Università degli studi dell'Aquila

Department of Information Engineering, Computer Science and Mathematics

UNIVERSITÀ DEGLI STUDI DELL'AQUILA

Ph.D. in Mathematics and Modelling, 38th Cycle

Probability and Mathematical Physics

APPLIED MARKOV CHAIN TREE THEOREM

Supervisors:

Prof. Alberici Diego
Prof. Aleandri Michele
Prof. Gabrielli Davide

Ph.D. Coordinator:

Prof. Amadori Debora

Ph.D. Candidate:

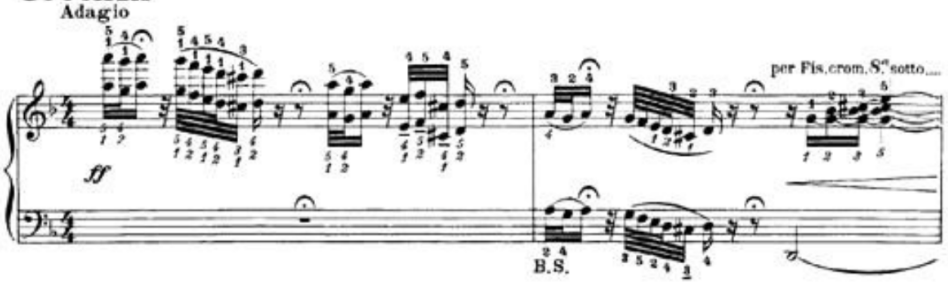
Pallotta Giulia

Roll Number: 289062

SSD Codes: MATH-03/B (Primary) MATH-04/A (Secondary)

TOCCATA

Adagio



Contents

Abstract

This doctoral thesis explores the multifaceted applications of the Markov chain tree Theorem in the study of stochastic processes, highlighting the deep connection between graph theory and the dynamics of continuous-time Markov chains. The work is organized into two main parts, each addressing asymptotic problems through a combinatorial lens.

In the First Part, we analyze the behavior of Markov chains on finite directed graphs where transition rates decay exponentially with respect to a positive parameter. We introduce the discrete counterpart of the Freidlin-Wentzell theory, formulating a discrete Hamilton-Jacobi equation for the large deviations functional of the invariant measure. Through a geometric characterization of viscosity solutions and the application of the Matrix tree Theorem, we classify all the solutions and establish a selection principle analogous to that observed in continuous diffusion processes.

The Second Part is devoted to the study of multiscale Markov chains and metastability phenomena. Leveraging the theory of trace processes, the thesis describes the decomposition of the limiting invariant measure through effective dynamics on reduced state spaces. By extending the Markov chain tree Theorem to chains that are not necessarily irreducible, the entire problem of metastability is framed within a purely combinatorial context based on the relationships between the weights of arborescences and directed forests.

Overall, the thesis demonstrates how the topological structure of the transition graph provides powerful and rigorous tools for resolving complex issues related to convergence and scale separation in stochastic systems.

Introduction

Some theorems in mathematics are so powerful that they establish unexpected connections between disparate fields, thereby indicating the tools best suited for further analysis. This is precisely the case of the Markov Chain Tree Theorem.

This result establishes a fundamental connection between the invariant measure of Markov chains and specific subgraphs of their associated transition graphs, known as spanning arborescences.

Moreover, the elegance of the theorem is reflected not only in its statement but also in the diversity of its proofs. Indeed, several alternative approaches exist, each highlighting a different mathematical structure: from matrix-based representations to directed graph techniques, as well as constructions based on liftings of Markov chains (see, for instance, [5, 27, 55]).

The purpose of this doctoral thesis is to illustrate the Markov Chain Tree Theorem in action. In the following chapters, we will build upon this result to study a sequence of irreducible Markov chains with exponentially decaying transition rates, to examine the large deviation limit of the invariant measure, to propose an extension to the non-irreducible case, and to apply the theorem to the analysis of metastability in multiscale Markov chains.

More specifically, this thesis is divided into two parts.

Part I, *Freidlin-Wentzell Theory on Metric Graphs*, presents the results from [4], a work carried out in collaboration with Prof. Davide Gabrielli (University of L'Aquila) and Prof. Michele Aleandri (LUISS University of Rome). Part II, *On the Limiting Invariant Measure of Multiscale Markov Chains*, describes some of the results obtained in [2], currently in preparation, that is a joint work with Prof. Diego Alberici and Prof. Davide Gabrielli, both at the University of L'Aquila.

Chapter 1 introduces the two parts of the thesis by presenting the common theoretical framework. We consider an irreducible continuous-time Markov chain X_t with transition digraph (V, E) and outline all the combinatorial objects that the reader will need to be familiar with throughout the thesis: arborescences, forests, maps, and inward- and outward-oriented unicyclic components. We also summarize the main notions of distances on directed graphs, in particular defining geodesic distances and resistance distances, thus connecting our results to the theory of electrical circuits. Finally, we focus on the statement and proof of the Markov Chain Tree Theorem for irreducible chains. Alongside this, we present two additional results that will be heavily used in the subsequent sections: the Harmonic Tree Formula, as proved in [55], and the ergodic theorem.

In Part I, we consider a sequence of finite irreducible Markov chains depending on a positive parameter N. The transition graph is a fixed, finite, strongly connected directed graph, while the transition rates decay exponentially with N, with a rate that may vary from edge to edge. The stationary equation that uniquely identifies the invariant measure for each N, at the exponential scale, in the limit as $N \to \infty$, reduces to a discrete equation for the large deviation rate functional of the invariant measure, which in general does not have a unique solution. In analogy with the continuous case of diffusions, we refer to this equation as a discrete Hamilton-Jacobi equation.

Similarly to the continuous setting, we introduce notions of viscosity supersolutions and viscosity subsolutions, and we provide a detailed geometric characterization of the solutions in terms of special faces of the polyhedron of Lipschitz functions on the transition graph. This construction parallels the weak KAM theory in a purely discrete framework. We also identify a special vanishing viscosity solution obtained in the limit from the combinatorial representation of the invariant measure provided by the Matrix Tree Theorem. This gives a selection principle on the set of solutions to the discrete Hamilton-Jacobi equation, obtained via the Freidlin-Wentzell minimal arborescences construction, thus elucidating a discrete analogue of what occurs in the continuous case.

Chapter 2 provides a brief overview, based on [19, 20, 24, 26, 27], of the known results in the continuous setting concerning diffusion processes. We briefly describe the large deviation principle, the Freidlin-Wentzell theory for the small-noise limit of a diffusion process, the associated Hamilton-Jacobi equations, and the weak KAM theory. Chapter 3 presents the new results in the discrete setting, as published in [4].

In Part II, we explore the asymptotic behavior of invariant measures in multiscale Markov chains. From a theoretical standpoint, multiscale Markov chains provide a precise description of how long-time behavior emerges from local equilibrium within

metastable sets. This approach has proven particularly effective in the study of interacting particle systems, statistical mechanics, and stochastic dynamics, where metastability plays a fundamental role. More generally, the analysis of multiscale Markov chains offers a unifying framework for understanding complex stochastic phenomena driven by rare events and scale separation.

In this section, we present the results obtained in the context of [2], concerning the decomposition of the limiting invariant measure into effective dynamics restricted to irreducible classes, leveraging trace process theory. We begin in Chapter 4 by focusing on the works [12, 13, 28, 40, 42, 43, 46, 47], which provide the definition of the trace process and a description of the metastability of a sequence of Markov chains. These results play a key role in our analysis.

Chapter 5 describes the multiscale behavior of irreducible chains and analyzes the limiting invariant measure. This analysis proceeds through the definition of an effective chain that operates on a reduced state space: the limiting invariant measure will concentrate only on certain states, allowing us to disregard transient states in the limit. The primary tool remains the Markov Chain Tree Theorem, now extended to chains that are not necessarily irreducible. The fact that the theorem applies both to the original process and to the trace process, through combinations of arborescences and forests, allows us to frame the entire problem within a purely combinatorial framework. This leads to special relations between the weights of trees and forests, which we summarize in the last sections of this thesis.

1

Digraphs and Continuous-Time Markov Chains

In this section, we aim to briefly outline the general framework of our work. The common foundation lies in the strong relationship between directed graphs and continuous-time Markov chains. Since a Markov chain can be naturally identified with its weighted transition graph, it is reasonable to seek results through a combinatorial perspective, in addition to the algebraic one.

1.1 Directed Graphs and Continuous-Time Markov Chains

We begin with some basic definitions from graph theory. This introductory section is intended to provide a general overview of the tools and concepts that will be used throughout the thesis. For a more comprehensive treatment, we refer the reader to [16] and [10].

1.1.1 Strongly Connected Digraphs

Definition 1. *A **directed graph** (or simply, a digraph)* (V, E) *consists of a non-empty finite set* V *of elements called* vertices*, and a finite set* E *of ordered pairs of distinct vertices, called* edges*.*

The *order* of the digraph (V, E) is the number of its vertices, denoted by $|V|$.

Visualizing a graph is often a powerful method for understanding its structure. In fact, one of the most effective ways to describe a small graph is simply to draw it (See Figure 1.1).

For each edge $(x, y) \in E$ we assign a positive number $r(x, y)$, which we call the **weight** of the edge (x, y). Henceforth, depending on the context and for the

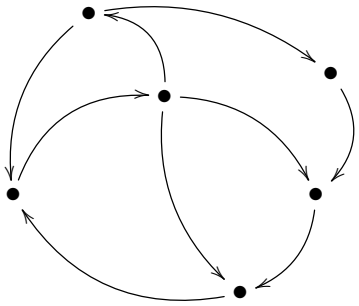


Figure 1.1: A digraph

sake of notational convenience, we shall refer to an edge either by its endpoints $(x, y) \in E$ or simply as $ein E$.

Definition 2. *We say that (V', E') is a **subgraph** of (V, E) if $V' \subset V$ and $E' \subset E$.*

If $V' = V$, then (V', E') is said to be a **spanning** subgraph of (V, E). We shall often construct new graphs from old ones by deleting or adding some vertices and edges.

In the following, (V, E) is always a finite digraph.

Definition 3. *A **path** from x_1 to x_k in (V, E) is an alternating sequence $\gamma_{x_1,x_k} = x_1 e_1 x_2 e_2 \dots x_{k-1} e_{k-1} x_k$ of distinct vertices $x_i \in V$ and edges $e_i := (x_i, x_{i+1}) \in E$ (such that the tail of e_i is x_i and the head of e_i is x_{i+1} for every $i = 1, \dots, k-1$).*

Given a weighted graph, the **weight** of a walk γ is defined as the product of the weight of every edge belonging to γ, namely:

$$r(\gamma_{xy}) = \prod_{e_i \in \gamma_{xy}} r(e_i).$$

A **geodesic path** between two vertices is the walk whose weight is the minimal one among all the other walks connecting the same endvertices.

If $\gamma_{x_1 x_k}$ is a path, $k \geq 3$ and $x_1 = x_k$, then $\gamma_{x_1 x_k}$ is a **cycle.**

Definition 4. *A digraph (V, E) is **strongly connected** if, for every pair x, y of distinct vertices in V there exists a path connecting x to y.*

For example, the digraph in Figure 1.1 is strongly connected.

From now on, unsless specified, we will consider finite, strongly connected digraphs.

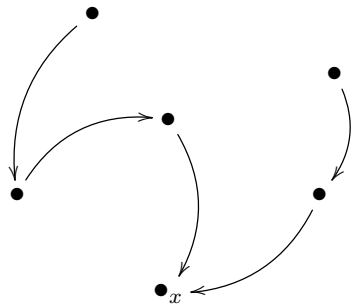


Figure 1.2: An arborescence rooted towards a vertex x

1.1.2 Arborescences and Directed Forests

Some general remarkable classes of graphs and their properties are needed to discuss: let us now move on to the definition of the key object of this thesis.

Definition 5. *Let (V, E) be a directed graph. A **spanning arborescence** τ directed towards $x \in V$ is a spanning subgraph of (V, E) such that:*

- *for each vertex $y \neq x$ there exists exactly one directed edge exiting from y and beloning to τ;*
- *for any $y \in V$ there exists one directed path from y to x in τ;*
- *there are no edges exiting from x.*

Let $\mathcal{T}_x$ be the set of arborescences spanning (V, E) directed toward $x \in V$ and $\mathcal{T}$ be the set of all spanning arborescences $\mathcal{T} = \bigcup_{x \in V} \mathcal{T}_x$.

If the graph is strongly connected, then $\mathcal{T}_x$ is non empty for any $x \in V$.

Just as in the case of paths, it is natural to define the **weight** of a spanning arborescence as follows:

$$r(\tau) := \prod_{(x,y)\in\tau} r(x, y)$$

for all $\tau \in \mathcal{T}$. It is also possible to assign a weight to a set of spanning arobrescences by simply summing the weights of the individual arborescences, for example:

$$r(\mathcal{T}) := \sum_{\tau\in\mathcal{T}} r(\tau).$$

The same notation will be used from now on for other sets: if $A \subseteq \mathcal{T}$, then $r(A) := \sum_{\tau\in A} r(\tau)$.

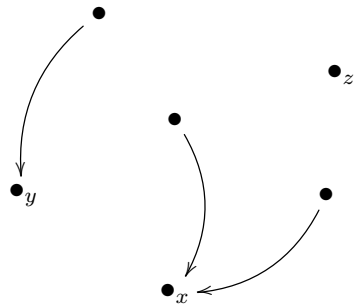


Figure 1.3: A forest with root set $\mathcal{S} = \{x, y, z\}$

Definition 6. *A **directed forest** f spanning (V, E) with root set $\mathcal{S}$ is a spanning subgraph of (V, E) that is the disjoint union of directed rooted arborescences $\{\tau_i\}_{i\in I}$, where each τ_i has a root $x_i \in V$. We denote by $\mathcal{S}$ the set of the roots of each constituent arborescence: $\mathcal{S} = \{x_i\}_{i\in I}$.*

We call $\mathcal{F}_\mathcal{S}$ the set of all spanning forests of (V, E) with root set $\mathcal{S}$, and $\mathcal{F}_\mathcal{S}(i \to j)$ the set of spanning forests of (V, E) such that $j \in \mathcal{S}$, $i \in V \setminus \mathcal{S}$ and i belongs to the arborescence rooted in j.

We shall also call $\mathcal{F}$ the set of all the spanning forests of (V, E). As before, we can assign to each forest $f \in \mathcal{F}$ a weight given by:

$$r(f) = \prod_{\tau\in f} r(\tau) = \prod_{e\in f} r(e),$$

where τ are the constituent arborescences of f, and e are the edges beloning to f.

As before, we can expand this definition to forests sets: we define for example $r(\mathcal{F}_\mathcal{S}) = \sum_{f\in\mathcal{F}_\mathcal{S}} r(f)$ and $r(\mathcal{F}_\mathcal{S}(i \to j)) = \sum_{f\in\mathcal{F}_\mathcal{S}(i\to j)} r(f)$.

Another class of subgraphs that will later be relevant is the following:

Maps and Unicyclic Components

Definition 7. *Let (V, E) be a digraph. A **map** $\mathfrak{f}$ is a spanning subgraph of (V, E) such that for each vertex $x \in V$ there exists exactly one edge $(x, y) \in \mathfrak{f}$ exiting from x.*

The name is due to the fact that a map $\mathfrak{f}$ is encoded by a function $\hat{\mathfrak{f}} : V \to V$ by considering belonging to $\mathfrak{f}$ only the $|V|$ edges $(V, (x, \hat{\mathfrak{f}}(x)_{x\in V}))$.

The last definition deals with subgraphs with only one cycle.

Definition 8. *A subgraph $\mathfrak{u}$ of a directed graph (V, E) is called an **inward oriented unicyclic component** , or simply **in-unicyclic component** , if*

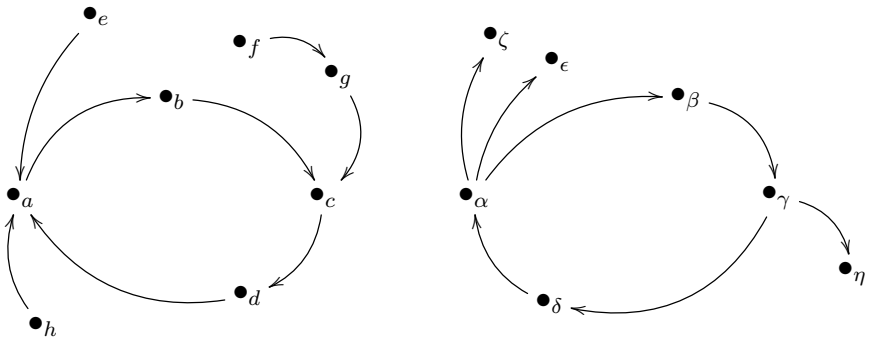


Figure 1.4: An inward oriented unicyclic component and an outward oriented unicyclic component.

- *there exists a unique directed cycle in* $\mathfrak{u}$*,*
- *for any vertex* x *belonging to* $\mathfrak{u}$ *outside the cycle there is a unique edge in* $\mathfrak{u}$ *exiting from* x *and a directed path from* x *to the unique cycle.*

A subgraph $\mathfrak{o}$ *of a digraph* (V, E) *is called an* ***outward oriented unicyclic component*** *, or simply* ***out-unicyclic component*** *, if*

- *there exists a unique directed cycle in* $\mathfrak{o}$*,*
- *for any vertex* x *belonging to* $\mathfrak{o}$ *outside the cycle there is a unique edge in* $\mathfrak{o}$ *entering in* x *and a directed path from the unique cycle to* x*.*

If a in-(out-)unicyclic component $\mathfrak{u}$ ($\mathfrak{o}$) is spanning, i.e. contains all the vertices in V, it is called an **in-(out-)unicyclic graph.** Call $\mathcal{U}^{\text{in}}$ and $\mathcal{U}^{\text{out}}$ the collection of all the in-unicyclic and out-unicyclic graphs, respectively.

On a strongly connected graph every node belongs to the cycle of at least one unicyclic graph and call $\mathcal{U}_x^{\text{in}}$ and $\mathcal{U}_x^{\text{out}}$ the set of in-unicyclic and out-unicyclic graphs with x belonging to the unique cycle.

The structure of maps is well known and we refer for example to [54] for the following basic fact: every map is the spanning disjoint union of in-unicyclic components. In particular, every in-unicyclic graph is a map.
Inverting the orientation of all the edges of a map, a graph with exactly one edge entering in each node is obtained, that is, a spanning disjoint union of out-unicyclic components. Moreover, for any spanning collections of spanning disjoint out-unicyclic components, changing the orientation of its edges it is obtained a map.

Directed Acyclic Graphs (DAG)

We conclude by focusing on a class of graphs with interesting topological properties.

Definition 9. *A digraph is acyclic, or a* ***directed acyclic graph (DAG)****, if it has no cycles.*

A DAG (V, E) induces a partial order on V by setting $x_i > x_j$ if there exists a directed path in the DAG from x_i to x_j. In particular, the following holds:

Theorem 1. *A relation is a partial order if and only if it is the directed path relation of a DAG.*

For the direct proof of this result and further properties of digraphs, see [10].

1.1.3 Distances

For a strongly connected digraph it makes sense to introduce the notion of distance between two vertices. There are various notions in the literature, and we choose to present the following definitions which will be useful during the discussion.

Geodesic Distance

Definition 10. *Let Γ_{xy} be the set of paths γ_{xy} connecting x and y. The* ***geodesic distance*** *between two vertices $x, y \in V$ is the weight of the minimal path in Γ_{xy}:*

$$d(x, y) = \min_{\gamma_{xy} \in \Gamma_{xy}} r(\gamma_{xy}).$$

Note that here we are using the notion of distance improperly, since in general $d(x, y)$ may not be symmetric. It is therefore necessary to extend the common definition of distance, according to [48].

Definition 11. *Let $\mathcal{M}$ be a nonempty set. The function $a : \mathcal{M} \times \mathcal{M} \to [0, \infty]$ is an* ***asymmetric distance function*** *if a satisfies:*

1. *$a(x, y) \geq 0$ and for each $x \in \mathcal{M}$, $a(x, x) = 0$,*
2. *$a(x, y) = a(y, x) = 0$ implies that there exists a zero-weighted path from x to y or that $x = y$,*
3. *$a(x, y) \leq a(x, z) + a(z, y)$ for each $x, y, z \in \mathcal{M}$.*

If we take $\mathcal{M} = V$, the geodesic distance d satisfies the definition of asymmetric distance function.

Resistance Distance

In network theory, *effective resistance* is defined as the total resistance of a circuit measured between two points, equivalent to the resistance of a single resistor that would produce the same voltage-current relationship. This concept allows one to reduce a complex network into a simpler, single-resistor model, taking into account the arrangement of all individual resistances, whether in series, parallel, or in more general configurations.

We first recall the theory for *undirected graphs* as developed in [39]. Further details and proofs concerning the undirected setting are provided in Appendix A. In this work, we extend the analysis to *directed graphs* (digraphs), where the orientation of edges dictates the direction of current flow.

Consider a digraph (V, E). We associate to it an electrical network in which each vertex $v \in V$ corresponds to a circuit node, and each directed edge $(x, y) \in E$ corresponds to a circuit branch containing a resistor with resistance equal to the edge weight $r(x, y)$. Current is assumed to flow consistently with the orientation of the digraph.

Effective resistance and arborescences. In this directed setting, the *effective resistance* between nodes is defined through the *preservation of the total weight of spanning arborescences*. Let $w \in V$ be a fixed root vertex. Denote by $\mathcal{T}_w$ the set of arborescences oriented toward w, and define their total weight as

$$r(\mathcal{T}_w) := \sum_{\tau \in \mathcal{T}_w} r(\tau),$$

where $r(\tau)$ is the product of the resistances along the edges of τ.

A *network reduction* is a transformation that produces a simpler network while preserving this total weight:

$$r(\mathcal{T}_w) = \hat{r}(\hat{\mathcal{T}}_w),$$

where $\hat{\mathcal{T}}_w$ denotes the set of arborescences in the reduced network, and $\hat{r}\hat{\mathcal{T}}_w$ is the corresponding reduced weight. In this sense, the reduced network is an *equivalent network*, analogous to the concept of a single effective resistor in classical circuits: it preserves the "resistance structure" between nodes in terms of arborescence weights, even in the presence of edge orientations.

Basic network transformations We now describe three basic transformations that reduce the network while preserving the total arborescence weight. These transformations correspond to familiar operations in electrical circuits:

1. **Parallel reduction:** Two edges $(x, y)_1$ and $(x, y)_2$ in parallel with resistances $r(x, y)_1$ and $r(x, y)_2$ are replaced by a single edge (x, y) of resistance $\hat{r}(x, y) =$

$r(x,y)_1 + r(x,y)_2$. Similarly for edges $(y,x)_1$ and $(y,x)_2$ in the opposite direction: $\hat{r}(y,x) = r(y,x)_1 + r(y,x)_2$.

2. **Series reduction:** Two edges (x,y) and (y,z), with y of degree 2, are replaced by a single edge (x,z) with

$$\hat{r}(x,z) = \frac{r(x,y)r(y,z)}{r(y,x) + r(y,z)},$$

and similarly in the opposite direction. This ensures that the total weight of arborescences remains invariant.

3. **Star-Delta transformation:** A degree-3 node s connected to nodes x, y, z can be removed and replaced by a complete subgraph among $\{x, y, z\}$, with reduced resistances computed to preserve arborescence weights. See Appendix A for details in the undirected case; the directed case preserves orientations in the same manner.

All this problems are currently polynomially solvable by means of determinant methods. Counting spanning trees can be solved efficiently for general graphs using the determinental technique outlined first by Kirchoff [38]. Spanning trees of an n-vertex graph are counted by computing the determinant of an $(n-1) \times (n-1)$ matrix with integer entries.

Mapping arborescences. For each transformation, we define a map $\phi : \mathcal{T}_w \longrightarrow \hat{\mathcal{T}}_w$ that associates each original arborescence to an arborescence in the reduced network. This mapping ensures that the external structure of arborescences is preserved and that $r(\mathcal{T}_w) = \hat{r}(\hat{\mathcal{T}}_w)$, providing a rigorous justification for the computed reduced resistances.

1. If the transformation is a parallel reduction on two nodes $x, y \in V$, ϕ associates

   - $\phi : \mathcal{T}_\omega^1 \longmapsto \hat{\mathcal{T}}_\omega^1$, where $\mathcal{T}_\omega^1 \subset \mathcal{T}_\omega$ is the set of arborescences containing one edge between $(x,y)_1$ and $(x,y)_2$, and $\hat{\mathcal{T}}_w^1 \subset \hat{\mathcal{T}}_w$ is the set of arborescences containing the edge (x,y). Let Σ_1 denote the spanning forests of (V,E) obtained by deleting the edges $(x,y)_1$ or $(x,y)_2$ from arborescences in $\mathcal{T}_\omega^1$, and observe that $\mathcal{T}_\omega^1 = (\Sigma_1 \cup (x,y)_1) \cup (\Sigma_1 \cup (x,y)_2)$ and $\hat{\mathcal{T}}_\omega^1 = \Sigma_1 \cup (x,y)$.
   - $\phi : \mathcal{T}_\omega^2 \longmapsto \hat{\mathcal{T}}_\omega^2$, where $\mathcal{T}_\omega^2 \subset \mathcal{T}_\omega$ denotes the set of arborescences containing one edge between $(y,x)_1$ and $(y,x)_2$, and $\hat{\mathcal{T}}_w^1 \subset \hat{\mathcal{T}}_w$ is the set of arborescences containing the edge (y,x). Let Σ_2 denote the spanning forests obtained by deleting the edges $(y,x)_1$ or $(y,x)_2$ from arborescences in $\mathcal{T}_\omega^2$. Again we notice that $\mathcal{T}_\omega^2 = (\Sigma_2 \cup (y,x)_1) \cup (\Sigma_2 \cup (y,x)_2)$ and $\hat{\mathcal{T}}_\omega^2 = \Sigma_2 \cup (y,x)$.

- $\phi : \mathcal{T}_\omega^3 \longmapsto \hat{\mathcal{T}}_\omega^3$, where $\mathcal{T}_\omega^3$ is the set of arborescences in $\mathcal{T}_w$ that contain no edges between x and y and $\hat{\mathcal{T}}_\omega^3$ is the set of arborescences in $\hat{\mathcal{T}}_w$ with no edges between x, y. Let Σ_3 denote those spanning arborescences.

The arborescences share the same external structure Σ_i for $i = 1, \ldots, 3$. If we impose that $r(\mathcal{T}_w) = \hat{r}(\hat{\mathcal{T}}_w)$, then we have, respectively:

$$\begin{aligned} r(\Sigma_1)(r(x,y)_1 + r(x,y)_2) &= r(\Sigma_1)(\hat{r}(x,y)) \\ r(\Sigma_2)(r(y,x)_1 + r(y,x)_2) &= r(\Sigma_2)(\hat{r}(y,x)) \\ r(\Sigma_3) &= r(\Sigma_3). \end{aligned}$$

As a consequence

$$\hat{r}(x,y) = r(x,y)_1 + r(x,y)_2 \text{ and } \hat{r}(y,x) = r(y,x)_1 + r(y,x)_2.$$

2. If the transformation is a series reduction, ϕ associates:

- $\phi : \mathcal{T}_\omega^1 \longmapsto \hat{\mathcal{T}}_\omega^1$, where $\mathcal{T}_\omega^1$ is the set of arborescences in $\mathcal{T}_\omega$ containing both the edges (x, y) and (y, z) and $\hat{\mathcal{T}}_\omega^1$ is the set of arborescences in $\hat{\mathcal{T}}_\omega$ containing the edge (x, z). Let Σ_1 denote the spanning forests of (V, E) obtained by deleting the edges (x, y) and (y, z) from arborescences of this type, and observe that $\mathcal{T}_\omega^1 = \Sigma_1 \cup ((x, y) \cup (y, z))$ and $\hat{\mathcal{T}}_\omega^1 = \Sigma_1 \cup (x, z)$.
- $\phi : \mathcal{T}_\omega^2 \longmapsto \hat{\mathcal{T}}_\omega^2$, where $\mathcal{T}_\omega^2$ is the set of arborescences in $\mathcal{T}_\omega$ containing both the edges (z, y) and (y, x) and $\hat{\mathcal{T}}_\omega^2$ is the set of arborescences in $\hat{\mathcal{T}}_\omega$ containing the edge (z, x). Let Σ_2 denote the spanning forests of (V, E) obtained by deleting the edges (z, y) and (y, x) from arborescences of this type, and observe that $\mathcal{T}_\omega^2 = \Sigma_2 \cup ((z, y) \cup (y, x))$ and $\hat{\mathcal{T}}_\omega^2 = \Sigma_2 \cup (z, x)$.
- $\phi : \mathcal{T}_\omega^3 \longmapsto \hat{\mathcal{T}}_\omega^3$, where $\mathcal{T}_\omega^3$ is the set of arborescences in $\mathcal{T}_w$ that contain one edge between (y, x) and (y, z) and $\hat{\mathcal{T}}_\omega^3$ is the set of arborescences in $\hat{\mathcal{T}}_w$ with no edges between x, y. Let Σ_3 denote the spanning forests of (V, E) obtained by deleting the edges (y, x) and (y, z) from arborescences of this type. We notice that $\mathcal{T}_\omega^3 = (\Sigma_3 \cup (y, x)) \cup (\Sigma_3 \cup (y, z))$ and $\hat{\mathcal{T}}_\omega^2 = \Sigma_3$.

The arborescences share the same external structure Σ_i for $i = 1, \ldots, 3$. If we impose that $\lambda\, r(\mathcal{T}_w) = \hat{r}(\hat{\mathcal{T}}_w)$, then we have, respectively:

$$\begin{aligned} \lambda r(\Sigma_1) r(x,y) r(y,z) &= r(\Sigma_1)(\hat{r}(x,z)) \\ \lambda r(\Sigma_2) r(z,y) r(y,x) &= r(\Sigma_2)(\hat{r}(z,x)) \\ \lambda r(\Sigma_3)(r(y,x) + r(y,z)) &= r(\Sigma_3). \end{aligned}$$

As a consequence $\lambda = \frac{1}{r(y,x)+r(y,z)}$ and so

$$\hat{r}(x,y) = \frac{r(x,y)r(y,z)}{r(y,x)+r(y,z)} \quad \text{and} \quad \hat{r}(y,x) = \frac{r(z,y)r(y,x)}{r(y,x)+r(y,z)}$$

3. For the directed Star-Delta transformation, we address the reader to section 5.8, in which we analyze a more general transformation.

1.1.4 Transition Graphs of Continuous-Time Markov Chains

There exists a natural one-to-one correspondence between finite weighted digraphs and continuous-time Markov chains, which we present below.

Let (V, E) be a finite and strongly connected weighted digraph, to which we associated with each edge $(x, y) \in E$ a positive weight $r(x, y) \geq 0$.

