Random processes with long memory

PhD thesis

Illés Horváth

Institute of Mathematics, Budapest University of Technology and Economics

Supervisor: prof. Bálint Tóth
Advisor: prof. Miklós Telek

2015

Contents

Acknowledgements

I would like to express my gratitude to my supervisor, prof. Bálint Tóth for introducing me to the field of probability theory and prof. Miklós Telek for introducing me into queuing theory. I thank both of them and Bálint Vető for the valuable discussions and collaboration.

I am grateful for the people at the Department of Stochastics at the Institute of Mathematics, and also the people at the MTA-BME Information Systems Research Group for providing an inspiring environment.

Last, but not least, I would like to thank my family for the support throughout.

Chapter 1

Introduction

Markov processes have been widely examined; the theory is well developed and applications are abundant. Different fields of applications include statistical mechanics, chemistry, economics, population dynamics and queueing theory.

As models become more and more complicated, a natural need arises to extend results available for Markov processes to systems where the Markov property does not fully hold, that is, to random processes with long memory. The exact nature of the memory in such systems can be very different; in mathematical physics, examples include interacting particle systems or a single particle moving in a random environment. In these cases, the memory corresponds to the state of the environment. In queueing theory, non-exponential service or interarrival times lead to M/G/1 and G/M/1 queues respectively; in such cases, the memory corresponds to the age of non-exponential clocks.

Non-Markovian behaviour can be handled using several different approaches. First, the state space may be extended to include more information about the process in order to make it Markovian. The difficulty of this approach is that the state space may end up being extremely large and difficult to handle. Nevertheless, this approach works for many physical systems, and theory has been constantly developed over the last decades. Two chapters of this thesis are related to this approach; Chapter 2 provides new theoretical tools called 'sector conditions" for such systems and Chapter 3 deals with a specific class of physical models (the so-called "true self-avoiding random walk").

For basic queueing models, matrix analytic methods are available, and direct calculations are also possible using Laplace–Stieltjes transform. These are established and straightforward methods [6], [44]. For more involved queuing models, another way to handle non-Markovian behaviour is via approximation by Markovian processes. General distributions may be approximated by specific classes of distributions that result in Markovian models. One of the most relevant classes of distributions for Markovian modelling is *phase-type distributions*. Chapter 4 deals with a question related to phase-type distributions.

Chapter 5 discusses a non-Markovian population model where generally-timed (non-exponential) transitions are allowed. The main goal is to find the mean-field limit of such a model (called *population generalized semi-Markov process*, PGSMP), and give a rigorous proof.

While all results in the present thesis deal with random processes with long memory, the results of Chapters 2 and 3 are fundamentally different from Chapters 4 and 5; Chapters 2 and 3 are based on the papers [28] and [29], coauthored with Bálint Tóth and Bálint Vető and require a background on operators in infinite-dimensional Hilbert spaces. Chapters 4 and 5 are based on the papers [26] and [23], which are coauthored with Miklós Telek and [23] also with Richard Hayden; Chapter 4 requires a background on matrix analysis, elementary functions and approximations, while Chapter 5 relies on Poisson representation and a number of classical probability concentration results.

The rest of this chapter gives an introduction and a varying level of setup to each of the four main topics.

1.1 Sector conditions

The theory of central limit theorems for additive functionals of Markov processes via martingale approximation was initiated in the mid-1980-s with applications to tagged particle diffusion in stochastic interacting particle systems and various models of random walks in random environment.

The Markov process is usually assumed to be in a stationary and ergodic regime. There are, however, also other types of related results, see e.g. [40], [14], which use partly different techniques.

In their celebrated 1986 paper [31], C. Kipnis and S. R. S. Varadhan proved a central limit theorem for the reversible case with no assumptions other than the strictly necessary ones, namely finiteness of the asymptotic variance of the properly scaled random variable. For an early non-reversible extension see [58] where the martingale approximation was applied to a particular model of random walk in random environment.

The theory has since been widely extended by Varadhan and collaborators to include processes with a varying degree of non-reversibility. Sufficient conditions for the central limit theorem are traditionally called *sector conditions*; for a detailed account of sector conditions and the different models they are applied to, see the surveys [47], [33] and [32].

In Chapter 2, we will discuss an improved version of the so-called *graded sector condition* [53], along with a new type of sector condition called the relaxed sector condition [28].

An application for the graded sector condition called the *true self-avoiding random walk* is given in Chapter 3; the graded sector condition guarantees Gaussian scaling limit in dimensions 3 and higher.

No application for the relaxed sector condition is given in the present thesis; however, an application is given in [34] for random walks in divergence-free random drift fields.

1.2 True self-avoiding random walk

The *'true' (or myopic) self-avoiding walk* model (TSAW) was introduced in the physics literature by Amit, Parisi and Peliti in [1]. This is a nearest neighbor non-Markovian random walk in $\mathbb{Z}^d$ which prefers to jump to those neighbors which were less visited in the past. Long memory effects are caused

by a path-wise self-repellence of the trajectories due to a push by the negative gradient of (softened) local time.

Let $t \mapsto X(t) \in \mathbb{Z}^d$ be a continuous time nearest neighbor jump process on the integer lattice $\mathbb{Z}^d$ whose law is given as follows:

$$\mathbf{P}\big(X(t+\mathrm{d}t) = y \,\big|\, \mathcal{F}_t, X(t) = x\big) = \mathbb{1}_{\{|\,x-y\,|=1\}} w(\ell(t,x) - \ell(t,y))\,\mathrm{d}t + o(\mathrm{d}t) \tag{1.1}$$

where

$$\ell(t,z) := \ell(0,z) + |\,\{0 \le s \le t : X(s) = z\}\,| \qquad z \in \mathbb{Z}^d \tag{1.2}$$

is the occupation time measure of the walk $X(t)$ with some initial values $\ell(0,z) \in \mathbb{R}$, $z \in \mathbb{Z}^d$, and the self-interaction rate function w is assumed to be increasing (more precisely formulated assumptions follow in Chapter 2). This is a continuous time version of the *'true' self-avoiding random walk* defined in [1].

Non-rigorous (but nevertheless convincing) scaling and renormalization group arguments suggest the following dimension-dependent asymptotic scaling behaviour (see e.g. [1], [45], [48]):

– In $d = 1$: $X(t) \sim t^{2/3}$ with intricate, non-Gaussian scaling limit.

– In $d = 2$: $X(t) \sim t^{1/2}(\log t)^{\zeta}$ and Gaussian (that is Wiener) scaling limit expected. (We note that actually there is some controversy in the physics literature about the value of the exponent ζ in the logarithmic correction.)

– In $d \ge 3$: $X(t) \sim t^{1/2}$ with Gaussian (i.e. Wiener) scaling limit expected.

In $d = 1$, for some particular cases of the model (discrete time TSAW with edge, rather than site repulsion and continuous time TSAW with site repulsion, as defined above), the limit theorem for $t^{-2/3}X(t)$ was established in [59], respectively, [61] with the truly intricate limiting distribution identified. The limit of the *process* $t \mapsto N^{-2/3}X(Nt)$ was constructed and analyzed in [62].

In $d = 2$, for the isotropic model exposed above, we expect the value $\zeta = 1/4$ in the logarithmic correction. For a modified, anisotropic version of the model where self-repulsion acts only in one spatial (say, the horizontal) direction, the exponent $\zeta = 1/3$ is expected. Superdiffusive lower bounds of order $t^{1/2}(\log\log t)^{1/2}$ for the isotropic case, respectively, of order $t^{1/2}(\log t)^{1/4}$ for the anisotropic case, have been proved for these two-dimensional models, cf. [60].

We address the $d \ge 3$ case in Chapter 3.

First, we identify a natural stationary (in time) and ergodic distribution of the environment (the local time profile) as seen from the moving particle. The main results are diffusive limits. For a wide class of self-interaction functions, we establish diffusive lower and upper bounds for the displacement and for a particular, more restricted class of interactions, we prove full CLT for the finite dimensional distributions of the displacement.

These results settle part of the conjectures in [1]. The proof of the CLT follows the non-reversible version of Kipnis – Varadhan theory. On the way to the proof, we slightly weaken the so-called *graded sector condition*.

A closely related model to the TSAW is the so-called *self-repelling Brownian polymer*, which is essentially the continuous-space counterpart of TSAW. For diffusive bounds for the self-repelling Brownian polymer in 1-dimension, see [56], and for dimensions $d \geq 3$, see [28] and the PhD thesis of Bálint Vető [64].

1.3 Phase-type distributions

Consider a continuous-time Markov chain on $n+1$ states with exactly one absorbing state. We assume that the initial probability distribution of the absorbing state is 0. Let X denote the time of absorption; its probability density function (pdf) is the following function $f : \mathbb{R}^+ \to \mathbb{R}^+$:

$$f(t) = -\alpha \mathbf{A} e^{t\mathbf{A}} \mathbf{1}, \quad t \geq 0, \tag{1.3}$$

where α is the initial row vector of size n (not including the absorbing state), and $\mathbf{A}$ is the vanishing infinitesimal generator; it is essentially the infinitesimal generator of the Markov chain, with the absorbing state removed. That is, $\mathbf{A}$ is a substochastic matrix of size $n \times n$, where the sum of row i is equal to the negative of the rate of absorption from state i. $\mathbf{1}$ is the column vector of size n whose elements are all equal to 1. The 0 initial probability of absorption corresponds to $\alpha\mathbf{1} = 1$; equivalently, X does not have a probability mass at 0.

Distributions that can be obtained in the above form are called *phase-type distributions*; the class of all such distributions will be denoted by PH. Phase-type distributions can be regarded as a generalization of exponential distributions (which correspond to $n = 1$ in the above definition) that can exhibit a wide range of behaviour while still being subject to Markovian modelling techniques due to the stochastic interpretation above.

Phase-type distributions can be used to approximate general distributions; PH is dense in total variation distance among all absolutely continuous positive distributions [6].

The pdf of a phase-type distribution is always analytic and takes the form

$$f(t) = \sum_i \sum_{j=1}^{n_i} c_{\lambda_i, j} t^{j-1} e^{-\lambda_i t}$$

where $-\lambda_i$ are the eigenvalues of $\mathbf{A}$, n_i is the multiplicity of λ_i and $c_{\lambda_i, j}$ are constants.

For a given f in PH, α and $\mathbf{A}$ are not unique; not even their dimensions are unique. Hence it makes sense to call the pair $(\alpha, \mathbf{A})$ a representation for f if (1.3) holds.

Before proceeding, we give the following precise definition of the class PH:

Definition 1. *The nonnegative random variable X with density function f_X is in the class PH if there exists a vector α of size n and a matrix $\mathbf{A}$ of size $n \times n$ for some finite n such that*

$$f_X(t) = -\alpha \mathbf{A} e^{t\mathbf{A}} \mathbf{1}, \quad t \geq 0,$$

and

- α *is nonnegative,*
- $\alpha\mathbf{1} = 1$, *where* $\mathbf{1}$ *denotes the column vector of size* n *whose elements are all equal to 1 (0 probability mass at zero),*
- $\mathbf{A}_{ij} \geq 0$ *for* $i \neq j$
- $\mathbf{A1}$ *is nonpositive, and*
- *the MC is eventually absorbed with probability 1.*

In this case, we will also say that f_X *is PH*$(\alpha, \mathbf{A})$*-distributed.*

Note that eventual absorption can also be characterized in a purely algebraic manner, based only on the position of nonzero elements in α and $\mathbf{A}$: for any index i for which there exists a sequence of indices $i_{-k}, \ldots, i_{-1}, i_0 = i$ such that $\alpha_{i_{-k}} > 0$ and $\mathbf{A}_{i_{-j}, i_{-j+1}} > 0$ for every $j = -k, \ldots, -1$ (that is, the Markov chain enters state i with a positive probability) there must exist a sequence $i = i_0, i_1, \ldots, i_l$ such that $\mathbf{A}_{i_{j-1}, i_j} > 0$ for every $j = 1, \ldots, l$ and $(\mathbf{A1})_{i_l} < 0$ (the Markov chain vanishes from state i with a positive probability).

The size n of α (and $\mathbf{A}$) is called the *order* of the representation. A matrix satisfying the above conditions will be called Markovian; similarly, a nonnegative vector will be called a Markovian vector. The states of the Markov chain are often called phases.

A *minimal PH representation* is defined simply as a PH representation of minimal order. Finding a minimal PH representation for a given PH distribution is generally very difficult; no method is available that always succeeds in finding a minimal PH representation.

The class of *matrix exponential functions* (ME) is defined as follows:

Definition 2. *A nonnegative random variable* X *with probability density function* f *is in the class ME if there exists a vector* α *of size* n *and a matrix* $\mathbf{A}$ *of size* $n \times n$ *for some finite* n *such that*

$$f(t) = -\alpha \mathbf{A} e^{t\mathbf{A}} \mathbf{1}, \quad t \geq 0,$$

In this case, we will also say that f *(and* X*) is ME*$(\alpha, \mathbf{A})$*-distributed.*

The difference of an ME pdf compared to a PH pdf is that we do not pose nonnegativity conditions on α and $\mathbf{A}$ (α and $\mathbf{A}$ are usually assumed to be real; that said, during calculations, complex numbers work just as well). If either α has negative or $\mathbf{A}$ has negative offdiagonal elements, the stochastic interpretation that X is the time of absorption of a Markov chain is no longer available. The condition $\int_0^\infty f(t)\mathrm{d}t = 1$ implies $\alpha\mathbf{1} = 1$.

Clearly, PH is a subclass of ME. Again it makes sense to define a minimal ME representation for any ME (or PH) distribution as an ME representation of minimal order. For any PH distribution X, the order of a minimal ME representation is a lower bound on the order of a minimal PH representation.

The order of the minimal ME representation (and an actual minimal ME representation) can be found easily (see Lemma 4.2). Further properties of minimal ME representations are examined in Chapter 4.

In practice, the lack of stochastic interpretation for matrix-exponential functions can be an issue. The nonnegativity of the pdf can not be taken for granted, and may have to be checked. Without a stochastic interpretation, stochastic simulations are not possible either.

Many approximation methods are insensitive of the signs of elements of α and $\mathbf{A}$, and may thus result in a matrix-exponential representation instead of a phase-type representation. It is often useful to transform ME representations into PH representations if possible.

The difference between the two classes is characterized due to O'Cinneide [46]. Before that, we need two more definitions.

Definition 3. *f satisfies the* positive density condition *if*

$$f(t) > 0 \quad \forall t > 0.$$

Note that the definition allows the density at 0 to be equal to 0.

Definition 4. *f satisfies the* dominant eigenvalue condition *if for some minimal ME representation $(\alpha, \mathbf{A})$ of f, $\mathbf{A}$ has a single eigenvalue with maximal real part.*

The dominant eigenvalue is always real to avoid oscillation of f around 0; the above definition excludes the case when a is the dominant real eigenvalue and there is a pair of complex eigenvalues with the same real part. However, the multiplicity of a may be higher than 1. We also remark that if the dominant eigenvalue condition holds for some minimal ME representation, it holds for all minimal ME representations of f. This is further discussed in Chapter 4.

Now we are ready to state O'Cinneide's characterization theorem.

Theorem 1.1. *[46] If f_X is ME(α, $\mathbf{A}$) distributed, then f_X has a finite dimensional PH(β, $\mathbf{B}$) representation iff the following two conditions hold:*

- *f_X satisfies the dominant eigenvalue condition and*
- *f_X satisfies the positive density condition.*

The main importance of the theorem is the sufficient direction; that is, if the dominant eigenvalue condition and the positive density condition hold, then a PH representation always exists. For the necessary direction, the positive density condition follows directly from the stochastic interpretation, and the dominant eigenvalue condition is essentially a consequence of the Perron–Frobenius theorem. Nevertheless, proofs for the necessary direction are also included in Chapter 4.

A possible interpretation of the theorem is that ME distributions that violate either the dominant eigenvalue condition or the positive density condition are on the "border" of ME, while PH is the interior of the set ME in some sense (we do not define these intuitive ideas more precisely). A pdf

from ME\PH may be approximated by a sequence of PH distributions; however, the order of those representations goes to infinity. From this, one may easily get the idea that ME distributions that violate either the dominant eigenvalue condition or the positive density condition are analogous to the time of absorption of a Markov chain on an infinite state space. This is *not* the case; the time of absorption of an infinite vanishing Markov chain still satisfies the positive density condition (see Lemma 4.9; the proof works for the infinite case as well).

The original proof of O'Cinneide for the sufficient direction of the theorem is rather involved, using geometric properties of certain subspaces of PH distributions in high-dimensional spaces. A quite different approach from Maier [38] uses Soittola's automata-theoretic algorithms [54].

Both [38] and [46] prove the characterization theorem, but use complex mathematical concepts, such as polytopes, or positive rational sequences.

The main contribution of Chapter 4 is a constructive proof for the sufficient part of the characterization theorem. We propose an explicit procedure for computing a PH representation of a matrix exponential function and showing that the procedure always terminates successfully if the matrix exponential function satisfies the positive density condition and the dominant eigenvalue condition.

Compared to existing results, one of the main advantages of the presented constructive proof is that it is rather elementary, using basic function and matrix theory and stochastic interpretation of Markov processes. It also links more recent results (such as the sparse monocyclic representation of [12]) to the characterization theorem.

1.4 Generalized semi-Markovian population models

A (homogeneous) Markov population model is defined as follows. Fix a positive integer N. Each of N individuals is inhabiting a state from a finite set $\mathcal{S}$. Each individual performs Markov transitions in continuous time: an individual in state i transitions to state j with rate r_{ij}^N. The rates may depend on the *global state of* the system; the global state of the system is defined as the total number of individuals in each state, that is, a vector $\mathbf{x}^N \in (\{0, 1, \ldots, N\})^{|\mathcal{S}|}$ with $x_1^N + \cdots + x_{|\mathcal{S}|}^N = N$. It is easy to see that the global state of the system $\mathbf{x}^N(t)$ is a continuous-time Markov chain.

We are interested in the behaviour of such a system for large values of N. A usual assumption is that a family of Markov population models is *density-dependent*; this means that the transition rates depend only on the *normalized global state* of the sytem, independent of N. The normalized global state of the system is defined as $\bar{\mathbf{x}}^N = \frac{\mathbf{x}^N}{N}$.

Density-dependence commonly occurs in real-life scenarios in the field of chemistry (chemical reaction speed may be affected by concentration), biology and many computer network applications. We will use a peer-to-peer software update model as a detailed example.

While the global state of the system is Markovian, an explicit analysis of this Markov chain is infeasible because the size of the state space increases exponentially in N.

The classic result of Kurtz [35] says that, upon some further regularity conditions (namely that r_{ij} are Lipschitz-continuous and the initial conditions converge), the evolution of a density-dependent

Markov population model converges to the solution of a system of ordinary differential equations (ODEs) as $N \to \infty$. The main advantage of Kurtz's approach is that the size of the system of equations is $|\mathcal{S}|$ regardless of N, thus avoiding the state-space explosion issue. Another consequence is that the limit is deterministic: for large values of N, the behaviour of the global state of the system is very close to deterministic. (Of course, on an individual level, it is still random.) The deterministic limit is called the *mean-field limit* of the system. A precise formulation of Kurtz's theorem will follow in Chapter 5.

Our main goal in Chapter 5 is to extend the mean-field methodology of Kurtz to a class of models where non-Markovian transitions are also allowed. We will define a class of *population generalized semi-Markov processes (PGSMP)*. The notation used here is different from the usual notation for PGSMPs, which has its roots in formal modelling and Petri nets; we will stick to a notation close to classic Markov-chain notation.

Just like for the Markov population model described above, a PGSMP has a finite local state space $\mathcal{S}$; each of N individuals is inhabiting a state from $\mathcal{S}$, but apart from each individual making Markov transitions, some of the states have a so-called *active clock*. When an individual enters a state with an active clock, a generally-timed clock starts. The distribution of the time before the clock goes off may depend on the state. Once the clock goes off, the individual makes a transition to another state.

The two main assumptions concerning active clocks are that in each state, there is either zero or one active clock, and that active clocks do not compete with Markovian transitions; that is, if state i has an active clock, all Markovian rates r_{ij} are 0. This assumption is usually referred to as *delay-only*, as the non-Markovian transitions cause delays of random length between Markovian transitions.

In Chapter 5, we formulate and prove a result analogous to Kurtz's theorem; the main difference is that the mean-field limit is the solution of a system of delayed differential equations (DDEs), where the evolution of the system depends not just on the current state of the system, but also on its entire past. The change from ODEs to DDEs corresponds to the fact that a "memory" has been introduced to the system by the generally-timed clocks.

The motivation for the mean-field approach is the same as in the Markov case — unsurprisingly, generalized semi-Markov process models with many components also become computationally intractable to explicit state techniques [10, 13] rapidly as a result of the familiar state-space explosion problem.

Numerical DDE solvers are also available, making this approach practically applicable; that said, our focus is the precise formulation and rigorous proof of the mean-field convergence.

Related work can be found in the biology and chemistry literature. Systems of DDEs have been derived to approximate stochastic models of reaction networks where deterministic delays are possible after reactions occur [3, 9, 51]. However, these models differ from those considered here in a number of critical ways; most importantly, the current presentation lacks the severe rigidity of models encountered in biology and chemistry, making it suitable for a much larger class of population models.

There has been a recent interest for PGSMPs in a general framework; closest related work is due to [24] and [5] which both deal with deterministic delay-only PGSMPs in different ways. Our

presentation is closer in spirit to [24], but the upgrade from deterministic delays to generally-timed delays calls for a careful and involved analysis.

The approach in [5] highlights the connection to ODE approximations of DDEs [39] which is directly analogous to the Erlang approximation of the delay in the PGSMP. The current approach, however, avoids any Erlang approximations whatsoever, proving the mean-field limit directly via probability concentration theorems.

Chapter 2

Sector conditions

In this chapter we give a short overview of the classic martingale approximation and central limit theorem *à la Kipnis – Varadhan* [31] and the sufficient conditions that guarantee central limit approximation called sector conditions (strong sector condition [63] and graded sector condition [53]). Then we will present an improved version of the graded sector condition, and we will also present a new condition, which we call the *relaxed sector condition (RSC)* that generalizes the strong sector condition (SSC) and the graded sector condition (GSC) in the case when the self-adjoint part of the infinitesimal generator acts diagonally in the grading. The main advantage being that the proof of the GSC in this case is more transparent and less computational than in the original versions.

An application for the improved graded sector condition called the *true self-avoiding random walk* is given in Chapter 3; the graded sector condition guarantees Gaussian scaling limit in dimensions 3 and higher.

No application for the relaxed sector condition is given in the present thesis; however, an application is given in [34] for random walks in divergence-free random drift fields.

2.1 Setup, abstract considerations

We recall the non-reversible version of the abstract Kipnis – Varadhan CLT for additive functionals of ergodic Markov processes, see [31] and [58].

Let $(\Omega, \mathcal{F}, \pi)$ be a probability space: Ω is the state space of a *stationary and ergodic* Markov process $t \mapsto \eta(t)$. We put ourselves in the Hilbert space $\mathcal{H} := \mathcal{L}^2(\Omega, \pi)$. Denote the *infinitesimal generator* of the semigroup of the process by G, which is a well-defined (possibly unbounded) closed linear operator on $\mathcal{H}$.

The adjoint G^* is the infinitesimal generator of the semigroup of the reversed (also stationary and ergodic) process $\eta^*(t) = \eta(-t)$. It is assumed that G and G^* have a *common core of definition* $\mathcal{C} \subseteq \mathcal{H}$.

We denote the *symmetric* and *antisymmetric* parts of the generators G, G^*, by

$$S := -\frac{1}{2}(G + G^*), \qquad A := \frac{1}{2}(G - G^*).$$

(We prefer to use the notation S for the positive semidefinite operator defined above, so the infinitesimal generator will be written as $G = -S + A$.) These operators are also extended from $\mathcal{C}$ by graph closure and it is assumed that they are well-defined self-adjoint, respectively, skew-self-adjoint operators:

$$S^* = S \geq 0, \qquad A^* = -A.$$

Summarizing: it is assumed that the operators G, G^*, S and A have a common dense core of definition $\mathcal{C}$. Note that $-S$ is itself the infinitesimal generator of a Markovian semigroup on $\mathcal{L}^2(\Omega, \pi)$, for which the probability measure π is reversible (not just stationary). We assume that $-S$ is itself ergodic:

$$\mathrm{Ker}(S) = \{c\mathbb{1} : c \in \mathbb{C}\}.$$

We shall restrict ourselves to the subspace of codimension 1, orthogonal to the constant functions.

In the sequel the operators $(\lambda I + S)^{\pm 1/2}$, $\lambda \geq 0$, will play an important role. These are defined by the spectral theorem applied to the self-adjoint and positive operator S. $\mathcal{C}$ is also a core for the operators $(\lambda I + S)^{1/2}$, $\lambda \geq 0$. The operators $(\lambda I + S)^{-1/2}$, $\lambda > 0$, are everywhere defined and bounded, with $\left\| (\lambda I + S)^{-1/2} \right\| \leq \lambda^{-1/2}$. The operator $S^{-1/2}$ is defined on

$$\mathrm{Dom}(S^{-1/2}) := \left\{ f \in \mathcal{H} : \left\| S^{-1/2} f \right\|^2 := \lim_{\lambda \to 0} \left\| (\lambda I + S)^{-1/2} f \right\|^2 \leq \infty \right\} = \mathrm{Ran}(S^{1/2}). \tag{2.1}$$

Let $f \in \mathcal{H}$, such that $(f, \mathbb{1}) = \int_\Omega f \, \mathrm{d}\pi = 0$. We ask about CLT/invariance principle for

$$N^{-1/2} \int_0^{Nt} f(\eta(s)) \, \mathrm{d}s \tag{2.2}$$

as $N \to \infty$.

Assume

$$f \in \mathrm{Ran}(S^{1/2}). \tag{2.3}$$

We shall refer to (2.3) as the H_{-1}-*condition*. From standard variational arguments (see e.g. [32], [47] and [53]) it follows that (2.3) is a sufficient condition for the diffusive upper bound:

$$\overline{\lim_{t \to \infty}} \, t^{-1} \mathbf{E} \left(\left(\int_0^t f(\eta(s)) \, \mathrm{d}s \right)^2 \right) \leq 2 \left\| S^{-1/2} f \right\|. \tag{2.4}$$

We denote by R_λ the resolvent of the semigroup $s \mapsto e^{sG}$:

$$R_\lambda := \int_0^\infty e^{-\lambda s} e^{sG} \mathrm{d}s = \big(\lambda I - G\big)^{-1}, \qquad \lambda > 0, \tag{2.5}$$

and given $f \in \mathcal{H}$ as above, we will use the notation

$$u_\lambda := R_\lambda f.$$

The following theorem is direct extension to the general non-reversible setup of the Kipnis – Varadhan Theorem from [31]. It yields the efficient martingale approximation of the additive functional (2.2). To the best of our knowledge this non-reversible extension appears first in [58].

Theorem KV. *With the notation and assumptions as before, if the following two limits hold in $\mathcal{H}$:*

$$\lim_{\lambda \to 0} \lambda^{1/2} u_\lambda = 0, \tag{2.6}$$

$$\lim_{\lambda \to 0} S^{1/2} u_\lambda =: v \in \mathcal{H}, \tag{2.7}$$

then

$$\sigma^2 := 2 \lim_{\lambda \to 0} (u_\lambda, f) = 2 \parallel v \parallel^2 \in [0, \infty),$$

exists, and there also exists a zero mean $\mathcal{L}^2$-martingale $M(t)$ adapted to the filtration of the Markov process $\eta(t)$, with stationary and ergodic increments and variance

$$\mathbf{E}\left(M(t)^2\right) = \sigma^2 t,$$

such that

$$\lim_{N \to \infty} N^{-1} \mathbf{E}\left(\big(\int_0^N f(\eta(s))\, \mathrm{d}s - M(N) \big)^2 \right) = 0.$$

In particular, if $\sigma > 0$, then the finite dimensional marginal distributions of the rescaled process $t \mapsto \sigma^{-1} N^{-1/2} \int_0^{Nt} f(\eta(s))\, \mathrm{d}s$ converge to those of a standard 1d Brownian motion.

Remarks. ∘ For the historical record it should be mentioned that the idea of martingale approximation and an early variant of this theorem under the much more restrictive condition $f \in \mathrm{Ran}(G)$, appears in [22]. For more exhaustive historical account and bibliography of the problem see the recent monograph [32].

∘ The reversible case, when $A = 0$, was considered in the celebrated paper [31]. In that case conditions (2.6) and (2.7) are equivalent. The proof of the Theorem KV in the reversible case relies on spectral calculus.

- Conditions (2.6) and (2.7) of Theorem KV are jointly equivalent to the following:

$$\lim_{\lambda,\lambda'\to 0}(\lambda+\lambda')(u_\lambda,u_{\lambda'})=0. \tag{2.8}$$

  Indeed, straightforward computations yield:

$$(\lambda+\lambda')(u_\lambda,u_{\lambda'})=\left\|\, S^{1/2}(u_\lambda-u_{\lambda'})\,\right\|^2+\lambda\,\|\,u_\lambda\,\|^2+\lambda'\,\|\,u_{\lambda'}\,\|^2\,.$$

- The non-reversible formulation appears – in discrete-time Markov chain, rather than continuous-time Markov process setup and with condition (2.8) – in [58] where it was applied, with bare hands computations, to obtain CLT for a particular random walk in random environment. Its proof mainly follows the original proof of the Kipnis – Varadhan theorem from [31] with the difference that spectral calculus is replaced by resolvent calculus.

- In continuous-time Markov process setup, it was formulated in [63] and applied to tagged particle motion in non-reversible zero mean exclusion processes. In this paper, the *strong sector condition (SSC)* was formulated, which, together with the H_{-1}-condition (2.3) on the function $f\in\mathcal{H}$, provide sufficient conditions for (2.6) and (2.7) of Theorem KV to hold.

- In [53], the so-called *graded sector condition (GSC)* was formulated and Theorem KV was applied to tagged particle diffusion in general (non-zero mean) non-reversible exclusion processes, in $d\geq 3$. The fundamental ideas related to the GSC have their origin partly in [36].

- For a list of applications of Theorem KV together with the SSC and GSC, see the surveys [47], [32], and for a more recent application of the GSC to the so-called *myopic self-avoiding walks* and *Brownian polymers*, see [28].

2.2 Sector conditions

In Subsection 2.2.1 we recall the SSC. In Subsection 2.2.2 we present an improved version of the GSC. In subsection 2.2.3 we formulate the RSC, then we show how the SSC and the diagonal version of the GSC follow in a very natural way from RSC. The main gain is in simplifying the proof of the diagonal GSC; the proof of the RSC may be called the "proof from the book".

