J Stat Phys (2017) 167:1355–1375
DOI 10.1007/s10955-017-1777-z

Metastable Distributions of Markov Chains with Rare Transitions

M. Freidlin[1] · **L. Koralov**[1]

Received: 26 July 2016 / Accepted: 29 March 2017 / Published online: 9 April 2017

Abstract In this paper we consider Markov chains X^ε_t with transition rates that depend on a small parameter ε. We are interested in the long time behavior of X^ε_t at various ε-dependent time scales $t = t(\varepsilon)$. The asymptotic behavior depends on how the point $(1/\varepsilon, t(\varepsilon))$ approaches infinity. We introduce a general notion of complete asymptotic regularity (a certain asymptotic relation between the ratios of transition rates), which ensures the existence of the metastable distribution for each initial point and a given time scale $t(\varepsilon)$. The technique of i-graphs allows one to describe the metastable distribution explicitly. The result may be viewed as a generalization of the ergodic theorem to the case of parameter-dependent Markov chains.

Keywords Markov chains · Metastable distributions · Large deviations

Mathematics Subject Classification 37A25 · 37A30 · 60J27 · 60J28 · 60F10

1 Introduction

Consider a family X^ε_t of Markov chains on a state space $S = \{1, \ldots, N\}$, where ε is a small parameter. The time may be continuous or discrete – we start by considering the case when $t \in \mathbb{R}^+$. The case of discrete time is similar and is briefly discussed in Sect. 6. Let $q_{ij}(\varepsilon)$, $i, j \in S$, $i \neq j$, be the transition rates, i.e.,

$$\mathrm{P}(X^\varepsilon_{t+\Delta} = j | X^\varepsilon_t = i) = q_{ij}(\varepsilon)(\Delta + o(\Delta)) \quad \text{as } \Delta \downarrow 0, \quad i \neq j. \tag{1.1}$$

✉ L. Koralov
koralov@math.umd.edu

M. Freidlin
mif@math.umd.edu

1 Department of Mathematics, University of Maryland, College Park, MD 20742, USA

We will be interested in the behavior of X_t^ε as $\varepsilon \downarrow 0$ and, simultaneously, $t = t(\varepsilon) \to \infty$. The results on the asymptotic behavior of X_t^ε can be viewed as a refinement of the ergodic theorem for Markov chains (which concerns the asymptotics with respect to the time variable only) and are obviously closely related to the spectral properties of the transition matrix (some of the earliest work on the eigenvalues of matrices with exponentially small entries can be found in [36]). The double limit at hand depends on how the point $(1/\varepsilon, t(\varepsilon))$ approaches infinity. Roughly speaking, one can divide the neighborhood of infinity into a finite number of domains such that $X_{t(\varepsilon)}^\varepsilon$ has a limiting distribution (which depends on the initial point) when $(1/\varepsilon, t(\varepsilon))$ approaches infinity without leaving a given domain. For different domains, these limits are different. These will be referred to as metastable distributions.

The precise description of these domains, i.e., the assumptions on the admissible growth rates of $t(\varepsilon)$, will be provided below. Intuitively, $t(\varepsilon)$ should be such there is a subset $S' \subseteq S$ with the property that $t(\varepsilon)$ is sufficiently large for averaging on S' to take place, yet not sufficiently large for the process to leave S'. The existence of such subsets, i.e., the ability to separate different time scales, is a consequence of the property of complete asymptotic regularity, formulated below.

Families of parameter-dependent Markov chains arise in many applications, such as various models of non-equilibrium statistical mechanics at low temperatures (see, e.g., [12,15,16,29] and references therein). Another important example leading to parameter-dependent Markov chains comes from the study of randomly perturbed dynamical systems [24]. If a dynamical system has N asymptotically stable attractors, each attractor can be associated with a state of a Markov chain, while the transition times between different states are due to large deviations and are determined by the action functional of the perturbed system. In this example, the transition rates are exponentially small (up to a pre-exponential factor) with respect to the size of the perturbation ε, i.e., the transition times are exponentially large. Some of the finer results on the transition times (including the pre-exponential factor) were obtained in [6,10].

As shown in [22,24], the long-time behavior of the perturbed process can typically be understood using the notion of the hierarchy of cycles. The hierarchy of cycles means, roughly speaking, that for each $0 \leq r \leq \rho$ the set of attractors is decomposed into disjoint subsets $C_1^r, \ldots, C_{n_r}^r$ (cycles of rank r), up to the maximal rank $\rho < N$. The individual attractors are the cycles of rank zero, they are combined in disjoint sets—cycles of rank one, those are combined in cycles of rank two, etc., until the cycle of maximal rank $\rho < N$ containing all the attractors. With probability close to one as $\varepsilon \downarrow 0$, the process goes from a neighborhood of one of the attractors in C_i^r to a small neighborhood of one the attractors of the next cycle C_j^r, thus remaining within a cycle of rank $r + 1$. The transition time between C_i^r and C_j^r is determined by the asymptotics, as $\varepsilon \downarrow 0$, of the transition rates between individual attractors that belong to the union of these cycles. The process leaves the cycle of rank $r + 1$ only after an exponentially large number of transitions between cycles of rank r.

In the generic case, for each λ (except a finite number of values) and the time scale $t(\varepsilon) \sim \exp(\lambda/\varepsilon)$, with probability close to one, the process can be found in a neighborhood of a particular attractor, which depends on the initial point. If only one invariant probability measure of the unperturbed system is concentrated on this attractor, then the distribution of $X_{t(\varepsilon)}^\varepsilon$ converges, as $\varepsilon \downarrow 0$, to this measure, which is called the metastable distribution for the given initial point and given λ.

The above description with the hierarchy of cycles and the meta-stable states is valid, however, only if the notion of the unique "next" cycle (and, consequently, unique meta-stable state) can be correctly defined. For example, when the unperturbed dynamics has certain symmetries (or "rough symmetries" as in [23]), the notion of "next" is not defined uniquely even for individual attractors (cycles of rank zero). A similar phenomenon was observed in [25], due not to symmetries but to degeneration of the unperturbed dynamics in a part of the phase space.

Since the hierarchies of cycles were introduced in [22], many different techniques have been developed for the study of metastability, some of which do not require the uniqueness of the 'next' state or cycle. Those include renormalization group ideas (as in [30–33]), the potential-theoretic approach (see [9–12,19] and references therein), and the martingale approach (see [4,5] and references therein). See also [13,28,34,35] for problems concerning simulated annealing, where a hierarchy construction also appears. Distribution of a process conditioned to stay within a trapping collection of states and the asymptotic distribution of the exit time were studied in [8,21]. (See also [1,2,17,20] for problems related to the distribution of a hitting time of a set in a rather general setting.) The distribution of the exit path from a set after the last visit to a given subset was studied in [14].

Yet another example leading to parameter-dependent Markov chains is provided by dynamical systems with heteroclinic networks. Namely, assume that there are finitely many stationary points with heteroclinic connections that together form a connected set. Assume that the entire network serves as an attractor for the dynamical system. The flow lines connecting pairs of stationary points can be associated with the states of a Markov chain. After a random perturbation of size ε, the transition times between the neighborhoods of such flow lines scale as powers of ε (up to a logarithmic factor). So the notions of the hierarchy of cycles and the meta-stable states could apply (at times that scale as powers of ε, rather than exponentially). However, the notion of "next" state may again be not defined uniquely, since the exit from a neighborhood of a heteroclinic connection (say, between stationary points A and B) can happen along either of the heteroclinic connections leading out of B. A detailed study of motion along heteroclinic networks is the subject of [3]. It should be stressed that the dynamics in this example, as well as in the case of asymptotically stable attractors discussed above, is only approximately described by a Markov chain of the type considered in the current paper. A reduction of the true dynamics to a finite-state Markov chain requires non-trivial analysis.

In the current paper, we introduce the notion of the hierarchy of Markov chains in a general setting. We show that it should replace the notion of the hierarchy of cycles. We do not require the transition rates between different states to scale exponentially (as in most models arising from statistical physics and in the study of randomly perturbed dynamical systems), but only assume that there is a simple asymptotic relation between the ratios of transition rates.

More precisely, we introduce the following definition.

Definition 1.1 The family X_t^ε is said to be *asymptotically regular* as $\varepsilon \downarrow 0$ if the following two conditions hold.

(a) The transition rates $q_{ij}(\varepsilon)$ are positive[1] for $\varepsilon > 0$ and all $i \neq j$.
(b) For each $i, j, k, l \in S$ satisfying $i \neq j, k \neq l$, the following finite or infinite limit exists

$$\lim_{\varepsilon \downarrow 0} (q_{ij}(\varepsilon)/q_{kl}(\varepsilon)) \in [0, \infty].$$

Loosely speaking, the property of asymptotic regularity implies that X_t^ε has a certain limiting structure when $\varepsilon \downarrow 0$. Namely, while asymptotic regularity does not allow one to control the transition times between different states, it allows one to analyze the order in which the states are visited, which is crucial for describing the limiting behavior of invariant measures of X_t^ε.

[1] The main result can be obtained even if the positivity assumption is relaxed, as mentioned in Sect. 6.

In Sect. 2, we will introduce the hierarchy of Markov chains. The hierarchy will be defined inductively by successively reducing the state space, i.e., combining the elements of the state space into subsets that will serve as states for the chain of higher rank. This construction generalizes the hierarchy of cycles notion of [22]. In order to perform an inductive step, we will require each chain appearing in the construction to be asymptotically regular. While this condition may seem not explicit, we will show that it holds if the transition rates of the original family satisfy a relatively simple condition, which we call *complete asymptotic regularity*. A similar condition was introduced in a recent paper [7], where a different technique for the hierarchy construction (based on the study of hitting times of states and appropriate clusters of states) was employed.

After constructing the hierarchy, we can associate an increasing sequence of time scales $f_n(i, \varepsilon)$, $1 \leq n \leq n^*$, with each initial state $i \in S$ of the original chain. These time scales satisfy $f_n(i, \varepsilon) \ll f_{n+1}(i, \varepsilon)$, $1 \leq n \leq n^* - 1$, and $f_{n^*}(i, \varepsilon) \equiv \infty$. Assuming that the process starts at i and $f_n(i, \varepsilon) \ll t(\varepsilon) \ll f_{n+1}(i, \varepsilon)$, we will show that $X^\varepsilon_{t(\varepsilon)}$ converges to a distribution $\nu(i, \cdot)$, called the metastable distribution for the initial state i at the time scale $t(\varepsilon)$. This is the main result of our paper.

In Sect. 3, we discuss some simple properties of asymptotically regular Markov chains. In Sect. 4, we formulate and prove the main theorem on the meta-stable behavior of the original process. In Sect. 5, we prove that a condition on the transition rates of the family of Markov chains guarantees that all the chains in the hierarchy are asymptotically regular. We briefly discuss a couple of generalizations in Sect. 6.

