

Large-Scale Dynamics of Stochastic Particle Systems

Stefan Grosskinsky

TCC, Warwick, 2017

These notes and other information about the course are available on
go.warwick.ac.uk/SGrosskinsky/teaching/tcc.html

Contents

Introduction	2
1 Basic theory	3
1.1 Markov processes	3
1.1.1 Path space characterization	3
1.1.2 Semigroups and resolvents	4
1.1.3 Generators and semigroups	7
1.1.4 Martingale characterization	10
1.2 Markov chains	12
1.2.1 Analytic description	12
1.2.2 Probabilistic construction	15
1.3 Stochastic particle systems	18
1.3.1 Simple examples of IPS	18
1.3.2 General facts on IPS	20
2 Stationary measures, symmetries and time reversal	23
2.1 Stationary and reversible measures, currents	23
2.2 Conservation laws and symmetries	26
2.3 Time reversal	31
3 Additive functionals of Markov processes	35
3.1 Ergodicity, typical behaviour, and fluctuations	35
3.2 Additive path functionals	38
3.3 Large deviations of additive functionals	41
3.4 Conditional path ensembles	44
3.5 Tilted path ensemble	46
3.6 LDPs for Markov chains	50
4 Hydrodynamic limits and macroscopic fluctuations	53
4.1 Heuristics on currents and conservation laws	53
4.2 Rigorous results on hydrodynamic limits	57
4.2.1 Reversible gradient systems	57
4.2.2 Weakly asymmetric systems	61
4.2.3 Asymmetric, attractive IPS	62
4.3 Equilibrium fluctuations of density fields	64
4.4 Macroscopic fluctuation theory	67
References	70

Introduction

Interacting particle systems (IPS) are mathematical models of complex phenomena involving a large number of interrelated components. There are numerous examples within all areas of natural and social sciences, such as traffic flow on motorways or communication networks, opinion dynamics, spread of epidemics or fires, genetic evolution, reaction diffusion systems, crystal surface growth, financial markets, etc. The central question is to understand and predict emergent behaviour on macroscopic scales, as a result of the microscopic dynamics and interactions of individual components. Qualitative changes in this behaviour depending on the system parameters are known as collective phenomena or phase transitions and are of particular interest.

In IPS the components are modeled as particles confined to a lattice or some discrete geometry. But applications are not limited to systems endowed with such a geometry, since continuous degrees of freedom can often be discretized without changing the main features. So depending on the specific case, the particles can represent cars on a motorway, molecules in ionic channels, or prices of asset orders in financial markets, to name just a few examples. In principle such systems often evolve according to well-known laws, but in many cases microscopic details of motion are not fully accessible. Due to the large system size these influences on the dynamics can be approximated as effective random noise with a certain postulated distribution. The actual origin of the noise, which may be related to chaotic motion or thermal interactions, is usually ignored. On this level the statistical description in terms of a random process where particles move and interact according to local stochastic rules is an appropriate mathematical model. It is neither possible nor required to keep track of every single particle. One is rather interested in predicting measurable quantities which correspond to expected values of certain observables, such as the growth rate of the crystalline surface or the flux of cars on the motorway. Although describing the system only on a mesoscopic level as explained above, stochastic particle systems are usually referred to as microscopic models and we stick to this convention. On a macroscopic scale, a continuum description of systems with a large number of particles is given by coarse-grained density fields, evolving in time according to a partial differential equation. The form of this equation depends on the particular application, and its mathematical connection to microscopic particle models is one of the fundamental questions in applied probability and statistical mechanics.

The main references for the module are:

- Chapters 3 and 4 in: T.M. Liggett: Continuous Time Markov Processes. AMS Graduate Studies in Mathematics 113 (2010)
- R. Chetrite, H. Touchette: Nonequilibrium Markov processes conditioned on large deviations. Ann. Inst. H. Poincaré **16**, 2005 (2015)
- L. Bertini, A. Faggionato, D. Gabrielli: Large deviations of the empirical flow for continuous time Markov chains. Ann. Inst. H. Poincaré **51**, 867 (2015)
- L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim: Macroscopic Fluctuation Theory. Reviews of Modern Physics **87**, 593-636 (2015)

1 Basic theory

1.1 Markov processes

The material in this section follows closely the presentation in [31], Chapter 3.

1.1.1 Path space characterization

We will study several continuous-time stochastic processes and introduce the basic general setting now. The **state space** E is a complete, separable¹ metric space. We denote by $\mathcal{M}(E)$ the set of measures and $\mathcal{M}_1(E)$ the set of normalized probability measures on E . The measurable structure on E is given by the Borel σ -algebra $\mathcal{B}(E)$. If E is a countable set we will usually use the discrete topology which is simply the powerset of $\mathcal{P}(E)$, as is the corresponding Borel σ -algebra.

A **continuous time stochastic process** on E is denoted by $(X_t : t \geq 0)$, i.e. a family of random variables X_t . The probability space Ω for the process is given by the path space of E -valued, right-continuous functions with left limits (also called càdlàg), denoted by

$$\Omega = D_E[0, \infty) := \{ \omega : [0, \infty) \rightarrow E \mid t \mapsto \omega_t \text{ càdlàg} \}. \quad (1.1)$$

The explicit construction of the random variables is then simply $X_t(\omega) = \omega_t$. We will often also restrict ourselves to compact time intervals $[0, T]$ with associated paths $D_E[0, T]$. The σ -algebra \mathcal{F} on Ω is the smallest such that the mapping $\omega \mapsto X_t(\omega)$ is measurable for all $t \geq 0$.

In general, the process $(X_t : t \geq 0)$ is then characterized by a probability measure \mathbb{P} on (Ω, \mathcal{F}) , with associated expectation denoted by \mathbb{E} . With ω drawn from this measure, the function $t \mapsto X_t(\omega)$ is called **sample path**. In the following we restrict ourselves to Markov processes, where given the present state, the future time evolution is independent of the past. To be able to consider different initial conditions for a given Markov process, we use the following definition.

Definition 1.1 A (*homogeneous*) *Markov process* on E consists of a collection $(\mathbb{P}^x : x \in E)$ of probability measures on $(D_E[0, \infty), \mathcal{F})$ and a right-continuous filtration $(\mathcal{F}_t : t \geq 0)$ on Ω with respect to which the random variables X_t are adapted. We further impose the following properties:

- (a) $\mathbb{P}^x[X_0 = x] = 1$ for all $x \in E$,
i.e. \mathbb{P}^x is normalized on all paths with initial condition $X_0 = x$.
- (b) The mapping $x \mapsto \mathbb{P}^x(A)$ is measurable for every $A \in \mathcal{F}$.
- (c) For all $x \in E, t > 0$ and all bounded measurable Y on Ω we have

$$\mathbb{E}^x[Y \circ \theta_t \mid \mathcal{F}_t] = \mathbb{E}^{X_t}[Y] \quad \mathbb{P}^x - a.s. ,$$

where $\theta_t \omega = \omega_{t+}$ denotes a time shift. (homogeneous Markov property)

Note that the Markov property as formulated in (c) implies that the process is (time-)homogeneous, excluding an explicit time dependence in the law \mathbb{P}^x . Markov processes can be generalized to be inhomogeneous (see e.g. [12]), but we will concentrate only on homogeneous processes. The condition in (b) allows to consider processes with general initial distributions $\mu \in \mathcal{M}_1(E)$ via

$$\mathbb{P}^\mu := \int_E \mathbb{P}^x \mu(dx) . \quad (1.2)$$

¹i.e. contains a countable, dense set

When we do not want to specify the initial condition for the process we will often only write \mathbb{P} .

Unless specified otherwise, we will always use the natural filtration generated by the process with

$$\mathcal{F}_t = \sigma\{X_s^{-1}(A) \mid s \leq t, A \in \mathcal{B}(E)\}, \quad (1.3)$$

accumulating the information concerning the process over time, and only that information.

1.1.2 Semigroups and resolvents

Using the homogeneous Markov property, it is easy to see that the transition kernel

$$P_t(x, dy) := \mathbb{P}^x[X_t \in dy] \quad (1.4)$$

fulfills the Chapman-Kolmogorov equation

$$P_{t+s}(x, A) = \int_{z \in E} P_t(x, dz)P_s(z, A) \quad \text{for all } A \in \mathcal{B}(E). \quad (1.5)$$

As usual, it is practical to use a weak characterization of the measure $P_t(x, \cdot)$ in terms of expectations of test functions. The suitable space of test functions in general depends on the process at hand, but there is a large class of processes fitting in the following framework. If the state space is compact $C(E)$ denotes the space of continuous real-valued functions on E (which are also bounded by compactness). If E is locally compact¹, $C(E)$ denotes the space of continuous real-valued functions vanishing at infinity. Together with the sup-norm $\|f\| = \|f\|_\infty = \sup_{x \in E} f(x)$ this is a Banach space in both cases. For each $t \geq 0$ we define the operator $P_t : C(E) \rightarrow C(E)$ by

$$(P_t f)(x) = P_t f(x) = \mathbb{E}^x[f(X_t)] = \int_E P_t(x, dy) f(y) \quad \text{for all } x \in E. \quad (1.6)$$

While P_t is well defined for all $f \in C(E)$, in general it is not guaranteed that the range of the operator is also $C(E)$. Processes which fulfill that property are called **Feller processes**. We will focus on this large class for now, and will see processes later where this approach has to be adapted. A (trivial) example of a non-Feller process is given by the translation semigroup on $E = [0, \infty)$ with $c > 0$,

$$P_t f(x) = f(x + ct), \quad x > 0, \quad P_t f(0) = f(0). \quad (1.7)$$

The corresponding process moves deterministically to the right with finite speed or is stuck in 0.

The Chapman-Kolmogorov equations imply that $(P_t : t \geq 0)$ has a semigroup structure, we summarize this and some further properties in the next result.

Proposition 1.1 Markov semigroups

Let $(X_t : t \geq 0)$ be a Feller process on E . Then the family $(P_t : t \geq 0)$ is a **Markov semigroup**, i.e.

$$(a) \quad P_0 f = f \quad \text{for all } f \in C(E) \quad (\text{identity at } t = 0)$$

¹i.e. every point $x \in E$ has a compact neighbourhood, such as \mathbb{R}^n or \mathbb{N}^n , n fixed. Other examples: $\{0, 1\}^{\mathbb{Z}}$ is compact w.r.t. product topology due to Tychonoff's theorem, $\mathbb{N}^{\mathbb{Z}}$ is not locally compact.

(b) $\lim_{t \searrow 0} P_t f = f$ for all $f \in C(E)$, (strong continuity)

(c) $P_{t+s} f = P_t P_s f$ for all $f \in C(E)$, $s, t \geq 0$, (semigroup/Markov property)

(d) $P_t f \geq 0$ for all non-negative $f \in C(E)$. (positivity)

(e) If E is compact, $P_t 1 = 1$ for all $t \geq 0$. If E is not compact, there exist $f_n \in C(E)$ so that $\sup_n \|f_n\| < \infty$ and $P_t f_n(x) \rightarrow 1$ for all $x \in E$ and $t \geq 0$.

(conservation of probability)

Proof. (a) $P_0 f(x) = \mathbb{E}^x(f(X_0)) = f(x)$ since $X_0 = x$ which is equivalent to (a) of Def. 1.1.
 (b) pointwise convergence follows immediately from right-continuity of paths and continuity of f . The required uniform convergence will be checked later making use of the resolvent which is introduced below.
 (c) follows from the Markov property of X_t (Def. 1.1(c))

$$\begin{aligned} P_{t+s} f(x) &= \mathbb{E}^x [f(X_{t+s})] = \mathbb{E}^x \left[\mathbb{E}^x [f(X_{t+s}) | \mathcal{F}_t] \right] = \mathbb{E}^x \left[\mathbb{E}^{X_t} [f(X_s)] \right] = \\ &= \mathbb{E}^x [(P_s f)(X_t)] = P_t P_s f(x). \end{aligned} \quad (1.8)$$

(d) is immediate by definition.

(e) $P_t 1 = \mathbb{E}^x[1] = \mathbb{E}^x[\mathbb{1}_{X_t}(E)] = 1$ since $X_t \in E$ for all $t \geq 0$. For non-compact E , $1 \notin C(E)$, but we can use pointwise converging indicator functions $f_n = \mathbb{1}_{A_n}$ for a sequence of compact sets $A_n \rightarrow E$ as $n \rightarrow \infty$. □

Remarks. One consequence of the semigroup property (c) is that P_s and P_t commute. Using (d) and (e), P_t is also *contractive*, i.e. for all $f \in C(E)$ with compact E

$$\|P_t f\| \leq \|P_t \|f\| \|P_t 1\| = \|f\|. \quad (1.9)$$

Strong continuity and contractivity imply that $t \mapsto P_t f$ is actually uniformly strongly continuous for all $t > 0$. Using also the semigroup property (c) we have for all $t > \epsilon > 0$ and $f \in C(E)$

$$\|P_t f - P_{t-\epsilon} f\| = \|P_{t-\epsilon} (P_\epsilon f - f)\| \leq \|P_\epsilon f - f\|, \quad (1.10)$$

which vanishes for $\epsilon \searrow 0$ and implies left-continuity. Right-continuity works analogously.

An important tool to study Markov semigroups is their Laplace transform, which is called the **resolvent** and given by

$$R_\lambda f := \int_0^\infty e^{-\lambda t} P_t f dt \quad \text{for all } \lambda > 0. \quad (1.11)$$

The integral is well defined since $t \mapsto e^{-\lambda t} P_t f$ is continuous and $\|e^{-\lambda t} P_t f\| \leq e^{-\lambda t} \|f\|$ by the contraction property. R_λ is a linear operator on $C(E)$ with

$$\|R_\lambda f\| \leq \frac{1}{\lambda} \|f\| \quad \text{and} \quad \lim_{\lambda \rightarrow \infty} \lambda R_\lambda f(x) = f(x) \text{ for all } x \in E.$$

The semigroup property translates into the **resolvent equation**

$$R_\lambda - R_\alpha = (\lambda - \alpha) R_\lambda R_\alpha \quad \text{for all } \lambda, \alpha > 0. \quad (1.12)$$

To see this write

$$\begin{aligned}
R_\lambda R_\alpha f &= \int_0^\infty e^{-\lambda t} \int_0^\infty e^{-\alpha s} P_t P_s f \, ds \, dt = \\
&= \int_0^\infty e^{-\lambda t} \int_0^\infty e^{-\alpha s} P_{s+t} f \, ds \, dt \\
&= \int_0^\infty P_r f \int_0^r e^{-\lambda t - \alpha(r-t)} \, dt \, dr = \int_0^\infty P_r f \frac{e^{-\lambda r} - e^{-\alpha r}}{\lambda - \alpha} \, dr.
\end{aligned}$$

One consequence of (1.12) is that R_λ and R_α commute, and that the range $\mathcal{R}(R_\lambda) \subseteq C(E)$ is independent of λ .

Remaining proof of Prop. 1.1(b). If $f = R_\lambda g \in \mathcal{R}(R_\lambda)$ then we can write

$$P_t f = \int_0^\infty e^{-\lambda s} P_{s+t} g \, ds = \int_t^\infty e^{-\lambda(r-t)} P_r g \, dr, \quad (1.13)$$

which converges uniformly to f as $t \searrow 0$. To justify the interchanges of integration, note that $P_t f(x)$ is uniformly bounded, continuous in t for each x and continuous in x for each t , so jointly measurable in x and t . Therefore, $\|P_t f - f\| \rightarrow 0$ for all f in the strong closure of $\mathcal{R}(R_\lambda)$, which is equal to the weak closure (see Corollary A.7 in [31]). Since $\lambda R_\lambda f$ converges pointwise to f as $\lambda \rightarrow \infty$ for every $f \in C(E)$, this can be used to define a weakly approximating sequence of f and the weak closure of $\mathcal{R}(R_\lambda)$ is the full space $C(E)$. \square

The semigroup $(P_t : t \geq 0)$ describes the time evolution of expected values of observables f on X for a given Markov process. In fact it weakly characterizes all **finite-dimensional distributions** $(X_{t_1}, \dots, X_{t_n})$ of the process for all $n \geq 1$ and $0 \leq t_1 < \dots < t_n$. For example, one dimensional distributions are given by the definition (1.6), and for two-dimensional distributions consider for $s < t$

$$\mathbb{E}^x [f(X_s)g(X_t)] = \mathbb{E}^x [f(X_s)\mathbb{E}^{X_s}[g(X_{t-s})]] = P_s(f P_{t-s}g)(x). \quad (1.14)$$

With Proposition 1.1 and the following result, semigroups indeed provide a full characterization of the process.

Theorem 1.2 *Suppose $(P_t : t \geq 0)$ is a Markov semigroup on $C(E)$. Then there exists a unique (Feller) Markov process $(X_t : t \geq 0)$ on E such that*

$$\mathbb{E}^x f(X_t) = P_t f(x) \quad \text{for all } f \in C(E), x \in E \text{ and } t \geq 0. \quad (1.15)$$

Proof. Denote by μ_{t_1, \dots, t_n} the finite dimensional distributions on E^n generated by the semigroup. The family of all these for $n \geq 1$ is consistent, in the sense that

$$\mu_{t_1, \dots, t_{n+1}}(A \times E) = \mu_{t_1, \dots, t_n}(A) \quad (1.16)$$

for all Borel sets $A \subseteq E^n$. (In the compact case this simply follows from $P_t 1 = 1$.) Then **Kolmogorov's extension theorem** guarantees the existence of a unique measure $\bar{\mathbb{P}}$ on $\bar{\Omega} := E^{[0, \infty)}$ so that the induced measure generated by the projection $\bar{\omega} \mapsto (\bar{\omega}_{t_1}, \dots, \bar{\omega}_{t_n})$ from $\bar{\Omega}$ to E^n is equal to μ_{t_1, \dots, t_n} for each $n \geq 1$ and $0 \leq t_1 < \dots < t_n$.

The main work is to show that there exists a right-continuous **version**¹ of the process $\bar{\mathbb{P}}$ with paths in $\Omega = D[0, \infty) \subseteq \bar{\Omega}$. This is slightly technical, and can be found in [31] Theorem 3.26. \square

¹ $(X_t : t \geq 0)$ and $(Y_t : t \geq 0)$ are versions, if for all $t \geq 0$, $\bar{\mathbb{P}}[X_t = Y_t] = 1$.

1.1.3 Generators and semigroups

Markov semigroups can be characterized by an **infinitesimal generator**, which is defined as

$$\mathcal{L}f := \lim_{t \searrow 0} \frac{P_t f - f}{t} \quad \text{for all } f \in \mathcal{D}_{\mathcal{L}}, \quad (1.17)$$

where the domain is given by

$$\mathcal{D}_{\mathcal{L}} = \{f \in C(E) : \text{the (strong) limit in (1.17) exists}\}. \quad (1.18)$$

In most cases of interest this is a proper subset of $C(E)$ and is related to the range of the resolvent. For any fixed $\lambda > 0$ the range $\mathcal{R}(R_{\lambda})$ is equal to the domain $\mathcal{D}_{\mathcal{L}}$ of the generator and we have $R_{\lambda} = (\lambda \mathbb{I} - \mathcal{L})^{-1}$. The inverse exists for all $\lambda > 0$ since the spectrum of \mathcal{L} is non-positive, as we will come back to later. We summarize this and further important properties of the generator in the next result.

Proposition 1.3 *With operator \mathcal{L} as defined in (1.17) we have for any $g \in C(E)$ and $\lambda > 0$*

$$f = \lambda R_{\lambda} g \quad \text{if and only if} \quad f \in \mathcal{D}_{\mathcal{L}} \quad \text{and satisfies} \quad f - \lambda^{-1} \mathcal{L}f = g. \quad (1.19)$$

Furthermore, \mathcal{L} is a **Markov generator**, i.e. it is a linear operator with the following properties:

- (a) $\mathcal{D}_{\mathcal{L}}$ is dense in $C(E)$.
- (b) For $f \in \mathcal{D}_{\mathcal{L}}$, $\lambda \geq 0$: $\inf_{x \in E} f(x) \geq \inf_{x \in E} (f - \lambda \mathcal{L}f)(x)$. (positivity)
- (c) The range $\mathcal{R}(\mathbb{I} - \lambda \mathcal{L}) = C(E)$ for sufficiently small $\lambda > 0$.
- (d) If E is compact, $1 \in \mathcal{D}_{\mathcal{L}}$ and $\mathcal{L}1 = 0$. If E is not compact, then there exist $f_n \in \mathcal{D}_{\mathcal{L}}$ so that $g_n = f_n - \lambda \mathcal{L}f_n$ satisfies $\sup_n \|g_n\| < \infty$ and $f_n, g_n \rightarrow 1$ pointwise. (conservation of probability)

Remarks. Note that with property (b)

$$f \in \mathcal{D}_{\mathcal{L}}, \lambda \geq 0 \text{ and } f - \lambda \mathcal{L}f = g \text{ imply } \|f\| \leq \|g\|. \quad (1.20)$$

So for sufficiently small $\lambda > 0$, $(\mathbb{I} - \lambda \mathcal{L})^{-1}$ is in fact an everywhere defined contraction that maps non-negative functions to non-negative functions, which is related to positivity of the semigroup. Property (d) ensures that $P_t 1 = 1$ as we will see below (in the compact case).

Proof. Suppose $f = \lambda R_{\lambda} g$ for some $\lambda > 0$ and $g \in C(E)$. Using the semigroup property and a change of variables as in the remaining proof of Prop. 1.1(b) we get

$$\begin{aligned} \frac{P_t f - f}{t} &= \lambda \frac{e^{\lambda t} - 1}{t} \int_t^{\infty} e^{-\lambda s} P_s g \, ds - \frac{\lambda}{t} \int_0^t e^{-\lambda s} P_s g \, ds \\ &\rightarrow \lambda^2 R_{\lambda} g - \lambda g = \lambda(f - g) \quad \text{as } t \searrow 0. \end{aligned} \quad (1.21)$$

Property (b) in Prop. 1.1 is used in the passage to the limit. This proves one direction in (1.19) as well as (c). Since $\lambda R_{\lambda} g \in \mathcal{D}_{\mathcal{L}}$ and $\lambda R_{\lambda} g \rightarrow g$ as $\lambda \rightarrow \infty$, $\mathcal{D}_{\mathcal{L}}$ is dense in $C(E)$ proving (a).

For $t > 0$ and $f \in \mathcal{D}_{\mathcal{L}}$ put

$$g_t := \left(1 + \frac{\lambda}{t}\right) f - \frac{\lambda}{t} P_t f = f - \lambda \frac{P_t f - f}{t}.$$

Then, as $t \searrow 0$, $g_t \rightarrow f - \lambda \mathcal{L}f$ and

$$\left(1 + \frac{\lambda}{t}\right) \inf_x f(x) \geq \frac{\lambda}{t} \inf_x P_t f(x) + \inf_x g_t(x) \geq \frac{\lambda}{t} \inf_x f(x) + \inf_x g_t(x),$$

which implies (b).

Now suppose $f - \lambda^{-1} \mathcal{L}f = g$ for some $f \in \mathcal{D}_{\mathcal{L}}$ and $\lambda > 0$. By the direction of (1.19) already proved, $h = \lambda R_{\lambda} g$ satisfies $h - \lambda^{-1} \mathcal{L}h = g$, so that $f = h$ by (1.20), finishing the proof of (1.19). (d) is immediate from the definition (1.17) and $P_t 1 = 1$ for compact E , with similar modification as before for non-compact E . \square

Theorem 1.4 (Hille-Yosida) *There is a one-to-one correspondence between Markov generators and semigroups on $C(E)$, given by (1.17) and*

$$P_t f := \lim_{n \rightarrow \infty} \left(\mathbb{I} - \frac{t}{n} \mathcal{L} \right)^{-n} f \quad \text{for } f \in C(E), t \geq 0. \quad (1.22)$$

Furthermore, if $f \in \mathcal{D}_{\mathcal{L}}$ then $P_t f \in \mathcal{D}_{\mathcal{L}}$ for all $t \geq 0$, is a continuously differentiable function of t , and satisfies

$$\frac{d}{dt} P_t f = P_t \mathcal{L}f = \mathcal{L}P_t f. \quad (1.23)$$

These are called the Kolmogorov **forward** and **backward equation**, respectively.

Proof. To show (1.23), note that

$$\frac{d}{dt} P_t f = \lim_{s \searrow 0} \frac{P_{t+s} f - P_t f}{s} = \lim_{s \searrow 0} P_t \frac{P_s f - f}{s} = \lim_{s \searrow 0} \frac{P_s(P_t f) - P_t f}{s} \quad (1.24)$$

provided that any of the limits exists, since the expressions inside the limits are identical. The middle limit exists, since $f \in \mathcal{D}_{\mathcal{L}}$ and P_t is a contraction, and therefore the others do as well. Due to the third limit $P_t f \in \mathcal{D}_{\mathcal{L}}$ and (1.23) holds. The middle expression in (1.23) is continuous in t , so $P_t f$ is continuously differentiable.

For (1.22), iterate (1.19) and (1.11) to get

$$\left(\mathbb{I} - \lambda^{-1} \mathcal{L} \right)^{-n} f = \lambda^n R_{\lambda}^n f = \int_0^{\infty} \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} T_s f ds.$$

Therefore,

$$\left(\mathbb{I} - \frac{t}{n} \mathcal{L} \right)^{-n} = \mathbf{E} [P_{t(\xi_1 + \dots + \xi_n)/n} f]$$

where ξ_1, ξ_2, \dots are auxiliary i.i.d. exponential random variables with mean 1 (and corresponding expectation \mathbf{E}). If $f \in \mathcal{D}_{\mathcal{L}}$,

$$\|P_t f - P_s f\| \leq \|\mathcal{L}f\| |t - s|$$

by (1.23), so

$$\left\| \left(\mathbb{I} - \frac{t}{n} \mathcal{L} \right)^{-n} f - P_t f \right\| \leq t \|\mathcal{L}f\| \mathbf{E} \left| \frac{\xi_1 + \dots + \xi_n}{n} - 1 \right|.$$

For $f \in \mathcal{D}_{\mathcal{L}}$ the result now follows from the law of large numbers. Since $\mathcal{D}_{\mathcal{L}}$ is dense in $C(E)$ and all operators involved are contractions, it holds for all $f \in C(E)$. \square

Remarks. $P_t f$ is the unique solution to the backward equation

$$\frac{d}{dt} u(t) = \mathcal{L} u(t) \quad \text{with initial condition} \quad u(0) = f, \quad (1.25)$$

so formally $P_t = \exp(\mathcal{L}t)$. So solutions to differential equations of the type (1.25) can be written as expectations of a stochastic process, $u(t, x) = P_t f(x) = \mathbb{E}^x[f(X_t)]$, provided that the operator on the right-hand side is a Markov generator. We will see a more general version of this idea later in the Feynman-Kac formula.

When \mathcal{L} is bounded (e.g. for finite E where \mathcal{L} is simply a matrix as discussed later), there are at least three ways of defining this exponential:

$$\sum_{n=0}^{\infty} \frac{(t\mathcal{L})^n}{n!}, \quad \lim_{n \rightarrow \infty} \left(\mathbb{I} + \frac{t}{n} \mathcal{L} \right)^n \quad \text{and} \quad \lim_{n \rightarrow \infty} \left(\mathbb{I} - \frac{t}{n} \mathcal{L} \right)^{-n}.$$

The last of these is the only one that makes sense also in case \mathcal{L} is unbounded.

Examples. A **jump process** on E with transition rates $c(x, dy)$ is given by the generator

$$\mathcal{L}f(x) = \int_E c(x, dy)(f(y) - f(x)) \quad (1.26)$$

defined (at least) for all $f \in C(E)$ with compact support. A special case is the **Poisson process** $PP(\lambda)$ with $E = \mathbb{N}_0$, $c(x, y) = \lambda \delta_{y, x+1}$ with rate $\lambda > 0$, which we discuss later in the context of Markov chains.

On $E = \mathbb{R}$ the operator $\mathcal{L}f = bf'$ for $b \in \mathbb{R}$ is a Markov generator with domain

$$\mathcal{D}_{\mathcal{L}} = \{f \in C(\mathbb{R}) : f' \in C(\mathbb{R})\}.$$

The corresponding semigroup is the (deterministic) translation $P_t f(x) = f(x + bt)$.

Standard Brownian motion on $E = \mathbb{R}$ is given by the generator $\mathcal{L}f = \frac{1}{2}f''$ with domain

$$\mathcal{D}_{\mathcal{L}} = \{f \in C(\mathbb{R}) : f', f'' \in C(\mathbb{R})\}. \quad (1.27)$$

The corresponding semigroup is given by the **heat kernel**

$$P_t f(x) = \int_{\mathbb{R}} p_t(x, y) f(y) dy, \quad \text{where} \quad p_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{(y-x)^2/(2t)} \quad (1.28)$$

is the pdf of the Gaussian $\mathcal{N}(x, t)$.

More generally, a **diffusion process** on $E = \mathbb{R}$ is a (Feller) Markov process with continuous paths. It is given by the generator

$$\mathcal{L}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x) \quad (1.29)$$

with the same domain (1.27), provided that $b, \sigma \in C(E)$. There is no Markov generator whose restriction to smooth functions is given by $\mathcal{L}f = f'''$, since the positivity property (b) in Prop. 1.3 is not fulfilled. The process can also be represented as a solution to the **stochastic differential equation** $dX_t = b(X_t) dt + \sigma(X_t) dB_t$, which we will not discuss in further detail.

Remark on domains. The domain of the generator is of great importance, and is not only a technical detail. For example, boundary conditions for processes are encoded in a restriction of the domain, not the action of the operator. For Brownian motion on $E = [0, \infty)$ with absorbing boundary condition in 0 has generator $\mathcal{L}^a f = \frac{1}{2}f''$ for functions in the restricted domain

$$\mathcal{D}_{\mathcal{L}} = \{f \in C([0, \infty)) : f', f'' \in C([0, \infty)), f''(0) = 0\}.$$

For reflecting boundary conditions, the domain would be

$$\mathcal{D}_{\mathcal{L}} = \{f \in C([0, \infty)) : f', f'' \in C([0, \infty)), f'(0) = 0\}.$$

In general by taking closures¹, a bounded linear operator is uniquely determined by its values on a dense set. But generators are usually unbounded such as the differential operators above, and there are simple counter examples related to Brownian motion with boundary conditions (see [31], Remark 3.57). But in general it is hard to fully characterize the domain of a generator, and it is desirable to uniquely define them on a suitable smaller set. We say that $D \subseteq \mathcal{D}_{\mathcal{L}}$ is a **core** for \mathcal{L} if the closure of $\mathcal{L}|_D = \mathcal{L}$. Therefore \mathcal{L} is determined by its values on D .