2.2.1 Strong sector condition

From abstract functional analytic considerations [31], it follows that the H_{-1}-condition (2.3) jointly with the following bound jointly imply (2.8), and hence the martingale approximation and CLT of Theorem KV:

$$\sup_{\lambda>0}\left\|\, S^{-1/2}Gu_\lambda\,\right\|<\infty. \tag{2.9}$$

Theorem SSC. *With notations as before, if there exists a constant $C < \infty$ such that for any $\varphi, \psi \in \mathcal{C}$, the common core of S and A,*

$$|\,(\psi, A\varphi)\,|^2 \leq C^2(\psi, S\psi)(\varphi, S\varphi) \tag{2.10}$$

then for any $f \in \mathcal{H}$ for which (2.3) *holds,* (2.9) *also follows. So for every function f for which* (2.3) *holds, the martingale approximation and CLT of Theorem KV applies automatically.*

Remark. ∘ Condition (2.10) is equivalent to requiring that the operator $S^{-1/2}AS^{-1/2}$ defined on the dense subspace $S^{1/2}\mathcal{C} := \{S^{1/2}\varphi : \varphi \in \mathcal{C}\}$ be bounded in norm by the constant C. Hence, by continuous extension, condition (2.10) is the same as

$$\left\| S^{-1/2}AS^{-1/2} \right\| \leq C < \infty. \tag{2.11}$$

2.2.2 Improved version of the graded sector condition

In the present section, we recall the non-reversible version of the Kipnis – Varadhan CLT for additive functionals of ergodic Markov processes and present an improved version of the *graded sector condition* of Sethuraman, Varadhan and Yau, [53].

We reformulate the graded sector condition from [47] and [32] in a somewhat enhanced version. Again, the next two conditions jointly imply (2.6) and (2.7) [31]:

$$f \in \mathrm{Ran}(S^{1/2}), \tag{2.12}$$

$$\sup_{\lambda>0} \left\| S^{-1/2}Gu_\lambda \right\| < \infty. \tag{2.13}$$

Assume that the Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega, \pi)$ is graded

$$\mathcal{H} = \overline{\oplus_{n=0}^{\infty} \mathcal{H}_n}, \tag{2.14}$$

and the infinitesimal generator is consistent with this grading in the following sense:

$$S = \sum_{n=0}^{\infty} \sum_{j=-r}^{r} S_{n,n+j}, \qquad S_{n,n+j} : \mathcal{H}_n \to \mathcal{H}_{n+j}, \qquad S^*_{n,n+j} = S_{n+j,n}, \tag{2.15}$$

$$A = \sum_{n=0}^{\infty} \sum_{j=-r}^{r} A_{n,n+j}, \qquad A_{n,n+j} : \mathcal{H}_n \to \mathcal{H}_{n+j}, \qquad A^*_{n,n+j} = -A_{n+j,n} \tag{2.16}$$

for some finite positive integer r. Here and in the sequel, the double sum $\sum_{n=0}^{\infty}\sum_{j=-r}^{r}\cdots$ is meant as
$\sum_{n=0}^{\infty}\sum_{j=-r}^{r} \mathbb{1}_{\{n+j\geq 0\}}\cdots$.

Theorem 2.1 (GSC). *Let the Hilbert space and the infinitesimal generator be graded in the sense specified above. Assume that there exists an operator $D = D^* \geq 0$ which acts diagonally on the grading*

of $\mathcal{H}$:

$$D = \sum_{n=0}^{\infty} D_{n,n}, \qquad D_{n,n} : \mathcal{H}_n \to \mathcal{H}_n \tag{2.17}$$

such that

$$0 \le D \le S. \tag{2.18}$$

Assume also that, with some $C < \infty$ and $2 \le \kappa < \infty$, the following bounds hold:

$$\left\| D_{n,n}^{-1/2}(S_{n,n} + A_{n,n})D_{n,n}^{-1/2} \right\| \le Cn^{\kappa}, \tag{2.19}$$

$$\left\| D_{n+j,n+j}^{-1/2} A_{n,n+j} D_{n,n}^{-1/2} \right\| \le \frac{n}{12r^2\kappa} + C, \qquad j = \pm 1, \dots, \pm r, \tag{2.20}$$

$$\left\| D_{n+j,n+j}^{-1/2} S_{n,n+j} D_{n,n}^{-1/2} \right\| \le \frac{n^2}{6r^3\kappa^2} + C, \qquad j = \pm 1, \dots, \pm r, \tag{2.21}$$

Under these conditions on the operators, for any function $f \in \oplus_{n=0}^{N} \mathcal{H}_n$, with some $N < \infty$, if

$$D^{-1/2} f \in \mathcal{H}, \tag{2.22}$$

then (2.12) *and* (2.13) *follow. As a consequence, the martingale approximation and CLT of Theorem KV hold.*

Remark 2.1. In the original formulation of the graded sector condition (see [53], [32] and [47]), the bound imposed in (2.21) on the symmetric part of the generator was of the same form as that imposed in (2.20) on the skew-symmetric part. We can go up to the bound of order n^2 (rather than of order n) in (2.21) due to decoupling of the estimates of the self-adjoint and skew self-adjoint parts. The proof follows the main lines of the original one with one extra observation which allows the enhancement mentioned above.

Proof. We present a proof following the main steps and notations used in [47] or [32]. The main difference, where we gain more in the upper bound imposed in (2.21) is in the bound (2.32). The expert should jump directly to comparing the bounds (2.31) and (2.32) to the bounds in the original.

Let

$$f = \sum_{n=0}^{N} f_n, \qquad u_\lambda = \sum_{n=0}^{\infty} u_{\lambda n}, \qquad f_n, u_{\lambda n} \in \mathcal{H}_n. \tag{2.23}$$

From (2.19), (2.20) and (2.21), it follows that

$$\left\| S^{-1/2} G u_\lambda \right\|^2 \le C \sum_n n^{2\kappa} \left\| D^{1/2} u_{\lambda n} \right\|^2 \tag{2.24}$$

with some $C < \infty$. So it suffices to prove that the right-hand side of (2.24) is bounded, uniformly in $\lambda > 0$.

Let

$$t(n) := n_1^{\kappa} \mathbb{1}_{\{0 \le n < n_1\}} + n^{\kappa} \mathbb{1}_{\{n_1 \le n \le n_2\}} + n_2^{\kappa} \mathbb{1}_{\{n_2 < n < \infty\}} \tag{2.25}$$

with the values of $0 < n_1 < n_2 < \infty$ to be fixed later, and define the bounded linear operator $T : \mathcal{H} \to \mathcal{H}$,

$$T \restriction_{\mathcal{H}_n} = t(n) I \restriction_{\mathcal{H}_n} . \tag{2.26}$$

In the end, n_1 will be large but fixed, and n_2 will go to ∞.

We start with the identity

$$\lambda(Tu_\lambda, Tu_\lambda) + (Tu_\lambda, STu_\lambda) = (Tu_\lambda, Tf) - (Tu_\lambda, [A,T]u_\lambda) + (Tu_\lambda, [S,T]u_\lambda) \tag{2.27}$$

obtained from the resolvent equation by manipulations. The key to the proof is controlling the order of the commutator terms on the right as precisely as possible. We point out here that separating the last two terms on the right-hand side rather than handling them jointly as $(Tu_\lambda, [T,G]u_\lambda)$ (as done in the original proof [47]) will allow for gain in the upper bound imposed in (2.21).

We get the following bounds via Schwarz:

$$\lambda(Tu_\lambda, Tu_\lambda) \geq 0, \tag{2.28}$$

$$\begin{aligned}(Tu_\lambda, STu_\lambda) &= \sum_n t(n)^2 (u_{\lambda n}, Su_{\lambda n}) \\ &= \sum_n t(n)^2 (S^{1/2}u_{\lambda n}, S^{1/2}u_{\lambda n}) \geq \sum_n t(n)^2 \left\| D^{1/2}u_{\lambda n} \right\|^2 ,\end{aligned} \tag{2.29}$$

$$\begin{aligned}(Tu_\lambda, Tf) &= \sum_n t(n)^2 (u_{\lambda n}, f_n) \\ &= \sum_n t(n)^2 \left(\frac{1}{\sqrt{2}} D^{1/2}u_{\lambda n}, \sqrt{2} D^{-1/2} f_n \right) \\ &\leq \frac{1}{4} \sum_n t(n)^2 \left\| D^{1/2}u_{\lambda n} \right\|^2 + \sum_n t(n)^2 \left\| D^{-1/2} f_n \right\|^2 .\end{aligned} \tag{2.30}$$

Now, the last two terms on the right-hand side of (2.27) follow. The second term (containing A) is treated just like in the original proof, the third term (containing S) slightly differently.

$$\begin{aligned}(Tu_\lambda, [A,T]u_\lambda) &= \frac{1}{2}(u_\lambda, (AT^2 - T^2A)u_\lambda) \\ &= \frac{1}{2} \sum_n \sum_{j=-r}^{r} \left(t(n)^2 - t(n+j)^2 \right) (u_{\lambda(n+j)}, A_{n,n+j} u_{\lambda n}) \\ &\leq \frac{1}{2} \sum_n \sum_{j=-r}^{r} \left| \, t(n)^2 - t(n+j)^2 \, \right| \left(\frac{n}{12 r^2 \kappa} + C \right) \left\| D^{1/2}u_{\lambda n} \right\| \left\| D^{1/2}u_{\lambda(n+j)} \right\| ,\end{aligned} \tag{2.31}$$

$$
\begin{aligned}
(Tu_\lambda, [S,T]u_\lambda) &= \frac{1}{2}(u_\lambda, (2TST - ST^2 - T^2S)u_\lambda) \qquad (2.32)\\
&= -\frac{1}{2}\sum_n \sum_{j=-r}^{r} \big(t(n) - t(n+j)\big)^2 (u_{\lambda(n+j)}, S_{n,n+j}u_{\lambda n})\\
&\le \frac{1}{2}\sum_n \sum_{j=-r}^{r} \big(t(n) - t(n+j)\big)^2 \left(\frac{n^2}{6r^3\kappa^2} + C\right) \left\| D^{1/2}u_{\lambda n} \right\| \left\| D^{1/2}u_{\lambda(n+j)} \right\|.
\end{aligned}
$$

Note the difference between the coefficients in the middle lines of (2.31), respectively, (2.32). Choosing n_1 sufficiently large, we get

$$
\begin{aligned}
&\sup_n \max_{-r\le j\le r} \frac{\left| t(n)^2 - t(n+j)^2 \right|}{t(n)^2} \left(\frac{n}{12r^2\kappa} + C\right)\\
\le &\sup_n \frac{\left| n^{2\kappa} - (n+r)^{2\kappa} \right|}{t(n)^2} \left(\frac{n}{12r^2\kappa} + C\right) \le \frac{1}{2(2r+1)} \qquad (2.33)
\end{aligned}
$$

since the main term in $\left| n^{2\kappa} - (n+r)^{2\kappa} \right|$ is $2r\kappa n^{2\kappa-1}$ and the main term in the entire expression is $\frac{1}{6r}$. Smaller order terms are arbitrarily small when n_1 is chosen large enough.

Similarly,

$$
\sup_n \max_{-r\le j\le r} \frac{\big(t(n) - t(n+j)\big)^2}{t(n)^2} \left(\frac{n^2}{6r^3\kappa^2} + C\right) \le \frac{1}{2(2r+1)}. \qquad (2.34)
$$

and hence, via another Schwarz,

$$
|\,(Tu_\lambda, [A,T]u_\lambda)\,| + |\,(Tu_\lambda, [S,T]u_\lambda)\,| \le \frac{1}{2}\sum_n t(n)^2 \left\| D^{1/2}u_{\lambda n} \right\|^2. \qquad (2.35)
$$

Putting (2.28), (2.29), (2.30) and (2.35) into (2.27), we obtain:

$$
\sum_n t(n)^2 \left\| D^{1/2}u_{\lambda n} \right\|^2 \le 4\sum_n t(n)^2 \left\| D^{-1/2}f_n \right\|^2 = 4\sum_{n=0}^{N} t(n)^2 \left\| D^{-1/2}f_n \right\|^2. \qquad (2.36)
$$

Finally, letting $n_2 \to \infty$, we get indeed (2.13) via (2.22) and (2.24). □

2.2.3 Relaxed sector condition

Let, as before, $\mathcal{C} \subset \mathcal{H}$ be a common core for the operators G, G^*, S and A. Note that for any $\lambda > 0$, $\mathcal{C} \subseteq \mathrm{Dom}((\lambda I + S)^{1/2})$ and the subspace

$$
(\lambda I + S)^{1/2}\mathcal{C} := \{(\lambda I + S)^{1/2}\varphi : \varphi \in \mathcal{C}\}
$$

is dense in $\mathcal{H}$. The operators

$$
B_\lambda : (\lambda I + S)^{1/2}\mathcal{C} \to \mathcal{H}, \quad B_\lambda := (\lambda I + S)^{-1/2}A(\lambda I + S)^{-1/2}, \qquad \lambda > 0, \qquad (2.37)
$$

are densely defined and skew-Hermitian, and thus closable. Actually it is the case that they are not only skew-Hermitian, but essentially skew-self-adjoint on $(\lambda I + S)^{1/2}\mathcal{C}$. Indeed, let $\chi \in \mathcal{C}$, $\varphi = (\lambda I + S)^{1/2}\chi$ and $\psi \in \mathcal{H}$, then

$$(\psi, (I \pm B_\lambda)\varphi) = ((\lambda I + S)^{-1/2}\psi, (\lambda I + S \pm A)\chi).$$

So, $\psi \perp \mathrm{Ran}(I \pm B_\lambda)$ implies $(\lambda I + S)^{-1/2}\psi \perp \mathrm{Ran}(\lambda I + S \pm A)$ and thus, since the operators $-S \pm A$ are Hille-Yosida-type, $(\lambda I + S)^{-1/2}\psi = 0$, and consequently $\psi = 0$ holds for any $\lambda > 0$. That is $\mathrm{Ran}(I \pm B_\lambda)$ is dense in $\mathcal{H}$. By slight abuse of notation we shall denote by the same symbol B_λ the skew-self-adjoint operators obtained by closure of the operators defined in (2.37).

The main point of the following theorem is that if there exists another skew-self-adjoint operator B, *formally* identified as

$$B := S^{-1/2} A S^{-1/2}, \tag{2.38}$$

and a sufficiently large subspace on which the sequence of operators B_λ converges pointwise (strongly) to B, as $\lambda \to 0$, then, the H_{-1}-condition (2.3) implies (2.6) and (2.7), and thus the martingale approximation and CLT of Theorem KV follow.

Theorem 2.2 (**Relaxed sector condition**). *Assume that there exist a subspace $\widetilde{\mathcal{C}} \subseteq \cap_{\lambda>0} \mathrm{Dom}(B_\lambda)$ which is still dense in $\mathcal{H}$ and an operator $B : \widetilde{\mathcal{C}} \to \mathcal{H}$ which is essentially skew-self-adjoint and such that for any vector $\varphi \in \widetilde{\mathcal{C}}$*

$$\lim_{\lambda \to 0} \| B_\lambda \varphi - B\varphi \| = 0. \tag{2.39}$$

Then, the H_{-1}-condition (2.3) *implies* (2.6) *and* (2.7)*, and thus the martingale approximation and CLT of Theorem KV follow.*

Remarks. ∘ Finding the appropriate subspace $\widetilde{\mathcal{C}}$ and defining the skew-Hermitian operator $B : \widetilde{\mathcal{C}} \to \mathcal{H}$ comes naturally. The difficulty in applying this criterion lies in proving that the operator B is not just skew-Hermitian, but actually skew-self-adjoint. That is, proving that

$$\overline{\mathrm{Ran}(I \pm B)} = \mathcal{H}. \tag{2.40}$$

This is the counterpart of *the basic criterion of self-adjointness*. See e.g. Theorem VIII.3. of [49]. Checking this is typically not easy in concrete cases.

∘ The statement and the proof of this theorem show close similarities with the Trotter-Kurtz theorem. See Theorem 2.12 in [37].

∘ Theorem SSC follows directly: In this case the operator B is actually *bounded* and thus automatically skew-self-adjoint, not just skew-Hermitian. In order to see (2.39) note that

$$B_\lambda = S^{1/2}(\lambda I + S)^{-1/2} B S^{1/2}(\lambda I + S)^{-1/2} \overset{\text{st.op.top.}}{\longrightarrow} B, \tag{2.41}$$

where $\overset{\text{st.op.top.}}{\longrightarrow}$ denotes convergence in the strong operator topology.

Proof. Since the operators B_λ, $\lambda > 0$, defined in (2.37) are a priori and the operator B is by assumption essentially skew-self-adjoint, we can define the following bounded operators (actually contractions):

$$\begin{aligned} &K_\lambda := (I - B_\lambda)^{-1}, && \| K_\lambda \| \leq 1, && \lambda > 0, \\ &K := (I - B)^{-1}, && \| K \| \leq 1. \end{aligned}$$

Hence, we can write the resolvent (2.5) as

$$R_\lambda = (\lambda + S)^{-1/2} K_\lambda (\lambda + S)^{-1/2}. \tag{2.42}$$

Lemma 2.3. *Assume that the sequence of bounded operators K_λ converges in the strong operator topology:*

$$K_\lambda \overset{\text{st.op.top.}}{\longrightarrow} K, \qquad \text{as} \qquad \lambda \to 0. \tag{2.43}$$

Then for any f satisfying the H_{-1}-condition (2.3)*,* (2.6) *and* (2.7) *hold.*

Proof. From the spectral theorem applied to the self-adjoint operator S, it is obvious that

$$\left\| \lambda^{1/2} (\lambda + S)^{-1/2} \right\| \leq 1, \qquad \lambda^{1/2} (\lambda + S)^{-1/2} \overset{\text{st.op.top.}}{\longrightarrow} 0, \tag{2.44}$$

$$\left\| S^{1/2} (\lambda + S)^{-1/2} \right\| \leq 1, \qquad S^{1/2} (\lambda + S)^{-1/2} \overset{\text{st.op.top.}}{\longrightarrow} I, \tag{2.45}$$

By condition (2.3) we can write

$$f = S^{1/2} g$$

with some $g \in \mathcal{H}$. Now, using (2.42), we get

$$\lambda^{1/2} u_\lambda = \lambda^{1/2} (\lambda + S)^{-1/2} K_\lambda (\lambda + S)^{-1/2} S^{1/2} g, \tag{2.46}$$

$$S^{1/2} u_\lambda = S^{1/2} (\lambda + S)^{-1/2} K_\lambda (\lambda + S)^{-1/2} S^{1/2} g. \tag{2.47}$$

From (2.43), (2.46), (2.47), (2.44) and (2.45), we readily get (2.6) and (2.7) with

$$v = Kg.$$

□

In the next lemma, we formulate a sufficient condition for (2.43) to hold. This is reminiscent of Theorem VIII.25(a) from [49]:

Lemma 2.4. *Let B_n, $n \in \mathbb{N}$, and $B = B_\infty$ be densely defined closed operators over the Hilbert space $\mathcal{H}$. Assume that*

(i) Some (fixed) $\mu \in \mathbb{C}$ is in the resolvent set of all operators B_n, $n \leq \infty$, and

$$\sup_{n \leq \infty} \left\| (\mu I - B_n)^{-1} \right\| < \infty. \tag{2.48}$$

(ii) There is a dense subspace $\widetilde{\mathcal{C}} \subseteq \mathcal{H}$ which is a core for B_∞ and $\widetilde{\mathcal{C}} \subseteq \mathrm{Dom}(B_n)$, $n < \infty$, such that for all $\widetilde{h} \in \widetilde{\mathcal{C}}$:

$$\lim_{n \to \infty} \left\| B_n \widetilde{h} - B \widetilde{h} \right\| = 0. \tag{2.49}$$

Then

$$(\mu I - B_n)^{-1} \overset{\text{st.op.top.}}{\longrightarrow} (\mu I - B)^{-1}. \tag{2.50}$$

Proof. Since $\widetilde{\mathcal{C}}$ is a core for the densely defined closed operator B and μ is in the resolvent set of B, the subspace

$$\widehat{\mathcal{C}} := \{\widehat{h} = (\mu I - B)\widetilde{h} \,:\, \widetilde{h} \in \widetilde{\mathcal{C}}\}$$

is dense in $\mathcal{H}$. Thus, for any $\widehat{h}$ from this dense subspace, we have

$$\left\{(\mu I - B_n)^{-1} - (\mu I - B)^{-1}\right\}\widehat{h} = (\mu I - B_n)^{-1}(B_n \widetilde{h} - B\widetilde{h}) \to 0,$$

due to (2.48) and (2.49). Using again (2.48), we conclude (2.50). □

Putting Lemmas 2.3 and 2.4 together, we obtain Theorem 2.2. □

As a direct consequence we formulate a version of Theorem GSC. The main advantage is actually in the proof: our proof is considerably less computational, more transparent and natural than the original one from [53], reproduced in a streamlined way in [47] and [32].

Assume the setup of Theorem GSC: the grading of the Hilbert space and the infinitesimal generator G acting consistently with the grading: (2.15) and (2.16). We assume that S is diagonal, that is, $S_{n,n+j} = 0$ for $j \neq 0$.

Proposition 1 (GSC from RSC). *If there exist two positive nondecreasing sequences d_n and c_n, such that*

$$d_n < \infty, \qquad \sum_{n=1}^{\infty} c_n^{-1} = \infty, \tag{2.51}$$

and for any $n, m \in \mathbb{N}$ and $\psi_m \in \mathcal{C}_m$, $\varphi_n \in \mathcal{C}_n$ the following bounds hold:

$$|\,(\psi_m, A_{m,n}\varphi_n)\,|^2 \leq (\delta_{m,n}d_n + (1-\delta_{m,n})c_n)\,(\psi_m, S_{m,m}\psi_m)(\varphi_n, S_{n,n}\varphi_n), \tag{2.52}$$

then the conditions of Theorem 2.2 hold with $\widetilde{\mathcal{C}} = \oplus_{n=1}^{\infty}\mathcal{H}_n$ (no closure!).

Proof. Let

$$\widetilde{\mathcal{C}} = \oplus_{n=1}^{\infty}\mathcal{H}_n.$$

Note that there is no closure of the orthogonal sum on the right hand side. Then the operator $B = S^{-1/2}AS^{-1/2}$ is defined on $\widetilde{\mathcal{C}}$ and is graded as

$$B = \sum_{\substack{m,n\geq 1 \\ |\,n-m\,|\leq r}} B_{m,n}, \qquad B_{m,n} : \mathcal{H}_n \to \mathcal{H}_m, \qquad B_{m,n} := S_{m,m}^{-1/2}A_{m,n}S_{m,m}^{-1/2}, \qquad B_{m,n}^* = -B_{n,m}.$$

Indeed, due to (2.52)

$$\|\,B_{n,m}\,\| \leq \delta_{m,n}d_n + (1-\delta_{m,n})c_n. \tag{2.53}$$

The operator $B : \widetilde{\mathcal{C}} \to \widetilde{\mathcal{C}}$ is clearly skew-Hermitian. In order to prove that it is actually essentially skew-self-adjoint we have to check (2.40).

For $\varphi \in \mathcal{H}$ we use the notation

$$\varphi = (\varphi_1, \varphi_2, \dots), \qquad \varphi^n := (\varphi_1, \varphi_2, \dots, \varphi_n, 0, 0, \dots).$$

In order to simplify the notation in the forthcoming argument we assume that $r = 1$. The cases with $r > 1$ are done exactly the same way, only notation becomes heavier.

Assume $\varphi \perp \mathrm{Ran}(I - B)$, then

$$0 = (\varphi, (I-B)\varphi^n) = \|\,\varphi^n\,\|^2 - (\varphi_{n+1}, B_{n+1,n}\varphi_n).$$

Hence, by (2.53) and letting n so large that $\|\,\varphi^n\,\|^2 \geq \|\,\varphi\,\|^2/2$,

$$\|\,\varphi_n\,\|^2 + \|\,\varphi_{n+1}\,\|^2 \geq \frac{2}{c_n}\,\|\,\varphi^n\,\|^2 \geq \frac{1}{c_n}\,\|\,\varphi\,\|^2\,.$$

Summing over n we obtain that $\varphi = 0$. This implies that $\mathrm{Ran}(I-B)$ is dense in $\mathcal{H}$. Identical argument works for $\mathrm{Ran}(I+B)$. This proves (2.40) and B is indeed essentially skew-self-adjoint on $\widetilde{\mathcal{C}}$.

Checking condition (2.49) is done exactly like in (2.41):

$$(B_\lambda)_{m,n} = S_{m,m}^{1/2}(\lambda I_{m,m} + S_{m,m})^{-1/2} B_{m,n} S_{n,n}^{1/2}(\lambda I_{n,n} + S_{n,n})^{-1/2}$$

$$\overset{\text{st.op.top.}}{\longrightarrow} B_{m,n},$$

as $\lambda \to 0$, since $\| B_{m,n} \| < \infty$ and $S_{m,m}^{1/2}(\lambda I_{m,m} + S_{m,m})^{-1/2} \overset{\text{st.op.top.}}{\longrightarrow} I_{m,m}$. □

2.3 Outlook

An application of the relaxed sector condition is given in [34] for stationary and ergodic random walks in divergence-free random drift fields. The RSC helps proving central limit theorem under more general conditions than previously and also simplifies the proof considerably.

For some models, the RSC might not be easy to verify directly. It would be interesting to formulate other consequences (apart from the graded sector condition) which are easier to check for specific models.

Chapter 3

Diffusive limits for "true" (or myopic) self-avoiding random walks in dimensions 3 and higher

The main structure of this chapter is as follows: In Section 3.1, we make the necessary formal preparations and formulate the main results of the chapter. In Section 3.2, the appropriate functional analytic formalism is prepared: the suitable $\mathcal{L}^2$ Hilbert spaces and the relevant linear operators are introduced and partly analyzed. Section 3.3 contains the so-called H_{-1}-bounds (that is: diffusive bounds) for the compensators appearing in the decomposition of displacement of the random motions investigated. In Section 3.4, the graded sector condition of Section 2.2.2 is verified. Section 3.5 gives an outlook.

We mention a very similar model, the *self-repelling Brownian polymer* model (SRBP), initiated in the probabilistic literature by Durrett and Rogers in [17]. It is effectively the continuous space-time counterpart of TSAW: a diffusion in $\mathbf{R}^d$ pushed by the negative gradient of the (mollified) occupation time measure of the process. The SRBP model can be handled using very similar machinery as the TSAW. This model is not part of the present thesis; for more details, see [28] and [64].

3.1 Formal setup and results

Let $t \mapsto X(t) \in \mathbb{Z}^d$ be a continuous time nearest neighbor jump process on the integer lattice $\mathbb{Z}^d$ whose law is defined by (1.1), with $w : \mathbb{R} \to (0, \infty)$ a fixed smooth "rate function" for which

$$\inf_{u \in \mathbb{R}} w(u) := \gamma > 0, \tag{3.1}$$

and denote by s and r its even, respectively, odd part:

$$s(u) := \frac{w(u) + w(-u)}{2} - \gamma, \qquad r(u) := \frac{w(u) - w(-u)}{2}. \tag{3.2}$$

Beside (3.1), we make the following assumptions: there exist constants $c > 0$, $\varepsilon > 0$ and $C < \infty$ such that

$$\inf_{u \in \mathbb{R}} r'(u) > c, \tag{3.3}$$

$$s(u) < C \exp\{(c - \varepsilon)u^2/2\}, \tag{3.4}$$

and, finally, we make the technical assumption that $r(\cdot)$ is an analytic function which satisfies:

$$\sum_{n=0}^{\infty} \left(\frac{2}{c - \varepsilon}\right)^{n/2} \left| r^{(n)}(0) \right| < \infty. \tag{3.5}$$

(The ε of (3.5) and the ε of (3.4) actually have different roles, but we can choose the smaller ε for both). Condition (3.1) is *ellipticity* which ensures that the jump rates of the random walk considered are *minorated* by an ordinary simple symmetric walk. Condition (3.3) ensures sufficient self-repellence of the trajectories and sufficient log-convexity of the stationary measure identified later. Conditions (3.4) and (3.5) are of technical nature and their role will be clarified later.

We consider the $d \geq 3$ cases. First, we identify a rather natural stationary and ergodic (in time) distribution of the environment (essentially: the local time profile) as seen from the position of the moving point. In this particular stationary and ergodic regime, we prove diffusive (that is of order t) bounds on the variance of $X(t)$ and *diffusive limit* (that is non-degenerate CLT with normal scaling) for the displacement.

It is natural to consider the local time profile as seen from the position of the random walker

$$\eta(t) = \big(\eta(t, x)\big)_{x \in \mathbb{Z}^d} \qquad \eta(t, x) := \ell(t, X(t) + x). \tag{3.6}$$

It is obvious that $t \mapsto \eta(t)$ is a c.a.d.l.a.g. Markov process on the state space

$$\Omega := \{\omega = \big(\omega(x)\big)_{x \in \mathbb{Z}^d} : \omega(x) \in \mathbb{R}, \ (\forall \varepsilon > 0) \lim_{|\, x \,| \to \infty} |\, x \,|^{-\varepsilon} \, |\, \omega(x) \,| = 0\}. \tag{3.7}$$

Note that we allow initial values $\ell(0, x) \in \mathbb{R}$ for the occupation time measure and thus $\ell(t, x)$ need not be non-negative. The group of spatial shifts

$$\tau_z : \Omega \to \Omega, \qquad \tau_z \omega(x) := \omega(z + x), \qquad z \in \mathbb{Z}^d \tag{3.8}$$

acts naturally on Ω.

Let

$$\mathcal{U} := \{e \in \mathbb{Z}^d : |\, e \,| = 1\}. \tag{3.9}$$

Throughout this chapter, we will denote by e the $2d$ unit vectors from $\mathcal{U}$ and by e_l, $l = 1, \ldots, d$, the unit vectors pointing in the positive coordinate directions.

The infinitesimal generator of the process $t \mapsto \eta(t)$, $f : \Omega \to \mathbb{R}$, is

$$Gf(\omega) = \sum_{e \in \mathcal{U}} w(\omega(0) - \omega(e))\big(f(\tau_e \omega) - f(\omega)\big) + \mathcal{D}f(\omega) \tag{3.10}$$

where the (unbounded) linear operator

$$\mathcal{D}f(\omega) := \frac{\partial f}{\partial \omega(0)}(\omega) \tag{3.11}$$

is well-defined for smooth cylinder functions.

The meaning of the various terms on the right-hand side of (3.10) is clear: the terms in the sum are due to the random shifts of the environment caused by the jumps of the random walker while the last term on the right-hand side is due to the deterministic linear growth of local time at the site actually occupied by the random walker.