2 Hierarchy of Markov Chains

2.1 Reduced Markov Chain

Given an asymptotically regular family of Markov chains X^ε_t with $N > 1$, we will construct a reduced Markov chain (later also referred to as the reduced Markov chain of rank one). First, we define a discrete-time Markov chain on S, which will be referred to as the *skeleton Markov chain* and denoted by Z_n. Its transition probabilities are defined by

$$P_{ij} = \lim_{\varepsilon \downarrow 0} (q_{ij}(\varepsilon) \Big/ \sum_{j' \neq i} q_{ij'}(\varepsilon)), \quad j \neq i; \quad P_{ii} = 0.$$

Observe that the above limit exists since the family of chains X^ε_t is asymptotically regular. Recall (see [18]) that the (finite) state space of a Markov chain can be uniquely decomposed into a disjoint union of ergodic classes and the sets consisting of individual transient states,

$$S = S_1 \bigcup \ldots \bigcup S_n. \tag{2.1}$$

Note that $n < N$ since each ergodic class of Z_n contains at least two states, which follows from the fact that for each i there is $j \neq i$ such that $P_{ij} > 0$.

Next, for each $1 \leq k \leq n$, we define the Markov chains $Y^{k,\varepsilon}_t$ by narrowing the state space S to S_k and retaining the same transition rates for $i, j \in S_k$ as in the original Markov chain X^ε_t. Let $\mu^k(i, \varepsilon)$ be the invariant measure of the state $i \in S_k$ for the chain $Y^{k,\varepsilon}_t$.

Finally, we define the *reduced Markov chain*. Its state space is the set $\{S_1, \ldots, S_n\}$. The transition rate between S_k and S_l, $k \neq l$, denoted by $Q_{kl}(\varepsilon)$, is defined by

$$Q_{kl}(\varepsilon) = \sum_{i \in S_k} \sum_{j \in S_l} \mu^k(i, \varepsilon) q_{ij}(\varepsilon). \tag{2.2}$$

2.2 Definition of the Hierarchy

We will use induction to define the reduced Markov chains $X_t^{r,\varepsilon}$, $0 \le r \le \rho$, with some $0 \le \rho < N$. The reduced Markov chain of rank zero (i.e., corresponding to $r = 0$) will coincide with X_t^ε, while the reduced Markov chain of rank ρ will be trivial (i.e., its state space will contain one element). The entire collection of reduced Markov chains will be referred to as the hierarchy. An example of a family of Markov chains with the hierarchy construction will be provided at the end of this section.

For each $0 \le r \le \rho - 1$, by partitioning the state space of $X_t^{r,\varepsilon}$ into n_{r+1} subsets (referred to as clusters), we will define Markov chains $Y_t^{r,k,\varepsilon}$, $1 \le k \le n_{r+1}$. The chain $Y_t^{r,k,\varepsilon}$ will be referred to as the k-th chain of rank r.

We set $n_0 = N$ and $S_1^0 = 1,\ldots, S_{n_0}^0 = n_0$. These are clusters of rank zero. The reduced Markov chain of rank zero, denoted by $X_t^{0,\varepsilon}$, is defined on the state space $S^0 = \{S_1^0, \ldots, S_{n_0}^0\}$ and has transition rates $Q_{ij}^0(\varepsilon) = q_{ij}(\varepsilon)$. Thus it coincides with the original Markov chain.

If $N = 1$, this results in the trivial hierarchy. If $N > 1$, we can apply the above construction of the reduced Markov chain. We set $n_1 = n$ and use the following notation

$$S = S_1^1 \bigcup \ldots \bigcup S_{n_1}^1.$$

for the decomposition of the original state space into ergodic classes and sets containing individual transient states for the skeleton chain. The sets $S_1^1, \ldots, S_{n_1}^1$ will be referred to as clusters of rank one. For each $1 \le k \le n_1$, we set $Y_t^{0,k,\varepsilon} = Y_t^{k,\varepsilon}$, which will be be referred to as the k-th Markov chain of rank zero. Let $\mu^{0,k}(i, \varepsilon)$ be the invariant measure of the state $S_i^0 \in S_k^1$ for the chain $Y_t^{0,k,\varepsilon}$ defined above.

The reduced Markov chain introduced above will also be referred to as the reduced Markov chain or rank one and denoted by $X_t^{1,\varepsilon}$. Its state space is the set $\{S_1^1, \ldots, S_{n_1}^1\}$. The transition rate between S_k^1 and S_l^1, $1 \le k, l \le n_1, k \ne l$, denoted by $Q_{kl}^1(\varepsilon)$, is defined, in conformance with (2.2), by

$$Q_{kl}^1(\varepsilon) = \sum_{i \in S_k^1} \sum_{j \in S_l^1} \mu^{0,k}(i, \varepsilon) Q_{ij}^0(\varepsilon). \tag{2.3}$$

If $X_t^{1,\varepsilon}$ is asymptotically regular, we can replicate the above construction, i.e., define the skeleton chain corresponding to $X_t^{1,\varepsilon}$ and partition $\{S_1^1, \ldots, S_{n_1}^1\}$ into clusters $S_1^2, \ldots, S_{n_2}^2$. For each $1 \le k \le n_2$, we can define the k-th Markov chain of rank one, denoted by $Y_t^{1,k,\varepsilon}$. The reduced Markov chain $X_t^{2,\varepsilon}$ of rank two is defined on the state space $\{S_1^2, \ldots, S_{n_2}^2\}$. The construction then continues inductively, assuming that all the reduced chains are asymptotically regular.

Let ρ be the first index such that $n_\rho = 1$. Observe that $\rho < N$ since $n_0 = N$ and $n_{r+1} < n_r$ for each r. Since $n_\rho = 1$, there is only one Markov chain of rank $\rho - 1$, and the reduced Markov chain of rank ρ is trivial – its state space consists of one element S_1^ρ. This completes the construction of the hierarchy.

The reduced Markov chain of rank r, $0 \le r \le \rho$, will be denoted by $X_t^{r,\varepsilon}$, its transition rates will be denoted by $Q_{kl}^r(\varepsilon)$, the k-th Markov chain of rank r, $0 \le r < \rho$, $1 \le k \le n_{r+1}$, will be denoted by $Y_t^{r,k,\varepsilon}$, and the invariant measure of a state $S_i^r \in S_k^{r+1}$ for this Markov chain will be denoted by $\mu^{r,k}(i, \varepsilon)$.

For a state $j \in S$, we'll write that $j \prec S_k^r$ if there is a sequence $j = j_0, j_1, \ldots, j_r = k$ such that

$$j = S_{j_0}^0 \in S_{j_1}^1 \ldots \in S_{j_r}^r. \tag{2.4}$$

For $1 \leq k \leq n_r$ and j such that $j \prec S_k^r$ does not hold, we will need the functions $\widetilde{Q}_{kj}^r(\varepsilon)$. These are defined inductively, namely, $\widetilde{Q}_{kj}^0(\varepsilon) = Q_{ij}^0(\varepsilon)$ if $S_k^0 = i$ and

$$\widetilde{Q}_{kj}^r(\varepsilon) = \sum_{i: S_i^{r-1} \in S_k^r} \mu^{r-1,k}(i, \varepsilon) \widetilde{Q}_{ij}^{r-1}(\varepsilon), \quad 1 \leq r \leq \rho - 1.$$

Intuitively, these serve as transition rates from a cluster of rank r to an individual state, although, unlike $Q_{kl}^r(\varepsilon)$, they don't correspond to transition rates of any of the Markov chains introduced above. Let us stress that the functions $\widetilde{Q}_{kj}^r(\varepsilon)$ were not used in the construction of the hierarchy. They will be needed, however, in order to describe the asymptotic behavior of X_t^ε (for example, in the case when the time needed for X_t^ε to exit the state j is longer than the time needed for $X_t^{r,\varepsilon}$ to exit the state S_k^r).

Let us stress that the inductive construction of the hierarchy is possible under the condition that all the reduced chains that appear at each step are asymptotically regular.

Definition 2.1 We will say that X_t^ε is *completely asymptotically regular* if for each $a > 0$ and each $(i_1, \dots, i_a)$, $(j_1, \dots, j_a)$, $(k_1, \dots, k_a)$, and $(l_1, \dots, l_a)$ the following finite or infinite limit exists

$$\lim_{\varepsilon \downarrow 0} \left(\frac{q_{i_1 j_1}(\varepsilon)}{q_{k_1 l_1}(\varepsilon)} \times \cdots \times \frac{q_{i_a j_a}(\varepsilon)}{q_{k_a l_a}(\varepsilon)} \right) \in [0, \infty], \tag{2.5}$$

provided that $i_1 \neq j_1, \dots, i_a \neq j_a, k_1 \neq l_1, \dots, k_a \neq l_a$.

Lemma 2.2 *Suppose that X_t^ε is completely asymptotically regular. Then the reduced chain is also completely asymptotically regular.*

This lemma will be proved in Sect. 5. For now, we observe that if X_t^ε is completely asymptotically regular, then, by Lemma 2.2, so are the reduced Markov chains that appear at each step of the inductive construction of the hierarchy, which implies that all of them are asymptotically regular.

