{m_j} \sum_{r_1,...,r_{K-1}} \prod_{j=0}^{K-2} \mathbb{P}_{r_j}\left(Y_0 = m_j, X_{\sigma_1} = r_{j+1},\, \tau_{m_j} < \sigma_1\right) \\
&\qquad\qquad\qquad \times \mathbb{P}_{r_{K-1}}\left(\tau_{m_{K-1}} < \sigma_1\right)\, \mathbb{P}_{m_{K-1}}(Y_1 = m_K) \\
&\qquad \geq \sum_{m_j} \sum_{r_1,...,r_{K-1}} \prod_{j=0}^{K-2} \mathbb{P}_{r_j}\left(Y_0 = m_j, X_{\sigma_1} = r_{j+1},\, \tau_{m_j} < \sigma_1\right) \\
&\qquad\qquad\qquad \times (1-\delta_{\max})\mathbb{P}_{m_{K-1}}(Y_1 = m_K) \\
&\qquad = (1-\delta_{\max}) \sum_{m_j} \sum_{r_1,...,r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j}\left(Y_0 = m_j, X_{\sigma_1} = r_{j+1},\, \tau_{m_j} < \sigma_1\right) \\
&\qquad\qquad \times \mathbb{P}_{r_{K-2}}\left(Y_0 = m_{K-2}, Y_1 = m_{K-1},\, \tau_{m_{K-2}} < \sigma_1\right)\, \mathbb{P}_{m_{K-1}}(Y_1 = m_K) \\
&\qquad = (1-\delta_{\max}) \sum_{m_j} \sum_{r_1,...,r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j}\left(Y_0 = m_j, X_{\sigma_1} = r_{j+1},\, \tau_{m_j} < \sigma_1\right) \\
&\qquad\qquad\qquad \times \mathbb{P}_{r_{K-2}}\left(Y_1 = m_{K-1},\, \tau_{m_{K-2}} < \sigma_1\right)\, \mathbb{P}_{m_{K-1}}(Y_1 = m_K)
\end{aligned}$$

$$
\begin{aligned}
&\geq (1-\delta_{\max}) \sum_{m_j} \sum_{r_1,\ldots,r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j}\left(Y_0 = m_j, X_{\sigma_1} = r_{j+1},\, \tau_{m_j} < \sigma_1\right) \\
&\qquad\qquad \times \mathbb{P}_{m_{K-2}}\left(Y_1 = m_{K-1}\right) \mathbb{P}_{r_{K-2}}(\tau_{m_{K-2}} < \sigma_1)\, \mathbb{P}_{m_{K-1}}(Y_1 = m_K) \\
&\geq (1-\delta_{\max})^2 \sum_{m_j} \sum_{r_1,\ldots,r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j}\left(Y_0 = m_j, X_{\sigma_1} = r_{j+1},\, \tau_{m_j} < \sigma_1\right) \\
&\qquad\qquad \times \mathbb{P}_{m_{K-2}}\left(Y_1 = m_{K-1}\right) \mathbb{P}_{m_{K-1}}(Y_1 = m_K) \\
&\;\;\vdots \\
&\geq (1-\delta_{\max})^{K-1} \sum_{m_j} \prod_{j=0}^{K-1} \mathbb{P}_{m_j}(Y_1 = m_{j+1}) \\
&\geq (1-\delta_{\max})^{K-1} \left[\eta_2 \wedge \eta_3\right]_K e^{-2K\varepsilon\beta},
\end{aligned}
$$

the last line following from Lemma 5.2.1. □

6. Comparison of Mixing Times

In this chapter, we want to compare the *mixing times* of the original process and its aggregated versions to show in particular that the relaxation (i.e. convergence to the equilibrium) of X can be mapped by the relaxation of $\overline{Y}$. For a Markov chain M with stationary distribution π the mixing time for $\varepsilon \geq 0$ is defined as

$$\tau_{\text{mix}}^{M}(\varepsilon) \;:=\; \inf\{n \geq 0 | \sup_{x} d_{TV}(\mathbb{P}_x^{M_k}, \pi) \;\leq\; \varepsilon \text{ for every } k \geq n\},$$

where $d_{TV}(\cdot\,,\cdot)$ denotes the *total variation (distance)* of two measures. In computer simulations, the mixing time is an important characteristic of a Markov chain, as it allows sampling from a distribution arbitrary close to the stationary distribution. This is used in many applications, for example in [63], where the author (in joint work with Alsmeyer and the theoretical chemists Heuer and Rubner) studies the number of feasible states in the hard core model with a fixed number of particles using the so called Widom-method.

For finite Markov chains, there are several methods to specify the mixing time, one of the most popular ones working with the eigenvalues of the transition matrix. Analytically, with these eigenvalues the exact rate of convergence to the stationary distribution can be calculated, but unfortunately there is no probabilistic explanation. We draw on various well known facts from the theory of spectral analysis of finite Markov chains and refer to the excellent textbook by Levin, Peres & Wilmer [39] for the general picture. The basic definitions are:

Definition 6.0.4. Let $\mathbf{P}$ be a finite stochastic matrix with stationary distribution π. If π is also reversible for $\mathbf{P}$, we call $(\mathbf{P}, \pi)$ a *reversible pair*. If $p(x,x) \geq \frac{1}{2}$ for every $x \in \mathcal{S}$, we call $\mathbf{P}$ *lazy*. For every reversible pair $(\mathbf{P}, \pi)$ we define

(a) the *spectral gap* as $\gamma := 1 - \lambda_2$, where

$$\lambda_2 \;:=\; \max\{\lambda \mid \lambda \text{ is eigenvalue of } \mathbf{P} \text{ and } \lambda < 1\},$$

(b) the *absolute spectral gap* as $\gamma_\star := 1 - \lambda_\star$, where

$$\lambda_\star \;:=\; \max\{|\lambda| \mid \lambda \text{ is eigenvalue of } \mathbf{P} \text{ and } \lambda < 1\},$$

(c) the *relaxation time* as $t_{rel} := \gamma_\star^{-1}$,

(d) the *Dirichlet form* for any real function f as

$$\mathcal{E}(f) \;:=\; \frac{1}{2} \sum_{x \neq y} \left(f(x) - f(y)\right)^2 \pi(x) p(x,y).$$

As we use some results various times, a *toolbox* on how these parameters are related with the mixing time of the Markov chain, most notably the variational characterization, is given in the Appendix, Section A.2.

6.1. Mixing Time of the Original Chain and some Functionals

6.1.1. The Original Chain

We start the investigation of the different mixing times by calculating the (absolute) spectral gap γ of the original process, depending of course on β. Since the holding probabilities of X are at least $\frac{1}{2}$, the spectral gap and the absolute spectral gap of X coincide.

Theorem 6.1.1. *Let $\Delta := E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) - E(m^{(\mathfrak{n}-1)})$ be the depth of the second deepest valley and γ the spectral gap of X. Then*

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln(\gamma) \;=\; -\Delta.$$

This is known for continuous-time processes on compact connected manifolds [31, Theorem 1.14] and smooth bounded domains in $\mathbb{R}^n$, $n \in \mathbb{N}$, [44, Theorem 2], as well as for a specific discrete-time Metropolis-process on a finite state space, where the proposal chain is symmetric and independent of β [30, Theorem 2.1]. We use the same technique as Holley & Stroock in [30] to extend this result to the perturbed Metropolis-algorithm of our chain and later also for several functionals of X. That the exponential rate of the spectral gap is given by the depth of the second deepest valley can be explained heuristically as follows: As β tends to infinity, the stationary distribution tends to $\delta_{m^{(\mathfrak{n})}}$. Thus, the crucial time for mixing in the worst case is given by the time needed to enter $V^{(\mathfrak{n}-1)}(m^{(\mathfrak{n})})$, which again is given by the time needed to leave $V^{(\mathfrak{n}-1)}(m^{(\mathfrak{n}-1)})$ (and not to leave $V^{(\mathfrak{n}-1)}(m^{(\mathfrak{n})})$).

Proof: We want to derive a lower bound of γ with the variational characterization by the Dirichlet form. Therefore, let $f : \mathcal{S} \to \mathbb{R}$ be some function with $Var_\pi(f) \neq 0$. This variance can be bounded by

$$\begin{aligned} Var_\pi(f) \;&=\; \sum_s \left(f(s) - \sum_r f(r)\pi(r) \right)^2 \pi(s) \\ &=\; \sum_s \left(\sum_r (f(s) - f(r))\pi(r) \right)^2 \pi(s) \\ &\leq\; \sum_s \sum_r \left(f(s) - f(r)\right)^2 \pi(r)\pi(s), \end{aligned} \tag{6.1}$$

using Jensen's Inequality. Now let $\gamma = (\gamma_0,...,\gamma_k) \in \Gamma^*(r,s)$ with $k := k(r,s) := |\gamma|$ be a minimal path from r to s. Note that, using again Jensen's Inequality,

$$\left(\sum_{i=1}^n a_i \right)^2 \;\leq\; n \sum_{i=1}^n a_i^2$$

for every sequence $(a_1,...,a_n) \in \mathbb{R}^n$, and therefore

$$\begin{aligned} (f(s) - f(r))^2 \pi(r)\pi(s) \;&=\; \left(\sum_{i=1}^k f(\gamma_i) - f(\gamma_{i-1}) \right)^2 \pi(r)\pi(s) \\ &\leq\; k \sum_{i=1}^k \left(f(\gamma_i) - f(\gamma_{i-1})\right)^2 \pi(r)\pi(s) \\ &\leq\; |\mathcal{S}| \sum_{i=1}^k \left(f(\gamma_i) - f(\gamma_{i-1})\right)^2 \left(\pi(\gamma_i) \wedge \pi(\gamma_{i-1})\right) \frac{\pi(r)\pi(s)}{\pi(\gamma_i) \wedge \pi(\gamma_{i-1})}. \end{aligned} \tag{6.2}$$

Since by Lemma 1.1.2,

$$\begin{aligned}\pi(\gamma_i) \wedge \pi(\gamma_{i-1}) \;&\geq\; \min_{1\leq i\leq k} \pi(\gamma_i) \;\geq\; \min_{1\leq i\leq k} \frac{1}{|\mathcal{S}|} e^{-\beta(E(\gamma_i)-E(m^{(\mathfrak{n})})+2|\mathcal{S}|\gamma_\beta)} \\ &=\; \frac{1}{|\mathcal{S}|} e^{-\beta(E(z^*(r,s))-E(m^{(\mathfrak{n})})+2|\mathcal{S}|\gamma_\beta)},\end{aligned}$$

we obtain, again using Lemma 1.1.2,

$$\begin{aligned}\frac{\pi(r)\pi(s)}{\pi(\gamma_i) \wedge \pi(\gamma_{i-1})} \;&\leq\; \frac{e^{-\beta(E(r)-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)} e^{-\beta(E(s)-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)}}{\frac{1}{|\mathcal{S}|} e^{-\beta(E(z^*(r,s))-E(m^{(\mathfrak{n})})+2|\mathcal{S}|\gamma_\beta)}} \\ &=\; |\mathcal{S}| e^{-\beta(E(r)+E(s)-E(z^*(r,s))-E(m^{(\mathfrak{n})})-6|\mathcal{S}|\gamma_\beta)} \\ &\leq\; \max_{r,s} |\mathcal{S}| e^{\beta(E(z^*(r,s))-E(r)-E(s)+E(m^{(\mathfrak{n})})-6|\mathcal{S}|\gamma_\beta)}.\end{aligned} \tag{6.3}$$

The above maximum is attained for $r = m^{(\mathfrak{n})}$ and $s = m^{(\mathfrak{n}-1)}$ or vice versa (in the case of $m^{(\mathfrak{n})} = m^{(1)}$, $m^{(\mathfrak{n}-1)}$ should denote an arbitrary state; in this case $\Delta = 0$): For $E(z^*(r,s)) - E(r) - E(s)$ to be positive, $r, s \notin z^*(r,s)$ and thus, for maximization, they have to be in different valleys. Among the local minima, $m^{(\mathfrak{n})}$ and $m^{(\mathfrak{n}-1)}$ provide the highest value.

Furthermore, for $E(\gamma_i) \geq E(\gamma_{i-1})$,

$$\begin{aligned}\pi(\gamma_i) \wedge \pi(\gamma_{i-1}) \;&\leq\; e^{-\beta(E(\gamma_i)-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)} \wedge e^{-\beta(E(\gamma_{i-1})-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)} \\ &=\; e^{-\beta(E(\gamma_i)-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)} \\ &=\; e^{-\beta(E(\gamma_{i-1})-E(m^{(\mathfrak{n})})+2|\mathcal{S}|\gamma_\beta)} e^{-\beta(E(\gamma_i)-E(\gamma_{i-1})+\gamma_\beta)} e^{5\beta|\mathcal{S}|\gamma_\beta} \\ &\leq\; |\mathcal{S}| \pi(\gamma_{i-1}) p(\gamma_{i-1}, \gamma_i) e^{5\beta|\mathcal{S}|\gamma_\beta},\end{aligned} \tag{6.4}$$

and analogously for $E(\gamma_i) < E(\gamma_{i-1})$,

$$\pi(\gamma_i) \wedge \pi(\gamma_{i-1}) \;\leq\; |\mathcal{S}| \pi(\gamma_i) p(\gamma_i, \gamma_{i-1}) e^{5\beta|\mathcal{S}|\gamma_\beta}.$$

Combining Equations (6.1), (6.2), (6.3), and (6.4), we conclude

$$\begin{aligned}Var_\pi(f) \;&\leq\; \sum_s \sum_r \left(f(s) - f(r)\right)^2 \pi(r)\pi(s) \\ &\leq\; \sum_s \sum_r |\mathcal{S}| \sum_{i=1}^{k(r,s)} \left(f(\gamma_i) - f(\gamma_{i-1})\right)^2 \left(\pi(\gamma_i) \wedge \pi(\gamma_{i-1})\right) \frac{\pi(r)\pi(s)}{\pi(\gamma_i) \wedge \pi(\gamma_{i-1})} \\ &\leq\; \sum_s \sum_r \sum_{i=1}^{k(r,s)} \left(f(\gamma_i) - f(\gamma_{i-1})\right)^2 |\mathcal{S}|^3 e^{\beta(\Delta - |\mathcal{S}|\gamma_\beta)} \\ &\qquad \times \left(\pi(\gamma_{i-1}) p(\gamma_{i-1}, \gamma_i) \mathbb{1}_{\{E(\gamma_i) \geq E(\gamma_{i-1})\}} + \pi(\gamma_i) p(\gamma_i, \gamma_{i-1}) \mathbb{1}_{\{E(\gamma_i) < E(\gamma_{i-1})\}}\right) \\ &\leq\; 2|\mathcal{S}|^4 \mathcal{E}(f) e^{\beta(\Delta - |\mathcal{S}|\gamma_\beta)},\end{aligned}$$

and therefore, recalling Proposition A.2.1 (d),

$$\gamma \;=\; \min_f \frac{\mathcal{E}(f)}{Var_\pi(f)} \;\geq\; \frac{1}{2} |\mathcal{S}|^{-5} e^{-\beta(\Delta - |\mathcal{S}|\gamma_\beta)}.$$

Finally,

$$\liminf_{\beta\to\infty} \frac{1}{\beta} \ln(\gamma) \;\geq\; -\Delta.$$

For the upper bound, we have to distinguish two cases, $\Delta > 0$ and $\Delta = 0$. If $\Delta > 0$, define

$$A \;:=\; \left\{ x \in \mathcal{S} \middle| E(z^*(m^{(\mathfrak{n}-1)}, x)) < E(z^*(m^{(\mathfrak{n})}, x)) \right\}.$$

As

$$E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n}-1)})) \;=\; E(m^{(\mathfrak{n}-1)}) \;<\; \Delta + E(m^{(\mathfrak{n}-1)}) \;=\; E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})),$$

A comprises $m^{(\mathfrak{n}-1)}$ and is therefore not empty. Due to a similar equation, namely

$$E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n})})) \;=\; E(m^{(\mathfrak{n})}) \;<\; E(m^{(\mathfrak{n}-1)}) \;<\; E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})),$$

we see furthermore that $m^{(\mathfrak{n})} \notin A$. Now let $x \in A$, $y \notin A$, $x \sim y$. Then

$$\begin{aligned} E(z^*(m^{(\mathfrak{n}-1)}, x)) \;<\; E(z^*(m^{(\mathfrak{n})}, x)) \;&\leq\; E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})) \vee E(z^*(m^{(\mathfrak{n}-1)}, x)) \\ &=\; E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})), \end{aligned}$$

and thus

$$\begin{aligned} E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) \;&\leq\; E(z^*(m^{(\mathfrak{n}-1)}, y)) \vee E(z^*(m^{(\mathfrak{n})}, y)) \\ &=\; E(z^*(m^{(\mathfrak{n}-1)}, y)) \\ &\leq\; E(z^*(m^{(\mathfrak{n}-1)}, x)) \vee E(z^*(x, y)) \\ &=\; E(z^*(m^{(\mathfrak{n}-1)}, x)) \vee E(x) \vee E(y) \\ &=\; E(z^*(m^{(\mathfrak{n}-1)}, x)) \vee E(y) \\ &=\; E(y). \end{aligned} \tag{6.5}$$

In particular $E(x) \leq E(y)$. We obtain for the Dirichlet form of $f = \mathbb{1}_A$

$$\mathcal{E}(\mathbb{1}_A) \;=\; \frac{1}{2} \sum_{x \neq y} \left(\mathbb{1}_A(x) - \mathbb{1}_A(y)\right)^2 \pi(x) p(x,y) \;=\; \sum_{x \in A \not\ni y} \pi(x) p(x,y),$$

and for the variance

$$\begin{aligned} Var_\pi(\mathbb{1}_A) \;&=\; \sum_x \left(\mathbb{1}_A(x) - \sum_y \mathbb{1}_A(y)\pi(y) \right)^2 \pi(x) \\ &=\; \sum_x \left(\mathbb{1}_A(x) - \pi(A)\right)^2 \pi(x) \\ &=\; \sum_x \left(\mathbb{1}_A(x) - 2\mathbb{1}_A(x)\pi(A) + \pi(A)^2\right) \pi(x) \\ &=\; \pi(A) - 2\pi(A)^2 + \pi(A)^2 \\ &=\; \pi(A)(1 - \pi(A)) \\ &=\; \pi(A)\pi(A^c) \\ &\geq\; \pi(m^{(\mathfrak{n}-1)})\pi(m^{(\mathfrak{n})}). \end{aligned}$$

Therefore, using Equation (6.5),

$$\begin{aligned}
\gamma &\leq \frac{\mathcal{E}(\mathbb{1}_A)}{Var_\pi(\mathbb{1}_A)} \\
&\leq \sum_{x\in A\not\ni y} \frac{\pi(x)p(x,y)}{\pi(m^{(\mathfrak{n})})\pi(m^{(\mathfrak{n}-1)})} \\
&\leq \sum_{x\in A\not\ni y} e^{-\beta(E(x)-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta+(E(y)-E(x))^+-\gamma_\beta-E(m^{(\mathfrak{n}-1)})+E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)} \\
&\leq \sum_{x\in A\not\ni y} e^{-\beta(E(y)-E(m^{(\mathfrak{n}-1)})-5|\mathcal{S}|\gamma_\beta)} \\
&\leq |\mathcal{S}|^2 e^{-\beta(E(z^*(m^{(\mathfrak{n}-1)},m^{(\mathfrak{n})}))-E(m^{(\mathfrak{n}-1)})-5|\mathcal{S}|\gamma_\beta)} \\
&= |\mathcal{S}|^2 e^{-\beta(\Delta-5|\mathcal{S}|\gamma_\beta)},
\end{aligned}$$

where of course the sum ranges over those $x \notin A \ni y$ with $p(x,y) > 0$. This gives the upper bound

$$\limsup_{\beta\to\infty} \frac{1}{\beta}\ln(\gamma) \;\leq\; -\Delta$$

in the considered case.

If $\Delta = 0$ and the set A from above is empty, define the set

$$B \;:=\; \left\{ x \in \mathcal{S} \,\middle|\, E(x) = \max_y E(y) \right\}.$$

Note that $m^{(\mathfrak{n})} \notin B$ and let $s \in B$ be arbitrary. In the same manner as above,

$$\begin{aligned}
\gamma &\leq \frac{\mathcal{E}(\mathbb{1}_B)}{Var_\pi(\mathbb{1}_B)} \\
&\leq \sum_{x\notin B\ni y} \frac{\pi(x)p(x,y)}{\pi(m^{(\mathfrak{n})})\pi(s)} \\
&\leq \sum_{x\notin B\ni y} e^{-\beta(E(x)-E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta+(E(y)-E(x))^+-\gamma_\beta-E(s)+E(m^{(\mathfrak{n})})-2|\mathcal{S}|\gamma_\beta)} \\
&\leq \sum_{x\notin B\ni y} e^{-\beta(E(y)-E(s)-5|\mathcal{S}|\gamma_\beta)} \\
&\leq |\mathcal{S}|^2 e^{5\beta|\mathcal{S}|\gamma_\beta},
\end{aligned}$$

where again the sum ranges over those $x \notin B \ni y$ with $p(x,y) > 0$. We conclude the upper bound

$$\limsup_{\beta\to\infty} \frac{1}{\beta}\ln(\gamma) \;\leq\; 0 \;=\; -\Delta$$

in this case as well. □

We state the results in the following subsections as corollaries of this theorem, not because they all are straight forward consequences of it, but because the used techniques are basically the same.

6.1.2. The Restricted Chain

Definition 6.1.2. Let $V = V^{(i)}(m)$ be a valley of some order $1 \leq i \leq \mathfrak{n}$ around some metastable $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$. Then the *restricted chain* $X_{|V}$ is a Markov chain on V with transition probabilities

$$p_{|V}(x,y) \;=\; \begin{cases} p(x,y), & x \neq y \\ p(x,x) + \sum_{z \notin V} p(x,z), & x = y \end{cases}$$

for $x, y \in V$.

In the same manner as for $\mathcal{S}$ (in Subsection 1.2.1), we identify the local minima $m_V^{(1)},...,m_V^{(k)}$ in V, ordered by increasing stability, and the depth $\Delta_V := E(z^*(m_V^{(k-1)}, m_V^{(k)})) - E(m_V^{(k-1)})$ of the second deepest (sub-)valley in V.

Corollary 6.1.3. *Let $V = V^{(i)}(m)$ by a valley of some order $1 \leq i < \mathfrak{n}$ around some metastable $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ and $m_V^{(1)},...,m_V^{(k)}$ the local minima in V, ordered by increasing stability. Let $\gamma_{|V}$ be the spectral gap of $X_{|V}$ and $\Delta_V := E(z^*(m_V^{(k-1)}, m_V^{(k)})) - E(m_V^{(k-1)})$ the depth of the second deepest (sub-)valley. Then*

$$\lim_{\beta \to \infty} \frac{1}{\beta} \ln(\gamma_{|V}) \;=\; -\Delta_V.$$

This statement is a direct consequence of Theorem 6.1.1, as the restricted chain acts on V in the same way as X acts on $\mathcal{S}$. In virtue of this corollary, the restricted chains mix exponentially faster than the original one, for the subvalleys are always less deep than the second deepest valley, that is, $\Delta_V < \Delta$ for every valley of order $i < \mathfrak{n}$. This is another definition of metastability (see [45]) since it entails a slow motion between the subsets compared to the motion within these subsets. Mimicking the proof from above, we see that for $\Delta_V > 0$

$$\begin{aligned} \sum_{x \notin A \ni y} \frac{\pi(x)p(x,y)}{\pi(m^{(k-1)})} \;&\leq\; \sum_{x \notin A \ni y} e^{-\beta(E(x) - E(m^{(k-1)}) - 2|\mathcal{S}|\gamma_\beta + (E(y)-E(x))^+ - \gamma_\beta)} \\ &\leq\; \sum_{x \notin A \ni y} e^{-\beta(E(x) \vee E(y) - E(m^{(k-1)}) - 3|\mathcal{S}|\gamma_\beta)} \\ &=\; \sum_{x \notin A \ni y} e^{-\beta(E(z^*(m^{(k-1)}, m^{(k)})) - E(m^{(k-1)}) - 3|\mathcal{S}|\gamma_\beta)} \\ &\leq\; |V|^2 e^{-\beta(\Delta_V - 3|\mathcal{S}|\gamma_\beta)}, \end{aligned} \tag{6.6}$$

with

$$A \;:=\; \left\{ x \in V \middle| E(z^*(m^{(k-1)}, x)) < E(z^*(m^{(k)}, x)) \right\}.$$

We will use this later.

6.1.3. Mixing to Quasi-Stationarity

The above corollary, most notably Equation (6.6), entails the following result on the convergence rate towards the quasi-stationary distribution on a valley V. As in Section 2.3, we denote the quasi-stationary distribution by ν and the second larges eigenvalue of $(p(x,y))_{x,y \in V}$ by $\lambda(V)$.

Corollary 6.1.4. *Let $V = V^{(i)}(m)$ be a valley of some order $1 \leq i < \mathfrak{n}$ around some metastable $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$ with $E(s_m) - E(m') > \Delta_V > 0$ for the second most stable minimum m' in V (after*

m). Let $|\lambda_2(V)|$ be the modulus of the (in terms of moduli) second largest eigenvalue of $(p(x,y))_{x,y\in V}$ and k the highest multiplicity of all eigenvalues with modulus equal to $|\lambda_2(V)|$. Then

$$\lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(1-\lim_{n\to\infty}\left(\max_{x\in V}d_{TV}(\mathbb{P}_x^{X_n|\zeta_0>n},\nu)\right)^{\frac{1}{n}}\right) \;=\; -\Delta_V.$$

As $\gamma = 1-\lim_{n\to\infty}\left(\max_{x\in\mathcal{S}}d_{TV}\left(\mathbb{P}_x^{X_n},\pi\right)\right)^{\frac{1}{n}}$, this is the analog of the previous result. It reveals that, when comparing with the restricted chain, the draining of mass does not affect the rate of convergence against the quasi-stationary distribution whenever $E(s_m)-E(m')\geq\Delta_V$. We will see in the proof that this assumption causes the draining to be of a sufficiently small rate.

Proof: In [17, Section 4], Darroch & Seneta argue that for every initial distribution μ, $x\in V$ and $n\to\infty$

$$\mathbb{P}_\mu(X_n=x|\zeta_0>n) \;=\; \nu(x)+\mathcal{O}\left(n^{k-1}\left(\frac{|\lambda_2(V)|}{\lambda(V)}\right)^n\right).$$

Therefore, for every $x\in V$,

$$d_{TV}(\mathbb{P}_x^{X_n|\zeta_0>n},\nu)^{\frac{1}{n}} \;=\; \mathcal{O}\left(n^{k-1}\left(\frac{|\lambda_2(V)|}{\lambda(V)}\right)^n\right)^{\frac{1}{n}} \;=\; \mathcal{O}\left(n^{(k-1)/n}\frac{|\lambda_2(V)|}{\lambda(V)}\right) \overset{n\to\infty}{\longrightarrow} \frac{|\lambda_2(V)|}{\lambda(V)}$$

and

$$\lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(1-\lim_{n\to\infty}\left(\max_{x\in V}d_{TV}(\mathbb{P}_x^{X_n|\zeta_0>n},\nu)\right)^{\frac{1}{n}}\right) \;=\; \lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(1-\frac{|\lambda_2(V)|}{\lambda(V)}\right),$$

if the limit exists. The rest of the proof is dedicated to the estimation of the ratio and the determination of the limit. The proof is rather long but based on a simple idea: For the lower bound, we use the Courant-Fischer-Theorem (see e.g. Theorem 4.2.12 in [32]), whereas for the upper bound we mimic the characterization of the second largest eigenvalue of an ergodic, reversible Markovian transition matrix by its Dirichlet form. We will end up with the left-hand side of Equation (6.6).

Let $Q:=(p(x,y))_{x,y\in V}$ and d be a right eigenfunction of Q to the eigenvalue $\lambda(V)>|\lambda_2(V)|$. By the Perron-Frobenius Theorem ([57, Theorem 1.1], original from 1912 in [25]), all entries of d are strictly positive, and we further assume $\sum_{x\in V}d(x)^2\pi(x)=1$. Then define $D_\pi:=diag(\pi(x),x\in V)$ and

$$A \;:=\; D_\pi^{1/2}QD_\pi^{-1/2},\quad a(x,y) \;=\; \pi(x)^{\frac{1}{2}}p(x,y)\pi(y)^{-\frac{1}{2}},\quad x,y\in V.$$

From the reversibility of $(\mathbf{P},\pi)$ we obtain

$$a(x,y) \;=\; \frac{\pi(x)p(x,y)}{\sqrt{\pi(x)\pi(y)}} \;=\; \frac{\pi(y)p(y,x)}{\sqrt{\pi(x)\pi(y)}} \;=\; a(y,x),$$

that is A is symmetric. Due to the spectral theorem for symmetric matrices (see for example [11, Ch. 7.6, Corollary 5]), all eigenvalues $(\lambda_j)_{1\leq j\leq|V|}$ are real and there is an orthonormal basis $(\phi_j)_{1\leq j\leq|V|}$ of $\left(\mathbb{R}^{|V|},\langle\cdot,\cdot\rangle\right)$ given by the eigenfunctions of A. As

$$A(D_\pi^{1/2}d) \;=\; D_\pi^{1/2}Qd \;=\; \lambda(V)(D_\pi^{1/2}d),\quad \text{and}\quad \langle D_\pi^{1/2}d,D_\pi^{1/2}d\rangle \;=\; \sum_x\pi(x)d(x)^2 \;=\; 1,$$

we fix $\phi_1\equiv D_\pi^{1/2}d$ and $\lambda_1\equiv\lambda(V)$. On the other hand, with $f_1:=d=D_\pi^{-1/2}\phi_1$ and $f_j:=D_\pi^{-1/2}\phi_j$ for $j\geq 2$, we have

$$Qf_j \;=\; (D_\pi^{-1/2}AD_\pi^{1/2})(D_\pi^{-1/2}\phi_j) \;=\; D_\pi^{-1/2}A\phi_j \;=\; D_\pi^{-1/2}\lambda_j\phi_j \;=\; \lambda_jf_j.$$

Therefore, $(f_j)_{1\le j\le|V|}$ are eigenfunctions of Q to the eigenvalues $(\lambda_j)_{1\le j\le|V|}$. In particular, $\lambda_2(V) = \lambda_2$ is real. As $\mathbf{P}$ is lazy, that is, $p(x,x) \ge \frac{1}{2}$ for every $x \in \mathcal{S}$, $(2Q - I)$ is substochastic and still aperiodic. Hence, all eigenvalues of $Q = \frac{1}{2}I + \frac{1}{2}(2Q - I)$ are nonnegative and $\lambda_2 \ge \lambda_j$ for every $j \ge 2$.