For example, $f \in C(E)$ with compact support are a core for the generator (1.26), or $f \in C^2(\mathbb{R})$ with compact support for Brownian motion.

1.1.4 Martingale characterization

Recall that the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ of a random variable X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ w.r.t. the sub- σ -algebra \mathcal{G} is a \mathcal{G} -measurable random variable which satisfies

$$\mathbb{E}[\mathbb{1}_G \mathbb{E}[X|\mathcal{G}]] = \int_G \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \int_G X d\mathbb{P} = \mathbb{E}[X \mathbb{1}_G] \quad \text{for all } G \in \mathcal{G}. \quad (1.30)$$

A martingale w.r.t. a filtration $(\mathcal{F}_t : t \geq 0)$ is a real-valued stochastic process $(M_t : t \geq 0)$ adapted to that filtration with

$$\mathbb{E}[|M_t|] < \infty \quad \text{and} \quad \mathbb{E}[M_t | \mathcal{F}_s] = M_s \quad \text{for all } t \geq s \geq 0. \quad (1.31)$$

Martingales have constant expectation and therefore no drift. In general, stochastic processes can be decomposed into a drift term and fluctuations around that, which are given by a martingale. The size of the fluctuations is characterized by the quadratic variation process of a martingale M_t written as $[M]_t$, which is the unique right-continuous and increasing process such that $M_t^2 - [M]_t$ is a local martingale.

Definition 1.2 Let L be a linear operator on $C(E)$ with domain \mathcal{D} . A measure \mathbb{P} on $D[0, \infty)$ is a solution to the *martingale problem* for L with initial condition $x \in E$, if

- (a) $\mathbb{P}[X_0 = x] = 1$, and
- (b) for all $f \in \mathcal{D}$, the process

$$M_t^f := f(X_t) - f(x) - \int_0^t Lf(X_s) ds \quad (1.32)$$

is a martingale w.r.t. \mathbb{P} and the natural filtration $(\mathcal{F}_t : t \geq 0)$.

¹A linear operator \mathcal{L} on $C(E)$ is closed if its graph $\{(f, \mathcal{L}f) : f \in \mathcal{D}_{\mathcal{L}}\}$ is a closed subset of $C(E) \times C(E)$.

Theorem 1.5 Let \mathcal{L} be a Markov generator and let $(\mathbb{P}^x : x \in E)$ be the corresponding unique Feller process. Then for each $x \in E$, \mathbb{P}^x is the unique solution to the martingale problem for \mathcal{L} with initial condition x .

On the other hand, suppose that for each $x \in E$ the solution to the martingale problem for a linear operator \mathcal{L} on $C(E)$ with initial condition x is unique. Then $(\mathbb{P}^x : x \in E)$ is a Feller process on E whose generator is \mathcal{L} .

'Proof'. (for details see [6], Sections I.5 and I.6, and [31], Section 3.4.2.)

Let $(X_t : t \geq 0)$ be a Markov process with generator \mathcal{L} . Then for all $s \leq t$ we can write

$$M_t^f = \underbrace{f(X_s) - f(X_0) + \int_0^s \mathcal{L}f(X_u) du}_{M_s^f} + f(X_t) - f(X_s) + \int_s^t \mathcal{L}f(X_u) du. \quad (1.33)$$

The forward equation (1.23) implies

$$\mathbb{E}[f(X_t) - f(X_s) | \mathcal{F}_s] = \mathbb{E}\left[\int_s^t \mathcal{L}f(X_u) du \middle| \mathcal{F}_s\right] \quad (1.34)$$

which leads to (1.31),

$$\mathbb{E}[M_t^f | \mathcal{F}_s] = M_s^f + \mathbb{E}\left[f(X_t) - f(X_s) + \int_s^t \mathcal{L}f(X_u) du\right] = M_s^f. \quad (1.35)$$

Also, for $f \in C(E)$ it can be shown that $\mathbb{E}(|M_t^f|) < \infty$ for all $t \geq 0$.

On the other hand, if (1.32) is a martingale, this implies

$$\frac{d}{dt} \mathbb{E}[M_t^f] = \frac{d}{dt} \mathbb{E}[f(X_t)] - \mathbb{E}[\mathcal{L}f(X_s)] = 0 \quad (1.36)$$

for all f , which uniquely identifies \mathcal{L} as the generator of the process using Theorem 1.4. \square

Uniqueness of the solution to the martingale problem is often hard to show, we will see a sufficient condition on the generator for a particular class of processes later. The quadratic variation of the martingale (1.32) is

$$[M^f]_t = \int_0^t \left(\mathcal{L}f^2(X_s) - 2f(X_s)\mathcal{L}f(X_s) \right) ds. \quad (1.37)$$

Note that $(f(X_t) : t \geq 0)$ itself is a martingale if and only if $\mathcal{L}f = 0$.

Examples. Let $(N_t : t \geq 0)$ be a Poisson process $PP(\lambda)$ with generator $\mathcal{L}f(x) = \lambda(f(x+1) - f(x))$. Then we have $\mathcal{L}x = \lambda$, $\mathcal{L}x^2 = \lambda(2x+1)$ and with Theorem 1.5 we get that

$$M_t = N_t - \lambda t \quad \text{is a martingale with quadratic variation} \quad [M]_t = \lambda t. \quad (1.38)$$

Therefore $(N_t - \lambda t)^2 - \lambda t$ is also a martingale.

For standard Brownian motion $(B_t : t \geq 0)$ with generator $\mathcal{L}f = \frac{1}{2}f''$ we have with Theorem 1.5

$$\begin{aligned} f(x) = x &\Rightarrow B_t \quad \text{itself is a martingale with quadratic variation } t, \\ f(x) = x^2 &\Rightarrow B_t^2 - t \quad \text{is a martingale.} \end{aligned} \quad (1.39)$$

In fact, for both processes the reverse is also true, i.e. on the given respective state space the Poisson process and Brownian motion are characterized by the above martingale properties.

For a diffusion process $(X_t : t \geq 0)$ with generator (1.29) we have $\mathcal{L}x = b(x)$ and $\mathcal{L}x^2 = 2xb(x) + \sigma^2(x)$, so we get with Theorem 1.5,

$$M_t = X_t - X_0 - \int_0^t b(X_s) ds = \int_0^t \sigma(X_s) dB_s \quad (1.40)$$

is a martingale, with quadratic variation $\int_0^t \sigma^2(X_s) ds$. The stochastic (Itô) integral on the r.h.s. is a result of the SDE representation of the process. $(X_t : t \geq 0)$ is itself a martingale if and only if $b(x) \equiv 0$.

1.2 Markov chains

1.2.1 Analytic description

Throughout this section let E be a countable set with the discrete topology $\mathcal{P}(E)$, so that every function $f : E \rightarrow \mathbb{R}$ is continuous. Markov processes $(X_t : t \geq 0)$ on E are called (continuous-time) **Markov chains**. The transition kernel is now simply a function in $p_t(x, y) = \mathbb{P}^x[X_t = y]$, and the associated semigroup can be interpreted as a (possibly infinite) matrix

$$P_t = (p_t(x, y) : x, y \in E) \quad \text{with} \quad p_0(x, y) = \delta_{x,y}. \quad (1.41)$$

The Chapman-Kolmogorov equations can be written as $P_{t+s} = P_t P_s$ and they imply in particular

$$p_{s+t}(x, x) \geq p_s(x, x) p_t(x, x) \quad \text{for all } x \in E \text{ and } s, t \geq 0.$$

By strong continuity (Prop.1.1(b)), this in turn implies that $p_t(x, x) > 0$ for all $t \geq 0$ and $x \in E$. Furthermore, $f(t) := -\log p_t(x, x)$ is a sub-additive function¹ with $f(0) = 0$, and therefore the derivative

$$c(x, x) := \frac{d}{dt} p_t(x, x) = \lim_{t \searrow 0} \frac{p_t(x, x) - 1}{t} = \lim_{t \searrow 0} \frac{-\log p_t(x, x)}{t} \in [-\infty, 0] \quad (1.42)$$

exists. A state $x \in E$ is called **instantaneous** if $|c(x, x)| = \infty$, and **absorbing** if $c(x, x) = 0$. See e.g. [31] Section 2.4 for an example by Blackwell where all states of a Markov chain are instantaneous. If $|c(x, x)| < \infty$ it can be shown that (cf. Theorem 2.14 in [31])

$$c(x, y) := \left. \frac{d}{dt} p_t(x, y) \right|_{t=0} \in [0, \infty) \quad (1.43)$$

exists for all $y \neq x$, and

$$\sum_{y \in E} c(x, y) \leq 0. \quad (1.44)$$

$c(x, y)$ is called the **transition rate** from state x to y , and $c(x, x) = 0$ implies $c(x, y) = 0$ also for all $x \neq y$. For a given process $(\mathbb{P}^x : x \in E)$ the rates can be identified via

$$\mathbb{P}^x[X_t = y] = c(x, y) t + o(t) \quad \text{as } t \searrow 0 \quad \text{for } x \neq y, \quad (1.45)$$

¹i.e. $f(t+s) \leq f(t) + f(s)$

and represent probabilities per unit time.

If strict inequality holds in (1.44) the chain is called **explosive**, and if equality holds we have

$$c(x) := -c(x, x) = \sum_{y \neq x} c(x, y) ,$$

which corresponds to the **total exit rate** from state x . With the definition (1.17) this yields for the generator of the Markov chain

$$\mathcal{L}f(x) = \lim_{t \searrow 0} \frac{P_t f - f}{t} = \sum_{y \in E} c(x, y) (f(y) - f(x)) , \quad (1.46)$$

for all $f \in C(E)$ such that the sum converges. Note that for Markov chains the generator is often written in terms of the **Q -matrix** (see e.g. [11] Section 2.1)

$$Q := (c(x, y) : x, y \in E) \quad \text{so that} \quad \mathcal{L}f(x) = Q|f\rangle(x) ,$$

where $Q|f\rangle$ denotes the standard matrix product with a column vector $|f\rangle$.

Example. For the simple random walk with state space $E = \mathbb{Z}$ we have

$$c(x, x+1) = p \quad \text{and} \quad c(x, x-1) = q , \quad (1.47)$$

while all other transition rates vanish. The generator is given by

$$\mathcal{L}f(x) = p(f(x+1) - f(x)) + q(f(x-1) - f(x)) , \quad (1.48)$$

and in the symmetric case $p = q$ it is proportional to the discrete Laplacian.

All Markov chains with non-instantaneous states have a generator (1.46), and the corresponding semigroup $P_t f = e^{\mathcal{L}t} f$ is given by a solution of the backward equation (1.25). If the chain is explosive, this solution is **sub-stochastic**, i.e.

$$\sum_{y \in E} p_t(x, y) < 1 \quad \text{for such } x \in E \text{ where inequality is strict in (1.44),}$$

and sample paths can leave the state space after a finite amount of time (disappear at infinity), as is discussed in the next subsection.

Note that Markov chains are Feller processes in the sense of the previous subsection, such that the equivalence of generators, semigroups and the process fully applies, if and only if there are no instantaneous states and the chain is not explosive. One sufficient condition for this is for E to be finite and therefore compact, in which case the domain of the generator is the full set $\mathcal{D}_{\mathcal{L}} = C(E)$ of all functions $f : E \rightarrow \mathbb{R}$. The same holds for infinite E , provided that the rates are uniformly bounded, i.e.

$$\bar{c} := \sup_{x \in E} c(x) < \infty . \quad (1.49)$$

This follows from the uniform bound since for every $f \in C(E)$

$$\|\mathcal{L}f\| = \sup_{x \in E} \mathcal{L}f(x) \leq 2\|f\| \sup_{x \in E} \sum_{y \in E} c(x, y) = 2\|f\| \bar{c} < \infty . \quad (1.50)$$

In particular, indicator functions $f = \mathbb{1}_x : E \rightarrow \{0, 1\}$ are always in $C(E)$ (since we use the discrete topology) and we have

$$\int_E P_t \mathbb{1}_x d\mu = \sum_{y \in E} (P_t \mathbb{1}_x)(y) \mu(y) = \mathbb{P}^\mu[X_t = x] =: \mu_t(x) \quad (1.51)$$

for the distribution at time t with $\mu_0(x) = \mu(x)$. Using this and (1.46) we get for the right-hand side of the backward equation (1.25) for all $x \in E$

$$\begin{aligned} \int_E \mathcal{L} P_t \mathbb{1}_x d\mu &= \sum_{y \in E} \mu(y) \sum_{z \in E} c(y, z) (P_t \mathbb{1}_x(z) - P_t \mathbb{1}_x(y)) = \\ &= \sum_{y \in E} \mu_t(y) \left(c(y, x) - \mathbb{1}_x(y) \sum_{z \in E} c(y, z) \right) = \\ &= \sum_{y \in E} \mu_t(y) \left(c(y, x) - \mu_t(x) \sum_{z \in E} c(x, z) \right), \end{aligned} \quad (1.52)$$

where in this particular computation only we use the convention $c(x, x) = 0$, $x \in E$ for simplicity. In summary, we get

$$\frac{d}{dt} \mu_t(x) = \sum_{y \neq x} \left(\mu_t(y) c(y, x) - \mu_t(x) c(x, y) \right), \quad \mu_0(x) = \mu(x). \quad (1.53)$$

This is called the **master equation**, with intuitive gain and loss terms into state x on the right-hand side. It makes sense only for countable E , and in that case it is actually equivalent to (1.25), since the indicator functions form a basis of $C(E)$.

If the state space is finite, the semigroup is simply given by the standard matrix exponential

$$P_t = e^{tQ} = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k.$$

Let $\lambda_i \in \mathbb{C}$, $i = 1, \dots, |E|$ be the eigenvalues of Q with corresponding left and right eigenvectors $\langle v_i |$ and $|w_i\rangle$, respectively. Assume that all λ_i are distinct, then eigenvectors are orthogonal and we can normalize them such that they form an orthonormal basis with $\langle v_i | w_j \rangle = \delta_{ij}$. Then

$$Q = \sum_{i=1}^{|E|} \lambda_i |w_i\rangle \langle v_i|, \quad \text{and} \quad P_t = \sum_{i=1}^{|E|} e^{\lambda_i t} |w_i\rangle \langle v_i|,$$

with matrices $|w_i\rangle \langle v_i|$ projecting on the eigenspace of λ_i . This also provides a solution to the master equation (1.53) written as a row vector $\langle \mu_t | = (\mu_t(x) : x \in E)$,

$$\langle \mu_t | = \langle \mu_0 | P_t = \sum_{i=1}^{|E|} \langle \mu_0 | w_i \rangle e^{\lambda_i t} \langle v_i |. \quad (1.54)$$

Since $Q|1\rangle = |0\rangle$, we know that $\lambda_1 = 0$ is an eigenvalue with $|w_1\rangle = |1\rangle$. The Perron-Frobenius theorem then implies that $\langle v_1 |$ has positive entries, and all other non-zero eigenvalues have negative real part $\text{Re}(\lambda_i) < 0$. Therefore, using $\langle \mu_0 | w_1 \rangle = 1$,

$$\langle \mu_t | \rightarrow \langle v_1 | \quad \text{as } t \rightarrow \infty, \quad (1.55)$$

which is the stationary distribution of the Markov chain. We will get back to stationary distributions in Section 2.

1.2.2 Probabilistic construction

Now we would like to get an understanding of the time evolution of a sample path and the role of the transition rates. For a process with $X_0 = x$, we denote by

$$W_x := \inf\{t \geq 0 : X_t \neq x\} \quad (1.56)$$

the **holding time** in state x . Its distribution is related to the total exit rate out of state x , $c(x) = \sum_{y \neq x} c(x, y)$, which we assume to be uniformly bounded (1.49). If $c(x) = 0$, x is an absorbing state and $W_x = \infty$ a.s. .

Proposition 1.6 *If $c(x) \in (0, \infty)$, $W_x \sim \text{Exp}(c(x))$ and $\mathbb{P}^x(X_{W_x} = y) = c(x, y)/c(x)$.*

Proof. W_x has the 'loss of memory' property

$$\mathbb{P}^x(W_x > s + t | W_x > s) = \mathbb{P}^x(W_x > s + t | X_s = x) = \mathbb{P}^x(W_x > t), \quad (1.57)$$

the distribution of the holding time W_x does not depend on how much time the process has already spent in state x . Thus

$$\mathbb{P}^x(W_x > s + t, W_x > s) = \mathbb{P}^x(W_x > s + t) = \mathbb{P}^x(W_x > s) \mathbb{P}^x(W_x > t). \quad (1.58)$$

Analogous to the Chapman-Kolmogorov equations, this is solved by an exponential (using also continuity in time) and implies that

$$\mathbb{P}^x(W_x > t) = e^{-\lambda t} \quad (\text{with initial condition } \mathbb{P}^x(W_x > 0) = 1). \quad (1.59)$$

The exponent is given by

$$\lambda = \left. \frac{d}{dt} \mathbb{P}^x(W_x > t) \right|_{t=0} = \lim_{t \searrow 0} \frac{\mathbb{P}^x(W_x > t) - 1}{t} = -c(x), \quad (1.60)$$

since with (1.45)

$$\mathbb{P}^x(W_x > 0) = 1 - \mathbb{P}^x(X_t \neq x) + o(t) = 1 - c(x)t + o(t). \quad (1.61)$$

Now, conditioned on a jump occurring in the time interval $[t, t + h)$ we have

$$\mathbb{P}^x(X_{t+h} = y | t \leq W_x < t + h) = \mathbb{P}^x(X_h = y | W_x < h) = \frac{\mathbb{P}^x(X_h = y)}{\mathbb{P}^x(W_x < h)} \rightarrow \frac{c(x, y)}{c(x)} \quad (1.62)$$

as $h \searrow 0$, using the Markov property and L'Hopital's rule with (1.45) and (1.60). With right-continuity of paths, this implies the second statement. \square

It can be shown that every Feller process and every Markov chain fulfils the **strong Markov property**. That is, if $Y_s(\omega)$ is jointly measurable on $[0, \infty) \times \Omega$, and τ is a stopping time¹, then for all $x \in E$

$$\mathbb{E}^x[Y_\tau \circ \theta_\tau | \mathcal{F}_\tau] = \mathbb{E}^{X_\tau}[Y_\tau] \quad \mathbb{P}^x - a.s. \quad \text{on } \{\tau < \infty\}.$$

¹i.e. $\{\tau(\omega) \leq t\} \in \mathcal{F}_t$

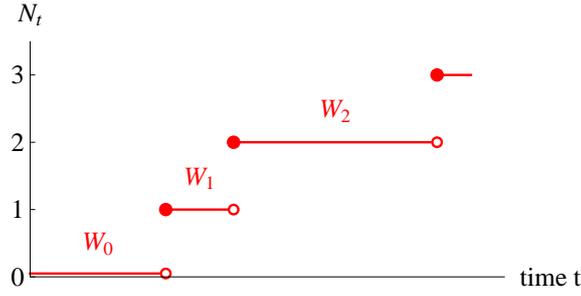


Figure 1: Sample path (càdlàg) of a Poisson process with holding times W_0, W_1, \dots

For a proof see e.g. [31], Theorem 1.68 (as well as Theorem 2.16 and Theorem 3.3). This enables us to construct sample paths of Markov chains iteratively, as is illustrated in Figure 1. Paths are piecewise constant, right-continuous functions, and jumps occur at **jump times**, defined recursively as

$$J_0 = 0 \quad \text{and} \quad J_{n+1} = \inf\{t > J_n : X_t \neq X_{J_n}\}.$$

Note that jump times are stopping times due to right-continuity of paths.

We summarize some important properties of exponential random variables, the proof of which can be found in any standard textbook. Let W_1, W_2, \dots be a sequence of independent exponentials $W_i \sim \text{Exp}(\lambda_i)$. Then $\mathbb{E}(W_i) = 1/\lambda_i$ and

$$\min\{W_1, \dots, W_n\} \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right). \quad (1.63)$$

The sum of iid exponentials with $\lambda_i = \lambda$ is Γ -distributed, i.e.

$$\sum_{i=1}^n W_i \sim \Gamma(n, \lambda) \quad \text{with PDF} \quad \frac{\lambda^n w^{n-1}}{(n-1)!} e^{-\lambda w}. \quad (1.64)$$

Example. The **Poisson process** $(N_t : t \geq 0)$ with rate $\lambda > 0$ (short $PP(\lambda)$) is a Markov chain with $X = \mathbb{N} = \{0, 1, \dots\}$, $N_0 = 0$ and $c(n, m) = \lambda \delta_{n+1, m}$.

For the Poisson process $(N_t : t \geq 0) \sim PP(\lambda)$ the holding times are iidrv's $W_i \sim \text{Exp}(\lambda)$ and we can write $N_t = \max\{n : \sum_{i=1}^n W_i \leq t\}$. This implies

$$\begin{aligned} \mathbb{P}[N_t = n] &= \mathbb{P}\left[\sum_{i=1}^n W_i \leq t < \sum_{i=1}^{n+1} W_i\right] = \int_0^t \mathbb{P}\left[\sum_{i=1}^n W_i \in ds\right] \mathbb{P}(W_{n+1} > t-s) = \\ &= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} e^{-\lambda(t-s)} ds = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \end{aligned} \quad (1.65)$$

so $N_t \sim \text{Poi}(\lambda t)$ has a Poisson distribution. Alternatively a Poisson process can be characterized by the following.

Proposition 1.7 $(N_t : t \geq 0) \sim PP(\lambda)$ if and only if it has stationary, independent increments, i.e.

$$N_{t+s} - N_s \sim N_t - N_0 \quad \text{and} \quad N_{t+s} - N_s \quad \text{independent of} \quad (N_u : u \leq s), \quad (1.66)$$

and for each t , $N_t \sim \text{Poi}(\lambda t)$.

Proof. By the loss of memory property and (1.65) increments have the distribution

$$N_{t+s} - N_s \sim Poi(\lambda t) \quad \text{for all } s \geq 0, \quad (1.67)$$

and are independent of N_s which is enough together with the Markov property.

The other direction follows by deriving the jump rates from the properties in (1.66) using (1.45).

□

Remember that for independent Poisson variables Y_1, Y_2, \dots with $Y_i \sim Poi(\lambda_i)$ we have $\mathbb{E}[Y_i] = Var[Y_i] = \lambda_i$ and

$$\sum_{i=1}^n Y_i \sim Poi\left(\sum_{i=1}^n \lambda_i\right). \quad (1.68)$$

With Prop. 1.7 this immediately implies that adding a finite number of independent Poisson processes $(N_t^i : t \geq 0) \sim PP(\lambda_i), i = 1, \dots, n$ results in a Poisson process, i.e.

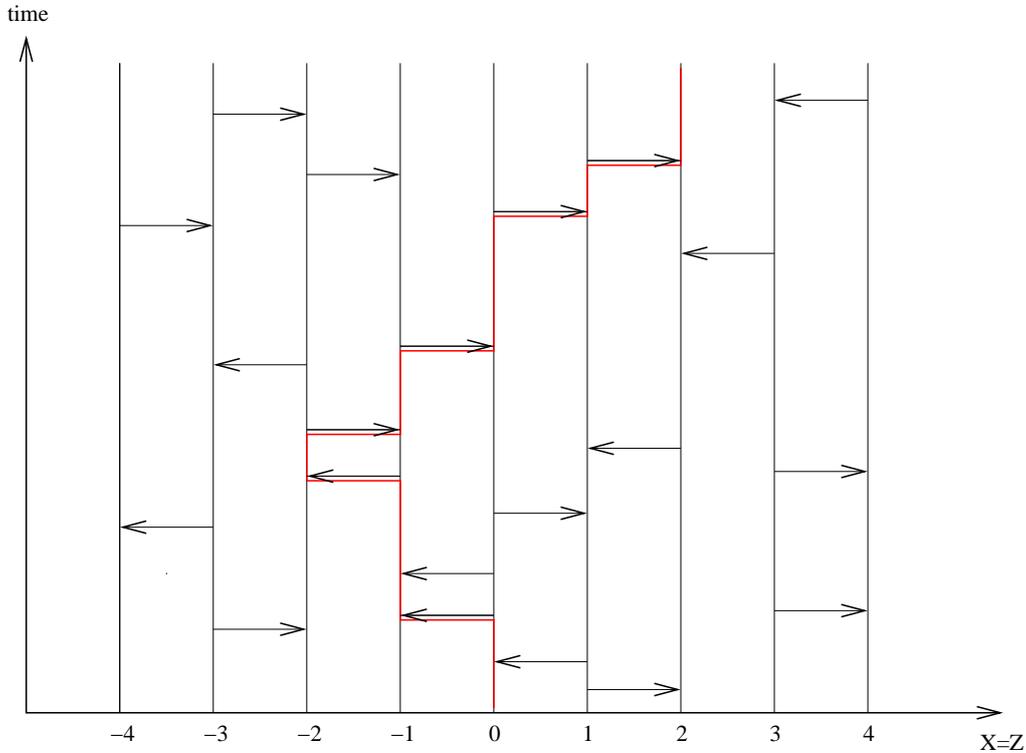
$$N_t = \sum_{i=1}^n N_t^i \Rightarrow (N_t : t \geq 0) \sim PP\left(\sum_{i=1}^n \lambda_i\right). \quad (1.69)$$

In the following we will identify a sample path of the Poisson process $t \mapsto (N_t(\omega))$ with the set of jump times $\{J_n(\omega) : n \in \mathbb{N}\}$ which is a locally finite subset of $[0, \infty)$. This second interpretation can be generalized to higher dimensions in the form of **Poisson point processes**.

Example. A continuous-time simple random walk $(X_t : t \geq 0)$ on $E = \mathbb{Z}$ with jump rates p to the right and q to the left is given by

$$X_t = R_t - L_t \quad \text{where } (R_t : t \geq 0) \sim PP(p), (L_t : t \geq 0) \sim PP(q). \quad (1.70)$$

The process can be constructed by the following graphical representation:



In each column the arrows $\rightarrow \sim PP(p)$ and $\leftarrow \sim PP(q)$ are independent Poisson (point) processes. Together with the initial condition, the trajectory of the process shown in red is then uniquely determined. An analogous construction is possible for a general Markov chain, which is a continuous time random walk on E with jump rates $c(x, y)$. In this way we can also construct interacting random walks and more general stochastic particle systems, as is shown in the next section.

Note that the graphical construction provides a sample $\omega \in D[0, \infty)$ of a path with distribution \mathbb{P}^x for each initial condition $x \in E$. It can also be used to construct a coupling of paths with different initial distributions. For the construction and the process to be well defined on a countably infinite state space E , a sufficient condition are uniformly bounded jump rates $\bar{c} < \infty$ (1.49) which also guarantee the Feller property of the process.

For unbounded rates the Markov chain may explode, i.e. sample paths leave the state space after exhibiting infinitely many jumps within finite time (see e.g. [11], Section 2.7). The explosion time is then given by the limit of jump times

$$J_\infty := \lim_{n \rightarrow \infty} J_n \in [0, \infty],$$

where $J_\infty = \infty$ corresponds to no explosion. A standard example is a pure birth process on $E = \mathbb{N}_0$ with rates $c(x, y) = \delta_{y, x+1}(x+1)^\alpha$, $\alpha > 0$. We can compute the expected explosion time

$$\mathbb{E}^0[J_\infty] = \mathbb{E}^0 \left[\sum_{n=0}^{\infty} W_n \right] = \sum_{n=0}^{\infty} (n+1)^{-\alpha},$$

which is finite if and only if $\alpha > 1$. In that case $J_\infty < \infty$ $\mathbb{P}^0 - a.s.$ and the process explodes.

1.3 Stochastic particle systems

1.3.1 Simple examples of IPS

For the stochastic particle systems (IPS for interacting particle systems) $(\eta_t : t \geq 0)$ introduced in this section the state space is of the form $E = \{0, 1\}^\Lambda$, with **particle configurations** $\eta = (\eta(i) : i \in \Lambda)$. $\eta(i) = 1$ means that there is a particle at site i and if $\eta(i) = 0$ site i is empty. The lattice Λ can be any countable set, typical examples we have in mind are regular lattices $\Lambda = \mathbb{Z}^d$, subsets of those, or the vertex set of a given graph.

If Λ is infinite E is uncountable, so we are not necessarily dealing with Markov chains in this section. But for the processes we consider the particles move/interact only locally and one at a time, so a description with jump rates still makes sense. More specifically, for a given $\eta \in E$ there are only countably many η' for which $c(\eta, \eta') > 0$. Define the configurations η^i and $\eta^{ij} \in E$ for $i \neq j \in \Lambda$ by

$$\eta^i(k) = \begin{cases} \eta(k) & , k \neq i \\ 1 - \eta(k) & , k = i \end{cases} \quad \text{and} \quad \eta^{ij}(k) = \begin{cases} \eta(k) & , k \neq i, j \\ \eta(k) - 1 & , k = i \\ \eta(k) + 1 & , k = j \end{cases}, \quad (1.71)$$

so that η^i corresponds to creation/annihilation of a particle at site i , and η^{ij} to motion of a particle from i to j . Then following standard notation we write for the corresponding jump rates

$$c(i, \eta) = c(\eta, \eta^i) \quad \text{and} \quad c(i, j, \eta) = c(\eta, \eta^{ij}). \quad (1.72)$$

All other jump rates including e.g. multi-particle interactions or simultaneous motion are zero. In the following, let $p(i, j) \geq 0$, $i, j \in \Lambda$, be transition rates of an irreducible continuous-time random walk on Λ .