In conclusion of this section, let us consider a simple example of a family of Markov chains and of the corresponding hierarchy. In this example, $n_0 = N = 7$, and the transition rates, arranged in a matrix without diagonal elements, are given by

$$q(\varepsilon) = Q^0(\varepsilon) = \begin{pmatrix} * & \varepsilon & \varepsilon^4 & e^{-1/\varepsilon} & e^{-1/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} \\ \varepsilon^4 & * & \varepsilon^2 & e^{-1/\varepsilon} & e^{-1/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} \\ \varepsilon^3 & \varepsilon^3 & * & e^{-1/\varepsilon} & e^{-1/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} \\ e^{-1/\varepsilon} & e^{-1/\varepsilon} & e^{-1/\varepsilon} & * & 1/|\ln(\varepsilon)| & e^{-2/\varepsilon} & e^{-2/\varepsilon} \\ e^{-1/\varepsilon} & e^{-1/\varepsilon} & e^{-1/\varepsilon} & 1/|\ln(\varepsilon)| & * & e^{-2/\varepsilon} & e^{-2/\varepsilon} \\ e^{-2/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} & * & \varepsilon \\ e^{-2/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} & e^{-2/\varepsilon} & \varepsilon & * \end{pmatrix}.$$

In Fig. 1, the solid arrows denote non-zero transitions for the skeleton Markov chain corresponding to $X_t^{0,\varepsilon}$. The skeleton Markov chain has three ergodic classes (clusters of rank one): $S_1^1 = \{1, 2, 3\}$, $S_2^1 = \{4, 5\}$, and $S_3^1 = \{6, 7\}$. Thus $n_1 = 3$. The transition rates for Markov chains of rank zero ($Y_t^{0,1,\varepsilon}$, $Y_t^{0,2,\varepsilon}$, and $Y_t^{0,3,\varepsilon}$) can be arranged into matrices

$$\begin{pmatrix} * & \varepsilon & \varepsilon^4 \\ \varepsilon^4 & * & \varepsilon^2 \\ \varepsilon^3 & \varepsilon^3 & * \end{pmatrix}, \quad \begin{pmatrix} * & 1/|\ln(\varepsilon)| \\ 1/|\ln(\varepsilon)| & * \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} * & \varepsilon \\ \varepsilon & * \end{pmatrix}.$$

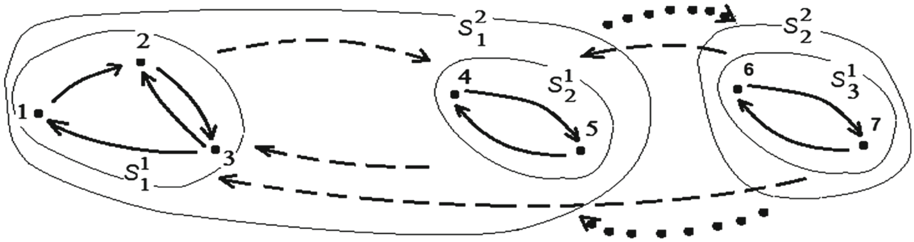

Fig. 1 In this example, $n_0 = N = 7$, $n_1 = 3$, $n_2 = 2$, $n_3 = 1$, and $\rho = 3$. The *solid arrows* denote non-zero transitions for the skeleton Markov chain corresponding to $X_t^{0,\varepsilon}$. The *dashed arrows* denote non-zero transitions for the skeleton Markov chain corresponding to $X_t^{1,\varepsilon}$. The *dotted arrows* denote non-zero transitions for the skeleton Markov chain corresponding to $X_t^{2,\varepsilon}$

Their invariant measures are easily seen to be given by

$$\mu^{0,1}(\cdot,\varepsilon) = (\varepsilon^2(1+o(1)), 2\varepsilon(1+o(1)), 1+o(1)), \quad \mu^{0,2}(\cdot,\varepsilon) = \left(\frac{1}{2},\frac{1}{2}\right),$$
$$\mu^{0,3}(\cdot,\varepsilon) = \left(\frac{1}{2},\frac{1}{2}\right).$$

(Here and below, each instance of $o(1)$ may represent a different function that tends to zero as $\varepsilon \downarrow 0$.) Thus, by (2.3), the transition rates for the reduced Markov chain of rank one are given by

$$Q^1(\varepsilon) = \begin{pmatrix} * & (2+o(1))e^{-1/\varepsilon} & (2+o(1))e^{-2/\varepsilon} \\ 3e^{-1/\varepsilon} & * & 2e^{-2/\varepsilon} \\ 3e^{-2/\varepsilon} & 2e^{-2/\varepsilon} & * \end{pmatrix}.$$

In Fig. 1, the dashed arrows denote non-zero transitions for the skeleton Markov chain corresponding to $X_t^{1,\varepsilon}$. This skeleton Markov chain has one ergodic class $S_1^2 = \{S_1^1, S_2^1\} = \{\{1,2,3\},\{4,5\}\}$ and one transient set $S_2^2 = \{S_3^1\} = \{\{6,7\}\}$, and thus $n_2 = 2$. The transition rates for the Markov chain $Y_t^{1,1,\varepsilon}$ of rank one can be arranged into the matrix

$$\begin{pmatrix} * & 2(1+o(1))e^{-1/\varepsilon} \\ 3e^{-1/\varepsilon} & * \end{pmatrix}.$$

Its invariant measure is given by

$$\mu^{1,1}(\cdot,\varepsilon) = ((3+o(1))/5, (2+o(1))/5).$$

The Markov chain $Y_t^{1,2,\varepsilon}$ is trivial since its state space contains only one element, and $\mu^{1,2}(1,\varepsilon) = 1$. The state space for the reduced Markov chain of rank two is the set $\{S_1^2, S_2^2\}$. Applying (2.3) to $X_t^{1,\varepsilon}$ instead of $X_t^{0,\varepsilon}$, we obtain that the transition rates for the reduced Markov chain of rank two are given by

$$Q_{12}^2 = \mu^{1,1}(1,\varepsilon)Q_{13}^1(\varepsilon) + \mu^{1,1}(2,\varepsilon)Q_{23}^1(\varepsilon) = (2+o(1))e^{-2/\varepsilon},$$
$$Q_{21}^2 = \mu^{1,2}(1,\varepsilon)Q_{31}^1(\varepsilon) + \mu^{1,2}(1,\varepsilon)Q_{32}^1(\varepsilon) = (5+o(1))e^{-2/\varepsilon}.$$

In Fig. 1, the dotted arrows denote non-zero transitions for the skeleton Markov chain corresponding to $X_t^{2,\varepsilon}$. This skeleton chain has one ergodic class, and thus $n_3 = 1$ and $\rho = 3$. There is only one Markov chain of rank two, $Y_t^{2,1,\varepsilon}$, which coincides with $X_t^{2,\varepsilon}$. The reduced

Markov chain of rank three is trivial—its state space consists of one element $S_1^3 = \{\{S_1^2, S_2^2\}\}$. This completes the construction of the hierarchy.

3 Asymptotically Regular Families of Markov Chains

In order to prepare for the discussion of metastability, we need several lemmas on asymptotically regular families of Markov chains. Let $\mu(i, \varepsilon)$ denote the invariant measure of the state i in an asymptotically regular Markov chain X_t^ε with $N > 1$. We'll be interested in the asymptotics of $\mu(i, \varepsilon)$ in the case when the skeleton chain has one ergodic class and no transient states (in which case $\rho = 1$ and no additional assumptions are required in order to define the hierarchy). Note that this condition is satisfied for each of the Markov chains $Y_t^{r,k,\varepsilon}$, $0 \le r \le \rho - 1$, $1 \le k \le n_{r+1}$, defined above, with the exception that the number of states for such a chain may be equal to one.

Let λ be the invariant measure of the skeleton chain Z_n. Define

$$T(i, \varepsilon) = \left(\sum_{j \in S, j \neq i} q_{ij}(\varepsilon) \right)^{-1}, \quad \bar{T}(\varepsilon) = \sum_{i' \in S} (\lambda(i') T(i', \varepsilon)).$$

Thus $T(i, \varepsilon)$ is the average of the exponentially distributed exit time of X_t^ε from the state i. The function $\bar{T}(\varepsilon)$ is the average time it takes X_t^ε to make one step, where the average is calculated with respect to the invariant measure for the skeleton chain.

The symbol $\sim$ is used below to denote asymptotic equivalence between positive functions, i.e., $f(\varepsilon) \sim g(\varepsilon)$ as $\varepsilon \downarrow 0$ if $\lim_{\varepsilon \downarrow 0}(f(\varepsilon)/g(\varepsilon)) = 1$.

Lemma 3.1 *Let X_t^ε be asymptotically regular and $N > 1$. Suppose that the skeleton chain has one ergodic class and no transient states. Then*

$$\mu(i, \varepsilon) \sim \lambda(i) T(i, \varepsilon)/\bar{T}(\varepsilon) \quad \text{as } \varepsilon \downarrow 0.$$

Proof Let $t_1^\varepsilon, t_2^\varepsilon, \ldots$ be the times of the jumps of the process X_t^ε. Let Z_n^ε be the discrete-time Markov chain obtained from X_t^ε by discretizing time, i.e., $Z_0^\varepsilon = X_0^\varepsilon$, $Z_n^\varepsilon = X_{t_n}^\varepsilon$, $n \ge 1$. Let P_{ij}^ε be the transition probabilities for Z_n^ε and λ^ε be its invariant measure. Then, from the definition of the skeleton chain, it follows that

$$\lim_{\varepsilon \downarrow 0} P_{ij}^\varepsilon = P_{ij}, \quad i \neq j, \tag{3.1}$$

where P_{ij} are the transition probabilities for Z_n. Since Z_n has one ergodic class and no transient states, this implies that

$$\lim_{\varepsilon \downarrow 0} \lambda^\varepsilon(i) = \lambda(i) > 0, \quad i \in S.$$

By the Law of Large Numbers for Markov chains, $\mu(i, \varepsilon)$ is equal to the asymptotic (as $t \to \infty$) proportion of time that the process X_t^ε spends in the state i. Therefore,

$$\mu(i, \varepsilon) = \lambda^\varepsilon(i) T(i, \varepsilon) \Big/ \sum_{i' \in S} (\lambda^\varepsilon(i') T(i', \varepsilon)).$$

Combining the latter two equalities, we obtain the result claimed in the lemma. □

From Lemma 3.1 and the asymptotic regularity of X^ε_t it follows that there is a limiting probability measure

$$\mu(i) = \lim_{\varepsilon \downarrow 0} \mu(i, \varepsilon) \in [0, 1], \quad i \in S. \tag{3.2}$$

Given two functions $t(\varepsilon), s(\varepsilon) : (0, \infty) \to (0, \infty)$, we'll write $s(\varepsilon) \ll t(\varepsilon)$ if $s(\varepsilon) = o(t(\varepsilon))$ as $\varepsilon \downarrow 0$. Let $\alpha(j, t)$ be the proportion of time, prior to t, that the process X^ε_t spends in j.

Lemma 3.2 *Let X^ε_t be asymptotically regular and $N > 1$. Suppose that the skeleton chain has one ergodic class and no transient states. Suppose that $t(\varepsilon)$ is such that $\bar{T}(\varepsilon) \ll t(\varepsilon)$. Then for each $i, j \in S$,*

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = \mu(j). \tag{3.3}$$

$$\mathrm{E}_i \alpha(j, t(\varepsilon)) \sim \mu(j, \varepsilon), \quad as\ \varepsilon \downarrow 0. \tag{3.4}$$

For each $c > 0$, there are $\delta(c) > 0$ and $\varepsilon_0 > 0$ such that

$$\mathrm{P}_i(\alpha(j, t(\varepsilon)) < (1 - c)\mu(j, \varepsilon)) \le e^{-\delta(c)t(\varepsilon)/\bar{T}(\varepsilon)}, \quad \varepsilon \le \varepsilon_0. \tag{3.5}$$

$$\mathrm{P}_i(\alpha(j, t(\varepsilon)) > (1 + c)\mu(j, \varepsilon)) \le e^{-\delta(c)t(\varepsilon)/\bar{T}(\varepsilon)}, \quad \varepsilon \le \varepsilon_0. \tag{3.6}$$

Here the subscript i stands for the initial location of the process.[2]

Proof Let $i^* \in S$ be such that $\mu(i^*) > 0$. Given $i \in S$, find a sequence $i_0, i_1, \ldots, i_k$ such that $i_0 = i, i_k = i^*, 0 \le k < N$, and $P_{i_0 i_1}, \ldots, P_{i_{k-1} i_k} > 0$, where the latter are the transition probabilities for the skeleton Markov chain. Then, examining the transition rates of X^ε_t, it is easy to see that

$$\mathrm{P}_i(Z^\varepsilon_0 = i_0, \ldots, Z^\varepsilon_k = i_k, \ \ t^\varepsilon_k < \bar{T}(\varepsilon) < t^\varepsilon_{k+1}) \ge a$$

for some positive constant a and all sufficiently small ε. In particular, $\mathrm{P}_i(X^\varepsilon_{\bar{T}(\varepsilon)} = i^*) \ge a$. Since i was arbitrary, this provides an upper bound on the speed of convergence of X^ε_t to the invariant distribution, i.e., $\mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) - \mu(j, \varepsilon) \to 0$ as $\varepsilon \downarrow 0$ if $\bar{T}(\varepsilon) \ll t(\varepsilon)$. Combined with (3.2), this implies (3.3).