Using the the **backward generator** :

$$\mathcal{L}_{x,y} := \begin{cases} \mathcal{L}_{xy} = r(x,y) & x \neq y, \\ \mathcal{L}_{xx} = -\sum_{y \in V} r(x,y) & x = y, \end{cases}$$

(V, E) becomes the transition graph of a Markov chain $X := (X_t)_{t \geq 0}$ with state space V and $t \in \mathbb{R}^+$, where the transition rates between a pair of states x and y correspond to the weight $r(x, y)$.

Since (V, E) is strongly connected, the Markov chain X is irreducible: recall, indeed, that a Markov chain is said to be irreducible when all states communicate with each other.

A classical result on Markov chains[1] tells us that if a Markov chain is irreducible, then it admits a unique stationary distribution π, which is the one that satisfies the balance equation with all elements positive

$$\pi(x) \sum_{y:(x,y)\in E} r(x,y) = \sum_{y:(y,x)\in E} \pi(y) r(y,x). \tag{1.1}$$

These are the equations characterizing the stationary distribution of an irreducible chain, but in the course of this discussion we are also interested in a combinatorial representation of the same, which we briefly present in the next section.

1.2 Markov Chain Tree Theorem

A very useful representation for the invariant measure of a finite irreducible Markov chain is given by the Markov Chain Tree Theorem.[2]

[1] For a comprehensive discussion, see [51]

[2] See [5, 27, 55]

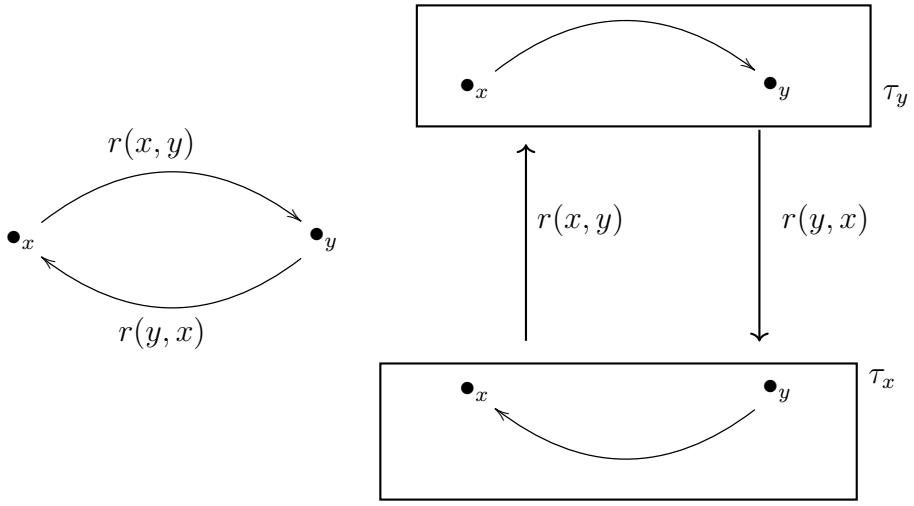


Figure 1.5: A graph (V, E) and its corresponding litfing graph $(\mathcal{T}, \hat{E})$

Theorem 2 (Markov Chain Tree Theorem)**.** *The invariant measure of an irreducible Markov chain with strongly connected transition graph (V, E) is given by*

$$\pi(x) = \frac{r(\mathcal{T}_x)}{r(\mathcal{T})} \tag{1.2}$$

for all $x \in V$.

A simple and elegant proof is provided in [5] by constructing a Markov chain in an enlarged state space, such that its projection is the original Markov chain on (V, E) with rates r. In particular, we consider the directed graph $(\mathcal{T}, \hat{E})$. The set of vertices coincides with the set of arborescences and $(\tau, \tau') \in \hat{E}$ if and only if $\tau \in \mathcal{T}_x$, $\tau' \in \mathcal{T}_y$, $(x, y) \in E$ and the arborescence τ' is obtained by $\tau' = \big[\tau \cup (x, y)\big] \setminus (y, z)$ where (y, z) is the unique edge exiting form y in τ. The rates $\hat{r}$ of $(\tau, \tau') \in \hat{E}$ with $\tau \in \mathcal{T}_x$ and $\tau' \in \mathcal{T}_y$ are given by $\hat{r}(\tau, \tau') := r(x, y)$. We call $\hat{X}$ the Markov chain associated with such rates. There is a natural projection map $\Pi : (\mathcal{T}, \hat{E}) \to (V, E)$ such that, for any $\tau \in \mathcal{T}_x$, $\Pi(\tau) = x$ and, for any $(\tau, \tau') \in \hat{E}$ with $\tau \in \mathcal{T}_x$ and $\tau' \in \mathcal{T}_y$, $\Pi(\tau, \tau') = (x, y)$.

Next result is from [5].

Theorem 3. *The Markov chain $\hat{X}$ is irreducible and has a unique invariant measure given by*

$$\hat{\mu}(\tau) = \frac{1}{\mathcal{Z}} r(\tau)\,, \tag{1.3}$$

where $\mathcal{Z} = r(\mathcal{T})$. The process $X := \Pi\left(\hat{X}\right)$, is a Markov chain on (V, E) with transition rates r and we deduce (1.2) *as a consequence.*

Proof. We give a sketch of the proof. The irreducibility of the Markov chain $\hat{X}$ is discussed in [5]. The graph $(\mathcal{T}, \hat{E})$ is Eulerian, since for each node τ there is a bijection among the outgoing and ingoing edges. Indeed, for each $\tau \in \mathcal{T}_x$, the outgoing edges in $(\mathcal{T}, \hat{E})$ are in correspondence with the outgoing edges from x in (V, E). Let $\tau \in \mathcal{T}_x$ and $(\tau, \tau') \in \hat{E}$ with $\tau' \in \mathcal{T}_y$. We have that $\tau \cup (x, y) = \mathfrak{u}_x \in \mathcal{U}_x^{\text{in}}$ so that there exists $\tau'' \in \mathcal{T}_z$ such that $\tau'' \cup (z, x) = \tau \cup (x, y) = \mathfrak{u}_x$. This correspondence $(\tau, \tau') \leftrightarrow (\tau'', \tau)$ gives the bijection among exiting and entering edges at τ, i.e. a bijection between exiting and entering elements of $\hat{E}$ on each $\tau \in \mathcal{T}$ (see [14] section 2.6 for more details). Moreover by (1.3) we have, up to a common multiplicative constant,

$$\hat{\mu}(\tau) r(x, y) = \hat{\mu}(\tau'') r(z, x) = r(\mathfrak{u}_x)\,. \tag{1.4}$$

Using the fact that the graph $(\mathcal{T}, \hat{E})$ is Eulerian, the stationarity of (1.3) follows. The last statements of the Theorem are direct consequences of the definitions. □

1.3 Harmonic Tree Formula

Consider now an irreducible continuous time Markov chain $(X_t)_{t \geq 0}$ with transition graph (V, E). Given a set $\mathcal{S} \subset V$, we denote by $P_N^{\mathcal{S}}(y|x)$ the probability that, starting from the point $x \in V \setminus \mathcal{S}$, the process enters the set $\mathcal{S}$ for the first time at the point $y \in \mathcal{S}$:

$$P^{\mathcal{S}}(y|x) := \mathbb{P}(X_{\bar{t}} = y | X_0 = x) \qquad \text{where } \bar{t} := \min\{t > 0 | X_t \in \mathcal{S}\},$$

where both $P^{\mathcal{S}}(\cdot)$ and $\mathbb{P}(\cdot)$ refer to X_t. Then we have the following representation:

Theorem 4 (Harmonic Tree formula)**.**

$$P_N^{\mathcal{S}}(y|x) = \frac{r(\mathcal{F}_{\mathcal{S}}(x \to y))}{r(\mathcal{F}_{\mathcal{S}})} \tag{1.5}$$

for $x \in V$ and $y \in \mathcal{S}$.

For a graphical proof of this result, see [55].

1.4 Ergodic Theorem

Finally, to conclude this survey of known results that will be used in the following chapters, we briefly recall the Ergodic Theorem for continuous-time Markov chains. The literature on this topic is extensive; therefore, we shall not dwell on the details of the proof here either, and instead refer the reader to [9, 51].

Theorem 5. *Let X_t be an irreducible, positive recurrent continuous-time Markov chain with state space V. Let π be its unique invariant measure. Then, for any bounded function $f : V \to \mathbb{R}$ and for any initial distribution*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s)ds = \sum_{x\in V} f(x)\pi(x) \qquad \textit{almost surely.} \tag{1.6}$$

In particular, we have that, almost everywhere

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{I}(X_s = x)ds = \mathbb{E}_\pi[\mathbb{I}(X_s = x)] = \pi(x), \tag{1.7}$$

where $\mathbb{E}_\pi$ denotes the expectation with respect to the invariant probability measure π, that is, the expectation of the process when initialized according to π.

Part I

Freidlin-Wentzell Theory on Graphs

2

Continuous Diffusions

Our analysis is rooted in the fundamental duality that arises in statistical mechanics between discrete and continuous systems. This dichotomy not only corresponds to distinct mathematical frameworks but also reflects different physical interpretations and modeling strategies. In particular, the well-established body of results in the context of continuous diffusions provides strong motivation to explore analogous structures and behaviors in the discrete setting.

In this section, we provide a brief overview of fundamental results concerning continuous diffusions, focusing particularly on Freidlin–Wentzell theory and weak KAM theory. We then present our contributions in the discrete case, which exhibit a strong parallel with their continuous counterparts.

2.1 Freidlin-Wentzell Theory

Before proceeding, let us briefly recall the concept of the Large Deviations Principle. For a more in-depth analysis, see [20].

Definition 12 (Rate Function)**.** *Let X be a Hausdorff topological space. An extended real-valued function $I : X \to [0, \infty]$ is called a* rate function *if it is lower semicontinuous; that is, for every $c \in [0, \infty]$, the level set $\{x \in X \mid I(x) \leq c\}$ is closed in X.*

Moreover, if all such level sets are compact, then I is called a good *rate function.*

Definition 13 (Large Deviation Principle)**.** *Let $n \geq 0$, and let $\{\mu_n\}_{n\in\mathbb{N}}$ be a family of probability measures on X. We say that μ_n satisfies the* Large Deviation Principle

(LDP) *with rate function* $I : X \to [0, \infty]$ *if*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} I(x) \quad \textit{for every closed set } F \subseteq X,$$
$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} I(x) \quad \textit{for every open set } G \subseteq X.$$

If the upper bound holds only for compact (rather than closed) sets F*, then the sequence* μ_n *is said to satisfy a* weak *LDP.*

2.1.1 Small-Noise Limit of a Diffusion Process

Consider a standard Brownian motion W_t in $\mathbb{R}^d$ starting at the origin, and a diffusion process X_t^ϵ taking values in $\mathbb{R}^d$, which is the unique strong solution to the stochastic differential equation:

$$dX_t^\epsilon = b(X_t^\epsilon)dt + \sqrt{\epsilon}dW_t;$$

with $t \geq 0$ and $X_0^\epsilon = 0$, where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a uniformly Lipschitz continuous vector field. Let $C_0 = C_0([0, T]; \mathbb{R}^d)$ be the Banach space of continuous paths starting at the origin, endowed with the supremum norm $||\cdot||_\infty$. Then the family of probability measures induced by $\{X^\epsilon\}_{\epsilon\geq 0}$ on C_0 satisfies the large deviations principle with a good rate function $I : C_0 \to \mathbb{R} \cup \{\infty\}$ given by the **Freidlin-Wentzell action functional** :

$$I_{0,T}(f) = \begin{cases} \frac{1}{2}\int_0^T |\dot{f}(t) - b(f(t))|^2 dt & \text{if } f \text{ lies in the Sobolev space } H^1([0, T]; \mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases}$$

The proof of these statements is well known, and for a comprehensive overview of the topic, we refer the reader to [27] and [19].

Freidlin-Wentzell Quasipotential

In this way, the principal term of the logarithmic asymptotics of probabilities of events concerning the process X_t^ϵ are reduced to the solution of variational problems, namely the minimization of the action functional over suitable classes of trajectories. In particular, for fixed $x, y \in \mathbb{R}^d$ and $t > 0$, we define the finite-time action by

$$V(t, x, y) = \min\{I_{0,t}(f) : f \in H^1([0, t]; \mathbb{R}^d), f(0) = x, f(t) = y\}.$$

Definition 14. *A point $0 \in \mathbb{R}^r$ is called an **asymptotically stable equilibrium point** of the deterministic dynamical system $\dot{x} = b(x)$ if $b(0) = 0$ and there exists a neighborhood U of 0 such that, for every $x_0 \in U$, the solution $x(t)$ of the system with initial condition $x(0) = x_0$ is defined for all $t \geq 0$ and satisfies $\lim_{t\to\infty} x(t) = 0$.*

Definition 15. *Assume that 0 is an asymptotically stable equilibrium point of the deterministic system $\dot{x} = b(x)$. The quasipotential with respect to 0 is defined by*

$$V(0,x) = \inf\left\{I_{0,T}(f) : T > 0,\ f \in H^1([0,T];\mathbb{R}^r),\ f(0) = 0,\ f(T) = x\right\}.$$

Invariant Measure

The asymptotic behavior of the invariant measure of the diffusion process X_t^ϵ as $\epsilon \to 0$ can also be described in terms of the quasipotential $V(0,x)$.

In order to guarantee the existence of an invariant measure for the perturbed diffusion process, we impose suitable conditions on the behavior of the vectorfield $b(x)$ at infinity.

Assumption 1. *We assume that outside a sufficiently large ball with center at the origin, the projection of $b(x)$ onto the position vector $r(x)$ of the point x is negative and separated from zero, that is, there exists a large number N such that $(b(x), r(x)) < -1/N$ for $|x| > N$.*

Theorem 6. *Let the point 0 be the unique stable equilibrium position of the non-perturbed system, let the space $\mathbb{R}^r$ be attracted to 0, and let Assumption 1 hold.*

Then the process X_t^ϵ admits an invariant measure μ^ϵ for every $\epsilon > 0$ and we have that

$$\lim_{\epsilon\to 0} \epsilon \log \mu^\epsilon(D) = -\left(\inf_{x\in D} V(0,x) - \inf_{y\in\mathbb{R}^r} V(0,y)\right), \tag{2.1}$$

for any domain $D \subset \mathbb{R}^r$ with compact boundary, where $V(0,x)$ is the quasipotential of $b(x)$ with respect to 0.

For an outline of the proof of this theorem, see Chapter 4 of [29].

2.2 Weak KAM Theory

In this section, we provide a brief introduction to the Weak KAM Theory, as it will be needed in later chapters, particularly in the study of action-minimizing solutions.

Weak KAM theory is a variational framework that connects viscosity solutions of the Hamilton–Jacobi equation with the dynamics of Lagrangian systems. The acronym "KAM" refers to the Kolmogorov–Arnold–Moser theorem, which deals

with the stability of quasi-periodic orbits in Hamiltonian systems. While the classical KAM theory requires strong regularity and non-degeneracy conditions, the "weak" version applies in broader settings and plays a fundamental role in the study of low-regularity or non-integrable systems.

Our aim is to give a concise overview of the essential statements and properties without entering into the technical details of the proofs. The complete development of the theory, including rigorous formulations and demonstrations, can be found in [26], [23], [50]. Therefore, we focus here solely on stating the results.

Let M be a compact C^∞ manifold without boundary. Denote by TM its tangent bundle, by T^*M its cotangent bundle, and by $\pi : TM \to M$ its canonical projection. A point $(x, v) \in TM$ is identified by $x \in M$ and $v \in T_xM = \pi^{-1}(x)$. With this notation, we have $\pi(x, v) = x$. The cotangent bundle is $\pi^* : T^*M \to M$. A point of T^*M is denoted by (x, p), where $x \in M$, and $p \in T_x^*M = L(T_xM \to \mathbb{R})$.

Definition 16 (Lagrangian)**.** *A continuous function $L : TM \to \mathbb{R}$ on the manifold M is called **Lagrangian.***

As a typical case of L, we can think of $L(x, v) = \frac{1}{2}g_x(v, v)$ where g is a Riemannian metric on M. The metric g allows one to identify the differential of a function with a vector field, thus yielding a well-defined notion of the gradient ∇ on M. For the definition of differential, local partial derivates and gradients over functions defined on manifolds, we address the reader to [22] and [21].

A *non-degenerate* Lagrangian L on the manifold M is a C^2 Lagrangian such that, for each $(x, v) \in TM$, the second partial derivative $\frac{\partial^2 L}{\partial v^2(x,v)}$ is a non degenerate quadratic form at each point of TM.

Definition 17 (Hamiltonian)**.** *If L is a non-degenerate Lagrangian on the manifold M, its corresponding **Hamiltonian** $H : T^*M \to \mathbb{R}$ is the function defined by*

$$H(x, p) = \sup_{x \in M} \{< p, v > -L(x, v)\},$$

where $< p, v >$ denotes the value of $v \in TM$ under the linear form $p \in T^*M$ and $p := \frac{\partial L}{\partial v}$.

Viscosity Subsolutions of the Hamilton-Jacobi Problem

Before proceeding, let us consider the Hamilton–Jacobi equation

$$\frac{\partial U}{\partial t}(t, x) + H(x, \nabla U(t, x)) = 0, \tag{2.2}$$

where $H : T^*M \to \mathbb{R}$ is the Hamiltonian associated to a non-degenerate Lagrangian $L : TM \to \mathbb{R}$. The unknown is a continuous function $U : [0, +\infty) \times M \to \mathbb{R}$, and ∇U represents the gradient of U at x.

Now let $g : M \to \mathbb{R}$ be a continuous function, and let us impose the following boundary conditions for all $x \in M$:

$$U(0, x) = g(x). \tag{2.3}$$

Definition 18 (Viscosity Solution)**.** *Let $U : [0, +\infty) \times M \to \mathbb{R}$ be a bounded and uniformly continuous function on $[0, +\infty) \times M$. We say that U is a **viscosity solution** of the Hamilton-Jacobi equation (2.2) with initial value (2.3) if for each function $\phi \in C^\infty((0, \infty) \times M)$,*

1. *U is a **viscosity subsolution:** if $U - \phi$ has a local maximum at a point $(t^*, x^*) \in (o, \infty) \times M$, then*

$$\frac{d\phi}{dt}(t^*, x^*) + H\left(x^*, \nabla\phi(x^*, t^*)\right) \leq 0,$$

and

2. *U is a **viscosity supersolution:** if $U - \phi$ has a local minimum at a point $(t^*, x^*) \in (o, \infty) \times M$, then*

$$\frac{d\phi}{dt}(t^*, x^*) + H\left(x^*, \nabla\phi(x^*, t^*)\right) \geq 0.$$

For a detailed discussion of the consistency and uniqueness of these solutions, together with their physical interpretation, see [24].

Tonelli Lagrangian and Dominated Curves

Definition 19 (Superlinear above Compact Subsets)**.** *We say that a Lagrangian $L : TM \to \mathbb{R}$ is **superlinear above compact subsets** if for every compact subset $K \subset \mathbb{R}$, and every $k \geq 0$, we can find a constant $c(K, k) > -\infty$ such that for all $(x, v) \in TM$ and $x \in K$,*

$$L(x, v) \geq k\|v\|_x + c(K, k),$$

where $\|\cdot\|_x$ is a fixed continuous norm on the vector bundle TM.

Definition 20 (Tonelli Lagrangian)**.** *A Lagrangian L on the manifold M is called a **Tonelli Lagrangian** if it satisfies the following conditions:*

1. *$L : TM \to \mathbb{R}$ is of class at least C^2.*
2. *For each $(x, v) \in TM$, the second partial derivative $\frac{\partial^2 L}{\partial v^2(x,v)}$ is positive definite as a quadratic form.*

3. L is superlinear above compact subsets of M.

From now on we will assume that we have a Tonelli Lagrangian L of class $C^r, r \geq 2$, on the manifold M. The corrisponding Hamiltonian H will be of class C^{r-1}.

Definition 21 (Dominated Curve)**.** *Let $u : U \to \mathbb{R}$ be a function defined on the open subset $U \subset M$. If $c \in \mathbb{R}$, we say that u is **dominated** by $L + c$ on U, which we denote by $u \prec L + c$, if for each continuous piecewise C^1 curve $\gamma : [a, b] \to U$ we have*

$$u(\gamma(b)) - u(\gamma(a)) \leq \int_a^b L(\gamma(s), \dot{\gamma}(s))ds + c(b - a).$$

Definition 22 ((u, L, c)-calibrated Curve)**.** *Let $u : U \to \mathbb{R}$ be a function and let $c \in \mathbb{R}$ be a constant, where U is an open subset of M. We say that the (continuous) piecewise C^1 curve $\gamma : I \to U$, defined on the interval $I \subset \mathbb{R}$ is (u, L, c)-calibrated, if for every $t \leq t^{'}$, we have*

$$u(\gamma(t^{'})) - u(\gamma(t)) = \int_t^{t^{'}} L(\gamma(s), \dot{\gamma}(s))ds + c(t^{'} - t).$$

Weak KAM Solutions

Definition 23 (Weak KAM solution)**.** *Let L be a Tonelli Lagrangian on the compact manifold M. A **weak KAM solution** of negative type (of a positive type) is a function $u : M \to \mathbb{R}$ for which there exists $c \in \mathbb{R}$ such that:*

1. *The function u is dominated by $L + c$;*
2. *For every $x \in M$ we can find a (u, L, c)-calibrated C^1 curve $\gamma : (-\infty, 0] \to M$ (resp. $\gamma : [0, +\infty) \to M$) with $\gamma(0) = x$.*

Definition 24 (Mañè's Critical Value)**.** *If L is a Tonelli Lagrangian on the connected compact manifold M, the Manè critical value of L is the constant $c[0]$ (or $c_L[0]$, if we need to precise the Lagrangian) defined by*

$$c[0] = \inf\{c \in \mathbb{R} | \exists u : M \to \mathbb{R}, u \prec L + c\}.$$

It can be proved [26] that $c[0] \geq -\inf_{x \in M} L(x, 0)$ and thus $c[0]$ of L is finite.

Definition 25 (Minimal action for a given time)**.** *If L is a Tonelli Lagrangian on the compact connected manifold M, for $t > 0$ fixed, we define the function $\mathcal{A}_t : M \times M \to \mathbb{R}$ by*

$$\mathcal{A}_t(x,y) = \inf_{\gamma} \int_0^t L(\gamma(s), \dot{\gamma}(s))ds, \tag{2.4}$$

where the infimum is taken over all the continuous piecewise C^1 curves $\gamma : [0,t] \to M$ with $\gamma(0) = x$ and $\gamma(t) = y$. This function is called the ***minimal action*** *to go from x to y in time t.*

Lax-Oleinik Semigroup

We call $\mathcal{F}(M, [-\infty, +\infty])$ the set of arbitrary functions from the manifold M to the set $[-\infty, +\infty]$ of extended real numbers. We now introduce a semi-group of non-linear operators $(\mathcal{V}_t)_{t\geq 0}$ from $\mathcal{F}(M, [-\infty, +\infty])$ into itself.

Definition 26 (Lax-Oleinik semigroup)**.** *Fix $u \in \mathcal{F}(M, [-\infty, +\infty])$ and $t > 0$. The function $\mathcal{V}_t u : M \to [-\infty, +\infty]$ is defined at $x \in M$ as*

$$\mathcal{V}_t u(x) = \inf_{\gamma} \{u(\gamma(0)) + \int_0^t L(\gamma(s), \dot{\gamma(s)})ds\}, \tag{2.5}$$

where the infimum is taken over all the absolutely continuous curves $\gamma : [0,t] \to M$ such that $\gamma(t) = x$.

By the definition of the minimal action (2.4), for $t > 0$, we have

$$\mathcal{V}_t u(x) = \inf_{y \in M} \{u(y) + \mathcal{A}_t(x,y)\}.$$

We will also set $\mathcal{V}_t u = u$. We call the family of maps $\mathcal{V}_t : \mathcal{F}(M, [-\infty, +\infty]) \to \mathcal{F}(M, [-\infty, +\infty])$, $t \in [0, +\infty)$ the **Lax-Oleinik semigroup** .

Theorem 7. *Suppose $L : TM \to \mathbb{R}$ is a Tonelli Lagrangian on the compact manifold M. Let $\mathcal{V}_t$ be the associated Lax-Oleinik semigroup. Let $u \in \mathcal{C}^0(M, \mathbb{R})$, then the continuous function $U : [0, +\infty) \times M \to \mathbb{R}$ defined by $U(t,x) = \mathcal{V}_t u(x)$ is a viscosity solution of*

$$\frac{\partial U}{\partial t}(t,x) + H(x, \nabla U(t,x)) = 0$$

on the open set $(0, +\infty) \times M$, where $H : T^*M \to \mathbb{R}$ is the Hamiltonian associated to L.

Let us now consider the time-independent Hamilton-Jacobi equation $H(x, d_x u) = c$ for some fixed $c \in \mathbb{R}$, where the unknown is a C^1 function $u : M \to \mathbb{R}$ and $d_x u$ is

its differential. We call $\underline{\mathcal{W}}$ the set of viscosity subsolution of such a problem. For $c \geq c[0]$, we define

$$S^c(x, y) = \sup_{u \in \underline{\mathcal{W}}} u(y) - u(x).$$

By using the Perron method (see [26] theorem 8.2.2), it follows that for each $x \in M$ the function $S^c(x, \cdot)$ is a viscosity subsolution of $H(y, d_y u) = c$ on M itself, and a viscosity solution on $M \setminus \{x\}$.

Definition 27 (Projected Aubry set)**.** *Let $H : T^*M \to \mathbb{R}$ is a continuous Hamiltonian. We define the projected Aury set C as the set of $x \in M$ such that $S^{c[0]}(x, \cdot)$ is a viscosity solution of $H(z, d_z u) = c[0]$.*

Theorem 8. *Any viscosity solution $u : M \to \mathbb{R}$ for $H(x, d_x u) = c[0]$ satisfies*

$$u(x) = \inf_{x_0 \in C} u(x_0) + S^{c[0]}(x_0, x) \qquad \forall x \in M.$$

3

Freidlin-Wentzell Solutions of Discrete Hamilton Jacobi Equations

The Freidlin-Wentzell theory has a basic structure that has been extended to generalized frameworks, both infinite-dimensional and finite-dimensional. In the discrete case of continuous-time Markov chains, the analogous framework is that of Markov chains with exponentially small transition rates, which are widely used and considered in statistical mechanics and metastability (see, for example, [52, 53, 58–60]). For Markov chains of this type, we have a large deviation principle for the invariant measure.

In this chapter, we present in detail the results obtained in [4]. In particular, within the framework of large deviations theory, we introduce a stationary discrete Hamilton-Jacobi equation on a finite directed weighted graph, which naturally arises from variational problems associated with the large deviation rate functional of Markov chains with exponentially small transition rates.

In the case of diffusion processes there is a rich and not trivial relation among large deviations and Hamilton-Jacobi equations, as illustrated for example in the references [25, 30, 32, 34], where non uniqueness, non differentiability, phase transitions and selection principles are intertwined. The main aim is to illustrate and study all these issues in the simplified discrete setting. In the discrete setting we can for example obtain a complete geometric characterization of viscosity solutions, that was instead difficult for example in reference [25], with the aim of giving a geometric characterization of the selection principle for the vanishing viscosity solution. Moreover the discrete version of the theory is interesting in itself with strong connections with other mathematical problems on graphs and challenging extensions as for example to the case of metric graphs and related stochastic processes on them.

We provide a detailed analysis of the structure of the viscosity solutions and

identify a special solution obtained as the limit of the unique invariant measure. The discrete theory parallels the weak KAM theory and Freidlin-Wentzell theory in the continuous setting. The discrete framework allows for a clearer view of formulas, constructions, and characterizations.

3.1 The Discrete Hamilton-Jacobi Equation

Let $\mathcal{G} = (V, E)$ be a finite, strongly connected directed graph. Assume there are no loops, i.e., no edges of the form (x, x). Consider a sequence of irreducible Markov chains with transition probabilities depending on a parameter N. Denote by $(r_N(x, y))_{(x,y)\in E}$ the rates of jumps from $x \in V$ to $y \in V$ with $(x, y) \in E$, and define $r_N(x) := \sum_{y:(x,y)\in E} r_N(x, y)$ as the total rate of jumps from $x \in V$.

Under these assumptions, the invariant measure is unique ([51]) Furthermore, the stationarity condition characterizing the invariant measure π_N is given by:

$$\pi_N(x) \sum_{y:(x,y)\in E} r_N(x, y) = \sum_{y:(x,y)\in E} \pi_N(y) r_N(y, x), \quad \forall x \in V. \tag{3.1}$$

We consider the case where the rates are exponentially small. This is a classic framework used for example in [27] to describe the effective fluctuations of diffusion processes and describe the behavior at very low temperatures of particle systems like a Glauber or an Ising-like dynamics. It is also a a general model of a multiscale Markov chain in a regime on which there are dynamic and static large deviations. Specifically, we assume that the following condition holds throughout the rest of the section.

Assumption 2. *([60]) The rates of jump $(r_N(x, y))_{(x,y)\in E}$ are exponentially small with N, namely the following limits exist and are finite*

$$\Delta(x, y) := -\lim_{N\to+\infty} N^{-1} \log r_N(x, y) \geq 0, \quad \forall x \in V.$$

Define a sequence of functions W_N by the relation $\pi_N(x) := e^{-NW_N(x)}$, then equation (3.1) uniquely defines $(W_N)_{N\in\mathbb{N}}$. The study of large deviations of $(\pi_N)_{N\in\mathbb{N}}$ is related to the study of the limit of the sequence $(W_N)_{N\in\mathbb{N}}$. The following lemma shows that the limit can be characterized as a solution of a discrete equation.