Definition 1.3 The **exclusion process (EP)** on E is characterized by the jump rates

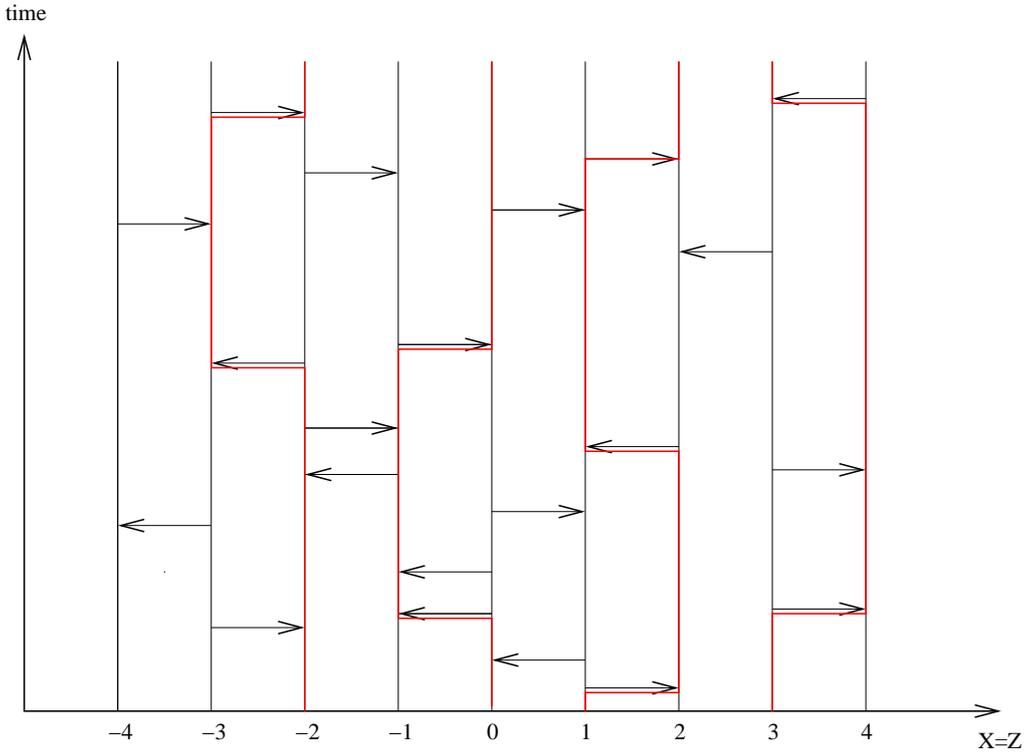
$$c(i, j, \eta) = p(i, j)\eta(i)(1 - \eta(j)) , \quad i, j \in \Lambda \quad (1.73)$$

where particles only jump to empty sites (exclusion interaction). If Λ is a regular lattice and $p(i, j) > 0$ only if i and j are nearest neighbours, the process is called **simple EP (SEP)**. If in addition $p(i, j) = p(j, i)$ for all $i, j \in \Lambda$ it is called **symmetric SEP (SSEP)** and otherwise **asymmetric SEP (ASEP)**.

Note that the presence of a direct connection (or directed edge) (i, j) is characterized by $p(i, j) > 0$, and irreducibility of $p(i, j)$ is equivalent to $(\Lambda, (p(i, j) : i, j \in \Lambda))$ being strongly connected graph. Particles only move and are not created or annihilated, therefore the number of particles in the system is conserved in time. In general such IPS are called **lattice gases**. The ASEP in one dimension $d = 1$ is one of the most basic and most studied models in IPS and nonequilibrium statistical mechanics (see e.g. [27] and references therein), and a common quick way of defining it is

$$10 \xrightarrow{p} 01 , \quad 01 \xrightarrow{q} 10 \quad (1.74)$$

where particles jump to the right (left) with rate p (q).



The graphical construction is analogous to the single particle process given above, with the additional constraint of the exclusion interaction.

Definition 1.4 The **contact process (CP)** on E is characterized by the jump rates

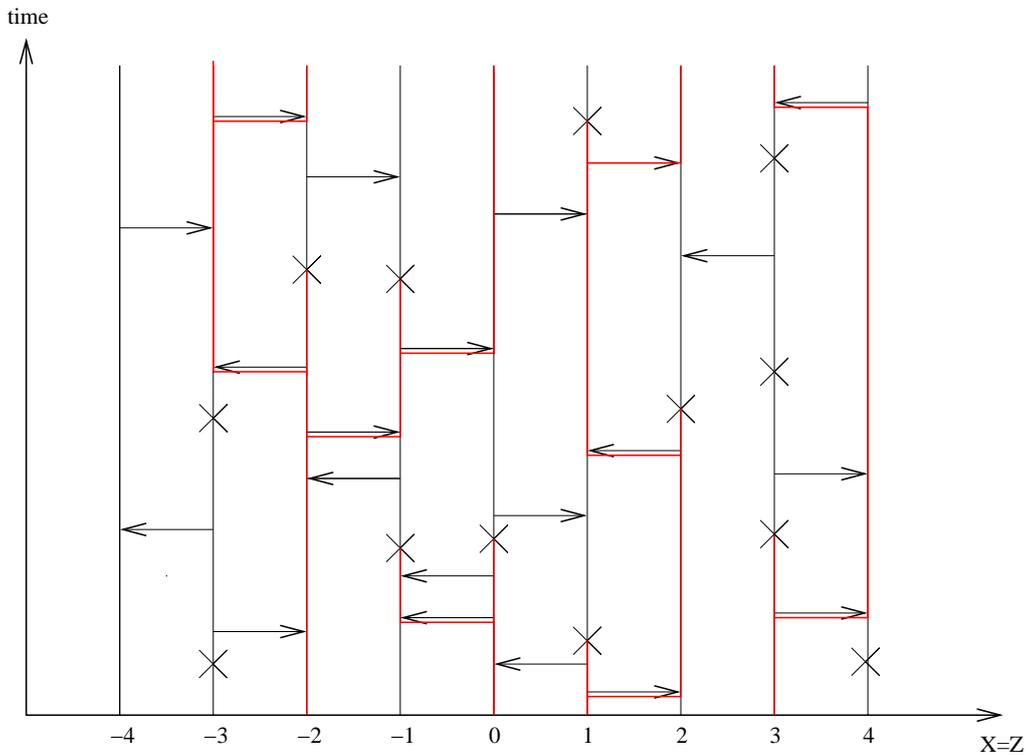
$$c(i, \eta) = \eta(i) + (1 - \eta(i)) \sum_{j \in E} p(j, i) \eta(j) \quad (1.75)$$

Particles can be interpreted as infected sites which recover with rate 1 and are infected independently with rate $p(j, i) > 0$ by an infected neighbour.

In contrast to the EP the CP does not have a conserved quantity like the number of particles, but it does have an absorbing state $\eta \equiv 0$, since there is no spontaneous infection. Often $p(i, j) \in \{0, \lambda\}$ with constant infection rate $\lambda > 0$ in case to individuals are connected. A compact notation for the CP is then

$$1 \xrightarrow{1} 0, \quad 01 \xrightarrow{\lambda} 11. \quad (1.76)$$

The graphical construction below contains now a third independent Poisson process $\times \sim PP(1)$ on each line marking the recovery events. The infection events are marked by the independent $PP(\lambda)$ Poisson processes \rightarrow and \leftarrow .



Systems with flip dynamics as above are often called **spin systems**, further examples include kinetic Ising models or voter models (see e.g. [31], Section 4).

1.3.2 General facts on IPS

For general IPS the local state space $\{0, 1\}$ is replaced by a general subset $S \subseteq \mathbb{N}$. Note that the state space $E = S^\Lambda$ is uncountable if Λ is countably infinite, but compact, provided that the local state space $S \subseteq \mathbb{N}$ is finite as in the two examples above, which will be useful later. The discrete

topology on the local state space S is simply given by the power set $\mathcal{P}(S)$, i.e. all subsets are open. The **product topology** σ on E is then given by the smallest topology such that all the canonical projections $\eta(x) : E \rightarrow S$ (occupation at a site x for a given configuration η) are continuous. That means that σ is generated by sets

$$\eta(i)^{-1}(U) = \{\eta : \eta(i) \in U\}, \quad U \subseteq S, \quad (1.77)$$

which are called **open cylinders**. Finite intersections of these sets

$$\{\eta : \eta(i_1) \in U_1, \dots, \eta(i_n) \in U_n\}, \quad n \in \mathbb{N}, U_i \subseteq S \quad (1.78)$$

are called **cylinder sets**, and any open set on E is a (finite or infinite) union of cylinder sets. Clearly if S is compact (i.e. finite), by **Tychonoff's theorem** any product of compact topological spaces is compact (w.r.t. the product topology). This holds for any countable lattice or vertex set Λ .

We can formally write down expressions for the generator similar to Markov chains (1.46). For a lattice gas with $S = \{0, 1\}$ (e.g. SEP) we have

$$\mathcal{L}f(\eta) = \sum_{x,y \in \Lambda} c(x,y,\eta)(f(\eta^{xy}) - f(\eta)), \quad (1.79)$$

and for spin systems like the CP

$$\mathcal{L}f(\eta) = \sum_{x \in \Lambda} c(x,\eta)(f(\eta^x) - f(\eta)). \quad (1.80)$$

For infinite lattices Λ convergence of the sums is an issue and we have to find a proper domain $D_{\mathcal{L}}$ of functions f for which they are finite.

Definition 1.5 For $E = S^\Lambda$ with $S \subseteq \mathbb{N}$, $f \in C(E)$ is a **cylinder function** if there exists a finite subset $\Delta_f \subseteq \Lambda$ such that

$$f(\eta) = f(\zeta) \quad \text{for all } \eta, \zeta \in E \quad \text{with } \eta(i) = \zeta(i) \text{ for all } i \in \Delta_f, \quad (1.81)$$

i.e. f depends only on a finite set of coordinates of a configuration (i.e. it is constant on particular cylinder sets). We write $C_0(E) \subseteq C(E)$ for the set of all cylinder functions.

Examples. The indicator function $\mathbb{1}_\eta$ is in general not a cylinder function (only on finite lattices, where it is also continuous), whereas the local particle number $\eta(i)$ or the product $\eta(i)\eta(j)$ are. These functions are important observables, and their expectations correspond to **local densities**

$$\rho(t, i) := \mathbb{E}^\mu(\eta_t(i)) \quad (1.82)$$

and **two-point correlation functions**

$$\theta(t, i, j) := \mathbb{E}^\mu(\eta_t(i)\eta_t(j)). \quad (1.83)$$

For $f \in C_0(E)$ the sum (1.80) contains only finitely many non-zero terms, so converges for any given η . However, we need $\mathcal{L}f$ to be finite w.r.t. the sup-norm of our Banach space $(C(E), \|\cdot\|)$. To assure this, we also need to impose some regularity conditions on the jump rates. For simplicity we will assume them to be of finite range as explained below. This is much more than is necessary, but it is easy to work with and fulfilled by all the examples we consider. In general, the independence of cylinder functions f and jump rates c on coordinates i outside a finite range $\Delta \subseteq \Lambda$ can be replaced by a weak dependence on coordinates $i \notin \Delta$ decaying with increasing Δ (see e.g. [6] Sections I.3 and VIII.0 for a more general discussion).

Definition 1.6 The jump rates of an IPS on $E = \{0, 1\}^\Lambda$ are said to be of **finite range** $R > 0$ if for all $i \in \Lambda$ there exists a finite $\Delta \subseteq \Lambda$ with $|\Delta| \leq R$ such that

$$c(i, \eta^k) = c(i, \eta) \quad \text{for all } \eta \in E \text{ and } k \notin \Delta. \quad (1.84)$$

in case of a spin system. For a lattice gas the same should hold for the rates $c(i, j, \eta)$ for all $j \in \Lambda$, with the additional requirement

$$\left| \{j \in \Lambda : c(i, j, \eta) > 0\} \right| \leq R \quad \text{for all } \eta \in E \text{ and } i \in \Lambda. \quad (1.85)$$

Proposition 1.8 *Under the condition of finite range jump rates, $\|\mathcal{L}f\| < \infty$ for all $f \in C_0(E)$. Furthermore, $C_0(E)$ is a core for the operators \mathcal{L} defined in (1.79) and (1.80), and their unique extensions are Markov generators in the sense of Prop. 1.3.*

Proof. Consider a spin system with rates $c(i, \eta)$ of finite range R . Then for each $i \in \Lambda$, $c(i, \eta)$ assumes only a finite number of values (at most 2^R), and therefore $\bar{c}(i) = \sup_{\eta \in E} c(i, \eta) < \infty$. Then we have for $f \in C_0(E)$, depending on coordinates in $\Delta_f \subseteq \Lambda$,

$$\begin{aligned} \|\mathcal{L}f\| &\leq 2\|f\| \sup_{\eta \in E} \sum_{i \in \Delta_f} c(i, \eta) \leq 2\|f\| \sum_{i \in \Delta_f} \sup_{\eta \in E} c(i, \eta) \leq \\ &\leq 2\|f\| \sum_{i \in \Delta_f} \bar{c}(i) < \infty, \end{aligned} \quad (1.86)$$

since the last sum is finite with finite summands. A similar computation works for lattice gases. The proof of the second statement is more involved, see e.g. [6], Theorem I.3.9. Among other points, this involves choosing a proper metric such that $C_0(E)$ is dense in $C(E)$, which is not the case for the one induced by the sup-norm. Details can also be found in Section 4.2 of [31]. \square

Generators are linear operators and Prop. 1.3 then implies that the sum of two or more generators is again a Markov generator, modulo technicalities regarding domains, which can be substantial on infinite lattices or for general Feller processes (see e.g. [10]). In that way we can define more general processes, e.g. a sum of (1.79) and (1.80) could define a contact process with nearest-neighbour particle motion. As mentioned before, on finite lattices with compact S we have a finite state space E , and the domain of all generators is simply $C(E)$.

2 Stationary measures, symmetries and time reversal

2.1 Stationary and reversible measures, currents

We are back to general Markov processes $(X_t : t \geq 0)$ on complete separable metric spaces E .

Definition 2.1 For a process $(P_t : t \geq 0)$ with initial distribution μ we denote by μ_t or $\mu P_t \in \mathcal{M}_1(E)$ the *distribution at time t* (previously μ_t), which is uniquely determined by

$$\int_E f d[\mu P_t] := \int_E P_t f d\mu \quad \text{for all } f \in C(E). \quad (2.1)$$

The notation μP_t is a convention from functional analysis, where one often writes

$$(\mu, P_t f) := \int_E P_t f d\mu = (P_t^\dagger \mu, f) = (\mu P_t, f). \quad (2.2)$$

The distribution μ is in fact evolved by the adjoint operator P_t^\dagger , and we usually denote it by $P_t^\dagger \mu = \mu P_t$ which is also reminiscent of the notation for Markov chains in (1.54). The fact that μP_t is uniquely specified by (2.1) is due to the standard duality of measures and functions $C(E)$ and the Riesz representation theorem (see e.g. [13], Theorem 2.14).

Definition 2.2 A measure $\mu \in \mathcal{M}_1(E)$ is **stationary** or **invariant** if $\mu P_t = \mu$ or, equivalently,

$$\int_E P_t f d\mu = \int_E f d\mu \quad \text{or shorter} \quad \mu(P_t f) = \mu(f) \quad \text{for all } f \in C(E). \quad (2.3)$$

The set of all invariant measures of a process is denoted by \mathcal{I} . A measure μ is called **reversible** if

$$\mu(f P_t g) = \mu(g P_t f) \quad \text{for all } f, g \in C(E). \quad (2.4)$$

To simplify notation here and in the following we use the standard notation $\mu(f) = \int_E f d\mu$ for integration, which is the expected value w.r.t. the measure μ on state space E . We use the symbol \mathbb{E} only for expectations on path space w.r.t. the measure \mathbb{P} .

Taking $g = 1$ in (2.4) we see that every reversible measure is also stationary.

Proposition 2.1 Consider a Feller process on state space E with generator \mathcal{L} . Then

$$\mu \in \mathcal{I} \quad \Leftrightarrow \quad \mu(\mathcal{L}f) = 0 \quad \text{for all } f \in \mathcal{D}_{\mathcal{L}} \text{ (or a suitable core)}, \quad (2.5)$$

and similarly

$$\mu \text{ is reversible} \quad \Leftrightarrow \quad \mu(f \mathcal{L}g) = \mu(g \mathcal{L}f) \quad \text{for all } f, g \in \mathcal{D}_{\mathcal{L}}. \quad (2.6)$$

Proof. The correspondence between semigroups and generators in the is given Hille-Yosida theorem in terms of limits in (1.17) and (1.22). By strong continuity of P_t in $t = 0$ and restricting to $f \in \mathcal{D}_{\mathcal{L}}$ we can re-write the conditions as

$$\mathcal{L}f := \lim_{n \rightarrow \infty} \underbrace{\frac{P_{1/n}f - f}{1/n}}_{:=h_n} \quad \text{and} \quad P_t f := \lim_{n \rightarrow \infty} \left(Id - \frac{t}{n} \mathcal{L} \right)^{-n} f. \quad (2.7)$$

Now $\mu \in \mathcal{I}$ implies that for all $n \in \mathbb{N}$

$$\mu(P_{1/n}f) = \mu(f) \quad \Rightarrow \quad \mu(h_n) = 0. \quad (2.8)$$

Then we have

$$\mu(\mathcal{L}f) = \mu\left(\lim_{n \rightarrow \infty} h_n\right) = \lim_{n \rightarrow \infty} \mu(h_n) = 0, \quad (2.9)$$

by bounded (or dominated) convergence, since h_n converges in $(C(E), \|\cdot\|)$ as long as $f \in \mathcal{D}_{\mathcal{L}}$, and we have $\mu(E) = 1$.

On the other hand, if $\mu(\mathcal{L}f) = 0$ for $f \in \mathcal{D}_L$ and we write $h = f - \lambda\mathcal{L}f$ then $\mu(f) = \mu(h)$. Rewriting this we have

$$\mu((\mathbb{I} - \lambda\mathcal{L})^{-1}h) = \mu(h), \quad \text{for all } h \in \mathcal{D}_L \text{ and } \lambda \geq 0.$$

Iterating this n times with $\lambda = 1/n$ and taking $n \rightarrow \infty$ implies $\mu \in \mathcal{I}$, since for all $t \geq 0$

$$\mu(P_t h) = \lim_{n \rightarrow \infty} \mu\left((\mathbb{I} - t\mathcal{L}/n)^{-n}\right) = \mu(h).$$

This finishes the proof of (2.5), a completely analogous argument works for the equivalence (2.6) on reversibility. \square

It is well known for Markov chains that on a finite state space there exists at least one stationary distribution, namely the left eigenvector $\langle v_1 | Q = \langle 0 |$ in (1.55). For IPS compactness of the state spaces E ensures a similar result.

Theorem 2.2 *For every Feller process with compact state space E we have:*

(a) \mathcal{I} is non-empty, (weakly) compact and convex.

(b) Suppose the weak limit $\mu = \lim_{t \rightarrow \infty} \pi P_t$ exists for some initial distribution $\pi \in \mathcal{M}_1(E)$, i.e.

$$\pi P_t(f) = \int_E P_t f d\pi \rightarrow \mu(f) \quad \text{for all } f \in C(E), \quad (2.10)$$

then $\mu \in \mathcal{I}$.

Proof. (a) Convexity of \mathcal{I} follows directly from two basic facts. Firstly, a convex combination of two probability measures $\mu_1, \mu_2 \in \mathcal{M}_1(E)$ is again a probability measure, i.e.

$$\nu := \lambda\mu_1 + (1 - \lambda)\mu_2 \in \mathcal{M}_1(E) \quad \text{for all } \lambda \in [0, 1]. \quad (2.11)$$

Secondly, the stationarity condition (2.5) is linear, i.e. if $\mu_1, \mu_2 \in \mathcal{I}$ then so is ν since

$$\nu(\mathcal{L}f) = \lambda\mu_1(\mathcal{L}f) + (1 - \lambda)\mu_2(\mathcal{L}f) = 0 \quad \text{for all } f \in C(E). \quad (2.12)$$

\mathcal{I} is a weakly closed subset of $\mathcal{M}_1(E)$ if we have

$$\mu_1, \mu_2, \dots \in \mathcal{I}, \quad \mu_n \rightarrow \mu \text{ weakly, implies } \mu \in \mathcal{I}. \quad (2.13)$$

But this is immediate by weak convergence, since for all $f \in C(E)$

$$\mu_n(\mathcal{L}f) = 0 \quad \text{for all } n \in \mathbb{N} \quad \Rightarrow \quad \mu(\mathcal{L}f) = \lim_{n \rightarrow \infty} \mu_n(\mathcal{L}f) = 0. \quad (2.14)$$

Under the topology of weak convergence $\mathcal{M}_1(E)$ is compact since E is compact¹, and therefore also $\mathcal{I} \subseteq \mathcal{M}_1(E)$ is compact since it is a closed subset of a convex set.

Non-emptiness: Define

$$\pi_n = \frac{1}{n} \int_0^n \pi T_s ds \quad \text{for some } \pi \in \mathcal{M}_1(E) .$$

Then we can show analogously to (b) below that

$$\pi_n(f) - \pi_n(T_t f) \rightarrow 0 \quad \text{as } n \rightarrow \infty . \quad (2.15)$$

Then by Prohorov's theorem, there exists a subsequence n_k such that $\pi_{n_k} \Rightarrow \mu$ for some $\mu \in \mathcal{M}_1(E)$. Since $T_t f \in C(E)$ we can pass to the limit in (2.15) along the subsequence to get $\mu(f) = \mu(T_t f)$ so that $\mu \in \mathcal{I}$.

(b) Let $\mu := \lim_{t \rightarrow \infty} \pi P_t$. Then $\mu \in \mathcal{I}$ since for all $f \in C(E)$,

$$\begin{aligned} \mu(P_s f) &= \lim_{t \rightarrow \infty} \int_E P_s f d[\pi P_t] = \lim_{t \rightarrow \infty} \int_E P_t P_s f d\pi = \\ &= \lim_{t \rightarrow \infty} \int_E P_{t+s} f d\pi = \lim_{t \rightarrow \infty} \int_E P_t f d\pi = \lim_{t \rightarrow \infty} \int_E f d[\pi P_t] = \mu(f) . \end{aligned} \quad (2.16)$$

□

Remarks. Note that we need to take a Césaro average in (2.15), since we only have convergence along a subsequence, and the limit of a weakly convergent subsequence πP_{t_k} with $t_k \nearrow \infty$ need not be stationary.

By the Krein Milman theorem (see e.g. [14], Theorem 3.23), compactness and convexity of $\mathcal{I} \subseteq \mathcal{M}_1(E)$ implies that \mathcal{I} is the closed convex hull of its extreme points \mathcal{I}_e , which are called **extremal invariant measures**. Every invariant measure can therefore be written as a convex combination of members of \mathcal{I}_e , so the extremal measures are the ones we need to find for a given process.

Definition 2.3 A Markov process with semigroup $(P_t : t \geq 0)$ is *ergodic* if

- (a) $\mathcal{I} = \{\mu\}$ is a singleton, and (unique stationary measure)
- (b) $\lim_{t \rightarrow \infty} \pi P_t = \mu$ for all $\pi \in \mathcal{M}_1(E)$. (convergence to equilibrium)

Phase transitions are related to the breakdown of ergodicity and in particular to non-uniqueness of stationary measures. This can be the result of the presence of absorbing states (e.g. CP), or of spontaneous symmetry breaking/breaking of conservation laws (e.g. SEP or VM) as is discussed later. On finite lattices, IPS are Markov chains which are known to have a unique stationary distribution under reasonable assumptions of non-degeneracy. Therefore, mathematically phase transitions occur only in infinite systems. Infinite systems are often interpreted/studied as limits of finite systems, which show traces of a phase transition by divergence or non-analytic behaviour of certain observables. In terms of applications, infinite systems are approximations or idealizations of real systems which may be large but are always finite, so results have to be interpreted with care. There is a well developed mathematical theory of phase transitions for reversible systems provided by the framework of Gibbs measures (see e.g. [7]). But for IPS which are in general

¹For more details on weak convergence see e.g. [20], Section 2.

non-reversible, the notion of phase transitions is not unambiguous.

Example for IPS. Consider an IPS with state space $E = \{0, 1\}^\Lambda$.

Definition 2.4 For a function $\rho : \Lambda \rightarrow [0, 1]$, ν_ρ is a *Bernoulli product measure* on E if for all $k \in \mathbb{N}$, $x_1, \dots, x_k \in \Lambda$ mutually different and $n_1, \dots, n_k \in \{0, 1\}$

$$\nu_\rho[\eta(i_1) = n_1, \dots, \eta(i_k) = n_k] = \prod_{l=1}^k \nu_{\rho(i_l)}^1[\eta(i_l) = n_l], \quad (2.17)$$

where the *single-site marginals* are given by

$$\nu_{\rho(i)}^1[\eta(i) = 1] = \rho(i) \quad \text{and} \quad \nu_{\rho(i)}^1[\eta(i) = 0] = 1 - \rho(i). \quad (2.18)$$

Remark. In other words, under ν_ρ the $\eta(i)$ are independent Bernoulli random variables $\eta(i) \sim Be(\rho(i))$ with **local density** $\rho(i) = \nu(\eta(i))$. The above definition can readily be generalized to non-Bernoulli product measures.

Now, consider the TASEP on the lattice $\Lambda = \mathbb{Z}$ with generator

$$\mathcal{L}f(\eta) = \sum_{i \in \Lambda} \eta(i)(1 - \eta(i+1))(f(\eta^{i,i+1}) - f(\eta)).$$

It can be shown (see e.g. [15], Theorem 2.1) that the homogeneous product measures ν_ρ for all $\rho \in [0, 1]$ are invariant for the process. In addition, there are absorbing states of the form $\eta_k(i) = \mathbb{1}_{\{k, k+1, \dots\}}(i)$ for all $k \in \mathbb{Z}$, where $\eta_k(i) = 0$ for $i < k$ and $\eta_k(i) = 1$ for $i \geq k$. Then all measures $\delta_{\eta_k} \in \mathcal{M}_1(E)$ concentrating on those absorbing states are also invariant. It can be shown (see e.g. [31], Section 4.5) that these are all extremal measures for the TASEP

$$\mathcal{I}_e = \{\nu_\rho : \rho \in [0, 1]\} \cup \{\delta_{\eta_k} : k \in \mathbb{Z}\}.$$

This is also true for the partially asymmetric exclusion process (PASEP), for the SSEP only the homogeneous product measures are invariant. The fact that \mathcal{I} is not a singleton is related to the conservation of mass, as explained in the next section.

2.2 Conservation laws and symmetries

Definition 2.5 For a given Feller process $(P_t : t \geq 0)$ a bounded¹ linear operator $\mathcal{T} : C(E) \rightarrow C(E)$ is called a **symmetry**, if it commutes with the semigroup. So for all $t \geq 0$ we have $P_t \mathcal{T} = \mathcal{T} P_t$, i.e.

$$P_t(\mathcal{T}f)(x) = \mathcal{T}(P_t f)(x), \quad \text{for all } f \in C(E), x \in E. \quad (2.19)$$

Proposition 2.3 For a Feller process with generator \mathcal{L} , a bounded linear operator $\mathcal{T} : C(E) \rightarrow C(E)$ is a symmetry iff $\mathcal{L}\mathcal{T} = \mathcal{T}\mathcal{L}$, i.e.

$$\mathcal{L}(\mathcal{T}f)(x) = \mathcal{T}(\mathcal{L}f)(x), \quad \text{for all } f \in \mathcal{D}_{\mathcal{L}} \text{ (or a suitable core)}. \quad (2.20)$$

We denote the set of all symmetries by $\mathcal{S}(\mathcal{L})$ or simply \mathcal{S} . The symmetries form a semigroup w.r.t. composition, i.e.

$$\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{S} \quad \Rightarrow \quad \mathcal{T}_1 \mathcal{T}_2 = \mathcal{T}_1 \circ \mathcal{T}_2 \in \mathcal{S}. \quad (2.21)$$

¹ $\mathcal{T} : C(E) \rightarrow C(E)$ is bounded if there exists $B > 0$ such that for all $f \in C(E)$, $\|\mathcal{T}f\| \leq B\|f\|$.

Proof. The first part is similar to the proof of Prop. 2.1 on stationarity. Note that since \mathcal{T} is bounded, $f \in \mathcal{D}_{\mathcal{L}}$ implies $\mathcal{T}f \in \mathcal{D}_{\mathcal{L}}$ (at least on a suitable core).

For the second part, note that composition of operators is associative. Then for $\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{S}$ we have

$$\mathcal{L}(\mathcal{T}_1\mathcal{T}_2) = (\mathcal{L}\mathcal{T}_1)\mathcal{T}_2 = (\mathcal{T}_1\mathcal{L})\mathcal{T}_2 = \mathcal{T}_1(\mathcal{L}\mathcal{T}_2) = (\mathcal{T}_1\mathcal{T}_2)\mathcal{L} \quad (2.22)$$

so that $\mathcal{T}_1\mathcal{T}_2 \in \mathcal{S}$. \square

Proposition 2.4 For a bijection $\tau : E \rightarrow E$ let $\mathcal{T}f := f \circ \tau$, i.e. $\mathcal{T}f(x) = f(\tau x)$ for all $x \in E$. Then \mathcal{T} is a symmetry for the process $(P_t : t \geq 0)$ iff

$$\mathbb{E}^x[f(\tau X_t)] = P_t(f \circ \tau) = (P_t f) \circ \tau = \mathbb{E}_{\tau x}[f(X_t)] \quad \text{for all } f \in C(E). \quad (2.23)$$

We call such \mathcal{T} (or equivalently τ) **simple symmetries**. Simple symmetries are invertible and form a group.

Proof. The first statement is immediate by the definition, \mathcal{T} is bounded since $\|f \circ \tau\| = \|f\|$ and obviously linear.