Next, we define a probability measure on Ω which will turn out to be stationary and ergodic for the Markov process $t \mapsto \eta(t)$. Let

$$R : \mathbb{R} \to [0, \infty), \qquad R(u) := \int_0^u r(v)\, \mathrm{d}v. \tag{3.12}$$

R is strictly convex and even. We denote by $\mathrm{d}\pi(\omega)$ *the unique centered Gibbs measure* (Markov field) on Ω defined by the conditional specifications for $\Lambda \subset \mathbb{Z}^d$ finite:

$$\mathrm{d}\pi\left(\omega_\Lambda \,\middle|\, \omega_{\mathbb{Z}^d \setminus \Lambda}\right) = Z_\Lambda^{-1} \exp\left\{ -\frac{1}{2} \sum_{\substack{x,y \in \Lambda \\ |\, x-y\,| = 1}} R(\omega(x) - \omega(y)) - \sum_{\substack{x \in \Lambda, y \in \Lambda^c \\ |\, x-y\,| = 1}} R(\omega(x) - \omega(y)) \right\} \mathrm{d}\omega_\Lambda \tag{3.13}$$

where ω_Λ is the Lebesgue measure on Λ. Note that the (translation invariant) Gibbs measure given by the specifications (3.13) exists only in three and more dimensions. For information about gradient measures of this type, see [20]. The measure π is invariant under the spatial shifts and the dynamical system $(\Omega, \pi, \tau_z : z \in \mathbb{Z}^d)$ is *ergodic*.

Remark on terminology: The notion of ergodicity is used throughout this chapter in two ways: ergodicity with respect to space shifts and ergodicity time-wise. By time-wise ergodicity of a stationary Markov process $t \mapsto \eta(t)$ on the state space (Ω, π) with a.s. c.a.d.l.a.g. sample paths, we mean ergodicity with respect to time-shifts of the stationary measure over the path space. This is equivalent to the fact that 0 is non-degenerate eigenvalue of the infinitesimal generator of the semigroup acting on the Hilbert space $\mathcal{L}^2(\Omega, \pi)$.

In the particular case when $r(u) = u$, $R(u) = u^2/2$, the measure $\mathrm{d}\pi(\omega)$ is the distribution of the massless free Gaussian field on $\mathbb{Z}^d$, $d \geq 3$ with expectations and covariances

$$\int_\Omega \omega(x)\, \mathrm{d}\pi(\omega) = 0, \qquad \int_\Omega \omega(x)\omega(y)\, \mathrm{d}\pi(\omega) = (-\Delta)^{-1}_{x,y} =: C(y - x) \tag{3.14}$$

where Δ is the lattice Laplacian: $\Delta_{x,y} = \mathbb{1}_{\{|\, x-y\,|=1\}} - 2d\mathbb{1}_{\{|\, x-y\,|=0\}}$. We will refer to this special setup as *the Gaussian case*.

We are ready now to formulate the results regarding the lattice model.

Proposition 2. *The probability measure $\pi(\omega)$ is stationary and ergodic for the Markov process $t \mapsto \eta(t) \in \Omega$.*

The law of large numbers for the displacement of the random walker follows:

Corollary 1. *For π-almost all initial profiles $\ell(0,\cdot)$, almost surely*

$$\lim_{t\to\infty} \frac{X(t)}{t} = 0. \tag{3.15}$$

For the proof, see Section 3.3.

The main results of this chapter refer to the diffusive scaling limit of the displacement.

Theorem 3.1. *(1) If conditions* (3.1), (3.3), (3.4) *and* (3.5) *hold for the rate function, then*

$$0 < \gamma \le \inf_{|\, e\,|=1} \varliminf_{t\to\infty} t^{-1}\mathbf{E}\left((e\cdot X(t))^2\right) \le \sup_{|\, e\,|=1} \varlimsup_{t\to\infty} t^{-1}\mathbf{E}\left((e\cdot X(t))^2\right) < \infty. \tag{3.16}$$

(2) Assume that

$$r(u) = u, \qquad s(u) = s_4u^4 + s_2u^2 + s_0, \tag{3.17}$$

and we also make the technical assumption that s_4/γ is sufficiently small. Then the matrix of asymptotic covariances

$$\sigma^2_{kl} := \lim_{t\to\infty} t^{-1}\mathbf{E}\left(X_k(t)X_l(t)\right) \tag{3.18}$$

exists and it is non-degenerate. The finite dimensional distributions of the rescaled displacement process

$$X_N(t) := N^{-1/2}X(Nt) \tag{3.19}$$

converge to those of a d dimensional Brownian motion with covariance matrix σ^2.

Remarks. (1) We do not strive to obtain optimal constants in our conditions. The upper bound imposed on the ratio s_4/γ which emerges from the computations in the proof in Section 3.4 is rather restrictive but far from optimal.

(2) A drawback of the "Kipnis-Varadhan approach" explored in this chapter is that CLT for the displacement is obtained *in probability with respect to the stationary initial distribution* π. This method doesn't provide so-called quenched results, valid for particular or for almost all initial conditions, in some natural sense.

(3) It is clear that, in dimensions $d \geq 3$, other stationary distributions of the process $t \mapsto \eta(t)$ must exist. In particular, due to transience of the process $t \mapsto X(t)$, the stationary measure (presumably) reached from starting with "empty" initial conditions $\ell(0,x) \equiv 0$ certainly differs from our π. Our methods and results are valid for the particular stationary distribution $\mathrm{d}\pi$.

3.2 Spaces and operators, general case

For simplicity, assume that $s(.)$ is a polynomial of even degree; we will upgrade the proof to general analytic s in part (c) of Lemma 3.3.

We put ourselves in the Hilbert space $\mathcal{H} := \mathcal{L}^2(\Omega, \pi)$ and define some linear operators. The following shift and difference operators will be used:

$$T_e f(\omega) := f(\tau_e \omega), \qquad \nabla_e := T_e - I, \qquad \Delta := \sum_{e \in \mathcal{U}} \nabla_e = -\frac{1}{2} \sum_{e \in \mathcal{U}} \nabla_e \nabla_{-e}. \tag{3.20}$$

Their adjoints are

$$T_e^* = T_{-e}, \qquad \nabla_e^* = \nabla_{-e}, \qquad \Delta^* = \Delta. \tag{3.21}$$

Occasionally, we shall also use the notation $\nabla_l := \nabla_{e_l}$.

We also define the multiplication operators

$$M_e f(\omega) := s(\omega(0) - \omega(e)) f(\omega), \tag{3.22}$$

$$N_e f(\omega) := r(\omega(0) - \omega(e)) f(\omega), \qquad N := \sum_{e \in \mathcal{U}} N_e. \tag{3.23}$$

These are unbounded self-adjoint operators. The following commutation relations hold:

$$M_e T_e - T_e M_{-e} = 0 = N_e T_e + T_e N_{-e}. \tag{3.24}$$

The (unbounded) differential operator $\mathcal{D}$ is defined in (3.11), a priori on the dense subspace of smooth cylinder functions. It is a closable operator and the subspace of smooth cylinder functions serve as a core for the closure. We denote the closed operator by the same symbol. Integration by parts on (Ω, π) yields

$$\mathcal{D} + \mathcal{D}^* = 2N. \tag{3.25}$$

Next, we express the infinitesimal generator (3.10) of the $\mathcal{L}^2$-semigroup of the Markov process $t \mapsto \eta(t)$. The infinitesimal generator will be acting on a dense domain within the Hilbert space $\mathcal{L}^2(\Omega, \pi)$. Denote

$$S := -\frac{1}{2}(G + G^*), \qquad A := \frac{1}{2}(G - G^*) \tag{3.26}$$

the self-adjoint, respectively, skew self-adjoint parts of the infinitesimal generator. Using (3.24) and (3.25), we readily obtain

$$S = -\gamma\Delta + S_1, \tag{3.27}$$

$$S_1 = -\sum_{e\in\mathcal{U}} M_e\nabla_e = \frac{1}{2}\sum_{e\in\mathcal{U}} \nabla_{-e}M_e\nabla_e, \tag{3.28}$$

$$A = \sum_{e\in\mathcal{U}} N_eT_e + \big(\mathcal{D} - N\big). \tag{3.29}$$

Note that both $-\gamma\Delta$ and S_1 are *positive operators*. Actually, $\gamma\Delta$ is the infinitesimal generator of the process of "scenery seen by the random walker" (in the so-called RW in random scenery) and $-S_1$ is the infinitesimal generator of "environment seen by the random walker in a symmetric RWRE".

It is also worth noting that, defining the unitary involution

$$Jf(\omega) := f(-\omega), \tag{3.30}$$

we get

$$JSJ = S, \quad JAJ = -A, \qquad JGJ = G^*. \tag{3.31}$$

Stationarity drops out: indeed, $G^*\mathbb{1} = 0$. Actually, (3.31) means slightly more than stationarity: the time-reversed and flipped process

$$t \mapsto \widetilde{\eta}(t) := -\eta(-t) \tag{3.32}$$

is equal in law to the process $t \mapsto \eta(t)$. This time reversal symmetry is called *Yaglom reversibility* and it appears in many models with physical symmetries. See e.g. [15], [66], [67].

Ergodicity: for the Dirichlet form of the process $t \mapsto \eta(t)$ we have

$$(f, -Gf) = (f, Sf) \geq \gamma(f, -\Delta f) = \frac{1}{2}\sum_{e\in\mathcal{U}} \| \nabla_e f \|^2, \tag{3.33}$$

and hence, $Gf = 0$ implies $\nabla_e f = 0$, $e \in \mathcal{U}$, which, in turn, by ergodicity of the shifts on (Ω, π), implies $f = \text{const.}\mathbb{1}$. Hence, Proposition 2. Corollary 1 will be proved in section 3.3.

3.2.1 Spaces and operators, the Gaussian case

Spaces

In the case where $r(u) = u$, the stationary measure defined by (3.13) is Gaussian, and we can build up the Gaussian Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega, \pi)$ and its unitary equivalent representations as Fock spaces in the usual way.

We use the following convention for normalization of Fourier transform

$$\widehat{u}(p) = \sum_{x \in \mathbb{Z}^d} e^{ip \cdot x} u(x), \qquad u(x) = (2\pi)^{-d} \int_{(-\pi,\pi]^d} e^{-ip \cdot x} \widehat{u}(p) \, \mathrm{d}p, \tag{3.34}$$

and the shorthand notation

$$\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Z}^{dn}, \qquad x_m = (x_{m1}, \dots, x_{md}) \in \mathbb{Z}^d, \tag{3.35}$$

$$\mathbf{p} = (p_1, \dots, p_n) \in (-\pi, \pi]^{dn}, \qquad p_m = (p_{m1}, \dots, p_{md}) \in (-\pi, \pi]^d, \tag{3.36}$$

$m = 1, \dots, n$.

We denote by $\mathcal{S}_n$, respectively, $\widehat{\mathcal{S}}_n$, the Schwartz space of symmetric test functions of n variables on $\mathbb{Z}^d$, respectively, on $(-\pi, \pi]^d$:

$$\mathcal{S}_n := \{u : \mathbb{Z}^{dn} \to \mathbb{C} : u \text{ of rapid decay}, u(\varpi \mathbf{x}) = u(\mathbf{x}), \varpi \in \mathrm{Perm}(n)\}, \tag{3.37}$$

$$\widehat{\mathcal{S}}_n := \{\widehat{u} : [-\pi, \pi]^{dn} \to \mathbb{C} : \widehat{u} \in C^\infty, \widehat{u}(\varpi \mathbf{p}) = \widehat{u}(\mathbf{p}), \varpi \in \mathrm{Perm}(n)\}. \tag{3.38}$$

In the preceding formulas $\mathrm{Perm}(n)$ denotes the symmetric group of permutations acting on the n indices.

As noted before, in the case of $r(u) = u$, the random variables $\big(\omega(x) : x \in \mathbb{Z}^d\big)$ form the *massless free Gaussian field* on $\mathbb{Z}^d$ with expectation and covariances given in (3.14). The Fourier transform of the covariances $\widehat{C} : [-\pi, \pi]^d \to \mathbb{C}$ is

$$\widehat{C}(p) = \widehat{D}(p)^{-1} \tag{3.39}$$

where $\widehat{D} : [-\pi, \pi]^d \to \mathbb{R}$ is the Fourier transform of the lattice Laplacian:

$$\widehat{D}(p) := \sum_{l=1}^{d} \big(1 - \cos(p_l)\big). \tag{3.40}$$

We endow the spaces $\mathcal{S}_n$, respectively, $\widehat{\mathcal{S}}_n$ with the following scalar products

$$\langle u, v \rangle := \sum_{\mathbf{x} \in \mathbb{Z}^{dn}} \sum_{\mathbf{y} \in \mathbb{Z}^{dn}} C(\mathbf{x} - \mathbf{y}) \overline{u(\mathbf{x})} v(\mathbf{y}), \tag{3.41}$$

$$\langle \widehat{u}, \widehat{v} \rangle := \int_{[-\pi,\pi]^{dn}} \widehat{C}(\mathbf{p}) \overline{\widehat{u}(\mathbf{p})} \widehat{v}(\mathbf{p}) \, \mathrm{d}\mathbf{p} \tag{3.42}$$

where

$$C(\mathbf{x} - \mathbf{y}) := \prod_{m=1}^{n} C(x_m - y_m), \qquad \widehat{C}(\mathbf{p}) := \prod_{m=1}^{n} \widehat{C}(p^m). \tag{3.43}$$

Let $\mathcal{K}_n$ and $\widehat{\mathcal{K}}_n$ be the closures of $\mathcal{S}_n$, respectively, $\widehat{\mathcal{S}}_n$ with respect to the Euclidean norms defined by these inner products. The Fourier transform (3.34) realizes an isometric isomorphism between the

Hilbert spaces $\mathcal{K}_n$ and $\widehat{\mathcal{K}}_n$.

These Hilbert spaces are actually the symmetrized n-fold tensor products

$$\mathcal{K}_n := \mathrm{symm}\big(\mathcal{K}_1^{\otimes n}\big), \qquad \widehat{\mathcal{K}}_n := \mathrm{symm}\big(\widehat{\mathcal{K}}_1^{\otimes n}\big). \tag{3.44}$$

Finally, the full Fock spaces are

$$\mathcal{K} := \overline{\oplus_{n=0}^{\infty}\mathcal{K}_n}, \qquad \widehat{\mathcal{K}} := \overline{\oplus_{n=0}^{\infty}\widehat{\mathcal{K}}_n}. \tag{3.45}$$

The Hilbert space of our true interest is $\mathcal{H} = \mathcal{L}^2(\Omega, \pi)$. This is itself a graded Gaussian Hilbert space

$$\mathcal{H} = \overline{\oplus_{n=0}^{\infty}\mathcal{H}_n} \tag{3.46}$$

where the subspaces $\mathcal{H}_n$ are isometrically isomorphic with the subspaces $\mathcal{K}_n$ of $\mathcal{K}$ through the identification

$$\phi_n : \mathcal{K}_n \to \mathcal{H}_n, \quad \phi_n(u) := \frac{1}{\sqrt{n!}} \sum_{\mathbf{x}\in\mathbb{Z}^{dn}} u(\mathbf{x})\, {:}\omega(x_1)\dots\omega(x_n){:}\ . \tag{3.47}$$

Here and in the rest of this chapter, we denote by ${:}X_1 \dots X_n{:}$ the *Wick product* of the jointly Gaussian random variables $(X_1, \dots, X_n)$; the Wick product is an orthogonalized product of the variables defined recursively [21]:

$${:}X_i{:} := X_i - \mathbf{E}\,(X_i) \tag{3.48}$$

and

$$\frac{\partial\, {:}X_1 \dots X_n{:}}{\partial X_i} = {:}X_1 \dots X_{i-1}X_{i+1}\dots X_n{:} \tag{3.49}$$

$$\mathbf{E}\,({:}X_1 \dots X_n{:}) = 0. \tag{3.50}$$

(For example, ${:}X_1X_2{:} = X_1X_2 - X_1\mathbf{E}\,(X_2) - X_2\mathbf{E}\,(X_1) + 2\mathbf{E}\,(X_1)\,\mathbf{E}\,(X_2) - \mathbf{E}\,(X_1X_2)$.)

As the graded Hilbert spaces

$$\mathcal{H} := \overline{\oplus_{n=0}^{\infty}\mathcal{H}_n}, \quad \mathcal{K} := \overline{\oplus_{n=0}^{\infty}\mathcal{K}_n}, \quad \widehat{\mathcal{K}} := \overline{\oplus_{n=0}^{\infty}\widehat{\mathcal{K}}_n} \tag{3.51}$$

are isometrically isomorphic in a natural way, we shall move freely between the various representations.

Operators

First, we give the action of the operators ∇_e, Δ, etc. introduced in Subsection 3.2 on the spaces $\mathcal{H}_n$, $\mathcal{K}_n$ and $\widehat{\mathcal{K}}_n$. The point is that we are interested primarily in their action on the space $\mathcal{L}^2(\Omega, \pi) = \overline{\oplus_{n=0}^{\infty}\mathcal{H}_n}$, but explicit computations in later sections are handy in the unitary equivalent representations over the space $\widehat{\mathcal{K}} = \overline{\oplus_{n=0}^{\infty}\widehat{\mathcal{K}}_n}$. The action of various operators over $\mathcal{H}_n$ will be given in terms of the Wick monomials ${:}\omega(x_1)\dots\omega(x_n){:}$ and it is understood that the operators are extended by linearity and

graph closure. For technical details see [30].

- The operators ∇_e, $e \in \mathcal{U}$ map $\mathcal{H}_n \to \mathcal{H}_n$, $\mathcal{K}_n \to \mathcal{K}_n$, $\widehat{\mathcal{K}}_n \to \widehat{\mathcal{K}}_n$, in turn, as follows:

$$\nabla_e :\omega(x_1)\dots\omega(x_n): =:\omega(x_1+e)\dots\omega(x_n+e): - :\omega(x_1)\dots\omega(x_n): , \tag{3.52}$$

$$\nabla_e u(\mathbf{x}) = u(x_1-e,\dots,x_n-e) - u(x_1,\dots,x_n), \tag{3.53}$$

$$\nabla_e \widehat{u}(\mathbf{p}) = \big(\exp\left(i\textstyle\sum_{m=1}^n p_m\cdot e\right) - 1\big)\widehat{u}(\mathbf{p}). \tag{3.54}$$

- The operator Δ maps $\mathcal{H}_n \to \mathcal{H}_n$, $\mathcal{K}_n \to \mathcal{K}_n$, $\widehat{\mathcal{K}}_n \to \widehat{\mathcal{K}}_n$, in turn, as follows:

$$\Delta :\omega(x_1)\dots\omega(x_n): = \sum_{e\in\mathcal{U}} :\omega(x_1+e),\dots,\omega(x_n+e): -2d :\omega(x_1)\dots\omega(x_n): , \tag{3.55}$$

$$\Delta u(\mathbf{x}) = \sum_{e\in\mathcal{U}} u(x_1+e,\dots,x_n+e) - 2du(\mathbf{x}), \tag{3.56}$$

$$\Delta\widehat{u}(\mathbf{p}) = -2\widehat{D}\left(\textstyle\sum_{m=1}^n p_m\right)\widehat{u}(\mathbf{p}). \tag{3.57}$$

- The operators $|\,\Delta\,|^{-1/2}\nabla_e$ map $\mathcal{H}_n \to \mathcal{H}_n$, $\mathcal{K}_n \to \mathcal{K}_n$, $\widehat{\mathcal{K}}_n \to \widehat{\mathcal{K}}_n$. There is no explicit expression for the first two. The action $\widehat{\mathcal{K}}_n \to \widehat{\mathcal{K}}_n$ is as follows:

$$|\,\Delta\,|^{-1/2}\nabla_e\widehat{u}(\mathbf{p}) = \frac{\exp\left(i\sum_{m=1}^n p_m\cdot e\right)-1}{\sqrt{2\widehat{D}\left(\sum_{m=1}^n p_m\right)}}\widehat{u}(\mathbf{p}). \tag{3.58}$$

  These are *bounded* operators with norm

$$\Big\|\,|\,\Delta\,|^{-1/2}\nabla_e\,\Big\| \le 1. \tag{3.59}$$

- The creation operators a_e^*, $e \in \mathcal{U}$ map $\mathcal{H}_n \to \mathcal{H}_{n+1}$, $\mathcal{K}_n \to \mathcal{K}_{n+1}$, $\widehat{\mathcal{K}}_n \to \widehat{\mathcal{K}}_{n+1}$, in turn, as follows:

$$a_e^* :\omega(x_1)\dots\omega(x_n): =:(\omega(0)-\omega(e))\omega(x_1)\dots\omega(x_n): , \tag{3.60}$$

$$a_e^* u(x_1,\dots,x_{n+1}) = \frac{1}{\sqrt{n+1}}\sum_{m=1}^{n+1}\big(\delta_{x_m,0}-\delta_{x_m,e}\big)u(x_1,\dots,\cancel{x_m},\dots,x_{n+1}), \tag{3.61}$$

$$a_e^*\widehat{u}(p_1,\dots,p_{n+1}) = \frac{1}{\sqrt{n+1}}\sum_{m=1}^{n+1}\big(e^{ip_m\cdot e}-1\big)\widehat{u}(p_1,\dots,\cancel{p_m},\dots,p_{n+1}). \tag{3.62}$$

  The creation operators a_e^* restricted to the subspaces $\mathcal{H}_n$, $\mathcal{K}_n$, respectively, $\widehat{\mathcal{K}}_n$ are bounded with operator norm

$$\|\,a_e^*\restriction_{\mathcal{H}_n}\,\| = \|\,a_e^*\restriction_{\mathcal{K}_n}\,\| = \Big\|\,a_e^*\restriction_{\widehat{\mathcal{K}}_n}\,\Big\| = \sqrt{2C(0)-C(e)-C(-e)}\sqrt{n+1}. \tag{3.63}$$

- The annihilation operators a_e, $e \in \mathcal{U}$ map $\mathcal{H}_n \to \mathcal{H}_{n-1}$, $\mathcal{K}_n \to \mathcal{K}_{n-1}$, $\widehat{\mathcal{K}}_n \to \widehat{\mathcal{K}}_{n-1}$, in turn, as

follows:

$$a_e :\omega(x_1)\dots\omega(x_n): = \sum_{m=1}^{n} \big(C(x_m+e)-C(x_m)\big) :\omega(x_1)\dots\cancel{\omega(x_m)}\dots\omega(x_n): , \tag{3.64}$$

$$a_e u(x_1,\dots,x_{n-1}) = \sqrt{n}\sum_{z\in\mathbb{Z}^d} \big(C(z+e)-C(z)\big)u(x_1,\dots,x_{n-1},z), \tag{3.65}$$

$$a_e \widehat{u}(p_1,\dots,p_{n-1}) = \sqrt{n}(2\pi)^{-d}\int_{[-\pi,\pi]^d} \big(e^{-iq\cdot e}-1\big)\widehat{C}(q)\widehat{u}(p_1,\dots,p_{n-1},q)\,\mathrm{d}q. \tag{3.66}$$

The annihilation operators a_e restricted to the subspaces $\mathcal{H}_n$, $\mathcal{K}_n$, respectively, $\widehat{\mathcal{K}}_n$ are bounded with operator norm

$$\| a_e \restriction_{\mathcal{H}_n} \| = \| a_e \restriction_{\mathcal{K}_n} \| = \Big\| a_e \restriction_{\widehat{\mathcal{K}}_n} \Big\| = \sqrt{2C(0)-C(e)-C(-e)}\sqrt{n}. \tag{3.67}$$

As the notation a_e^* and a_e suggests, these operators are adjoints of each other.

In order to express the infinitesimal generator in the Gaussian case, two more observations are needed. Both follow from standard facts in the context of Gaussian Hilbert spaces or Malliavin calculus. First, the operator of multiplication by $\omega(0)-\omega(e)$, acting on $\mathcal{H}$, is $a_e^*+a_e$. Hence, the multiplication operators M_e and N_e defined in (3.22) and (3.23), in the Gaussian case, are

$$N_e = a_e^* + a_e, \qquad M_e = s(a_e^*+a_e). \tag{3.68}$$

Second, from the formula of *directional derivative* in $\mathcal{H}$, it follows that

$$\mathcal{D} = \sum_{e\in\mathcal{U}} a_e. \tag{3.69}$$

The identity (3.69) is checked directly on Wick polynomials and extends by linearity.

Using these identities, after simple manipulations, we obtain

$$S_1 = \frac{1}{2}\sum_{e\in\mathcal{U}} \nabla_{-e} s(a_e^*+a_e)\nabla_e, \tag{3.70}$$

$$A = \sum_{e\in\mathcal{U}} \nabla_{-e} a_e - \sum_{e\in\mathcal{U}} a_e^* \nabla_{-e} =: A_- - A_+. \tag{3.71}$$

Note that

$$A_\pm : \mathcal{H}_n \to \mathcal{H}_{n\pm1}, \qquad S_1 : \mathcal{H}_n \to \oplus_{j=-q}^{q} \mathcal{H}_{n+2j} \tag{3.72}$$

where $2q$ is the degree of the even polynomial $s(u)$.

3.3 Diffusive bounds

We now prove Corollary 1 and part (1) of Theorem 3.1.

We write the displacement $X(t)$ in the standard martingale + compensator decomposition:

$$X(t) = N(t) + M(t) + \int_0^t \overline{\varphi}(\eta(s))\,\mathrm{d}s + \int_0^t \widetilde{\varphi}(\eta(s))\,\mathrm{d}s. \tag{3.73}$$

Here, $N(t)$ is the martingale part due to the jump rates γ and $M(t)$ is the martingale part due to the jump rates $w - \gamma$.

The compensators are

$$\overline{\varphi} : \Omega \to \mathbb{R}^d, \qquad \overline{\varphi}_l(\omega) = s(\omega(0) - \omega(e_l)) - s(\omega(0) - \omega(-e_l)), \tag{3.74}$$

$$\widetilde{\varphi} : \Omega \to \mathbb{R}^d, \qquad \widetilde{\varphi}_l(\omega) = r(\omega(0) - \omega(e_l)) - r(\omega(0) - \omega(-e_l)). \tag{3.75}$$

From this representation (and ergodicity of the process $\eta(t)$, cf. Proposition 2), Corollary 1 drops out for free. Indeed, $N(t)$ and $M(t)$ are martingales with stationary and ergodic increments, thus almost surely $N(t)/t \to 0$, $M(t)/t \to 0$, and due to Birkhoff's individual ergodic theorem the corresponding ergodic averages of the integrals on the right hand side of (3.73) also converge to 0, as $t \to \infty$ [52].

Note that, since $s(\cdot)$ is *even*, $\overline{\varphi}_l$, $l = 1, \ldots, d$ are actually *gradients*:

$$\overline{\varphi}_l = \nabla_l \psi_l \quad \text{where} \quad \psi_l : \Omega \to \mathbb{R}, \quad \psi_l(\omega) := s(\omega(0) - \omega(-e_l)). \tag{3.76}$$

The *diffusive lower bound* follows simply from *ellipticity* (3.1). The martingale $N(t)$ in the decomposition (3.73) is uncorrelated with the other terms. Hence, the lower bound in (3.16).

The main point is the *diffusive upper bound* which is more subtle. Since the martingale terms in (3.73) scale diffusively [52], we only need to prove diffusive upper bound for the compensators. From standard variational arguments, it follows (see e.g. [32], [47] and [53]) that

$$\overline{\lim_{t\to\infty}}\, t^{-1}\mathbf{E}\left((\int_0^t \varphi(\eta(s))\,\mathrm{d}s)^2\right) \le 2(\varphi, S^{-1}\varphi). \tag{3.77}$$

In our particular case, from (3.27), it follows that it is sufficient to prove upper bounds on $(\overline{\varphi}, (-\Delta)^{-1}\overline{\varphi})$ and $(\widetilde{\varphi}, (-\Delta)^{-1}\widetilde{\varphi})$. The first one drops out from (3.76):

$$(\overline{\varphi}_l, (-\Delta)^{-1}\overline{\varphi}_l) = (\nabla_l\psi_l, (-\Delta)^{-1}\nabla_l\psi_l) \le \|\,\psi_l\,\|^2 = \mathbf{E}\left(s(\omega(0) - \omega(e_l))^2\right). \tag{3.78}$$

We need

$$\mathbf{E}\left(s(\omega(0) - \omega(e_l))^2\right) < \infty. \tag{3.79}$$

Dropping the index l, denote

$$Z(\lambda) := \mathbf{E}\left(\exp\{\lambda(\omega(0)-\omega(e))^2\}\right) \in [1,\infty]. \tag{3.80}$$

In Lemma 3.2 below, we formulate a direct consequence of Brascamp-Lieb inequality which will be used for proving the diffusive bound on the second integral on the right-hand side of (3.73).

Lemma 3.2. *For any smooth cylinder function $F:\Omega\to\mathbb{R}$ and $0\le\lambda<c/2$:*

$$\begin{aligned}\frac{1}{Z(\lambda)}\mathbf{E}\left(F(\omega)^2\exp\{\lambda(\omega(0)-\omega(e))^2\}\right) \qquad (3.81)\\ \le \frac{1}{c}\frac{1}{Z(\lambda)}\mathbf{E}\left(\sum_{x,y\in\mathbb{Z}^d}\partial_x F(\omega)(-\Delta)^{-1}_{xy}\partial_y F(\omega)\exp\{\lambda(\omega(0)-\omega(e))^2\}\right)\\ +\frac{1}{Z(\lambda)^2}\mathbf{E}\left(F(\omega)\exp\{\lambda(\omega(0)-\omega(e))^2\}\right)^2.\end{aligned}$$

∂_x denotes $\frac{\partial}{\partial\omega(x)}$.

Proof. We apply Brascamp–Lieb inequality as stated in e.g. Proposition 2.1 in [4] with the measure $\mathbf{E}_\lambda(\cdot)=\frac{1}{Z(\lambda)}\mathbf{E}\left(\cdot\exp\{\lambda(\omega(0)-\omega(e))^2\}\right)$. The measure $\mathbf{E}_\lambda(\cdot)$ has density $\exp\{-V\}$ where $V(\omega)=-\lambda(\omega(0)+\omega(e))^2+\frac{1}{2}\sum_{|x-y|=1}R(\omega(x)-\omega(y))$ is convex (and thus Brascamp-Lieb applicable) for $\lambda<c/2$.