For the lower bound, by the Courant-Fischer-Theorem [32, Theorem 4.2.12],

$$\lambda_2 = \min_{x\in\mathbb{R}^{|V|}} \max_{0\neq y\in\mathbb{R}^{|V|}, y\perp x, \|y\|=1} \langle y, Ay\rangle.$$

But $A = D_\pi^{1/2} Q D_\pi^{-1/2} = D_\pi^{1/2}(\mathbf{P}_{|V} - (I - D))D_\pi^{-1/2}$ with $D = diag(p(x,V), x \in V)$, and $\pi_{|V} = \pi/\pi(V)$ is the stationary distribution of $X_{|V}$. Therefore,

$$\begin{aligned}
\langle y, Ay\rangle &= \langle y, D_\pi^{1/2}\mathbf{P}_{|V} D_\pi^{-1/2} y\rangle - \langle y, D_\pi^{1/2}(I-D)D_\pi^{-1/2} y\rangle \\
&= \langle y, D_{\pi_{|V}}^{1/2}\mathbf{P}_{|V} D_{\pi_{|V}}^{-1/2} y\rangle - \langle y, (I-D)y\rangle \\
&\le \langle y, D_{\pi_{|V}}^{1/2}\mathbf{P}_{|V} D_{\pi_{|V}}^{-1/2} y\rangle.
\end{aligned}$$

In the same manner as before, it can be shown that $D_{\pi_{|V}}^{1/2}\mathbf{P}_{|V} D_{\pi_{|V}}^{-1/2}$ is symmetric with the same eigenvalues as $\mathbf{P}_{|V}$, which are positive and the second largest of which is given by $1 - \gamma_{|V}$. Therefore,

$$\begin{aligned}
\lambda_2 &= \min_{x\in\mathbb{R}^{|V|}} \max_{0\neq y\in\mathbb{R}^{|V|}, y\perp x, \|y\|=1} \langle y, Ay\rangle \\
&\le \min_{x\in\mathbb{R}^{|V|}} \max_{0\neq y\in\mathbb{R}^{|V|}, y\perp x, \|y\|=1} \langle y, D_{\pi_{|V}}^{1/2}\mathbf{P}_{|V} D_{\pi_{|V}}^{-1/2} y\rangle \qquad (6.7)\\
&= 1 - \gamma_{|V} \\
&= 1 - e^{-\beta(\Delta_V + o(1))}
\end{aligned}$$

as $\beta \to \infty$.

For the upper bound, define a scalar product $\langle\cdot,\cdot\rangle_\pi$ on $\mathbb{R}^{|V|}$ by

$$\langle f, g\rangle_\pi := \sum_{x\in V} f(x)g(x)\pi(x).$$

With respect to this scalar product, the eigenfunctions $(f_j)_{1\le j\le|V|}$ are orthonormal, for

$$\delta_{ij} = \langle \phi_i, \phi_j\rangle = \langle D_\pi^{1/2} f_i, D_\pi^{1/2} f_j\rangle = \langle f_i, f_j\rangle_\pi$$

for every $1 \le i, j \le |V|$. Furthermore, since $(\phi_j)_{1\le j\le|V|}$ is a basis, so is $(f_j)_{1\le j\le|V|}$. Now we define the Dirichlet form of (Q,π), though π is not stationary for Q. For any function $f : V \to \mathbb{R}$ let

$$\begin{aligned}
\mathcal{E}_Q(f) &:= \frac{1}{2}\sum_{x,y\in V} (f(x) - f(y))^2 \pi(x) q(x,y) \\
&= \frac{1}{2}\left(\sum_{x,y\in V} f(x)^2\pi(x)p(x,y) + \sum_{x,y\in V} f(y)^2\pi(x)p(x,y)\right) - \sum_{x,y\in V} f(x)f(y)\pi(x)p(x,y) \\
&= \sum_{x\in V} f(x)^2\pi(x)p(x,V) - \sum_{x\in V} f(x)(Qf)(x)\pi(x) \\
&= \langle f, Df\rangle_\pi - \langle f, Qf\rangle_\pi.
\end{aligned}$$

In particular, for $f = \sum_{j=1}^{|V|} a_j f_j$, $a_j \in \mathbb{R}$, $1 \leq j \leq |V|$, with $\langle f, f_1 \rangle_\pi = 0$ (that is $a_1 = 0$), this yields

$$\begin{aligned}
\mathcal{E}_Q(f) &= \langle f, Df \rangle_\pi - \langle f, Qf \rangle_\pi \\
&= \langle f, Df \rangle_\pi - \sum_{j=1}^{|V|} a_j \langle f, Qf_j \rangle_\pi \\
&= \langle f, Df \rangle_\pi - \sum_{j=2}^{|V|} a_j \lambda_j \langle f, f_j \rangle_\pi \qquad (6.8) \\
&= \langle f, Df \rangle_\pi - \sum_{j=2}^{|V|} a_j^2 \lambda_j \\
&\geq \langle f, Df \rangle_\pi - \lambda_2 \sum_{j=2}^{|V|} a_j^2 \\
&= \langle f, Df \rangle_\pi - \lambda_2 \langle f, f \rangle_\pi .
\end{aligned}$$

Let as before $f := \mathbb{1}_A$ with

$$A := \left\{ x \in V | E(z^*(m_V^{(k-1)}, x)) < E(z^*(m_V^{(k)}, x)) \right\}.$$

Again, $m_V^{(1)}, ..., m_V^{(k)}$ are the local minima in V, ordered by increasing stability. Note that $m_V^{(k-1)} \in A$, $m_V^{(k)} \notin A$, for $\Delta_V > 0$. Since $\langle f, f_1 \rangle_\pi \neq 0$ in general, define

$$g := f - \frac{\langle f, f_1 \rangle_\pi}{\langle 1, f_1 \rangle_\pi},$$

which satisfies

$$\langle g, f_1 \rangle_\pi = \langle f, f_1 \rangle_\pi - \langle 1, f_1 \rangle_\pi \frac{\langle f, f_1 \rangle_\pi}{\langle 1, f_1 \rangle_\pi} = 0.$$

Equation (6.8) applied to f resp. g yields

$$\mathcal{E}_Q(f) = \mathcal{E}_Q(g) \geq \langle g, Dg \rangle_\pi - \lambda_2 \langle g, g \rangle_\pi$$

or, equivalently,

$$\frac{\mathcal{E}_Q(f)}{\langle g, Dg \rangle_\pi} \geq 1 - \lambda_2 \frac{\langle g, g \rangle_\pi}{\langle g, Dg \rangle_\pi}.$$

A simple calculation shows that

$$\begin{aligned}
\langle g, Dg \rangle_\pi &= \sum_{x \in A} \pi(x) p(x, V) \left(1 - \left(\frac{\langle f, f_1 \rangle_\pi}{\langle 1, f_1 \rangle_\pi} \right) \right)^2 + \sum_{x \in V \setminus A} \pi(x) p(x, V) \left(\frac{\langle f, f_1 \rangle_\pi}{\langle 1, f_1 \rangle_\pi} \right)^2 \\
&\geq (\pi(A) \wedge \pi(V \setminus A)) \min_{x \in V} p(x, V) \left(\left(1 - \frac{\langle f, f_1 \rangle_\pi}{\langle 1, f_1 \rangle_\pi} \right)^2 + \left(\frac{\langle f, f_1 \rangle_\pi}{\langle 1, f_1 \rangle_\pi} \right)^2 \right) \\
&\geq (\pi(A) \wedge \pi(V \setminus A)) \min_{x \in V} p(x, V) \frac{1}{2} \\
&\geq \left(\pi \left(m_V^{(k-1)} \right) \wedge \pi \left(m_V^{(k)} \right) \right) \frac{\min_{x \in V} p(x, V)}{2} \\
&= \pi \left(m_V^{(k-1)} \right) \frac{\min_{x \in V} p(x, V)}{2}.
\end{aligned}$$

Here we used the fact that $(1-x)^2+x^2 = 2\left(\frac{1}{2}-x\right)^2+\frac{1}{2} \geq \frac{1}{2}$ for any $x \in [0,1]$, and that f_1 is strictly positive, which yields $\langle f, f_1\rangle_\pi = \sum_{x\in A} f_1(x)\pi(x) \leq \sum_{x\in V} f_1(x)\pi(x) = \langle 1, f_1\rangle_\pi$. Now we conclude that $\min_{x\in V} p(x,V)$ tends to 1 as β tends to infinity, so that, using Equation (6.6),

$$\begin{aligned} 1-\lambda_2\frac{\langle g,g\rangle_\pi}{\langle g,Dg\rangle_\pi} &\leq \frac{\mathcal{E}_Q(f)}{\langle g,Dg\rangle_\pi} \\ &\leq 4\cdot\frac{\mathcal{E}_Q(\mathbb{1}_A)}{\pi\left(m_V^{(k-1)}\right)} \\ &= 4\sum_{x\notin A\ni y}\frac{\pi(x)p(x,y)}{\pi\left(m_V^{(k-1)}\right)} \\ &\leq 4|V|^2 e^{-\beta(\Delta_V-3|\mathcal{S}|\gamma_\beta)}. \end{aligned} \tag{6.9}$$

We further infer

$$\langle g,g\rangle_\pi = \pi(A)\left(1-\frac{\langle f,f_1\rangle_\pi}{\langle 1,f_1\rangle_\pi}\right)^2 + \pi(V\backslash A)\left(\frac{\langle f,f_1\rangle_\pi}{\langle 1,f_1\rangle_\pi}\right)^2 \geq \frac{\pi\left(mV^{(k-1)}\right)}{2},$$

and, using $g(x)\in[-1,1]$, for $\langle f,f_1\rangle_\pi \leq \langle 1,f_1\rangle_\pi$,

$$\begin{aligned} \langle g,Dg\rangle_\pi &= \langle g,g\rangle_\pi - \langle g,(I-D)g\rangle_\pi \\ &= \langle g,g\rangle_\pi - \sum_{x\in\partial V} g(x)^2\pi(x)p(x,N^{(i)}) \\ &\geq \langle g,g\rangle_\pi - \sum_{x\in\partial V}\pi(x)p(x,N^{(i)}) \\ &\geq \langle g,g\rangle_\pi - \sum_{x\in\partial V}\sum_{y\in N^{(i)},y\sim x} e^{-\beta(E(x)-E(m^{(\mathfrak{n})})+(E(y)-E(x))^+ +o(1))} \\ &\geq \langle g,g\rangle_\pi - e^{-\beta(E(s_m)-E(m^{(\mathfrak{n})})+o(1))} \end{aligned}$$

since $E(y)\vee E(x) = E(y) \geq E(s_m)$ for every $x\in\partial V$, $y\in\partial^+ V$ with $x\sim y$. A combination of both estimates yields

$$\begin{aligned} \frac{\langle g,Dg\rangle_\pi}{\langle g,g\rangle_\pi} &\geq 1-2\frac{e^{-\beta(E(s_m)-E(m^{(\mathfrak{n})})+o(1))}}{\pi(m^{(k-1)})} \\ &\geq 1-2e^{-\beta(E(s_m)-E(m^{(\mathfrak{n})})-E(m^{(k-1)})+E(m^{(\mathfrak{n})})+o(1))} \\ &= 1-e^{-\beta(E(s_m)-E(m^{(k-1)})+o(1))}. \end{aligned} \tag{6.10}$$

Via Insertion of Equation (6.10) in Equation (6.9) we obtain

$$\begin{aligned} \lambda_2 &\geq \left(1-e^{-\beta(\Delta_V+o(1))}\right)\left(1-e^{-\beta(E(s_m)-E(m^{(k-1)})+o(1))}\right) \\ &= 1-e^{-\beta(\Delta_V+o(1))}\left(1-e^{-\beta(E(s_m)-E(m^{(k-1)})+o(1))}+e^{-\beta(E(s_m)-E(m^{(k-1)})-\Delta_V+o(1))}\right) \\ &= 1-e^{-\beta(\Delta_V+o(1))}(1+o(1)), \end{aligned}$$

as by assumption $\Delta_V < E(s_m)-E(m^{(k-1)})$. Together with Equation (6.7) we infer

$$\lambda_2 = 1-e^{-\beta(\Delta_V+o(1))}(1+o(1)).$$

Using now that $\lambda(V) = 1 - e^{-\beta(E(s_m)-E(m)+o(1))}$ by Proposition 2.3.3, this yields

$$\begin{aligned} 1 - \frac{\lambda_2(V)}{\lambda(V)} &= 1 - \frac{1 - e^{-\beta(\Delta_V+o(1))}(1+o(1))}{1 - e^{-\beta(E(s_m)-E(m)+o(1))}} \\ &= \frac{(1+o(1))e^{-\beta(\Delta_V+o(1))} - e^{-\beta(E(s_m)-E(m)+o(1))}}{1 - e^{-\beta(E(s_m)-E(m)+o(1))}} \\ &= e^{-\beta(\Delta_V+o(1))}\frac{(1+o(1)) - e^{-\beta(E(s_m)-E(m)-\Delta_V+o(1))}}{1 - e^{-\beta(E(s_m)-E(m)+o(1))}}, \end{aligned}$$

and therefore, as asserted,

$$\begin{aligned} &\lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(1 - \lim_{n\to\infty}\left(\max_{x\in V} d_{TV}(\mathbb{P}_x^{X_n|\zeta_0>n}, \nu)\right)^{\frac{1}{n}}\right) \\ &\quad= \lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(1 - \frac{|\lambda_2(V)|}{\lambda(V)}\right) \\ &\quad= \lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(e^{-\beta(\Delta_V+o(1))}\frac{(1+o(1)) - e^{-\beta(E(s_m)-E(m)-\Delta_V+o(1))}}{1 - e^{-\beta(E(s_m)-E(m)+o(1))}}\right) \\ &\quad= -\Delta_V \end{aligned}$$

since $E(s_m) - E(m) > E(s_m) - E(m^{(k-1)}) > \Delta_V$ by assumption. □

6.1.4. The Hit Chain

Another interesting process derived from X is the hit chain on $\mathcal{S}^{(i)}$, which will be used later for the comparison of the mixing times of X and Y.

Definition 6.1.5. The *hit chain* $\overline{Z}$ on $\mathcal{S}^{(i)}$ has the transition matrix

$$p_{\mathcal{S}^{(i)}}(m, m') = \mathbb{P}_m(X_{\tau_{\mathcal{S}^{(i)}}} = m'), \quad m, m' \in \mathcal{S}^{(i)}.$$

Recall that by definition $\tau_{\mathcal{S}^{(i)}} \geq 1$. Due to holding probabilities of at least $\frac{1}{2}$, the absolute spectral gap and the spectral gap coincide for $\overline{Z}$ as well. We will prove that the asymptotic behavior of the convergence rate as a function of β is the same for $\overline{Z}$ and X. This indicates that the relaxation time of X is dominated by the mass-transport between valleys, given by the AC, and not within valleys.

Corollary 6.1.6. *Let Δ be defined as in Theorem 6.1.1, satisfying $\Delta > 0$, and denote the spectral gap of $\overline{Z}$ by $\gamma^{\overline{Z}}$. Then*

$$\lim_{\beta\to\infty}\frac{1}{\beta}\ln(\gamma^{\overline{Z}}) = -\Delta.$$

As already mentioned, the denotation as a corollary is justified by being proved basically with the same argument as Theorem 6.1.1.

Proof: First, we cite Lemma A.2.4 to see that $\pi_{\mathcal{S}^{(i)}} := (\pi(x)/\pi(\mathcal{S}^{(i)}))_{x\in\mathcal{S}^{(i)}}$ is reversible for $\overline{Z}$ and that $\gamma^{\overline{Z}} \geq \gamma$. Thus,

$$\liminf_{\beta\to\infty}\frac{1}{\beta}\ln(\gamma^{\overline{Z}}) \geq -\Delta.$$

Using the same argument as before, define

$$A := \left\{m \in \mathcal{S}^{(i)} | E(z^*(m^{(\mathfrak{n}-1)}, m)) < E(z^*(m^{(\mathfrak{n})}, m))\right\}.$$

Then $m^{(\mathfrak{n}-1)} \in A$, $m^{(\mathfrak{n})} \notin A$. Now let $m \in A$, $n \in \mathcal{S}^{(i)} \backslash A$ and note that

$$\begin{aligned} E(z^*(m^{(\mathfrak{n}-1)}, m)) \;<\; E(z^*(m^{(\mathfrak{n})}, m)) \;&\leq\; E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})) \vee E(z^*(m^{(\mathfrak{n}-1)}, m)) \\ &=\; E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})). \end{aligned}$$

Thus,

$$\begin{aligned} E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) \;&\leq\; E(z^*(m^{(\mathfrak{n}-1)}, n)) \vee E(z^*(m^{(\mathfrak{n})}, n)) \;=\; E(z^*(m^{(\mathfrak{n}-1)}, n)) \\ &\leq\; E(z^*(m^{(\mathfrak{n}-1)}, m)) \vee E(z^*(m, n)) \;=\; E(z^*(m, n)), \end{aligned} \tag{6.11}$$

and $E(m) \leq E(z^*(m^{(\mathfrak{n}-1)}, m)) < E(z^*(m^{(\mathfrak{n})}, m^{(\mathfrak{n}-1)})) \leq E(z^*(m, n))$. We conclude that $m \notin z^*(m, n)$ and therefore, for this choice of m and n,

$$\begin{aligned} p_{\mathcal{S}^{(i)}}(m, n) \;=\; \mathbb{P}_m(X_{\tau_{\mathcal{S}^{(i)}}} = n) \;&\leq\; \mathbb{P}_m(\tau_n < \tau_m) \\ &\leq\; e^{-\beta(E(z^*(m,n)) - E(m) + o(1))} \\ &\leq\; e^{-\beta(E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) - E(m) + o(1))}. \end{aligned}$$

In the by now familiar manner,

$$\begin{aligned} \gamma^{\overline{Z}} \;&\leq\; \frac{\mathcal{E}_{\mathcal{S}^{(i)}}(\mathbb{1}_A)}{Var_{\pi_{\mathcal{S}^{(i)}}}(\mathbb{1}_A)} \\ &\leq\; \sum_{m \in A \not\ni n} \frac{\pi_{\mathcal{S}^{(i)}}(m) p_{\mathcal{S}^{(i)}}(m, n)}{\pi_{\mathcal{S}^{(i)}}(m^{(\mathfrak{n})}) \pi_{\mathcal{S}^{(i)}}(m^{(\mathfrak{n}-1)})} \\ &\leq\; \sum_{m \in A \not\ni n} \frac{\pi(m) p_{\mathcal{S}^{(i)}}(m, n)}{\pi(m^{(\mathfrak{n})}) \pi(m^{(\mathfrak{n}-1)})} \\ &\leq\; \sum_{m \in A \not\ni n} e^{-\beta(E(m) - E(m^{(\mathfrak{n})}) + E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) - E(m) - E(m^{(\mathfrak{n}-1)}) + E(m^{(\mathfrak{n})}) + o(1))} \\ &\leq\; \sum_{m \in A \not\ni n} e^{-\beta(E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) - E(m^{(\mathfrak{n}-1)}) + o(1))} \\ &\leq\; |\mathcal{S}|^2 e^{-\beta(\Delta + o(1))}, \end{aligned}$$

from which we conclude

$$\limsup_{\beta \to \infty} \frac{1}{\beta} \ln(\gamma^{\overline{Z}}) \;\leq\; -\Delta,$$

and thus, together with the lower bound, the assertion. □

6.1.5. A Markovian Macroscopic Process

As a last functional, we regard the Markov chain given by the transition probabilities of the AC in the stationary regime, and show that it mixes asymptotically with the same rate as X, too. In addition to the previous subsections, this is another indication for $\overline{Y}$ describing the long-range part of X in an appropriate way.

Corollary 6.1.7. *Let Δ as defined in Theorem 6.1.1 be positive and $1 \leq i < \mathfrak{n}$ some fixed aggregation level. Write $V(m) := V^{(i)}(m)$ for every $m \in \mathcal{S}^{(i)}$. Let*

$$q(m, n) \;:=\; \frac{1}{\pi(V(m))} \sum_{x \in V(m)} \sum_{y \in V(n)} \pi(x) p(x, y)$$

be the transition probabilities of the AC in the stationary regime with spectral gap γ_Q. Then $(Q, \overline{\pi})$ with $\overline{\pi}(\cdot) := \pi(V(\cdot))$ is a reversible pair and

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln(\gamma_Q) \;=\; -\Delta.$$

Proof: The reversibility is a consequence of the obvious symmetry of $\sum_{x\in V(m)} \sum_{y\in V(n)} \pi(x)p(x,y)$ in m and n. For the second assertion, let $f : \mathcal{S} \to \mathbb{R}$ be some function which is constant on every $V(m)$, $m \in \mathcal{S}^{(i)}$, and $\mathcal{E}_Q$ the Dirichlet form with respect to $(Q, \overline{\pi})$. Then

$$\begin{aligned}
\mathcal{E}_Q(f) \;&:=\; \sum_{m\neq n} (f(m) - f(n))^2 \overline{\pi}(m) q(m,n) \\
&=\; \sum_{m\neq n} (f(m) - f(n))^2 \sum_{x\in V(m)} \sum_{y\in V(n)} \pi(x)p(x,y) \\
&=\; \sum_{m\neq n} \sum_{x\in V(m)} \sum_{y\in V(n)} (f(x) - f(y))^2 \pi(x)p(x,y) \\
&=\; \mathcal{E}(f)
\end{aligned}$$

and

$$\begin{aligned}
Var_{\overline{\pi}}(f) \;&:=\; \sum_m \left(f(m) - \sum_n f(n)\overline{\pi}(n) \right)^2 \overline{\pi}(m) \\
&=\; \sum_m f(m)^2 \overline{\pi}(m) - \left(\sum_n f(n)\overline{\pi}(n) \right)^2 \\
&=\; \sum_{x\in\mathcal{S}} f(x)^2 \pi(x) - \left(\sum_{x\in\mathcal{S}} f(x)\pi(x) \right)^2 \\
&=\; Var_\pi(f).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\gamma_Q \;=\; \inf_{f:\mathcal{S}^{(i)}\to\mathbb{R}, Var_{\overline{\pi}}(f)\neq 0} \frac{\mathcal{E}_Q(f)}{Var_{\overline{\pi}}(f)} \;&=\; \inf_{f:\mathcal{S}^{(i)}\to\mathbb{R}, Var_\pi(f)\neq 0} \frac{\mathcal{E}(f)}{Var_\pi(f)} \\
&\geq\; \inf_{f:\mathcal{S}\to\mathbb{R}, Var_\pi(f)\neq 0} \frac{\mathcal{E}(f)}{Var_\pi(f)} \\
&=\; \gamma,
\end{aligned}$$

and

$$\liminf_{\beta\to\infty} \frac{1}{\beta} \ln(\gamma_Q) \geq \lim_{\beta\to\infty} \frac{1}{\beta} \ln(\gamma) \;=\; -\Delta.$$

Furthermore, define again

$$A \;:=\; \left\{ m \in \mathcal{S}^{(i)} | E(z^*(m^{(\mathfrak{n}-1)}, m)) < E(z^*(m^{(\mathfrak{n})}, m)) \right\}.$$

We know already by Equation (6.11) that $E(z^*(m^{(\mathfrak{n}-1)}, m^{(\mathfrak{n})})) \;\leq\; E(z^*(m,n))$ for $m \in A$, $n \in \mathcal{S}^{(i)}\backslash A$. In addition, if $q(m,n) > 0$, then at least one of the two states is non-assigned and by Lemma 1.3.10,

$$E(z^*(m,n)) = E(m) \vee E(n) \leq E(x) \vee E(y)$$

for every two states $x \in V^{(i)}(m)$ and $y \in V^{(i)}(n)$. We obtain

$$\begin{aligned}
\gamma_Q &\leq \frac{\mathcal{E}(\mathbb{1}_A)}{Var_{\overline{\pi}}(\mathbb{1}_A)} \\
&\leq \sum_{m \in A \not\ni n} \frac{\overline{\pi}(m) q(m,n)}{\overline{\pi}(m^{(\mathfrak{n})}) \overline{\pi}(m^{(\mathfrak{n}-1)})} \\
&\leq \sum_{m \in A \not\ni n} \sum_{x \in V(m)} \sum_{y \in V(n)} \frac{\pi(x) p(x,y)}{\pi(m^{(\mathfrak{n})}) \pi(m^{(\mathfrak{n}-1)})} \\
&\leq \sum_{m \in A \not\ni n} \sum_{x \in V(m)} \sum_{y \in V(n)} e^{-\beta(E(x)-E(m^{(\mathfrak{n})})+(E(y)-E(x))^+-E(m^{(\mathfrak{n}-1)})+E(m^{(\mathfrak{n})})+o(1))} \\
&\leq \sum_{m \in A \not\ni n} \sum_{x \in V(m)} \sum_{y \in V(n)} e^{-\beta(E(z^*(m,n))-E(m^{(\mathfrak{n}-1)})+o(1))} \\
&\leq |\mathcal{S}|^2 e^{-\beta(E(z^*(m^{(\mathfrak{n}-1)},m^{(\mathfrak{n})}))-E(m^{(\mathfrak{n}-1)})+o(1))} \\
&= |\mathcal{S}|^2 e^{-\beta(\Delta+o(1))},
\end{aligned}$$

where the sum of course ranges over those $n \notin A \ni m$ with $q(m,n) > 0$. This proves

$$\limsup_{\beta \to \infty} \frac{1}{\beta} \ln(\gamma_Q) \leq -\Delta$$

and thus, together with the lower bound, the assertion. □

6.2. The General Problem

The basic idea for the derivation of a relation between the mixing times of X and Y (the precise definition of the mixing time of Y is given below) is to study the hit chain $\overline{Z}$ on $\mathcal{S}^{(i)}$ and its embedded jump chain Z: As we can relate the spectral gaps of X and $\overline{Z}$ as well as the mixing times of Y and Z due to the fact that both chains converge to the same Markov chain, it remains to compare the mixing times or gaps of $\overline{Z}$ and Z. This is a special case of the following

General Problem: Given a finite Markov chain and a decelerated version of it (for the exact definition see below), are there relations between the two associated mixing times?

In the literature, there are some results on comparisons of Markov chains in terms of their spectral gaps, most notably by Diaconis & Saloff-Coste in [20] and [21]. We can use them in the case where the decelerated version is again Markovian to obtain catchy relations between the spectral gaps. We will also study this natural problem in the case where the slowed version is only semi-Markovian. The final outcome will be formulas for the mixing time of Y in terms of the mixing time of X in Section 6.5.

Definition 6.2.1. Let $\overline{M}$ be an irreducible finite Markov chain with transition matrix $\overline{\mathbf{P}}$. With the jump epochs $\sigma_0 \equiv 0$ and $\sigma_{n+1} := \inf\{k > \sigma_n | \overline{M}_k \neq \overline{M}_{k-1}\}$, $n \geq 0$, we define the *embedded jump chain M of $\overline{M}$* by $M_n := \overline{M}_{\sigma_n}$, $n \geq 0$.

The embedded jump chain is again Markovian with transition matrix $\mathbf{P}$ given by

$$p(x,y) = \sum_{n \geq 0} \overline{p}(x,x)^n \, \overline{p}(x,y) = \frac{\overline{p}(x,y)}{1-\overline{p}(x,x)}, \quad x \neq y,$$

and $p(x,x) = 0$, that is,

$$\mathbf{P} \;=\; diag\left(\frac{1}{1-\overline{p}(x,x)}, x \in \mathcal{S}\right)\overline{\mathbf{P}} - diag\left(\frac{\overline{p}(x,x)}{1-\overline{p}(x,x)}, x \in \mathcal{S}\right).$$

Conversely, given a finite Markov chain M with transition matrix $\mathbf{P}$ and a sequence of sojourn times $(T_n)_{n\geq 0}$ on $\mathbb{N}_{>0}$, we can define a decelerated version: $\overline{M}$ visits the same states as M but stays in the k-th state for T_k time units, $k \geq 0$.

Definition 6.2.2. Let M be an irreducible finite Markov chain on $\mathcal{S}$ with transition matrix $\mathbf{P}$ and $(T_n)_{n\geq 0}$ a sequence of random variables on $\mathbb{N}_{>0}$. We define the *(by $(T_n)_{n\geq 0}$) decelerated version $\overline{M}$ of M* as the random process given by

$$\overline{M}_n \;:=\; \sum_{k=0}^{n} M_k \mathbb{1}_{\{T_0+\ldots+T_{k-1}\leq n<T_0+\ldots+T_k\}}.$$

Let $Geo(p)$ be the geometric distribution on $\mathbb{N}_{>0}$ with parameter $p \in [0,1]$. Due to the memoryless property of this distribution, the decelerated process is Markovian if

$$\mathbb{P}\left((T_n)_{n\geq 0} \in \cdot\,|M\right) \;=\; \prod_{n\geq 0}\mathbb{P}(T_n \in \cdot\,|M) \;=\; \prod_{n\geq 0} Geo(1-p_{M_n})$$

for some $(p_x)_{x\in\mathcal{S}} \in [0,1]^{\mathcal{S}}$. Note that in particular the different sojourn times at a given state x are independent and identically distributed whenever this state is reached by M. Thus, in the Markovian case it suffices to specify those geometric distributions as $(T_x)_{x\in\mathcal{S}}$. The transition matrix $\overline{\mathbf{P}}$ of $\overline{M}$ is then given by

$$\begin{aligned}\overline{p}(x,x) \;&=\; \mathbb{P}(T_x > 1) + \mathbb{P}(T_x = 1)p(x,x) \;=\; p_x + (1-p_x)p(x,x),\\ \overline{p}(x,y) \;&=\; p(x,y)\mathbb{P}(T_x = 1) \;=\; p(x,y)(1-p_x), \quad x \neq y,\end{aligned} \tag{6.12}$$

that is

$$\overline{\mathbf{P}} \;=\; diag\left(1-p_x, x\in\mathcal{S}\right)\mathbf{P} + diag\left(p_x, x\in\mathcal{S}\right).$$

For Markov chains M and $\overline{M}$ with $p(x,x) = 0$ for every $x \in \mathcal{S}$ we have: M is the embedded jump chain of $\overline{M}$ if and only if $\overline{M}$ is a decelerated chain of M. Of course, M can have more than one decelerated Markov chain (by varying the holding parameters $(p_x)_{x\in\mathcal{S}}$), as well as different Markov chains can have the same embedded jump chain.

6.3. Markovian Decelerated Chains

In this section, we focus on the case where $\overline{M}$ is a *Markovian* decelerated version of a finite Markov chain M. Before we give first statements on relations between the spectral gaps of a M and $\overline{M}$, we want to illustrate why deceleration not always increases the mixing time as one would expect at first glance.