In general compositions of symmetries are symmetries according to Prop. 2.3, and if $\tau_1, \tau_2 : E \rightarrow E$ are simple symmetries then the composition $\tau_1 \circ \tau_2 : E \rightarrow E$ is also a simple symmetry. A simple symmetry τ is a bijection, so it has an inverse τ^{-1} . Then we have for all $f \in C(E)$ and all $t \geq 0$

$$(P_t(f \circ \tau^{-1})) \circ \tau = P_t(f \circ \tau^{-1} \circ \tau) = P_t f \quad (2.24)$$

since $\tau \in \mathcal{S}$. Composing with τ^{-1} leads to

$$(P_t(f \circ \tau^{-1})) \circ \tau \circ \tau^{-1} = P_t(f \circ \tau^{-1}) = (P_t f) \circ \tau^{-1}, \quad (2.25)$$

so that τ^{-1} is also a simple symmetry. \square

Example. For the ASEP on $\Lambda = \mathbb{Z}$ the translations $\tau_i : E \rightarrow E$ for $i \in \Lambda$, defined by

$$(\tau_i \eta)(j) = \eta(j - i) \quad \text{for all } j \in \Lambda \quad (2.26)$$

are simple symmetries. This can be easily seen since the jump rates are invariant under translations, i.e. we have for all $i, j \in \Lambda$

$$\begin{aligned} c(i, i+1, \eta) &= p\eta(i)(1 - \eta(i+1)) = p\eta(i+j-j)(1 - \eta(i+1+j-j)) = \\ &= c(i+j, i+1+j, \tau_j \eta). \end{aligned} \quad (2.27)$$

An analogous relation holds for jumps to the left with rate $c(i, i-1, \eta) = q\eta(i)(1 - \eta(i-1))$. Note that the family $\{\tau_i : i \in \Lambda\}$ forms a group. The same symmetry holds for the ASEP on $\Lambda_L = \mathbb{Z}/L\mathbb{Z}$ with periodic boundary conditions, where there are only L distinct translations τ_i for $i = 0, \dots, L-1$ (since e.g. $\tau_L = \tau_0$ etc.). The argument using invariance of the jump rates can be made more general.

Proposition 2.5 Consider a Feller process characterized by jump rates $c(x, y)$ such as a Markov chain or a IPS. Then a bijection $\tau : E \rightarrow E$ is a simple symmetry iff

$$c(x, y) = c(\tau x, \tau y) \quad \text{for all } x, y \in E. \quad (2.28)$$

Proof. Assuming the invariance of the jump rates, we have for all $f \in C_0(E)$ and $x \in E$

$$\begin{aligned} (\mathcal{L}(\mathcal{T}f))(x) &= (\mathcal{L}(f \circ \tau))(x) = \sum_{y \in E} c(x, y)(f(\tau y) - f(\tau x)) = \\ &= \sum_{y \in E} c(\tau x, \tau y)(f(\tau y) - f(\tau x)) = \sum_{z \in E} c(\tau x, z)(f(z) - f(\tau x)) = \\ &= (\mathcal{L}f)(\tau x) = (\mathcal{T}(\mathcal{L}f))(x), \end{aligned} \quad (2.29)$$

where the identity in the second line just comes from relabelling the sum which is possible since τ is bijective. On the other hand, $\mathcal{L}T = T\mathcal{L}$ implies that from the above that

$$\sum_{y \in E} c(x, y)(f(\tau y) - f(\tau x)) = \sum_{y \in E} c(\tau x, \tau y)(f(\tau y) - f(\tau x)). \quad (2.30)$$

Since this holds for all $f \in C_0(E)$ and $x \in E$ it uniquely determines that $c(x, y) = c(\tau x, \tau y)$ for all $x, y \in E$ with $x \neq y$. In fact, if there existed x, y for which this is not the case, we can plug $f = \mathbb{1}_{\tau y}$ into (2.30) which yields a contradiction. For fixed η both sums then contain only a single term, so this is even possible on infinite lattices even though $\mathbb{1}_{\tau \zeta}$ is not a cylinder function². \square

Proposition 2.6 For an observable $g \in C(E)$ define the multiplication operator $\mathcal{T}_g := g\mathbb{I}$ via

$$\mathcal{T}_g f(x) = g(x) f(x) \quad \text{for all } f \in C(E), x \in E. \quad (2.31)$$

Then \mathcal{T}_g is a symmetry for the process $(X_t : t \geq 0)$ iff $g(X_t) = g(X_0)$ for all $t > 0$. In that case \mathcal{T}_g (or equivalently g) is called a **conserved quantity**.

Proof. First note that \mathcal{T}_g is linear and bounded since $\|f\| \leq \|g\|\|f\|$. If $g(X_t) = g(X_0)$ we have for all $t > 0$, $f \in C(E)$ and $x \in E$

$$(P_t(\mathcal{T}_g f))(x) = \mathbb{E}^x(g(X_t) f(X_t)) = g(x)(P_t f)(x) = \mathcal{T}_g(P_t f)(x). \quad (2.32)$$

On the other hand, if \mathcal{T}_g is a symmetry the above computation implies that for all (fixed) $t > 0$

$$\mathbb{E}^x(g(X_t) f(X_t)) = \mathbb{E}^x(g(x) f(X_t)). \quad (2.33)$$

Since this holds for all $f \in C(E)$ the value of $g(X_t)$ is uniquely specified by the expected values to be $g(x)$ since g is continuous (cf. argument in (2.30)). \square

Remarks. If $g \in C(E)$ is a conserved quantity then so is $h \circ g$ for all $h : \mathbb{R} \rightarrow \mathbb{R}$ provided that $h \circ g \in C(E)$.

A subset $A \subseteq E$ is called **invariant** if $X_0 \in A$ implies $X_t \in A$ for all $t > 0$. Then $g = \mathbb{1}_A$ is a conserved quantity iff A is invariant. In general, every level set

$$E_l = \{x \in E : g(x) = l\} \subseteq E \quad \text{for all } l \in \mathbb{R}, \quad (2.34)$$

for a conserved quantity $g \in C(E)$ is invariant.

²So the function $\eta \mapsto \mathcal{L}\mathbb{1}_{\tau \zeta}(\eta)$ would in general not be well defined since it is given by an infinite sum for $\eta = \tau \zeta$. But here we are only interested in a single value for $\eta \neq \zeta$.

Examples. For the ASEP on $\Lambda_L = \mathbb{Z}/L\mathbb{Z}$ (discrete torus with periodic boundary conditions) the number of particles $\Sigma_L(\eta) := \sum_{i \in \Lambda_L} \eta(i)$ is conserved. The level sets of this integer valued function are the subsets

$$E_{L,N} = \{\eta : \Sigma_L(\eta) = N\} \quad \text{for } N = 0, \dots, L. \quad (2.35)$$

In particular the indicator functions $\mathbb{1}_{E_{L,N}}$ are conserved quantities.

Proposition 2.7 $g \in C(E)$ is a conserved quantity if and only if $\mathcal{L}g = 0$ and $\mathcal{L}g^2 = 0$.

Proof. Follows from using the martingale characterization, since

$$M_t^g := g(X_t) - g(X_0) - \int_0^t \mathcal{L}g(X_s) ds$$

is a martingale with quadratic variation given by

$$[M^g]_t = \int_0^t (\mathcal{L}g^2(X_s) - 2g(X_s)\mathcal{L}g(X_s)) ds.$$

Then $g(X_t) \equiv g(X_0)$ if and only if the quadratic variation vanishes. \square

The most important result of this section is the connection between symmetries and stationary measures. For a measure μ and a symmetry \mathcal{T} we define the measure $\mu\mathcal{T}$ via

$$(\mu\mathcal{T})(f) = \int_E f d\mu\mathcal{T} := \int_E \mathcal{T}f d\mu = \mu(\mathcal{T}f) \quad \text{for all } f \in C(E), \quad (2.36)$$

analogous to the definition of μP_t in Def. 2.1.

Theorem 2.8 For a Feller process $(P_t : t \geq 0)$ with state space E we have

$$\mu \in \mathcal{I}, \mathcal{T} \in \mathcal{S} \quad \Rightarrow \quad \frac{1}{\mu\mathcal{T}[E]} \mu\mathcal{T} \in \mathcal{I}, \quad (2.37)$$

provided that the normalization $\mu\mathcal{T}[E] \in (0, \infty)$.

Proof. For $\mu \in \mathcal{I}$ and $\mathcal{T} \in \mathcal{S}$ we have for all $t \geq 0$ and $f \in C(E)$

$$(\mu\mathcal{T})P_t(f) = \mu(\mathcal{T}P_t f) = \mu(P_t \mathcal{T}f) = \mu P_t(\mathcal{T}f) = \mu(\mathcal{T}f) = \mu\mathcal{T}(f). \quad (2.38)$$

With $\mu\mathcal{T}[E] \in (0, \infty)$, $\mu\mathcal{T}$ can be normalized and $\frac{1}{\mu\mathcal{T}[E]} \mu\mathcal{T} \in \mathcal{I}$. \square

Remarks. For $\mu \in \mathcal{I}$ it will often be the case that $\mu\mathcal{T} = \mu$ so that μ is invariant under some $\mathcal{T} \in \mathcal{S}$ and not every symmetry generates a new stationary measure. For ergodic processes $\mathcal{I} = \{\mu\}$ is a singleton, so μ has to respect all the symmetries of the process, i.e. $\mu\mathcal{T} = \mu$ for all $\mathcal{T} \in \mathcal{S}$.

If $\mathcal{T}_g = g\mathbb{1}$ is a conserved quantity, then $\mu\mathcal{T}_g = g\mu$, i.e.

$$\mu\mathcal{T}_g[A] = \int_A g(\eta) \mu(d\eta) \quad \text{for all measurable } A \subseteq E. \quad (2.39)$$

So g is the **density** of $\mu\mathcal{T}_g$ w.r.t. μ and one also writes $g = \frac{d\mu\mathcal{T}_g}{d\mu}$. This implies also that $\mu\mathcal{T}_g$ is **absolutely continuous** w.r.t. μ (short $\mu\mathcal{T}_g \ll \mu$), which means that for all measurable A , $\mu[A] = 0$ implies $\mu\mathcal{T}_g[A] = 0$ ¹.

For an invariant set $A \subseteq E$ and the conserved quantity $g = \mathbb{1}_A$ we have $\mu\mathcal{T}_g = \mathbb{1}_A\mu$. If $\mu[A] \in (0, \infty)$ the measure of Theorem (2.8) can be written as a *conditional measure*

$$\frac{1}{\mu\mathcal{T}_g[E]} \mu\mathcal{T}_g = \frac{\mathbb{1}_A}{\mu[A]} \mu =: \mu[\cdot|A] \quad (2.40)$$

concentrating on the set A , since the normalization is $\mu\mathcal{T}_g[E] = \mu(\mathbb{1}_A) = \mu[A]$.

Examples. The homogeneous product measures ν_ρ , $\rho \in [0, 1]$ are invariant under the translations τ_i , $i \in \Lambda$ for all translation invariant lattices with $\tau_i\Lambda = \Lambda$ such as $\Lambda = \mathbb{Z}$ or $\Lambda = \mathbb{Z}/L\mathbb{Z}$. But the blocking measures ν_i for $\Lambda = \mathbb{Z}$ are not translation invariant, and in fact $\nu_i = \nu_0 \circ \tau_{-i}$, so the family of blocking measures is generated from a single one by applying translations.

For $\Lambda_L = \mathbb{Z}/L\mathbb{Z}$ we have the invariant sets

$$E_{L,N} = \left\{ \eta \in E_L : \sum_{x \in \Lambda_L} \eta(x) = N \right\}$$

for a fixed number of particles $N = 0, \dots, L$. Since the ASEP is an irreducible Markov chain on $E_{L,N}$ it has a unique stationary measure $\pi_{L,N}$. Using the above remark we can write $\pi_{L,N}$ as a conditional product measure ν_ρ (which is also stationary). For all $\rho \in (0, 1)$ we have (by uniqueness of $\pi_{L,N}$)

$$\pi_{L,N} = \nu_\rho[\cdot|E_{L,N}] = \frac{\mathbb{1}_{E_{L,N}}}{\nu_\rho[E_{L,N}]} \nu_\rho, \quad (2.41)$$

where $\nu_\rho[E_{L,N}] = \binom{L}{N} \rho^N (1-\rho)^{L-N}$ is binomial (see previous section). Therefore we can compute explicitly

$$\pi_{L,N}[\eta] = \begin{cases} 0 & , \eta \notin E_{L,N} \\ \frac{\rho^N (1-\rho)^{L-N}}{\binom{L}{N} \rho^N (1-\rho)^{L-N}} = 1/\binom{L}{N} & , \eta \in E_{L,N} \end{cases}, \quad (2.42)$$

and $\pi_{L,N}$ is uniform on $E_{L,N}$, and in particular independent of ρ . We can write the product measures ν_ρ as convex combinations

$$\nu_\rho = \sum_{N=0}^L \binom{L}{N} \rho^N (1-\rho)^{L-N} \pi_{L,N}, \quad (2.43)$$

but this is not possible for the $\pi_{L,N}$ since they concentrate on irreducible subsets $E_{L,N} \subsetneq E_L$. Thus for the ASEP on $\Lambda_L = \mathbb{Z}/L\mathbb{Z}$ we have

$$\mathcal{I}_e = \{ \pi_{L,N} : N = 0, \dots, L \} \quad (2.44)$$

given by the canonical measures. So for each value of the conserved quantity Σ_L we have an extremal stationary measure and these are the only elements of \mathcal{I}_e . The latter follows from

$$E_L = \bigcup_{N=0}^L E_{L,N} \quad \text{and} \quad \text{irreducibility on each } E_{L,N}. \quad (2.45)$$

¹In fact, absolute continuity and existence of a density are equivalent by the Radon-Nikodym theorem (see e.g. [10] Thm. 2.10).

In fact, suppose that for some $\lambda \in (0, 1)$ and $\mu_1, \mu_2 \in \mathcal{I}$

$$\pi_{L,N} = \lambda\mu_1 + (1 - \lambda)\mu_2 . \quad (2.46)$$

Then for all measurable $Y \subseteq E$ with $Y \cap E_{L,N} = \emptyset$ we have

$$0 = \pi_{L,N}(Y) = \lambda\mu_1(Y) + (1 - \lambda)\mu_2(Y) , \quad (2.47)$$

which implies that $\mu_1(Y) = \mu_2(Y) = 0$. So $\mu_1, \mu_2 \in \mathcal{I}$ concentrate on $E_{L,N}$ and thus $\mu_1 = \mu_2 = \pi_{L,N}$ by uniqueness of $\pi_{L,N}$ on $E_{L,N}$. So the conservation law provides a decomposition of the state space E_L into irreducible non-communicating subsets.

2.3 Time reversal

In this section we only consider processes on a fixed finite time interval $[0, T]$. For each path $(\omega_t : t \in [0, T])$ on $D_E([0, T])$ define the time reversal $\mathcal{R} : D_E([0, T]) \rightarrow D_E([0, T])$ via

$$(\mathcal{R}\omega)_t := \omega_{(T-t)+} \quad \text{for all } t \in [0, T] , \quad (2.48)$$

so that $\mathcal{R}\omega \in D_E([0, T])$ is the time-reversal of the path ω . Note that of course time-reversal is an invertible transformation on path space with $\mathcal{R} \circ \mathcal{R} = \mathbb{I}$. Then for a given process \mathbb{P} on $D_E([0, T])$ we can define the path measure $\mathbb{P} \circ \mathcal{R}$, which is the law of the time reversed process, and ask the question if this is again a Markov process. We have to be careful and precise with initial conditions here. Under the law \mathbb{P}^x time reversed paths will start in distribution $(\mathcal{R}X)_0 \sim \mu_T$ and end up in $(\mathcal{R}X)_T = x$, which is obviously not Markovian (and strange). There are two ways of making sense of this question.

Time-reversal of stationary processes.

Let \mathbb{P} be an ergodic Feller process on E with semigroup $(P_t : t \geq 0)$ and unique stationary measure $\mu \in \mathcal{M}_1(E)$. Consider the set of test functions square integrable w.r.t. μ ,

$$L^2(E, \mu) = (f \in C(E) : \mu(f^2) < \infty) . \quad (2.49)$$

With the inner product $\langle f, g \rangle_\mu = \mu(fg)$ the closure of this (w.r.t. the metric given by the inner product) is a Hilbert space, and the semigroup P_t , for all $t \geq 0$ are bounded linear operators on $L^2(E, \mu)$. They are uniquely defined by their values on $C(E) \cap L^2(E, \mu)$, which is a dense subset of the closure of $L^2(E, \mu)$. Therefore they have an *adjoint* operator P_t^* , uniquely defined by

$$\langle P_t^* f, g \rangle_\mu = \mu(g P_t^* f) = \mu(f P_t g) = \langle f, P_t g \rangle_\mu \quad \text{for all } f, g \in L^2(E, \mu) , \quad (2.50)$$

Analogously, one can define an adjoint generator \mathcal{L}^* on a domain in $L^2(E, \mu)$ which forms a core for \mathcal{L} and \mathcal{L}^* . The domains require some thought on a case-by-case basis but is often straightforward. Note that the adjoint operators on the self-dual Hilbert space $L^2(E, \mu)$ are not the same as the adjoints L^\dagger and P_t^\dagger in (2.2) on $\mathcal{M}_1(E)$, which evolve the probability measures.

Proposition 2.9 *Consider the stationary Feller process \mathbb{P}^μ with initial condition $\mu_0 = \mu$. Then the **reversed path measure** $\mathbb{P}_{\mathcal{R}}^\mu := \mathbb{P}^\mu \circ \mathcal{R}$ is again a stationary Feller process with measure μ . Its semigroup and generator are given by the adjoint operators P_t^* and \mathcal{L}^* .*

Proof. To compute the action of the adjoint operator note that for all $g \in L^2(E, \mu)$

$$\begin{aligned} \mu(gP_t^*f) &= \int_E fP_tg d\mu = \mathbb{E}^\mu[f(\eta_0)g(\eta_t)] = \mathbb{E}^\mu\left[\mathbb{E}^\mu[f(\eta_0)|\eta_t]g(\eta_t)\right] = \\ &= \int_E \mathbb{E}^\mu[f(\eta_0)|\eta_t = \zeta]g(\zeta) \mu(d\zeta) = \mu\left(g\mathbb{E}^\mu[f(\eta_0)|\eta_t = \cdot]\right), \end{aligned} \quad (2.51)$$

where the identity between the first and second line is due to μ being the stationary measure. Since this holds for all $g \in L^2(E, \mu)$ it implies, again using stationarity

$$P_t^*f(\eta) = \mathbb{E}^\mu[f(\eta_0)|\eta_t = \eta] = \mathbb{E}^\mu[f(\eta_{T-t})|\eta_T = \eta] = \mathbb{E}_{\mathcal{R}}^\mu[f(\eta_t)|\eta_0 = \eta]. \quad (2.52)$$

So the adjoint operators ($P_t^* : t \in [0, T]$) describe the evolution of the time-reversed process $\mathbb{P}^\mu \circ \mathcal{R}$. With $g = 1$ in (2.50) we have immediately $\mu(P_t^*f) = \mu(f)$, so the reversed process is also stationary with measure μ . Similarly, it can be shown that ($P_t^* : t \geq 0$) is actually a semigroup with the adjoint generator \mathcal{L}^* . This includes some technicalities with domains of definition, see e.g. [19] and references therein. \square

The process \mathbb{P}^μ on $D([0, T])$ is called

$$\textbf{time-reversible} \quad \text{if} \quad \mathbb{P}^\mu \circ \mathcal{R} = \mathbb{P}^\mu, \quad (2.53)$$

With Prop. 2.9 this is the case if and only if $\mathcal{L} = \mathcal{L}^*$, i.e. $\mu(f\mathcal{L}g) = \mu(g\mathcal{L}f)$ with (2.50) (analogously for the semigroup). Therefore time-reversibility is equivalent to μ being reversible, i.e. \mathcal{L} and P_t being **self-adjoint** operators w.r.t. the measure μ as in (2.4) and (2.6).

For Markov chains with countable E the rates of $\mathbb{P}^\mu \circ \mathcal{R}$ are given by

$$c_{\mathcal{R}}(x, y) = \frac{\mu[y]}{\mu[x]} c(y, x) \quad \text{for all } x, y \in E. \quad (2.54)$$

The process is reversible, if and only if the **detailed balance** conditions are fulfilled,

$$\mu[x]c(x, y) = \mu[y]c(y, x) \quad \text{for all } x, y \in E, \quad (2.55)$$

which of course also implies $c_{\mathcal{R}}(x, y) = c(x, y)$.

In general, the dynamics of a process can be decomposed into a reversible (or symmetric) and a non-reversible (or anti-symmetric) part. This can be formulated on the level of the generators as $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_a$ with

$$\mathcal{L}_s := \frac{1}{2}(\mathcal{L} + \mathcal{L}^*) \quad \text{and} \quad \mathcal{L}_a := \frac{1}{2}(\mathcal{L} - \mathcal{L}^*). \quad (2.56)$$

Note that \mathcal{L}_s generates a reversible Feller process with measure μ , being a positive linear combination of generators with the same stationary measure. \mathcal{L}_a does in general not lead to a positive semigroup and does not generate a process, and we have $\mathcal{L}_a^* = -\mathcal{L}_a$. If μ is reversible for the original process \mathcal{L} , then obviously $\mathcal{L}_s = \mathcal{L}$ and $\mathcal{L}_a = 0$.

Examples. A simple example is a simple random walk on the torus $\mathbb{Z}/(L\mathbb{Z})$ with rates $c(x, y) = p\delta_{y, x+1} + q\delta_{y, x-1}$. Since the generator given by the Q -matrix is doubly stochastic we have $\langle 1|Q = \langle 1|$ and $\mu[x] = 1/L$ is the unique stationary measure. The time reversed stationary process is then simply a random walk with rates $c_{\mathcal{R}}(x, y) = q\delta_{y, x+1} + p\delta_{y, x-1}$, which is equal

to $c(x, y)$ if and only if $p = q$. The symmetric part of the dynamics is a random walk with rates $(p + q)/2$. In this example, the anti-symmetric part can also be interpreted as a totally asymmetric random walk with direction $\text{sgn}(p - q)$ and rates $|p - q|/2$.

Another interesting example is a random walk with closed or reflecting boundary conditions on $E = \{1, \dots, L\}$, which has a non-uniform stationary measure and turns out to be reversible for all $p, q > 0$. Analogous results hold for exclusion processes on finite lattices Λ equal to the above state spaces for random walks. For a fixed number of particles N , the unique stationary measure $\mu_{L,N}$ is uniform if Λ is a torus and non-uniform for other boundary conditions.

A second approach for stationary processes is based on the fact that their path measure is invariant under time shifts, and their definition can be extended to negative times on the path space $D(-\infty, \infty)$, using time translations $(\theta_T \omega)_s = \omega_{s+T}$. $\mathbb{P} \circ \theta_T$ can then be defined for paths on $D([-T, \infty))$, starting at arbitrary negative times $-T < 0$. For stationary processes, the induced law $\mathbb{P}^\mu \circ \theta_T = \mathbb{P}^\mu$ on $D([-T, \infty))$ is identical to the original, and the sequence $\mathbb{P}^\mu \circ \theta_T$ converges to the law \mathbb{P}^μ on $D(-\infty, \infty)$, which is then

$$\text{time-translation invariant, i.e. } \mathbb{P}^\mu \circ \theta_t = \mathbb{P}^\mu \text{ for all } t \in \mathbb{R}. \quad (2.57)$$

Then time reversal can simply be defined as $(\mathcal{R}\omega)_t := \omega_{(-t)_+}$ for all $t \in \mathbb{R}$, and the time reversed process is again stationary on $D(-\infty, \infty)$ with distribution μ and generator \mathcal{L}^* as defined above. As before, the path measure is then

$$\text{time-reversible, i.e. } \mathbb{P}^\mu \circ \mathcal{R} = \mathbb{P}^\mu \text{ iff } \mu \text{ is reversible}. \quad (2.58)$$

Time-reversal of non-stationary processes.

Consider a Markov chain with jump rates $c(x, y)$ on the time interval $[0, T]$ and recall the master equation (1.53)

$$\frac{d}{dt} \mu_t[x] = \sum_{y \neq x} \mu_t[y] c(y, x) - \sum_{y \neq x} \mu_t[x] c(x, y) \quad (2.59)$$

with arbitrary (non-stationary) initial condition μ_0 . Then the marginals of the time-reversed process $\mathbb{P}^\mu \circ \mathcal{R}$ are given by μ_{T-t} , and we can write

$$\begin{aligned} \frac{d}{dt} \mu_{T-t}[x] &= - \sum_{y \neq x} \mu_{T-t}[y] c(y, x) + \sum_{y \neq x} \mu_{T-t}[x] c(x, y) \\ &= \sum_{y \neq x} \mu_{T-t}[y] c_{\mathcal{R}}(y, x; T, t) - \sum_{y \neq x} \mu_{T-t}[x] c_{\mathcal{R}}(x, y; T, t) \end{aligned} \quad (2.60)$$

using reversed rates,

$$c_{\mathcal{R}}(x, y; T, t) = \frac{\mu_{T-t}[y]}{\mu_{T-t}[x]} c(y, x) \text{ for all } x \neq y, \quad (2.61)$$

which now depend on T and t . Note that the diagonal element of the Q -matrix

$$c_{\mathcal{R}}(x, x; T, t) = \sum_{y \neq x} \frac{\mu_{T-t}[y]}{\mu_{T-t}[x]} c(y, x) \neq c(x, x) \quad (2.62)$$

is also time dependent. Using the homogeneous Markov property of the original chain, a similar computation can be done for the transition kernel $P_t(x, y)$ and the reversed process $\mathbb{P}^\mu \circ \mathcal{R}$ is

indeed a time-inhomogeneous Markov chain with rates (2.61). If the process is stationary with $\mu_t \equiv \mu$ for all $t \in [0, T]$ then we recover (2.54) and the reversed chain is stationary and time-homogeneous.

A similar result holds also for jump processes on continuous state spaces, where $\mu_{T-t}[x]$ has to be replaced by the PDF with respect to the uniform (Borel) measure on E (e.g. Lebesgue on \mathbb{R}^n).

There are many examples of processes which do not have a stationary measure, but for which the uniform measure μ on E is actually invariant, i.e. $P_t\mu = \mu$ for all $t \geq 0$ but cannot be normalized since $\mu[E] = \infty$. Simple examples are simple random walks on $E = \mathbb{Z}$ with counting measure μ , or diffusion processes with constant coefficients on $E = \mathbb{R}$ with Lebesgue measure μ . In general, state spaces of such processes are transitive, i.e. generated by a symmetry group τ , such as translations generating $E = \mathbb{Z}, \mathbb{Z}^d$ or subsets $\mathbb{Z}/L\mathbb{Z}$ with periodic boundary conditions. Then any process $(P_t : t \geq 0)$ on E that has the same symmetry, i.e. $P_t\tau = \tau P_t$, has uniform invariant measure μ . Then for paths on $D[0, T]$ we can define the **modified time reversal**

$$\bar{\mathcal{R}}\omega := \mathcal{T}_{\omega_T}^{-1}\mathcal{R}\omega$$

where the path is time-reversed and shifted back to the origin point ω_0 . Then the process $\mathbb{P}_{\bar{R}}^x$ for each $x \in E$ is a time-homogeneous Markov process with the generator \mathcal{L}^* which is the adjoint operator w.r.t. the uniform measure μ (i.e. simply the transpose for Q -matrices).

Extension to negative times does not apply here since the processes involved are not stationary.

3 Additive functionals of Markov processes

3.1 Ergodicity, typical behaviour, and fluctuations

Consider a Markov process $(X_t : t \geq 0)$ on the state space E . We will investigate results like the law of large numbers and central limit theorem for ergodic averages of observables for the process.

Theorem 3.1 (LLN) *Let μ be an extremal stationary measure of $(X_t : t \geq 0)$ and $f \in C(E)$ with $\mu(f^2) < \infty$. Then*

$$\frac{1}{t} \int_0^t f(X_s) ds \rightarrow \mu(f) \quad \mathbb{P}^\mu - a.s. , \quad (3.1)$$

and in $L^2(\mathbb{P}^\mu)$, i.e. $\mathbb{E}^\mu \left[\left| \frac{1}{t} \int_0^t f(X_s) ds - \mu(f) \right|^2 \right] \rightarrow 0$.

If $(X_t : t \geq 0)$ is ergodic, the same holds for all initial conditions $x \in E$.

Proof. For the $\mathbb{P}^\mu - a.s.$ statement see [10], Section 20, convergence in $L^2(\mu)$ will follow from Theorem 3.5 (which we also do not prove...).

Theorem 3.1 is the analogue of the strong law of large numbers, and the fluctuations around the typical behaviour are described by a central limit theorem result.

Lemma 3.2 (Martingale CLT) *Let $(M_t : t \geq 0)$ be a square integrable martingale on the path space $D_{\mathbb{R}}[0, \infty)$ w.r.t. some given filtration. We assume that the increments are stationary, i.e. for all $t \geq 0$, $n \geq 1$ and $0 \leq s_0 < \dots < s_n$*

$$(M_{s_1} - M_{s_0}, \dots, M_{s_n} - M_{s_{n-1}}) \sim (M_{t+s_1} - M_{t+s_0}, \dots, M_{t+s_n} - M_{t+s_{n-1}}) , \quad (3.2)$$

and that the quadratic variation $[M]_t$ converges as

$$\mathbb{E}^\mu \left[\left| \frac{[M]_t}{t} - \sigma^2 \right| \right] \rightarrow 0 \quad \text{for some } \sigma^2 > 0 . \quad (3.3)$$

Then $M_t/\sqrt{t} \rightarrow \mathcal{N}(0, \sigma^2)$ converges in distribution to a Gaussian.

Proof. Similar to the proof of the classical CLT for sums of i.i.d. random variables, it involves a Taylor expansion of a characteristic function $\log \mathbb{E}^\mu [e^{i\theta M_t/\sqrt{t}}]$ which converges to that of a Gaussian. For details, see [21], Chapter 1 and 2.