Then

$$\mathbf{Var}_\lambda(F)\le\mathbf{E}_\lambda\left(\sum_{x,y}\partial_x F(\omega)(V'')^{-1}_{xy}\partial_y F(\omega)\right), \tag{3.82}$$

or, equivalently,

$$\begin{aligned}\frac{1}{Z(\lambda)}\mathbf{E}\left(F(\omega)^2\exp\{\lambda(\omega(0)-\omega(e))^2\}\right)-\frac{1}{Z(\lambda)^2}\mathbf{E}\left(F(\omega)\exp\{\lambda(\omega(0)-\omega(e))^2\}\right)^2\le \qquad (3.83)\\ \frac{1}{Z(\lambda)}\mathbf{E}\left(\sum_{x,y}\partial_x F(\omega)(V'')^{-1}_{xy}\partial_y F(\omega)\exp\{\lambda(\omega(0)-\omega(e))^2\}\right).\end{aligned}$$

Since $V''\ge\inf R''(-\Delta)\ge c(-\Delta)$, rearranging gives (3.81). □

In order to prove (3.79), choose $F(\omega)=\omega(0)-\omega(e)$ in (3.81) and note that the second term on the right-hand side of the inequality vanishes. We get

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}Z(\lambda)\le\frac{\beta}{c}Z(\lambda) \tag{3.84}$$

with some explicit constant $\beta<\infty$. Hence, for $\lambda\in[0,c/2)$,

$$Z(\lambda)\le C\exp\left(\frac{\lambda\beta}{c}\right)<\infty. \tag{3.85}$$

Now, (3.79) follows from (3.4) and (3.85).

We also need to get

$$(\widetilde{\varphi}_l, (-\Delta)^{-1}\widetilde{\varphi}_l) < \infty. \tag{3.86}$$

We fix $l = 1$ and drop the subscript. Denote

$$C(x) := \mathbf{E}\left(\widetilde{\varphi}(\omega)\widetilde{\varphi}(\tau_x\omega)\right), \qquad \widehat{C}(p) := \sum_{x\in\mathbb{Z}^d} e^{ip\cdot x}C(x), \;\; p \in [-\pi,\pi]^d. \tag{3.87}$$

The bound (3.86) is equivalent to the infrared bound

$$\int_{[-\pi,\pi]^d} \frac{\widehat{C}(p)}{\widehat{D}(p)}\,\mathrm{d}p < \infty \tag{3.88}$$

where $\widehat{D}$ is the Fourier transform of the lattice Laplacian, defined in (3.40). Since $d \geq 3$, $\int_{[-\pi,\pi]^d} \frac{1}{\widehat{D}(p)}\,\mathrm{d}p < \infty$ and it is sufficient to prove

$$\sup_{p\in[-\pi,\pi]^d} \Big|\, \widehat{C}(p) \,\Big| < \infty. \tag{3.89}$$

Lemma 3.3. *(a) Let $f : \mathbf{R} \to \mathbf{R}$ be smooth and denote*

$$C(x) := \mathbf{Cov}\big(f(\omega(0)-\omega(e)), f(\omega(x)-\omega(x+e))\big), \tag{3.90}$$

$$C'(x) := \mathbf{Cov}\big(f'(\omega(0)-\omega(e)), f'(\omega(x)-\omega(x+e))\big), \tag{3.91}$$

$$m' := \mathbf{E}\left(f'(\omega(0)-\omega(e))\right). \tag{3.92}$$

Then

$$\sup_{p\in[-\pi,\pi]^d} \Big|\, \widehat{C}(p) \,\Big| \leq (cd)^{-1} \sup_{p\in[-\pi,\pi]^d} \Big|\, \widehat{C}'(p) \,\Big| + c^{-1}(m')^2. \tag{3.93}$$

(b) Let

$$C_{nm}(x) := \mathbf{Cov}\big((\omega(0)-\omega(e))^n, (\omega(x)-\omega(x+e))^m\big). \tag{3.94}$$

Then

$$\sup_{p\in[-\pi,\pi]^d} \Big|\, \widehat{C}_{nm}(p) \,\Big| \leq (Z((c-\varepsilon)/2))^2 n!m! \left(\frac{2}{c-\varepsilon}\right)^{(n+m)/2} \tag{3.95}$$

where ε

(c) If r is an analytic function and it satisfies condition (3.5)*, then*

$$\sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}(p) \Big| < \infty. \tag{3.96}$$

Proof. (a) We apply (3.81) with $\lambda = 0$ and

$$F(\omega) := \sum_{x\in\mathbb{Z}^d} \alpha(x) f(\omega(x) - \omega(x+e)) \tag{3.97}$$

where $\alpha : \mathbb{Z}^d \to \mathbb{R}$ is finitely supported and $\sum_{z\in\mathbb{Z}^d} \alpha(z) = 0$. Straightforward computations yield

$$\mathbf{E}\left(F(\omega)^2\right) = \sum_{x,y\in\mathbb{Z}^d} \alpha(x) C(x-y) \alpha(y) \le c^{-1} \sum_{x,y\in\mathbb{Z}^d} \alpha(x) \Gamma(x-y) \big(C'(x-y) + (m')^2\big) \alpha(y) \tag{3.98}$$

where Γ is the matrix

$$\Gamma := \nabla_1 (-\Delta^{-1}) \nabla_1 \tag{3.99}$$

well-defined in any dimension. Its Fourier transform is

$$\widehat{\Gamma}(p) = \frac{1-\cos p_1}{\widehat{D}(p)}. \tag{3.100}$$

The bound (3.98) is equivalent to

$$\widehat{C}(p) \le c^{-1} \left(\widehat{\Gamma} * \widehat{C}'(p) + (m')^2 \widehat{\Gamma}(p) \right). \tag{3.101}$$

Convolution is meant periodically in $[-\pi,\pi]^d$. Hence,

$$\begin{aligned} \sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}(p) \Big| &\le c^{-1} \sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}'(p) \Big| \int_{[-\pi,\pi]^d} \widehat{\Gamma}(p)\, \mathrm{d}p + c^{-1}(m')^2 \sup_{p\in[-\pi,\pi]^d} \widehat{\Gamma}(p) \\ &\le (cd)^{-1} \sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}'(p) \Big| + c^{-1}(m')^2. \end{aligned} \tag{3.102}$$

(b) We apply (3.93) to the function $f(u) = u^n$ and use the notation

$$m_n := \mathbf{E}\left((\omega(0) - \omega(e))^n\right) \tag{3.103}$$

to get

$$\sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}_{nn}(p) \Big| \le (cd)^{-1} n^2 \sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}_{n-1,n-1}(p) \Big| + c^{-1} n^2 m_{n-1}^2. \tag{3.104}$$

Induction on n yields

$$\sup_{p\in[-\pi,\pi]^d} \Big| \widehat{C}_{nn}(p) \Big| \le \sum_{k=1}^{n} \frac{(n!)^2}{((n-k)!)^2} \frac{m_{n-k}^2}{c^k d^{k-1}}. \tag{3.105}$$

By (3.85), we have the finiteness of

$$\mathbf{E}\left(\exp\left((c-\varepsilon)(\omega(0)-\omega(e))^2\right)/2\right) = Z((c-\varepsilon)/2) < \infty. \tag{3.106}$$

Hence, by expanding the exponential,

$$m_n = \mathbf{E}\left((\omega(0)-\omega(e))^n\right) \leq Z((c-\varepsilon)/2)\frac{2^{n/2}\lfloor n/2\rfloor!}{(c-\varepsilon)^{n/2}} \tag{3.107}$$

follows. We neglect the fact that the odd moments are 0 by symmetry. Combining the last inequality with (3.105), we obtain

$$\sup_{p\in[-\pi,\pi]^d}\left|\,\widehat{C}_{nn}(p)\,\right| \leq (Z((c-\varepsilon)/2))^2(n!)^2(c-\varepsilon)^{-n}\sum_{k=1}^{n}\frac{(\lfloor(n-k)/2\rfloor!)^2}{((n-k)!)^2}\frac{2^{n-k}}{d^{k-1}} \tag{3.108}$$

$$\leq (Z((c-\varepsilon)/2))^2(n!)^2\left(\frac{2}{c-\varepsilon}\right)^n, \tag{3.109}$$

which proves (3.95) for $n = m$. The constant $2/(c-\varepsilon)$ is far from optimal here, but the order $(n!)^2$ is the best one can get with this argument.

The general case $n \neq m$ follows by Schwarz's inequality.

(c) By power series expansion of the analytic function r, and using (3.75), (3.95) and (3.5), one gets

$$\left|\,\widehat{C}(p)\,\right| \leq 4\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\left|\,r^{(n)}(0)\,\right|}{n!}\frac{\left|\,r^{(m)}(0)\,\right|}{m!}\left|\,\widehat{C}_{nm}(p)\,\right| \tag{3.110}$$

$$\leq 4(Z(c-\varepsilon/2))^2\left(\sum_{n=0}^{\infty}\left|\,r^{(n)}(0)\,\right|\left(\frac{2}{c-\varepsilon}\right)^{n/2}\right)^2 < \infty.$$

□

Finally, we apply Lemma 3.3 to prove (3.89) and the diffusive bound.

3.4 Checking the graded sector condition

Next we prove the second part of Theorem 3.1. We have to show that the martingale approximation of Theorem KV is valid for the integrals in on the right-hand side of (3.73). We apply the *graded sector condition* formulated in Theorem 2.1 with $D = \gamma\,|\,\Delta\,|$ and the operators S and A given in graded form in (3.70) and (3.71). (2.18) clearly holds and (2.22) was already proved in Section 3.3. We still need to verify conditions (2.19), (2.20) and (2.21).

Checking (2.21) is straightforward: If $s(u)$ is an even polynomial of degree $2q$, then using in turn

(3.59), (3.63) and (3.67), we obtain

$$\left\| \, |\Delta|^{-1/2} \nabla_{-e} s(a_e + a_e^*) \nabla_e |\Delta|^{-1/2} \restriction_{\mathcal{H}_n} \right\| \leq \| \, s(a_e + a_e^*) \restriction_{\mathcal{H}_n} \| \leq cn^q + C \tag{3.111}$$

with the constant c proportional to the leading coefficient in the polynomial $s(u)$ and $C < \infty$. Hence, if $q = 2$ (that is: $s(u)$ quartic polynomial) and the leading coefficient is sufficiently small, then (2.21) follows. The bound (2.19) with $\kappa = 2$ also drops out from (3.111).

Finally, we check (2.20). By (3.59),

$$\left\| \, |\Delta|^{-1/2} a_{-e}^* \nabla_e |\Delta|^{-1/2} \restriction_{\mathcal{H}_n} \right\| \leq \left\| \, |\Delta|^{-1/2} a_e^* \restriction_{\mathcal{H}_n} \right\|. \tag{3.112}$$

We prove

$$\left\| \, |\Delta|^{-1/2} a_e^* \restriction_{\mathcal{H}_n} \right\| \leq C n^{1/2} \tag{3.113}$$

with some finite constant C.

For $\widehat{u} \in \mathcal{H}_n$,

$$\begin{aligned} &| \, \Delta \, |^{-1/2} \, a_e^* \widehat{u}(p_1, \ldots, p_{n+1}) \\ &\qquad = \frac{1}{\sqrt{n+1}} \frac{1}{\sqrt{\widehat{D}(\sum_{m=1}^{n+1} p_m)}} \sum_{m=1}^{n+1} \left(e^{ip_m \cdot e} - 1 \right) \widehat{u}(p_1, \ldots, \cancel{p_m}, \ldots, p_{n+1}). \end{aligned} \tag{3.114}$$

Hence,

$$\begin{aligned} &\left\| \, | \, \Delta \, |^{-1/2} \, a_e^* \widehat{u} \, \right\|^2 \\ &= \frac{1}{n+1} \int_{(-\pi,\pi]^{d(n+1)}} \frac{1}{\widehat{D}(\sum_{m=1}^{n+1} p_m)} \\ &\qquad \times \left| \sum_{m=1}^{n+1} \left(e^{ip_m \cdot e} - 1 \right) \widehat{u}(p_1, \ldots, \cancel{p_m}, \ldots, p_{n+1}) \right|^2 \prod_{m=1}^{n+1} \frac{1}{\widehat{D}(p_m)} \, \mathrm{d}p_1 \ldots \mathrm{d}p_{n+1} \\ &\leq (n+1) \int_{(-\pi,\pi]^{d(n+1)}} \frac{1}{\widehat{D}(\sum_{m=1}^{n+1} p_m)} \\ &\qquad \times \left| \, e^{ip_{n+1} \cdot e} - 1 \, \right|^2 | \, \widehat{u}(p_1, \ldots, p_n) \, |^2 \prod_{m=1}^{n+1} \frac{1}{\widehat{D}(p_m)} \, \mathrm{d}p_1 \ldots \mathrm{d}p_{n+1} \\ &= (n+1) \int_{(-\pi,\pi]^{dn}} | \, \widehat{u}(p_1, \ldots, p_n) \, |^2 \prod_{m=1}^{n} \frac{1}{\widehat{D}(p_m)} \\ &\qquad \times \left(\int_{(-\pi,\pi]^d} \frac{\left| \, e^{ip_{n+1} \cdot e} - 1 \, \right|^2}{\widehat{D}(p_{n+1})} \frac{1}{\widehat{D}(\sum_{m=1}^{n+1} p_m)} \mathrm{d}p_{n+1} \right) \mathrm{d}p_1 \ldots \mathrm{d}p_n. \end{aligned} \tag{3.115}$$

Schwarz's inequality and symmetry was used. Note that on the right-hand side, for the innermost

term, since $d \geq 3$, we have

$$\int_{(-\pi,\pi]^d} \frac{\left| e^{ip_{n+1}\cdot e} - 1 \right|^2}{\widehat{D}(p_{n+1})} \frac{1}{\widehat{D}(\sum_{m=1}^{n+1} p_m)} \mathrm{d}p_{n+1} \leq C^2. \tag{3.116}$$

Hence,

$$\left\| \left| \Delta \right|^{-1/2} a_e^* \widehat{u} \right\|^2 \leq C^2 (n+1) \left\| \widehat{u} \right\|^2, \tag{3.117}$$

and (3.113) follows. This proves (2.20). □

3.5 Outlook

This is the first rigorous result for central limit theorem for the TSAW. That said, the class of rate functions for which part (2) of Theorem 3.1 applies is somewhat restricted. It seems reasonable to conjecture that martingale approximation and the central limit theorem are valid for a larger class of rate functions.

Proving central limit theorem for finite-dimensional distributions of the rescaled process is not quite invariance principle (convergence to Brownian motion). For the full invariance principle, *tightness* is also necessary, which was not addressed here. Proving tightness is generally a technical affair, that has been done for a number of models; for more on this topic, we refer to [32].

Chapter 4

A constructive proof of the phase-type characterization theorem

4.1 Preliminaries

We continue from 1.3. As mentioned, an ME representation is not unique; we show how different ME representations can be transformed into each other.

Theorem 4.1. *[7] Let ME($\alpha, \mathbf{A}$) of order n and ME($\gamma, \mathbf{G}$) of order m be two ME representations that correspond to pdf's $f_X(t)$ and $f_Y(t)$, respectively.*

If there exists a matrix $\mathbf{W}$ of cardinality $n \times m$ such that

$$\gamma = \alpha\mathbf{W}, \quad \mathbf{AW} = \mathbf{WG}, \quad \mathbf{1}_n = \mathbf{W}\mathbf{1}_m,$$

then ME($\alpha, \mathbf{A}$) $\equiv$ ME($\gamma, \mathbf{G}$) (that is, $f_X(t) = f_Y(t)$).

Proof.

$$f_Y(t) = -\gamma\mathbf{G}e^{t\mathbf{G}}\mathbf{1}_m = -\alpha\mathbf{WG}e^{t\mathbf{G}}\mathbf{1}_m = -\alpha\mathbf{A}e^{t\mathbf{A}}\mathbf{W}\mathbf{1}_m = -\alpha\mathbf{A}e^{t\mathbf{A}}\mathbf{1}_n = f_X(t).$$

□

Theorem 4.1 will be used as a representation transformation tool. The size of column vector $\mathbf{1}$ is explicitly indicated in the theorem as a subscript.

Next, we turn to additional properties of minimal ME representations.

In a minimal ME representation, there are no "extra" or "redundant" eigenvalues in matrix $\mathbf{A}$. More precisely a minimal ME representation has the following properties [57]:

P1) All Jordan blocks of $\mathbf{A}$ have different eigenvalues.

P2) All eigenvalues contribute to the distribution with maximal multiplicity. For example, a Jordan block of size n_i corresponding to eigenvalue $-\lambda_i$ results in the terms $\sum_{j=1}^{n_i} c_{\lambda_i,j} x^{j-1} e^{-\lambda_i x}$ in $f_X(x)$, where $c_{\lambda_i,n_i} \neq 0$.

P3) α is not orthogonal to any of the right-eigenvectors of $\mathbf{A}$.

P4) $\mathbf{1}$ is not orthogonal to any of the left-eigenvectors of $\mathbf{A}$.

P5) The Jordan block structures of all minimal ME representations of an ME distribution are identical.

These properties are explained further in Section 4.4. Based on these properties, a minimal ME representation can be obtained directly from f_X. If f takes the form

$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} c_{\lambda_i,j} x^{j-1} e^{-\lambda_i x}$$

where λ_i are different and $c_{\lambda_i,n_i} \neq 0$, then we will consider the following representation $(\alpha, \mathbf{A})$:

$$\mathbf{A} = \begin{pmatrix} \mathbf{J}_1 & 0 & \dots & & 0 \\ 0 & \mathbf{J}_2 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & & 0 & \mathbf{J}_m \end{pmatrix},$$

where

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & \dots & & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & & 0 & \lambda_i \end{pmatrix}.$$

and $\mathbf{J}_i$ is of size n_i. α can be calculated by solving

$$-\alpha e^{\mathbf{A}x} \mathbf{A} \mathbf{1} = f_X(x);$$

this equation can be solved because the left-hand side contains all the terms $x^{j-1} e^{-\lambda_i x}$ up to $j \leq n_i$ for $i = 1, \dots, m$.

Lemma 4.2. *The representation $(\alpha, \mathbf{A})$ is minimal for f_X.*

The proof is essentially due to properties P1-P5 and the fact that no Jordan block of size smaller than n_i can represent the term $x^{n_i-1} e^{-\lambda_i x}$. Section 4.4 elaborates more on this topic.

If ME representation $(\alpha, \mathbf{A})$ is minimal then there are some straightforward necessary conditions for vector α and matrix $\mathbf{A}$ to define a valid distribution according to Definition 2:

C1) The eigenvalues of $\mathbf{A}$ have negative real part (to avoid divergence of $f_X(x)$ as $x \to \infty$).

C2) There is a real eigenvalue of $\mathbf{A}$ with maximal real part (to avoid oscillations to negative values as $x \to \infty$).

C3) $\alpha\mathbf{1} = 1$ (normalizing condition which ensures $\int_0^\infty f_X(x)\mathrm{d}x = 1$).

C4) If for all $i \in \{0, 1, \ldots, j-1\}$, $-\alpha\mathbf{A}^i\mathbf{1} = 0$, then $-\alpha\mathbf{A}^j\mathbf{1} \geq 0$ (to avoid decreasing behavior around $f(x) = 0$; note that $-\alpha\mathbf{A}^i\mathbf{1}$ is the i-th derivative of f at $x = 0$).

If any of these necessary conditions are violated then the tuple consisting of the vector α and matrix $\mathbf{A}$ does not define a valid ME distribution. Note that non-minimal ME representations might contain any additional eigenvalues, including, for example, positive ones.

Properties P1-P5 ensure that if C1-C4 hold for one minimal representation ME(α, $\mathbf{A}$), they hold for all equivalent minimal ME representations. Additionally, if the dominant eigenvalue has multiplicity higher than 1, then it belongs to a Jordan-block whose size is equal to the multiplicity of the dominant eigenvalue.

The rest of this Chapter is structured as follows. Section 4.2 contains the proof for the sufficient direction of Theorem 1.1 along with the constructive algorithm. Section 4.3 contains a worked example that demonstrates the algorithm step-by-step. Section 4.4 contains the proofs for the necessary direction. Section 4.5 gives an outlook.

4.2 Procedure and proof

Our main goal is an algorithm that provides a constructive proof for the sufficient direction of Theorem 1.1, that is, given that the dominant eigenvalue condition and the positive density condition hold for ME(α, $\mathbf{A}$), find a PH-representation equivalent to ME(α, $\mathbf{A}$); in other words, find a vector-matrix pair (β, $\mathbf{B}$) where β and $\mathbf{B}$ are Markovian and define the same distribution as ME(α, $\mathbf{A}$).

This section is devoted to the algorithmic construction, also stating and proving the theorems used along the way.

4.2.1 Sketch of the algorithm

The algorithm consists of five main steps. Steps 1 and 2 are preparatory, and Step 5 is just correction related to Step 2.

- Step 1. We find an equivalent minimal ME representation $(\alpha_1, \mathbf{A}_1)$ for $(\alpha, \mathbf{A})$ if it is not minimal by eliminating any "extra" eigenvalues of $\mathbf{A}$, which does not contribute to the pdf (the pdf will be denoted by f_X). We refer to Lemma 4.2 and [7] for a different, computationally stable method of finding a minimal ME representation.

- Step 2. This step applies only if density is zero at 0, that is, $f_X(0) = 0$. This step is essentially what may be called "deconvolution": we represent f_X as the convolution of some f_Y matrix

exponential density function with $f_Y(0) > 0$ and an appropriate Gamma (Erlang) distribution Erlang(k, μ) (see Lemma 4.3); if f_Y has a Markovian representation, then it gives a straightforward Markovian representation for f_X as well (see Lemma 4.4). Thus we only need to find a Markovian representation for f_Y (and the corresponding representation, which is obtained from Lemma 4.3), where $f_Y(0) > 0$. If this step is applied, Steps 3 and 4 are applied for f_Y instead of f_X, and we switch back to f_X in Step 5.

- Step 3. An equivalent ME representation (γ, $\mathbf{G}$) is given with Markovian matrix $\mathbf{G}$, while γ may still have negative elements. The main tool of this step is the so-called monocyclic structure (with Feedback-Erlang blocks). Typically, the size of $\mathbf{G}$ is larger than that of $\mathbf{A}_1$ (because each pair of complex conjugate eigenvalues is represented with at least 3 phases); that said, $\mathbf{G}$ is a sparse matrix with a simple block bi-diagonal structure. For this step only the dominant eigenvalue condition is necessary.

- Step 4. γ and $\mathbf{G}$ are transformed further into β and $\mathbf{B}$ where β is Markovian (and the Markovity of $\mathbf{B}$ is also preserved) essentially by adding an "Erlang-tail" (a number of sequentially connected exponential phases with identical rates) of proper order and rate to the monocyclic structure described by the Markovian matrix $\mathbf{G}$. The main mathematical tool of this step is the approximation of elementary functions. The skeleton of this step is composed of the following elements:

  - Find τ such that $\gamma e^{\tau \mathbf{G}} > 0$ (element-wise). Such τ always exists if the dominant eigenvalue and the positive density conditions hold and $\mathbf{G}$ has Feedback-Erlang structure. We remark that if $\mathbf{G}$ does not have Feedback-Erlang structure, such a τ may not exist, even if $\mathbf{G}$ is Markovian. This is further explained after Lemma 4.8.
  - Find λ' such that
  $$\gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{\tau\lambda} > 0 \quad \forall \lambda \geq \lambda'$$
  which is always possible since $\left\| \gamma(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\tau\lambda} - \gamma e^{\tau \mathbf{G}} \right\| \to 0$ as $\lambda \to \infty$.
  - Let $\epsilon = \inf_{t \in (0,\tau)} f_X(t)$. $\epsilon > 0$ because of the positive density condition and the result of Step 2. Find λ'' such that
  $$\left| -\gamma e^{\mathbf{G}\tau} \mathbf{G}\mathbf{1} + \gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{\tau\lambda} \mathbf{G}\mathbf{1} \right| < \epsilon \quad \forall \lambda \geq \lambda''.$$
  This ensures that $-\gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^k \mathbf{G}\mathbf{1} > 0$ for $k = 1, \ldots, n$ where $n = \tau\lambda''$. This is always possible when $\epsilon > 0$.
  - Extend the $(\gamma, \mathbf{G})$ representation with an Erlang tail of rate $\lambda = \max(\lambda', \lambda'')$ and order $n = \lceil \lambda\tau \rceil$.

- Step 5. If Step 2 was applied, at this point we have a Markovian representation for f_Y. To switch back to f_X, we use Lemma 4.4. If Step 2 was not applied, Step 5 does not apply either.

4.2.2 Step 1: Minimal ME representation

Starting from ME representation $(\alpha, \mathbf{A})$, we can obtain a minimal ME representation $(\alpha_1, \mathbf{A}_1)$ with the application of a representation minimization method. A minimal ME representation can be obtained through several approaches. One possibility is directly from the pdf $f(x) = -\alpha \mathbf{A} e^{\mathbf{A}x} \mathbf{1}$ as in Lemma 4.2. Another, computationally stable order reduction method is the Staircase method from [7], which uses singular value decomposition. In any case, the minimal ME representation $(\alpha_1, \mathbf{A}_1)$ enjoys properties P1-P5.

There are two important properties that can be determined from a minimal ME representation (or the density function directly). These are the value and the multiplicity of the dominant eigenvalue and the validity of the dominant eigenvalue condition. We denote the dominant eigenvalue (which is real and negative) by $-\lambda_1$ and its multiplicity by n_1. Indeed, λ_1 and n_1 determine the asymptotic rate of decay of the pdf: it decays like $c_{\lambda_1,n_1} t^{n_1-1} e^{-\lambda_1 t}$, where c_{λ_1,n_1} is a positive constant, more precisely

$$\lim_{t\to\infty} \frac{f(t)}{t^{n_1-1}e^{-\lambda_1 t}} = c_{\lambda_1,n_1}.$$

4.2.3 Step 2: Positive density at zero

In the case when $\mathrm{ME}(\alpha_1, \mathbf{A}_1)$ is such that $f_X(t) > 0$ for positive values of t, but $f_X(0) = 0$, then based on the following lemma, we represent $\mathrm{ME}(\alpha_1, \mathbf{A}_1)$ as the convolution of an Erlang distribution and a matrix exponential distribution $\mathrm{ME}(\alpha_2, \mathbf{A}_1)$ whose density is positive at 0. Note that the generator remains the same (thus the dominant eigenvalue condition is automatically preserved for f_Y).

Lemma 4.3. *If* $f_X(t) = -\alpha_1 e^{t\mathbf{A}_1} \mathbf{A}_1 \mathbf{1}$ *is a matrix exponential pdf with*

$$f_X(t) > 0 \quad \forall t > 0, \qquad f_X^{(i)}(t)\Big|_{t=0} = 0 \quad i = 0, \ldots, l-1, \qquad f_X^{(l)}(t)\Big|_{t=0} > 0, \tag{4.1}$$

then f_X *can be written in the form*

$$f_X = f_Y * g(l, \mu, \cdot),$$

for some large enough μ*, where* $g(l, \mu, t) = \frac{\mu^l t^{l-1} e^{-\mu t}}{(l-1)!}$ *is the Erlang*(l, μ) *pdf,* $*$ *denotes convolution and* $f_Y(t)$ *is a matrix exponential function with*

$$f_Y(t) > 0 \quad \forall t \geq 0$$

and ME representation (α_2, A_1)*.*

Proof. The intuitive behavior of the convolution of the pdf of a non-negative r.v. (Y) and the Erlang(l, μ) pdf is the following: assume $f_Y(0) > 0$; for large values of μ, the Erlang pdf decays rapidly, so the function f_Y is very close to f_X, except around 0, since convolution of a pdf f_Y with an Erlang(l, μ) pdf increases the multiplicity of 0 by l. Lemma 4.3 utilizes this relation in the opposite direction.

f_Y can be calculated in the Laplace-transform domain as follows. The Laplace-transform of the Erlang(l, μ) pdf is

$$f^*_{l,\mu}(s) = \left(\frac{\mu}{s+\mu}\right)^l.$$

Denote by $f^*_X(s)$ and $f^*_Y(s)$ the Laplace-transform of f_X and f_Y, respectively. Then from $f_X(t) = f_Y(t) * f_{l,\mu}(t)$ we have $f^*_X(s) = f^*_Y(s) \cdot \left(\frac{\mu}{s+\mu}\right)^l$, and so

$$f^*_Y(s) = f^*_X(s)\left(\frac{s+\mu}{\mu}\right)^l = f^*_X(s)\left(1+\frac{s}{\mu}\right)^l.$$

For $l = 1$, the inverse transform of $f^*_X(s)\left(1+\frac{s}{\mu}\right)$ gives

$$f_Y(x) = f_X(t) + \frac{1}{\mu}\left(f'_X(t) + f_X(0)\right) = f_X(t) + \frac{1}{\mu}f'_X(t).$$

For $l > 1$, induction (or the binomial theorem) gives

$$f_Y(t) = \sum_{i=0}^{l}\binom{l}{i}\frac{1}{\mu^i}f^{(i)}_X(t) = -\alpha_1\sum_{i=0}^{l}\binom{l}{i}\left(\frac{\mathbf{A}_1}{\mu}\right)^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1}.$$

The fact that $f_Y(t)$ is a matrix-exponential function is clear from the above formula. Also, it has a representation of the form ME$(\alpha_2, \mathbf{A}_1)$ where

$$\alpha_2 = -\alpha_1\sum_{i=0}^{l}\binom{l}{i}\left(\frac{\mathbf{A}_1}{\mu}\right)^i.$$

That said, we still have to check the positivity of f_Y. For technical reasons, this is carried out in 3 parts: first around 0, then for the tail, then for the main body of the function.

We fix a value $\delta > 0$ (independent from μ) such that

$$f^{(l)}_X(t) > 0, \quad t \in (0, \delta].$$

This is possible since $f^{(l)}_X(0) > 0$ and $f^{(l)}_X$ is continuous. This in turn implies by integration that

$$f^{(i)}_X(t) \geq 0, \quad t \in (0, \delta].$$

for every $i = l, l-1, \ldots, 1, 0$ and thus

$$f_Y(t) = \sum_{i=0}^{l}\binom{l}{i}\frac{1}{\mu^i}f^{(i)}_X(t) > 0 \quad t \in (0, \delta).$$

This holds for any value of μ.

We examine the tail of f_Y next. Recall that as $t \to \infty$, $f_X(t)$ decays as $c_{\lambda_1,n_1}t^{n_1-1}e^{-\lambda_1 t}$ where

$c_{\lambda_1,n_1} > 0$.

$$f_Y(t) - f_X(t) = \sum_{i=0}^{l} \binom{l}{i} \frac{1}{\mu^i} f_X^{(i)}(t) - f_X(t) = -\alpha_2 \sum_{i=1}^{l} \binom{l}{i} \left(\frac{\mathbf{A}_1}{\mu}\right)^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1}. \tag{4.2}$$

Since $e^{t\mathbf{A}_1}$ decays with rate $t^{n_1-1}e^{-\lambda_1 t}$,

$$-\alpha_1 \binom{l}{i} {\mathbf{A}_1}^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1} \sim c_i t^{n_1-1}e^{-\lambda_1 t}$$

for each $i = 1, \ldots, l$ for some constants c_i.