Lemma 1 (HJ equation)**.** *Consider the sequence of probability measures $(\mu_N)_{N\in\mathbb{N}}$ solutions to* (3.1) *and assume that there exist the limits*

$$W(x) := -\lim_{N\to+\infty} N^{-1} \log \mu_N(x), \quad \forall x \in V.$$

Then W satisfies the following set of equations

$$W(x) + \min_{y:(x,y)\in E} \{\Delta(x, y)\} = \min_{y:(x,y)\in E} \{W(y) + \Delta(y, x)\}, \quad \forall x \in V. \tag{3.2}$$

Proof. Since in (3.1) there are finite sums, considering $\lim_{N\to+\infty} \frac{1}{N}\log(\cdot)$ on both sides, using Assumption 2, we obtain the result by the following basic fact. Given a finite collection $\left(a_N^i\right)$, with $i = 1, \dots k$, of sequences such that there exist the limits $a^i := \lim_{N\to+\infty} \frac{\log a_N^i}{N}$. Then we have

$$\lim_{N\to+\infty} \frac{1}{N} \log \left(\sum_{i=1}^{k} a_N^i \right) = \max_i a^i \,.$$

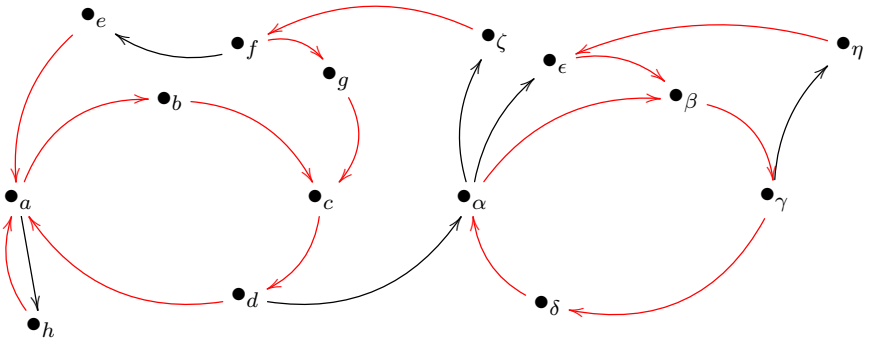
The proof is elementary and this fact is sometimes called the basic principle of large deviations ("the largest exponent wins", see for example [20] formulas I.1, I.2) equation (3.2) holds. □

This is precisely what occurs in the continuous case of diffusions, where the HJ equation is derived as the vanishing noise limit of the stationary equation, see [29]. In line with the classical FW theory for diffusions, we refer to the limiting equation (3.2) as a *discrete Hamilton–Jacobi equation* (discrete HJ equation).

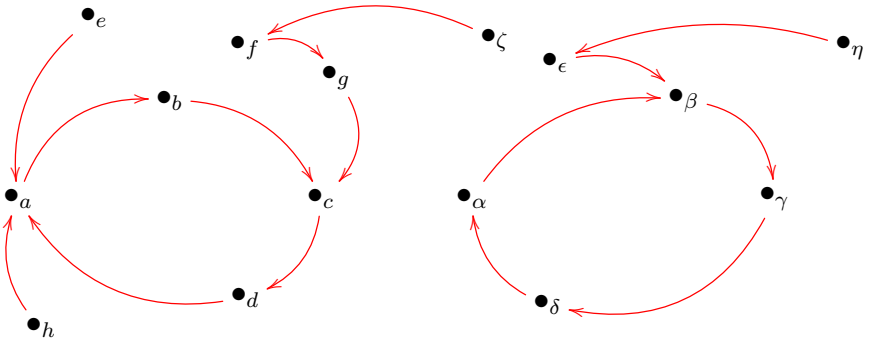
It turns out that, unlike equation (3.1), the discrete HJ equation does not always admit a unique solution, as occurs in the continuous setting. An interesting feature of the discrete framework is that it allows for a detailed characterization of the set of solutions. To achieve this, we introduce an additional assumption.

Assumption 3. *For any $x \in V$ there is exactly one $(x, y) \in E$ such that $\Delta(x, y) = 0$. Call $E_0 \subseteq E$ the set of edges $(x, y) \in E$ having $\Delta(x, y)$ equal to zero (Figure 3.1).*

This is a technical assumption that allows to give a simple and complete carachterization of the solutions of the HJ equation. In particular the edges in E_0 are those across which the chain jumps at zero cost and represent the discrete counterpart of critical points and ω limit sets. Under this assumption the geometric structure of (V, E_0) simplifies and elementary cycles play the same role of the stable equilibrium points for continuous dynamics. In the general case, however, there may be more variations, making a full characterization of solutions to the discrete HJ equations more challenging. The meaning of the assumption is that at each vertex of the graph there is always exactly one typical escape direction that is across the unique edge having zero cost. In absence of fluctuations the chain will follow the unique typical edge so that our model can be considered a perturbation of a discrete deterministic, apart the holding times, dynamical system. We are excluding also cases on which there is a vertex for which the total rate of jump is exponentially small, being an actractive point modulo fluctuations.



(a) A graph (V, E) satisfying Assumption 2. The edges with Δ equal to zero are highlighted in red.



(b) The corresponding subgraph (V, E_0).

Figure 3.1

By Assumptions 2 and 3, it follows that the function $W : V \to \mathbb{R}$ of Lemma 1 has to satisfy following discrete HJ equation

$$W(x) = \min_{y:(y,x)\in E} \left\{ W(y) + \Delta(y,x) \right\} , \qquad \forall x \in V . \tag{3.3}$$

Remark 1. *Here we are considering continuous time Markov chains but similar results hold in the case of discrete time Markov chains. We are studying the behavior of the system at an exponential scale, and t is possible to associate to any discrete time Markov chain with exponentially small transition probabilities a continuous time chain having the same asymptotic rates. Some of the related mathematical structure, as for example the sample path Large Deviations, will have a different structure, but the discrete Hamilton Jacobi equation will be the same. Since for discrete time Markov chains the transition probabilities are normalized to one, this will corresponds to the fact that for each $x \in V$ there will be at least an edge in E_0 exiting from x. This implies in particular that the HJ equation in the discrete time case will be always of the form* (3.3)*. The Markov Chain Tree Theorem holds also in the discrete time case and we have therefore that the discrete time case, corresponds to a special case of the continuous time one. We will briefly discuss the connection with the discrete weak KAM theory in section 3.5 in terms of discrete time Markov chains.*

3.2 Freidlin-Wentzell Quasipotential

Let τ be an arborescence as in Definition 5. Given a weight function $Q : E \to \mathbb{R}$, to any arborescence τ the following weights are associated

$$\begin{cases} Q(\tau) := \sum_{e\in\tau} Q(e) , \\ \mathcal{P}_Q(\tau) := \prod_{e\in\tau} Q(e) . \end{cases} \tag{3.4}$$

In the formulas above, the products and sums are taken over all directed edges e that belong to the arborescence τ. For example, if the weight function is r_N, then $\mathcal{P}_{r_N}(\tau) = \prod_{e\in\tau} r_N(e)$; and if the weight function is Δ, then $\Delta(\tau) = \sum_{e\in\tau} \Delta(e)$. A similar notation is used for subgraphs that are not arborescences.

A cycle in $\mathcal{G}$ is a sequence $(z_1, \dots, z_n)$ of distinct vertices $z_i \in V$ such that $(z_i, z_{i+1}) \in E$ where $z_{n+1} \equiv z_1$.

With this notation, the Markov chain tree theorem (Theorem 1.2) can be restated as follows: the invariant measure of an irreducible Markov chain is given by

$$\pi_N(x) = \frac{\sum_{\tau\in\mathcal{T}_x} \mathcal{P}_{r_N}(\tau)}{\sum_{\tau\in\mathcal{T}} \mathcal{P}_{r_N}(\tau)} . \tag{3.5}$$

The Matrix Tree Theorem allows for the immediate deduction of a Large Deviations principle for the sequence of invariant measures.

Lemma 2. *The sequence of invariant measures* $(\pi_N)_{N\in\mathbb{N}}$ *defined in* (3.5) *satisfies a Large Deviations principle with rate functional*

$$\mathcal{FW}(x) := \lim_{N\to+\infty} -\frac{1}{N}\log \pi_N(x) = \inf_{\tau\in\mathcal{T}_x} \Delta(\tau) - \inf_{\tau\in\mathcal{T}} \Delta(\tau)\,, \qquad x \in V\,. \tag{3.6}$$

The function $\mathcal{FW}$ *is called* FW-quasipotential *and it is a solution to the discrete HJ equation* (3.2)*.*

Proof. Formula (3.5) together with Assumption 2 implies immediately (again by the basic principle of large deviations "the largest exponent wins", as discussed in the proof of lemma 1) that the sequence $(\pi_N)_{N\in\mathbb{N}}$ satisfies (3.6). The function $\mathcal{FW} : V \to \mathbb{R}^+$ is the LD rate functional of the invariant measure. By Lemma 1, $\mathcal{FW}$ is a solution of equation (3.2). Note that the additive constant in (3.6) implies that $\inf_{x\in V} \mathcal{FW}(x) = 0$. □

3.3 Viscosity Solutions

Observe that the discrete HJ equations (3.2) and (3.3) form a collection of equations that must be satisfied at each node of the graph. In general, this discrete HJ equation does not have a unique solution.

Solutions to (3.3) can be naturally characterized—similarly to viscosity solutions for continuous HJ equations—using subsolution and supersolution conditions. Differently from the continuous setting where the notion of viscosity solutions introduces a class of weak solutions that do not satisfy the equation in classical sense, in the discrete setting we are not enlarging the set of solutions of (3.3), that is algebraic in nature and has a well defined set of solutions. What happens is that, in order to have a characterization of the solutions of (3.3), it is very useful to introduce a notion of subsolution and one of supersolution. The set of sub/supersolutions are easier to be characterized and the set of solutions is the intersection of the two. The definitions of sub/supersolutions parallel the continuous definitions and indeed could be very useful in comparisons and scaling limits.

Definition 28. *A function* $W : V \to \mathbb{R}$ *is a* subsolution *of equation* (3.3) *if the following inequality holds for all* $(x, y) \in E$*:*

$$W(x) \le W(y) + \Delta(y, x)\,, \qquad (x, y) \in E\,. \tag{3.7}$$

A function $W : V \to \mathbb{R}$ *is a* supersolution *of* (3.3) *if, for any* $x \in V$*, there exists* $y^* = y^*(x) \in V$ *such that* $(y^*, x) \in E$ *and*

$$W(x) = W(y^*) + \Delta(y^*, x)\,, \qquad x \in V\,. \tag{3.8}$$

Let $E_W \subseteq E$ *be the set of edges where equality* (3.8) *is verified and we call it the* Skeleton *of* W*. A function* W *is a* solution *of* (3.3) *if it is both a subsolution and a supersolution.*

Call $\underline{\mathcal{W}}$ *and* $\overline{\mathcal{W}}$ *the set of subsolutions and supersolutions, respectively, and call* $\mathcal{W} := \underline{\mathcal{W}} \cap \overline{\mathcal{W}}$ *the set of solutions.*

Since adding an arbitrary constant to a solution leads to another solution, it is natural to consider the set of normalized solutions, which we refer to as

$$\mathcal{W}_0 := \left\{ W \in \mathcal{W} : \inf_{x \in V} W(x) = 0 \right\}. \tag{3.9}$$

As already observed in the proof of Lemma 2, the $\mathcal{FW}$ solution belongs to $\mathcal{W}_0$.

Moreover, notice that By Assumption 3, the graph (V, E_0) is a map. Call $\mathcal{C}$ the set of cycles of the map (V, E_0), then $1 \le |\mathcal{C}| \le \left\lceil \frac{|V|}{2} \right\rceil$ ($\lceil \cdot \rceil$ represents the integer part). The value one is obtained in the case of one in-unicyclic graph and the second bound follows since there are no loops (edges of the type (x, x)).

For any $C_i \in \mathcal{C}$, $i = 1, \dots, |\mathcal{C}|$, call $\mathfrak{u}_i$ the associated in-unicyclic component of (V, E_0) whose vertices are the *basin of attraction of* C_i, i.e. the set of $x \in V$ such that $d(x, C_i) = 0$. For simplicity of notation we write also $i \in \mathcal{C}$ to denote the cycle C_i.

3.3.1 Charachterization of $\underline{\mathcal{W}}$ and $\overline{\mathcal{W}}$

The Set of Subsolutions $\underline{\mathcal{W}}$

Recall that a discrete pseudo quasimetric is a function $d : V \times V \to [0, +\infty)$ that satisfies

- triangle inequality: $d(x, z) \le d(x, y) + d(y, z)$, for any $x, y, z \in V$,
- point inequality: $d(x, x) = 0$, for any $x \in V$.

It may be not symmetric, i.e. there may exist $x, y \in V$ such that $d(x, y) \neq d(y, x)$, and not separable, i.e. there may exist $x, y \in V$, with $x \neq y$, such that $d(x, y) = 0$.

The weight functions $\Delta : E \to [0, +\infty)$ induce on the graph (V, E) a pseudo quasimetric d defined as

$$d(x, y) := \inf_{\gamma_{x,y} \in \mathcal{R}_{x,y}} \Delta(\gamma_{x,y}) \,, \tag{3.10}$$

where $\mathcal{R}_{x,y}$ is the set of directed paths $\gamma_{x,y}$ from x to y: a directed path from x to y is given by $\gamma_{x,y} := (x_0, x_1, \dots x_n)$ such that $x_i \in V$, $(x_i, x_{i+1}) \in E$, for all i and $x_0 = x, x_n = y$. Call $\Delta(\gamma_{x,y})$ the length of the path $\gamma_{x,y}$ and observe that $d(x, y) = d(y, x) = 0$ when x, y belong to the same cycle in the map (V, E_0).

Given $U, S \subseteq V$, define

$$d(U, y) := \inf_{x \in U} d(x, y), \quad d(x, S) := \inf_{y \in S} d(x, y), \quad d(U, S) := \inf_{x \in U, y \in S} d(x, y)$$

the distance between a set and a vertex, a vertex and a set, and two sets, respectively.

Call respectively $\bar\gamma_{x,y}, \bar\gamma_{U,y}, \bar\gamma_{U,S}$ the geodetic paths achieving the minimum in the above definitions (one of the paths chosen arbitrarily in the case of several minimizers).

Given a vertex x, it belongs to the exiting basin of attraction of the cycle $C_i \in \mathcal{C}$ if $d(C_i, x) \le d(C_j, x)$ for any $j \neq i$. In the case where there is equality among two or more cycles, the node x is assigned to the basin of attraction of an arbitrary cycle among the minimizers.

Definition 29. *Call* Lip_1 *the set of Lipschitz functions with Lipschitz constant one that are identified by the set of* $|V| \times (|V| - 1)$ *inequalities*

$$W(x) - W(y) \le d(y, x) \,, \qquad \forall x, y \in V \,. \tag{3.11}$$

The following result links subsolutions and Lipschitz functions.

Lemma 3. *The set of subsolutions and the set of Lipschitz functions define the same polyedron, i.e.* $\underline{\mathcal{W}} = \mathrm{Lip}_1$.

Proof. Let W be a subsolution and let $\gamma_{yx} = (y = x_0, x_1, \dots, x_n = x)$ be a path from y to x. Observe that, for any i, $W(x_{i+1}) - W(x_i) \le \Delta(x_i, x_{i+1})$ and the left hand side of the sum over i of this inequalities is telescopic. Inequality (3.11) follows by taking the infimum over all paths. Conversely if (3.11) holds then, for any $(y, x) \in E$,

$$W(x) - W(y) \le d(y, x) \le \Delta(y, x).$$

□

Since a function in Lip_1 must be constant along the cycles in $\mathcal{C}$, the space Lip_1 is a $|V| - \sum_{C_i \in \mathcal{C}} (|C_i| - 1)$ dimensional polyhedron.

Remark 2. *Every polyedron $\mathfrak{P}$ can be written as the Minkowski sum $\mathfrak{P} = \mathfrak{C} + \mathfrak{B}$, where $\mathfrak{C}$ is a convex cone and $\mathfrak{B}$ is a polytope, i.e. a bounded polyedron, see for example section 5.3 of [57]. In the case of Lip_1, the cone $\mathfrak{C}$ is reduced to a single line corresponding to functions that are multiple of the identity. This characteristic is also inherited by the solutions of the discrete HJ equations. For this reason, we focus solely on the normalized solutions by adding a suitable constant function. We do not delve further into these discrete geometric features.*

The Set of Super Solutions $\overline{\mathcal{W}}$

Consider an out-unicyclic component $\mathfrak{o}$ such that its unique cycle is one of the cycles in $\mathcal{C}$ and define the function $\omega_{\mathfrak{o}} : V \to \mathbb{R}$ as follows:

- $\omega_{\mathfrak{o}}(x) = 0$ for any x in the unique cycle,
- $\omega_{\mathfrak{o}}(x) = 0$ for any $x \notin \mathfrak{o}$,
- the values of $\omega_{\mathfrak{o}}(x)$ for the remaining $x \in \mathfrak{o}$ are uniquely determined by the condition $\omega_{\mathfrak{o}}(a) - \omega_{\mathfrak{o}}(b) = \Delta(b, a)$, for any $(b, a) \in \mathfrak{o}$.

The third condition uniquely identify the values $\big(\omega_{\mathfrak{o}}(x)\big)_{x\in\mathfrak{o}}$ since for each $x \in \mathfrak{o}$ there is a unique path γ going from the cycle to x and then $\omega_{\mathfrak{o}}(x) = \sum_{e\in\gamma} \Delta(e)$. Define also the indicator function of the nodes belonging to an out-cyclic component $\mathbb{I}_{\mathfrak{o}} : V \to \{0, 1\}$ by fixing $\mathbb{I}_{\mathfrak{o}}(x) = 1$ when $x \in \mathfrak{o}$ and zero otherwise.

A collection $(\mathfrak{o}_i)_{i\in I}$ of out-unicyclic components is called *spanning* if the cycle of each component $\mathfrak{o}_i$ is an element of $\mathcal{C}$, all the components are disjoint and each $x \in V$ belongs to one component.

Proposition 1. *A function $W : V \to \mathbb{R}$ is an element of $\overline{\mathcal{W}}$ the set of supersolutions of the Hamilton Jacobi equation* (3.3)*, if and only if there exists a spanning collection of out-unicyclic components $(\mathfrak{o}_i)_{i\in I}$ and some values of the parameters $(\lambda_i)_{i\in I}$, with $\lambda_i \in \mathbb{R}$, such that*

$$W = \sum_{i\in I} \left(\omega_{\mathfrak{o}_i} + \lambda_i \mathbb{I}_{\mathfrak{o}_i}\right) . \tag{3.12}$$

Proof. Consider a function W as in (3.12). Note that the terms of the sum are functions with disjoint support. Since the collection $(\mathfrak{o}_i)_{i\in I}$ is spanning, any $x \in V$ belongs to exactly one $\mathfrak{o}_i$ and for each $x \in \mathfrak{o}_i$ there is exactly one $(y, x) \in \mathfrak{o}_i$. By the disjoint support property we have $W(x) - W(y) = \omega_{\mathfrak{o}_i}(x) - \omega_{\mathfrak{o}_i}(y) = \Delta(y, x)$, so that the supersolution condition is satisfied at each $x \in V$.

Conversely suppose that W is a supersolution. In the skeleton E_W there is at least one edge entering on each vertex of V. For each vertex in V remove arbitrarly from E_W all but one edge entering. As discussed below definition 28, the

resulting graph is a spanning collection of out-unicyclic components $(\mathfrak{o}_i)_{i\in I}$. The corresponding cycles can be only the elements $C_i \in \mathcal{C}$ and there are no other cycles, indeed $W(x) = W(y)$ for each $x, y \in C_i$. Define $\lambda_i := W(x)$, $x \in C_i$, $\forall i \in I$. Then, for vertex $z \in \mathfrak{o}_i$, $W(z) = \omega_{\mathfrak{o}_i}(z) + \lambda_i$ and since the supports are disjoint formula (3.12) holds. □

3.3.2 Freidlin-Wentzell Solutions

In Lemma 2, we proved that $\mathcal{FW}$ is a solution of the discrete HJ equation (3.3). It is also interesting and non-trivial to deduce this fact directly by inserting $\mathcal{FW}$ in (3.2).

Proposition 2. *The FW-quasipotential in* (3.6) *is always an element of* $\mathcal{W}_0$*, the set of normalized solutions of* (3.2).

Proof. The validity of (3.2) is checked for each node. Fix $x \in V$ and let $y^* \in V$ be such that $(x, y^*) \in E_0$ and $\tau^* \in \mathcal{T}_x$ be the minimal arborescence in (3.6). Then $\Delta(x, y^*) = 0$ and $\Delta(\tau^*)$ is minimal in $\mathcal{T}_x$.

From the setup, $\gamma^* := (x, y^*) \cup \tau^* \in \mathcal{U}_x^{\text{in}}$ and moreover

$$\Delta(\gamma^*) = \min_{\gamma\in\mathcal{U}_x^{\text{in}}} \Delta(\gamma) = \Delta(\tau^*)\,. \tag{3.13}$$

In fact, any $\gamma \in \mathcal{U}_x^{\text{in}}$ can be written as $\gamma = (x, y) \cup \tau$ for a suitable $y \in V$ and $\tau \in \mathcal{T}_x$ and $\Delta(\gamma) = \Delta(x, y) + \Delta(\tau)$. Since both $\Delta(x, y^*)$ and $\Delta(\tau^*)$ are minimal (respectively on $(x, y) \in E$ and $\mathcal{T}_x$) we proved (3.13).

For any $y \in V$, up to a constant, $\mathcal{FW}(y) = \Delta(\tau)$ for a suitable $\tau \in \mathcal{T}_y$ and moreover when $(y, x) \in E$ the graph $(y, x) \cup \tau \in \mathcal{U}_x^{\text{in}}$. By (3.13)

$$\mathcal{FW}(y) + \Delta(y, x) = \Delta((y, x) \cup \tau) \geq \Delta(\gamma^*) = \mathcal{FW}(x)\,, \qquad \forall (y, x) \in E\,, \tag{3.14}$$

that means $\mathcal{FW} \in \underline{\mathcal{W}}$. It remains to prove that there exists an $(y', x) \in E$ such that the equality holds in (3.14), i.e. $\mathcal{FW} \in \overline{\mathcal{W}}$. Consider (y', x) the unique edge entering in x and belonging to the unique cycle in $\gamma^* \in \mathcal{U}_x^{\text{in}}$ (recall that γ^* is carachterized by (3.13)).

Since $\gamma^* \setminus (y', x) \in \mathcal{T}_{y'}$

$$\mathcal{FW}(y') + \Delta(y', x) \leq \Delta(\gamma^* \setminus (y', x)) + \Delta(y', x) = \Delta(\gamma^*) = \mathcal{FW}(x)\,. \tag{3.15}$$

Last inequality together with the inequality (3.14), implies the equality. □

3.4 The Geometric Structure of $\mathcal{W}$

In this section we give some general representations of the set of solutions of the discrete HJ equation and discuss their geometric characterizations.

3.4.1 Minimal Solutions

Proposition 2 establishes that there exists at least one solution of (3.3). The following proposition indicates that if there are two or more solutions, they can be combined to form others.

Proposition 3. *Suppose that* $\{W_j\}_{j\in J} \subseteq \mathcal{W}$ *is a family of solutions of the discrete HJ equation* (3.3)*. Then the function* W_J*, defined as*

$$W_J(x) := \inf_{j\in J}\{W_j(x)\}, \quad \forall x \in V.$$

is a solution of the discrete HJ equation. If moreover $\{W_j\}_{j\in J} \subseteq \mathcal{W}_0$ *then* $W_J \in \mathcal{W}_0$*,*

Proof. For all $x \in V$:

$$\begin{aligned}
\min_{y:(y,x)\in E}\{W_J(y) + \Delta(y,x)\} &= \min_{y:(y,x)\in E}\left\{\inf_{j\in J}\{W_j(y)\} + \Delta(y,x)\right\} \\
&= \min_{y:(y,x)\in E}\left\{\inf_{j\in J}\{W_j(y) + \Delta(y,x)\}\right\} \\
&= \inf_{j\in J}\left\{\min_{y:(y,x)\in E}\{W_j(y) + \Delta(y,x)\}\right\} \\
&= \inf_{j\in J}\{W_j(x)\} \\
&= W_J(x).
\end{aligned}$$

The forth equality is obtained using that W_j is solution for each $j \in J$. The second statement follows by the fact that $0 \leq \inf_x W_J(x) \leq \inf_x W_j(x) = 0$ so that $W_J \in \mathcal{W}_0$. □

Corollary 1. *The discrete HJ equation* (3.3) *admits among the normalized solution a minimal one, defined as*

$$W_m := \inf\{W : W \in \mathcal{W}_0\}. \tag{3.16}$$

Proof. By Proposition 2, $W_m \in \mathcal{W}_0$ and, by definition, $W_m \leq W$ for any $W \in \mathcal{W}_0$. □

Proposition 3 can be easily extended to the solutions of the general discrete HJ equation (3.2).

3.4.2 Quasipotentials

For each subset C of V, the distance between set C and a node x defines a function that measures the minimum cost to move from a node in C to the node x. If C is a cycle in $\mathcal{C}$ this function is a normalized solution to the discrete HJ equation.

Definition 30. *For each cycle $C \in \mathcal{C}$ of (V, E_0), define the corresponding* quasipotential *based on C as*

$$W_C(x) := d(C, x)\,, \qquad x \in V\,. \tag{3.17}$$

Observe that, by definition, $W_C(x) = 0$ for any $x \in C$ and $W_C(x) > 0$ for any $x \notin C$.

Fix $C \in \mathcal{C}$ and consider the edges $\bar{\gamma}_C := \cup_{\{x \in V \cap C^c\}} \bar{\gamma}_{C,x}$. Call $(V, \bar{\gamma}_C^*)$, the directed graph obtained removing arbitrarily from $\bar{\gamma}_C$, for each node $x \notin C$ all the entering edges except one and finally adding the edges of the cycle C.

Lemma 4. *The directed graph $(V, \bar{\gamma}_C^*) = \mathfrak{o}$ is a spanning out-unicyclic graph with unique cycle C. Moreover, $W_C(x) = \omega_{\mathfrak{o}}(x)$.*

Proof. Since it is constructed using geodetic, the directed graph $(V, \bar{\gamma}_C^*)$ has a unique cycle that is C and, for each vertex $x \in V$ outside the cycle, there is exactly one edge entering to x, and this means that it is a spanning out-unicyclic component. The equality $W_C(x) = \omega_{\mathfrak{o}}(x)$ follows by definition. □

Proposition 4. *For each $C \in \mathcal{C}$ the* quasipotential W_C *based on C belongs to $\mathcal{W}_0$.*

Proof. By Proposition 1 and Lemma 4, W_C is a supersolution. For the subsolution condition, suppose that there exists $(y, x) \in E$ such that $W_C(x) > W_C(y) + \Delta(y, x)$. By definition of W_C previous inequality is impossible when at least one among x, y belong to C; we assume therefore that $x, y \notin C$. Define the path $\tilde{\gamma}_{C,x} = \bar{\gamma}_{C,y} \cup (y, x)$. Then it follows that

$$d(C, x) = \Delta(\bar{\gamma}_{C,x}) = W_C(x) > W_C(y) + \Delta(y, x) = \Delta\big(\tilde{\gamma}_{C,x}\big)\,,$$

which is impossible since $\bar{\gamma}_{C,x}$ is a geodetic from C to x. The normalization follows since $W_C \geq 0$ and $W_C(x) = 0$ for any $x \in C$. □

This is a special situation of a more general case when, given a collection of spanning out-unicyclic components $(\mathfrak{o}_i)_{i \in I}$, for each $x \in \mathfrak{o}_i$ the unique path from the cycle C_i to x is a geodetic path $\overline{\gamma}_{C_i,x}$. An out-unicyclic component of this type is called *geodetic out-unicyclic component.*

The following simple lemma will be useful for the characterization.

Lemma 5. *Take $V \in \mathcal{W}$. The associated family of spanning out-unicyclic components $(\mathfrak{o}_i)_{i\in I}$ in Proposition 1 is composed by geodetic out-unicyclic components.*

Proof. Assume by contraddiction that $\mathfrak{o}_i$ is not geodetic. There exists an $x \in \mathfrak{o}_i \cap C_i^c$ whose path in $\mathfrak{o}_i$ from C_i is not the geodetic one. This implies that $V(x) - V(y) > d(C_i, x)$, with $y \in C_i$, and by Lemma 3, V is not a subsolution. □

The following theorem gives a complete characterization of solutions of (3.3) in terms of the parameters $\lambda = (\lambda_i)_{i\in\mathcal{C}} \in \mathbb{R}^{|\mathcal{C}|}$.

Theorem 9. *The set $\mathcal{W}$ of the solutions to the discrete HJ equation* (3.3) *coincides with the family of functions*

$$W_\lambda = \min_{i\in\mathcal{C}} \left\{ W_{C_i} + \lambda_i \right\}, \tag{3.18}$$

where the vector $\lambda = (\lambda_i)_{i\in\mathcal{C}}$ belongs to the polyhedron $\mathrm{Lip}_1^{\mathcal{C}}$ in $\mathbb{R}^{|\mathcal{C}|}$ defined by

$$\lambda_i - \lambda_j \le d(C_j, C_i)\,. \tag{3.19}$$

Proof. For any i, $W_{C_i} + \lambda_i$ is a solution then, by Proposition 3, W_λ it is also a solution for any $\lambda \in \mathbb{R}^{|\mathcal{C}|}$. It remains to show that every solution can be written as in formula (3.18) with λ statisfying (3.19).

Take a solution $W \in \mathcal{W}$. Since W is both a supersolution and subsolution, it can be written as (3.12) with $(\mathfrak{o}_i)_{i\in I}$ a spanning geodetic out-unicyclic components. Since any solution has to be constant on each cycle in $\mathcal{C}$, adding all the edges of the cycles in $\mathcal{C}$ we can assume that the set of labels I of the components coincides with $\mathcal{C}$. We call $\lambda_i = W(x)$, where $x \in C_i$. Inequality (3.19) is then satisfied because W is in Lip_1:

$$\lambda_i - \lambda_j = W(x) - W(y) \le d(y, x) = d(C_j, C_i),$$

where $x \in C_i$ and $y \in C_j$. By definition we have $W(x) = W_{C_i}(x) + \lambda_i$ for any $x \in \mathfrak{o}_i$. The final step in the proof is to show that, for any $x \in \mathfrak{o}_i$ and $j \neq i$, it holds: $W_{C_i}(x) + \lambda_i \le W_{C_j}(x) + \lambda_j$. Take $y \in C_j$. Since W is subsolution, $W(x) - W(y) \le d(C_j, x)$ that is equivalent to $W(x) \le \lambda_j + W_{C_j}(x)$.

□

Observe now that the function d induces a discrete quasimetric among the cycles $\mathcal{C}$ in (V, E_0) defined by $d(C, C')$, for any $C, C' \in \mathcal{C}$. Under the Assumption 3, the distance $d(C, C') > 0$, for any $C \neq C'$. Condition (3.19) says that the parameters $(\lambda_i)_{i\in\mathcal{C}}$ coincide with the set $\mathrm{Lip}_1^{\mathcal{C}}$ of the Lipschitz functions on $\mathcal{C}$ with respect to the quasimetric d.