Let $n(\varepsilon) = t(\varepsilon)/\bar{T}(\varepsilon)$. Let $\beta(j, t)$ be the amount of time, prior to t, that the process spends in j. From the large deviation estimates for the Markov chains (see Chap. 8 of [26]) it easily follows that for $c' \in (0, 1)$ there are $\delta(c') > 0$ and $\varepsilon_0 > 0$ such that

$$\mathrm{P}_i\left(t^\varepsilon_{[(1-c')n(\varepsilon)]} > t(\varepsilon)\right) < e^{-\delta(c')n(\varepsilon)}, \quad \varepsilon \le \varepsilon_0. \tag{3.7}$$

$$\mathrm{P}_i\left(t^\varepsilon_{[(1+c')n(\varepsilon)]} < t(\varepsilon)\right) < e^{-\delta(c')n(\varepsilon)}, \quad \varepsilon \le \varepsilon_0, \tag{3.8}$$

Moreover, for $c' \in (0, c \wedge 1)$, there are $\delta(c, c') > 0$ and $\varepsilon_0 > 0$ such that

$$\mathrm{P}_i\left(\beta(j, t^\varepsilon_{[(1-c')n(\varepsilon)]}\right) < (1 - c)\mu(j, \varepsilon)t(\varepsilon)) < e^{-\delta(c,c')n(\varepsilon)}, \quad \varepsilon \le \varepsilon_0, \tag{3.9}$$

while for $c' > 0$ satisfying $(1 + c)(1 - c')/(1 + c') > 1$, there are $\delta(c, c') > 0$ and $\varepsilon_0 > 0$ such that

$$\mathrm{P}_i\left(\beta(j, t^\varepsilon_{[(1+c')n(\varepsilon)]}) > (1 + c)\mu(j, \varepsilon)t(\varepsilon)\right) < e^{-\delta(c,c')n(\varepsilon)}, \quad \varepsilon \le \varepsilon_0, \tag{3.10}$$

[2] Formula (3.3) can be improved to $\mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) \sim \mu(j, \varepsilon)$ as $\varepsilon \downarrow 0$, but we don't need it here.

We obtain (3.5) by combining (3.7) and (3.9). We obtain (3.6) by combining (3.8) and (3.10).

From (3.6) and the strong Markov property of the process, it follows that

$$\mathrm{P}_i(\alpha(j, t(\varepsilon)) > k(1+c)\mu(j, \varepsilon)) \leq e^{-k\delta(c)t(\varepsilon)/\bar{T}(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0,$$

for each $k \in \mathbb{N}$. Combined with (3.5), this immediately implies (3.4). □

Next, let us consider the behavior of an asymptotically regular chain that is stopped when it enters a non-empty set $E \subseteq S$. Let $\sigma = \inf\{t : X_t^\varepsilon \in E\}$, $\tau = \min\{n : Z_n \in E\}$, $\tau' = \min\{n : Z_n^\varepsilon \in E\}$.

Lemma 3.3 *Let X_t^ε be asymptotically regular and $N > 1$. Suppose that the skeleton chain has one ergodic class and no transient states. Let E be a non-empty subset of S. Then for each $i \in S$ and $j \in E$,*

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X_\sigma^\varepsilon = j) = \mathrm{P}_i(Z_\tau = j). \tag{3.11}$$

If $t(\varepsilon)$ is such that $T(i, \varepsilon) \ll t(\varepsilon)$ for $i \notin E$, then for each $i \in S$ and $j \in E$,

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i\left(\sigma > \frac{t(\varepsilon)}{2}\right) = 0. \tag{3.12}$$

Proof Recall that Z_n^ε is the discrete-time Markov chain obtained from X_t^ε by discretizing time. Then

$$\mathrm{P}_i(X_\sigma^\varepsilon = j) = \mathrm{P}_i(Z_{\tau'}^\varepsilon = j) \to \mathrm{P}_i(Z_\tau = j) \ \text{ as } \varepsilon \downarrow 0,$$

where the convergence follows from (3.1). Now let us prove (3.12). Given $\delta > 0$, find k such that $\mathrm{P}_i(\tau > k) < \delta$. From the convergence of Z_n^ε to Z_n, it then follows that

$$\mathrm{P}_i(\tau' > k) < \delta$$

for all sufficiently small ε. From the condition $T(i, \varepsilon) \ll t(\varepsilon)$ for $i \notin E$, it now follows that

$$\mathrm{P}_i\left(\sigma > \frac{t(\varepsilon)}{2}\right) < 2\delta$$

for all sufficiently small ε, which implies (3.12). □

Remark The quantity $\mathrm{P}_i(X_\sigma^\varepsilon = j)$ can be represented in terms of i-graphs (see Chap. 6 of [24]). Such a representation could be used as an alternative way to prove Lemma 3.3.

4 Metastable Distributions for Completely Asymptotically Regular Families

4.1 Formulation of the Result

Suppose that X_t^ε is completely asymptotically regular. In this section, we'll study the distribution of $X_{t(\varepsilon)}^\varepsilon$ at time scales $t(\varepsilon)$ that vary with ε as $\varepsilon \downarrow 0$. The initial state X_0^ε is assumed to be fixed.

To give a clearer picture, we first formulate the result in a particular case, with the general case to follow.

Theorem 4.1 *Let X_t^ε be asymptotically regular and $N > 1$. Suppose that the skeleton chain has one ergodic class and no transient states. Suppose that for each $1 \le i \le N$ either $t(\varepsilon) \ll T(i,\varepsilon)$ or $t(\varepsilon) \gg T(i,\varepsilon)$. Then for each i there is a probability measure $\nu(i,\cdot)$ on S such that the following limit exists*

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = \nu(i,j). \tag{4.1}$$

The measure $\nu(i,\cdot)$ will be referred to as the metastable distribution for the initial state i at the time scale $t(\varepsilon)$.

Proof Let $E = \{i \in S : t(\varepsilon) \ll T(i,\varepsilon)\}$. If $E = \varnothing$, then the result follows from (3.3), i.e., the limiting measure is the invariant measure, for each initial state. If $E \neq \varnothing$, then by Lemma 3.3 there is a measure $\nu(i,\cdot)$ concentrated on E such that

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{\sigma \wedge t(\varepsilon)} = j) = \nu(i,j).$$

Notice also that for $i \in E$ and $\delta > 0$,

$$\mathrm{P}_i(t_1^\varepsilon \le t(\varepsilon)) < \delta$$

for all sufficiently small ε, as follows from the condition that $t(\varepsilon) \ll T(i,\varepsilon)$ for $i \in E$. Here t_1^ε is the time of the first jump. Since δ was arbitrary, using the last two relations and the strong Markov property of the process, we obtain (4.1).

Now let us formulate the result in the general case, i.e., without assuming that the skeleton chain has one ergodic class. For $0 \le r \le \rho$, the *inverse transition rate* of S_i^r is defined as

$$T^r(i,\varepsilon) = \left(\sum_{j \neq i} Q^r_{ij}(\varepsilon)\right)^{-1}, \tag{4.2}$$

where the sum is over $1 \le j \le n_r$, $j \neq i$. (If $S_i^r \in S_k^{r+1}$ and $S_k^{r+1} \neq \{S_i^r\}$, then the sum in the right hand side of (4.2) is asymptotically equivalent to the sum that extends only over such $j \neq i$ for which $S_j^r \in S_k^{r+1}$.) Note that $T^r(i,\varepsilon) \equiv +\infty$ if and only if $r = \rho$. Our main result is the following.

Theorem 4.2 *Let X_t^ε be completely asymptotically regular and $N > 1$. Suppose that for each $0 \le r \le \rho$ and each $1 \le m \le n_r$ either $t(\varepsilon) \ll T^r(m,\varepsilon)$ or $t(\varepsilon) \gg T^r(m,\varepsilon)$. Then there is a family of probability measures $\nu(i,\cdot)$, $i \in S$, on S such that*

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = \nu(i,j). \tag{4.3}$$

Remark The condition on $t(\varepsilon)$ stated in Theorem 4.2 requires that we compare $t(\varepsilon)$ with each of the time scales $T^r(m,\varepsilon)$, $0 \le r \le \rho$, $1 \le m \le n_r$. It is not difficult to show (compare with [22]) that for each i, a finite sequence of time scales $f_n(i,\varepsilon)$, $1 \le n \le n^*$, can be selected from this collection such that $f_n(i,\varepsilon) \ll f_{n+1}(i,\varepsilon)$, $1 \le n \le n^* - 1$, and $f_{n^*}(i,\varepsilon) = T^\rho(1,\varepsilon) \equiv \infty$. The condition on $t(\varepsilon)$ in Theorem 4.2 can be relaxed to where we require only that there is $1 \le n \le n^* - 1$ such that $f_n(i,\varepsilon) \ll t(\varepsilon) \ll f_{n+1}(i,\varepsilon)$.

It is important that the metastable distributions $\nu(i,\cdot)$ and the time scales $T^r(m,\varepsilon)$ can be expressed in terms of limits of invariant measures of certain completely regular Markov chains, as will be seen below. The invariant measure of a process X_t (whether ε-dependent or not) can be described using the notion of i-graphs (see Chap. 6 of [24]).

Namely, let X_t be an ergodic Markov chain on a state space E with transition rates q_{ij}. A directed graph with the set of vertices E is called an i-graph if no edges start at i, exactly one edge starts at each $k \in E \setminus \{i\}$, and for each $k \in E \setminus \{i\}$ there is a sequence of directed edges leading from k to i. Let G_i be the set of all i-graphs. Then the (non-normalized) invariant measure of the process X_t is given by

$$M(i) = \sum_{g \in G_i} \prod_{(m \to n) \in g} q_{mn},$$

where the product is taken over all directed edges $(m \to n) \in g$. If X_t^ε is a completely regular Markov chain, this expression allows one to calculate the limiting invariant distribution as $\varepsilon \downarrow 0$. Using this representation, one can also calculate the main terms of transition probabilities for Markov chains of higher ranks.