Select K_1 such that

$$\left| \frac{-\alpha_1 \binom{l}{i} {\mathbf{A}_1}^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1}}{t^{n_1-1}e^{-\lambda_1 t}} \right| \leq 2|c_i| \quad \forall t > K_1$$

for $i = 1, \ldots, k$, Then

$$|f_Y(t) - f_X(t)| \leq \sum_{i=1}^{l} \frac{2|c_i|}{\mu^i} t^{n_1-1}e^{-\lambda_1 t}.$$

Note that K_1 is also independent from μ.

The constant $\sum_{i=1}^{l} \frac{2|c_i|}{\mu^i}$ is decreasing in μ and goes to 0. Select μ_0 such that

$$\sum_{i=1}^{l} \frac{2|c_i|}{\mu^i} \leq \frac{1}{2} c_{\lambda_1,n_1} \quad \forall \mu > \mu_0.$$

Select K_2 such that

$$f_X(t) \geq \frac{1}{2} c_{\lambda_1,n_1} t^{n_1-1}e^{-\lambda_1 t} \quad \forall t > K_2.$$

Set $K = \max(K_1, K_2)$. At this point, δ and K are fixed (independently of μ), and for any $\mu > \mu_0$ it holds that

$$f_Y(t) \geq f_X(t) - |f_Y(t) - f_X(t)| \geq \frac{1}{2} c_{\lambda_1,n_1} x^{n_1-1}e^{-\lambda_1 t} - \frac{1}{2} c_{\lambda_1,n_1} t^{n_1-1}e^{-\lambda_1 t} = 0 \quad \forall t > K.$$

We now have positivity of f_Y at $[0, \delta]$ and $[K, \infty]$. For $[\delta, K]$, we use the formula (4.2) again, and note that

$$\sup_{t\in[\delta,K]} \left| \alpha_1 \sum_{i=1}^{l} \binom{l}{i} \left(\frac{\mathbf{A}_1}{\mu}\right)^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1} \right| \leq \sum_{i=1}^{l} \left(\frac{1}{\mu}\right)^i \sup_{t\in[\delta,K]} \left| \alpha_1 \binom{l}{i} \mathbf{A}_1^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1} \right|,$$

where $\sup_{t\in[\delta,K]} \left|\alpha_1 \binom{l}{i} \mathbf{A}_1^i \mathbf{A}_1 e^{t\mathbf{A}_1}\mathbf{1}\right|$ is finite for each $i = 1, \ldots, l$, while $\frac{1}{\mu^i} \to 0$, so there exists a μ_1 such that for any $\mu > \mu_1$,

$$|f_Y(t) - f_X(t)| \leq \inf_{t\in[\delta,K]} f_X(t),$$

which is positive due to the positive density condition (specifically that f_X is strictly positive on a bounded closed interval not containing 0).

Selecting any $\mu > \max(\mu_1, \mu_2)$ finishes the lemma.

□

The representation $(\alpha_2, \mathbf{A}_1)$ can be constructed either from f_Y via Lemma 4.2 or by calculating α_2 from the appropriate linear equations.

Lemma 4.3 and the following composition ensures that $f_X(t)$ and $f_Y(t)$ have a Markovian representation and satisfy the dominant eigenvalue condition at the same time if $\mu > \lambda_1$.

Lemma 4.4. *If $f_Y(t)$ is ME distributed with representation $(\alpha_2, \mathbf{A}_1)$ of order m and $\mu > \lambda_1$ then*

$$f_X(t) = f_Y(t) * g(l, \mu, t),$$

is ME distributed with initial vector $\beta = \{1, 0, 0, \ldots, 0\}$ and generator matrix

$$\mathbf{B} = \begin{pmatrix} -\mu & \mu & & \\ & \ddots & \ddots & \\ & & -\mu & \mu\alpha_2 \\ & & & \mathbf{A}_1 \end{pmatrix},$$

where the first l blocks of the matrix are of size one and the last block is of size m. Additionally, if $(\alpha_2, \mathbf{A}_1)$ is a Markovian representation then $(\beta, \mathbf{B})$ is Markovian as well.

Proof. In the case $(\alpha_2, \mathbf{A}_1)$ is a PH representation, the following probabilistic argument works: based on the structure of $\mathbf{B}$, the time to leave the first l phases is Erlang(l, μ) distributed and the time spent in the set of phases from $l+1$ to m is ME$(\alpha_2, \mathbf{A}_1)$ distributed. Technically, this argument does not apply if either α_2 or $\mathbf{A}_1$ are not Markovian, but the equality

$$f_X = f_Y * g(l, \mu, \cdot)$$

nevertheless holds, since it is independent of the actual representation of f_Y and f_X.

Also, this lemma will be applied only in Step 5, where a PH representation of f_Y is already available. □

Based on Lemma 4.3 and 4.4 it remains to prove that the matrix exponential density function $f(x)$ with $f(0) > 0$ satisfying the dominant eigenvalue and the positive density conditions has a Markovian representation.

4.2.4 Step 3: Markovian generator

The aim of this subsection is to transform the ME representation $(\alpha_1, \mathbf{A}_1)$ to a representation $(\gamma, \mathbf{G})$ where $\mathbf{G}$ is Markovian. If Step 2 was applied, the transformation is carried out for $(\alpha_2, \mathbf{A}_1)$ instead of

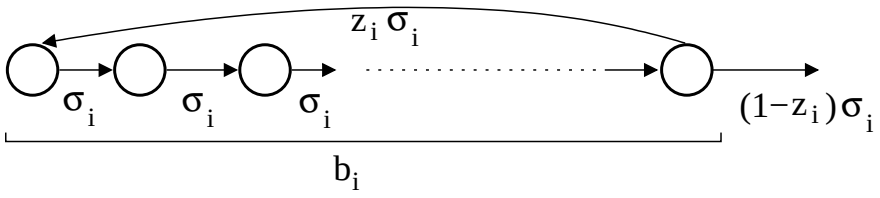

Figure 4.1: FE-diagonal block.

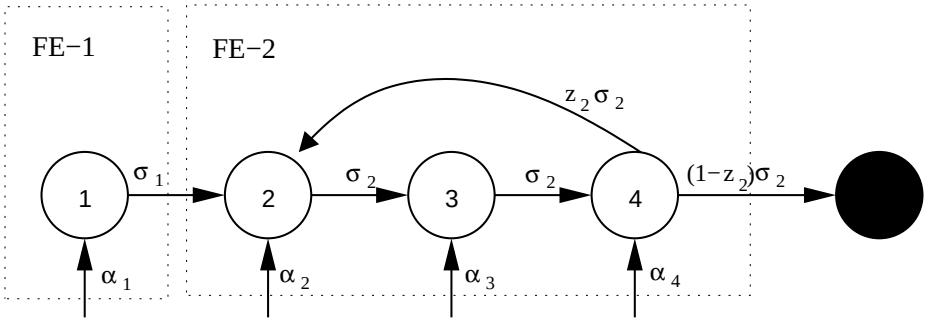

Figure 4.2: FE-diagonal representation of a generator with a real eigenvalue (σ_1) and a pair of complex ones.

$(\alpha_1, \mathbf{A}_1)$; nevertheless, we will stick to the use of the notation $(\alpha_1, \mathbf{A}_1)$ throughout the subsection. For matrix $\mathbf{G}$, we apply the matrix structure proposed in [43]. It is a block bi-diagonal matrix structure, where each block represents a real eigenvalue or a pair of complex conjugate eigenvalues of $\mathbf{A}_1$. The blocks associated with real eigenvalue $-\lambda_i$ ($-\lambda_i < 0$) are of size one, the diagonal element is $-\lambda_i$ and the first sub-diagonal element is λ_i. The blocks associated with complex eigenvalues are composed by Feedback-Erlang (FE) blocks.

Definition 5. *[43] A* Feedback-Erlang (FE) block *with parameters (b, σ, z) is a chain of b states with transition rate σ and one transition from the bth state to the first state, with rate $z\sigma$ (c.f. Figure 4.1). The probability $z \in [0, 1)$ is called the feedback probability.*

A FE block (b, σ, z) with length $b = 1$ and $z = 0$ corresponds to a real eigenvalue $-\sigma$ and is referred to as *degenerate* FE blocks. Matrix $\mathbf{G}$ contains as many FE blocks (degenerate or non-degenerate) associated with a real eigenvalue or a pair of complex conjugate eigenvalues as the multiplicity of the eigenvalue. A non-degenerate FE block where b is odd has a real eigenvalue and $(b-1)/2$ complex conjugate eigenvalue pairs. A non-degenerate FE block where b is even has 2 real eigenvalues and $(b-2)/2$ complex conjugate eigenvalue pairs. In both cases the eigenvalues are equidistantly located on a circle in the complex plane around $-\sigma$. The dominant eigenvalue of the FE block (the one with the largest real part) with parameters (b, σ, z) is always real and equals to $r = -\sigma\left(1 - z^{\frac{1}{b}}\right)$ [43]. Denote the eigenvalues of matrix $\mathbf{A}_1$ by $-\lambda_j$; the dominant eigenvalue (which is real) is $-\lambda_1$. The FE blocks representing the eigenvalues are composed as follows

- if λ_j is real, the corresponding FE block is a degenerate block; thus the parameters are:

$$\sigma_j = \lambda_j, \quad b_j = 1, \quad z_j = 0,$$

- if $\lambda_j = a_j \pm \mathbf{i}c_j$ $(a_j > \lambda_1 > 0, c_j > 0)$ is a complex conjugate pair, the parameters are:

$$\begin{aligned} b_j &= \left\lceil \frac{2\pi}{\pi - 2\arctan\left(\dfrac{c_j}{-\lambda_1 + a_j}\right)} \right\rceil, \\ \sigma_j &= \frac{1}{2}\left(-2a_j - c_j \tan\frac{\pi}{b_j} + c_j \cot\frac{\pi}{b_j}\right), \\ z_j &= \left(1 - \left(-a_j - c_j \tan\frac{\pi}{b_j}\right)/(2\sigma_j)\right)^{b_j}, \end{aligned}$$

  where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

This construction of the FE blocks ensures that λ_1 remains the dominant eigenvalue of matrix $\mathbf{G}$, that is, the dominant eigenvalue of any FE block (r_j) is less than $-\lambda_1$ except the one(s) associated with $-\lambda_1$.

Connecting the obtained FE blocks such that the exit transition of an FE block (whose rate is $\lambda_j(1 - z_j)$, see Figure 4.1, in case of non-degenerate FE block and λ_j in case of a degenerate one) is connected to the first state of the next FE block composes a block bi-diagonal matrix (c.f. Figure 4.2). The obtained matrix $\mathbf{G}$ is Markovian and its Jordan form contains all Jordan blocks of matrix $\mathbf{A}_1$. We order the FE blocks such that the first n_1 FE blocks are the n_1 degenerate FE blocks associated with $-\lambda_1$. The order of the rest of the FE blocks is irrelevant. The FE blocks based finite Markovian representation of the eigenvalues of $\mathbf{A}_1$ is always feasible when the dominant eigenvalue condition holds. If there was a pair of complex conjugate eigenvalues $a_j \pm \mathbf{i}c_j$ which violates the dominant eigenvalue condition such that $a_j = \lambda_1$ then the denominator of b_j would be zero.

Figure 4.2 depicts an example of a Markovian generator which is the monocyclic representation of a generator with a dominant real eigenvalue $(-\lambda_1 = -\sigma_1)$ and a pair of complex conjugate eigenvalues in FE-diagonal form. In this representation there are two FE blocks, one of length $b_1 = 1$ with rate σ_1, and one of length $b_2 = 3$ with rate σ_2 and feedback probability z_2. The associated generator matrix is

$$\mathbf{G} = \left(\begin{array}{c|ccc} -\sigma_1 & \sigma_1 & 0 & 0 \\ \hline 0 & -\sigma_2 & \sigma_2 & 0 \\ 0 & 0 & -\sigma_2 & \sigma_2 \\ 0 & z\sigma_2 & 0 & -\sigma_2 \end{array}\right).$$

In order to find an equivalent representation of $\mathrm{ME}(\alpha_1, \mathbf{A}_1)$ with matrix $\mathbf{G}$ we need to compute vector γ, for which $\mathrm{ME}(\alpha_1, \mathbf{A}_1) \equiv \mathrm{ME}(\gamma, \mathbf{G})$, with the help of Theorem 4.1. Let n and m $(n \leq m)$ be

the order of $\mathbf{A}_1$ and $\mathbf{G}$, respectively. Compute matrix $\mathbf{W}$ of size $n \times m$ as the unique solution [43] of

$$\mathbf{A}_1\mathbf{W} = \mathbf{W}\mathbf{G}, \qquad \mathbf{W}\mathbf{1} = \mathbf{1},$$

and based on $\mathbf{W}$ vector γ is

$$\gamma = \alpha_1\mathbf{W}.$$

Since $\mathbf{G}$ is Markovian, the obtained $(\gamma, \mathbf{G})$ representation is already a PH representation if γ is non-negative, but this is not necessarily the case. The case when γ has negative elements is considered in the following subsection.

4.2.5 Step 4: Markovian vector

At this point in the algorithm, the ME distribution is described by representation $(\gamma, \mathbf{G})$ of order u which has a block bi-diagonal, Markovian matrix $\mathbf{G}$, and a vector γ with at least one negative element. In the next step we extend the $(\gamma, \mathbf{G})$ representation with an additional n phases in the following way.

$$\mathbf{B} = \begin{pmatrix} \mathbf{G} & -\mathbf{G}\mathbf{1} & & \\ & -\lambda & \lambda & \\ & & \ddots & \ddots \\ & & & -\lambda \end{pmatrix}, \tag{4.3}$$

where $\mathbf{B}$ is of order $u+n$ (the size of the upper left block of $\mathbf{B}$ is u, the remaining n blocks are of size one). $-\mathbf{G}\mathbf{1}$ is a non-negative column vector of size u. Due to the structural properties of $\mathbf{G}$ it contains exactly one non-zero element, which is the last element and it contains the exit rate from the last FE block. The transformation matrix $\mathbf{W}$ of size $u \times (u+n)$, which transforms from representation $(\gamma, \mathbf{G})$ to representation $(\beta, \mathbf{B})$ is the unique solution of $\mathbf{G}\mathbf{W} = \mathbf{W}\mathbf{B}$, $\mathbf{W}\mathbf{1}_{u+n} = \mathbf{1}_n$. Fortunately, due to the special structure of matrix $\mathbf{B}$, $\mathbf{W}$ is rather regular.

Lemma 4.5. $\mathbf{W}$ *has the following form:*

$$\mathbf{W} = \left(\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^n \,\middle|\, \left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-1} \tfrac{-\mathbf{G}\mathbf{1}}{\lambda} \,\middle|\, \left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-2} \tfrac{-\mathbf{G}\mathbf{1}}{\lambda} \,\middle|\, \ldots \,\middle|\, \tfrac{-\mathbf{G}\mathbf{1}}{\lambda} \right),$$

where the size of the first block is $u \times u$*, the size of each remaining block is* $1 \times u$*.*

Proof. Substituting this expression of $\mathbf{W}$ into $\mathbf{G}\mathbf{W} = \mathbf{W}\mathbf{B}$ and $\mathbf{W}\mathbf{1}_{u+n} = \mathbf{1}_n$ results in identities. □

Our goal is to find n and λ such that $\beta = \gamma\mathbf{W}$ is Markovian (that is non-negative), where

$$\gamma\mathbf{W} = \left(\gamma\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^n \,\middle|\, \gamma\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-1} \tfrac{-\mathbf{G}\mathbf{1}}{\lambda} \,\middle|\, \gamma\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-2} \tfrac{-\mathbf{G}\mathbf{1}}{\lambda} \,\middle|\, \ldots \,\middle|\, \gamma\, \tfrac{-\mathbf{G}\mathbf{1}}{\lambda} \right). \tag{4.4}$$

The first block of this vector is of size u and the remaining n blocks are of size 1. We need to prove that this vector is nonnegative for an appropriate pair (λ, n).

Theorem 4.6. *There exists a pair (λ, n) such that $\gamma\mathbf{W}$ is strictly positive.*

The rest of this subsection is devoted to proving Theorem 4.6. We assume everything that was done so far, for example that the dominant eigenvalue condition and the positive density condition hold, the density is positive at zero and also that the matrix $\mathbf{G}$ is Markovian and in FE block form such that the degenerate FE block(s) representing the dominant eigenvalue $-\lambda_1$ are the first one(s). First we present a heuristic argument, then the formal proof.

Heuristic argument

λ and n are typically chosen to be large (see [43]). However, finding an appropriate pair is not as simple as choosing some large λ and a large n. For each n, the set of appropriate values of λ forms a finite interval. If n is large enough, this interval is nonempty, but – without further considerations – it is impossible to identify this interval (or even one element of it). Vice versa, for each λ there is a finite set of appropriate values for n. This means that the naive algorithm of increasing the values of n and λ – without further considerations – may possibly never yield an appropriate pair. For this reason, we instead propose a different parametrization, which takes the dependence between n and λ into account better.

Let $\tau = n/\lambda$. τ turns out to be a value interesting in its own right. The ME pdf resulting from the pair $(\gamma\mathbf{W}, \mathbf{B})$ has a term coming from the first block of $\mathbf{B}$ and it has n terms coming from the Erlang-tail. We argue that the terms coming from the Erlang-tail can be regarded as an approximation of the original pdf on the interval $[0, \tau]$, while the term coming from the first block is some sort of correction that makes the approximation exactly equal to the original pdf. Each of the terms in the Erlang-tail contribute an Erlang pdf with rate λ and order $k \in [1, \ldots, n]$ to the pdf. The Erlang(λ, k) pdf is concentrated around the point $\frac{k}{\lambda} = \frac{k\tau}{n}$. These points are situated along the interval $[0, \tau]$ in an equidistant way with distance $\frac{1}{\lambda}$.

The weight (initial probability) of the Erlang pdf centered around the point $\frac{k}{n}\tau$ is

$$\gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G1}}{\lambda} \approx \gamma e^{\frac{k\tau}{n}\mathbf{G}} \frac{-\mathbf{G1}}{\lambda} = \frac{1}{\lambda} f_X\left(\frac{k\tau}{n}\right),$$

which means that the weights are approximately equal to samples of the original pdf at points $\frac{k\tau}{n}$, $k \in [1, \ldots, n]$ divided by λ, resulting in a pdf that is approximately equal to the original along the interval $[0, \tau]$.

The first block of $\gamma\mathbf{W}$ is different. From the form of $\mathbf{B}$ it is clear that the contribution of the first block is concentrated after the point τ; the role of this block is essentially to make a correction in the interval $[\tau, \infty]$, where the previous Erlang-approximation does not hold.

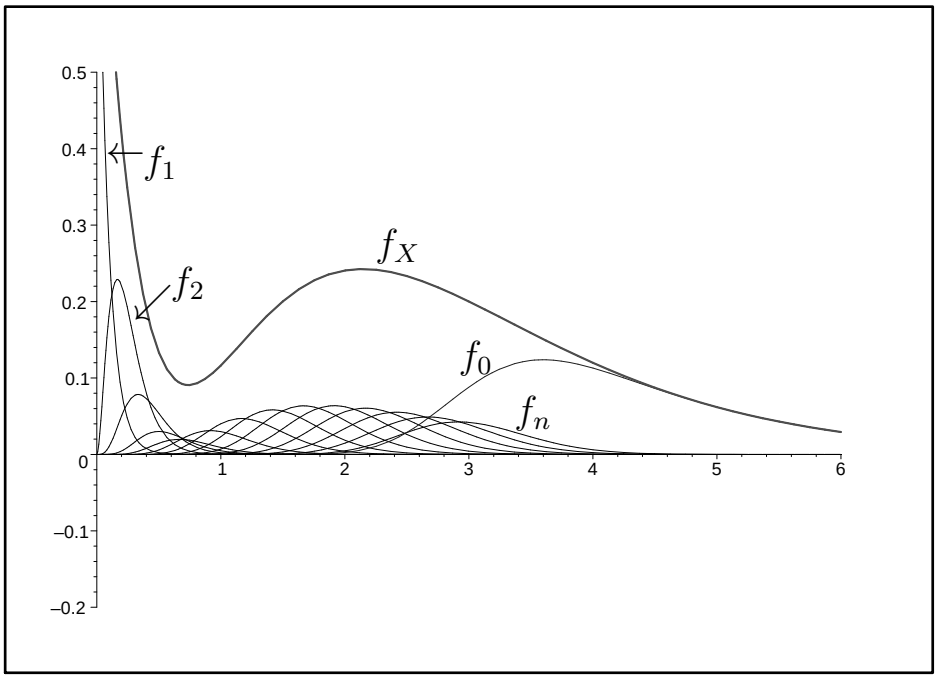

Figure 4.3: Erlang pdf's approximating the original one

Altogether the previous argument can be depicted nicely in Figures 4.3 and 4.4. We denote

$$f_k(t) = \gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G1}}{\lambda} \; g(k, \lambda, t), \quad k = 0, \ldots, n-1$$

the approximating Erlang terms and

$$f_0(t) = \gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^n e^{t\mathbf{G}}(-\mathbf{G1}) * g(n, \lambda, t)$$

the correction term. In Figure 4.3, the approximating Erlang terms roughly follow the graph of f_X, while f_0 is concentrated after τ. (The values are $\tau = 3, \lambda = 12$ and $n = 36$; to make the figure visually apprehensible, only some of the approximating Erlang functions were included with slightly increased weights.)

The value of λ controls how concentrated the approximating Erlang pdf's are and also controls how close their weights are to the sampling of the original pdf. Given that $f_X(t) > 0$ for $t \geq 0$, this means that for *any* choice of τ, the Erlang-approximation has positive weights if λ is large enough. The choice of τ is only important to make sure that the weights assigned to the correction term are also positive. Figure 4.4 shows an example where λ is too small (notably $\lambda = 4$). In this case, some of the approximating Erlang functions have negative coefficients.

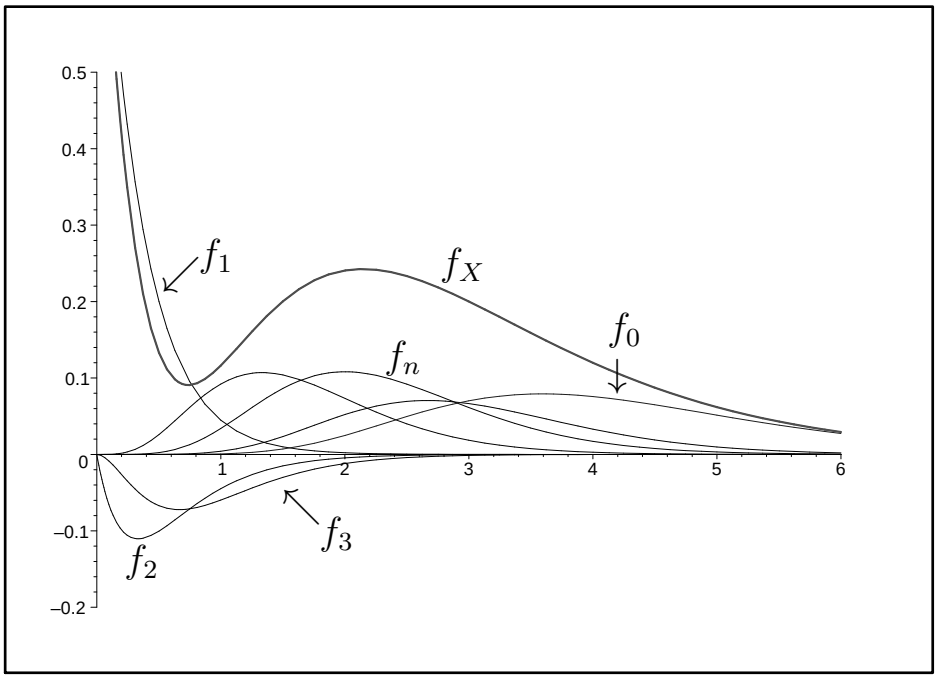

Figure 4.4: If λ is too small, some Erlang pdf's are negative

Formal proof

Before the actual proof, some results are stated as standalone lemmas.

The first one is essentially a real approximation, so we state it in that form too, along with the matrix version which is useful for our purposes. The norms we will stick to in this Chapter are: $\|.\|_1$ for row vectors, $\|.\|_\infty$ for column vectors and $\|.\|_\infty$ for matrices (which happens to be the induced norm for the vector norm $\|.\|_\infty$ when multiplying a column vector with a matrix from the left, and the induced matrix norm of the vector norm $\|.\|_1$ when multiplying a row vector with a matrix from the right).

Lemma 4.7. *i) For any fixed $r > 0$ and positive integer n,*

$$\sup_{|z|\leq r} \left| e^z - \left(1 + \frac{z}{n}\right)^n \right| \leq \frac{r^2 e^r}{2n},$$

and the supremum is obtained at $z = r$.

ii) For any $\mathbf{H}$ square matrix,

$$\left\| e^{\mathbf{H}} - \left(\mathbf{I} + \frac{\mathbf{H}}{n}\right)^n \right\| \leq \frac{r^2 e^r}{2n},$$

where $r = \|\mathbf{H}\|$.

Proof. We will prove part *i)* first.

We will begin by showing that the supremum is obtained for $z = r$.

Series expansion gives

$$e^z - \left(1 + \frac{z}{n}\right)^n = \sum_{k=0}^{\infty} \frac{z^k}{k!} B(n,k),$$

where

$$B(n,k) = \begin{cases} 1 - \frac{n(n-1)\ldots(n-k+1)}{n^k} & \text{if } k \le n \\ 1 & \text{if } k > n \end{cases}$$

Note the following properties of $B(n,k)$:

$$0 \le B(n,k) \le 1 \ \forall\, n,\ k; \qquad \lim_{n\to\infty} B(n,k) = 0 \ \forall\, k.$$

For every z with $|z| \le r$, we have

$$\left|e^z - \left(1 + \frac{z}{n}\right)^n\right| = \left|\sum_{k=0}^{\infty} \frac{z^k}{k!} B(n,k)\right| \le \sum_{k=0}^{\infty} \frac{|z|^k}{k!} B(n,k) \le \sum_{k=0}^{\infty} \frac{r^k}{k!} B(n,k) = \left|e^r - \left(1 + \frac{r}{n}\right)^n\right|.$$

Notice that the series expansion ensures $e^r - \left(1 + \frac{r}{n}\right)^n > 0$, so we only need an upper bound on $e^r - \left(1 + \frac{r}{n}\right)^n$. Using the basic inequalities

$$\ln(1+x) \ge x - \frac{x^2}{2} \ (x \ge 0) \quad \text{and} \quad e^x \ge 1 + x \ (x \in \mathbb{R})$$

we get that

$$e^r - \left(1 + \frac{r}{n}\right)^n = e^r - e^{n\ln(1+r/n)} \le e^r - e^{r - r^2/(2n)} = e^r \left(1 - e^{-r^2/(2n)}\right) \le e^r \left(1 - \left(1 - \frac{r^2}{2n}\right)\right) = e^r \frac{r^2}{2n}.$$

We note that this estimate is asymptotically sharp as $n \to \infty$.

For part *ii)*, we use the series expansion again:

$$\begin{aligned} \left\|e^{\mathbf{H}} - \left(1 + \frac{\mathbf{H}}{n}\right)^n\right\| &= \left\|\sum_{k=0}^{\infty} \frac{\mathbf{H}^k}{k!} B(n,k)\right\| \le \sum_{k=0}^{\infty} \frac{\|\mathbf{H}\|^k}{k!} B(n,k) \le \\ &\le \sum_{k=0}^{\infty} \frac{r^k}{k!} B(n,k) = e^r - \left(1 + \frac{r}{n}\right)^n \le \frac{r^2 e^r}{2n}, \end{aligned}$$

where $r = \|\mathbf{H}\|$. □

We state one more lemma. It identifies the main terms in each of the columns of $e^{t\mathbf{G}}$ when $\mathbf{G}$ is in FE block form.

Lemma 4.8.

$$\begin{aligned}
&\left(e^{t\mathbf{G}}\right)_{1j} \sim C_j t^{j-1} e^{-\lambda_1 t} \quad \text{if } 1 \le j \le n_1 \\
&\left(e^{t\mathbf{G}}\right)_{1j} \sim C_j t^{n_1-1} e^{-\lambda_1 t} \quad \text{if } n_1 < j \le u, \\
&\lim_{t\to\infty} \frac{\left(e^{t\mathbf{G}}\right)_{ij}}{\left(e^{t\mathbf{G}}\right)_{1j}} = 0 \quad \text{if} \quad 2 \le i \le u, \quad 1 \le j \le u,
\end{aligned}$$

where C_j denote positive (combinatorial) constants and $f(t) \sim g(t)$ denotes that $\lim_{t\to\infty} f(t)/g(t) = 1$. The last relation means that the first row dominates all other rows as t tends to infinity.

Proof. According to the FE block composition of $\mathbf{G}$ it has the following block structure

$$\mathbf{G} = \left[\begin{array}{c|c} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \hline 0 & \mathbf{G}_{22} \end{array}\right], \tag{4.5}$$

where

$$\mathbf{G}_{11} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_1 & \lambda_1 & \dots & 0 \\ \vdots & & & & \\ 0 & & \dots & 0 & -\lambda_1 \end{bmatrix}, \quad \mathbf{G}_{12} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 0 \\ \lambda_1 & 0 & \dots & 0 \end{bmatrix},$$

and $\mathbf{G}_{22}$ contains the rest of the FE blocks. The size of $\mathbf{G}_{11}$ is denoted by n_1 (which is the multiplicity of the dominant eigenvalue $-\lambda_1$) and the size of $\mathbf{G}_{22}$ by n_2. Let

$$\mathbf{H} = \mathbf{G} + \lambda_1 \mathbf{I},$$

and accordingly $\mathbf{H}_{11} = \mathbf{G}_{11} + \lambda_1\mathbf{I}, \mathbf{H}_{12} = \mathbf{G}_{12}$ and $\mathbf{H}_{22} = \mathbf{G}_{22} + \lambda_1\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix of appropriate size. From $\mathbf{H} = \mathbf{G} + \lambda_1\mathbf{I}$, it follows that

$$e^{t\mathbf{G}} = e^{-\lambda_1 t} e^{t\mathbf{H}},$$

and it is enough to investigate the dominant row of $e^{t\mathbf{H}}$. In the rest of the proof, $(.)_{11}, (.)_{12}, (.)_{22}$ denote the corresponding matrix blocks (*not* single elements). The eigenvalues of $\mathbf{H}_{22}$ have negative real parts. Their real parts are less than or equal to $\lambda_1 - \Re(\lambda_2)$, where $-\lambda_2$ is the eigenvalue with the second largest real part.