The $\mathcal{FW}$ solution can be identified as a special solution of the form (3.18). Let $\mathcal{FW}^{\mathcal{C}}$ be the minimal weight of arborescences on the complete graph with vertices $\mathcal{C}$ and weights d. We have the following representation of the FW solution $\mathcal{FW}$

Proposition 5. *The solution $\mathcal{FW}$ in* (3.6) *corresponds to a solution as in* (3.18) *with $\lambda_i = \mathcal{FW}^{\mathcal{C}}(C_i)$, i.e. we have*

$$\mathcal{FW}(x) = \min_{C \in \mathcal{C}} \Big\{ W_C(x) + \mathcal{FW}^{\mathcal{C}}(C) \Big\}. \tag{3.20}$$

Proof. For $x \in C' \in \mathcal{C}$, it is easy to show by contraddiction that the minimum in (3.20) is obtained for $C = C'$. Moreover, by definition $W_{C'}(x) = 0$. We have then that, up to a constant, the right hand side of (3.20) coincides with $\mathcal{FW}$ on all the vertices belonging to the cycles in $\mathcal{C}$. Since by Theorem 9 the values λ completely identify the solution, then the right hand side of (3.20) coincides with $\mathcal{FW}$. □

The above result implies that the minimal arborescence $\tau \in \mathcal{T}_x$ is obtained by joining minimal arborescences among the cycles in $\mathcal{C}$. It incurs zero cost for paths from vertices to the cycle of the basin of attraction and requires a single geodetic path $\gamma_{C',x}$ for some $C' \in \mathcal{C}$.

3.4.3 Faces of the Polyhedron Lip_1

We describe some features of the polyhedron Lip_1, and identify the solutions of (3.3) with a subset of the relative boundary of such polyhedron.

The polyhedron Lip_1 is identified by the inequalities (3.11), but as shown in Lemma 3 we can restrict to the inequalities $W(x) - W(y) \leq \Delta(y, x)$ for $(y, x) \in E$. Furthermore, the inequalities associated to edges such that $\Delta(y, x) > d(y, x)$ are redundant. The relative boundary of Lip_1 is the region where at least one of the effective inequalities is indeed an equality. The relative boundary is the union of faces of different dimensions.

A face is identified therefore by the edges of E where equality takes place and this means that to each face, there is an associated subgraph $\mathfrak{f} \subseteq \mathcal{G}$ that shares the same vertices of $\mathcal{G}$ and has special features. Of course equality has to be always satisfied on edges of the cycles in $\mathcal{C}$ so that for any i we have $C_i \subseteq \mathfrak{f}$. A complete geometric carachterization of faces may be difficult: we have for example that any directed path on $\mathfrak{f}$ has to be a geodetic; we just identify some special faces. Let g be a function belonging to a face in the relative boundary of Lip_1 associated the subgraph $\mathfrak{f}$. Ignore the orientation of arrows in $\mathfrak{f}$; since equality takes place on edges, the values of g on vertices belonging to the same connected component of the unoriented subgraph are completely determined by the value of one single vertex. This means that the dimension of the face corresponds to the number of components of the induced subgraph. This dimension can be decresead by one considering the invariance of the polyhedron by a global addition.

Consider $\mathfrak{f} = (\mathfrak{u}_i)_{i \in \mathcal{C}}$, the collection of in-unicyclic components that are the basin of attaction of the cycles in (V, E_0). Note that any path in $\mathfrak{f}$ is geodetic

and moreover each edge $e \in \mathfrak{f}$ is such that $\Delta(e) = 0$. The subgraph $\mathfrak{f}$ has $|\mathcal{C}|$ components and corresponds to a special $|\mathcal{C}|$ dimensional face of Lip_1 that contains all the functions g that are constant on each basin of attraction $\mathfrak{u}_i$. The values λ_i that the function g assumes on each basin must satisfy (3.19). We call such face the *minimal* $|\mathcal{C}|$ *face*.

Consider now $\mathfrak{f} = (\mathfrak{o}_i)_{i\in\mathcal{C}}$ a spanning collection of out-unyciclic geodetic components. Any subgraph $\mathfrak{f}$ of this type has $|\mathcal{C}|$ components and corresponds to a $|\mathcal{C}|$ dimensional face of Lip_1. A function g that belongs to one of such faces assumes the values λ_i on the cycles in $\mathcal{C}$ and the values on all the remaining vertices in $\mathfrak{o}_i$ are completely determined by $(\lambda_i)_{i\in\mathcal{C}}$. The values $(\lambda_i)_{i\in\mathcal{C}}$ have to satisfy additional constraints with respect to (3.19), depending on the geometry of $\mathfrak{f}$; indeed there could also be in some cases no values of $(\lambda_i)_{i\in\mathcal{C}}$. The new constraints arise from the fact that, fixing $(\lambda_i)_{i\in\mathcal{C}}$, the function is naturally Lipschitz inside each component, but it is not guaranteed to be globally Lipschitz.. We call faces of this type *maximal* $|\mathcal{C}|$ *faces*.

Theorem 10. *The set $\mathcal{W}$ of the solutions of the Hamilton Jacobi equation* (3.3) *coincides with the union of the maximal $|\mathcal{C}|$ dimensional faces of the polyhedron* Lip_1. *Moreover for any $\lambda = (\lambda_i)_{i\in\mathcal{C}} \in \mathrm{Lip}_1^{\mathcal{C}}$ there exist, among the functions in* Lip_1 *assuming the values λ_i for all $C_i \in \mathcal{C}$, a minimal function g_m and a maximal one g_M; moreover g_m belongs to the minimal $|\mathcal{C}|$ dimensional face and g_M belongs to a maximal $|\mathcal{C}|$ dimensional face and indeed coincides with W_λ.*

Proof. Given a solution $W \in \mathcal{W}$, the corresponding skeleton identifies the spanning outgoing geodetic family $(\mathfrak{o}_i)_{i\in\mathcal{C}}$ of a $\mathfrak{f}$ that corresponds to a maximal $|\mathcal{C}|$ dimensional face of Lip_1 to which W belongs. Conversely given W that belongs to a maximal $|\mathcal{C}|$ dimensional face identified by $\mathfrak{f}$, then: W is a subsolution since is one Lipschitz and it is also a supersolution with the skeleton corresponding to the edges of $\mathfrak{f}$.

For the second part of the statement fix $\lambda \in \mathrm{Lip}_1^{\mathcal{C}}$. For any function $W \in \mathrm{Lip}_1$, for any $x \in V$ and for any $i \in \mathcal{C}$ we have $W(x) - \lambda_i \le d(C_i, x)$. This implies

$$W(x) \le \inf_i \left\{\lambda_i + d(C_i, x)\right\} = W_\lambda(x)\,.$$

The unique function g_m that belongs to the minimal $|\mathcal{C}|$ dimensional face and assumes the values λ on the cycles is the function that is constantly equal to λ_i on each vertex of $\mathfrak{u}_i$ for any $i \in \mathcal{C}$. Since the one Lipschitz condition imposes that any function in Lip_1 that takes the values $\lambda \in \mathrm{Lip}_1^{\mathcal{C}}$ on the cycles has necessarily to assume on $\mathfrak{u}_i$ a value bigger or equal to λ_i, for any i, the proof is complete. □

An interesting question is to find a useful geometric carachterization of $\mathcal{FW}$ as an element of the relative boundary of Lip_1.

3.5 Discrete Weak KAM Theory

In this section, we briefly outline how the above results parallel the general structure of the so-called weak KAM theory (see, for example, [23, 26, 50], to which we refer for comparison) in a purely discrete setting.

The key step in establishing this correspondence is to prove a sample path large deviations principle. The corresponding rate functional plays the role of the action. Although it is not natural to identify a Lagrangian and a corresponding Hamiltonian, we can define a discrete-time Lax-Oleinik semigroup and develop the parallel analysis. It turns out that it is easier and more natural to discuss this relation in terms of discrete time Markov chains. As explained in remark 1 the associated discrete HJ equation is the same.

We consider a family of discrete time Markov chain $\left(X_i^N\right)_{i\in\mathbb{N}}$, depending on the parameter N, assuming values on V and such that

$$\lim_{N\to+\infty} -\frac{1}{N}\log P_N(x,y) = \Delta(x,y) \geq 0\,,$$

where P_N are the transition probabilities. This is exactly the framework of [52, 53, 58–60]. Consider a finite trajectory $\left(X_1^N, \dots X_k^N\right)$; by the assumption (3) we have that the limiting behavior described by the variables $\left(X_1^\infty, \dots X_k^\infty\right)$ is purely deterministic. When $X_i^\infty = x$ then $X_{i+1}^\infty = y$ where y is the unique element of V such that $\Delta(x,y) = 0$. The Markov chains X^N are therefore random perturbations of the deterministic dynamical system X^∞. The behavior of X^∞ is such that after a finite transient period, the system reach one of the cycles in $\mathcal{C}$ and moves then periodically around it. A sample path large deviations is easy to compute and it is one of the main instruments used in [52, 53, 58–60]. Consider a fixed sequence $x := (x_1, \dots, x_k)$ we have then

$$\begin{aligned}
&\lim_{N\to+\infty} -\frac{1}{N}\log \mathbb{P}\Big((X_1^N, \dots, X_k^N) = (x_1, \dots, x_k)\Big) \\
&= \lim_{N\to+\infty} -\frac{1}{N}\sum_{i=1}^{k} \log P_N(x_i, x_{i+1}) = \sum_{i=1}^{k} \Delta(x_i, x_{i+1})\,.
\end{aligned}$$

We have therefore that the exponential cost to observe a finite trajectory x is given by an action functional $\mathcal{A}$ on finite trajectories and defined by

$$\mathcal{A}(x) := \sum_{i=1}^{k} \Delta(x_i, x_{i+1})\,. \tag{3.21}$$

The functional (3.21) acts as an action, assigning a cost to every trajectory of the system. An interesting feature of the action (3.21) is that it does not depend

on the parameter T. For this reason, even though the stochastic dynamics are in continuous time, it is natural to define a corresponding discrete-time Lax-Oleinik semigroup.

Given a function $\mathcal{V}_0 : V \to \mathbb{R}$ we define an iterated function $\mathcal{V}_n$ at time n as

$$\mathcal{V}_n(x_n) := \inf_{x_0,\dots,x_{n-1}} \left\{ \mathcal{V}_0(x_0) + \sum_{i=0}^{n-1} \Delta(x_i, x_{i+1}) \right\} = \inf_y \left\{ \mathcal{V}_{n-1}(y) + \Delta(y, x_n) \right\} . \tag{3.22}$$

This defines our discrete-time Lax-Oleinik semigroup. The fixed points of (3.22) are the solutions of the discrete HJ equation (3.3). This framework is natural for our discrete setting, as there is no straightfowrward definition of the Hamiltonian corresponding to the action (3.21).

Based on the definition (3.22), we note that our discrete contructions parallel the weak KAM theory, including discrete versions of all the related machinery and geometric constructions. We do not provide a detailed comparison here.

3.6 Examples and Remarks

3.6.1 The Reversible Case

As an example, we consider the reversible case: in this situation, if $(x, y) \in E$ then also $(y, x) \in E$. The detailed balance characterizing the invariant measure π_N in the reversible case is given by

$$\pi_N(x) r_N(x, y) = \pi_N(y) r_N(y, x) , \qquad (x, y) \in E . \tag{3.23}$$

The equivalent of Lemma 1 for the equation (3.23) gives the conditions

$$W(x) + \Delta(x, y) = W(y) + \Delta(y, x), \quad (x, y) \in E. \tag{3.24}$$

Equations (3.24) have a unique solution (up to an arbitrary additive constant) if the following condition, called *gradient condition*, is satisfied: $\delta(x, y) := \Delta(x, y) - \Delta(y, x)$ is a discrete vector field of gradient type, namely antisymmetric $\delta(x, y) = -\delta(y, x)$. If the rates are reversible (i.e. (3.23) is satisfied) the gradient condition is satisfied and there is a unique solution to (3.24); if the gradient condition is not satisfied there are clearly no solutions to (3.24).

The gradient condition corresponds to requiring that, for any cycle $C = (x_0, \dots, x_n)$, it holds $\sum_{(x,y)\in C} \delta(x, y) = 0$. In this case the unique, up to an arbitrary constant, solution to (3.24) is

$$\mathcal{FW}(x) := \sum_{e \in \gamma_{x^*,x}} \delta(e) , \forall x \in V, \tag{3.25}$$

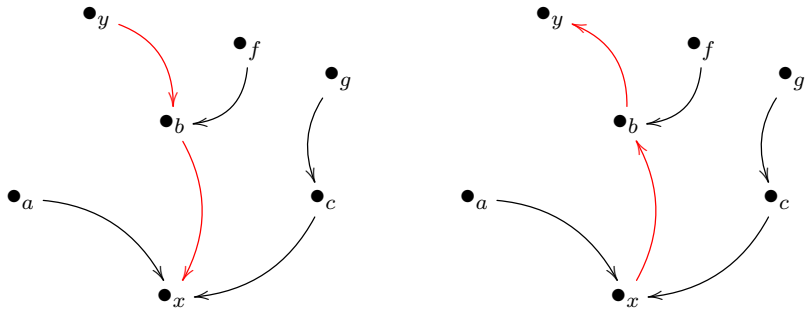


Figure 3.2: The switch of orientation from an arborescence oriented toward x to an arborescence oriented toward y.

where x^* is an arbitrary fixed node and $\gamma_{x^*,x}$ is also an arbitrary path. The independence of the definition (3.25) from the initial point and the arbitrary path follows by classic homological arguments since δ is antisymmetric and the sum along any cycle is zero. Considering different paths and inverting the direction of one of them you get a cycle and deduce the invariance. Changing the initial point just change the values by a global additive constant. We used the notation $\mathcal{FW}$ since (3.25) coincides, up to a constant, with the Freidlin-Wentzell solution. Moreover, $\mathcal{FW}(y) - \mathcal{FW}(x) = \sum_{e\in\gamma_{x,y}} \delta(e)$.

In the reversible case the Assumption 3 imposes a strict condition on the graph (V, E_0).

Lemma 6. *Given a sequence of reversible Markov chains, Assumption 3 imposes that the map (V, E_0) contains only cycles of length 2, i.e. $|C| = 2$ for any $C \in \mathcal{C}$.*

Proof. Consider $C \in \mathcal{C}$. Recalling the gradient condition and the fact that $\Delta(e) = 0$ for $e \in C$, we have

$$0 = \sum_{e\in C} \delta(e) = -\sum_{e\in C} \Delta(-e)\,,$$

where $-e$ is the edge obtained from e by switching its orientation. The above equalities can be true only if $\Delta(e) = \Delta(-e) = 0$ for any $e \in C$ but this implies that for each node in C there are at least two edges in E_0 unless $|C| = 2$. □

Remark 3. *In the reversible case, there is a bijection between $\mathcal{T}_x$ and $\mathcal{T}_y$ for any $x, y \in V$. The map from elements in $\mathcal{T}_x$ to $\mathcal{T}_y$ involves inverting the orientation of the edges in the unique directed path from y to x in $\mathcal{T}_x$. Similarly, the inverse map involves inverting the orientation of the edges in the unique path from y to x in $\mathcal{T}_y$ (see Figure 3.2).*

Proposition 6. *The minimal weight arborescences $\tau_x^* \in \mathcal{T}_x$ and $\tau_y^* \in \mathcal{T}_y$ are related by the bijection in remark 3. In particular all the minimal weights arborescences $(\tau_z^*)_{z\in V}$ have the same unoriented support, differ just by the orientation and $\Delta\left(\tau_y^*\right) - \Delta\left(\tau_x^*\right) = \sum_{e\in\gamma_{x,y}} \delta(e)$.*

Proof. Given $\tau_x \in \mathcal{T}_x$ and $\tau_y \in \mathcal{T}_y$ two elements linked by the bijection, by definition, we have

$$\Delta\left(\tau_y\right) - \Delta\left(\tau_x\right) = \sum_{e\in\gamma_{x,y}} \delta(e) = \mathcal{FW}(y) - \mathcal{FW}(x). \tag{3.26}$$

This implies that the weights of the arborescences in $\mathcal{T}_y$ are the same as those in $\mathcal{T}_x$, up to a global shift, and the minimizers are therefore mapped bijectively. Consequently, every pair of minimizers in $(\mathcal{T}_z)_{z\in V}$ is related simply by the inversion of some arrows. The last statement follows from the fact that (3.26) holds for any pair of arborescences related by the bijection. □

In the reversible case, we can characterize the maximal $|\mathcal{C}|$ dimensional face to which the Friedlin Wentzell solution belongs. Recall that a maximal $|\mathcal{C}|$ dimensional face is indentified by a spanning collection of geodetic out-unicyclic components.

Proposition 7. *In the reversible case the solution $\mathcal{FW}$ belongs to the maximal $|\mathcal{C}|$ dimensional face of* Lip_1 *associated to the spanning out-unicyclic geodetic components obtained reversing all the edges of the map (V, E_0).*

Proof. We need to show that the skeleton of $\mathcal{FW}$ is obtained by reversing the edges in (V, E_0). Recall that $\mathcal{FW}$ solves (3.24), then for any $(x, y \in E)$ we have

$$\mathcal{FW}(x) + \Delta(x, y) = \mathcal{FW}(y) + \Delta(y, x)\,.$$

Fix $x \in V$ and consider the infimum over all y such that $(x, y) \in E$ of the left hand side of the above equation. We have that the minimum is achieved for the unique $(x, y^*) \in E_0$. This leads to the equality $\mathcal{FW}(x) = \mathcal{FW}(y^*) + \Delta(y^*, x)$ which is equivalent to say that (y^*, x) belongs to the skeleton of $\mathcal{FW}$. □

3.6.2 The Ring

Consider $\mathcal{G}$ as a finite ring with k nodes , where $V = \mathbb{Z}/(k\mathbb{Z})$ and $E = \{(x, x\pm1), x \in V\}$. In this case, we have a simple and compact representation of the $\mathcal{FW}$ solution, which shares a similar structure with the continuous one, see [25]. Let $\delta : E \to \mathbb{R}$ be the discrete vector field, i.e. such that $\delta(x, y) = -\delta(y, x)$, defined by $\delta(x, x + 1) := \Delta(x, x + 1) - \Delta(x + 1, x)$, with the remaining values defined by antisymmetry. Define the function $S : \mathbb{N} \to \mathbb{R}$ by setting $S(x) := \sum_{y=1}^{x-1} \delta(y, y + 1)$. In the previous formulas the function δ is computed on edges of $\mathcal{G}$ so that the sums

in the arguments are modulo k and $y+k=y$. If $S(k+1) := \sum_{y=1}^{k} \delta(y, y+1) = 0$ we have that $S(y+k) = S(y)$ so that S is a periodic function of period k, and this corresponds to the reversible case (since the gradient condition is satisfied).

We have

$$\mathcal{FW}(x) = \min_{x \le y < x+k} \Big\{ S(x) - S(y) - \Delta(y, y+1) \Big\}. \tag{3.27}$$

The function $\mathcal{FW}$ defined in (3.27) is defined on $\mathbb{N}$ but can be easily shown to be periodic of period k so that can be naturally interpreted as a function on V. To prove that (3.27) coincides with (3.6) up to a constant, we have just to add and subtract on the right hand side of (3.27) the constant $\sum_{z=1}^{k} \Delta(z, z+1)$, getting that the right hand side of (3.27) coincides with $\min_{\tau \in \mathcal{T}_x} \Big\{ \Delta(\tau) - \sum_{z=1}^{k} \Delta(z, z+1) \Big\}$.

In the reversible case, when S is periodic, we may write

$$\mathcal{FW}(x) = S(x) + \min_{x \le y < x+k} \Big\{ -S(y) - \Delta(y, y+1) \Big\}.$$

The second term on the right hand side does not depend on x so that $\mathcal{FW}$ coincides with S up to an additive constant.

The minimization formula (3.27) is very similar to the corresponding continuous one for diffusions on a circle and features a simple and interesting geometric interpretation, as in [25]. In the continuous case it is easier to describe but in the discrete setting there is a similar construction. Consider f a continuous function on the unit circle $\mathbb{R}/\mathbb{Z}$ and define the function $S : \mathbb{R} \to \mathbb{R}$ as $S(x) := \int_0^x f(y)dy$. Taking

$$\mathcal{FW}(x) := \inf_{y \in [x, x+1]} \Big\{ S(x) - S(y) \Big\}, \tag{3.28}$$

then Theorem 2.3 in [25] summarize some general properties of the function $\mathcal{FW}$, including the fact that it is one-periodic, i.e. it is a function on the circle, coincides with S up to an additive constant if $\int_0^1 f(y)dy = 0$, and can be practically constructed by the so called *sunshine transformation* (see figure 3.3). Formula (3.28) can be interpreted as a minimal continuous arborescences construction, as suggested by the remark 1 in [25]. For the definitions of continuous arborescences on metric graphs and a continuous version of the Matrix Tree Theorem for diffusions on metric graphs, we refer to [3]. In the case of the continuous ring, a continuos arborescence directed torward a point x is obtained cutting the ring on another point, say y, and orienting the two portions of the segment identified by x to direct torward x. Formula (3.28) coincides up to an additive constant with

$$\mathcal{FW}(x) = \min_{y \in [x, x+1]} \Big\{ \int_x^y [-f(z)]_+ dz + \int_y^{x+1} [f(z)]_+ dz \Big\} \tag{3.29}$$

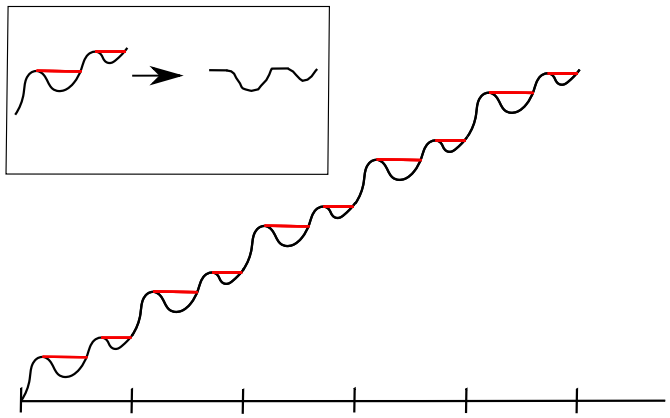


Figure 3.3: An example of the sunshine transformation for the case of a continuous function S (in black) with $\int_0^1 f(y)dy > 0$. The light is assumed to arrive horizontally from the left; parts of the graph below the horizontal red line are in shadow, while the remaining parts are illuminated. The sunshine transformation retains the shadowed parts of the graph and replace the illuminated parts with horizontal segments, shifting them vertically to ensure that the final transformed graph remains continuous. An example of this transformation is illustrated on a portion of the graph inside the rectangle at the upper left. The transformed graph is periodic of period one. In the case $\int_0^1 f(y)dy < 0$, the construction is performed with the light arriving horizontally from the right. For $\int_0^1 f(y)dy = 0$, the graph is already periodic, every point of the graph is in shadow, and the transformation does not modify the graph.

where $[\cdot]_+$ is the positive part. This suggests that solutions of HJ equations associated to diffusions on metric graphs can be obtained by continuous minimal arborescences constructions.

3.6.3 Vanishing Viscosity Solutions

In the continuous setting, viscosity solutions of HJ equations can be obtained by a vanishing viscosity limit. In the case of a diffusion process on a ring, see for example [25, 29], the stationary HJ equation associated to the rate functional of the invariant measure is

$$H(x, W_x) = W_x(W_x - f(x)) = 0\,. \tag{3.30}$$

In general, the viscosity solution of (3.30) is not unique and a natural selection principle for a solution is to consider the stochastic process behind and look at (3.30) as a limit of the stationary Kolmorogov equation. This means to consider the small noise limit $\epsilon \to 0$ of the stationary condition

$$\epsilon \left(e^{-\epsilon^{-1}W^\epsilon(x)}\right)_{xx} - \left(f(x)e^{-\epsilon^{-1}W^\epsilon(x)}\right)_x = 0\,. \tag{3.31}$$

The unique smooth solution W^ϵ can be found explicitely and converges, in the limit as ϵ goes to zero, to the solution (3.28), i.e. $W^\epsilon \to \mathcal{FW}$. The same solution is selected by the classic vanishing viscosity limit obtained by considering the equation

$$H(x, W^\epsilon_x) = W^\epsilon_x\,(W^\epsilon_x - f) = \epsilon W^\epsilon_{xx}\,. \tag{3.32}$$

Indeed, the unique smooth solution of the previous equation can be computed explicitely by the change of variables $\psi^\epsilon = 1/W^\epsilon_x$ that linearizes the equation, and then it can be shown that again $W^\epsilon \to \mathcal{FW}$.

We observe that the solution $\mathcal{FW}$ to the discrete HJ equation (3.3) can also be considered a *vanishing viscosity solution*. The stationary condition (3.1) written substituting e^{-NW^N} to π_N can be written as

$$e^{-NW^N(x)} \sum_{y:(x,y)\in E} r_N(x,y) = \sum_{y:(y,x)\in E} e^{-NW^N(y)} r_N(y,x)\,, \qquad \forall x \in V \tag{3.33}$$

By Lemma 2 we have that $\lim_{N\to+\infty} W^N = \mathcal{FW}$.

3.6.4 Bidimensional Case

In this subsection we show how the FW-solution is an inner point of the polytope identified by $\mathcal{W}$.

Take $V = \{1, 2, 3, 4\}$ and let E be the set of edges depicted in figure 3.4. Suppose that $\Delta(1,2) = \Delta(2,1) = \Delta(3,4) = \Delta(4,3) = 0$ and that, without loss of generality, $0 < \Delta(2,3) < \Delta(4,1)$. The graph (V, E_0) contains two cycles, C_1 and C_2. By Theorem 9 there exists a pair $(\lambda_1, \lambda_2) \in \mathrm{Lip}_1^2$ such that any solution to the

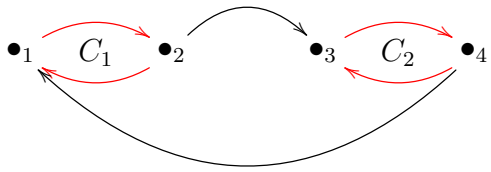


Figure 3.4

discrete HJ equation (3.3) can be written as

$$W_\lambda(x) = \min\{W_{C_1}(x) + \lambda_1; W_{C_2}(x) + \lambda_2\},$$

where

	1	2	3	4
W_{C_1}	0	0	$\Delta(2,3)$	$\Delta(2,3)$
W_{C_2}	$\Delta(4,1)$	$\Delta(4,1)$	0	0

The FW-solution is given by

$$\mathcal{FW}(1) = \mathcal{FW}(2) = \Delta(4,1) - \Delta(2,3); \qquad \mathcal{FW}(3) = \mathcal{FW}(4) = 0.$$

By a straightforward calculation, we obtain that taking $\lambda_1 = \Delta(4,1) - \Delta(2,3)$ and $\lambda_2 = 0$, we have $W_\lambda = \mathcal{FW}$. This choice of parameters shows that the FW-solution is indeed an inner point of the polytope identified by the normalized solution $\mathcal{W}_0$, see figure 3.5.

The grey area in figure 3.5 is the set Lip_1^2 and $\mathcal{W}_0$ is indicated by the thicker segments on the axis.

3.7 Conclusions and Perspectives

The final section is dedicated to outlining potential extensions and unresolved questions. We discuss only general ideas and proposals.

We discussed Markov chains that satisfy Assumption 3, this has been done mainly for reasons of simplicity since in this case the structure of the graph (V, E_0) is simple and has a canonical form. Many of our arguments works also in the general setting when assumption 3 does not hold, but in this case the structure of the graph (V, E_0) can be somehow arbitrary and this is the discrete counterpart of the equilibrium points in the continuous setting. In particular, in the general case,

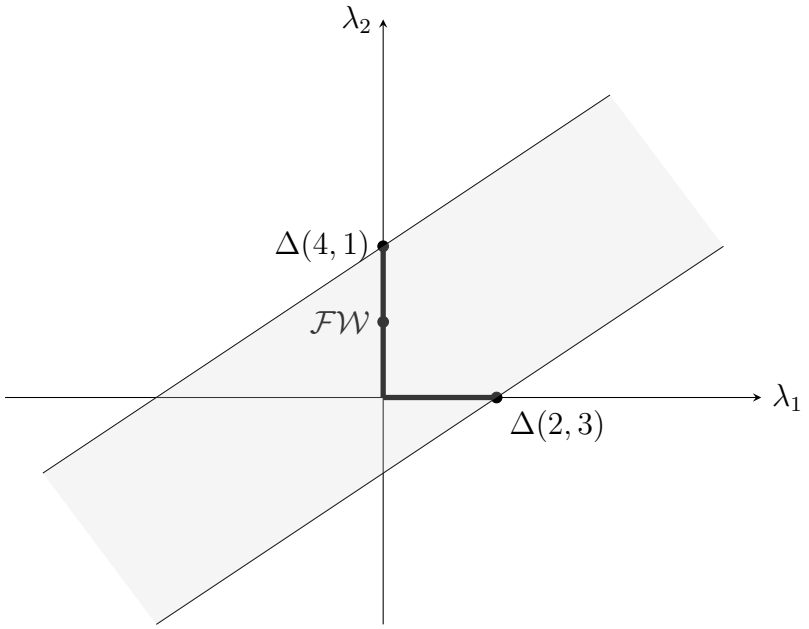


Figure 3.5

equation (3.2) has to be used, and it is not clear if it is always possible to get a simple and complete carachterization of the solutions.

Another interesting issue is to study the asymptotic behavior of the invariant measure of a multiscale Markov chain using the Markov chain tree theorem; in this paper we considered the case of exponential decaying rates and considered just the exponential asymptotic behavior but an interesting issue is to analyze the full asymptotic behavior and how HJ equations are related. This problem is related to the renormalization group contruction in [52, 53, 58–60].

The natural extension and generalization of our result is to consider not just graphs but metric graphs having diffusions along the metric edges with possibly a sticky behavior on the nodes, as in [3]. The natural scaling to be considered is the small noise one, which can involve in different ways also the behavior on the nodes. The analytical problems obtained are within the general theory of HJ equations on metric graphs [17] for which more delicate and involved definitions of solutions are necessary. Several different scaling limit related to problems of this type seems to be interesting.

An interesting problem is to study the discrete HJ equation of the lifted Markov chain described in Section 1.2 and its relation to the discrete HJ equation on the original graph. Is there a simple continuous version of such a lifting?

Finally it is very interesting to investigate the application of the discrete HJ equations to problems involving metastability, like in [52,53,58–60], and the relation

with problems and algorithms for optimization of paths on graph [57].

Part II

On the invariant measure of multiscale Markov chains

4

Trace Processes

In this second part, we focus on the limiting behavior of the invariant measure in multiscale Markov chains. Our analysis is grounded in the theory of trace processes, as developed in [40, 42, 43, 46, 47], which offers a comprehensive framework for studying the dynamics of Markov chains restricted to subsets of the state space. Trace processes are particularly useful in capturing the effective behavior of many complex systems,in which the state space can be naturally decomposed into regions in which the process remains trapped for long periods of time, with rare transitions occurring between them. This phenomenon, commonly referred to as metastable behavior, cannot be adequately captured without an explicit multiscale description.

