Translation Invariant Exclusion Processes (Book in Progress)

Department of Mathematics
University of Wisconsin
Madison, WI 53706-1388

December 11, 2008

Contents

PART I Preliminaries

1 Markov chains and Markov processes

This section serves several purposes. To prepare the reader for the construction issues of the exclusion process that will be addressed in Section 2, we discuss here the construction of countable state Markov chains, first in discrete and then in continuous time. The treatment is far from complete, so prior familiarity with these topics is necessary. Motivated by these examples, in Section 1.3 we discuss the general definition of a Markov process as a family of probability measures $\{P^x\}$ on path space, indexed by initial states x. A brief section introduces Poisson processes which are a key building block of interacting Markov processes. The last section on harmonic functions discusses the coupling technique for Markov chains and proves some results for later use.

1.1 Discrete-time Markov chains

A stochastic process in most general terms is a collection of random variables $\{X_j : j \in J\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, indexed by some index set J. If the stochastic process represents the temporal evolution of some random system, the index set is a discrete or continuous set of time points, for example $J = \mathbf{Z}_+ = \{0, 1, 2, \ldots\}$ or $J = \mathbf{R}_+ = [0, \infty)$. However, much more exotic index sets are quite natural. For example, for a point process on the Euclidean space $\mathbf{R}^d$, J would be the collection of all Borel subsets of $\mathbf{R}^d$.

The key feature of the definition is that the random variables X_j are defined on a common probability space. This enables us to talk about probabilities of events that involve several or even infinitely many variables simultaneously. This is what the theory is all about.

Among the first stochastic processes one meets is the discrete-time, countable state space Markov chain with time-homogeneous transition probabilities. Let S be a finite or countable set, the state space of the process. A *stochastic matrix* is a matrix $(p(x,y))_{x,y \in S}$ of nonnegative numbers that satisfy

$$\sum_{y \in S} p(x,y) = 1 \qquad \text{for all } x \in S. \tag{1.1}$$

If S is infinite, the matrix $p(x,y)$ is an infinite matrix. Suppose $\{X_n : n \in \mathbf{Z}_+\}$ are random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Then X_n is a *Markov chain* with transition probability $p(x,y)$ if, for all $n \geq 0$ and all choices of states $x_0, x_1, \ldots, x_{n-1}, x, y \in S$,

$$P[X_{n+1} = y \,|\, X_n = x, X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0] = p(x,y). \tag{1.2}$$

This condition makes sense for all $x_0, x_1, \ldots, x_{n-1}, x$ for which the conditioning event has positive probability. Condition (1.2) expresses the idea that, given the present state x, the future evolution is entirely independent of the past evolution $x_0, x_1, \ldots, x_{n-1}$. This notion is called the *Markov property*. *Time homogeneity* is the property that the right-hand side of (1.2) does not depend on n.

Sometimes one may be faced with the task of using definition (1.2) to check that a given process is Markovian. But the more natural question goes the other way around. Given a stochastic matrix $p(x, y)$ and an initial state $x_0 \in S$, does there exist a Markov chain X_n with transition probability $p(x, y)$ and such that $X_0 = x_0$ almost surely? This question is nontrivial because we are asked to construct an *infinite* collection of random variables X_n that are in a special relationship with each other.

To get finitely many random variables $(X_0, \ldots, X_m)$ with the required relationship, we answer immediately as follows. Let $\Omega = S^{m+1}$ be the space of $(m+1)$-vectors with entries in S, and let $\mathcal{F}$ be the collection of all subsets of Ω. For each $\omega = (s_0, s_1, \ldots, s_m) \in \Omega$, define its probability by

$$P(\omega) = \mathbf{1}_{\{x_0\}}(s_0)p(s_0, s_1)p(s_1, s_2)\cdots p(s_{m-1}, s_m). \tag{1.3}$$

Define the random variables by $X_i(\omega) = s_i$ for $0 \le i \le m$. The Markov property (1.2) for $0 \le n < m$ is built into the model.

Analogously, it is natural to construct the infinite process $(X_n)_{0 \le n < \infty}$ on the sequence space $\Omega = S^{\mathbf{Z}_+}$, whose elements are infinite sequences $\omega = (s_0, s_1, s_2, \ldots)$ from S, and take again the *coordinate random variables* $X_n(\omega) = s_n$. The product σ-algebra $\mathcal{F}$ on Ω is generated by cylinder sets. Cylinder sets are events that constrain only finitely many coordinates. With a countable state space it suffices to consider sets of the type

$$\{\omega : (s_0, \ldots, s_m) = (u_0, \ldots, u_m)\} = \{\omega : X_0(\omega) = u_0, \ldots, X_m(\omega) = u_m\}.$$

Let $\mathcal{F}_0$ be the class of such sets obtained by letting $(u_0, \ldots, u_m)$ vary over all finite vectors with S-valued entries.

But now it is impossible to explicitly write down the probability of every event in $\mathcal{F}$. Extending formula (1.3) to an infinite sequence $\omega = (s_n)_{0 \le n < \infty}$ is useless because the answer would be 0 in most cases. Ω is now an uncountable space so we cannot expect to define a measure on it by giving the values of singletons $P\{\omega\}$.

We can write down probabilities of cylinder events, and this is the first step towards a solution of the construction problem. Define a function P^x on $\mathcal{F}_0$ by

$$P^x\{X_0 = u_0, X_1 = u_1, \ldots, X_m = u_m\} = \mathbf{1}_{\{x\}}(u_0)p(u_0, u_1)p(u_1, u_2)\cdots p(u_{m-1}, u_m). \tag{1.4}$$

The second step comes from an extension theorem, which says that consistent finite-dimensional distributions always come from a measure on the infinite-dimensional space. We state Kolmogorov's extension theorem in a form sufficiently general for our needs.

Theorem 1.1 Kolmogorov's Extension Theorem. *Suppose S is a complete separable metric space, $I = \{i_1, i_2, i_3, \ldots\}$ a countable index set, and $\Omega = S^I$ the space of functions ω from I into S. Let $\mathcal{F}$ be the product σ-algebra on Ω, which is by definition the smallest σ-algebra that contains all sets of the type $\{\omega : \omega(i) \in B\}$ for Borel sets $B \subseteq S$ and $i \in I$.*

Suppose that for each n we are given a probability measure μ_n on the space S^n. Assume that the collection $\{\mu_n\}$ is consistent in this sense: for each n and Borel set $A \subseteq S^n$,

$$\mu_{n+1}\{(s_1, \ldots, s_{n+1}) \in S^{n+1} : (s_1, \ldots, s_n) \in A\} = \mu_n(A).$$

Then there exists a probability measure P on $(\Omega, \mathcal{F})$ such that for all finite n,

$$\mu_n(A) = P\{\omega \in \Omega : (\omega(i_1), \ldots, \omega(i_n)) \in A\}.$$

Kolmogorov's theorem guarantees that for each starting state x, a probability measure P^x exists on the infinite product space such that cylinder probabilities are given by (1.4). If we want the initial state X_0 to be random with a distribution μ, we put on Ω the measure P^μ defined by $P^\mu(A) = \sum_x \mu(x) P^x(A)$ for events $A \in \mathcal{F}$. Thus a Markov process with transition probability $p(x, y)$ exists for every choice of intial distribution μ.

1.2 Continuous-time Markov chains

Next we construct a Markov chain X_t in continuous time $0 \le t < \infty$, but still on a countable state space S. Since S is countable, the chain has to move in jumps, it cannot move continuously. Thus the evolution must be of the following form: a random amount of time spent in a state x, a jump to a new randomly chosen state y, a random amount of time spent in state y, a jump to a randomly chosen state z, and so on. Given an initial state, a description of the process has to provide (1) the probability distributions of the random holding times at different states; and (2) the mechanism for choosing the next state when a jump occurs.

(1) The Markov property stipulates that the distribution of the time till the next jump can only depend on the current location x. It cannot depend on the time already spent at x. This memoryless property forces the waiting time at x to be exponentially distributed, and we let $c(x)^{-1}$ be its mean. Then $c(x)$ is the *rate* of jumping from state x. Whenever the chain is at x, the remaining time T before the next jump has exponential tail $P[T > t] = e^{-c(x)t}$.

(2) When the chain jumps, the Markov property dictates that the choice of next state depends only on the current state x. Thus the jumps are described by a stochastic matrix $p(x, y)$ where $p(x, y)$ is the probability that the next state after x is y.

This suggests that to construct a continuous-time Markov chain X_t with parameters $c(x)$ and $p(x, y)$, we take a discrete-time Markov chain Y_n with transition matrix $p(x, y)$, and adjust the holding times to produce the correct exponentially distributed times with means $c(x)^{-1}$.

Let $x \in S$ be a given initial state. Let $(\Omega, \mathcal{H}, \mathbf{P}^x)$ be a probability space on which are defined a discrete-time Markov chain Y_n with transition matrix $p(u, v)$ and initial state x, and independently of (Y_n), a sequence of exponentially distributed i.i.d. random variables $(\tau_j)_{0 \le j < \infty}$ with common mean $E\tau_j = 1$. To construct such a probability space, let $(\Omega_1, \mathcal{H}_1, P_1^x)$ be a probability space for (Y_n) and $(\Omega_2, \mathcal{H}_2, P_2)$ a probability space for (τ_j), and take $(\Omega, \mathcal{H}, \mathbf{P}^x)$ to be the product probability space:

$$(\Omega, \mathcal{H}, \mathbf{P}^x) = (\Omega_1 \times \Omega_2, \mathcal{H}_1 \otimes \mathcal{H}_2, P_1^x \otimes P_2).$$

The sequence of states that the continuous-time chain X_t visits is $x = Y_0, Y_1, Y_2, Y_3, \ldots$ Define the holding times by $\sigma_n = c(Y_n)^{-1}\tau_n$. Given Y_n, the variable σ_n is independent of $(\sigma_k, Y_k)_{0 \le k \le n-1}$ and has exponential distribution with mean $c(Y_n)^{-1}$. Now define $T_0 = 0$ and $T_n = \sigma_0 + \cdots + \sigma_{n-1}$ for $n \ge 1$, and then

$$X_t = Y_n \text{ for } T_n \le t < T_{n+1}, \text{ for } n = 0, 1, 2, \ldots \tag{1.5}$$

In words, X_t spends time σ_n at state Y_n, and then jumps to state Y_{n+1}. X_t is defined for all times $0 \le t < \infty$ if $T_n \nearrow \infty$ as $n \nearrow \infty$. This happens almost surely if for example there is a constant C_0 such that $c(x) \le C_0$ for all $x \in S$. We assume this throughout our discussion. Note that in (1.5) we specifically chose the path $t \mapsto X_t$ to be right-continuous.

This construction can be repeated for each starting state x. Define the *transition probability* by $p_t(x, y) = \mathbf{P}^x[X_t \in y]$. One can prove the following property for all time points $0 \le t_0 < t_1 < t_2 < \cdots < t_n$ and states $x_0, x_1, x_2, \ldots, x_n$:

$$\begin{aligned} \mathbf{P}^x[X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_{n-1}} = x_{n-1}, X_{t_n} = x_n] \\ = p_{t_0}(x, x_0) p_{t_1 - t_0}(x_0, x_1) \cdots p_{t_n - t_{n-1}}(x_{n-1}, x_n). \end{aligned} \tag{1.6}$$

See Chapter 5 in [29] for a proof. (1.6) implies the Markov property, namely that

$$\mathbf{P}^x[X_{t_n} = x_n \,|\, X_{t_{n-1}} = x_{n-1}, X_{t_{n-2}} = x_{n-2}, \ldots, X_{t_0} = x_0] = p_{t_n - t_{n-1}}(x_{n-1}, x_n) \tag{1.7}$$

whenever the conditioning makes sense.

In Section 1.1 the discrete-time chain X_n was constructed on the sequence space $\Omega = S^{\mathbf{Z}_+}$, whose sample points are the paths of the process. We do the same for the continuous-time chain. Let D_S be the space of functions ξ from $[0, \infty)$ into S with the property that at each $t \in [0, \infty)$, ξ is continuous from the right, and has a limit from the left. Precisely, at each $t \in [0, \infty)$,

$$\xi(t) = \lim_{s \searrow t} \xi(s), \quad \text{and the limit} \quad \xi(t-) = \lim_{s \nearrow t} \xi(s) \quad \text{exists.} \tag{1.8}$$

Such functions are called RCLL functions, and also cadlag functions (the corresponding French acronym).

Let $\mathcal{F}$ be the σ-algebra on D_S generated by the coordinate mappings $\xi \mapsto \xi(t)$, $t \ge 0$. We can think of $X_\cdot = (X_t : 0 \le t < \infty)$ as a D_S-valued random variable defined on $(\Omega, \mathcal{H}, \mathbf{P}^x)$, and let P^x be its distribution. Then P^x is the probability measure on $(D_S, \mathcal{F})$ defined by

$$P^x(A) = \mathbf{P}^x\{X_\cdot \in A\}$$

for events $A \in \mathcal{F}$. This defines a family $\{P^x\}$ of probability measures on D_S, indexed by states $x \in S$. E^x stands for expectation under the measure P^x. The transition probability can be expressed as

$$p_t(x, y) = P^x[\xi(t) = y]. \tag{1.9}$$

We wish to express the simple Markov property (1.7) in a more abstract and powerful form. Let $\{\theta_t : t \ge 0\}$ be the *shift maps* on the path space D_S, defined by $\theta_t\xi(s) = \xi(t+s)$. The effect of the map θ_t is to restart the path at time t. For an event $A \in \mathcal{F}$, the inverse image

$$\theta_t^{-1}A = \{\xi \in D_S : \theta_t\xi \in A\}$$

is the event "A happens from time t onwards." Let $\mathcal{F}_t = \sigma\{\xi(s) : 0 \le s \le t\}$ be the σ-algebra on D_S generated by coordinates up to time t. Then for all events $A \in \mathcal{F}$ and all $x \in S$,

$$P^x[\,\theta_t^{-1}A \,|\, \mathcal{F}_t\,](\xi) = P^{\xi(t)}(A) \tag{1.10}$$

for P^x-almost every ξ. The object on the left-hand side is the conditional probability of an event that concerns the future from time t onwards, conditioned on the past up to time t. It is a random variable on the space D_S, in other words a measurable function of a path ξ which we indicated explicitly. Measurability of $x \mapsto P^x(A)$ on the right-hand side is automatic because on a countable space, all functions are measurable. To derive (1.10) from (1.6), check $E^x[\mathbf{1}_B \cdot \mathbf{1}_A \circ \theta_t] = E^x[\mathbf{1}_B \cdot P^{\xi(t)}(A)]$ first for cylinder events $A \in \mathcal{F}$ and $B \in \mathcal{F}_t$, and then extend to all events by the π-λ-theorem A.1.

Markov property (1.10) expresses the idea that conditioning on the entire past and looking forward from time t onwards amounts to restarting the process, with the current state $\xi(t)$ as the new initial state.

As the last issue, we look at the infinitesimal behavior of the process. In countable state spaces one can express everything in terms of point probabilities [as in (1.9) for example], but in more general spaces this is no longer possible. The alternative is to look at expectations of functions on the state space, so we adopt this practice now.

Define a linear operator L on bounded functions f on S by

$$Lf(x) = c(x) \sum_{y \in S} p(x, y)[f(y) - f(x)]. \tag{1.11}$$

This operator encodes the jump rules of the chain, and reads as follows: starting from state x, the next jump arrives at rate $c(x)$, and when the jump happens, the new state y is selected with probability $p(x,y)$. This jump causes the value of f to change by $f(y)-f(x)$. Rigorously speaking, $Lf(x)$ is the infinitesimal expected change in $f(\xi(t))$, in the sense of the next theorem. L is called the *generator*, or the *infinitesimal generator*, of the Markov chain.

Theorem 1.2 *Assume $c(x) \le C_0$ for all $x \in S$ and let f be a bounded function on S. First, we have the strong continuity at $t = 0$,*

$$\lim_{t\to 0} \sup_{x\in S} |E^x[f(\xi(t))] - f(x)| \;=\; 0. \tag{1.12}$$

Second, the expectation $E^x[f(\xi(t))]$ can be differentiated with respect to t at $t = 0$, uniformly in $x \in S$. Precisely,

$$\lim_{t\to 0} \sup_{x\in S} \left| \frac{E^x[f(\xi(t))] - f(x)}{t} - Lf(x) \right| \;=\; 0. \tag{1.13}$$

We leave the proof of Theorem 1.2 as an exercise, because in Section 2.3 we go through the details of the same result for the more complicated case of the exclusion process. This is a valuable exercise, because it requires some basic estimation in the simplest of situations.

The infinitesimal rates can be expressed in terms of a matrix $Q = (q(x,y))_{x,y\in S}$ defined by $q(x,y) = c(x)p(x,y)$ for $x \ne y$ and $q(x,x) = -\sum_{y:y\ne x} q(x,y)$. Even if originally $p(x,x) > 0$ so that jumps from x to x are permitted, Q ignores this possibility and records only the rates of jumps to genuinely new states. The generator can be equivalently expressed as

$$Lf(x) = \sum_{y\in S} q(x,y)[f(y) - f(x)]. \tag{1.14}$$

Combining $c(x)p(x,y)$ into a single factor $q(x,y)$ represents a change in perspective. Earlier the chain moved in two stages: first the random clock rings at rate $c(x)$, and then a new state y is selected with probability $p(x,y)$. We can equivalently attach to each possible move $x \curvearrowright y$ $(y \ne x)$ a Poisson clock with rate $q(x,y)$, and undertake that jump whose clock rings first. After the jump all clocks are reset. The equivalence of these descriptions follows from properties of Poisson point processes (see Proposition 1.5 below).

We can also write $Lf = Qf$ when we think of $f = (f(x))_{x\in S}$ as a column vector, and interpret Qf as matrix multiplication. In particular, taking $f = \mathbf{1}_{\{y\}}$ in (1.13) gives

$$\frac{d}{dt}\, p_t(x,y)\Big|_{t=0} = q(x,y).$$

1.3 General definitions for Markov processes

Motivated by the continuous-time Markov chain example, we now state some general definitions. Let Y be a metric space, and D_Y the space of RCLL functions ω from $[0, \infty)$ into Y. Measurability on Y will mean Borel measurability, and on D_Y with respect to the coordinate σ-algebra $\mathcal{F}$. In case Y is separable, $\mathcal{F}$ is the Borel σ-algebra of a separable metric on D_Y (see Section A.2.2). On the space D_Y, let $X_\cdot = (X_t : t \ge 0)$ be the coordinate process defined by $X_t(\omega) = \omega(t)$, and $\mathcal{F}_t = \sigma\{X_s : 0 \le s \le t\}$ the σ-algebra generated by coordinates up to time t. The shift maps $\theta_t : D_Y \to D_Y$ are defined by $\theta_t\omega(s) = \omega(s+t)$.

Definition 1.3 *A Markov process is a collection $\{P^x : x \in Y\}$ of probability measures on D_Y with these properties:*

(a) $P^x\{\omega \in D_Y : \omega(0) = x\} = 1$.

(b) *For each $A \in \mathcal{F}$, the function $x \mapsto P^x(A)$ is measurable on Y.*

(c) $P^x[\theta_t^{-1}A|\mathcal{F}_t](\omega) = P^{\omega(t)}(A)$ *for P^x-almost every ω, for every $x \in Y$ and $A \in \mathcal{F}$.*

Requirement (a) in the definition says that x is the initial state under the measure P^x. Requirement (b) is for technical purposes. Requirement (c) is the Markov property. We write E^x for expectation under the measure P^x.

To start the process with a distribution μ other than a point mass δ_x, put on D_Y the measure P^μ defined by

$$P^\mu(A) = \int_Y P^x(A)\,\mu(dx) \quad \text{for } A \in \mathcal{F}.$$

The *transition probability* $p(t, x, dy)$ is defined for $t \ge 0$, $x \in Y$, and Borel sets $B \subseteq Y$ by

$$p(t, x, B) = P^x\{X_t \in B\}. \tag{1.15}$$

The *Chapman-Kolmogorov* equations

$$p(s+t, x, B) = \int_Y p(s, y, B)\,p(t, x, dy) \tag{1.16}$$

are a consequence of the Markov property.

For bounded measurable functions f on Y and $t \ge 0$, define a new function $S(t)f$ on Y by

$$S(t)f(x) = E^x[f(X_t)] = \int_Y f(y)\,p(t, x, dy). \tag{1.17}$$

Measurability of $S(t)f$ follows from part (b) of Definition 1.3. Define the supremum norm on functions by

$$\|f\|_\infty = \sup_{x \in Y} |f(x)|. \tag{1.18}$$

Then

$$\|S(t)f\|_\infty \le \|f\|_\infty, \tag{1.19}$$

so $S(t)$ maps bounded measurable functions into bounded measurable functions. By the linearity of integration,

$$S(t)(\alpha f + \beta g) = \alpha S(t)f + \beta S(t)g \tag{1.20}$$

for scalars α, β and functions f, g. This says that $S(t)$ is a linear operator on bounded measurable functions. Finally, by the Markov property,

$$\begin{aligned} S(s+t)f(x) &= E^x[f(X_{s+t})] = E^x[E^x\{f(X_{s+t})\,|\,\mathcal{F}_s\}\,] \\ &= E^x[E^{X_s}\{f(X_t)\}\,] = E^x[S(t)f(X_s)] = S(s)S(t)f(x). \end{aligned}$$

Thus the operators $\{S(t) : t \ge 0\}$ form a *semigroup*, which means that $S(0) = I$ and $S(s+t) = S(s)S(t)$. Property (1.19) says that the operators $S(t)$ contract distances among functions, so we call $\{S(t)\}$ a *contraction semigroup*.

Note that the probability measures $\{P^x\}$ are uniquely determined by the semigroup $\{S(t)\}$. First, the semigroup $\{S(t)\}$ determines the transition probabilities $p(t, x, dy)$ via (1.17). Second, finite dimensional distributions under P^x are computed as iterated integrals of the transition probabilities:

$$\begin{aligned} &E^x[\Phi(X_{t_1}, X_{t_2}, \dots, X_{t_n})] \\ &\quad = \int_Y \int_Y \cdots \int_Y \Phi(x_1, x_2, \dots, x_n) \\ &\qquad\qquad p(t_n - t_{n-1}, x_{n-1}, dx_n) \cdots p(t_2 - t_1, x_1, dx_2)\, p(t_1, x, dx_1) \end{aligned}$$

for any time points $0 \le t_1 < t_2 < \cdots < t_n$ and any bounded function Φ product measurable on Y^n. Finally, the measure P^x is uniquely determined by its finite dimensional distributions, by the π-λ-theorem A.1.

There is a convenient freedom of language in the theory. Depending on which point of view is fruitful for the occasion, one can talk about a Markov process in terms of random variables X_t, in terms of a semigroup $\{S(t)\}$ on a function space, or in terms of probability measures $\{P^x\}$ on a path space.

Let $C_b(Y)$ be the space of bounded continuous functions on Y. The Markov process $\{P^x\}$ is a *Feller process* if $C_b(Y)$ is closed under the semigroup action. In other words, if $S(t)f \in C_b(Y)$ for all $f \in C_b(Y)$ and $t \ge 0$. Equivalently, if the transition probability $p(t, x, dy)$ is weakly continuous as a function of x for each fixed $t \ge 0$.

All our examples will be Feller processes. Since a probability measure on a metric space is uniquely determined by the integrals of bounded continuous functions, (1.17) shows that for a Feller process, the semigroup action on $C_b(Y)$ is sufficient to determine transition probabilities, and thereby the whole process. Thus for Feller processes it is convenient to consider the semigroup on the space $C_b(Y)$, which is what we shall do.

A strengthening of the Markov property concerns the admission of certain random times t in property (c) of Definition 1.3. A random variable $\tau : D_Y \to [0, \infty]$ is a *stopping time* if $\{\tau \le t\} \in \mathcal{F}_t$ for each $t < \infty$. The σ-algebra of events known at time τ is

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}_t \text{ for all } t < \infty\}. \tag{1.21}$$

The random shift θ_τ on D_Y is defined by $\theta_\tau \omega(s) = \omega(\tau(\omega) + s)$.

Proposition 1.4 *(Strong Markov property) Suppose $\{P^x\}$ is a Feller process on D_Y, and τ is a stopping time. Then*

$$P^x[\, \theta_\tau^{-1} A \,|\, \mathcal{F}_\tau \,](\omega) = P^{\omega(\tau)}(A)$$

for P^x-almost every ω such that $\tau(\omega) < \infty$, for every $x \in Y$ and $A \in \mathcal{F}$.

Proof. We outline the somewhat lengthy proof. Check first that the result holds in case the values of τ can be arranged in an increasing sequence. This argument is the same as the proof of the strong Markov property for discrete-time Markov chains, see for example Section 5.2 in [11].

To handle the general case, we prove that

$$E^x[\mathbf{1}_B \mathbf{1}_{\{\tau<\infty\}} \cdot f \circ \theta_\tau] = E^x[\mathbf{1}_B \mathbf{1}_{\{\tau<\infty\}}\, E^{X_\tau}(f)] \tag{1.22}$$

for an arbitrary event $B \in \mathcal{F}_\tau$, and for a function f on D_Y of the form $f(\omega) = \prod_{i=1}^m f_i(\omega(t_i))$ where $f_1, \dots, f_m \in C_b(Y)$ and $0 \le t_1 < \cdots < t_m$. Let us argue why this suffices for the conclusion. By taking bounded limits of the functions f_i we can extend the validity of (1.22) to $f = \mathbf{1}_A$ for cylinder events of the type

$$A = \{\omega \in D_Y : \omega(t_1) \in A_1, \dots, \omega(t_m) \in A_m\}$$

for closed sets $A_i \subseteq Y$. Such events form a π-system and generate $\mathcal{F}$. The class of events A such that (1.22) is valid for $f = \mathbf{1}_A$ is a λ-system. Thus by the π-λ-theorem (Theorem A.1), (1.22) is valid for $f = \mathbf{1}_A$ for all $A \in \mathcal{F}$.

Now to prove (1.22). Use the Feller continuity of the transition probability to check, by induction on m, that $E^x(f)$ is a bounded continuous function of x. Set $\tau_n = 2^{-n}([2^n\tau] + 1)$. Check that τ_n is a stopping time, $\{\tau_n < \infty\} = \{\tau < \infty\}$, that $\tau_n \searrow \tau$ as $n \to \infty$, and $\mathcal{F}_\tau \subseteq \mathcal{F}_{\tau_n}$. We already know the conclusion for discrete stopping times, hence (1.22) is valid for τ_n:

$$E^x[\mathbf{1}_B \mathbf{1}_{\{\tau<\infty\}}\, f \circ \theta_{\tau_n}] = E^x[\mathbf{1}_B \mathbf{1}_{\{\tau<\infty\}}\, E^{X_{\tau_n}}(f)]. \tag{1.23}$$

Let $n \to \infty$ and check that (1.23) becomes (1.22) in the limit. This follows from the right-continuity of the paths ω. ∎

1.4 Poisson processes

Poisson processes on $[0, \infty)$ are central examples of continuous-time Markov chains, and also a building block of the interacting processes we construct in Section 2.

A *homogeneous Poisson process* with rate $r \in (0, \infty)$ is a Markov chain N_t on the state space $\mathbf{Z}_+ = \{0, 1, 2, 3, \ldots\}$ of nonnegative integers, whose rate matrix Q is given by $q(j, j+1) = r$ and $q(j, j) = -r$ for all $j \in \mathbf{Z}_+$. In words, N_t marches upward one step at a time, and the waiting time between each step is exponentially distributed with mean r^{-1}.

To construct N_t by the method of the previous section, introduce a deterministic, discrete time chain $Y_j = j$, and holding times $\{\sigma_j\}$ which are now i.i.d. exponential with mean r^{-1}. Set again $T_n = \sigma_0 + \cdots + \sigma_{n-1}$. With initial state $N_0 = 0$, (1.5) becomes

$$\begin{aligned} N_t &= \sum_{n=0}^{\infty} n \cdot \mathbf{1}_{[T_n, T_{n+1})}(t) = \sum_{n=1}^{\infty} \sum_{j=1}^{n} \mathbf{1}_{[T_n, T_{n+1})}(t) = \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} \mathbf{1}_{[T_n, T_{n+1})}(t) \\ &= \sum_{j=1}^{\infty} \mathbf{1}_{[T_j, \infty)}(t) = \sum_{j=1}^{\infty} \mathbf{1}_{(0,t]}(T_j). \end{aligned}$$

The last formulation suggests a different point of view. Regard the $\{T_j\}$ as random points on $(0, \infty)$, and let N_t be the number of these points in the interval $(0, t]$. A natural generalization is to define the *random counting measure* $N(\cdot)$ by

$$N(B) = \sum_{j=1}^{\infty} \mathbf{1}_B(T_j) \tag{1.24}$$

for Borel sets $B \subseteq (0, \infty)$. Then $N_t = N(0, t]$ is the special case where $B = (0, t]$. The process $\{N(B) : B \in \mathcal{B}(0, \infty)\}$ is uniquely determined by these two properties:

(a) Let $|B|$ denote the Lebesgue measure of B. If $|B| < \infty$, $N(B)$ is Poisson distributed with mean $r|B|$, in other words

$$P[N(B) = k] = \frac{e^{-r|B|}(r|B|)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

While if $|B| = \infty$, $P[N(B) = \infty] = 1$.

(b) For pairwise disjoint Borel sets $B_1, B_2, \ldots, B_m$, the random variables $N(B_1)$, $N(B_2)$, $\ldots$, $N(B_m)$ are independent.

For a proof that $\{N(B)\}$ defined by (1.24) has the above properties, see Section 4.8 in [29]. This makes $N(\cdot)$ into a Poisson random measure with mean measure r times Lebesgue measure.

We give an alternative construction of $\{N(B)\}$ that is useful for many proofs. This construction satisfies immediately properties (a)–(b) above. It works for arbitrary σ-finite mean measures on general spaces, which the "renewal definition" (1.24) does not do.

Let $(0,\infty) = \cup_{i=1}^{\infty} U_i$ be a decomposition of $(0,\infty)$ as a union of pairwise disjoint, bounded intervals. For each i, let $\{w_k^i\}_{1\le k<\infty}$ be i.i.d. random variables, uniformly distributed in the interval U_i, and such that all the random variables $\{w_k^i\}_{1\le i,k<\infty}$ are independent. Let $\{K_i\}$ be independent Poisson distributed random variables, independent of the $\{w_k^i\}$, with means $EK_i = r|U_i|$. The random point set that gives the Poisson process is the set

$$\mathcal{T} = \{w_k^i : i = 1,2,3,\ldots,\ 1 \le k \le K_i\},$$

or in terms of the random counting measure,

$$N(B) = \sum_{i=1}^{\infty} \sum_{k=1}^{K_i} \mathbf{1}_B(w_k^i). \tag{1.25}$$

We shall alternate freely between different points of view of the Poisson process on $(0,\infty)$: as an ordered sequence of jump times $0 < T_1 < T_2 < T_3 < \cdots$, as the random set $\mathcal{T} = \{T_1, T_2, T_3, \ldots\}$, or as the counting function $N_t = |\mathcal{T} \cap (0,t]\,|$. In Section 2 Poisson processes on $(0,\infty)$ serve as the random clocks in the construction of the exclusion process. These basic properties will be needed.

Proposition 1.5 (a) *Suppose $\{\mathcal{T}_j\}$ is a family of mutually independent Poisson point processes on $(0,\infty)$ with rates r_j, respectively, and $r = \sum r_j < \infty$. Let $\mathcal{T} = \cup_j \mathcal{T}_j$. Then $\mathcal{T}$ is a Poisson point process with rate r. For any time point $0 < s < \infty$, the first point of $\mathcal{T}$ after s comes from $\mathcal{T}_j$ with probability r_j/r.*

(b) *Let $\mathcal{T}$ be a Poisson point process with rate r, and let $\{p_i\}$ be a probability distribution on $\mathbf{N}$. To each point $t \in \mathcal{T}$, assign independently a mark $Y_t \in \mathbf{N}$ with probabilities $P[Y_t = i] = p_i$. Set $\mathcal{T}_i = \{t \in \mathcal{T} : Y_t = i\}$. Then $\{\mathcal{T}_i\}$ are mutually independent Poisson point processes with rates $\{p_i r\}$.*

The proof of Proposition 1.5 as left as an exercise. The Feller continuity of a single Poisson process $N(t)$ is immediate because its state space $\mathbf{Z}_+$ has the discrete topology. Later we need to consider a countably infinite family $\bar{N}(t) = \{N_i(t) : i \in I\}$ of Poisson processes, indexed by a subset I of a some square lattice $\mathbf{Z}^d$. The state space of $\bar{N}(\cdot)$ is $\mathbf{Z}_+^I$, which is a Polish space with its product metric. Feller continuity is true again, and so in particular the strong Markov property holds.

We conclude this section with an alternative construction of the continuous-time Markov chain X_t of Section 1.2. This construction is better because it simultaneously constructs the chains from all initial states on a single probability space. Such a simultaneous construction of several processes is a *coupling*. This construction is the same as the graphical representation of the exclusion process in Section 2, except that here there is no interaction.

The probability space is $(\Omega, \mathcal{H}, \mathbf{P})$ on which are defined independent Poisson point processes $\{\mathcal{T}_{(x,y)} : (x,y) \in S^2, x \neq y\}$ on $(0,\infty)$. The rate of $\mathcal{T}_{(x,y)}$ is $q(x,y) = c(x)p(x,y)$. Perform the following mental construction. To each $x \in S$ attach a time axis $[0,\infty)$, to create the product space $S \times [0,\infty)$. For each $t \in \mathcal{T}_{(x,y)}$, create an arrow $((x,t),(y,t))$ that emanates from (x,t) and points to (y,t). For each initial state x, we define a path (t, X_t^x) for $0 \le t < \infty$ through the space $S \times [0,\infty)$ that moves at rate 1 along a time axis, and instantaneously jumps along any arrow it encounters (but only in the correct direction).

Given an initial state x, define the path X_t^x explicitly as follows: Set $T_0 = 0$, $y_0 = x$, and $X_0^x = y_0$. Let T_1 be the first time t when an arrow emanates from (y_0, t), and suppose this arrow points to (y_1, t). Then define

$$X_t^x = y_0 \text{ for } T_0 < t < T_1, \text{ and } X_{T_1}^x = y_1.$$

Now repeat the same step. Let T_2 be the first time t after T_1 that an arrow emanates from (y_1, t), and suppose this arrow points to (y_2, t). Then continue defining the evolution:

$$X_t^x = y_1 \text{ for } T_1 < t < T_2, \text{ and } X_{T_2}^x = y_2.$$

Continuing in this manner, we obtain a sequence of times $0 = T_0 < T_1 < T_2 < T_3 < \cdots$ and states $x = y_0, y_1, y_2, y_3, \ldots$ with the property that no arrows emanate from (y_i, t) for $T_i \le t < T_{i+1}$, and $((y_i, T_{i+1}), (y_{i+1}, T_{i+1}))$ is an arrow for each i. The path is defined by $X_t^x = y_i$ for $T_i \le t < T_{i+1}$.

Proposition 1.6 *The path X_t^x defined above is a Markov process. It has the following property: After a jump to a state (say) x, the holding time in x is exponentially distributed with mean $c(x)^{-1}$ and independent of the past, and the next state y is selected with probability $p(x,y)$, independently of everything else.*

Proof. We first prove that X_t^x is a Markov process. Let $\overline{\mathcal{T}} = \{\mathcal{T}_{(x,y)}\}$ represent the entire family of Poisson point processes. Let $\mathcal{H}_t$ be the σ-algebra of the Poisson processes on the time interval $(0,t]$. Time shifts θ_s act on Poisson processes as they did on the path space in Section 1.3. In terms of the counting function, $\theta_s N_t = N_{s+t}$. The effect on the random measure or the random set is to restart the counting from time s:

$$\theta_s N(t,u] = \theta_s N_u - \theta_s N_t = N(s+t, s+u], \quad \text{and} \quad \theta_s \mathcal{T} = \{t - s : t \in \mathcal{T}, t > s\}.$$

Think of the construction of X_t^x as a family of maps G_t, so that $X_t^x = G_t(x, \overline{\mathcal{T}})$ constructs the state X_t^x from the inputs x (the initial state) and $\overline{\mathcal{T}}$. Let

$$p_t(x,y) = \mathbf{P}[X_t^x = y] = \mathbf{P}[G_t(x, \overline{\mathcal{T}}) = y].$$

The construction of X^x_{s+t} can be done in two stages, first from time 0 to s, and then from s to $s+t$. This restarting of the construction can be expressed as $X^x_{s+t} = G_t(X^x_s, \theta_s\overline{\mathcal{T}})$. Hence

$$\mathbf{P}[X^x_{s+t} = y \,|\, \mathcal{H}_s](\omega) = \mathbf{P}[G_t(X^x_s, \theta_s\overline{\mathcal{T}}) = y \,|\, \mathcal{H}_s](\omega) = p_t(X^x_s(\omega), y).$$

The last step is a consequence of several points. X^x_s is $\mathcal{H}_s$-measurable while $\theta_s\overline{\mathcal{T}}$ is independent of $\mathcal{H}_s$, because $\theta_s\overline{\mathcal{T}}$ depends only on Poisson points in (s,∞) and Poisson points in disjoint sets are independent. We can apply a basic property of conditional expectations: if Y is $\mathcal{B}$-measurable and $\sigma(Z)$ is independent of $\mathcal{B}$, then $E[\varphi(Y,Z)\,|\,\mathcal{B}] = g(Y)$ where $g(y) = E[\varphi(y,Z)]$. Then note that $\theta_s\overline{\mathcal{T}}$ has the same distribution as $\overline{\mathcal{T}}$.

The equation above implies that X^x_t is Markovian with transition probability $p_t(x,y)$, because the past $\{X^x_u : 0 \le u \le s\}$ is $\mathcal{H}_s$-measurable.

Next we check by induction the second part of the statement of the Proposition. As in the construction, suppose $X^x_{T_n} = y_n$. The construction from time T_n onwards is given by $X^x_{T_n+t} = G_t(y_n, \theta_{T_n}\overline{\mathcal{T}})$. T_n is a stopping time for the Poisson processes, so by the strong Markov property, $\theta_{T_n}\overline{\mathcal{T}}$ is independent of $\mathcal{H}_{T_n}$ and distributed as $\overline{\mathcal{T}}$. This is because $\mathcal{T}$ is a function of $(N_t - N_0 : t > 0)$, and hence independent of N_0 by the Poisson construction. The state y_n is $\mathcal{H}_{T_n}$-measurable, so the restarted Poisson processes $\theta_{T_n}\overline{\mathcal{T}}$ are independent of that too.

Let $z = y_n$. Apply Proposition 1.5(a) to the Poisson processes $\{\theta_{T_n}\mathcal{T}_{(z,y)} : y \in S\}$ and $\theta_{T_n}\mathcal{T}_z = \cup_y \theta_{T_n}\mathcal{T}_{(z,y)}$ restarted at time T_n. $\theta_{T_n}\mathcal{T}_z$ has rate $c(z)$, so $T_{n+1} - T_n$ has exponential distribution with mean $c(z)^{-1}$. The arrow that emanates from (z, T_{n+1}) points to (y, T_{n+1}) if the first jump time came from $\theta_{T_n}\mathcal{T}_{(z,y)}$, which happens with probability $p(z,y)$. ∎

1.5 Harmonic functions for Markov chains

This section introduces the important coupling technique for Markov chains, and collects results about harmonic functions needed later. For a stochastic matrix $p(x,y)$, the m-step transition probabilities $p^{(m)}(x,y)$ are obtained from the mth power of the matrix, inductively by

$$p^{(m)}(x,y) = \sum_{z\in S} p^{(m-1)}(x,z)p(z,y).$$

A function h on S is *harmonic* for the transition $p(x,y)$ if $h(x) = \sum_y p(x,y)h(y)$ for all x, and the sum on the right is defined. Then by induction on n,

$$h(x) = E^x[h(X_1)] = E^x[h(X_n)]$$

for the discrete-time chain X_n with transition $p(x,y)$.

In general, a *coupling* of two stochastic processes X_n and Y_n is a realization of the two processes on the same probability space. One studies the joint process (X_n, Y_n) to learn something about the marginal processes. A coupling is *successful* if with probability 1 the processes X_n and Y_n eventually stay together. In other words, there exists an almost surely finite random N such that $X_n = Y_n$ for all $n \geq N$.

Lemma 1.7 *Suppose that two copies of the Markov chain with transition $p(x, y)$ can be coupled successfully for any pair of starting states (x, y). Then every bounded harmonic function for this transition is constant.*

Proof. Let h be bounded harmonic, and fix two states x, y. We shall couple two versions of the Markov chain, one started at x and the other at y, to show that $h(x) = h(y)$. Let (X_n, Y_n) be a successful coupling with starting state $(X_0, Y_0) = (x, y)$, defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let $N(\omega)$ be the a.s. finite random time such that $X_n(\omega) = Y_n(\omega)$ for $n \geq N(\omega)$. Since both X_n and Y_n are Markov chains with transition $p(u, v)$,

$$\begin{aligned}
|h(x) - h(y)| &= |E^x[h(X_n)] - E^y[h(X_n)]| \\
&= |\mathbf{E}[h(X_n)] - \mathbf{E}[h(Y_n)]| = |\mathbf{E}[h(X_n) - h(Y_n)]| \leq \mathbf{E}|h(X_n) - h(Y_n)| \\
&\leq 2\|h\|_\infty \mathbf{P}(X_n \neq Y_n) \leq 2\|h\|_\infty \mathbf{P}(N > n).
\end{aligned}$$

Since $N < \infty$ a.s., letting $n \nearrow \infty$ shows $h(x) = h(y)$. ∎

According to the standard Markov chain definition, a transition probability $p(x, y)$ is *irreducible* if for all states x and y there exists an m so that $p^{(m)}(x, y) > 0$. For interacting systems, the following more inclusive definitions of irreducibility are sometimes useful. The first way to relax the definition is to require that for some m,

$$p^{(m)}(x, y) + p^{(m)}(y, x) > 0.$$

We can relax this further by permitting each step of the path from x to y to be traversed in either direction.

$$\begin{aligned}
&\text{For every pair of states } x \text{ and } y \text{ there exists a finite sequence} \\
&\text{of states } x = x^{(1)}, x^{(2)}, \ldots, x^{(k)} = y \text{ such that} \\
&p(x^{(j)}, x^{(j+1)}) + p(x^{(j+1)}, x^{(j)}) > 0 \text{ for } j = 0, \ldots, k-1.
\end{aligned} \tag{1.26}$$

A discrete time random walk on the countable state space $S = \mathbf{Z}^d$ is a Markov chain X_n that can be represented as a sum $X_n = x + \xi_1 + \cdots + \xi_n$ for i.i.d. step variables ξ_k. Equivalently, the transition probability is translation invariant in the sense

$$p(x, y) = p(0, y - x), \tag{1.27}$$

and then $\{p(0, x) : x \in S\}$ is the step distribution.

Theorem 1.8 *Constants are the only bounded harmonic functions for a random walk on* $\mathbf{Z}^d$ *that is irreducible in the sense* (1.26).

Proof. It suffices to construct a successful coupling started from $(X_0, Y_0) = (x, y)$ for each pair $x \neq y$ such that $p(x, y) > 0$. Then $h(x) = h(y)$ if $p(x, y) + p(y, x) > 0$, and by the irreducibility assumption (1.26) any pair x, y can be connected by a finite sequence $x = x^{(1)}, x^{(2)}, \ldots, x^{(k)} = y$ such that $h(x^{(j)}) = h(x^{(j+1)})$. We may assume that $p(0, 0) > 0$, for otherwise the original transition can be replaced by $\tilde{p}(x, y) = \frac{1}{2}p(x, y) + \frac{1}{2}\mathbf{1}_{\{x=y\}}$ which has the same harmonic functions as $p(x, y)$.

Fix $x \neq y$ such that $p(x, y) > 0$. The joint process (X_n, Y_n) will be a Markov chain on the state space

$$\mathcal{X} = \{(u, v) \in \mathbf{Z}^d \times \mathbf{Z}^d : v - u \text{ is an integer multiple of } y - x\}$$

started from (x, y). Let $\beta = p(0, 0) \wedge p(x, y) \in (0, 1/2]$. The upper bound of 1/2 comes from $p(0, 0) + p(0, y - x) \leq 1$.