From the series expansion of $e^{t\mathbf{H}}$

$$e^{t\mathbf{H}} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{H}^n,$$

and from the block triangular structure of $\mathbf{H}$ we have that the upper left block is

$$(e^{t\mathbf{H}})_{11} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{H}_{11}^n,$$

where $(t\mathbf{H}_{11})^n$ can be calculated explicitly:

$$(t\mathbf{H}_{11})^n = \begin{bmatrix} 0 & \dots & 0 & (\lambda_1 t)^n & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & (\lambda_1 t)^n & \dots & 0 \\ \vdots & & & & & & \\ 0 & & & \dots & & & (\lambda_1 t)^n \\ 0 & & & & & & 0 \\ \vdots & & & & & & \\ 0 & & & \dots & & & 0 \end{bmatrix}$$

with the nonzero elements being at positions $(1, n+1), (2, n+2), \dots$. Specifically, $\mathbf{H}_{11}^n$ is 0 for $n \geq n_1$, so the sum $\sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{H}_{11}^n$, is actually finite, and from the above form it is clear that $(e^{t\mathbf{H}})_{11}$ is upper diagonal, dominated by its first row, which of course also dominates $(e^{t\mathbf{H}})_{21} = 0$.

The rest of the proof is devoted to the elements of $(e^{t\mathbf{H}})_{12}$ and $(e^{t\mathbf{H}})_{22}$. For that, we need to examine $(e^{t\mathbf{H}})_{12}$.

$$(e^{t\mathbf{H}})_{12} = \sum_{n=0}^{\infty} \frac{t^n}{n!}(\mathbf{H}^n)_{12}.$$

Here,

$$(\mathbf{H}^n)_{12} = \sum_{k=0}^{n-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{n-k-1}$$

since $\mathbf{H}$ is an upper block bi-diagonal matrix. Thus

$$\begin{aligned} (e^{t\mathbf{H}})_{12} &= \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{k=0}^{n-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{n-k-1} \\ &= \sum_{k=0}^{\infty} (\mathbf{H}_{11})^k \mathbf{H}_{12} \sum_{n=k+1}^{\infty} \frac{t^n}{n!} (\mathbf{H}_{22})^{n-k-1} \\ &= \sum_{k=0}^{n_1-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} \sum_{n=k+1}^{\infty} \frac{t^n}{n!} (\mathbf{H}_{22})^{n-k-1}. \end{aligned}$$

Again, the sum over k is finite.

The inner sum can be calculated as

$$\sum_{n=k+1}^{\infty} \frac{1}{n!} t^{n-k-1} = x^{-k-1} \sum_{n=k+1}^{\infty} \frac{1}{n!} t^n = t^{-k-1} \left(e^t - \sum_{l=0}^{k} \frac{t^l}{l!} \right),$$

and accordingly,

$$\sum_{n=k+1}^{\infty} \frac{t^n}{n!} (\mathbf{H}_{22})^{n-k-1} = (\mathbf{H}_{22})^{-(k+1)} \left(e^{t\mathbf{H}_{22}} - \mathbf{I} - t\mathbf{H}_{22} - \dots - \frac{(t\mathbf{H}_{22})^k}{k!} \right).$$

Putting it all together, we obtain that

$$(e^{t\mathbf{H}})_{12} = \sum_{k=0}^{n_1-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \left(e^{t\mathbf{H}_{22}} - \mathbf{I} - t\mathbf{H}_{22} - \cdots - \frac{(t\mathbf{H}_{22})^k}{k!} \right).$$

The form of $(\mathbf{H}_{11})^k \mathbf{H}_{12}$ guarantees that for each k

$$(\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \left(e^{t\mathbf{H}_{22}} - \mathbf{I} - t\mathbf{H}_{22} - \cdots - \frac{(t\mathbf{H}_{22})^k}{k!} \right)$$

has a single nonzero row, with $k = n_1 - 1$ corresponding to the first row being nonzero, $k = n_1 - 2$ to the second etc. Within each row, the main term is

$$-(\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \frac{(t\mathbf{H}_{22})^k}{k!} = -\frac{t^k}{k!} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-1}.$$

Specifically, the main term in each element of the first row is of order t^{n_1-1}, and the order in the other rows within the block $(e^{t\mathbf{H}})_{12}$ is smaller.

We need to calculate $\mathbf{H}_{22}^{-1}$. It can be calculated either via Cramer's rule (which allows for calculating the constants C_j explicitly, but is left to the reader), or by using the following identity:

$$\mathbf{H}_{22}^{-1} = -\int_{tt=0}^{\infty} e^{t\mathbf{H}_{22}} \mathrm{d}t = -\int_{t=0}^{\infty} e^{t\lambda_1} \cdot e^{t\mathbf{G}_{22}} \mathrm{d}t.$$

The integral exists because all eigenvalues of $\mathbf{H}_{22}$ have negative real part. $e^{\lambda_1 t}$ is a positive function ("weight") and $e^{t\mathbf{G}_{22}}$ contains the transition probabilities of a CTMC, so all elements of $e^{t\mathbf{G}_{22}}$ are positive for all $t > 0$. Thus all elements of $\mathbf{H}_{22}^{-1}$ are negative, and the single nonzero row of $-(\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \frac{(t\mathbf{H}_{22})^k}{k!}$ is strictly positive.

Finally, since the block $(\mathbf{H})_{22}$ has eigenvalues with negative real part, the elements of $(e^{t\mathbf{H}})_{22}$ decay exponentially, so they are of course dominated by the first row of $(e^{t\mathbf{H}})_{12}$. □

Note that the last part of Lemma 4.8 is stated as $\frac{(e^{t\mathbf{G}})_{ij}}{(e^{t\mathbf{G}})_{1j}} \to 0$; in fact, the elements $\left(e^{t\mathbf{G}}\right)_{ij}$ are in a form similar to $\left(e^{t\mathbf{G}}\right)_{1j}$, just with either the same exponential term and lower degree polynomial terms, or lower exponent (and in this case, the polynomial term does not matter). The actual exponents and polynomial terms, along with the constants C_j can be calculated explicitly from the proof of Lemma 4.8, but will not be used.

We emphasize that Lemma 4.8 relies heavily on the monocyclic structure of $\mathbf{G}$, notably on the fact that the upper bi-diagonal elements (elements $(1,2), (2,3), \dots$) of the matrix are strictly positive.

Now we are ready to prove Theorem 4.6.

Proof of Theorem 4.6.

We assume that the matrix exponential density function f_X associated with representation $(\gamma, \mathbf{G})$ satisfies $f_X(0) > 0$, the dominant eigenvalue and the positive density conditions, and that $\mathbf{G}$ is in

monocyclic block structure with the first block corresponding to the dominant eigenvalue λ_1.

First we show that the first coordinate of γ, denoted by γ_1, is positive.

If $\gamma_1 = 0$, then the multiplicity of $-\lambda_1$ is $n_1 - 1$ according to the structure of matrix $\mathbf{G}$ (see (4.5) in the proof of Lemma 4.8, which is in conflict with the fact that the multiplicity of $-\lambda_1$ in the minimal ME representation is n_1.

$f_X(t)$ is dominated by the first row of $e^{t\mathbf{G}}$ for large values of x and consequently the sign of $f_X(t)$ is determined by γ_1. The elements of $e^{t\mathbf{G}}$ are transient probabilities of the Markov chain with generator $\mathbf{G}$, consequently they are non-negative. The elements of the first row of $e^{t\mathbf{G}}$ are strictly positive for $t > 0$ because the FE-blocks are connected that way that all states are reachable from the first state (cf. Figure 4.2). According to Lemma 4.8 $f_X(t)$ is dominated by the first row of $e^{t\mathbf{G}}$ for large values of t and consequently the sign of $f_X(t)$ is determined by γ_1. More precisely, Lemma 4.8 implies that

$$0 < f_X(t) = \gamma(-\mathbf{G})e^{t\mathbf{G}}\mathbf{1} \sim C\lambda_1\gamma_1 t^{n_1-1}e^{-\lambda_1 t}$$

where $C = \sum_{j \geq n_1} C_j > 0$ and $\lambda_1 > 0$.

Next we show that there exists a τ such that $\gamma e^{\tau\mathbf{G}}$ is positive.

For the first row of $\gamma e^{\mathbf{G}t}$ we have

$$\begin{aligned} \left(\gamma e^{t\mathbf{G}}\right)_{1j} &\sim C_j\gamma_1 t^{j-1}e^{-\lambda_1 t} && \text{if } j < n_1, \\ \left(\gamma e^{t\mathbf{G}}\right)_{1j} &\sim C_j\gamma_1 t^{n_1-1}e^{-\lambda_1 t} && \text{if } n_1 \leq j \leq u, \end{aligned}$$

from Lemma 4.8. Thus $\gamma e^{t\mathbf{G}}$ is positive if t is large enough. For a constructive procedure to find τ, one can double t starting from n_1/λ_1 as long as $\min(\gamma e^{t\mathbf{G}}) < 0$. It is not necessary to find the smallest t for which $\gamma e^{t\mathbf{G}}$ is nonnegative.

After that we show that there exists λ' such that $\gamma(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\lambda\tau} > 0$ for $\lambda \geq \lambda'$.

Apply Lemma 4.7 with $\mathbf{H} = \mathbf{G}\tau$ and $n = \lambda\tau$ to get that

$$\left\| \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^{\lambda\tau} - e^{\mathbf{G}\tau} \right\| \to 0$$

as $\lambda \to \infty$, and consequently

$$\left\| \gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^{\lambda\tau} - \gamma e^{\mathbf{G}\tau} \right\| \to 0,$$

meaning that $\gamma(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\lambda\tau}$ is also strictly positive if λ is large enough. Let $\epsilon_1 = \min(\gamma e^{\mathbf{G}\tau})$; in accordance with Lemma 4.7, define λ' as the solution of

$$\|\gamma\| \frac{(g\tau)^2 e^{g\tau}}{2\lambda\tau} = \epsilon_1. \tag{4.6}$$

where $g = \|\mathbf{G}\|$. Then $\gamma(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\lambda\tau} > 0$ for $\lambda > \lambda'$, because the left-hand side is a strictly monotone

decreasing function of λ. Note that λ' is explicitly computable from (4.6).

Next we investigate the sign of the rest of the elements of vector $\gamma\mathbf{W}$. We apply Lemma 4.7 again, this time for $\mathbf{H} = \frac{k\mathbf{G}}{\lambda}$ and $n = k$ to get

$$\left\| e^{\frac{k\mathbf{G}}{\lambda}} - \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \right\| \leq e^{\frac{kg}{\lambda}} \frac{(kg)^2}{2k\lambda^2} \leq e^{\tau g} \frac{\tau g^2}{2\lambda}$$

uniformly in $0 \leq k \leq \lambda\tau$.

Let $\epsilon_2 = \inf_{0\leq t\leq\tau} f_X(t) = \inf_{0\leq t\leq\tau} \gamma e^{t\mathbf{G}}(-\mathbf{G})\mathbf{1}$ (actually, we do not need the exact value of ϵ_2; any smaller value works as well). Since $f_X(0) > 0$ as a result of Step 3 in Section 4.2.3, ϵ_2 is strictly positive, due to the positive density condition. Let V_k be the k-th coordinate of $\gamma\mathbf{W}$ associated with the Erlang tail in (4.4); that is,

$$V_k = \gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G}\mathbf{1}}{\lambda}.$$

Then

$$\left|\lambda V_k - f_X(\tfrac{k}{\lambda})\right| = \left| \gamma \left[e^{\frac{k\mathbf{G}}{\lambda}} - \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \right] \mathbf{G}\mathbf{1} \right| \leq \|\gamma\| \left\| e^{\frac{k\mathbf{G}}{\lambda}} - \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \right\| \|\mathbf{G}\|\|\mathbf{1}\| \leq \|\gamma\| e^{\tau g} \frac{\tau g^2}{2\lambda} g \|\mathbf{1}\|.$$

Define λ'' as the solution of

$$\|\gamma\| e^{\tau g} \frac{\tau g^2}{2\lambda} g \|\mathbf{1}\| = \epsilon_2. \tag{4.7}$$

λ'' is also explicitly computable. (Note that $\|\mathbf{1}\| = 1$). For all $\lambda > \lambda''$ we have $V_k > 0$ because $f_X(\frac{k}{\lambda}) \geq \epsilon_2$ and the difference between λV_k and $f_X(\frac{i}{\lambda})$ is less than ϵ_2.

Putting these together, we get that for τ and $\lambda = \max(\lambda', \lambda'')$ both parts of the vector $\gamma\mathbf{W}$, that is, $\gamma\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^n$ and $\gamma\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G}\mathbf{1}}{\lambda}$ for $k = 0, 1, \ldots, n-1$, are positive where $n = \lceil \tau\lambda \rceil$ and the obtained representation is indeed Markovian.

4.2.6 Step 5: Correction related to Step 2

If Step 2 was applied, $(\beta, \mathbf{B})$ is actually a Markovian representation for f_Y; Lemma 4.4 ensures that

$$\beta' = \{1, 0, 0, \ldots, 0\}$$

$$\mathbf{B}' = \begin{pmatrix} -\mu & \mu & & \\ & \ddots & \ddots & \\ & & -\mu & \mu\beta \\ & & & \mathbf{B} \end{pmatrix}$$

is a Markovian representation for $f_X(t) = f_Y(x) * g(l, \mu, t)$.

4.3 Worked example

Let us consider the following ME representation.

$$\alpha = \frac{102}{139}\begin{pmatrix} 1 & 1 & -\frac{1}{3} & \frac{2}{3} & -\frac{5}{2} & \frac{12}{17} & \frac{14}{17} \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 & -3 & -5 \end{pmatrix},$$

then

$$f(t) = -\alpha \mathbf{A} e^{t\mathbf{A}} \mathbf{1} = \frac{102}{139}\left(te^{-t} + e^{-t} + e^{-3t} - 10e^{-4t} + e^{-5t}\left(8\cos(3t) + 4\sin(3t)\right)\right).$$

The eigenvalues of A are -1 (with multiplicity 2), $-3, -4, -5+3i, -5-3i$ and 1. The eigenvalue 1 is redundant: the corresponding right-eigenvector is orthogonal to α, thus it does not appear in the pdf. It is eliminated during Step 1.

After Step 1, a minimal ME representation is obtained:

$$\alpha_1 = \frac{102}{139}\begin{pmatrix} 1 & 1 & \frac{1}{3} & -\frac{5}{2} & \frac{13+\mathbf{i}}{17} & \frac{13-\mathbf{i}}{17} \end{pmatrix},$$

$$\mathbf{A}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5+3\mathbf{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & -5-3\mathbf{i} \end{pmatrix}.$$

Since $f(0) = 0$, Step 2 needs to be applied.

$$f(0) = 0 \qquad f'(0) = 7 > 0,$$

so the value of l in Lemma 4.3 is $l = 1$. Setting $\mu = 10$, the transformed pdf after Step 2 (borrowing the notation f_Y from Lemma 4.3) is

$$f_Y(t) = \frac{102}{139}\left(\frac{9}{10}te^{-t} + e^{-t} + \frac{7}{10}e^{-3t} - 6e^{-4t} + \frac{13+\mathbf{i}}{5}e^{(-5+3\mathbf{i})t} + \frac{13-\mathbf{i}}{5}e^{(-5-3\mathbf{i})t}\right)$$

and the corresponding representation for f_Y is

$$\alpha_2 = \frac{102}{139}\left(\begin{array}{cccccc} \frac{9}{10} & 1 & \frac{2}{5} & -\frac{7}{15} & \frac{-2-\mathbf{i}}{5} & \frac{-2+\mathbf{i}}{5} \end{array} \right),$$

$$\mathbf{A}_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5+3\mathbf{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & -5-3\mathbf{i} \end{pmatrix}.$$

From now on, we work with this representation. In Step 3, the eigenvalue pair $5 \pm 3\mathbf{i}$ is represented by a feedback-Erlang block. The order of this pair is $b = 4$, and the corresponding FE-block is

$$\begin{pmatrix} -5 & 5 & 0 & 0 \\ 0 & -5 & 5 & 0 \\ 0 & 0 & -5 & 5 \\ \frac{81}{125} & 0 & 0 & -5 \end{pmatrix}.$$

Step 3 results in the representation

$$\gamma = \frac{102}{139}\left(\begin{array}{cccccccc} \frac{315}{2176} & \frac{10733}{21760} & \frac{6641}{32640} & \frac{8399}{21760} & \frac{147}{680} & -\frac{67}{272} & -\frac{45}{1088} & \frac{225}{1088} \end{array} \right),$$

$$\mathbf{G} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -5 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -5 & 5 \\ 0 & 0 & 0 & 0 & \frac{81}{125} & 0 & 0 & -5 \end{pmatrix}.$$

Since γ still contains negative elements, Step 4 needs to be applied.

Following the algorithm in the proof of Theorem 4.6, we obtain the following values:

- $\tau = 0.5$ (from $\gamma e^{\mathbf{G}\tau} > 0$),
- $g = \|G\| = 10$,
- $\|\gamma\| < 1.5$,
- $\epsilon_1 > 0.05$ (for $\tau = 0.5$),
- $\lambda' = 112000$ from (4.6),
- $\epsilon > 0.069$, and thus $\lambda'' = 806600$ from (4.7).

This means that applying Step 4 with $\lambda = 806600$ and $n = \tau\lambda = 403300$ we obtain a Markovian representation for f_Y in the form of (4.3).

Finally, Step 5 applies, so by Lemma 4.4 with $\mu = 10$ and $k = 1$, we obtain a Markovian representation for the original ME($\alpha, \mathbf{A}$). The representation is of order 403309. Note that the order of this PH representation is *very* far from minimal.

The point where the order of the representation gets very large is the calculation of λ in Step 4; however, the estimates (4.6) and (4.7) are not sharp in the sense that a much smaller value might also be suitable. For the worked example, $\lambda = 32$ is also suitable (this can be checked directly through (4.4)) and gives a PH representation of order 25. In practice, it might be better to search for the smallest possible λ instead of relying on (4.6) and (4.7) (which nevertheless provide an upper bound for the search).

We also note that the value of τ is also subject to optimization; for the worked example, $\tau = 13/30$ and $\lambda = 270/13$ yields $n = 9$ for a representation of order 18. However, this is still not necessarily a minimal PH representation; we do not pursue a minimal representation any further.

4.4 Proofs for the necessary direction

Definition 6. *The PH representation (α, $\mathbf{A}$) is* redundant *if it contains at least one state which cannot be visited by the Markov chain with initial distribution α and generator $\mathbf{A}$. Otherwise (α, $\mathbf{A}$) is* non-redundant.

If the ME representation (α, $\mathbf{A}$) is redundant then it is possible to identify and eliminate the redundant states in the following way. Consider the vector $-\alpha\mathbf{A}^{-1}$. The stochastic interpretation of its ith coordinate is the mean time spent in state i before absorption. If the ith element of vector $-\alpha\mathbf{A}^{-1}$ is zero then state i is redundant and the associated elements can be deleted from vector α and matrix $\mathbf{A}$ without changing the distribution of time till absorption.

Lemma 4.9. *If X is PH(α,$\mathbf{A}$) distributed, then the positive density condition holds, that is,*

$$f_X(t) > 0 \qquad \forall t > 0.$$

Proof. According to the previous remark, we may assume that (α,$\mathbf{A}$) is non-redundant; then there is a path from every state with positive initial probability to the absorbing state and every state belongs to one of those paths. Consequently, the Markov chain is in state j at time t with positive probability, for any time $t > 0$ and for any state j. Let state i be a transient state from where the absorption rate g_i is positive. Then

$$f_X(t) = \alpha e^{\mathbf{A}t}(-\mathbf{A})\mathbf{1} = \sum_{j=1}^{n} \Pr(Z(t) = j)g_j \geq \Pr(Z(t) = i)g_i > 0,$$

where $Z(t)$ denotes the underlying Markov chain. □

Lemma 4.10. *If X is PH(α,$\mathbf{A}$) distributed, then the dominant eigenvalue condition holds.*

Before proving Lemma 4.10, we elaborate on the form of a minimal representation. Let $ME(\gamma, \mathbf{G})$ be a minimal ME representation for X. Consider its pdf using the Jordan decomposition of $\mathbf{G}$ ($\mathbf{G} = \mathbf{PJP}^{-1}$)

$$f_X(t) = -\gamma \mathbf{P}\mathbf{J}e^{t\mathbf{J}}\mathbf{P}^{-1}\mathbf{1} = \sum_{i=1}^{l} -\gamma \mathbf{P}_i \mathbf{J}_i e^{t\mathbf{J}_i}\mathbf{P}'_i\mathbf{1},$$

where $\mathbf{J}_i$ denotes the Jordan-block corresponding to the eigenvalue $-\lambda_i$ and $\mathbf{P}_i$ denotes the submatrix of $\mathbf{P}$ containing only the columns corresponding to $\mathbf{J}_i$. $\mathbf{P}'_i$ denotes the submatrix of $\mathbf{P}^{-1}$ that contains only the rows corresponding to $\mathbf{J}_i$ (thus $\mathbf{P}_i$ is of size $n \times n_i$, where n_i is the multiplicity of $-\lambda_i$ and n is the size of $\mathbf{G}$, and $\mathbf{P}'_i$ is of size $n_i \times n$). In $\mathbf{P}_i$, the first column of each block is the (unique, up to a constant factor) right eigenvector $\mathbf{v}_i$ corresponding to that eigenvalue and the other columns are generalized eigenvectors. Similarly in $\mathbf{P}'_i$, the last row of each block is the (unique, up to a constant factor) left eigenvector $\mathbf{u}_i$ corresponding to that eigenvalue and the rest of the rows are generalized eigenvectors. If $i \neq j$, then $\mathbf{P}'_i\mathbf{P}_j = \mathbf{0}$.

The dominant term of $e^{t\mathbf{J}_i}$ is equal to $\frac{t^{n_i-1}e^{-\lambda_i t}}{(n_i-1)!}$ (where n_i denotes the size of $\mathbf{J}_i$), and it is situated in the upper right corner. Within $-\gamma \mathbf{P}_i \mathbf{J}_i e^{t\mathbf{J}_i}\mathbf{P}'_i\mathbf{1}$ this dominant term is obtained exactly when taking

$$-\gamma \mathbf{v}_i \mathbf{J}_i e^{t\mathbf{J}_i}\mathbf{u}_i\mathbf{1} = (\gamma \mathbf{v}_i)\lambda_i \frac{t^{n_i-1}e^{-\lambda_i t}}{(n_i-1)!}(\mathbf{u}_i\mathbf{1}).$$

If any of the coefficients $(\gamma \mathbf{v}_i)$ and $(\mathbf{u}_i\mathbf{1})$ were 0, this term would vanish. Properties P3 and P4 ensure that this is not the case, in other words, all eigenvalues contribute to the pdf with maximal multiplicity (Property P2).

This allows us to prove the DEC for any (possibly non-minimal) Markovian representation $(\alpha, \mathbf{A})$ by proving that there exists a real eigenvalue of $\mathbf{A}$ that is strictly greater than the real part of all other eigenvalues AND this eigenvalue contributes to the pdf with maximal multiplicity.

The proof of Lemma 4.10 is based essentially on the Perron–Frobenius lemma. We begin by citing the Perron–Frobenius lemma along with a necessary definition, see for example [42].

Definition 7. *An $n \times n$ matrix $\mathbf{M}$ is* reducible *if there exists a nontrivial partition $I \cup J$ of $\{1, 2, \ldots, n\}$ such that*

$$\mathbf{M}_{ij} = 0 \qquad \forall i \in I, j \in J.$$

Otherwise, $\mathbf{M}$ is irreducible.

In case $\mathbf{M}$ is the transient generator of a PH distribution, then irreducibility means that each state can be reached from any other state before absorption, in this case we say that $\mathbf{M}$ has a single communicating class. If the Markov chain defined by $\mathbf{M}$ has multiple communicating classes, they correspond to a partition of the states as in the above definition.

Theorem 4.11 (Perron–Frobenius)**.** *If the irreducible matrix $\mathbf{M}$ has nonnegative elements, then there exists a positive eigenvalue ν_1 of $\mathbf{M}$ such that*

- ν_1 *has multiplicity* 1,
- $\nu_1 \geq |\nu_i| \, \forall i$ *where* v_i *denote the eigenvalues of* $\mathbf{M}$*, and*
- *the corresponding right-eigenvector* $\mathbf{v}_1$ *is strictly positive (note that* $\mathbf{v}_1$ *is unique up to a constant factor; it can be chosen such that* $\mathbf{v}_1$ *is strictly positive).*

See Theorem 3 in [55] for a short, self-contained proof or Chapter 8 in [42] for a more detailed discussion. Note that the same conclusion holds for the left-eigenvector $\mathbf{u}_1$ as well. Note that the fact that ν_1 is positive with multiplicity 1 and $\nu_1 \geq |\nu_i|$ mean that $\Re(\nu_i) < \nu_1$ for $i \neq 1$.

Proof of Lemma 4.10.

In case $\mathbf{A}$ has a single communicating class we apply Theorem 4.11 to the matrix $\mathbf{M} = \mathbf{A} + \omega\mathbf{I}$, where $\omega = \max_i |a_{ii}|$. Given that the matrix $\mathbf{A}$ is Markovian, $\mathbf{M}$ is nonnegative with the same eigenvectors and the eigenvalues shifted by ω. The dominant eigenvalue ν_1 of $\mathbf{M}$ corresponds to the dominant eigenvalue $-\lambda_1$ of $\mathbf{A}$, that is $\nu_1 = -\lambda_1 + \omega$ and the same relation holds for the other eigenvectors. Clearly for $i \neq 1$

$$\Re(\nu_i) < \nu_1 \quad \Longrightarrow \quad \Re(-\lambda_i) < -\lambda_1.$$

If $\mathbf{A}$ has a single communicating class then Theorem 4.11 guarantees that the multiplicity of $-\lambda_1$ is 1; this means that the unique dominant term in the pdf is $(\alpha\mathbf{v}_1)\lambda_1 e^{-\lambda_1 t}(\mathbf{u}_1\mathbf{1})$. Strict positivity of $\mathbf{v}_1$ and $\mathbf{u}_1$ ensure $\alpha\mathbf{v}_1 > 0$ and $\mathbf{u}_1\mathbf{1} > 0$, so indeed λ_1 contributes to the pdf with multiplicity 1.

If $\mathbf{A}$ has several communicating classes, the states can be renumbered such that $\mathbf{A}$ is an upper block triangular matrix, where each diagonal block corresponds to a communicating class and the blocks above the diagonal correspond to transitions between classes. The diagonal blocks are denoted by $\mathbf{B}_1, \ldots, \mathbf{B}_k$. The eigenvalues of $\mathbf{A}$ are the union of the eigenvalues associated with these diagonal blocks. Each $\mathbf{B}_i$ is itself the generator of a transient Markov chain, and, since $\mathbf{B}_i$ is also irreducible, Theorem 4.11 can be applied to each of them (technically, it is applied for $M_i = B_i + \omega_i I$ for a large enough ω_i). It follows that each of these blocks (communicating classes) has its own dominant eigenvalue such that within that class, the real parts of all other eigenvalues are strictly smaller. It follows directly that the largest eigenvalue of $\mathbf{A}$ (denoted by $-\lambda_1$) is real and has $-\lambda_1 > \Re(-\lambda_i)$ for all $\lambda_i \neq \lambda_1$.

However, as opposed to the single class case, the multiplicity of $-\lambda_1$ may be higher than 1. Also, there may be several eigenvectors corresponding to $-\lambda_1$. This means that in order to calculate the contribution of $-\lambda_1$ to the pdf, we need to be slightly more meticulous. The proof is essentially a transformation of the matrix $\mathbf{A}$ to a form that is similar to the Jordan form (but not the same), while preserving some nonnegativity of $\mathbf{A}$ and α (where it is important). We also present a numerical example at the end of this section to demonstrate the steps of the proof.

Let $\mathbf{Q}_i\mathbf{J}_i\mathbf{Q}_i^{-1} = \mathbf{B}_i$ be the Jordan decomposition of $\mathbf{B}_i$. We assume that the first block of $\mathbf{J}_i$ is the single dominant eigenvalue of $\mathbf{B}_i$; Theorem 4.11 thus guarantees that the first column of $\mathbf{Q}_i$, which is the corresponding right eigenvector, is strictly positive, and the first row of $\mathbf{Q}_i^{-1}$, which is

the corresponding left eigenvector, is also strictly positive. Create the transformation matrix

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{Q}_2 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & & \dots & 0 & \mathbf{Q}_k \end{bmatrix}.$$

Then $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ is an upper triangular matrix that contains the eigenvalues of $\mathbf{A}$ in its diagonal. Applying this transformation to the pdf, we get

$$f_X(t) = -\alpha\mathbf{A}e^{t\mathbf{A}}\mathbf{1} = -(\alpha\mathbf{Q})(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})e^{t(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})}(\mathbf{Q}^{-1}\mathbf{1}).$$

Take all rows and columns of $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ that have $-\lambda_1$ in the diagonal. Denote this submatrix by $\mathbf{B}$. The submatrix $\mathbf{B}$ is responsible for the whole contribution of $-\lambda_1$. $\mathbf{B}$ can be calculated as

$$\mathbf{B} = \mathbf{R}\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}\mathbf{R}^T$$

where $\mathbf{R}$ is a $n_1 \times n$ binary matrix (whose elements are either 0 or 1) where n_1 is the multiplicity of the dominant eigenvalue in $\mathbf{A}$ and n is the size of $\mathbf{A}$; row i in $\mathbf{R}$ is equal to the unit vector $\mathbf{e}_j$ if the i-th instance of $-\lambda_1$ in the diagonal of $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ is at coordinate j, j. $(\alpha\mathbf{Q})$ is strictly positive on the coordinates corresponding to $\mathbf{B}$ since the dominant eigenvectors of $\mathbf{Q}_i$ are strictly positive and the block of α associated with $\mathbf{Q}_i$ is nonnegative and different from 0 (if it was 0 then PH(α,$\mathbf{A}$) would be redundant). Similarly, $(\mathbf{Q}^{-1}\mathbf{1})$ is strictly positive on the coordinates corresponding to $\mathbf{B}$.

Finally, we argue that we can identify the dominant term in $e^{t\mathbf{B}}$ and see that it has a positive coefficient. This is done directly instead of transforming $\mathbf{B}$ to Jordan form. To this end, note that the offdiagonal elements of $\mathbf{B}$ are nonnegative since $\mathbf{A}$ originally contained nonnegative elements above the diagonal, which were then multiplied by the strictly positive dominant left and right eigenvectors of each block $\mathbf{B}_i$.