In order to analyze the invariant measure in this context, we will make use of the Markov chain tree theorem, along with the forest-based representation of absorption probabilities, both of which were outlined in the introduction. These combinatorial tools provide an explicit characterization of the invariant measure and its asymptotic behavior, even in the presence of multiple interacting time scales.

We begin by briefly reviewing the definition and main properties of trace processes, in the context of metastable Markov chains. For a more in-depth analysis, see [42].

4.1 Definition of Trace Processes

Let V be a finite set, and let X_t be a continuous-time, irreducible Markov chain taking values in V.

Denote by $r(x, y)$, for $x \neq y \in V$, the jump rates of the chain, and by $r(x) = \sum_{y \in V} r(x, y)$ the total rate at which the process leaves state x. Let π be the unique invariant measure associated with X_t.

We denote by $D([0,\infty), V)$ the space of CADLAG trajectories $\omega : [0,\infty) \to V$, endowed with the Skorohod topology (see [15]). For each $x \in V$, let $\mathbb{P}_x$ be the probability measure on $D([0,\infty), V)$ corresponding to the law of the process X_t started from x, and let $\mathbb{E}_x$ denote the associated expectation.

Let $S \subset V$ be a non-empty, proper subset. We define T_t^S as the total amount of time the process spends in S over the interval $[0,t]$:

$$T_t^S := \int_0^t \chi_S(X_s) ds,$$

where χ_S represents the indicator function of the set S:

$$\chi_S(x) = \begin{cases} 1 \text{ if } x \in S \\ 0 \text{ otherwise.} \end{cases}$$

Denote by Q_t^S the generalized inverse of the additive functional T_t^S:

$$Q_t^S := \sup\{s \geq 0 : T_s^S \leq t\}. \tag{4.1}$$

The irreducibility guarantees that for all $t > 0$, Q_t^S is finite almost surely.

Q_t^S corresponds to a 'time change' that only counts the time spent in S. When X_t is outside S, the clock is paused.

The process T_t^S is continuous. It increases linearly when $X_t \in S$, and is constant whenever $X_t \in V \setminus S$.

Definition 31. *The trace of the chain X_t on the set S, denoted by X_t^S, is*

$$X_t^S := X_{Q_t^S}.$$

Taking the trace of the process corresponds to a change in the time scale: when the process reaches $V \setminus S$, time is effectively paused until X_t re-enters S, at which point the clock resumes. In particular, X_t^S takes values in the set S.

It can be proven (see [12]) that X_t^S is an irreducible, continuous-time, S-valued Markov chain. Let $\tau_1 = \inf\{t \geq 0 : X_t \neq X_0\}$ be the time of the first jump of X_t. Define the *hitting time* of a state $y \in V$ as $H_y := \inf\{t \geq 0 : X_t = y\}$ and the *return time* to the set S as $H_S^+ := \inf\{t \geq \tau_1 : X_t \in S\}$. The jump rates of X_t^S are given by:

$$r^S(x,y) := r(x)\mathbb{P}_x[H_S^+ = H_y], \qquad x, y \in S, x \neq y. \tag{4.2}$$

The unique invariant measure $\pi_S(x)$ of X_t^S is the measure π conditioned to S:

$$\pi_S(x) = \frac{\pi(x)}{\pi(S)}, \qquad x \in S. \tag{4.3}$$

Moreover, π_S is reversible if so is π (see again [12]).

5

On the Limiting Invariant Measure of Multiscale Markov Chains

Multiscale Markov chains play a central role in the modeling of stochastic systems whose dynamics evolve on widely separated time scales.

From a theoretical standpoint, multiscale Markov chains allow for a precise description of how long-time behavior arises from local equilibrium within metastable sets. This approach has proven particularly effective in the study of interacting particle systems, statistical mechanics, and stochastic dynamics, where metastability plays a fundamental role. More generally, the analysis of multiscale Markov chains provides a unifying framework for understanding complex stochastic phenomena driven by rare events and scale separation. This is a well-established topic in the literature (see, e.g., [11, 31, 37, 41–44, 47]). Indeed, the recursive structure we obtain in section 5.4 is consistent with those encountered in those contexts.

In particular, this part of the thesis shows the results obtained in [2], concerning the decomposition of the limiting invariant measure of multiscale chains.

5.1 Notation and combinatorial tools

Let $G = (V, E)$ be a finite strongly connected digraph with set of vertices V and set of edges E. To every edge $(x, y) \in E$ we associate a positive rate $r_N(x, y) > 0$ depending on a parameter $N > 0$.

We call X_t^N a continuous time Markov chain on (V, E) and with transition rates r_N. Since for finite N the chain is irreducible, it admits a unique strictly positive invariant measure π_N that is characterized by the stationarity condition

$$\pi_N(x) \sum_{y:(x,y)\in E} r_N(x, y) = \sum_{y:(y,x)\in E} \pi_N(y) r_N(y, x)\,, \qquad \forall x \in V. \tag{5.1}$$

Our aim is to determine the limiting invariant measure

$$\pi := \lim_{N\to\infty} \pi_N \,, \tag{5.2}$$

under suitable assumptions on the rates. Our approach is based on the combinatorial representation of π_N given by the Matrix tree theorem.

We consider the case when the Markov chain X_t^N has a multiscale behavior. This is formalized in the following basic properties, to which additional hypotheses will be added when necessary.

Assumption 4. *We assume that for each pair of edges $(x,y),(w,z) \in E$ the limit*

$$\lim_{N\to+\infty} \frac{r_N(x,y)}{r_N(w,z)} \tag{5.3}$$

exists and is finite, zero or $+\infty$.

This assumption decomposes the set E into disjoint equivalence classes $E = \cup_i E_i$ where the index $i \in \{1,\dots,F\}$ is an integer number. The edges (x,y) and (w,z) belong to the same equivalence class when the limit (5.3) is finite and strictly positive. We label the equivalence classes with natural numbers since they have a natural total order relation: we define $E_j > E_i$ when $j > i$ and this happens when the limit (5.3) is $+\infty$ for every $(x,y) \in E_j$ and $(w,z) \in E_i$. In this case we say that the edges in E_j are "faster" than those in E_i, since the corresponding transition rates are much bigger. The equivalence class E_F contains the rates that are faster than any other and we call the edges in E_F the *fast edges*.

5.1.1 Equivalence classes, DAG and invariant measures

Let us briefly recall the classic equivalence classes in Markov chain theory from a graph-theoretic perspective. Given a digraph $(\mathfrak{V},\mathfrak{E})$ we divide $\mathfrak{V}$ into equivalence classes according to *communication*. In this subsection $(\mathfrak{V},\mathfrak{E})$ is a generic digraph, not necessarily strongly connected. The construction introduced here will be applied in the following to several digraphs, and not necessarily to the original transition graph.

Let us denote by R_N a positive weight associated to each edge of the digraph. We say that x, y belong to the same equivalence class $Q \subseteq \mathfrak{V}$ if there exists a directed path in $(\mathfrak{V},\mathfrak{E})$ both from x to y and from y to x. The subgraph $(Q, \mathfrak{E}[Q])$, where $\mathfrak{E}[Q] \subseteq \mathfrak{E}$ are the edges of $\mathfrak{E}$ that have both extrema in Q, is a strongly connected subgraph. We denote by $(Q, \mathfrak{E}[Q], R_N)$ the weighted digraph where the weight of the edges in $\mathfrak{E}[Q]$ is again R_N. In the theory of Markov chains such equivalence classes are called *irreducible classes*. Let $\mathcal{Q}$ be the collection of all the

equivalence classes and let $(\mathcal{Q}, \mathcal{E})$ be a directed graph such that $(Q, Q') \in \mathcal{E}$ if there exists an $x \in Q$ and a $y \in Q'$ such that $(x, y) \in \mathfrak{E}$. By construction, the digraph $(\mathcal{Q}, \mathcal{E})$ is a *directed acyclic graph* (DAG), i.e. it does not contain directed cycles. A DAG naturally induces a partial order on $\mathcal{Q}$; we say that $Q' \leq Q$ if there exists a directed path from Q to Q' in $(\mathcal{Q}, \mathcal{E})$. Since the DAG is finite, it has at least a local minimum, i.e. a $Q \in \mathcal{Q}$ such that there are not $Q' \in \mathcal{Q}$ different from Q such that $Q' \leq Q$. Since there are no edges in $\mathcal{E}$ exiting from a locally minimal Q, such equivalence classes in the theory of Markov chain are called *closed*, or recurrent. We call $\mathcal{M} \subseteq \mathcal{Q}$ the collection of the equivalence classes corresponding to local minima.

Recall that the equivalence classes M, Q are subsets of $\mathfrak{V}$ while $\mathcal{Q}, \mathcal{M}$ are set of disjoint subsets of $\mathfrak{V}$. We denote by $\mathrm{Su}(\mathcal{M})$ the *support* of $\mathcal{M}$ defined by $\mathrm{Su}(\mathcal{M}) := \cup_{M \in \mathcal{M}} M \subseteq \mathfrak{V}$. A similar notation is used also for $\mathcal{Q}$.

As follows by the classic theory of Markov chains, the set $\mathcal{I}$ of invariant measures of a Markov chain associated to the weighted digraph $(\mathfrak{V}, \mathfrak{E}, R_N)$ is a $|\mathcal{M}|$ dimensional convex subset of $\mathcal{P}(\mathfrak{V})$, the set of probability measures on $\mathfrak{V}$, and in particular it coincides with the simplex

$$\mathcal{I} = \left\{ \sum_{M \in \mathcal{M}} c_M \mu_M^N \,, 0 \leq c_M \leq 1 \,, \sum_{M \in \mathcal{M}} c_M = 1 \right\} , \tag{5.4}$$

where μ_M^N is the unique invariant measure of the irreducible Markov chain associated to the strongly connected weighted graph $(M, \mathfrak{E}[M], R_N)$ with $M \in \mathcal{M}$. Note that each μ_M^N can be naturally interpreted as an element of $\mathcal{P}(M)$ as well as of $\mathcal{P}(\mathfrak{V})$, where we recall that we denote by $\mathcal{P}(S)$ the probability measures on the set S. In particular we have that, for any $\mu \in \mathcal{I}$, $\mu\left(\mathrm{Su}(\mathcal{M})\right) = 1$.

Since it will be used in the following, we prove a simple proposition for the case $|\mathcal{M}| = 1$ that generalizes Theorem 1.2 and that does not seem to be available in the literature.

Proposition 8. *Given a Markov chain with transition graph $(\mathfrak{V}, \mathfrak{E})$ and rates R_N, there exists an unique minimal equivalence class, i.e. $|\mathcal{M}| = 1$, and an unique invariant measure if and only if there exists at least one rooted arborescence in $(\mathfrak{V}, \mathfrak{E})$. In this case the unique invariant measure is given by formula* (1.2).

Proof. If $|\mathcal{M}| = 1$ by (5.4) the invariant measure is unique. Moreover, the acyclic graph $(\mathcal{Q}, \mathcal{E})$ has an unique local minimum M so that from each $Q \neq M$ there is at least one exiting edge in $\mathcal{E}$. For each node of this type, choose arbitrarily one outgoing arrow. By definition the result is necessarily an arborescence in $(\mathcal{Q}, \mathcal{E})$ directed torward M. For each (Q, Q') belonging to such an arborescence, consider

an edge $(x, x') \in \mathfrak{E}$ with $x \in Q$ and $x' \in Q'$, which necessarily exists. We selected in this way one vertex x for each $Q \notin \mathcal{M}$. Consider now, on the strongly connected graph $(Q, \mathfrak{E}[Q])$, a directed arborescence oriented torward x. Considering altogether all the edges, we obtain a directed arborescence in $(\mathfrak{V}, \mathfrak{E})$.

Conversely, if there exists at least one directed arborescence in $(\mathfrak{V}, \mathfrak{E})$, say τ_x, then there can exist at most one locally minimal equivalence class in $(\mathcal{Q}, \mathcal{E})$, namely the class containing x. This follows from the fact that, from any other vertex $y \in \mathfrak{V}$, there exists a directed path leading to x.

Once that the set of arborescences is not empty, then the denominator in (1.2) is not zero and the formula is well defined. Inserting (1.2) into (5.1) (written with the rates R_N), we obtain on both sides $R_N(\mathcal{U}_x)$ where $\mathcal{U}_x$ is the set of spanning unicyclic graphs that contain one single cycle to which x belongs (in some cases, the cycle may be just the vertex x) and from each node outside of the cycle there is one single edge exiting and a directed path torwards the unique cycle. Unless the Markov chain is irreducible, i.e. $M = V$, the invariant measure is not strictly positive. □

The proof of the above proposition implies directly the following result

Corollary 2. *The number of spanning arborescenses in a DAG $(\mathcal{Q}, \mathcal{E})$ such that $|\mathcal{M}| = 1$ is given by $\prod_{Q \in \mathcal{Q}} \left(d_Q^+\right)$, where d_Q^+ is the outgoing degree of the vertex Q.*

Proof. As in the proof of Proposition 8, all the arborescences are obtained selecting arbitrarily an exiting edge for each $Q \notin \mathcal{M}$, and this gives the final result. □

5.2 Assumptions and the limiting invariant measure

We introduce an additional assumption in order to prove our results. This type of assumption was introduced in [42], see also [28]. The following assumptions are imposed on the original weighted graph (V, E, r_N).

Assumption 5. *The rates $(r_N)_{N \in \mathbb{N}}$ are such that for any $n \in \mathbb{N} > 0$, considering two collections of n distinct edges $\{e_1, \dots e_n\}$ and $\{f_1 \dots f_n\}$, the limit*

$$\lim_{N \to \infty} \frac{\prod_i^n r_N(e_i)}{\prod_i^n r_N(f_i)}, \tag{5.5}$$

exists and is either finite and strictly positive, zero or $+\infty$.

A second stronger assumption, exactly the one introduced in [42] and [28], is the following

Assumption 6. *The rates $(r_N)_{N\in\mathbb{N}}$ are such that, considering for each edge $e \in E$ two integer numbers $m_e \geq 0$ and $m'_e \geq 0$ such that $\sum_{e\in E} m_e = \sum_{e\in E} m'_e$, the limit*

$$\lim_{N\to\infty} \frac{\prod_{e\in E} r_N(e)^{m_e}}{\prod_{e\in E} r_N(e)^{m'_e}}, \tag{5.6}$$

exists and is either finite and strictly positive, zero or $+\infty$.

Since any arborescence has the same number of edges, i.e. $|V|-1$, we have that, under assumption 5, for any pair $\tau, \tau' \in \mathcal{T}$ there exists the limit $\lim_{N\to+\infty} \frac{r_N(\tau)}{r_N(\tau')}$ (positive, zero or $+\infty$). This means that, as for the edges, we can divide the arborescences into equivalence classes and we call $\mathcal{T}^F$ the fast one. Note that to identify the fast class of arborescences and the corresponding scale is not an easy task and it is essentially the main aim of this paper. We would like to stress that fast arborescences are not necessarily those consisting solely of fast edges (this occurs only when such arborescences exist), but rather those belonging to the fastest equivalence class of arborescences.

Proposition 9 (Existence of the limiting invariant measure)**.** *Let X_t^N be a Markov chain satisfying Assumption 5, then there exists the limit $\pi := \lim_{N\to\infty} \pi_N$ and it is a probability measure.*

Proof. The fact that π is a probability measure follows by the fact that our graph is finite and we can exchange the limit with the sum. We have

$$\pi_N(x) = \frac{r_N(\mathcal{T}_x)}{r_N(\mathcal{T})} = \frac{r_N(\mathcal{T}_x^F)}{r_N(\mathcal{T}^F)} \frac{\left[1 + \frac{r_N(\mathcal{T}_x \setminus \mathcal{T}_x^F)}{r_N(\mathcal{T}_x^F)}\right]}{\left[1 + \frac{r_N(\mathcal{T} \setminus \mathcal{T}^F)}{r_N(\mathcal{T}^F)}\right]}. \tag{5.7}$$

Here $\mathcal{T}_x^F$ denotes the arborescences rooted at x that belong to the fastest equivalence classe of arborescences having a nontrivial intersection with $\mathcal{T}_x$. We have that $\mathcal{T}_x^F$ is not empty but may not contain arborescences in $\mathcal{T}^F$.

Let $\mathcal{G}, \mathcal{G}^F$ be two collection of arborescences such that $\mathcal{G}^F$ contains at least one fast arborescence. Let $g \in \mathcal{G}$ be one among the fastest arborescences in $\mathcal{G}$ and let $g^* \in \mathcal{G}^F$ be a fast arborescence. We have then

$$\frac{r_N(\mathcal{G})}{r_N(\mathcal{G}^F)} = \frac{r_N(g)}{r_N(g^*)} \frac{\left(\sum_{h\in\mathcal{G}} \frac{r_N(h)}{r_N(g)}\right)}{\left(\sum_{h^*\in\mathcal{G}^F} \frac{r_N(h^*)}{r_N(g^*)}\right)}. \tag{5.8}$$

By Assumption 5 the two terms inside the parenthesis converge to a finite value. Moreover, the term $\frac{r_N(g)}{r_N(g^*)}$ converges to a finite value if g is a fast arborescence, and

to zero otherwise. Therefore, (5.8) is converging to a finite value, that may be zero, when $N \to \infty$.

We can deduce that all the terms $\frac{r_N\left(\mathcal{T}_x \backslash \mathcal{T}_x^F\right)}{r_N\left(\mathcal{T}_x^F\right)}$, $\frac{r_N\left(\mathcal{T} \backslash \mathcal{T}^F\right)}{r_N\left(\mathcal{T}^F\right)}$ and $\frac{r_N\left(\mathcal{T}_x^F\right)}{r_N\left(\mathcal{T}^F\right)}$ in equation (5.7) converge and moreover the limit of the first two is zero while the third one may converge either to zero or to a finite value. Therefore we have

$$\lim_{N\to+\infty} \pi_N(x) = \lim_{N\to+\infty} \frac{r_N\left(\mathcal{T}_x\right)}{r_N\left(\mathcal{T}\right)} = \lim_{N\to+\infty} \frac{r_N\left(\mathcal{T}_x^F\right)}{r_N\left(\mathcal{T}^F\right)}\,, \tag{5.9}$$

and the limit exists and is a probability measure.

□

The argument above shows that, in computing the limit of π_N, only fast arborescences contribute. Consequently, if we can show that a class of arborescences is not fast, it can be neglected in the determination of the limit of π_N.

5.3 The basic iterative step and the main result

Recall that E_F is the equivalence class containing the fastest edges, i.e. the equivalence class of edges E_j with the maximal value of j. We define also (V, E_F) as the *fast subgraph*. The fast subgraph is not necessarily strongly connected and we can construct the DAG of the equivalence classes, as illustrated above; we call $(\mathcal{Q}, \mathcal{E})$ this DAG and we denote by $Q \in \mathcal{Q}$ a generic equivalence class. We recall that Q is an equivalence class with respect to communication using fast edges only. Later on, we will add some upper indices to denote that this is the first step of an iterative procedure. Graphs constructed at iteration number i will have an upper index (i). We call $\mathcal{M}$ the set of local minima of the partial order induced by the DAG and M a generic element of $\mathcal{M}$.

We now proceed in two steps. First, we define a *reduced dynamics* on the vertices of V belonging to the minimal classes of the DAG; this corresponds to the trace process as defined in [42]. Then, using the reduced dynamics, we introduce an *effective dynamics*, which is a Markov chain on the equivalence classes in $\mathcal{M}$.

5.3.1 Reduced dynamics

For notational convenience, let $\hat{V}$ denote the support of $\mathcal{M}$, i.e., the set of all vertices belonging to the minimal classes.

$$\hat{V} := \bigcup_{M\in\mathcal{M}} M = \mathrm{Su}\left(\mathcal{M}\right) \subseteq V\,.$$

It is worth noticing that $\hat{V}$ is the set of recurrent points in the fast graph (V, E_F), while $V \setminus \hat{V}$ is the set of transient points in the fast graph.

Let us introduce the symbol $P_N^S(y \mid z)$, where $S \subset V$ and $y \in S$ while $z \notin S$. This symbol denotes the probability that, starting from the point $z \in V \setminus S$, the process X_t^N enters the set S for the first time at point y:

$$P_N^S(y \mid z) := \mathbb{P}(X_T^N = y \,|\, X_0^N = z) \quad \text{where } T := \inf\{t > 0 \,|\, X_t^N \in S\}\,. \tag{5.10}$$

We define the following reduced rates for a Markov chain restricted to $\hat{V}$.

Definition 32. *For every $x, y \in \hat{V}$ we define the* reduced rates *as*

$$\hat{r}_N(x,y) := r_N(x,y) + \sum_{z \in V \setminus \hat{V}} r_N(x,z) P_N^{\hat{V}}(y|z)\,, \qquad x, y \in \hat{V}\,. \tag{5.11}$$

The absorption probabilities (5.10) have a representation in terms of spanning forests [55] as:

$$P_N^{\hat{V}}(y|z) \;=\; \frac{r_N\big(\mathcal{F}_{\hat{V}}(z \to y)\big)}{r_N(\mathcal{F}_{\hat{V}})} \tag{5.12}$$

where , given $S \subseteq V$, $\mathcal{F}_S$ denotes the collection of directed spanning forests of the digraph (V, E) with $|S|$ components and with root set S. For $z \notin S$ and $y \in S$, we denote by $\mathcal{F}_S(z \to y)$ the subset of forests where the vertex z belongs to the component rooted at y. Recall that the weight of a set of forests, or more generally of any family of subgraphs, is defined by:

$$r_N(\mathcal{F}_{\mathcal{S}}) = \sum_{f \in \mathcal{F}_{\mathcal{S}}} \prod_{(x,y) \in f} r_N(x,y)\,, \tag{5.13}$$

and recall that any forest $f \in \mathcal{F}_S$ has the same number of edges, that is $|V| - |S|$.

Substituting (5.12) into (5.11) yields, by direct inspection,

$$\hat{r}_N(x,y) \;=\; \frac{r_N\big(\mathcal{F}_{\hat{V}\setminus\{x\}}(x \to y)\big)}{r_N(\mathcal{F}_{\hat{V}})}\,. \tag{5.14}$$

It can be seen [42] that $\hat{r}_N$ are the transition rates of a reduced chain, $\hat{X}_t^N$, which coincides with the trace process of X_t^N on $\hat{V}$. By equation 5.11, it is clear that the rates $\hat{r}_N$ are irreducible, since r_N are irreducible rates, and we denote by $\hat{\pi}_N$ the associated invariant measure on $\hat{V}$.

The rates and transition graph of the reduced dynamics can be obtained by successive iterations of the star-delta reduction presented in section 5.8. Performing in sequence a reduction removing one single node in each step and doing this for all

the nodes in $V \setminus \hat{V}$ we obtain the chain with weigtehd transition graph $\left(\hat{V}, \hat{E}, \hat{r}_N\right)$. We now show that Assumption 6 is preserved under the reduction operation. To this end, it suffices to prove this property for a reduction that removes a single node.

Proposition 10. *If r_N satisfies Assumption 6, then so does $\hat{r}_N$. In particular, the reduction procedure* (5.11) *preserves the validity of Assumption 6.*

Proof. Consider two vectors of integer numbers $\underline{m} = (m_e)_{e\in E} \in A$ and $\underline{m}' = (m'_e)_{e\in E} \in B$, where A, B are sets of vectors such that $\sum_{e\in E} m(e) = \sum_{e\in E} m'(e) = K$ for a given fixed constant K. Consider also two positive coefficients $c(\underline{m}), c'(\underline{m}') > 0$. Since we are assuming the validity of (6) there exist $\underline{m}^* \in A$ and $\underline{m}'' \in B$ such that $\lim_{N\to+\infty} \frac{\prod_{e\in E} r_N(e)^{m_e^*}}{\prod_{e\in E} r_N(e)^{m_e}} > 0$ for any $\underline{m} \in A$ and $\lim_{N\to+\infty} \frac{\prod_{e\in E} r_N(e)^{m_e''}}{\prod_{e\in E} r_N(e)^{m'_e}} > 0$ for any $\underline{m}' \in B$. We have therefore

$$\frac{\sum_{\underline{m}\in A} c(\underline{m}) \prod_{e\in E} r(e)^{m_e}}{\sum_{\underline{m}'\in B} c(\underline{m}') \prod_{e\in E} r(e)^{m'_e}} = \left(\frac{c(\underline{m}^*) \prod_{e\in E} r(e)^{m_e^*}}{c(\underline{m}'') \prod_{e\in E} r(e)^{m_e''}}\right) \frac{\sum_{\underline{m}\in A} \left(\frac{c(\underline{m}) \prod_{e\in E} r(e)^{m_e}}{c(\underline{m}^*) \prod_{e\in E} r(e)^{m_e^*}}\right)}{\sum_{\underline{m}'\in B} \left(\frac{c(\underline{m}') \prod_{e\in E} r(e)^{m'_e}}{c(\underline{m}'') \prod_{e\in E} r(e)^{m_e''}}\right)}. \tag{5.15}$$

The first term inside the parentheses on the right-hand side converges, while the second term converges to a finite value. Therefore, their product converges, and consequently the limit of the left-hand side exists as $N \to +\infty$.

The proof is now concluded by observing that, using formula (5.59), an expression of the form $\frac{\prod_{\hat{e}\in\hat{E}} \hat{r}_N(\hat{e})^{\hat{m}_{\hat{e}}}}{\prod_{\hat{e}\in\hat{E}} \hat{r}_N(\hat{e})^{\hat{m}'_{\hat{e}}}}$ for any pair of vectors of integers $\underline{\hat{m}} = (\hat{m}_{\hat{e}})_{\hat{e}\in\hat{E}}$ and $\underline{\hat{m}}' = (\hat{m}'_{\hat{e}})_{\hat{e}\in\hat{E}}$ such that $\sum_{\hat{e}\in\hat{E}} \hat{m}(\hat{e}) = \sum_{\hat{e}\in\hat{E}} \hat{m}'(\hat{e})$ can be written as the left hand side of (5.15).

The result for formula (5.11) follows by performing a finite sequence of single-node removals.

□

5.3.2 Effective dynamics

We now illustrate the first step of an iteration defining the effective rates $\bar{r}_N$ of an irreducible Markov chain on the state space $\mathcal{M}$. These rates are defined using the reduced rates. The subsequent steps of the iteration follow the same procedure, and we will later introduce indices to denote the iteration level. Before giving the formal definition, we need to introduce some notation.

Given $M \in \mathcal{M}$, we call μ_M^N the invariant measure of the chain having weighted transition graph $\left(M, E_F[M], r_N\right)$. This means that the transition graph is the

strongly connected subgraph formed by the fast edges connecting vertices in M, and that the rates coincide with r_N. We call also $\mu_M := \lim_{N\to+\infty} \mu_M^N$, the corresponding limiting measure, that exists by the same argument from Assumption 5. Notice that μ_M is strictly positive, since the latter transition graph is irreducibile.

We are now ready to define the effective rates

Definition 33. *We call* effective rates *the transition rates of an* effective Markov chain $\bar{X}_t^N$ *with state space* $\mathcal{M}$ *that are defined by*

$$\bar{r}_N(M_1, M_2) = \sum_{y_1\in M_1, y_2\in M_2} \mu_{M_1}^N(y_1)\hat{r}_N(y_1, y_2) \qquad \forall M_1, M_2 \in \mathcal{M}. \tag{5.16}$$

We call effective graph *the transition graph* $\left(\mathcal{M}, \bar{E}\right)$ *of this effective Markov chain.*

From 5.16 it is clear that the rates $\bar{r}_N$ are irreducible, as we already noticed that r_N are irreducible and μ_M^N is strictly positive. Moreover, by the same argument as in Proposition 10 if the rates r_N satisfy assumption 6, the effective rates (5.16) also satisfy 6 . The chain $\bar{X}_t^N$ is therefore a multiscale irreducible Markov chain too. If we call $\bar{\pi}_N$ its invariant measure, we have that this converges to a limiting measure, that we denote by $\bar{\pi}$.

We can now state one of our main result.

Theorem 11. *Let* $\pi^N(x)$ *be the unique invariant measure for each finite* $N > 0$ *of a Markov chain on* V *whose transition rates satisfy assumptions 4 and 5. Then we have*

$$\pi = \lim_{N\to\infty} \pi^N = \sum_{M\in\mathcal{M}} \bar{\pi}(M)\mu_M. \tag{5.17}$$

Note that, as a consequence of this theorem, we have that for any $x \in V \setminus \hat{V}$, $\pi(x) = 0$. This is a direct consequence of Lemma 7 in the following.

5.4 The recursion and the forest representation

We now illustrate how (5.17) defines a recursive structure that is geometrically encoded by a tree. Since the construction is based on an iteration of the procedure just described, we change the notation in this section to clarify the iterative process.

5.4.1 The iterative notation

The construction presented in this section coincides with that introduced in [13, 13, 40, 42, 46] for different problems. In the present setting, however, we apply

it to the invariant measure and provide an independent proof based on directed arborescences and forests.

We label by (0) the initial transition graph (V, E), which in this section is denoted by $\left(\mathcal{M}^{(0)}, \bar{E}^{(0)}\right)$, and we denote by $\bar{r}_N^{(0)}$ the original rates r_N. Similarly, we label by (1) the transition graph obtained through the construction of the effective dynamics described in the previous section. Accordingly, $\mathcal{M}^{(1)}$ corresponds to what was previously denoted by $\mathcal{M}$, and $\bar{E}^{(1)}$ to what was previously denoted by $\bar{E}$. The resulting effective graph is therefore $\left(\mathcal{M}^{(1)}, \bar{E}^{(1)}\right)$. The associated transition rates, obtained after the first iteration of the procedure, are denoted by $\bar{r}_N^{(1)}$.

To summarize, we have the following equivalences:

$$\begin{cases} \left(\mathcal{M}^{(0)}, \bar{E}^{(0)}, \bar{r}_N^{(0)}\right) = (V, E, r_N) \ , \\ \left(\mathcal{M}^{(1)}, \bar{E}^{(1)}, \bar{r}_N^{(1)}\right) = \left(\mathcal{M}, \bar{E}, \bar{r}_N\right) \ . \end{cases}$$

We are now going to define for each integer (i) a weighted transition graph $\left(\mathcal{M}^{(i)}, \bar{E}^{(i)}\bar{r}_N^{(i)}\right)$ that is obtained from $\left(\mathcal{M}^{(i-1)}, \bar{E}^{(i-1)}, \bar{r}_N^{(i-1)}\right)$ by the same construction that gave $\left(\mathcal{M}^{(1)}, \bar{E}^{(1)}, \bar{r}_N^{(1)}\right)$ from $\left(\mathcal{M}^{(0)}, \bar{E}^{(0)}, \bar{r}_N^{(0)}\right)$.