Define the joint transition for $u \neq v$ by

$$\begin{aligned}
p((u, v), (u + w, v + w)) &= p(0, w) \quad \text{for } w \neq 0, y - x, \\
p((u, v), (u, v + y - x)) &= \beta, \\
p((u, v), (u + y - x, v)) &= \beta, \\
p((u, v), (u, v)) &= p(0, 0) - \beta, \\
p((u, v), (u + y - x, v + y - x)) &= p(x, y) - \beta,
\end{aligned}$$

and for $u = v$ by $p((u, u), (u + w, u + w)) = p(0, w)$. The chain with these transitions stays in $\mathcal{X}$.

Let B_n be the integer defined by $Y_n - X_n = B_n(y - x)$. Then B_n is a Markov chain on $\mathbf{Z}$ with transitions $p(k, k \pm 1) = \beta$ and $p(k, k) = 1 - 2\beta$ for $k \neq 0$, and $p(0, 0) = 1$. By the recurrence of one-dimensional simple symmetric random walk, eventually B_n will be absorbed at 0, which is the same as saying that eventually $X_n = Y_n$. ∎

For a continuous-time Markov chain X_t, a function h is harmonic if $h(x) = E^x h(X_t)$ for all starting states x and $t \geq 0$. In the special case where all clocks ring at the same rate c, and the new state is chosen according to a probability kernel $p(x, y)$, the transition probability $p_t(x, y) = P^x[X_t = y]$ of the continuous-time chain can be written down explicitly:

$$p_t(x, y) = \sum_{n=0}^{\infty} \frac{e^{-ct}(ct)^n}{n!} p^{(n)}(x, y). \tag{1.28}$$

The index n in the sum represents the number of jumps that the chain has experienced up to time t.

Exercise 1.1 Suppose X_t has transition probabilities given by (1.28). Show that a function h is harmonic for X_t iff it is harmonic for the discrete-time transition $p(x, y)$.

Theorem 1.8 and the above exercise give the following corollary.

Corollary 1.9 *Suppose X_t has transition probabilities given by* (1.28)*, and $p(x, y)$ is translation invariant in the sense of* (1.27) *and irreducible in the sense of* (1.26)*. Then every bounded harmonic function for X_t is constant.*

Finally, we check that certain limits are harmonic functions.

Lemma 1.10 *Let X_t be a continuous-time Markov chain on a countable state space S, and assume jump rates are uniformly bounded. Suppose there is a sequence of times $t_j \nearrow \infty$ and a bounded function g such that the limit $h(x) = \lim_{j\to\infty} E^x g(X_{t_j})$ exists for all $x \in S$. Then h is a harmonic function, in other words $h(x) = E^x h(X_t)$ for all $x \in S$ and $t \geq 0$.*

Proof. The generator of X_t is of the form

$$Lf(x) = \sum_y c(x)r(x, y)[f(y) - f(x)]$$

where $c(x)$ is the rate at which X_t jumps from state x, and $r(x, y)$ is the probability of choosing y as the next state. Assume r is a Markov transition, in other words $\sum_y r(x, y) = 1$ for each fixed x. The assumption is that $c(x) \leq c$ for a constant c. By introducing additional "dummy" jumps from x to x, we can make all clocks ring at uniform rate c. Then the new jump probabilities are

$$p(x, y) = \begin{cases} c^{-1}c(x)r(x, y) & \text{if } y \neq x \\ 1 - c^{-1}c(x)(1 - r(x, x)) & \text{if } y = x. \end{cases}$$

The transition probability $p_t(x, y)$ of X_t can then be expressed as

$$p_t(x, y) = \sum_{n=0}^{\infty} \frac{e^{-ct}(ct)^n}{n!} p^{(n)}(x, y).$$

To show $h(x) = \sum_y p_s(x, y)h(y)$, first by boundedness and Chapman-Kolmogorov,

$$\begin{aligned} \sum_z p_s(x, z)h(z) &= \lim_{j\to\infty} \sum_{z,y} p_s(x, z)p_{t_j}(z, y)g(y) \\ &= \lim_{j\to\infty} \sum_y p_{s+t_j}(x, y)g(y). \end{aligned}$$

Then

$$
\begin{aligned}
&\left|\sum_y p_s(x,y)h(y) - h(x)\right| \\
\le\ & \lim_{j\to\infty} \sum_y |p_{s+t_j}(x,y) - p_{t_j}(x,y)| \cdot |g(y)| \\
\le\ & \lim_{j\to\infty} \|g\|_\infty \sum_{n=0}^{\infty} \left| \frac{e^{-cs-ct_j}(cs+ct_j)^n}{n!} - \frac{e^{-ct_j}(ct_j)^n}{n!} \right| \sum_y p^{(n)}(x,y) \\
\le\ & \lim_{j\to\infty} \|g\|_\infty \sum_{n=0}^{\infty} \frac{e^{-ct_j}(ct_j)^n}{n!} \left| e^{-cs}\left(1+\frac{s}{t_j}\right)^n - 1 \right|.
\end{aligned}
$$

To see that this last line tends to 0 as $j \nearrow \infty$, think of the sum as $E\phi(t_j, Y_j)$ where Y_j is Poisson(ct_j) distributed, and $\phi(t,n) = |e^{-cs}(1+s/t)^n - 1|$. First check that $\phi(t_j, Y_j) \to 0$ in probability by showing that

$$\lim_{j\to\infty} P\{|Y_j - ct_j| \ge \delta t_j\} = 0$$

for every $\delta > 0$, and that, given $\varepsilon > 0$, there exists $\delta > 0$ such that $\phi(t,n) \le \varepsilon$ if $t > \delta^{-1}$ and $|n - ct| \le \delta t$. Second, since $\phi(t,n) \le e^{ns/t} + 1$, direct computation shows that $\sup_j E[\phi(t_j, Y_j)^2] < \infty$. These suffice for $E\phi(t_j, Y_j) \to 0$. We leave the details as exercise. ∎

Exercise 1.2 Prove Theorem 1.2. The key is to decompose $E^x[f(X_t)]$ according to how many jumps the Markov chain experienced in $(0, t]$.

Exercise 1.3 Prove Proposition 1.5 for Poisson processes.

Exercise 1.4 Fill in the details for the proof of Lemma 1.10. Look at a periodic example to show that Lemma 1.10 is not necessarily true for a discrete-time Markov chain.

Exercise 1.5 Suppose $p(x,y)$ is symmetric and translation invariant, in other words $p(x,y) = p(y,x) = p(0, y-x)$. Let X_t and Y_t be independent copies of the continuous time Markov chain with jump rates $p(x,y)$. Let $Z_t = X_t - Y_t$. Prove that the process $(Z_t : t \ge 0)$ has the same distribution as the process $(X_{2t} : t \ge 0)$. In other words, the difference $X_t - Y_t$ is the same as the original process, run at twice the speed.

Hint: Let $p_t(x,y)$ be the common transition probability of X_t and Y_t. Show that

$$
\begin{aligned}
&P^{x,y}[Z_{t_1} = z_1, Z_{t_2} = z_2, \ldots, Z_{t_n} = z_n] \\
&\qquad = p_{2t_1}(x-y, z_1)\, p_{2(t_2-t_1)}(z_1, z_2) \cdots p_{2(t_n-t_{n-1})}(z_{n-1}, z_n)
\end{aligned}
$$

for any $0 \le t_1 < t_2 < \cdots < t_n$, using induction on n.

Notes

It is not possible to rigorously define models of stochastic processes without some measure theory. Product measure spaces are especially important for probability theory because the product construction corresponds to the probabilistic notion of independence. Sources for the measure theory needed here are for example the appendix of [11], or any of the standard real analysis textbooks, such as [17]. A proof of Kolmogorov's Extension Theorem for an arbitrary index set can be found in Chapter 12 of [10]. Section 1.5 is from Liggett's monograph [27]. The reader is referred to [6] for a comprehensive treatment of Poisson point processes.

2 Construction of the exclusion process

Assume given a transition probability $p(x,y)$ on the lattice $S = \mathbf{Z}^d$, in other words nonnegative numbers that satisfy $\sum_{y \in S} p(x,y) = 1$ for each x. Our standing assumptions are that $p(x,y)$ is

$$\text{translation invariant: } p(x,y) = p(0, y-x) \tag{2.1}$$

and

$$\text{finite range: there exists a finite set } B^p \subseteq S \text{ such that } p(0,x) = 0 \text{ for } x \notin B^p. \tag{2.2}$$

We wish to construct a Markov process that corresponds to the following idea. Particles are distributed on the points of S (we call these points *sites*) subject to the restriction that no two particles occupy the same site. Each particle waits an exponentially distributed random time with mean 1, independently of the other particles, and then attempts to jump. If the particle is at x, it chooses a new site y with probability $p(x,y)$. If site y is vacant, this particle moves to y and site x becomes empty. If site y was already occupied, the jump is cancelled and the particle remains at x. In either case, the particle resumes waiting for another exponentially distributed random time, independent of the past and the rest of the system, after which it attempts a new jump to a new randomly chosen target site y. All particles are going through this cycle of waits and jump attempts. The random waiting times and choices of target site are mutually independent and independent of the rest of the system. The interaction between the particles happens through the exclusion rule which stipulates that jumps to already occupied sites are not permitted. Without this rule all the particles would be simply moving as independent Markov chains on S with jump rates $p(x,y)$.

Note that because the waiting time distribution is continuous, with probability one no two particles ever attempt to jump at the same time, so no conflicts arise between two particles attempting to jump to the same vacant site.

We can assume that $p(0,0) = 0$. Otherwise we could define a new kernel by $\widetilde{p}(0,0) = 0$ and $\widetilde{p}(0,x) = p(0,x)/(1-p(0,0))$ for $x \neq 0$. This eliminates jump attempts from x to x that have no effect on the configuration, and runs the process faster by a factor of $(1-p(0,0))^{-1}$.

We define the state of the system to keep track of the occupied and vacant sites. For each $x \in S$, let $\eta(x) = 1$ if x is occupied, and $\eta(x) = 0$ if x is empty. Thus the state is a configuration $\eta = (\eta(x) : x \in S)$ of 0's and 1's, and the state space is the product space $X = \{0,1\}^S$. The goal of this section is to rigorously construct a Markov process $\eta_t = (\eta_t(x))_{x \in S}$ that operates according to the description given above. The state space X is uncountable, so existence of this process does not follow from our earlier construction of countable state Markov chains.

2.1 Graphical representation of the exclusion process

Let $S_p^2 = \{(x,y) \in S^2 : p(x,y) > 0\}$ be the set of pairs of sites between which jump attempts can happen. Let $(\Omega, \mathcal{H}, \mathbf{P})$ be a probability space on which is defined a family $\{\mathcal{T}_{(x,y)} : (x,y) \in S_p^2\}$ of mutually independent Poisson point processes on the time line $[0,\infty)$. Poisson process $\mathcal{T}_{(x,y)}$ is homogeneous with rate $p(x,y)$. The jump times of $\mathcal{T}_{(x,y)}$ are the random times at which we will attempt to move a particle from x to y.

As explained in Section 1.4, we can switch freely between representing a Poisson process as the random set $\mathcal{T}_{(x,y)}$, as the random measure $N_{(x,y)}(B) = |\mathcal{T}_{(x,y)} \cap B|$ for Borel sets $B \subseteq [0,\infty)$, or as the counting function $N_{(x,y)}(t) = N_{(x,y)}((0,t])$.

Let

$$\mathcal{T}_x = \bigcup_y \mathcal{T}_{(x,y)} \qquad \text{and} \qquad \mathcal{T}'_x = \bigcup_y \left(\mathcal{T}_{(x,y)} \cup \mathcal{T}_{(y,x)}\right). \tag{2.3}$$

$\mathcal{T}_x$ is the set of times when a particle attempts to jump out of x, if x is occupied. $\mathcal{T}_x$ is a Poisson process of rate $\sum_y p(x,y) = 1$. $\mathcal{T}'_x$ is the set of all times when a jump either in or out of x can happen. $\mathcal{T}'_x$ is a Poisson process of rate

$$\sum_y (p(x,y) + p(y,x)) = \sum_y p(x,y) + \sum_y p(0, x-y) = 2,$$

where we used the translation invariance assumption.

According to Proposition 1.5, attaching the independent Poisson processes $\{\mathcal{T}_{(x,y)}\}$ to edges is equivalent to attaching a single Poisson point process $\mathcal{T}_x$ of rate 1 to each site x, and then assigning each $t \in \mathcal{T}_x$ to a particular edge (x,y) with probability $p(x,y)$. For our discussion it is convenient to have the Poisson processes $\{\mathcal{T}_{(x,y)}\}$ given at the outset. So informally, instead of having one alarm clock at x and then flipping a $p(x,y)$-coin after the clock rings, we attach clocks to all edges (x,y) and react whenever one of them rings.

Fix a sample point $\omega \in \Omega$, in other words a realization $\{\mathcal{T}_{(x,y)}\}$ of the Poisson processes. By discarding a set of $\mathbf{P}$-probability zero, we may assume that

> each $\mathcal{T}'_x$ has only finitely many jump times in every bounded interval $(0,T]$, and no two distinct processes $\mathcal{T}_{(x,y)}$ and $\mathcal{T}_{(x',y')}$ have a jump time in common. (2.4)

Assume given an initial state $\eta \in X$.

The term "graphical representation" refers to the following space-time picture. Put the lattice $S = \mathbf{Z}^d$ on the horizontal axis. (It may be necessary to take $d = 1$ to make drawing feasible!) To each $x \in S$ attach a vertical time axis oriented upward. At each jump time t of $\mathcal{T}_{(x,y)}$, draw an arrow from (x,t) to (y,t). Put the initial particle configuration at level $t = 0$ on the sites of S. After the process starts, all particles move vertically upward at a steady rate 1. When a particle encounters an arrow at (x,t) pointing to (y,t), it moves instantaneously along the arrow from (x,t) to (y,t) in case y is vacant at time t (in other words, if there is no particle at (y,t) blocking the arrow). This way each particle traces a trajectory in the space-time picture, moving vertically upward at rate 1 and occasionally sideways along an arrow.

A problem arises. Suppose we compute the value $\eta_t(0)$ for some $t > 0$. This value is potentially influenced by $\eta_s(x)$, $0 \le s \le t$, from all sites x that interacted with 0 during the time interval $(0,t]$. This means those x for which $\mathcal{T}_{(x,0)}$ or $\mathcal{T}_{(0,x)}$ had a jump time during $(0,t]$. But to know $\eta_s(x)$, $0 \le s \le t$, for these x, we have to consider what happened at all sites y that interacted with any one of these x-sites. And so on. How does the construction ever get off the ground? For the Markov chain we could wait for the first jump time before anything happened. But now, if the initial state η has infinitely many particles, infinitely many of them attempt to jump in every nontrivial time interval $(0,\varepsilon)$.

The percolation argument

To get around this difficulty we use a percolation argument due to T. Harris. This guarantees that for a short, but positive, deterministic time interval $[0,t_0]$, the entire set S decomposes into disjoint finite components that do not interact during $[0,t_0]$. In each finite component the evolution η_t ($0 \le t \le t_0$) can be constructed because only finitely many jump times need to be considered.

For $0 \le s < t$, let $\mathcal{G}_{s,t}$ be the undirected random graph with vertex set S and edge set $\mathcal{E}_{s,t}$ defined by

$$\{x,y\} \in \mathcal{E}_{s,t} \text{ iff } \mathcal{T}_{(x,y)} \text{ or } \mathcal{T}_{(y,x)} \text{ has a jump time in } (s,t]. \tag{2.5}$$

Each edge $\{x,y\}$ is present in $\mathcal{G}_{s,t}$ with probability $1 - e^{-(t-s)(p(x,y)+p(y,x))}$, independently of the other edges. Quite obviously, as s is fixed and t increases edges are only added to the graph, never removed.

The connection with the exclusion evolution is that in order to compute the evolution $\eta_s(x)$ for $0 \le s \le t$, only those sites that lie in the same connected component as x in the graph $\mathcal{G}_{0,t}$ are relevant.

Lemma 2.1 *If t_0 is small enough, the random graph $\mathcal{G}_{0,t_0}$ has almost surely only finite connected components.*

Proof. Let $B_* = B^p \cup (-B^p)$, the symmetrized version of the set B^p in the finite range assumption (2.2), $R = \max_{x \in B_*} |x|_\infty$ the radius and $k_* = |B_*|$ the cardinality of the set B_*. By (2.2), $\mathcal{T}_{(x,y)}$ cannot have any jump times unless $y - x \in B^p$. Hence if $\{x,y\} \in \mathcal{E}_{0,t}$ then necessarily $y - x$ and $x - y$ lie in B_*, and so $|y - x|_\infty \le R$.

We first show that if $t_0 > 0$ is fixed small enough, then with probability 1 the connected component containing the origin is finite.

Suppose a site y with $|y|_\infty \geq L$ lies in the same connected component as 0. Then there must exist an open path $0 = x_0, x_1, \ldots, x_m = y$ in the graph, and since each edge $\{x_i, x_{i+1}\}$ can span distance at most R, $m \geq L/R$. A given sequence of sites $0 = x_0, x_1, \ldots, x_m$ such that $x_{i+1} - x_i \in B_*$ is an open path in the graph $\mathcal{G}_{0,t_0}$ with probability

$$\prod_{i=0}^{m-1} \left(1 - e^{-t_0(p(x_i,x_{i+1})+p(x_{i+1},x_i))}\right) \leq (1 - e^{-2t_0})^m.$$

The number of possible paths $0 = x_0, x_1, \ldots, x_m$ starting at the origin is k_*^m, as each successive vertex x_{i+1} must be chosen from among the k_* elements of the set $x_i + B_*$. Pick t_0 small enough so that $k_*(1 - e^{-2t_0}) < 1$. This choice is deterministic and positive. Then

$$\sum_{m=1}^{\infty} \mathbf{P}[\text{there is an open path of length } m \text{ in the graph starting at } 0]$$
$$\leq \sum_{m=1}^{\infty} k_*^m (1 - e^{-2t_0})^m < \infty.$$

By Borel-Cantelli, almost surely there is a finite upper bound on the length of the longest open path starting at 0. And consequently there is a finite L such that 0 is not connected to any vertex y such that $|y|_\infty \geq L$.

Finally, by the assumption (2.1) of translation invariance, the joint distribution of the edge indicator variables $\mathbf{1}[\{x, y\} \in \mathcal{E}_{0,t_0}]$ is translation invariant. Consequently the probability that a vertex x lies in an infinite connected component is the same for each x. By the above proof, this probability is zero. ∎

To construct the process η_t for $0 \leq t \leq t_0$, we imagine doing it separately in each finite connected component of the graph $\mathcal{G}_{0,t_0}$. Ignoring everything outside a particular connected component and considering only the time interval $[0, t_0]$ is by assumption (2.4) the same as having only finitely many edges in the entire graph $\mathcal{G}_{0,t_0}$.

Construction with finitely many jump times

Suppose the finitely many jump attempts happen at times $0 < \tau_1 < \tau_2 < \cdots < \tau_n$. Let (x_1, y_1), (x_2, y_2), $\ldots$, (x_n, y_n) be the pairs of sites such that $\tau_k \in \mathcal{T}_{(x_k,y_k)}$ for $k = 1, \ldots, n$. Recall that the initial state is η. Set

$$\eta_t = \eta_0 \text{ for } 0 \leq t < \tau_1.$$

If x_1 is occupied and y_1 is vacant at τ_1- (this means: immediately before time τ_1), move a particle from x_1 to y_1 at time τ_1, and so set

$$\eta_{\tau_1} = \eta_{\tau_1-}^{x_1,y_1}.$$

We introduced this notation: $\eta^{x,y}$ is the state that results from η after interchanging $\eta(x)$ and $\eta(y)$, in other words

$$\eta^{x,y}(z) = \begin{cases} \eta(y), & z = x \\ \eta(x), & z = y \\ \eta(z), & z \neq x, y. \end{cases} \tag{2.6}$$

If x_1 is vacant or y_1 is occupied at τ_1-, no change happens, and we set

$$\eta_{\tau_1} = \eta_{\tau_1-}.$$

We have defined the process on the time interval $[0, \tau_1]$.

The general step: Suppose the state η_t has been defined for $0 \leq t \leq \tau_k$. Define

$$\eta_t = \eta_{\tau_k} \text{ for } \tau_k < t < \tau_{k+1}.$$

If $\eta_{\tau_{k+1}-}(x_{k+1}) = 1$ and $\eta_{\tau_{k+1}-}(y_{k+1}) = 0$, set

$$\eta_{\tau_{k+1}} = \eta_{\tau_{k+1}-}^{x_{k+1},y_{k+1}}$$

while if $\eta_{\tau_{k+1}-}(x_{k+1}) = 0$ or $\eta_{\tau_{k+1}-}(y_{k+1}) = 1$, set

$$\eta_{\tau_{k+1}} = \eta_{\tau_{k+1}-}.$$

This step is repeated until the construction is done on the time interval $[0, t_0]$ for a particular connected component. Since the connected component is finite, by assumption (2.4) we reach time t_0 after finitely many updating steps. Then this construction is repeated for each connected component.

Construction for all time

Given an arbitrary initial configuration η, we can now construct the evolution η_t for $0 \leq t \leq t_0$, for almost every realization of the Poisson processes $\{\mathcal{T}_{(x,y)}\}$. Once the evolution is constructed up to time t_0, take the state η_{t_0} as the new starting state, and use the Poisson processes restricted to the time interval $(t_0, 2t_0]$. Lemma 2.1 guarantees again that with probability one, the connected components are finite for the random graph $\mathcal{G}_{t_0,2t_0}$, and thus the construction can be extended from time t_0 to $2t_0$. Continue this way, and conclude that the evolution η_t can be constructed for all time ($0 \le t < \infty$), for an arbitrary initial configuration η, and for all but an exceptional $\mathbf{P}$-null set of jump time processes $\{\mathcal{T}_{(x,y)}\}$.

We seem to have deviated from the description at the very beginning of this section. There we specified that each particle waits an exponentially distributed time, and then attempts to jump. However, in our construction the Poisson clocks $\mathcal{T}_{(x,y)}$ are attached to directed edges (x, y) between sites. This apparent difference can be cleared by the memoryless property of the exponential distribution. Let $0 < T_1 < T_2 < T_3 < \cdots$ be the times of the jump attempts experienced by a particle initially located at x_0. Let x_k be its location at time T_k, either the new location after a successful jump, or the old location after a suppressed jump. The T_k's are stopping times for the Poisson processes. By the strong Markov property, the Poisson processes start anew at time T_k independently of the past. The past includes the choices of states $x_1, \ldots, x_k$. So from the perspective of this particle, after time T_k the clocks $\mathcal{T}_{(x_k,y)}$ ring at rates $p(x_k, y)$ independently of everything in the past. So it is as if the particle were carrying his own clocks.

2.2 Stirring particle construction for the symmetric case

Suppose the jump probabilities are symmetric: $p(x, y) = p(y, x)$. Then instead of arrows in the graphical construction, we can use undirected edges. Start again by putting in the vertical time axes $\{x\} \times [0, \infty)$ for $x \in S$. Take a realization of mutually independent Poisson processes $\{\mathcal{T}_{\{x,y\}}\}$ indexed by unordered pairs $\{x, y\}$ of distinct sites, with rate $p(x, y) = p(y, x)$ for $\mathcal{T}_{\{x,y\}}$. For each jump time t of $\mathcal{T}_{\{x,y\}}$, connect the space-time points (x, t) and (y, t) with a horizontal undirected edge.

Let us say there is a path between $(x, 0)$ and (y, t) in the space-time graph if there is a sequence of times $0 = s_0 < s_1 < \cdots < s_k < s_{k+1} = t$ and sites $x = x_0, x_1, \ldots, x_k = y$ such that

(a) no horizontal edges touch the open vertical segments $\{x_i\} \times (s_i, s_{i+1})$ for $0 \le i \le k$ and

(b) each $\{(x_i, s_{i+1}), (x_{i+1}, s_{i+1})\}$ for $0 \le i \le k - 1$ is a horizontal edge.

The rules for a path leave no choices. Starting from $(x, 0) = (x_0, 0)$, the path proceeds forward in time until it encounters the first horizontal edge $\{(x_0, s_1), (x_1, s_1)\}$, which forces the path to jump to (x_1, s_1). Subsequently the path proceeds forward in time along the segment $\{x_1\} \times (s_1, s_2)$, until the edge $\{(x_1, s_2), (x_2, s_2)\}$ forces the path to jump to (x_2, s_2). And so on. Hence starting from $(x, 0)$ and proceeding to level $S \times \{t\}$, the endpoint (y, t) is uniquely determined, because there is only one path from $(x, 0)$ to level $S \times \{t\}$. Similarly one can traverse the same path backwards in time, from (y, t) to $(x, 0)$. These paths define bijective maps between $S \times \{0\}$ and $S \times \{t\}$.

To construct the exclusion process η_s for $0 \le s \le t$, place the initial particle configuration on level $S \times \{0\}$, and let the particles follow the paths to the level $S \times \{t\}$. This has the effect of swapping the contents of sites x and y at each jump time of $\mathcal{T}_{\{x,y\}}$. If x and y were both empty, nothing happened. If x and y were both occupied, the particles at x and y traded places, but no change occurred in the occupation variables $\eta(x)$ and $\eta(y)$. If x was occupied and y empty, then the particle at x jumped to y. Or vice versa. This jump happened at rate $p(x, y) = p(y, x)$, as it should have. Since this representation swaps particles when both sites are occupied, we call it the stirring representation.

We introduce notation for the backward paths. $X_s^y \in S$ is the position after time s of a walker who starts at (y, t) and follows the path from (y, t) down to level $S \times \{0\}$. The walker proceeds backward in time at rate 1, but jumps instantaneously across horizontal edges. He reads his own time forward, so $X_0^y = y$ and his location in the space-time graph is actually $(X_s^y, t - s)$ for $0 \le s \le t$. If the path connects (y, t) and $(x, 0)$, then $X_t^y = x$. The stirring representation can now be expressed as

$$\eta_t(y) = \eta_0(X_t^y). \tag{2.7}$$

Notice for later use that the backward paths also represent the exclusion process with rates $p(x, y) = p(y, x)$. If we start an exclusion process with finitely many occupied sites $\{y_1, \dots, y_n\}$, then the occupied set at time t is $\{X_t^{y_1}, \dots, X_t^{y_n}\}$.

2.3 Properties of the construction

To prove necessary measurability and continuity properties of the construction, we need to be specific about the metrics on the various spaces. The state space $X = \{0, 1\}^S$ of the exclusion process is metrized with the product metric

$$d(\eta, \zeta) = \sum_{x \in S} 2^{-|x|_\infty} |\eta(x) - \zeta(x)|. \tag{2.8}$$

Convergence $d(\eta^j, \eta) \to 0$ is equivalent to saying that, for any finite set $A \subseteq S$, there exists j_0 such that $\eta^j(x) = \eta(x)$ for all $x \in A$ and $j \ge j_0$. Under this metric X is a compact space. D_X is the space of X-valued right-continuous functions $\eta_\cdot$ on $[0, \infty)$ with left limits. Let $\mathcal{F}$ be the σ-algebra on D_X generated by coordinate mappings. This is the same as the Borel σ-algebra of the Skorokhod topology (Section A.2.2).

Think of the Poisson processes in terms of right-continuous counting functions $\omega = (N_{(x,y)}(\cdot))_{(x,y) \in S_p^2}$ where $N_{(x,y)}(t) = |\mathcal{T}_{(x,y)} \cap (0, t]|$ counts the number of Poisson jump times. The path $N_{(x,y)}(\cdot)$ is an element of $D_{\mathbf{Z}_+}$. The value $\omega(t) = (N_{(x,y)}(t))_{(x,y) \in S_p^2}$ is an element of the product space $U = \mathbf{Z}_+^{S_p^2}$. U is a Polish space under the product metric

$$d_U(\mathbf{m}, \mathbf{n}) = \sum_{(x,y)\in S_p^2} 2^{-|x|_\infty - |y|_\infty} |m_{(x,y)} - n_{(x,y)}| \wedge 1$$

where $\mathbf{m} = (m_{(x,y)})_{(x,y)\in S_p^2}$, and similarly $\mathbf{n}$, denote elements of U. The path $\omega = (\omega(t) : t \geq 0)$ is an element of the space D_U, metrized by the Skorokhod metric. This is defined as in (A.4) with d_U in place of ρ. The probability space of the Poisson processes is thus $(\Omega, \mathcal{H}, \mathbf{P}) = (D_U, \mathcal{B}(D_U), \mathbf{P})$.

Let Ω_0 be the set of paths ω that satisfy (2.4) and for which the random graphs $\mathcal{G}_{kt_0,(k+1)t_0}$, $k = 0, 1, 2, \dots$, have finite connected components. Ω_0 is a Borel subset of D_U and has probability 1 (exercise).

In the previous section we constructed the exclusion process as a function of an initial state $\eta \in X$ and a sample point $\omega \in \Omega_0$. We express this dependence on (η, ω) by writing $\eta_t^\eta(\omega)$ for the state of the process at time $t \geq 0$, and $\eta_t^\eta(x;\omega)$ for the occupation variable at site x. The process with initial state η is denoted by $\eta_\cdot^\eta$. The first item to check is that $\eta_\cdot^\eta(\omega)$ is an element of the path space D_X.

Lemma 2.2 *For $(\eta, \omega) \in X \times \Omega_0$, the function $t \mapsto \eta_t^\eta(\omega)$ is right-continuous and has left limits at all time points.*

Proof. Fix a radius $r > 0$, and consider the cube $B = \{x \in S : |x|_\infty \leq r\}$. By (2.4), the Poisson process $\cup_{x\in B}\mathcal{T}'_x$ contains only finitely many points in (t, T) for any $0 \leq t < T < \infty$. Fix t, and pick $\delta > 0$ so that $\cup_{x\in B}\mathcal{T}'_x$ has no jump times in $(t, t+\delta)$. Then no changes happen in the cube B during time interval $(t, t+\delta)$, whence $\eta_t^\eta(x;\omega) = \eta_s^\eta(x;\omega)$ for $x \in B$ and $t \leq s < t + \delta$, and so

$$\begin{aligned} d(\eta_t^\eta(\omega), \eta_s^\eta(\omega)) &\leq \sum_{x:|x|_\infty > r} 2^{-|x|_\infty} \leq \sum_{n=r+1}^{\infty} 2^{-n} \left|\{x : |x|_\infty = n\}\right| \\ &\leq \sum_{n=r+1}^{\infty} 2^{-n} 2d(2n+1)^{d-1}. \end{aligned}$$

The last expression can be made less than ε by fixing r large enough at the outset. This shows that $d(\eta_t^\eta(\omega), \eta_s^\eta(\omega)) \to 0$ as $s \searrow t$.

Considering left limits, we cannot rule out the possibility of a jump time $t \in \mathcal{T}'_x$ at some site x. Then $\eta_{t-\delta}^\eta(x;\omega)$ can differ from $\eta_t^\eta(x;\omega)$ for all small $\delta > 0$. But for any site x, there is a $\delta_x > 0$ such that $\mathcal{T}'_x$ has no jump times in $(t - \delta_x, t)$. Consequently $\eta_s^\eta(x;\omega) = \eta_u^\eta(x;\omega)$

for $s, u \in (t - \delta_x, t)$, and the limit $\eta^\eta_{t-}(x;\omega) = \lim_{s \nearrow t} \eta^\eta_s(x;\omega)$ exists. Since convergence in X is coordinatewise, this implies that the limit $\eta^\eta_{t-}(\omega) = \lim_{s \nearrow t} \eta^\eta_s(\omega)$ exists. ∎

Now that we know $\eta^\eta_\cdot(\omega)$ is an element of D_X, we need to check that it is continuous in η and measurable in ω. Both will follow from the next lemma.

Lemma 2.3 *The path $\eta^\eta_\cdot(\omega)$ is a continuous D_X-valued function of $(\eta, \omega) \in X \times \Omega_0$.*

Proof. We sketch the proof and leave details as an exercise. We use the notation established in Section A.2.2.

Fix $(\eta, \omega) \in X \times \Omega_0$. Fix an arbitrary cube $A \subseteq S$ and $T < \infty$. Let k be such that $(k-1)t_0 < T \le kt_0$ where t_0 is the number given by Lemma 2.1. There exists a finite set $B \subseteq S$ such that, if $\zeta = \eta$ on B, then $\eta^\zeta_t(x;\omega) = \eta^\eta_t(x;\omega)$ for all $x \in A$ and $0 \le t \le T$. Existence of such a B can be shown inductively on k. Let $C_{(k-1)t_0, kt_0}(x)$ denote the connected component containing x in the graph $\mathcal{G}_{(k-1)t_0, kt_0}$. If $k = 1$, take

$$B = \bigcup_{x \in A} C_{0,t_0}(x).$$

Suppose a finite B exists for $k = n$ and any finite A. Set

$$A' = \bigcup_{x \in A} C_{nt_0,(n+1)t_0}(x),$$

and take the set B corresponding to A' and $k = n$. Thus a finite set B exists for A and $k = n + 1$. The point of the inductive argument is that, if a state η_{nt_0} at time nt_0 is given, the evolution $\{\eta_t(x) : nt_0 \le t \le (n+1)t_0\}$ depends on η_{nt_0} only through the values $\{\eta_{nt_0}(y) : y \in C_{nt_0,(n+1)t_0}(x)\}$.

Let $\varepsilon > 0$. Let $\omega' = (N'_{(x,y)}(\cdot))$ be another element of Ω_0. If $\delta > 0$ is chosen small enough, then $s(\omega, \omega') < \delta$ guarantees that for some $\lambda \in \Lambda$, $\gamma(\lambda) < \varepsilon$ and for all edges (x, y) incident to the set B,

$$N_{(x,y)}(t) = N'_{(x,y)}(\lambda(t)) \quad \text{for } 0 \le t \le T.$$

Shrink δ further if necessary so that $d(\eta, \zeta) < \delta$ forces $\eta = \zeta$ on the set B.

Now we have, for $(\zeta, \omega') \in X \times \Omega_0$ such that $d(\eta, \zeta) + s(\omega, \omega') < \delta$, for $x \in A$ and $0 \le t \le T$,

$$\eta^\eta_t(x;\omega) = \eta^\zeta_t(x;\omega) = \eta^\zeta_t(x;\omega' \circ \lambda) = \eta^\zeta_{\lambda(t)}(x;\omega').$$

In the first step we can replace η by ζ because only the initial values on B matter for the state at time t on the set A. Next we can replace ω by $\omega' \circ \lambda$ because these two paths agree on all jumps that influence the set B up to time T. The last step follows because the number of jumps the path $\omega'(\lambda(\cdot))$ has in time interval $(s,t]$ is the same as ω' has in $(\lambda(s),\lambda(t)]$. Then

$$\sup_{0\le t\le T} d(\eta^\eta_t(\omega),\eta^\zeta_{\lambda(t)}(\omega')) = \sup_{0\le t\le T}\sum_x 2^{-|x|_\infty}|\eta^\eta_t(x;\omega)-\eta^\zeta_{\lambda(t)}(x;\omega')| \le \sum_{x\notin A} 2^{-|x|_\infty}.$$

This last quantity can be made arbitrarily small by choosing A large enough. By Lemma A.2, we have shown that if $d(\eta,\eta^{(n)})+s(\omega,\omega^{(n)})\to 0$, then

$$s(\eta^{\eta^{(n)}}_\cdot(\omega^{(n)}),\eta^\eta_\cdot(\omega))\to 0$$

as $n\to\infty$. ∎

The lemma above implies that $\eta^\eta_\cdot$ is a measurable mapping from Ω_0 into D_X. Let P^η be the probability measure on $(D_X,\mathcal{F})$ defined by

$$P^\eta(A) = \mathbf{P}\{\omega : \eta^\eta_\cdot(\omega)\in A\} \tag{2.9}$$

for events $A\in\mathcal{F}$. Let $\mathcal{F}_t=\sigma\{\eta_s : 0\le s\le t\}$ be the σ-algebra generated by the coordinates over time interval $[0,t]$.

Theorem 2.4 *The collection $\{P^\eta : \eta\in X\}$ of probability measures on D_X is a Markov process, in other words*

(a) $P^\eta[\eta_0=\eta]=1$.

(b) *For each $A\in\mathcal{F}$, the function $\eta\mapsto P^\eta(A)$ is measurable.*

(c) $P^\eta[\eta_{t+\cdot}\in A|\mathcal{F}_t]=P^{\eta_t}(A)$ *P^η-almost surely, for every $\eta\in X$ and $A\in\mathcal{F}$.*

Proof. (a) is clear since by definition $\eta^\eta_0(\omega)=\eta$.

(b) Let $\mathcal{L}$ be the class of sets $A\in\mathcal{F}$ for which $\eta\mapsto P^\eta(A)$ is measurable. It is a λ-system. Let $\mathcal{P}$ be the class of finite product sets $A=\{\eta_\cdot\in D_X : \eta_{t_1}\in H_1,\eta_{t_2}\in H_2,\dots,\eta_{t_k}\in H_k\}$ for finite k and Borel subsets $H_1,H_2,\dots,H_k$ of X. $\mathcal{P}$ is closed under intersections, and generates the σ-algebra $\mathcal{F}$ of D_X. By the π-λ-theorem, it suffices to show that $\eta\mapsto P^\eta(A)$ is measurable for $A\in\mathcal{P}$.

The function

$$F(\eta,\omega)=\mathbf{1}_{H_1}(\eta^\eta_{t_1}(\omega))\mathbf{1}_{H_2}(\eta^\eta_{t_2}(\omega))\cdots\mathbf{1}_{H_k}(\eta^\eta_{t_k}(\omega))$$

is jointly measurable, because $(\eta,\omega)\mapsto\eta^\eta_\cdot(\omega)$ is measurable by Lemma 2.3, and the coordinate projections $\eta_\cdot\mapsto\eta_t$ are measurable on D_X. Hence by Fubini's theorem, the function

$$\int_\Omega F(\eta,\omega)\,\mathbf{P}(d\omega)=\int_\Omega \mathbf{1}_{H_1}(\eta^\eta_{t_1}(\omega))\mathbf{1}_{H_2}(\eta^\eta_{t_2}(\omega))\cdots\mathbf{1}_{H_k}(\eta^\eta_{t_k}(\omega))\,\mathbf{P}(d\omega)=P^\eta(A)$$

is a measurable function of η.

(c) Let us write $\eta_{[0,t]}$ to denote the function $\eta_\cdot \in D_X$ restricted to the time interval $[0,t]$. $D_X[0,t]$ is the space of right-continuous functions on $[0,t]$ with left limits, again with the σ-algebra generated by coordinate projections. We need to show that, for any measurable set $B \subseteq D_X[0,t]$,

$$E^\eta[\mathbf{1}_A(\eta_{t+\cdot})\mathbf{1}_B(\eta_{[0,t]})] = E^\eta[P^{\eta_t}(A)\mathbf{1}_B(\eta_{[0,t]})].$$

The argument returns to the construction. Write $G : (\eta,\omega) \mapsto \eta^\eta_\cdot(\omega)$ for the measurable map from $X \times \Omega_0$ into D_X that constructs the process from an initial state η and the Poisson processes $\omega = \{\mathcal{T}_{(x,y)}\}$. Let θ_t denote time shift of the Poisson processes: $\theta_t\omega$ is obtained by restarting ω at time t. To clarify, if $\mathcal{T}_{(x,y)}$ has jump times

$$0 < s_1 < s_2 < \cdots < s_k < s_{k+1} < \cdots ,$$

and $s_{m-1} \le t < s_m$, then $\theta_t\mathcal{T}_{(x,y)}$ has jump times

$$0 < s_m - t < s_{m+1} - t < s_{m+2} - t < \cdots$$

The restriction $\omega_{[0,t]}$ is independent of $\theta_t\omega$, because Poisson processes on disjoint sets are independent.

The evolution $\eta^\eta_{t+\cdot}(\omega)$ from time t onwards can be constructed by evolving the state η^η_t with the restarted Poisson processes $\theta_t\omega$. In other words, $\eta^\eta_{t+\cdot}(\omega) = G(\eta^\eta_t, \theta_t\omega)$. Now

$$\begin{aligned}
E^\eta[\mathbf{1}_A(\eta_{t+\cdot})\mathbf{1}_B(\eta_{[0,t]})] &= \int \mathbf{1}_A(\eta^\eta_{t+\cdot}(\omega))\mathbf{1}_B(\eta^\eta_{[0,t]}(\omega))\,\mathbf{P}(d\omega) \\
&= \int \mathbf{1}_A(G(\eta^\eta_t(\omega),\theta_t\omega))\mathbf{1}_B(\eta^\eta_{[0,t]}(\omega))\,\mathbf{P}(d\omega) \\
&= \int \mathbf{P}(d\omega)\,\mathbf{1}_B(\eta^\eta_{[0,t]}(\omega)) \int \mathbf{P}(d\widetilde{\omega})\,\mathbf{1}_A(G(\eta^\eta_t(\omega),\theta_t\widetilde{\omega})) \\
&= \int \mathbf{P}(d\omega)\,\mathbf{1}_B(\eta^\eta_{[0,t]}(\omega)) \int \mathbf{P}(d\widetilde{\omega})\,\mathbf{1}_A(G(\eta^\eta_t(\omega),\widetilde{\omega})) \\
&= \int \mathbf{P}(d\omega)\,\mathbf{1}_B(\eta^\eta_{[0,t]}(\omega))\,P^{\eta^\eta_t(\omega)}(A) \\
&= E^\eta[\mathbf{1}_B(\eta_{[0,t]})P^{\eta_t}(A)].
\end{aligned}$$

By the independence of $\omega_{[0,t]}$ and $\theta_t\omega$, we integrated separately over them above, and emphasized this by writing $\widetilde{\omega}$ for the second integration variable. Note that the restricted evolution $\eta^\eta_{[0,t]}(\omega)$ depends only on $\omega_{[0,t]}$. And also that $\theta_t\widetilde{\omega}$ has the same distribution as $\widetilde{\omega}$, hence the shift could be dropped in the inside integral. ∎

Let $C(X)$ be the space of continuous functions on X. Since X is compact, all continuous functions on it are bounded.

Lemma 2.5 *The exclusion process is a Feller process, in other words for any $f \in C(X)$ and $t > 0$, $E^\eta[f(\eta_t)]$ is a continuous function of the initial state η.*

Proof. $E^\eta[f(\eta_t)] = \mathbf{E}[f(\eta_t^\eta)]$. By Lemma 2.3 the integrand is a continuous function of η. The expectation is then continuous in η by the bounded convergence theorem. ∎

A consequence of this lemma is that the strong Markov property is valid for the exclusion process. The operators $S(t)$ are defined on the space $C(X)$ by $S(t)f(\eta) = E^\eta[f(\eta_t)]$. The previous lemma guarantees that $S(t)f \in C(X)$ for $f \in C(X)$.

Finally we look at the infinitesimal evolution of the process. For Markov chains, the infinitesimal expected evolution was described by the generator L defined by (1.11), in the sense of the time derivative given in (1.13). Our goal is to find a similar description for the exclusion process.

Suppose f is a *cylinder function* on X. This means that there exists a finite set $A^f = \{x_1, \dots, x_m\} \subseteq S$ and a function $\tilde{f}$ on $\{0,1\}^m$ such that $f(\eta) = \tilde{f}(\eta(x_1), \dots, \eta(x_m))$. Another term for a cylinder function is *local function.* Let $\mathcal{G}_{0,t}$ be the random graph defined by (2.5), and t_0 given by Lemma 2.1. Let C_t be the union of the connected components in $\mathcal{G}_{0,t}$ that intersect A^f. By Lemma 2.1 C_t is almost surely finite for $0 \le t \le t_0$. To compute $f(\eta(t))$ for $0 \le t \le t_0$, only the initial values $\{\eta(x) : x \in C_{t_0}\}$ and the finitely many Poisson jump times in $\cup_{x \in C_{t_0}} \mathcal{T}'_x \cap [0, t_0]$ need to be inspected.

Fix a cube A in S that contains A^f, and is large enough so that the distance from every point of A^f to any point outside A is at least $3R$. Here R is the upper bound on the distance $|x-y|_\infty$ between the endpoints of an edge $\{x,y\} \in \mathcal{E}_{0,t}$, as introduced in the proof of Lemma 2.1. Define the event

$$H_t = \{C_t \subseteq A\}. \tag{2.10}$$

Let N_t be the number of jump times in $\cup_{x \in A} \mathcal{T}'_x \cap [0,t]$.