Example 6.3.1. Let M be a random walk on a complete graph without loops on $n \geq 3$ vertices and $\overline{M}$ a decelerated version with geometric sojourn times such that $\overline{M}$ is the random walk with loops and probability $p \in (0,1)$ to stay put in any vertex. Let

$$\mathbf{P} \;=\; \begin{pmatrix} 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1}\\ \frac{1}{n-1} & 0 & \cdots & \frac{1}{n-1}\\ \vdots & & \ddots & \vdots\\ \frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0 \end{pmatrix} \quad \text{resp.} \quad \overline{\mathbf{P}} \;=\; p\mathbf{I} + (1-p)\mathbf{P}$$

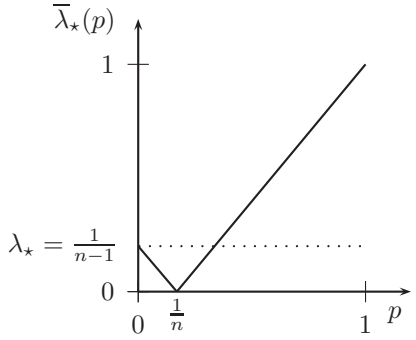

Figure 6.1.: Plot of the second largest eigenvalue (in modulus) of the decelerated transition matrix $\overline{\mathbf{P}}$ in Example 6.3.1 as a function of the holding probability p.

be the transition matrices of M resp. $\overline{M}$. Then

$$\begin{aligned} v_1 &= \left(-1, \frac{1}{n-1}, ..., \frac{1}{n-1}\right) \\ v_2 &= \left(\frac{1}{n-1}, -1, \frac{1}{n-1}, ..., \frac{1}{n-1}\right) \\ &\vdots \\ v_{n-1} &= \left(\frac{1}{n-1}, ..., \frac{1}{n-1}, -1, \frac{1}{n-1}\right) \end{aligned}$$

are left eigenvectors of $\mathbf{P}$ to the eigenvalue $\frac{-1}{n-1}$. Furthermore, they are linearly independent, which can be seen by an inductive argument (see Lemma A.3.1 in the Appendix). In particular, $\lambda = \frac{-1}{n-1}$ is an $(n-1)$-fold eigenvalue. Another eigenvector is given by $v_n = (1, ..., 1)$ with eigenvalue 1. Thus, $\lambda_\star = \frac{1}{n-1}$.

Since $\overline{\mathbf{P}} = p\mathbf{I} + (1-p)\mathbf{P}$, every eigenvector of $\mathbf{P}$ with eigenvalue λ is also an eigenvector of $\overline{\mathbf{P}}$ with eigenvalue $p + \lambda(1-p)$. Therefore, the set of eigenvalues of $\overline{\mathbf{P}}$ is given by $\{1, p - \frac{1-p}{n-1}\}$ with

$$\overline{\lambda}_\star = \left| p - \frac{1-p}{n-1} \right| = \frac{|np-1|}{n-1}.$$

Depending on the sojourn distribution, that is, on p, the absolute value of the second largest eigenvalue thus can be both: arbitrarily close to 1 as well as to 0. It is not even monotone in p, see Figure 6.1. In fact, in the case $p = \frac{1}{n}$, the transition matrix is given by

$$\overline{\mathbf{P}} = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix},$$

which equilibrates within one step. The crux in this example is that λ_2 is negative and the absolute value of the average $p \cdot 1 + (1-p) \cdot \lambda_2$ is less than $|\lambda_2|$ when choosing p properly.

Proposition 6.3.2. *Let $\overline{M}$ and M be two finite Markov chains satisfying* (6.12) *(for example M is the embedded jump chain of $\overline{M}$).*

(a) $\overline{\pi}$ is stationary for $\overline{M} \Leftrightarrow (\pi(x))_{x\in\mathcal{S}} := \left(\frac{1}{N}\overline{\pi}(x)(1-p_x)\right)_{x\in\mathcal{S}}$ is stationary for M, where $N := \sum_x \overline{\pi}(x)(1-p_x)$ is the normalization constant.

(b) $\overline{M}$ is reversible with respect to $\overline{\pi} \Leftrightarrow M$ is reversible with respect to π.

(c) If both chains are reversible with respect to $\overline{\pi}$ reps. π, then

$$\frac{\gamma}{\max_x \mathbb{E}(T_x)} \;=\; \min_x (1-p_x)\gamma \;\le\; \overline{\gamma} \;\le\; \max_x (1-p_x)\gamma \;=\; \frac{\gamma}{\min_x \mathbb{E}(T_x)} \;\le\; \gamma.$$

Proof: The proof of the reversibility and stationarity just requires an easy calculation. For every $x \neq y$,

$$\pi(x)p(x,y) \;=\; \frac{\overline{\pi}(x)(1-p_x)}{N}\frac{\overline{p}(x,y)}{1-p_x} \;=\; \frac{\overline{\pi}(x)\overline{p}(x,y)}{N}.$$

We infer that the left-hand side is symmetric in x and y if and only if the right-hand side is symmetric. Furthermore, for every y,

$$\begin{aligned} N \cdot \sum_x \pi(x)p(x,y) \;&=\; \sum_{x\neq y} \overline{\pi}(x)\overline{p}(x,y) + \overline{\pi}(y)(1-p_y)p(y,y) \\ &=\; \sum_{x\neq y} \overline{\pi}(x)\overline{p}(x,y) + \overline{\pi}(y)(\overline{p}(y,y)-p_y) \\ &=\; \sum_x \overline{\pi}(x)\overline{p}(x,y) - \overline{\pi}(y)p_y. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_x \pi(x)p(x,y) \;=\; \pi(y) \quad &\Leftrightarrow \quad \sum_x \overline{\pi}(x)\overline{p}(x,y) - \overline{\pi}(y)p_y \;=\; \overline{\pi}(y)(1-p_y) \\ &\Leftrightarrow \quad \sum_x \overline{\pi}(x)\overline{p}(x,y) \;=\; \overline{\pi}(y). \end{aligned}$$

For part (c) let $\overline{\mathcal{E}}$ and $\mathcal{E}$ be the Dirichlet forms of $\overline{M}$ and M. Then we obtain for every $f : \mathcal{S} \to \mathbb{R}$

$$\mathcal{E}(f) \;=\; \frac{1}{2}\sum_{x\neq y} (f(x)-f(y))^2\,\pi(x)p(x,y) \;=\; \frac{1}{2N}\sum_{x\neq y} (f(x)-f(y))^2\,\overline{\pi}(x)\overline{p}(x,y) \;=\; \frac{1}{N}\,\overline{\mathcal{E}}(f).$$

Thus, with Lemma A.2.3, we obtain

$$\gamma \;\le\; \frac{1}{N}\max_x \frac{N\overline{\pi}(x)}{\overline{\pi}(x)(1-p_x)}\overline{\gamma} \;=\; \max_x \frac{1}{1-p_x}\overline{\gamma},$$

and, using $N\mathcal{E}(f) = \overline{\mathcal{E}}(f)$ for every f and the same lemma,

$$\overline{\gamma} \;\le\; N\max_x \frac{\overline{\pi}(x)(1-p_x)}{N\overline{\pi}(x)}\gamma \;=\; \max_x (1-p_x)\gamma.$$

□

We note the following special case:

Corollary 6.3.3. *Let $\overline{M}$ and M be two reversible finite Markov chains satisfying* (6.12) *and $p_x = p$ for every $x \in \mathcal{S}$. Then*

$$(1-p)\gamma \;=\; \overline{\gamma}.$$

Hence, the deceleration of a reversible finite Markov chain to a Markov chain by independent geometric sojourn times $(T_x)_{x\in\mathcal{S}}$ decreases the spectral gap, in fact by at most the factor $(\max_x \mathbb{E}(T_x))^{-1}$ and at least the factor $(\min_x \mathbb{E}(T_x))^{-1}$. This result would be even more satisfactory, if it held for the relaxation times as well. But unfortunately, all statements above are made for the spectral gap (or equivalently for λ_2), and do not generally hold for the absolute spectral gap (or equivalently $\lambda_\star$) which defines the relaxation and mixing time. Moreover, the ergodicity (in particular the aperiodicity) of $\overline{M}$ carries not necessarily over to M. On the other hand, if $\overline{p}(x,x) \geq \frac{1}{2}$ for every $x \in \mathcal{S}$, we obtain results for the absolute spectral gap without additional work.

Theorem 6.3.4. *Let $\overline{M}$ and M be two reversible finite Markov chains satisfying* (6.12) *with $\overline{p}(x,x) \geq \frac{1}{2}$ for every $x \in \mathcal{S}$. Then*

$$\min_x (1-p_x)\gamma_\star \;\leq\; \overline{\gamma}_\star. \tag{6.13}$$

If additionally $p(x,x) \geq \frac{1}{2}$ for every $x \in \mathcal{S}$, then

$$\overline{\gamma}_\star \;\leq\; \max_x (1-p_x)\gamma_\star \;\leq\; \gamma_\star.$$

Proof: These bounds follow directly from Proposition 6.3.2 by observing $\gamma_\star = 1-\lambda_\star \leq 1-\lambda_2 = \gamma$ in general, and $\gamma_\star = \gamma$ if $\min_x p(x,x) \geq \frac{1}{2}$ as well as $\overline{\gamma}_\star = \overline{\gamma}$ if $\min_x \overline{p}(x,x) \geq \frac{1}{2}$ (Proposition A.2.1). □

Remark 6.3.5. (a) Obviously, the above theorem provides also a statement for two different decelerated Markovian versions of the same reversible finite Markov chain M: Let $(T_x)_{x\in\mathcal{S}}$ and $(S_x)_{x\in\mathcal{S}}$ be to two different families of waiting times with $T_x \leq_{st} S_x$ for every $x \in \mathcal{S}$, that is, $p_x^T \leq p_x^S$ for every $x \in \mathcal{S}$. Denote the two decelerated versions by $\overline{M}^S$ and $\overline{M}^T$, and

$$p_x \;:=\; \frac{p_x^S - p_x^T}{1-p_x^T} \;=\; 1 - \frac{1-p_x^S}{1-p_x^T} \;=\; 1 - \frac{\mathbb{E}(T_x)}{\mathbb{E}(S_x)}.$$

The transition probabilities satisfy

$$\begin{aligned}
p_x + (1-p_x)\overline{p}^T(x,x) &= \frac{p_x^S - p_x^T}{1-p_x^T} + \left(1 - \frac{p_x^S - p_x^T}{1-p_x^T}\right)\left(p_x^T + (1-p_x^T)p(x,x)\right) \\
&= \frac{p_x^S - p_x^T}{1-p_x^T}\left(1 - p_x^T - (1-p_x^T)p(x,x)\right) + p_x^T + (1-p_x^T)p(x,x) \\
&= \left(p_x^S - p_x^T\right)(1-p(x,x)) + p_x^T(1-p(x,x)) + p(x,x) \\
&= p_x^S\,(1-p(x,x)) + p(x,x) \\
&= p_x^S + (1-p_x^S)p(x,x) \\
&= \overline{p}^S(x,x)
\end{aligned}$$

for every $x \in \mathcal{S}$, and

$$(1-p_x)\overline{p}^T(x,y) \;=\; \left(1 - \frac{p_x^S - p_x^T}{1-p_x^T}\right)(1-p_x^T)p(x,y) \;=\; (1-p_x^S)p(x,y) \;=\; \overline{p}^S(x,y)$$

for $x \neq y$. Thus, $\overline{M}^S$ is a decelerated version of $\overline{M}^T$. If $\overline{p}^T(x,x) \geq \frac{1}{2}$ and $\overline{p}^S(x,x) \geq \frac{1}{2}$ for every $x \in \mathcal{S}$, then

$$\min_x \frac{\mathbb{E}(T_x)}{\mathbb{E}(S_x)}\gamma_\star^T \;=\; \min_x (1-p_x)\gamma_\star^T \;\leq\; \gamma_\star^S \;\leq\; \max_x (1-p_x)\gamma_\star^T \;=\; \max_x \frac{\mathbb{E}(T_x)}{\mathbb{E}(S_x)}\gamma_\star^T \;\leq\; \gamma_\star^T.$$

As the two decelerated chains are ergodic due to the positive holding probabilities, this yields that $\overline{M}^T$ mixes faster than $\overline{M}^S$.

(b) In the special case where $p_x = p$ for every x, Equation (6.13) (even without assuming holding probabilities of at least 1/2) is also a consequence of the convexity of $\lambda_\star = \boldsymbol{\lambda}_\star(\mathbf{P})$ as a function of $\mathbf{P}$, giving the second largest eigenvalue in terms of moduli (see [14]). Namely,

$$\begin{aligned} \overline{\lambda}_\star &= \boldsymbol{\lambda}_\star((1-p)\mathbf{P} + p\mathbf{I}) \leq (1-p)\boldsymbol{\lambda}_\star(\mathbf{P}) + p = (1-p)\lambda_\star + p \\ \Rightarrow \quad \overline{\gamma}_\star &= 1 - \overline{\lambda}_\star \geq 1 - (1-p)\lambda_\star - p = (1-p)(1-\lambda_\star) = (1-p)\gamma_\star. \end{aligned}$$

The assumption on M having holding probabilities of at least 0.5 is way more restrictive than the same assumption for $\overline{M}$ because it completely rules out the case where M is the embedded jump chain. The main problem with bounding $\gamma_\star$ from below by $\overline{\gamma}_\star$ without this assumption is the possibility of $\mathbf{P}$ having negative eigenvalues and $\lambda_\star = |\lambda_n|$. This can stand in any relation to $\overline{\lambda}_\star$, for example given by $(1-p)\lambda_2 + p$ in the special case $p_x \equiv p \geq 0.5$. With a trick, we can handle this problem by passing over to the $(2n)^{\text{th}}$ product $\mathbf{P}^{2n}$ for an appropriate n. This squared matrix has only positive eigenvalues so that the absolute spectral gap equals the spectral gap, which in turn can be related to the spectral gap of $\overline{\mathbf{P}}$ via Dirichlet forms. This relation can be pulled back to the original absolute spectral gap.

Lemma 6.3.6. *Let M be the embedded jump chain of a reversible, ergodic finite Markov chain $\overline{M}$ with stationary distribution $\overline{\pi}$ and $\overline{p}(x,x) = p \geq \frac{1}{2}$ for every $x \in \mathcal{S}$. Furthermore, let n be the well defined natural number*

$$n := \min\left\{m \in \mathbb{N} | p^{(2m)}(x,y) > 0 \text{ for every } x \neq y \text{ with } p(x,y) > 0\right\}$$

and denote by $\mathcal{E}^{(2n)}$ the Dirichlet form of $\mathbf{P}^{2n}$. Then

$$\gamma_\star \geq \frac{1}{2n(1-p)} \min_{x \neq y : p(x,y) > 0} \frac{p^{(2n)}(x,y)}{p(x,y)} \, \overline{\gamma}_\star.$$

Proof: In order to determine the Dirichlet form $\mathcal{E}^{(2n)}$, we have to identify the associated stationary distribution. Obviously, if $(\overline{\mathbf{P}}, \overline{\pi})$ is a reversible pair, so is $(\mathbf{P}, \overline{\pi})$, by Proposition 6.3.2, and therefore also $(\mathbf{P}^{2n}, \overline{\pi})$. Proceeding to the Dirichlet forms, define the system parameter

$$C := \min_{x \neq y : p(x,y) > 0} \frac{p^{(2n)}(x,y)}{p(x,y)},$$

which is positive due to the finiteness of $\mathcal{S}$ and the positivity of $p^{(2n)}(x,y)$ for all $x \neq y$ with $p(x,y) > 0$. With this, the Dirichlet form $\mathcal{E}^{(2n)}$ can be bounded via

$$\begin{aligned} \mathcal{E}^{(2n)}(f) &= \frac{1}{2} \sum_{x \neq y} (f(x) - f(y))^2 \overline{\pi}(x) p^{(2n)}(x,y) \\ &\geq \frac{1}{2} \sum_{x \neq y, \overline{p}(x,y) > 0} (f(x) - f(y))^2 \overline{\pi}(x) \overline{p}(x,y) \frac{p^{(2n)}(x,y)}{(1-p)p(x,y)} \\ &\geq \frac{1}{2} \sum_{x \neq y} (f(x) - f(y))^2 \overline{\pi}(x) \overline{p}(x,y) \frac{C}{1-p} \\ &= \overline{\mathcal{E}}(f) \frac{C}{1-p} \end{aligned} \tag{6.14}$$

for any $f : \mathcal{S} \to \mathbb{R}$. Thus, as before,

$$\gamma_\star^{(2n)} = \gamma^{(2n)} \geq \frac{C}{1-p} \, \overline{\gamma} = \frac{C}{1-p} \, \overline{\gamma}_\star$$

since $\mathbf{P}^{2n}$ and $\overline{\mathbf{P}}$ have only positive eigenvalues. It remains to relate $\gamma_\star^{(2n)}$ to $\gamma_\star$. For this we use the fact that $\lambda_\star^{(2n)} = \lambda_\star^{2n}$, and therefore

$$\gamma_\star \;=\; 1-\lambda_\star \;=\; \frac{1-\lambda_\star^{2n}}{\sum_{i=0}^{2n-1}\lambda_\star^i} \;\geq\; \frac{1}{2n}\gamma_\star^{(2n)} \;\geq\; \frac{C}{2n(1-p)}\overline{\gamma}_\star.$$

□

Using Remark 6.3.5 (a), we can extend Lemma 6.3.6 to a Markov chain $\overline{M}$ and its embedded jump chain M, where $\overline{p}(x,x)$ is not constant. We will reduce this problem to the just solved one by comparing $\overline{M}$ with a decelerated version $\widetilde{M}$ of M with constant holding probabilities $p \leq \min_x \overline{p}(x,x)$.

Theorem 6.3.7. *Let M be the embedded jump chain of a reversible finite chain $\overline{M}$ with stationary distribution $\overline{\pi}$ and $\overline{p}(x,x) \geq \frac{1}{2}$ for every $x \in \mathcal{S}$. Furthermore, define*

$$\begin{aligned} n \;&:=\; \min\left\{m \in \mathbb{N} | p^{(2m)}(x,y) > 0 \text{ for every } x \neq y \text{ with } p(x,y) > 0\right\}, \\ C \;&:=\; \min_{x \neq y : p(x,y) > 0} \frac{p^{(2n)}(x,y)}{p(x,y)}, \end{aligned}$$

and let $\mathcal{E}^{(2n)}$ be the Dirichlet form of $\mathbf{P}^{(2n)}$. Then, with $p := \min_x \overline{p}(x,x) \geq \frac{1}{2}$, it holds true that

$$\gamma_\star \;\geq\; \frac{C}{2n(1-p)}\overline{\gamma}_\star.$$

Proof: Let $\widetilde{M}$ be the decelerated version of M with holding probabilities p and $\widetilde{\gamma}_\star$ its absolute spectral gap. Then, by Lemma 6.3.6,

$$\gamma_\star \;\geq\; \frac{C}{2n(1-p)}\widetilde{\gamma}_\star. \tag{6.15}$$

With a look at the holding probabilities, we see that $p \leq \overline{p}(x,x)$ for every x, which ensures according to Remark 6.3.5 (a) that $\widetilde{\gamma}_\star \geq \overline{\gamma}_\star$. This leads in combination with (6.15) to

$$\gamma_\star \;\geq\; \frac{C}{2n(1-p)}\overline{\gamma}_\star.$$

□

The inequality may be strict when the constant C in the above theorem is too small. This is the case in Example 6.3.1 with $n = 3$ and $p = \frac{1}{2}$. There, $\gamma_\star = 1-\lambda_\star = \frac{1}{2}$, $\overline{\gamma}_\star = 1-\overline{\lambda}_\star = \frac{3}{4}$ and $C = \frac{1}{2}$, $\min\left\{m \in \mathbb{N} | p^{(2m)}(x,y) > 0 \text{ for every } x \neq y \text{ with } p(x,y) > 0\right\} = 1$. In this case Theorem 6.3.7 only ensures $\left(\frac{1}{2} =\right) \gamma_\star \geq \frac{\frac{1}{2}}{2 \cdot 1 \cdot \frac{1}{2}} \cdot \frac{3}{4} = \frac{3}{8}$.

6.4. Semi-Markovian Decelerated Chains

In this section, we study decelerated versions $M^T := \overline{M}$ not necessarily Markovian but still *semi-Markovian* (compare Definition 3.2.1). Note that here (unlike Definition 3.2.1) T_n is not the sojourn time in the state M_{n-1} but in M_n. Hence, we assume the sojourn times to be still conditionally independent given M and T_n to depend on M only via M_n and M_{n+1}.

6.4.1. Independent Sojourn Times

Since the above concepts lose their applicability, we start from scratch in the special case of sojourn times being independent of M, i.e. $(T_n)_{n\geq 0}$ is a family of i.i.d. random variables. We launch this subsection by verifying that with M also M^T has a stationary distribution to which it converges in total variation whenever $\mathbb{E}T_0 < \infty$. To this end we bound the total variation in terms of the tails of the number of jumps up to time n.

Definition 6.4.1. For a family $(T_n)_{n\geq 0}$ of random variables on $\mathbb{N}_{>0}$, define for any $n \geq 0$

$$N^T(n) \;:=\; \inf\{k \geq 0 | T_0 + ... + T_k > n\} \;=\; \sup\{k \geq 0 | T_0 + ... + T_{k-1} \leq n\}, \quad \sup\emptyset \;=\; 0.$$

If $M^T := \overline{M}$ is a by $(T_n)_{n\geq 0}$ decelerated version of a finite Markov chain M, than $N^T(n)$ is the *number of jumps of M^T up to time n.*

Although M^T is not Markovian, we are able to define a mixing time with respect to the limiting distribution π by

$$\tau^{M^T}_{\text{mix}}(\varepsilon) \;:=\; \inf\{n \geq 0 | \sup_x d_{TV}(\mathbb{P}^{M^T_k}_x, \pi) \leq \varepsilon \text{ for every } k \geq n\}, \quad \varepsilon > 0.$$

Lemma 6.4.2. *Let M be an ergodic finite Markov chain with respect to π, M^T a decelerated version with i.i.d. sojourn times independent of M, and $N^T(n)$ the number of jumps of M^T up to time $n \geq 0$. Then π is also stationary for M^T and*

(a) $\max_x d_{TV}(\mathbb{P}^{M^T_n}_x, \pi) \;\leq\; \max_x d_{TV}(\mathbb{P}^{M_K}_x, \pi) + \mathbb{P}_x(N^T(n) < K)$ *for every* $0 \leq K \leq n$.

(b) $\tau^{M^T}_{\text{mix}}(2\varepsilon) \leq \frac{\mathbb{E}(T_0)}{\varepsilon}\tau^{M}_{\text{mix}}(\varepsilon)$ *for every* $\varepsilon > 0$ *if* $\mathbb{E}(T_0) < \infty$.

Proof: First, we note that, by the assumed independence, the n-step transition probabilities of M^T are given by

$$\mathbb{P}_x\left(M^T_n = y\right) \;=\; \sum_{k=0}^{n} \mathbb{P}_x\left(M_k = y, N^T(n) = k\right) \;=\; \sum_{k=0}^{n} \mathbb{P}_x\left(M_k = y\right)\mathbb{P}\left(N^T(n) = k\right). \tag{6.16}$$

Therefore,

$$\mathbb{P}_\pi\left(M^T_n = y\right) \;=\; \sum_x \pi(x)\mathbb{P}_x\left(M^T_n = y\right) \;=\; \sum_x \sum_{k=0}^{n} \pi(x)\mathbb{P}_x\left(M_k = y\right)\mathbb{P}\left(N^T(n) = k\right) \;=\; \pi(y),$$

and thus, the stationarity of π.

(a) With this characterization of the transition probabilities, we observe

$$\begin{aligned}
\max_x d_{TV}(\mathbb{P}^{M^T_n}_x, \pi) \;&=\; \max_x \frac{1}{2}\sum_y \left|\mathbb{P}_x(M^T_n = y) - \pi(y)\right| \\
&=\; \max_x \frac{1}{2}\sum_y \left|\sum_{k=0}^{n} \mathbb{P}_x(M_k = y)\mathbb{P}(N^T(n) = k) - \pi(y)\right| \\
&=\; \max_x \frac{1}{2}\sum_y \left|\sum_{k=0}^{n} \left(\mathbb{P}_x(M_k = y) - \pi(y)\right)\mathbb{P}(N^T(n) = k)\right| \\
&\leq\; \max_x \frac{1}{2}\sum_y \sum_{k=0}^{n} \left|\mathbb{P}_x(M_k = y) - \pi(y)\right|\mathbb{P}(N^T(n) = k) \\
&=\; \sum_{k=0}^{n} \max_x d_{TV}(\mathbb{P}^{M_k}_x, \pi)\mathbb{P}(N^T(n) = k).
\end{aligned}$$

Finally noting that, for every $n \geq 0$,

$$d_{TV}(\mathbb{P}_x^{M_{n+1}}, \pi) \;\leq\; d_{TV}(\mathbb{P}_x^{M_n}, \pi),$$

we infer

$$\max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) \;\leq\; \mathbb{P}(N^T(n) < K) + \max_x d_{TV}(\mathbb{P}_x^{M_K}, \pi)$$

for every $K \leq n$.

(b) In particular, for $K = \tau_{\text{mix}}^M(\varepsilon) \leq n$ we find

$$\mathbb{P}(N^T(n) < K) \;=\; \mathbb{P}\left(T_0 + \ldots + T_{\tau_{\text{mix}}^M(\varepsilon)-1} > n\right) \;\leq\; \frac{\mathbb{E}\left(T_0 + \ldots + T_{\tau_{\text{mix}}^M(\varepsilon)-1}\right)}{n} \;\leq\; \frac{\tau_{\text{mix}}^M(\varepsilon)}{n}\mathbb{E}(T_0),$$

which is less than or equal to some $\varepsilon \in (0,1]$, if

$$n \;\geq\; \frac{\mathbb{E}(T_0)}{\varepsilon}\tau_{\text{mix}}^M(\varepsilon).$$

Thus, for such n,

$$\max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) \;\leq\; \mathbb{P}(N^T(n) < \tau_{\text{mix}}^M(\varepsilon)) + \max_x d_{TV}(\mathbb{P}_x^{M_{\tau_{\text{mix}}^M(\varepsilon)}}, \pi) \;\leq\; 2\varepsilon,$$

which finishes the proof. □

In particular, this proves that the total variation of $\mathbb{P}_x^{M_n^T}$ and π vanishes as n goes to ∞. This first bound is yet very rough and we can give a better one making use of the coupling method. For the existence of a maximal coupling of two random processes X and Y with $d_{TV}(\mathbb{P}_x^{X_n}, \mathbb{P}_y^{Y_n}) \to 0$ as $n \to \infty$ see [27, Theorem 3.3] or [28, Theorem 4]. If a maximal coupling is given by a bivariate process (X^1, X^2), then its *coupling time* is defined as

$$\inf\left\{n \geq 0 | X_m^1 = X_m^2 \text{ for every } m \geq n\right\}.$$

Proposition 6.4.3. *Let M be a reversible, ergodic finite Markov chain with respect to π and M^T a decelerated version with i.i.d. sojourn times independent of M. Then, for every $x \in \mathcal{S}$, there exists a maximal coupling of $\mathbb{P}_x^M$ and π^∞ with coupling time $T_{C(M),\max}^x$. Every coupling time $T_{C(M),\max}^x$, $x \in \mathcal{S}$, has a finite second moment and*

$$\tau_{\text{mix}}^{M^T}(\varepsilon) \;\leq\; \frac{\mathbb{E}(T_0)}{\varepsilon}\left(\tau_{\text{mix}}^M(\varepsilon) + \max_x \sqrt{\varepsilon \mathbb{E}(T_{C(M),\max}^x)^2}\right)$$

for every $\varepsilon > 0$.

Note that

$$\tau_{\text{mix}}^M(\varepsilon) + \max_x \sqrt{\varepsilon \mathbb{E}(T_{C(M),\max}^x)^2} \;=\; \tau_{\text{mix}}^M(\varepsilon) + o(1)$$

as $\varepsilon \to 0$.