The matrix $\lambda_1\mathbf{I}+\mathbf{B}$ is strictly upper triangular, thus nilpotent; this implies that the series expansion

$$e^{t(\lambda_1\mathbf{I}+\mathbf{B})} = \sum_{k=0}^{\infty} \frac{(t(\lambda_1\mathbf{I}+\mathbf{B}))^k}{k!}$$

is actually a finite sum, and $e^{t(\lambda_1\mathbf{I}+\mathbf{B})}$ is a polynomial of t. The dominant term in $e^{t\mathbf{B}}$ is equal to the last nonzero term of this polynomial, multiplied by $e^{-\lambda_1 t}$. The coefficient of this term is necessarily positive since $(\lambda_1\mathbf{I}+\mathbf{B})$ and thus powers of $(\lambda_1\mathbf{I}+\mathbf{B})$ do not have negative elements.

Consequently, we have proved that λ_1 contributes to the pdf

$$f_X(t) = -\alpha\mathbf{A}e^{t\mathbf{A}}\mathbf{1} = -(\alpha\mathbf{A})(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})e^{t(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})}(\mathbf{Q}^{-1}\mathbf{1}).$$

with maximal multiplicity and with a positive coefficient, and the DEC holds.

Example 1. *Let*

$$\mathbf{A} = \left[\begin{array}{ccc|ccc|cc|cc} -4 & 1 & 1 & 0 & 0.2 & 0.4 & 0 & 0 & 0 & 0.4 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & -3 & 0 & 0 & 0 & 0 & 0.2 & 0.4 & 0.2 \\ \hline 0 & 0 & 0 & -4 & 3 & 0.2 & 0.2 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 & 0.2 & 0.2 & 0 & 0.2 \\ \hline 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & -8 & 2 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & -7 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{array}\right].$$

$\mathbf{A}$ *has 5 communicating classes:* $\mathbf{B}_1$ *has size 3 and dominant eigenvalue* -1, $\mathbf{B}_2$, $\mathbf{B}_3$ *and* $\mathbf{B}_4$ *are of size 2 and their dominant eigenvalues are* $-1, -1$ *and* -4 *respectively;* $\mathbf{B}_5$ *is of size 1 with dominant eigenvalue* -1. *Thus* $\lambda_1 = 1$.

$$\mathbf{Q} = \left[\begin{array}{ccc|cc|cc|cc|c} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Notice that in $\mathbf{Q}$, *the first column in each block is strictly positive. Even though it is not displayed in this example,* $\mathbf{Q}$ *(and* $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$*) may contain complex numbers, but only in rows and columns*

corresponding to non-dominant eigenvalues.

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \begin{bmatrix} \boxed{-1} & 0 & 0 & \boxed{0.05} & 0.05 & \boxed{0.10} & -0.10 & 0.25 & 0.20 & \boxed{0.15} \\ 0 & -3 & 0 & 0.10 & 0.10 & 0.20 & -0.20 & 0.50 & 0.40 & 0.30 \\ 0 & 0 & -5 & -0.15 & 0.30 & -0.15 & -0.30 & 0.25 & 0.20 & -0.25 \\ \boxed{0} & 0 & 0 & \boxed{-1} & 0 & \boxed{0.25} & 0.15 & 0.35 & 0 & \boxed{0.15} \\ 0 & 0 & 0 & 0 & -5 & -0.05 & 0.05 & -0.15 & -0.40 & 0.05 \\ \boxed{0} & 0 & 0 & \boxed{0} & 0 & \boxed{-1} & 0 & 0.20 & 0.30 & \boxed{0} \\ 0 & 0 & 0 & 0 & 0 & 0 & -3 & -0.20 & -0.30 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 9/35 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -11 & -6/35 \\ \boxed{0} & 0 & 0 & \boxed{0} & 0 & \boxed{0} & 0 & 0 & 0 & \boxed{-1} \end{bmatrix}.$$

The rows and columns that include the dominant eigenvalue are marked and so

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \mathbf{R}\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}\mathbf{R}^T = \begin{bmatrix} -1 & 0.05 & 0.10 & 0.15 \\ 0 & -1 & 0.25 & 0.15 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

The last nonzero power of the nilpotent matrix $\lambda_1\mathbf{I} + \mathbf{B}$ *is*

$$(\lambda_1\mathbf{I} + \mathbf{B})^2 = \begin{bmatrix} 0 & 0 & 0.00125 & 0.0075 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

whose nonzero elements are all positive.

4.5 Outlook

The algorithm presented in this chapter is constructive, but the resulting PH representation is usually not minimal. Finding a minimal PH representation generally seems to be out of reach at the moment. When searching for relatively small representations, heuristic arguments also perform reasonably well. For a heuristic argument that improves the FE block structure of Step 3, see [27]. We also refer to the program package BuTools [8] developed specifically for PH/ME distributions and two classes of related processes called Markov arrival processes (MAPs)/rational arrival processes (RAPs). For discussion and results on MAPs and RAPs, see [7]. The package includes a number of iterative methods for finding Markovian representations.

For results on various transformation techniques, see [41].

For computationally efficient methods to carry out some of the calculations presented here, see

[50].

We also mention that the order of the representation is not the only option for measuring how "good" a representation is. Another possibility is to consider the total number of nonzero elements in the representation. Yet another option is to examine the running time when simulating the actual Markov process defined by the PH representation. For more on this topic, see [25].

Chapter 5

Mean-field limit for population models with generally-timed transitions

The chapter is structured as follows: Section 5.1 gives a setup for the classic Kurtz theorem for density-dependent Markov population models including a short, self-contained proof. Section 5.2 gives a proper setup for the generalized semi-Markov population models, stating the main theorem of this chapter. Section 5.3 provides an example. Section 5.4 contains the proof for the mean-field convergence.

5.1 Markov population models

We begin by formulating the classic result of Kurtz; the main goal of this section is to familiarize the reader with the notation and some of the techniques before tackling the semi-Markovian setup.

The notations used are the same as in Section 1.4; N is the population size, each individual is inhabiting a state from a finite set $\mathcal{S}$ and each individual in state i performs Markov transitions to other states j with rate r_{ij}^N. The global state of the system is the total number of individuals in each state, that is, a vector $\mathbf{x}^N \in \{0,1,\dots,N\}^{(S)}$ with $x_1^N + \cdots + x_{|\mathcal{S}|}^N = N$. We are interested in the evolution of the system in time.

We give a construction of the above Markov population model in terms of Poisson processes. Let $P_{ij}(.)$ be independent Poisson-processes with rate 1 for every $i \neq j \in \mathcal{S}$. The *Poisson-representation* of $\mathbf{x}^N(t)$ is

$$x_i^N(t) = x_i^N(0) - \sum_{j:j\neq i} P_{ij}\left(\int_0^t x_i^N(u) r_{ij}^N(\mathbf{x}^N(u))\mathrm{d}u\right) + \sum_{j:j\neq i} P_{ji}\left(\int_0^t x_j^N(u) r_{ji}^N(\mathbf{x}^N(u))\mathrm{d}u\right) \tag{5.1}$$

for $i = 1, \ldots, |\mathcal{S}|$.

To see that the above formula describes the same model as in Section 1.4, note that the *total rate* of a jump from i to j for any individual in the system is $x_i^N r_{ij}^N(\mathbf{x}^N)$ if the system is in state $\mathbf{x}^N$. The idea behind the above representation is that to track the evolution of x_i^N it suffices to keep track of the total number of transitions to and from state i, it does not matter which individual made the actual transition.

We aim to rewrite the formula to the normalized version of the process. Let $\bar{\mathbf{x}}^N(t) = \frac{\mathbf{x}^N(t)}{N}$. The density-dependent assumption means that

$$\frac{x_i^N}{N} r_{ij}^N(\mathbf{x}^N) = r_{ij}\left(\frac{\mathbf{x}^N}{N}\right)$$

for some common r_{ij} (independent of N), where $r_{ij} : [0,1]^{\mathcal{S}} \to \mathbb{R}$. The inclusion of the term x_i as a multiplier on the left hand side means that r_{ij} corresponds to the *aggregate rate* of jumps from i to j instead of the "individual rate"; the aggregate rate is proportional to the number of individuals in state i. This notation of r_{ij} serves to make the formulas more compact. Also, r_{ij} needs to be defined only on the subset $x_1 + \cdots + x_{|\mathcal{S}|} = 1$, which is an $(|\mathcal{S}| - 1)$-dimensional simplex.

Accordingly, the normalized version of (5.1) is

$$\bar{x}_i^N(t) = \bar{x}_i^N(0) - \sum_{j:j\neq i} \frac{1}{N} P_{ij}\left(N \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) + \sum_{j:j\neq i} \frac{1}{N} P_{ji}\left(N \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) \tag{5.2}$$

for $i = 1, \ldots, |\mathcal{S}|$.

The Poisson representation provides an inherent coupling of the Markov population model for different values of N; however, this coupling is technical without any deeper meaning.

We further assume that r_{ij} are *Lipschitz-continuous* with some common Lipschitz-constant R (R will be assumed to also be an upper bound on r_{ij}), and that the initial conditions converge to some $v(0)$. In most applications, $\bar{\mathbf{x}}^N(0)$ are deterministic, but we may allow random initial conditions as well, and thus the convergence of the initial condition is formulated as

$$\lim_{N\to\infty} \mathbf{P}(\|\mathbf{v}(0) - \bar{\mathbf{x}}^N(0)\| > \epsilon) = 0 \quad \forall \varepsilon > 0.$$

($\|.\|$ denotes $\|.\|_\infty$; however, since on a $|\mathcal{S}|$-dimensional space, all reasonable norms define the same topology, the actual choice of the norm is not that important.)

The mean-field limit is defined by the following system of ODEs:

$$\dot{v}_i(t) = -\sum_{j:j\neq i} r_{ij}(\mathbf{v}(t)) + \sum_{j:j\neq i} r_{ji}(\mathbf{v}(t))$$

for $i = 1, \ldots, |\mathcal{S}|$ with initial condition $\mathbf{v}(0)$, or, written in integral form:

$$v_i(t) = v_i(0) - \sum_{j:j\neq i} \int_0^t r_{ij}(\mathbf{v}(u))\mathrm{d}u + \sum_{j:j\neq i} \int_0^t r_{ji}(\mathbf{v}(u))\mathrm{d}u \tag{5.3}$$

for $i = 1, \ldots, |\mathcal{S}|$ with initial condition $\mathbf{v}(0)$.

The assumption that r_{ij} are Lipschitz continuous guarantees that the solution of (5.3) uniquely exists.

Theorem (Kurtz)**.** *[35] Under the assumptions and setup given above, we have, for any $T > 0$ and $\epsilon > 0$:*

$$\lim_{N\to\infty} \mathbf{P}\left\{ \sup_{t\in[0,T]} \|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| > \epsilon \right\} = 0$$

Proof. Define the auxiliary process $\mathbf{y}^N(t)$ via

$$y_i^N(t) := v_i(0) - \sum_{j:j\neq i} \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u + \sum_{j:j\neq i} \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u$$

and denote

$$D_i^N(T) = \sup_{t\in[0,T]} |\bar{x}_i^N(t) - y_i^N(t)|.$$

We use the estimate

$$|\bar{x}_i^N(t) - v_i(t)| \leq |\bar{x}_i^N(t) - y_i^N(t)| + |y_i^N(t) - v_i(t)| \leq D_i^N(T) + |y_i^N(t) - v_i(t)|. \tag{5.4}$$

We claim that $D_i^N(T)$ goes to 0 in probability as $N \to \infty$.

$$\begin{aligned} |x_i^N(t) - y_i^N(t)| \leq & |x_i^N(0) - v_i(0)| \\ & + \sum_{j:j\neq i} \left| \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ij}\left(N \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u \right) \right| \\ & + \sum_{j:j\neq i} \left| \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ji}\left(N \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u \right) \right|; \end{aligned}$$

the first term on the right hand side goes to 0 in probability by our assumptions, and for the second and third term, note that

$$\sup_{t\in[0,T]} \left| \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ji}\left(N \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u \right) \right| \leq \frac{1}{N} \sup_{s\in[0,RT]} |P_c(Ns) - Ns|,$$

which goes to 0 almost surely by the functional strong law of large numbers (FSLLN) for the Poisson process ([65], Section 3.2).

Next we take the last term in (5.4):

$$|y_i^N(t) - v_i(t)| \leq \sum_{j:j\neq i} \int_0^t |r_{ij}(\mathbf{v}(u)) - r_{ij}(\bar{\mathbf{x}}^N(u))|\mathrm{d}u + \sum_{j:j\neq i} \int_0^t |r_{ji}(\mathbf{v}(u)) - r_{ji}(\bar{\mathbf{x}}^N(u))|\mathrm{d}u$$
$$\leq 2 \sum_{j:j\neq i} \int_0^t R\|\mathbf{v}(u) - \bar{\mathbf{x}}^N(u))\|\mathrm{d}u = 2R|\mathcal{S}| \int_0^t \|\mathbf{v}(u) - \bar{\mathbf{x}}^N(u))\|\mathrm{d}u.$$

Then

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \leq \max_{i\in\mathcal{S}} D_i^N(T) + 2R|\mathcal{S}| \int_0^t \|\bar{\mathbf{x}}^N(u) - \mathbf{v}(u)\|\mathrm{d}u \tag{5.5}$$

and an application of Grönwall's lemma ([19], page 498) readily yields

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \leq \max_{i\in\mathcal{S}} D_i^N(T) \exp(2R|\mathcal{S}|T),$$

proving Kurtz's theorem.

□

5.2 Population generalized semi-Markov processes

We extend the model defined in the previous section. Again, N is the population size, each individual is inhabiting a state from a finite set $\mathcal{S}$ and each individual in state i performs Markov transitions from transitions to other states j with rate r_{ij}^N. The global state of the system is the total number of individuals in each state, that is, a vector $\mathbf{x}^N \in \{0, 1, \ldots, N\}^{(S)}$ with $x_1^N + \cdots + x_{|\mathcal{S}|}^N = N$.

We also assume that the density-dependent aggregate rates $r_{ij}(\mathbf{x}) = x_i r_{ij}^N(N\mathbf{x})$ are Lipschitz-continuous with common Lipschitz-constant R. R will be assumed to also be an upper bound on r_{ij}.

We also include generally-timed transitions. Partition the states into $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$, where $\mathcal{S}_0$ contains the states where no generally-timed transitions are allowed, while $\mathcal{S}_1$ contains the states where generally-timed transitions are allowed (these will be called *active states*). For each $i \in \mathcal{S}_1$, a distribution function F_i is given; our only assumption on F_i is that it is concentrated on $[0, \infty)$.

Whenever an individual enters a state $i \in \mathcal{S}_1$, it generates a random time according to F_i independent of everything (we say that the clock is initialized). After that time has elapsed, the individual makes a transition to some other state (these types of transitions will be called non-Markovian transitions). Upon leaving state i, the clock is disabled and will be resampled according to F_i if the individual returns to i later.

To summarize: a single individual may have at most one active clock at any given time; however, there is no restriction on the total number of simultaneously active clocks in the entire system.

We have a number of assumptions. We assume that the system is delay-only, that is, if i is an active state then $r_{ij} = 0 \, \forall j$.

We also assume that the non-Markovian transition from an active state i always targets the same state; we will formulate this by saying that the distribution $\mathbf{p}^i$ of the target state is deterministic: p^i_j is equal to 1 for a unique $j \in \mathcal{S}$ and 0 otherwise. We also assume that the target state is non-active, that is, $j \in \mathcal{S}_0$. This is a technical restriction ensuring that non-Markovian transitions do not follow each other directly. However, it is not a modelling restriction as the state space may be reconfigured so the generally-timed transition is followed by a (very fast) Markovian transition sampling from an arbitrary distribution. We set $p^i_j = 0$ for all other pairs (i, j).

We also assume that the initial state of the system is concentrated on $\mathcal{S}_1$; in other words, no generally-timed clocks are active initially. For a (simpler) model where general initial condition is examined, see [5].

We are looking to construct the above model via Poisson-representation. Let $P_{ij}(.)$ be independent Poisson-processes with rate 1 for every $i \neq j \in \mathcal{S}$. Let $\{T^{ij}_k\}_{k=1}^{\infty}$ be mutually independent sequences of identically-distributed random variables distributed according to F_i for each $i \in \mathcal{S}_1$ and $j \in \mathcal{S}_0$. (Taking a separate sample for each j will make formulating the Poisson-representation easier.)

The Poisson-representation of $\bar{\mathbf{x}}^N(t)$ is

$$\begin{aligned}\bar{x}^N_i(t) =& \bar{x}^N_i(0) - \sum_{j:j\neq i} \frac{1}{N} P_{ij}\left(N\int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) + \sum_{j:j\neq i} \frac{1}{N} P_{ji}\left(N\int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\\ &+ \sum_{h\in\mathcal{S}_0}\sum_{j\in\mathcal{S}_1}\int_{z=0}^t p^i_j \mathbf{1}\left(T^{ji}_{P_{hj}\left(N\int_0^z r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)} \leq t-z\right)\frac{1}{N}\mathrm{d}P_{hj}\left(N\int_0^z r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\\ &- \sum_{h\in\mathcal{S}_0}\sum_{j:j\in\mathcal{S}_0}\int_{z=0}^t p^j_i \mathbf{1}\left(T^{ij}_{P_{hi}\left(N\int_0^z r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)} \leq t-z\right)\frac{1}{N}\mathrm{d}P_{hi}\left(N\int_0^z r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\end{aligned} \tag{5.6}$$

for $i \in \mathcal{S}$.

The first of the last two terms in the formula (5.6) should be understood as follows. If i is active, the term is 0 by our assumptions on p^i_j. If i is inactive, consider an active state j with $p^i_j = 1$ and an inactive state h. If a Markov transition from h to j occurs at time z, a non-Markovian clock distributed according to F_j starts. The clock samples from the list $\{T^{ji}_k\}_{k=1}^{\infty}$; to ensure that a new k is used for each clock, k is set to $P_{hj}\left(\int_0^z r_{hj}((\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)$ (which increases with each arrival of P_{hj}). When the indicator variable is 1, the clock has already set off before time t and needs to be counted among the actual transitions; when the indicator variable is 0, the clock has not yet set off by time t, so the corresponding $j \to i$ transition has not yet occurred, and the contribution of the integral is 0.

Conversely, the last term of (5.6) is nonzero only if i is active; consider an inactive state j with $p^j_i = 1$ and an inactive state h. If a Markov transition from h to i occurs at time z, a non-Markovian clock distributed according to F_i starts. The clock samples from the list $\{T^{ij}_k\}_{k=1}^{\infty}$ with $k = P_{ij}\left(\int_0^z r_{ij}((\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)$ (which increases with each arrival of P_{ij}). When the indicator variable is 1, the clock has already set off before time t and needs to be counted among the actual transitions (which decrease the number of individuals in state i, hence the negative sign); when the indicator variable is 0, the clock has not yet set off by time t, so the corresponding $i \to j$ transition has not yet

occurred.

The mean-field limit is defined by the following delayed differential equations (written in integral form):

$$\begin{aligned} v_i(t) =& v_i(0) - \sum_{j:j\neq i} \int_0^t r_{ij}(\mathbf{v}(u))\mathrm{d}u + \sum_{j:j\neq i} \int_0^t r_{ji}(\mathbf{v}(u))\mathrm{d}u \\ &+ \sum_{h\in\mathcal{S}_0} \sum_{j\in\mathcal{S}_1} \int_{u=0}^t p_j^i F_j(t-u) r_{hj}(\mathbf{v}(u))\mathrm{d}u \\ &- \sum_{h\in\mathcal{S}_0} \sum_{j:j\in\mathcal{S}_0} \int_{u=0}^t p_i^j F_i(t-u) r_{hi}(\mathbf{v}(u))\mathrm{d}u \end{aligned} \tag{5.7}$$

for $i \in \mathcal{S}$.

Lipschitz continuity of r_{ij} guarantees that the solution of (5.7) uniquely exists.

Again, we assume convergence of the initial condition:

$$\lim_{N\to\infty} \mathbf{P}(\|\mathbf{v}(0) - \bar{\mathbf{x}}^N(0)\| > \epsilon) = 0 \quad \forall \varepsilon > 0.$$

Theorem 5.1. *Under the assumptions and setup given above, we have, for any $T > 0$ and $\epsilon > 0$:*

$$\lim_{N\to\infty} \mathbf{P}\left\{ \sup_{t\in[0,T]} \|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| > \epsilon \right\} = 0$$

5.3 Example: peer-to-peer software update

In this section, we derive the system of DDEs for a simple example model of a peer-to-peer software update process.

We consider two general types of nodes in this model which we term *old* and *updated*. Old nodes are those running an old software version and new nodes are those which have been updated to a new version. Nodes alternate between being *on* and *off*; when an old node turns on, it searches for an update in peer-to-peer fashion, with the probability of successfully finding an update being proportional to the number of nodes already updated. If it does not find an update, it gives up after a timeout. After that, it stays on for some time and then eventually turns off. New nodes do not search for updates, just alternate between on and off. We assume that the off time of a node is random and Pareto-distributed. So, nodes have five possible local states: updated nodes can be on and off, which we denote by a and b, respectively. Old nodes can be on (c), off (e) or in a state representing an old node which is on but has given up seeking updates (d). In the notation of Section 5.2, the set of local states is thus $\mathcal{S} := \{a, b, c, d, e\}$ with $\mathcal{S}_0 = \{a, c, d\}$, $\mathcal{S}_1 = \{d, e\}$. The local behaviour of a node is depicted in Figure 5.1.

In this example, we consider all transitions to be Markovian except for the transitions bringing nodes from their off state into their on state, which have cumulative distribution function $F(s)$.

The DDEs corresponding to this model are:

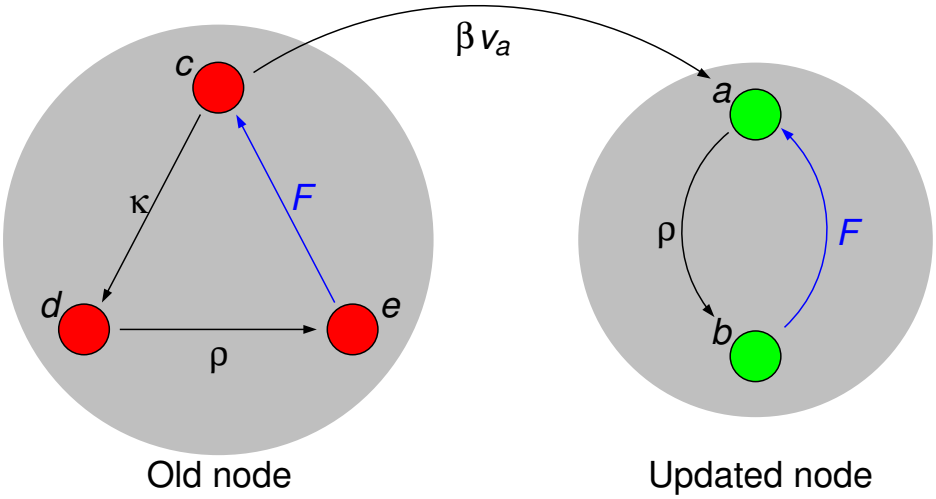

Figure 5.1: Representation of the behaviour of a single node in the delay-only software update model.

$$
\begin{aligned}
\dot{v}_a(t) &= \beta v_a(t) v_c(t) - \rho v_a(t) + \rho \int_0^t v_a(t-s) \mathrm{d}F(s) \\
\dot{v}_b(t) &= \rho v_a(t) - \rho \int_0^t v_a(t-s) \mathrm{d}F(s) \\
\dot{v}_c(t) &= -\kappa v_c(t) - \beta v_a(t) v_c(t) + \rho \int_0^t v_d(t-s) \mathrm{d}F(s) \\
\dot{v}_d(t) &= \kappa v_c(t) - \rho v_d(t) \\
\dot{v}_e(t) &= \rho v_d(t) - \rho \int_0^t v_d(t-s) \mathrm{d}F(s)
\end{aligned}
$$

5.4 Proof of the mean-field convergence

Proof of Theorem 5.1. Define the auxiliary process $\mathbf{y}^N(t)$ via

$$
\begin{aligned}
y_i^N(t) :=& v_i(0) - \sum_{j:j\neq i} \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u)) \mathrm{d}u + \sum_{j:j\neq i} \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u)) \mathrm{d}u \\
&+ \sum_{h\in\mathcal{S}_0} \sum_{j\in\mathcal{S}_1} \int_{u=0}^t p_j^i F_j(t-u) r_{hj}(\bar{\mathbf{x}}^N(u)) \mathrm{d}u \\
&- \sum_{h\in\mathcal{S}_0} \sum_{j:j\in\mathcal{S}_0} \int_{u=0}^t p_i^j F_i(t-u) r_{hi}(\bar{\mathbf{x}}^N(u)) \mathrm{d}u \qquad (5.8)
\end{aligned}
$$

for $i \in \mathcal{S}$.

Then

$$|\bar{x}_i^N(t) - v_i(t)| \leq |\bar{x}_i^N(t) - y_i^N(t)| + |y_i^N(t) - v_i(t)|$$

for any $i \in \mathcal{S}$.

Denote

$$D_i^N(T) = \sup_{t\in[0,T]} |\bar{x}_i^N(t) - y_i^N(t)|$$

We estimate $\|\mathbf{y}^N(t) - \mathbf{v}(t)\|$ by

$$\begin{aligned}
|y_i^N(t) - v_i(t)| &\leq \sum_{j:j\neq i} \int_0^t |r_{ij}(\bar{\mathbf{x}}^N(u)) - r_{ij}(\mathbf{v}(u))|\mathrm{d}u + \sum_{j:j\neq i} \int_0^t |r_{ji}(\bar{\mathbf{x}}^N(u)) - r_{ji}(\mathbf{v}(u))|\mathrm{d}u \\
&+ \sum_{h\in\mathcal{S}_0} \sum_{j\in\mathcal{S}_1} \int_{u=0}^t p_j^i F_j(t-u)|r_{hj}(\bar{\mathbf{x}}^N(u)) - r_{hj}(\mathbf{v}(u))|\mathrm{d}u \\
&+ \sum_{h\in\mathcal{S}_0} \sum_{j:j\in\mathcal{S}_0} \int_{u=0}^t p_i^j F_i(t-u)|r_{hi}(\bar{\mathbf{x}}^N(u)) - r_{hi}(\mathbf{v}(u))|\mathrm{d}u \\
&\leq ZR \int_0^t \|\mathbf{x}^N(u) - \mathbf{v}(u)\|\mathrm{d}u
\end{aligned}$$

where

$$Z := |\mathcal{S}_0| + |\mathcal{S}_0| + |\mathcal{S}_0| \cdot |\mathcal{S}_1| + |\mathcal{S}_0|^2$$

and $\|.\|$ is the maximum norm on $\mathbb{R}^{\mathcal{S}}$. We aim to show that $D_i^N(T) \to 0$ in probability as $N \to \infty$ for each $i \in \mathcal{S}$; once we have that, we have

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \leq \max_{i\in\mathcal{S}} D_i^N(T) + ZR \int_0^t \|\bar{\mathbf{x}}^N(u) - \mathbf{v}(u)\|\mathrm{d}u \tag{5.9}$$

and an application of Grönwall's lemma ([19], page 498) readily yields

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \leq \max_{i\in\mathcal{S}} D_i^N(T) \exp(ZRT),$$

proving the theorem.

It now remains to show that for each $i \in \mathcal{S}$, $D_i^N(T) \to 0$ in probability as $N \to \infty$.

To see this note that:

$$
\begin{aligned}
D_i^N(T) \leq |v_i(0) - \bar{x}_i^N(0)| &+ \sum_{j:j\neq i} \sup_{t\in[0,T]} \left| \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ij}\left(N\int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \\
&+ \sum_{j:j\neq i} \sup_{t\in[0,T]} \left| \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ji}\left(N\int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| + \\
&+ \sum_{h\in\mathcal{S}_0}\sum_{j\in\mathcal{S}_1} p_j^i \left| \int_{u=0}^t F_j(t-u) r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \right. \\
&\left. - \int_{u=0}^t \mathbf{1}\left(T^{ji}_{P_{hj}\left(N\int_0^z r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)} \leq t-z\right) \frac{1}{N}\mathrm{d}P_{hj}\left(N\int_0^z r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \\
&+ \sum_{h\in\mathcal{S}_0}\sum_{j\in\mathcal{S}_0} p_i^j \left| \int_{u=0}^t F_i(t-u) r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \right. \\
&\left. - \int_{u=0}^t \mathbf{1}\left(T^{ij}_{P_{hi}\left(N\int_0^z r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)} \leq t-z\right) \frac{1}{N}\mathrm{d}P_{hi}\left(N\int_0^z r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \qquad (5.10)
\end{aligned}
$$

for $i \in \mathcal{S}$

The first term in (5.10) converges to 0 in probability per our assumptions. The second and third terms in (5.10) are essentially the same; they are handled in the following lemma.

Lemma 5.2. *For any $i, j \in \mathcal{S}$*

$$
\sup_{t\in[0,T]} \left| \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ji}\left(N\int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \to 0
$$

almost surely as $N \to \infty$.

Remark. The lemma states almost sure convergence; this makes sense because the coupling provided by the Poisson representation puts the PGSMP for different values of N in the same probability space. That said, convergence in probability is enough for our purposes.

Proof of Lemma 5.2. By the Lipschitz-condition, $0 \leq \int_0^t r_{ij}^N(\bar{\mathbf{x}}^N(u))\mathrm{d}u \leq RT$ for any $t \in [0,T]$ and thus

$$
\sup_{t\in[0,T]} \left| \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u - \frac{1}{N} P_{ji}\left(N\int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \leq \frac{1}{N} \sup_{s\in[0,RT]} |P_c(Ns) - Ns|,
$$

which goes to 0 almost surely by the functional strong law of large numbers (FSLLN) for the Poisson process ([65], Section 3.2). □

What remains is to prove that the last two terms in (5.10) go to 0 in probability.

Before proceeding, we change the notation a bit. h, i and j will be fixed from now on until the end of this section. We will thus drop them from notation and use $F = F_j, r = r_{hj}, T_k = T_k^{ji}$ and $P = P_{hj}$ in (5.10) for the first of the last two terms in (5.10) (and, correspondingly, $F = F_i, r = r_{hi}, T_k = T_k^{ij}$ and $P = P_{hi}$ for the last term in (5.10)).

We will also use the shorthand

$$J^N(t) = P\left(N\int_0^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)$$

for the (measure generated by the) Poisson process.