This means that the Markov dynamics at iteration $(i+1)$ is obtained as an effective dynamics on the local minima of the DAG associated with fast communication in the dynamics at iteration (i). By the same arguments used for the first iteration, we conclude that if the rates $\bar{r}_N^{(i)}$ satisfy Assumption 6, then also the rates $\bar{r}_N^{(i+1)}$ satisfy Assumption 6. For each (i), we call $\bar{r}_N^{(i),F}$ the fast rates and $\mathcal{M}^{(i+1)}$ the local minimal equivalence classes with respect to communication on $\left(\mathcal{M}^{(i)}, \bar{E}^{(i),F}\right)$.

The elements of $\mathcal{M}^{(0)}$ are elements of V, while the elements of $\mathcal{M}^{(1)}$ are subsets of V. Given two different elements $M_1^{(1)}, M_2^{(1)} \in \mathcal{M}^{(1)}$, they are subsets of V and moreover disjoint. Likewise, elements of $\mathcal{M}^{(2)}$ are subsets of $\mathcal{M}^{(1)}$, so that an element $M^{(2)} \in \mathcal{M}^{(2)}$ is given by $M^{(2)} = \left\{M_i^{(1)}\right\}$, where $M_i^{(1)}$ are disjoint substets of V. We may then associate to $M^{(2)}$ a subset of V by defining $\mathrm{Su}\left(M^{(2)}\right) := \cup_i M_i^{(1)} \subseteq V$. By construction we have that $\mathrm{Su}\left(M_i^{(2)}\right) \cap \mathrm{Su}\left(M_j^{(2)}\right) = \emptyset$, when $M_i^{(2)} \neq M_j^{(2)}$. This construction can be iterated at every level. In particular, each element $M^{(i)} \in \mathcal{M}^{(i}$ is a subset of $\mathcal{M}^{(i-1)}$; different elements are disjoint subsets. Moreover, we can associate to any M^i a subset of V defined by $\mathrm{Su}\left(M^{(i)}\right) := \cup_{M^{(i-1)} \in M^{(i)}} \mathrm{Su}\left(M^{(i-1)}\right) \subseteq V$.

For each (i) we denote by $\bar{\pi}_N^{(i)} \in \mathcal{P}\left(\mathcal{M}^{(i)}\right)$ the invariant measure of the chain

with transition rates $\bar{r}_N^{(i)}$, and by $\bar{\pi}^{(i)} \in \mathcal{P}\left(\mathcal{M}^{(i)}\right)$ the corresponding limiting measure.

Given $M^{(i+1)} \in \mathcal{M}^{(i+1)}$ with $i \geq 0$, we call $\mu_{M^{(i+1)}}^{(i),N} \in \mathcal{P}\left(M^{(i+1)}\right)$ the invariant measure of the irreducible Markov chain with weigthed transition graph $\left(M^{(i+1)}, \bar{E}_N^{(i),F}\left[M^{(i+1)}\right], \bar{r}_N^{(i),F}\right)$. We remark once again that $\mu_{M^{(i+1)}}^{(i),N}$ is a positive probability measure on $\mathcal{M}^{(i)}$, whose support is concentrated on the elements belonging to the equivalence class $M^{(i+1)} \in \mathcal{M}^{(i+1)}$. We call $\mu_{M^{(i+1)}}^{(i)}$ the limiting value of $\mu_{M^{(i+1)}}^{(i),N}$.

5.4.2 The recursive formula

Using the iterative notation, formula (5.17) reads

$$\bar{\pi}^{(0)} = \lim_{N\to\infty} \bar{\pi}_N^{(0)} = \sum_{M^{(1)}\in\mathcal{M}^{(1)}} \bar{\pi}^{(1)}\left(M^{(1)}\right) \mu_{M^{(1)}}^{(0)} \,. \tag{5.18}$$

Since the relation between the reduced chains at iterations (i) and $(i+1)$ is always the same, and Assumption 6 is always satisfied, the proof of Theorem 11 which yields formula (5.18), can be applied at any iteration (i). This then gives the corresponding recursive formula.

$$\bar{\pi}^{(i)} = \sum_{M^{(i+1)}\in\mathcal{M}^{(i+1)}} \bar{\pi}^{(i+1)}\left(M^{(i+1)}\right) \mu_{M^{(i+1)}}^{(i)} \,, \qquad i \geq 0 \,. \tag{5.19}$$

The number of possible iterations is finite; let us denote it by k. The iterative procedure stops at the first iteration (i) for which $\left|\mathcal{M}^{(i)}\right| = 1$, so that there is a single element $M^{(i)} \in \mathcal{M}^{(i)}$ with $\bar{\pi}_N^{(i)}\left(M^{(i)}\right) = 1$. The number of iteration k is therefore defined as $k = \inf\left\{i \,:\, |\mathcal{M}^{(i)}| = 1\right\}$. The iteration can also be formally extended to higher indices, but we then have $\left|\mathcal{M}^{(j)}\right| = 1$ for all $j \geq k$.

5.4.3 The forest

In this subsection, using formula (5.19), we are going to obtain a representation formula for the limiting measure π, which in this section is denoted by $\bar{\pi}^{(0)}$, in terms of weighted forests. We begin by defining the forest and the associated weights, and then explain how the measure can be derived from them.

For each index (i) with $i \geq 1$, we denote by $\mathcal{Q}^{(i)}$ the equivalence classes in $\mathcal{M}^{(i-1)}$ with respect to fast communication, determined by the fast rates $\bar{r}_N^{(i-1)}$.

The nodes of the forest are the elements of $\cup_{j=0}^{k}\mathcal{M}^{(j)}$. These nodes can be naturally organized into levels: the nodes at level j are precisely the elements of $\mathcal{M}^{(j)}$. The forest is then constructed by adding edges between nodes.

Edges connect only nodes belonging to consecutive levels. More precisely, an edge is drawn from $M^{(i)} \in \mathcal{M}^{(i)}$ to $M^{(i+1)} \in \mathcal{M}^{(i+1)}$ if $M^{(i)} \in M^{(i+1)}$, or equivalently, if $\mathrm{Su}\left(M^{(i)}\right) \subseteq \mathrm{Su}\left(M^{(i+1)}\right)$.

In this way, for each node M_J belonging to $\mathcal{M}^{(j)}$, $j \neq k$, if $M^{(j)} \in M^{(j+1)}$ for some $M^{(j+1)} \in \mathcal{M}^{(j+1)}$, then there is exactly one outgoing edge to a node at the next higher level. Otherwise, if $M^{(j)} \in Q^{(j+1)}$ for some $Q^{(j+1)} \in \mathcal{Q}^{(j+1)} \setminus \mathcal{M}^{(j+1)}$, there is no edge to a higher-level node. Similarly, there are no outgoing edges from the unique element in $\mathcal{M}^{(k)}$, since there are no nodes at a higher level.

This construction thus produces a directed forest, with edges oriented toward higher levels. For each connected component, the unique node at the highest level is called the *root*. In particular, there is a single component whose root is $M^{(k)}$.

The forest is weighted: we assign a weight to each edge as follows. To the edge $\left(M^{(i+1)}, M^{(i)}\right)$ we associate the weight $\bar{\mu}^{(i)}_{M^{(i+1)}}\left(M^{(i)}\right)$. Note that for each node, the sum of the weights of all edges connecting it to nodes at the lower level is equal to one. This follows from the fact that $\sum_{M^{(i)} \in M^{(i+1)}} \bar{\mu}^{(i)}_{M^{(i+1)}}\left(M^{(i)}\right) = 1$.

Since all iterative steps are identical and the rates satisfy Assumption 6 at each iteration, Lemma 7 implies that $\bar{\pi}^{(i)}\left(Q^{(i)}\right) = 0$ for any $Q^{(i)} \notin \mathcal{M}^{(i)}$.

Consequently, $\bar{\pi}^{(0)}_{N}\left(M^{(0)}\right) = 0$ for any $M^{(0)}$ such that $\mathrm{Su}\left(M^{(0)}\right) \not\subseteq \mathrm{Su}\left(M^{(k)}\right)$. Moreover, by definition, $\bar{\pi}^{(k)}\left(M^{(k)}\right) = 1$ for the unique element of $\mathcal{M}^{(k)}$. It follows that the recursion (5.19) has a unique solution, which provides a formula for the limiting invariant measure $\pi = \bar{\pi}^{(0)}$.

Given $M^{(0)} \in \mathcal{M}^{(0)}$, there exists a unique path connecting this node to the root of the component to which it belongs. If this root is not $M^{(k)}$, then $\bar{\pi}^{(0)}\left(M^{(0)}\right) = 0$. If the root of the component containing $M^{(0)}$ is $M^{(k)}$, which occurs, for example, when $\mathrm{Su}\left(M^{(0)}\right) \subseteq \mathrm{Su}\left(M^{(k)}\right)$, we call $\gamma_{M^{(0)}}$ the unique path from $M^{(0)}$ to the root in the forest we have constructed. We then have the following formula.

Corollary 3. *The unique solution of the recursion in Theorem 11 is given by*

$$\bar{\pi}^{(0)}\left(M^{(0)}\right) = \begin{cases} \prod_{\left(M^{(i)}, M^{(i+1)}\right) \in \gamma_{M^{(0)}}} \mu^{(i)}_{M^{(i+1)}}\left(M^{(i)}\right) & Su\left(M^{(0)}\right) \subseteq Su\left(M^{(k)}\right) \\ 0 & Su\left(M^{(0)}\right) \cap Su\left(M^{(k)}\right) = \emptyset\,. \end{cases} \tag{5.20}$$

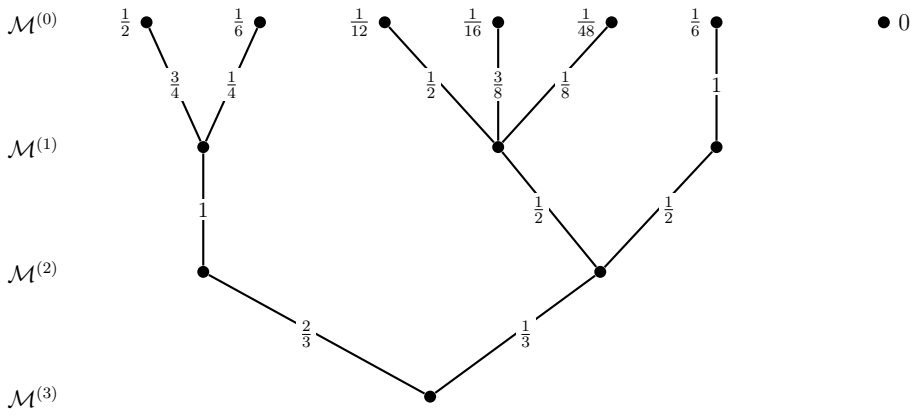


Figure 5.1: An example of a forest with two components and $k = 3$. The corresponding weight is written near each edge. Above the elements in $\mathcal{M}^{(0)}$, the values of $\bar{\pi}^{(0)}$, are shown according to formula (5.20).

.

Using lemma 7 the proof of the above corollary is simple and consist just in iterating the recursion. We illustrate in Figure 5.1 the construction and the formula.

5.5 Some estimates

In this section we establish some estimates that will be needed later.

Lemma 7. *It holds that* $\lim_{N\to+\infty} \pi_N \left(V \setminus \hat{V}\right) = 0$.

Proof. By the Markov chain tree theorem, we need to prove

$$\lim_{N\to\infty} \frac{\sum_{z\in V\setminus\hat{V}} r_N\left(\mathcal{T}_z\right)}{r_N\left(\mathcal{T}\right)} = 0\,. \tag{5.21}$$

Since we have a finite sum, this follows by

$$\lim_{N\to\infty} \frac{r_N\left(\mathcal{T}_z\right)}{r_N\left(\mathcal{T}\right)} = 0\,, \qquad \forall\ z \in V \setminus \hat{V}\,,$$

and again, since we have a finite sum, it is enough to prove

$$\lim_{N\to\infty} \frac{r_N\left(\tau_z\right)}{r_N\left(\mathcal{T}\right)} = 0\,, \qquad \forall\ z \in V \setminus \hat{V}\,,\ \forall \tau_z\,.$$

For any $z \in Q$, with $Q \in \mathcal{Q} \setminus \mathcal{M}$, there exists by definition at least one edge, denoted by (z, z'), exiting from z and belonging to E_F. For any arborescence $\tau_z \in \mathcal{T}_z$, the graph $\tau_z \cup (z, z')$ is a spanning unicyclic graph.

If all the edges of the unique cycle belong to E_F, we remove one of them, thereby obtaining an arborescence, and iterate this procedure until the first time that we reach an arborescence directed torward a vertex $w \in Q$ from which there exists an edge $(w, w') \in E_F$ exiting from w, and such that w' and w belong to different equivalent classes in $\mathcal{Q}$.

This is possible because the procedure outlined above allows one to transform any arborescence into any other within a strongly connected digraph (which in this case is (Q, E_F)), and because for any $Q \notin \mathcal{M}$ there must exist an edge $(w, w') \in E_F$ such that $w \in Q$ and $w' \notin Q$.

When such an edge (w, w') is added to the arborescence, the resulting unique cycle must contain at least one edge that does not belong to to E_F (by definition of the equivalence classes). Removing this edge, we obtain a new arborescence τ_{z^*} with the following properties: if the original arborescence τ_z contains n_i edges on E_i, for $i = 1, \dots, F$ then τ_{z^*} contains $n_F + 1$ fast edges, while all the other n_i remain unchanged except for a single index i^*, for which there are $n_{i^*} - 1$ edges in E_{i^*}. As a consequence

$$\frac{r_N(\tau_z)}{r_n(\mathcal{T})} \leq \frac{r_N(\tau_z)}{r_N(\tau_{z^*})} \xrightarrow[N\to\infty]{} 0\,,$$

and the convergence is due to the fact that in the ratio we can pair in numerator and denominator edges belonging to the same equivalence classes (giving a finite value in the limit) with an extra fast edge in the denominator paired with a slowed one in the numerator getting zero in the limit.

□

The above lemma says that for any $z \in V \setminus \hat{V}$ we have $\mathcal{T}_z \cap \mathcal{T}^F = \emptyset$. Therefore, all arborescences of this type can be neglected in the computation of the limit of π_N. We now consider another class.

Lemma 8. *Let $\tau \in \mathcal{T}_x$ with $x \in \hat{V}$ and in particular $x \in M \in \mathcal{M}$. Then, if τ contains edges of the type (z, z'), with $z \in M$ and $z' \notin M$, then $\tau \notin \mathcal{T}^F$. Moreover, if, for any $M' \in \mathcal{M}$ with $M' \neq M$, τ contains more than one single edge of the type (z, z'), with $z \in M'$ and $z' \notin M'$, then $\tau \notin \mathcal{T}^F$. Finally, if for any $M'' \in \mathcal{M}$ there exists an edge (z, z') in τ with $z, z' \in M''$ and $(z, z') \notin E_F$, then $\tau \notin \mathcal{T}^F$.*

Proof. To prove the statement, it suffices to show that, given an arborescence τ having at least one of the features described above, it is possible to construct another arborescence τ' such that

$$\lim_{N\to+\infty} \frac{r_N(\tau)}{r_N(\tau')} = 0\,. \tag{5.22}$$

For simplicity we consider the case when there are two edges of the type (z, z'), with $z \in M'$ and $z' \notin M'$ and $M' \neq M$. The other cases can be discussed similarly.

Recall that, since $M' \in \mathcal{M}$, the outgoing edges are necessarily slow, i.e. they do not belong to E_F.

Since the root $x \notin M'$, and since there are only two edges, denoted by (z, z') and (w, w'), exiting from M', this means that $\tau \cap (M', E[M'])$ is a spanning forest of $\big(M', E\,[M']\big)$ with two components, one rooted at z and the other one rooted at w.

Since $\big(M', E_F\,[M']\big)$ is strongly connected we can for example remove the slow edge (z, z') and all the edges of the component rooted at z, and consider instead only edges in $E_F\,[M']$ in such a way that we have only one spanning arborescence rooted at w. This can be shown easily since $\big(M', E_F\,[M']\big)$ is strongly connected. The global directed graph obtained in this way is again a $\tau' \in \mathcal{T}_x$. Moreover, (5.22) holds since we remove a collection of edges, among which at least one is slow, and replace them with the same number of edges, all of which are fast. A similar argument applies to prove the remaining statements of the lemma.

□

5.6 First iteration

Let us fix some notation:

- Given $M \in \mathcal{M}$, we call $\mathcal{T}\left[M, E^F\right]$ the spanning arborescences of $\big(M, E_F\,[M]\big)$, and by $\mathcal{T}_x\left[M, E^F\right]$ the ones rooted at $x \in M$. By the above discussion, they are the same of $\mathcal{T}^F\,[M]$ and $\mathcal{T}_x^F\,[M]$, the fast arborescences restricted to the digraphs $\big(M, E\,[M]\big)$.
- We call $\hat{T}^F$ the fast spanning arborescences on $\left(\hat{V}, \hat{E}\right)$, and likewise we denote by $\hat{T}_x^F$ those rooted at $x \in \hat{V}$.
- We call $\bar{\mathcal{T}}$ the spanning arborescences of the effective rates on $\big(\mathcal{M}, \bar{E}\big)$ and $\bar{\mathcal{T}}_M$ the ones rooted at M.

The following lemma is established in [42]. We complement it with a combinatorial characterization.

Lemma 9. *The invariant measure $\hat{\pi}_N$ of the reduced Markov chain with rates $\hat{r}_N$ in* (5.11), (5.13) *is given by*

$$\hat{\pi}_N(x) = \frac{\pi_N(x)}{c_N} \qquad x \in \hat{V}\,, \tag{5.23}$$

where the constant $c_N = \sum_{z \in \hat{V}} \pi_N(z)$ is such that $\lim_{N \to +\infty} c_N = 1$.

There are several interesting combinatorial interpretations of the constant c_N; see for example Proposition 5.70.

Proof. Recall that we denote by $\hat{\mathcal{T}}_y$ the set of spanning arborescences of the reduced graph oriented towards $y \in \hat{V}$. Formula (5.23) follows by the Ergodic Theorem.

Indeed, if we consider $X_t^N \in V$ and we set $\hat{X}_t^N = X_{T(t)}^N \in \hat{V}$, where $T(t) = |\{s \leq t : X_t^N \in \hat{V}\}|$, by Ergodic Theorem we write for almost all the trajectories X^N:

$$\pi^N(x) = \lim_{t\to\infty} \frac{\int_0^t \mathbb{I}(X_s = x)ds}{t}\,. \tag{5.24}$$

Again, by Ergodic Theorem and since $T(t) \to \infty$ as $t \to \infty$, we have a.e.

$$\hat{\pi}^N(x) = \lim_{t\to\infty} \frac{\int_0^{T(t)} \mathbb{I}(\hat{X}_{T(s)}^N = x)ds}{T(t)}\,, \tag{5.25}$$

by multiplying both numerator and denominator by t, we therefore have:

$$\hat{\pi}^N(x) = \lim_{t\to\infty} \frac{\int_0^t \mathbb{I}(X_s^N = x)ds}{t}\frac{t}{T(t)} = \pi^N(x)\left(\pi^N\left(\hat{V}\right)\right)^{-1},$$

and this is (5.23).

Finally, by lemma 7, we have that

$$c_N = \sum_{y\in\hat{V}} \pi^N(y) = \frac{\sum_{y\in\hat{V}} r_N(\mathcal{T}_y)}{r_N(\mathcal{T})} \xrightarrow[N\to\infty]{} 1\,.$$

□

5.6.1 Proof of Theorem 11

We can now prove our main result.

a) If $x \in V \setminus \hat{V}$, by Lemma 7, we have that $\pi(x) = 0$

b) Let now $x \in \hat{V}$. By lemma 9, we have

$$\pi(x) = \lim_{N\to+\infty} \frac{\hat{r}_N\left(\hat{\mathcal{T}}_x\right)}{\hat{r}_N\left(\hat{\mathcal{T}}\right)}\,, \qquad x \in \hat{V}\,. \tag{5.26}$$

Since we can restrict the computation to fast arborescences, we can consider in the numerator of (5.26), when $x \in M$,

$$\hat{r}_N\left(\hat{\mathcal{T}}_x\right) = \left\{ r_N\left(\mathcal{T}_x^F\left[M\right]\right) \sum_{\bar{\tau}\in\bar{\mathcal{T}}_M} \prod_{(M_i,M_j)\in\bar{\tau}} \left[\sum_{\substack{y_i\in M_i \\ y_j\in M_j}} \hat{r}_N(y_i,y_j) r_N\left(\mathcal{T}_{y_i}^F\left[M_i\right]\right) \right] \right\} (1+o_N(1)) \tag{5.27}$$

where we recall that $\bar{\mathcal{T}}_{M_i}$ denotes the set of arborescences on the effective graph oriented towards M_i. Note that in the limit, all the negligible-weight arborescences highlighted in the lemmas 8 and 9 can be disregarded.

Before proceeding, let us observe that, given two minimal classes M_i, M_j, by the definition of effective rates (5.16) and by matrix tree Theorem 1.2,

$$\bar{r}_N\left(M_i,M_j\right) r_N\left(\mathcal{T}^F\left[M_i\right]\right) = \sum_{y_i\in M_i, y_j\in M_j} \hat{r}_N(y_i,y_j) r_N(\mathcal{T}_{y_i}^F\left[M_i\right]). \tag{5.28}$$

By equation (5.28), we can rewrite (5.27) as:

$$\hat{r}_N\left(\hat{\mathcal{T}}_x\right) = \left\{ r_N\left(\mathcal{T}_x^F\left[M\right]\right) \sum_{\bar{\tau}\in\bar{\mathcal{T}}_M} \prod_{(M_i,M_j)\in\bar{\tau}} \left[\bar{r}_N\left(M_i,M_j\right) r_N\left(\mathcal{T}^F\left[M_i\right]\right) \right] \right\} (1+o_N(1)) \tag{5.29}$$

and since in every arborescence $\bar{\tau}\in\bar{\mathcal{T}}_M$ there is exactly one effective edge going out from each class $M_i \neq M$, (5.29) rewrites:

$$\hat{r}_N\left(\hat{\mathcal{T}}_x\right) = r_N\left(\mathcal{T}_x^F\left[M\right]\right) \bar{r}_N\left(\bar{\mathcal{T}}_M\right) \prod_{M_i\neq M} r_N\left(\mathcal{T}^F\left[M_i\right]\right) (1+o(1)) \tag{5.30}$$

By applying the same approach to the denominator, equation (5.27) then becomes

$$\pi_N(x) = \frac{r_N\left(\mathcal{T}_x^F\left[M\right]\right) \bar{r}_N\left(\bar{\mathcal{T}}_M\right) \prod_{M_i\neq M} r_N\left(\mathcal{T}^F\left[M_i\right]\right)}{\bar{r}_N\left(\bar{\mathcal{T}}\right) \prod_{M'\in\mathcal{M}} r_N\left(\mathcal{T}^F[M']\right)} (1+o_N(1)). \tag{5.31}$$

Now, dividing both the numerator and the denominator by $\prod_{M'\in\mathcal{M}} r_N\left(\mathcal{T}^F[M']\right)$ and using Theorem 1.2, we obtain

$$\pi_N(x) = \frac{\mu_N^M(x)\bar{r}_N\left(\bar{\mathcal{T}}_M\right)}{\bar{r}_N\left(\bar{\mathcal{T}}\right)} (1+o_N(1)) = \mu_N^M(x)\bar{\pi}_N\left(M\right)(1+o_N(1)) \tag{5.32}$$

Finally, by Assumption 6, the limit exist and we have

$$\pi(x) = \bar{\pi}\left(M\right)\mu^M(x). \tag{5.33}$$

5.7 Examples

In this section, some concrete models are presented with the aim of illustrating the application of the main theorems discussed in this work. Although an exhaustive analysis of the phenomenological consequences in explicit models is intended for future dedicated research, these examples serve to demonstrate how the novel insights provided by our theorems allow for a systematic approach to complex problems.

Beyond the cases examined here, the theoretical framework developed in this work may find fruitful application in other explicit models driven by rare events, where for instance the jump rates are intrinsically multiscale.

5.7.1 The case $k = 1$

The simplest possible case is when $k = 1$. This corresponds to the case when the DAG constructed by the equivalence classes of $\left(\mathcal{M}^{(0)}, \bar{E}_F^{(0)}\right)$ has a unique minimal class. In this case, by Proposition 8 there exists at least an arborescence that has all fast edges. A fast arborescence of this type can only be rooted at some $x \in M^{(1)}$, which is the unique element of $\mathcal{M}^{(1)}$. In computing the limiting value of $\bar{\pi}_N^{(0)}$, we can use the Markov chain tree theorem considering only the arborescences whose edges are all fast. For such an arborescence, restricting it to $\left(M^{(1)}, \bar{E}_F^{(0)}[M^{(1)}]\right)$ yields a spanning arborescence of this subgraph; otherwise, the original arborescence could not have been fast. We therefore have

$$\begin{cases} \bar{r}_N^{(0)}\left(\mathcal{T}_x^F\right) = \bar{r}_N^{(0)}\left(\mathcal{T}_x^F\left[M^{(1)}\right]\right)\bar{r}_N^{(0)}(\bar{\mathcal{F}}^F_{M^{(1)}})\,, \\ \bar{r}_N^{(0)}\left(\mathcal{T}^F\right) = \bar{r}_N^{(0)}\left(\mathcal{T}^F\left[M^{(1)}\right]\right)\bar{r}_N^{(0)}(\bar{\mathcal{F}}^F_{M^{(1)}})\,, \end{cases} \tag{5.34}$$

where $\bar{r}_N^{(0)}(\bar{\mathcal{F}}^F_{M^{(1)}})$ is a common factor related to the sum of all the weight of fast spanning forests directed torward $M^{(1)}$. As a result, we obtain

$$\bar{\pi}^{(0)}(x) = \lim_{N\to+\infty} \frac{\bar{r}_N^{(0)}\left(\mathcal{T}_x^F\left[M^{(1)}\right]\right)\bar{r}_N^{(0)}(\bar{\mathcal{F}}^F_{M^{(1)}})}{\bar{r}_N^{(0)}\left(\mathcal{T}^F\left[M^{(1)}\right]\right)\bar{r}_N^{(0)}(\bar{\mathcal{F}}^F_{M^{(1)}})} = \mu^{(0)}_{M^{(1)}}(x)\,. \tag{5.35}$$

5.7.2 A simple case with $k = 2$

Let us start with a simple example where all the computations can be done explicitely and $k = 2$, so that the forest construction only has 3 levels. We consider the Markov chain associated with the graph (V, E) in figure 5.2 with $V = \{x, y, z\}$.

There are only two equivalence classes of edges E_1, E_2 with E_1 the slow ones and E_2 the fast ones. In the figure, fast edges are drawn as straight lines, while slow edges are represented with wavy lines. More precisely, the rates are defined as follows: consider a function $c : V \times V \to (0, \infty)$ and define:

$$\begin{aligned} r_N(z,x) &= c(z,x), \\ r_N(z,y) &= c(z,y), \\ r_N(y,x) &= \frac{1}{N} c(y,x), \\ r_N(x,z) &= \frac{1}{N} c(x,z). \end{aligned}$$

so that the fast rates are $O_N(1)$, while the slow ones are $O(1/N)$. The graph is so simple that the Markov chain tree theorem can be applied directly. We have

$$\begin{cases} r_N\left(\mathcal{T}_x\right) = \frac{1}{N} c(y,x)[c(z,x) + c(z,y)], \\ r_N\left(\mathcal{T}_y\right) = \frac{1}{N} c(x,z) c(z,y), \\ r_N\left(\mathcal{T}_z\right) = \frac{1}{N^2} c(y,x) c(x,z). \end{cases}$$

We have for the state x, and likewise for the other ones

$$\pi^N(x) = \frac{r_N(\mathcal{T}_x)}{r_N(\mathcal{T}_x) + r_N(\mathcal{T}_y) + r_N(\mathcal{T}_z)} \tag{5.36}$$

$$= \frac{\frac{1}{N} c(y,x)[c(z,x) + c(z,y)]}{\frac{1}{N} c(y,x)[c(z,x) + c(z,y)] + \frac{1}{N} c(x,z) c(z,y) + \frac{1}{N^2} c(y,x) c(x,z)}, \tag{5.37}$$

and therefore

$$\pi(x) = \lim_{N \to \infty} \pi^N(x) = \frac{c(y,x)[c(z,x) + c(z,y)]}{c(y,x)[c(z,x) + c(z,y)] + c(x,z) c(z,y)}. \tag{5.38}$$

As shown in Figure 5.2, the recurrent nodes of (V, E_F) are $\hat{V} = \{x, y\}$ and the reduced rates can be computed directly:

$$\hat{r}_N(x,y) = r_N(x,z) P^{\{x,y\}}(y|z) = \frac{1}{N} c(x,z) \frac{c(z,y)}{c(z,x) + c(z,y)} \tag{5.39}$$

$$\hat{r}_N(y,x) = r_N(y,x) = \frac{1}{N} c(y,x). \tag{5.40}$$

To construct the effective dynamics, we observe that $\mathcal{Q}^{(1)} = \big\{\{x\}, \{y\}, \{z\}\big\}$ and $\mathcal{M}^{(1)} = \big\{\{x\}, \{y\}\big\}$, since every equivalence class consists of a single point. Accordingly, $\mu^N_{\{x\}}$ and $\mu^N_{\{y\}}$ are delta measures concentrated at x and y, respectively. We then have $\bar{r}_N\left(\{x\}, \{y\}\right) = \hat{r}_N(x,y)$ and $\bar{r}_N\left(\{y\}, \{x\}\right) = \hat{r}_N(y,x)$

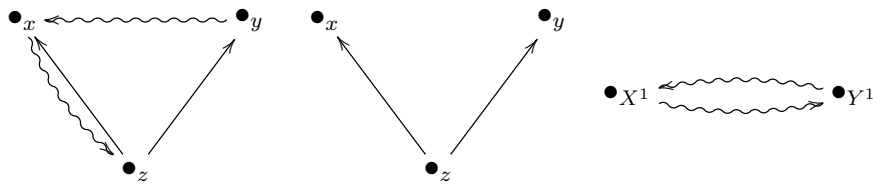


Figure 5.2: The graph (V, E), the fast digraph (V, E_F) (that in this case coincides with the DAG) and the first iteration effective graph $\left(\hat{V}, \hat{E}\right)$. Rates $O_N(1)$ are drawn by straight lines while rates $O(1/N)$ are drawn with wavy lines.