Proof: Since M is ergodic, $\lambda_\star$ is bounded away from 1. Therefore, there exists $\varepsilon > 0$ with $\lambda_\star + \varepsilon < 1$ and $N \in \mathbb{N}$ such that

$$\left(\max_{x \in \mathcal{S}} d_{TV}(\mathbb{P}_x^{M_n}, \pi)\right)^{\frac{1}{n}} \;\leq\; \lambda_\star + \varepsilon$$

for every $n \geq N$. This further implies

$$\begin{aligned}\mathbb{E}((T^x_{C(M),\max})^2) &= \sum_{n\geq 0}(2n+1)\mathbb{P}(T^x_{C(M),\max} > n)\\ &= \sum_{n\geq 0}(2n+1)d_{TV}(\mathbb{P}^{M_n}_x, \pi)\\ &\leq \sum_{n=0}^{N-1}(2n+1)d_{TV}(\mathbb{P}^{M_n}_x, \pi) + \sum_{n\geq N}(2n+1)(\lambda_\star+\varepsilon)^n\\ &< \infty\end{aligned}$$

by the generalized ratio test. For the decelerated process, the bound

$$d_{TV}(\mathbb{P}^{M^T_n}_x, \pi) \leq \mathbb{P}(T^x_{C(M^T)} > n)$$

holds true for every coupling of $\mathbb{P}^{M^T}_x$ and π^∞ with coupling time $T^x_{C(M^T)}$. We choose the following one: the waiting times of the two marginals are exactly the same ones (distributed like T_0) and the coupling of the visited states is the maximal coupling of $\mathbb{P}^M_x$ and π^∞ above. More precisely, let $\mathbb{Q}_x$ be the maximal coupling of $\mathbb{P}^M_x$ and π^∞, that is,

$$\begin{aligned}\mathbb{Q}_x(\{(y_n)_{n\geq 0}\}\times\mathcal{S}^{\mathbb{N}}) &= \mathbb{P}_x(M=(y_n)_{n\geq 0})\\ \mathbb{Q}_x(\mathcal{S}^{\mathbb{N}}\times\{(z_n)_{n\geq 0}\}) &= \mathbb{P}_\pi(M=(z_n)_{n\geq 0})\end{aligned}$$

for each $\{(y_n)_{n\geq 0}\}$, $\{(z_n)_{n\geq 0}\} \in \mathcal{S}^{\mathbb{N}}$. Then the described coupling of $\mathbb{P}^{M^T}_x$ and π^∞ is a measure $\mathbb{Q}^T_x$ with

$$\mathbb{Q}^T_x\left(\{y\}\times\{z\}\right) = \mathbb{Q}_x\left(\{y'\}\times\{z'\}\right)\cdot\mathbb{P}\left((T_n)_{n\geq 0}=t\right)\mathbb{1}_{\{t=s\}}$$

for every

$$y=(\underbrace{y_0,...,y_0}_{t_0 \text{ times}},\underbrace{y_1,...,y_1}_{t_1 \text{ times}},...),\quad z=(\underbrace{z_0,...,z_0}_{s_0 \text{ times}},\underbrace{z_1,...,z_1}_{s_1 \text{ times}},...)$$

with induced jumps $y'=(y_0,y_1,...)$, $z'=(z_0,z_1,...)$ and sojourn times $t=(t_0,t_1,...)$ and $s=(s_0,s_1,...)$. Because of the assumed independence, $\mathbb{Q}^T_x$ satisfies

$$\mathbb{Q}^T_x(\{y\}\times\mathcal{S}^{\mathbb{N}}) = \mathbb{P}_x(M=y')\mathbb{P}((T_n)_{n\geq 0}=t) = \mathbb{P}_x(M_n=y_n, T_n=t_n, n\geq 0) = \mathbb{P}_x(M^T=y)$$

and

$$\mathbb{Q}^T_x(\mathcal{S}^{\mathbb{N}}\times\{z\}) = \mathbb{P}_\pi(M=z')\mathbb{P}((T_n)_{n\geq 0}=s) = \mathbb{P}_\pi(M_n=z_n, T_n=s_n, n\geq 0) = \mathbb{P}_\pi(M^T=z).$$

Thus, it is a coupling of $\mathbb{P}^{M^T}_x$ and π^∞. Its coupling time $T^x_{C(M^T)}$ is given by

$$T^x_{C(M^T)} = \sum_{i=0}^{T^x_{C(M),\max}-1} T_i \quad \text{a.s.} \tag{6.17}$$

Therefore, again using the independence, we can bound the total variation by

$$\begin{aligned} d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) &\leq \mathbb{P}\left(\sum_{i=0}^{T^x_{C(M),\max}-1} T_i > n\right) \\ &= \sum_{k\geq 1} \mathbb{P}\left(\sum_{i=0}^{k-1} T_i > n, T^x_{C(M),\max} = k\right) \\ &= \sum_{k\geq 1} \mathbb{P}\left(\sum_{i=0}^{k-1} T_i > n\right) \mathbb{P}(T^x_{C(M),\max} = k) \\ &\leq \frac{\mathbb{E}(T_0)}{n} \sum_{k\geq 1} k\mathbb{P}(T^x_{C(M),\max} = k) \\ &= \frac{\mathbb{E}(T_0)}{n} \mathbb{E}(T^x_{C(M),\max}). \end{aligned}$$

Written in terms of the mixing time, this gives

$$\tau_{\text{mix}}^{M^T}(\varepsilon) \leq \frac{\mathbb{E}(T_0)}{\varepsilon} \max_x \mathbb{E}(T^x_{C(M),\max}). \tag{6.18}$$

The remainder of the proof is to express the right-hand side of Equation (6.18), that is, the moments of $T^x_{C(M),\max}$, in terms of the mixing time of M. By the formulas derived in [51, Equation (10)] and already used in Section 2.2, we can bound the tail probabilities of the coupling time for $n \leq \mathbb{E}(T^x_{C(M),\max})$ by

$$d_{TV}(\mathbb{P}_x^{M_n}, \pi) = \mathbb{P}(T^x_{C(M),\max} > n) \geq \frac{\left(\mathbb{E}T^x_{C(M),\max} - n\right)^2}{\mathbb{E}(T^x_{C(M),\max})^2},$$

which is larger than some $\varepsilon > 0$ if

$$\mathbb{E}(T^x_{C(M),\max}) - \sqrt{\varepsilon \mathbb{E}(T^x_{C(M),\max})^2} \geq n.$$

This yields

$$\tau_{\text{mix}}^{M}(\varepsilon) \geq \mathbb{E}(T^x_{C(M),\max}) - \sqrt{\varepsilon \mathbb{E}(T^x_{C(M),\max})^2}$$

for every $\varepsilon > 0$ and every x. Via insertion in equation (6.18), we obtain

$$\tau_{\text{mix}}^{M^T}(\varepsilon) \leq \frac{\mathbb{E}(T_0)}{\varepsilon} \max_x \mathbb{E}(T^x_{C(M),\max}) \leq \frac{\mathbb{E}(T_0)}{\varepsilon} \left(\tau_{\text{mix}}^{M}(\varepsilon) + \max_x \sqrt{\varepsilon \mathbb{E}(T^x_{C(M),\max})^2}\right)$$

for every $\varepsilon > 0$. □

The above proof can not be generalized to waiting times depending on M due to the unavailability of this specific coupling. When the waiting times of the two marginals have different distributions, the jumps occur at different times. Equation (6.17) (or a sufficient version with "$\leq$" instead of "$=$"), which is the main ingredient of the proof, would no longer be valid. A typical trajectory of the coupling process (X, X') would look like in Figure 6.2 where the embedded jump chain couples at time 1, but the decelerated version couples at time $s_0 + s_1$, which is larger than both sojourn times in the initial states.

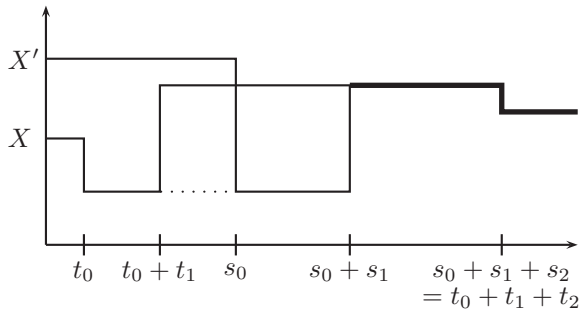

Figure 6.2.: Typical trajectory of a coupling process (X, X') with different sojourn times (t_0, t_1,... and s_0, s_1,...) in the two marginals.

Nor did we find a corresponding coupling for the other direction, that is, starting with a maximal coupling of $\mathbb{P}_x^{M^T}$ and π^∞ with coupling time $T_{C(M^T),\max}$, a coupling of $\mathbb{P}_x^M$ and π^∞ with coupling time $T^x_{C(M)}$ that satisfies

$$\mathbb{E}\, T^x_{C(M)} \cdot \mathbb{E}\, T_0 \;\leq\; \mathbb{E}\, T^x_{C(M^T),\max},$$

the integrated version of Equation (6.17).

As mentioned before, the concepts of the previous section can not be applied without further ado. Nevertheless, the spectrum of $(\mathbb{P}_x(M_n^T = y))_{x,y}$ is still linked with the mixing time as we will see. Again, we restrict ourselves to embedded processes with $\lambda_\star = \lambda_2 > 0$ for the same reasons as before, once more illustrated by an example later on.

Proposition 6.4.4. *Let M be an ergodic, reversible finite Markov chain with respect to π that satisfies $\lambda_\star = \lambda_2 > 0$ and M^T a decelerated version with i.i.d. sojourn times independent of M. Then $\pi(x)\mathbb{P}_x(M_n^T = y) = \pi(y)\mathbb{P}_y(M_n^T = x)$ for every x, y and each $n \geq 0$. Moreover, define*

$$\lambda_\star^{T,n} \;:=\; \max\left\{|\lambda| < 1 | \lambda \text{ is eigenvalue of } (\mathbb{P}_x(M_n^T = y))_{x,y}\right\}.$$

(a) For every $n \geq 0$ with $\mathbb{P}(T_0 \leq n) > 0$,

$$\frac{\lambda_\star^{T,n}}{2} \;\leq\; \max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) \;\leq\; \frac{\lambda_\star^{T,n}}{\pi_{min}},$$

where $\pi_{min} := \min_x \pi(x)$. That is, if $\lim_{n\to\infty} \sqrt[n]{\lambda_\star^{T,n}}$ exists and is less than 1, then it is the geometric rate of convergence of the law of M_n^T towards its stationary distribution as $n \to \infty$.

(b) For every $n \geq 0$,

$$\lambda_\star^{T,n} \;=\; \mathbb{E}(\lambda_\star^{N^T(n)}) \;\geq\; \lambda_\star^n.$$

That is, $\sqrt[n]{\lambda_\star^{T,n}} \geq \lambda_\star$, which is the rate of convergence of the law of M towards its stationary distribution.

Proof: The reversibility is an easy consequence of equation (6.16), namely

$$\begin{aligned}\pi(x)\mathbb{P}_x(M_n^T = y) &= \sum_{k=0}^{n} \pi(x)\mathbb{P}_x(M_k = y)\mathbb{P}(N^T(n) = k) \\ &= \sum_{k=0}^{n} \pi(y)\mathbb{P}_y(M_k = x)\mathbb{P}(N^T(n) = k) \\ &= \pi(y)\mathbb{P}_y(M_n^T = x).\end{aligned}$$

(a) Fix $n \geq 0$ with $\mathbb{P}(T_0 \leq n) > 0$ and define the transition matrix

$$\mathbf{Q} := (\mathbb{P}_x(M_n^T = y))_{x,y}.$$

It is reversible with respect to π and has second largest eigenvalue (in modulus) $\lambda_\star^{T,n}$. Furthermore, it is irreducible and aperiodic. For this to see, choose, first, $k \in \mathbb{N}$ with $\mathbb{P}_x(M_{k'} = y) > 0$ for every $k' \geq k$ and any x, y (which exists by the ergodicity of M) and, second, $m := \max\{m' \leq n | \mathbb{P}(T_0 = m') > 0\}$. Then $n = k''m + l$ for some $k'' \geq 1$ and $l < m$, and

$$\begin{aligned}\mathbf{Q}^{(k')}(x,y) &= \sum_{x_1,\ldots,x_{k'-1}} \mathbb{P}_x(M_{k''m+l}^T = x_1)\mathbb{P}_{x_1}(M_{k''m+l}^T = x_2) \cdot \ldots \cdot \mathbb{P}_{x_{k'-1}}(M_{k''m+l}^T = y) \\ &\geq \sum_{x_1,\ldots,x_{k'-1}} \mathbb{P}_x(M_{k''} = x_1)\mathbb{P}_{x_1}(M_{k''} = x_2) \cdot \ldots \cdot \mathbb{P}_{x_{k'-1}}(M_{k''} = y) \cdot \mathbb{P}(T_0 = m)^{k'(k''+1)} \\ &\geq \mathbb{P}_x(M_{k'k''} = y)\mathbb{P}(T_0 = m)^{k'(k''+1)} \\ &> 0\end{aligned}$$

for every $k' \geq k$ and any x, y. Hence, any pair x, y is connected and returns to any x are possible at times $k, k+1,\ldots$, ensuring the ergodicity. Now part (a) is a consequence of results from Chapter 12.2 in [39]. There,

$$\left|\frac{\mathbf{P}(x,y)}{\rho(y)} - 1\right| \leq \frac{\lambda_\star}{\rho_{min}} \quad \text{and} \quad \lambda_\star \leq \max_x \sum_y |\mathbf{P}(x,y) - \rho(y)|$$

is proved for every reversible and ergodic finite Markov chain with transition matrix $\mathbf{P}$, stationary distribution ρ, second largest eigenvalue $\lambda_\star$ (in modulus) and each x, y. Applying this to $\mathbf{Q}$, the first assertion follows by

$$d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) = \sum_{y:\mathbf{Q}(x,y)>\pi(y)} \pi(y)\left(\frac{\mathbf{Q}(x,y)}{\pi(y)} - 1\right) \leq \max_y \left|\frac{\mathbf{Q}(x,y)}{\pi(y)} - 1\right| \leq \frac{\lambda_\star^{T,n}}{\pi_{min}}$$

for any $x \in \mathcal{S}$, and the second by

$$\frac{\lambda_\star^{T,n}}{2} \leq \frac{1}{2}\max_x \sum_y |\mathbf{Q}(x,y) - \pi(y)| = \max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi).$$

For part (b), we have seen in (6.16) that the matrix $(\mathbb{P}_x(M_n^T = y))_{x,y}$ can be written as a weighted sum of powers of $\mathbf{P}$ with $(\mathbb{P}(N^T(n) = k)_{0 \leq k \leq n})$ being the weights. Thus, if λ is an eigenvalue of $\mathbf{P}$, then

$$\sum_{k=0}^{n} \mathbb{P}(N^T(n) = k)\lambda^k = \mathbb{E}(\lambda^{N^T(n)})$$

is an eigenvalue of $(\mathbb{P}_x(M_n^T = y))_{x,y}$. By assumption, $\lambda_\star = \lambda_2 > 0$, and therefore the second largest eigenvalue of this matrix is $\mathbb{E}(\lambda_\star^{N^T(n)})$, which is larger than $\lambda_\star^n$ since $N^T(n) \leq n$ a.s. □

For the special case of almost surely constant sojourn times we obtain:

Corollary 6.4.5. *Let M be an ergodic, reversible finite Markov chain with respect to π that satisfies $\lambda_\star = \lambda_2 > 0$ and M^T a decelerated version with sojourn times $T_n \equiv k$ a.s. for every $n \geq 0$ and some $k \in \mathbb{N}$. Then $N^T(n) = \lfloor \frac{n}{k} \rfloor$ a.s., and therefore $\lambda_\star^{T,n} = \lambda_\star^{\lfloor n/k \rfloor}$ and*

$$\left(\max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi)\right)^{1/n} \overset{n\to\infty}{\longrightarrow} \lambda_\star^{1/k}.$$

So in fact, just as in Section 6.3, the mixing performance of the decelerated process is decelerated when $\lambda_\star = \lambda_2$, for example, if M is lazy. Compared with Example 6.3.1, we have good *general* insight in this necessary condition here, as without it we would lose control of $\lambda_\star^{T,n} = \max_{\lambda \neq 1} \left|\mathbb{E}(\lambda^{N^T(n)})\right|$.

Example 6.4.6. We take another look at the Markovian case. The sojourn times are independent $Geo(1-p)$ variables, so that $T_0 + \ldots + T_k - (k+1) \sim NB(k+1, 1-p)$, $k \geq 1$. This yields

$$\mathbb{P}(N^T(n) = 0) \;=\; \mathbb{P}(T_0 > n) \;=\; p^n,$$

and for $1 \leq k \leq n$,

$$\begin{aligned}
\mathbb{P}(N^T(n) = k) &= \mathbb{P}(T_0 + \ldots + T_{k-1} \leq n < T_0 + \ldots + T_k) \\
&= \sum_{i=k}^{n} \mathbb{P}(T_0 + \ldots + T_{k-1} = i, T_k > n - i) \\
&= \sum_{i=k}^{n} \mathbb{P}(T_0 + \ldots + T_{k-1} = i)\mathbb{P}(T_k > n - i) \\
&= \sum_{i=k}^{n} \binom{i-1}{i-k} (1-p)^k p^{i-k} p^{n-i} \\
&= \sum_{i=0}^{n-k} \binom{i+k-1}{i} (1-p)^k p^{n-k} \\
&= \binom{n}{k} (1-p)^k p^{n-k},
\end{aligned}$$

where the last equality can be proved via induction for $n \geq k$. This proves $N^T(n) \sim B(n, 1-p)$, with moment generating function

$$\mathbb{E}(\lambda^{N^T(n)}) \;=\; (\lambda(1-p) + p)^n.$$

Therefore, $\overline{\lambda}_\star = \sqrt[n]{\lambda_\star^{T,n}} = \max_{\lambda \neq 1}(\lambda(1-p) + p)$, where the maximum ranges over the eigenvalues of M. As in Example 6.3.1, we observe that, given $\lambda_\star$, p has to be suitably chosen to for $\overline{\lambda}_\star \geq \lambda_\star$.

In order to give bounds on the mixing time, we have to bound the limit of the rates $\sqrt[n]{\lambda_\star^{T,n}}$, the existence of which is not clear in general. The above example disproves this rates to converge against $\lambda_\star^{1/\mathbb{E}(T_0)}$ in general.

Proposition 6.4.7. *Let M be an ergodic, reversible finite Markov chain with respect to π that satisfies $\lambda_\star = \lambda_2 > 0$ and M^T a decelerated version with i.i.d. and integrable sojourn times independent of M.*

(a) It holds true that

$$\liminf_{n\to\infty} \sqrt[n]{\lambda_\star^{T,n}} \;\geq\; \lambda_\star^{\frac{1}{\mathbb{E}(T_0)}} \quad \textit{and} \quad \liminf_{\varepsilon\to 0} \frac{\tau_{\text{mix}}^{M^T}(\varepsilon)}{\tau_{\text{mix}}^{M}(\varepsilon)} \;\geq\; \mathbb{E}(T_0).$$

(b) If $\mathbb{E}(a^{T_0}) < \infty$ for some $a > 1$ with $\mathbb{E}(a^{T_0 - \mathbb{E}(T_0)/(1-\delta)}) < \lambda_\star$ for some $\delta \in (0,1)$, then

$$\limsup_{n\to\infty} \sqrt[n]{\lambda_\star^{T,n}} \;\leq\; \lambda_\star^{(1-\delta)/\mathbb{E}(T_0)} \quad \textit{and} \quad \limsup_{\varepsilon\to 0} \frac{\tau_{\text{mix}}^{M^T}(\varepsilon)}{\tau_{\text{mix}}^{M}(\varepsilon)} \;\leq\; \frac{\mathbb{E}(T_0)}{1-\delta}.$$

Proof: We note for both parts that, by Proposition A.2.1 (b),

$$\lambda_\star^{1+\delta(N)} \;\leq\; \max_x d_{TV}(\mathbb{P}_x^{M_n}, \pi)^{\frac{1}{n}} \;\leq\; \lambda_\star^{1-\delta(N)}$$

for any $n \geq N$ and some $\delta(N)$ such that $\delta(N) \to 0$ as $N \to \infty$. Thus, for $n \geq N$,

$$\begin{aligned}
\max_x d_{TV}(\mathbb{P}_x^{M_n}, \pi) \;&\leq\; \lambda_\star^{n(1-\delta(N))} \;\leq\; \varepsilon \quad \text{if} \quad n \;\geq\; \frac{\ln(\varepsilon)}{\ln(\lambda_\star)} \cdot \frac{1}{1-\delta(N)}, \\
\max_x d_{TV}(\mathbb{P}_x^{M_n}, \pi) \;&\geq\; \lambda_\star^{n(1+\delta(N))} \;\geq\; \varepsilon \quad \text{if} \quad n \;\leq\; \frac{\ln(\varepsilon)}{\ln(\lambda_\star)} \cdot \frac{1}{1+\delta(N)}.
\end{aligned}$$

Note that for any $\varepsilon > 0$ there is some $N_1 := N_1(\varepsilon)$ such that

$$\frac{\ln(\varepsilon)}{\ln(\lambda_\star)} \cdot \frac{1}{(1+\delta(N_1))} \;\geq\; N_1$$

and $N_1(\varepsilon) \to \infty$ as $\varepsilon \to 0$. This yields for $\varepsilon \to 0$

$$\tau_{mix}^{M}(\varepsilon) \;\begin{cases} \leq \; \dfrac{\ln(\varepsilon)}{\ln(\lambda_\star)} \cdot \dfrac{1}{1-\delta(N_1(\varepsilon))} \\ \geq \; \dfrac{\ln(\varepsilon)}{\ln(\lambda_\star)} \cdot \dfrac{1}{1+\delta(N_1(\varepsilon))} \end{cases} \;=\; \frac{\ln(\varepsilon)}{\ln(\lambda_\star)} \cdot \frac{1}{1+o(1)}. \tag{6.19}$$

Turning to (a), it is known from renewal theory that

$$\begin{aligned}
\frac{n}{N^T(n)} \;&\leq\; \frac{T_0 + \ldots + T_{N^T(n)}}{N^T(n)} \;\to\; \mathbb{E}(T_0) \quad \text{a.s.} \\
\frac{n}{N^T(n)} \;&\geq\; \frac{T_0 + \ldots + T_{N^T(n)-1}}{N^T(n)} \;\to\; \mathbb{E}(T_0) \quad \text{a.s.},
\end{aligned}$$

and thus $\frac{1}{n} N^T(n) \to \mathbb{E}(T_0)^{-1}$ a.s. as $n \to \infty$. Using this, a lower bound on $\liminf_{n\to\infty} \sqrt[n]{\lambda_\star^{T,n}}$ is given by an application of Jensen's Inequality and Fatou's Lemma, viz.

$$\liminf_{n\to\infty} \sqrt[n]{\lambda_\star^{T,n}} \;=\; \liminf_{n\to\infty} \sqrt[n]{\mathbb{E}(\lambda_\star^{N^T(n)})} \;\geq\; \liminf_{n\to\infty} \mathbb{E}(\lambda_\star^{N^T(n)/n}) \;\geq\; \lambda_\star^{\frac{1}{\mathbb{E}(T_0)}} \;>\; 0.$$

Now, for any $\varepsilon > 0$,

$$\max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) \geq \frac{1}{2}\lambda_\star^{T,n} > \varepsilon \quad \text{if} \quad \mathbb{P}(T_0 \leq n) > 0 \text{ and } n < \frac{\ln(2\varepsilon)}{\ln\left(\sqrt[n]{\lambda_\star^{T,n}}\right)}$$

by Proposition 6.4.4. Therefore, denote by $N_2 := N_2(\varepsilon)$ the smallest $n \in \mathbb{N}$ violating this, that is,

$$N_2 \;:=\; \min\left\{ n \geq 1 \;\middle|\; \mathbb{P}(T_0 \leq n) > 0, \text{ and } n \geq \frac{\ln(2\varepsilon)}{\ln\left(\sqrt[n]{\lambda_\star^{T,n}}\right)} \right\},$$

which is finite since $\max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) \to 0$ as $n \to \infty$. Choose $\delta := \delta(\varepsilon)$ large enough to ensure $\sqrt[N_2]{\lambda_\star^{T,N_2}} \geq \lambda_\star^{(1+\delta)/\mathbb{E}(T_0)}$, and observe that $\delta(\varepsilon)$ tends to 0 as ε tends to 0. Then

$$\begin{aligned}\tau_{\text{mix}}^{M^T}(\varepsilon) \;\geq\; N_2 \;&\geq\; \frac{\ln(2\varepsilon)}{\ln\left(\sqrt[N_2]{\lambda_\star^{T,N_2}}\right)} \;\geq\; \frac{\ln(2\varepsilon)}{\ln\left(\lambda_\star^{\frac{1+\delta}{\mathbb{E}(T_0)}}\right)} \;=\; \frac{\mathbb{E}(T_0)}{1+\delta} \cdot \frac{\ln(2\varepsilon)}{\ln(\lambda_\star)} \\ &=\; \frac{\mathbb{E}(T_0)}{1+\delta}\tau_{\text{mix}}^{M}(\varepsilon)(1+o(1)) + \mathcal{O}(1)\end{aligned}$$

and therefore,

$$\liminf_{\varepsilon \to 0} \frac{\tau_{\text{mix}}^{M^T}(\varepsilon)}{\tau_{\text{mix}}^{M}(\varepsilon)} \;\geq\; \mathbb{E}(T_0).$$

Part (b) can be proved by giving an upper bound for the limes superior of the sequence of roots of eigenvalues. But first we prove the following technical assertion which we will need thereafter: If $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ are two positive sequences with $\limsup_{n\to\infty} \sqrt[n]{a_n} \leq a, \limsup_{n\to\infty} \sqrt[n]{b_n} \leq a$, then $\limsup_{n\to\infty} \sqrt[n]{a_n + b_n} \leq a$. To see this, take some arbitrary $\varepsilon > 0$ and $N \in \mathbb{N}$ large enough to ensure $\sqrt[n]{a_n} \leq (1+\varepsilon)a$ and $\sqrt[n]{b_n} \leq (1+\varepsilon)a$ for any $n \geq N$. Then it holds true for $n \geq N$ that

$$\sqrt[n]{a_n + b_n} \;\leq\; \sqrt[n]{2}(1+\varepsilon)a \;\to\; (1+\varepsilon)a$$

as $n \to \infty$. For the lower bound, take $\varepsilon < 0$.

Writing

$$C(n) \;:=\; \left\lfloor \frac{n(1-\delta)}{\mathbb{E}(T_0)} \right\rfloor,$$

under the given assumptions, we obtain with the exponential Chebychev Inequality

$$\mathbb{P}\left(T_1 + ... + T_{C(n)} > n\right) \;\leq\; \frac{\mathbb{E}\left(a^{T_1+...+T_{C(n)}}\right)}{a^n} \;\leq\; \mathbb{E}\left(a^{T_0}\right)^{C(n)} a^{-n}$$

and

$$\begin{aligned}\sqrt[n]{\mathbb{P}\left(T_1 + ... + T_{C(n)} > n\right)} \;&\leq\; \mathbb{E}\left(a^{T_0}\right)^{\frac{1}{n}C(n)} a^{-1} \\ &\xrightarrow{n\to\infty}\; \mathbb{E}\left(a^{T_0}\right)^{(1-\delta)/\mathbb{E}(T_0)} a^{-1} \\ &=\; \mathbb{E}\left(a^{T_0 - \mathbb{E}(T_0)/(1-\delta)}\right)^{(1-\delta)/\mathbb{E}(T_0)} \\ &\leq \lambda_\star^{(1-\delta)/\mathbb{E}(T_0)}.\end{aligned}$$

Furthermore, $\limsup_{n\to\infty} \lambda_\star^{C(n)/n} = \lambda_\star^{(1-\delta)/\mathbb{E}(T_0)}$. Thus, using the special structure of $N^T(n)$, we

obtain

$$\begin{aligned}
\limsup_{n\to\infty} \sqrt[n]{\mathbb{E}\left(\lambda_\star^{N^T(n)}\right)} &= \limsup_{n\to\infty} \sqrt[n]{\mathbb{E}\left(\lambda_\star^{N^T(n)}\mathbb{1}_{\{N^T(n)<C(n)\}} + \lambda_\star^{N^T(n)}\mathbb{1}_{\{N^T(n)\geq C(n)\}}\right)} \\
&\leq \limsup_{n\to\infty} \sqrt[n]{\mathbb{P}\left(N^T(n) < C(n)\right) + \lambda_\star^{C(n)}} \\
&= \limsup_{n\to\infty} \sqrt[n]{\mathbb{P}\left(T_1 + \ldots + T_{C(n)} > n\right) + \lambda_\star^{C(n)}} \\
&= \lambda_\star^{(1-\delta)/\mathbb{E}(T_0)}.
\end{aligned}$$

Similar to part (a), for any given $\varepsilon > 0$,

$$\max_x d_{TV}(\mathbb{P}_x^{M_n^T}, \pi) \leq \frac{\lambda_\star^{T,n}}{\pi_{min}} < \varepsilon \quad \text{if} \quad \mathbb{P}(T_0 \leq n) > 0 \text{ and } n > \frac{\ln(\varepsilon\pi_{min})}{\ln\left(\sqrt[n]{\lambda_\star^{T,n}}\right)}.$$

Therefore, denote now by $N_3 := N_3(\varepsilon)$ the largest $n \in \mathbb{N}$ violating this inequality, that is,

$$N_3 \;:=\; \max\left\{ n \geq 1 \;\middle|\; \mathbb{P}(T_0 \leq n) > 0, \text{ and } n \leq \frac{\ln(\pi_{min}\varepsilon)}{\ln\left(\sqrt[n]{\lambda_\star^{T,n}}\right)} \right\}.$$

Since $\sqrt[n]{\lambda_\star^{T,n}}$ is bounded away form 1 and 0 as $n \to \infty$, the above set is nonempty for sufficiently small ε and the maximum exists. Choose $\delta_2 := \delta_2(\varepsilon)$ large enough to ensure $\sqrt[N_3]{\lambda_\star^{T,N_3}} \leq \lambda_\star^{(1-\delta-\delta_2)/\mathbb{E}(T_0)}$, and observe that $\delta_2(\varepsilon)$ tends to 0 as ε tends to 0. Then

$$\begin{aligned}
\tau_{\text{mix}}^{M^T}(\varepsilon) \;\leq\; N_3 + 1 \;\leq\; \frac{\ln(\pi_{min}\varepsilon)}{\ln\left(\sqrt[N_3]{\lambda_\star^{T,N_3}}\right)} + 1 \;&\leq\; \frac{\ln(\pi_{min}\varepsilon)}{\ln\left(\lambda_\star^{(1-\delta-\delta_2)/\mathbb{E}(T_0)}\right)} + 1 \\
&= \frac{\mathbb{E}(T_0)}{1-\delta-\delta_2}\,\frac{\ln(\pi_{min}\varepsilon)}{\ln(\lambda_\star)} + 1 \\
&= \frac{\mathbb{E}(T_0)}{1-\delta-\delta_2}\,\tau_{\text{mix}}^{M}(\varepsilon)(1+o(1)) + \mathcal{O}(1),
\end{aligned}$$

and therefore, proving the remaining part of (b),

$$\limsup_{\varepsilon\to 0} \frac{\tau_{\text{mix}}^{M^T}(\varepsilon)}{\tau_{\text{mix}}^{M}(\varepsilon)} \;\leq\; \frac{\mathbb{E}(T_0)}{1-\delta}.$$

□

Again regarding the special case of almost surely constant sojourn times equal to $k \in \mathbb{N}$, we find with the same arguments that

$$\lim_{\varepsilon\to 0} \frac{\tau_{\text{mix}}^{M^T}(\varepsilon)}{\tau_{\text{mix}}^{M}(\varepsilon)} \;=\; k.$$

If the condition $\lambda_\star = \lambda_2 > 0$ is violated, one can still hope for M^T being also a decelerated version of some $\widetilde{M}$, where $\widetilde{M}$ is a decelerated version of M and Markovian, satisfying the condition above - for example $\widetilde{M}$ being a lazy version of M. Then a combination of the above proposition and Theorem 6.3.7 provides a relation between the mixing times of M and M^T.

In other cases, a simplification may be achieved by changing the family of sojourn times for the deceleration. How this affects the spectral gap is content of the following easy corollary.

Corollary 6.4.8. *Let M be an ergodic, reversible finite Markov chain with respect to π that satisfies $\lambda_\star = \lambda_2 > 0$. If $(T_n)_{n\geq 0}$ and $(S_n)_{n\geq 0}$ are two families of waiting times, both independent of M with $T_0 \leq_{st} S_0$, then*

$$\lambda_\star^{T,n} \;\leq\; \lambda_\star^{S,n}.$$

Thus, the process decelerated by the stochastically smaller waiting times mixes faster than the process decelerated by the stochastically larger waiting time.

Proof: Via standard calculations (see e.g. Corollary A.2 in [2]) we obtain

$$\begin{aligned}
\lambda_\star^{T,n} &= \mathbb{E}(\lambda_\star^{N^T(n)}) \\
&= 1-(1-\lambda_\star)\sum_{k\geq 0}\lambda_\star^k \mathbb{P}(N^T(n)>k) \\
&= 1-(1-\lambda_\star)\sum_{k\geq 0}\lambda_\star^k \mathbb{P}(T_0+...+T_k\leq n) \\
&\leq 1-(1-\lambda_\star)\sum_{k\geq 0}\lambda_\star^k \mathbb{P}(S_0+...+S_k\leq n) \\
&= \lambda_\star^{S,n}.
\end{aligned}$$

□

6.4.2. General Sojourn Times

Having studied sojourn times independent of the jump chain to some extent so far, we want to proceed with the more general case when the sojourn times depend on M in such a way that M^T is a semi-Markov chain. Providing M is ergodic, M^T is aperiodic and the sojourn times have finite moments, there is still a convergence to a distribution π^T,

$$\mathbb{P}_x\left(M_n^T = y\right) \overset{n\to\infty}{\longrightarrow} \pi^T(y) \;=\; \frac{\pi(y)\mathbb{E}_y(T_0)}{\mathbb{E}_\pi(T_0)}, \tag{6.20}$$

where $\mathbb{E}_\pi(T_0) = \sum_{x\in\mathcal{S}}\pi(x)\mathbb{E}_x(T_0)$ is the normalization constant (see [6, Chapter 4]). This limiting distribution is no longer stationary, that is,

$$\mathbb{P}_{\pi^T}^{M_n^T} \;\neq\; \pi^T,\; n\geq 1,$$

in general. Es a simple example, regard the random walk M on $\{1,2,3\}$ without loops and sojourn times $T_i \equiv i$ for the states $i\in\{1,2,3\}$. But, as $(M_n^T, T_0+...+T_{N^T(n)}-n)_{n\geq 0}$ is a homogeneous Markov chain with stationary distribution $\mu(x,s) = \pi^T(x)\mathbb{P}_x(T_0\geq s)/\mathbb{E}_x(T_0)$, $x\in\mathcal{S}$, $s\geq 1$ (see Lemma A.3.2),

$$\mathbb{P}_\mu^{M_n^T} \;=\; \pi^T \quad \text{for every } n\geq 0.$$

This reveals that stationarity can only be achieved when specifying also the *residual sojourn times* $T_0+...+T_{N^T(n)}-n$. The distribution of this residual sojourn time must equal the biased distribution of $T_{N^T(n)}$.