Using the above notation, either of the last two terms in (5.10) (without the finite summations $\sum_{h\in\mathcal{S}_0}\sum_{j\in\mathcal{S}_1}\cdot$ and the bounded constant p_i^j) simplifies to

$$\left|\int_0^t F(t-u)r(\bar{\mathbf{x}}^N(u))\,\mathrm{d}u - \int_0^t \mathbf{1}\left(T_{J^N(u)}\le t-u\right)\frac{1}{N}\,\mathrm{d}J^N(u)\right| \tag{5.11}$$

We note that

$$\begin{aligned}
&\left|\int_0^t \mathbf{1}\left(T_{J^N(u)}\le t-u\right)\frac{1}{N}\,\mathrm{d}J^N(u) - \int_0^t F(t-u)r(\bar{\mathbf{x}}^N(u))\,\mathrm{d}u\right| \le \\
&\left|\int_0^t F(t-u)\frac{1}{N}\,\mathrm{d}J^N(u) - \int_0^t F(t-u)r(\bar{\mathbf{x}}^N(u))\,\mathrm{d}u\right| \\
+&\left|\int_0^t \mathbf{1}\left(T_{J^N(u)}\le t-u\right)\frac{1}{N}\,\mathrm{d}J^N(u) - \int_0^t F(t-u)\frac{1}{N}\,\mathrm{d}J^N(u)\right|.
\end{aligned} \tag{5.12}$$

The first term on the right hand side will be dealt with in Lemma 5.3 and the second in Lemma 5.4.

We have some more preparations first. From Lemma 5.2 we already have that

$$\sup_{t\in[0,T]}\left|\frac{1}{N}P\left(N\int_0^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) - \int_0^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \to 0$$

almost surely as $N\to\infty$. As a direct consequence of this, we also have

$$\sup_{s,t\in[0,T]}\left|\frac{1}{N}P\left(N\int_s^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) - \int_s^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \to 0$$

almost surely since

$$\sup_{s,t\in[0,T]}\left|\int_s^t \cdot\right| = \sup_{s,t\in[0,T]}\left|\int_0^t \cdot - \int_0^s \cdot\right| \le 2\sup_{t\in[0,T]}\left|\int_0^t \cdot\right|$$

Also as a preparation, we have

$$\sup_{t\in[0,T]}\int_0^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u \le \sup_{t\in[0,T]}\int_0^t R\|\bar{\mathbf{x}}^N(u)\|\mathrm{d}u \le \sup_{t\in[0,T]} Rt = RT$$

independent of N, again using $\|\bar{\mathbf{x}}^N\| \le N$ and $r(\bar{\mathbf{x}}) \le R\|\bar{\mathbf{x}}\|$. Lemma 5.2 then also implies $\frac{1}{N}\int_0^t \mathrm{d}J^N(u) \le RT + \varepsilon_N$, where $\varepsilon_N \to 0$ almost surely as $N\to\infty$.

Lemma 5.3.

$$\sup_{t\in[0,T]}\left|\int_0^t F(t-u)\frac{1}{N}\mathrm{d}J^N(u)-\int_0^t F(t-u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right|\to 0$$

almost surely as $N\to\infty$.

Proof. Let $\varepsilon>0$ be fixed. Write

$$F(t-u)=g_{t,\varepsilon}(u)+h_{t,\varepsilon}(u),$$

where $g=g_{t,\varepsilon}$ is a piecewise constant function with $0\le g(u)\le 1$ and $\|h\|_\infty\le\varepsilon$. Their exact definition is as follows. Take the $\varepsilon, 2\varepsilon,\dots$ quantiles of $F(t-u)$ ($F(t-u)$ as a function of u is nonincreasing between 0 and 1); that is, let $u_k=\inf\{u:\ F(t-u)\le k\varepsilon\}$. Some of these u_k's may be equal if F has discontinuities. The number of distinct quantiles is certainly no more than $\lceil\varepsilon^{-1}\rceil$, independent of N and t.

Let g be the piecewise constant function

$$g(u)=F(t-u_k)\quad\text{if}\quad u\in(u_{k-1},u_k],$$

so $g(u)\le F(t-u)$. The choice of u_k's guarantees that $h(u)=F(t-u)-g(u)\le\varepsilon$.

Then we can write

$$\begin{aligned}&\left|\int_0^t F(t-u)\frac{1}{N}\mathrm{d}J^N(u)-\int_0^t F(t-u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right|\le\\&\left|\int_0^t g(u)\frac{1}{N}\mathrm{d}J^N(u)-\int_0^t g(u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right|+\\&\left|\int_0^t h(u)\frac{1}{N}\mathrm{d}J^N(u)-\int_0^t h(u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right|\end{aligned}$$

Since g is piecewise constant,

$$\begin{aligned}
&\left|\int_0^t g(u)\frac{1}{N}\mathrm{d}J^N(u) - \int_0^t g(u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| = \\
&\frac{1}{N}\left|\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil} g(u_k)\left(J^N(u_k) - J^N(u_{k-1}) - \int_{u_{k-1}}^{u_k} Nr(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \leq \\
&\frac{1}{N}\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil}\left|g(u_k)\left(J^N(u_k) - J^N(u_{k-1}) - \int_{u_{k-1}}^{u_k} Nr(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\right| \leq \\
&\frac{1}{N}\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil}\left|J^N(u_k) - J^N(u_{k-1}) - \int_{u_{k-1}}^{u_k} Nr(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \leq \\
&\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil}\sup_{s,t\in[0,T]}\left|\frac{1}{N}P\left(N\int_s^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) - \int_s^t r(\bar{\mathbf{x}}^N(u))\right| = \\
&\lceil\varepsilon^{-1}\rceil\cdot\sup_{s,t\in[0,T]}\left|\frac{1}{N}P\left(N\int_s^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) - \int_s^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \to 0
\end{aligned}$$

almost surely as $N\to\infty$ since ε is independent of N.

Since $\|h\|_\infty \leq \varepsilon$, we have

$$\begin{aligned}
&\left|\int_0^t h(u)\frac{1}{N}\mathrm{d}J^N(u) - \int_0^t h(u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \leq \\
&\left|\int_0^t h(u)\frac{1}{N}\mathrm{d}J^N(u)\right| + \left|\int_0^t h(u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \leq \\
&\frac{\varepsilon}{N}\left|\int_0^t \mathrm{d}J^N(u)\right| + \varepsilon\left|\int_0^t r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \leq \varepsilon(2RT+\varepsilon_N),
\end{aligned}$$

independent of t (with $\varepsilon_N\to 0$ almost surely as $N\to\infty$).

Letting $\varepsilon\to 0$ proves

$$\sup_{t\in[0,T]}\left|\int_0^t F(t-u)\frac{1}{N}\mathrm{d}J^N(u) - \int_0^t F(t-u)r(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right| \to 0$$

almost surely as $N\to\infty$. □

Lemma 5.4.

$$\sup_{t\in[0,T]}\frac{1}{N}\left|\int_0^t \mathbf{1}\left(T_{J^N(u)} \leq t-u\right)\mathrm{d}J^N(u) - \int_0^t F(t-u)\mathrm{d}J^N(u)\right| \to 0$$

almost surely as $N\to\infty$.

Proof. Let $\varepsilon > 0$ be fixed. Also fix t for now. We want to prove

$$\mathbf{P}\left(\frac{1}{N}\left|\int_0^t \mathbf{1}\left(T_{J^N(u)} \le t-u\right) \mathrm{d}J^N(u) - \int_0^t F(t-u)\mathrm{d}J^N(u)\right| > \varepsilon\right)$$

is exponentially small in N via Azuma's inequality [11, 2]. Once we have that, we can apply Borel–Cantelli lemma (see e.g. [16] Chapter 2.3) to conclude that for any fixed ϵ, the above event happens only finitely many times, which is equivalent to almost sure convergence to 0. To apply Azuma, we need to write the above integral as a martingale with bounded increments. The measure $\mathrm{d}J^N(u)$ is concentrated on points u where P has an arrival at $N\int_0^u r(\bar{\mathbf{x}}^N(z))\mathrm{d}z$. Let we denote these random points by $u_1, u_2, \ldots$ (in fact, the whole sequence depends on the value of N; we will keep N fixed as long as we use this sequence, and not include N in the notation). The integral only has contributions from these points; it is natural to write (using a slightly different notation)

$$\begin{aligned} S_l &:= \left(\mathbf{1}\left(T_1 \le t-u_1\right) - F(t-u_1)\right) + \cdots + \left(\mathbf{1}\left(T_l \le t-u_l\right) - F(t-u_l)\right) \\ M_N &:= P\left(N\int_0^t r(\bar{\mathbf{x}}^N(z))\mathrm{d}z\right) \end{aligned}$$

so that

$$\int_0^t \mathbf{1}\left(T_{J^N(u)} \le t-u\right) \mathrm{d}J^N(u) - \int_0^t F(t-u)\mathrm{d}J^N(u) = S_{M_N}.$$

We first resolve the difficulty that M_N is in fact random.

$$\begin{aligned} &\mathbf{P}\left(\frac{1}{N}\left|\int_0^t \mathbf{1}\left(T^e_{J^N(u)\le t-u}\right) \mathrm{d}J^N(u) - \int_0^t F(t-u)\mathrm{d}J^N(u)\right| > \varepsilon\right) = \\ &\mathbf{P}\left(\left|\frac{S_{M_N}}{N}\right| > \varepsilon\right) = \sum_{l=0}^{\infty} \mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon,\, M_N = l\right) \le \\ &\sum_{l=0}^{2RTN} \mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right) + \sum_{2RTN+1}^{\infty} \mathbf{P}(M_N = l) = \\ &\sum_{l=0}^{2RTN} \mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right) + \mathbf{P}(M_N > 2RTN). \end{aligned}$$

The sum was cut at $2RTN$ because M_N is stochastically dominated by a Poisson distribution with parameter RTN, so $\mathbf{P}(M_N > 2RTN)$ is exponentially small due to Cramér's large deviation theorem (see e.g. Theorem II.4.1 in [18]):

$$\mathbf{P}(M_N > 2RTN) \le e^{-RTN(2\ln 2 - 1)}.$$

(The Cramér rate function of the Poisson-distribution with parameter λ is $I(x) = x\ln(x/\lambda) - x + \lambda$.)

To apply Azuma to each of the terms $\mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right)$, we also need to check that S_l is indeed a martingale with bounded increments. To set it up properly as a martingale, note that $\{u_l\}$ is an increasing sequence of stopping times, so the filtration $\{\mathcal{F}_l\}$ is well-defined; $\mathcal{F}_l$ contains all the

information known up to time u_l, including the values of all of the non-Markovian clocks that started by the time u_l.

S_l has bounded increments, since

$$|\mathbf{1}(\{T_l \leq t - u_l) - F(t - u_l)| \leq 1.$$

The last step to apply Azuma is that we need to check that S_l is a martingale with respect to $\mathcal{F}_l$. It is clearly adapted, and

$$\mathbf{E}(\mathbf{1}\left(T_{l+1} \leq t - u_{l+1}\right)|\mathcal{F}_l) = \mathbf{E}(\mathbf{E}(\mathbf{1}\left(T_{l+1} \leq t - u_{l+1}\right)|\mathcal{F}_l, u_{l+1})|\mathcal{F}_l) =$$
$$\mathbf{E}(\mathbf{P}(T_{l+1} \leq t - u_{l+1}|\mathcal{F}_l, u_{l+1})|\mathcal{F}_l) = \mathbf{E}(F(t - u_{l+1})|\mathcal{F}_l)$$

shows that it is a martingale as well. (In the last step, we used the fact that u_{l+1} is measurable with respect to $\sigma\{\mathcal{F}_l \cup \{u_{l+1}\}\}$ while T_{l+1} is independent from it.)

We have everything assembled to apply Azuma's inequality (Theorem 5.2 in [11]):

$$\mathbf{P}\left(|S_l - \mathbf{E}\left(S_l\right)| > \lambda\right) \leq 2e^{-\frac{\lambda^2}{2l}}$$

and thus (considering $\mathbf{E}\left(S_l\right) = 0$)

$$\sum_{l=0}^{2RTN} \mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right) \leq \sum_{l=0}^{2RTN} 2e^{-\frac{\varepsilon^2 N^2}{2l}} \leq$$
$$\leq 2RTN \cdot 2e^{-\frac{\varepsilon^2 N^2}{4RTN}} = 4RTNe^{-\frac{\varepsilon^2 N}{4RT}}.$$

In the last inequality, we estimated each term in the sum by the largest one, which is for $l = 2RN$.

The estimate obtained is

$$\mathbf{P}\left(\frac{1}{N}\left|\int_0^t \mathbf{1}\left(T_{J^N(u) \leq t-u}\right) \mathrm{d}J^N(u) - \int_0^t F(t-u)\mathrm{d}J^N(u)\right| > \varepsilon\right) \leq$$
$$4RTNe^{-\frac{\varepsilon^2 N}{4RT}} + e^{-RTN(2\ln 2 - 1)}.$$

Remember that t was fixed; we need to upgrade this estimate into an estimate that is valid for $\sup_{t\in[0,T]}(.)$ before applying Borel–Cantelli lemma. We do this by partitioning the interval $[0, T]$ into N subintervals uniformly, and then controlling what happens *at* the partition points and *between* the partition points separately. For the former, we apply the previous estimate. Let

$$t_m := \frac{mT}{N},\ m = 0, 1, \ldots N,$$

then

$$\mathbf{P}\left(\max_{0\le m\le N}\frac{1}{N}\left|\int_0^{t_m}\mathbf{1}\left(T_{J^N(u)\le t-u}\right)\mathrm{d}J^N(u)-\int_0^{t_m}F(t-u)\mathrm{d}J^N(u)\right|>\varepsilon\right)\le$$
$$(N+1)\left(4RTNe^{-\frac{\varepsilon^2 N}{4RT}}+e^{-RTN(2\ln 2-1)}\right),$$

which is still summable.

Now we turn our attention to the intervals $[t_m, t_{m+1}]$. Since

$$\int_0^t\mathbf{1}\left(T_{J^N(u)\le t-u}\right)\mathrm{d}J^N(u)\quad\text{and}\quad\int_0^t F(t-u)\mathrm{d}J^N(u)$$

are both increasing in t, we only have to check that neither of them increases by more than εN over an interval $[t_m, t_{m+1}]$.

Let m be fixed. We handle the two integrals separately. First, for

$$\int_0^t F(t-u)\mathrm{d}J^N(u),$$

we have

$$\int_0^{t_{m+1}}F(t_{m+1}-u)\mathrm{d}J^N(u)-\int_0^{t_m}F(t_m-u)\mathrm{d}J^N(u)=$$
$$\int_0^{t_m}F(t_{m+1}-u)-F(t_m-u)\mathrm{d}J^N(u)+\int_{t_m}^{t_{m+1}}F(t_{m+1}-u)\mathrm{d}J^N(u)\le$$
$$\int_0^{t_m}F(t_{m+1}-u)-F(t_m-u)\mathrm{d}J^N(u)+\int_{t_m}^{t_{m+1}}1\mathrm{d}J^N(u).$$

The second term is equal to $J^N(t_{m+1})-J^N(t_m)$. By the Lipschitz-condition, this is stochastically dominated by $Z\sim\text{Poisson}(RT)$ since the length of the interval is T/N, and thus

$$\mathbf{P}\left(\frac{1}{N}\int_{t_m}^{t_{m+1}}F(t_{m+1}-u)\mathrm{d}J^N(u)>\varepsilon\right)\le\mathbf{P}\left(\frac{Z}{N}>\varepsilon\right)=\mathbf{P}\left(\frac{Z}{\varepsilon}>N\right).$$

Note that the right hand side is summable in N, its sum being equal to the expectation of $\left\lceil\frac{Z}{\varepsilon}\right\rceil$.

To estimate the other term, note that

$$u\in[t_{l-1},t_l]\Longrightarrow F(t_{m+1}-u)-F(t_m-u)\le F(t_{m+1}-t_{l-1})-F(t_m-t_l)=$$
$$F(t_{m+1}-t_{l-1})-F(t_{m+1}-t_l)+F(t_{m+1}-t_l)-F(t_m-t_l),$$

which gives

$$\int_0^{t_m} F(t_{m+1}-u)-F(t_m-u)\mathrm{d}J^N(u)=$$
$$\sum_{l=1}^{m}\int_{t_{l-1}}^{t_l} F(t_{m+1}-u)-F(t_m-u)\mathrm{d}J^N(u)\leq$$
$$\sum_{l=1}^{m}\int_{t_{l-1}}^{t_l} F(t_{m+1}-t_{l-1})-F(t_m-t_l)\mathrm{d}J^N(u)=$$
$$\sum_{l=1}^{m}(F(t_{m+1}-t_{l-1})-F(t_m-t_l))(J^N(t_l)-J^N(t_{l-1})).$$

We use two things here: the fact that $(J^N(t_l)-J^N(t_{l-1}))$ is stochastically dominated by Poisson(RT) and the fact that the sum

$$\sum_{l=1}^{m}(F(t_{m+1}-t_{l-1})-F(t_m-t_l))=\sum_{l=1}^{m}F(t_{m-l+2})-F(t_{m-l})=$$
$$F(t_{m+1})+F(t_m)-F(1)-F(0)\leq 2$$

is telescopic. This means that the whole sum can be stochastically dominated by Poisson$(2RT)$ (note that the number of clocks starting at each interval is not independent, but because of the Lipschitz-condition, we may still use independent Poisson variables when stochastically dominating the sum). Using the notation $Z\sim$ Poisson(RT) again, we get that

$$\sum_{N=1}^{\infty}\mathbf{P}\left(\frac{2Z}{N}>\varepsilon\right)=\sum_{N=1}^{\infty}\mathbf{P}\left(\frac{2Z}{\varepsilon}>N\right)\leq\frac{2RT}{\varepsilon}+1.$$

(In fact, $\mathbf{P}\left(\frac{2Z}{\varepsilon}>N\right)$ goes to 0 superexponentially in N.)

The last term to estimate is the increment of

$$\int_0^t \mathbf{1}\left(T_{J^N(u)}\leq t-u\right)\mathrm{d}J^N(u).$$

between t_m and t_{m+1}, e.g. the number of clocks expiring between t_m and t_{m+1}.

Partition the clocks according to which interval $[t_{l-1},t_l]$ they started in. The number of clocks starting in $[t_{l-1},t_l]$ is stochastically dominated by $Z\sim$ Poisson(RT) by the Lipschitz-condition, and for each such clock, the probability that it goes off in $[t_m,t_{m+1}]$ is less than or equal to $F(t_{m+1})-F(t_{l-1})$. This implies that the number of the clocks starting in $[t_{l-1},t_l]$ and going off in $[t_m,t_{m+1}]$ is stochastically dominated by $W_{m,l}\sim$ Poisson$(RT(F(t_{m+1})-F(t_{l-1})))$. The total number of clocks going off in $[t_m,t_{m+1}]$ is stochastically dominated by Poisson$(RT\sum_{l=1}^{m}(F(t_{m+1})-F(t_{l-1})))$, where the familiar telescopic sum appears in the parameter. (Once again, the Lipschitz-condition was used implicitly.) So the total number of clocks going off in $[t_m,t_{m+1}]$ is stochastically dominated by Poisson$(2RT)$, which means we arrive at the also familiar $\mathbf{P}\left(\frac{2Z}{\varepsilon}>N\right)$ value, which we already examined and proved

to be summable in N.

Putting it together, we get that

$$\mathbf{P}\left(\sup_{t\in[0,T]}\frac{1}{N}\left|\int_0^t \mathbf{1}\left(T_{J^N(u)}\leq t-u\right)\mathrm{d}J^N(u)-\int_0^t F(t-u)\mathrm{d}J^N(u)\right|>\varepsilon\right)\leq C_{N,\varepsilon}$$

where

$$\sum_{N=1}^{\infty} C_{N,\varepsilon}<\infty,$$

so the Borel–Cantelli lemma gives almost sure convergence as $N\to\infty$.

□

With Lemmas 5.2-5.4 finished, the proof of Theorem 5.1 is complete.

□

5.5 Outlook

We have presented a model of PGSMPs where individuals can enable both Markovian transitions that compete with each other (as opposed to delay-only PGSMPs), and we have given a rigorous proof for the mean-field convergence.

That said, there are a number of questions left open. A natural generalization is the *race case*, when non-Markovian transitions compete with Markovian transitions; when a Markovian transition occurs in an active state, it disables the non-Markovian clock. The race case is subject to further research; we expect that the mean-field limit convergence holds for such systems as well, and the limit is again described by the solution of appropriate DDE's.

Taking one step further, another generalization would be the case when non-Markovian clocks are not disabled upon Markovian transitions. Such systems would be widely applicable, yet, to the best of our knowledge, no rigorous proof has been given for mean-field convergence for any non-trivial model of this kind.

A question in a different direction is second order approximation, that is, fluctuations around the mean-field limit. For the original Markov population model, the fluctuations are Gaussian, with the covariances satisfying a system of ordinary differential equations [35] [19]. For PGSMPs, it is reasonable to expect Gaussian fluctuations, with the covariance satisfying a system of delayed differential equations instead, but no results are available yet. The technology presented in this chapter seems to be geared more towards first-order approximation.

Bibliography

[1] D. Amit, G. Parisi, and L. Peliti. Asymptotic behavior of the 'true' self-avoiding walk. *Phys. Rev. B*, 27:1635–1645, 1983.

[2] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

[3] Roberto Barbuti, Giulio Caravagna, Andrea Maggiolo-Schettini, and Paolo Milazzo. Delay stochastic simulation of biological systems: a purely delayed approach. *Transactions on Computational Systems Biology XIII*, 6575:61–84, January 2011.

[4] S. G. Bobkov and M. Ledoux. From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. and Funct. Anal.*, 10:1028–1052, 2000.

[5] Luca Bortolussi and Jane Hillston. Fluid approximation of ctmc with deterministic delays. *Int. Conf. on Quantitative Evaluation of Systems*, pages 53–62, 2012.

[6] L. Breuer and D. Baum. An Introduction to Queueing Theory and Matrix-Analytic Methods Springer, 2005

[7] Peter Buchholz and Miklós Telek. On minimal representation of rational arrival processes. *Annals of Operations Research*, 202(1):35–58, 2013.

[8] BuTools program package for phase type (PH) / matrix exponential (ME) distributions and Markov arrival processes (MAPs) / rational arrival processes (RAPs). `http://webspn.hit.bme.hu/ telek/tools/butools/butools.html`

[9] Giulio Caravagna and Jane Hillston. Bio-PEPAd: A non-Markovian extension of Bio-PEPA. *Theoretical Computer Science*, 419:26–49, 2012.

[10] Hoon Choi, Vidyadhar Kulkarni, Kishor Trivedi, and Marco Ajmone Marsan. Transient analysis of deterministic and stochastic petri nets. In *Application and Theory of Petri Nets*, volume 691, pages 166–185. Springer Berlin / Heidelberg, 1993.

[11] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics*, 3(1):79–127, 2006.

[12] C. Commault and S. Mocanu. Phase-type distributions and representations: some open problems for system theory. *Int. J. Control*, 76(6):566–580, 2003.

[13] David R. Cox. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(03):433–441, July 1955.

[14] C. Cuny and M. Peligrad. Central limit theorem started at a point for additive functionals of reversible Markov chains. *Journal of Theoretical Probability*, 25:171–188, 2012.

[15] R. L. Dobrushin, Yu. M. Suhov, and J. Fritz. A. N. Kolmogorov – the founder of the theory of reversible Markov processes. *Russian Math. Surveys*, 43:6:157–182, 1988.

[16] R. Durrett. *Probability: Theory and Examples*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2010.

[17] R. T. Durrett and L. C. G. Rogers. Asymptotic behavior of Brownian polymers. *Probab. Theory Rel. Fields*, 92:337–349, 1992.

[18] R. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Classics in Mathematics. Springer, 2005.

[19] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 2005.

[20] T. Funaki. *Stochastic Interface Models. Lectures on Probability Theory and Statistics, Lecture Notes in Mathematics*, volume 1869, 103–274, 2005.

[21] J. Glimm and A. Jaffe. Quantum physics, a functional integral point of view. Springer, 1981

[22] M. I. Gordin and B. A. Lifshits. Central limit theorem for stationary Markov processes. *Dokladi Akademii Nauk SSSR*, 239:766–767, 1978.

[23] R. Hayden, I. Horváth, and M. Telek. Mean field for performance models with generally distributed-timed transitions. In W. Sanders G. Norman, editor, *11th International Conference on Quantitative Evaluation of Systems, QEST*, volume 8657 of *Lecture Notes in Computer Science*, pages 90–105. Springer, 2014.

[24] R. A. Hayden. Mean field for performance models with deterministically-timed transitions. In *9th International Conference on Quantitative Evaluation of Systems, QEST*, pages 63–73, Sept 2012.

[25] G. Horváth, P. Reinecke, M. Telek and K. Wolter Heuristic representation optimization for efficient generation of PH-distributed random variates *Annals of Operations Research*, 2014, DOI: 10.1007/s10479-014-1736-x

[26] I. Horváth and M. Telek. A constructive proof of the phase-type characterization theorem. *Stochastic Models*, to appear, 2015.

[27] I. Horváth and M. Telek. A heuristic procedure for compact Markov representation of phase-type distributions. *Proceedings of the 8th International Conference on Performance Evaluation Methodologies and Tools (2014)*, pages 195–202, 2014.

[28] I. Horváth, B. Tóth, and B. Vető. Diffusive limits for "true" (or myopic) self-avoiding random walks and self-repellent Brownian polymers in dimensions 3 and higher. *Probability Theory and Related Fields*, 153(3-4), 2012.

[29] I. Horváth, B. Tóth, and B. Vető. Relaxed sector condition. *Bulletin of the Institute of Mathematics, Academia Sinica (N.S.)*, 7:463–476, 2012.

[30] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[31] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes with applications to simple exclusion. *Communications in Mathematical Physics*, 106:1–19, 1986.

[32] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov Processes – Time Symmetry and Martingale Approximation*, volume 345 of *Grundlehren der mathematischen Wissenschaften*, Springer, 2012.

[33] T. Komorowski and S. Olla. On the sector condition and homogenization of diffusions with a gaussian drift. *Journal of Functional Analysis*, 197:179–211, 2003.

[34] G. Kozma and B. Tóth. Central limit theorem for random walks in divergence-free random drift field: H_{-1} suffices. preprint, 2014, `http://arxiv.org/abs/1411.4171`

[35] Thomas G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, 1978.

[36] C. Landim and H-T. Yau. Fluctuation-dissipation equation of asymmetric simple exclusion processes. *Probability Theory and Related Fields*, 108:321–356, 1997.

[37] T. L. Liggett. *Interacting Particle Systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften*.

[38] R. S. Maier. The algebraic construction of phase-type distributions. *Commun. Stat., Stochastic Models*, 7(4):573–602, 1991.

[39] S. Maset. Numerical solution of retarded functional differential equations as abstract cauchy problems. *J. Comput. Appl. Math.*, 161(2):259–282, 2003.

[40] M. Maxwell and M. Woodroofe. Central limit theorems for additive functionals of Markov chains. *Annals of Probability*, 28:713–724, 2000.

[41] A. Mészáros, G. Horváth and M. Telek. Representation transformations for finding markovian representations. *Analytical and Stochastic Modeling Techniques and Applications*, volume 7984 of *Lecture Notes in Computer Science*, pages 277–291, Springer, 2013.

[42] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2004.

[43] S. Mocanu and C. Commault. Sparse representations of phase-type distributions. *Commun. Stat., Stochastic Models*, 15(4):759–778, 1999.

[44] M. F. Neuts. *Matrix Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, 1981.

[45] S. P. Obukhov and L. Peliti. Renormalisation of the "true" self-avoiding walk. *J. Phys. A*, 16:L147–L151, 1983.

[46] Colm Art O'Cinneide. Characterization of phase-type distributions. *Communications in Statistics. Stochastic Models*, 6(1):1–57, 1990.

[47] S. Olla. Central limit theorems for tagged particles and for diffusions in random environment. In É. Pardoux F. Comets, editor, *Milieux aléatoires*, volume 12 of *Panoramas et Synthses*. Societé Mathématique de France, Paris, 2002.

[48] L. Peliti and L. Pietronero. Random walks with memory. *Riv. Nuovo Cimento*, 10:1–33, 1987.

[49] M. Reed and B. Simon. *Methods of Modern Mathematical Physics Vol 1, 2.*, volume 1, 2. Academic Press New York, 1975.

[50] P. Reinecke and M. Telek. Does a given vector-matrix pair correspond to a ph distribution? *Performance Evaluation* 81:40–51, 2014.

[51] Robert Schlicht and Gerhard Winkler. A delay stochastic process with applications in molecular biology. *Journal of Mathematical Biology*, 57(5):613–48, 2008.

[52] D. J. Scott. Central Limit Theorem for martingales and for processes with stationary increments using a Skorokhod representaion approach. In *Advances in Applied Probability*, volume 5, 1:119–137, 1973.

[53] S. Sethuraman, S. R. S. Varadhan, and H-T. Yau. Diffusive limit of a tagged particle in asymmetric simple exclusion processes. *Communications in Pure and Applied Mathematics*, 53:972–1006, 2000.

[54] M. Soittola. Positive rational sequences. *Theoretical Computer Science*, 2:317–322, 1976.

[55] Kotaro Suzumura. Perron-frobenius theorem on non-negative square matrices: An elementary proof. *Hitotsubashi Journal of Economics*, 24(2):137–141, 1983.

[56] P. Tarrès, B. Tóth and B. Valkó. Diffusivity bounds for 1d Brownian polymers. Ann. Probab., 40(2):695–713, 2012.

[57] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9-12):1153–1168, 2007.

[58] B. Tóth. Persistent random walk in random environment. *Probability Theory and Related Fields*, 71:615–625, 1986.

[59] B. Tóth. The 'true' self-avoiding walk with bond repulsion on $\mathbb{Z}$: limit theorems. *Ann. Probab.*, 23:1523–1556, 1995.

[60] B. Tóth and B. Valkó. Superdiffusive bounds on self-repellent Brownian polymers and diffusion in the curl of the gaussian free field in d=2. *Journal of Stat. Phys.*, 147(1):113–131, 2012.

[61] B. Tóth and B. Vető. Continuous time 'true' self-avoiding random walk on $\mathbb{Z}$. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 8:59–75, 2011.

[62] B. Tóth and W. Werner. The true self-repelling motion. *Probab. Theory Rel. Fields*, 111:375–452, 1998.

[63] S. R. S. Varadhan. Self-diffusion of a tagged particle in equilibrium of asymmetric mean zero random walks with simple exclusion. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, 31:273–285, 1996.

[64] B. Vető. Asymptotic behaviour of random walks with long memory. PhD Thesis, 2011.

[65] Ward Whitt. Internet supplement to Stochastic-Process Limits, 2002.

[66] A. M. Yaglom. On the statistical treatment of Brownian motion (russian). *Dokl. Akad. Nauk SSSR*, 56:691–694, 1947.

[67] A. M. Yaglom. On the statistical reversibility of Brownian motion (russian). *Mat. Sbornik*, 24:457–492, 1949.