Corollary 3.3 (CLT) *Let μ be an extremal stationary measure of $(X_t : t \geq 0)$ and $f \in C(E)$ with $\mu(f) = 0$ and $\mu(f^2) < \infty$. In addition, assume that there exists a solution of the Poisson equation, i.e.*

$$-\mathcal{L}g = f \quad \text{for some } g \in \mathcal{D}_{\mathcal{L}} , \quad (3.4)$$

such that also $g^2 \in \mathcal{D}_{\mathcal{L}}$. Then we have convergence in distribution to a Gaussian

$$\frac{1}{\sqrt{t}} \int_0^t f(X_s) ds \rightarrow \mathcal{N}(0, 2\langle g, f \rangle_\mu) . \quad (3.5)$$

Proof. Since g and g^2 are in the domain of the generator,

$$M_t^g = g(X_t) - g(X_0) - \int_0^t (\mathcal{L}g)(X_s) ds = g(X_t) - g(X_0) + \int_0^t f(X_s) ds \quad (3.6)$$

is a martingale with quadratic variation

$$[M^g]_t = \int_0^t [(\mathcal{L}g^2)(X_s) - 2g(X_s)(\mathcal{L}g)(X_s)] ds .$$

μ is stationary, and exchanging expectation and integration implies

$$\mathbb{E}^\mu [[M^g]_t] = 2t \langle g, -\mathcal{L}g \rangle_\mu , \quad (3.7)$$

with $\mathbb{E}^\mu [\mathcal{L}g^2] = 0$. Since g is the solution of (3.4) we have

$$\frac{1}{\sqrt{t}} \int_0^t f(X_s) ds = \frac{M_t^g}{\sqrt{t}} + \frac{g(X_0) - g(X_t)}{\sqrt{t}} .$$

$g^2 \in \mathcal{D}_L$ in particular implies $g \in L^2(\mu)$, and therefore $(g(X_0) - g(X_t))/\sqrt{t} \rightarrow 0$ in $L^2(\mathbb{P}^\mu)$ as $t \rightarrow \infty$. Since $(X_t : t \geq 0)$ is a stationary process under \mathbb{P}^μ , the increments of the martingale (3.6) are stationary, and with (3.7) obviously (3.3) is fulfilled with $\sigma^2 = 2\langle g, f \rangle_\mu$. \square

Remarks. The corollary also holds without the assumption that g^2 is in the domain of the generator, where an additional approximation argument to show (3.3) has to be made. But the functions we will consider later all fulfill this assumption anyway.

Note that the corollary implies a weak version of the LLN in Theorem 3.1. Unlike the general validity of the LLN, convergence to Gaussians can only be shown for observables of the type (3.4).

From now on we consider μ to be an extremal reversible measure of the process $(X_t : t \geq 0)$ and restrict to functions in $L^2(\mu)$. This implies that the generator \mathcal{L} is self-adjoint, i.e. $\langle f, \mathcal{L}g \rangle_\mu = \langle g, \mathcal{L}f \rangle_\mu$, and therefore has real spectrum. Since associated semigroups P_t are contracting (1.9), the spectrum is also non-positive, with 0 being the largest eigenvalue with corresponding constant eigenfunction 1.

In general, the spectrum of \mathcal{L} is given by the complement of the **resolvent set**, which contains all $\lambda \in \mathbb{R}$ such that the resolvent equation $\lambda f - \mathcal{L}f = g$ has a unique solution $f \in L^2(\mu)$ for any fixed $g \in L^2(\mu)$. In that case $g = (\lambda \mathbb{I} - \mathcal{L})^{-1} f = R_\lambda f$ is given by the resolvent as given in (1.11), which is well defined for all $\lambda > 0$. The spectrum can be decomposed into the point spectrum consisting of the eigenvalues of \mathcal{L} , and a continuous part of the spectrum, for details see e.g. [14].

Definition 3.1 For a process with generator \mathcal{L} and reversible measure μ the **Dirichlet form** for $f \in \mathcal{D}_\mathcal{L}$ is defined as

$$\mathcal{E}_\mu(f) := \langle f, -\mathcal{L}f \rangle_\mu , \quad (3.8)$$

and the **spectral gap** is

$$\lambda_{\text{gap}} := \inf_{f \in \mathcal{D}_\mathcal{L}} \left\{ \frac{\mathcal{E}_\mu(f)}{\text{Var}_\mu(f)} : \text{Var}_\mu(f) > 0 \right\} = \inf_{f \in \mathcal{D}_\mathcal{L}} \left\{ \mathcal{E}_\mu(f) : \mu(f^2) = 1, \mu(f) = 0 \right\} . \quad (3.9)$$

If $\lambda > 0$, the inverse of the spectral gap is called the **relaxation time**

$$t_{\text{rel}} := 1/\lambda_{\text{gap}} . \quad (3.10)$$

The second expression in (3.9) follows from the fact that the Dirichlet form and variance do not change when adding constants to the function f , since $\mathcal{L}1 = 0$. Useful inequalities for all $f \in L^2(\mu)$ resulting from the definition are

$$\text{Var}(f) \leq t_{\text{rel}} \mathcal{E}_\mu(f) \quad \text{and} \quad -\mathcal{E}_\mu(f) \leq -\lambda_{\text{gap}} \text{Var}(f), \quad (3.11)$$

where the first is called **Poincaré inequality**. If λ is an eigenvalue of \mathcal{L} with corresponding eigenfunction f , then $\mathcal{E}_\mu(f) = -\lambda \mu(f^2)$. This implies that the spectral gap is given by the modulus of the largest, non-zero eigenvalue, since the infimum is taken over non-constant functions with $\text{Var}(f) > 0$. This can easily be shown under the assumption that the eigenfunctions form a basis of $L^2(\mu)$. Analogously to (1.54) for Markov chains, the gap characterizes the decay of temporal correlations as summarized in the following result.

Proposition 3.4 *Let μ be a reversible, extremal measure of a Markov process with generator \mathcal{L} and semigroup $(P_t : t \geq 0)$. Then*

$$\mathbb{E}^\mu[f(X_t)f(X_0)] - \mu(f)^2 = \text{Var}_\mu(P_t f) \leq e^{-2\lambda t} \text{Var}_\mu(f) \quad \text{for all } f \in L^2(\mu). \quad (3.12)$$

Proof. Let $u(t) = \text{Var}_\mu(P_t f) = \mu((P_t f - \mu(f))^2)$. Then, using (3.11)

$$u'(t) = -2\mathcal{E}_\mu(P_t f - \mu(f)) \leq -2\lambda u(t),$$

which implies $u(t) \leq e^{-2\lambda t} u(0)$. Since $u(0) = \text{Var}_\mu(f)$ this is (3.12). \square

Examples. Consider the Laplacian Δ on $L^2(\mathbb{R}, \mu)$ with uniform Lebesgue measure μ . This is not normalized, but otherwise reversible for Δ which is generating Brownian motion on \mathbb{R} . It is well known that Δ has continuous spectrum $(-\infty, 0]$ (see e.g. [14]) and no eigenvalues. We have $\Delta 1 = 0$, but the constant function $1 \notin L^2(\mathbb{R}, \mu)$. Therefore correlations decay subexponentially fast in time, such as polynomial decay for Brownian motion as given by the heat kernel (1.28).

The discrete Laplacian Δ on $\ell^2(\mathbb{Z}, \mu)$ with uniform counting measure μ generates the continuous-time random walk. Using that the shift operator S on ℓ^2 has spectrum $\{z \in \mathbb{C} : |z| = 1\}$ given by the unit circle, and $\Delta = S + S^{-1} - 2\mathbb{I}$, the spectrum of Δ is now $[-4, 0]$. This restriction of the spectrum can be associated to a cut-off of high frequency oscillations due to the lattice structure. Considering the Laplacian on the discrete torus $\mathbb{Z}/(L\mathbb{Z})$ cuts off further low frequency oscillations (infrared cut-off). Due to finite state space it has a pure point spectrum and the spectral gap is L^{-2} , consistent with the CLT scaling for the simple random walk.

For conservative IPS with finite range jumps on lattices of size $|\Lambda| = L$, the gap usually decays with the system size also as L^{-2} for reversible systems. Intuitively, this is related to the fact that for correlations to decay, mass has to be moved distances of order L in a symmetric fashion.

Further remarks. For asymmetric, non-reversible IPS the decay of correlations is not necessarily characterized by a result like Proposition 3.4 using the same definition for λ_{gap} . It depends heavily on boundary conditions, and for example for the ASEP on a periodic 1D lattice it is known that the relaxation time scales like $L^{3/2}$. The gap for the symmetric part \mathcal{L}_s of the generator (2.56) provides a (usually bad) lower bound for the gap (upper bound for the relaxation time).

On infinite lattices the gap of conservative IPS is usually 0 related to the properties of the Laplacian above. On the other hand, for spin systems IPS correlations can decay through local reactions much faster, and the gap is usually bounded below independently of the system size. Interesting

examples include the contact process (in particular the supercritical case) and kinetically constrained models such as the east model, and in particular the connection with mixing times (see e.g. [16, 17, 18] for further details and references).

The gap or relaxation time can also be used to bound the convergence of an additive path functional to its expectation, as is summarized in the next result. This also implies the L^2 convergence in Theorem 3.1.

Theorem 3.5 *Let μ be an extremal reversible measure of $(X_t : t \geq 0)$ and $f \in C(E)$ with $\mu(f) = 0$ and $\mu(f^2) < \infty$. Then*

$$\mathbb{E}^\mu \left[\sup_{0 \leq t \leq T} \left(\int_0^t f(X_s) ds \right)^2 \right] \leq 24\mu(f^2)Tt_{\text{rel}}. \quad (3.13)$$

Proof. In [21], Chapter 2, a more general result is shown (Lemma 2.3) which applies also to non-reversible processes. There exists a universal constant $C > 0$ such that

$$\mathbb{E}^\mu \left[\sup_{0 \leq t \leq T} \left(\int_0^t f(X_s) ds \right)^2 \right] \leq CT\|f\|_{-1}^2, \quad (3.14)$$

where the \mathcal{H}_{-1} -norm is given by the Legendre transform of the Dirichlet form

$$\|f\|_{-1}^2 = \sup_g \{2\langle f, g \rangle_\mu - \mathcal{E}_\mu(g)\}.$$

Note that $\mathcal{E}_\mu(g)$ is a semi-norm (where also constant functions have norm 0), and the supremum is performed over a common core of the generator \mathcal{L} and its adjoint \mathcal{L}^* .

For reversible processes the \mathcal{H}_{-1} -norm can be related to the spectral gap. For details of the proof see [21], Chapter 2, and [22], Section 3.

3.2 Additive path functionals

The integrated observables in the previous section can be interpreted as path functionals of the process which are additive in time. Other important functionals include counters of jumps along a path of a jump process, which cannot be written as integrals of functions on state space. This is relevant for Markov chains (MCs) on a countable state space E , recalling the generator

$$\mathcal{L}f(x) = \sum_{y \neq x} c(x, y)(f(y) - f(x)), \quad (3.15)$$

but also for conservative IPS such as exclusion processes. They have state space $E = S^\Lambda$, considering $S \subseteq \mathbb{N}_0$ finite, and generator

$$\mathcal{L}f(\eta) = \sum_{i, j \in \Lambda} c(i, j, \eta)(f(\eta^{ij}) - f(\eta)), \quad (3.16)$$

and can be interpreted as interacting MCs on Λ . As before we focus on finite range, uniformly bounded jump rates. Results in this section can be formulated completely analogously for spin systems counting the number of spin flips. The common feature of these jump processes is that from a given state x they can jump to a countable number of other states y , characterized by jump rates $c(x, y)$ with a generator of the form (3.15). For IPS the sum is a-priori uncountable, but contains only countably many non-zero terms. Similar results could be formulated for jump processes on continuous state space with rates $c(x, dy)$.

Definition 3.2 For a given path $\omega \in D[0, \infty)$ and $x, y \in E$ we define the **empirical current** up to time $t > 0$ as

$$\mathcal{J}_t(x, y)[\omega] = \sum_{0 \leq s \leq t} \delta(\omega_s(y) - \omega_{s-}(y) - \omega_s(x) + \omega_{s-}(x), 2) \in \mathbb{N}_0. \quad (3.17)$$

For MCs and also IPS with generator (3.16) the above sum is a $\mathbb{P} - a.s.$ finite.

For given coefficients $a(x, y) \in \mathbb{R}$, we define the **additive path functional**

$$A_t[\omega] = \sum_{x, y \in E} a(x, y) \mathcal{J}_t(x, y)[\omega] \quad (3.18)$$

as a linear combination of low counters.

Note that for an IPS on an infinite lattice Λ the sum in (3.18) is uncountable, and one has to make sure it contains only countable many terms and converges. For example, consider the ASEP with particles jumping left and right with rates p and q , respectively. Then choosing $a(\eta, \zeta) = \delta_{\zeta, \eta^{i, i+1}} - \delta_{\zeta, \eta^{i+1, i}}$, A_t/t measures the net particle current across a particular bond $(i, i+1)$ up to time t . This corresponds to uncountably many coefficients $a(\eta, \zeta) \neq 0$, but $\mathbb{P} - a.s.$ only finitely many $\mathcal{J}_t(x, y) > 0$ for arbitrary $t > 0$. Other examples include the activity where each jump contributes positively. In finite systems one can also count the total particle current or activity across all bonds.

Lemma 3.6 Consider a jump process $(X_t : t \geq 0)$ with generator (3.15). Then

$$M_t^{(x, y)} := \mathcal{J}_t(x, y) - \int_0^t c(x, y) \delta_{X_s, x} ds \quad (3.19)$$

is a martingale with quadratic variation $[M^{(x, y)}]_t = \int_0^t c(x, y) \delta_{X_s, x} ds$.

Proof. It is convenient to consider the joint Markov process $(X_t, \mathcal{J}_t(x, y))$ on the state space $E \times \mathbb{N}_0$ which has generator

$$\mathcal{L}^{(x, y)} f(z, j) = \sum_{z' \in E} c(z, z') (f(z', j + \delta_{x, z} \delta_{y, z'}) - f(z, j)). \quad (3.20)$$

Then with the test function $f(z, j) = j$ we get $\mathcal{L}^{(x, y)} f(z, j) = c(x, y) \delta_{z, x}$, which implies (3.19) with the martingale characterization in Theorem 1.5. We further compute

$$\mathcal{L}^{(x, y)} f(z, j)^2 - 2f(z, j) \mathcal{L}^{(x, y)} f(z, j) = c(x, y) \delta_{z, x},$$

which implies the expression for the quadratic variation. \square

Theorem 3.7 Consider jump process $(X_t : t \geq 0)$ with generator (3.15) and extremal stationary measure μ . Then we have the law of large numbers

$$\frac{1}{t} \mathcal{J}_t(x, y) \rightarrow \mu[x]c(x, y) \quad \mathbb{P}^\mu - a.s. \text{ and in } L^2(\mathbb{P}^\mu). \quad (3.21)$$

If the process is ergodic the same holds for arbitrary initial condition.

If the process is reversible w.r.t. μ , the fluctuations around the limit can be bounded as

$$\mathbb{E}^\mu \left[\sup_{0 \leq t \leq T} \left(\mathcal{J}_t(x, y) - t\mu[x]c(x, y) \right)^2 \right] \leq T \left(4\mu[x]c(x, y) + 24\mu[x]c(x, y)^2 t_{\text{rel}} \right). \quad (3.22)$$

Proof. Using (3.19) we can write

$$\frac{1}{t} \mathcal{J}_t(x, y) - \mu[x]c(x, y) = \frac{1}{t} M_t^{(x, y)} - \frac{1}{t} \int_0^t \left(c(x, y) \delta_{X_s, x} - \mu[x]c(x, y) \right) ds. \quad (3.23)$$

With the quadratic variation of the martingale $M_t^{(x, y)}$ we have

$$\frac{1}{t^2} \mathbb{E}^\mu \left[(M_t^{(x, y)})^2 \right] = \frac{1}{t} \mu[x]c(x, y) \rightarrow 0,$$

which implies that $M_t^{(x, y)}/t \rightarrow 0$, $\mathbb{P}^\mu - a.s.$ and in $L^2(\mu)$. Then the LLN (3.21) follows applying Theorem 3.1 to the second term of (3.23).

Using (3.23) and Minkowski's inequality we can bound

$$\sup_{0 \leq t \leq T} \left(\mathcal{J}_t(x, y) - t\mu[x]c(x, y) \right)^2 \leq \sup_{0 \leq t \leq T} (M_t^{(x, y)})^2 + \sup_{0 \leq t \leq T} \left(\int_0^t \bar{c}(x, y, X_s) ds \right)^2,$$

where we write $\bar{c}(x, y, X_s) = c(x, y) \delta_{X_s, x} - \mu[x]c(x, y)$ for the standardized rates which are a mean-zero function of X_s . For the first term we can use Doob's inequality¹, and for the second we use Theorem 3.5 to get (3.22). \square

Analogously to Theorem 3.5, there exists a bound to (3.22) also for non-reversible processes. We can also use (3.23) to derive a CLT-type result. We can use Lemma 3.2 for the martingale and an adaption of Corollary 3.3 for the second part, since the integrand is of the form $\mathcal{L}^{(x, y)} g$ for the joint process $(X_t, \mathcal{J}_t(x, y))$. Both lead to Gaussian limits in distribution, which are, however, correlated, and this does not lead to a general result for the limiting variance.

Note also that for IPS on infinite lattices $\mu[\eta] = 0$ and the above results are trivial if we count the empirical current between particular configurations which is simply 0, $\mathbb{P} - a.s.$. Analogous results of course hold for general path functionals A_t due to linearity, for example counting the particle current in a lattice gas between sites (i, j) we put choose as discussed before

$$A_t^{(i, j)} := \sum_{\eta, \zeta \in E} \delta_{\zeta, \eta^{i, j}} - \delta_{\zeta, \eta^{j, i}} \mathcal{J}_t(\eta, \zeta).$$

Then analogously to Lemma 3.6 we have the martingale

$$M_t^{(i, j)} := A_t^{(i, j)} - \int_0^t c(i, j, \eta_s) ds,$$

and we get as in Theorem 3.7,

$$\frac{1}{t} A_t^{(i, j)} \rightarrow \mu(c(i, j, \cdot)) \quad \mathbb{P}^\mu - a.s. \text{ and in } L^2(\mathbb{P}^\mu), \quad (3.24)$$

with an analogous statement for the fluctuations as in (3.22).

Example. For the TASEP with rates $c(i, i+1, \eta) = \eta(i)(1 - \eta(i+1))$ we get from (3.24) for the asymptotic current under the stationary product measure ν_ρ (see Def. 2.4)

$$\frac{1}{t} A_t^{(i, i+1)} \rightarrow \nu_\rho(c(i, i+1, \cdot)) = \rho(1 - \rho).$$

¹For a p -integrable martingale $(M_t : t \geq 0)$ with $p > 1$ we have $\mathbb{E}[\sup_{0 \leq t \leq T} M_t^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[M_T^p]$.

It is further known [32] that the fluctuations of the current,

$$\frac{A_t^{(i,i+1)} - t\rho(1-\rho)}{\sqrt{t}} \rightarrow \mathcal{N}(0, \rho(1-\rho)|1-2\rho|)$$

have a Gaussian limit for all densities except $\rho = 1/2$. In that case negative correlations in the two contributions in (3.23) cancel out and the variance vanishes. The current fluctuations occur on a lower order scale of $t^{1/3}$, which is a trademark of processes in the so-called KPZ universality class (see [23] for a recent review).

3.3 Large deviations of additive functionals

In this section we collect some general results on large deviations of path functionals for a general Feller process \mathbb{P} , where we do not specify the initial condition. The functional $A_t[\omega] \in \mathbb{R}$ can be arbitrary, as long as it is \mathbb{P} -*a.s.* finite for all $t \geq 0$ and fulfills LLN with $A_t \rightarrow a \in \mathbb{R}$ as $t \rightarrow \infty$. Additive path functionals introduced earlier are particular examples, but now we use the convention to include the factor $1/t$ in the definition of A_t as is standard in large deviation theory.

Definition 3.3 The family of real-valued random variables ($A_t : t \geq 0$) is said to satisfy a **large deviation principle (LDP)** with respect to \mathbb{P} , if there exists a lower semi-continuous¹ function $I : \mathbb{R} \rightarrow [0, \infty]$ with $I \neq \infty$, such that

$$\liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}[A_t \in C] \geq \inf_{a \in C} I(a) \quad \text{for all closed sets } C \subseteq \mathbb{R}, \quad (3.25)$$

and

$$\limsup_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}[A_t \in O] \leq \inf_{a \in O} I(a) \quad \text{for all open sets } O \subseteq \mathbb{R}. \quad (3.26)$$

The function I is called the **rate function**, and if I has compact level sets it is called a **good rate function**.

The LDP implies that events of the form $A_t \in [a, a + da]$ have exponentially small probability in leading order as $t \rightarrow \infty$, which in compact notation is often written as

$$\mathbb{P}[A_t \in [a, a + da]] \asymp e^{-tI(a)} da.$$

In general, LDPs can hold with different **speeds** $b_t \nearrow \infty$ where above probability scales like $e^{-b_t I(a)} da$. But this can be absorbed in a re-parametrization of A_t , and for standardized time-additive functionals we expect $b_t = t$ to be the correct speed anyway.

In the following we assume that the **scaled cumulant generating function (SCGF)**

$$\theta(k) := \lim_{t \rightarrow \infty} \theta_t(k) := \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}[e^{ktA_t}] \in (-\infty, \infty] \quad (3.27)$$

is well defined (the limit exists) as an extended real number for all $k \in \mathbb{R}$. We obviously have $\theta(0) = 0$, and assume that the origin is in the interior of the essential domain, i.e.

$$0 \in \mathcal{D}_\theta^\circ \quad \text{where} \quad \mathcal{D}_\theta := \{k \in \mathbb{R} : \theta(k) < \infty\}. \quad (3.28)$$

¹ I is lower semi-continuous in a if for all $\epsilon > 0$ there exists a neighbourhood U of a such that $I(x) \geq I(a) - \epsilon$.

Definition 3.4 The **Legendre-Fenchel transform** of $\theta(k)$ is given by

$$\theta^*(a) := \sup_{k \in \mathbb{R}} (k a - \theta(k)) . \quad (3.29)$$

$k \in \mathbb{R}$ is an **exposed point** of θ if for some $a \in \mathbb{R}$

$$k a - \theta(k) > l a - \theta(l) \quad \text{for all } l \neq k ,$$

and such a is called an **exposing hyperplane**. For all $k \in \mathbb{R}$ the **subdifferential** of θ is given by the set of all exposing hyperplanes

$$\partial\theta(k) := \{a \in \mathbb{R} : k a - \theta(k) > l a - \theta(l) \text{ for all } l \neq k\} . \quad (3.30)$$

Some basic properties of θ and its Legendre transform are summarized in the next result, in particular including convexity of θ . Note that if the point $k \in \mathbb{R}$ is not exposed, then simply $\partial\theta(k) = \emptyset$. If θ has kinks, the subdifferential can also be an interval, if θ is differentiable in k then simply $\partial\theta(k) = \{\theta'(k)\}$. We also have $\theta'(0) = \lim_{t \rightarrow \infty} \mathbb{E}[A_t]$ corresponding to the asymptotic expectation of the functional.

Proposition 3.8 *Under assumption (3.27), both θ and θ^* are convex functions on \mathbb{R} , and $(\theta^*)^* = \theta$. Furthermore, θ^* is non-negative, lower-semicontinuous and additionally assuming (3.28) has compact level sets, so can be a good rate function for an LDP.*

If $a \in \partial\theta(k)$, then $\theta^(a) = k a - \theta(k)$.*

Proof. Convexity of θ follows by Hölder's inequality, since for all $\alpha \in [0, 1]$, $k_1, k_2 \in \mathbb{R}$ and $t \geq 0$ we have

$$\begin{aligned} \theta_t(\alpha k_1 + (1 - \alpha)k_2) &= \frac{1}{t} \log \mathbb{E} \left[(e^{k_1 t A_t})^\alpha (e^{k_2 t A_t})^{1-\alpha} \right] \\ &\leq \frac{1}{t} \log \left(\mathbb{E} [e^{k_1 t A_t}]^\alpha \mathbb{E} [e^{k_2 t A_t}]^{1-\alpha} \right) = \alpha \theta_t(k_1) + (1 - \alpha) \theta_t(k_2) . \end{aligned}$$

Then θ is convex as the limit of convex functions θ_t . θ^* is convex simply by the definition of the Legendre transform,

$$\begin{aligned} \alpha \theta^*(a_1) + (1 - \alpha) \theta^*(a_2) &= \\ &= \sup_{k \in \mathbb{R}} (\alpha k a_1 - \alpha \theta(k)) + \sup_{k \in \mathbb{R}} ((1 - \alpha) k a_2 - (1 - \alpha) \theta(k)) \\ &\geq \sup_{k \in \mathbb{R}} ((\alpha a_1 + (1 - \alpha) a_2) k - \theta(k)) = \theta^*(\alpha a_1 + (1 - \alpha) a_2) . \end{aligned} \quad (3.31)$$

$\theta(0) = 0$ implies that $\theta^*(a) \geq 0 a - \theta(0) = 0$ is non-negative for all $a \in \mathbb{R}$.

To see that θ^* is lower semi-continuous, fix a sequence $a_n \rightarrow a$. Then for all $k \in \mathbb{R}$

$$\liminf_{a_n \rightarrow a} \theta^*(a_n) \geq \liminf_{a_n \rightarrow a} (k a_n - \theta(k)) = k a - \theta(k) ,$$

and thus

$$\liminf_{a_n \rightarrow a} \theta^*(a_n) \geq \sup_{k \in \mathbb{R}} (k a - \theta(k)) = \theta^*(a) ,$$

which implies lower semi-continuity.

Since $0 \in \mathcal{D}_\theta^\circ$, there is a closed ball $\bar{B}_0(\delta) \subseteq \mathcal{D}_\theta^\circ$ for some radius $\delta > 0$ such that $M := \sup_{k \in \bar{B}_0(\delta)} \theta(k) < \infty$, because the convex function θ is continuous in \mathcal{D}_θ° . Therefore,

$$\theta^*(a) \geq \sup_{k \in \bar{B}_0(\delta)} (k a - \theta(k)) \geq \delta |a| - M .$$

Thus, for every $\alpha < \infty$ the level set $\{a : \theta^*(a) \leq \alpha\}$ is bounded.

If $k a - \theta(k) > l a - \theta(l)$ for all $l \neq k$, then by definition $\theta^*(a) = k a - \theta(k)$. \square

Definition 3.5 A convex function $\theta : \mathbb{R} \rightarrow (-\infty, \infty]$ is **essentially smooth** if:

- \mathcal{D}_θ° is non-empty,
- θ is differentiable throughout \mathcal{D}_θ° ,
- θ is steep, i.e. $\lim_{n \rightarrow \infty} |\theta'(k_n)| = \infty$ for all sequences k_n in \mathcal{D}_θ° converging to a boundary point of \mathcal{D}_θ° .

Theorem 3.9 Gärtner-Ellis. *Under assumption (3.27) we have*

$$\begin{aligned} \liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}[A_t \in C] &\geq \inf_{a \in C} \theta^*(a) \quad \text{for all closed } C \subseteq \mathbb{R} , \\ \limsup_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}[A_t \in O] &\leq \inf_{a \in O \cap G} \theta^*(a) \quad \text{for all open } O \subseteq \mathbb{R} , \end{aligned} \quad (3.32)$$

where G is the set of exposed points of θ^* whose exposing hyperplane belongs to \mathcal{D}_θ° .

If θ^* is strictly convex, or alternatively, θ is essentially smooth and lower semi-continuous, then the LDP in Definition 3.3 holds with good rate function θ^* .

Proof. see [28], Chapter 2.

In general, the large deviation rate function as given in Definition 3.3 does not have to be convex. Its Legendre transform is always convex by definition, and given by the scaled cumulant generating function (3.27).

Corollary 3.10 *Let $(A_t : t \geq 0)$ be a sequence of path functionals of a Feller process $(X_t : t \geq 0)$ that obey a LLN. Then $(A_t : t \geq 0)$ satisfies an LDP w.r.t. \mathbb{P} with good rate function*

$$I(a) = \theta^*(a) = \sup_{k \in \mathbb{R}} (k a - \theta(k)) \quad (3.33)$$

given by the Legendre transform of the SCGF $\theta(k)$ as defined in (3.27), provided it fulfills the regularity assumptions of Theorem 3.9.

As mentioned earlier, since additive path functionals are roughly linear in time, we expect non-trivial results for rate functions of LDPs with speed t . For general functionals obeying a LLN $A_t \rightarrow \bar{a}$ we expect an LDP with speed t in any case, but it may be trivial. If the speed t is too low, i.e. actual fluctuations are even more unlikely, we have $I(\bar{a}) = 0$ and $I(a) = \infty$ for all $a \neq \bar{a}$. If the speed t is too high all values appear typical on the exponential scale with $I(a) \equiv 0$.

3.4 Conditional path ensembles

To quantitatively study large deviations of path functionals we have to get a handle on the scaled cumulant generating function $\theta(k)$ (3.27). One popular approach is to define conditional path ensembles on a compact time interval $[0, T]$, and investigate their behaviour as $T \rightarrow \infty$ which will be dominated by the principal eigenvalue of a modified generator. As in the previous section we consider a general Feller process $(X_t : 0 \leq t \leq T)$ with path measure \mathbb{P} , and a general functional $A_T[\omega]$, which exhibits a non-trivial LDP with speed T .

Definition 3.6 Conditional path measures

The **canonical path measure** for the process $X_t|_{\{A_T=a\}}$ on path space $D([0, T])$ is given by

$$d\mathbb{P}_c[\omega] := \frac{\mathbb{1}_{[a, a+da]}(A_T(\omega))}{\mathbb{P}[A_T \in [a, a+da]]} d\mathbb{P}[\omega], \quad (3.34)$$

and the **grand-canonical path measure** or **Feynman-Kac transform** is given by

$$d\mathbb{P}_{gc}[\omega] := \frac{e^{TkA_T(\omega)}}{\mathbb{E}[e^{TkA_T}]} d\mathbb{P}[\omega]. \quad (3.35)$$

The **tilted path measure** is given by an un-normalized version of the grand-canonical one,

$$d\mathbb{P}_k[\omega] := e^{TkA_T(\omega)} d\mathbb{P}[\omega]. \quad (3.36)$$

The first two are both probability measures on path space $D([0, T])$, the canonical one conditions on a particular value of the functional A_T with normalization given by the (large deviation) probability $\mathbb{P}[A_T \in [a, a+da]]$, and the grand-canonical one re-weights the path ensemble according to the value of A_T , and the normalization is related to the SCGF. Both measures do in general not define Markov processes, but for large T are equivalent to Markov processes, as is shown in [29], Section V. We will cover some aspects of process equivalence in the following. A standard way to measure the difference between probability measures is to use the relative entropy.