For every fixed y, the convergence in Equation 6.20 is a consequence of Blackwell's renewal theorem for delayed renewal processes (see [1, Theorem 2.4.1]), applied to the renewal process which counts the occurrences of the state $y\in\mathcal{S}$ (resp. the Markov Renewal Theorem [1, Theorem 9.2.6]). In [41, Chapter II. (p. 30)], Lindvall argues that the convergence in Blackwell's renewal theorem is exponentially fast with rate at most $\rho<1$ if for the i.i.d. renewal times $(S_n)_{n\geq 1}$, the delay waiting

time S_0 and a random variable S_0' with the biased distribution $\mathbb{P}(S_0' = k) = \frac{1}{\mathbb{E}(S_1)}\mathbb{P}(S_1 > k)$, $k \geq 0$, the respective exponential moments

$$\mathbb{E}\left(\rho^{-S_1}\right), \quad \mathbb{E}\left(\rho^{-S_0}\right) \quad \text{and} \quad \mathbb{E}\left(\rho^{-S_0'}\right)$$

are finite. For the specific delayed renewal process used for the convergence of the distribution of M^T, the i.i.d. renewal times are given by the recurrence times of the fixed state y and the delay waiting time equals the time needed to reach y from the initial state x. Therefore, we conclude the following conditions for geometric ergodicity of M^T.

Proposition 6.4.9. *Let M^T be an irreducible and aperiodic semi-Markov chain with ergodic embedded Markov chain M. For every $x \in \mathcal{S}$ define*

$$\tau_x \ := \ \inf\{n \geq 1 | M_n = x\} \quad \textit{and} \quad \tau_x^T \ := \ T_0 + ... + T_{\tau_x - 1}.$$

If $\max_{r,s} \mathbb{E}_r\left(\rho^{-T_0}|M_1 = s\right) < \infty$ for some $\rho < 1$ and $\mathbb{E}_x\left(\max_{r,s}\mathbb{E}_r(\rho^{-T_0}|M_1 = s)^{\tau_y}\right) < \infty$ for every $x, y \in \mathcal{S}$, then

$$\mathbb{P}_x\left(M_n^T = y\right) - \pi^T(y) \ = \ o(\rho^n)$$

for every $x, y \in \mathcal{S}$.

Proof: Fix $x, y \in \mathcal{S}$ and let θ_y^T be a random variable on $\mathbb{N}_0$ with the size biased distribution of τ_y^T that is given by $\mathbb{P}(\theta_y^T = k) = \mathbb{P}(\tau_y^T > k)/\mathbb{E}(\tau_y^T)$, $k \geq 0$. In view of the previous remarks, it remains to show that the hitting times τ_y^T under $\mathbb{P}_x$ (delay), the hitting times τ_y^T under $\mathbb{P}_y$ (renewal times), and the biased version θ_y^T under $\mathbb{P}_y$ have finite exponential moments, that is,

$$\mathbb{E}_x\left(\rho^{-\tau_y^T}\right) \vee \mathbb{E}_y\left(\rho^{-\tau_y^T}\right) \ < \ \infty \quad \text{and} \quad \mathbb{E}_y\left(\rho^{-\theta_y^T}\right) \ = \ \sum_{k\geq 0} \rho^{-k} \frac{1}{\mathbb{E}_y(\tau_y^T)} \mathbb{P}_y\left(\tau_y^T > k\right) \ < \ \infty.$$

Assuming for the moment the existence of the exponential moment of τ_y^T under $\mathbb{P}_y$, its first moment is finite as well. Furthermore,

$$\sum_{l\geq 1}\sum_{k=0}^{l-1} \rho^{-k}\mathbb{P}_y\left(\tau_y^T = l\right) \ = \ \sum_{l\geq 1} \frac{\rho^{-l}-1}{\rho^{-1}-1}\mathbb{P}_y\left(\tau_y^T = l\right)$$

possesses the summable majorant $f(l) = \rho^{-l}\mathbb{P}_y\left(\tau_y^T = l\right)/(\rho^{-1}-1)$, as

$$\sum_{l\geq 1} \frac{\rho^{-l}}{\rho^{-1}-1}\mathbb{P}_y\left(\tau_y^T = l\right) \ = \ \frac{1}{\rho^{-1}-1}\mathbb{E}_y\left(\rho^{-\tau_y^T}\right) \ < \ \infty.$$

We conclude that

$$\infty \ > \ \sum_{l\geq 1}\sum_{k=0}^{l-1} \rho^{-k}\mathbb{P}_y\left(\tau_y^T = l\right) \ = \ \sum_{k\geq 0}\sum_{l>k} \rho^{-k}\mathbb{P}_y\left(\tau_y^T = l\right) \ = \ \sum_{k\geq 0} \rho^{-k}\mathbb{P}_y\left(\tau_y^T > k\right).$$

But the assumed existence of the exponential moment of τ_y^T under $\mathbb{P}_x$ and $\mathbb{P}_y$ is easily proved by

$$\begin{aligned}
\mathbb{E}_z\left(\rho^{-\tau_y^T}\right) \ &= \ \mathbb{E}_z\left(\mathbb{E}_z\left(\rho^{-(T_0+...+T_{\tau_y-1})}|M\right)\right) \\
&= \ \mathbb{E}_z\left(\prod_{i=0}^{\tau_y-1}\mathbb{E}_z\left(\rho^{-T_i}|M_i, M_{i+1}\right)\right) \\
&\leq \ \mathbb{E}_z\left(\max_{r,s}\mathbb{E}_r(\rho^{-T_0}|M_1 = s)^{\tau_y}\right) \\
&< \ \infty,
\end{aligned}$$

where z is either x or y. □

Hence, there is geometric ergodicity if the sojourn distributions are suitably chosen, but the obtained rate is far from being optimal. Nevertheless, in analogy to Corollary 6.4.8, for two families of sojourn times $(T_n)_{n\geq 0}$ an $(S_n)_{n\geq 0}$ with $\mathbb{P}_x(T_0 \leq n|M_1 = y) \geq \mathbb{P}_x(S_0 \leq n|M_1 = y)$ for every $x, y \in \mathcal{S}$,

$$\mathbb{E}_x(\rho^{-T_0}|M_1 = y) \;\leq\; \mathbb{E}_x(\rho^{-S_0}|M_1 = y)$$

for every $x, y \in \mathcal{S}$ and $\rho < 1$. Thus, the finiteness of these moments for the sojourn times $(S_n)_{n\geq 0}$ ensures the finiteness of these moments for the sojourn times $(T_n)_{n\geq 0}$. Moreover, it ensures the geometric ergodicity of M^T as well as an upper bound for the rate of this convergence, given the aperiodicity of M^T and $\mathbb{E}_x\left(\max_{y,z}\mathbb{E}_y(\rho^{-S_0}|M_1 = z)^{\tau_y}\right) < \infty$ for every $x, y \in \mathcal{S}$.

In addition to the above convergence performance which hearkens back on Blackwells renewal theorem, we can also use the Markov chain $(M_n^T, T_0 + ... + T_{N^T(n)} - n)_{n\geq 0}$ to obtain convergence rate results.

Lemma 6.4.10. *Let M^T be a semi-Markov chain with integrable sojourn times $(T_n)_{n\geq 0}$, limiting distribution π^T and embedded Markov chain M with stationary distribution π. Let as before $\mu(x, s) = \pi^T(x)\mathbb{P}_x(T_0 \geq s)/\mathbb{E}_x(T_0)$, $x \in \mathcal{S}$, $s \geq 1$, and write $\mu_{0,x} := \delta_x \otimes \mathbb{P}_x^{T_0}$, $x \in \mathcal{S}$, as well as $S_n := T_0 + ... + T_{N^T(n)}$, $n \geq 0$. Then, for every $n \geq 0$,*

$$d_{TV}(\mathbb{P}_x^{M_n^T}, \pi^T) \;\leq\; d_{TV}(\mathbb{P}_{\mu_{0,x}}^{(M_n^T, S_n - n)}, \mu).$$

Proof: Specifying the remaining sojourn time $S_n - n$ for every $n \geq 0$ yields

$$\begin{aligned} d_{TV}(\mathbb{P}_x^{M_n^T}, \pi^T) \;&=\; \frac{1}{2}\sum_y \left|\sum_{s\geq 1}\mathbb{P}_x(M_n^T = y, S_n - n = s) - \frac{\pi^T(x)}{\mathbb{E}_x(T_0)}\mathbb{P}_x(T_0 \geq s)\right| \\ &\leq\; \frac{1}{2}\sum_x\sum_{s\geq 1}\left|\mathbb{P}_{\mu_{0,x}}(M_n^T = x, S_n - n = s) - \mu(x, s)\right| \\ &=\; d_{TV}(\mathbb{P}_{\mu_{0,x}}^{(M_n^T, S_n - n)}, \mu). \end{aligned}$$

□

6.5. Aggregated Processes

We return to our model and the problem of relating the mixing time of the original chain X and the aggregated resp. accelerated aggregated chains $\overline{Y}^{(i)}$ resp. $Y^{(i)}$. It is intuitively clear that the mixing time is not much affected by aggregating only very few states. So, for small i, we expect the processes $\overline{Y}^{(i)}$ and X to have almost the same mixing times. But by aggregating more and more states, this strong relation will decrease until at aggregation level $i = \mathfrak{n}$ the mixing time of $\overline{Y}^{(\mathfrak{n})}$ is 1. This proves the behavior of the mixing time to depend strongly on i.

In the following, fix some $1 \leq i \leq \mathfrak{n}$ and write $Y := Y^{(i)}$, $\overline{Y} := \overline{Y}^{(i)}$ and $V(m) := V^{(i)}(m)$ for $m \in \mathcal{S}^{(i)}$ for ease of notation.

6.5.1. The AAC

We start by giving explicit bounds on the mixing time of the AAC in terms of the mixing time of X, the first being expected to be smaller than the latter because the time is accelerated. As explained at the beginning of this chapter, we want to use that Y and the embedded jump chain Z of the hit chain $\overline{Z}$ on $\mathcal{S}^{(i)}$ do not differ much for large β. We formalize this as follows:

Lemma 6.5.1. *For every $\varepsilon > 0$ and every $n \geq 1$, there exists some β large enough such that $\max_{m \in \mathcal{S}^{(i)}} d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) \leq \varepsilon$.*

Proof: To bound the total variation, we first bound the difference between the transition probabilities of Y and Z. By Theorem 2.1.3,

$$\mathbb{P}_r(Y_1 = s, \sigma_2 < \tau_s) \;\leq\; \sum_{x \in V(s)} \mathbb{P}_x(\sigma_1 < \tau_s) \;\stackrel{\beta \to \infty}{\longrightarrow}\; 0$$

for every $r, s \in \mathcal{S}^{(i)}$. Furthermore,

$$\mathbb{P}_r(Z_1 \neq Y_1) \;=\; \sum_{x \in \mathcal{N}(r) \setminus N^{(i)}} \mathbb{P}_r(X_{\sigma_1} = x) \mathbb{P}_x(\sigma_1 < \tau_{\mathfrak{m}(x)}) \;\stackrel{\beta \to \infty}{\longrightarrow}\; 0$$

for $r \in N^{(i)}$ and $\mathbb{P}_r(Z_1 \neq Y_1) = 0$ for $r \in \mathcal{S}^{(i)} \setminus N^{(i)}$. Therefore,

$$\begin{aligned}
\mathbb{P}_r(Y_1 = s) \;&=\; \mathbb{P}_r(Y_1 = s, \sigma_2 < \tau_s) + \mathbb{P}_r(Y_1 = s, \sigma_2 > \tau_s) \\
&\leq\; o(1) + \mathbb{P}_r(Z_1 = s), \\
\mathbb{P}_r(Z_1 = s) \;&=\; \mathbb{P}_r(Z_1 = s = Y_1) + \mathbb{P}_r(Z_1 = s \neq Y_1) \\
&\leq\; \mathbb{P}_r(Y_1 = s) + \mathbb{P}_r(Z_1 \neq Y_1) \\
&=\; \mathbb{P}_r(Y_1 = s) + o(1),
\end{aligned}$$

as β tends to infinity. Using the results of Section 3.2, we further infer

$$\begin{aligned}
\mathbb{P}_m(Y_k = s | Y_{k-1} = r) \;&=\; \sum_{x \in V(r)} \frac{\mathbb{P}_m(Y_k = s | X_{\sigma_{k-1}} = x) \mathbb{P}_m(X_{\sigma_{k-1}} = x)}{\mathbb{P}_m(Y_{k-1} = r)} \\
&=\; \sum_{x \in V(r)} \frac{(\mathbb{P}_r(Y_1 = s) + o(1)) \mathbb{P}_m(X_{\sigma_{k-1}} = x)}{\mathbb{P}_m(Y_{k-1} = r)} \\
&=\; \mathbb{P}_r(Y_1 = s) + o(1) \\
&=\; \mathbb{P}_r(Z_1 = s) + o(1)
\end{aligned}$$

as $\beta \to \infty$ for every $m, s \in \mathcal{S}^{(i)}$, $k \geq 1$ and $r \in \mathcal{S}^{(i)}$ with $\mathbb{P}_m(Y_{k-1} = r) > 0$. Hence, there exists some β large enough such that

$$|\mathbb{P}_m(Y_k = s | Y_{k-1} = r) - \mathbb{P}_r(Z_1 = s)| \;\leq\; \frac{2\varepsilon}{n |\mathcal{S}^{(i)}|} \tag{6.21}$$

for any $1 \leq k \leq n$. This yields

$$\max_{r : \mathbb{P}_m(Y_{k-1} = r) > 0} \sum_s |\mathbb{P}_m(Y_k = s | Y_{k-1} = r) - \mathbb{P}_r(Z_1 = s)| \;\leq\; \frac{2\varepsilon}{n}$$

for every $1 \leq k \leq n$ and β large enough (depending on ε and n). With this and in consideration of

$\mathbb{P}_m(Y_{k-1}=r)>0$ if and only if $\mathbb{P}_m(Z_{k-1}=r)>0$, we find

$$\begin{aligned}
&d_{TV}(\mathbb{P}_m^{Y_k},\mathbb{P}_m^{Z_k})\\
&\quad= \frac{1}{2}\sum_s |\mathbb{P}_m(Y_k=s)-\mathbb{P}_m(Z_k=s)|\\
&\quad= \frac{1}{2}\sum_s \Big| \sum_{r:\mathbb{P}_m(Y_{k-1}=r)>0} \mathbb{P}_m(Y_k=s|Y_{k-1}=r)\mathbb{P}_m(Y_{k-1}=r)\\
&\qquad\qquad - \mathbb{P}_m(Z_k=s|Z_{k-1}=r)\mathbb{P}_m(Z_{k-1}=r)\Big|\\
&\quad\le \frac{1}{2}\sum_{r:\mathbb{P}_m(Y_{k-1}=r)>0}\sum_s \mathbb{P}_m(Y_k=s|Y_{k-1}=r)|\mathbb{P}_m(Y_{k-1}=r)-\mathbb{P}_m(Z_{k-1}=r)|\\
&\qquad+\frac{1}{2}\sum_{r:\mathbb{P}_m(Y_{k-1}=r)>0}\sum_s \mathbb{P}_m(Z_{k-1}=r)|\mathbb{P}_m(Z_k=s|Z_{k-1}=r)-\mathbb{P}_m(Y_k=s|Y_{k-1}=r)|\\
&\quad\le \frac{1}{2}\sum_{r:\mathbb{P}_m(Y_{k-1}=r)>0} |\mathbb{P}_m(Y_{k-1}=r)-\mathbb{P}_m(Z_{k-1}=r)|\\
&\qquad+\frac{1}{2}\max_{r:\mathbb{P}_m(Y_{k-1}=r)>0}\sum_s |\mathbb{P}_m(Z_k=s|Z_{k-1}=r)-\mathbb{P}_m(Y_k=s|Y_{k-1}=r)|\\
&\quad\le d_{TV}(\mathbb{P}_m^{Y_{k-1}},\mathbb{P}_m^{Z_{k-1}})+\frac{\varepsilon}{n}
\end{aligned}$$

for any $1\le k\le n$. In conclusion,

$$d_{TV}(\mathbb{P}_m^{Y_n},\mathbb{P}_m^{Z_n}) \;\le\; d_{TV}(\mathbb{P}_m^{Y_{n-1}},\mathbb{P}_m^{Z_{n-1}})+\frac{\varepsilon}{n} \;\le\; \ldots \;\le\; \varepsilon$$

for every $m\in\mathcal{S}^{(i)}$. □

Since the processes Y and Z behave quite similar as $\beta\to\infty$, so do the limiting distributions. For results on the mixing time, we identify these limiting distributions and specify their distance. Let π^* be the stationary distribution of Z, which is an irreducible Markov chain and aperiodic if $\mathbb{P}_m(Z_2=m')>0$ for every pair (m,m') with $\mathbb{P}_m(Z_1=m')>0$. This ensures the existence of and convergence to π^*. From Lemma A.2.4 we know that $\pi(\cdot)/\pi(\mathcal{S}^{(i)})$ is the stationary distribution of $\overline{Z}$, and thus, by Proposition 6.3.2 (a),

$$\pi^*(m) \;=\; \frac{1}{K}\pi(m)\left(1-\mathbb{P}_m(\overline{Z}_1=m)\right), \quad m\in\mathcal{S}^{(i)},$$

where K is the normalization constant. More precisely, we have for $m\in N^{(i)}$

$$\pi(m)\left(1-\mathbb{P}_m(\overline{Z}_1=m)\right) \;=\; \pi(m)\left(1-p(m,m)-\mathcal{R}(m)\right),$$

where

$$\mathcal{R}(m) \;:=\; \mathbb{P}_m\left(X_1\in V(m') \text{ for some } m'\in\mathcal{S}^{(i)}\backslash N^{(i)}, \tau_{m'}>\zeta_0, X_{\zeta_0}=m\right) \;\to\; 0$$

as $\beta \to \infty$, and for $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$

$$\begin{aligned}
&\pi(m)(1-\mathbb{P}_m(\overline{Z}_1 = m)) \\
&\quad = \pi(m)p(m,N^{(i)}) + \sum_{n\geq 1} \sum_{x_1,\ldots,x_n \in V(m)\backslash\{m\}} \sum_{x\in N^{(i)}} \pi(m)p(m,x_1)\cdot\ldots\cdot p(x_n,x) \\
&\quad = \pi(m)p(m,N^{(i)}) + \sum_{n\geq 1} \sum_{x_1,\ldots,x_n \in V(m)\backslash\{m\}} \sum_{x\in N^{(i)}} p(x_n,x_{n-1})\cdot\ldots\cdot p(x_1,m)\pi(x_n)p(x_n,x) \\
&\quad = \pi(m)p(m,N^{(i)}) + \sum_{x'\in\partial V(m)} \sum_{n\geq 1} \mathbb{P}_{x'}(\zeta_0 > \tau_m = n) \sum_{x\in N^{(i)}} \pi(x')p(x',x) \\
&\quad = \pi(m)p(m,N^{(i)}) + \sum_{x'\in\partial V(m)} \mathbb{P}_{x'}(\zeta_0 > \tau_m \geq 1)\pi(x')p(x',N^{(i)}) \\
&\quad = \sum_{x'\in\partial V(m)} \mathbb{P}_{x'}(\zeta_0 > \tau_m^0)\pi(x')p(x',N^{(i)}).
\end{aligned}$$

Proposition 6.5.2. *Let $\mathbb{P}_m(Z_2 = m') > 0$ for every pair (m,m') with $\mathbb{P}_m(Z_1 = m') > 0$. For every $m, m' \in \mathcal{S}^{(i)}$,*

$$\mathbb{P}_m(Y_n = m') \overset{n\to\infty}{\longrightarrow} \frac{1}{K'} \sum_{x\in\partial V(m')} \pi(x)(1-p(x,V(m'))),$$

where K' is a normalization constant.

The assumption may seem quite unintuitive but it ensures the ergodicity of the Markov chains $(X_{\sigma_n})_{n\geq 0}$ (needed in the proof) as well as the ergodicity of Z. Furthermore, in view of an application of Theorem 6.3.7, this precise assumptions appears to be more convenient than the ergodicity of the two chains

Proof: In this proof, if not otherwise specified, x and y are elements of $N^{(i)}$ or $\partial V(m)$ for some $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$, that is, possible values of $(X_{\sigma_n})_{n\geq 0}$, and $\mathfrak{m}(x) := m$ if $x \in \partial V(m)$. Since the assumption ensures the ergodicity of $(X_{\sigma_n})_{n\geq 0}$, it suffices to prove that $\{\pi(x)(1-p(x,V(\mathfrak{m}(x))))\}_x$ is stationary for $(X_{\sigma_n})_{n\geq 0}$. Then the normalized version is the unique limiting distribution and

$$\mathbb{P}_m(Y_n = m') = \sum_{x\in\partial V(m')} \mathbb{P}_m(X_{\sigma_n} = x) \overset{n\to\infty}{\longrightarrow} \frac{1}{K'} \sum_{x\in\partial V(m')} \pi(x)(1-p(x,V(m'))).$$

For the stationarity, first, let $y \notin N^{(i)}$. Then

$$\mathbb{P}_x(X_{\sigma_1} = y) = \frac{p(x,y)}{1-p(x,x)}$$

for $x \in N^{(i)}$ and zero otherwise. Thereby,

$$\begin{aligned}
\sum_x \mathbb{P}_x(X_{\sigma_1} = y)\pi(x)(1-p(x,V(\mathfrak{m}(x)))) &= \sum_{x\in N^{(i)}} \frac{p(x,y)}{1-p(x,x)}\pi(x)(1-p(x,x)) \\
&= \pi(y)p(y,N^{(i)}) \\
&= \pi(y)\left(1-p\left(y,V\left(\mathfrak{m}\left(y\right)\right)\right)\right).
\end{aligned}$$

For $y \in N^{(i)}$, observe first that

$$\begin{aligned}
&\sum_{x \notin N^{(i)}} \mathbb{P}_x(X_{\sigma_1} = y)\pi(x)(1 - p(x, V(\mathfrak{m}(x)))) \\
&\quad = \sum_{x \notin N^{(i)}} \left(\pi(x)p(x,y)p(x,N^{(i)}) + \sum_{n \geq 1} \sum_{x_1,\ldots,x_n \in V(\mathfrak{m}(x))} \pi(x)p(x,x_1) \cdot \ldots \cdot p(x_n,y)p(x,N^{(i)}) \right) \\
&\quad = \sum_{x \notin N^{(i)}} \Bigg(\pi(y)p(y,x)p(x,N^{(i)}) \\
&\qquad\quad + \sum_{n \geq 1} \sum_{\substack{x_1,\ldots,x_{n-1},x' \in V(m(x)) \\ x'' \in N^{(i)}}} \pi(y)p(y,x')p(x',x_{n-1}) \cdot \ldots \cdot p(x_1,x)p(x,x'') \Bigg) \\
&\quad = \pi(y) \sum_{x' \notin N^{(i)}} p(y,x') \left(\mathbb{P}_{x'}(\zeta_0 = 1) + \sum_{n \geq 1} \mathbb{P}_{x'}(\zeta_0 = n+1) \right) \\
&\quad = \pi(y)(1 - p(y, N^{(i)})).
\end{aligned}$$

This provides the announced stationarity, for

$$\begin{aligned}
&\sum_x \mathbb{P}_x(X_{\sigma_1} = y)\pi(x)(1 - p(x, V(\mathfrak{m}(x)))) \\
&\qquad = \sum_{x \in N^{(i)} \setminus \{y\}} p(x,y)\pi(x) + \sum_{x \notin N^{(i)}} \mathbb{P}_x(X_{\sigma_1} = y)\pi(x)(1 - p(x, V(\mathfrak{m}(x)))) \\
&\qquad = \pi(y)p(y, N^{(i)} \setminus \{y\}) + \pi(y)(1 - p(y, N^{(i)})) \\
&\qquad = \pi(y)(1 - p(y,y)).
\end{aligned}$$

□

Hence, though Y is in general not Markovian, it has a limiting distribution, denoted by π^Y hereafter. Therefore, we are able to define a mixing time

$$\tau^Y_{\mathrm{mix}}(\varepsilon) \;:=\; \inf\left\{ k \geq 0 \,|\, \max_m d_{TV}(\mathbb{P}^{Y_n}_m, \pi^Y) \leq \varepsilon \text{ for every } n \geq k \right\}.$$

Next, we show that for every $\varepsilon > 0$ there exists some β large enough such that $d_{TV}(\pi^*, \pi^Y) \leq \varepsilon$.

Proposition 6.5.3. *For the limiting distributions π^* and π^Y of $\overline{Z}$ and Y it holds true that*

$$d_{TV}(\pi^*, \pi^Y) \;\to\; 0 \quad \textit{as} \quad \beta \to \infty.$$

Proof: For $m \in N^{(i)}$,

$$\begin{aligned}
\pi^*(m) - \pi^Y(m) \;&=\; \frac{\pi(m)\,(1 - p(m,m) - \mathcal{R}(m))}{K} - \frac{\pi(m)\,(1 - p(m,m))}{K'} \\
&=\; \pi(m)\,(1 - p(m,m)) \left(\frac{1 - \frac{\mathcal{R}(m)}{1-p(m,m)}}{K} - \frac{1}{K'} \right),
\end{aligned}$$

and for $m \in \mathcal{S}^{(i)} \backslash N^{(i)}$

$$\begin{aligned}\pi^*(m) - \pi^Y(m) &= \sum_{x \in \partial V(m)} \frac{\mathbb{P}_x(\zeta_0 > \tau_m^0)\pi(x)p(x, N^{(i)})}{K} - \sum_{x \in \partial V(m)} \frac{\pi(x)p(x, N^{(i)})}{K'} \\ &= \sum_{x \in \partial V(m)} \pi(x)p(x, N^{(i)}) \left(\frac{\mathbb{P}_x(\zeta_0 > \tau_m^0)}{K} - \frac{1}{K'} \right).\end{aligned}$$

Therefore,

$$\begin{aligned}d_{TV}(\pi^*, \pi^Y) &= \frac{1}{2} \sum_{m \in \mathcal{S}^{(i)}} |\pi^*(m) - \pi^Y(m)| \\ &\leq \frac{1}{2} \sum_{m \in \mathcal{S}^{(i)}} \sum_{x \in \partial V(m)} \pi(x) \left(1 - p(x, V(\mathfrak{m}(x)))\right) \\ &\qquad \times \max \left\{ \max_{m \in N^{(i)}} \left| \frac{1 - \frac{\mathcal{R}(m)}{1-p(m,m)}}{K} - \frac{1}{K'} \right|, \max_{\substack{m \in \mathcal{S}^{(i)} \backslash N^{(i)} \\ x \in \partial V(m)}} \left| \frac{\mathbb{P}_x(\zeta_0 > \tau_m^0)}{K} - \frac{1}{K'} \right| \right\} \\ &= \frac{1}{2} \max \left\{ \max_{m \in N^{(i)}} \left| \frac{K' \left(1 - \frac{\mathcal{R}(m)}{1-p(m,m)}\right)}{K} - 1 \right|, \max_{\substack{m \in \mathcal{S}^{(i)} \backslash N^{(i)} \\ x \in \partial V(m)}} \left| \frac{K' \mathbb{P}_x(\zeta_0 > \tau_m^0)}{K} - 1 \right| \right\}.\end{aligned}$$

But we see for every $m \in N^{(i)}$,

$$\left| \frac{K' \left(1 - \frac{\mathcal{R}(m)}{1-p(m,m)}\right)}{K} - 1 \right| \leq \left| \frac{K'}{K} - 1 \right| + \frac{K'}{K} \underbrace{\frac{\mathcal{R}(m)}{1 - p(m,m)}}_{\to 0 \text{ as } \beta \to \infty}$$

since $p^*(m,m) < 1$ and $\mathcal{R}(m) \to 0$ as $\beta \to \infty$ for every $m \in N^{(i)}$. In a similar vein, for every $m \in \mathcal{S}^{(i)} \backslash N^{(i)}, x \in \partial V(m)$,

$$\left| \frac{K' \mathbb{P}_x(\zeta_0 > \tau_m^0)}{K} - 1 \right| \leq \left| \frac{K'}{K} - 1 \right| + \frac{K'}{K} \underbrace{\mathbb{P}_x(\zeta_0 < \tau_m^0)}_{\to 0 \text{ as } \beta \to \infty}.$$

We are left with the proof of $K'/K \to 1$ as $\beta \to \infty$. This is a consequence of

$$\begin{aligned}K' - K &= \sum_{m \in N^{(i)}} \pi(m)\mathcal{R}(m) + \sum_{m \in \mathcal{S}^{(i)} \backslash N^{(i)}} \sum_{x \in \partial V(m)} \pi(x)(1 - p(x, V(m)))\mathbb{P}_x(\zeta_0 < \tau_m^0) \\ &\leq K \cdot \left(\left(\max_{m \in N^{(i)}} \frac{\mathcal{R}(m)}{1 - p(m,m) - \mathcal{R}(m)} \right) \vee \left(\max_{\substack{m \in \mathcal{S}^{(i)} \backslash N^{(i)} \\ x \in \partial V(m)}} \frac{\mathbb{P}_x(\zeta_0 < \tau_m^0)}{\mathbb{P}_x(\zeta_0 > \tau_m^0)} \right) \right)\end{aligned}$$

and

$$\frac{K'}{K} - 1 = \frac{K' - K}{K} \leq \left(\max_{m \in N^{(i)}} \frac{\mathcal{R}(m)}{1 - p(m,m) - \mathcal{R}(m)} \right) \vee \left(\max_{\substack{m \in \mathcal{S}^{(i)} \backslash N^{(i)} \\ x \in \partial V(m)}} \frac{\mathbb{P}_x(\zeta_0 < \tau_m^0)}{\mathbb{P}_x(\zeta_0 > \tau_m^0)} \right) \to 0$$

as $\beta \to \infty$. □

The next proposition summarizes the above results. It is preliminary to the main result, stated in Theorem 6.5.5, but also of interest on its own right.