In the first step, we decompose the probability space Ω four ways into $H_t \cap \{N_t = 0\}$, $H_t \cap \{N_t = 1\}$, $H_t \cap \{N_t > 1\}$, and H_t^c. Then we note that $\{N_t = 0\}$ and $\{N_t = 1\}$ are subsets of H_t, because with only one edge in $\mathcal{G}_{0,t}$ incident to the set A, A^f cannot be connected to anything outside A. Fix $0 \le t \le t_0$ and a starting state η.

$$\begin{aligned} S(t)f(\eta) - f(\eta) &= E^\eta[f(\eta_t) - f(\eta_0)] \\ &= \mathbf{E}[(f(\eta_t^\eta) - f(\eta))\mathbf{1}_{\{N_t=0\}}] + \mathbf{E}[(f(\eta_t^\eta) - f(\eta))\mathbf{1}_{\{N_t=1\}}] \\ &\quad + \mathbf{E}[(f(\eta_t^\eta) - f(\eta))\mathbf{1}_{H_t}\mathbf{1}_{\{N_t>1\}}] + \mathbf{E}[(f(\eta_t^\eta) - f(\eta))\mathbf{1}_{H_t^c}]. \end{aligned} \tag{2.11}$$

The first term after the equality sign is zero because $\eta_t^\eta = \eta$ when no jumps happen in A. On the event $\{N_t = 1\}$, there is a single jump time which can influence $f(\eta_t^\eta) - f(\eta)$. Let

$$\beta = \sum_{x \in A,\, y \in S} p(x,y) + \sum_{x \in S \setminus A,\, y \in A} p(x,y),$$

the finite rate of the Poisson process $\cup_{x\in A}\mathcal{T}'_x$. Then $\mathbf{P}[N_t = 1] = \beta t e^{-\beta t}$. Given that $N_t = 1$, the unique jump time occurred in $\mathcal{T}_{(x,y)}$ with probability $\beta^{-1}p(x,y)$ for $(x,y) \in [A\times S]\cup[(S\setminus A)\times A]$. Given that the unique jump time occurred in $\mathcal{T}_{(x,y)}$,

$$f(\eta_t^\eta) - f(\eta) = \eta(x)(1-\eta(y))[f(\eta^{x,y}) - f(\eta)]$$

because the (x,y)-jump has no effect unless $\eta(x)$ is occupied and $\eta(y)$ vacant. Recall the definition of $\eta^{x,y}$ from (2.6). Thus the second term of (2.11) equals

$$te^{-\beta t}\sum_{x,y\in S} p(x,y)\eta(x)(1-\eta(y))[f(\eta^{x,y}) - f(\eta)].$$

We summed over all pairs (x,y) because $f(\eta^{x,y})-f(\eta)$ is zero for $(x,y)\notin [A\times S]\cup[(S\setminus A)\times A]$.

The last two terms in (2.11) are treated as error terms. The absolute value of the third term is bounded by

$$2\|f\|_\infty \mathbf{P}[N_t > 1] = 2\|f\|_\infty(1 - e^{-\beta t} - \beta t e^{-\beta t}) \le 2\|f\|_\infty \beta^2 t^2.$$

The fourth term we estimate as we did in the proof of Lemma 2.1. On the complement of H_t there must exist an open path in $\mathcal{G}_{0,t}$ of length at least 3 edges, starting from some $x\in A^f$. Recall that k_* is the cardinality of the symmetrized set $B^p\cup(-B^p)$ where $B^p = \{x : p(0,x) > 0\}$ is the finite support of the jump probability. Fix $t_1 \in (0,t_0)$ so that $k_*(1-e^{-2t}) \le 1/2$ for $t \le t_1$. Then for $t\le t_1$, the fourth term in (2.11) is bounded above by

$$\begin{aligned}2\|f\|_\infty \mathbf{P}[C_t \notin A] &\le 2\|f\|_\infty |A^f| \sum_{n\ge 3} k_*^n (1-e^{-2t})^n \le 4\|f\|_\infty |A^f| k_*^3 (1-e^{-2t})^3 \\ &\le 32\|f\|_\infty |A^f| k_*^3 t^3.\end{aligned}$$

Define an operation L on cylinder functions f on X by

$$Lf(\eta) = \sum_{x,y} p(x,y)\eta(x)(1-\eta(y))[f(\eta^{x,y}) - f(\eta)]. \tag{2.12}$$

We call L the *generator* of the exclusion process. The sum has only finitely many nonzero terms, and so $Lf \in C(X)$ for cylinder functions f. We have the bound

$$\|Lf\|_\infty \le 2\|f\|_\infty\Bigg(\sum_{x\in A^f,\, y\in S} p(x,y) + \sum_{x\in S\setminus A^f,\, y\in A^f} p(x,y)\Bigg) \le 4|A^f|\cdot\|f\|_\infty. \tag{2.13}$$

We can summarize the estimation made thus far.

$$\begin{aligned}&\sup_{\eta\in X} |S(t)f(\eta) - f(\eta) - tLf(\eta)| \\ &\qquad \le t(1-e^{-\beta t})\cdot 4|A^f|\cdot\|f\|_\infty + 2\|f\|_\infty \beta^2 t^2 + 32\|f\|_\infty |A^f| k_*^3 t^3 \\ &\qquad \le C(f)t^2,\end{aligned} \tag{2.14}$$

where $C(f)$ is a constant that depends only on f. Two conclusions from this.

Proposition 2.6 (a) *For cylinder functions f, the function $t \mapsto S(t)f(\eta)$ is differentiable at $t = 0$,*

$$\frac{d}{dt}S(t)f(\eta)\bigg|_{t=0} = Lf(\eta),$$

and the limit holds uniformly in η:

$$\lim_{t\to 0}\sup_{\eta\in X}\left|\frac{S(t)f(\eta)-f(\eta)}{t} - Lf(\eta)\right| = 0. \tag{2.15}$$

(b) *For all $f \in C(X)$ we have this uniform continuity at $t = 0$:*

$$\lim_{t\to 0}\sup_{\eta\in X}|S(t)f(\eta) - f(\eta)| = 0. \tag{2.16}$$

Proof. (a) is immediate from (2.14). For cylinder functions f, (b) follows from (2.13) and (2.14). To complete the proof of (b), we argue that cylinder functions are dense in $C(X)$, in other words for any $f \in C(X)$ and $\varepsilon > 0$ there exists a cylinder function g such that $\sup_\eta |f(\eta) - g(\eta)| < \varepsilon$.

As a continuous function on a compact space, f is uniformly continuous. Pick $\delta > 0$ so that $|f(\eta) - f(\zeta)| < \varepsilon$ for all $\eta, \zeta \in X$ such that $d(\eta, \zeta) < \delta$. Fix a finite set $V \subseteq S$ such that $\sum_{x\notin V} 2^{-|x|_\infty} < \delta$. For each η, define $\eta^0 \in X$ by

$$\eta^0(x) = \begin{cases} \eta(x) & \text{for } x \in V, \\ 0 & \text{for } x \notin V. \end{cases}$$

By the choice of V, $d(\eta, \eta^0) < \delta$ for all $\eta \in X$. Now $g(\eta) = f(\eta^0)$ defines a cylinder function that is uniformly within ε of f. ∎

For continuous time, countable state Markov chains, and for the exclusion process, we have naturally arrived at semigroups $S(t)$ of operators on bounded continuous functions, and an infinitesimal description of $S(t)$ given by an operator L. In the next section we turn to study these notions in an abstract setting. This yields results that are applicable to a wide range of Markov processes.

Exercise 2.1 Extend the construction to more general transition probabilities $p(x, y)$. For example, drop the translation invariance assumption. Replace the finite range assumption with a bound on the tail of $p(x, y)$. Chapter I of Liggett's monograph [27] constructs the process for a general countable index set S under the assumption

$$\sup_y \sum_x p(x, y) < \infty.$$

The construction in [27] is based on semigroup theory, and is analytic rather than probabilistic. Can you show that the graphical representation is well-defined in this case?

Exercise 2.2 Show that the event Ω_0 defined above Lemma 2.2 is measurable and has probability 1.

Exercise 2.3 Show by example that Lf is not defined for all continuous functions $f \in C(X)$. Show that as a mapping of functions, L is not continuous even among cylinder functions. For example, it is possible to define a sequence of cylinder functions f_n such that $\|f_n\|_\infty \to 0$, but yet $\|Lf_n\|_\infty \to \infty$.

Exercise 2.4 Derive the generator of the symmetric exclusion process from the stirring particle representation of Section 2.2, repeating the steps that led to (2.14). You should arrive at

$$Lf(\eta) = \sum_{\{x,y\}} p(x,y)[f(\eta^{x,y}) - f(\eta)] \tag{2.17}$$

where the sum is over unordered pairs $\{x,y\}$ of sites. Check that this is the same as L defined by (2.12). In Exercise 4.5 below you use martingales to check that this equality of generators guarantees that the two constructions produce the same Markov process.

Exercise 2.5 Let f be a cylinder function on X. Use (2.14) to show that

$$M_t = f(\eta_t) - \int_0^t Lf(\eta_s)\,ds$$

is a martingale with respect to the filtration $\mathcal{F}_t$.

Here is a way to carry this out. The goal is to show

$$E\left[f(\eta_t) - f(\eta_s) - \int_s^t Lf(\eta_u)\,du \,\middle|\, \mathcal{F}_s \right] = 0.$$

Partition $[s,t]$ into m subintervals $[s_i, s_{i+1}]$ of common length $\delta = s_{i+1} - s_i$. The goal becomes

$$E\left[\sum_i \left\{ E\big[f(\eta_{s_{i+1}})\big|\mathcal{F}_{s_i}\big] - f(\eta_{s_i}) - \int_{s_i}^{s_{i+1}} Lf(\eta_u)\,du \right\} \,\middle|\, \mathcal{F}_s \right] = 0.$$

Rewrite each innermost term as

$$\begin{aligned} &E\big[f(\eta_{s_{i+1}})\big|\mathcal{F}_{s_i}\big] - f(\eta_{s_i}) - \int_{s_i}^{s_{i+1}} Lf(\eta_u)\,du \\ &\quad = E^{\eta_{s_i}}\big[f(\eta_\delta)\big] - f(\eta_{s_i}) - \delta Lf(\eta_{s_i}) - \delta\big(Lf(\eta_{s_{i+1}}) - Lf(\eta_{s_i})\big) \\ &\qquad + \int_{s_i}^{s_{i+1}} \big(Lf(\eta_{s_{i+1}}) - Lf(\eta_u)\big)\,du. \end{aligned}$$

Sum over i. Apply (2.14). Let $\delta \to 0$, and use the right-continuity of the paths η_s in the time variable.

After this exercise the reader will appreciate the ease with which the conclusion follows from semigroup theory (Exercise 3.1 in Section 3.2).

3 Semigroups and generators

We cover here a minimum of semigroup theory. The goal is to have the $\int Lf\,d\nu = 0$ criterion for equilibrium distributions, and to prove some technical lemmas needed in later sections.

3.1 Some generalities about Banach spaces

A *norm* on a real vector space $\mathcal{X}$ is a function $\|\cdot\|$ from $\mathcal{X}$ into nonnegative reals that satisfies these properties, for vectors $f, g \in \mathcal{X}$ and real numbers α:

(a) $\|f + g\| \le \|f\| + \|g\|$ (the triangle inequality)
(b) $\|\alpha f\| = \alpha\|f\|$
(c) $\|f\| > 0$ iff $f \neq 0$.

A vector space with a norm is called a normed vector space. Such a space is a metric space with distance $d(f, g) = \|f - g\|$. A *Banach space* is a normed vector space in which the metric is complete. This means that every Cauchy sequence converges. In other words, if $\{f_n\}$ is a sequence in $\mathcal{X}$ such that

$$\lim_{n\to\infty} \sup_{m>n} \|f_n - f_m\| = 0$$

then there exists a vector $f \in \mathcal{X}$ such that $\|f_n - f\| \to 0$ as $n \to \infty$.

Example 3.1 The space $\mathbf{R}^d$ of real d-vectors $x = (x_1, \dots, x_d)$ is a Banach space with any norm $|x|_p = (|x_1|^p + \cdots + |x_d|^p)^{1/p}$ for $1 \le p < \infty$, and also with the norm $|x|_\infty = \max_i |x_i|$. The case $p = 2$ is the Euclidean norm.

Example 3.2 For any metric space S, the space $C_b(S)$ of bounded continuous functions on S with the supremum norm $\|f\|_\infty = \sup_{x\in S} |f(x)|$ is a Banach space. The completeness of $C_b(S)$ follows from the fact that a uniform limit of continuous functions is continuous. Note that in case S is compact, then all continuous functions are bounded and $C_b(S) = C(S)$, the space of all continuous functions.

The *dual space* $\mathcal{X}^*$ of a Banach space $\mathcal{X}$ is by definition the space of all continuous linear functions from $\mathcal{X}$ into $\mathbf{R}$. (Or into $\mathbf{C}$, if one works with complex scalars.) Elements of $\mathcal{X}^*$ are also called functionals, and the value of $v^* \in \mathcal{X}^*$ applied to $f \in \mathcal{X}$ can be denoted by $v^*(f)$ or $\langle v^*, f\rangle$. $\mathcal{X}^*$ is also a Banach space, with norm

$$\|v^*\| = \sup\{\langle v^*, f\rangle : f \in \mathcal{X}, \|f\| \le 1\}. \tag{3.1}$$

This gives the useful inequality

$$|\langle v^*, f\rangle| \le \|v^*\| \, \|f\|.$$

A fundamental fact is that $\mathcal{X}^*$ is sufficiently rich to separate points on $\mathcal{X}$, namely if $f, g \in \mathcal{X}$ and $\langle v^*, f\rangle = \langle v^*, g\rangle$ for all $v^* \in \mathcal{X}^*$, then necessarily $f = g$.

A *linear operator* on $\mathcal{X}$ is a linear map A whose domain

$$\mathcal{D}(A) = \{f \in \mathcal{X} : \; Af \text{ is defined } \}$$

and range

$$\mathcal{R}(A) = \{Af : f \in \mathcal{D}(A)\}$$

are linear subspaces of $\mathcal{X}$. The graph

$$\mathcal{G}(A) = \{(f, Af) : f \in \mathcal{D}(A)\}$$

is a linear subspace of the product space $\mathcal{X} \times \mathcal{X}$. $\mathcal{X} \times \mathcal{X}$ is also a Banach space with norm $\|(f,g)\| = \|f\| + \|g\|$. A is a *closed* linear operator if $\mathcal{G}(A)$ is a closed subspace of $\mathcal{X} \times \mathcal{X}$. Equivalently, if $f_n \to f$ and $Af_n \to g$, then $f \in \mathcal{D}(A)$ and $g = Af$.

A is a *bounded linear operator* on $\mathcal{X}$ if its domain is all of $\mathcal{X}$, and its operator norm

$$\|A\| = \sup\{\|Af\| : f \in \mathcal{X}, \|f\| \le 1\} \tag{3.2}$$

is finite. As above, then $\|Af\| \le \|A\| \, \|f\|$ for all f. An operator defined on all of $\mathcal{X}$ is bounded iff it is continuous. A is a *contraction* if $\|A\| \le 1$.

Example 3.3 Linear transformations defined on finite-dimensional spaces are given by matrices, and always have finite operator norm and so are bounded linear transformations. Note the different notions of boundedness: a linear map cannot be bounded in the sense that $\sup_f \|Af\|$ is finite, unless it is trivial.

On infinite dimensional spaces it becomes natural to consider unbounded operators that are defined only on a subspace. For example, on $C_b(\mathbf{R})$ differentiation $Df = f'$ is an unbounded linear operator whose domain is the subspace of functions with bounded continuous derivatives. It is a closed operator, because if $f_n \to f$ and $f_n' \to g$ boundedly and uniformly, then passing to the limit in $f_n(x) - f_n(0) = \int_0^x f_n'(y)\, dy$ shows that $f' = g$.

An example of a bounded linear operator on $C_b(\mathbf{R})$ is

$$Af(x) = \int_{\mathbf{R}} p(t,x,y) f(y) dy$$

where $p(t,x,y) = (2\pi t)^{-1/2} \exp(-(x-y)^2/2t)$ is the transition probability function for Brownian motion, or the Gaussian kernel. Since $\int p(t,x,y)dy = 1$, A is a contraction.

Parts of calculus work fine for Banach space valued functions. Let $u : [a, b] \to \mathcal{X}$. Say u is (Riemann) integrable if the limit

$$\lim_{\|\Delta\| \to 0} \sum_{i=1}^{n} u(s_i)(t_i - t_{i-1}) \tag{3.3}$$

exists, where

$$\Delta = \{a = t_0 < t_1 < \cdots < t_n = b\}$$

is a partition of $[a, b]$, $\|\Delta\| = \max(t_i - t_{i-1})$ is the mesh of the partition, and the $\{s_i\}$ are arbitrary points such that $s_i \in [t_{i-1}, t_i]$ for each i. The limit is denoted by $\int_a^b u(t)dt$. Differentiability has to be interpreted in the Banach space norm. To say that $u'(t) = g$ for some $g \in \mathcal{X}$ means that

$$\lim_{h \to 0} \left\| \frac{u(t+h) - u(t)}{h} - g \right\| = 0.$$

At the endpoints $u'(a)$ and $u'(b)$ are defined as one-sided derivatives in the obvious way. Let us say the function $u(t)$ is continuously differentiable on $[a, b]$ if the derivative $u'(t)$ exists and is itself continuous as an $\mathcal{X}$-valued function from $[a, b]$ into $\mathcal{X}$.

Lemma 3.4 (a) *The integral $\int_a^b u(t)dt$ exists for every continuous function $u : [a, b] \to \mathcal{X}$. It satisfies*

$$\left\| \int_a^b u(t)dt \right\| \le \int_a^b \|u(t)\| \, dt. \tag{3.4}$$

(b) *Suppose K is a closed linear operator on $\mathcal{X}$ with domain $\mathcal{D}(K)$. Let $u : [a, b] \to \mathcal{X}$. Assume that u actually maps $[a, b]$ into $\mathcal{D}(K)$, and that both $u(t)$ and $Ku(t)$ are continuous functions of t. Then $\int_a^b u(t)dt$ also lies in $\mathcal{D}(K)$, and*

$$K \int_a^b u(t)dt = \int_a^b Ku(t)dt. \tag{3.5}$$

(c) *Suppose $u : [a, b] \to \mathcal{X}$ is continuous and continuously differentiable on $[a, b]$. Then*

$$\int_a^b u'(t)dt = u(b) - u(a). \tag{3.6}$$

Proof. (a) Existence of the integral can be proved as in calculus. A continuous function u defined on a compact interval $[a, b]$ is uniformly continuous. So given $\varepsilon > 0$, we may choose an integer $M > 0$ so that $\|u(s) - u(t)\| < \varepsilon/(b - a)$ whenever $|s - t| \le \delta = (b - a)/M$. Let

$\{t_i\}$ be an arbitrary partition with mesh $\|\Delta\| \le \delta$ and $\{s_i\}$ points chosen from the partition intervals. We shall compare the Riemann sum

$$S = \sum_i u(s_i)(t_i - t_{i-1})$$

to the sum

$$R = \sum_k u(r_k)\delta$$

formed with the partition $\{r_k = k(b-a)/M\}_{0\le k\le M}$. Let $\{t'_j\}$ be the common refinement of the partitions $\{t_i\}$ and $\{r_k\}$, in other words, as point sets $\{t'_j\} = \{t_i\} \cup \{r_k\}$. Let s'_j equal s_i for that i which satisfies $[t_{i-1}, t_i] \supseteq [t'_{j-1}, t'_j]$. Then

$$S = \sum_j u(s'_j)(t'_j - t'_{j-1}).$$

The point s'_j may lie outside $[t'_{j-1}, t'_j]$, but it is within δ of t'_j. Hence S is within ε of the sum

$$S' = \sum_j u(t'_j)(t'_j - t'_{j-1}).$$

For each j, choose k so that $[t'_{j-1}, t'_j] \subseteq [r_{k-1}, r_k]$. Then t'_j is within δ of r_k, and it follows that S' is within ε of R.

To summarize, for any partition Δ with $\|\Delta\| \le \delta$, the Riemann sum S is within 2ε of the (fixed) Riemann sum R. Consequently any two Riemann sums from partitions with mesh at most δ differ by at most 4ε. Thus the upper and lower limits in (3.3) differ by at most 4ε. Since $\varepsilon > 0$ was arbitrary, the upper and lower limits in (3.3) must actually coincide, and so we have proved that the limit exists.

To show (3.4): by the triangle inequality

$$\left\| \sum_i u(s_i)(t_i - t_{i-1}) \right\| \le \sum_i \|u(s_i)\| (t_i - t_{i-1}).$$

As $\|\Delta\| \to 0$, this inequality turns into (3.4).

(b) Fix a sequence of partitions $\Delta^{(n)} = \{t_i\}$ whose mesh tends to zero as $n \to \infty$, and fix points $\{s_i\}$ from the partition intervals. Let $f_n = \sum_i u(s_i)(t_{i-1} - t_i)$ be the Riemann sum for partition $\Delta^{(n)}$. By part (a), $f_n \to f = \int_a^b u(t)dt$ as $n \to \infty$. By (a) again and the assumption that $t \mapsto Ku(t)$ is continuous, the integral $g = \int_a^b Ku(t)dt$ exists and equals the limit

$$g = \lim_{n\to\infty} \sum_i Ku(s_i)(t_{i-1} - t_i) = \lim_{n\to\infty} K \sum_i u(s_i)(t_{i-1} - t_i) = \lim_{n\to\infty} Kf_n$$

where we used the linearity of K.

To summarize, $(f_n, Kf_n) \to (f, g)$. By the closedness of K, $g = Kf$ which is the conclusion (3.5).

(c) For this part we rely on the fact that the dual of a Banach space separates points. Fix $v^* \in \mathcal{X}^*$, and define a real-valued function $\varphi(t) = \langle v^*, u(t)\rangle$. First observe that $\varphi'(t) = \langle v^*, u'(t)\rangle$, for

$$\begin{aligned} & \limsup_{h\to\infty} \left| \frac{\varphi(t+h) - \varphi(t)}{h} - \langle v^*, u'(t)\rangle \right| \\ = \; & \limsup_{h\to\infty} \left| \left\langle v^*, \frac{u(t+h) - u(t)}{h} - u'(t) \right\rangle \right| \\ \leq \; & \limsup_{h\to\infty} \|v^*\| \cdot \left\| \frac{u(t+h) - u(t)}{h} - u'(t) \right\| = 0 \end{aligned}$$

by the assumption of differentiability of u. As a composite of two continuous functions, namely $u'(t)$ and v^*, $\varphi'(t)$ is continuous. Thus the fundamental theorem of calculus gives

$$\int_a^b \langle v^*, u'(t)\rangle dt = \int_a^b \varphi'(t) dt = \varphi(b) - \varphi(a) = \langle v^*, u(b) - u(a)\rangle.$$

Applying part (b) to the first member above gives

$$\left\langle v^*, \int_a^b u'(t) dt \right\rangle = \langle v^*, u(b) - u(a)\rangle.$$

The integral $\int_a^b u'(t)dt$ exists because we assumed that $u'(t)$ is continuous on $[a, b]$. Since the equality above holds for all $v^* \in \mathcal{X}^*$, (3.6) follows. ∎

3.2 The generator of a semigroup

Let $S(t)$ be a bounded linear operator on $\mathcal{X}$ for each $t \geq 0$. $\{S(t)\}$ is a semigroup if $S(0) = I$ and $S(s + t) = S(s)S(t)$. $\{S(t)\}$ is a strongly continuous semigroup if $\|S(t)f - f\| \to 0$ as $t \to 0$ for every $f \in \mathcal{X}$. If each $S(t)$ is a contraction, then $\{S(t)\}$ is a contraction semigroup.

Lemma 3.5 *Suppose $\{S(t)\}$ is a strongly continuous contraction semigroup on $\mathcal{X}$. Then for every $f \in \mathcal{X}$, $S(t)f$ is a uniformly continuous function of $t \in [0, \infty)$ into $\mathcal{X}$.*

Proof. For $t, h \geq 0$,

$$\|S(t + h)f - S(t)f\| = \|S(t)(S(h)f - f)\| \leq \|S(h)f - f\|$$

and for $0 \le h \le t$,

$$\|S(t-h)f - S(t)f\| = \|S(t-h)(S(h)f - f)\| \le \|S(h)f - f\|.$$

In both cases the right-hand side vanishes as $h \to 0$, and the bounds are uniform in t. ∎

The *generator* (also called the *infinitesimal generator*) of a semigroup $\{S(t)\}$ is the operator L defined by

$$Lf = \lim_{t \to 0} \frac{S(t)f - f}{t} \tag{3.7}$$

with domain $\mathcal{D}(L)$ consisting of those $f \in \mathcal{X}$ for which this limit exists (convergence has to be in the norm of $\mathcal{X}$).

Example 3.6 In Sections 1.2 and 2 we constructed the semigroups for continuous-time Markov chains on countable state spaces, and for the finite range exclusion process. We checked that these semigroups were strongly continuous. They are contraction semigroups by virtue of their definition in terms of integration against probability measures.

For a Markov chain on a countable state space, the generator is a bounded operator given by (1.11) and its domain is the entire space $C_b(S)$.

For the exclusion process the generator is the unbounded operator given by (2.12), and its domain is a subspace of $C(X)$. By (2.15) the domain $\mathcal{D}(L)$ contains the cylinder functions.

Lemma 3.7 *Suppose $\{S(t)\}$ is a strongly continuous contraction semigroup on $\mathcal{X}$ with generator L.*

(a) *For all $f \in \mathcal{X}$ and $t > 0$, $\int_0^t S(s)f\,ds \in \mathcal{D}(L)$ and*

$$S(t)f - f = L\int_0^t S(s)f\,ds. \tag{3.8}$$

(b) *For all $f \in \mathcal{D}(L)$ and $t \ge 0$, $S(t)f \in \mathcal{D}(L)$ and*

$$\frac{d}{dt}S(t)f = LS(t)f = S(t)Lf. \tag{3.9}$$

(c) *For all $f \in \mathcal{D}(L)$ and $t \ge 0$,*

$$S(t)f - f = \int_0^t LS(s)f\,ds = \int_0^t S(s)Lf\,ds. \tag{3.10}$$

Proof. (a) Note that a bounded linear operator is automatically closed. So applying (3.5) and easily verified additivity properties of the Banach space valued integral, we get

$$\begin{aligned}
&\frac{S(h)-I}{h}\int_0^t S(s)f\,ds = \frac{1}{h}\int_0^t S(s+h)f\,ds - \frac{1}{h}\int_0^t S(s)f\,ds \\
= \;& \frac{1}{h}\int_h^{t+h} S(s)f\,ds - \frac{1}{h}\int_0^t S(s)f\,ds = \frac{1}{h}\int_t^{t+h} S(s)f\,ds - \frac{1}{h}\int_0^h S(s)f\,ds \\
&\longrightarrow S(t)f - f \qquad \text{as } h \searrow 0,
\end{aligned}$$

by Lemma 3.5. This checks that $\int_0^t S(s)f\,ds \in \mathcal{D}(L)$ and proves (3.8).

(b) Fix $t \geq 0$ and let $h > 0$. Algebraic manipulation gives

$$\frac{S(t+h)f - S(t)f}{h} = \frac{S(h)-I}{h}S(t)f = S(t)\frac{S(h)-I}{h}f. \tag{3.11}$$

Let $h \searrow 0$. By assumption $h^{-1}(S(h)-I)f \to Lf$. Since $S(t)$ is a continuous map on $\mathcal{X}$, the last term of (3.11) converges to $S(t)Lf$. This forces the middle term to converge too, which implies that $S(t)f \in \mathcal{D}(L)$ and $LS(t)f = S(t)Lf$. Convergence of the leftmost member then says that $S(t)f$ is differentiable from the right, and the derivative is the one given in (3.9).

It remains to check differentiability from the left. Let still $h > 0$.

$$\begin{aligned}
&\frac{S(t-h)f - S(t)f}{-h} - S(t)Lf \\
= \;& S(t-h)\left(\frac{S(h)f - f}{h} - Lf\right) + S(t-h)Lf - S(t)Lf,
\end{aligned}$$

from which, using contractivity,

$$\begin{aligned}
&\left\|\frac{S(t-h)f - S(t)f}{-h} - S(t)Lf\right\| \\
\leq \;& \left\|\frac{S(h)f - f}{h} - Lf\right\| + \|S(t-h)Lf - S(t)Lf\|.
\end{aligned}$$

The last line vanishes as $h \searrow 0$, the first term by the assumption $f \in \mathcal{D}(L)$, the second by Lemma 3.5. This proves differentiability of $S(t)f$ from the left.

(c) $S(t)Lf$ is a continuous function of t, and so this follows from (3.6). ∎

Here is a probabilistic application of the foregoing.

Exercise 3.1 Suppose X_t is a Feller continuous Markov process on a metric space Y, and suppose $S(t)f(x) = E^x[f(X_t)]$ is a strongly continuous semigroup on $C_b(Y)$. Let L be the generator and $f \in \mathcal{D}(L)$. Show that

$$M_t = f(X_t) - f(X_0) - \int_0^t Lf(X_s)\,ds$$

is a right-continuous mean zero martingale. See Exercise A.2 for the measurability of the integral term on the path space.

Corollary 3.8 *If L is the generator of a strongly continuous contraction semigroup on $\mathcal{X}$, then $\mathcal{D}(L)$ is dense in $\mathcal{X}$ and L is a closed operator.*

Proof. Let $f \in \mathcal{X}$. By Lemma 3.7(a), $t^{-1}\int_0^t S(s)f\,ds$ lies in $\mathcal{D}(L)$ for each $t > 0$. By the strong continuity of the semigroup $t^{-1}\int_0^t S(s)f\,ds \to f$ as $t \searrow 0$, and consequently $\mathcal{D}(L)$ is dense in $\mathcal{X}$.

To show that L is a closed operator, suppose $(f_j, Lf_j) \to (f, g)$ in $\mathcal{X} \times \mathcal{X}$ for some sequence $\{f_j\}$ of elements of $\mathcal{D}(L)$. By (3.10)

$$S(t)f_j - f_j = \int_0^t S(s)Lf_j\,ds.$$

Let $j \to \infty$. Note that by (3.4) and the contraction property,

$$\left\| \int_0^t S(s)Lf_j\,ds - \int_0^t S(s)g\,ds \right\| \le \int_0^t \|S(s)(Lf_j - g)\|\;ds \le t\,\|Lf_j - g\|\,.$$

Thus in the limit we obtain

$$S(t)f - f = \int_0^t S(s)g\,ds.$$

which implies that $f \in \mathcal{D}(L)$ and $Lf = g$. ∎

The above were basic properties of the generator. Next a special formula that we need in Section 5.3.

Proposition 3.9 *Let $\{S(t)\}$ and $\{T(t)\}$ be strongly continuous contraction semigroups on $\mathcal{X}$ with generators L and M, respectively. Assume that L and M are bounded operators, so that in particular, their domains are the whole space $\mathcal{X}$ and they are continuous mappings. Then for any $f \in \mathcal{X}$,*

$$S(t)f - T(t)f = \int_0^t T(t-r)(L-M)S(r)f\,dr. \tag{3.12}$$

Proof. Note first that, by the continuity of the semigroups (Lemma 3.5) and the assumed continuity of L and M, the integral is well-defined by Lemma 3.4(a).

We want to find $(d/ds)T(t-s)S(s)f$ for $s \in (0,t)$. First we differentiate from the right. Let $h > 0$.

$$\begin{aligned}
& \frac{1}{h}\left\{T(t-s-h)S(s+h)f - T(t-s)S(s)f\right\} \\
= \; & T(t-s-h)\frac{S(s+h)f - S(s)f}{h} + \frac{T(t-s-h) - T(t-s)}{h}S(s)f \\
= \; & T(t-s-h)\frac{1}{h}\int_0^h S(r)LS(s)f\,dr + \frac{T(t-s-h) - T(t-s)}{h}S(s)f \\
= \; & T(t-s-h)LS(s)f - \frac{T(t-s-h) - T(t-s)}{-h}S(s)f \\
& \qquad + T(t-s-h)\frac{1}{h}\int_0^h \left(S(r)LS(s)f - LS(s)f\right)dr
\end{aligned}$$

Let $h \searrow 0$, and consider the three last terms. The first one converges to $T(t-s)LS(s)f$ by the continuity of the semigroup. The second one converges to $T(t-s)MS(s)f$ by (3.9), since by assumption all elements of $\mathcal{X}$ are in $\mathcal{D}(M)$. By contractivity and (3.4), the norm of the last term is bounded above by

$$\frac{1}{h}\int_0^h \|S(r)LS(s)f - LS(s)f\|\,dr$$

which vanishes as $h \searrow 0$ by the continuity of the semigroup.

We have differentiated from the right and obtained $(d/ds+)T(t-s)S(s)f = T(t-s)(L-M)S(s)f$. We leave the similar calculation for the left derivative as an exercise, and consider

$$\frac{d}{ds}T(t-s)S(s)f = T(t-s)(L-M)S(s)f \tag{3.13}$$

proved.

Finally we check that this derivative is a continuous function of s:

$$\begin{aligned}
& \|T(t-s-h)(L-M)S(s+h)f - T(t-s)(L-M)S(s)f\| \\
\leq \; & \|T(t-s-h)(L-M)(S(s+h)f - S(s)f)\| + \|[T(t-s-h) - T(t-s)](L-M)S(s)f\| \\
\leq \; & \|(L-M)(S(s+h)f - S(s)f)\| + \|[T(t-s-h) - T(t-s)](L-M)S(s)f\|
\end{aligned}$$

where in the last step we used contractivity. Both terms vanish as $h \to 0$, by the continuity of the semigroups. The first term also needs the continuity of the mapping $L - M$. Now (3.12) follows from (3.6). ∎

3.3 The resolvent and cores

Recall that for a matrix A, a scalar λ is an eigenvalue iff the matrix $\lambda - A$ is singular. (When we mean a matrix or an operator, a scalar λ stands for λI.) The set of eigenvalues is the *spectrum* $\sigma(A)$ of the matrix. The *resolvent set* $\rho(A)$ is the set of λ for which $\lambda - A$ is invertible. The resolvent set is an important concept also in operator theory on infinite dimensional spaces. The complement of the resolvent set is still called the spectrum, but eigenvalues alone do not account for the entire spectrum.

For any closed linear operator L on $\mathcal{X}$, we say that $\lambda \in \rho(L)$ if $\lambda - L$ is a one-to-one map on $\mathcal{D}(L)$, its range $\mathcal{R}(\lambda - L)$ is the entire space $\mathcal{X}$, and the inverse operator $(\lambda - L)^{-1}$ is a bounded linear operator on $\mathcal{X}$. The bounded operator $R_\lambda = (\lambda - L)^{-1}$ is called the *resolvent* of L.

Proposition 3.10 *Let L be the generator of a strongly continuous contraction semigroup $\{S(t)\}$ on $\mathcal{X}$. Then $(0, \infty) \subseteq \rho(L)$, and for all $f \in \mathcal{X}$ and $\lambda > 0$,*

$$(\lambda - L)^{-1} f = \int_0^\infty e^{-\lambda t} S(t) f \, dt. \tag{3.14}$$

Remark. To make sense of the right-hand side of (3.14), extend the Banach space valued integral to infinite intervals by

$$\int_0^\infty u(t)\, dt = \lim_{b \to \infty} \int_0^b u(t)\, dt \tag{3.15}$$

provided the limit exists. To see that the required limit in (3.14) exists, note that for $0 < a < b$, by the contraction property,

$$\left\| \int_0^a e^{-\lambda t} S(t) f \, dt - \int_0^b e^{-\lambda t} S(t) f \, dt \right\| \le \int_a^b e^{-\lambda t} \|S(t) f\| \, dt \le \|f\| \frac{e^{-\lambda a}}{\lambda}$$

which vanishes as $a \nearrow \infty$. Thus by the completeness of $\mathcal{X}$, the limit $\lim_{b\to\infty} \int_0^b e^{-\lambda t} S(t) f \, dt$ exists. Lemma 3.4(a)–(b) extend to integrals over an infinite interval.

Proof of Proposition 3.10. Define the operator B_λ on $\mathcal{X}$ by

$$B_\lambda f = \int_0^\infty e^{-\lambda t} S(t) f \, dt.$$

B is well-defined for all $f \in \mathcal{X}$ by the remark above, and by the linearity of the integral it is a linear operator. Since

$$\|B_\lambda f\| \le \int_0^\infty e^{-\lambda t} \|S(t) f\| \, dt \le \lambda^{-1} \|f\|, \tag{3.16}$$

we see that B_λ is a bounded linear operator on $\mathcal{X}$.

Using (3.5) extended to the infinite interval, and linearity, we can see that $B_\lambda f \in \mathcal{D}(L)$:

$$\begin{aligned}\frac{S(h)-I}{h} B_\lambda f &= \frac{1}{h}\int_0^\infty e^{-\lambda t}[S(t+h)f - S(t)f]\,dt \\ &= \frac{e^{\lambda h}-1}{h}\int_0^\infty e^{-\lambda t} S(t)f\,dt - e^{\lambda h}\,\frac{1}{h}\int_0^h e^{-\lambda t}S(t)f\,dt \\ &\longrightarrow \lambda B_\lambda f - f \qquad \text{as } h \to 0.\end{aligned}$$

This gives $LB_\lambda f = \lambda B_\lambda f - f$, or equivalently

$$(\lambda - L)B_\lambda f = f \qquad \text{for all } f \in \mathcal{X}. \tag{3.17}$$

This says that B_λ is a right inverse for $\lambda - L$. To get the other half, take $f \in \mathcal{D}(L)$. By (3.9) L commutes with the semigroup, so applying (3.5) again, we see that B_λ and L commute:

$$\begin{aligned}B_\lambda Lf &= \int_0^\infty e^{-\lambda t}S(t)Lf\,dt = \int_0^\infty L[e^{-\lambda t}S(t)f]\,dt \\ &= L\int_0^\infty e^{-\lambda t}S(t)f\,dt = LB_\lambda f.\end{aligned}$$

Thus (3.17) gives also

$$B_\lambda(\lambda - L)f = f \qquad \text{for all } f \in \mathcal{X}. \tag{3.18}$$

It remains to oberve that we have checked everything. (3.17) implies that $\mathcal{R}(\lambda - L) = \mathcal{X}$, (3.18) that $\lambda - L$ is one-to-one, and together that $(\lambda - L)^{-1} = B_\lambda$, a bounded operator. ∎

Corollary 3.11 *The generator L of a strongly continuous contraction semigroup has the following property called dissipativity: for all $f \in \mathcal{D}(L)$ and $\lambda > 0$,*

$$\|\lambda f - Lf\| \geq \lambda \|f\|. \tag{3.19}$$

Proof. By (3.16), $\|(\lambda - L)^{-1}g\| \leq \lambda^{-1}\|g\|$ for all $g \in \mathcal{X}$. Take $g = (\lambda - L)f$. ∎

The definition of the domain $\mathcal{D}(L)$ in terms of the existence of the limit in (3.7) is not very useful, because for complicated Markov processes verification of this limit is not easy. Fortunately the precise domain need not be often known. Instead, it is sufficient to identify a suitable smaller subspace of $\mathcal{X}$ which carries all the relevant information. For a closed linear operator L, a linear subspace $\mathcal{Y}$ of $\mathcal{D}(L)$ is a *core* if the graph of L is the closure of the graph of L restricted to $\mathcal{Y}$. We express this by saying that L is the closure of its restriction to $\mathcal{Y}$. Explicitly, the requirement is that for each $f \in \mathcal{D}(L)$ there exists a sequence $g_n \in \mathcal{Y}$ such that $g_n \to f$ and $Lg_n \to Lf$.

Proposition 3.12 *Let L be the generator of a strongly continuous contraction semigroup $\{S(t)\}$ on $\mathcal{X}$. Let $\mathcal{Y}_0$ and $\mathcal{Y}_1$ be dense subspaces of $\mathcal{X}$ such that $\mathcal{Y}_0 \subseteq \mathcal{Y}_1 \subseteq \mathcal{D}(L)$, and $S(t)f \in \mathcal{Y}_1$ for each $f \in \mathcal{Y}_0$ and $t \geq 0$. Then $\mathcal{Y}_1$ is a core for L.*

Proof. Fix $\lambda > 0$. First we show that $\{(\lambda - L)f : f \in \mathcal{Y}_1\}$ is dense in $\mathcal{X}$, by showing that each $f \in \mathcal{Y}_0$ is the limit of $(\lambda - L)f_n$ for a sequence $\{f_n\} \subseteq \mathcal{Y}_1$. So fix $f \in \mathcal{Y}_0$, and define

$$f_n = \frac{1}{n} \sum_{k=0}^{n^2-1} e^{-\lambda k/n} S(k/n) f \in \mathcal{Y}_1.$$

We leave it to the reader to check that for any $g \in \mathcal{X}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n^2-1} e^{-\lambda k/n} S(k/n) g = \int_0^\infty e^{-\lambda t} S(t) g \, dt = (\lambda - L)^{-1} g. \tag{3.20}$$

Apply this to $g = (\lambda - L)f$ to get

$$\lim_{n\to\infty} (\lambda - L) f_n = (\lambda - L)^{-1} (\lambda - L) f = f.$$

We have shown that $\{(\lambda - L)f : f \in \mathcal{Y}_1\}$ is dense in $\mathcal{X}$.

Next we show that L is the closure of its restriction to $\mathcal{Y}_1$. Let $f \in \mathcal{D}(L)$. Let $g = \lambda f - Lf$. Find $f_n \in \mathcal{Y}_1$ such that $(\lambda - L)f_n \to g$. Since it converges, the sequence $\{(\lambda - L)f_n\}$ must be a Cauchy sequence. By dissipativity

$$\|f_n - f_m\| \leq \lambda^{-1} \|(\lambda - L)f_n - (\lambda - L)f_m\|,$$

and so $\{f_n\}$ is also a Cauchy sequence in $\mathcal{X}$. By completeness, f_n converges to some $h \in \mathcal{X}$. And then

$$Lf_n = \lambda f_n - (\lambda - L)f_n \to \lambda h - g.$$

Now we have $(f_n, Lf_n) \to (h, \lambda h - g)$, so by the closedness of L, $h \in \mathcal{D}(L)$ and $Lh = \lambda h - g = \lambda h - \lambda f + Lf$. The latter gives $(\lambda - L)f = (\lambda - L)h$. But by dissipativity $\lambda - L$ is one-to-one, and so we conclude $h = f$. To summarize, we have shown $(f_n, Lf_n) \to (f, Lf)$. Since $f \in \mathcal{D}(L)$ was arbitrary and $\{f_n\} \subseteq \mathcal{Y}_1$, we have shown that L is the closure of its restriction to $\mathcal{Y}_1$, and thereby that $\mathcal{Y}_1$ is a core for L. ∎

We present here a lemma useful for uniqueness questions, and whose proof utilizes the resolvent. If L and M are two operators on $\mathcal{X}$, we say that M is an extension of L if the graph of L is a subset of the graph of M. In other words, $\mathcal{D}(L) \subseteq \mathcal{D}(M)$ and $M = L$ on $\mathcal{D}(L)$.

Proposition 3.13 *Suppose L and M are generators of strongly continuous contraction semigroups $S(t)$ and $T(t)$, respectively, and M is an extension of L. Then $M = L$ and $S(t) = T(t)$ for all $t \geq 0$.*

Proof. To show $M = L$, it suffices to show that $\mathcal{D}(M) \subseteq \mathcal{D}(L)$. Let $f \in \mathcal{D}(M)$. Since the range of $\lambda - L$ is the entire space $\mathcal{X}$, we can find $g \in \mathcal{D}(L)$ such that $(\lambda - M)f = (\lambda - L)g$. But since M extends L, $Mg = Lg$, and we get $(\lambda - M)f = (\lambda - M)g$. By dissipativity $\lambda - M$ is one-to-one, and so $f = g$. This implies that $f \in \mathcal{D}(L)$.

Since $M = L$, also $(\lambda - M)^{-1} = (\lambda - L)^{-1}$, and we get for all $\lambda > 0$ and $f \in \mathcal{X}$ that

$$\int_0^\infty e^{-\lambda t} T(t) f \, dt = (\lambda - M)^{-1} f = (\lambda - L)^{-1} f = \int_0^\infty e^{-\lambda t} S(t) f \, dt.$$

Consequently, for any bounded linear functional $v^* \in \mathcal{X}^*$,

$$\int_0^\infty e^{-\lambda t} \langle v^*, T(t)f - S(t)f \rangle \, dt = 0$$

for all $\lambda > 0$. By Lemma A.19 in the appendix, a bounded continuous function is uniquely determined by its Laplace transform, hence it follows that $\langle v^*, T(t)f - S(t)f \rangle = 0$ for all t. Since v^* is arbitrary, we can conclude that $T(t)f = S(t)f$ for all t. ∎

Exercise 3.2 Provide the argument for the left derivative for (3.13).