Definition 3.7 For two probability measures μ_1, μ_2 on Ω , the **relative entropy** is given by

$$H(\mu_1; \mu_2) := \begin{cases} \int_{\Omega} \log \frac{d\mu_1}{d\mu_2}(\omega) d\mu_1[\omega] & , \text{ if } \mu_1 \ll \mu_2 \\ \infty & , \text{ if } \mu_1 \not\ll \mu_2 \end{cases}. \quad (3.37)$$

Here we use the standard notation $\mu_1 \ll \mu_2$ for **absolut continuity** of μ_1 w.r.t. μ_2 , i.e. for each measurable $A \subseteq \Omega$, $\mu_2(A) = 0$ implies that $\mu_1(A) = 0$. This implies existence of the **Radon-Nikodym derivative** $\frac{d\mu_1}{d\mu_2}$, which is a measurable function on Ω such that

$$d\mu_1[\omega] = \frac{d\mu_1}{d\mu_2}(\omega) d\mu_2[\omega].$$

For $\mu_1 \ll \mu_2$ with $q := \frac{d\mu_1}{d\mu_2}$ we have $d\mu_1 = q d\mu_2$ and we can write

$$H(\mu_1; \mu_2) = \int_{\Omega} q(\omega) \log q(\omega) d\mu_2.$$

Since $x \log x$ is a convex function, $H(\mu_1; \mu_2)$ is convex in μ_1 for given μ_2 . By Jensen's inequality we have

$$H(\mu_1; \mu_2) \geq \mu_2(q) \log \mu_2(q) = \mu_1[\Omega] \log \mu_1[\Omega] = 0 ,$$

with equality if and only if $h = 1 - \mu_2 - a.s.$ and $\mu_1 = \mu_2$. Relative entropy is not symmetric and therefore not a metric, but due to non-negativity can still be used to quantify the difference between probability measures. For more details related to relative entropy see [8], Appendix 1.

To quantify the distance between canonical and grand-canonical path ensembles, we define the **specific relative entropy**

$$h(a, k) := \lim_{T \rightarrow \infty} \frac{1}{T} H(\mathbb{P}_c; \mathbb{P}_{gc}) . \quad (3.38)$$

Proposition 3.11 Equivalence of ensembles

Under the conditions of Corollary 3.10 the limit in (3.38) is well-defined and

$$h(a, k) = I(a) + \theta(k) - k a .$$

Furthermore, the **equivalence of ensembles** holds, i.e. for all $a \in \mathbb{R}$ such that $I(a) < \infty$ we have

$$\inf_{k \in \mathbb{R}} h(a, k) = I(a) + \theta(k^*) - k^* a = 0 , \quad (3.39)$$

where the minimizer k^* is such that $\theta'(k^*) = a$.

Proof. First note that $\mathbb{P}_c \ll \mathbb{P}_{gc}$ since $\mathbb{P}_c \ll \mathbb{P}$ as a conditional version of \mathbb{P} , and $\mathbb{P} \ll \mathbb{P}_{gc}$. Therefore the Radon-Nikodym derivative

$$\frac{d\mathbb{P}_c}{d\mathbb{P}_{gc}}(\omega) = \frac{d\mathbb{P}_c/d\mathbb{P}(\omega)}{d\mathbb{P}_{gc}/d\mathbb{P}(\omega)} = \frac{\mathbb{E}[e^{TkA_T}] e^{-Tka}}{\mathbb{P}[A_T \in [a, a + da]]} \mathbb{1}_{[a, a+da]}(A_T(\omega))$$

is given by a simple function which is constant on $\{A_T \in [a, a + da]\}$ and vanishes outside. This leads to

$$H(\mathbb{P}_c; \mathbb{P}_{gc}) = \mathbb{E}_c \left[\log \frac{d\mathbb{P}_c}{d\mathbb{P}_{gc}} \right] = \log \mathbb{E}[e^{TkA_T}] - k a - \log \mathbb{P}[A_T \in [a, a + da]] ,$$

and together with the LDP in Corollary 3.10 this implies that $h(a, k) = I(a) + \theta(k) - k a$. Taking the infimum over k with the definition of the Legendre transform in (3.29) leads to

$$\inf_{k \in \mathbb{R}} h(a, k) = I(a) - \theta^*(a) = 0 .$$

The conditions of Corollary 3.10 imply that $\theta(k)$ is differentiable on its essential domain \mathcal{D}_θ^o , and therefore the infimum in (3.39) is attained at $k^* \in \mathbb{R}$ such that

$$\theta'(k^*) = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{\mathbb{E}[T A_T e^{Tk^* A_T}]}{\mathbb{E}[e^{Tk^* A_T}]} = \mathbb{E}_{gc}[A_T] = a .$$

□

Therefore, an LDP for the path functional implies the asymptotic equivalence of canonical and grand-canonical path measures as $T \rightarrow \infty$, provided that the grand-canonical measure \mathbb{P}_{gc} is tilted such that the functional expectation is equal to the conditional value a under \mathbb{P}_c .

3.5 Tilted path ensemble

In this section we consider a Feller jump process $(X_t : t \geq 0)$ on the general state space E with jump rates $c(x, dy)$ and generator

$$\mathcal{L}f(x) = \int_E c(x, dy) [f(y) - f(x)] .$$

We will go back to particular additive path functionals as introduced in Sections 3.1 and 3.2, using a more general notation

$$A_t = \frac{1}{t} \int_0^t h(X_s) ds + \frac{1}{t} \sum_{0 \leq s \leq t: X_{s-} \neq X_s} a(X_{s-}, X_s) . \quad (3.40)$$

We have $h \in C(E)$, and $a \in C(E \times E)$ is now counting the contribution of jumps on path space. For general $a(x, y)$ we have to assume that \mathbb{P} -a.s. there are only finitely many jumps in any given time interval $[0, t]$, which excludes IPS on infinite lattices. For the latter, the sum (3.40) can still be well-defined for particular choices of a corresponding to e.g. the measurement of local currents as discussed previously.

In the following we analyze the tilted path measure (3.36), which has actually a Markovian structure as we will see, but does not conserve probability since in general

$$\mathbb{P}_k[\Omega] = \mathbb{E}[e^{TkA_T}] \neq 1 .$$

Trying to incorporate the action of the tilting e^{TkA_T} with A_T given in (3.40), we define the **tilted generator**

$$\mathcal{L}_k f(x) := \int_E c(x, dy) [e^{ka(x,y)} f(y) - f(x)] + k h(x) f(x) . \quad (3.41)$$

With each jump from x to y the path ensemble is re-weighted by a factor $e^{ka(x,y)}$ corresponding to the second term in (3.40), and the integral conditioning leads to an additional diagonal term. We split the operator \mathcal{L}_k into a conservative part $\tilde{\mathcal{L}}_k$ and diagonal parts,

$$\mathcal{L}_k f(x) := \underbrace{\int_E c(x, dy) e^{ka(x,y)} [f(y) - f(x)]}_{:= \tilde{\mathcal{L}}_k f(x)} + f(x) \underbrace{\left(\int_E c(x, dy) [e^{ka(x,y)} - 1] + k h(x) \right)}_{= (\mathcal{L}_k 1)(x)} .$$

One interpretation of the backward equation (1.23) in the Hille-Yosida Theorem 1.4 is, that the semigroup $u(t, x) := P_t f(x) = \mathbb{E}^x [f(X_t)]$ provides a solution of the evolution equation

$$\frac{d}{dt} u(t, x) = \mathcal{L}u(t, x) , \quad u(0, x) = f(x) .$$

The Feynman-Kac formula provides a probabilistic representation of the solution to an equation where the generator has additional diagonal terms as above.

Theorem 3.12 Feynman-Kac formula

Suppose that $(X_t : t \geq 0)$ is a Feller process on the state space E with generator \mathcal{L} , path measure \mathbb{P} and expectation \mathbb{E} , and take $f \in \mathcal{D}_{\mathcal{L}}$ and $g \in C(E)$. Define the operator

$$P_t^g f(x) := \mathbb{E}^x \left[f(X_t) \exp \left(\int_0^t g(X_s) ds \right) \right] .$$

Then $u(t, \cdot) = P_t^g f \in \mathcal{D}_{\mathcal{L}}$ for each $t \geq 0$ and $u(t, x)$ solves the equation

$$\frac{d}{dt}u(t, x) = \mathcal{L}u(t, x) + g(x)u(t, x), \quad u(0, x) = f(x). \quad (3.42)$$

$P_t^h = e^{t(\mathcal{L}+g)}$, $t \geq 0$ has a semi-group structure and is called the **Feynman-Kac semigroup** generated by $\mathcal{L} + g$. On compact time intervals $[0, T]$ it characterizes a (non-conservative) path measure \mathbb{P}_g with

$$d\mathbb{P}_g[\omega] = \exp\left(\int_0^T g(\omega_s) ds\right) d\mathbb{P}[\omega]. \quad (3.43)$$

Proof. Consider the joint process (X_t, Y_t) on $E \times \mathbb{R}$ with $Y_t = \int_0^t g(X_s) ds$. This is Markov with generator

$$\tilde{\mathcal{L}}F(x, y) = \mathcal{L}F(x, y) + \partial_y F(x, y)g(x).$$

Now, use the backward equation for the special test function $F(x, y) = f(x)e^y$ with $\partial_y F(x, y) = F(x, y)$ to get (3.42). See [31], Section 3.4.6 for an alternative proof with technical details including operator domains.

To see the semigroup structure, note that $P_0^g = \mathbb{I}$, and using the tower property of the expectation we get analogously to the standard case for each $0 < u < t$

$$\begin{aligned} P_t^g f(x) &= \mathbb{E}^x \left[\mathbb{E}^x \left[f(X_t) e^{\int_0^u g(X_s) ds + \int_u^t g(X_s) ds} \middle| \mathcal{F}_u \right] \right] \\ &= \mathbb{E}^x \left[e^{\int_0^u g(X_s) ds} \mathbb{E}^{X_u} \left[f(X_{t-u}) e^{\int_0^{t-u} g(X_s) ds} \right] \right] = P_u^g P_{t-u}^g f(x). \end{aligned} \quad (3.44)$$

Analogously to the conservative case, the statement on the Radon-Nikodym derivative of the path measure then follows from the definition of P_t^g . \square

Applying this in the above setting, the non-conservative Feynman-Kac semigroup generated by the tilted operator \mathcal{L}_k (3.41) can be written as

$$e^{t\mathcal{L}_k} f(x) = \bar{\mathbb{E}}_k^x \left[f(X_t) \exp\left(\int_0^t (\mathcal{L}_k 1)(X_s) ds\right) \right]. \quad (3.45)$$

The process generated by the conservative operator $\bar{\mathcal{L}}_k f(x)$ is a Markov jump process with tilted transition rates $c_k(x, dy) := c(x, dy)e^{ka(x,y)}$ and we denote the corresponding path measure and expectation by \mathbb{P}_k and $\bar{\mathbb{E}}_k$, respectively. With (3.43) we have on $[0, T]$

$$d\mathbb{P}_k[\omega] = \exp\left(\int_0^T (\mathcal{L}_k 1)(\omega_s) ds\right) d\bar{\mathbb{P}}_k[\omega]. \quad (3.46)$$

The tilted rates $c_k(x, dy)$ are absolutely continuous w.r.t. $c(x, dy)$ and vice versa, since $e^{ka(x,y)}$ is strictly positive. The Radon-Nikodym derivatives are simply given by

$$\frac{dc_k(x, \cdot)}{dc(x, \cdot)}(y) = e^{ka(x,y)} \quad \text{and} \quad \frac{dc(x, \cdot)}{dc_k(x, \cdot)}(y) = e^{-ka(x,y)}, \quad (3.47)$$

so the tilted and the original process with generator \mathcal{L} have the same set of allowed transitions. In this situation the corresponding path measures are absolutely continuous, and a formula for the Radon-Nikodym derivative is provided by the following result.

Theorem 3.13 Girsanov formula

Let \mathbb{P} and $\bar{\mathbb{P}}$ be path measures of two Feller jump processes on the same state space E on the compact time interval $[0, T]$, with rates $c(x, dy)$ and $\bar{c}(x, dy)$, respectively. We denote by $c(x) = \int_E c(x, dy)$ and $\bar{c}(x) = \int_E \bar{c}(x, dy)$ the total exit rates, and assume that $\bar{c}(x, dy) \ll c(x, dy)$ with corresponding Radon-Nikodym derivative $d\bar{c}(x, \cdot)/dc(x, \cdot)$. Then

$$d\bar{\mathbb{P}}[\omega] = \exp \left(\int_0^T (c(\omega_t) - \bar{c}(\omega_t)) dt + \sum_{0 \leq t \leq T} \log \frac{d\bar{c}(\omega_{t-}, \cdot)}{dc(\omega_{t-}, \cdot)}(\omega_t) \right) d\mathbb{P}[\omega].$$

Proof. by direct computation, see e.g. [8], Appendix 1, Proposition 2.6.

Note that in our setting, the difference of the total exit rates can be written as

$$\int_E c(x, dy) - \int_E \bar{c}(x, dy) = \int_E c(x, dy) (1 - e^{ka(x,y)}) = -(\mathcal{L}_k 1)(x) + k h(x).$$

Using further the simple expression (3.47) for derivatives of the jump rates, we get

$$d\bar{\mathbb{P}}_k[\omega] = \exp \left(\int_0^T (k h(\omega_t) - (\mathcal{L}_k 1)(\omega_t)) dt + \sum_{0 \leq t \leq T} ka(\omega_{t-}, \omega_t) \right) d\mathbb{P}[\omega] \quad (3.48)$$

Combining (3.46) and (3.48), the main result of this section is then the following.

Theorem 3.14 *The tilted path measure $\mathbb{P}_k = e^{TkAt} \mathbb{P}$ describes a non-conservative Markov jump process with generator \mathcal{L}_k given in (3.41) and associated Feynman-Kac semigroup (3.45).*

Proof. With (3.46) and (3.48) we get a cancellation of one integral term which yields

$$\frac{d\mathbb{P}_k}{d\mathbb{P}}(\omega) = \frac{d\mathbb{P}_k}{d\bar{\mathbb{P}}_k}(\omega) \frac{d\bar{\mathbb{P}}_k}{d\mathbb{P}}(\omega) = \exp \left(k \int_0^T h(\omega_t) dt + k \sum_{0 \leq t \leq T} a(\omega_{t-}, \omega_t) \right)$$

for the path measure generated by \mathcal{L}_k , which coincides with tilted measure defined in (3.36). \square

The operator \mathcal{L}_k has positive 'off-diagonal' entries, but their sum does not necessarily cancel with the 'diagonal' entries as for conservative operators. Still, the **Perron-Frobenius theorem** implies that for finite state spaces (where \mathcal{L}_k is a matrix) and irreducible dynamics, there exists a unique principal eigenvalue with maximal real part, which is real and the corresponding eigenvectors have positive entries. This can be extended to infinite-dimensional compact operators, see [29], Section III.B for details and further references.

Assume that $\Lambda_k \in \mathbb{R}$ is the unique principal eigenvalue of \mathcal{L}_k with corresponding left and right eigenvectors $\ell_k(dy)$ and $r_k(x)$, respectively, and positive spectral gap $\Delta_k > 0$. Then for every $f \in \mathcal{D}_{\mathcal{L}}$ the associated semigroup can be written as

$$P_t^k f(x) = e^{t\mathcal{L}_k} f(x) = e^{t\Lambda_k} \left(r_k(x) \int_E f(y) \ell_k(dy) + O(e^{-t\Delta_k}) \right),$$

if we use the normalization $\int_E \ell_k(dy) = 1$ and $\int_E r_k(x) \ell_k(dx) = 1$. If we further assume that the initial distribution μ is such that

$$\alpha := \int_E r_k(x) \mu(dx) \in (0, \infty), \quad (3.49)$$

we get

$$\mathbb{E}^\mu [e^{tkA_t}] = \mathbb{E}_k^\mu [1] = \int_E P_t^k 1(x) \mu(dx) = e^{t\Lambda_k} \left(\alpha + O(e^{-t\Delta_k}) \right).$$

With (3.27) this implies that the SCGF

$$\theta(k) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}_k^\mu [1] = \Lambda_k \quad (3.50)$$

is given by the principal eigenvalue of \mathcal{L}_k . The probabilistic meaning of the normalized left eigenvector is the limiting stationary distribution of the process under the grand-canonical path measure (3.35)

$$\ell_k(dy) = \lim_{t \rightarrow \infty} \frac{\mathbb{E}^\mu [e^{tkA_t} \delta_{X_t}(dy)]}{\mathbb{E}^\mu [e^{tkA_t}]} = \lim_{t \rightarrow \infty} \mathbb{P}_{gc}^\mu [X_t \in dy],$$

which holds independently of initial distributions μ that fulfill (3.49). So the principal eigenvalue and eigenvector of the tilted generator characterize the most important properties of the grand-canonical ensemble, which is equivalent to the canonical ensemble, and determine the large deviation rate function for the additive path functional A_t .

Cloning algorithms.

In order to sample from the non-conservative tilted path measure, one can use a cloning-type algorithm where an ensemble of trajectories is run in parallel. To understand the dynamics we re-write the tilted generator (3.41) once again as

$$\begin{aligned} \mathcal{L}_k f(x) &= \int_E c(x, dy) (e^{ka(x,y)} f(y) - f(x)) + kh(x) f(x) \\ &= \int_E c(x, dy) \sum_{n=0}^{\infty} q_n(k, x, y) (nf(y) - f(x)) \\ &\quad + \begin{cases} |kh(x)| (0 - f(x)) & , \text{ if } kh(x) < 0 \\ |kh(x)| (2f(x) - f(x)) & , \text{ if } kh(x) > 0 \end{cases} . \end{aligned} \quad (3.51)$$

Here we turn the real path weights into randomized integer weights, which can be interpreted as the number of clones or offspring produced at a jump event or due to the diagonal term. For jumps, $q_n(k, x, y)$ is a probability distribution on integers $n \in \mathbb{N}_0$ which is the number of copies the process is replaced with after the jump from x to y . If $ka(x, y) < 0$ we want to suppress these jumps, and the simplest choice is simply to use a Bernoulli distribution with $q_1(k, x, y) = e^{ka(x,y)} = 1 - q_0(k, x, y)$, so that the process is killed with probability $1 - e^{ka(x,y)}$. If $ka(x, y) > 0$ jumps from x to y are encouraged, and after the jump the process is replaced with a number of copies that has mean $\sum_{n=0}^{\infty} q_n(k, x, y) n = e^{ka(x,y)} > 1$. A simple and quite realistic choice would be to use a geometric distribution. But in fact any distribution on positive integers with the right mean is allowed (including e.g. also Poisson), and can be chosen to optimize the performance of the algorithm. A similar interpretation applies to the diagonal term, where we run additional dynamics with rates $|kh(x)|$, that lead to killing events if $kh(x) < 0$, or simple cloning events where one new copy is created if $kh(x) > 0$. Of course the number of copies could also be randomized for this term. In fact there are many other ways to set up similar cloning rules that are consistent with the generator \mathcal{L}_k , see e.g. [33] for more details.

As is implied by (3.50),

$$\mathbb{E}_k[1] \propto e^{t\Lambda_k} \quad \text{as } t \rightarrow \infty,$$

and since usually $\Lambda_k \neq 0$ the number of clones would decrease or increase exponentially on average. This leads to obvious computational problems and one usually runs a clone ensemble of constant size, say M . Whenever a clone is killed it is replaced by the copy of another uniformly chosen one. If several copies are created in an event, they replace uniformly chosen ones from the ensemble. Since this replacement mechanism is unbiased, the clone ensemble represents a discrete approximation of the grand-canonical path measure. The growth factor of the population in each clone i , where $m_i \geq -1$ clones have been created (random number depending on the path), is given by $(M + m_i)/M$. Denoting by $N(t)$ the total (random) number of clone events up to time t , with (3.50) we expect

$$\theta(k) = \Lambda_k = \lim_{M \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^{N(t)} \log(M + m_i)/M = \lim_{M \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{1}{Mt} \sum_{i=1}^{N(t)} m_i. \quad (3.52)$$

For the last identity we use $m_i \ll M$ as $M \rightarrow \infty$. For large finite ensemble sizes M and simulation times t this provides a converging approximation scheme for $\theta(k)$.

3.6 LDPs for Markov chains

As a first simple example consider a Poisson process $(N_t : t \geq 0) \sim PP(\lambda)$ with rate $\lambda > 0$. Then N_t/t fulfills an LDP with good rate function

$$I(a; \lambda) := \lambda - a + a \log(a/\lambda) \quad \text{for all } a \geq 0. \quad (3.53)$$

To see this, simply compute the SCGF $\theta(k)$, using that $N_t \sim Poi(\lambda t)$ to get

$$\mathbb{E}[e^{kN_t}] = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{kn} e^{-\lambda t} = \exp[\lambda t(e^k - 1)], \quad (3.54)$$

and therefore

$$\theta(k) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}[e^{kN_t}] = \lambda(e^k - 1). \quad (3.55)$$

This is smooth for all $k \in \mathbb{R}$, so the Gärtner-Ellis Theorem 3.9 applies and Legendre transformation leads to $I(a; \lambda)$ as given in (3.53). Note that $I(a; \lambda)$ is strictly convex for all $a \geq 0$ with $I(0; \lambda) = \lambda$ and $I(\lambda; \lambda) = 0$.

Following results in [30], we now describe a joint LDP for the empirical measure and flow of a continuous Markov chain $(X_t : t \geq 0)$ with state space E and rates $c(x, y)$. We assume the latter are uniformly bounded, i.e. $c(x) < R < \infty$ for all $x \in E$, excluding explosion, and that the chain is ergodic with unique stationary measure π . Both functionals are simple measure-valued generalizations of additive functionals considered before.

Definition 3.8 For given $T > 0$ the **empirical measure** $\mu_T : D_E([0, T]) \rightarrow \mathcal{M}_1(E)$ is given by

$$\mu_T = \frac{1}{T} \int_0^T \delta_{X_t} dt, \quad (3.56)$$

and the **empirical flow** $\mathcal{Q}_T : D_E([0, T]) \rightarrow L^1_+(E)$ is given by

$$\mathcal{Q}_T = \frac{1}{T} \sum_{t \in [0, T]: X_{t-} \neq X_t} \delta_{(X_{t-}, X_t)}. \quad (3.57)$$

Here $\delta_{(x,y)}$ is the Dirac measure on $E \times E$, and we denote by $L^1_+(E)$ all non-negative, summable sequences on $[0, \infty)^{E \times E}$, with associated L^1 -norm $\|\cdot\|_1$.

Note that μ_T and $T\mathcal{Q}_T$ are measure-valued, with the latter counting simultaneously the empirical currents $\mathcal{J}_T(x, y)$ between states $x, y \in E$ as introduced earlier in Definition 3.2. Also $c(x, y) = 0$ obviously implies $\mathcal{Q}_T(x, y) = 0$.

By the law of large numbers given in Theorem 3.1, $\mu_T[x] \rightarrow \pi[x]$, $\mathbb{P} - a.s.$ as $T \rightarrow \infty$ for each $x \in E$. Since we use the discrete topology on E this implies also weak convergence $\mu_T(f) \rightarrow \pi(f)$, $\mathbb{P} - a.s.$ for all $f \in C(E)$, which implies to $\mathcal{P} \circ \mu_T^{-1} \rightarrow \delta_\pi$ on $\mathcal{M}_1(E)$ under the topology of weak convergence.

For a given flow $\mathcal{Q} \in L^1_+(E)$ we denote its **divergence** $\text{div} \mathcal{Q} : L^1_+ \rightarrow \mathbb{R}$ as

$$\text{div} \mathcal{Q}(x) = \sum_{y \in E} \mathcal{Q}(x, y) - \sum_{y \in E} \mathcal{Q}(y, x) \quad \text{for all } x \in E, \quad (3.58)$$

measuring the net flow out of state x . For each $\mu \in \mathcal{M}_1(E)$ we associate the **flow** $\mathcal{Q}^\mu \in L^1_+$ defined as

$$\mathcal{Q}^\mu(x, y) = \mu(x) c(x, y) \quad \text{for all } x, y \in E. \quad (3.59)$$

Therefore $\text{div} \mathcal{Q}^\mu = 0$ if and only if $\mu = \pi$ is stationary. With Theorem 3.7 we have the LLN $\mathcal{Q}_T(x, y) \rightarrow \mathcal{Q}^\pi(x, y)$, $\mathbb{P} - a.s.$ as $T \rightarrow \infty$ for all $x, y \in E$. Note that for finite E this implies a LLN on L^1_+ w.r.t. the norm $\|\cdot\|_1$, while for infinite E in general a stronger topology is required also to formulate an LDP. One choice is to use a bounded version of the weak* topology on $L^1(E)$, with full details given in Section 2.3 of [30]. In general, $L^1(E)$ is the dual to $C(E \times E)$ of functions vanishing at infinity, and the **weak* topology** on $L^1(E)$ is the smallest topology such that all maps $\mathcal{Q} \mapsto \langle \mathcal{Q}, f \rangle \in \mathbb{R}$ with $f \in C(E \times E)$ are continuous.

We need one more concept before stating the main result. Recalling the definition of the Dirichlet form (3.8), for the symmetric part \mathcal{L}_s (2.56) of the generator for a general Markov chain with rates $c(x, y)$ this gives

$$\mathcal{E}_\pi(f) = \frac{1}{4} \sum_{x, y \in E} (\pi[x]c(x, y) + \pi[y]c(y, x)) (f(y) - f(x))^2, \quad f \in L^2(E, \pi). \quad (3.60)$$

The process satisfies a **logarithmic Sobolev inequality**, if there exists a constant $C_{LS} \in (0, \infty)$ such that for any $\mu \in \mathcal{M}_1(E)$

$$H(\mu; \pi) \leq C_{LS} \mathcal{E}_\pi(\sqrt{\mu/\pi}). \quad (3.61)$$

Note that the density $\mu/\pi : E \rightarrow [0, \infty)$ is well defined since $\pi[x] > 0$ for all $x \in E$.

Theorem 3.15 *Consider an ergodic Markov chain with uniformly bounded rates, satisfying (3.61) and that the graph (E, G) of allowed transitions with $G = \{(x, y) \in E \times E : c(x, y) > 0\}$ is locally finite. Then, as $T \rightarrow \infty$, the family of measure-valued observables*

$$((\mu_T, \mathcal{Q}_T) : T \geq 0) \quad \text{on } \mathcal{M}_1(E) \times L^1_+(E) \quad \text{satisfies an LDP w.r.t. } \mathbb{P}^x, \quad (3.62)$$

uniformly in x on compact subsets of E with good and convex rate function

$$I(\mu, \mathcal{Q}) := \begin{cases} \sum_{(y,z) \in G} I(\mathcal{Q}(y, z); \mathcal{Q}^\mu(y, z)) & , \text{ if } \operatorname{div} \mathcal{Q} = 0 \\ \infty & , \text{ otherwise} \end{cases} , \quad (3.63)$$

which is determined by the rate functions (3.53) of Poisson processes, where we set $I(q; 0) = \infty$ for all $q > 0$.

Idea of the proof. Note that by the graphical construction of the Markov chain, the integer-valued random field $\{T\mathcal{Q}_T(y, z) : (y, z) \in G\}$ under \mathbb{P}^x is stochastically dominated (see Definition 4.4) by the field $\{N_T^{y,z} : (y, z) \in G\}$ of iid Poisson random variables with parameters $c(y, z)T$. Due to independence, the rate function for the latter is simply given by the sum of rate functions $I(q; c(y, z)T)$ for individual Poisson processes given in (3.53). The connection is then based on an extension of the **contraction principle**:

Let \mathcal{A} and \mathcal{B} be two complete, separable metric spaces and $(A_t : t \geq 0)$ a family on \mathcal{A} satisfying an LDP with rate function $I_A : \mathcal{A} \rightarrow [0, \infty]$ w.r.t. a measure \mathbb{P} . Let $F : \mathcal{A} \rightarrow \mathcal{B}$ be a continuous function. Then $(F(A_t) : t \geq 0)$ satisfies an LDP on \mathcal{B} w.r.t. $\mathbb{P} \circ F^{-1}$ with rate function

$$I_B(b) = \inf \{ I_A(a) : a \in \mathcal{A}, F(a) = b \} .$$

If I_A is a good rate function on \mathcal{A} , so is I_B on \mathcal{B} .

The empirical measure μ leads to a thinning of $PP(c(y, z))$ processes, resulting in a modified parameter $\mathcal{Q}^\mu(y, z) = \mu[y]c(y, z)$. Furthermore, the asymptotic flows \mathcal{Q} have to be divergence free, corresponding to conservation of probability of the Markov chain. These two conditions lead to a mapping, which has to be shown to be continuous in a proper setting. \square

Note that of course $I(\mu, \mathcal{Q}^\mu) = 0$ if and only if $\operatorname{div} \mathcal{Q}^\mu = 0$, which is equivalent to $\mu = \pi$ being stationary. A further contraction using the mapping

$$(\mu, \mathcal{Q}) \mapsto \mu \quad \text{on } \mathcal{M}_1(E) \times L_+^1(E) \rightarrow \mathcal{M}_1(E) ,$$

yields the classical **Donsker-Varadhan LDP** for the empirical measures (see e.g. [28]).

As mentioned earlier, the above result holds for a bounded weak* topology on the flow space $L^1(E)$. A more general version on the flow space $[0, \infty]^G$ endowed with the product topology can be formulated with a less explicit construction of the rate function. For details on this aspect, on the proof and also on further properties of the rate function (3.63) see [30].

4 Hydrodynamic limits and macroscopic fluctuations

Hydrodynamic limits are scaling limits of the dynamics of conservative IPS on large space and time scales, where the system is completely described by the local density field $\rho(s, u)$ as a function of a continuous space coordinate $u \in \mathbb{R}$ and time $s \geq 0$. This is the solution of a conservation law (PDE) of the following form

$$\partial_s \rho(u, s) + \partial_u j(u, s) = 0 \quad \text{and} \quad j(u, s) = J(\rho(u, s), u, s), \quad (4.1)$$

where we only consider the one-dimensional case for simplicity. $j(u, s)$ is the local current determined by the second relation called the **constitutive equation**. For IPS this usually takes the form

$$J(\rho, s) = -D(\rho)\partial_u \rho + \chi(\rho)E(u, s), \quad (4.2)$$

with **diffusion coefficient** $D(\rho) \geq 0$, **mobility** $\chi(\rho) > 0$ and **external field** $E(u, s) \in \mathbb{R}$. In d dimensions, both are symmetric, non-negative definite matrices. The two equations can be rewritten as

$$\partial_s \rho(u, s) + \partial_u (\chi(\rho(u, s))E(u, s)) = \partial_u (D(\rho(u, s))\partial_u \rho(u, s)), \quad (4.3)$$

and since χ and D depend on ρ this is in general non-linear which is the source of interesting phenomena. For systems on finite domains boundary conditions have to be specified, for now we will focus simply on periodic boundary conditions or systems on infinite domains. In the following we will investigate how to derive (4.3) from IPS, starting out with some heuristics to fix the main ideas.