Proposition 6.5.4. *Given the previous notation, let $\mathbb{P}_m(Z_2 = m') > 0$ whenever $\mathbb{P}_m(Z_1 = m') > 0$, and let the parameter Δ in Theorem 6.1.1 be positive.*

(a) Then $\frac{2(1-p)}{C}\gamma_\star^Z \geq \gamma_\star^{\overline{Z}} \geq \gamma_\star^X$, where

$$\begin{aligned} C &:= \min\left\{ \frac{\mathbb{P}_m(Z_2 = m')}{\mathbb{P}_m(Z_1 = m')} \Big| m \neq m', \mathbb{P}_m(Z_1 = m') > 0 \right\} \\ p &:= \min_m \mathbb{P}_m(\overline{Z}_1 = m). \end{aligned}$$

(b) Furthermore, in the low-temperature limit,

$$\limsup_{\beta\to\infty} \frac{1}{\beta} \ln\left(\min_m (1 - \mathbb{P}_m(\overline{Z}_1 = m))\gamma_\star^Z\right) \;\leq\; \lim_{\beta\to\infty} \frac{1}{\beta} \ln\left(\gamma_\star^{\overline{Z}}\right) \;=\; \lim_{\beta\to\infty} \frac{1}{\beta} \ln\left(\gamma_\star^X\right)$$

and

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln\left(\min_m (1 - \mathbb{P}_m(\overline{Z}_1 = m))\right) \;=\; -\max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} \left(E(s_m) - E(m)\right).$$

(c) Finally, let $\varepsilon, \delta > 0$. For every (n, β) with $d_{TV}(\pi^*, \pi^Y) \leq \delta$, $\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) \leq \delta$ and $n \geq \tau_{mix}^Z(\varepsilon)$,

$$\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y) \leq \varepsilon + 2\delta.$$

For every (n, β) with $d_{TV}(\pi^*, \pi^Y) \leq \delta$, $\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) \leq \delta$ and $n \geq \tau_{mix}^Y(\varepsilon)$,

$$\max_m d_{TV}(\mathbb{P}_m^{Z_n}, \pi^*) \leq \varepsilon + 2\delta.$$

Proof: The first inequality in part (a) is a consequence of Theorem 6.3.7 above, for the second we refer to Lemma A.2.4. The first inequality in part (b) follows from Theorem 6.3.4. In all these parts we used the fact that X and $\overline{Z}$ have holding probabilities of at least $\frac{1}{2}$ in every state. The remaining equality in the first part of (b) is a combination of Theorem 6.1.1 and Corollary 6.1.6. For the second line in (b), note that for $m \in N^{(i)}$,

$$1 - \mathbb{P}_m(\overline{Z}_1 = m) \;=\; 1 - p(m, m) - \mathcal{R}(m),$$

which converges to a positive limit, whereas for $m \in \mathcal{S}^{(i)}\setminus N^{(i)}$,

$$1 - \mathbb{P}_m(\overline{Z}_1 = m) \;=\; \mathbb{P}_m(\zeta_0 < \tau_m),$$

which converges to zero. Therefore, as $\beta \to \infty$,

$$\begin{aligned} \min_{m\in\mathcal{S}^{(i)}} (1 - \mathbb{P}_m(\overline{Z}_1 = m)) \;&=\; \min_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} \mathbb{P}_m(\zeta_0 < \tau_m) \;=\; \min_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} e^{-\beta(E(s_m)-E(m)+o(1))} \\ &=\; e^{-\beta(\max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m))+o(1))}. \end{aligned}$$

To prove part (c), Lemma 6.5.1 and Proposition 6.5.3 yield for every $n \geq \tau_{\text{mix}}^Z(\varepsilon)$ and β large enough (depending on n and δ),

$$\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y) \;\leq\; \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) + \max_m d_{TV}(\mathbb{P}_m^{Z_n}, \pi^*) + d_{TV}(\pi^*, \pi^Y) \;\leq\; \varepsilon + 2\delta$$

and analogously for $n \geq \tau_{\text{mix}}^Y(\varepsilon)$ and β large enough, again depending on n and δ,

$$\max_m d_{TV}(\mathbb{P}_m^{Z_n}, \pi^*) \;\leq\; \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) + \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y) + d_{TV}(\pi^Y, \pi^*) \;\leq\; \varepsilon + 2\delta.$$

□

Theorem 6.5.5. *Given the previous notation, let $\mathbb{P}_m(Z_2 = m') > 0$ whenever $\mathbb{P}_m(Z_1 = m') > 0$, and let the parameter Δ in Theorem 6.1.1 be positive. Then*

(a) $\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^*) \leq \varepsilon + 2\delta$ *for every* $\varepsilon, \delta > 0$ *and any* (n, β) *such that*

$$n \;\geq\; \ln\left(\frac{1}{\varepsilon \min_x \pi^*(x)}\right)\left(\frac{1}{\ln\left(\frac{1}{2\varepsilon}\right)}\tau_{\text{mix}}^X(\varepsilon) + 1\right)\frac{2(1-p)}{C},$$

$d_{TV}(\pi^*, \pi^Y) \leq \delta$, *and* $\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) \leq \delta$.

(b) If n and β both tend to infinity in such a way that $d_{TV}(\pi^Y, \pi^*) \leq \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y)$ and $\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) \leq \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y)$ for each pair (n, β), then

$$\limsup_{n,\beta\to\infty} \frac{1}{\beta} \ln\left(1 - \sqrt[n]{\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y)}\right) \;\leq\; -\Delta + \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \left(E(s_m) - E(m)\right).$$

Proof: (a) The statement is a combination of Proposition 6.5.4 and Proposition A.2.1 (c). Namely, by Proposition 6.5.4 (c), the statement holds true for $n \geq \tau_{mix}^Z(\varepsilon)$ and

$$\begin{aligned}
\tau_{\text{mix}}^Z(\varepsilon) \;&\leq\; \ln\left(\frac{1}{\varepsilon \min_x \pi^*(x)}\right)\frac{1}{\gamma_\star^Z} \\
&\leq\; \ln\left(\frac{1}{\varepsilon \min_x \pi^*(x)}\right)\frac{2(1-p)}{C\gamma_\star^{\overline{Z}}} \\
&\leq\; \ln\left(\frac{1}{\varepsilon \min_x \pi^*(x)}\right)\frac{2(1-p)}{C\gamma_\star^X} \\
&\leq\; \ln\left(\frac{1}{\varepsilon \min_x \pi^*(x)}\right)\left(\frac{1}{\ln\left(\frac{1}{2\varepsilon}\right)}\tau_{\text{mix}}^X(\varepsilon) + 1\right)\frac{2(1-p)}{C}.
\end{aligned}$$

(b) Under the given assumptions,

$$\begin{aligned}
3\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y) \;&\geq\; \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \mathbb{P}_m^{Z_n}) + \max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y) + d_{TV}(\pi^*, \pi^Y) \\
&\geq\; \max_m d_{TV}(\mathbb{P}_m^{Z_n}, \pi^*)
\end{aligned}$$

so that

$$\begin{aligned}
&\limsup_{n,\beta\to\infty} \frac{1}{\beta} \ln\left(1 - \sqrt[n]{\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y)}\right) \\
&\qquad= \limsup_{n,\beta\to\infty} \frac{1}{\beta} \ln\left(1 - \sqrt[n]{3\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y)}\right) \\
&\qquad\leq \limsup_{n,\beta\to\infty} \frac{1}{\beta} \ln\left(1 - \sqrt[n]{\max_m d_{TV}(\mathbb{P}_m^{Z_n}, \pi^*)}\right) \\
&\qquad= \limsup_{\beta\to\infty} \frac{1}{\beta} \ln\left(1 - \lambda_\star^Z\right) \\
&\qquad= \limsup_{\beta\to\infty} \frac{1}{\beta} \ln\left(\gamma_\star^Z\right) \\
&\qquad\leq -\lim_{\beta\to\infty} \frac{1}{\beta} \ln\Big(\min_m(1 - \mathbb{P}_m(\overline{Z}_1 = m))\Big) + \lim_{\beta\to\infty} \frac{1}{\beta} \ln\left(\gamma_\star^X\right) \\
&\qquad\leq -\Delta + \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \left(E(s_m) - E(m)\right).
\end{aligned}$$

□

Here is an informal interpretation of this theoretic result under the assumption that the above bound is significant, that is, $\Delta - \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m)) > 0$: By the above, as n and β tend to infinity, $\max_m d_{TV}(\mathbb{P}_m^{Y_n}, \pi^Y)$ is approximately bounded below by

$$\left(1 - e^{-\beta(\Delta - \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m))+o(1))}\right)^n.$$

Hence, as $\varepsilon \to 0$ and $n, \beta \to \infty$,

$$\begin{aligned}
\tau_{mix}^Y(\varepsilon) &\geq \frac{\ln(\varepsilon)}{\ln\left(1 - e^{-\beta(\Delta - \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m))+o(1))}\right)} \\
&\approx \frac{\ln(\varepsilon)}{-e^{-\beta(\Delta - \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m))+o(1))}} \\
&= \frac{-\ln(\varepsilon)}{e^{-\beta(\Delta+o(1))}\max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} e^{\beta((E(s_m)-E(m))+o(1))}} \\
&= \frac{-\ln(\varepsilon)}{\gamma_\star^X \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} \mathbb{E}_m(\zeta_0)} \\
&\geq \frac{\tau_{mix}^X(\varepsilon)}{\max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} \mathbb{E}_m(\zeta_0)} \cdot \frac{\ln(\varepsilon)}{\ln(\varepsilon\pi_{min})}.
\end{aligned}$$

That is, the mixing time of X is at most the mixing time of Y times the maximal expected sojourn time in a metastable state. If, otherwise, $\Delta - \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m)) < 0$, the bound

$$\tau_{mix}^Y(\varepsilon) \geq 1 \geq -\ln(\varepsilon) e^{\beta(\Delta - \max_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}}(E(s_m)-E(m))+o(1))}$$

is trivially true for every fixed $\varepsilon > 0$ and $\beta \to \infty$.

6.5.2. The AC

Much is known on how to relate the mixing time of a Markov chain with the mixing time of its restriction to given subsets and of certain Markov chains representing the transitions between those subsets (see e.g. [42], [43], [21] or [22]). For example, let M be an ergodic, reversible Markov chain with stationary distribution π and $\mathcal{S}_1$,...,$\mathcal{S}_l$ such a partition of the state space that $\overline{Y}_n := \sum_{k=1}^l k\mathbb{1}_{\{M_n\in\mathcal{S}_k\}}$, $n \geq 0$, is again Markovian (in this case M is called *lumpable* and $\sum_{y\in\mathcal{S}_j} p(x,y) = \sum_{y\in\mathcal{S}_j} p(x',y)$ for every $x, x' \in \mathcal{S}_k$, $1 \leq k \leq l$) with transition probabilities $q(k,j) = \sum_{y\in\mathcal{S}_j} p(x,y)$ for some $x \in \mathcal{S}_k$. The ergodicity and reversibility of M ensure the ergodicity and reversibility of $\overline{Y}$ with respect to $\overline{\pi}(k) := \pi(\mathcal{S}_k)$, $1 \leq k \leq l$. Furthermore, easy calculations show for the corresponding Dirichlet forms and variances that $\mathcal{E}_M(g) = \mathcal{E}_{\overline{Y}}(f)$ and $Var_\pi(g) = Var_{\overline{\pi}}(f)$ for any function $f : \{1,...,l\} \to \mathbb{R}$ and $g : \mathcal{S} \to \mathbb{R}$, $g(x) := \sum_{k=1}^l f(k)\mathbb{1}_{\{x\in\mathcal{S}_k\}}$. Thus, for the corresponding spectral gaps $\gamma(M)$ and $\gamma(\overline{Y})$ it holds true that $\gamma(M) \leq \gamma(\overline{Y})$.

But since the chains studied in the above-mentioned works do not coincide with the AC, the letter not being Markovian in general, relations of X and $\overline{Y}$ apparently have not yet been studied. We begin with an example where the total variation distance is calculated numerically to illustrate the general picture.

Example 6.5.6. (a) Let M a birth and death chain on $\mathcal{S} := \{1,...,N\}$ with

$$\begin{aligned}
p(i,i+1) &= p(N,N) = p \quad \text{for } 1 \leq i \leq N-1, \\
p(i,i-1) &= p(1,1) = 1-p \quad \text{for } 2 \leq i \leq N.
\end{aligned}$$

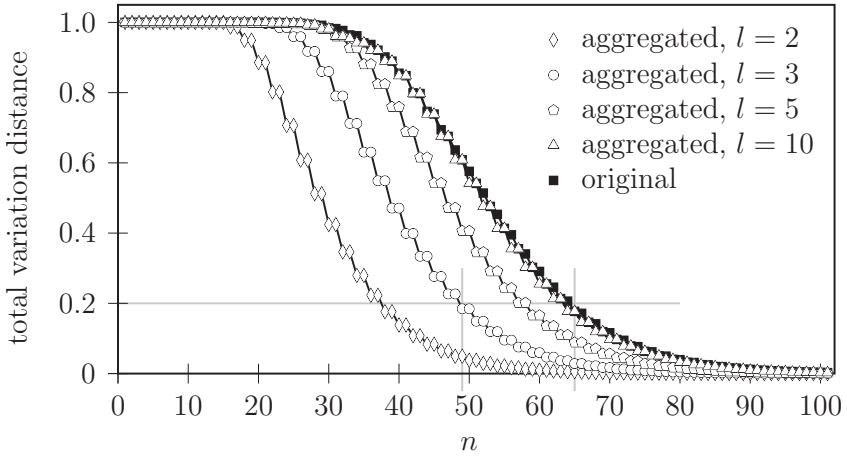

Figure 6.3.: Total variation distance between the distribution of the Markov chain with transition probabilities given in Example 6.5.6 (a) and its stationary distribution (squares) as well as the total variation distance between the aggregated chains for $l = 2$ (diamond), $l = 3$ (circle), $l = 5$ (pentagon) and $l = 10$ (triangle) and their limiting distributions. Grey lines mark the height 0.2 and the length 49 resp. 65.

For $l \in \mathbb{N}$, $N/l \in \mathbb{N}$, define subsets $(\mathcal{S}_k)_{1\le k\le l}$ of $\mathcal{S}$ by $\mathcal{S}_k := \left\{\frac{k-1}{l}N + 1, \frac{k}{l}N\right\}, 1 \le k \le l$. On these segments, we can define the aggregated chain $\overline{Y}_n := \sum_{k=1}^{l} k\mathbb{1}_{\{X_n\in\mathcal{S}_k\}}$ with limiting distribution $(\pi(\mathcal{S}_k))_{1\le k\le l}$. For $N = 30$, $p = 0.25$ and initial distribution $\lambda = \delta_N$, Figure 6.3 shows the total variation distance between the original chain resp. the aggregated chains (for $l \in \{2, 3, 5, 10\}$) and their limiting distributions evolving with n. The reader should notice that the higher the level of aggregation (that is, the smaller l), the faster the convergence.

(b) In part (a) there is a constant drift towards one end of the state space. Now we introduce on every segment a drift towards its center by changing some of the transition probabilities, namely

$$p(i, i+1) \;=\; 1-p \quad \text{and} \quad p(i, i-1) \;=\; p(1,1) \;=\; p$$

for $i = \frac{k-1}{l}N + j$, $1 \le k \le l$, $1 \le j \le \frac{N}{2l}$. Then the aggregated chain is still mixing faster but the difference is very small compared to the mixing time in total. In Figure 6.4 we see, again for $N = 30$, $p = 0.25$ and initial distribution $\lambda = \delta_N$, the total variation distance between the original chain resp. the aggregated chain for $l = 3$ and their limiting distributions. Regarding for example $\varepsilon = 0.2$, we have in the original chain a mixing time of 2054 and in the aggregated chain of 1857, giving a relative difference of 0.1. Without drift we have for $l = 3$ and $\varepsilon = 0.2$ a relative difference of $(65-49)/65 = 0.25$. Heuristically, this can be explained by the following: If we had two independent versions of the process starting in different states, due to the drift it would take quite a while for the corresponding ACs to meet. But once they meet, the original processes would comparably fast fall down to the center of that $\mathcal{S}_k$ and meet as well.

We want to record the obvious observation that $\overline{Y}$ is stationary under $\mathbb{P}_\pi$, where π is the stationary distribution of X.

Lemma 6.5.7. *For every $n \ge 0$ and $m \in \mathcal{S}^{(i)}$ the distribution of $\overline{Y}_n$ under $\mathbb{P}_\pi$ is given by*

$$\mathbb{P}_\pi(\overline{Y}_n = m) \;=\; \pi(V(m)) \;=:\; \overline{\pi}(m).$$

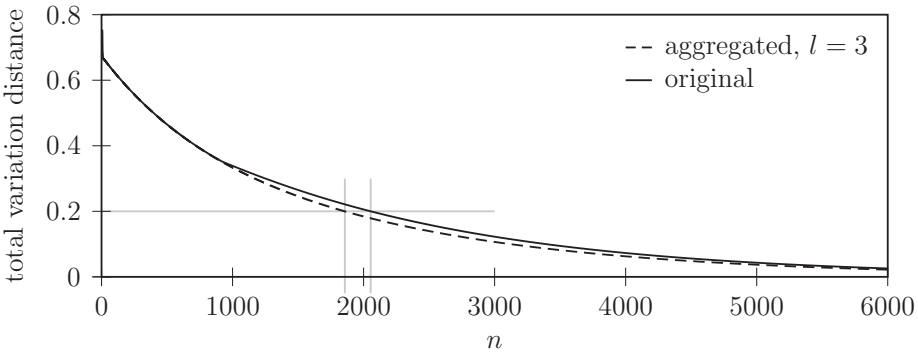

Figure 6.4.: Total variation distance between the distribution of the Markov chain with transition probabilities given in Example 6.5.6 (b) and its stationary distribution (solid line) as well as the total variation distance between the aggregated chain for $l = 3$ and its limiting distribution (dashed line). Grey lines mark the height 0.2 and the length 1857 resp. 2054.

So both processes converge towards some equilibrium distribution and the following lemma shows that the mixing time of the AC is smaller than the mixing time of the original process.

Lemma 6.5.8. *For every initial state $x \in \mathcal{S}$, the total variation satisfies*

$$d_{TV}(\mathbb{P}_x^{\overline{Y}_n}, \overline{\pi}) \;\leq\; d_{TV}(\mathbb{P}_x^{X_n}, \pi).$$

Proof: We just need the maximum-notation of the total variation to show

$$\begin{aligned} d_{TV}(\mathbb{P}_x^{\overline{Y}_n}, \overline{\pi}) &= \max_{A \subset \mathcal{S}^{(i)}} \left| \mathbb{P}_x\left(\overline{Y}_n \in A\right) - \overline{\pi}(A) \right| \\ &= \max_{A \subset \mathcal{S}^{(i)}} \left| \mathbb{P}_x\left(X_n \in \bigcup_{m \in A} V(m)\right) - \overline{\pi}\left(\bigcup_{m \in A} V(m)\right) \right| \\ &\leq \max_{A \subset \mathcal{S}} |\mathbb{P}_x(X_n \in A) - \pi(A)| \\ &= d_{TV}(\mathbb{P}_x^{X_n}, \pi). \end{aligned}$$

□

In addition to this analytic proof, Lemma 6.5.8 can be verified by a coupling argument, as a coupling of $\mathbb{P}_x^X$ and $\mathbb{P}_\pi^X$ induces a coupling of $\mathbb{P}_x^{\overline{Y}}$ and $\mathbb{P}_\pi^{\overline{Y}}$, the coupling time of the former being almost surely larger than the coupling time of the latter.

Consequently, if X is equilibrated up to some $\varepsilon \geq 0$, then $\overline{Y}$ is equilibrated up to ε as well. The converse is not that obvious since the stationarity of $\overline{Y}$ does not lead to a stationarity of X. This is illustrated in the following example, where it is also seen that at least in some cases it takes not that much more time for X to equilibrate once $\overline{Y}$ is equilibrated.

Example 6.5.9. Let M be a Birth and Death Process on $\{1, 2, 3, 4\}$ with

$$\begin{aligned} p(i, i+1) &= p(4,4) = p \quad \text{for } 1 \leq i \leq 3, \\ p(i, i-1) &= p(1,1) = 1-p \quad \text{for } 2 \leq i \leq 4. \end{aligned}$$

The stationary distribution π for which $(\mathbf{P}, \pi)$ is a reversible pair is proportional to

$$\left(\frac{p}{1-p}\right)^{i-1}, \quad 1 \leq i \leq 4.$$

We partition this state space into $\mathcal{S}_1 := \{1,2\}$ and $\mathcal{S}_2 := \{3,4\}$. Furthermore, let λ be some initial distribution with

$$\lambda(1) \;=\; a\pi(\mathcal{S}_1), \quad \lambda(2) \;=\; (1-a)\pi(\mathcal{S}_1), \quad \lambda(3) \;=\; b\pi(\mathcal{S}_2), \quad \lambda(4) \;=\; (1-b)\pi(\mathcal{S}_2).$$

with $a, b \in [0,1]$. Then $\lambda(1) + \lambda(2) = \pi(\mathcal{S}_1)$, $\lambda(3) + \lambda(4) = \pi(\mathcal{S}_2)$, and

$$\begin{aligned}(1-p)\lambda(3) \;&=\; (1-p)b\pi(\mathcal{S}_2) \;=\; (1-p)b\left(\frac{p}{1-p}\right)^2\pi(\mathcal{S}_1) \;=\; \frac{bp^2}{1-p}\pi(\mathcal{S}_1),\\ p\lambda(2) \;&=\; p(1-a)\pi(\mathcal{S}_1) \;=\; p(1-a)\left(\frac{1-p}{p}\right)^2\pi(\mathcal{S}_2) \;=\; \frac{(1-a)(1-p)^2}{p}\pi(\mathcal{S}_2).\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathbb{P}_\lambda(M_1 = 1) \;&=\; (1-p)\lambda(1) + (1-p)\lambda(2) \;=\; (1-p)\pi(\mathcal{S}_1)\\ \mathbb{P}_\lambda(M_1 = 2) \;&=\; p\lambda(1) + (1-p)\lambda(3) \;=\; \left(a + \frac{bp}{1-p}\right)p\pi(\mathcal{S}_1)\\ \mathbb{P}_\lambda(M_1 = 3) \;&=\; p\lambda(2) + (1-p)\lambda(4) \;=\; \left(\frac{(1-a)(1-p)}{p} + 1 - b\right)(1-p)\pi(\mathcal{S}_2)\\ \mathbb{P}_\lambda(M_1 = 4) \;&=\; p\lambda(3) + p\lambda(4) \;=\; p\pi(\mathcal{S}_2).\end{aligned}$$

That is, $\mathbb{P}_\lambda(M_n \in \mathcal{S}_k) = \pi(\mathcal{S}_k)$, $k \in \{1,2\}$, for any $n \geq 0$, if and only if $b = (1-a)(1-p)/p$. Hence, with this parameter in the initial distribution, the aggregated chain is in its equilibrium ab initio, whereas the original chain is not. We want to study how long it takes for the original chain to reach equilibrium. As

$$\begin{aligned}\mathbb{P}_\lambda(M_1 = 1) \;&=\; (1-p)\pi(\mathcal{S}_1) \;=\; (1-p)(\pi(1)+\pi(2)) \;=\; (1-p)\left(\pi(1) + \frac{p}{1-p}\pi(1)\right) \;=\; \pi(1),\\ \mathbb{P}_\lambda(M_1 = 2) \;&=\; \pi(\mathcal{S}_1) - \pi(1) \;=\; \pi(2),\\ \mathbb{P}_\lambda(M_1 = 3) \;&=\; (1-p)\pi(\mathcal{S}_2) \;=\; (1-p)(\pi(3)+\pi(4)) \;=\; (1-p)\left(\pi(3) + \frac{p}{1-p}\pi(3)\right) \;=\; \pi(3),\\ \mathbb{P}_\lambda(M_1 = 4) \;&=\; \pi(\mathcal{S}_2) - \pi(3) \;=\; \pi(4),\end{aligned}$$

we can conclude for this example that once the aggregated chain is in its equilibrium, it takes at most one more step for the original chain to reach its equilibrium.

Via the following method, it is possible to derive an upper bound on the mixing time of X in terms of the mixing time of $\overline{Y}$ for every $\varepsilon > 0$. Let more generally M be an irreducible, ergodic finite Markov chain on a set $\mathcal{S} = \sum_{k=1}^{l} \mathcal{S}_k$ with stationary distribution π. Then

$$\pi(s) \;=\; \sum_{k=1}^{l} \overline{\pi}_k \pi^{(k)}(s) \quad \text{with} \quad \pi^{(k)}(s) \;=\; \frac{\pi(s)}{\pi(\mathcal{S}_k)}\mathbb{1}_{\mathcal{S}_k}(s)$$

for every $1 \leq k \leq l$, and $\overline{\pi}_k = \pi(\mathcal{S}_k) = \lim_{n\to\infty} \mathbb{P}_x(M_n \in \mathcal{S}_k)$ for every $x \in \mathcal{S}$. A similar result holds true for the distribution of M at time n, namely

$$\mathbb{P}_x(M_n = s) \;=\; \sum_{k=1}^{l} \mathbb{P}_x(M_n = s, M_n \in \mathcal{S}_k) \;=\; \sum_{k=1}^{l} \mathbb{P}_x(M_n = s | M_n \in \mathcal{S}_k)\mathbb{P}_x(M_n \in \mathcal{S}_k)$$

for every $x \in \mathcal{S}$ and $n \geq 0$ with $\mathbb{P}_x(M_n \in \mathcal{S}_k) > 0$. Given this decomposition and the aggregated chain $\overline{Y}_n := \sum_{k=1}^{l} k \mathbb{1}_{\{M_n \in \mathcal{S}_k\}}$, $n \geq 0$, we can write the total variation distance between $\mathbb{P}_x^{M_n}, n \geq |\mathcal{S}|$, and π as

$$\begin{aligned}
d_{TV}(\mathbb{P}_x^{M_n}, \pi) &= \frac{1}{2} \sum_{s \in \mathcal{S}} |\mathbb{P}_x(M_n = s) - \pi(s)| \\
&= \frac{1}{2} \sum_{s \in \mathcal{S}} \left| \sum_{k=1}^{l} \left(\mathbb{P}_x(\overline{Y}_n = k) \mathbb{P}_x(M_n = s | M_n \in \mathcal{S}_k) - \overline{\pi}_k \pi^{(k)}(s) \right) \right| \\
&\leq \frac{1}{2} \sum_{s \in \mathcal{S}} \sum_{k=1}^{l} \left| \mathbb{P}_x(\overline{Y}_n = k) \mathbb{P}_x(M_n = s | M_n \in \mathcal{S}_k) - \overline{\pi}_k \pi^{(k)}(s) \right| \qquad (6.22) \\
&\leq \frac{1}{2} \sum_{s \in \mathcal{S}} \sum_{k=1}^{l} \Big(\left| \mathbb{P}_x(\overline{Y}_n = k) - \overline{\pi}_k \right| \mathbb{P}_x(M_n = s | M_n \in \mathcal{S}_k) \\
&\qquad + \overline{\pi}_k \left| \pi^{(k)}(s) - \mathbb{P}_x(M_n = s | M_n \in \mathcal{S}_k) \right| \Big) \\
&= d_{TV}(\mathbb{P}_x^{\overline{Y}_n}, (\overline{\pi}_k)_{1 \leq k \leq l}) + \sum_{k=1}^{l} \overline{\pi}_k d_{TV}(\mathbb{P}_x^{M_n | M_n \in \mathcal{S}_k}, \pi^{(k)}) \\
&= d_{TV}(\mathbb{P}_x^{\overline{Y}_n}, (\overline{\pi}_k)_{1 \leq k \leq l}) + \max_{1 \leq k \leq l} d_{TV}(\mathbb{P}_x^{M_n | M_n \in \mathcal{S}_k}, \pi^{(k)}).
\end{aligned}$$

Therefore, we can give a bound on the mixing time of M in terms of the mixing times of $\overline{Y}$ and $\mathbb{P}_x(M_n \in \cdot | M_n \in \mathcal{S}_k)_{n \geq |\mathcal{S}|}$ for $1 \leq k \leq l$, $x \in \mathcal{S}$. More precisely:

Proposition 6.5.10. *Let M an irreducible, ergodic finite Markov chain on a set $\mathcal{S} = \sum_{k=1}^{l} \mathcal{S}_k$ with stationary distribution π, and $\overline{Y}_n := \sum_{k=1}^{l} k \mathbb{1}_{\{M_n \in \mathcal{S}_k\}}$, $n \geq 0$. Let $\varepsilon > 0$, $\delta > 0$.*

$$d_{TV}(\mathbb{P}_x^{M_n}, \pi) \leq \varepsilon + \delta$$

for $x \in \mathcal{S}$ and every $n \geq \max\left\{ |\mathcal{S}|, \tau_{\text{mix}}^{\overline{Y}}(\varepsilon), \max_k \tau_{mix}^{(M_n | M_n \in \mathcal{S}_k)_{n \geq |\mathcal{S}|}}(\delta) \right\}$.

The above estimation

$$d_{TV}(\mathbb{P}_x^{M_n}, \pi) \leq d_{TV}(\mathbb{P}_x^{\overline{Y}_n}, (\overline{\pi}_k)_{1 \leq k \leq l}) + \max_{1 \leq k \leq l} d_{TV}(\mathbb{P}_x^{M_n | M_n \in \mathcal{S}_k}, \pi^{(k)}) \qquad (6.23)$$

is similar to the one obtain by Madras & Randall in [42], resp. Martin & Randall [43]:

Theorem 6.5.11 (Theorem 4.1 in [43])**.** *In the above notation, let $\mathbf{P}_{|\mathcal{S}_k}$ for every $1 \leq k \leq l$ be the restricted Markov chain on $\mathcal{S}_k$ (in canonical extension of Definition 6.1.2) with spectral gap $\gamma_{|\mathcal{S}_k}$ and Q be the limiting transition probabilities of $\overline{Y}$ with spectral gap γ_Q as in Corollary 6.1.7. Then*

$$\gamma \geq \frac{1}{2} \gamma_Q \min_{1 \leq k \leq l} \gamma_{|\mathcal{S}_k}.$$

Note that in the case where $\overline{Y}$ itself is again Markovian (that is $\mathbf{P}$ is lumpable with respect to $\mathcal{S}_1$,...,$\mathcal{S}_l$), γ_Q is the spectral gap of $\overline{Y}$.