Exercise 3.3 Extend Lemma 3.4(a)–(b) to integrals over an infinite interval.

Exercise 3.4 Prove (3.20).

Exercise 3.5 To develop intuition for the semigroup material, here are deterministic examples simple enough so that everything can be explicitly calculated.

(a) Fix $a \in \mathbf{R}$, and consider the ordinary differential equation $x'(t) = a$ on $\mathbf{R}$. Its solutions are $x(t) = x(0) + at$. The natural semigroup on functions is

$$S(t)f(x) = f(x + at).$$

On what Banach space is this semigroup strongly continuous? The generator is $Lf = af'$. Given a suitable function g and $\lambda > 0$, consider the linear o.d.e.

$$-af' + \lambda f = g.$$

There is no initial condition, but you can show that there is a unique bounded solution, given by

$$f(x) = \int_0^\infty e^{-\lambda t} g(x + at)\, dt.$$

You have discovered the resolvent formula $(\lambda - L)^{-1} g = \int_0^\infty e^{-\lambda t} S(t) g\, dt$.

(b) An easier case to consider is $S(t)x = xe^{-at}$ for $x \in \mathbf{R}$, where $a > 0$ is fixed.

Notes

Most of this section is from Ethier and Kurtz [13], Chapter 1. We omitted the main result of semigroup theory, namely the Hille-Yosida theorem, because we have no need for it.

4 Applications of semigroups

4.1 Invariant probability measures

4.1.1 The general situation

Let $\{P^x\}$ be a Feller continuous Markov process as defined in Section 1.3. Assume $S(t)f(x) = E^x[f(X_t)]$ defines a strongly continuous, contraction semigroup $\{S(t)\}$ on the Banach space $C_b(Y)$. The space $\mathcal{M}_1(Y)$ of probability measures on Y is also a metric space under the Prohorov metric. Convergence in $\mathcal{M}_1(Y)$ is the familiar notion of weak convergence of probability distributions, characterized by

$$\mu_n \to \mu \quad \text{iff} \quad \int f\,d\mu_n \to \int f\,d\mu \quad \text{for all } f \in C_b(Y). \tag{4.1}$$

The semigroup $S(t)$ acts naturally on the space $\mathcal{M}_1(Y)$, in a way that is dual to the action on functions. For $\mu \in \mathcal{M}_1(Y)$ and $t \geq 0$, define $\mu S(t) \in \mathcal{M}_1(Y)$ by

$$\int f\,d[\mu S(t)] = \int S(t)f\,d\mu \quad \text{for all } f \in C_b(Y). \tag{4.2}$$

A little more explicitly, for Borel subsets B of Y,

$$\mu S(t)(B) = \int P^x[X_t \in B]\mu(dx). \tag{4.3}$$

Probabilistically speaking, $\mu S(t)$ is the probability distribution of X_t, when the initial distribution of the process is μ.

We say that μ is *invariant* for the process X_t if

$$\mu S(t) = \mu \qquad \text{for all } t \geq 0. \tag{4.4}$$

We write $\mathcal{I}$ for the set of invariant probability measures when it is clear from the context which process is under discussion. Alternative terms for invariant probability measures are *invariant distributions* and *equilibrium distributions*.

Invariance implies that if the initial state X_0 has probability distribution μ, then so does X_t at all later times $t \geq 0$. And furthermore, the process $(X_t)_{0\leq t<\infty}$ is *stationary*, which means that the distribution of the shifted process $(X_{s+t})_{0\leq t<\infty}$ is the same as the distribution of the original process. Invariant measures are a key component in a description of the long term behavior of a Markov process. This we know well from the theory of discrete-time Markov chains.

We want a convenient way of checking whether a given measure is invariant. For a discrete-time Markov chain on a countable state space S the requirement (4.4) reduces to a single equation

$$\mu\{y\} = \sum_{x \in S} \mu\{x\} p(x, y) \qquad \text{for all } y \in S, \tag{4.5}$$

or $\mu = \mu P$, if we think of μ as a row vector, $P = (p(x, y))_{x,y \in S}$ is the transition matrix, and μP is matrix multiplication. For a continuous-time process it would not suffice to check (4.4) for any finite set of time points. But we can use the generator to obtain a single equation.

Theorem 4.1 *Let L be the generator of the strongly continuous contraction semigroup $S(t)$ on $C_b(Y)$ defined by a Markov process X_t. Let μ be a probability measure on Y. Let $\mathcal{Y}$ be any core for L. Then μ is invariant for X_t iff*

$$\int Lf \, d\mu = 0 \qquad \textit{for all } f \in \mathcal{Y}. \tag{4.6}$$

One possible choice for the core $\mathcal{Y}$ is of course the domain $\mathcal{D}(L)$ itself.

Proof. Suppose first that μ is invariant and $f \in \mathcal{D}(L)$. Since $t^{-1}(S(t)f - f) \to Lf$ boundedly and uniformly as $t \to 0$ [this is what convergence in $C_b(Y)$ means], we can take the limit of integrals and get

$$\begin{aligned} \int Lf \, d\mu &= \lim_{t \to 0} \int \frac{S(t)f - f}{t} \, d\mu = \lim_{t \to 0} \frac{1}{t} \left\{ \int f \, d[S(t)\mu] - \int f \, d\mu \right\} \\ &= \lim_{t \to 0} \frac{1}{t} \left\{ \int f \, d\mu - \int f \, d\mu \right\} = 0. \end{aligned}$$

Conversely, assume $\int Lg \, d\mu = 0$ for all g in a core $\mathcal{Y}$. By the definition of a core, for any $f \in \mathcal{D}(L)$ there are $g_n \in \mathcal{Y}$ such that $Lg_n \to Lf$ boundedly and uniformly. Consequently $\int Lf \, d\mu = 0$ for all $f \in \mathcal{D}(L)$.

Fix $f \in \mathcal{D}(L)$. By Lemma 3.7(b), $S(t)f \in \mathcal{D}(L)$ for all $t \geq 0$. Integrate the equality $S(t)f - f = \int_0^t LS(s)f \, ds$ against μ and use Fubini's theorem to get

$$\int S(t)f \, d\mu - \int f \, d\mu = \int_0^t \left\{ \int L[S(s)f] \, d\mu \right\} ds = 0. \tag{4.7}$$

Consequently

$$\int f \, d[\mu S(t)] = \int f \, d\mu \tag{4.8}$$

for all $f \in \mathcal{D}(L)$. By Corollary 3.8 $\mathcal{D}(L)$ is dense in $C_b(Y)$, so (4.8) holds for all $f \in C_b(Y)$. This implies that $\mu S(t) = \mu$. ∎

Example 4.2 This criterion works well for a continuous-time Markov chain on a countable state space S. As for the discrete-time chain, we get a single matrix equation: μ is invariant iff

$$\mu Q = 0 \tag{4.9}$$

where Q is the rate matrix defined in Section 1.2.

$\mathcal{I}$ is a convex set, which means that $\beta\mu + (1-\beta)\nu \in \mathcal{I}$ for all $\mu, \nu \in \mathcal{I}$ and $0 < \beta < 1$. In case of a Feller process, $\mathcal{I}$ is closed in the weak topology of $\mathcal{M}_1(Y)$, because the map $\mu \mapsto \mu S(t)$ is continuous. A measure $\mu \in \mathcal{I}$ is an *extremal* invariant measure, or an *extreme point* of $\mathcal{I}$, if it cannot be expressed as a convex combination of two distinct measures in $\mathcal{I}$. This means that if $\mu = \beta\nu' + (1-\beta)\nu''$ for some $\nu', \nu'' \in \mathcal{I}$ and $0 < \beta < 1$, then necessarily $\nu' = \nu'' = \mu$. $\mathcal{I}_e$ denotes the set of extreme points of $\mathcal{I}$.

Consider the special case of a Feller process on a compact state space. Then by Corollary A.14 of Choquet's theorem, $\mu \in \mathcal{I}$ iff there exists a probability measure Γ on $\mathcal{I}_e$ such that

$$\mu = \int_{\mathcal{I}_e} \nu \, \Gamma(d\nu). \tag{4.10}$$

Thus knowing $\mathcal{I}_e$ is equivalent to knowing the entire collection $\mathcal{I}$. This result is applicable to exclusion processes.

Before turning to the exclusion process, we insert here a general lemma for later use. Let us say a measurable subset A of the state space is *closed* for the process X_t if $P^x[X_t \in A] = 1$ for all $x \in A$. Note that this does not require any topological closedness.

Lemma 4.3 *Suppose A is a closed set for the Markov process X_t, and μ is an invariant measure. Suppose $0 < \mu(A) < 1$. Then both $\mu_1 = \mu(\cdot\,|A)$ and $\mu_2 = \mu(\cdot\,|A^c)$ are invariant measures for X_t.*

An extremal invariant measure μ must have $\mu(A) = 0$ or 1 for every set A that is closed for X_t.

Proof. Consider a function $f \geq 0$. Observe first that, by the assumption on A,

$$\begin{aligned}\mathbf{1}_A(x)S(t)f(x) &= \mathbf{1}_A(x)E^x[f(X_t)] = \mathbf{1}_A(x)E^x[\mathbf{1}_A(X_t)f(X_t)] \\ &\leq E^x[\mathbf{1}_A(X_t)f(X_t)] = S(t)(\mathbf{1}_A f)(x).\end{aligned}$$

Using this and the invariance of μ,

$$\begin{aligned}&\int S(t)f\,d\mu_1 = \frac{1}{\mu(A)}\int \mathbf{1}_A S(t)f\,d\mu \\ &\quad\leq \frac{1}{\mu(A)}\int S(t)(\mathbf{1}_A f)\,d\mu = \frac{1}{\mu(A)}\int \mathbf{1}_A f\,d\mu = \int f\,d\mu_1.\end{aligned}$$

Supposing now that $0 \leq f \leq 1$, we can apply this to both f and $1-f$ to get

$$\int S(t)f\,d\mu_1 \leq \int f\,d\mu_1 \quad \text{and} \quad \int S(t)(1-f)\,d\mu_1 \leq \int (1-f)\,d\mu_1.$$

This implies, since $S(t)1 = 1$, that

$$\int S(t)f\,d\mu_1 = \int f\,d\mu_1$$

and thereby the invariance of μ_1. The invariance of μ_2 now follows from noting first that

$$\mu = \beta\mu_1 + (1-\beta)\mu_2 \text{ for } \beta = \mu(A), \tag{4.11}$$

and then by the invariance of μ and μ_1:

$$\begin{aligned}(1-\beta)\int S(t)f\,d\mu_2 &= \int S(t)f\,d\mu - \beta\int S(t)f\,d\mu_1 \\ &= \int f\,d\mu - \beta\int f\,d\mu_1 = (1-\beta)\int f\,d\mu_2.\end{aligned}$$

Since μ_1 and μ_2 are invariant, (4.11) shows that μ cannot be extremal. The last statement of the lemma follows. ∎

4.1.2 Checking invariance for the exclusion process

Let $\mathcal{I}$ be the set of probability measures on $X = \{0,1\}^S$, $S = \mathbf{Z}^d$, that are invariant for the exclusion process. In this section we improve Theorem 4.1 for the exclusion process, by showing that it suffices to check condition (4.6) for cylinder functions. Let $\mathcal{C}$ be the class of cylinder functions on X, in other words, functions that depend on only finitely many coordinates.

Theorem 4.4 *Let L be the generator of the exclusion process η_t, defined by* (2.12). *Let μ be a probability measure on X. Then $\mu \in \mathcal{I}$ iff*

$$\int Lf\,d\mu = 0 \qquad \textit{for all cylinder functions } f \in \mathcal{C}. \tag{4.12}$$

By Theorem 4.1, this will follow from showing that

Proposition 4.5 *$\mathcal{C}$ is a core for the generator L of the exclusion process.*

To prove Proposition 4.5, we introduce a class of functions larger than the cylinder functions, and with the property that this class is closed under the semigroup evolution. Let

$$\Delta_f(x) = \sup\{|f(\eta) - f(\zeta)| : \eta(y) = \zeta(y) \text{ for all } y \neq x\}. \tag{4.13}$$

Let $D(X)$ be the class of functions $f \in C(X)$ for which $\sum_x \Delta_f(x) < \infty$. Note that if η and ζ agree outside a finite or countably infinite set A, then

$$|f(\eta) - f(\zeta)| \leq \sum_{x \in A} \Delta_f(x).$$

Lemma 4.6 *$S(t)f \in D(X)$ for all $f \in D(X)$.*

Proof. By iterating the semigroup $S(t)$, it suffices to show that $S(t)f \in D(X)$ for $0 < t \leq t_0$. Let $C_{0,t}(x)$ be the connected component that contains x in the random graph $\mathcal{G}_{0,t}$. By Lemma 2.1, $C_{0,t}(x)$ is almost surely finite for every x and $t \leq t_0$. Suppose η and ζ are two configurations that differ only at site x. Using the original construction of Section 2.1 performed on the probability space $(\Omega, \mathcal{H}, \mathbf{P})$ of the Poisson processes $\{\mathcal{T}_{(x,y)}\}$, construct two processes η_t^η and η_t^ζ, the first one started from η and the second from ζ. Both η_t^η and η_t^ζ are exclusion processes in their own right, and they are coupled so that they use the same Poisson jump time processes. Note that $\eta_t^\eta(y) = \eta_t^\zeta(y)$ for y outside $C_{0,t}(x)$.

$$\begin{aligned}|S(t)f(\eta) - S(t)f(\zeta)| &= |\mathbf{E}f(\eta_t^\eta) - \mathbf{E}f(\eta_t^\zeta)| \leq \mathbf{E}|f(\eta_t^\eta) - f(\eta_t^\zeta)| \\ &\leq \mathbf{E}\left[\sum_{y \in C_{0,t}(x)} \Delta_f(y)\right] = \sum_y \Delta_f(y)\mathbf{P}[y \in C_{0,t}(x)].\end{aligned}$$

The bound above is uniform over η and ζ that agree outside $\{x\}$, and hence is a bound on $\Delta_{S(t)f}(x)$. Since $y \in C_{0,t}(x)$ iff $x \in C_{0,t}(y)$,

$$\sum_x \Delta_{S(t)f}(x) \leq \sum_y \Delta_f(y) \sum_x \mathbf{P}[x \in C_{0,t}(y)] = \mathbf{E}|C_{0,t}(0)| \cdot \sum_y \Delta_f(y),$$

where we used translation invariance to say that the expected size $\mathbf{E}|C_{0,t}(x)|$ is independent of x. It remains to observe that $\mathbf{E}|C_{0,t}(0)|$ is finite. This follows by the type of estimation used in the proof of Lemma 2.1. For $|C_{0,t}(0)| > (2k+1)^d$ to be possible, 0 must be connected to a site y such that $|y|_\infty > k$. Then there must be an open path in the graph $\mathcal{G}_{0,t}$ starting at 0 and with at least k/R edges. We get the estimate

$$\mathbf{P}[\,|C_{0,t}(0)| > (2k+1)^d] \leq \sum_{m \geq k/R} \theta^m \leq b\theta^{k/R}$$

where $\theta = k_*(1 - e^{-2t}) < 1$ and $b = (1 - \theta)^{-1}$. The reader can check that this is strong enough to give

$$\mathbf{E}|C_{0,t}(0)| = \sum_{n=0}^{\infty} \mathbf{P}[\,|C_{0,t}(0)| > n] < \infty. \qquad \blacksquare$$

Note that for $f \in D(X)$, the function

$$Lf(\eta) = \sum_{x,y} p(x,y)\eta(x)(1 - \eta(y))[f(\eta^{x,y}) - f(\eta)]$$

exists, since the sum on the right converges uniformly in η. This alone does not imply that $f \in \mathcal{D}(L)$ because membership in the domain is defined by the existence of the limit in (3.7). To get this, we approximate f with a cylinder function.

Lemma 4.7 *Suppose $f \in D(X)$. There exist cylinder functions f_n such that $f_n \to f$ and $Lf_n \to Lf$ uniformly. Consequently $f \in \mathcal{D}(L)$.*

Proof. Once we have the convergence statements, $f \in \mathcal{D}(L)$ follows from the closedness of the operator L (proved in Corollary 3.8), because we have shown that cylinder functions lie in the domain of L.

Recall the definition of R as the maximal distance $|x - y|_\infty$ between two sites with $p(x,y) > 0$. For a cube $A \subseteq S$, let

$$A^c(R) = \{x \in S : |x - y|_\infty \le R \text{ for some } y \in A^c\}.$$

Let $\varepsilon > 0$. Fix a cube A such that

$$\sum_{x \in A^c(R)} \Delta_f(x) < \varepsilon.$$

Note that then also $|f(\eta) - f(\zeta)| < \varepsilon$ for all η and ζ that agree on A. Pick a larger cube B such that $B \supseteq A$. Define a cylinder function g by $g(\eta) = f(\eta^0)$ where $\eta^0 \in X$ is defined by

$$\eta^0(x) = \begin{cases} \eta(x), & x \in B \\ 0, & x \notin B. \end{cases} \tag{4.14}$$

Abbreviate $c(x,y,\eta) = p(x,y)\eta(x)(1 - \eta(y))$.

$$\begin{aligned} Lg(\eta) &= \sum_{x,y \in S} c(x,y,\eta)[g(\eta^{x,y}) - g(\eta)] \\ &= \sum_{x,y \in A} c(x,y,\eta)[g(\eta^{x,y}) - g(\eta)] \\ &\qquad + \left(\sum_{x \in A,\, y \in S\setminus A} + \sum_{x \in S\setminus A,\, y \in S} \right) c(x,y,\eta)[g(\eta^{x,y}) - g(\eta)]. \end{aligned}$$

Decompose $Lf(\eta)$ in exactly the same way. Note that

$$|f(\eta^{x,y}) - f(\eta)| \ \vee \ |g(\eta^{x,y}) - g(\eta)| \le \Delta_f(x) + \Delta_f(y).$$

Then we can bound as follows:

$$\begin{aligned}
\|Lf - Lg\|_\infty \le \sup_\eta & \left| \sum_{x,y\in A} c(x,y,\eta)[f(\eta^{x,y}) - g(\eta^{x,y}) - f(\eta) + g(\eta)] \right| \\
&+ 2\left(\sum_{x\in A,\, y\in S\setminus A} + \sum_{x\in S\setminus A,\, y\in S} \right) p(x,y)\,(\Delta_f(x) + \Delta_f(y)) \\
&\le 2\sum_{x,y\in A} p(x,y)\,\|f-g\|_\infty + 8\sum_{x\in A^c(R)} \Delta_f(x) \\
&\le 2|A|\cdot\|f-g\|_\infty + 8\varepsilon.
\end{aligned}$$

We used the fact that $p(x,y) = 0$ unless x and y are at most distance R apart, and the translation invariance $p(x,y) = p(0, y-x)$. Keep cube A fixed, and choose cube B large enough so that $\|f-g\|_\infty < \varepsilon/|A|$.

Since $\varepsilon > 0$ was arbitrary, we have shown that (f, Lf) can be uniformly approximated by (g, Lg) for cylinder functions g. ∎

Proof of Proposition 4.5. $D(X)$ is dense in $C(X)$ because $\mathcal{C}$ is dense and $\mathcal{C} \subseteq D(X)$. By the lemmas above, $D(X)$ lies in the domain $\mathcal{D}(L)$ and is closed under the semigroup action. Thus by Proposition 3.12, $D(X)$ is a core for L. But then Lemma 4.7 implies that $\mathcal{C}$ is also a core, because now an arbitrary (f, Lf) for $f \in \mathcal{D}(L)$ can be uniformly approximated by (g, Lg) for some $g \in D(X)$, which in turn by Lemma 4.7 can be uniformly approximated by (h, Lh) for some $h \in \mathcal{C}$. ∎

For $0 \le \rho \le 1$, let ν_ρ denote the Bernoulli measure on X with density ρ, defined by

$$\nu_\rho\{\eta : \eta(x) = 1 \text{ for all } x \in A,\ \eta(y) = 0 \text{ for all } y \in B\} = \rho^{|A|}(1-\rho)^{|B|} \tag{4.15}$$

for any two disjoint finite sets of sites A and B. Equivalently, under ν_ρ the occupation variables $\{\eta(x)\}$ are i.i.d. with mean ρ.

Exercise 4.1 Check that Bernoulli measures ν_ρ are invariant for the translation invariant, finite range exclusion process. This will be established as part of Theorem 6.1 below.

Exercise 4.2 *Equilibrium distributions for independent random walks.* Let initial occupation variables $\{\eta_0(x) : x \in \mathbf{Z}^d\}$ be given, with values in $\mathbf{Z}_+$. At site x, put $\eta_0(x)$ particles.

Let all particles execute independent random walks, so that each particle jumps at rate one and new sites are chosen according to a fixed translation-invariant jump probability $p(x, y) = p(0, y - x)$. Let $\eta_t(x)$ denote the number of particles at site x at time t. Show that if $\{\eta_0(x)\}$ are i.i.d Poisson distributed, then so are $\{\eta_t(x)\}$.

Hint. Do not try to use advanced machinery. Just compute a joint Laplace transform

$$E[\exp\{-\lambda_1\eta_t(x_1) - \lambda_2\eta_t(x_2) - \cdots - \lambda_m\eta_t(x_m)\}].$$

Count how many particles have moved from site y to site x_i during $[0, t]$, and use independence.

4.2 Uniqueness results

Return again to the general setting of a Feller process $\{P^x\}$ on the path space D_Y, with a strongly continuous contraction semigroup $S(t)f(x) = E^x[f(X_t)]$ and generator L on $C_b(Y)$. It is a consequence of Lemma 3.7(c) that for any f in the domain of L,

$$M_t = f(X_t) - \int_0^t Lf(X_s)\,ds \tag{4.16}$$

is a martingale with respect to the filtration $\mathcal{F}_t = \sigma\{X_s : 0 \le s \le t\}$ (Exercise 3.1). It is useful to know that these martingales actually characterize the Markov process. Recall that $P^\mu = \int P^x\,\mu(dx)$ denotes the probability measure on D_Y under which the Markov process $X.$ has initial distribution μ.

Theorem 4.8 *Let μ be a probability measure on Y. Let $\mathcal{Y}$ be any core for the generator L. Suppose P is a probability measure on D_Y with these properties:*

(a) *$P[X_0 \in B] = \mu(B)$ for all Borel sets $B \subseteq Y$.*

(b) *M_t is a martingale under the measure P, for all $f \in \mathcal{Y}$.*

Then $P = P^\mu$.

Proof. We write E for expectation under measure P, and as before, E^μ for expectation under measure P^μ.

First we observe that (b) can be strengthened to work for all f in the domain of L. For given $f \in \mathcal{D}(L)$, $s < t$, and an event $A \in \mathcal{F}_s$, find $g_n \in \mathcal{Y}$ such that $g_n \to f$ and $Lg_n \to Lf$ boundedly and uniformly. That this is possible is the definition of a core. By (b),

$$E\left[\mathbf{1}_A\left(g_n(X_t) - \int_0^t Lg_n(X_u)\,du\right)\right] = E\left[\mathbf{1}_A\left(g_n(X_s) - \int_0^s Lg_n(X_u)\,du\right)\right].$$

Let $n \to \infty$. Bounded convergence replaces g_n by f so that

$$E\left[\mathbf{1}_A\left(f(X_t) - \int_0^t Lf(X_u)\,du\right)\right] = E\left[\mathbf{1}_A\left(f(X_s) - \int_0^s Lf(X_u)\,du\right)\right].$$

Since $A \in \mathcal{F}_s$ is arbitrary and since $M_s = f(X_s) - \int_0^s Lf(X_u)\,du$ is $\mathcal{F}_s$-measurable, this says that $E[M_t|\mathcal{F}_s] = M_s$. In other words that M_t is a martingale.

Let $g \in C_b(Y)$ and $\lambda > 0$. By Proposition 3.10 there exists $f \in \mathcal{D}(L)$ such that

$$(\lambda - L)f = g.$$

By the martingale property, for $0 \le s < t$,

$$E\left[f(X_t) - \int_s^t Lf(X_u)\,du \,\middle|\, \mathcal{F}_s\right] = f(X_s). \tag{4.17}$$

Multiply this by $\lambda e^{-\lambda t}$ and integrate over $t \in [s, \infty)$:

$$E\left[\int_s^\infty e^{-\lambda t}\lambda f(X_t)\,dt - \int_s^\infty e^{-\lambda t} Lf(X_t)\,dt \,\middle|\, \mathcal{F}_s\right] = e^{-\lambda s} f(X_s)$$

which gives

$$E\left[\int_s^\infty e^{-\lambda t} g(X_t)\,dt \,\middle|\, \mathcal{F}_s\right] = e^{-\lambda s} f(X_s). \tag{4.18}$$

Measure P^μ also satisfies hypotheses (a)–(b), so (4.18) is valid also when E is replaced by E^μ.

We prove by induction on n that P and P^μ have the same finite-dimensional distributions on n variables $(X_{s_1}, \ldots, X_{s_n})$ for any $0 \le s_1 < \cdots < s_n$. This suffices for $P = P^\mu$ as measures on D_Y.

First set $s = 0$ in (4.18) and take expectations of both sides with respect to P. This and assumption (a) give

$$\int_0^\infty e^{-\lambda t} E[g(X_t)]\,dt = E[f(X_0)] = \int f\,d\mu.$$

The same step applies to P^μ too, and we have

$$\int_0^\infty e^{-\lambda t} E[g(X_t)]\,dt = \int_0^\infty e^{-\lambda t} E^\mu[g(X_t)]\,dt.$$

Both measures P and P^μ live on the path space D_Y and so both expectations $E[g(X_t)]$ and $E^\mu[g(X_t)]$ are right-continuous functions of t. By the uniqueness of Laplace transforms (Lemma A.19),

$$E[g(X_t)] = E^\mu[g(X_t)]$$

for all t. This is valid for all $g \in C_b(Y)$, and so we have shown that X_t has the same distribution under P and P^μ, for any $t \geq 0$.

Now assume that the vector $(X_{s_1}, \ldots, X_{s_n})$ has the same distribution under P and P^μ, for any n time points $0 \leq s_1 < \cdots < s_n$. Take $s = s_n$ in (4.18), let $h_1, \ldots, h_n$ be arbitrary bounded continuous functions on Y, and multiply (4.18) by

$$\prod_{i=1}^{n} h(X_{s_i}).$$

Notice that this product is $\mathcal{F}_{s_n}$-measurable and so can be taken inside the conditional expectation on the left-hand side of (4.18). Take expectation under P. Take the expectation inside the t-integral. This gives

$$\int_{s_n}^{\infty} e^{-\lambda t} E\left[g(X_t) \prod_{i=1}^{n} h(X_{s_i}) \right] dt = e^{-\lambda s_n} E\left[f(X_{s_n}) \prod_{i=1}^{n} h(X_{s_i}) \right]. \tag{4.19}$$

The same step applies to P^μ too, and we have

$$\int_{s_n}^{\infty} e^{-\lambda t} E^\mu\left[g(X_t) \prod_{i=1}^{n} h(X_{s_i}) \right] dt = e^{-\lambda s_n} E^\mu\left[f(X_{s_n}) \prod_{i=1}^{n} h(X_{s_i}) \right]. \tag{4.20}$$

By the induction assumption, the right-hand sides of (4.19) and (4.20) agree. Hence so do the left-hand sides. Use again the uniqueness of Laplace transforms. Note that having

$$\int_{s_n}^{\infty} e^{-\lambda t} u(t)\, dt = \int_{s_n}^{\infty} e^{-\lambda t} v(t)\, dt$$

for all $\lambda > 0$ guarantees that $u = v$ on $[s_n, \infty)$ which is what we need here. So we have

$$E\left[g(X_t) \prod_{i=1}^{n} h(X_{s_i}) \right] = E^\mu\left[g(X_t) \prod_{i=1}^{n} h(X_{s_i}) \right]$$

for any set $0 \leq s_1 < \cdots < s_n < t$ of $n+1$ time points. ∎

Let us emphasize two things: The measure P in the theorem was not even assumed to represent a Markov process. This is useful because checking the Markov property of a given process may be tricky. Second, the core $\mathcal{Y}$ can of course be the entire domain of L.

Next we show that the action of the generator on cylinder functions uniquely specifies the exclusion process among Feller processes on X with strongly continuous semigroups. This point is worth making because we defined the exclusion process by directly constructing the probability measures $\{P^\eta\}$ on the path space. This construction alone does not rule out the possibility that some other Markov process $\{\widetilde{P}^\eta\}$ has the same generator.

So let L be the exclusion generator (2.12). Suppose a Markov process $\{\widetilde{P}^\eta\}$ exists on the path space D_X with these properties: the semigroup $T(t)f(\eta) = \widetilde{E}^\eta[f(\eta_t)]$ is a strongly continuous contraction semigroup on $C(X)$ with generator M, cylinder functions lie in the domain of M, and $Mf = Lf$ for cylinder functions f.

Proposition 4.9 *$P^\eta = \widetilde{P}^\eta$ for all $\eta \in X$.*

Proof. This can be viewed as a corollary of Theorem 4.8. Here is an alternative argument from semigroup theory.

First observe that, since both L and M are closed, $\mathcal{D}(L) \subseteq \mathcal{D}(M)$. For if $f \in \mathcal{D}(L)$, we can find a sequence f_n from the core $\mathcal{C}$ (for L) such that $(f_n, Lf_n) \to (f, Lf)$. And then $f \in \mathcal{D}(M)$ and $Mf = Lf$. Thus we know M must be an extension of L, and by Proposition 3.13 the semigroups agree. Then, as explained in Section 1.3, the transition probabilities agree, and consequently the probability measures P^η and $\widetilde{P}^\eta$ on path space must agree. ∎

Exercise 4.3 Equation (4.17) is valid almost surely, for any fixed $s < t$. To obtain (4.18) from it through the integration step we in principle involve uncountably many t-values. Give a rigorous justification for the almost sure validity of (4.18) for any fixed s.

Exercise 4.4 Solve Exercise 1.5 using Theorem 4.8.

Exercise 4.5 Check that the graphical representation of Section 2.1 and the stirring particle construction of Section 2.2 produce the same process, when the jump kernel is symmetric.

PART II Convergence to equilibrium

5 Symmetric exclusion processes

We have constructed the exclusion process under the assumptions of translation invariance and finite range of the jump probability. These meant that $p(x, y) = p(0, y - x)$ for all $x, y \in S = \mathbf{Z}^d$, and that the set $\{x : p(0, x) > 0\}$ is finite. In this chapter we make two additional assumptions, namely

$$\text{symmetry: } p(x, y) = p(y, x) \text{ for all } x, y \in S,$$

and

$$\text{irreducibility: for all } x, y \in S \text{ there exists } n > 0 \text{ such that } p^{(n)}(x, y) > 0.$$

Note that by symmetry we are actually not assuming anything more than the relaxed irreducibility condition (1.26). The finite range condition will not be explicitly used in this chapter. We write $\mathcal{M}_1$ for the space of probability measures on the state space $X = \{0, 1\}^S$ of the exclusion process.

This chapter proves the following two theorems under the assumptions mentioned above.

Theorem 5.1 *The class $\mathcal{I}$ of invariant measures is precisely the class of exchangeable measures on $X = \{0, 1\}^S$.*

Recall the definition of the Bernoulli measure ν_ρ for $0 \le \rho \le 1$. Under ν_ρ the coordinates $\{\eta(x)\}$ are independent with common distribution

$$\nu_\rho\{\eta : \eta(x) = 1\} = \rho \quad \text{and} \quad \nu_\rho\{\eta : \eta(x) = 0\} = 1 - \rho.$$

By de Finetti's Theorem A.16 the exchangeable measures are exactly the mixtures of Bernoulli measures. So Theorem 5.1 can be equivalently stated by saying that the extreme points of $\mathcal{I}$ are the Bernoulli measures, or $\mathcal{I}_e = \{\nu_\rho : 0 \le \rho \le 1\}$.

The second theorem concerns convergence to equilibrium from a special class of initial distributions, namely the translation invariant ergodic ones. The spatial *translations*, or *shifts*, are bijective maps θ_x defined on X by $\theta_x\eta(y) = \eta(x + y)$ for all $x, y \in S$. They are continuous, hence measurable. As usual, maps on a space act on measures through composition with inverses, so for any $\mu \in \mathcal{M}_1$, the measure $\mu \circ \theta_x^{-1}$ is defined by

$$\mu \circ \theta_x^{-1}(A) = \mu(\theta_x^{-1}A) = \mu\{\eta : \theta_x\eta \in A\} \tag{5.1}$$

for measurable sets $A \subseteq X$. A measure μ is *translation invariant* if $\mu = \mu \circ \theta_x^{-1}$ for all $x \in S$. Let $\mathcal{S}$ denote the set of translation invariant probability measures on X. An event A is translation invariant if $\theta_x^{-1} A = A$ for all x. A measure $\mu \in \mathcal{S}$ is *ergodic* if $\mu(A) = 0$ or 1 for every translation invariant event A. The Bernoulli measures are a basic example of ergodic measures in $\mathcal{S}$.

If μ is the initial distribution of the process, let $\mu_t = \mu S(t)$ be the distribution of η_t, the state at time t.

Theorem 5.2 *Suppose μ is a translation invariant and ergodic probability measure on X with density $\rho = \mu\{\eta(x) = 1\}$. Then μ_t converges weakly to ν_ρ as $t \to \infty$.*

The proof of Theorem 5.1 splits into two parts, depending on whether the jump kernel $p(x, y)$ is recurrent or transient. For an irreducible discrete-time Markov chain X_n with transition $p(x, y)$, recurrence is equivalent to

$$P^x[X_n = y \text{ for infinitely many } n \geq 1] = 1 \qquad \text{for all states } x \text{ and } y,$$

while transience is equivalent to

$$P^x[X_n = y \text{ for only finitely many } n \geq 1] = 1 \qquad \text{for all states } x \text{ and } y.$$

We begin with the key notion of duality.

5.1 Duality

In general, two Markov processes z_t and w_t with state spaces Z and W are in duality with respect to a bounded measurable function H on $Z \times W$, if for all states $z \in Z$ and $w \in W$,

$$E^z H(z_t, w) = E^w H(z, w_t).$$

On the left the w-argument is fixed while E^z denotes expectation over the random z_t with initial state z. And similarly on the right. For symmetric exclusion we obtain this type of relationship between two versions of the same process. The second member of the duality is taken to be an exclusion process with finitely many particles. The benefit of the duality is that questions about the general process can be converted into questions about the finite exclusion process. This is an improvement because the latter is a countable state Markov chain.

Let Y denote the set of all finite subsets of S. It is a countable set. Equation

$$A = \{x \in S : \eta(x) = 1\}$$

maps bijectively between configurations $\eta \in X$ with finitely many particles [meaning that $\sum_x \eta(x) < \infty$] and sets $A \in Y$. So we can regard an exclusion process with finitely many particles as a countable state Markov chain with state space Y. Let A_t denote the process, in other words the set of occupied sites at time t.

The next theorem expresses the self-duality of symmetric exclusion, a key tool for its analysis.

Theorem 5.3 *For $\eta \in X$ and $A \in Y$,*

$$P^\eta\{\eta_t(x) = 1 \textit{ for all } x \in A\} = P^A\{\eta(x) = 1 \textit{ for all } x \in A_t\}. \tag{5.2}$$

Proof. Fix a realization of the Poisson clocks $\{\mathcal{T}_{\{x,y\}}\}$ for the graphical construction for the time interval $[0, t]$ used in Section 2.2. Let X_s^y be the stirring particles that march backward in time, as defined in Section 2.2. According to (2.7) we can construct η_t by $\eta_t(x) = \eta(X_t^x)$. Let us construct the evolution A_s $(0 \le s \le t)$ backwards in time as suggested in the last paragraph of Section 2.2, so that $A_t = \{X_t^x : x \in A\}$. Then $\eta_t(x) = 1$ for all $x \in A$ iff $\eta(X_t^x) = 1$ for all $x \in A$ iff $\eta(x) = 1$ for all $x \in A_t$. ∎

For a probability measure μ on X, define a function $\widehat{\mu}$ on Y by

$$\widehat{\mu}(A) = \mu\{\eta : \eta(x) = 1 \text{ for all } x \in A\}. \tag{5.3}$$

Note that two probability measures μ and ν on X are equal if

$$\mu\{\eta = 1 \text{ on } A\} = \nu\{\eta = 1 \text{ on } A\}$$

for all $A \in Y$. In other words, $\mu = \nu$ iff $\widehat{\mu} = \widehat{\nu}$. The function $\widehat{\mu}$ can be used to conveniently express the duality.

Corollary 5.4 *For $\mu \in \mathcal{M}_1$, let $\mu_t = \mu S(t)$ be the distribution of η_t when the initial distribution is μ. Then for $A \in Y$,*

$$\widehat{\mu_t}(A) = E^A \widehat{\mu}(A_t). \tag{5.4}$$

Proof. Integrate (5.2) against μ:

$$\begin{aligned}
\widehat{\mu_t}(A) &= \mu_t\{\eta = 1 \text{ on } A\} = \int_X P^\eta\{\eta_t = 1 \text{ on } A\}\mu(d\eta) = \int_X P^A\{\eta = 1 \text{ on } A_t\}\mu(d\eta) \\
&= E^A \int_X \mathbf{1}\{\eta = 1 \text{ on } A_t\}\mu(d\eta) = E^A \widehat{\mu}(A_t). \qquad \blacksquare
\end{aligned}$$

Duality converts the question about invariant measures of η_t into a question about harmonic functions of the finite exclusion process A_t. For $\nu \in \mathcal{M}_1$, ν is invariant for the exclusion process η_t iff $\nu = \nu_t$ iff $\widehat{\nu} = \widehat{\nu_t}$. By (5.4) this last statement is equivalent to having $\widehat{\nu}(A) = E^A \widehat{\nu}(A_t)$ for all $A \in Y$. We conclude that

$$\nu \text{ is invariant for } \eta_t \text{ iff } \widehat{\nu} \text{ is harmonic for } A_t. \tag{5.5}$$

We get the following intermediate results toward a characterization of $\mathcal{I}$.

Proposition 5.5 *All exchangeable measures lie in $\mathcal{I}$.*

Proof. According to Exercise A.6 in Section A.6, if ν is exchangeable, then $\widehat{\nu}(A)$ depends only on $|A|$. Since $|A_t| = |A|$ (particles are neither created nor destroyed in an exclusion process), $\widehat{\nu}(A_t) = \widehat{\nu}(A)$. So $\widehat{\nu}$ is not only harmonic for A_t but even almost surely constant in time. ∎

Proposition 5.6 *In order to prove that $\mathcal{I}$ is exactly the class of exchangeable measures, it suffices to prove this statement: If f is a bounded harmonic function for A_t, then $f(A)$ depends only on $|A|$.*

Proof. It only remains to show that all invariant measures are necessarily exchangeable. So let ν be invariant. Then by (5.5) $\widehat{\nu}$ is harmonic for A_t. If the statement about harmonic functions of A_t is proved, then we know $\nu\{\eta = 1 \text{ on } A\} = \widehat{\nu}(A)$ depends only on $|A|$. According to Exercise A.6 this property characterizes exchangeability. Thus ν must be exchangeable. ∎

5.2 Proof of Theorem 5.1 in the recurrent case

By Propositions 5.5 and 5.6, Theorem 5.1 will follow from proving this proposition:

Proposition 5.7 *Suppose the transition probability $p(x, y)$ is recurrent. Let f be a bounded function on the space Y of finite subsets of S. If f is harmonic for the finite exclusion process A_t, then $f(A)$ depends only on $|A|$, the size of the set.*

Proof. It suffices to show that, for any two n-sets A and B that have $n-1$ points in common, the processes A_t and B_t started from A and B can be coupled successfully. For then, as in Chapter 1,

$$|f(A) - f(B)| = |E^A f(A_t) - E^B f(B_t)| \le E|f(A_t) - f(B_t)| \le 2\|f\|_\infty P(A_t \ne B_t) \to 0.$$

Since any two n-sets can be transformed into each other by replacing one point at a time, and since n is arbitrary, it follows that $f(A) = f(B)$ for any two sets A and B with the same number of points.

The constructions used in this proof do not utilize the graphical representation. They are more basic and require only construction of certain countable state Markov chains.

A single particle exclusion process is simply a random walk. So in the case $n = 1$, we can write $A_t = \{X_t\}$ and $B_t = \{Y_t\}$ where X_t and Y_t are random walks with jump rates $p(x, y)$. Define the coupled process (X_t, Y_t) so that X_t and Y_t move independently until they meet for the first time, and after that they move together. By symmetry, $Z_t = X_t - Y_t$ is a random walk with twice the rate but the same jump probability $p(x, y)$ (Exercise 1.5). Hence Z_t is recurrent and therefore almost surely eventually hits 0. This is the same as saying that almost surely there is a time $t \geq 0$ at which $X_t = Y_t$, so the coupling is successful. We leave as an exercise the precise definition of the coupled process.

Now for the case $n > 1$. We shall describe a process (C_t, X_t, Y_t) where C_t is an $(n-1)$-set of points in S, and X_t and Y_t are S-valued but outside C_t. The state space of this process is

$$\mathcal{Z} = \{(C, x, y) \in Y \times S \times S : |C| = n - 1,\ x \notin C,\ y \notin C\}.$$

This process will realize a coupling (A_t, B_t) through the formulas $A_t = C_t \cup \{X_t\}$ and $B_t = C_t \cup \{Y_t\}$. In particular, this coupling of A_t and B_t will be successful if eventually $X_t = Y_t$.

The process (C_t, X_t, Y_t) is a countable state Markov chain with generator G of the standard form

$$Gf(\mathbf{z}) = \sum_{\mathbf{w} \in \mathcal{Z}} r(\mathbf{z}, \mathbf{w})[f(\mathbf{w}) - f(\mathbf{z})].$$

Here we denoted generic elements of $\mathcal{Z}$ by $\mathbf{z}$ and $\mathbf{w}$, and $r(\mathbf{z}, \mathbf{w})$ is the rate of jumping from state $\mathbf{z}$ to $\mathbf{w}$. Below we describe these jump rates case by case. For $u \in C$ and $v \notin C$, write $C_{u,v} = (C \setminus \{u\}) \cup \{v\}$ for the effect of removing point u and adding v to the set C.

Case 1. Suppose the current state is $\mathbf{z} = (C, x, y)$ with $x \neq y$. In each case below, $u \in C$ and $v \notin C \cup \{x, y\}$. Then

$$\begin{array}{llll}
\text{the jump to state } \mathbf{w} = & (C_{u,v}, x, y) & \text{happens at rate } r(\mathbf{z}, \mathbf{w}) = & p(u, v) \\
& (C, v, y) & & p(x, v) \\
& (C, y, y) & & p(x, y) \\
& (C, x, v) & & p(y, v) \\
& (C, x, x) & & p(y, x) \\
& (C_{u,x}, u, y) & & p(u, x) = p(x, u) \\
& (C_{u,y}, x, u) & & p(u, y) = p(y, u).
\end{array}$$

Case 2. Suppose the current state is $\mathbf{z} = (C, x, x)$. Let $u \in C$ and $v, y \notin C \cup \{x\}$. Then

$$\begin{array}{llll} \text{the jump to state } \mathbf{w} = & (C_{u,v}, x, x) & \text{happens at rate } r(\mathbf{z}, \mathbf{w}) = & p(u, v) \\ & (C_{u,x}, u, u) & & p(u, x) = p(x, u) \\ & (C, y, y) & & p(x, y). \end{array}$$

Reading off the rates of the marginal processes, we see that X_t and Y_t are random walks with jump rates $p(x, y)$ that evolve independently until they meet, after which they stay together. Hence again $Z_t = X_t - Y_t$ is a random walk with twice the rates until it is absorbed at the origin. Eventually $Z_t = X_t - Y_t = 0$ by the recurrence assumption, so the coupling is successful.