4.2 Proof of the Main Result

We start with a lemma that gives the same result as Theorem 4.2, but under an additional assumption on the structure of the hierarchy. Let

$$\mu^{r,k}(i) = \lim_{\varepsilon \downarrow 0} \mu^{r,k}(i, \varepsilon) \text{ if } S_i^r \in S_k^{r+1}; \quad \mu^{r,k}(i) = 0 \text{ if } S_i^r \notin S_k^{r+1}, \tag{4.4}$$

where the existence of the limit is guaranteed by Lemma 3.1 applied to the chain $Y_t^{r,k,\varepsilon}$. For $-1 \le r < \rho$ and $j \prec S_k^{r+1}$ such that $j = S_{j_0}^0 \in S_{j_1}^1 \ldots \in S_{j_{r+1}}^{r+1} = S_k^{r+1}$, define

$$\nu^r(j) = \mu^{0,j_1}(j_0)\mu^{1,j_2}(j_1) \ldots \mu^{r,j_{r+1}}(j_r), \tag{4.5}$$

where the right hand side is defined to be one if $r = -1$. For $i \in S$, let $r(i)$ be the minimal value of r such that there is k with $i \prec S_k^{r+1}$ and $t(\varepsilon) \ll T^{r+1}(k, \varepsilon)$. Let $L(i) = \{l : S_l^{r(i)} \in S_k^{r(i)+1}$ and $t(\varepsilon) \ll T^{r(i)}(l, \varepsilon)\}$. Obviously, $L(i) = \varnothing$ if $r(i) = -1$ since then there are no l for which $S_l^{r(i)}$ is defined. As will be seen below, the meaning of $L(i)$ is the following: if $l \in L(i)$ and the process X_t^ε starts at i and enters the set $\{j : j \prec S_l^{r(i)}\}$, then it will not leave this set prior to time $t(\varepsilon)$ with probability that tends to one. The following lemma provides the description of the metastable distribution in the case when $L(i) = \varnothing$.

Lemma 4.3 *Let the assumptions made in Theorem 4.2 hold. Let k be such that $i \prec S_k^{r(i)+1}$. Suppose that $L(i) = \varnothing$. If j is such that there are $j_0, \ldots, j_{r(i)}$ with*

$$j = S_{j_0}^0 \in S_{j_1}^1 \ldots \in S_{j_{r(i)}}^{r(i)},$$

then

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X_{t(\varepsilon)}^\varepsilon = j) = \nu^{r(i)}(j). \tag{4.6}$$

If $j \prec S_k^{r(i)+1}$ does not hold, then

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X_{t(\varepsilon)}^\varepsilon = j) = 0. \tag{4.7}$$

Moreover, assume that $\bar{t}(\varepsilon)$ satisfies $c_1 t(\varepsilon) \le \bar{t}(\varepsilon) \le c_2 t(\varepsilon)$, where $0 < c_1 < c_2$. Then the limits in (4.6) and (4.7) exist when $t(\varepsilon)$ is replaced by $\bar{t}(\varepsilon)$ and are uniform with respect to the choice of the function $\bar{t}(\varepsilon)$.

Proof The last statement easily follows by noticing that $t(\varepsilon)$ can be replaced by $\bar{t}(\varepsilon)$ in all the arguments below. Therefore, we'll focus on proving (4.6) and (4.7).

Let us start by briefly explaining the main idea of the proof. First, let us "reduce" the state space of X_t^ε by clumping all the states i' with the property that $i' \prec S_m^{r(i)}$ into a single state (recall the definition of '$\prec$' from (2.4)). The resulting process is well-approximated by the Markov chain $Y_t^{r(i),k,\varepsilon}$, where k is such that $X_0^\varepsilon = i \prec S_k^{r(i)+1}$. Let us observe the process on a time scale $s(\varepsilon) \sim t(\varepsilon)$, such that $s(\varepsilon)$ is slightly smaller than $t(\varepsilon)$. Take m such that $j \prec S_m^{r(i)}$. From the properties of $Y_t^{r(i),k,\varepsilon}$ it then follows that $\lim_{\varepsilon\downarrow 0} \mathrm{P}_i(X_{s(\varepsilon)}^\varepsilon \prec S_m^{r(i)}) = \mu^{r(i),k}(m)$, yielding the last factor in the expression (4.5). Next, we can consider the process X_t^ε on the time scale $t(\varepsilon) - s(\varepsilon)$ starting at a point $i' = X_{s(\varepsilon)}^\varepsilon \prec S_m^{r(i)}$. (In other words, we condition on the event that $X_{s(\varepsilon)}^\varepsilon \prec S_m^{r(i)}$ and use the Markov property. The complementary event can be seen not to contribute much to the probability that $X_{t(\varepsilon)}^\varepsilon = j$.) Provided that $s(\varepsilon)$ is chosen appropriately, the problem of identifying the limiting distribution at the time scale $t(\varepsilon) - s(\varepsilon)$ is similar to the original one, but with $r(i') = r(i) - 1$. Iterating the argument $r(i)$ times, we'll get the desired distribution. Let us now make the above arguments formal.

For $0 \le r < \rho$, define the process $\bar{Y}_t^{r,\varepsilon}$ via

$$\bar{Y}_t^{r,\varepsilon} = S_m^r \quad \text{if} \quad X_t^\varepsilon \prec S_m^r.$$

This is not necessarily a Markov process. However, below we'll see that on certain time scales it is close, in a certain sense, to the Markov process $Y_t^{r,k,\varepsilon}$, where k is such that $X_0^\varepsilon = i \prec S_k^{r+1}$. Suppose that $s(\varepsilon)$ is such that $s(\varepsilon) \ll T^{r+1}(k,\varepsilon)$, yet $T^r(l,\varepsilon) \ll s(\varepsilon)$ whenever $S_l^r \in S_k^{r+1}$. We claim (and will prove below) that

$$\lim_{\varepsilon\downarrow 0} \mathrm{P}(\bar{Y}_{s(\varepsilon)}^{r,\varepsilon} = S_l^r | X_0^\varepsilon = i) = \mu^{r,k}(l) \tag{4.8}$$

for each $1 \le l \le n_r$, where $\mu^{r,k}$, defined in (4.4), is the limit, as $\varepsilon \downarrow 0$, for the invariant measures of the processes $Y_t^{r,k,\varepsilon}$.

Recall that $t(\varepsilon) \ll T^{r(i)+1}(k,\varepsilon)$, while $T^{r(i)}(l,\varepsilon) \ll t(\varepsilon)$ whenever $S_l^{r(i)} \in S_k^{r(i)+1}$. We'll see that we can choose $s(\varepsilon) < t(\varepsilon)$ with the following properties:

(a) $s(\varepsilon) \ll T^{r(i)+1}(k,\varepsilon)$, while $T^{r(i)}(l,\varepsilon) \ll s(\varepsilon)$ whenever $S_l^{r(i)} \in S_k^{r(i)+1}$.
(b) For each l such that $\mu^{r(i),k}(l) > 0$, we have $t(\varepsilon) - s(\varepsilon) \ll T^{r(i)}(l,\varepsilon)$, while $T^{r'}(l',\varepsilon) \ll t(\varepsilon) - s(\varepsilon)$ whenever $S_{l'}^{r'}$ is such that $S_{l'}^{r'} \prec S_l^{r(i)}$ and there is $j \prec S_l^{r(i)}$ for which $j \prec S_{l'}^{r'}$ does not hold. (Here $S_{l'}^{r'} \prec S_l^{r(i)}$ means that $S_{l'}^{r'} \in S_{l_1}^{r'+1} \in \ldots \in S_{l_m}^{r'+m} \in S_l^{r(i)}$ for some $m, l_1, \ldots, l_m$.)

The former property means that $s(\varepsilon)$ satisfies the same asymptotic bounds as $t(\varepsilon)$. The latter property means that $t(\varepsilon) - s(\varepsilon)$ is not sufficiently large for $\bar{Y}_t^{r(i),\varepsilon}$ to escape from $S_l^{r(i)}$, yet sufficiently large for $\bar{Y}_t^{r',\varepsilon}$ with $r' < r$ to make many transitions between different states contained (in the sense of $\prec$) in $S_l^{r(i)}$.

Assuming that $S_k^{r(i)+1}$ contains more than one state, we can apply Lemma 3.1 to the chain $Y_t^{r(i),k,\varepsilon}$ and conclude that the quantities $T^{r(i)}(l,\varepsilon)$ with l such that $\mu^{r(i),k}(l) > 0$ are all asymptotically equivalent, up to multiplicative constants. Also note that $T^{r'}(l',\varepsilon) \ll T^{r(i)}(l,\varepsilon)$ whenever $S_{l'}^{r'}$ is such that $S_{l'}^{r'} \prec S_l^{r(i)}$ and there is $j \prec S_l^{r(i)}$ for which $j \prec S_{l'}^{r'}$ does not hold. This implies the existence of $s(\varepsilon)$ with the desired properties.

By property (a), from (4.8) with $r = r(i)$, it follows that $X_{s(\varepsilon)}^\varepsilon \prec S_l^{r(i)}$ with probability close to $\mu^{r(i),k}(l)$ for l such that $\mu^{r(i),k}(l) > 0$. Then, by the Markov property, $X_{s(\varepsilon)}^\varepsilon$ can be taken as a new starting point for the process, which will be studied on the time interval $t(\varepsilon) - s(\varepsilon)$. By property (b), for this new time scale and the new initial point, all the assumptions of Lemma 4.3 are satisfied, with the exception that the value of $r(i)$ is reduced by one if $S_l^{r(i)}$ contains only one element and by more than one if $S_l^{r(i)}$ contains only one element. Thus, conditioning on the event that $\bar{Y}_{s(\varepsilon)}^{r(i),\varepsilon} = S_l^r$, with l such that $j \prec S_l^{r(i)}$, gives

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X_{t(\varepsilon)}^{\varepsilon} = j) = \mu^{r(i),k}(l) \lim_{\varepsilon \downarrow 0} \mathrm{P}_{i'}(X_{t(\varepsilon)-s(\varepsilon)}^{\varepsilon} = j), \tag{4.9}$$

provided that the limit in the right hand side is defined and does not depend on $i' \prec S_l^{r(i)}$. Iterating this argument up to $r(i)+1$ times gives (4.6). It should be noted that if $S_l^{r(i)}$ contains only one element, then $r(i') < r(i) - 1$ in (4.9). This does not pose a problem, since the iterations can still be continued with the limit appearing in the right hand side of (4.9), while $r(i) - r(i') - 1$ consecutive factors in (4.5) will be equal to one. It remains to prove (4.8).