In terms of the variation distance, with $\lambda_Q = 1 - \gamma_Q$, $\lambda_{|\mathcal{S}_k} = 1 - \gamma_{|\mathcal{S}_k}$, this theorem of Martin & Randall states that

$$\begin{aligned}
\max_x d_{TV}(\mathbb{P}_x^{M_n}, \pi) &\leq \left((1 + o(1)) \left(1 - \frac{1}{2} \gamma_Q \min_{1 \leq k \leq l} \gamma_{|\mathcal{S}_k} \right) \right)^n \\
&= \left(\frac{1 + o(1)}{2} \right)^n \left(1 + \lambda_Q + \max_{1 \leq k \leq l} \lambda_{|\mathcal{S}_k} - \lambda_Q \max_{1 \leq k \leq l} \lambda_{|\mathcal{S}_k} \right)^n
\end{aligned}$$

as $n \to \infty$. Furthermore, since

$$\frac{1}{2}(1+a+b-ab) \;=\; \frac{1}{2}(1+a)(1-b)+b \;\geq\; b$$

for every $0 \leq a < b \leq 1$, we find

$$(1+o(1))\left(\frac{1}{2}(1+a+b-ab)\right)^n \;\geq\; (1+o(1))b^n \;=\; \left(1+\frac{a^n}{b^n}\right)b^n \;=\; b^n + a^n$$

for n sufficiently large. Therefore,

$$\begin{aligned}
&\left(\frac{1+o(1)}{2}\right)^n \left(1+\lambda_Q+\max_{1\leq k\leq l}\lambda_{|\mathcal{S}_k}-\lambda_Q\max_{1\leq k\leq l}\lambda_{|\mathcal{S}_k}\right)^n \\
&\qquad\geq\; \left((\lambda_Q)^n+\left(\max_{1\leq k\leq l}\lambda_{|\mathcal{S}_k}\right)^n\right)(1+o(1))^{n-1} \\
&\qquad=\; \left(\max_x d_{TV}\left(\delta_x Q^n,(\overline{\pi}_k)_{1\leq k\leq l}\right)+\max_{1\leq k\leq l}\max_{x\in\mathcal{S}_k} d_{TV}\left(\delta_x \mathbf{P}^n_{|\mathcal{S}_k},\pi^{(k)}\right)\right)(1+o(1))
\end{aligned}$$

as $n \to \infty$ whenever $\lambda_Q \neq \max_{1\leq k\leq l}\lambda_{|\mathcal{S}_k}$. In our situation, as $\beta \to \infty$, $\lambda_Q = 1-e^{-\beta(\Delta+o(1))}$ differs from this maximum, as $\lambda_{|\mathcal{S}_k} = 1-e^{-\beta(\Delta_V+o(1))}$ whenever $\mathcal{S}_k = V := V(m)$ for some $m \in \mathcal{S}^{(i)}\backslash N^{(i)}$ and $\lambda_{|\mathcal{S}_k} := 0$ for $\mathcal{S}_k = \{s\}$ with $s \in N^{(i)}$. To compare this with our result, recall that Q contains the limiting transition probabilities of $\overline{Y}$ and, furthermore, the restricted transition probabilities $\mathbf{P}_{|\mathcal{S}_k}$ equal in the same sense the limiting transition probabilities of $\mathbb{P}_x(M_n \in \cdot\,|M_n \in \mathcal{S}_k)$. Namely, for n sufficiently large,

$$\begin{aligned}
\mathbb{P}_x(M_n = s|M_n \in \mathcal{S}_k) \;&=\; \sum_{r\in\mathcal{S}}\frac{\mathbb{P}_x(M_{n-1}=r,M_n=s)}{\mathbb{P}_x(M_n\in\mathcal{S}_k)} \\
&=\; \sum_{r\in\mathcal{S}_k\backslash\{s\}}\frac{\mathbb{P}_x(M_{n-1}=r)}{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}\frac{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}{\mathbb{P}_x(M_n\in\mathcal{S}_k)}p(r,s) \\
&\qquad+\frac{\mathbb{P}_x(M_{n-1}=s)}{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}\frac{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}{\mathbb{P}_x(M_n\in\mathcal{S}_k)}p(s,s) \\
&\qquad+\frac{\mathbb{P}_x(M_{n-1}=s)}{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}\sum_{y\notin\mathcal{S}_k}\frac{\mathbb{P}_x(M_{n-1}=y)}{\mathbb{P}_x(M_{n-1}=s)}\frac{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}{\mathbb{P}_x(M_n\in\mathcal{S}_k)}p(y,s) \\
&=\; \sum_{r\in\mathcal{S}_k}\mathbb{P}_x(M_{n-1}=r|M_{n-1}\in\mathcal{S}_k)p^{(x,n)}_{|\mathcal{S}_k}(r,s),
\end{aligned}$$

where

$$p^{(x,n)}_{|\mathcal{S}_k}(r,s) \;=\; \frac{\mathbb{P}_x(M_{n-1}\in\mathcal{S}_k)}{\mathbb{P}_x(M_n\in\mathcal{S}_k)}\left(p(r,s)+\sum_{y\notin\mathcal{S}_k}\frac{\mathbb{P}_x(M_{n-1}=y)}{\mathbb{P}_x(M_{n-1}=s)}p(y,s)\mathbb{1}_{\{r=s\}}\right),\; r,s\in\mathcal{S}_k.$$

Since M converges in distribution,

$$p^{(x,n)}_{|\mathcal{S}_k}(r,s) \;\overset{n\to\infty}{\to}\; p(r,s)+p(s,\mathcal{S}_k^c)\mathbb{1}_{\{r=s\}} \;=\; p_{|\mathcal{S}_k}(r,s)$$

for every $r,s \in \mathcal{S}_k$. Thus, the result of Martin & Randall coincides with ours whenever the processes $\overline{Y}$ and $\mathbb{P}_x(M_n \in \cdot\,|M_n \in \mathcal{S}_k)_{n\geq|\mathcal{S}|}$ are Markovian and gives the analogous statement for the limiting Markov chains.

Regarding again Equation (6.23) and recalling Lemma 6.5.8, for the author it seems reasonable that $\overline{Y}$ mixes *asymptotically with the same rate* as X (on a logarithmic scale) because of several reasons:

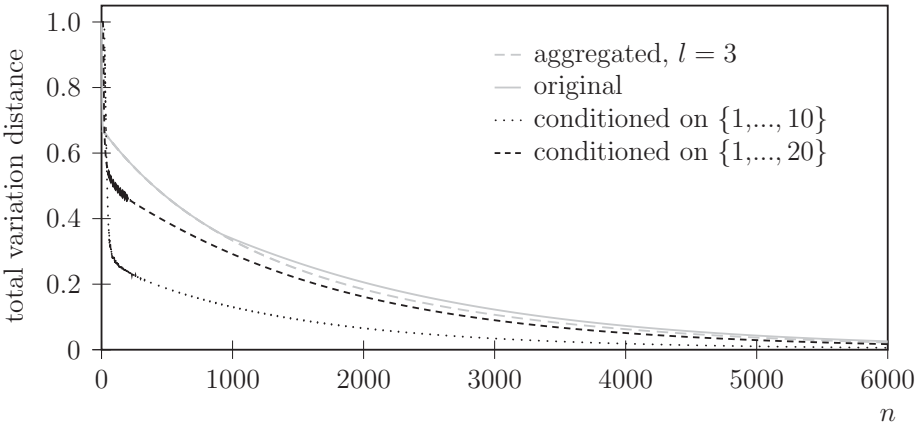

Figure 6.5.: Total variation distance between the distribution of the Markov chain with transition probabilities given in Example 6.5.6 (b) and its stationary distribution (solid, light gray line); the total variation distance between the aggregated chain for $l = 3$ and its stationary distribution (dashed, light gray line), and the total variation distance between the process conditioned on being in $\{1,...10\}$ (dotted) resp. in $\{1,...,20\}$ (dashed) and their conditional stationary distribution.

Limiting Chains: According to Corollary 6.1.7, the limiting Markov chain of $\overline{Y}$ mixes with the same rate as X. According to Corollary 6.1.3, the limiting Markov chain of $\mathbb{P}_x(M_n \in \cdot\, | M_n \in \mathcal{S}_k)_{n \geq |\mathcal{S}|}$ mixes with a rate smaller than the one of X.

Time-Inhomogeneous Chains: We can understand $\mathbb{P}_x(M_n \in \cdot\, | M_n \in \mathcal{S}_k)_{n \geq |\mathcal{S}|}$ as the distribution of a time-inhomogeneous Markov chain (whose transition probabilities depend also on x). In this case there are bounds on the mixing rates in terms of the mixing rate of the limiting chain (see e.g. the Main Theorem in [33]). This precise bound is to rough for our situation but can be improved at least in special cases.

Variational Characterization: Recall the proof of Theorem 6.1.1, where the function f which minimizes the ratio of Dirichlet form and variance (and therefore specifies the spectral gap) is constant on every valley. Hence, only pairs of states from *different* valleys contribute to the spectral gap. This indicates that the mixing *between* valleys is slower than the mixing *within* valleys.

We conclude with the example form above, studied in regard of the mixing performance of the distribution $\mathbb{P}_x(M_n \in \cdot\, | M_n \in \mathcal{S}_k)_{n \geq |\mathcal{S}|}$.

Example 6.5.12. Regard again Example 6.5.6 (b). For $N = 30$ and $l = 3$ we have three subspaces, each consisting of a birth and death process with drift towards the center of the subspace. Due to this drift, in terms of energies the subspaces are valleys around local minima. Figure 6.5 shows the total variation distance between the original resp. the aggregated chain and their limiting distributions (compare Figure 6.4) and, most notably, the variation distance between $\mathbb{P}_N(X_n \in \cdot\, | X_n \in \{1,...,10\})_{n \geq 21}$ resp. $\mathbb{P}_N(X_n \in \cdot\, | X_n \in \{1,...,20\})_{n \geq 11}$ and their limiting distributions. What should be noticed is that, first, the process conditioned on one subvalley mixes much faster than the aggregated chain and, second, even the process conditioned on the union of two subvalleys mixes faster than the aggregated process.

7. Comparison of Hitting and Cover Times

7.1. Cover Time

Another characterizing quantity is the *cover time* of a process, defined by

$$\tau_{cov} \;:=\; \inf\{n \geq 0 | \{X_0,...,X_n\} = \mathcal{S}\} \;=\; \max_{s\in\mathcal{S}} \tau_s.$$

As usual, fix $1 \leq i \leq \mathfrak{n}$ and write $\overline{Y} := \overline{Y}^{(i)}$ and $V(m) := V^{(i)}(m)$ for every $m \in \mathcal{S}^{(i)}$. Defining the cover times τ^X_{cov} and $\tau^{\overline{Y}}_{cov}$ for the processes X and $\overline{Y}$, we obviously have

$$\tau^X_{cov} \;\geq\; \tau^{\overline{Y}}_{cov} \quad \text{a.s.}$$

since by the time every state has been visited, every valley must have been visited.

Now we want to show that, nevertheless, the cover time of $\overline{Y}$ is of the same order as the cover time of X. Obviously, if there is a valley $V(m)$ containing some state s with an energy higher than the energy barrier arising from the MB-definition, then this state will most likely not be visited during a stay in $V(m)$. Thus, the above conjecture can only hold true if the valleys coincide with their inner part $V_<$. As every finite trajectory stays most likely beyond the energy level given by the trap model, it is no big restriction to assume the energy landscape to satisfy

(A) $V(m) = V_<(m)$ for every $m \in \mathcal{S}^{(i)}$.

Recall that for $m \in \mathcal{S}^{(i)}$,

$$\tau^{(i)}_{V(m)} \;:=\; \inf\left\{k \geq 1 | Y^{(i)}_k = m\right\}.$$

Theorem 7.1.1. *Let the energy function satisfy (A). Let m_1 be a metastable state with the deepest valley at level i and let it be unique. Define, for $m \in \mathcal{S}^{(i)}$,*

$$\begin{aligned} B_m \;&:=\; \left\{\text{for every } s \in V(m),\ \text{there exists some } \sigma_{\tau^{(i)}_{V(m)}} \leq n < \sigma_{\tau^{(i)}_{V(m)}+1} \text{ such that } X_n = s\right\} \\ B \;&:=\; \bigcap_{m\in\mathcal{S}^{(i)}} B_m. \end{aligned}$$

Then $\mathbb{P}_x(B) \to 1$ as $\beta \to \infty$ for every initial state $x \in \mathcal{S}$, and

$$\frac{\mathbb{E}_x\left(\tau^X_{cov}\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\right)} \;=\; \frac{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\right)} \;+\; \frac{\mathbb{E}_x\left(\left(\tau^X_{cov} - \tau^{\overline{Y}}_{cov}\right)\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\right)}$$

with

$$\frac{\mathbb{E}_x\left(\left(\tau^X_{cov} - \tau^{\overline{Y}}_{cov}\right)\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\right)} \;\leq\; \frac{\mathbb{P}_x\left(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m_1\right) + o(1)}{\mathbb{P}_x\left(\overline{Y}_{\tau^{\overline{Y}}_{cov}} \neq m_1\right) + o(1)} \tag{7.1}$$

as $\beta \to \infty$.

Since in most energy landscapes the numerator of the right-hand side of (7.1) is close to 0 and therefore the denominator is close to 1, this confirms the following heuristic: Regard the event that during the first visit of every valley it is completely sampled. Then the difference between the two cover times is in average smaller than the cover time of $\overline{Y}$ and the fraction becomes very small when β is large. The proof relies on the fact that, first, with high probability all states in a given valley are visited during the stay in this valley and that, second, once every valley is visited, for X to cover the whole state space it just has to scan the last valley. But scanning this valley needs less time than leaving it and scanning all the previous valleys needs more time than leaving the deepest of them. Both expected times can be expressed with the help of Theorem 2.2.2.

Proof: The first part uses the well known fact that $\mathbb{P}_r(\tau_s > \sigma_1) \to 0$ as $\beta \to \infty$ for every r and s in the same valley. Thus,

$$\mathbb{P}_x(B) \;\geq\; \prod_{m\in\mathcal{S}^{(i)}} \min_{r\in V(m)} \mathbb{P}_r\left(\tau_s < \sigma_1 \text{ for every } s \in V(m)\right) \;\to\; 1.$$

For the second part, let m_2 be the metastable state with the second deepest valley at level i. Notice that, as $\beta \to \infty$,

$$\begin{aligned}
\mathbb{E}_x\left((\tau^X_{cov} - \tau^{\overline{Y}}_{cov})\mathbb{1}_B\right) &= \sum_{m\in\mathcal{S}^{(i)}} \mathbb{E}_x\left((\tau^X_{cov} - \tau^{\overline{Y}}_{cov})\mathbb{1}_B | \overline{Y}_{\tau^{\overline{Y}}_{cov}} = m\right) \mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m) \\
&\leq \sum_{m\in\mathcal{S}^{(i)}} \max_{r\in V(m)} \mathbb{E}_r(\sigma_1)\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m) \\
&= \sum_{m\in\mathcal{S}^{(i)}\setminus N^{(i)}} e^{\beta(E(s_m)-E(m)+o(1))}\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m) + \sum_{m\in N^{(i)}} \frac{\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m)}{1-p(m,m)} \\
&\leq e^{\beta(E(s_{m_1})-E(m_1)+o(1))}\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m_1) \\
&\quad + e^{\beta(E(s_{m_2})-E(m_2)+o(1))}\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} \neq m_1)
\end{aligned}$$

by use of Theorem 2.2.2. Analogously,

$$\mathbb{E}_x(\tau^{\overline{Y}}_{cov}) \;\geq\; e^{\beta(E(s_{m_1})-E(m_1)+o(1))}\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} \neq m_1) + e^{\beta(E(s_{m_2})-E(m_2)+o(1))}\mathbb{P}_x(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m_1).$$

Combining these two estimates and using the abbreviations

$$\begin{aligned}
a &:= e^{\beta(E(s_{m_1})-E(m_1)+o(1))} \\
b &:= e^{\beta(E(s_{m_2})-E(m_2)+o(1))} \\
p &:= \mathbb{P}_x\left(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m_1\right) \\
1-p &:= \mathbb{P}_x\left(\overline{Y}_{\tau^{\overline{Y}}_{cov}} \neq m_1\right),
\end{aligned}$$

we obtain

$$\frac{\mathbb{E}_x\left(\left(\tau^X_{cov} - \tau^{\overline{Y}}_{cov}\right)\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\right)} \;\leq\; \frac{pa+(1-p)b}{(1-p)a+pb} \;=\; \frac{p+(1-p)\frac{b}{a}}{1-p+p\frac{b}{a}}.$$

Since $\frac{b}{a} \to 0$ as $\beta \to \infty$, this gives for large β

$$\frac{\mathbb{E}_x\left(\left(\tau^X_{cov} - \tau^{\overline{Y}}_{cov}\right)\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{cov}\right)} \;\leq\; \frac{p+o(1)}{1-p+o(1)} \;=\; \frac{\mathbb{P}_x\left(\overline{Y}_{\tau^{\overline{Y}}_{cov}} = m_1\right)+o(1)}{\mathbb{P}_x\left(\overline{Y}_{\tau^{\overline{Y}}_{cov}} \neq m_1\right)+o(1)}.$$

□

7.2. Hitting Time

The same intuition holds true for the *hitting time* of the over all minimum $m^{(\mathfrak{n})}$. By the time X hits this state, $\overline{Y}$ as well had hit it, thus, $\tau^X_{m^{(\mathfrak{n})}} \geq \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}$ a.s. By the first time the valley around this minimum is visited, it just needs another few steps for X to decline to $m^{(\mathfrak{n})}$. Of course X may have to cross several sub-valleys which may take some time, but leaving a sub-valley is exponentially faster than leaving the valleys visited before since the depth is smaller.

Before giving a result similar to the above one for the cover time, we want to note the following:

Proposition 7.2.1. *For every $x \in \mathcal{S}$ it holds true that*

$$\mathbb{P}_x\left(\tau^X_{m^{(\mathfrak{n})}} - \tau^{\overline{Y}}_{m^{(\mathfrak{n})}} > \sigma_1'\right) \;\to\; 0$$

as $\beta \to \infty$, where σ_1' should denote the time spend in $V(m^{(\mathfrak{n})})$.

This proposition of course is an easy consequence of the fact that starting anywhere in the valley, the minimum will be reach a.a.s. earlier than the valley is left again.

Theorem 7.2.2. *Define*

$$B \;:=\; \left\{E(X_k) \leq E(z^*(X_n, m^{(\mathfrak{n})})) \text{ for every } \tau^{\overline{Y}}_{m^{(\mathfrak{n})}} \leq n \leq k \leq \tau^X_{m^{(\mathfrak{n})}}\right\},$$

and let m^* be the deepest sub-valley in $V(m^{(\mathfrak{n})})$ and $D(m)$ the depth $E(s_m) - E(m)$ of a valley $V(m)$. Then $\mathbb{P}_x(B) \to 1$ as $\beta \to \infty$ for every initial state $x \in \mathcal{S}$, and

$$\frac{\mathbb{E}_x\left(\tau^X_{m^{(\mathfrak{n})}} \mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)} \;=\; \frac{\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}} \mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)} \;+\; \frac{\mathbb{E}_x\left(\left(\tau^X_{m^{(\mathfrak{n})}} - \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right) \mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)}$$

with

$$\frac{\mathbb{E}_x\left(\left(\tau^X_{m^{(\mathfrak{n})}} - \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right) \mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)} \;\to\; 0$$

as $\beta \to \infty$ as long as

$$\liminf_{\beta\to\infty} \mathbb{P}_x\left(D(\overline{Y}_n) \geq D(m^*) + \varepsilon \text{ for some } 0 \leq n < \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right) \;>\; 0$$

for some $\varepsilon > 0$.

Proof: B is the event that once the valley around $m^{(\mathfrak{n})}$ is visited, the process will not visit states with higher energy than necessary to reach $m^{(\mathfrak{n})}$. Similar to the proof of Theorem 2.1.2, it holds true that

$$\begin{aligned}
\mathbb{P}_x(B^c) \;&\leq\; \sum_{r \in V(m^{(\mathfrak{n})})} \mathbb{P}_r\left(E(X_k) > E(z^*(r, m^{(\mathfrak{n})})) \text{ for some } 0 \leq k \leq \tau_{m^{(\mathfrak{n})}}\right) \\
&=\; \sum_{r \in V(m^{(\mathfrak{n})})} \mathbb{P}_r\left(\tau_s < \tau_{m^{(\mathfrak{n})}} \text{ for some } s \text{ with } E(s) > E(z^*(r, m^{(\mathfrak{n})}))\right) \\
&\to\; 0
\end{aligned}$$

as $\beta \to \infty$, and thus B occurs a.a.s.

To estimate the time needed to decline to $m^{(\mathfrak{n})}$ once $V(m^{(\mathfrak{n})})$ is entered, let r_0 be the state via which this valley is entered at time $\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}$. For any $r \in V(m^{(\mathfrak{n})})$ we write $r \leftrightarrow_B r_0$ if, on the set B, r is reachable from r_0 and r_0 is reachable from r. That is $r \leftrightarrow_B r_0$, if there is a path $\gamma = (\gamma_0,...,\gamma_k) \in \Gamma(r_0, r)$ with $E(\gamma_j) \le E(z^*(\gamma_i, m^{(\mathfrak{n})}))$ for every $0 \le i, j \le k$. With this, let

$$S_{r_0} \;:=\; \Big\{ r \in V(m^{(\mathfrak{n})}) \,|\, r \leftrightarrow_B r_0 \Big\}$$

be the set of states which, on B, can be reached from r_0 and from which r_0 can be reached. Observe that in particular

$$E(z^*(r, r_0)) \;\le\; E(z^*(r, m^{(\mathfrak{n})})) \wedge E(z^*(r_0, m^{(\mathfrak{n})}))$$

and

$$E(z^*(r, m^{(\mathfrak{n})})) \;\le E(z^*(r, s)) \vee E(z^*(s, m^{(\mathfrak{n})})) \;=\; E(z^*(s, m^{(\mathfrak{n})})) \;\le\; E(z^*(r, m^{(\mathfrak{n})}))$$

holds true for every $r, s \in S_{r_0}$. Thus, $E(z^*(r, m^{(\mathfrak{n})})) = E(z^*(s, m^{(\mathfrak{n})}))$ for every $r, s \in S_{r_0}$. The process samples S_{r_0} until some $r_1 \notin S_{r_0}$ is reached. But $r_1 \notin S_{r_0}$ provides $E(z^*(r_0, r_1)) > E(z^*(r_1, m^{(\mathfrak{n})}))$. In this case

$$\begin{aligned} E(z^*(r_1, m^{(\mathfrak{n})})) \;&<\; E(z^*(r_1, r_0)) \\ &\le\; E(z^*(r_1, r)) \vee E(z^*(r, r_0)) \\ &\le\; E(z^*(r_1, r)) \vee E(z^*(r, m^{(\mathfrak{n})})) \\ &\le\; E(z^*(r_1, r)) \vee E(z^*(r_1, m^{(\mathfrak{n})})) \\ &=\; E(z^*(r_1, r)) \end{aligned}$$

for every $r \in S_{r_0}$. Thus, once reaching $r_1 \notin S_{r_0}$, on the set B the set S_{r_0} will not be visited again. Starting the above procedure again in r_1, inductively we decent down to $r_k = m^{(\mathfrak{n})}$ in a finite number k of steps. For determining $\mathbb{E}_{r_0}(\tau^X_{m^{(\mathfrak{n})}} \mathbb{1}_B)$ we therefore have to determine the expected time the process spends in every S_{r_i} defined above. We have to distinguish two different types of sets S_{r_i}:

(1) $E(r) = E(r_i)$ for every $r \in S_{r_i}$,

(2) there is some $r \in S_{r_i}$ with $E(r) < E(r_i)$.

In the first case, for every $r \in S_{r_i}$ there is a non-increasing path leaving S_{r_i} in some $r_{i+1} \notin S_{r_i}$ with $E(z^*(r_{i+1}, m^{(\mathfrak{n})})) < E(z^*(r_i, m^{(\mathfrak{n})}))$. Therefore, the expected residence time in S_{r_i} converges to a finite constant when β tends to ∞.

The second case is more complicated. When there are local minima in S_{r_i}, then there are subvalleys which must be left and whose residence times converge to ∞. In the following we will argue that nevertheless the residence time in S_{r_i} is at most $\exp(\beta(E(s_{m^*}) - E(m^*) + o(1)))$, what should not be surprising.

To this end, define on $S_{r_i} \backslash z^*(r_i, m^{(\mathfrak{n})})$ the relation

$$r \sim_{r_i} s \;:\Leftrightarrow\; E(z^*(r, s)) < E(z^*(r_i, m^{(\mathfrak{n})})).$$

Since $E(r) < E(z^*(r_i, m^{(\mathfrak{n})}))$ for $r \in S_{r_i} \backslash z^*(r_i, m^{(\mathfrak{n})})$ the relation $\sim_{r_i}$ is reflexive. Obviously it is symmetric and the transience arises from

$$E(z^*(r, z)) \;\le\; E(z^*(r, s)) \vee E(z^*(s, z)) \;<\; E(z^*(r_i, m^{(\mathfrak{n})}))$$

for $r \sim_{r_i} s \sim_{r_i} z$. Thus, $S_{r_i} \backslash z^*(r_i, m^{(\mathfrak{n})})$ separates into equivalence classes of $\sim_{r_i}$ with $E(z^*(r, s)) = E(z^*(r_i, m^{(\mathfrak{n})}))$ for r and s in different equivalence classes. Every such equivalence class $[m]$ is a subvalley of $V(m^{(\mathfrak{n})})$ of some order $l(m)$, centered around its minimum m, with the essential saddles

between them being all of the height $E(z^*(r_i, m^{(\mathfrak{n})}))$. Thus, $l(m)$ can chosen to be the same level for every equivalence class in $S_{r_i} \backslash z^*(r_i, m^{(\mathfrak{n})})$ but varies with r_i. Since the essential saddle between these subvalleys is of the same hight as the essential saddle to $m^{(\mathfrak{n})}$, the AAC $Y^{(l(m))}$ will leave this set of minima and saddles almost surely in a finite time whose expectation is bounded. Therefore, as $\beta \to \infty$, the time needed for X to leave S_{r_i} is bounded above by the time $Y^{(l(m))}$ needs to leave this set times the maximal time to leave a single subvalley which is neglected by $Y^{(l(m))}$. That is why the expected time for the descent from r_0 to $m^{(\mathfrak{n})}$ on the set B can be bounded by

$$\mathbb{E}_{r_0}\left(\left(\tau^X_{m^{(\mathfrak{n})}} - \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)\mathbb{1}_B\right) \;\leq\; \mathcal{O}(1)\cdot \max_{m\in V(m^{(\mathfrak{n})})\backslash\{m^{(\mathfrak{n})}\}} \mathbb{E}_m(\sigma_1^{(l(m))}) \;\leq\; e^{\beta(E(s_{m^*})-E(m^*)+o(1))}$$

as $\beta \to \infty$. Define, for $\varepsilon > 0$,

$$C \;:=\; C(\varepsilon) \;:=\; \left\{D(\overline{Y}_n) \geq D(m^*) + \varepsilon \text{ for some } 0 \leq n < \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right\}.$$

As in the previous theorem, we obtain

$$\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right) \;\geq\; e^{\beta(D(m^*)+\varepsilon+o(1))}\mathbb{P}_x(C) + e^{\beta(D(m')+o(1))}\left(1-\mathbb{P}_\lambda(C)\right),$$

where m' is the valley with the lowest depth. Hence, we have for the quotient

$$\begin{aligned}
\frac{\mathbb{E}_x\left(\left(\tau^X_{m^{(\mathfrak{n})}} - \tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)\mathbb{1}_B\right)}{\mathbb{E}_x\left(\tau^{\overline{Y}}_{m^{(\mathfrak{n})}}\right)} \;&\leq\; \frac{e^{\beta(D(m^*)+o(1))}}{e^{\beta(D(m^*)+\varepsilon+o(1))}\mathbb{P}_x(C) + e^{\beta(D(m')+o(1))}(1-\mathbb{P}_x(C))} \\
&\leq\; \frac{1}{e^{\beta(\varepsilon+o(1))}\mathbb{P}_x(C) + e^{\beta(D(m')-D(m^*)+o(1))}(1-\mathbb{P}_x(C))} \\
&\to\; 0
\end{aligned}$$

as $\beta \to \infty$ as long as $\liminf_{\beta\to\infty} \mathbb{P}_x(C(\varepsilon)) > 0$ for some $\varepsilon > 0$. □

A. Auxiliary Results and Tools

A.1. Auxiliaries for Chapter 2

Lemma A.1.1. *Let $(\mathbf{P}, \pi)$ be a reversible pair on $\mathcal{S}$. For any one-dimensional subgraph $\Delta = (\omega_0,...,\omega_k)$ of $\mathcal{S}$ and corresponding transition matrix $\widetilde{\mathbf{P}} = \mathbf{P}_{|\Delta}$, we have*

$$\mathbb{P}_{\omega_0}(\tau_{\omega_0} > \tau_{\omega_k}) \;\geq\; \widetilde{\mathbb{P}}_{\omega_0}(\tau_{\omega_0} > \tau_{\omega_k}) \;=\; \left(\sum_{i=1}^{k} \frac{\pi(\omega_0)}{\pi(\omega_i)} \frac{1}{p(\omega_i, \omega_{i-1})}\right)^{-1}.$$

Proof: We just prove the second inequality, as its proof (unlike that of the first inequality) is not given in [12]. The restricted process $\widetilde{\mathbb{P}}$ is a birth-and-death process whose transition probabilities are given by

$$\begin{aligned}
\tilde{p}(\omega_i, \omega_{i+1}) &= p(\omega_i, \omega_{i+1}), && 0 \leq i \leq k-1,\\
\tilde{p}(\omega_i, \omega_{i-1}) &= p(\omega_i, \omega_{i-1}), && 1 \leq i \leq k,\\
\tilde{p}(\omega_i, \omega_i) &= 1 - \tilde{p}(\omega_i, \omega_{i+1}) - \tilde{p}(\omega_i, \omega_{i-1}), && 1 \leq i \leq k-1,\\
\tilde{p}(\omega_0, \omega_0) &= 1 - \tilde{p}(\omega_0, \omega_1),\\
\tilde{p}(\omega_k, \omega_k) &= 1 - \tilde{p}(\omega_k, \omega_{k-1}).
\end{aligned} \tag{A.1}$$

It is a well-known fact in Markov chain theory and can be verified easily that it has stationary measure

$$\tilde{\pi}(\omega_0) \;=\; 1 \quad \text{and} \quad \tilde{\pi}(\omega_i) \;=\; \frac{\tilde{p}(\omega_0, \omega_1) \cdot ... \cdot \tilde{p}(\omega_{i-1}, \omega_i)}{\tilde{p}(\omega_1, \omega_0) \cdot ... \cdot \tilde{p}(\omega_i, \omega_{i-1})} \quad \text{for } 1 \leq i \leq k,$$

and further satisfies

$$\tilde{\pi}(\omega_i) \;=\; \frac{\pi(\omega_i)}{\pi(\omega_0)}, \quad 0 \leq i \leq k. \tag{A.2}$$

Furthermore, the probability of hitting ω_k before ω_0 when starting in ω_1 may explicitly be calculated by solving the system of linear equations given by

$$\begin{aligned}
\mathbb{P}_{\omega_i}(\tau^0_{\omega_k} < \tau^0_{\omega_0}) \;=\; &\tilde{p}(\omega_i, \omega_i)\mathbb{P}_{\omega_i}(\tau^0_{\omega_k} < \tau^0_{\omega_0}) + \tilde{p}(\omega_i, \omega_{i+1})\mathbb{P}_{\omega_{i+1}}(\tau^0_{\omega_k} < \tau^0_{\omega_0})\\
&+ \tilde{p}(\omega_i, \omega_{i-1})\mathbb{P}_{\omega_{i-1}}(\tau^0_{\omega_k} < \tau^0_{\omega_0})
\end{aligned}$$

for $1 \leq i \leq k-1$ and $\mathbb{P}_{\omega_0}(\tau^0_{\omega_k} < \tau^0_{\omega_0}) = 0$, $\mathbb{P}_{\omega_k}(\tau^0_{\omega_k} < \tau^0_{\omega_0}) = 1$. One gets

$$\widetilde{\mathbb{P}}_{\omega_1}(\tau_{\omega_0} > \tau_{\omega_k}) \;=\; \widetilde{\mathbb{P}}_{\omega_1}(\tau^0_{\omega_0} > \tau^0_{\omega_k}) \;=\; \left(1 + \sum_{i=1}^{k-1} \frac{\tilde{p}(\omega_0, \omega_1)}{\tilde{\pi}(\omega_i)\tilde{p}(\omega_i, \omega_{i+1})}\right)^{-1}.$$

Using that $(\mathbf{P}, \pi)$ is a reversible pair, we now obtain

$$\begin{aligned}
\widetilde{\mathbb{P}}_{\omega_0}(\tau_{\omega_0} > \tau_{\omega_k}) &= \tilde{p}(\omega_0,\omega_1)\widetilde{\mathbb{P}}_{\omega_1}(\tau_{\omega_0} > \tau_{\omega_k}) \\
&= \left(\frac{1}{\tilde{p}(\omega_0,\omega_1)}\left(1+\sum_{i=1}^{k-1}\frac{\tilde{p}(\omega_0,\omega_1)}{\tilde{\pi}(\omega_i)\tilde{p}(\omega_i,\omega_{i+1})}\right)\right)^{-1} \\
&= \left(\sum_{i=0}^{k-1}\frac{\pi(\omega_0)}{\pi(\omega_i)p(\omega_i,\omega_{i+1})}\right)^{-1} \\
&= \left(\sum_{i=0}^{k-1}\frac{\pi(\omega_0)}{\pi(\omega_{i+1})p(\omega_{i+1},\omega_i)}\right)^{-1} \\
&= \left(\sum_{i=1}^{k}\frac{\pi(\omega_0)}{\pi(\omega_i)p(\omega_i,\omega_{i-1})}\right)^{-1},
\end{aligned}$$

where (A.1) and (A.2) have been utilized for the last identity. □

Proposition A.1.2 (see Corollary 1.6 in [12]). *Given $I \subset \mathcal{S}$ and distinct $x, z \in \mathcal{S}\backslash I$,*

$$\mathbb{P}_x(\tau_z < \tau_I) = \frac{\mathbb{P}_x(\tau_z < \tau_{I\cup\{x\}})}{\mathbb{P}_x(\tau_{I\cup\{z\}} < \tau_x)}$$

holds true.