To make this reasoning rigorous, we apply the martingale characterization Theorem 4.8 along the following lines. By Exercise 3.1 the process

$$M_t = f(C_t, X_t, Y_t) - \int_0^t Gf(C_s, X_s, Y_s)\, ds \tag{5.6}$$

is a martingale for any $f \in C_b(\mathcal{Z})$. Let ϕ be a bounded continuous function on S, and take $f(C, x, y) = \phi(x - y)$. Check that for this f, $Gf(C, x, y) = H\phi(x - y)$ where

$$H\phi(z) = \mathbf{1}\{z \neq 0\} \cdot \sum_{v \in S: v \neq z} 2p(z, v)[\phi(v) - \phi(z)].$$

Define the process $Z_t = X_t - Y_t$. Eq. (5.6) becomes

$$M_t = \phi(Z_t) - \int_0^t H\phi(Z_s)\, ds. \tag{5.7}$$

Now observe that H is precisely the generator of a random walk on S that jumps with rates $2p(x, y)$ until it is absorbed at the origin. We could construct such a random walk as was done in Section 1.2, and then derive the generator H from this construction as in Theorem 1.2. Since we have shown that (5.7) is a martingale for every $\phi \in C_b(S)$, Z_t is exactly this random walk by Theorem 4.8.

Using the same argument one checks that $A_t = C_t \cup \{X_t\}$ behaves exactly as a finite exclusion process, jumping from A to $A_{a,b}$ with rate $p(a, b)$ for any $a \in A$ and $b \notin A$. Similarly for $B_t = C_t \cup \{Y_t\}$. To summarize, we have constructed a successful coupling of A_t and B_t and thereby completed the proof. ∎

5.3 Comparison with independent walks

In the transient case we cannot hope to couple successfully as we did in the proof of Proposition 5.7. Instead, we approach the finite exclusion process by comparing it with a system where the walks are independent and not subject to the exclusion rule.

Let $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ be a vector of independent random walks on $S = \mathbf{Z}^d$ with jump rates 1 and jump probabilities $p(x, y)$. Let $U(t)$ denote the semigroup for $\mathbf{X}(t)$, explicitly given by

$$U(t)f(\mathbf{x}) = \sum_{\mathbf{y} \in S^n} \prod_{i=1}^{n} p_t(x_i, y_i) f(\mathbf{y}) \tag{5.8}$$

where we wrote $\mathbf{x} = (x_1, \ldots, x_n)$ for an n-vector of sites from S, and similarly for $\mathbf{y}$. The transition probabilities of the individual walks $X_i(t)$ that appear in the formula above are given by

$$p_t(x, y) = \sum_{n=0}^{\infty} \frac{e^{-t} t^n}{n!} p^{(n)}(x, y). \tag{5.9}$$

The generator of $U(t)$ is the bounded operator U on $C_b(S^n)$ given by

$$Uf(\mathbf{x}) = \sum_{i=1}^{n} \sum_{y \in S} p(x_i, y)[f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n) - f(\mathbf{x})]. \tag{5.10}$$

For the comparison with $U(t)$, it is convenient to encode a finite exclusion process with n particles by keeping track of the vector $\mathbf{x}$ of particle locations. The state space for this process is

$$T = \{\mathbf{x} \in S^n : \ x_i \neq x_j \text{ for } i \neq j \ \} \tag{5.11}$$

and the generator

$$Vf(\mathbf{x}) = \sum_{i=1}^{n} \sum_{y \in S \setminus \{x_1, \ldots, x_n\}} p(x_i, y)[f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n) - f(\mathbf{x})]. \tag{5.12}$$

We write $V(t)$ for the semigroup of the n particle exclusion on the space T.

The connection between $V(t)$ and our earlier notation A_t is the natural one, as can be seen by constructing them together with the graphical representation. Let

$$\mathbf{w}(t) = (w_1(t), \ldots, w_n(t))$$

denote the process on T with generator V and initial state $\mathbf{w} = (w_1, \ldots, w_n)$. Let A_t be the finite exclusion process with initial set $A = \{w_1, \ldots, w_n\}$. If both processes follow the same Poisson clocks in the graphical representation, $A_t = \{w_1(t), \ldots, w_n(t)\}$ for all $t \geq 0$. In particular, given a bounded function h defined on n-sets $A \in Y$, define $f(\mathbf{x}) = h(\{x_1, \ldots, x_n\})$. If A and $\mathbf{w}$ are as above, then

$$E^A h(A_t) = V(t)f(\mathbf{w}). \tag{5.13}$$

The comparison of $V(t)$ and $U(t)$ will come in the following somewhat abstract form. Let us say a bounded, symmetric function $f(x,y)$ on S^2 is positive definite if

$$\sum_{x,y} f(x,y)\psi(x)\psi(y) \geq 0 \tag{5.14}$$

for absolutely summable real functions ψ on S. Absolute summability means that $\sum |\psi(x)| < \infty$ and guarantees that the sum in (5.14) is well-defined. A bounded symmetric function of n variables is positive definite if the condition is satisfied for any two variables, holding the other $n-2$ variables fixed.

Proposition 5.8 *For every bounded symmetric positive definite function f on S^n and every $\mathbf{x} \in T$, $V(t)f(\mathbf{x}) \leq U(t)f(\mathbf{x})$.*

Proof. Directly from (5.10) and (5.12) for $\mathbf{x} \in T$

$$\begin{aligned} Uf(\mathbf{x}) \; - \; Vf(\mathbf{x}) & \\ = \sum_{1\leq i,j\leq n} & p(x_i,x_j)[f(x_1,\ldots,x_{i-1},x_j,x_{i+1},\ldots,x_n) - f(\mathbf{x})] \\ = \frac{1}{2} \sum_{1\leq i,j\leq n} & p(x_i,x_j)[f(x_1,\ldots,x_{i-1},x_j,x_{i+1},\ldots,x_n) \\ & +f(x_1,\ldots,x_{j-1},x_i,x_{j+1},\ldots,x_n) - 2f(\mathbf{x})]. \end{aligned}$$

To get the last form, take 1/2 times the middle expression, and another 1/2 times the middle expression after first interchanging i and j in the sum and then using symmetry to replace $p(x_j,x_i)$ by $p(x_i,x_j)$. The last expression in brackets is as in (5.14) with $\psi(x) = \delta_{x_i}(x) - \delta_{x_j}(x)$. Thus $Uf - Vf \geq 0$ on T for bounded symmetric positive definite f.

Next we show that operating with $U(t)$ preserves the property of positive definiteness of a function f.

$$\begin{aligned} \sum_{x_j,x_k\in S} \psi(x_j)\psi(x_k)U(t)f(\mathbf{x}) &= \sum_{x_j,x_k\in S} \psi(x_j)\psi(x_k) \sum_{\mathbf{y}\in S^n} \left\{ \prod_{i=1}^{n} p_t(x_i,y_i) \right\} f(\mathbf{y}) \\ = \sum_{i\neq j,k} \sum_{y_i\in S} \left\{ \prod_{i\neq j,k} p_t(x_i,y_i) \right\} & \sum_{y_j,y_k\in S} \left(\sum_{x_j} \psi(x_j)p_t(x_j,y_j) \right) \left(\sum_{x_k} \psi(x_k)p_t(x_k,y_k) \right) f(\mathbf{y}). \end{aligned}$$

Let $\varphi(y) = \sum_x \psi(x)p_t(x,y)$. Then the last line above is a positive linear combination of terms $\sum_{y_j,y_k\in S} \varphi(y_j)\varphi(y_k)f(\mathbf{y})$, each nonnegative by the positive definiteness of f. Thus the entire last line is nonnegative, and so $U(t)f$ is positive definite. That $U(t)$ preserves symmetry is evident from (5.8), and boundedness is preserved because $U(t)$ is a contraction semigroup.

Now we may apply the first conclusion to the function $U(s)f$, and claim that $(U - V)U(s)f \geq 0$ on T. The semigroup $V(t)$ preserves nonegativity of functions, and so

$$V(t-s)(U-V)U(s)f \geq 0$$

on T. Finally, by Proposition 3.9,

$$U(t)f - V(t)f = \int_0^t V(t-s)(U-V)U(s)f\, ds \geq 0$$

on T. ∎

Corollary 5.9 *Let* $\mathbf{x} = (x_1, \ldots, x_n) \in T$ *be an* n*-vector with distinct entries, and consider also the* n*-set* $A = \{x_1, \ldots, x_n\} \in Y$*. Then for* $\mu \in \mathcal{M}_1$ *with* $\mu_t = \mu S(t)$*,*

$$\widehat{\mu}_t(A) \leq E^{\mathbf{x}} \widehat{\mu}\{X_1(t), \ldots, X_n(t)\}.$$

Proof. For all vectors $\mathbf{x} \in S^n$, let

$$f(\mathbf{x}) = \mu\{\eta(x_1) = \cdots = \eta(x_n) = 1\} = \widehat{\mu}\{x_1, \ldots, x_n\}.$$

Note that this definition is not problematic even if some of the points x_i coincide. The function f is bounded, symmetric, and positive definite as can be seen from this calculation:

$$\begin{aligned}\sum_{x,y} f(x, y, x_3, \ldots, x_n)\psi(x)\psi(y) &= \int \sum_{x,y} \eta(x)\eta(y)\psi(x)\psi(y) \left\{\prod_{i=3}^{n} \eta(x_i)\right\} \mu(d\eta) \\ &= \int \left(\sum_{x,y} \eta(x)\psi(x)\right)^2 \left\{\prod_{i=3}^{n} \eta(x_i)\right\} \mu(d\eta) \geq 0.\end{aligned}$$

By duality (5.4), (5.13), and Proposition 5.8,

$$\begin{aligned}\widehat{\mu}_t(A) &= E^A \widehat{\mu}(A_t) = V(t)f(\mathbf{x}) \leq U(t)f(\mathbf{x}) \\ &= E^{\mathbf{x}} f(X_1(t), \ldots, X_n(t)) = E^{\mathbf{x}} \widehat{\mu}\{X_1(t), \ldots, X_n(t)\}.\end{aligned}$$ ∎

5.4 Proof of Theorem 5.1 in the transient case

Again our goal is to prove that bounded harmonic functions for the finite exclusion process depend only on the size of the set. The transition probability $p(x, y)$ is transient, in addition to the assumptions of translation invariance and irreducibility.

As before, $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ denotes a vector of independent random walks on $S = \mathbf{Z}^d$, each with jump rates $p(x, y)$. Recall that T defined by (5.11) is the set of n-vectors with distinct entries, and serves as the state space for n particle exclusion. $U(t)$ denotes the semigroup of $\mathbf{X}(t)$, and $V(t)$ the semigroup of n particle exclusion. For $\mathbf{x} = (x_1, \ldots, x_n) \in S^n$, let

$$\begin{aligned} g(\mathbf{x}) &= P^{\mathbf{x}}[\mathbf{X}(t) \notin T \text{ for some } t \geq 0] \\ &= P^{\mathbf{x}}[X_i(t) = X_j(t) \text{ for some } 1 \leq i < j \leq n \text{ and } t \geq 0]. \end{aligned} \tag{5.15}$$

Lemma 5.10 (a) *For* $\mathbf{x} \in S^n$, $U(t)g(\mathbf{x}) \to 0$ *as* $t \to \infty$.

(b) *For* $\mathbf{x} \in T$, $V(t)g(\mathbf{x}) \to 0$ *as* $t \to \infty$.

Proof. For $x, y \in S$, define

$$w(x, y) = P^{(x,y)}[X_1(t) = X_2(t) \text{ for some } t > 0].$$

We check that $w(x, y)$ is positive definite in the sense of definition (5.14). First fix $0 < t_1 < t_2 < \cdots < t_m$, and consider the function

$$\widetilde{w}(x, y) = P^{(x,y)}[X_1(t_i) = X_2(t_i) \text{ for some } 1 \leq i \leq m].$$

If we can show $\sum_{x,y} \widetilde{w}(x, y)\psi(x)\psi(y) \geq 0$, then by letting $\{t_i\}$ increase into a countable dense subset of $(0, \infty)$, in the limit we recover $\sum_{x,y} w(x, y)\psi(x)\psi(y) \geq 0$.

$$\begin{aligned} &\sum_{x,y} \psi(x)\psi(y)P^{(x,y)}[X_1(t_i) = X_2(t_i) \text{ for some } 1 \leq i \leq m] \\ &\quad = \sum_{x,y} \psi(x)\psi(y) \sum_{i=1}^{m} P^{(x,y)}[X_1(t_i) = X_2(t_i), X_1(t_j) \neq X_2(t_j) \text{ for } j = i+1, \ldots, m] \\ &\quad = \sum_{x,y} \psi(x)\psi(y) \sum_{i=1}^{m} \sum_{w} P^{x}[X_1(t_i) = w]P^{y}[X_2(t_i) = w] \\ &\qquad\qquad \times P^{(w,w)}[X_1(t_j - t_i) \neq X_2(t_j - t_i) \text{ for } j = i+1, \ldots, m] \\ &\quad = \sum_{i=1}^{m} \sum_{w} \left(\sum_{x} \psi(x)P^{x}[X_1(t_i) = w] \right)^2 \\ &\qquad\qquad \times P^{(w,w)}[X_1(t_j - t_i) \neq X_2(t_j - t_i) \text{ for } j = i+1, \ldots, m] \\ &\quad \geq 0. \end{aligned}$$

We have shown that $w(x,y)$ is positive definite. It is also bounded and symmetric. For $\mathbf{x} \in S^n$, let

$$G(\mathbf{x}) = \sum_{1 \le i < j \le n} w(x_i, x_j).$$

G inherits the properties of boundedness, symmetry and positive definiteness from w.

Again use the fact that by the symmetry of $p(x,y)$, $Z(t) = X_1(t) - X_2(t)$ is a random walk with twice the rates of $X_1(t)$ (Exercise 1.5). By the transience assumption $Z(t)$ has a last visit to the origin almost surely, and so for all $x, y \in S$,

$$\lim_{t\to\infty} P^{(x,y)}[X_1(s) = X_2(s) \text{ for some } s > t] = 0. \tag{5.16}$$

Recalling that $U(t)$ represented the semigroup of $\mathbf{X}(t)$,

$$\begin{aligned} U(t)G(\mathbf{x}) &= E^{\mathbf{x}} G(\mathbf{X}(t)) \\ &= \sum_{1\le i<j\le n} E^{\mathbf{x}} w(X_i(t), X_j(t)) \\ &= \sum_{1\le i<j\le n} E^{\mathbf{x}}[P^{(X_i(t),X_j(t))}\{X_i(s) = X_j(s) \text{ for some } s > 0\}] \\ &= \sum_{1\le i<j\le n} P^{(x_i,x_j)}\{X_i(s) = X_j(s) \text{ for some } s > t\}. \end{aligned}$$

By (5.16) we conclude that

$$\lim_{t\to\infty} U(t)G(\mathbf{x}) = 0 \quad \text{for } \mathbf{x} \in S^n. \tag{5.17}$$

By Proposition 5.8 $V(t)G \le U(t)G$ on T, so

$$\lim_{t\to\infty} V(t)G(\mathbf{x}) = 0 \quad \text{for } \mathbf{x} \in T. \tag{5.18}$$

The conclusions of the lemma now follow from (5.17), (5.18) and the observation that $g(\mathbf{x}) \le G(\mathbf{x})$. ∎

Proposition 5.11 *Suppose f is a bounded function on T such that $V(t)f = f$ for all $t \ge 0$. Then f must be constant.*

Proof. For $n = 1$ the exclusion is just the irreducible random walk, and we already know bounded harmonic functions of an irreducible random walk are constant. So assume $n \ge 2$. Multiplication by a constant and adding a constant do not change boundedness or being harmonic, so we may assume $0 \le f \le 1$. We may couple $\mathbf{X}(t)$ and the exclusion process so that they agree until the first time $\mathbf{X}(t)$ leaves T. (Detailed specification of this coupling left as an exercise.) Consequently

$$|V(t)f(\mathbf{x}) - U(t)f(\mathbf{x})| \leq g(\mathbf{x}) \text{ for } \mathbf{x} \in T. \tag{5.19}$$

By assumption $V(t)f = f$ on T, and so

$$|f(\mathbf{x}) - U(t)f(\mathbf{x})| \leq g(\mathbf{x}) \text{ for } \mathbf{x} \in T. \tag{5.20}$$

Extend f to S^n by setting $f = 0$ on $S^n \setminus T$. Since $g = 1$ on $S^n \setminus T$, the above inequality extends:

$$|f(\mathbf{x}) - U(t)f(\mathbf{x})| \leq g(\mathbf{x}) \text{ for } \mathbf{x} \in S^n. \tag{5.21}$$

Next, for $s, t \geq 0$,

$$|U(s)f(\mathbf{x}) - U(s+t)f(\mathbf{x})| \leq U(s)|f - U(t)f|(\mathbf{x}) \leq U(s)g(\mathbf{x}), \ \mathbf{x} \in S^n.$$

Letting $s \to \infty$ and using Lemma 5.10(a) shows that the limit of $U(s)f(\mathbf{x})$ exists as $s \to \infty$, for all $\mathbf{x} \in S^n$. This limit is harmonic for the random walk by Lemma 1.10, hence constant because $\mathbf{X}(t)$ is an irreducible random walk on $\mathbf{Z}^{dn}$. So for some constant $b \in [0,1]$, $U(t)f(\mathbf{x}) \to b$ as $t \to \infty$. Pass to this limit in (5.21) to get

$$|f(\mathbf{x}) - b| \leq g(\mathbf{x}) \text{ for } \mathbf{x} \in S^n. \tag{5.22}$$

Again since f is assumed harmonic for $V(t)$, for $\mathbf{x} \in T$ we have

$$|f(\mathbf{x}) - b| = |V(t)f(\mathbf{x}) - b| = |V(t)(f-b)(\mathbf{x})| \leq V(t)|f-b|(\mathbf{x}) \leq V(t)g(\mathbf{x}).$$

Finally, let $t \to \infty$ and apply Lemma 5.10(b) to conclude that $f = b$ on T. ∎

Corollary 5.12 *Suppose h is a bounded function on Y, and harmonic for the finite exclusion process A_t. Then $h(A)$ depends only on $|A|$.*

Proof. Let $A, B \in Y$ be such that $|A| = |B| = n$. We want to show that $h(A) = h(B)$. For $\mathbf{x} = (x_1, \ldots, x_n) \in T$, define $f(\mathbf{x}) = h(\{x_1, \ldots, x_n\})$. By the harmonicity of h and (5.13), $V(t)f = f$, and by Proposition 5.11 f must be constant. ∎

We have proved that a bounded harmonic function for the finite exclusion can depend only on the number of particles. By Proposition 5.6, we have proved Theorem 5.1 for the transient case.

5.5 Proof of Theorem 5.2

In the proof of Theorem 5.2 we can now handle both the recurrent and the transient case simultaneously.

Let μ be a spatially invariant and ergodic probability measure on X, and $\rho = \mu\{\eta(x) = 1\}$ the density of μ. To prove $\mu_t \to \nu_\rho$ as $t \to \infty$, it suffices to show that an arbitrary sequence $t_j \nearrow \infty$ has a subsequence t_{j_k} such that $\mu_{t_{j_k}} \to \nu_\rho$. This suffices because if μ_t were not to converge to ν_ρ weakly, there would have to exist a bounded continuous function ψ on X, a sequence of times $t_j \nearrow \infty$, and $\delta > 0$ such that for all j,

$$\left| \int_X \psi \, d\mu_{t_j} - \int_X \psi \, d\nu_\rho \right| > \delta.$$

However, this would be contradicted along $\{j_k\}$ if $\mu_{t_{j_k}} \to \nu_\rho$.

So let $t_j \nearrow \infty$ be an arbitrary sequence of times. By the compactness of the space $\mathcal{M}_1$ of probability measures on X, we can pick a subsequence, which we denote by $\{t_k\}$, such that a weak limit $\mu_{t_k} \to \nu$ exists. We need to show that $\nu = \nu_\rho$.

For any $A \in Y$,

$$\widehat{\nu}(A) = \nu\{\eta = 1 \text{ on } A\} = \lim_{k\to\infty} \mu_{t_k}\{\eta = 1 \text{ on } A\} = \lim_{k\to\infty} \widehat{\mu_{t_k}}(A) = \lim_{k\to\infty} E^A \widehat{\mu}(A_{t_k})$$

where the last equality follows from duality (5.4). By Lemma 1.10, $\widehat{\nu}$ is harmonic for A_t. Consequently $\widehat{\nu}(A)$ depends only on $|A|$. For the recurrent case this comes from Proposition 5.7, and for the transient case from Corollary 5.12. By Exercise A.6 in Section A.6, ν is an exchangeable measure. By de Finetti's theorem A.16, there exists a probability measure γ on $[0, 1]$ such that

$$\nu = \int_{[0,1]} \nu_\alpha \, \gamma(d\alpha).$$

It remains to show that $\gamma = \delta_\rho$, the point mass at ρ. Cylinder functions are continuous on X, hence their expectations converge under weak convergence.

$$\begin{aligned} \nu\{\eta(0) = 1\} &= \lim_{k\to\infty} \mu_{t_k}\{\eta(0) = 1\} = \lim_{k\to\infty} \widehat{\mu_{t_k}}(\{0\}) \\ &= \lim_{k\to\infty} E^{\{0\}} \widehat{\mu}(A_{t_k}) = \lim_{k\to\infty} E^0 \widehat{\mu}\{X(t_k)\} = \rho, \end{aligned}$$

where we used duality, represented the one-particle exclusion started from $A_0 = \{0\}$ by the random walk $X(t)$, and noted that $\widehat{\mu}\{x\} = \rho$ for any singleton $\{x\}$. This implies

$$\int_{[0,1]} \alpha \, \gamma(d\alpha) = \rho. \tag{5.23}$$

Similarly for $x \neq 0$, by duality, Corollary 5.9, Fubini's theorem, and Corollary A.23,

$$\begin{aligned} &\nu\{\eta(0) = \eta(x) = 1\} = \lim_{k\to\infty} \widehat{\mu_{t_k}}(\{0, x\}) \\ \leq\ & \lim_{k\to\infty} E^{(0,x)} \widehat{\mu}\{X_1(t_k), X_2(t_k)\} = \lim_{k\to\infty} E^{(0,x)} \mu\{\eta(X_1(t_k)) = \eta(X_2(t_k)) = 1\} \\ =\ & \lim_{k\to\infty} \int E^{(0,x)}[\eta(X_1(t_k))\eta(X_2(t_k))]\, \mu(d\eta) = \rho^2. \end{aligned}$$

This says

$$\int_{[0,1]} \alpha^2\, \gamma(d\alpha) \leq \rho^2. \tag{5.24}$$

By (5.23), (5.24), and Schwarz inequality,

$$\rho = \int \alpha\, \gamma(d\alpha) \leq \left(\int \alpha^2\, \gamma(d\alpha) \right)^{1/2} \leq \rho.$$

Equality in the Schwarz inequality forces the function to be almost surely constant. Thus γ is concentrated on the point ρ. The proof of Theorem 5.2 is complete.

Exercise 5.1 Construct explicitly the coupling used in the proof of Proposition 5.11 between the random walk $\mathbf{X}(t)$ and the finite exclusion process.

Notes

This chapter came from section VIII.1 in Liggett's monograph [27] and from Chapter 3 in Liggett's lectures [26].

6 Equilibrium distributions without symmetry assumptions

6.1 Equilibrium product distributions

In this section we drop the symmetry assumption, and $p(x, y)$ satisfies the assumptions we used for constructing the exclusion process, namely translation invariance $p(x, y) = p(0, y-x)$ and finite range.

For a function $\alpha : S \to [0, 1]$, let ν_α be the product probability measure on X defined by

$$\nu_\alpha\{\eta : \eta(x) = 1 \text{ for all } x \in A,\, \eta(y) = 0 \text{ for all } y \in B\} = \prod_{x\in A} \alpha(x) \prod_{y\in B}(1 - \alpha(y)) \qquad (6.1)$$

for any two disjoint finite sets of sites A and B. The case of constant $\alpha(x) = \rho$ is the Bernoulli measure ν_ρ defined earlier in (4.15).

Theorem 6.1 (a) *Suppose $\pi(x)$ is a positive function on S that satisfies $\pi(x)p(x, y) = \pi(y)p(y, x)$ for all $x, y \in S$. Let $\alpha(x) = \pi(x)/(1 + \pi(x))$. Then $\nu_\alpha \in \mathcal{I}$.*

(b) *The Bernoulli measures are invariant for the exclusion process with any translation invariant $p(x, y)$.*

Proof. By Theorem 4.4, we need to check that $\int Lf\, d\nu_\alpha = 0$ for an arbitrary cylinder function f. By the finite range assumption on $p(x, y)$, we can fix a finite set $A \subseteq S$ such that

$$Lf(\eta) = \sum_{x,y\in A} p(x, y)\eta(x)(1 - \eta(y))[f(\eta^{x,y}) - f(\eta)],$$

and then

$$\begin{aligned}\int Lf\, d\nu_\alpha \;&=\; \sum_{x,y\in A} p(x, y) \int \eta(x)(1 - \eta(y))f(\eta^{x,y})\, \nu_\alpha(d\eta)\\ &\qquad - \sum_{x,y\in A} p(x, y) \int \eta(x)(1 - \eta(y))f(\eta)\, \nu_\alpha(d\eta).\end{aligned}$$

Separate one (x, y)-term from the first sum above. To manipulate the coordinates explicitly, introduce the notation $\eta = (\eta', \eta(x), \eta(y))$ where η' contains all the coordinates outside $\{x, y\}$. Write ν'_α for the marginal distribution of η', and ν^x_α for the marginal distribution of $\eta(x)$. Since ν_α is a product measure, we can integrate separately over distinct coordinates, and reason as follows.

$$\begin{aligned}
&\int \nu_\alpha(d\eta)\,\eta(x)(1-\eta(y))f(\eta^{x,y})\\
&\quad= \int \nu'_\alpha(d\eta') \int \nu^x_\alpha(d\eta(x)) \int \nu^y_\alpha(d\eta(y))\,\eta(x)(1-\eta(y))f(\eta',\eta(y),\eta(x))\\
&\quad= \int \nu'_\alpha(d\eta')\,\alpha(x)(1-\alpha(y))\,f(\eta',0,1)\\
&\quad= \frac{\alpha(x)(1-\alpha(y))}{\alpha(y)(1-\alpha(x))}\int \nu'_\alpha(d\eta')\,\alpha(y)(1-\alpha(x))\,f(\eta',0,1)\\
&\quad= \frac{\alpha(x)(1-\alpha(y))}{\alpha(y)(1-\alpha(x))}\int \nu_\alpha(d\eta)\,\eta(y)(1-\eta(x))f(\eta).
\end{aligned}$$

Substitute this back above, combine the sums and rearrange terms to get

$$\begin{aligned}
\int Lf\,d\nu_\alpha = \int \nu_\alpha(d\eta)\,f(\eta) \sum_{x,y\in A} p(x,y)\\
\times\left\{\frac{\alpha(x)(1-\alpha(y))}{\alpha(y)(1-\alpha(x))}\eta(y)(1-\eta(x)) \;-\; \eta(x)(1-\eta(y))\right\}.
\end{aligned} \tag{6.2}$$

In case (a) the sum inside the integral equals

$$\begin{aligned}
&\sum_{x,y\in A} p(x,y)\frac{\pi(x)}{\pi(y)}\eta(y)(1-\eta(x)) \;-\; \sum_{x,y\in A} p(x,y)\eta(x)(1-\eta(y))\\
&\quad= \sum_{x,y\in A} p(y,x)\eta(y)(1-\eta(x)) \;-\; \sum_{x,y\in A} p(x,y)\eta(x)(1-\eta(y)) \;=\; 0
\end{aligned}$$

where the last equality follows because the sums differ only by a relabeling of the summation indices. This proves case (a).

For case (b) we have to do someting different to get the sum in (6.2) to vanish. We restrict the process to a finite cube in S with periodic boundary conditions. This means that jumps of particles out of the cube are directed to a site inside the cube, and jumps from outside the cube are completely eliminated. Fix a positive integer k, and let A be the cube $A = [-k,k]^d \cap \mathbf{Z}^d$. Define the new jump probability for $x, y \in A$ by

$$p_A(x,y) = \sum_{w\in\mathbf{Z}^d} p(x, y + (2k+1)w).$$

Note that $\{y+(2k+1)\mathbf{Z}^d : y \in A\}$ partitions $\mathbf{Z}^d$ into disjoint subsets. This gives $\sum_{y\in A} p_A(x,y) = 1$ so that $p_A(x,y)$ is a well-defined jump probability on A. Using translation invariance, for any $z \in A$,

$$\sum_{x\in A} p_A(x,z) = \sum_{x\in A}\sum_{w\in\mathbf{Z}^d} p(x, z+(2k+1)w) = \sum_{x\in A}\sum_{w\in\mathbf{Z}^d} p(0, z-x+(2k+1)w) = 1.$$

The last equality follows because, by the symmetry $A = -A$, $\{-x + (2k+1)\mathbf{Z}^d : x \in A\}$ is also a partition of $\mathbf{Z}^d$, and this is not changed by translating each set by z.

Fix the cube A large enough so that $p_A(x,y) = p(x,y)$ for any x, y such that $f(\eta^{x,y}) \neq f(\eta)$, so that

$$Lf(\eta) = \sum_{x,y\in A} p_A(x,y)\eta(x)(1-\eta(y))[f(\eta^{x,y}) - f(\eta)].$$

We can repeat the earlier calculations down to (6.2) with $p(x,y)$ replaced by $p_A(x,y)$. Now $\alpha(x) = \rho$, and we get

$$\begin{aligned}
&\int Lf\, d\nu_\alpha \\
&= \int \nu_\alpha(d\eta)\, f(\eta) \sum_{x,y\in A} p_A(x,y) \left\{ \frac{\alpha(x)(1-\alpha(y))}{\alpha(y)(1-\alpha(x))}\eta(y)(1-\eta(x)) \; - \; \eta(x)(1-\eta(y)) \right\} \\
&= \int \nu_\alpha(d\eta)\, f(\eta) \sum_{x,y\in A} p_A(x,y) \left\{\eta(y) - \eta(x)\right\} \\
&= \int \nu_\alpha(d\eta)\, f(\eta) \left\{ \sum_{y\in A}\eta(y)\sum_{x\in A} p_A(x,y) \; - \; \sum_{x\in A}\eta(x)\sum_{y\in A} p_A(x,y) \right\} \\
&= 0.
\end{aligned}$$

This proves part (b). ∎

The condition $\pi(x)p(x,y) = \pi(y)p(y,x)$ says that π is *reversible* for the transition probability $p(x,y)$. Suppose π is a probability measure on S. Then π is invariant for the Markov chain with transition $p(x,y)$, and furthermore the process with marginal distribution π is reversible, meaning that $\{X_{-n}\}$ has the same distribution as $\{X_n\}$.

A transition matrix that satisfies $\sum_x p(x,y) = 1$ for all y is called *doubly stochastic.* This is the hypothesis needed for part (b) of the above theorem. Translation invariance is stronger.

Example 6.2 Consider the one-dimensional nearest-neighbor exclusion process with jump kernel $p(x,x+1) = p$, $p(x,x-1) = q = 1-p$, and $p(x,y) = 0$ for $y \neq x \pm 1$. If $p = 1/2$ this process is symmetric, and we already know that the Bernoulli measures are all the extremal invariant measures. If $p \neq 1/2$ this is an *asymmetric* process. By part (b) of the above theorem, the Bernoulli measures are again invariant. But now part (a) also gives something nontrivial. One can check that $\pi(x) = c(p/q)^x$ is reversible for $p(x,y)$, where c is any constant. By the theorem, the product measure ν_α with $\alpha(x) = cp^x/(cp^x + q^x)$ is invariant. Thus interestingly, even though the transition mechanism for this exclusion process is translation invariant, it has invariant measures that are not translation invariant, because under ν_α the distribution of the coordinate $\eta(x)$ varies with x.

6.2 Translation invariant equilibrium distributions

In Theorem 5.1 we characterized the entire set $\mathcal{I}$ of invariant distributions for symmetric exclusion processes. In this section we shall do less for the general translation invariant finite range exclusion process. Namely, we characterize the set $\mathcal{I} \cap \mathcal{S}$ of translation invariant equilibrium distributions. The example of the p, q-asymmetric exclusion above shows that there can be nonstationary equilibrium distributions, so the result is incomplete.

The duality arguments of the symmetric case are not applicable now. The main technique in this section is coupling of two exclusion processes. We already encountered this briefly in the proof of Lemma 4.6. The *basic coupling* means that two exclusion processes η_t and ζ_t are constructed as explained in Section 2.1, with one set of Poisson processes $\{\mathcal{T}_{(x,y)}\}$ that govern the jump attempts of both processes. The effect of this is that individually η_t and ζ_t are both exclusion processes, but their jump attempts are synchronized. At a jump time in Poisson process $\mathcal{T}_{(x,y)}$, the state (η, ζ) transforms into $(\eta^{x,y}, \zeta)$, $(\eta, \zeta^{x,y})$, or $(\eta^{x,y}, \zeta^{x,y})$, depending on whether an η-particle, a ζ-particle, or both can jump from x to y at time t. By considering the different possibilities, one sees that in this coupling an η-particle and a ζ-particle jump together whenever possible.

This construction defines a process (η_t, ζ_t) with state space X^2, path space D_{X^2}, and all the properties proved in Section 2.3. Its generator is

$$
\begin{aligned}
\widetilde{L}f(\eta,\zeta) &= \sum_{x,y\in S} p(x,y)\,\eta(x)(1-\eta(y))\zeta(x)(1-\zeta(y))\,[f(\eta^{x,y},\zeta^{x,y}) - f(\eta,\zeta)] \\
&\quad + \sum_{x,y\in S} p(x,y)\,\eta(x)(1-\eta(y))\,\mathbf{1}\{\zeta(x)=0 \text{ or } \zeta(y)=1\}\,[f(\eta^{x,y},\zeta) - f(\eta,\zeta)] \\
&\quad + \sum_{x,y\in S} p(x,y)\,\zeta(x)(1-\zeta(y))\,\mathbf{1}\{\eta(x)=0 \text{ or } \eta(y)=1\}\,[f(\eta,\zeta^{x,y}) - f(\eta,\zeta)].
\end{aligned}
\tag{6.3}
$$

Write $\widetilde{S}(t)$ for the semigroup of the process (η_t, ζ_t).

We define a partial order on X in a coordinatewise fashion, by $\eta \geq \zeta$ iff $\eta(x) \geq \zeta(x)$ for all $x \in S$. Relative to this order, the basic coupling has an important monotonicity property: if initially $\eta_0 \geq \zeta_0$, then $\eta_t \geq \zeta_t$ holds with probability one for all time $t \geq 0$. (Proof by considering the possible effects of a jump.) This property is expressed by saying that the exclusion process is *attractive*. See Section A.3 for discussion of this order relation and its definition for measures on X.

Let us write $\widetilde{\mathcal{I}}$ for the set of invariant probability measures of the process (η_t, ζ_t) constructed by the basic coupling. And similarly, $\widetilde{\mathcal{I}}_e$ for the extremal members of $\widetilde{\mathcal{I}}$, and $\widetilde{\mathcal{S}}$ for the translation-invariant measures on the space $X^2 = (\{0,1\} \times \{0,1\})^{\mathbf{Z}^d}$.

It is instructive to consider how discrepancies behave in the basic coupling. Let us say there is a positive discrepancy at site x at time t if $\eta_t(x) > \zeta_t(x)$, and a negative discrepancy if $\eta_t(x) < \zeta_t(x)$. As the processes evolve under the coupling, discrepancies move around but are never created. Discrepancies of the opposite type annihilate each other when they land on the same site. To see this, consider the effect of a jump time $t \in \mathcal{T}_{(x,y)}$ on a state (η_{t-}, ζ_{t-}) such that $(\eta_{t-}(x), \eta_{t-}(y)) = (1,0)$, $(\zeta_{t-}(x), \zeta_{t-}(y)) = (0,1)$. After the jump the processes agree on $\{x, y\}$.

Suppose we make an irreducibility assumption which enables the discrepancies to mix around. Then it is reasonable to expect that in the long run discrepancies of the opposite type cannot coexist. In particular, an invariant measure should not give such an occurrence positive probability. This we prove now.

Assume that $p(x,y)$ is translation invariant, finite range, and has this irreducibility property:

$$\text{for all } (x,y) \text{ there exists } n \text{ such that } p^{(n)}(x,y) + p^{(n)}(y,x) > 0. \tag{6.4}$$

This property includes examples such as $p(x, x+1) = 1$ where not all spatial directions are permitted for jumps.

Proposition 6.3 *Let $\widetilde{\nu} \in \widetilde{\mathcal{I}} \cap \widetilde{\mathcal{S}}$. Then for all sites $x, y \in S$,*

$$\widetilde{\nu}\{(\eta,\zeta) : \eta(x) = \zeta(y) = 0,\ \eta(y) = \zeta(x) = 1\} = 0.$$

Proof. Fix x, and apply the generator $\widetilde{L}$ to the cylinder function $f(\eta,\zeta) = \mathbf{1}\{\eta(x) \neq$

$\zeta(x)\}$.

$$\begin{aligned}\widetilde{L}f(\eta,\zeta) &= \sum_y p(x,y)\mathbf{1}\{\eta(x)=\zeta(x)=1,\, \eta(y)\neq\zeta(y)\} \\ &\quad +\sum_y p(y,x)\mathbf{1}\{\eta(x)=\zeta(x)=0,\, \eta(y)\neq\zeta(y)\} \\ &\quad -\sum_y p(x,y)\mathbf{1}\{\eta(x)\neq\zeta(x),\, \eta(y)=\zeta(y)=0\} \\ &\quad -\sum_y p(y,x)\mathbf{1}\{\eta(x)\neq\zeta(x),\, \eta(y)=\zeta(y)=1\} \\ &\quad -\sum_y (p(x,y)+p(y,x))\mathbf{1}\{\eta(x)=\zeta(y)\neq\eta(y)=\zeta(x)\}.\end{aligned}$$

The plus terms contain the ways in which a discrepancy can be moved to x. For example, in the first sum x is occupied for both η and ζ before the jump, and then at rate $p(x,y)$ a particle is moved from x to y in either η or ζ but not in both. The two first minus terms count the ways a discrepancy can be moved out of x. The last term counts the ways a discrepancy at x can be annihilated.

Now take expectation under $\widetilde{\nu}$. Since $\widetilde{\nu}\in\widetilde{\mathcal{I}}$, $\int \widetilde{L}f\, d\widetilde{\nu} = 0$, so the expectation of the left-hand side vanishes. On the right-hand side, use the translation invariance of $p(x,y)$ and $\widetilde{\nu}$. Then the expectation of the first term equals

$$\begin{aligned}&\sum_y p(x,y)\widetilde{\nu}\{\eta(x)=\zeta(x)=1,\, \eta(y)\neq\zeta(y)\} \\ &\quad =\sum_y p(0,y-x)\widetilde{\nu}\{\eta(0)=\zeta(0)=1,\, \eta(y-x)\neq\zeta(y-x)\} \\ &\quad =\sum_z p(0,z)\widetilde{\nu}\{\eta(0)=\zeta(0)=1,\, \eta(z)\neq\zeta(z)\}.\end{aligned}$$

Similarly, the expectation of the fourth term equals

$$\begin{aligned}&\sum_y p(y,x)\widetilde{\nu}\{\eta(x)\neq\zeta(x),\, \eta(y)=\zeta(y)=1\} \\ &\quad =\sum_y p(0,x-y)\widetilde{\nu}\{\eta(x-y)\neq\zeta(x-y),\, \eta(0)=\zeta(0)=1\} \\ &\quad =\sum_z p(0,z)\widetilde{\nu}\{\eta(z)\neq\zeta(z),\, \eta(0)=\zeta(0)=1\}.\end{aligned}$$

We see that the expectations of the first and fourth term cancel each other on the right-hand side. Similarly, the expectations of the second and third term cancel each other.

We conclude from all this that for every pair (x, y) such that $p(x, y) + p(y, x) > 0$,

$$\widetilde{\nu}\{\eta(x) = \zeta(y) \neq \eta(y) = \zeta(x)\} = 0. \tag{6.5}$$

By the assumption (6.4), to finish the proof we show by induction the following statement: (6.5) holds for any pair (x, y) such that $p^{(m)}(x, y) + p^{(m)}(y, x) > 0$ for some $1 \le m \le n$. We have already proved the $n = 1$ case. Let us assume that the result is true for $n - 1$, and prove it for n.

Let us employ the following shorthand for events involving η and ζ. For any sites $x_1, \dots, x_m$, and any vectors $(a_1, \dots, a_m)$ and $(b_1, \dots, b_m)$ of 0's and 1's,

$$\begin{aligned} &\begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{bmatrix} \\ &= [\, x_1\, x_2 \,\cdots\, x_m \mid a_1\, a_2 \,\cdots\, a_m \mid b_1\, b_2 \,\cdots\, b_m \,] \\ &= \{(\eta, \zeta) : \eta(x_1) = a_1, \eta(x_2) = a_2, \dots, \eta(x_m) = a_m, \\ &\qquad \zeta(x_1) = b_1, \zeta(x_2) = b_2, \dots, \zeta(x_m) = b_m\}. \end{aligned}$$

In this notation, the induction assumption is that

$$\widetilde{\nu} \begin{bmatrix} x & y \\ 1 & 0 \\ 0 & 1 \end{bmatrix} + \widetilde{\nu} \begin{bmatrix} x & y \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = 0$$

for any pair (x, y) such that $p^{(m)}(x, y) + p^{(m)}(y, x) > 0$ for some $1 \le m \le n - 1$.

Now for the induction step. Suppose $p^{(n)}(x, y) + p^{(n)}(y, x) > 0$. Let us suppose $p^{(n)}(x, y) > 0$. Otherwise exchange x and y, which is permissible because the event in (6.5) is not affected by such an exchange.

We shall show $\widetilde{\nu}[\, x\ y \mid 1\ 0 \mid 0\ 1 \,] = 0$ and leave the analogous argument for $\widetilde{\nu}[\, x\ y \mid 0\ 1 \mid 1\ 0 \,] = 0$ as an exercise. Find states $x = x_0, x_1, \dots, x_n = y$ such that $p(x_i, x_{i+1}) > 0$ for $i = 0, \dots, n-1$. Let $\mathbf{a} = (a_1, \dots, a_{n-1})$ and $\mathbf{b} = (b_1, \dots, b_{n-1})$ represent $(n-1)$-tuples of 0's and 1's. Decompose the event $[\, x\ y \mid 1\ 0 \mid 0\ 1 \,]$ as

$$\begin{bmatrix} x & y \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \bigcup_{\mathbf{a}, \mathbf{b}} \begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} & x_n \\ 1 & a_1 & \cdots & a_{n-1} & 0 \\ 0 & b_1 & \cdots & b_{n-1} & 1 \end{bmatrix}.$$

By the induction assumption, any term with some $(a_i, b_i) = (1, 0)$ or $(0, 1)$ will have zero $\widetilde{\nu}$ measure. Let V be the set of pairs $(\mathbf{a}, \mathbf{b})$ such that $(a_i, b_i) = (0, 0)$ or $(1, 1)$ for each i. Then

$$\widetilde{\nu} \begin{bmatrix} x & y \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \sum_{(\mathbf{a}, \mathbf{b}) \in V} \widetilde{\nu} \begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} & x_n \\ 1 & a_1 & \cdots & a_{n-1} & 0 \\ 0 & b_1 & \cdots & b_{n-1} & 1 \end{bmatrix}.$$

It remains to show that each term on the right is zero. Consider first the case where $(a_1, b_1) = (0,0)$. Abbreviate

$$A = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_{n-1} & x_n \\ 1 & 0 & a_2 & \cdots & a_{n-1} & 0 \\ 0 & 0 & b_2 & \cdots & b_{n-1} & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_{n-1} & x_n \\ 0 & 1 & a_2 & \cdots & a_{n-1} & 0 \\ 0 & 0 & b_2 & \cdots & b_{n-1} & 1 \end{bmatrix}.$$

Event B results from A after a jump from x_0 to x_1. Let G be the event that in time $(0,t]$, there is one jump time in $\mathcal{T}_{(x_0,x_1)}$, and no jump times in any other $\mathcal{T}_{(u,v)}$ such that either u or v is among $x_0, \dots, x_n$. Note two things:

$$\{(\eta_0, \zeta_0) \in A\} \cap G \subseteq \{(\eta_t, \zeta_t) \in B\},$$

and

$$\mathbf{P}(G) = p(x_0,x_1) t e^{-p(x_0,x_1)t} \cdot e^{-\beta t} > 0$$

where

$$\beta = \sum_{i=0}^{n} \sum_{y} p(x_i, y) + \sum_{y \notin \{x_i\}} \sum_{j=0}^{n} p(y, x_j) - p(x_0, x_1).$$

Now estimate

$$\widetilde{\nu}(B) = \widetilde{\nu}\widetilde{S}(t)(B) \geq \mathbf{P}^{\widetilde{\nu}}(\{(\eta_0,\zeta_0) \in A\} \cap G) = \widetilde{\nu}(A)\, \mathbf{P}(G).$$

Above we used first the invariance of $\widetilde{\nu}$. Then we wrote $\mathbf{P}^{\widetilde{\nu}}$ for the probability measure on the probability space where the processes (η_t, ζ_t) and the Poisson processes are defined. Finally, it is a property of the construction that the Poisson jump time processes $\{\mathcal{T}_{(u,v)}\}$ are independent of the initial condition (η_0, ζ_0).