For $i \in S$ and $-1 \le r < \rho$, let k be such that $i \prec S_k^{r+1}$. Let $A^r = \{j : j \prec S_k^{r+1}\}$. For $0 \le r < \rho$, let $\tilde{X}_t^{r,\varepsilon}$ be the Markov chain on the state space A^r, whose transition rates agree with those of the chain X_t^ε (i.e., $\tilde{X}_t^{r,\varepsilon}$ is obtained from X_t^ε by disallowing transitions outside the specified state space).

For $j \prec S_k^{r+1}$ such that $j = S_{j_0}^0 \in S_{j_1}^1 \ldots \in S_{j_{r+1}}^{r+1}$ with $j = j_0$, $k = j_{r+1}$, let

$$M^r(j, \varepsilon) = \mu^{0,j_1}(j_0, \varepsilon)\mu^{1,j_2}(j_1, \varepsilon) \ldots \mu^{r,j_{r+1}}(j_r, \varepsilon).$$

For $0 \le r < \rho$, let $\alpha^r(j, s(\varepsilon))$ be the proportion of time, prior to $s(\varepsilon)$, that the process $\tilde{X}_t^{r,\varepsilon}$ spends in j. Let

$$\bar{T}^r(\varepsilon) = \sum_{l : S_l^r \in S_k^{r+1}} \mu^{r,k}(l) T^r(l, \varepsilon).$$

Define the process $\tilde{Y}_t^{r,\varepsilon}$ via

$$\tilde{Y}_t^{r,\varepsilon} = S_m^r \quad \text{if} \quad \tilde{X}_t^{r,\varepsilon} \prec S_m^r.$$

Recall that $s(\varepsilon)$ satisfies $s(\varepsilon) \ll T^{r+1}(k, \varepsilon)$ and $T^r(l, \varepsilon) \ll s(\varepsilon)$ whenever $S_l^r \in S_k^{r+1}$. We claim that for each $0 \le r < \rho$ and $s(\varepsilon)$ satisfying these conditions, the following asymptotic relations hold. For each $i, j \in A^r$ and l such that $S_l^r \in S_k^{r+1}$,

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}(\tilde{Y}_{s(\varepsilon)}^{r,\varepsilon} = S_l^r | \tilde{X}_0^{r,\varepsilon} = i) = \mu^{r,k}(l). \tag{4.10}$$

For each $c > 0$, there are $\delta(c) > 0$ and a function $\varphi(\varepsilon) > 0$ such that $\lim_{\varepsilon \downarrow 0} \varphi(\varepsilon) = +\infty$ and

$$\mathrm{P}_i(\alpha^r(j, s(\varepsilon)) < (1 - c)M^r(j, \varepsilon)) \le e^{-\delta(c)\varphi(\varepsilon)}, \tag{4.11}$$

$$\mathrm{P}_i(\alpha^r(j, s(\varepsilon)) > (1 + c)M^r(j, \varepsilon)) \le e^{-\delta(c)\varphi(\varepsilon)}. \tag{4.12}$$

When $r = 0$, (4.10)–(4.12) follow from Lemma 3.2. Next, let us sketch the inductive step, i.e., the proof that relations (4.10)–(4.12) hold for a fixed $r > 0$, assuming that they hold for all the smaller values of r.

The process $\tilde{Y}_t^{r,\varepsilon}$ is close to the Markov process $Y_t^{r,k,\varepsilon}$ in the following sense.

(a) If $l \neq m$ and $S_l^r, S_m^r \in S_k^{r+1}$, $i \prec S_l^r$, then

$$\mathrm{P}(\text{the first jump of } \tilde{Y}_t^{r,\varepsilon} \text{ is to } S_m^r | \tilde{X}_0^{r,\varepsilon} = i) \sim Q_{lm}^r(\varepsilon) \quad \text{as } \varepsilon \downarrow 0.$$

(b) If β^ε is the random time till the first transition of $\tilde{Y}_t^{r,\varepsilon}$, then

$$\mathrm{E}(\beta^\varepsilon|\tilde{X}_0^{r,\varepsilon} = i) \sim T^r(l,\varepsilon), \quad \text{as } \varepsilon \downarrow 0,$$

and there are $\delta > 0$ and $\varepsilon_0 > 0$ such that

$$\mathrm{P}(\beta^\varepsilon \geq \lambda T^r(l,\varepsilon)|\tilde{X}_0^{r,\varepsilon} = i) \leq e^{-\delta\lambda}, \quad \lambda \geq 2, \quad \varepsilon \in (0,\varepsilon_0].$$

The validity of (a) and (b) easily follows by examining the Markov chain $\tilde{X}_t^{r-1,\varepsilon}$ and utilizing the fact that (4.11)–(4.12) hold with r replaced by $r-1$.

Lemma 3.2 (with $s(\varepsilon)$ instead of $t(\varepsilon)$) can be applied to the process $Y_t^{r,k,\varepsilon}$. However, we are interested a similar result for the process $\tilde{Y}_t^{r,\varepsilon}$. It is not difficult to see that conditions (a) and (b) on the transition probabilities and transition times are sufficient for the proof of Lemma 3.2 to go through and for the result to be valid for the process $\tilde{Y}_t^{r,\varepsilon}$ (compare with Sect. 6.3 of [24]). Thus we have (4.10). Moreover, for $i \prec S_l^r$ and each $c > 0$, there are $\delta(c) > 0$ and a function $\varphi(\varepsilon) > 0$ such that $\lim_{\varepsilon\downarrow 0}\varphi(\varepsilon) = +\infty$ and

$$\mathrm{P}(\tilde{\alpha}^r(l,s(\varepsilon)) < (1-c)\mu^{r,k}(l,\varepsilon)|\tilde{X}_0^{r,\varepsilon} = i) \leq e^{-\delta(c)\varphi(\varepsilon)},$$
$$\mathrm{P}(\tilde{\alpha}^r(l,s(\varepsilon)) > (1+c)\mu^{r,k}(l,\varepsilon)|\tilde{X}_0^{r,\varepsilon} = i) \leq e^{-\delta(c)\varphi(\varepsilon)},$$

where $\tilde{\alpha}^r(l,s(\varepsilon))$ is the proportion of time, prior to $s(\varepsilon)$, that the process $\tilde{Y}_t^{r,\varepsilon}$ spends in the state S_l^r. Together with (4.11)–(4.12) for $r-1$ instead of r, these are easily seen to control the proportion of time, prior to $s(\varepsilon)$, that $\tilde{X}_t^{r,\varepsilon}$ spends in j, thus yielding (4.11)–(4.12) for r.

From (4.11)–(4.12) it easily follows that

$$\mathrm{P}_i(X_t^\varepsilon \prec S_k^{r+1} \text{ for all } t \leq s(\varepsilon)) \to 1 \quad \text{as } \varepsilon \downarrow 0.$$

Thus (4.8) follows from (4.10). □

Proof of Theorem 4.2. It remains to consider the case when $L(i) \neq \varnothing$ and explain how to identify the metastable distributions. Let $l(i)$ be such that $i \prec S_{l(i)}^r$, where $r = r(i)$. Observe that $l(i) \notin L(i)$. Recall that the state space of $Y_t^{r,k,\varepsilon}$ is S_k^{r+1}. We define a new state space $\widehat{S}_k^{r+1}$ by removing all the states S_l^r, $l \in L(i)$, from S_k^{r+1} and adjoining the set $E = \{j : j \prec S_l^r \text{ for some } l \in L(i)\}$. On this new state space, we define the Markov chain $\widehat{Y}_t^{r,k,\varepsilon}$. Its transition rates are defined as follows. If $l, m \notin L(i)$, $l \neq m$, are such that $S_l^r, S_m^r \in S_k^{r+1}$, then

$$\widehat{Q}_{S_l^r S_m^r}(\varepsilon) = Q_{lm}^r(\varepsilon).$$

The transition rates between $S_l^r \subset S_k^{r+1}$, $l \notin L(i)$, and $j \in E$ are defined as

$$\widehat{Q}_{S_l^r j}(\varepsilon) = \widehat{Q}_{lj}^r(\varepsilon),$$

where the quantity in the right hand side has been defined after the construction of the hierarchy. The transition rate from $j \in E$ to any other state is zero, i.e., E is the terminal set. It is not difficult to show that $\widehat{Y}_t^{r,k,\varepsilon}$ satisfies the assumptions of Lemma 3.3, i.e., it coincides with an asymptotically regular Markov chain stopped upon entering E. (The proof of this statement is the same as the proof of Lemma 2.2.) Therefore, by (3.11), there is a probability measure $\eta(i,\cdot)$ on E such that

$$\lim_{\varepsilon\downarrow 0}\lim_{t\to\infty}\mathrm{P}(\widehat{Y}_t^{r,k,\varepsilon} = j|\widehat{Y}_0^{r,k,\varepsilon} = S_{l(i)}^r) = \eta(i,j), \quad j \in E.$$

We'll use induction on the value of $r(i)$ to prove the following statements.

(a) Given two positive constants $c_1 < c_2$, the limits $\lim_{\varepsilon\downarrow 0} \mathrm{P}_i(X^\varepsilon_{\bar{t}(\varepsilon)} = j)$ exist and are uniform with respect to the choice of function $\bar{t}(\varepsilon)$ satisfying $c_1 t(\varepsilon) \le \bar{t}(\varepsilon) \le c_2 t(\varepsilon)$.

(b) Suppose that $j \prec S^{r(i)}_m$, where $m \in L(i)$. Then

$$\lim_{\varepsilon\downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = \sum_{i' \prec S^{r(i)}_m} \eta(i, i') \lim_{\varepsilon\downarrow 0} \mathrm{P}_{i'}(X^\varepsilon_{t(\varepsilon)} = j).$$

(c) If $L(i) \neq \varnothing$ and j is such that $j \prec S^{r(i)}_m$ does not hold for any $m \in L(i)$, then

$$\lim_{\varepsilon\downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = 0.$$

These statements imply Theorem 4.2 (statement (a)) and explain how to identify the metastable distributions. Indeed, if $L(i) = \varnothing$, then the metastable distribution is given by Lemma 4.3. If $L(i) \neq \varnothing$, then, by statements (b) and (c), $\nu(i, j)$ is either equal to zero or is equal to a linear combination of the quantities $\nu(i', j)$ with $r(i') < r(i)$. The values of $\nu(i', j)$ can be found from Lemma 4.3 (when $L(i') = \varnothing$), or again expressed in terms of metastable distributions with different initial points using statements (b) and (c). This recursive procedure can be continued until all the resulting initial points j satisfy $L(j) = \varnothing$, in which case Lemma 4.3 can be applied (which will happen in no more than $r(i) + 1$ steps since $L(j) = \varnothing$ whenever $r(j) = -1$).