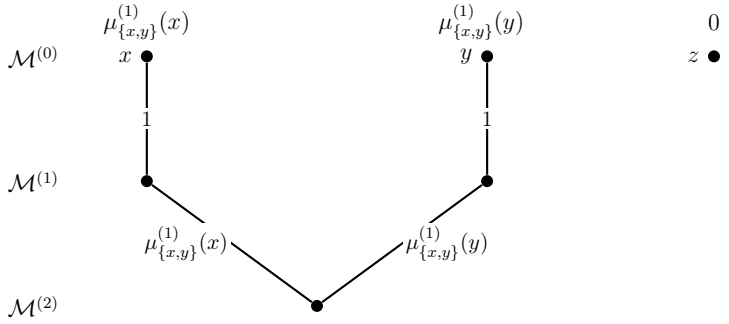


Figure 5.3: The weighted forest with 2 components and $k = 2$ associated to the example in section 5.7.2

.

All the effective rates are at the same scale, so that $\left(\bar{V}, \bar{E}^F\right)$ is irrecucible, $|\mathcal{M}^{(2)}| = 1$ and contains the only element $M^{(2)} = \{x\} \cup \{y\}$. The probability measure $\mu^{(1,N)}_{\{x,y\}} \in \mathcal{P}(\{x\} \cup \{y\})$ can be explicitely computed from the invariant measure of the chain with rates (5.39) (this is because in this special case $\bar{r} = \hat{r}$), and therefore we have

$$\mu_{M^{(2)}}(\{x\}) = \mu^{(1,N)}_{M^{(2)}}(\{x\}) = \frac{c(y,x)[c(z,x) + c(z,y)]}{c(y,x)[c(z,x) + c(z,y)] + c(x,z)c(z,y)}\,,$$

and $\mu^{(1,N)}_{M^{(2)}}(\{y\}) = 1 - \mu^{(1,N)}_{M^{(2)}}(\{x\})$.

So, as $N \to \infty$, we have

$$\pi(x) = \mu^{(1)}_{\{x,y\}}(\{x\})$$

that is the result obtained directly in (5.38).

5.7.3 A product state space

We discuss a class of models that, in special cases, can be imagined as discrete counterparts of coupled diffusions.

Consider the finite vertex set $V = \mathcal{X} \times \mathcal{Y}$. The rates are related to a symmetric edge-weight function $r_N : (\mathcal{X} \times \mathcal{Y})^2 \to [0, \infty)$. The parameter $N \gg 1$ controls the weights, giving rise to two distinct time scales:

$$r_N\big((x,y),\,(x',y')\big) \;=\; \begin{cases} w_y(x,x') & \text{if } x \neq x' \text{ and } y = y' \\ \frac{1}{N}\, w_x(y,y') & \text{if } x = x' \text{ and } y \neq y' \\ 0 & \text{if } x \neq x' \text{ and } y \neq y' \end{cases} \tag{5.41}$$

Transitions in the x-direction are *fast*, while transitions in the y-direction are *slow*.

We assume that the full graph $(\mathcal{X} \times \mathcal{Y},\, r_N)$ is *irreducible* for every $N \geq 1$, hence the associated Markov process has a unique invariant measure π_N.

In order to underline a parallel with the continuous case in [1], we consider the case in which, for every choice of $y \in \mathcal{Y}$, the graph $(\mathcal{X}, w_y)$ is *irreducible* and the associated Markov process is *reversible* with respect to the unique invariant measure π_y:

$$\pi_y(x)\;\; w_y(x,x') \;=\; \pi_y(x')\;\; w_y(x',x) \tag{5.42}$$

for all $x, x' \in \mathcal{X}$.

Similarly, we assume that for every choice of $x \in \mathcal{X}$ the graph $(\mathcal{Y},\, w_x)$ is *irreducible* and the associated Markov process is *reversible* with respect to the unique invariant measure π_x:

$$\pi_x(y)\;\; w_x(y,y') \;=\; \pi_x(y')\;\; w_x(y',y) \tag{5.43}$$

for all $y, y' \in \mathcal{Y}$. Note that the two scale chain (5.41) is not necessarily reversible.

There are $|\mathcal{Y}|$ equivalence classes in $\mathcal{M}^{(1)}$ that can be labeled by the elements $y \in \mathcal{Y}$, i.e. we have $M_y^{(1)} := \{(x,y), x \in \mathcal{X}\}$.

The effective rates for the slow variables are

$$\bar{r}_N^{(1)}\left(M_y^{(1)}, M_{y'}^{(1)}\right) \;:=\; \frac{1}{N} \sum_{x \in \mathcal{X}} \pi_y(x)\; w_x(y,y') \tag{5.44}$$

for all $y, y' \in \mathcal{Y}$. All the effective rates are at the same scale so that they are all fast; the effective graph's irreducibility follows by the irreducibility of $(\mathcal{Y},\, w_x)$ and the positivity of the invariant measure π_y. This means that $\mathcal{M}^{(2)}$ contains one single element. The unique invariant measure of the effective process with rates (5.44) does not depend on N and coincides with $\overline{\pi}^{(1)}$.

We obtain therefore, by our result with index of iteration $k = 2$, that

$$\lim_{N \to \infty} \pi_N(x,y) \;=\; \pi_y(x)\; \bar{\pi}^{(1)}\left(M_y^{(1)}\right) . \tag{5.45}$$

for every $x \in \mathcal{X}\,,\;\; y \in \mathcal{Y}$.

Even if the fast dynamics on $\mathcal{X}$ and the slow one on $\mathcal{Y}$ are reversible, the global dynamics (5.41) and the effective one (5.44) are in general not reversible. In order to compare it with the continuous framework [3] we give conditions to have reversibility of the effective dinamics. This is important since it coincides with the case where the invariant measures can be usually computed.

Proposition 11. *Suppose that there exist $a_x, b_y > 0$ such that*

$$a_x \, \pi_x(y) \; = \; b_y \, \pi_y(x) \tag{5.46}$$

for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Then the effective process (5.44) is reversible with respect to its unique invariant measure

$$\bar{\pi}^{(1)} \left(M_y^{(1)} \right) \; = \; \frac{b_y}{Z} \, , \quad y \in \mathcal{Y} \, , \tag{5.47}$$

where $Z > 0$ is the suitable normalization constant.

Proof. Let us verify that the effective rates and the prescribed measure satisfy the detailed balance equation:

$$\begin{aligned} \bar{\pi}^{(1)} \left(M_y^{(1)} \right) \; \bar{r}_N^{(1)} \left(M_y^{(1)}, M_{y'}^{(1)} \right) &\overset{(5.44),(5.47)}{=} \frac{b_y}{Z} \sum_{x \in \mathcal{X}} \pi_y(x) \, w_x(y, y') \; = \\ &\overset{(5.46)}{=} \frac{1}{Z} \sum_{x \in \mathcal{X}} a_x \, \pi_x(y) \, w_x(y, y') \; = \\ &\overset{(5.43)}{=} \frac{1}{Z} \sum_{x \in \mathcal{X}} a_x \, \pi_x(y') \, w_x(y', y) \; = \\ &\overset{(5.47),(5.44)}{=} \bar{\pi}^{(1)} \left(M_{y'}^{(1)} \right) \; \bar{r}_N^{(1)} \left(M_{y'}^{(1)}, M_y^{(1)} \right) . \quad \square \end{aligned}$$

Corollary 4. *Let $\left(\pi_y \in \mathcal{P}(\mathcal{X}) \right)_{y \in \mathcal{Y}}$, $\bar{\mu} \in \mathcal{P}(\mathcal{Y})$ be some given probability measures. Then there exist irreducible transition rates $w_y(x, x')$ and $w_x(y, y')$ such that:*

i) for every $y \in \mathcal{Y}$, the rates $w_y(x, x')$ are reversible with respect to the invariant measure π_y;

ii) for every $x \in \mathcal{X}$, the rates $w_x(y, y')$ are reversible with respect to the invariant measure defined by

$$\pi_x(y) \; := \; \frac{1}{Z_x} \, \pi_y(x) \; \bar{\mu}(y) \tag{5.48}$$

where $Z_x > 0$ is the suitable normalization;

iii) the effective transition rates defined as in (5.44) are reversible with respect to the invariant measure $\bar{\pi}^{(1)} \left(M_y^{(1)} \right) = \bar{\mu}(y)$.

Proof. Define the measures $\pi_x \in \mathcal{P}(\mathcal{Y})$ as in (5.48). Parts *i)* and *ii)* are straightforward since, given a probability measure, it is always possible to construct a reversible Markov chain with respect to that measure, for instance by employing the Metropolis–Hastings algorithm.

Now, in order to prove part *iii)*, set

$$b_y := \bar{\mu}(y) \quad , \quad a_x := Z_x = \sum_{y \in \mathcal{Y}} \pi_y(x)\, \bar{\mu}(y). \tag{5.49}$$

Then

$$a_x\, \pi_x(y) \;=\; \pi_y(x)\, \bar{\mu}(y) \;=\; \pi_y(x)\, b_y\,. \tag{5.50}$$

Therefore by Proposition 11 the effective rates $\bar{r}_N^{(1)}\left(M_y^{(1)}, M_{y'}^{(1)}\right) := \sum_{x \in \mathcal{X}} \pi_y(x)\, w_x(y, y')$ are reversible with respect to the invariant measure $\bar{\pi}^{(1)}\left(M_y^{(1)}\right) = \bar{\mu}(y)\,.$ □

Remark 4. *Let us choose*

$$\pi_y(x) := \frac{1}{Z_y}\, e^{-\beta_1 V(x,y)} \quad , \quad \overline{\pi}^{(2)}(y) := \frac{1}{\overline{Z}^{(2)}}\, e^{-\beta_2 F(y)} \tag{5.51}$$

for $\beta_1, \beta_2 > 0$ *,* $V : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$*, and* $F(y) := -\frac{1}{\beta_1} \log Z_y$ *. The measure* (5.48) *provided by Corollary 4 is*

$$\pi_x^{(2)}(y) \;=\; C_x\, e^{-\beta_1 V(x,y) - (\beta_2 - \beta_1) F(y)} \tag{5.52}$$

where C_x *is the suitable normalization.*

We notice that, in the case of continuous diffusions, it follows that

$$\int \nabla_y V(x,y)\, \pi_y(x)\, \dot{x} = \nabla_y F(y). \tag{5.53}$$

As a consequence, also the simpler measure

$$\tilde{\pi}_x^{(2)}(y) \;=\; \tilde{C}_x\, e^{-\beta_2 V(x,y)} \tag{5.54}$$

would do the job.

5.7.4 Boundary driven exclusion process

Consider the boundary driven exclusion process. In this section, we consider only the simplest case; however, in the future we plan to develop this example further, as much deeper combinatorial results can be obtained through higher-order expansions and the derivation of large deviation results. We consider a one-dimensional segment with n nodes, where each node can accommodate at most one particle.

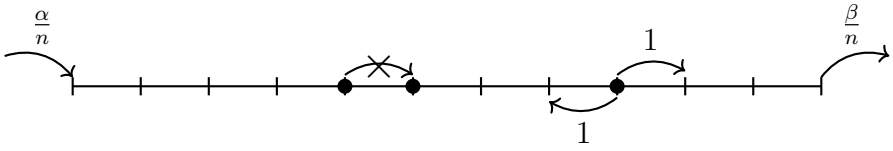


Figure 5.4: The rates of a one-dimensional boundary driven exclusion process in weak contact with external reservoirs.

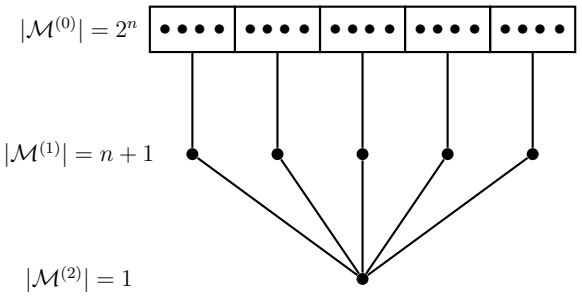


Figure 5.5: The recursive forest associated to the boundary driven exclusion process, that is a two-levels tree.

.

We denote by η a generic configuration of particles and $\eta \in V = \mathcal{M}^{(0)} = \{0,1\}^n$. We have that $\eta(x) = 0, 1$ with $x = 1, \ldots, n$, where $\eta(x) = 1$ means that there is a particle at site x while instead $\eta(x) = 0$ means that the site x is empty. The dynamics is described as follows. Particles jump to a nearest-neighbor site at rate 1, provided that the site is empty, so that the exclusion rule (allowing at most one particle per site) is preserved. The system is in contact with boundary sources, which evolve slowly compared to the bulk dynamics. In particular, at left boundary, i.e. $x = 1$, a particle is created at rate α/N when the site is empty, while at the right boundary, i.e. $x = n$, the particles are destroyed at rate β/N. Here α, β are some fixed positive constant. The rules of the dynamics is illustrated in figure 5.4. This dynamics is usually called the boundary driven exclusion process in contact with weak reservoirs.

The edges of the transition graph are of two types. The fast edges have unit weight and correspond to particle jumps that preserve the total number of particles. The slow edges are of order $O(1/N)$ and correspond to creation and annihilation of particles.

We denote these sets respectively by E_F and E_S, and we write $E = E_S \cup E_F$. The forest corresponding to the iterative construction is indeed a tree and it is illustrated in Figure 5.5.

After the first iteration, we obtain a DAG only composed by $n+1$ local minima

and we have $\mathcal{M}^{(1)} = \left\{M_j^{(1)}, j = 0, \dots, n\right\}$, where $M_j^{(1)} = \left\{\eta : \sum_{x=1}^n \eta(x) = j\right\}$. Since $\hat{V} = V$, the reduced rates are equal to the original ones, i.e. $\hat{r} = r$.

The fast dynamics restricted to the equivalence classes $M_j^{(1)}$ corresponds to the simple exclusion dynamics on the interval with closed boundaries. We then have that the invariant measures do not depend on N and moreover $\mu^{(0),N}_{M_j^{(1)}} = \mu^{(0)}_{M_j^{(1)}} = \mu_j$, where μ_j is the uniform probability measure on $M_j^{(1)}$. The effective dynamics is such that $\bar{r}_N^{(1)}\left(M_i^{(1)}, M_j^{(1)}\right) = 0$ if $|i-j| \neq 1$. Otherwise we have

$$\begin{cases} \bar{r}_N^{(1)}\left(M_i^{(1)}, M_{i+1}^{(1)}\right) = \frac{\alpha}{N}\mu_i(\eta(1) = 0) = \left(1 - \frac{i}{n}\right)\frac{\alpha}{N} \\ \bar{r}_N^{(1)}\left(M_i^{(1)}, M_{i-1}^{(1)}\right) = \frac{\beta}{N}\mu_i(\eta(n) = 1) = \frac{i}{n}\frac{\beta}{N}\,. \end{cases} \tag{5.55}$$

The effective dynamics is a random walk on the integer numbers from 0 to n that increases or decreases its value by one, with rates given in (5.55). This Markov chain is reversible, and its invariant measure does not depend on N and can be directly computed, obtaining:

$$\bar{\pi}^{(1)}\left(M_i^{(1)}\right) = \binom{n}{i}\left(\frac{\alpha}{\alpha+\beta}\right)^i\left(\frac{\beta}{\alpha+\beta}\right)^{n-i}, \tag{5.56}$$

that is a Binomial distribution of parameters n and $\frac{\alpha}{\alpha+\beta}$. where Z is the normalization factor. As a consequence, we can write the limiting measure as

$$\pi(\eta) = \sum_{j=0}^n \bar{\pi}^{(1)}\left(M_j^{(1)}\right)\mu_j(\eta) = \mathrm{Ber}_{\frac{\alpha}{\alpha+\beta}}(\eta)\,, \tag{5.57}$$

that is a Bernoulli product measure on $\{0,1\}^n$ of parameter $\left(\frac{\alpha}{\alpha+\beta}\right)$.

We can indeed obtain the above result in this case by a direct identification of the fast arborescences. Consider $\eta \in M_j^{(1)}$. The structure of the fast arborescences $\mathcal{T}_\eta^F$ can be deduced from Lemma 8 as follows. Consider an arborescence τ and restrict it to a fast equivalence class, i.e. $\tau\left[M_j^{(1)}\right]$. In general, the restriction of a spanning arborescence to a subgraph is a spanning forest. For the fast arborescences $\tau \in \mathcal{T}_\eta^F$, it holds that, for any j, $\tau\left[M_j^{(1)}\right]$ is itself a spanning arborescence.

When $j > i$, it must be a spanning arborescence rooted at a configuration $\eta' \in M_j^{(1)}$, such that $\eta'(n) = 1$. When $j < i$, the arborescence is rooted on a $\eta' \in M_j^{(1)}$ such that $\eta'(1) = 0$. When $j = i$, the arborescence is rooted on η.

From the equivalence classes $M_j^{(1)}$ with $j > i$, there is only one slow edge exiting from the root and going toward $M_{j-1}^{(1)}$, while for $j < i$ the slow edges go torward $M_{j+1}^{(1)}$.

Fix a configuration with i particles $\eta \in M_i^{(1)}$. Any fast arborescece $\mathcal{T}_\eta^F$, when restricted to $M_j^{(1)}$, is an arborescence rooted on a configuration of particles having j particles, one of those is in the lattice site n. All the edges of this subarborescence have weight 1. The number of configurations on which this subarborescence can be rooted are $\binom{n-1}{j-1}$ and by symmetry, for each root, the number of arborescences in $\left(M^{(j)}, E_F\right)$ is the same, and let us call it $\mathcal{N}_j$. Moreover, in $\mathcal{T}_\eta^F$ there is one single edge exiting from $M_j^{(1)}$, with $j > i$, and it is the unique edge connecting the root of the subarborescence $\mathcal{T}_\eta^F \left[M_j^{(1)}\right]$ to $M_{j-1}^{(1)}$. The latter has a weight that is $\frac{\beta}{N}$.

Likewise, for $j < i$ we have a similar situation, but the number of roots in $M_j^{(1)}$ with $j < i$ is $\binom{n-1}{j}$ and the weight of the unique edge going from $M_j^{(1)}$ to $M_{j+1}^{(1)}$ is $\frac{\alpha}{N}$.

We finally obtain that $\mathcal{T}_\eta^F \left[M_i^{(1)}\right]$ is an arborescence rooted at η.

From this we deduce that, for $\eta \in M_i^{(1)}$, we have

$$r_N\left(\mathcal{T}_\eta^F\right) = \frac{\alpha^i \beta^{n-i}}{N^n} \left(\prod_{k=0}^{n-1} \binom{n-1}{k}\right) \prod_{\ell=0}^{n} \mathcal{N}_\ell^n, \tag{5.58}$$

that is a common factor that does not depend on the configuration η considered times the factor $\alpha^{\sum_i \eta(i)} \beta^{n-\sum_i \eta(i)}$ that gives a product Bernoulli measure with parameter $\frac{\alpha}{\alpha+\beta}$.

5.8 Star-Delta reduction for directed graphs

We start discussing the reduction in the case of the complete graph and then consider the general case. Let $K_{n+1} := (V, E, r_N)$ be a complete weighted directed graph with $n+1$ vertices $V = \{x_1, ..., x_n, s\}$. One of them, s, is the one we are going to remove. To each edge in $(x, y) \in E$ we associate a nonegative rate $r_N(x, y) \geq 0$, that is the weight of the graph. We split the edge set E as the disjoint union of two subsets: $E = E^b \bigcup E^i$ that are respectively the boundary edges and the internal ones:

- E^b, consisting of edges of type (x_i, x_j), with $x_i, x_j \in V \setminus \{s\}$, and
- E^i, consisting of edges of type (x_i, s) or (s, x_i), with $x_i \in V \setminus \{s\}$.

We describe now $\hat{K}_n := (\hat{V}, \hat{E}, \hat{r}_N)$ a weighted directed complete graph, that we call the Wye-Delta reduction of K_{n+1}. The set of vertices is $\hat{V} := V \setminus \{s\} =$

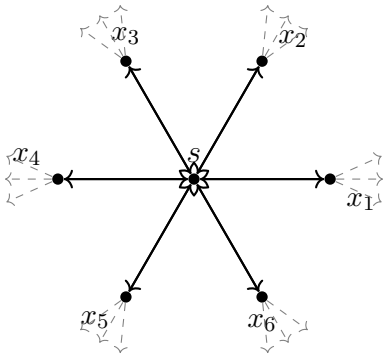


Figure 5.6: The graph K_{n+1}: the figure shows only the edges in E^i highlighted

$\{x_1, ..., x_n\}$ and the set of edges is $\hat{E} = E^b$. The reduced rates associated to the edges in $\hat{E}$ are defined as:

$$\hat{r}_N(x_i, x_j) := r_N(x_i, x_j) + \frac{r_N(x_i, s) r_N(s, x_j)}{\sum_{x_k \in \hat{V}} r_N(s, x_k)} \,. \tag{5.59}$$

Remark 5. *It will be sometimes useful and natural to write the above rates as*

$$\hat{r}_N(x_i, x_j) = r_N(x_i, x_j) + r^*_N(x_i, x_j) \,, \tag{5.60}$$

and to consider $\hat{K}_n$ as a multigraph having two different edges going from x_i to x_j having respectively weights associated to the two terms on the right hand side of (5.60)*. We call $\hat{\mathbb{K}}_n$ this multigraph. In particular we identify the splitting in* (5.60) *as a splitting of the edges of $\hat{K}_n$, and we have in this case that the edges of $\hat{\mathbb{K}}_n$ are $E^b \cup E^*$. Any pair of vertices $x_i, x_j \in \hat{\mathbb{K}}_n$ are connected by an edge in E^b with weight $r_N(x_i, x_j)$, and by an edge in E^* with weight $r^*_N(x_i, x_j)$. We call $\hat{\mathbb{R}}_N$ the weight of the multigraph defined in this way.*

We call respectively $\mathcal{T}$ and $\hat{\mathcal{T}}$ the arborescences of the original and of the reduced complete graphs. Similarly, we identify any other quantity associated to $\hat{K}_n$ using the hat symbol.

Let us illustrate the construction, considering without loss of generality the vertex x_1. We can classify a spanning arborescence $\tau \in \mathcal{T}_{x_1}$, according with the geometric structure of the forest $\hat{f} := \tau \cap E^b$; notice that $\hat{f}$ is a spanning directed forest of $\hat{K}_n$ with one component directed torwards x_1. This directed forest contains only edges in E^b. We may have several cases, depending on the number of components. In general, the forest $\hat{f}$ has k components, that we call $(C_i)_{i=1}^k$, i.e. we have $\hat{f} = \cup_{j=1}^k C_j$.

Each component is a directed arborescence (not spanning), and without loss of generality we call C_1 the one rooted at x_1. The number k can be equal to one, in

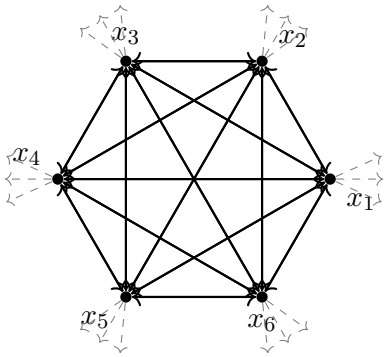


Figure 5.7: The multi-graph $\hat{\mathbb{K}}_n$

the case that τ was containing just one edge of E^i exiting from s, or can assume every natural number up to n, that corresponds to the case when τ was containing only edges in E^i. Indeed, the number of components k coincides exactly with $\left|\tau \cap E^i\right|$.

Calling $c_i := |C_i|$, we have that $\sum_{i=1}^{k} c_i = n$. Since τ is directed torward x_1, among these k edges belonging to E^i, only one is directed exiting from s and is therefore connecting s to the component C_1. We call $\mathcal{T}_{x_1,k}$ the arborescences directed torwards x_1, and such that the corresponding directed forest f has k components.

Recall that $\mathcal{T}_{x_1}$ is the set of all the spanning arborescences of K_{n+1} rooted in x_1. We call $r_N^{out}(s) := \sum_{i=1}^{n} r_N(s, x_i)$ the exiting rate from the node s. We have the following result: note that the corresponding result for non directed graphs [18] has a stronger statement that can not be extended in the directed case.

Lemma 10. *For any $x_i \in \hat{K}_n$ we have*

$$\hat{r}_N(\hat{\mathcal{T}}_{x_i}) = \frac{r_N(\mathcal{T}_{x_i})}{r_N^{out}(s)}$$

Proof. Without loss of generality, we prove it for the vertex x_1. Observe that the edges in $\hat{E}$ can be considered both edges of $\hat{K}_n$ or K_{n+1}. We call $\hat{\mathcal{F}}(x_{i_1}, \ldots, x_{i_\ell})$ a spanning forest in $\hat{K}_n$ having ℓ components, rooted at $x_{i_1}, \ldots, x_{i_\ell}$. We use a similar notation for spanning forests in K_{n+1}.

Note that $\hat{\mathcal{F}}(x_{i_1}, \ldots, x_{i_\ell})$ is both a subgraph of $\hat{K}_n$ and of K_{n+1}, so that we can compute both $r_N\left(\hat{\mathcal{F}}(x_{i_1}, \ldots, x_{i_\ell})\right)$ and $\hat{r}_N\left(\hat{\mathcal{F}}(x_{i_1}, \ldots, x_{i_\ell})\right)$. We now compute $r_N(\mathcal{T}_{x_1})$, by separating the contributions according to the classification of arborescences introduced above.

Given $\hat{f} \in \hat{\mathcal{T}}_{x_1}$, all the arborescences in $\mathcal{T}_{x_1,1}$ (recall the definition given before lemma 10) that gives such an $\hat{f}$ are of the form $\hat{f} \cup (s, x_i)$, for some $x_i \in \hat{K}_n$. This

means that the weight of all arborescences of this type, with respect to the weight function r_N, is $r_N(\hat{f})\sum_{i=1}^n r_N(s,x_i) = r_N(\hat{f})r_N^{out}(s)$. Summing over all the elements in $\hat{\mathcal{T}}_{x_1}$ we obtain

$$r_N\left(\mathcal{T}_{x_1,1}\right) = r_N\left(\hat{\mathcal{T}}_{x_1}\right) r_N^{out}(s)\,. \tag{5.61}$$

Given $\hat{f} \in \hat{\mathcal{F}}(x_1,x_{i_2})$ all the $\tau \in \mathcal{T}_{x_1}$ that are associated to $\hat{f}$ are of the form $\hat{f}\cup(x_{i_2},s)\cup(s,y)$ for some $y \in C_1$. This is because transforming the arborescence into $\hat{f}$ requires exactly two edges to be removed: one leaving s and one entering it. This means that the weight of all such arborescences is $r_N\left(\hat{\mathcal{F}}(x_1,x_{i_2})\right) r_N(x_{i_2},s)\sum_{y\in C_1} r_N(s,y)$. Summing over all elements in $\hat{\mathcal{F}}(x_1,x_{i_2})$, we obtain

$$r_N(\mathcal{T}_{x_1,2}) = \sum_{x_{i_2}\neq x_1}\sum_{\hat{f}\in\hat{\mathcal{F}}(x_1,x_{i_2})} r_N(\hat{f})r_N(x_{i_2},s)\sum_{y\in C_1} r_N(s,y)\,, \tag{5.62}$$

where we recall that $C_1 = C_1(\hat{f})$ is the component containing x_1.

In a similar way, for arborescences belonging to $\mathcal{T}_{x_1,k}$

$$r_N(\mathcal{T}_{x_1,k}) = \sum_{\left\{\{x_{i_2},\ldots,x_{i_k}\}\in\mathcal{S}_{x_1}^{k-1}\right\}}\ \sum_{\left\{\hat{f}\in\hat{\mathcal{F}}(x_1,x_{i_2},\ldots,x_{i_k})\right\}} r_N(\hat{f})\prod_{j=1}^{k-1} r_N(x_{i_j},s)\sum_{y\in C_1} r_N(s,y)\,, \tag{5.63}$$

where by $\mathcal{S}_{x_1}^{k-1}$ we denote the collection of subsets of $\hat{K}_n$ containing exactly $k-1$ elements, all different from x_1. We have finally that $r_N(\mathcal{T}_{x_1}) = \sum_{k=1}^n r_N(\mathcal{T}_{x_1,k})$, where the terms in the right hand side are given by (5.65).

In order to compute $\hat{r}_N(\hat{\mathcal{T}}_{x_1})$, it is convenient to use the splitting (5.60) and the multigraph of remark 5.

The total weight associated to $\hat{\mathcal{T}}_{x_1}$ in $\hat{K}_n$ is the same as the total weight associated to all the arborescence directed torwards x_1 on the multigraph $\hat{M}_n$, and, with the different weights, associated to the different families of edges. In formulas, recalling that we denote by $\hat{\mathbb{R}}_N$ the weights on $\hat{\mathbb{K}}_n$ (that, as we saw, can be of two different types), and denoting by $\hat{\mathbb{T}}_{x_1}$ the arborescences directed torwards x_1 in $\hat{\mathbb{K}}_n$, we have $\hat{r}_N\left(\hat{\mathcal{T}}_{x_1}\right) = \hat{\mathbb{R}}_N\left(\hat{\mathbb{T}}_{x_1}\right)$.

If we consider a directed arborescence on the multigraph, i.e. an element of $\hat{\mathbb{T}}_{x_1}$, and consider just the edges belonging to E^b, we obtain a spanning directed forest in $\hat{K}_n$, that may have $k = 1,\ldots,n$ components. We call $\hat{\mathbb{T}}_{x_1,k}$ the arborescenses of $\hat{\mathbb{K}}_n$ associated to forests with k components. We have $\hat{r}_N(\hat{\mathcal{T}}_{x_1}) = \sum_{k=1}^n \hat{\mathbb{R}}_N(\hat{\mathbb{T}}_{x_1,k})$.

In the case of $k=1$, since every element of $\hat{\mathbb{T}}_{x_1,1}$ has all the edges in E^b, we simply have that $\hat{\mathbb{T}}_{x_1,1} = \hat{\mathcal{T}}_{x_1}$; but the edges are weighted now with r_N in the multigraph, and not with $\hat{r}_N$, so that $\hat{\mathbb{R}}_N\left(\hat{\mathbb{T}}_{x_1,1}\right) = r_N\left(\hat{\mathcal{T}}_{x_1}\right)$.

In the case of $k = 2$, we have

$$\hat{\mathbb{R}}_N(\hat{\mathbb{T}}_{x_1,2}) = \sum_{x_{i_2} \neq x_1} \sum_{\hat{f} \in \hat{\mathcal{F}}(x_1, x_{i_2})} r_N(\hat{f}) \sum_{y \in C_1} r_N^*(x_{i_2}, y)\,. \tag{5.64}$$

In the general case,

$$\hat{\mathbb{R}}_N(\hat{\mathbb{T}}_{x_1,k}) = \sum_{\left\{\{x_{i_2},\ldots,x_{i_k}\} \in \mathcal{S}_{x_1}^{k-1}\right\}} \sum_{\left\{\hat{f} \in \hat{\mathcal{F}}(x_1, x_{i_2},\ldots,x_{i_k})\right\}} r_N(\hat{f}) \sum_{\left\{y_{i_2},\ldots,y_{i_k} \in T^k\right\}} \prod_{j=1}^{k-1} r_N^*(x_{i_j}, y_{i_j})\,, \tag{5.65}$$

where, as before, $\mathcal{S}_{x_1}^{k-1}$ is the collection of subsets of $\hat{K}_n$ with k elements, all different from x_1.