Proof: We first show that, for all $u \in [0,1]$,

$$\left(\sum_{t\geq 0} u^t\,\mathbb{P}_x(\tau_z = t < \tau_I)\right)\left(1-\sum_{s\geq 0} u^s\,\mathbb{P}_x(\tau_x = s \leq \tau_{I\cup\{z\}})\right) = \sum_{t\geq 0} u^t\,\mathbb{P}_x(\tau_z = t < \tau_{I\cup\{x\}}), \quad \text{(A.3)}$$

which in fact follows directly from

$$\begin{aligned}
&\sum_{s,t\geq 0} u^{t+s}\,\mathbb{P}_x(\tau_x = s \leq \tau_{I\cup\{z\}})\,\mathbb{P}_x(\tau_z = t < \tau_I) \\
&\qquad = \sum_{t\geq 0} u^t\left(\mathbb{P}_x(\tau_z = t < \tau_I) - \mathbb{P}_x(\tau_z = t < \tau_{I\cup\{x\}})\right). \qquad \text{(A.4)}
\end{aligned}$$

To see the latter, we compute

$$\begin{aligned}
&\sum_{s,t\geq 0} u^{t+s}\,\mathbb{P}_x(\tau_x = s \leq \tau_{I\cup\{z\}})\,\mathbb{P}_x(\tau_z = t < \tau_I) \\
&\qquad = \sum_{t\geq 0} u^t \sum_{s\leq t}\mathbb{P}_x(\tau_z = t-s < \tau_I)\,\mathbb{P}_x(\tau_x = s \leq \tau_{I\cup\{z\}}) \\
&\qquad = \sum_{t\geq 0} u^t \sum_{s\leq t}\mathbb{P}_x(\tau_z = t < \tau_I | \tau_x = s \leq \tau_{I\cup\{z\}})\,\mathbb{P}_x(\tau_x = s < \tau_{I\cup\{z\}}) \\
&\qquad = \sum_{t\geq 0} u^t \sum_{s\leq t}\mathbb{P}_x(s = \tau_x \leq \tau_z = t < \tau_I) \\
&\qquad = \sum_{t\geq 0} u^t\,\mathbb{P}_x(\tau_x \leq \tau_z = t < \tau_I),
\end{aligned}$$

which proves (A.4), for

$$\mathbb{P}_x(\tau_z = t < \tau_I) - \mathbb{P}_x(\tau_z = t < \tau_{I\cup\{x\}}) \;=\; \mathbb{P}_x(\tau_x \le \tau_z = t < \tau_I).$$

Upon choosing $u = 1$ in (A.3), we obtain for the left-hand side

$$\begin{aligned}\left(\sum_{t\ge 0}\mathbb{P}_x(\tau_z = t < \tau_I)\right)\left(1 - \sum_{t\ge 0}\mathbb{P}_x(\tau_x = t \le \tau_{I\cup\{z\}})\right) \;&=\; \mathbb{P}_x(\tau_z < \tau_I)\Big(1 - \mathbb{P}_x(\tau_x \le \tau_{I\cup\{z\}})\Big)\\ &=\; \mathbb{P}_x(\tau_z < \tau_I)\,\mathbb{P}_x(\tau_x > \tau_{I\cup\{z\}}),\end{aligned}$$

while the right-hand side equals

$$\sum_{t\ge 0}\mathbb{P}_x(\tau_z = t < \tau_{I\cup\{x\}}) \;=\; \mathbb{P}_x(\tau_z < \tau_{I\cup\{x\}}).$$

Since $\mathbb{P}_x(\tau_x > \tau_{I\cup\{z\}})$ must be positive, the assertion follows. □

A.2. Toolbox for Spectral Gap Analysis

Proposition A.2.1. *Let $(\mathbf{P}, \pi)$ be a reversible pair of a finite Markov chain M.*

(a) Is $p(x,x) \ge \frac{1}{2}$ for every x, then all eigenvalues of $\mathbf{P}$ are positive and $\lambda_2 = \lambda_\star$.

(b) Is M ergodic, then $\big(\max_x d_{TV}(\mathbb{P}_x^{M_n}, \pi)\big)^{\frac{1}{n}} \to \lambda_\star$ as $n \to \infty$.

(c) Is M ergodic and $\lambda_\star > 0$, then $(t_{rel} - 1)\log\left(\frac{1}{2\varepsilon}\right) \le \tau_{\mathrm{mix}}(\varepsilon) \le \log\left(\frac{1}{\varepsilon\pi_{min}}\right) t_{rel}$ for every $\varepsilon > 0$, where $\pi_{min} := \min_x \pi(x)$.

(d) $\gamma = \min_f \frac{\mathcal{E}(f)}{Var_\pi(f)}$, where $Var_\pi(f) := \sum_s (f(s) - \sum_r f(r)\pi(r))^2\pi(s)$ and the minimum ranges over all functions $f : \mathcal{S} \to \mathbb{R}$ with $Var_\pi(f) \ne 0$ (variational characterization).

Whereas (a) is trivial ($\mathbf{P}$ can be written as $\frac{1}{2}\tilde{\mathbf{P}} + \frac{1}{2}I$, for the transition matrix $\tilde{\mathbf{P}} = 2\mathbf{P} - I$) for the proof of the remaining parts we just give references: (b) is Corollary 12.6, (c) is Theorem 12.3 and Theorem 12.4 in [39]. (d) is Lemma 13.12 ibidem.

There is the following generalization of part (b) for non-reversible Markov chains by Fill:

Proposition A.2.2 (Theorem 2.1 in [23])**.** *Let M be an ergodic finite Markov chain with transition matrix $\mathbf{P}$ and stationary distribution π. Then all eigenvalues of the* multiplicative reversiblization *$\sum_z p(x,z)p(y,z)\pi(y)/\pi(z)$, $x, z \in \mathcal{S}$, are positive, the second largest of which we denote by ρ. Furthermore, $d_{TV}(\mathbb{P}_x^{X_n}, \pi) \le \frac{\rho^{n/2}}{\pi_{min}}$ for every $x \in \mathcal{S}$ and $n \ge 0$.*

Furthermore, with the Dirichlet form a comparison of the gaps of two chains is possible. Again, a proof can be found in [39, Lemma 13.22].

Lemma A.2.3 (see Lemma 13.22 in [39])**.** *Let M and $\widetilde{M}$ be two reversible finite Markov chains on $\mathcal{S}$ with stationary distributions π and $\widetilde{\pi}$. If $\mathcal{E}(f) \le c\widetilde{\mathcal{E}}(f)$ for every f, then*

$$\gamma \;\le\; \widetilde{\gamma}\cdot c\max_x \frac{\widetilde{\pi}(x)}{\pi(x)}.$$

Lemma A.2.4 (see Theorem 13.20 in [39])**.** *Let M be a reversible finite Markov chain on $\mathcal{S}$ with stationary distribution π and spectral gap γ. Let $A \subset \mathcal{S}$ be non-empty and let γ_A be the spectral gap of the hit chain on A with transition probabilities $p_A(x,y) = \mathbb{P}_x(M_{\tau_A} = y)$, $x, y \in A$. Then $(\mathbf{P}_A, (\pi(\cdot \cap A))/\pi(A))$ is a reversible pair and $\gamma_A \ge \gamma$.*

A.3. Auxiliaries for Chapter 6

Lemma A.3.1. *For $n \geq 3$, the vectors*

$$\begin{aligned} v_1 &= \left(-1, \frac{1}{n-1}, \ldots, \frac{1}{n-1}\right) \\ v_2 &= \left(\frac{1}{n-1}, -1, \frac{1}{n-1}, \ldots, \frac{1}{n-1}\right) \\ &\vdots \\ v_{n-1} &= \left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}, -1, \frac{1}{n-1}\right) \end{aligned}$$

are linearly independent.

Proof: We use an inductive argument: For $n = 3$ it is an easy calculation that $(-1, \frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{2}, -1, \frac{1}{2})$ are linearly independent. Thus, assume the independence to hold true for some $n \geq 3$. In particular the system of linear equations

$$\underbrace{\left[\begin{array}{ccccc|c} -1 & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} & 0 \\ \frac{1}{n-1} & -1 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} & 0 \\ \frac{1}{n-1} & \frac{1}{n-1} & -1 & \cdots & \frac{1}{n-1} & 0 \\ \vdots & & & \ddots & \vdots & \vdots \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & -1 & 0 \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} & 0 \end{array}\right]}_{n-1 \text{ many}} \Bigg\} \; n \text{ many} \tag{A.5}$$

exhibits only the trivial solution. Passing over to $n+1$, the system of linear equations in question is

$$\underbrace{\left[\begin{array}{ccccc|c} -1 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \\ \frac{1}{n} & -1 & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \\ \frac{1}{n} & \frac{1}{n} & -1 & \cdots & \frac{1}{n} & 0 \\ \vdots & & & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & -1 & 0 \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \end{array}\right]}_{n \text{ many}} \Bigg\} \; n+1 \text{ many}$$

and changes to

$$\left[\begin{array}{ccccc|c} -1 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \\ 0 & \frac{1-n^2}{n^2} & \frac{n+1}{n^2} & \cdots & \frac{n+1}{n^2} & 0 \\ 0 & \frac{n+1}{n^2} & \frac{1-n^2}{n^2} & \cdots & \frac{n+1}{n^2} & 0 \\ \vdots & & & \ddots & \vdots & \vdots \\ 0 & \frac{n+1}{n^2} & \frac{n+1}{n^2} & \cdots & \frac{1-n^2}{n^2} & 0 \\ 0 & \frac{n+1}{n^2} & \frac{n+1}{n^2} & \cdots & \frac{n+1}{n^2} & 0 \end{array}\right] \longrightarrow \underbrace{\left[\begin{array}{ccccc|c} -1 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \\ 0 & -1 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} & 0 \\ 0 & \frac{1}{n-1} & -1 & \cdots & \frac{1}{n-1} & 0 \\ \vdots & & & \ddots & \vdots & \vdots \\ 0 & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & -1 & 0 \\ 0 & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} & 0 \end{array}\right]}_{n-1 \text{ many}} \Bigg\} \; n \text{ many}$$

when, first, adding n^{-1} times the first row to all other rows and, second, multiplying those other rows with $-\frac{n^2}{1-n^2}$. Neglecting the first row and column, this is the system in (A.5), of which we know that it has only the trivial solution. Thus, the whole system itself exhibits just the trivial solution, and the eigenvectors are linearly independent. □

Lemma A.3.2. *Let M^T be a semi-Markov chain on $\mathcal{S}$ with sojourn times $(T_n)_{n\geq 0}$ and embedded Markov chain M whose transition matrix we denote by $\mathbf{P}$. Write $S_n := T_0 + ... + T_{N^T(n)}$, $n \geq 0$. Then the process $(M_n^T, S_n - n)_{n\geq 0}$ is a homogeneous Markov chain with transition matrix*

$$q((x,s),(y,t)) \;=\; \begin{cases} \mathbb{1}_{\{x=y,t=s-1\}}, & s>1 \\ p(x,y)\mathbb{P}_y(T_0=t), & s=1, \end{cases}$$

and stationary distribution $\mu(x,k) = \pi^T(x)\mathbb{P}_x(T_0 \geq k)/\mathbb{E}_x(T_0)$, $x \in \mathcal{S}$, $k \geq 1$.

Proof: Let $m_0,...,m_{n+1} \in \mathcal{S}$ and $s_0,...,s_{n-1}, s, t \in \mathbb{N}$. If $S_n - n = s > 1$, then $N^T(n+1) = N^T(n)$ and therefore,

$$\begin{aligned} &\mathbb{P}\left(M_{n+1}^T = m_{n+1}, S_{n+1} - (n+1) = t | M_n^T = m_n,..., M_0^T = m_0, S_n - n = s,..., S_0 - 0 = s_0\right) \\ &\quad= \mathbb{P}\left(M_{N^T(n)} = m_{n+1}, S_n - (n+1) = t | M_{N^T(n)} = m_n, S_n - n = s\right) \\ &\quad= \mathbb{1}_{\{m_{n+1}=m_n, t=s-1\}}. \end{aligned}$$

If $S_n - n = s = 1$, then $N^T(n+1) = N^T(n) + 1$ and

$$\begin{aligned} &\mathbb{P}\left(M_{n+1}^T = m_{n+1}, S_{n+1} - (n+1) = t | M_n^T = m_n,..., M_0^T = m_0, S_n - n = 1,..., S_0 - 0 = s_0\right) \\ &\quad= \mathbb{P}\left(M_{N^T(n)+1} = m_{n+1}, S_n + T_{N^T(n)+1} - (n+1) = t | M_{N^T(n)} = m_n, S_n - n = 1\right) \\ &\quad= \mathbb{P}\left(M_{N^T(n)+1} = m_{n+1}, T_{N^T(n)+1} = t | M_{N^T(n)} = m_n, S_n - n = 1\right) \\ &\quad= \mathbb{P}\left(M_{N^T(n)+1} = m_{n+1}, T_{N^T(n)+1} = t | M_{N^T(n)} = m_n\right) \\ &\quad= p(m_n, m_{n+1})\mathbb{P}_{m_{n+1}}(T_0 = t). \end{aligned}$$

It remains to show the stationarity of μ, which is obtained by

$$\begin{aligned} \sum_x \sum_s q((x,s),(y,t))\mu(x,s) \;&=\; \sum_x \sum_{s>1} \mathbb{1}_{\{x=y,t=s-1\}}\mu(x,s) + \sum_x p(x,y)\mathbb{P}_y(T_0=t)\mu(x,1) \\ &=\; \mu(y,t+1) + \frac{1}{\mathbb{E}_\pi(T_0)} \sum_x p(x,y)\mathbb{P}_y(T_0=t)\pi(x)\mathbb{P}_x(T_0 \geq 1) \\ &=\; \frac{1}{\mathbb{E}_\pi(T_0)} \left(\pi(y)\mathbb{P}_y(T_0 \geq t+1) + \mathbb{P}_y(T_0=t)\sum_x p(x,y)\pi(x)\right) \\ &=\; \frac{\pi(y)}{\mathbb{E}_\pi(T_0)}\mathbb{P}_y(T_0 \geq t). \end{aligned}$$

□

List of Symbols

$(\beta_n)_{n\geq 0}$	a proper annealing schedule, page 37
β	system-parameter representing the inverse temperature, page 11
Δ	depth of the second deepest valley of E, page 70
Δ_V	depth of the second deepest (sub)-valley in V, page 74
$\eta_{1,\varepsilon}$	connectivity parameter, page 60
$\eta_{2,\varepsilon}$	connectivity parameter, page 60
$\eta_{3,\varepsilon}$	connectivity parameter, page 60
γ	spectral gap, page 69
$\Gamma(r,s)$	set of finite self-avoiding paths from r to s having positive probability, page 13
$\Gamma^*(r,s)$	set of minimal paths from r to s, page 13
$\gamma^{\overline{Z}}$	spectral gap of $\overline{Z}$, page 79
γ_β	system-parameter, page 11
$\gamma_{\lvert V}$	spectral gap of the (to a valley V) restricted chain, page 74
$\gamma_\star$	absolute spectral gap, page 69
$\lambda(V)$	Perron-Frobenius eigenvalue of $(p(x,y))_{x,y\in V}$, page 33
$\lambda_\star^{T,n}$	convergence rate of the by T decelerated chain, page 93
λ_2	second largest eigenvalue of a stochastic matrix, page 69
$\lambda_\star$	in terms of moduli second largest eigenvalue of a stochastic matrix, page 69
$\mathcal{E}(h)$	Dirichlet form of h, page 25
$\mathcal{S}^{(i)}$	state space of the AAC and AC consisting of the metastable states which are not attracted until level i and the non-assigned states at level i, page 36
$\mathcal{V}_n$	path-dependent MB, page 3
$\mathfrak{n}$	umber of metastable states at level 1, page 13
ν	quasi-stationary distribution, page 33
$\overline{M}$	decelerated version of the Markov chain M, page 83
$\overline{Y}^{(i)}$	aggregated chain at level i, page 36
$\overline{Z}$	hit chain on $\mathcal{S}^{(i)}$, page 79
$\partial V^{(i)}(m)$	inner boundary of the valley $V^{(i)}(m)$, page 20
$\partial^+ V^{(i)}(m)$	outer boundary of the valley $V^{(i)}(m)$, page 20
$\mathbf{P}_{\lvert V}$	transition matrix of the (to a valley V) restricted chain, page 74
$\mathbf{P}_{\mathcal{S}^{(i)}}$	transition matrix of $\overline{Z}$, page 79
π^T	limiting distribution of the semi-Markov chain M^T, page 99
$\sigma_n^{(i)}$	sojourn times of the aggregated chain at level i, page 36
τ_A	recurrence time (≥ 1) of the set A, page 7
τ_A^0	recurrence time (≥ 0) of the set A, page 7
$\tau_{\text{mix}}^M(\varepsilon)$	mixing time of the Markov chain M, page 69
τ_{cov}	cover time, page 117

$\tau^{(i)}_{V(m)}$	recurrence time (≥ 1) of the AAC, page 61
$\tilde{\varepsilon}(x, y, z, \beta)$	bound for the probability to reach z before y when starting in x, page 23
$\varepsilon(x, y, z, \beta)$	bound for the probability to reach z before y when starting in x, page 23
$\widehat{\tau}^{(i)}$	recurrence time (≥ 1) of the asymptotic jump chain, page 63
$\widehat{p}(r, s)$	transition probabilities of the asymptotic jump chain, page 38
$\widehat{Y}^{(i)}$	asymptotic jump chain, page 40
$\widetilde{\mathbf{P}}$	transition matrix of the Markov chain restricted to some subgraph Δ, page 23
$\xi^{(i)}_n$	entrance times, page 28
$\zeta^{(i)}_n$	exit times, page 28
$a \vee b$	maximum of a and b, page 7
$a \wedge b$	minimum of a and b, page 7
E	energy function, page 11
$I(s, m)$	cumulative activation energy, page 22
$I(s_1,..., s_n)$	cumulative activation energy along the path $s_1,..., s_n$, page 22
$l(i)$	minimal level at which $m^{(i)}$ becomes attracted, page 14
$M^{(i)}$	set of metastable states at level i, page 13
$m_0 \uparrow\!\downarrow m_1$	m_0 and m_1 are connected via an uphill-downhill-path, page 46
$N^T(n)$	number of jumps of the decelerated chain up to time n, page 89
$N^{(i)}$	set of non-assigned states at level i, page 15
$p^*(r, s)$	limiting transition probabilities, page 11
$s \rightsquigarrow m$	s is attracted by m, page 14
s_m	set of minimal states on the outer boundary of the valley $V^{(i)}(m)$, page 28
t_{rel}	relaxation time, page 69
$V^{(i)}(m)$	valley of order i containing $m \in M^{(i)}$, page 15
$V^{(i)}(s)$	valley of order i containing $s \in \mathcal{S}^{(i)}$, page 36
$V^{(i)}_<(m)$	inner part of the valley of order i around m, page 14
$X_{\|V}$	to a valley V restricted chain, page 74
$Y^{(i)}$	accelerated aggregated chain at level i, page 36
$z^*(r, s)$	essential saddle between r and s, page 13
a.a.s.	asymptotically almost surely, page 37

Bibliography

[1] G. Alsmeyer. *Erneuerungstheorie*. Teubner Skripten zur Mathematischen Stochastik. B. G. Teubner, Stuttgart, 1991. Analyse stochastischer Regenerationsschemata.

[2] G. Alsmeyer. *Wahrscheinlichkeitstheorie*. Skripten zur Mathematischen Statistik Nr. 30, 2005. 4. Auflage.

[3] G. Alsmeyer and A. Winkler. Metabasins - a state space aggregation for highly disordered energy landscapes. *Theory Stoch. Process.*, 18(34):3–44, 2012 (no.1).

[4] P. Atkins and J. De Paula. *Physikalische Chemie*. Wiley-VCH, 2006. 4. vollständig überarbeitete Auflage.

[5] V. S. Barbu and N. Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications*, volume 191 of *Lecture Notes in Statistics*. Springer, New York, 2008.

[6] V. S. Barbu and N. Limnios. Some algebraic methods in semi-Markov chains. In *Algebraic methods in statistics and probability II*, volume 516 of *Contemp. Math.*, pages 19–35. Amer. Math. Soc., Providence, RI, 2010.

[7] J. Beltrán and C. Landim. Metastability of reversible finite state markov processes. *Stochastic Processes and their Applications*, 121(8):1633 – 1677, 2011.

[8] A. Bianchi and A. Gaudillière. Metastable states, quasi-stationary and soft measures, mixing time asymprtotics via variational principles. *ArXiv e-prints*, Mar. 2011.

[9] K. Binder and A. P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.*, 58:801–976, Oct 1986.

[10] J. P. Bouchaud. Weak ergodicity breaking and aging in disordered systems. *Journal de Physique I France*, 2:1705–1713, 1992.

[11] A. Bovier. Metastability: a potential theoretic approach. In *International Congress of Mathematicians. Vol. III*, pages 499–518. Eur. Math. Soc., Zürich, 2006.

[12] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theory Related Fields*, 119(1):99–161, 2001.

[13] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and low lying spectra in reversible Markov chains. *Comm. Math. Phys.*, 228(2):219–255, 2002.

[14] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing Markov chain on a graph. *SIAM Rev.*, 46(4):667–689, 2004.

[15] M. Cassandro, A. Galves, E. Olivieri, and M. E. Vares. Metastable behavior of stochastic dynamics: a pathwise approach. *J. Statist. Phys.*, 35(5-6):603–634, 1984.

[16] O. Catoni and R. Cerf. The exit path of a Markov chain with rare transitions. *ESAIM Probab. Statist.*, 1:95–144 (electronic), 1995/97.

[17] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probability*, 2:88–100, 1965.

[18] F. den Hollander. Metastability under stochastic dynamics. *Stochastic Process. Appl.*, 114(1):1–26, 2004.

[19] F. den Hollander. Three lectures on metastability under stochastic dynamics. In *Methods of contemporary mathematical statistical physics*, volume 1970 of *Lecture Notes in Math.*, pages 223–246. Springer, Berlin, 2009.

[20] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.

[21] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[22] A. Farago. Speeeding up markov chain monte carlo algorithms. In *Proceedings of the 2006 International Conference on Foundations of Computer Science, Las Vegas, Nevada, USA*, pages 102–108, 2006.

[23] J. A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, 1(1):pp. 62–87, 1991.

[24] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, New York, second edition, 1998.

[25] G. Frobenius. Über Matrizen aus nicht negativen Elementen. *Berl. Ber.*, pages 456–477, 1912.

[26] M. Goldstein. Viscous liquids and the glass transition: A potential energy barrier picture. *The Journal of Chemical Physics*, 51(9):3728–3739, 1969.

[27] S. Goldstein. Maximal coupling. *Z. Wahrsch. Verw. Gebiete*, 46(2):193–204, 1978/79.

[28] D. Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.

[29] A. Heuer. Exploring the potential energy landscape of glass-forming systems: from inherent structures via metabasins to macroscopic transport. *J. Phys.: Condens. Matter*, 20:56pp, 2008.

[30] R. Holley and D. Stroock. Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.*, 115(4):553–569, 1988.

[31] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[32] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.

[33] C. C. Huang, D. Isaacson, and B. Vinograde. The rate of convergence of certain nonhomogeneous Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(2):141–146, 1976.

[34] W. Huisinga, S. Meyn, and C. Schütte. Phase transitions and metastability in Markovian and molecular systems. *Ann. Appl. Probab.*, 14(1):419–458, 2004.

[35] T. Kaijser. A limit theorem for partially observed Markov chains. *Ann. Probability*, 3(4):677–696, 1975.

[36] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.

[37] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. The University Series in Undergraduate Mathematics. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960.

[38] J. Ledoux. A geometric invariant in weak lumpability of finite Markov chains. *J. Appl. Probab.*, 34(4):847–858, 1997.

[39] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

[40] T. M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, New York, 1985.

[41] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[42] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Ann. Appl. Probab.*, 12(2):581–606, 2002.

[43] R. A. Martin and D. Randall. Sampling adsorbing staircase walks using a new markov chain decomposition method. In *In Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 492–502. IEEE, 2000.

[44] P. Mathieu. Spectra, exit times and long time asymptotics in the zero-white-noise limit. *Stochastics Stochastics Rep.*, 55(1-2):1–20, 1995.

[45] E. Meerbach, C. Schütte, and A. Fischer. Eigenvalue bounds on restrictions of reversible nearly uncoupled Markov chains. *Linear Algebra Appl.*, 398:141–160, 2005.

[46] C. E. Mortimer. *Chemie*. Georg Thieme Verlag, Stuttgart, 1996. 6. Auflage.

[47] T. Okushima, T. Niiyama, K. S. Ikeda, and Y. Shimizu. Graph-based analysis of kinetics on multidimensional potential-energy surfaces. *Phys. Rev. E*, 80(3):036112, Sep 2009.

[48] E. Olivieri and E. Scoppola. Markov chains with exponentially small transition probabilities: first exit problem from a general domain. I. The reversible case. *J. Statist. Phys.*, 79(3-4):613–647, 1995.

[49] E. Olivieri and E. Scoppola. Markov chains with exponentially small transition probabilities: first exit problem from a general domain. II. The general case. *J. Statist. Phys.*, 84(5-6):987–1041, 1996.

[50] O. Penrose and J. Lebowitz. Rigorous treatment of metastable states in the van der waals-maxwell theory. *Journal of Statistical Physics*, 3:211–236, 1971. 10.1007/BF01019851.

[51] V. V. Petrov. On lower bounds for tail probabilities. *J. Statist. Plann. Inference*, 137(8):2703–2705, 2007.

[52] P. K. Pollett. Quasi-stationary distributions: A bibliography. *online*, Version 2012/06/11. http://www.maths.uq.edu.au/∼pkp/papers/qsds/qsds.html.

[53] O. Rubner and A. Heuer. From elementary steps to structural relaxation: A continuous-time random-walk analysis of a supercooled liquid. *Phys. Rev. E*, 78(1):011504, Jul 2008.

[54] M. Rudemo. State estimation for partially observed Markov chains. *J. Math. Anal. Appl.*, 44:581–611, 1973.

[55] E. Scoppola. Renormalization group for Markov chains and application to metastability. *J. Statist. Phys.*, 73(1-2):83–121, 1993.

[56] E. Scoppola. Metastability for Markov chains: a general procedure based on renormalization group ideas. In *Probability and phase transition (Cambridge, 1993)*, volume 420 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 303–322. Kluwer Acad. Publ., Dordrecht, 1994.

[57] E. Seneta. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York, 2006. Revised reprint of the second (1981) edition.

[58] Y. Serin and Z. M. Avsar. Markov decision processes with restricted observations: finite horizon case. *Naval Res. Logist.*, 44(5):439–456, 1997.

[59] W. Stadje. The evolution of aggregated Markov chains. *Statist. Probab. Lett.*, 74(4):303–311, 2005.

[60] F. H. Stillinger. A topographic view of supercooled liquids and glass formation. *Science*, 267, Issue 5206:1935–1939.

[61] N. G. van Kampen. *Stochastic processes in physics and chemistry*, volume 888 of *Lecture Notes in Mathematics*. North-Holland Publishing Co., Amsterdam, 1981.

[62] L. B. White, R. Mahony, and G. D. Brushe. Lumpable hidden Markov models—model reduction and reduced complexity filtering. *IEEE Trans. Automat. Control*, 45(12):2297–2306, 2000.

[63] A. Winkler, G. Alsmeyer, O. Rubner, and A. Heuer. Entropy of the hard-core model. *Phys. Rev. E*, 82:021502, Aug 2010.

Lebenslauf