By induction $\widetilde{\nu}(B) = 0$. Since $\mathbf{P}(G) > 0$, we must have $\widetilde{\nu}(A) = 0$.

The remaining case is of the type

$$A = \begin{bmatrix} x_0 & x_1 & \cdots & x_{k-1} & x_k & x_{k+1} & \cdots & x_{n-1} & x_n \\ 1 & a_1 & \cdots & a_{k-1} & 1 & 0 & \cdots & 0 & 0 \\ 0 & b_1 & \cdots & b_{k-1} & 1 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

for some $k < n$. Let

$$B = \begin{bmatrix} x_0 & x_1 & \cdots & x_{k-1} & x_k & x_{k+1} & \cdots & x_{n-1} & x_n \\ 1 & a_1 & \cdots & a_{k-1} & 0 & 0 & \cdots & 0 & 1 \\ 0 & b_1 & \cdots & b_{k-1} & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}.$$

To turn A into B, let G be the event that clocks $\mathcal{T}_{(x_k,x_{k+1})}$, $\mathcal{T}_{(x_{k+1},x_{k+2})}$, $\dots$, $\mathcal{T}_{(x_{n-1},x_n)}$ ring in order, and no other clocks that affect $x_0, \dots, x_n$ ring during a fixed time interval $(0,t]$. Repeat the earlier argument. ∎

Corollary 6.4 *Let* $\widetilde{\nu} \in \widetilde{\mathcal{I}} \cap \widetilde{\mathcal{S}}$. *Then* $\widetilde{\nu}(\{\eta \geq \zeta\} \cup \{\zeta \geq \eta\}) = 1$.

Proof. A configuration (η, ζ) outside $\{\eta \geq \zeta\} \cup \{\zeta \geq \eta\}$ has discrepancies of both types. But

$$\begin{aligned}&\widetilde{\nu}\{\text{discrepancies of both type coexist}\}\\ &\qquad \leq \sum_{x,y \in S} \widetilde{\nu}\{\eta(x) = \zeta(y) = 0,\, \eta(y) = \zeta(x) = 1\} = 0\end{aligned}$$

by Proposition 6.3. ∎

Next we consider starting the joint process (η_t, ζ_t) with a translation invariant initial probability distribution $\widetilde{\gamma}$ on the space X^2. It is important that the distribution $\widetilde{\gamma}\widetilde{S}(t)$ at later times retains this translation invariance.

Lemma 6.5 *Suppose* $\widetilde{\gamma}$ *is a translation invariant probability measure on* X^2. *Then* $\widetilde{\gamma}\widetilde{S}(t)$ *is translation invariant for each* $t \geq 0$.

Proof. Think of the initial configurations and the Poisson point processes together as a process indexed by $x \in \mathbf{Z}^d$ in this sense:

$$(\eta_0, \zeta_0, \omega) = \{\eta_0(x), \zeta_0(x), (\mathcal{T}_{(x,y)} : y \in S, p(x,y) > 0) : x \in \mathbf{Z}^d\}.$$

The distribution of this process is invariant under translations of x.

Let g be the map that constructs the occupation numbers at the origin at time t from the initial configurations and the Poisson point processes: $(\eta_t(0), \zeta_t(0)) = g(\eta, \zeta, \omega)$. Then $(\eta_t(x), \zeta_t(x)) = g(\theta_x \eta, \theta_x \zeta, \theta_x \omega)$. By Lemma A.10 the process $\{\eta_t(x), \zeta_t(x) : x \in \mathbf{Z}^d\}$ is stationary. ∎

Lemma 6.6 *Let* $\mu, \nu \in (\mathcal{I} \cap \mathcal{S})_e$. *Then there exists* $\widetilde{\nu} \in (\widetilde{\mathcal{I}} \cap \widetilde{\mathcal{S}})_e$ *with marginals* μ *and* ν.

Proof. Let $\widetilde{\gamma} = \mu \otimes \nu$, a translation invariant probability measure on X^2. The space of probability measures on X^2 is compact, by the compactness of X^2. Hence there is a sequence $t_n \nearrow \infty$ along which the probability measures

$$t_n^{-1} \int_0^{t_n} \widetilde{\gamma}\widetilde{S}(t)\, dt$$

converge weakly to a measure $\widetilde{\mu}$.

Since μ and ν are invariant for the exclusion process, $\widetilde{\gamma}\widetilde{S}(t)$ has marginals μ and ν for all $t \geq 0$. By Lemma 6.5 above $\widetilde{\gamma}\widetilde{S}(t)$ is translation invariant for all $t \geq 0$. Both these properties are preserved by time averaging and weak convergence, so $\widetilde{\mu}$ has them also.

The invariance of $\widetilde{\mu}$ under the semigroup comes from the fact that the process is Feller continuous. For then the operation $\widetilde{S}(s)$ is weakly continuous on measures, and we get

$$\begin{aligned}\widetilde{\mu}\widetilde{S}(s) &= \left(\lim_{n\to\infty} t_n^{-1}\int_0^{t_n} \widetilde{\gamma}\widetilde{S}(t)\,dt\right)\widetilde{S}(s) = \lim_{n\to\infty}\left\{\left(t_n^{-1}\int_0^{t_n} \widetilde{\gamma}\widetilde{S}(t)\,dt\right)\widetilde{S}(s)\right\}\\ &= \lim_{n\to\infty} t_n^{-1}\int_0^{t_n} \widetilde{\gamma}\widetilde{S}(s+t)\,dt = \lim_{n\to\infty} t_n^{-1}\int_s^{s+t_n} \widetilde{\gamma}\widetilde{S}(t)\,dt\\ &= \lim_{n\to\infty} t_n^{-1}\int_0^{t_n} \widetilde{\gamma}\widetilde{S}(t)\,dt = \widetilde{\mu}.\end{aligned}$$

Thus $\widetilde{\mu}$ is an invariant probability measure for the joint process.

We now know $\widetilde{\mu} \in \widetilde{\mathcal{I}} \cap \widetilde{\mathcal{S}}$. If $\widetilde{\mu}$ is an extreme point of this set, we are done and we can take $\widetilde{\nu} = \widetilde{\mu}$. Otherwise by Corollary A.14 there is a probability measure Γ on $(\widetilde{\mathcal{I}} \cap \widetilde{\mathcal{S}})_e$ such that

$$\widetilde{\mu} = \int_{(\widetilde{\mathcal{I}}\cap\widetilde{\mathcal{S}})_e} \widetilde{\nu}\,\Gamma(d\widetilde{\nu}).$$

Let $\widetilde{\nu}_1, \widetilde{\nu}_2$ be the marginals of $\widetilde{\nu}$. They lie in $\mathcal{I} \cap \mathcal{S}$. The integral gives

$$\mu = \int_{(\widetilde{\mathcal{I}}\cap\widetilde{\mathcal{S}})_e} \widetilde{\nu}_1\,\Gamma(d\widetilde{\nu}) \quad \text{and} \quad \nu = \int_{(\widetilde{\mathcal{I}}\cap\widetilde{\mathcal{S}})_e} \widetilde{\nu}_2\,\Gamma(d\widetilde{\nu}).$$

By assumption μ and ν are extreme points of $\mathcal{I} \cap \mathcal{S}$, hence we must have $\widetilde{\nu}_1 = \mu$ and $\widetilde{\nu}_2 = \nu$ for Γ-almost every $\widetilde{\nu}$. Now we can pick any $\widetilde{\nu}$ in the support of Γ. ∎

We are ready for the main theorem, which says that the exchangeable measures make up all the translation invariant equilibrium distributions for the class of processes considered.

Theorem 6.7 *For a finite range, translation invariant exclusion process with irreducibility property* (6.4), $(\mathcal{I} \cap \mathcal{S})_e = \{\nu_\rho : 0 \le \rho \le 1\}$.

Proof. Each ν_ρ lies in $\mathcal{I} \cap \mathcal{S}$. Also, ν_ρ is ergodic, so it cannot be a convex combination of two distinct translation invariant measures. In particular, it must be an extreme point of the set $\mathcal{I} \cap \mathcal{S}$.

Let $\nu \in (\mathcal{I} \cap \mathcal{S})_e$. Let $\rho \in [0,1]$. By Lemma 6.6, we can find $\widetilde{\nu} \in (\widetilde{\mathcal{I}} \cap \widetilde{\mathcal{S}})_e$ with marginals $\widetilde{\nu}_1 = \nu$ and $\widetilde{\nu}_2 = \nu_\rho$. Consider these two events in the space X^2:

$$A = \{\eta \le \zeta\} \qquad \text{and} \qquad B = \{\eta \ge \zeta\}.$$

Suppose $0 < \widetilde{\nu}(A) < 1$. Since A is closed for the process (η_t, ζ_t), $\widetilde{\nu}(\cdot\,|A)$ and $\widetilde{\nu}(\cdot\,|A^c)$ are equilibrium distributions by Lemma 4.3. Also, events A and A^c are translation invariant, so

$\widetilde{\nu}(\cdot\,|A)$ and $\widetilde{\nu}(\cdot\,|A^c)$ are translation invariant measures. Since $\widetilde{\nu}$ is a convex combination of $\widetilde{\nu}(\cdot\,|A)$ and $\widetilde{\nu}(\cdot\,|A^c)$, this contradicts the extremality of $\widetilde{\nu}$.

This same reasoning works for B also, and we conclude that both events A and B have measure 0 or 1 under $\widetilde{\nu}$. By Corollary 6.4 one of these events must have probability 1. Depending on which set has full measure, we conclude that either $\nu \le \nu_\rho$ or $\nu \ge \nu_\rho$.

To summarize: for each $0 \le \rho \le 1$, either $\nu \le \nu_\rho$ or $\nu \ge \nu_\rho$. Since $\nu_{\rho_1} \le \nu_{\rho_2}$ for $\rho_1 \le \rho_2$, the number $\rho_0 = \inf\{\rho : \nu \le \nu_\rho\}$ satisfies $\nu_\rho \le \nu \le \nu_\lambda$ for $\rho < \rho_0 < \lambda$. Thus by Lemma A.7, $\nu = \nu_{\rho_0}$. ∎

In the next chapter we address the question of convergence towards an equilibrium ν_ρ from a spatially ergodic initial distribution. We can solve this question only for the one-dimensional lattice $S = \mathbf{Z}$. We generalize the treatment in a different direction, by permitting more than one particle per site. This introduces new problems. For example we cannot explicitly describe even the translation invariant equilibrium distributions in all cases.

Notes

This chapter is from Section 4 in Liggett's lectures [26].

7 Asymmetric K-exclusion processes in one dimension

7.1 The K-exclusion process

In this chapter we generalize the exclusion process to allow more than one particle per site. The bound on the number of particles a site can hold is a fixed positive integer K. As before, jumps are governed by mutually independent Poisson jump time processes $\{\mathcal{T}_{(x,y)}\}$ with given rates $p(x,y)$. The new jump rule is that if $t \in \mathcal{T}_{(x,y)}$, then at time t one particle is moved from site x to site y provided that, after this move, site y has at most K particles. The occupation variables $\eta(x)$ now take values $0, 1, \ldots, K$, and the state space of the process is $X = \{0, 1, \ldots, K\}^S$ with $S = \mathbf{Z}^d$ as before. The generator of the process is

$$Lf(\eta) = \sum_{x,y} p(x,y)\mathbf{1}\{\eta(x) \geq 1,\, \eta(y) \leq K-1\}[f(\eta^{x,y}) - f(\eta)] \tag{7.1}$$

where

$$\eta^{x,y}(z) = \begin{cases} \eta(x)-1, & z = x \\ \eta(y)+1, & z = y \\ \eta(z), & z \notin \{x,y\} \end{cases} \tag{7.2}$$

is the configuration that results from moving a single particle from x to y. The process described by this generator is the K-exclusion process.

The construction of Section 2.1 and the properties proved in Sections 2.3 and 4.1.2 can all be repeated for a translation invariant finite range transition probability $p(x,y)$. We leave these as an extended exercise for the reader. For the case $K = 1$ Theorem 6.1 showed that Bernoulli measures are equilibrium measures for the translation invariant process. When $K > 1$ we cannot explicitly write down translation invariant equilibrium distributions, except in the symmetric case. See Exercises 7.1 and 7.2 at the end of this section.

The assumptions for the results of this section include the standing assumptions of translation invariance

$$p(x,y) = p(0, y-x),$$

and finite range:

$$p(0,x) = 0 \text{ for } |x|_\infty > R.$$

The new assumptions are that the lattice is one dimensional, so $S = \mathbf{Z}$, and for convenience the assumption

$$p(0,1) > 0 \tag{7.3}$$

that gives us the irreducibility we need. Of course, we could just as well assume $p(0,-1) > 0$ and then switch left and right in the proof.

Let $\mathcal{I}$ be the set of equilibrium probability measures of the process, and $\mathcal{S}$ the set of translation invariant probability measures on $X = \{0, 1, \ldots, K\}^{\mathbf{Z}}$. Our objective is to say as much as we can about the set $(\mathcal{I} \cap \mathcal{S})_e$, and to prove that if started from a spatially ergodic initial distribution, the distribution of the state converges weakly as $t \to \infty$.

As already indicated above, in general we do not know what extreme elements of $\mathcal{I} \cap \mathcal{S}$ look like. So our goal is an existence theorem that says $(\mathcal{I} \cap \mathcal{S})_e$ is indexed by density. Let

$$H = \left\{ \int \eta(0)\, \nu(d\eta) : \nu \in (\mathcal{I} \cap \mathcal{S})_e \right\} \tag{7.4}$$

be the set of densities of the extreme measures in $\mathcal{I} \cap \mathcal{S}$. Here is the main result, proved for the one dimensional, translation invariant, finite range K-exclusion process that satisfies assumption (7.3). K is an arbitrary positive integer.

Theorem 7.1 *The set H is a closed subset of $[0, K]$, and for each $\rho \in H$ there is a unique measure $\nu_\rho \in (\mathcal{I} \cap \mathcal{S})_e$ such that $\int \eta(0)\, d\nu_\rho = \rho$. Dependence on ρ is monotone: $\nu_{\rho_1} \leq \nu_{\rho_2}$ for $\rho_1 < \rho_2$ in H. The measures in $(\mathcal{I} \cap \mathcal{S})_e = \{\nu_\rho : \rho \in H\}$ are spatially ergodic.*

Suppose the process is started with an initial distribution μ that is translation invariant and ergodic, and has density $\rho = \int \eta(0)\, d\mu$. Let $\mu_t = \mu S(t)$ be the distribution of the state of the process at time t. The measures μ_t converge weakly as $t \to \infty$, and the limit depends on ρ as follows.

(i) *If $\rho \in H$, then $\mu_t \to \nu_\rho$ as $t \to \infty$.*

(ii) *If $\rho \notin H$, let ρ_* and ρ^* be the closest densities below and above ρ in H. Precisely,*

$$\rho_* = \sup\{h \in H : h < \rho\} \quad \textit{and} \quad \rho^* = \inf\{h \in H : h > \rho\}.$$

Let $\alpha = (\rho^* - \rho)/(\rho^* - \rho_*)$. Then $\mu_t \to \alpha \nu_{\rho_*} + (1 - \alpha)\nu_{\rho^*}$ as $t \to \infty$.

For two special cases we can give a complete result.

Theorem 7.2 *Suppose $K = 1$, so we are discussing a one-dimensional exclusion process. Then $(\mathcal{I} \cap \mathcal{S})_e$ is the set of Bernoulli measures $\{\nu_\rho : 0 \leq \rho \leq 1\}$ defined by (4.15). For a spatially ergodic initial distribution μ with density $\rho = \int \eta(0)\, d\mu$, $\mu_t \to \nu_\rho$ as $t \to \infty$.*

Proof. The characterization of $(\mathcal{I} \cap \mathcal{S})_e$ is a special case of Theorem 6.7. Thus $H = [0, 1]$, and the convergence follows from case (i) of Theorem 7.1. ∎

The other special case is the symmetric case $p(x, y) = p(y, x)$. For $b > 0$ let λ^b be the product measure on X with single coordinate marginal distribution

$$\lambda^b\{\eta(x) = j\} = c^{-1} b^j,\ j = 0, 1, \ldots, K,\ x \in \mathbf{Z}, \tag{7.5}$$

where $c = 1 + b + b^2 + \cdots + b^K$ is the normalization constant that makes this a probability measure. The family $\{\lambda^b : b > 0\}$ forms the extremal translation invariant equilibria for the symmetric K-exclusion. The parameter b does not correspond to density, so we change the notation to be consistent with our earlier statements. For $\rho \in [0, K]$, define $b = b(\rho)$ by

$$\rho = \frac{b + 2b^2 + 3b^3 + \cdots + Kb^K}{1 + b + b^2 + b^3 + \cdots + b^K} = \int \eta(0)\, \lambda^b(d\eta).$$

Let $\nu_\rho = \lambda^{b(\rho)}$. Then $\int \eta(0)\, d\nu_\rho = \rho$.

Theorem 7.3 *Suppose K is arbitrary and $p(x,y)$ is symmetric. Then $(\mathcal{I} \cap \mathcal{S})_e = \{\nu_\rho : 0 \leq \rho \leq K\}$, with ν_ρ as defined above. For a spatially ergodic initial distribution μ with density $\rho = \int \eta(0)\, d\mu$, $\mu_t \to \nu_\rho$ as $t \to \infty$.*

Proof. According to Exercise 7.1, the measures ν_ρ are equilibrium measures. Since they are translation invariant, they are members of $\mathcal{I} \cap \mathcal{S}$. Being i.i.d. they are ergodic, and so extreme points of the set $\mathcal{S}$. Thus no ν_ρ can be a nontrivial convex combination of any translation invariant measures, in particular not of elements of $\mathcal{I} \cap \mathcal{S}$. So the ν_ρ's must be extreme elements of $\mathcal{I} \cap \mathcal{S}$. The densities of these ν_ρ's cover the entire interval $[0, K]$, and so $H = [0, K]$. By the uniqueness statement in Theorem 7.1, there are no other extreme measures in $\mathcal{I} \cap \mathcal{S}$. The convergence follows from case (i) of Theorem 7.1. ∎

The main open problem left here in one dimension is the existence of spatially ergodic equilibrium distributions for all densities, in the general nonsymmetric case. In higher dimension there is nothing like Theorem 7.1 for the nonsymmetric case. The remainder of this chapter proves Theorem 7.1.

Exercise 7.1 Suppose $p(x,y)$ is finite range, translation invariant, and symmetric. Check that the i.i.d. product measure λ^b with marginals defined by (7.5) is invariant for the Markov process with generator (7.1). This result does not require dimension $d = 1$.

Exercise 7.2 Show that if $p(x,y)$ is not symmetric, no i.i.d. product measure can be invariant. Here is a suggestion: Suppose an i.i.d. product measure μ is invariant. First let $f(\eta) = f_0(\eta(0))$ be an arbitrary function of a single coordinate, and show that $\int Lf\, d\mu = 0$ forces μ to be of the type λ^b for some b. Next take

$$f(\eta) = \mathbf{1}\{\eta(u) = 1, \eta(v) = K\}$$

for two sites u and v, and show that $\int Lf\, d\mu = 0$ forces $p(0, u - v) = p(0, v - u)$.

7.2 Proof of Theorem 7.1

This section utilizes the basic coupling of two copies of the K-exclusion process, and it works in principle exactly as in Section 6.2. The state space for the coupled process (η_t, ζ_t) is X^2. Their jumps are governed by the same Poisson clocks $\{\mathcal{T}_{(x,y)}\}$. We leave the explicit formula for the generator of the joint process as an exercise, in the style of (6.3). The semigroup for the coupled process is $\widetilde{S}(t)$.

Fix a translation invariant and ergodic probability measure $\widetilde{\mu}$ on X^2. This is the initial distribution of the coupled process (η_t, ζ_t). The coupled process is constructed as a function of the $\widetilde{\mu}$-distributed initial configuration (η_0, ζ_0) and the Poisson jump time processes, as explained in Section 2.1. The Poisson processes are represented by a sample point $\omega = (\mathcal{T}_{(x,y)} : (x,y) \in S_p^2)$ from a probability space $(\Omega, \mathcal{H}, \mathbf{P})$. Recall that $S_p^2 = \{(x,y) : p(x,y) > 0\}$. The Poisson processes are independent of the initial configuration. So the process $(\eta_\cdot, \zeta_\cdot)$ is constructed on the product space $(X^2 \times \Omega, \mathcal{B}(X^2) \otimes \mathcal{H}, \widetilde{\mu} \otimes \mathbf{P})$.

For expectations on this probability space we write P and E instead of the longer $\widetilde{\mu} \otimes \mathbf{P}$. P and E are also used to refer to the distribution of the coupled process $(\eta_\cdot, \zeta_\cdot)$ on its path space D_{X^2}. The distribution of the state (η_t, ζ_t) at time t is $\widetilde{\mu}_t = \widetilde{\mu}\widetilde{S}(t)$.

Lemma 7.4 *The probability measure $\widetilde{\mu}_t$ is translation invariant and ergodic for each $t \geq 0$.*

Proof. Think of the initial configurations and the Poisson point processes together as a process indexed by $x \in \mathbf{Z}$ in this sense:

$$(\eta_0, \zeta_0, \omega) = \{\eta_0(x), \zeta_0(x), (\mathcal{T}_{(x,y)} : y \in \mathbf{Z}, p(x,y) > 0) : x \in \mathbf{Z}\}.$$

The distribution of this process is the measure $\widetilde{\mu} \otimes \mathbf{P}$ on the space $X^2 \times \Omega$. As x varies, the collections of Poisson processes $(\mathcal{T}_{(x,y)} : y \in \mathbf{Z}, p(x,y) > 0)$ form an i.i.d. sequence. Thus by Lemma A.9 $(\eta_0, \zeta_0, \omega)$ is an ergodic process indexed by $\mathbf{Z}$.

Let g be the map that constructs the occupation numbers at the origin at time t from the initial configurations and the Poisson point processes: $(\eta_t(0), \zeta_t(0)) = g(\eta_0, \zeta_0, \omega)$. Then $(\eta_t(x), \zeta_t(x)) = g(\theta_x\eta_0, \theta_x\zeta_0, \theta_x\omega)$. By Lemma A.10 the process $\{\eta_t(x), \zeta_t(x) : x \in \mathbf{Z}\}$ is ergodic. ∎

Define positive and negative discrepancies by

$$\beta_t^+(x) = (\eta_t(x) - \zeta_t(x))^+ \text{ and } \beta_t^-(x) = (\eta_t(x) - \zeta_t(x))^-,$$

and denote discrepancies of both kinds by $\xi_t(x) = \beta_t^+(x) + \beta_t^-(x)$. Put $\gamma_t(x) = \eta_t(x) \wedge \zeta_t(x)$. The joint process (η_t, ζ_t) can be recovered from the process $(\gamma_t, \beta_t^+, \beta_t^-)$ by

$$\eta_t = \gamma_t + \beta_t^+ \qquad \text{and} \qquad \zeta_t = \gamma_t + \beta_t^-.$$

The dynamics can be described entirely in terms of $(\gamma_t, \beta_t^+, \beta_t^-)$ without any reference to (η_t, ζ_t), as follows.

(a) The γ-particles jump according to the K-exclusion rules, obeying the Poisson clocks $\{\mathcal{T}_{(x,y)}\}$.

(b) If two discrepancy particles of opposite sign are on the same site, they immediately merge and produce one γ-particle.

(c) The discrepancy particles are invisible to the γ-particles, and behave as second class particles relative to the γ-particles, which means two things:

(c.i) When a jump time $t \in \mathcal{T}_{(x,y)}$ happens, if a γ-particle is present at x at time $t-$, only a γ-particle attempts to jump. A discrepancy particle at x may attempt the $x \curvearrowright y$ jump only if no γ-particle is present at x at time $t-$.

(c.ii) If a γ-particle jumps to y, and after the jump there are $K+1$ particles at y, a discrepancy particle is moved from y to x. There must be a discrepancy particle at y, because otherwise y already had K γ-particles at time $t-$, and no jump $x \curvearrowright y$ could have happened.

By (b) each site contains discrepancy particles of at most one type, so there is no ambiguity about which type of discrepancy particle might move from x to y in (c.i) or from y to x in (c.ii).

A consequence of Lemma 7.4 is that $E[\eta_t(x)] = E[\eta_t(0)]$ for all $x \in \mathbf{Z}$, and similarly for the densities of the other particles and discrepancies.

Lemma 7.5 *The densities $E[\eta_t(0)]$ and $E[\zeta_t(0)]$ are constant in time. The discrepancy densities $E[\beta_t^\pm(0)]$ and $E[\xi_t(0)]$ are nonincreasing functions of t.*

Proof. Fix $s < t$, and imagine restarting the process at time s, from state (η_s, ζ_s). Find sites

$$0 < z_1 < z_2 < z_3 < \cdots < z_\ell \nearrow \infty$$

with this property: if $\mathcal{T}_{(x,y)}$ has a jump time during $(s,t]$, then for each ℓ either both x and y lie in the interval $\{-z_\ell, \ldots, z_\ell\}$, or neither does. Such a sequence of sites exists with probability 1. Here is a way to see this. Recall definition (2.3) of $\mathcal{T}'_x$ and the definition of R as the maximal range of a single jump. For each $k > 0$, there is a fixed positive probability that the Poisson process

$$\bigcup_{Rk \le x < R(k+1)} \mathcal{T}'_x \cup \mathcal{T}'_{-x}$$

has no jump times in $(s,t]$. Consequently this happens for infinitely many $k > 0$, say for $0 < k_1 < k_2 < k_3 < \cdots$. For such k there can be no jump across the interval $\{Rk, \ldots, Rk+R-1\}$ or across the interval $\{-Rk-R+1, \ldots, -Rk\}$, because the maximum range of a jump is R. Take $z_\ell = Rk_\ell$.

By the ergodic theorem,

$$E[\xi_t(0)] = \lim_{n\to\infty} \frac{1}{2n+1} \sum_{x=-n}^{n} \xi_t(x) \qquad \text{almost surely,} \tag{7.6}$$

and similarly at time s. Since discrepancies are not created and possibly only annihilated, and since the interval $\{-z_\ell, \ldots, z_\ell\}$ does not exchange discrepancies with the rest of the system during $(s, t]$,

$$\sum_{x=-z_\ell}^{z_\ell} \xi_t(x) \;\le\; \sum_{x=-z_\ell}^{z_\ell} \xi_s(x). \tag{7.7}$$

Divide by $2z_\ell + 1$ and let $\ell \nearrow \infty$. Passing to the limit in (7.6) along a random subsequence $n = z_\ell$ does not alter the conclusion, and hence in the limit $E[\xi_t(0)] \le E[\xi_s(0)]$. Same argument applies to $\beta_t^{\pm}$ as well.

For η_t and ζ_t there is equality in (7.7) because particles are neither created nor annihilated. The ergodic limit shows that their densities are constant in time. ∎

Let I be a finite interval of sites in $\mathbf{Z}$, and $B \subseteq X^2$ the event that I contains discrepancies of opposite sign. The proof of the next lemma uses the assumption $p(0,1) > 0$.

Lemma 7.6 *$\widetilde{\mu}_t(B) \to 0$ as $t \to \infty$.*

Proof. The probability $\widetilde{\mu}_t(B)$ is not altered by translating I, so we may assume that $I = \{0, 1, \ldots, m-1\}$ for some m. Fix $0 < T < \infty$. For any time point s, let A_s be the following event on the Poisson jump time processes: during $(s, s+T]$, this sequence of jump times happens, and no other clock rings that involves any site in I:

Round 1: K jump times in $\mathcal{T}_{(m-2,m-1)}$.

Round 2: K jump times in $\mathcal{T}_{(m-3,m-2)}$, followed by K jump times in $\mathcal{T}_{(m-2,m-1)}$.

Round 3: K jump times in $\mathcal{T}_{(m-4,m-3)}$, followed by K jump times in $\mathcal{T}_{(m-3,m-2)}$, followed by K jump times in $\mathcal{T}_{(m-2,m-1)}$.

... And so on, until in the last round:

Round $m-1$: K jump times in $\mathcal{T}_{0,1}$, followed by K jump times in $\mathcal{T}_{1,2}$, followed by K jump times in $\mathcal{T}_{2,3}$, ..., followed by K jump times in $\mathcal{T}_{m-2,m-1}$.

After round k in this scheme, particles initially at sites $\{m-k-1, \ldots, m-1\}$ are packed to the right end of the interval I, and after the last round $m-1$, all particles initially in I are packed to the right end of the interval I. In particular, after the last round there cannot be discrepancies of opposite types in I. Let $\delta_0 = \mathbf{P}(A_s) > 0$. By the temporal invariance of the Poisson processes, this quantity is the same for all s. The assumption $p(0,1) > 0$ in (7.3) was made to facilitate the definition of A_s.

Let

$$I_x = I + x = \{x, x+1, \dots, x+m-1\} \qquad \text{for } x \in \mathbf{Z}.$$

The spatially shifted event $A_{s,x} = \theta_x^{-1} A_s$ has the same effect on the interval I_x as A_s has on I described above, and the same probability $\delta_0 = \mathbf{P}(A_{s,x})$. $\theta_x^{-1}B$ is the the event that interval I_x contains discrepancies of opposite sign.

To get a contradiction, suppose there exists a sequence $t_n \nearrow \infty$ and $\delta_1 > 0$ such that

$$\widetilde{\mu}_{t_n}(B) \geq \delta_1. \tag{7.8}$$

We may assume that $t_{n+1} > t_n + T$ for all n, by dropping terms from the sequence if necessary.

Fix t_n for the moment. As in the proof of the previous lemma, find sites $z_\ell \nearrow \infty$ such that the system in $\{-z_\ell, \dots, z_\ell\}$ does not interact with the outside during time interval $(t_n, t_n + T]$.

To explain the next inequality, note that if interval $I_x \subseteq \{-z_\ell, \dots, z_\ell\}$ contains opposite discrepancies at time t_n, and if event $A_{t_n,x}$ happens, then at least two discrepancies were annihilated in I_x during $(t_n, t_n + T]$. Event $A_{t_n,x}$ contains the requirement that I_x does not interact with the outside during $(t_n, t_n + T]$. So for disjoint I_x and I_y the annihilated discrepancies were distinct, and can be counted separately. No discrepancies migrate to $\{-z_\ell, \dots, z_\ell\}$ from the outside during $(t_n, t_n + T]$. Let $J = [(2z_\ell + 1)/m] - 1$ be the number of disjoint m-length intervals that fit in $\{-z_\ell, \dots, z_\ell - m\}$. Let $x \in \{0, \dots, m-1\}$. Let us write $A(s,x)$ for $A_{s,x}$ to cut down on subscripts. The discrepancy balance from time t_n to $t_n + T$ satisfies this inequality.

$$\sum_{y=-z_\ell}^{z_\ell} \xi_{t_n+T}(y) \leq \sum_{y=-z_\ell}^{z_\ell} \xi_{t_n}(y) - 2 \sum_{0 \leq j < J} \mathbf{1}_B \left(\theta_{-z_\ell+x+jm}(\eta_{t_n}, \zeta_{t_n})\right) \mathbf{1}_{A(t_n, -z_\ell+x+jm)}.$$

As j varies through $0, \dots, J-1$, the intervals $I_{-z_\ell+x+jm}$ are disjoint and lie inside $\{-z_\ell, \dots, z_\ell - m + x\}$, so annihilated discrepancies are not counted more than once. Next, average over $x \in \{0, \dots, m-1\}$ to get

$$\begin{aligned}
\sum_{y=-z_\ell}^{z_\ell} \xi_{t_n+T}(y) &\leq \sum_{y=-z_\ell}^{z_\ell} \xi_{t_n}(y) - \frac{2}{m} \sum_{x=0}^{m-1} \sum_{0 \leq j < J} \mathbf{1}_B \left(\theta_{-z_\ell+x+jm}(\eta_{t_n}, \zeta_{t_n})\right) \mathbf{1}_{A(t_n, -z_\ell+x+jm)} \\
&\leq \sum_{y=-z_\ell}^{z_\ell} \xi_{t_n}(y) - \frac{2}{m} \sum_{y=-z_\ell}^{z_\ell - m} \mathbf{1}_B(\theta_y \eta_{t_n}, \theta_y \zeta_{t_n}) \mathbf{1}_{A(t_n, y)}.
\end{aligned}$$

Divide by $2z_\ell + 1$ and let $\ell \nearrow \infty$. Ergodicity applies to all terms. Note in particular that the events $\{(\theta_y \eta_{t_n}, \theta_y \zeta_{t_n}) \in B\}$ are independent of the events $A_{t_n,y}$, because the latter only depend on Poisson clocks in time interval $(t_n, t_n + T]$. So here is another application of Lemma A.9. In the limit

$$E[\xi_{t_n+T}(0)] \le E[\xi_{t_n}(0)] - \frac{2}{m}\widetilde{\mu}_{t_n}(B)\mathbf{P}(A_{t_n}) \le E[\xi_{t_n}(0)] - \frac{2\delta_0\delta_1}{m}$$

where we used (7.8) and $\delta_0 = \mathbf{P}(A_{t_n})$. By Lemma 7.5 and the assumption $t_{n+1} > t_n + T$ we get

$$E[\xi_{t_{n+1}}(0)] \le E[\xi_{t_n+T}(0)] \le E[\xi_{t_n}(0)] - \frac{2\delta_0\delta_1}{m}.$$

However, $E[\xi_{t_n}(0)]$ is obviously bounded by 0 and K so we cannot subtract from it a fixed constant $2\delta_0\delta_1/m$ infinitely many times along the sequence t_n, which the above inequality suggests. This contradiction indicates that (7.8) must be false, and so $\widetilde{\mu}_t(B) \to 0$. ∎

Let $H_{t,x}$ be the event that at time t there is a discrepancy at x, and the next discrepancy to the right of site x is of the opposite sign. By spatial invariance, $P(H_{t,x}) = P(H_{t,0})$ for all x.

Corollary 7.7 *$P(H_{t,0}) \to 0$ as $t \to \infty$.*

Proof. Suppose there exist a subsequence $t_n \nearrow \infty$ and $\delta > 0$ such that $P(H_{t_n,0}) \ge \delta$. Fix $t = t_n$ for the moment. Let Y_t be the nearest site to the right of 0 that contains a discrepancy at time t. Then

$$E[\mathbf{1}_{H_{t,0}} Y_t] \le 1$$

by Lemma A.11 applied to the spatially stationary process

$$\{\mathbf{1}[\text{there is a discrepancy at } x \text{ at time } t] : x \in \mathbf{Z}\}.$$

Pick $N > 2/\delta$. Let G_t be the event that there is at least one discrepancy in the interval $\{1, \dots, N\}$ at time t. Then

$$1 \ge E[\mathbf{1}_{H_{t,0}} \mathbf{1}_{G_t^c} Y_t] \ge N \cdot E[\mathbf{1}_{H_{t,0}} \mathbf{1}_{G_t^c}]$$

from which

$$P(H_{t,0} \cap G_t) = P(H_{t,0}) - P(H_{t,0} \cap G_t^c) \ge \delta - \frac{1}{N} \ge \delta/2.$$

This bound is valid for all $t = t_n$. However,

$$H_{t,0} \cap G_t \subseteq \{\text{the interval } \{0, \dots, N\} \text{ contains discrepancies of opposite type}\},$$

and for a fixed N the probability of the latter event vanishes as $t \to \infty$ by Lemma 7.6. This contradiction shows that $P(H_{t,0}) \to 0$ as $t \to \infty$. ∎

We come to the main step of the proof which says that, in the presence of sufficiently many positive discrepancies, all negative discrepancies will be annihilated. The initial distribution of the coupled process is still a fixed, spatially ergodic probability measure $\widetilde{\mu}$ on X^2.

Proposition 7.8 *Suppose $E[\eta_0(x)] = \rho_1$, $E[\zeta_0(x)] = \rho_2$, and $\rho_1 \geq \rho_2$. Then*

$$\lim_{t\to\infty} E[(\eta_t(x) - \zeta_t(x))^-] = 0. \tag{7.9}$$

Proof. By Lemma 7.5 the densities of discrepancies are nonincreasing functions of t. Thus to get a contradiction, we may assume that for some $\delta > 0$, $E[(\eta_t(x) - \zeta_t(x))^-] \geq \delta$ for all $t \geq 0$. But then, since

$$0 \leq \rho_1 - \rho_2 = E[\eta_t(x) - \zeta_t(x)] = E[(\eta_t(x) - \zeta_t(x))^+] - E[(\eta_t(x) - \zeta_t(x))^-],$$

it follows that

$$E[(\eta_t(x) - \zeta_t(x))^+] \geq \delta \quad \text{and} \quad E[(\eta_t(x) - \zeta_t(x))^-] \geq \delta \quad \text{for all } t \geq 0. \tag{7.10}$$

We shall derive a contradiction from (7.10).

We track individual discrepancies. At time 0 assign integer labels separately to the positive and negative discrepancies. Each discrepancy retains its label throughout its lifetime. Let $w_j^+(t)$ denote the location of the positive discrepancy with label j at time t, and similarly for $w_j^-(t)$. Assume the initial labeling is nondecreasing, in other words $w_i^\pm(0) \leq w_j^\pm(0)$ for $i < j$.

The life of a discrepancy ends when it merges with a discrepancy of opposite sign to create a γ-particle. Let

$$\tau_j^+ = \inf\{t \geq 0 : w_j^+(t) \text{ meets a discrepancy of the opposite type}\}$$

be the time when positive discrepancy w_j^+ ceases to exist. Of course, w_j^+ may never meet a negative discrepancy, and lives forever. Then $\tau_j^+ = \infty$ and w_j^+ is *immortal*. Same conventions for negative discrepancies, with expiration time τ_j^- for discrepancy w_j^-.

We need to be specific about which discrepancy at a site is affected by a jump if more than one discrepancy is present. We stipulate that if the interaction is to the *left* of x, then it involves the discrepancy at x with the smallest index. If the interaction is to the *right*, then the discrepancy at x with the largest index is affected. For example, suppose $t \in \mathcal{T}_{(x,y)}$, $x < y$, and at time $t-$ we have $\gamma_{t-}(x) = 0$, $\beta_{t-}^+(x) > 0$, and $\beta_{t-}^-(y) > 0$. Then at time t the positive discrepancy with the largest index at x jumps to y, and merges with the negative discrepancy with the smallest index at y to create a γ-particle. Note that a positive discrepancy was allowed to jump because there were no γ-particles at x at time $t-$.

Despite this rule, discrepancies do not stay ordered, unless only nearest-neighbor jumps are permitted, which means that $p(x, y) = 0$ for $y \neq x \pm 1$.

Let

$$\beta_{0,t}^{\pm}(x) = \sum_j \mathbf{1}\{w_j^{\pm}(0) = x,\, \tau_j^{\pm} > t\}$$

denote the occupation variables at time 0 of those discrepancies that are still alive at time t. For fixed t, the spatial process

$$\{\beta_{0,t}^{+}(x), \beta_{0,t}^{-}(x) : x \in \mathbf{Z}\}$$

is translation invariant and ergodic. This follows by the argument of Lemma 7.4. The labeling of the discrepancies may appear to confound the issue. But again we can write

$$(\beta_{0,t}^{+}(x), \beta_{0,t}^{-}(x)) = F_{t,x}(\eta_0, \zeta_0, \omega),$$

where the map $F_{t,x}$ first labels the discrepancies, then constructs the evolution up to time t from the graphical representation, following the conventions on discrepancies enunciated above, and returns the occupation numbers at site x. Whatever the details of this map, if the entire picture $(\eta_0, \zeta_0, \omega)$ is translated so that x becomes the origin, then the discrepancies at x are moved to the origin, and the entire evolution is similarly translated. This just says that

$$F_{t,x}(\eta_0, \zeta_0, \omega) = F_{t,0}(\theta_x\eta_0, \theta_x\zeta_0, \theta_x\omega). \tag{7.11}$$

Translation invariance and ergodicity now follow from Lemma A.10.

Let

$$g^{\pm}(t) = E\beta_{0,t}^{\pm}(0)$$

be the density of these discrepancies at time 0. By the ergodic theorem, also

$$g^{\pm}(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{x=1}^{n} \beta_{0,t}^{\pm}(x) \qquad \text{almost surely.} \tag{7.12}$$

Assumption (7.10) implies that $g^{\pm}(t) \geq \delta$. We can see this by repeating the argument used in the proof of Lemma 7.5 above. Suppose $\{-z_\ell, \dots, z_\ell\}$ is a portion of the lattice that does not interact with the outside during time interval $(0, t]$. Since discrepancies alive at time t were also alive at time 0,

$$\sum_{x=-z_\ell}^{z_\ell} \beta_{0,t}^{\pm}(x) = \sum_{x=-z_\ell}^{z_\ell} \beta_t^{\pm}(x).$$

Dividing by $2z_\ell + 1$ and letting $\ell \to \infty$ gives

$$g^{\pm}(t) = E[\beta_t^{\pm}(x)] = E[(\eta_t(x) - \zeta_t(x))^{\pm}] \geq \delta$$

by assumption (7.10).

Define the initial occupation numbers of immortal discrepancies by

$$\beta^{\pm}_{0,\infty}(x) = \lim_{t\to\infty} \beta^{\pm}_{0,t}(x).$$

The limit exists by monotonicity. The translation property (7.11) is preserved to the limit and becomes

$$(\beta^{+}_{0,\infty}(x), \beta^{-}_{0,\infty}(x)) = F_{\infty,x}(\eta_0, \zeta_0, \omega) = F_{\infty,0}(\theta_x\eta_0, \theta_x\zeta_0, \theta_x\omega),$$

so the process

$$\{\beta^{+}_{0,\infty}(x), \beta^{-}_{0,\infty}(x) : x \in \mathbf{Z}\}$$

is spatially translation invariant and ergodic. The ergodic theorem gives the almost sure limits

$$g^{\pm}(\infty) = E\beta^{\pm}_{0,\infty}(0) = \lim_{n\to\infty} \frac{1}{n}\sum_{x=1}^{n} \beta^{\pm}_{0,\infty}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{x=-n}^{-1} \beta^{\pm}_{0,\infty}(x). \tag{7.13}$$

The equality of the left and right limits is a special case of (A.9) in the appendix.

By monotone convergence,

$$g^{\pm}(\infty) = E\beta^{\pm}_{0,\infty}(0) = \lim_{t\to\infty} E\beta^{\pm}_{0,t}(0) = \lim_{t\to\infty} g^{\pm}(t) \geq \delta. \tag{7.14}$$

From (7.13) and (7.14) we conclude that at time 0, with probability one there are infinitely many immortal discrepancies of both sign on both sides of the origin. We are now ready to complete the proof of the proposition for the case of a nearest-neighbor process, where discrepancies of opposite sign cannot jump over each other.

Completion of the proof of Proposition 7.8 for a nearest-neighbor process.

The assumption is that $p(x, y) = 0$ for $y \neq x \pm 1$. Then a pair of immortal discrepancies is never switched around. Intuitively speaking, by (7.13) and (7.14) there must be a fixed positive density of immortal negative discrepancies followed by a positive discrepancy. This contradicts Corollary 7.7. Here is the rigorous argument.