The notation $\sum_{\left\{y_{i_2},\ldots,y_{i_k} \in T^k\right\}}$ denotes a sum over configurations of points $\left\{y_{i_2}, \ldots, y_{i_k}\right\}$ in $\hat{K}_n$, not necessarily distinct, and such that if we consider a graph having as vertices the k components $C_1, \ldots, C_k$ of $\hat{f}$ and we draw an arrow from C_ℓ to C_m, when $y_{i_\ell} \in C_m$, then we obtain an arborescence directed torward C_1.

Recalling the form of the rates r_N^* from (5.59) and (5.60), using the following lemma 12, we obtain that for each k we have $r_N\left(\mathcal{T}_{x_1,k}\right) = \hat{\mathbb{R}}_N(\hat{\mathbb{T}}_{x_1,k}) r_N^{out}(s)$. This implies the statement of the Theorem.

□

Proposition 12. *Given $C_1, \ldots, C_k$ a partition of $\hat{K}_n$, then we have*

$$\sum_{\left\{y_{i_2},\ldots,y_{i_k} \in T^k\right\}} \prod_{j=1}^{k-1} r_N(s, y_{i_j}) = \left(\sum_{y \in C_1} r_N(s, y)\right) \left(r_N^{out}(s)\right)^{k-2} \tag{5.66}$$

Proof. Consider an abstract, complete directed graph A_k having as vertices the connected components, and weight to the edge (C_i, C_j) equal to $w(C_i, C_j) = \sum_{y \in C_j} r_N(s, y) := r_j$. Note that the weight associated to an edge depends just on the arriving node. In particular, all the edges entering in C_j have the same weight r_j. Note also that $\sum_{j=1}^{k} r_j = r_N^{out}(s)$.

Let $\mathcal{T}_{C_1}^A$ be the set of arborescences directed torwards C_1. The left hand side of (5.66) coincides with $w\left(\mathcal{T}_{C_1}^A\right)$. By the directed version of the matrix tree Theorem, we can write the total weigth of the arborescences directed torward C_1 as a determinant of a Laplacian matrix [61]. More precisely, consider a $k \times k$ matrix L with rows and columns indicized by the vertices C_i. The elements $L_{i,i}$ along the diagonal corresponds to the total weights of edges exiting from C_i, and in our case we have $L_{i,i} = r_N^{out}(s) - r_1$. Outside of the diagonal, we have that $L_{i,j} = -w(C_i, C_j) = -r_j$. Let L^1 be the $(k-1) \times (k-1)$ matrix obtained by

removing the first line and the first colimn from L. The directed version of the matrix tree Theorem [61] says that $w\left(\mathcal{T}_{C_1}^A\right) = \det L^1$. We have that the matrix L^1 is the difference between a diagonal matrix and a rank-one matrix and can be written as

$$L^1 = r_N^{out}(s)\mathbb{I} - \underline{1} r^T \,, \tag{5.67}$$

where $\mathbb{I}$ is the $(k-1)\times(k-1)$ identity matrix, $\underline{1}$ is a $k-1$ dimensional column vector having all the elements equal to one, and r^T is the $k-1$ dimensional row vector given by $r^T = (r_2, r_3, \ldots, r_k)$. The determinant of L^1 is the product of the corresponding eigenvalues that can be easily computed due to the special form of the matrix. In particular, we have that $\underline{1}$ is an eigenvector of (5.67) with eigenvalue r_1, and then if we consider a basis of the $k-2$ dimensional plane orthogonal to the vector r, we obtain $k-2$ eigenvectors with eigenvalues all equal to $r_N^{out}(s)$. The $k-1$ eigenvectors we found are linearly independent. We obtain therefore

$$\det L^1 = r_1 \left(r_N^{out}(s)\right)^{k-2} \,, \tag{5.68}$$

that coincides with the right hand side of (5.66), and this concludes the proof. □

5.8.1 The general theorem

Consider as usual (V, E) to be a finite strongly connected digraph where $V = S \cup \hat{V}$, $S \neq \emptyset$. (V, E) is the transition graph of a Markov chain with transition rates r_N. We consider the trace process on the set $\hat{V}$. The corresponding reduced graph is $(\hat{V}, \hat{E})$, and the reduced rates are

$$\hat{r}_N(x_i, x_j) := r_N(x_i, x_j) + \sum_{s\in S} r_N(x_i, s) P_N^{\hat{V}}(x_j|s) \,, \qquad x_i, x_j \in \hat{V} \,,$$

where we recall that $P_N^{\hat{V}}(x_j|s)$ is the probability that a Markov chain having rates r_N and starting at s hits $\hat{V}$ for the first time in x_j. As already discussed, by the Matrix-tree theorem we can write:

$$\hat{r}_N(x_i, x_j) = \frac{r_N(\mathcal{F}_{\hat{V}\setminus\{x_i\}}(x_i \to x_j))}{r_N(\mathcal{F}_{\hat{V}})}. \tag{5.69}$$

Theorem 12. *Let $x \in \hat{V}$. Then*

$$r_N(\mathcal{T}_x) = r_N(\mathcal{F}_{\hat{V}})\, \hat{r}_N(\hat{\mathcal{T}}_x). \tag{5.70}$$

where we recall that $\mathcal{F}_{\hat{V}}$ denotes the collection of spanning forests of (V, E) with root set $\hat{V}$.

Proof. We prove the statement by induction. First of all, we observe that lemma 10 is written for the complete graph, but the proof holds also in the case that some rates are equal to zero: it corresponds to the statement of this Theorem in the case that $|S| = 1$.

Assume that (5.70) holds after the removal of $|S| - 1$ nodes. We will show that the statement then also holds for $|S|$. Consider $y \in S$, and call $\tilde{V} = \hat{V} \cup \{y\}$ and $\tilde{r}, \tilde{T}_x$ respectively the reduced rates and arborescences. By the inductive hypothesis we have for $x \in \hat{V}$

$$r_N(\mathcal{T}_x) = r_N\left(\mathcal{F}_{\tilde{V}}\right) \tilde{r}_N\left(\tilde{T}_x\right) . \tag{5.71}$$

We now apply the inductive hypothesis, considering the removal of a single node, to the second term on the right-hand side above, and we obtain

$$\tilde{r}_N\left(\tilde{T}_x\right) = \left(\sum_{z \in \hat{V}} \tilde{r}_N(y, z)\right) \hat{r}_N\left(\hat{T}_x\right) . \tag{5.72}$$

Inserting (5.72) in (5.71) and using (5.69), we get

$$r_N(\mathcal{T}_x) = \left(\sum_{z \in \hat{V}} r_N\left(\mathcal{F}_{\hat{V}}(y \to z)\right)\right) \hat{r}_N\left(\hat{T}_x\right) , \tag{5.73}$$

and, since this is clearly true,

$$\sum_{z \in \hat{V}} r_N\left(\mathcal{F}_{\hat{V}}(y \to z)\right) = r_N(\mathcal{F}_{\hat{V}}) ,$$

the proof concludes.

□

5.9 Conclusions and Perspectives

The result obtained in Theorem 11 is consistent with the results presented in [13, 13, 40, 42, 44, 46]. It is therefore natural to expect that the objects defined in [46] [31], like the capacities and the rate functional, admit a combinatorial representation in terms of forests, that would be reasonabily similar to the combinatorial expression of the Drazin-Inverse presented in [35].

Furthermore, these results hint at a deeper connection that warrants further investigation. It remains a distinct possibility that Theorem 11, and the relationship between the original chain and the trace process, could be interpreted in terms of couplings, specifically within the framework of intertwining duality relations, as seen in [6] and [49].

One of the central contributions of this research is that it does not rely on the assumption of reversibility. This approach contributes to the ongoing effort within the field to generalize 'electrical network tools' beyond their traditional reversible constraints, building upon the theoretical foundations recently laid by authors such as [45] [8], [7], [56], [33], [36].

We also ask whether the assumption of complete asymptotic regularity (Definition 5) is truly necessary for the existence of the limit of the invariant measure, or whether this condition can be relaxed. It is noteworthy that the two parts of this thesis (as well as our first and second papers, [2, 4]) are connected by a remarkably similar underlying structure. Indeed, in the first work we again consider a sequence of Markov chains with multiscale transition rates, in which the set of fast edges satisfies an additional condition, namely Assumption 2. This observation gives us reason to believe that this assumption may be removed, thereby allowing for a significant extension of the results obtained in our first work [4].

The main results can be further pushed and generalised in different non trivial directions. Among the simplest possible natural extensions, there is the setting, still of finite state spaces, but with cardinality growing with N.

Moreover, this combinatorial approach, relying on arborescences and forests, provides a natural framework for investigating the first-order corrections to the invariant measure in the limit.

Appendices

A

The Reversible Case: Reductions on Undirected Graphs

We now present the known results concerning reduction theory for undirected graphs. For a more detailed treatment, we refer the reader to [18]. We emphasize that undirected graphs can be interpreted as transition graphs of reversible Markov chains; consequently, these results may also be understood from a probabilistic perspective as reduction results for reversible Markov chains.

A.1 Planar Graph Transformations

A natural method to develop graph-theoretic algorithms is to develop simplifying transformations that, when applied repeatedly to a graph, reduce it in polynomially many steps to a trivial graph such as a single edge. We examine algorithms of this type.

We start with n-vertex graph $G = (V, E)$, with no loops and possibly with parallel edges. Consider the following four basic transformations:

- *Pendant edge deletion*: A pendant edge $e \in E$ (incident with a vertex of degree 1) can be deleted along with the degree 1 vertex.
- *Parallel reduction*: Two edges e and f in parallel (each having the same two endvertiees) can be replaced by a single edge e' having the same two endvertices.
- *Series reduction*: Two edges $e = (x, y)$ and $f = (y, z)$ for which y is a degree 2 vertex can be replaced by a single edge $e' = (x, z)$, with y being deleted.

- *Star-Delta transformation*: If w is a degree 3 vertex (a star) adjacent to three vertices x, y, z by edges e, f, g respectively, vertex w and edges e, f, g can be deleted and replaced by edges $e' = (y, z)$, $f' = (x, z)$ and $g' = (x, z)$.
- *Delta-Star transformation*: If $e' = (y, z)$, $f' = (x, z)$ and $g' = (x, y)$ are edges of G (a delta), they can be deleted and replaced by adding a new vertex w adjacent to x, y, z.

G is called a Delta-Star-Delta-reducible graph if it can be reduced to a single edge by repeated application of the six transformations above, in any order.

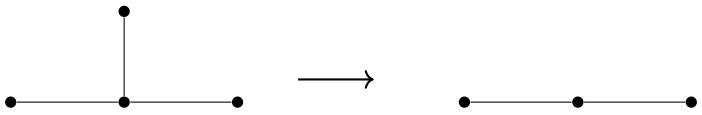


Figure A.1: Pendant edge deletion

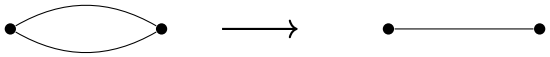


Figure A.2: Parallel edge reduction

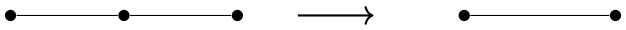


Figure A.3: Series edge reduction

A.2 Spanning Trees

Let G be a Delta-Star-Delta-reducible graph. For each edge $e \in E$ we associate a weight $r(e)$.

Definition 34. *A spanning tree $\tau \in G$ is a spanning connected subraph of G with no loops.*

Let us denote by $\mathcal{T}$ the set of spanning trees of G. We assign a weight r to each $\tau \in \mathcal{T}$ as follows:

$$r(\tau) = \prod_{e \in \tau} r(e).$$

Similarly, we assign a weight to the set $\mathcal{T}$: $r(\mathcal{T}) = \sum_{\tau \in \mathcal{T}} r(\tau)$.

We next show how these edge weights are updated under each of the six basic transformations. We first observe that if, for any single edge $e \in E$, the inclusion

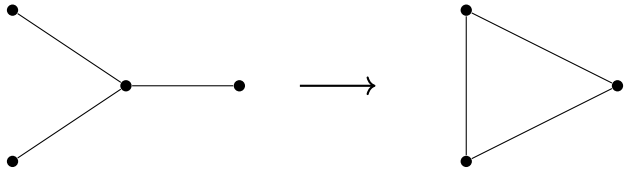


Figure A.4: Star-Delta reduction

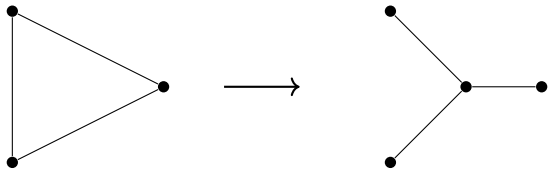


Figure A.5: Delta-Star reduction

and exclusion weights are both multiplied by a factor δ, the effect on the total weight is likewise a factor of δ. We therefore maintain a single *overall scale* $S := \prod_{e \in E} r(e)$.

Suppose, then, that G is the graph prior to the transformation, with normalized weights on the edges and overall scale d.

- If the transformation is the deletion of a pendant edge e with weight $r(e)$, observe that e is in every spanning tree of G, and hence the resulting reduced graph's overall scale is multiplied by $r(e)$, with other edge weights remaining unchanged.

- If the transformation is a parallel reduction on two edges e, f with weights $r(e)$ and $r(f)$, observe that every spanning tree contains at most one between e or f. Hence the replacement edge e' will have weight $r(e) + r(f)$, and the remaining edge weights and overall scale remain unchanged.

- If the transformation is a series reduction on two edges e, f with weights $r(e)$ and $r(f)$, observe that every spanning tree contains at least one between e and f. There are $r(e)r(f)$ ways to include both e and f, and $r(e)+r(f)$ ways to exclude one of them; there is no way to exclude both and get a spanning tree. Then normalizing as before, the replacement edge e' has weight $\frac{r(e)r(f)}{r(e)+r(f)}$, the overall scale is multiplied by $r(e) + r(f)$, and again no other edge weights are changed.

Let us now consider the last two redcutions. Let G_Y be a graph with scale d, having a vertex w of degree 3 (a star) adjacent to three vertices x, y, z via edges e, f, g of weights $r(e), r(f), r(g)$, respectively. Let G_Δ be the same underlying graph as G_Y, but with vertex w and edges e, f, g removed and replaced by $e' = (y, z)$, $f' = (x, z)$ and $g' = (x, y)$, having weights $r(e'), r(g'), r(f')$ respectively; G_Δ has overall scale δ.

Now consider the total weight of G_Y and G_Δ. A spanning tree τ of G_Y is one of five types, according to the structure of the forests $f = \tau \setminus \{e, f, g\}$:

1. τ' has x, y, z in the same component;
2. τ' has two components with x and y in the same component;
3. τ' has two components with x and z in the same component;
4. τ' has two components with y and z in the same component; or
5. τ' has three components, containing x, y and z respectively.

The same applies to a tree τ spanning G_Δ, depending upon the structure of $f = \tau \setminus \{e', f', g'\}$.

For $i = 1 \dots 5$, let Σ_i, let Σ_i be the sum over all forests f of $G_Y \setminus \{e, f, g\} = G_\Delta \setminus \{e', f', g'\}$ of type i above of the product of the weights of the edges in f. Then observe that the total weight of the spanning trees in G_Y is

$$d\Bigg[\Big(r(e)+r(f)+r(g)\Big)\Sigma_1 + r(g)\Big(r(e)+r(f)\Big)\Sigma_2 + r(f)\Big(r(e)+r(g)\Big)\Sigma_3 + \\ + r(e)\Big(r(f)+r(g)\Big)\Sigma_4 + r(e)r(f)r(g)\Sigma_5\Bigg],$$

while the total weight of the spanning trees of G_Δ is

$$\delta\Bigg[\Sigma_1 + \Big(r(e')+r(f')\Big)\Sigma_2 + \Big(r(e')+r(g')\Big)\Sigma_3 + \Big((f')+r(g')\Big)\Sigma_4 + \\ + \Big(r(e')r(f')+r(e')r(g')+r(f')r(g')\Big)\Sigma_5\Bigg].$$

Equating the coefficients of Σ_1 through Σ_5 gives the following system of five equations:

$$\begin{cases} d\Big(r(e)+r(f)+r(g)\Big) = \delta, \\ dr(e)\Big(r(f)+r(g)\Big) = \delta\Big(r(f')+r(g')\Big), \\ dr(f)\Big(r(e)+r(g)\Big) = \delta\Big(r(e')+r(g')\Big), \\ dr(g)\Big(r(e)+r(f)\Big) = \delta\Big(r(e')+r(f')\Big), \\ dr(e)r(f)r(g) = \delta\Big(r(e')r(f')+r(e')r(g')+r(f')r(g')\Big). \end{cases} \tag{A.1}$$

Given $r(e), r(f), r(g), d$, we have five equations to solve for the four unknowns $r(e'), r(f'), r(g'), \delta$ and given $r(e'), r(f'), r(g'), \delta$ we have five equations to solve for the four unknowns $r(e), r(f), r(g), d$.

Given $r(e), r(f), r(g), d$ from the first equation of (A.1) we obtain

$$\delta = \frac{d}{r(e) + r(f) + r(g)}.$$

Then, using the second, third and fourth equation of of (A.1), we can compute:

$$\begin{pmatrix} r(e') \\ r(f') \\ r(g') \end{pmatrix} = \frac{r(e)r(f)r(g)}{r(e) + r(f) + r(g)} \begin{pmatrix} \frac{1}{r(e)} \\ \frac{1}{r(f)} \\ \frac{1}{r(g)} \end{pmatrix}. \tag{A.2}$$

It remains to verify that the last equations in (A.1) is satisfied:

$$\delta\Big(r(e')r(f') + r(e')r(g') + r(f')r(g')\Big) = \delta r(e')r(f')r(g')\left(\frac{1}{r(e')} + \frac{1}{r(f')} + \frac{1}{r(g')}\right),$$

which simplifies to $dr(e)r(f)r(g)$ as required.

Similarly, in the case of the Delta-Star transformation, we are given $r(e'), r(f'), r(g')$ and δ. Since $r(e) = \frac{dr(e)r(f)r(g)}{dr(f)r(g)}$, $r(f) = \frac{dr(e)r(f)r(g)}{dr(e)r(g)}$ and $r(g) = \frac{dr(e)r(f)r(g)}{dr(e)r(f)}$, we obtain:

$$\begin{pmatrix} r(e) \\ r(f) \\ r(g) \end{pmatrix} = \Big(r(e')r(f') + r(e')r(g') + r(f')r(g')\Big) \begin{pmatrix} \frac{1}{r(e')} \\ \frac{1}{r(f')} \\ \frac{1}{r(g')} \end{pmatrix}, \tag{A.3}$$

and as a consequence

$$d = \frac{\delta}{\Big(r(e')r(f') + r(e')r(g') + r(f')r(g')\Big)\left(\frac{1}{r(e')} + \frac{1}{r(f')} + \frac{1}{r(g')}\right)},$$

that satisfies from the second to the fifth equation of (A.1). In order to verify the first one, just observe that

$$r(e) + r(f) + r(g) = \Big(r(e')r(f') + r(e')r(g') + r(f')r(g')\Big)\left(\frac{1}{r(e')} + \frac{1}{r(f')} + \frac{1}{r(g')}\right);$$

and the other edge weights remain unchanged.

Acknowledgments

I would like to first thank Prof. Davide Gabrielli for the fundamental role he played throughout my three years of doctoral studies. I am grateful to him for finding research topics that truly sparked my interest, for always striving for what was best for me, and for carefully selecting scientific events and meetings that were genuinely aligned with my interests and growth.

I thank Prof. Diego Alberici for guiding me through the most technical and computational aspects of this work, for the enormous amount of time he dedicated to me (often sacrificing entire afternoons) and for patiently engaging with my digressions. I am also grateful to him for having been, over time, first a tutor and then a friend.

I thank Dr. Michele Aleandri for the time and energy he devoted to me, and for the many valuable pieces of advice he has given me over the years.

I would like to express my sincere gratitude to the Referees for their time and insightful comments. Their feedback was instrumental in improving this thesis and prompted me to consider new and important research perspectives.

I would also like to thank Professors Landim, Tsagkarogiannis, Ferrari, Camilli, Aragona, Palladino and Dr. Nozzi for their essential scientific advice. A special mention goes to Professor Travaglini, whose profound passion for research has been truly inspiring.

Finally, I would like to thank all those who have been close to me during these three years, in particular my family, Newton, my friends, and my colleagues at IC Balsorano. A special mention goes to all my students, who have been both a challenge and a joy during this demanding final stage of my studies. A heartfelt thought goes to Danilo, whose memory remains a constant presence.

Bibliography

[1] D. Alberici, N. Macris, and E. Mingione. On the convergence to the non-equilibrium steady state of a Langevin dynamics with widely separated time scales and different temperatures.

[2] Diego Alberici, Davide Gabrielli, and Giulia Pallotta. On the limiting invariant measure of multiscale Markov chains. *In preparation*, May 2026.

[3] Michele Aleandri, Davide Gabrielli, and Matteo Colangeli. A combinatorial representation for the invariant measure of diffusion processes on metric graphs. *Latin American Journal of Probability and Mathematical Statistics*, 18, 02 2020.

[4] Michele Aleandri, Davide Gabrielli, and Giulia Pallotta. Freidlin-Wentzell solutions of discrete Hamilton-Jacobi equations. *Electronic Journal of Probability*, 31, 01 2026.

[5] V. Anantharam and P. Tsoucas. A proof of the Markov chain tree theorem. *Statist. Probab. Lett.*, 8(2):189–192, 1989.

[6] Luca Avena, Fabienne Castell, Alexandre Gaudillière, and Clothilde Mélot. Intertwining wavelets or multiresolution analysis on graphs through random forests. *Applied and Computational Harmonic Analysis*, 48(3):949–992, 2020.

[7] Luca Avena and Alexandre Gaudillière. Two applications of random spanning forests. *Journal of Theoretical Probability*, 31, 12 2018.

[8] Marton Balazs and Áron Folly. Electric network for non-reversible Markov chains. *The American Mathematical Monthly*, 123, 05 2014.

[9] Paolo Baldi. *Calcolo delle probabilità e statistica*. Collana di istruzione scientifica. Serie di matematica. McGraw-Hill libri Italia, Milano, 2 edition, 1998.

[10] Jørgen Bang Jensen. *Digraphs: Theory, Algorithms and Applications*. Springer Publishing Company, Incorporated, 2nd edition, 2008.

[11] Johel Beltrán and Claudio Landim. Tunneling and metastability of continuous time Markov chains II: the nonreversible case. *Journal of Statistical Physics*, 149(1):59–84, 2012.

[12] Jesús Beltrán and Claudio Landim. Tunneling and metastability of continuous time Markov chains. *Journal of Statistical Physics*, 140:1–50, 10 2009.

[13] Lorenzo Bertini, Davide Gabrielli, and Claudio Landim. Metastable Γ-expansion of finite state Markov chains level two large deviations rate functions. *Annals of Applied Probability*, 34(4):3820–3869, August 2024.

[14] Philippe Biane and Guillaume Chapuy. Laplacian matrices and spanning trees of tree graphs. *Ann. Fac. Sci. Toulouse Math. (6)*, 26(2):235–261, 2017.

[15] Patrick Billingsley. *Measure and Probability*. John Wiley and Sons: New York, 1995.

[16] Béla Bollobás. *Modern Graph Theory*. Graduate Texts in Mathematics 184. Springer-Verlag New York, 1 edition, 1998.

[17] Fabio Camilli and Claudio Marchi. A comparison among various notions of viscosity solution for Hamilton–Jacobi equations on networks. *Journal of Mathematical Analysis and Applications*, 407(1):112–118, 2013.

[18] Charles J. Colbourn, J.Scott Provan, and Dirk Vertigan. A new approach to solving three combinatorial enumeration problems on planar graphs. *Discrete Applied Mathematics*, 60(1):119–129, 1995.

[19] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

[20] Frank den Hollander. *Large deviations*. Fields Institute monographs. American Mathematical Society, United States, 2000.

[21] Manfredo P. do Carmo. *Differential Geometry of Curves and Surfaces*. Dover Publications, Mineola, New York, revised and updated second edition edition, 2016.

[22] Manfredo Perdigao do Carmo. *Riemannian Geometry*. Birkhauser, Boston, 1992.

[23] Lawrence C. Evans. *Weak KAM Theory and Partial Differential Equations*, pages 123–154. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[24] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010.

[25] Alessandra Faggionato and Davide Gabrielli. A representation formula for large deviations rate functionals of invariant measures on the one dimensional torus. *Annales de l Institut Henri Poincaré Probabilités et Statistiques*, 48, 01 2010.

[26] Albert Fathi. Weak KAM Theorem in Lagrangian Dynamics. 88, 01 2014. Monograph.

[27] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984. Translated from the Russian by Joseph Szücs.

[28] Mark Freidlin. Metastable distributions of Markov chains with rare transitions. *Journal of Statistical Physics*, 2017.

[29] M.I. Freidlin, J. Szücs, and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. Springer, 2012.

[30] Yuan Gao and Jian-Guo Liu. A selection principle for weak KAM solutions via Freidlin-Wentzell large deviation principle of invariant measures. *SIAM J. Math. Anal.*, 55:6457–6495, 2022.

[31] Alexandre Gaudillière and Claudio Landim. A Dirichlet principle for non-reversible Markov chains and some recurrence theorems. *Probability Theory and Related Fields*, 158(3-4):883–932, 2014.

[32] Robert Graham and Tamás Tél. Weak-noise limit of Fokker-Planck models and nondifferentiable potentials for dissipative dynamical systems. *Physical review. A, General physics*, 31 2:1109–1122, 1985.

[33] Jonathan Hermon. Some inequalities for reversible Markov chains and branching random walks via spectral optimization. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 58, 08 2022.

[34] H.R. Jauslin. Nondifferentiable potentials for nonequilibrium steady states. *Physica A: Statistical Mechanics and its Applications*, 144(1):179–191, None 1987.

[35] Faezeh Khodabandehlou, Christian Maes, Irene Maes, and Karel Netočný. The vanishing of excess heat for nonequilibrium processes reaching zero ambient temperature. *Annales Henri Poincaré*, 25(7):3371–3403, 2024.

[36] S. Kim and C. Landim. Γ-expansion of the measure-current large deviations rate functional of non-reversible finite-state Markov chains. *Stochastic Processes and their Applications*, 2025. In press.

[37] Claude Kipnis and Claudio Landim. *Scaling Limits of Interacting Particle Systems*, volume 320 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1999.

[38] Gustav R. Kirchhoff. Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Annalen der Physik*, 148:497–508.

[39] Douglas Klein and Milan Randic. Resistance distance. *Journal of Mathematical Chemistry*, 12:81–95, 12 1993.

[40] C. Landim. A topology for limits of Markov chains. *Stochastic Processes and their Applications*, 125(3):1058–1088, 2015.

[41] C. Landim. Metastability from the large deviations point of view: A Γ-expansion of the level two large deviations rate functional of non-reversible finite-state Markov chains. *Stochastic Processes and their Applications*, 165(C):275–315, None 2023.

[42] Claudio Landim. Metastable Markov chains. *Probability Surveys*, 16:143–227, 2019.

[43] Claudio Landim, Michail Loulakis, and Mustapha Mourragui. Metastable Markov chains: from the convergence of the trace to the convergence of the finite-dimensional distributions. *arXiv: Probability*, 2017.

[44] Claudio Landim, Michail Loulakis, and Mustapha Mourragui. Metastable Markov chains: from the convergence of the trace to the convergence of the finite-dimensional distributions. *Electronic Journal of Probability*, 23:1–34, 2018.

[45] Claudio Landim, D. Marcondes, and I. Seo. Non-equilibrium functional inequalities for finite Markov chains. *Preprint*, January 2026.

[46] Claudio Landim, Ricardo Misturini, and Federico Sau. Full Γ-expansion of reversible Markov chains level two large deviations rate functionals. *The Annals of Applied Probability*, 34(6):5578–5614, December 2024.

[47] Claudio Landim and T. Xu. Metastability of finite state Markov Chains: a recursive procedure to identify slow variables for model reduction. *ALEA : Latin American Journal of Probability and Mathematical Statistics*, 2016.

[48] Andrea Mennucci. On asymmetric distances. *Analysis and Geometry in Metric Spaces [electronic only]*, 1, 01 2013.

[49] Laurent Miclo and Pierre Patie. On intertwining relations between Ehrenfest, Yule and Ornstein-Uhlenbeck processes. *Lecture Notes in Mathematics - Springer-verlag-*, 11 2019.

[50] Hiroyoshi Mitake and Kohei Soga. Weak KAM theory for discount Hamilton-Jacobi equations and its application. 11 2016.

[51] J. R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, UK, 1997.

[52] E. Olivieri and E. Scoppola. Markov chains with exponentially small transition probabilities: First exit problem from a general domain. I. the reversible case. *Journal of Statistical Physics*, 79(3):613–647, 1995.

[53] E. Olivieri and E. Scoppola. Markov chains with exponentially small transition probabilities: First exit problem from a general domain. II. The general case. *Journal of Statistical Physics*, 84(5-6):987–1041, September 1996.

[54] Jim Pitman. *Combinatorial stochastic processes. Ecole d'Eté de Probabilités de Saint-Flour XXXII – 2002*, volume 1875. 01 2006.

[55] Jim Pitman and Wenpin Tang. Tree formulas, mean first passage times and Kemeny's constant of a Markov chain. *Bernoulli*, 24(3):1942–1972, 2018.

[56] André Schlichting and Martin Slowik. Poincaré and logarithmic Sobolev constants for metastable Markov chains via capacitary inequalities. *The Annals of Applied Probability*, 29(6):3438–3488, 2019.

[57] Alexander Schrijver. Combinatorial optimization: Polyhedra and efficiency. *Journal of Computer and System Sciences - JCSS*, B, 01 2003.

[58] E. Scoppola. Renormalization and graph methods for Markov chains. In S. Albeverio, R. Figari, E. Orlandi, and A. Teta, editors, *Advances in Dynamical Systems and Quantum Physics*, pages 261–280. World Scientific, 1995.

[59] Elisabetta Scoppola. Renormalization group for Markov chains and application to metastability. *Journal of Statistical Physics*, 73(1):83–121, 1993.

[60] Elisabetta Scoppola. *Metastability for Markov Chains: A General Procedure Based on Renormalization Group Ideas*, pages 303–322. Springer Netherlands, Dordrecht, 1994.

[61] Richard P. Stanley. *Enumerative Combinatorics, Volume 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.