4.1 Heuristics on currents and conservation laws

Consider a lattice gas on the compact state space $E = S^\Lambda$ with compact $S \subseteq \mathbb{N}_0$ with generator (1.79)

$$\mathcal{L}f(\eta) = \sum_{x, y \in \Lambda} c(x, y, \eta) (f(\eta^{x,y}) - f(\eta)). \quad (4.4)$$

As before we assume the jump rates to be of finite range as given in Definition (1.6), and we denote lattice coordinates as $x, y, z \in \Lambda$ in all of this section. Recall the forward equation from Theorem 1.4

$$\frac{d}{dt} P_t f = P_t \mathcal{L}f \quad \text{which holds for all } f \in C^0(E). \quad (4.5)$$

Integrating w.r.t. the initial distribution μ , and writing $\mu_t := \mu P_t$ as before we have

$$\frac{d}{dt} \mu_t(f) = \frac{d}{dt} \mu(P_t f) = \mu(P_t \mathcal{L}f) = \mu_t(\mathcal{L}f). \quad (4.6)$$

For the bounded cylinder function $f(\eta) = \eta(x)$ we get with (4.4)

$$\mathcal{L}\eta(x) = \sum_{y \in \Lambda} (c(y, x, \eta) - c(x, y, \eta)) = -\text{div} j(x, \eta), \quad (4.7)$$

with notation explained in the following definition.

Definition 4.1 We denote the **instantaneous particle current** across the directed edge (x, y) as

$$j(x, y, \eta) := c(x, y, \eta) - c(y, x, \eta) , \quad (4.8)$$

and its **divergence** is defined similarly to (3.58) as

$$\operatorname{div} j(x, \eta) := \sum_{y \in \Lambda} j(x, y, \eta) . \quad (4.9)$$

The **average particle density** at site x at time t is given by

$$\rho(x, t) := \mathbb{E}^\mu(\eta_t(x)) = \mu_t(\eta(x)) , \quad (4.10)$$

and the **average particle current** across (x, y) by

$$j(x, y, t) := \mathbb{E}[j(x, y, \eta_t)] = \mu_t(j(x, y, \cdot)) . \quad (4.11)$$

With (4.7), the forward equation (4.6) for $f(\eta) = \eta(x)$ can then be written as

$$\frac{d}{dt} \rho(x, t) + \operatorname{div} j(x, t) = 0 , \quad (4.12)$$

which is also called the **lattice continuity equation**. The divergence of the average current is defined analogously to (4.9). Note that in contrast to the flow $\mathcal{Q}(x, y)$ in (3.58) which is non-negative, $j(x, y, \eta) \in \mathbb{R}$ denotes already the *net* current across the directed bond (x, y) , which leads to the difference in the definitions of the divergence. It describes the time evolution of the density $\rho(x, t)$ in terms of higher order (two-point) correlation functions which determine the current. The form of this equation implies that the total particle density in closed systems is conserved, i.e. on the finite lattice $\Lambda_L = \mathbb{Z}/L\mathbb{Z}$ with periodic boundary conditions we have

$$\frac{d}{dt} \sum_{x \in \Lambda_L} \rho(x, t) = - \sum_{x \in \Lambda_L} \operatorname{div} j(x, t) = 0 . \quad (4.13)$$

In general, the density is locally conserved and a change in certain domain $\Delta \subseteq \Lambda$ only occurs through flux through the boundary.

Example. For the one-dimensional ASEP on $\Lambda = \mathbb{Z}$ or $\Lambda_L = \mathbb{Z}/L\mathbb{Z}$ we have for all $x \in \Lambda$

$$j(x, x+1, t) = p\mu_t(1_x 0_{x+1}) - q\mu_t(0_x 1_{x+1}) , \quad (4.14)$$

with $j(x, x-1, t) = -j(x-1, x, t)$ and all other currents vanish. Here we have used the simplifying notation $\mu_t(\eta(x)(1-\eta(y))) = \mu_t(1_x 0_y)$. This leads to the divergence

$$\operatorname{div} j(x, t) = p\mu_t(1_x 0_{x+1}) - q\mu_t(0_x 1_{x+1}) + q\mu_t(0_{x-1} 1_x) - p\mu_t(1_{x-1} 0_x) . \quad (4.15)$$

Symmetric case For $p = q = 1$ this simplifies significantly. Adding and subtracting an auxiliary term we get

$$\begin{aligned} j(x, x+1, t) &= \mu_t(1_x 0_{x+1}) + \mu_t(1_x 1_{x+1}) - \mu_t(1_x 1_{x+1}) - \mu_t(0_x 1_{x+1}) = \\ &= \mu_t(1_x) - \mu_t(1_{x+1}) = \rho(x, t) - \rho(x+1, t) . \end{aligned} \quad (4.16)$$

So the current is given by the lattice derivative of the density, and (4.12) turns into a closed equation

$$\frac{d}{dt}\rho(x, t) = \Delta_x \rho(x, t) := \rho(x - 1, t) - 2\rho(x, t) + \rho(x + 1, t). \quad (4.17)$$

Thus the particle density of the SSEP behaves like the probability density of a single simple random walk with jump rates $p = q = 1$. The fact that we get a closed equation for particle density (one-point correlation function) is related to the concept of **duality**, which is explained in detail in [6], Chapter II and in Chapter IIIV for the excusion process. In general, duality allows to study the dynamics of correlation functions for one model in terms of the dynamics of another dual model, which is often simpler than the original one.

To describe the behaviour on large scales we scale the lattice constant by a factor of $1/L$ and embed it in the continuum, i.e. $\frac{1}{L}\Lambda \subseteq \mathbb{R}$ or $\frac{1}{L}\Lambda_L \subseteq \mathbb{T} = \mathbb{R}/\mathbb{Z}$ for the torus. Using the macroscopic space variable $u = x/L \in \mathbb{R}, \mathbb{T}$ we assume that there exists a smooth function $y \mapsto \tilde{\rho}(\cdot, t)$ for all $t \geq 0$ such that

$$\tilde{\rho}(x/L, t) := \rho(x, t) \quad \text{for all } x \in \Lambda, \quad (4.18)$$

which we call the **macroscopic density field**. A Taylor expansion around the point $y = x/L$ then yields

$$\rho(x \pm 1, t) = \tilde{\rho}(u \pm \frac{1}{L}, t) = \tilde{\rho}(u, t) \pm \frac{1}{L}\partial_u \tilde{\rho}(u, t) + \frac{1}{2L^2}\partial_u^2 \tilde{\rho}(u, t) + O(1/L^3). \quad (4.19)$$

The lattice Laplacian in (4.17) is then given by

$$\Delta_x \rho(x, t) = \frac{1}{L^2}\partial_u^2 \tilde{\rho}(u, t) + O(1/L^4), \quad (4.20)$$

since first and third order terms vanish due to symmetry. In order to get a non-degenerate equation in the limit $L \rightarrow \infty$, we have to scale time as $s = t/L^2$. This corresponds to speeding up the process by a factor of L^2 , in order to see diffusive motion of the particles on the scaled lattice. Using both in (4.17) we obtain in the limit $L \rightarrow \infty$

$$\partial_s \rho(u, s) = \partial_u^2 \rho(u, s) \quad (\text{heat/diffusion equation}), \quad (4.21)$$

where we drop the notation $\tilde{\rho}$ and simply use again ρ for the density in macroscopic time. Note that this equation also describes the diffusion of independent particles on large scales, since the SSEP can also be defined as simply swapping occupation numbers with rate 1 (irrespective of occupation). In the framework of (4.3) we have $D(\rho) = 1$ and no external driving field $E \equiv 0$. From a PDE point of view, (4.21) is a **parabolic** equation, for which the initial value problem has a maximum principle and a unique smooth solution for all $t > 0$ even for non-smooth initial data. For further details see e.g. [8], Chapter 4.

Asymmetric case. In the asymmetric case with $p \neq q$ we do not get a closed equation as in (4.17) and the constitutive equation for the current is more complicated. If we use a stationary measure $\mu_t = \mu$ in the continuity equation (4.12) we get

$$0 = \frac{d}{dt}\mu(1_x) = j(x - 1, x) - j(x, x + 1), \quad (4.22)$$

which implies that the **stationary current** $j(x, x + 1) := p\mu(1_x 0_{x+1}) - q\mu(0_x 1_{x+1})$ is site-independent. Since we know the stationary measures for the ASEP we can compute it explicitly. For the homogeneous product measure $\mu = \nu_\rho$ we get

$$j(x, x + 1) := p\nu_\rho(1_x 0_{x+1}) - q\nu_\rho(0_x 1_{x+1}) = (p - q)\rho(1 - \rho) =: \phi(\rho), \quad (4.23)$$

which is actually just a function of the total particle density $\rho \in [0, 1]$. We can use this to arrive at a scaling limit of the continuity equation for the asymmetric case $p \neq q$. We use the same space scaling $u = x/L$ as above and write

$$\operatorname{div} j(x, t) = \frac{1}{L} \partial_u \tilde{j}(u, t) + o\left(\frac{1}{L}\right), \quad (4.24)$$

with a similar notation \tilde{j} for the current on the macroscopic scale as above for the density. In the asymmetric case the first order terms in the spatial derivative do not vanish and we have to scale time as $s = t/L$, speeding up the process only by a factor L to see motion with a drift. In the limit $L \rightarrow \infty$ this leads to the **hyperbolic conservation law**

$$\partial_s \rho(u, s) + \partial_u j(u, s) = 0, \quad (4.25)$$

where we can define the macroscopic current $j(u, s)$ as (assuming that the limit exists)

$$\tilde{j}(u, s) := \lim_{L \rightarrow \infty} j([uL] - 1, [uL], sL). \quad (4.26)$$

Since we effectively take microscopic time $t = sL \rightarrow \infty$ in that definition, it is plausible to use

$$j(u, s) = \phi(\rho(u, s)), \quad (4.27)$$

i.e. the stationary current corresponding to the local density $\rho(u, s)$ gives the constitutive equation for the current. This is equivalent to the process becoming locally stationary in the limit $L \rightarrow \infty$, the only (slowly) varying quantity remaining on a large scale is the macroscopic density field. In general, for a lattice gas with stationary product measures ν_ρ indexed by the density **Local stationarity** (also called **local equilibrium**) w.r.t. a density profile $\rho(u, s)$ holds if

$$\mu_{sL} \circ \tau_{-[uL]} \rightarrow \nu_{\rho(u, s)} \quad \text{weakly (locally), as } L \rightarrow \infty. \quad (4.28)$$

By local weak convergence of the translated distributions μ_{Ls} we mean

$$\mu_{Ls} \circ \tau_{-[uL]}(f) \rightarrow \nu_{\rho(u, s)}(f) \quad \text{for all } f \in C^0(E). \quad (4.29)$$

Note that the limit $\nu_{\rho(u, s)}$ is a now measure on the infinite lattice, \mathbb{Z} in our case. This can also hold for stationary measures ν_ρ which are not of product form, as long as they can be defined on the infinite lattice, i.e. they should be Gibbs measures.

For the ASEP, local equilibrium holds if $\rho(u, s)$ is a solution of the conservation law

$$\partial_s \rho(u, s) + \partial_u \phi(\rho(u, s)) = 0 \quad \text{where } \phi(\rho) = (p - q)\rho(1 - \rho). \quad (4.30)$$

This has been established rigorously for the ASEP and more general IPS in a so-called **hydrodynamic limit**, which we come back to in the next subsection. In the framework of (4.3) we have now a mobility $\chi(\rho) = \rho(1 - \rho)$ with a constant external field $E = p - q$ and vanishing diffusion $D = 0$. In terms of PDEs, (4.30) is a **hyperbolic conservation law**, the solution theory of which is much more complex than for parabolic equations. Solutions can develop singularities even for

smooth initial data, leading to the concept of weak solutions, which are in general not unique without additional selection criteria. For details see e.g. [26], and also the next subsection.

Weakly asymmetric case. In order to get non-zero drift and diffusion on the same scale, one has to consider a weakly asymmetric version of the IPS. For the ASEP we can choose e.g.

$$p = 1 + \frac{E}{2L}, \quad q = 1 - \frac{E}{2L} \quad \text{with a constant external field } E \in \mathbb{R}.$$

Then the above arguments with space and time scales $u = x/L$, $s = t/L^2$ will lead to a scaling limit of the form (4.3) with $D = 1$, $\chi(\rho) = \rho(1 - \rho)$ and external field E .

4.2 Rigorous results on hydrodynamic limits

We consider lattice gases on the state space $E = S^\Lambda$ where $S \subseteq \mathbb{N}_0$ is a connected set including 0, and $\Lambda = \mathbb{Z}$ or $\mathbb{Z}/(L\mathbb{Z})$ as before where the rescaled lattices Λ/L are embedded in \mathbb{R} or the unit torus \mathbb{T} , respectively. In the previous section we presented some heuristics to derive a candidate PDE for a hydrodynamic limit, which was based on computations for expected particle densities. Rigorous hydrodynamic limits have been established in a yet stronger sense, corresponding to weak laws of large numbers for the evolution of empirical density profiles.

Definition 4.2 For each $t \geq 0$ we define the **empirical measure**

$$\pi_t^L[du] := \frac{1}{L} \sum_{x \in \Lambda} \eta_t(x) \delta_{x/L}[du] \in \mathcal{M}_+(\mathbb{R}) \text{ or } \mathcal{M}_+(\mathbb{T}), \quad (4.31)$$

and the measure-valued process $(\pi_t^L : t \geq 0)$ is called the **empirical process**.

The π_t^L describe the discrete particle densities on \mathbb{R} , \mathbb{T} . They are (random) measures depending on the configurations η_t and for $A \subseteq \mathbb{R}, \mathbb{T}$ we have

$$\pi_t^L(A) = \frac{1}{L} \left(\# \text{ of particles in } A \cap \frac{1}{L}\Lambda \text{ at time } t \right). \quad (4.32)$$

In the following two subsections we will discuss how to derive rigorously parabolic equations for a certain class of reversible systems and hyperbolic equations for a certain class of asymmetric processes.

4.2.1 Reversible gradient systems

In this section we derive parabolic equations for lattice gases on the macroscopic time scale $s = t/L^2$, which corresponds to speeding up the microscopic process by a factor of L^2 , multiplying the generator of the process. Recall the generator

$$\mathcal{L}f(\eta) = \sum_{x,y \in \Lambda} c(x,y,\eta) (f(\eta^{x,y}) - f(\eta)).$$

with bounded rates of finite range, and the instantaneous current (4.8)

$$j(x,y,\eta) = c(x,y,\eta) - c(y,x,\eta).$$

We assume that the model is translation invariant and has a (possibly bounded) family of reversible stationary measures ν_ρ indexed by the density $\rho \in [0, \max S]$ or $[0, \infty)$. These may not be of product form, then they should correspond to limiting (Gibbs) measures on the infinite lattice \mathbb{Z} . We have seen in the previous section that the stationary expectation of the current is translation invariant, and many systems in fact satisfy the following (stronger) property, which facilitates computation of the diffusion coefficient.

Definition 4.3 A translation invariant lattice gas on a regular lattice with instantaneous current $j(x, y, \eta)$ is of **gradient type**, if there exists a positive local function $h \in C_0(E)$ and $r : \Lambda \rightarrow [0, \infty)$ such that for all $x, y \in \Lambda$

$$j(x, y, \eta) = r(y - x)(h(\tau_x \eta) - h(\tau_y \eta)) \quad \text{and} \quad r(y - x) = r(x - y). \quad (4.33)$$

For one-dimensional Λ , we denote by $m_k(r) = \sum_{z \in \Lambda} z^k r(z) \in (0, \infty)$ the k -th moment of r for all $k \geq 0$. In higher dimensions there is an obvious generalization of the following result where the diffusivity is replaced by a diagonal matrix where each element is multiplied with the corresponding moment. The above definition can be generalized to a less restrictive condition, as can be found in [8], Section 4.2. The sign convention for h and r is important to get well-posed limit equations with positive-definite diffusivity.

Theorem 4.1 Consider a sequence of lattice gases $(\eta_t^L : t \geq 0)$ satisfying the above conditions, with initial distributions μ^L such that

$$\pi_0^L[du] \rightarrow \rho(u, 0) du \quad \text{weakly, in probability as } L \rightarrow \infty, \quad (4.34)$$

i.e. there exists a limiting density $\rho(\cdot, 0) \in (0, \infty)$ w.r.t. Lebesgue measure on \mathbb{R} or \mathbb{T} . Then

$$\pi_{sL}^L[du] \rightarrow \rho(u, s) du \quad \text{weakly, in probability as } L \rightarrow \infty, \quad (4.35)$$

where $\rho(u, s)$ is a solution of the parabolic equation (4.3) with vanishing mobility $\chi = 0$,

$$\partial_s \rho(u, s) = \frac{1}{2} m_2(r) \partial_u (D(\rho(u, s)) \partial_u \rho(u, s)) = \frac{1}{2} m_2(r) \partial_u^2 \Phi(\rho(u, s)) \quad (4.36)$$

on \mathbb{R} or \mathbb{T} with initial condition $\rho(u, 0)$. The diffusion coefficient is given through the gradient condition (4.33) by expectation w.r.t. the stationary measures

$$D(\rho) = \Phi'(\rho) \quad \text{with} \quad \Phi(\rho) = \nu_\rho(h). \quad (4.37)$$

As usual weak convergence means that for every $g \in C_b(\mathbb{R}), C_b(\mathbb{T})$ with compact support

$$\pi_{sL}^L(g) = \frac{1}{L} \sum_{x \in \Lambda} g(x/L) \eta_{sL}(x) \rightarrow \int_{\mathbb{R}, \mathbb{T}} g(u) \rho(u, s) du. \quad (4.38)$$

The left-hand side is random, and convergence holds in probability, i.e. for all $\epsilon > 0$

$$\mathbb{P}^\mu \left(\left| \frac{1}{L} \sum_{x \in \Lambda} g(x/L) \eta_{sL}(x) - \int_{\mathbb{R}, \mathbb{T}} g(y) \rho(u, s) du \right| > \epsilon \right) \rightarrow 0 \quad \text{as } L \rightarrow \infty. \quad (4.39)$$

Proof. We give an outline of the proof for following the presentation in [35], for a full proof see [8], Chapters 4 and 5, and also [36], Section 8. For simplicity we focus on periodic boundary conditions and nearest neighbour dynamics. In a first step a usual strategy is to show existence of

a limit for the sequence of empirical measures $\pi_t^L[du]$ for each $t > 0$ using a standard compactness argument. On the torus \mathbb{T} this is very simple, since the total density

$$\pi_t^L[\mathbb{T}] = \pi_0^L[\mathbb{T}] = \int_{\mathbb{T}} \rho(u, 0) du =: \rho < \infty$$

is conserved and finite under the initial conditions. Therefore $\pi_t^L[du]$ is a sequence in the set of bounded measures on the compact set \mathbb{T} , which is itself compact. Then the martingale problem is used to characterize the limit process as a weak solution of the equation (4.36), for which uniqueness has (in this case) been established in the PDE literature.

Our starting point for this main part is the martingale characterization of the process speeded up by L^2 . This implies with $\mathcal{L}\eta(x)$ given in (4.7) that

$$\eta_t(x) - \eta_t(0) = -L^2 \int_0^t \operatorname{div} j(x, \eta_s) ds + M_t(x),$$

where $M_t(x)$ is a martingale. Note that t here denotes a macroscopic time. Multiplying with a smooth test function $\psi : \mathbb{T} \rightarrow \mathbb{R}$, dividing by L and summing over x we can write this in terms of the empirical measure (4.31)

$$\int_{\mathbb{T}} \psi(u) \pi_t^L[du] - \int_{\mathbb{T}} \psi(u) \pi_0^L[du] = -L \int_0^t \sum_{x \in \Lambda} \operatorname{div} j(x, \eta_s) \psi(x/L) ds + o(1). \quad (4.40)$$

The $o(1)$ term on the right is the combined contribution of the martingales, which can be shown vanish as $L \rightarrow \infty$. Using the gradient condition (4.33) we get for the main contribution

$$L \sum_{x \in \Lambda} \sum_{z \in \Lambda} r(z) (\psi(x/L) - \psi((x+z)/L)) \int_0^t h(\tau_x \eta_s) ds.$$

Taylor expansion of the smooth function ψ around x/L implies to leading order

$$\left(\sum_{z \in \Lambda} \frac{z^2}{2} r(z) \right) \frac{1}{L} \sum_{x \in \Lambda} \psi''(x/L) \int_0^t h(\tau_x \eta_s) ds,$$

where the first order contribution vanishes due to symmetry $r(z) = r(-z)$. In general, h is a non-linear function and the summation is therefore not given by an average w.r.t. the empirical measure (as is the case for the SSEP). Using again smoothness of ψ we can introduce an auxiliary sum to get

$$\frac{1}{L} \sum_{x \in \Lambda} \int_0^t \frac{1}{|\Delta_x|} \sum_{y \in \Delta_x} h(\tau_y \eta_s) \psi''(x/L) ds,$$

where Δ_x is a microscopically large and macroscopically small volume around $x \in \Lambda$ on an intermediate scale $1 \ll |\Delta_x| \ll L$. The central part of the proof is now to show a **Replacement Lemma**, substituting the spatial average of the gradient function

$$\frac{1}{|\Delta_x|} \sum_{y \in \Delta_x} h(\tau_y \eta) \quad \text{by} \quad \Phi \left(\frac{1}{|\Delta_x|} \sum_{y \in \Delta_x} \eta(y) \right), \quad (4.41)$$

i.e. by its stationary expectation as given in (4.37) under the local density. This is equivalent to establishing a local equilibrium for the process, and allows to write a 'closed form' of equation

(4.40) in terms of the empirical measure up to vanishing correction terms. Taking the limit $L \rightarrow \infty$, this implies that every limit point $\rho(u, t)du$ of the empirical measures satisfies

$$\int_{\mathbb{T}} \psi(u) \rho(u, s) du - \int_{\mathbb{T}} \psi(u) \rho(u, 0) du = \int_0^t ds \int_{\mathbb{T}} \Phi(\rho(u, s)) \psi''(u) du .$$

This is the weak form of a parabolic PDE, which are known to have unique smooth solutions for $t > 0$ even for non-smooth initial data. This implies that the limit point of the sequence $\pi_t^L[du]$ is unique, and in particular absolutely continuous w.r.t. Lebesgue measure for all $t \geq 0$ (under assumption on the initial condition). \square

Examples. The SSEP is a gradient model with $h(\eta) = \eta(0)$ and $r(z) = \delta_{|z|,1}$, since as in (4.16)

$$j(x, x+1, \eta) = \eta(x)(1 - \eta(x+1)) - \eta(x+1)(1 - \eta(x)) = \eta(x) - \eta(x+1) .$$

With stationary product Bernoulli measures ν_ρ we have $\nu_\rho(h) = \rho$ and consequently $D = 1$ consistent with (4.21).

The **zero-range process** (ZRP) is a lattice gas on the state space $E = \mathbb{N}_0^\Lambda$ with the generator

$$\mathcal{L}f(\eta) = \sum_{x,y \in \Lambda} p(x,y) g(\eta(x)) (f(\eta^{x,y}) - f(\eta)) , \quad (4.42)$$

where $p(x,y)$ is an irreducible transition kernel on Λ . The jump rates $g : \mathbb{N}_0 \rightarrow [0, \infty)$ depend on the particle configuration only through the departure site of a jump (zero-range), and to avoid degeneracies one usually assumes $g(n) = 0 \Leftrightarrow n = 0$. For symmetric, translation invariant dynamics with $p(x,y) = p(y,x) = r(y-x)$ this system is obviously gradient with

$$j(x,y,\eta) = r(y-x)(g(\eta(x)) - g(\eta(y))) ,$$

which leads to a diffusivity

$$D(\rho) = \frac{d}{d\rho} \Phi(\rho) \quad \text{with} \quad \Phi(\rho) := \nu_\rho(g) .$$

Another nice fact about the ZRP is that the ν_ρ are product measures with explicit formulas for the marginals. In contrast to the exclusion process, g is in general a non-linear gradient function and therefore the replacement lemma (4.41) is non-trivial and requires additional regularity assumptions on the rates. For the ZRP, due to the unbounded local state space a cut-off argument has to be made in the proof of the replacement lemma, which requires the jump rates $g(n)$ to grow at most linearly in n .

There is an alternative strategy of proof called the **relative entropy method**, which also covers reversible non-gradient models, but the stationary measures have to be of product form. It is based on a Gronwall-type argument for the relative entropy between the empirical measure and the local equilibrium $\nu_{\rho(u,s)}$ as given in (4.28), where $\rho(u,s)$ is a weak solution of the limit equation. This is essentially a contraction-type argument with a unique fixed point given by the local equilibrium, so it implies uniqueness of the weak solution of the limit equation, but requires that this is smooth and in fact a strong solution. This method is therefore limited to the reversible case with parabolic limit equations. Details can be found in [8], Chapter 6.

4.2.2 Weakly asymmetric systems

In general, non-reversible systems even with weak asymmetry are not of gradient form. The diffusion coefficient can in general be very hard to compute, and is linked to the microscopic dynamics by a **Green-Kubo** formula, see [3], Section II.2.2. But if the model has a family ν_ρ of stationary measures indexed by the density, which is the same for symmetric and asymmetric dynamics (and therefore also for weak asymmetry), then the diffusion coefficient is related to the mobility via the **Einstein relation**

$$D(\rho) = \chi(\rho)F''(\rho). \quad (4.43)$$

Here $F(\rho)$ is the equilibrium **free energy density**, which for systems with a single conserved quantity is simply given by the relative entropy between ν_ρ and an a-priori measure $\nu_{\bar{\rho}}$ with some fixed density $\bar{\rho}$. The choice of $\bar{\rho}$ is arbitrary and only changes the constant and linear part of F , which are irrelevant for the second derivative. As we will see in the next section, the mobility is determined by a simple stationary expectation of the instantaneous current and is often easy to calculate explicitly. In particular, Einstein relations hold for gradient models as a consequence of the following result.

Proposition 4.2 *Consider a translation invariant lattice gas on a regular lattice of gradient type as given in Definition 4.3 with rates $c(x, y, \eta)$ and reversible distribution π . Then π is stationary for any asymmetric version of the process with rates $q(y-x)c(x, y, \eta)$, where q is a translation invariant, irreducible transition kernel on Λ .*

Proof. Consider the generator of the asymmetric version

$$\mathcal{L}f(\eta) = \sum_{x, y \in \Lambda} q(y-x)c(x, y, \eta)(f(\eta^{xy}) - f(\eta)),$$

where detailed balance holds for the original rates as

$$\frac{\pi[\eta]}{\pi[\eta^{x,y}]} = \frac{c(y, x, \eta^{xy})}{c(x, y, \eta)} \quad \text{for all } \eta \in E \text{ and } x, y \in \Lambda.$$

Note that for uncountable E the left-hand side has to be interpreted as a Radon-Nikodym derivative. Then the standard computation gives

$$\begin{aligned} \pi(\mathcal{L}f) &= \int_E f(\eta) d\pi[\eta] \sum_{x, y \in \Lambda} q(y-x) \left(\frac{\pi[\eta^{yx}]}{\pi[\eta]} c(x, y, \eta^{yx}) - c(x, y, \eta) \right) \\ &= \int_E f(\eta) d\pi[\eta] \sum_{x, y \in \Lambda} q(y-x) (c(y, x, \eta) - c(x, y, \eta)), \end{aligned} \quad (4.44)$$

where we use a change of variables in the first, and detailed balance in the second line. Now we use the gradient condition (4.33) and another change of variables to get

$$\pi(\mathcal{L}f) = \int_E f(\eta) d\pi[\eta] \sum_{x \in \Lambda} h(\tau_x \eta) \sum_{z \in \Lambda} r(z) (q(z) - q(-z)) = 0,$$

since $r(z) = r(-z)$ and the z -summation vanishes. \square

Examples. For the simple exclusion process we have stationary Bernoulli product measures, so

$$F(\rho) = \rho \log \frac{\rho}{\bar{\rho}} + (1 - \rho) \log \frac{1 - \rho}{1 - \bar{\rho}} = \rho \log \rho + (1 - \rho) \log(1 - \rho) + \log 2 ,$$

if we choose $\bar{\rho} = 1/2$ for simplicity, where the constant $\log 2$ is irrelevant. This leads to

$$F''(\rho) = \frac{1}{\rho} + \frac{1}{1 - \rho} = \frac{1}{\rho(1 - \rho)} ,$$

which is consistent with (4.43) with $D = 1$ and $\chi(\rho) = \rho(1 - \rho)$.

For zero-range processes it can be shown that $F'(\rho) = \log \Phi(\rho)$, such that $F''(\rho) = \Phi'(\rho)/\Phi(\rho)$, which is again consistent with (4.43) with $D(\rho) = \Phi'(\rho)$ and $\chi(\rho) = \Phi(\rho)$.