For $0 \leq t \leq \infty$, let

$$L_{t,x} = \inf\left\{ n \geq 1 : \sum_{y=x}^{x+n-1} \beta^{+}_{0,t}(y) \geq 1 \right\}$$

be the smallest n such that at time 0 the interval $\{x, \ldots, x+n-1\}$ contains a positive discrepancy that survives up to time t, or is immortal in case $t = \infty$. By (7.13) and (7.14), $L_{\infty,x}$ is an almost surely finite random variable. By the bound $L_{t,x} \leq L_{\infty,x}$ and by translation invariance, it is possible to fix an integer $0 < u < \infty$ so that $P[L_{t,x} > u] < \delta/2K$ for all $x \in \mathbf{Z}$ and $t \geq 0$. Note also that

$$P\{\beta^-_{0,t}(0) \geq 1\} \geq \frac{1}{K} E\beta^-_{0,t}(0) \geq \delta/K.$$

By spatial ergodicity,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{x=1}^{n} \mathbf{1}\{\beta^-_{0,t}(x) \geq 1\}\mathbf{1}\{L_{t,x} \leq u\} = P(\{\beta^-_{0,t}(0) \geq 1\} \cap \{L_{t,0} \leq u\}) \geq \frac{\delta}{2K} \tag{7.15}$$

almost surely.

Fix a sample point of the underlying probability space for which the limit (7.15) holds. Let $1 \leq x_0 < x_1 < x_2 < x_3 < \cdots$ be the points x for which

$$\mathbf{1}\{\beta^-_{0,t}(x) \geq 1\}\mathbf{1}\{L_{t,x} \leq u\} = 1.$$

Let $w^-_{i(r)}(0) = x_r$, $r = 0, 1, 2, 3, \ldots$, be negative discrepancies that live past time t. If there is more than one such negative discrepancy at some x_r, pick the one with the smallest index. Consider every uth such discrepancy:

$$w^-_{i(0)}, w^-_{i(u)}, w^-_{i(2u)}, w^-_{i(3u)}, \ldots$$

Since

$$x_{ku} + u = w^-_{i(ku)}(0) + u \leq w^-_{i((k+1)u)}(0)$$

and $L_{x_{ku},t} \leq u$, at time 0 there must be a positive discrepancy between $w^-_{i(ku)}(0)$ and $w^-_{i((k+1)u)}(0)$ that survives past time t. Since the order of discrepancies is preserved under nearest-neighbor jumps, in the interval $\{w^-_{i(ku)}(t), \ldots, w^-_{i((k+1)u)}(t) - 1\}$ is at least one negative discrepancy for which the next discrepancy to the right is positive.

The limit (7.15) translates into

$$\lim_{n\to\infty} \frac{1}{n} \max\{r : x_r \leq n\} \geq \delta/(2K),$$

and this in turn into

$$\lim_{n\to\infty} \frac{1}{n} \max\{k : x_{ku} \leq n\} \geq \delta/(2Ku). \tag{7.16}$$

Let $H^-_{t,x}$ be the event that at time t there is a negative discrepancy at x, and the next discrepancy to the right is positive. By ergodicity

$$P(H^-_{t,0}) = \lim_{n\to\infty} \frac{1}{n} \sum_{x=1}^{n} \mathbf{1}\{H^-_{t,x}\} \qquad \text{almost surely.} \tag{7.17}$$

Choose again sites $z_\ell \nearrow \infty$ with the property that no particles or discrepancies jump over z_ℓ during time $(0, t]$. The limit (7.17) is valid also along the subsequence $n = z_\ell$. Let

$$m = \inf\{k \geq 0 : w^-_{i(ku)}(t) \geq 1\}. \tag{7.18}$$

Since the event $H^-_{t,x}$ happens at least once between each pair $w^-_{i(ku)}(t)$ and $w^-_{i((k+1)u)}(t)$,

$$\begin{aligned} \sum_{x=1}^{z_\ell} \mathbf{1}\{H^-_{t,x}\} &\geq \max\{k : w^-_{i(ku)}(t) \leq z_\ell\} - m = \max\{k : w^-_{i(ku)}(0) \leq z_\ell\} - m \\ &= \max\{k : x_{ku} \leq z_\ell\} - m. \end{aligned} \tag{7.19}$$

Divide by z_ℓ and let $\ell \nearrow \infty$. Combining with (7.16) and (7.17) gives $P(H^-_{t,0}) \geq \delta/(2Ku)$.

This bound is valid for all $t \geq 0$. We have contradicted Corollary 7.7. This concludes the proof of Proposition 7.8 for the nearest-neighbor case.

Completion of the proof of Proposition 7.8 for the general case.

Now we deal with the possibility that the ordering of discrepancies is not preserved. Say that a jump time t is a *switching time* if at time t the order of two discrepancies of opposite sign is reversed and neither one is annihilated. In other words, t is a switching time if for some indices i and j, either

$$w^-_j(t-) < w^+_i(t-) \quad \text{and} \quad w^-_j(t) > w^+_i(t),$$

or same statement holds with reversed inequalities, and $\tau^-_j \wedge \tau^+_i > t$. Let us say that discrepancies w^-_j and w^+_i are involved in a *switch* when the above event happens. Note that as many as $K(R-1)$ negative discrepancies may be involved in a switch at a particular switching time, if a positive discrepancy jumps over $R-1$ sites all filled with negative discrepancies. Let $k^-_j(t)$ be the number of switches experienced by discrepancy w^-_j during time interval $(0, t]$.

For $M \in \mathbf{N}$, $x \in \mathbf{Z}$ and $t < \infty$, define the event

$$G^M_{t,x} = \bigcup_j \{w^-_j(0) = x,\, \tau^-_j > t,\, k^-_j(t) \leq M\}.$$

$G^M_{t,x}$ is the event that at time 0 there is a negative discrepancy at site x which is alive at time t and has experienced at most M switches by time t. As discrepancies can be annihilated and switches only accumulate, $G^M_{t,x}$ decreases with t. By translation invariance $P(G^M_{t,x}) = P(G^M_{t,0})$ for all x. We prove an intermediate lemma.

Lemma 7.9 $P(G^M_{t,0}) \to 0$ *as* $t \to \infty$.

Proof. To get a contradiction, assume there exists a $\delta_1 > 0$ such that $P(G^M_{t,0}) \geq \delta_1$ for all $t \geq 0$. This proof essentially repeats the argument given above in the nearest-neighbor case. For $0 \leq t \leq \infty$, set

$$L'_{t,x} = \inf\left\{ n \geq 1 : \sum_{y=x}^{x+n-1} \beta^+_{0,t}(y) \geq 2M+1 \right\}.$$

By (7.13) and (7.14), $L'_{\infty,x}$ is an almost surely finite random variable. By the bound $L'_{t,x} \leq L'_{\infty,x}$ and by translation invariance, it is possible to fix an integer $0 < u < \infty$ so that $P[L'_{t,x} > u] < \delta_1/2$ for all $x \in \mathbf{Z}$ and $t \geq 0$. By spatial ergodicity,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{x=1}^{n} \mathbf{1}\{G^M_{t,x}\}\mathbf{1}\{L'_{t,x} \leq u\} = P(G^M_{t,0} \cap \{L'_{t,0} \leq u\}) \geq \delta_1/2 \quad \text{almost surely.} \tag{7.20}$$

Fix a sample point of the underlying probability space for which the limit (7.20) holds. Let $1 \leq x_0 < x_1 < x_2 < x_3 < \cdots$ be the points x for which $\mathbf{1}\{G^M_{t,x}\}\mathbf{1}\{L'_{t,x} \leq u\} = 1$. Let $w^-_{i(r)}(0) = x_r$, $r = 0, 1, 2, 3, \ldots$, be negative discrepancies that live past time t but experience at most M switches up to time t. (If there is more than one such negative discrepancy at some x_r, pick the one with the smallest index.) Again, consider the subsequence

$$w^-_{i(0)}, w^-_{i(u)}, w^-_{i(2u)}, w^-_{i(3u)}, \ldots$$

Since

$$x_{ku} + u = w^-_{i(ku)}(0) + u \leq w^-_{i((k+1)u)}(0)$$

and $L'_{x_{ku},t} \leq u$, at time 0 there are at least $2M+1$ positive discrepancies between $w^-_{i(ku)}(0)$ and $w^-_{i((k+1)u)}(0)$ that survive past time t. Discrepancies $w^-_{i(ku)}$ and $w^-_{i((k+1)u)}$ experience together at most $2M$ switches. So at least one positive discrepancy remains between them at time t. In particular, their order must have been preserved up to time t, meaning that $w^-_{i(ku)}(t) < w^-_{i((k+1)u)}(t)$. Consequently, in the interval $\{w^-_{i(ku)}(t), \ldots, w^-_{i((k+1)u)}(t) - 1\}$ is at least one negative discrepancy for which the next discrepancy to the right is positive.

(7.16) now becomes

$$\lim_{n\to\infty} \frac{1}{n} \max\{k : x_{ku} \leq n\} \geq \delta_1/(2u).$$

Equations (7.17)–(7.19) can be repeated verbatim. The conclusion is

$$P(H^-_{t,0}) = \lim_{\ell\to\infty} \frac{1}{z_\ell} \sum_{x=1}^{z_\ell} \mathbf{1}\{H^-_{t,x}\} \geq \delta_1/(2u).$$

This bound is valid for all $t \geq 0$ and contradicts Corollary 7.7. Lemma 7.9 is proved. ∎

We conclude the proof of Proposition 7.8. The idea is this. By the lemma above, a positive density of immortal negative discrepancies forces the density of switches to increase without bound. But after each switch there is an opportunity to annihilate a pair of discrepancies. These two points contradict each other.

Fix $0 < T < \infty$. Define the event $A_{s,x}$ as in the beginning of the proof of Lemma 7.6 for the interval $I_x = \{x, \dots, x+R\}$. $A_{s,x}$ is an event that involves only the Poisson processes

$$\bigcup_{y: x \leq y \leq x+R} \mathcal{T}'_y \cap (s, s+T]$$

and its probability $\delta_1 = \mathbf{P}(A_{s,x}) > 0$ is independent of (s, x). When $A_{s,x}$ occurs, the process in I_x does not interact with the rest of the process during $(s, s+T]$, and at time $s+T$ there are discrepancies of only one type in I_x.

For $m > 0$, let $N^*_m(t)$ be the total number of jump times that cross either site m or $-m$ during $(0, t]$. In terms of counting functions,

$$N^*_m(t) = \sum \{N_{(x,y)}(t) : x \leq -m \leq y,\ y \leq -m \leq x,\ x \leq m \leq y, \text{ or } y \leq m \leq x\}.$$

$N^*_m(\cdot)$ is a Poisson process with rate bounded by $4R$. $N^*_m(t)$ is an upper bound on the number of discrepancies that migrate into the interval $\{-m, \dots, m\}$ during $(0, t]$.

Let

$$\mathcal{T}''_x = \bigcup_{y: x < y \leq x+R} \mathcal{T}_{(x,y)} \cup \mathcal{T}_{(y,x)}$$

be the Poisson process of jump times at which x interacts with a site y to its right. Each jump time in the entire system lies in exactly one $\mathcal{T}''_x$. Define the random variable

$$Q_x(t) = \sum_{s \in \mathcal{T}''_x : s \leq t} \mathbf{1}\{s \text{ is a switching time}\} \mathbf{1}\{A_{s,x}\}.$$

The key observation is that if $s \in \mathcal{T}''_x$ is a switching time, at time s the interval I_x contains at least one pair of discrepancies of opposite sign. If $A_{s,x}$ happens, this pair is annihilated. The jumps that happen in I_x during $(s, s+T]$ in the event $A_{s,x}$ are all nearest-neighbor jumps, hence no switches happen in I_x during $(s, s+T]$. This implies that no annihilation is counted twice in the sum that defines Q_x. Thus $2Q_x(t-T)$ is a lower bound on the number of discrepancies that have been annihilated during $(0, t]$ following a switching time in $\mathcal{T}''_x$.

Next we observe that for distinct $x < x'$ the sums for Q_x and $Q_{x'}$ do not count the same annihilation. Suppose that for $s \in \mathcal{T}''_x$ and $s' \in \mathcal{T}''_{x'}$

$$\mathbf{1}\{s \text{ is a switching time}\} \mathbf{1}\{A_{s,x}\} = 1 \quad \text{and} \quad \mathbf{1}\{s' \text{ is a switching time}\} \mathbf{1}\{A_{s',x'}\} = 1.$$

There is potential interference only if $x < x' \le x + R$ and $|s - s'| < T$. If $s < s' \le s + T$ then in fact s' cannot be a switching time because x' is only involved in nearest-neighbor jumps during $[s', s+T]$. If $s' < s \le s' + T$ then the switching jump at time s happened in $\{x, \dots, x'-1\}$ because during event $A_{s',x'}$ the interval $I_{x'}$ does not interact with the outside. Consequently event $A_{s,x}$ annihilates a pair of discrepancies in $\{x, \dots, x'-1\}$ which must be distinct from those annihilated by $A_{s',x'}$. The case $s = s'$ need not be considered because we use only those realizations of the random clocks that have no simultaneous jump times.

We get this inequality for the discrepancy balance in the interval $\{-m, \dots, m\}$.

$$\sum_{x=-m}^{m} \xi_t(x) \le \sum_{x=-m}^{m} \xi_0(x) - 2 \sum_{x=-m}^{m-R} Q_x(t-T) + N_m^*(t). \tag{7.21}$$

Let $\kappa(t)$ be the expected number of switching times among jump times in $\mathcal{T}_x'' \cap (0, t]$. It is independent of x by spatial invariance. Let $0 < \sigma_1 < \sigma_2 < \sigma_3 < \cdots$ be the jump times in $\mathcal{T}_x''$. Let $\mathcal{F}_t$ denote the filtration of the coupled process (η_t, ζ_t) and the Poisson clocks. The σ_k's are stopping times for the filtration $\mathcal{F}_t$. The event $\{\sigma_k \text{ is a switching time}\}$ lies in the σ-algebra $\mathcal{F}_{\sigma_k}$ of events that are known by time σ_k. The event $A_{\sigma_k,x}$ is a future event. By the strong Markov property the Poisson processes restart independently of the past, and so

$$P(A_{\sigma_k,x}|\mathcal{F}_{\sigma_k}) = \mathbf{P}(A_{0,x}) = \delta_1 \qquad \text{almost surely.}$$

We get

$$\begin{aligned} EQ_x(t-T) &= \sum_{k=1}^{\infty} E\left[\mathbf{1}\{\sigma_k \le t-T\}\mathbf{1}\{\sigma_k \text{ is a switching time}\}\mathbf{1}\{A_{\sigma_k,x}\}\right] \\ &= \sum_{k=1}^{\infty} E\left[\mathbf{1}\{\sigma_k \le t-T\}\mathbf{1}\{\sigma_k \text{ is a switching time}\}P(A_{\sigma_k,x}|\mathcal{F}_{\sigma_k})\right] \\ &= \delta_1 \kappa(t-T). \end{aligned}$$

Take expectations in (7.21), divide by $2m+1$ and let $m \nearrow \infty$ to get

$$E[\xi_t(0)] \le E[\xi_0(0)] - 2\delta_1 \kappa(t-T). \tag{7.22}$$

We show that Lemma 7.9 together with the positive density of immortal negative discrepancies leads to $\kappa(t) \nearrow \infty$, which contradicts (7.22). Let $K^*(x, y, t)$ be the total number of switching times that happen within the space interval $\{x, \dots, y\}$ during $(0, t]$. By spatial ergodicity,

$$\kappa(t) = \lim_{m\to\infty} \frac{1}{2m+1} K^*(-m, m, t) \qquad \text{almost surely.}$$

Recall that $k_j^-(t)$ is the number of switches experienced by discrepancy w_j^- during $(0,t]$. Fix again a sequence of sites $z_\ell \nearrow \infty$ with the property that no exchange of particles or discrepancies happens across $\pm z_\ell$ during time interval $(0,t]$. Recall that a particular switching time affects no more than $K(R-1)$ negative discrepancies. Then

$$K^*(-z_\ell, z_\ell, t) \geq \frac{1}{K(R-1)} \sum_{x=-z_\ell}^{z_\ell} \sum_j k_j^-(t) \, \mathbf{1}\{w_j^-(0) = x, \, \tau_j^- > t\}.$$

The sum on the right above counts the switches experienced by those negative discrepancies that survive past time t in the interval $\{-z_\ell, \ldots, z_\ell\}$. Develop this further:

$$\begin{aligned} K^*(-z_\ell, z_\ell, t) \; &\geq \; \frac{M}{K(R-1)} \sum_{x=-z_\ell}^{z_\ell} \sum_j \mathbf{1}\{w_j^-(0) = x, \, \tau_j^- > t, \, k_j^-(t) > M\} \\ &\geq \; \frac{M}{K(R-1)} \sum_{x=-z_\ell}^{z_\ell} \Biggl(\sum_j \mathbf{1}\{w_j^-(0) = x, \, \tau_j^- > t\} \\ &\qquad\qquad - \sum_j \mathbf{1}\{w_j^-(0) = x, \, \tau_j^- > t, \, k_j^-(t) \leq M\} \Biggr) \\ &\geq \; \frac{M}{K(R-1)} \sum_{x=-z_\ell}^{z_\ell} \left(\beta_{0,t}^-(x) - K \mathbf{1}\{G_{t,x}^M\} \right). \end{aligned}$$

Divide by $2z_\ell + 1$, let $\ell \nearrow \infty$ and use ergodicity to get

$$\kappa(t) \geq \frac{M}{K(R-1)} \left(E\beta_t^-(0) - KP(G_{t,0}^M) \right).$$

Let $t \nearrow \infty$ and use assumption (7.10) and Lemma 7.9 to get

$$\liminf_{t \to \infty} \kappa(t) \geq \frac{M\delta}{K(R-1)}.$$

The quantities δ, K and R are fixed. We may take M and t large enough to contradict (7.22).

We have reached a contradiction, and thereby disproved (7.10). The proof of Proposition 7.8 is complete. ∎

This proposition was the main technical point of the proof. For the remainder of the section, $\widetilde{\mu}$ no longer denotes a fixed initial distribution for the coupled process. Recall that $S(t)$ is the semigroup of the one-dimensional K-exclusion process, and $\widetilde{S}(t)$ the semigroup of the coupled process.

Corollary 7.10 *Let μ_1 and μ_2 be two translation invariant ergodic probability measures on X with densities $\rho_1 = \int \eta(0)\,d\mu_1$ and $\rho_2 = \int \eta(0)\,d\mu_2$.*

(a) *Suppose $\rho_1 \geq \rho_2$. Suppose $t_n \nearrow \infty$ is a subsequence along which the limits $\mu_1 S(t_n) \to \bar{\mu}_1$ and $\mu_2 S(t_n) \to \bar{\mu}_2$ exist. Then $\bar{\mu}_1 \geq \bar{\mu}_2$.*

(b) *Suppose $\rho_1 = \rho_2$, and $t_n \nearrow \infty$ is a subsequence along which the limit $\mu_1 S(t_n) \to \bar{\mu}_1$ exists. Then also $\mu_2 S(t_n) \to \bar{\mu}_1$.*

Proof. (a) Let $\widetilde{\mu}$ be a spatially ergodic measure on X^2 with η-marginal μ_1 and ζ-marginal μ_2. Such a measure exists because the ergodic decomposition of the translation invariant measure $\mu_1 \otimes \mu_2$ must be supported on such measures. Let $\widetilde{\gamma}$ be any limit point of the sequence $\widetilde{\mu}\widetilde{S}(t_n)$. The marginals of $\widetilde{\gamma}$ are $\bar{\mu}_1$ and $\bar{\mu}_2$, and by Proposition 7.8 $\widetilde{\gamma}\{\eta \geq \zeta\} = 1$. This says $\bar{\mu}_1 \geq \bar{\mu}_2$.

(b) Let $\bar{\mu}_2$ be any limit point of the sequence $\mu_2 \widetilde{S}(t_n)$. By part (a), $\bar{\mu}_1 \geq \bar{\mu}_2$ and $\bar{\mu}_1 \leq \bar{\mu}_2$. By Lemma A.5 $\bar{\mu}_2 = \bar{\mu}_1$. Since this is true for all limit points, and we have compactness, the convergence follows. ∎

For the next lemma, let

$$A_n(\eta) = \frac{1}{2n+1} \sum_{x=-n}^{n} \eta(x)$$

denote the average density in the interval $\{-n, \ldots, n\}$. Let $A_\infty(\eta) = \lim_{n\to\infty} A_n(\eta)$ for those $\eta \in X$ for which the limit exists. The limit exists almost surely under any translation invariant measure, and is almost surely constant under any ergodic measure. As a first step toward proving the ergodicity of extreme elements of $\mathcal{I} \cap \mathcal{S}$, we show that A_∞ is constant under such a measure.

Lemma 7.11 *Suppose $\nu \in (\mathcal{I} \cap \mathcal{S})_e$ with density $\int \eta(0)\,d\nu = \rho$. Then ν-almost surely the limit $A_\infty(\eta) = \rho$.*

Proof. Let $\{-z_\ell, \ldots, z_\ell\}$, $z_\ell \nearrow \infty$, be lattice intervals that do not interact with the outside during time interval $(s,t\,]$. As in the proof of Lemma 7.5, we can define $z_\ell = Rk_\ell$, $\ell = 1, 2, 3, \ldots$, where $1 \leq k_1 < k_2 < k_3 < \cdots$ are the indices k at which the event

$$U_k = \left\{ \sum_{Rk \leq x < R(k+1)} \sum_{y \in \mathbf{Z}} \Big(N_{(x,y)}(s,t\,] + N_{(-x,y)}(s,t\,] \Big) = 0 \right\}$$

happens. By the i.i.d. property of the Poisson processes,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{U_k} = \mathbf{P}(U_0) > 0.$$

By Exercise 7.3 this implies

$$\lim_{\ell\to\infty} \frac{z_{\ell+1} - z_\ell}{z_\ell} = 0. \tag{7.23}$$

By particle conservation

$$\sum_{x=-z_\ell}^{z_\ell} \eta_s(x) = \sum_{x=-z_\ell}^{z_\ell} \eta_t(x).$$

Together with (7.23), this implies that the limiting density $A_\infty(\eta_s)$ exists iff $A_\infty(\eta_t)$ exists, and the limits coincide. Here is the argument. For any n, define $\ell = \ell(n)$ by $z_\ell \le n < z_{\ell+1}$. Then

$$\frac{2z_\ell + 1}{2n+1} \cdot \frac{1}{2z_\ell + 1} \sum_{x=-z_\ell}^{z_\ell} \eta_s(x) \le \frac{1}{2n+1} \sum_{x=-n}^{n} \eta_s(x) \le \frac{2z_{\ell+1} + 1}{2n+1} \cdot \frac{1}{2z_{\ell+1} + 1} \sum_{x=-z_{\ell+1}}^{z_{\ell+1}} \eta_s(x).$$

The ratio $(2z_\ell + 1)/(2z_{\ell+1} + 1) \to 1$ by (7.23). Consequently the limits

$$\frac{1}{2z_\ell + 1} \sum_{x=-z_\ell}^{z_\ell} \eta_s(x) \to A_\infty(\eta_s) \qquad \text{and} \qquad \frac{1}{2n+1} \sum_{x=-n}^{n} \eta_s(x) \to A_\infty(\eta_s)$$

are equivalent. The same holds for η_t.

For $0 \le r \le K$ define the event

$$B = \{\eta \in X : A_\infty(\eta) \text{ exists and } A_\infty(\eta) \le r\}.$$

We have shown that this event is closed for the Markov process η_t in the sense of Lemma 4.3. Also, B is translation invariant. Thus if $0 < \nu(B) < 1$, the conditioned measures $\nu(\cdot\,|B)$ and $\nu(\cdot\,|B^c)$ are elements of $\mathcal{I} \cap \mathcal{S}$. Then ν cannot be extreme.

Thus $\nu(B) = 0$ or 1 for any choice of r in the definition of B. It follows that ν-almost surely A_∞ must equal its mean $\int A_\infty \, d\nu = \rho$. ∎

Recall the definition (7.4) of H, the set of densities of extreme measures of $\mathcal{I} \cap \mathcal{S}$.

Proposition 7.12 (a) *Let $\rho \in H$. Then there is a unique measure $\nu_\rho \in (\mathcal{I} \cap \mathcal{S})_e$ with density $\int \eta(0)\, d\nu_\rho = \rho$. For every spatially ergodic μ with density $\int \eta(0)\, d\mu = \rho$, we have the convergence $\mu S(t) \to \nu_\rho$.*

(b) *The measure ν_ρ is spatially ergodic.*

(c) *If $\rho_1 < \rho_2$ are in H, then $\nu_{\rho_1} \le \nu_{\rho_2}$.*

Proof. (a) Fix an ergodic measure μ on X with density $\int \eta(0)\, d\mu = \rho$. Let $\bar{\mu}$ be any limit point of $\mu S(t)$, and $\{t_n\}$ any sequence such that

$$\mu S(t_n) \to \bar{\mu}. \tag{7.24}$$

Let $\bar{\nu}$ be any measure in $(\mathcal{I}\cap\mathcal{S})_e$ with density ρ, which exists by the assumption $\rho \in H$. Let

$$\bar{\nu} = \int_{\mathcal{S}_e} \lambda\, \Gamma(d\lambda)$$

be the ergodic decomposition of $\bar{\nu}$. From Lemma 7.11,

$$1 = \bar{\nu}\{A_\infty = \rho\} = \int_{\mathcal{S}_e} \lambda\{A_\infty = \rho\}\, \Gamma(d\lambda).$$

It follows that Γ-almost every λ has density ρ. And then from (7.24) and Corollary 7.10(b) that $\lambda S(t_n) \to \bar{\mu}$ for Γ-almost every λ. Take a test function $f \in C(X)$. By the invariance of $\bar{\nu}$,

$$\begin{aligned} \int f\, d\bar{\nu} &= \lim_{n\to\infty} \int S(t_n) f\, d\bar{\nu} = \lim_{n\to\infty} \int_{\mathcal{S}_e} \left\{ \int S(t_n) f\, d\lambda \right\} \Gamma(d\lambda) \\ &= \int_{\mathcal{S}_e} \left\{ \lim_{n\to\infty} \int S(t_n) f\, d\lambda \right\} \Gamma(d\lambda) = \int f\, d\bar{\mu}. \end{aligned}$$

This tells us two things. First, there can be only one measure in $(\mathcal{I}\cap\mathcal{S})_e$ with density ρ, because an arbitrary such measure turns out to equal $\bar{\mu}$. Let $\bar{\nu} = \nu_\rho$ be this unique measure. Second, $\mu S(t) \to \nu_\rho$ because an arbitrary limit point $\bar{\mu}$ is equal to ν_ρ.

(b) By adjusting the argument above, we get the ergodicity of ν_ρ. Let again $\nu_\rho = \int_{\mathcal{S}_e} \lambda\, \Gamma(d\lambda)$ be the ergodic decomposition. For an ergodic λ, $\lambda S(t)$ is also ergodic by the proof of Lemma 7.4. Since ν_ρ is invariant,

$$\nu_\rho = \nu_\rho S(t) = \int_{\mathcal{S}_e} \lambda S(t)\, \Gamma(d\lambda).$$

Thus the distribution of $\lambda S(t)$ under Γ is also an ergodic decomposition for ν_ρ. By the uniqueness of the ergodic decomposition, $\lambda S(t)$ has the same distribution as λ. And $\lambda S(t) \to \nu_\rho$ for Γ-almost every λ by the argument for part (a). All this implies that Γ has to be in fact supported on the singleton $\{\nu_\rho\}$, and then ν_ρ itself must be ergodic. To see this last point, take a bounded continuous function Ψ on the space $\mathcal{M}_1$ of probability measures on X, and pass to the $t \to \infty$ limit in this integral:

$$\int \Psi(\lambda)\, \Gamma(d\lambda) = \int \Psi(\lambda S(t))\, \Gamma(d\lambda) \xrightarrow[t\to\infty]{} \Psi(\nu_\rho).$$

(c) The inequality $\nu_{\rho_1} \le \nu_{\rho_2}$ for $\rho_1 < \rho_2$ in H follows from Corollary 7.10(a). ∎

Proposition 7.13 (a) *For any ergodic μ, all limit points of $\mu S(t)$ as $t \to \infty$ lie in $\mathcal{I} \cap \mathcal{S}$.*

(b) *The set H is closed.*

(c) *Suppose μ is ergodic with density $\rho \notin H$. Let*

$$\rho_* = \sup\{h \in H : h < \rho\} \quad \text{and} \quad \rho^* = \inf\{h \in H : h > \rho\}.$$

Let $\alpha = (\rho^* - \rho)/(\rho^* - \rho_*)$. Then $\mu S(t) \to \alpha \nu_{\rho_*} + (1-\alpha)\nu_{\rho^*}$ as $t \to \infty$.

Proof. (a) Let $\bar{\mu}$ be a limit point, realized along the sequence $\mu S(t_n) \to \bar{\mu}$. $\mu S(s)$ is also an ergodic measure with the same density as μ, so by Corollary 7.10(b) $\mu S(t_n + s) \to \bar{\mu}$. Consequently

$$\bar{\mu} = \lim_{n\to\infty} \mu S(t_n + s) = \left\{ \lim_{n\to\infty} \mu S(t_n) \right\} S(s) = \bar{\mu} S(s),$$

where we used the semigroup property $S(t_n + s) = S(t_n)S(s)$ and the continuity of the operator $S(s)$ on measures. This says $\bar{\mu} \in \mathcal{I}$. The other part $\bar{\mu} \in \mathcal{S}$ follows because $\mu S(t_n) \in \mathcal{S}$ for all t_n, and $\mathcal{S}$ is a weakly closed set by the continuity of the spatial translations.

(b)–(c) Suppose μ is an ergodic measure with density $\rho \notin H$. Let $\bar{\mu}$ be a limit point as above. By Choquet's theorem A.13 there exists a probability measure Λ on $(\mathcal{I} \cap \mathcal{S})_e$ such that

$$\bar{\mu} = \int_{(\mathcal{I}\cap\mathcal{S})_e} \lambda \, \Lambda(d\lambda).$$

Let $M(\lambda) = \int \eta(0)\, d\lambda$ denote the function that maps a probability measure on X to its density. Define the probability measure γ on $[0, K]$ by

$$\gamma(B) = \Lambda(M^{-1}(B)) = \Lambda\left\{\lambda : \textstyle\int \eta(0)\, d\lambda \in B\right\}.$$

We claim that γ is supported on the two points $\{\rho_*, \rho^*\}$, which in particular then have to be elements of H.

Suppose $\gamma[h_2, K] > 0$ for some $h_2 > \rho^*$. Recall the definition of A_∞ and Lemma 7.11.

$$\begin{aligned} \bar{\mu}\{A_\infty \geq h_2\} &= \int \lambda\{A_\infty \geq h_2\}\, \Lambda(d\lambda) = \int_{M(\lambda)\in[h_2,K]} \lambda\{A_\infty \geq h_2\}\, \Lambda(d\lambda) \\ &= \Lambda\{\lambda : M(\lambda) \in [h_2, K]\} = \gamma[h_2, K] > 0. \end{aligned}$$

Pick $h_1 \in H$ such that $\rho^* \leq h_1 < h_2$. By Corollary 7.10(a) $\bar{\mu} \leq \nu_{h_1}$. $A_n(\eta)$ is an increasing cylinder function, converges to $A_\infty(\eta)$ $\bar{\mu}$-almost surely, and so for $h = (h_1 + h_2)/2$,

$$\bar{\mu}\{A_\infty \geq h_2\} \leq \liminf_{n\to\infty} \bar{\mu}\{A_n \geq h\} \leq \liminf_{n\to\infty} \nu_{h_1}\{A_n \geq h\} = 0.$$

This contradicts what we obtained a moment earlier. Consequently $\gamma[h_2, K] = 0$ for all $h_2 > \rho^*$, which implies $\gamma(\rho^*, K] = 0$. An analogous argument gives $\gamma[0, \rho_*) = 0$. And since by assumption $(\rho_*, \rho^*) \cap H = \emptyset$, we conclude that $\gamma(\{\rho_*, \rho^*\}^c) = 0$.

The actual γ-masses on the points $\{\rho_*, \rho^*\}$ are determined by the density, namely by

$$\begin{aligned}\rho &= \int \eta(0)\, d\bar{\mu} = \int M(\lambda)\, \Lambda(d\lambda) = \int_{[0,K]} h\, \gamma(dh) = \rho_* \gamma\{\rho_*\} + \rho^* \gamma\{\rho^*\} \\ &= \alpha\rho_* + (1-\alpha)\rho^*.\end{aligned}$$

We have proved (b). To see this, suppose ρ is a limit point of H but not in H. Then ρ must equal either ρ_* or ρ^*. Find an ergodic measure μ with density ρ (for example, an i.i.d. measure) and let $\bar{\mu}$ be a limit point of $\mu S(t)$ (it exists by compactness). The above argument shows that the measure γ on $[0, K]$ must be supported on $\{\rho_*, \rho^*\}$. Thereby ρ must lie in H.

We have also proved (c). Since an arbitrary limit point $\bar{\mu}$ was identified as $\alpha\nu_{\rho_*} + (1-\alpha)\nu_{\rho^*}$, the measures $\mu S(t)$ converge to this limit. ∎

Propositions 7.12 and 7.13 complete the proof of Theorem 7.1.

Exercise 7.3 Let $1 \leq z_1 < z_2 < z_3 < \cdots$ be positive integers, R a positive integer, and assume

$$\frac{1}{n} \cdot \max\{\ell : z_\ell \leq Rn\} \to c > 0 \quad \text{as } n \to \infty.$$

Show that $\ell^{-1} z_\ell \to R/c$.

Notes

The ideas for this section come from the unpublished manuscript of Ekhaus and Gray [12].

For symmetric K-exclusion in arbitrary dimension with an irreducible jump kernel, Keisling [23] proved that $(\mathcal{I} \cap \mathcal{S})_e = \{\nu_\rho : 0 \leq \rho \leq K\}$ and that the process converges to a mixture of the ν_ρ's from a spatially invariant initial distribution. Convergence to an extremal ν_ρ from a spatially ergodic initial distribution has not been proved. Keisling's proof is a generalization of one given by Liggett for $K = 1$ exclusion, and relies on couplings and generator calculations.

PART III Hydrodynamic limits

Hydrodynamic limits describe the behavior of an interacting system over large space and time scales when the initial particle density varies in space. In Part II we discussed the convergence over time of processes with translation invariant initial distributions. The translation invariance was preserved under the evolution. The percolation argument showed that the global particle density was preserved. In other words, the dynamics preserves a constant density profile. Here we ask what happens over time to a nonconstant density profile.

For the reader who is new to hydrodynamic limits, Exercises 8.2 to 8.5 at the end of Chapter 8 give useful experience with simple situations.

8 Symmetric exclusion process

In this chapter we look at the large scale behavior of the symmetric exclusion process on the lattice $\mathbf{Z}^d$ with jump kernel $p(u,v)$. The assumptions on $p(u,v)$ are those used in the construction in Chapter 2. Namely, in addition to symmetry

$$p(u,v) = p(v,u),$$

we assume translation invariance

$$p(u,v) = p(0, v-u),$$

and finite range:

$$p(0,v) = 0 \text{ for } |v|_1 > R$$

for some fixed finite number R. Chapter 2 gave two constructions of this symmetric exclusion process, the graphical representation in Section 2.1 that works for all exclusion processes, and the stirring particle construction in Section 2.2 that was special for the symmetric case. Both constructions produce a process $\eta_t = (\eta_t(u) : u \in \mathbf{Z}^d)$ with state space $X = \{0,1\}^{\mathbf{Z}^d}$ and generator

$$Lf(\eta) = \tfrac{1}{2} \sum_{u,v \in \mathbf{Z}^d} p(u,v)[f(\eta^{u,v}) - f(\eta)] \tag{8.1}$$

where

$$\eta^{u,v}(w) = \begin{cases} \eta(v), & w = u \\ \eta(u), & w = v \\ \eta(w), & w \notin \{u,v\} \end{cases} \tag{8.2}$$

is the configuration that results from exchanging the contents at sites u and v. Formula (8.1) for the generator is not exactly like either of the formulas (2.12) or (2.17) given in Chapter 2, but can be deduced from them.

Assume we have a sequence η_t^n, $n = 1, 2, 3, \ldots$, of these symmetric exclusion processes. These processes may be defined on separate probability spaces or on one common probability space. All statements involve only distributions so the exact definition of the processes is immaterial for the discussion. For each n, we shall use the generic P for the probability measure on the probability space on which process η_t^n is defined.

The hydrodynamic limit concerns the occupation measure

$$\pi_t^n = n^{-d} \sum_{u \in \mathbf{Z}^d} \eta_t^n(u) \delta_{u/n} \tag{8.3}$$

of the exclusion process. The symbol δ_x represents a unit mass placed at the point x in $\mathbf{R}^d$. In probabilistic terms, δ_x is the degenerate probability distribution defined for Borel sets A by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Thus as a measure π_t^n places mass n^{-d} at point u/n in $\mathbf{R}^d$ if site u in $\mathbf{Z}^d$ is occupied by a particle at time t. Technically speaking π_t^n is a random Radon measure on $\mathbf{R}^d$. The term Radon measure means that π_t^n is a nonnegative Borel measure whose total mass may be infinite but which gives finite mass to bounded sets.

The space of Radon measures on $\mathbf{R}^d$ is denoted by $\mathbf{M}$. $\mathbf{M}$ is a Polish space with its so-called vague topology. (See Section A.10 in the appendix.) Convergence in this topology (vague convergence) is defined in terms of convergence of integrals of compactly supported continuous functions. The same convergence can also be defined by test functions $\phi \in C_c^\infty(\mathbf{R}^d)$, the space of compactly supported infinitely differentiable functions. So our goal is to derive laws of large numbers for averages of the type

$$\int_{\mathbf{R}^d} \phi(x)\, \pi_t^n(dx) = n^{-d} \sum_{u \in \mathbf{Z}^d} \eta_t^n(u) \phi\big(\tfrac{u}{n}\big) \tag{8.4}$$

for test functions $\phi \in C_c^\infty(\mathbf{R}^d)$. However, as it stands, the average above lacks a crucial ingredient. Since we shrunk lattice distance to n^{-1}, a symmetric random walk moves only a distance of order $t^{1/2}n^{-1}$ in a fixed time t. Thus to see any motion in the limit $n \to \infty$, the space scaling has to be matched with the time scaling $n^2 t$. See Exercise 8.6 in this context.

The basic hypothesis for the hydrodynamic limit is the existence of a macroscopic density profile at time 0. Let $0 \le \rho_0(x) \le 1$ be a given bounded measurable function on $\mathbf{R}$. Assume that for all $\phi \in C_c^\infty(\mathbf{R}^d)$ and all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ \left| n^{-d} \sum_{u\in\mathbf{Z}^d} \eta_0^n(u)\phi\big(\tfrac{u}{n}\big) - \int_{\mathbf{R}^d} \rho_0(x)\phi(x)\,dx \right| > \varepsilon \right\} = 0. \tag{8.5}$$

We also make an ellipticity hypothesis on the random walk kernel $p(u, v)$. The term ellipticity is borrowed here from p.d.e. theory (see Exercise 8.8). The assumption is that the possible jump directions span the d-dimensional space. Precisely,

$$\begin{aligned} &\text{for every } x \in \mathbf{R}^d \text{ such that } x \neq 0 \text{ there exists a} \\ &u \in \mathbf{Z}^d \text{ such that } u \cdot x \neq 0 \text{ and } p(0, u) > 0. \end{aligned} \tag{8.6}$$

Write $u = (u_1, \dots, u_d)$ to represent a site $u \in \mathbf{Z}^d$ in terms of its coordinates. Define the covariance matrix $\Sigma = (\sigma_{i,j})_{1\le i,j\le d}$ of the jump kernel $p(0, u)$ by

$$\sigma_{i,j} = \sum_{u\in\mathbf{Z}^d} p(0, u)u_iu_j.$$

Assumption (8.6) is equivalent to the requirement that Σ is nonsingular, or equivalently, that Σ has strictly positive eigenvalues (Exercise 8.7). If assumption (8.6) fails, the additive subgroup of $\mathbf{Z}^d$ generated by the support of $p(0, u)$ is isomorphic to $\mathbf{Z}^k$ for some $k < d$, and the problem can be restated in a lower dimensional space. (See Proposition P1 in Section 7 of [43].)

The result we are looking for is a weak law of large numbers for $\pi^n_{n^2t}$, of the general type that appears in assumption (8.5). The density function $\rho(x, t)$ that will appear in the theorem has an explicit formula. Define the function $\rho(x, t)$ on $\mathbf{R}^d \times [0, \infty)$ by

$$\rho(x, 0) = \rho_0(x), \tag{8.7}$$

and for $t > 0$ by

$$\rho(x, t) = (2\pi t)^{-d/2}(\det \Sigma)^{-1/2} \int_{\mathbf{R}^d} \rho_0(y) \exp\big\{-\tfrac{1}{2t}(x - y) \cdot \Sigma^{-1}(x - y)\big\}\, dy. \tag{8.8}$$

The function in the exponent is the quadratic form

$$(x - y) \cdot \Sigma^{-1}(x - y) = \sum_{1\le i,j\le d} (x_i - y_i)(\Sigma^{-1})_{i,j}(x_j - y_j).$$

The function ρ has the following properties. On $\mathbf{R}^d \times (0, \infty)$ it is infinitely differentiable and satisfies the partial differential equation $\rho_t = \frac{1}{2}A\rho$, where the differential operator A is defined by

$$A\rho = \sum_{1\le i,j\le d} \sigma_{i,j}\rho_{x_i,x_j}. \tag{8.9}$$

The initial density function ρ_0 is the vague limit of $\rho(x,t)$ as $t \to 0$, meaning that

$$\int \phi(x)\rho(x,t)dx \longrightarrow \int \phi(x)\rho_0(x)dx \qquad \text{as } t \to 0, \text{ for } \phi \in C_c(\mathbf{R}^d).$$

If ρ_0 is assumed continuous to begin with, then $\rho(x,t)$ is continuous all the way to the $t=0$ boundary. In either case we regard ρ as a solution of the initial value problem

$$\rho_t = \tfrac{1}{2}A\rho \text{ on } \mathbf{R}^d \times (0,\infty), \quad \rho(x,0) = \rho_0(x) \text{ for } x \in \mathbf{R}^d. \tag{8.10}$$

The point of this chapter is that this partial differential equation gives the macroscopic description of the dynamics whose particle-level decription is the symmetric exclusion process. The mathematically rigorous form of this idea is the next law of large numbers, called a hydrodynamic limit of the particle system.

Theorem 8.1 *Suppose a sequence of symmetric exclusion processes η_t^n satisfies assumption* (8.5). *Assume their common jump kernel $p(u,v)$ satisfies* (8.6). *Define the function ρ by* (8.7)–(8.8). *Then for every $t \geq 0$, $\phi \in C_c^\infty(\mathbf{R}^d)$, and $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\biggl\{ \biggl| n^{-d} \sum_{u\in\mathbf{Z}^d} \eta^n_{n^2t}(u)\phi\bigl(\tfrac{u}{n}\bigr) - \int_{\mathbf{R}^d} \rho(x,t)\phi(x)\,dx \biggr| \geq \varepsilon \biggr\} = 0. \tag{8.11}$$

Behind this theorem is convergence of the entire empirical measure process $(\pi^n_{n^2t} : t \geq 0)$. Let $D_{\mathbf{M}}$ denote the space of paths $\alpha : [0,\infty) \to \mathbf{M}$ that are right-continuous and have left limits everywhere, endowed with the Skorokhod topology described in Section A.2.2. Let $s_{\mathbf{M}}$ denote the Skorokhod metric on $D_{\mathbf{M}}$, defined as in recipe (A.4) in terms of the vague metric $d_{\mathbf{M}}$ on $\mathbf{M}$. Define a specific path $\bar\alpha$ in $D_{\mathbf{M}}$ by

$$\bar\alpha(t,dx) = \rho(x,t)dx$$

with ρ as defined by (8.7)–(8.8). This path $\bar\alpha$ is in fact continuous, not only right-continuous.

Define the time-scaled empirical measure process $\bar\pi^n_\cdot$ by $\bar\pi^n_t = \pi^n_{n^2t}$ for $t \geq 0$. The right-continuity of $t \mapsto \bar\pi^n_t$ follows from the right-continuity of integrals of compactly supported test functions. These integrals are finite sums of the type (8.4), and their right-continuity in t follows from the right-continuity of the exclusion process η^n_t itself. Consequently the path $\bar\pi^n_\cdot$ is an element of $D_{\mathbf{M}}$.