It remains to prove properties (a)–(c). If $r(i) = -1$, then $L(i) = \varnothing$, and (a)-(c) trivially hold. If $L(i) = \varnothing$, then (a) holds by Lemma 4.3, while (b) and (c) are trivial. Now assume that $L(i) \neq \varnothing$. Let σ be the first time when the process X^ε_t reaches E. We claim that

$$\lim_{\varepsilon\downarrow 0} \mathrm{P}_i(X^\varepsilon_\sigma = j) = \eta(i, j), \quad j \in E, \tag{4.13}$$

$$\lim_{\varepsilon\downarrow 0} \mathrm{P}_i(\sigma > \frac{t(\varepsilon)}{2}) = 0. \tag{4.14}$$

Note that Lemma 4.3 concerns the distribution of X^ε_t at time $t(\varepsilon)$, while here we are interested in the distribution at time σ. The role of the process $Y^{r,k,\varepsilon}_t$ is now played by $\widehat{Y}^{r,k,\varepsilon}_t$, and the main difference is that we need to use Lemma 3.3 instead of Lemma 3.2. Otherwise, the proof of (4.13)–(4.14) relies on the same arguments as those provided in the proof of Lemma 4.3, and thus we do not provide it here.

Let $\sigma' = t(\varepsilon) - \sigma \wedge t(\varepsilon)$. Using the strong Markov property of the process, we write

$$\mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = \sum_{i' \in S} \mathrm{P}_i(X^\varepsilon_{\sigma\wedge t(\varepsilon)} = i')\mathrm{E}_i \mathrm{P}_{i'}(X^\varepsilon_{\sigma'} = j).$$

From (4.13)–(4.14) it follows that the factor $\mathrm{P}_i(X^\varepsilon_{\sigma\wedge t(\varepsilon)} = i')$ in each of the terms has a limit $\varepsilon \downarrow 0$. The limit is nonzero only for i' such that $i' \prec S^r_l$ for some $l \in L(i)$ and is equal to $\eta(i, i')$. Moreover, if $l \neq m$ (where m is such that $j \prec S^{r(i)}_m$), then the factor $\mathrm{E}_i \mathrm{P}_{i'}(X^\varepsilon_{\sigma'} = j)$ tends to zero as $\varepsilon \downarrow 0$. Indeed,

$$\lim_{\varepsilon\downarrow 0} \mathrm{P}_i\left(\sigma' \in \left[\frac{t(\varepsilon)}{2}, t(\varepsilon)\right]\right) = 1 \tag{4.15}$$

by (4.14), while $r(i') < r(i)$. Therefore, by the inductive assumption and properties (a) and (c), $\lim_{\varepsilon \downarrow 0} \mathrm{E}_i \mathrm{P}_{i'}(X^\varepsilon_{\sigma'} = j) = 0$ if $i' \prec S^r_l$, $l \in L(i)$, and $l \neq m$. Therefore,

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = \sum_{i' \prec S^{r(i)}_m} \eta(i, i') \lim_{\varepsilon \downarrow 0} \mathrm{E}_i \mathrm{P}_{i'}(X^\varepsilon_{\sigma'} = j), \tag{4.16}$$

provided the limits in the right hand side exist. Similarly, if no m with $j \prec S^{r(i)}_m$ exists, then

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{t(\varepsilon)} = j) = 0,$$

which proves property (c). By (4.15) and property (a) applied to i' instead of i, the limits in the right hand side of (4.16) can be written as

$$\lim_{\varepsilon \downarrow 0} \mathrm{E}_i \mathrm{P}_{i'}(X^\varepsilon_{\sigma'} = j) = \lim_{\varepsilon \downarrow 0} \mathrm{P}_{i'}(X^\varepsilon_{t(\varepsilon)} = j).$$

By (4.16), this implies property (b). Finally, property (a) follows from (b) if we observe that the arguments leading to the proof of (b) can be applied to $\bar{t}(\varepsilon)$ instead of $t(\varepsilon)$. □

5 Complete Asymptotic Regularity

In this section, we prove Lemma 2.2. We start with the following simple lemma.

Lemma 5.1 *Suppose that $a_1(\varepsilon), \ldots, a_n(\varepsilon)$ and $b_1(\varepsilon), \ldots, b_{n'}(\varepsilon)$ are positive functions such that there are limits*

$$\lim_{\varepsilon \downarrow 0} \frac{a_i(\varepsilon)}{b_{i'}(\varepsilon)} \in [0, \infty], \quad 1 \le i \le n, \ 1 \le i' \le n'.$$

Let $A(\varepsilon) = a_1(\varepsilon) + \cdots + a_n(\varepsilon)$, $B(\varepsilon) = b_1(\varepsilon) + \cdots + b_{n'}(\varepsilon)$. Then there is the limit

$$\lim_{\varepsilon \downarrow 0} \frac{A(\varepsilon)}{B(\varepsilon)} \in [0, \infty].$$

Proof For each i', there is the limit

$$\lim_{\varepsilon \downarrow 0} \frac{A(\varepsilon)}{b_{i'}(\varepsilon)} = \lim_{\varepsilon \downarrow 0} \frac{a_1(\varepsilon)}{b_{i'}(\varepsilon)} + \cdots + \lim_{\varepsilon \downarrow 0} \frac{a_n(\varepsilon)}{b_{i'}(\varepsilon)} \in [0, \infty].$$

Therefore, there is the limit

$$\lim_{\varepsilon \downarrow 0} \frac{A(\varepsilon)}{B(\varepsilon)} = (\lim_{\varepsilon \downarrow 0} \frac{b_1(\varepsilon)}{A(\varepsilon)} + \cdots + \lim_{\varepsilon \downarrow 0} \frac{b_{n'}(\varepsilon)}{A(\varepsilon)})^{-1} \in [0, \infty].$$

□

Proof of Lemma 2.2 The case $N = 1$ is trivial. Let us assume that $N > 1$. Recall the decomposition (2.1) of S into ergodic classes and transient states. To prove the lemma, we need to show that for $a \ge 1$ there is the limit

$$\lim_{\varepsilon \downarrow 0} \left(\frac{Q_{k_1 l_1}(\varepsilon)}{Q_{m_1 n_1}(\varepsilon)} \times \cdots \times \frac{Q_{k_a l_a}(\varepsilon)}{Q_{m_a n_a}(\varepsilon)} \right) \in [0, \infty], \tag{5.1}$$

provided that $k_1 \neq l_1, \ldots, k_a \neq l_a, m_1 \neq n_1, \ldots, m_a \neq n_a$. We will repeatedly use Lemma 5.1, which will allow us to replace each of the factors above by simpler expressions.

First consider a factor of the form Q_{kl}/Q_{mn} under the assumption that $S_k = \{i\}$ and $S_m = \{i'\}$, i.e., S_k and S_m have only one element each. In this case

$$\frac{Q_{kl}(\varepsilon)}{Q_{mn}(\varepsilon)} = \frac{\sum_{j\in S_l} q_{ij}(\varepsilon)}{\sum_{j'\in S_n} q_{i'j'}(\varepsilon)}.$$

Thus, by Lemma 5.1, it is sufficient to prove the existence of the limit in (5.1) with all such factors replaced by those of the form $q_{ij}/q_{i'j'}$.

Next consider a factor of the form Q_{kl}/Q_{mn} under the assumption that one of the sets S_k and S_m (say, S_k) has at least two elements, while the other one has one element, i.e., $S_m = \{i'\}$. Then the chain $Y_t^{k,\varepsilon}$ satisfies the assumptions of Lemma 3.1, and therefore, by (2.3),

$$\frac{Q_{kl}(\varepsilon)}{Q_{mn}(\varepsilon)} = \frac{\sum_{i\in S_k}\sum_{j\in S_l} \mu^k(i,\varepsilon) q_{ij}(\varepsilon)}{\sum_{j'\in S_n} q_{i'j'}(\varepsilon)} \sim \frac{\sum_{i\in S_k}\sum_{j\in S_l} \lambda(i)T(i,\varepsilon)q_{ij}(\varepsilon)}{\sum_{j'\in S_n} q_{i'j'}(\varepsilon)\sum_{i''\in S_k}(\lambda(i'')T(i'',\varepsilon))},$$

where λ and T are the invariant measure for the skeleton chain and the inverse transition rate, respectively, for the chain $Y_t^{k,\varepsilon}$. Thus, by Lemma 5.1, it is sufficient to prove the existence of the limit in (5.1) with Q_{kl}/Q_{mn} replaced by $T(i,\varepsilon)q_{ij}(\varepsilon)(T(i'',\varepsilon)q_{i'j'}(\varepsilon))^{-1}$, where $i, i'' \in S_k$. By the definition of T,

$$\frac{T(i,\varepsilon)q_{ij}(\varepsilon)}{T(i'',\varepsilon)q_{i'j'}(\varepsilon)} = \frac{\sum_{b'\in S_k, b'\neq i''} q_{i''b'}(\varepsilon) q_{ij}(\varepsilon)}{\sum_{b\in S_k, b\neq i} q_{ib}(\varepsilon) q_{i'j'}(\varepsilon)}.$$

By Lemma 5.1, each such expression can be replaced by $(q_{i''b'}(\varepsilon)q_{ij}(\varepsilon))/(q_{ib}(\varepsilon)q_{i'j'}(\varepsilon))$.

The final case, when S_k and S_m have at least two elements each, is treated similarly, resulting in Q_{kl}/Q_{mn} being replaced by a product of three factors of the form $q_{ij}/q_{i'j'}$. Thus we see that each factor in (5.1) can be replaced by either one, two, or three factors of the form $q_{ij}/q_{i'j'}$. Therefore, the limit in (5.1) exists since the original chain is completely asymptotically regular. □

6 Remarks and Generalizations

(A) One could replace the complete asymptotic regularity by a somewhat weaker assumption that also implies the asymptotic regularity of all the reduced chains appearing in the inductive construction of the hierarchy. We will say that an asymptotically regular family of Markov chains X_t^ε satisfies Condition (A) if for each $0 \le a \le N-2$ the following finite or infinite limit exists

$$\lim_{\varepsilon\downarrow 0}\left(\frac{q_{k_1 l_1}(\varepsilon)}{q_{k_1 n_1}(\varepsilon)} \times \cdots \times \frac{q_{k_a l_a}(\varepsilon)}{q_{k_a n_a}(\varepsilon)} \times \frac{q_{k'l'}(\varepsilon)}{q_{m'n'}(\varepsilon)}\right) \in [0,\infty],$$

provided that $k_1,\ldots,k_a$ are all distinct, $k', m' \notin \{k_1,\ldots,k_a\}$, and $k_1 \neq l_1,\ldots,k_a \neq l_a$, $k_1 \neq n_1,\ldots,k_a \neq n_a$, $k' \neq l'$, $m' \neq n'$. The proof of the following lemma is similar to that of Lemma 2.2, and so we don't provide it here.