The above martingale method can also be adapted to cover weakly asymmetric systems, where the symmetric dynamics is of gradient type. Consider a perturbation of the jump rates of a homogeneous gradient system of the form

$$c(x, y, \eta)(1 + q(y - x)/L) .$$

Then the right-hand side of (4.40) has an additional contribution of the form

$$\sum_{x \in \Lambda} \sum_{z \in \Lambda} q(z) (\psi(x/L) - \psi((x + z)/L)) \int_0^t c(x, x + z, \eta_s) ds .$$

Taylor expansion of smooth ψ up to first order, and a similar block average to replace $c(x, x + z, \eta_s)$ by its expectation under density $\rho(u, s)$, will then lead to a contribution to the constitutive equation (4.2) for the current of the form

$$E\chi(\rho) = \sum_{z \in \Lambda} zq(z) \nu_\rho(c(x, x + z, \cdot))$$

with constant field E . Note that $\nu_\rho(c(x, x + z, \cdot))$ is independent of x due to translation invariance, and if ν_ρ is of product form it often is even independent of z (see next section). Then all explicit spatial information of the jump rates can be separated into the field and we have

$$E = \sum_{z \in \Lambda} zq(z) \quad \text{and} \quad \chi(\rho) = \nu_\rho(c(x, x + z, \cdot)) \text{ for all } x, z \in \Lambda .$$

4.2.3 Asymmetric, attractive IPS

For asymmetric processes the previously discussed martingale and relative entropy methods do not apply. One problem is that the scaling limit on time scale $s = t/L$ is now given by a hyperbolic conservation law, the weak solutions of which can develop shocks and are only unique under additional selection criteria. But under the assumption of stochastic monotonicity (also called attractivity), there is in fact a very general hydrodynamic limit result. Again we consider lattice gases on the same state space $E = S^\Lambda$ where $S \subseteq \mathbb{N}_0$ is a connected set including 0, and $\Lambda = \mathbb{Z}$ or $\mathbb{Z}/(L\mathbb{Z})$ as before. Note that E admits a partial order where $\eta \leq \zeta$ if $\eta_x \leq \zeta_x$ for all $x \in \Lambda$.

Definition 4.4 A function $f \in C(E)$ is called **increasing** if $\eta \leq \zeta$ implies $f(\eta) \leq f(\zeta)$. A process with semigroup $(P_t : t \geq 0)$ is called **monotone** or **attractive** if $\eta \leq \zeta$ implies $P_t f(\eta) \leq P_t f(\zeta)$ for all $t \geq 0$ and all increasing f . Two probability measures $\mu_1, \mu_2 \in \mathcal{M}_1(E)$ are **stochastically ordered** $\mu_1 \leq \mu_2$ if $\mu_1(f) \leq \mu_2(f)$ for all increasing f .

Theorem 4.3 (Strassen) Suppose $\mu_1, \mu_2 \in \mathcal{M}_1(E)$. Then $\mu_1 \leq \mu_2$ if and only if there exists a coupling $\mu \in \mathcal{M}_1(E \times E)$ such that $\mu[., E] = \mu_1$ and $\mu[E, .] = \mu_2$, and

$$\mu(\{\eta : \eta^1 \leq \eta^2\}) = 1 \quad (\text{i.e. } \eta^1 \leq \eta^2 \text{ } \mu - \text{a.s.}). \quad (4.45)$$

Proof. \Leftarrow : Suppose such a coupling μ exists. If $f \in C(E)$ is increasing then $f(\eta^1) \leq f(\eta^2)$ $\mu - \text{a.s.}$ and writing $\tau^i : E \times E \rightarrow E$ for the projection on the i -th coordinate $\tau^i(\eta) = \eta^i$, we have

$$\mu_1(f) = \mu(f \circ \tau^1) \leq \mu(f \circ \tau^2) = \mu_2(f), \quad (4.46)$$

so that $\mu_1 \leq \mu_2$.

\Rightarrow : involves a construction of the coupling on a probability space, see e.g. [2], Theorem 2.4, p. 72. \square

Using Strassen's theorem for path space distributions, it follows that for a monotone process $\eta \leq \zeta$ implies that $\mathbb{P}^\eta \leq \mathbb{P}^\zeta$ in the sense that there exists a coupling \mathbb{P} defining a joint process $((\eta_t, \zeta_t) : t \geq 0)$ such that $\eta_t(x) \leq \zeta_t(x)$ $\mathbb{P} - \text{a.s.}$ for all $x \in \Lambda$ and $t \geq 0$. If initial distributions are ordered as $\mu_1 \leq \mu_2$, this also implies that corresponding stationary distributions of a monotone process are ordered, i.e.

$$\nu_1 := \lim_{t \rightarrow \infty} \mu_1 P_t \leq \lim_{t \rightarrow \infty} \mu_2 P_t =: \nu_2.$$

Therefore, for any IPS that has a family of stationary distributions ν_ρ indexed by the density, $\rho_1 \leq \rho_2$ implies $\nu_{\rho_1} \leq \nu_{\rho_2}$. This can be shown by starting coupled processes in an ordered initial condition which is in the domain of attraction of the stationary measures.

Proposition 4.4 Consider an IPS on $E = S^\Lambda$ as given above with generator of the form

$$\mathcal{L}f(\eta) = \sum_{x,y} p(x,y)b(\eta(x), \eta(y))(f(\eta^{xy}) - f(\eta)) \quad (4.47)$$

where $p(x,y)$ is an irreducible transition kernel on Λ and $b(n,m) \geq 0$ with $b(n,m) = 0$ if and only if $n = 0$ or $m = \max S < \infty$. Then the IPS is monotone if and only if $b(n,m)$ is non-decreasing in n and non-increasing in m (also called **misanthrope process**).

Proof. If b fulfills this assumption, one can construct a **basic coupling** of two processes with initial conditions $\eta \leq \zeta$ where particles jump together with maximal rate. They jump from x to y in both processes

$$\text{simultaneously with rate } p(x,y)b(\eta(x), \eta(y)) \wedge b(\zeta(x), \zeta(y)),$$

with additional jumps of only η particles at rate

$$p(x,y)\left(b(\eta(x), \eta(y)) - (b(\eta(x), \eta(y)) \wedge b(\zeta(x), \zeta(y)))\right) \geq 0,$$

and analogously for ζ -particles. The monotonicity assumptions on b then imply that these dynamics conserves the partial order. If b violates this assumption, one can find an increasing test function f for which $P_t f(\eta) > P_t f(\zeta)$ even though $\eta \leq \zeta$.

For details see [34] and references therein. \square

For example, the exclusion process is of the above form and monotone, and the zero-range process is monotone if and only if the rate $g(n)$ is non-decreasing.

Theorem 4.5 Consider a sequence of translation invariant, asymmetric IPS $(\eta_t^L : t \geq 0)$ with generator of the form (4.47) on the lattice $\frac{1}{L}\mathbb{Z}$ or $\frac{1}{L}\mathbb{Z}/L\mathbb{Z}$ and translation invariant $p(x, y) = q(y - x)$ with finite first moment $m = \sum_{z \in \Lambda} zq(z) \in (-\infty, \infty)$. Assume that the initial distribution μ has a limiting density $\rho(u, 0)$ as in (4.34), and that the system has a family of translation invariant stationary measures ν_ρ on $S^{\mathbb{Z}}$ with $\rho \in [0, \max S]$. Then as $L \rightarrow \infty$

$$\pi_{sL}^L(du) \rightarrow \rho(\cdot, s) du \quad \text{weakly, in probability,} \quad (4.48)$$

where $\rho(u, s)$ is a solution of the hyperbolic equation (4.3) with vanishing diffusion $D = 0$,

$$\partial_s \rho(u, s) + m \partial_u (\chi(\rho(u, s))) = 0 \quad (4.49)$$

on \mathbb{R} or \mathbb{T} with initial condition $\rho(u, 0)$. If ν_ρ are of product form, the external field is given by the first moment m and the mobility by

$$\chi(\rho) = \nu_\rho(b) = \sum_{m, n \in S} \nu_\rho^1(m) \nu_\rho^1(n) b(m, n).$$

Note that weakly asymmetric IPS are also covered by the method based on monotonicity for asymmetric models. If the stationary measures are known the mobility is easy to compute, and if the measures are stationary independently of symmetry of $p(x, y)$, the diffusion coefficient of weakly asymmetric versions is given by the Einstein relation (4.43).

Relevant results and references can be found in [8] Chapter 8. The above result was first proved in [24] for the TASEP, and in [25] for a more general class of models using attractivity. See also [34] for a more recent reference. For a theory of entropy solutions of hyperbolic conservation laws see e.g. [26].

4.3 Equilibrium fluctuations of density fields

While the hydrodynamic limit provides a law of large numbers result for the density field, in this section we are interested in studying the space-time fluctuations of that field. As before for additive functionals we will focus on equilibrium fluctuations. As before, consider a spatially homogeneous lattice gas on the state space $E = S^\Lambda$ where $S \subseteq \mathbb{N}_0$ is a connected set including 0, and we focus on $\Lambda = \mathbb{Z}$ for simplicity. Results can also be formulated for periodic boundary conditions, but then additional care has to be taken about the sequence of stationary measures. We consider the dynamics in a homogeneous stationary state with density $\rho > 0$ given by a translation invariant measure ν_ρ on E , which should have decaying correlations such that

$$\text{cov}(\rho) := \sum_{x \in \mathbb{Z}} \left(\nu(\eta(0)\eta(x)) - \rho^2 \right) \in (0, \infty). \quad (4.50)$$

Embedding the rescaled lattice \mathbb{Z}/L in \mathbb{R} with scaling parameter $L \rightarrow \infty$, the fluctuations around ν_ρ are then characterized by the **density field**

$$Y^L[du] = \frac{1}{\sqrt{L}} \sum_{x \in \mathbb{Z}} (\eta(x) - \rho) \delta_{x/L}[du] \in \mathcal{M}(\mathbb{R}). \quad (4.51)$$

This is a signed measure on \mathbb{R} , and can be characterized in a weak form as

$$Y^L(G) = \frac{1}{\sqrt{L}} \sum_{x \in \mathbb{Z}} G(x/L) (\eta(x) - \rho) \quad (4.52)$$

for any smooth G decaying sufficiently fast at infinity. The fluctuation field is a function on state space E , and we can define the process Y_t^L in complete analogy, simply evaluating the field at η_t . For a large class of processes, the diffusive scaling limit Y_t of the process $Y_{tL^2}^L$ is given by a **generalized Ornstein-Uhlenbeck process** characterized by the SDE

$$dY_t(G) = Y_t(\mathcal{A}G)dt + dB_t^G . \quad (4.53)$$

Here \mathcal{A} usually takes the form of an elliptic second order differential operator with constant coefficients, simply $\mathcal{A}G = \sigma^2 G''$ in one space dimension where $\sigma^2 > 0$ is related to the spatial part of the dynamics. By the **fluctuation-dissipation relation** the Brownian motions B^G depend linearly on the test function G and satisfy

$$\mathbb{E}[(B_t^G)^2] = \text{cov}(\rho)t \int_{\mathbb{R}} \sigma^2 G''(u) du , \quad (4.54)$$

fixing the quadratic variation. The scaling limit (4.53) is a result of the martingale decomposition of underlying process using Itô's formula, which takes a particularly simple form for reversible systems. Convergence holds in law, i.e. for the stationary process characterized by \mathbb{P}^{ν_ρ} we have

$$\mathbb{P}^{\nu_\rho} \circ (Y^L)^{-1} \rightarrow \mathbb{P}$$

where \mathbb{P} is the law of the \mathbb{R} -valued limit process (4.53) with corresponding expectation \mathbb{E} used in (4.54).

In the following, we sketch the derivation of this result for a symmetric exclusion process with generator

$$\mathcal{L}f(\eta) = \sum_{x,y \in \Lambda} p(y-x)\eta(x)(1-\eta(y))(f(\eta^{xy}) - f(\eta)) ,$$

with translation invariant dynamics given by $p(y-x)$ which is also symmetric, i.e. $p(-z) = p(z)$ for all $z \in \mathbb{Z}$. A simple computation reveals

$$\begin{aligned} \mathcal{L}Y^L(G) &= \frac{1}{\sqrt{L}} \sum_{x,y \in \mathbb{Z}} p(y-x)\eta(x)(1-\eta(y))(G(y/L) - G(x/L)) \\ &= \frac{1}{\sqrt{L}} \sum_{x,y \in \mathbb{Z}} p(y-x)\eta(x)(G(y/L) - G(x/L)) , \end{aligned} \quad (4.55)$$

where we used symmetry of p in the second line. Using smoothness of G and Taylor expansion we get to leading order

$$\mathcal{L}Y^L(G) \simeq \frac{1}{L^{5/2}} \sum_{x \in \mathbb{Z}} \sigma^2 G''(x/L)(\eta(x) - \rho) = \frac{1}{L^2} Y^L(\sigma^2 G''),$$

where $\sigma^2 = \sum_{z \in \Lambda} z^2 p(z)$, the first order term drops out due to symmetry of p , and we can add the constant $-\rho$ since G vanishes at infinity. Therefore we get for the process using Itô's formula

$$Y_t^L(G) - Y_0^L(G) = \int_0^t \mathcal{L}Y_s^L(G) ds + M_t^L(G) = \frac{1}{L^2} \int_0^t Y_s^L(\sigma^2 G'') ds + M_t^L(G) \quad (4.56)$$

where $M_t^L(G)$ is a martingale with quadratic variation

$$[M^L(G)]_t = \int_0^t \left(\mathcal{L}(Y_s^L(G))^2 - 2Y_s^L(G)\mathcal{L}Y_s^L(G) \right) ds . \quad (4.57)$$

Speeding up time by a factor L^2 the factor in (4.56) is absorbed in the generator, and the martingale CLT (Lemma 1.32) implies that $M_t^L(G)$ converges to a Brownian motion B^G as $L \rightarrow \infty$. This leads to an integrated version of (4.53) in the scaling limit,

$$Y_t(G) - Y_0(G) = \sigma^2 \int_0^t Y_s(G'') + B_t^G .$$

The quadratic variation of B_t^G can be identified from a scaling limit of (4.57) with speeded up generator $L^2\mathcal{L}$ and leads to

$$\mathbb{E}[(B_t^G)^2] = \rho(1 - \rho)t \int_{\mathbb{R}} \sigma^2 G''(u) du .$$

This is consistent with the fluctuation-dissipation relation (4.54), where under the simple product stationary measure ν_ρ for the exclusion process the covariance (4.50) $\text{cov}(\rho) = \rho(1 - \rho)$ is simply given by the variance of a $\text{Be}(\rho)$ random variable. Since the process is stationary, we get for the limiting equilibrium fluctuations

$$\mathbb{E}[Y_t(G)^2] = \mathbb{E}[Y_0(G)^2] = \rho(1 - \rho) \int_{\mathbb{R}} G''(u)^2 du ,$$

which can easily be seen taking expectation under the stationary product measure ν_ρ of

$$Y^L(G)^2 = \frac{1}{L} \sum_{x,y \in \mathbb{Z}} G(x/L) G(y/L) (\eta(x) - \rho) (\eta(y) - \rho) .$$

Due to the product structure only diagonal terms contribute, and the equilibrium fluctuations have the structure of spatial white noise, with the time evolution given by a generalized Ornstein-Uhlenbeck process.

The above proof can be adapted to a large class of reversible lattice gases of gradient type, where the Brownian noise in the SDE (4.53) results from the martingale part of the decomposition (4.56), and the drift from the integral part. For asymmetric non-gradient systems often a similar limit result can be shown, but the integral part in (4.56) contributes to the noise in the limit equation (4.53), and the proof is much more complicated. For details see [21], Chapter 7.

4.4 Macroscopic fluctuation theory

In the previous subsections we have investigated the typical large scale dynamic behaviour of IPS, and in this section we will look at large deviations of current and density profiles. A comprehensive macroscopic fluctuation theory has recently been developed [5], but has only been made rigorous for particular examples of processes, including the exclusion process. Recall that the typical large scale behaviour from (4.1) and (4.2) is given by

$$\partial_t \rho(u, t) + \partial_u j(u, t) = 0 \quad \text{and} \quad j(u, t) = J(\rho(u, t), u, t), \quad (4.58)$$

where we write $t \geq 0$ for the macroscopic time in this subsection. The local current $j(u, t)$ is determined by the constitutive equation

$$J(\rho, u, t) = -D(\rho) \partial_u \rho + \chi(\rho) E(u, t), \quad (4.59)$$

with diffusion coefficient, mobility and external field as defined at the beginning of this section. The models we consider should be of gradient type with a generator

$$\mathcal{L}f(\eta) = \sum_{x \in \Lambda} \sum_{|z|=1} c(x, x+z, \eta) (f(\eta^{x, x+z}) - f(\eta)),$$

where we assume nearest neighbour dynamics for simplicity, and set $\Lambda = \mathbb{Z}/L\mathbb{Z}$. For translation invariant gradient models with a gradient function h and a family of translation invariant stationary measures ν_ρ , $\rho \in [0, \max S]$, the macroscopic transport parameters are connected to the microscopic rates via

$$D(\rho) = \Phi'(\rho), \quad \Phi(\rho) := \nu_\rho(h) \quad \text{and} \quad E\chi(\rho) = \sum_{z \in \Lambda} \nu_\rho(c(x, x+z, \cdot)),$$

where the field $E(u, t) \equiv E$ is constant. We have shown this in previous sections for the ASEP and ZRP, for which it has been established rigorously, but this connection should hold for general weakly asymmetric versions of gradient models, but I could not find this in the literature...

For simplicity we restrict the following presentation to the case $E \equiv 0$. We consider a parabolic scaling with $u = x/L$ and speeding up the process by L^2 . In that scaling fields E can be included with weak asymmetries of the microscopic rates, as seen in an earlier section. All conservative IPS fulfill the conservation law (4.58) on large scales, and the particular typical behaviour of an IPS is characterized by the constitutive equation. The strategy to compute the asymptotic probability or cost to see an atypical density/current profile $\rho(u, t)$, $j(u, t)$ is again to tilt the path measure \mathbb{P} of the IPS, such that the rare event becomes typical. This can be done by an auxiliary (additional) external field $F(u, t)$, which is determined by the constitutive equation

$$j(u, t) = -D(\rho(u, t)) \partial_u \rho(u, t) + \chi(\rho(u, t)) F(u, t) \quad (4.60)$$

such that the prescribed $j(u, t)$ is the typical current. The action of the macroscopic field $F(u, t)$ on the microscopic rates of the process can be implemented as a weak asymmetry of the form

$$c_F(x, x \pm 1, \eta) := c(x, x \pm 1, \eta) e^{F_{x, x \pm 1}(t)} \quad \text{where} \quad F_{x, x \pm 1}(t) = \pm \frac{1}{2L} F(x/L, t).$$

Note that since $e^{F_{x, x \pm 1}(t)} = 1 \pm \frac{1}{2L} F(x/L, t)$ to leading order, this is a weak asymmetry that does not change the mobility of the process, but only contributes to the field.

Now we can use Girsanov's formula to compute the path measure \mathbb{P}_F under the change of rates on the compact time interval $[0, T]$. For a path ω we denote the associated jump times where a particle jumps from x to $x+z$ by $\tau_k^{x, x+z}$, $k = 1, 2, \dots$, $\mathcal{J}_T(x, x+z)(\omega)$, where $\mathcal{J}_T(x, x+z)(\omega)$ is the total number of such jumps up to time T introduced in (3.17). Then with \mathbb{P} and \mathbb{P}_F denoting the path measures of the process speeded up by a factor of L^2 , we get

$$\begin{aligned} d\mathbb{P}[\omega] = \exp \left(\int_0^T \sum_{x \in \Lambda} \sum_{|z|=1} L^2 c(x, x+z, \eta_t) (e^{F_{x, x+z}(t)} - 1) dt \right. \\ \left. - \sum_{x \in \Lambda} \sum_{|z|=1} \sum_{k=1}^{\mathcal{J}_T(x, x+z)(\omega)} F_{x, x+z}(\tau_k^{x, x+z}) \right) d\mathbb{P}_F[\omega]. \end{aligned} \quad (4.61)$$

The second term can be written in terms of the **empirical current**, which we introduce analogously to the empirical process/density as

$$j^L(du, dt) := \frac{1}{L} \sum_{x \in \Lambda} \left(d\mathcal{J}_t(x, x+1) - d\mathcal{J}_t(x, x-1) \right) \delta_{x/L}(du).$$

Since $\mathcal{J}_t(x, y)$ is a jump process this is an atomic measure on $\mathbb{T} \times [0, T]$. This leads to

$$\sum_{x \in \Lambda} \sum_{|z|=1} \sum_{k=1}^{\mathcal{J}_T(x, x+z)(\omega)} F_{x, x+z}(\tau_k^{x, x+z}) = \frac{L}{2} \int_0^T \int_{\mathbb{T}} j^L(du, dt) F(u, t). \quad (4.62)$$

Expanding $e^{F_{x, x+z}(t)}$ up to second order and using the anti-symmetry $F_{x, x-1}(t) = -F_{x, x+1}(t)$ and the gradient condition we get for the first term in (4.61)

$$\begin{aligned} & \int_0^T dt \sum_{x \in \Lambda} \sum_{|z|=1} L^2 c(x, x+z, \eta_t) \left(F_{x, x+z}(t) + \frac{1}{2} F_{x, x+z}(t)^2 \right) = \\ & = \int_0^T dt \sum_{x \in \Lambda} \frac{L^2}{2} h(\tau_x \eta_t) \sum_{|z|=1} \underbrace{\left(F_{x, x+z}(t) - F_{x+z, x}(t) \right)}_{\approx \partial_u F(u, t)/L^2} \\ & \quad + \frac{L^2}{2} \sum_{x \in \Lambda} \sum_{|z|=1} \left(c(x, x+z, \eta_t) + c(x+z, x, \eta_t) \right) F_{x, x+z}(t)^2 \\ & \approx L \int_0^T dt \int_{\mathbb{T}} \left(\frac{1}{2} \pi_t^L(h \circ \tau_{[L \cdot]}) \partial_u F(u, t) \right. \\ & \quad \left. + \frac{1}{4} F(u, t)^2 \pi_t^L(c([L \cdot], [L \cdot] + z, \cdot) + c([L \cdot] + z, [L \cdot], \cdot)) \right). \end{aligned} \quad (4.63)$$

Now we use the local equilibrium assumption and again an auxiliary spatial average to replace

$$\begin{aligned} \pi_t^L(h \circ \tau_{[L \cdot]}) & \approx \nu_{\rho(u, t)}(h) = \Phi(\rho(u, t)) \\ \pi_t^L(c([L \cdot], [L \cdot] + z, \cdot) + c([L \cdot] + z, [L \cdot], \cdot)) & \approx \nu_{\rho(u, t)}(c(x, x+1, \cdot) + c(x+1, x, \cdot)) = 2\chi(\rho(u, t)), \end{aligned}$$

and plugging this into the above yields

$$L \int_0^T dt \int_{\mathbb{T}} du \left(\frac{1}{2} \Phi(\rho(u, t)) \partial_u F(u, t) + \frac{1}{4} F(u, t)^2 \chi(\rho(u, t)) \right).$$

Combining this with (4.62) we get

$$\frac{d\mathbb{P}}{d\mathbb{P}_F}(\omega) \approx \exp \left(L \int_0^T \int_{\mathbb{T}} \frac{1}{2} (j(u, t) du dt - j^L(du, dt)(\omega)) F(u, t) - \frac{1}{4} F(u, t)^2 \chi(\rho(u, t)) du dt \right), \quad (4.64)$$

where we have also used (4.60) to get

$$\int_{\mathbb{T}} \phi \partial_u F du = \int_{\mathbb{T}} D \partial_u \rho F du = \int_{\mathbb{T}} (\chi F^2 - j F) du,$$

omitting the arguments to simplify notation. The probability of the fluctuation is then given by

$$\begin{aligned} \mathbb{P} \left[(\pi_t^L(du), j^L(du, dt)) \approx (\rho(u, t) du, j(u, t) du dt), t \in [0, T] \right] \\ \asymp \mathbb{E}_F \left[\frac{d\mathbb{P}}{d\mathbb{P}_F} \mathbb{1}_{(\pi_t^L(du), j^L(du, dt)) \approx (\rho(u, t) du, j(u, t) du dt)} \right]. \end{aligned} \quad (4.65)$$

Under \mathbb{E}_F , the event in the indicator function is typical, and we only have to evaluate the expectation of the Radon-Nikodym derivative (4.64). The only random part is the empirical current $j^L(du, dt)$ which we replace by its expectation to get

$$\begin{aligned} \mathbb{P} \left[(\pi_t^L(du), j^L(du, dt)) \approx (\rho(u, t) du, j(u, t) du dt), t \in [0, T] \right] \\ \asymp \exp \left(- \frac{L}{4} \int_0^T \int_{\mathbb{T}} F(u, t)^2 \chi(\rho(u, t)) du dt \right). \end{aligned} \quad (4.66)$$

With $E \equiv 0$ the constitutive equation for the typical current is simply

$$J(u, t) = -D(\rho(u, t)) \partial_u \rho(u, t) = j(u, t) - \chi(\rho(u, t)) F(u, t)$$

and we can use this to replace F in the above formula. In conclusion we arrive at

$$\mathbb{P}^\eta \left[(\pi_t^L(du), j^L(du, dt)) \approx (\rho(u, t) du, j(u, t) du dt), t \in [0, T] \right] \asymp \exp \left(- LI_{[0, T]}(\rho, j) \right)$$

with large deviation rate function

$$I_{[0, T]}(\rho, j) = \frac{1}{4} \int_0^T dt \int_{\mathbb{T}} du (j(u, t) - J(u, t))^2 / \chi(\rho(u, t)),$$

as $L \rightarrow \infty$. Here the initial conditions η have to be consistent with the initial density profile $\rho(u, 0)$, and if the pair $(\rho(u, t), j(u, t))$ does not fulfill the continuity equation (4.58), then the rate function is ∞ .

Note that this corresponds to a large deviation principle on the rather complex space of measures on $\mathbb{T} \times [0, \infty)$, and rigorous results on this are not easy to obtain. Results like this have been proven for the exculsion processes and can be found in [8], Chapter 10.

References

- [1] F. Spitzer: Interaction of Markov Processes. *Adv. Math.* **5**, 246-290 (1970)
- [2] T.M. Liggett: *Stochastic Interacting Systems*, Springer (1999)
- [3] H. Spohn: *Large Scale Dynamics of Interacting Particles*, Springer (1991)
- [4] T.M. Liggett: *Interacting Particle Systems - An Introduction*, ICTP Lecture Notes **17** (2004) (available at publications.ictp.it/lns/vol17/vol17toc.html)
- [5] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim: *Macroscopic Fluctuation Theory*. *Reviews of Modern Physics*, Volume **87**, 593-636 (2015)
- [6] T.M. Liggett: *Interacting Particle Systems*, Springer (1985)
- [7] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter (1988)
- [8] C. Kipnis, C. Landim: *Scaling limits of Interacting Particle Systems*, Springer (1999)
- [9] K. Yosida: *Functional Analysis*, 6th edition, Springer (1980)
- [10] O. Kallenberg: *Foundations of modern probability*, 2nd edition, Springer (2002)
- [11] J.R. Norris: *Markov Chains*, Cambridge (1997)
- [12] G. Grimmett, D. Stirzaker: *Probability and Random Processes*, 3rd edition, Oxford (2001)
- [13] W. Rudin: *Real and complex analysis*, McGraw-Hill (1987)
- [14] W. Rudin: *Functional analysis*, 2nd edition, McGraw-Hill (1991)
- [15] P. Chleboun, S. Grosskinsky: *Condensation in stochastic particle systems with stationary product measures*, *J. Stat. Phys.* **154**(1-2), 432465 (2014)
- [16] L. Saloff-Coste. *Random Walks on Finite Groups*. In: H. Kesten, ed. *Probability on Discrete Structures*, *Encyclopaedia of Mathematical Sciences* (110), Springer, 2003.
- [17] D. Levin, Y. Peres and E. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.
- [18] N. Berestycky. *Mixing times of Markov chains*. lecture notes available at www.statslab.cam.ac.uk/~beresty/teach/Mixing/mc.html
- [19] D.-Q. Jiang, F.-X. Zhang: *The Green-Kubo formula and power spectrum of reversible Markov processes*, *J. Math. Phys.* **44**(10), 4681-4689 (2003)
- [20] P. Billingsley: *Convergence of probability measures*, 2nd edition, Wiley (1999)
- [21] T. Komorowski, C. Landim, S. Olla: *Fluctuations in Markov Processes*, Springer (2012)
- [22] J. Beltrán, C. Landim: *A martingale approach to metastability*, *Probab. Theory Relat. Fields* **161**, 267-307 (2015)

- [23] J. Quastel: Introduction to KPZ, Current Developments in Mathematics 2011, 125194, Int. Press, Somerville, MA (2012)
- [24] H. Rost: Non-Equilibrium Behaviour of a Many Particle Process: Density Profile and Local Equilibria. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **58**, 41-53 (1981)
- [25] F. Rezakhanlou: Hydrodynamic limit for attractive particle systems on \mathbb{Z}^d . *Commun. Math. Phys.* **140**, 417-448 (1991)
- [26] C.M. Dafermos: *Hyperbolic Conservation Laws in Continuum Physics*. Springer (2000)
- [27] O. Golinelli, K. Mallick: The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics. *J. Phys. A: Math. Gen.* **39**, 12679-12705 (2006)
- [28] A. Dembo, O. Zeitouni: *Large Deviations Techniques and Applications*. Springer (1998)
- [29] R. Chetrite, H. Touchette: Nonequilibrium Markov processes conditioned on large deviations. *Ann. Inst. H. Poincaré* **16**, 2005 (2015)
- [30] L. Bertini, A. Faggionato, D. Gabrielli: Large deviations of the empirical flow for continuous time Markov chains. *Ann. Inst. H. Poincaré* **51**, 867 (2015)
- [31] T.M. Liggett: *Continuous Time Markov Processes*. AMS Graduate Studies in Mathematics 113 (2010)
- [32] P.A. Ferrari, L.R.G. Fontes: Current fluctuations for the asymmetric simple exclusion process. *Ann. Probab.* **22**(2), 820-832 (1994)
- [33] C. Giardiná, J. Kurchan, V. Lecomte, J. Tailleur: Simulating rare events in dynamical processes. *J. Stat. Phys.* **145**, 787-811 (2011)
- [34] C. Bahadoran, H. Guiol, K. Ravishankar and E. Saada: Euler Hydrodynamics of One-Dimensional Attractive Particle Systems. *Ann. Probab.* **34**(4), 1339–1369 (2006)
- [35] L. De Carlo and D. Gabrielli: Totally asymmetric limit for models of heat conduction. arXiv:1611.07566
- [36] T. Seppalainen: *Translation Invariant Exclusion Processes (Book in Progress)*. <http://www.math.wisc.edu/~seppalai/excl-book/etusivu.html>