Lemma 6.1 *If X_t^ε satisfies Condition (A), then the reduced chain also satisfies Condition (A).*

If X_t^ε satisfies Condition (A), then, by Lemma 6.1, so do the reduced Markov chains that appear at each step of the inductive construction of the hierarchy, which implies that all of them are asymptotically regular.

(B) Next, let us mention that all the above analysis can be easily adapted to the case of discrete-time Markov chains. Instead of (1.1), we assume that X_n^ε, $n \geq 0$, is a Markov chain with transition probabilities $p_{ij}(\varepsilon)$, $i, j \in S$. We impose an additional assumption that $p_{ii}(\varepsilon) \geq c > 0$ for all $i \in S$, $\varepsilon > 0$. This is needed in order for the discrete time analogue of Lemma 3.2 to remain valid.

The definitions of asymptotic regularity and complete asymptotic regularity remain the same as in the continuous time case, with transition rates $q_{ij}(\varepsilon)$ replaced by transition probabilities $p_{ij}(\varepsilon)$. The reduced Markov chains the chains chains $Y_t^{r,k,\varepsilon}$ can be still defined in continuous time, simply replacing $q_{ij}(\varepsilon)$ by $p_{ij}(\varepsilon)$ in all the definitions. The definition of the inverse transition rates (4.2) remains the same. The theorem on metastable distributions now takes the following form.

Theorem 6.2 *Let $n : (0, \infty) \to \mathbb{N}$ be such that for each $0 \leq r \leq \rho - 1$ and each $1 \leq i \leq n_r$ either $n(\varepsilon) \ll T^r(i, \varepsilon)$ or $n(\varepsilon) \gg T^r(i, \varepsilon)$. Then there is a family of probability measures $\nu(i, \cdot)$, $i \in S$, on S such that*

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_i(X^\varepsilon_{n(\varepsilon)} = j) = \nu(i, j).$$

The proof of this theorem is identical to that of Theorem 4.2.

(C) Finally, consider an example of a completely asymptotically regular family of Markov chains. Given numbers α_{ij}, β_{ij}, and γ_{ij}, $i \neq j$, such that $\alpha_{ij} > 0$, assume that the transition rates satisfy

$$q_{ij}(\varepsilon) \sim \alpha_{ij}\varepsilon^{\beta_{ij}} \exp(-\gamma_{ij}\varepsilon^{-1}), \quad \text{as } \varepsilon \downarrow 0, \quad i \neq j. \tag{6.1}$$

It is clear that these functions satisfy (2.5), and therefore the corresponding family is completely asymptotically regular. Replacing ε by $|\ln \tilde{\varepsilon}|^{-1}$, we obtain functions

$$\tilde{q}_{ij}(\tilde{\varepsilon}) \sim \alpha_{ij}|\ln \tilde{\varepsilon}|^{-\beta_{ij}}(\tilde{\varepsilon})^{\gamma_{ij}}, \quad \text{as } \varepsilon \downarrow 0, \quad i \neq j, \tag{6.2}$$

which also satisfy (2.5). The systems discussed in the Introduction lead to Markov chains with transition rates that satisfy either (6.1) or (6.2), with the exception that the condition $\alpha_{ij} > 0$ may be violated, i.e., some of the coefficients may be equal to zero. In fact, this positivity condition (or condition (a) in the definition of asymptotic regularity) are not that crucial. If i and j are such that $\alpha_{ij} = 0$, the transition rates can be re-defined for those (i, j) by taking $\alpha_{ij} = 1$ and γ_{ij} sufficiently large, resulting in a completely asymptotically regular family with the same metastable behavior as the original one.

As we already mentioned, the reduction of dynamics on an infinite state space (as in the examples provided in the Introduction) to a finite-state Markov chain often requires non-trivial steps. In particular, the example with heteroclinic networks formally fits the framework of our paper (i.e., the scaling of transition rates satisfy (6.2)). However, in order to approximate the genuine dynamics by a finite-state Markov chain, one needs to (a) interpret individual heteroclinic connections as states, (b) introduce stopping times appropriately, (c) carefully calculate the transition times and transition probabilities, and (d) deal with the fact that the process obtained this way is only approximately Markov. This program is currently being carrying out (see [3]), and will likely involve the results of this paper for a part of the analysis.

Acknowledgements We are grateful to three anonymous referees who read the paper very thoroughly and suggested many important improvements. While working on this article, M. Freidlin was supported by NSF Grant DMS-1411866 and L. Koralov was supported by NSF Grant DMS-1309084.

References

1. Aldous, D., Brown, M.: Inequalities for rare events in time-reversible Markov chains. I. Stochastic inequalities (Seattle, WA, 1991), 1–16, IMS Lecture Notes Monogr. Ser., 22, Inst. Math. Statist., Hayward, CA (1992)
2. Aldous, D., Brown, M.: Inequalities for rare events in time-reversible Markov chains II. Stoch. Process. Appl. **44**(1), 15–25 (1993)
3. Bakhtin, Y., Pajor-Gyulai, Z.: Metastability and cycle structure in strictly attracting noisy heteroclinic networks (in preparation)
4. Beltran, J., Landim, C.: Tunneling and metastability of continuous time Markov chains. J. Stat. Phys. **140**(6), 1065–1114 (2010)
5. Beltran, J., Landim, C.: A martingale approach to metastability. Probab. Theory Relat. Fields **161**(1–2), 267–307 (2015)
6. Berglund, N., Gentz, B.: The Eyring–Kramers law for potentials with nonquadratic saddles. Markov Process. Relat. Fields **16**(3), 549–598 (2010)
7. Betz, V., Le Roux, S.: Multi-scale metastable dynamics and the asymptotic stationary distribution of perturbed Markov chains. Stoch. Process. Appl. **126**(11), 3499–3526 (2016)
8. Bianchi, A., Gaudilliere, A.: Metastable states, quasi-stationary distributions and soft measures. Stoch. Process. Appl. **126**(6), 1622–1680 (2016)
9. Bovier, A., Eckhoff, M., Gayrard, V., Klein, M.: Metastability in stochastic dynamics of disordered meanfield models. Probab. Theory Relat. Fields **119**(1), 99–161 (2001)
10. Bovier, A., Eckhoff, M., Gayrard, V., Klein, M.: Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. J. Eur. Math. Soc. **6**(4), 399–424 (2004)
11. Bovier, A., Eckhoff, M., Gayrard, V., Klein, M.: Metastability and low lying spectra in reversible Markov chains. Commun. Math. Phys. **228**, 219–255 (2002)
12. Bovier, A., den Hollander, F.: Metastability. A Potential-Theoretical Approach. Springer, New York (2015)
13. Catoni O.: Simulated annealing algorithms and Markov chains with rare transitions. In: Seminaire de Probabilites, XXXIII, volume 1709 of Lecture Notes in Math., pp. 69–119. Springer, Berlin (1999)
14. Catoni, O., Cerf, R.: The exit path of a Markov chain with rare transitions. ESAIM Probab. Stat. 1: 95–144 (1995/1997)
15. Cirillo, E., Nardi, F.: Metastability for a stochastic dynamics with a parallel heat bath updating rule. J. Stat. Phys. **110**(1–2), 183–217 (2003)
16. Cirillo, E., Nardi, F.: Relaxation height in energy landscapes: an application to multiple metastable states. J. Stat. Phys. **150**(6), 1080–1114 (2013)
17. Cirillo, E., Nardi, F., Sohier, J.: Metastability for general dynamics with rare transitions: escape time and critical configurations. J. Stat. Phys. **161**, 365–403 (2015)
18. Doob, J.L.: Stochastic Processes. Wiley-Interscience. Revised edition (January 25, 1990)
19. Eckhoff, M.: Precise asymptotics of small eigenvalues of reversible diffusions in the metastable regime. Ann. Prob. **33**(1), 244–299 (2005)
20. Fernandez, R., Manzo, F., Nardi, F., Scoppola, E.: Asymptotically exponential hitting times and metastability: a pathwise approach without reversibility. Electron. J. Probab. **20**(122), 1–37 (2015)
21. Fernandez, R., Manzo, F., Nardi, F., Scoppola, E., Sohier, J.: Conditioned, quasi-stationary restricted measures and escape from metastable states. Ann. Appl. Prob. **26**, 760–793 (2016)
22. Freidlin, M.I.: Sublimiting distributions and stabilization of solutions of parabolic equations with a small parameter. Soviet Math. Dokl. **18**(4), 1114–1118 (1977)
23. Freidlin, M.I.: On stochastic perturbations of systems with rough symmetry. Hierarchy of Markov chains. J. Stat. Phys. **157**(6), 1031–1045 (2014)
24. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems, 3rd edn. Springer, Berlin (2012)
25. Freidlin, M.I., Koralov, L., Wentzell, A.D.: On the behavior of diffusion processes with traps. To appear in Annals of Probability
26. Hennion, H., Herve, L.: Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness. Springer, Berlin (2001)
27. Holmes-Cerfon, M., Gortler, S.J., Brenner, M.P.: A geometrical approach to computing free-energy landscapes from short-ranged potentials. Proc. Natl. Acad. Sci. **110**(1), E5–E14 (2013)
28. Huang, C., Sheu, S.: Singular perturbed Markov chains and exact behaviors of simulated annealing processes. J. Theor. Probab. **5**(2), 223–249 (1992)
29. Manzo, F., Nardi, F.R., Olivieri, E., Scoppola, E.: On the essential features of metastability: tunneling time and critical configurations. J. Stat. Phys. **115**(1/2), 591–642 (2004)

30. Olivieri, E., Scoppola, E.: Markov chains with exponentially small transition probabilities: first exit problem from a general domain. I. The reversible case. J. Stat. Phys. **79**(3/4), 613–647 (1995)
31. Olivieri, E., Scoppola, E.: Markov chains with exponentially small transition probabilities: first exit problem from a general domain. II. The general case. J. Stat. Phys. **84**(5/6), 987–1041 (1996)
32. Oliveiri, E., Vares, M.E.: Large Deviations and Metastability. Cambridge University Press, Cambridge (2005)
33. Scoppola, E.: Renormalization group for Markov chains and application to metastability. JSP **73**, 83–121 (1993)
34. Trouve, A.: Cycle decomposition and simulated annealing. SIAM J. Control. Optim. **34**, 966–986 (1996)
35. Trouve, A.: Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms. Ann. Inst. H. Poinc. **32**, 299–348 (1996)
36. Wentzell, A.D.: On the asymptotics of eigenvalues of matrices with elements of order $\exp(-V_{ij}/(2\varepsilon^2))$. Soviet Math. Dokl. **13**(1), 65–68 (1972)