Theorem 8.2 *Under the assumptions of Theorem 8.1, $\bar\pi^n_\cdot$ converges to $\bar\alpha$ in probability, as $n \to \infty$. In other words, for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\bigl\{ s_{\mathbf{M}}(\bar\pi^n_\cdot, \bar\alpha) \geq \varepsilon \bigr\} = 0. \tag{8.12}$$

Theorem 8.1 will be obtained as a corollary of Theorem 8.2. To prove Theorem 8.2 we show that the process $\bar{\pi}^n_\cdot$ converges weakly in the space $D_{\mathbf{M}}$ to a weak solution of the initial value problem (8.10). This last notion we have not yet defined. Let us call a vaguely continuous Radon measure-valued path $\alpha : [0, \infty) \to \mathbf{M}$ a *weak solution* of (8.10) if the initial condition

$$\alpha(0, dx) = \rho_0(x)dx \tag{8.13}$$

is satisfied, and if for all $\phi \in C_c^\infty(\mathbf{R}^d)$,

$$\alpha(t, \phi) - \alpha(0, \phi) - \tfrac{1}{2} \int_0^t \alpha(s, A\phi)\, ds = 0. \tag{8.14}$$

The notation $\alpha(t, \phi)$ is a shorthand for the integral,

$$\alpha(t, \phi) = \int_{\mathbf{R}^d} \phi(x)\, \alpha(t, dx).$$

In Section A.11 in the appendix we show that $\bar{\alpha}$ is the unique weak solution, subject to a boundedness assumption. Thus $\bar{\pi}^n_\cdot$ converges weakly to a degenerate limit, which implies convergence in probability.

8.1 Proof of Theorems 8.1 and 8.2

The main technical part of the proof consists in showing that a certain martingale vanishes in the $n \to \infty$ limit. To estimate the mean square $E[M_t^2]$ of a martingale M_t, one seeks to find and control an increasing process A_t such that $M_t^2 - A_t$ is a martingale. General theorems about the existence of such processes can be found in Section 1.5 of [22] and in Section 2.6 in [13]. Here we consider only the basic martingales associated with Markov processes.

A martingale lemma for Markov processes

Assume for the moment a general Markov process setting. Y is a Polish space. $\{P^x\}$ is a Feller process on the path space D_Y, with a strongly continuous contraction semigroup $S(t)f(x) = E^x[f(X_t)]$ and generator L on $C_b(Y)$. X_t is the coordinate process. For any function f in the domain of L,

$$M_t = f(X_t) - \int_0^t Lf(X_s)\, ds \tag{8.15}$$

is a martingale with respect to the filtration $\mathcal{F}_t = \sigma\{X_s : 0 \le s \le t\}$ (Exercise 3.1).

Lemma 8.3 *Suppose both f and f^2 lie in the domain of the generator L. Then*

$$V_t = M_t^2 - \int_0^t \big\{ L(f^2)(X_s) - 2f(X_s)Lf(X_s) \big\}\, ds$$

is a martingale with respect to the filtration $\mathcal{F}_t$.

Proof. The hypothesis implies that f, f^2, Lf and $L(f^2)$ are bounded. So integrability is not an issue anywhere. Abbreviate

$$\gamma(s) = L(f^2)(X_s) - 2f(X_s)Lf(X_s).$$

Start by considering a small time increment.

$$\begin{aligned}
E\big[M_{s+\delta}^2 - M_s^2 \big| \mathcal{F}_s\big] &= E\big[\big(M_{s+\delta} - M_s\big)^2 \big| \mathcal{F}_s\big] \\
&= E\Bigg[\bigg(f(X_{s+\delta}) - f(X_s) - \int_s^{s+\delta} Lf(X_r)\,dr\bigg)^2 \,\Bigg|\, \mathcal{F}_s\Bigg] \\
&= E\big[\big(f(X_{s+\delta}) - f(X_s)\big)^2 \big| \mathcal{F}_s\big] - 2E\bigg[\big(f(X_{s+\delta}) - f(X_s)\big)\int_s^{s+\delta} Lf(X_r)\,dr \bigg| \mathcal{F}_s\bigg] \\
&\qquad + E\Bigg[\bigg(\int_s^{s+\delta} Lf(X_r)\,dr\bigg)^2 \,\Bigg|\, \mathcal{F}_s\Bigg] \\
&\equiv A_1(s, s+\delta) - 2A_2(s, s+\delta) + A_3(s, s+\delta).
\end{aligned}$$

That last identity is the definition of the conditional expectations $A_i(s, s+\delta)$, $i = 1, 2, 3$. The main term is the first one.

$$\begin{aligned}
A_1(s, s+\delta) &= E\big[\big(f(X_{s+\delta}) - f(X_s)\big)^2 \big| \mathcal{F}_s\big] \\
&= E\big[f(X_{s+\delta})^2 \big| \mathcal{F}_s\big] + f(X_s)^2 - 2f(X_s)E\big[f(X_{s+\delta}) \big| \mathcal{F}_s\big] \\
&= E\bigg[f(X_s)^2 + \int_s^{s+\delta} L(f^2)(X_r)\,dr \bigg| \mathcal{F}_s\bigg] + f(X_s)^2 \\
&\qquad - 2f(X_s)E\bigg[f(X_s) + \int_s^{s+\delta} Lf(X_r)\,dr \bigg| \mathcal{F}_s\bigg] \\
&= E\bigg[\int_s^{s+\delta} L(f^2)(X_r)\,dr \bigg| \mathcal{F}_s\bigg] - 2E\bigg[f(X_s)\int_s^{s+\delta} Lf(X_r)\,dr \bigg| \mathcal{F}_s\bigg] \\
&= E\bigg[\int_s^{s+\delta} \big\{L(f^2)(X_r) - 2f(X_r)Lf(X_r)\big\}\,dr \bigg| \mathcal{F}_s\bigg] \\
&\qquad + 2E\bigg[\int_s^{s+\delta} \big(f(X_r) - f(X_s)\big)Lf(X_r)\,dr \bigg| \mathcal{F}_s\bigg] \\
&\equiv E\bigg[\int_s^{s+\delta} \gamma(r)\,dr \bigg| \mathcal{F}_s\bigg] + 2A_4(s, s+\delta).
\end{aligned}$$

The last line above defines the quantity $A_4(s, s+\delta)$. Along the way we used martingales of the type (8.15) for both f and f^2. The term $A_4(s, s+\delta)$ is further expressed as follows:

$$\begin{aligned}
A_4(s, s+\delta) &= E\Big[\int_s^{s+\delta} \big(f(X_r) - f(X_s)\big) Lf(X_r)\, dr \Big| \mathcal{F}_s\Big] \\
&= E\Big[\int_s^{s+\delta} \big(f(X_{s+\delta}) - f(X_s)\big) Lf(X_r)\, dr \Big| \mathcal{F}_s\Big] \\
&\qquad - E\Big[\int_s^{s+\delta} \big(f(X_{s+\delta}) - f(X_r)\big) Lf(X_r)\, dr \Big| \mathcal{F}_s\Big] \\
&\equiv A_2(s, s+\delta) - A_5(s, s+\delta).
\end{aligned}$$

Given $s < t$, let m be a positive integer, $\delta = \frac{t-s}{m}$, and $s_i = s + i\delta$ for $i = 0, \ldots, m$. Use the estimates from above on each partition interval $[s_i, s_{i+1}]$.

$$\begin{aligned}
E\big[M_t^2 - M_s^2 \big| \mathcal{F}_s\big] &= E\Big[\sum_{i=0}^{m-1} E\big(M_{s_{i+1}}^2 - M_{s_i}^2 \big| \mathcal{F}_{s_i}\big) \Big| \mathcal{F}_s\Big] \\
= E\Big[\int_s^t \gamma(r)\, dr \Big| \mathcal{F}_s\Big] &+ E\Big[\sum_{i=0}^{m-1} A_3(s_i, s_{i+1}) \Big| \mathcal{F}_s\Big] - 2E\Big[\sum_{i=0}^{m-1} A_5(s_i, s_{i+1}) \Big| \mathcal{F}_s\Big].
\end{aligned}$$

It remains to show that the last two conditional expectations vanish as $m \to \infty$ or equivalently $\delta \to 0$. The uniform bound

$$|A_3(s_i, s_{i+1})| \le \delta^2 \|Lf\|_\infty^2$$

takes care of the next to last term. And finally,

$$\begin{aligned}
E\Big[\sum_{i=0}^{m-1} A_5(s_i, s_{i+1}) \Big| \mathcal{F}_s\Big] &= E\Big[\sum_{i=0}^{m-1} \int_{s_i}^{s_{i+1}} \big(f(X_{s_{i+1}}) - f(X_r)\big) Lf(X_r)\, dr \Big| \mathcal{F}_s\Big] \\
&= E\Big[\int_0^t \sum_{i=0}^{m-1} \mathbf{1}_{(s_i, s_{i+1}]}(r) \big(f(X_{s_{i+1}}) - f(X_r)\big) Lf(X_r)\, dr \Big| \mathcal{F}_s\Big].
\end{aligned}$$

The integrand

$$\sum_{i=0}^{m-1} \mathbf{1}_{(s_i, s_{i+1}]}(r) \big(f(X_{s_{i+1}}) - f(X_r)\big) Lf(X_r)$$

is uniformly bounded, also over the index m. As $m \to \infty$ it converges to zero for each fixed r and path of the process, by the right continuity of the path $r \mapsto X_r$. Consequently the conditional expectation converges to zero almost surely as $m \to \infty$. Here we need a conditional dominated convergence theorem. The reader can derive such a theorem without too much trouble in the same way that the unconditional dominated convergence theorem is proved. Start with the conditional monotone convergence theorem (property 1.1c in Section 4.1 of [11]), prove a conditional Fatou Lemma, and from that a conditional dominated convergence theorem.

We have proved that

$$E\big[M_t^2 - M_s^2 \big| \mathcal{F}_s\big] = E\bigg[\int_s^t \gamma(r)\,dr\bigg|\mathcal{F}_s\bigg]$$

which is equivalent to the conclusion of the lemma. ∎

Exercise* 8.1 Check that the statement of the lemma would not change if we defined the martingale as

$$M_t = f(X_t) - f(X_0) - \int_0^t Lf(X_s)\,ds. \tag{8.16}$$

Return to the proof of Theorem 8.1

For the proof we move the discussion to the path space D_X of the exclusion process. Let P^n be the distribution of the nth exclusion process $(\eta_t^n : t \ge 0)$. E^n denotes expectation under measure P^n. As before, η_t denotes the coordinate process on the space D_X. The empirical measures π_t^n are **M**-valued measurable functions on D_X, defined in terms of test functions ϕ by

$$\pi_t^n(\phi) = n^{-d} \sum_{u\in\mathbf{Z}^d} \eta_t(u)\phi\big(\tfrac{u}{n}\big).$$

The distribution of this process $\pi_\cdot^n$ under the measure P^n is the same as the distribution of the process $\pi_\cdot^n$ earlier defined by (8.3) in terms of process η_t^n.

Fix a test function $\phi \in C_c^\infty(\mathbf{R}^d)$. The average

$$f(\eta) = n^{-d} \sum_{u\in\mathbf{Z}^d} \eta(u)\phi\big(\tfrac{u}{n}\big). \tag{8.17}$$

is a cylinder function of η. (Dependence on n will be suppressed from the notation f.) With this f, define the martingale

$$M_n(t) = f(\eta_t) - f(\eta_0) - \int_0^t Lf(\eta_s)\,ds. \tag{8.18}$$

We first find the main contribution to the integrand $Lf(\eta_s)$. Let $\eta \in X$ be arbitrary.

$$\begin{aligned}
Lf(\eta) &= \frac{1}{2}\sum_{u,z} p(0,z)[f(\eta^{u,u+z}) - f(\eta)] \\
&= \frac{n^{-d}}{2}\sum_{u,z} p(0,z)\big\{\eta(u+z)\phi\big(\tfrac{u}{n}\big) + \eta(u)\phi\big(\tfrac{u+z}{n}\big) - \eta(u+z)\phi\big(\tfrac{u+z}{n}\big) - \eta(u)\phi\big(\tfrac{u}{n}\big)\big\} \\
&= \frac{n^{-d}}{2}\sum_{u,z} p(0,z)\eta(u)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} + \frac{n^{-d}}{2}\sum_{u,z} p(0,z)\eta(u+z)\big\{\phi\big(\tfrac{u}{n}\big) - \phi\big(\tfrac{u+z}{n}\big)\big\} \\
&= \frac{n^{-d}}{2}\sum_{u,z} p(0,z)\eta(u)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} + \frac{n^{-d}}{2}\sum_{u,z} p(0,z)\eta(u)\big\{\phi\big(\tfrac{u-z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} \\
&= \frac{n^{-d}}{2}\sum_{u} \eta(u)\sum_{z} p(0,z)\big\{\phi\big(\tfrac{u+z}{n}\big) + \phi\big(\tfrac{u-z}{n}\big) - 2\phi\big(\tfrac{u}{n}\big)\big\}. \qquad (8.19)
\end{aligned}$$

Taylor's theorem (see for example page 378 in [45]) gives the expansion

$$\begin{aligned}
\phi\Big(\frac{u \pm z}{n}\Big) = \phi\Big(\frac{u}{n}\Big) \pm \nabla\phi\Big(\frac{u}{n}\Big)\cdot\frac{z}{n} + \frac{1}{2}\sum_{1\le i,j\le d}\phi_{x_i,x_j}\Big(\frac{u}{n}\Big)\frac{z_i z_j}{n^2} \\
\pm \frac{1}{6}\sum_{1\le i,j,k\le d}\phi_{x_i,x_j,x_k}\Big(\frac{u \pm \theta^{\pm} z}{n}\Big)\frac{z_i z_j z_k}{n^3}
\end{aligned}$$

for some numbers $0 < \theta^{\pm} < 1$. Substitute this into the sum on line (8.19). Cancel the $\phi\big(\frac{u}{n}\big)$ and $\nabla\phi\big(\frac{u}{n}\big)$ terms. The second order terms contribute

$$\frac{n^{-d-2}}{2}\sum_{u}\eta(u)\sum_{1\le i,j\le d}\Big(\sum_{z} p(0,z)z_i z_j\Big)\phi_{x_i,x_j}\big(\tfrac{u}{n}\big) = \frac{n^{-d-2}}{2}\sum_{u}\eta(u)A\phi\big(\tfrac{u}{n}\big).$$

Hence from above comes

$$\begin{aligned}
Lf(\eta) &= \frac{n^{-d-2}}{2}\sum_{u}\eta(u)A\phi\big(\tfrac{u}{n}\big) \\
&+ \frac{n^{-d-3}}{6}\sum_{u}\eta(u)\sum_{1\le i,j,k\le d}\sum_{z} p(0,z)z_i z_j z_k\Big\{\phi_{x_i,x_j,x_k}\Big(\frac{u+\theta^{+}z}{n}\Big) - \phi_{x_i,x_j,x_k}\Big(\frac{u-\theta^{-}z}{n}\Big)\Big\}.
\end{aligned}$$

Since $p(0,z)$ has finite support in $\mathbf{Z}^d$, ϕ has bounded support in $\mathbf{R}^d$, and $\theta^{\pm} \in (0,1)$, there is a constant C such that the sum

$$\sum_{1\le i,j,k\le d}\sum_{z} p(0,z)z_i z_j z_k\Big\{\phi_{x_i,x_j,x_k}\Big(\frac{u+\theta^{+}z}{n}\Big) - \phi_{x_i,x_j,x_k}\Big(\frac{u-\theta^{-}z}{n}\Big)\Big\}$$

is nonzero for at most Cn^d sites u. Furthermore, this sum is uniformly bounded over u because all derivatives of ϕ are bounded by the standing assumption $\phi \in C_c^\infty(\mathbf{R}^d)$. Combining all this gives the estimate

$$Lf(\eta) = \frac{n^{-d-2}}{2} \sum_u \eta(u) A\phi\big(\tfrac{u}{n}\big) + O(n^{-3}) \tag{8.20}$$

uniformly over η. Replace η by η_t, and note that $f(\eta_t)$ is precisely the integral $\pi_t^n(\phi)$, while

$$Lf(\eta_s) = \frac{n^{-d-2}}{2} \sum_u \eta_s(u) A\phi\big(\tfrac{u}{n}\big) + O(n^{-3}) = \frac{n^{-2}}{2} \pi_s^n(A\phi) + O(n^{-3}). \tag{8.21}$$

Replace t by $n^2 t$ in definition (8.18) and substitute in (8.21). Change variables in the time integral to get

$$\pi_{n^2 t}^n(\phi) - \pi_0^n(\phi) - \tfrac{1}{2} \int_0^t \pi_{n^2 s}^n(A\phi)\, ds = M_n(n^2 t) + O(tn^{-1}). \tag{8.22}$$

The error $O(tn^{-1})$ results from integrating $O(n^{-3})$ over the time interval $[0, n^2 t]$, and is uniform over the path space D_X. Note how the change of variable cancelled the powers of n in front of the time integral.

We come to the point where we show the vanishing of the martingale.

Lemma 8.4 *For any* $0 < T < \infty$,

$$\lim_{n\to\infty} E^n\Big[\sup_{0 \le t \le T} M_n^2(n^2 t)\Big] = 0. \tag{8.23}$$

Proof. We estimate

$$\gamma(s) = L(f^2)(\eta_s) - 2f(\eta_s)Lf(\eta_s)$$

for the function f in (8.17). First note that for a general f in the domain of the generator,

$$\begin{aligned} &L(f^2)(\eta) - 2f(\eta)Lf(\eta) \\ &\quad = \frac{1}{2} \sum_{u,z} p(0,z) \big\{ f(\eta^{u,u+z})^2 - f(\eta)^2 - 2f(\eta)[f(\eta^{u,u+z}) - f(\eta)] \big\} \\ &\quad = \frac{1}{2} \sum_{u,z} p(0,z) \big(f(\eta^{u,u+z}) - f(\eta) \big)^2. \end{aligned}$$

For the particular f defined by (8.17) and for fixed u and z, Lipscitz continuity of ϕ gives

$$\begin{aligned} &f(\eta^{u,u+z}) - f(\eta) \\ &\quad = n^{-d} \big\{ \eta(u+z)\phi\big(\tfrac{u}{n}\big) + \eta(u)\phi\big(\tfrac{u+z}{n}\big) - \eta(u+z)\phi\big(\tfrac{u+z}{n}\big) - \eta(u)\phi\big(\tfrac{u}{n}\big) \big\} \\ &\quad = n^{-d} \big(\eta(u) - \eta(u+z) \big) \big\{ \phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big) \big\} \\ &\quad = O\big(n^{-d-1} |z|_1 \big). \end{aligned}$$

This estimate is applied to those sites u such that either $\frac{u}{n}$ lies in the support of ϕ, or $\frac{u+z}{n}$ lies in the support of ϕ for some z such that $p(0,z) > 0$. Let Λ_n be the set of sites u for which this happens. For u outside Λ_n the sum

$$\sum_z p(0,z)\big(f(\eta^{u,u+z}) - f(\eta)\big)^2$$

vanishes for all η. $|\Lambda_n| \le Cn^d$ as noted earlier, by the bounded support of ϕ and the finite range of $p(0,z)$. Consequently

$$\begin{aligned}\gamma(s) &= \frac{1}{2}\sum_{u\in\Lambda_n}\sum_z p(0,z)\big(f(\eta_s^{u,u+z}) - f(\eta_s)\big)^2\\ &\le Cn^{-2d-2}\sum_{u\in\Lambda_n}\sum_z p(0,z)|z|_1^2\\ &\le Cn^{-d-2}.\end{aligned}$$

Above we used the common practice of letting C denote a constant whose actual value may change from line to line, but which does not depend on the parameter n. The "final C" incorporates the moment $\sum p(0,z)|z|_1^2$, the size of the support of ϕ, the Lipschitz constant of ϕ, and miscellaneous factors such as the $\frac{1}{2}$ that was initially at the front of the generator. Note that $\gamma(s)$ is nonnegative so we have

$$|\gamma(s)| \le Cn^{-d-2} \tag{8.24}$$

uniformly over time s and over the path space.

Next apply Doob's maximal inequality to the martingale $M_n(t)$, then the fact that

$$V_n(t) = M_n^2(t) - \int_0^t \gamma(s)\,ds$$

is a martingale (Lemma 8.3), and finally the bound (8.24) on $\gamma(s)$. These steps lead to

$$E^n\big[\sup_{0\le t\le T} M_n^2(n^2t)\big] \le 4E^n\big[M_n^2(n^2T)\big] = 4E^n\left[\int_0^{n^2T}\gamma(s)\,ds\right] \le CTn^{-d}.$$

Doob's maximal inequality for discrete-time martingales is a standard part of probability texts, see for example Section 4.4 in [11]. From discrete time it extends readily to right-continuous martingales in continuous time, as we have here. This extension can be found in Section 1.3 in [22].

We have proved (8.23). ∎

Let $\bar\pi^n_t = \pi^n_{n^2t}$ denote the time-scaled empirical measure. Rewrite (8.22) in the form

$$\bar\pi^n_t(\phi) - \bar\pi^n_0(\phi) - \tfrac{1}{2}\int_0^t \bar\pi^n_s(A\phi)\,ds = M_n(n^2t) + O(tn^{-1}). \tag{8.25}$$

By (8.23) the right-hand side of this equation vanishes as $n \to \infty$. Comparison with (8.14) shows that the measure $\bar\pi^n_t$ is on its way to becoming a weak solution of the p.d.e. The rest is technical. We need to find the right context to establish convergence. Weak convergence at process level comes naturally from the martingale estimate (8.23).

The process $\bar\pi^n_\cdot = (\bar\pi^n_t : t \ge 0)$ has its paths in the space $D_{\mathbf{M}}$ of RCLL functions from $[0,\infty)$ into $\mathbf{M}$. Let Q^n be the distribution of this process, defined for Borel sets $B \subseteq D_{\mathbf{M}}$ by

$$Q^n(B) = P^n\{\bar\pi^n_\cdot \in B\}.$$

E^{Q^n} denotes expectation under Q^n. We write $\alpha = (\alpha(t) : t \ge 0)$ for a generic element of $D_{\mathbf{M}}$.

Lemma 8.5 *The sequence of probability measures $\{Q^n\}$ is tight on $D_{\mathbf{M}}$.*

Proof. By Theorem A.3 we need to check two things.

(a) Compact containment: for each time $0 \le t < \infty$ and $\varepsilon > 0$ there exists a compact set $K \subseteq \mathbf{M}$ such that

$$\inf_n P^n\{\pi^n_{n^2t} \in K\} > 1 - \varepsilon. \tag{8.26}$$

(b) Modulus of continuity: for every $\varepsilon > 0$ and $0 < T < \infty$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} P^n\{w'(\bar\pi^n_\cdot, \delta, T) \ge \varepsilon\} \le \varepsilon. \tag{8.27}$$

The modulus of continuity is defined by

$$w'(\alpha,\delta,T) = \inf_{\{t_i\}} \sup\{d_{\mathbf{M}}(\alpha(s),\alpha(t)) : s,t \in [t_{i-1},t_i) \text{ for some } i\}$$

where the infimum is over finite partitions $0 = t_0 < t_1 < \cdots < t_{N-1} < T \le t_N$ that satisfy $\min_{1\le i\le N}(t_i - t_{i-1}) > \delta$. The metric $d_{\mathbf{M}}$ on Radon measures is defined by

$$d_{\mathbf{M}}(\mu,\nu) = \sum_{j=1}^\infty 2^{-j}\Big(1 \wedge \Big|\int \phi_j\,d\mu - \int \phi_j\,d\nu\Big|\Big)$$

with an appropriately chosen sequence of functions $\phi_j \in C_c^\infty(\mathbf{R}^d)$ (Section A.10).

Part (a). Compact containment is immediate. Let K be the set of Radon measures μ such that for each $k \in \mathbf{N}$,

$$\mu([-k,k]^d) \le (2k+1)^d.$$

K is a compact subset of $\mathbf{M}$ by Proposition A.25. $(2kn+1)^d$ points of $n^{-1}\mathbf{Z}^d$ lie in the cube $[-k,k]^d$. With at most n^{-d} mass per point, the empirical measure of the exclusion process satisfies

$$\pi^n_{n^2t}([-k,k]^d) \le n^{-d}(2kn+1)^d \le (2k+1)^d. \tag{8.28}$$

Thus $P^n\{\pi^n_{n^2t} \in K\} = 1$.

Part (b). We can do with the simpler modulus of continuity

$$w(\alpha,\delta,T) = \sup\{d_{\mathbf{M}}(\alpha(s),\alpha(t)) : s,t \in [0,T], |s-t| \le \delta\}.$$

We leave it to the reader to check that

$$w'(\alpha,\delta,T) \le w(\alpha,2\delta,T+2\delta) \le w(\alpha,2\delta,T+1),$$

the last inequality valid if $\delta \le 1/2$. Further, by the definition (A.20) of the metric $d_{\mathbf{M}}$,

$$\begin{aligned} w(\alpha,2\delta,T+1) &= \sup_{\substack{|t-s|\le 2\delta \\ s,t\in[0,T+1]}} \sum_{j=1}^{\infty} 2^{-j}\Big(1 \wedge |\alpha(s,\phi_j) - \alpha(t,\phi_j)|\Big) \\ &\le 2^{-m} + \sum_{j=1}^{m} \sup_{\substack{|t-s|\le 2\delta \\ s,t\in[0,T+1]}} |\alpha(s,\phi_j) - \alpha(t,\phi_j)|. \end{aligned}$$

Combining these inequalities, we have for $0 < \delta \le 1/2$,

$$E^n[w'(\bar{\pi}^n_\cdot,\delta,T)] \le 2^{-m} + \sum_{j=1}^{m} E^n\Big[\sup_{\substack{|t-s|\le 2\delta \\ s,t\in[0,T+1]}} |\bar{\pi}^n_s(\phi_j) - \bar{\pi}^n_t(\phi_j)|^2\Big]^{1/2}.$$

We show that each term in the last sum vanishes as first $n \to \infty$ and then $\delta \to 0$. Then we have

$$\limsup_{\delta\searrow 0}\limsup_{n\to\infty} E^n[w'(\bar{\pi}^n_\cdot,\delta,T)] \le 2^{-m}.$$

Since m can be taken arbitrarily large, (8.27) follows by Chebychev's inequality.

So it remains to show that for fixed $0 < T < \infty$ and $\phi \in C_c^\infty(\mathbf{R}^d)$

$$\lim_{\delta\searrow 0}\limsup_{n\to\infty} E^n\Big[\sup_{\substack{|s-t|\le\delta \\ s,t\in[0,T]}} |\pi^n_{n^2t}(\phi) - \pi^n_{n^2s}(\phi)|^2\Big] = 0. \tag{8.29}$$

From (8.22)

$$\pi^n_{n^2t}(\phi) - \pi^n_{n^2s}(\phi) = M_n(n^2t) - M_n(n^2s) + \tfrac{1}{2}\int_s^t \pi^n_{n^2r}(A\phi)\,dr + O(n^{-1}),$$

and we get the bound

$$\begin{aligned}
&\sup_{\substack{|s-t|\le\delta\\ s,t\in[0,T]}} |\pi^n_{n^2t}(\phi) - \pi^n_{n^2s}(\phi)|^2\\
&\qquad \le \sup_{\substack{|s-t|\le\delta\\ s,t\in[0,T]}} 4\big(M_n(n^2t) - M_n(n^2s)\big)^2 + \sup_{|s-t|\le\delta}\left(\int_s^t \pi^n_{n^2r}(A\phi)\,dr\right)^2 + O(n^{-2})\\
&\qquad \le \sup_{0\le t\le T} 16M_n^2(n^2t) + C^2\delta^2 + O(n^{-2}).
\end{aligned}$$

The constant C is a uniform bound on $\pi^n_{n^2r}(A\phi)$ which exists because $A\phi$ is bounded and compactly supported. An application of (8.23) finishes the proof of (8.29). ∎

Since $\{Q^n\}$ is a tight sequence of distributions, to establish weak convergence it is enough to show that all limit points coincide. Let Q be an arbitrary limit point of $\{Q^n\}$, and fix a subsequence n_j such that $Q^{n_j} \to Q$ weakly on the space $D_{\mathbf{M}}$. Let $C_{\mathbf{M}}$ be the subset of continuous paths in the Skorohod space $D_{\mathbf{M}}$.

Lemma 8.6 *$Q(C_{\mathbf{M}}) = 1$.*

Proof. For paths $\alpha \in D_{\mathbf{M}}$, let

$$G(\alpha) = \sup_{t\ge0} e^{-t} d_{\mathbf{M}}(\alpha(t), \alpha(t-)).$$

$G(\alpha) = 0$ iff $\alpha \in C_{\mathbf{M}}$. Lemma A.2 can be used to show that G is a continuous function on $D_{\mathbf{M}}$ (Exercise 8.9). G is bounded by the definition of the metric $d_{\mathbf{M}}$. Consequently

$$E^Q\big[G\big] = \lim_{j\to\infty} E^{Q^{n_j}}\big[G\big] = \lim_{j\to\infty} E^{n_j}\big[G(\bar\pi^{n_j}_\cdot)\big]. \tag{8.30}$$

Since the metric $d_{\mathbf{M}}$ is bounded by one,

$$G(\bar\pi^n_\cdot) \le \sup_{0\le t\le T} d_{\mathbf{M}}(\bar\pi^n_t, \bar\pi^n_{t-}) + e^{-T} \le \sup_{0\le s,t\le T:|t-s|\le\delta} d_{\mathbf{M}}(\bar\pi^n_t, \bar\pi^n_s) + e^{-T}.$$

Reasoning as in the previous proof, limit (8.29) implies that the limit on the right in (8.30) is zero. Consequently $Q\{G=0\} = 1$, in other words Q-almost every path is continuous. ∎

Lemma 8.7 *Let $\phi \in C_c^\infty(\mathbf{R}^d)$, $\delta > 0$ and $0 < T < \infty$. The set*

$$H = \left\{ \alpha \in D_{\mathbf{M}} : \sup_{0 \le t < T} \left| \alpha(t, \phi) - \alpha(0, \phi) - \tfrac{1}{2} \int_0^t \alpha(s, A\phi)\, ds \right| \le \delta \right\} \tag{8.31}$$

is closed in the path space $D_{\mathbf{M}}$.

Proof. The function $s \mapsto \alpha(s, A\phi)$ is a real-valued RCLL path, hence bounded on $[0, T]$ by property (A.8) of Skorohod space. Consequently the integral $\int_0^t \alpha(s, A\phi)\, ds$ is finite for any path $\alpha \in D_{\mathbf{M}}$.

Now we prove that H is a closed set. Note the strict inequality $t < T$ in the supremum in the definition of H. Suppose $\alpha_j \in H$ and $\alpha_j \to \alpha$ in $D_{\mathbf{M}}$. By Lemma A.2 there exist strictly increasing Lipschitz bijections $\lambda_j : [0, \infty) \to [0, \infty)$ such that $\gamma(\lambda_j) \to 0$ and

$$\sup_{0 \le t \le T} d_{\mathbf{M}}\big(\alpha_j(\lambda_j(t)), \alpha(t)\big) \to 0. \tag{8.32}$$

Fix $t \in [0, T)$. The goal is to show

$$\left| \alpha(t, \phi) - \alpha(0, \phi) - \tfrac{1}{2} \int_0^t \alpha(s, A\phi)\, ds \right| \le \delta. \tag{8.33}$$

Once this is true for an arbitrary $t \in [0, T)$, $\alpha \in H$ has been verified.

The limit $\gamma(\lambda_j) \to 0$ implies that λ_j converges to the identity function uniformly on compact intervals. In particular, $\lambda_j(t) < T$ for large enough j. Then by the assumption $\alpha_j \in H$,

$$\left| \alpha_j(\lambda_j(t), \phi) - \alpha_j(0, \phi) - \tfrac{1}{2} \int_0^{\lambda_j(t)} \alpha_j(s, A\phi)\, ds \right| \le \delta. \tag{8.34}$$

It remains to show that the left-hand side of (8.34) converges to the left-hand side of (8.33). By limit (8.32), $\alpha_j(\lambda_j(s), \phi) \to \alpha(s, \phi)$ for both $s = 0$ and $s = t$. Note that $\lambda_j(0) = 0$ by assumption.

To get the convergence

$$\int_0^{\lambda_j(t)} \alpha_j(s, A\phi)\, ds \xrightarrow[j \to \infty]{} \int_0^t \alpha(s, A\phi)\, ds, \tag{8.35}$$

first change variables in the integral:

$$\int_0^{\lambda_j(t)} \alpha_j(s, A\phi)\, ds = \int_0^t \alpha_j(\lambda_j(s), A\phi) \lambda_j'(s)\, ds.$$

The derivative λ'_j is defined Lebesgue a.e. and the change of variable is valid by the Lipschitz continuity of λ_j (Exercise 8.10). The limit $\gamma(\lambda_j) \to 0$ implies that $\lambda'_j(s) \to 1$ in $L^\infty[0,T]$. Consequently

$$\alpha_j(\lambda_j(s), A\phi)\lambda'_j(s) \xrightarrow[j\to\infty]{} \alpha(s, A\phi)$$

for all $0 \le s \le t$. An application of inequality (A.21) from the appendix and (8.32) give, for some m and large enough j,

$$|\alpha_j(\lambda_j(s), A\phi)| \le \|A\phi\|_\infty \cdot \Big\{\alpha(s,\phi_m) + 2^m d_{\mathbf{M}}\big(\alpha_j(\lambda_j(s)), \alpha(s)\big)\Big\}.$$

This shows that $\alpha_j(\lambda_j(s), A\phi)$ is uniformly bounded over $0 \le s \le T$ and large enough j, and thus dominated convergence implies (8.35). ∎

We are in a position to complete the proofs of Theorems 8.1 and 8.2. From (8.22) and (8.23) we get the limit

$$\lim_{n\to\infty} Q^n(H) = \lim_{n\to\infty} P\bigg\{\sup_{0\le t<T}\bigg|\pi^n_{n^2 t}(\phi) - \pi^n_0(\phi) - \tfrac{1}{2}\int_0^t \pi^n_{n^2 s}(A\phi)\,ds\bigg| \le \delta\bigg\} = 1.$$

Then by weak convergence and closedness of H,

$$Q(H) \ge \limsup_{j\to\infty} Q^{n_j}(H) = 1.$$

This is true for an arbitrarily small $\delta > 0$ and an arbitrarily large $T < \infty$. So for a fixed $\phi \in C_c^\infty(\mathbf{R}^d)$ Q-almost every path α satisfies

$$\alpha(t,\phi) - \alpha(0,\phi) - \tfrac{1}{2}\int_0^t \alpha(s, A\phi)\,ds = 0 \tag{8.36}$$

for all $0 < t < \infty$. Next, apply this to a countable set of functions $\phi_j \in C_c^\infty(\mathbf{R}^d)$ such that for an arbitrary $\phi \in C_c^\infty(\mathbf{R}^d)$, ϕ and $A\phi$ are uniform limits of ϕ_{j_k} and $A\phi_{j_k}$ for a subsequence $\{\phi_{j_k}\}$ supported on a common compact set. (Exercise 8.11 shows that a countable set satisfying this property exists.) By taking these limits in (8.36), we obtain (8.36) Q-almost surely, simultaneously for all $\phi \in C_c^\infty(\mathbf{R}^d)$. By hypothesis (8.5), $\alpha(0,dx) = \rho_0(x)dx$ for Q-almost every path α. We conclude that Q-almost every path is continuous and a weak solution of the initial value problem (8.10).

At this point we apply the uniqueness theorem for the initial value problem (8.10). The same reasoning that gave (8.28) also implies that the closed set

$$L_k = \{\alpha \in D_{\mathbf{M}} : \alpha(t, x + (-k,k)^d) \le (2k-1)^d \text{ for all } x \in \mathbf{R}^d \text{ and all } t \ge 0\}$$

satisfies $Q^n(L_k)$=1 for all n and k. By weak convergence $Q(L_k) = 1$. (See Exercise 8.12 for the closedness of L_k.) Thus the boundedness assumption of Theorem A.28 is satisfied Q-almost surely, namely that

$$\alpha(t, B(x,r)) \leq Cr^d$$

for all balls of sufficiently large radius r. By Theorem A.28, Q is supported by the unique path $\bar{\alpha}(t,dx) = \rho(x,t)dx$, where $\rho(x,t)$ is defined by (8.7)–(8.8).

Since Q was an arbitrary weak limit point of the tight sequence $\{Q^n\}$, this sequence actually converges weakly to the degenerate distribution $Q = \delta_{\bar{\alpha}}$. Weak convergence to a degenerate distribution implies convergence in probability, so we have the convergence $\bar{\pi}^n_\cdot \to \bar{\alpha}$ in probability, in the Skorokhod topology of the space $D_{\mathbf{M}}$. Theorem 8.2 is proved.

Finally, we derive the statement (8.11) which we restate in the following form. Given $\phi \in C_c^\infty(\mathbf{R}^d)$, define the Borel set

$$B = \{\alpha \in D_{\mathbf{M}} : |\alpha(t,\phi) - \bar{\alpha}(t,\phi)| \geq \varepsilon\}.$$

The requirement is then $\lim_{n\to\infty} Q^n(B) = 0$. If we can show that $Q(\partial B) = 0$ where ∂B is the topological boundary of B, then $\lim_{n\to\infty} Q^n(B) = Q(B)$ and the conclusion follows from $Q(B) = 0$. (Here we use again a basic property of weak convergence of probability measures.) It suffices to show that $\bar{\alpha}$ cannot be a limit of a sequence $\alpha_j \notin B$. For then $\bar{\alpha}$ cannot be a boundary point of B, and consequently $Q(\partial B) = \delta_{\bar{\alpha}}(\partial B) = 0$. Two facts finish this. Convergence $\alpha_j \to \bar{\alpha}$ in the Skorokhod topology implies that $\alpha_j(t) \to \bar{\alpha}(t)$ at each continuity point t of $\bar{\alpha}$ (by Lemma A.2). We already know $\bar{\alpha}$ is a continuous path. Thus any sequence α_j convergent to $\bar{\alpha}$ has $\alpha_j(t,\phi) \to \bar{\alpha}(t,\phi)$, and consequently such a sequence cannot lie outside B.

We have proved (8.11) and thereby completed the proof of Theorem 8.1.

8.2 The gradient condition

A key stage in the hydrodynamic limit of the symmetric exclusion process was the appearance of the difference operator

$$A_n\phi\left(\tfrac{u}{n}\right) \equiv \sum_z p(0,z)\left\{\phi\left(\tfrac{u+z}{n}\right) + \phi\left(\tfrac{u-z}{n}\right) - 2\phi\left(\tfrac{u}{n}\right)\right\}$$

on line (8.19) in the calculation for Lf, where f was the empirical average

$$f(\eta) = n^{-d} \sum_{u\in\mathbf{Z}^d} \eta(u)\phi\left(\tfrac{u}{n}\right). \tag{8.37}$$

A_n is the lattice version of the differential operator A defined by (8.9). $n^2 A_n\phi$ converges to $A\phi$ as $n \to \infty$. That $A_n\phi$ is of order n^{-2} is crucially important for the calculation because it can absorb the factor n^2 produced by the time scaling. In the previous section this happened in the passage from (8.21) to (8.22). In (8.21) there is still an extra factor of n^{-2} in front of $\pi^n_s(A\phi)$, but in (8.22) this factor has disappeared as a result of scaling time by n^2.

Let us abstract the situation to find the property of the exclusion process that made this happen, so that we might generalize the approach to other processes. Let us consider a particle system with state space $X = \{0, 1, \dots, K\}^{\mathbf{Z}^d}$ for some finite K, or $X = \mathbf{Z}_+^{\mathbf{Z}^d}$. Let $c(u, v, \eta)$ be the rate of moving one particle from site u to site v when the present configuration is η. Assume that construction issues of the particle system can be resolved. The generator on $C_b(X)$ is

$$Lf(\eta) = \sum_{u,v} c(u, v, \eta)[f(\eta^{u,v}) - f(\eta)]$$

where

$$\eta^{u,v}(w) = \begin{cases} \eta(u) - 1, & w = u \\ \eta(v) + 1, & w = v \\ \eta(w), & w \notin \{u, v\} \end{cases}$$

is the configuration that results from moving one particle from site u to v. Of course, for this to make sense, it must be that $c(u, v, \eta) = 0$ if a jump from u to v is impossible.

Fix a test function $\phi \in C_c^\infty(\mathbf{R}^d)$ and define $f(\eta)$ as in (8.37) above. Let us attempt to repeat the calculation that led to (8.19) for the symmetric exclusion. For the third equality below replace the summation index z by $-z$ in the second sum. For the fourth equality replace u by $u + z$ in the second sum, for each fixed z.

$$\begin{aligned} Lf(\eta) &= \sum_{u,z} c(u, u+z, \eta)\big\{f(\eta^{u,u+z}) - f(\eta)\big\} \\ &= n^{-d} \sum_{u,z} c(u, u+z, \eta)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} \\ &= \frac{n^{-d}}{2} \sum_{u,z} c(u, u+z, \eta)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} - \frac{n^{-d}}{2} \sum_{u,z} c(u, u-z, \eta)\big\{\phi\big(\tfrac{u}{n}\big) - \phi\big(\tfrac{u-z}{n}\big)\big\} \\ &= \frac{n^{-d}}{2} \sum_{u,z} c(u, u+z, \eta)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} - \frac{n^{-d}}{2} \sum_{u,z} c(u+z, u, \eta)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} \\ &= \frac{n^{-d}}{2} \sum_{u,z} \big(c(u, u+z, \eta) - c(u+z, u, \eta)\big)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\}. \end{aligned} \tag{8.38}$$

At this point we have only one lattice derivative $\phi\big(\frac{u+z}{n}\big) - \phi\big(\frac{u}{n}\big)$ which is of order n^{-1}. We need to take differences for a second time to get another factor of n^{-1}. The factor

$$j(u, u+z, \eta) = c(u, u+z, \eta) - c(u+z, u, \eta)$$

arranged into the sum is the net flux of particles from site u to site $u+z$, namely, the flux from u to $u+z$ minus the flux from $u+z$ to u. We make the following assumption: there is a finitely supported function $r : \mathbf{Z}^d \to \mathbf{R}$ and a bounded cylinder function $h : X \to \mathbf{R}$ such that

$$j(u, u+z, \eta) = r(z)\big(h(\theta_u \eta) - h(\theta_{u+z}\eta)\big) \tag{8.39}$$

for all $u, z \in \mathbf{Z}^d$ and $\eta \in X$. This condition is called the *gradient condition* since it requires that the microscopic flux be a gradient of some other function. The spatial translations θ_u on the space X are defined by $\theta_u \eta(v) = \eta(u+v)$. Let us check (8.39) for symmetric exclusion. By $p(0,z) = p(0,-z)$,

$$\begin{aligned} j(u, u+z, \eta) &= p(0,z)\eta(u)(1-\eta(u+z)) - p(0,-z)\eta(u+z)(1-\eta(u)) \\ &= p(0,z)\big(\eta(u) - \eta(u+z)\big), \end{aligned} \tag{8.40}$$

clearly of the gradient form.

Continue the calculation from line (8.38) and utilize assumption (8.39).

$$\begin{aligned} Lf(\eta) &= \frac{n^{-d}}{2}\sum_{u,z}\big(c(u,u+z,\eta) - c(u+z,u,\eta)\big)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} \\ &= \frac{n^{-d}}{2}\sum_{u,z} r(z)\big(h(\theta_u\eta) - h(\theta_{u+z}\eta)\big)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} \\ &= \frac{n^{-d}}{2}\sum_{u,z} r(z)h(\theta_u\eta)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} + \frac{n^{-d}}{2}\sum_{u,z} r(z)h(\theta_{u+z}\eta)\big\{\phi\big(\tfrac{u}{n}\big) - \phi\big(\tfrac{u+z}{n}\big)\big\} \\ &= \frac{n^{-d}}{2}\sum_{u,z} r(z)h(\theta_u\eta)\big\{\phi\big(\tfrac{u+z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} + \frac{n^{-d}}{2}\sum_{u,z} r(z)h(\theta_u\eta)\big\{\phi\big(\tfrac{u-z}{n}\big) - \phi\big(\tfrac{u}{n}\big)\big\} \\ &= \frac{n^{-d}}{2}\sum_{u} h(\theta_u\eta)\sum_{z} r(z)\big\{\phi\big(\tfrac{u+z}{n}\big) + \phi\big(\tfrac{u-z}{n}\big) - 2\phi\big(\tfrac{u}{n}\big)\big\}. \end{aligned} \tag{8.41}$$

On the last line we have the second order lattice difference operator

$$B_n\phi\big(\tfrac{u}{n}\big) = \sum_{z} r(z)\big\{\phi\big(\tfrac{u+z}{n}\big) + \phi\big(\tfrac{u-z}{n}\big) - 2\phi\big(\tfrac{u}{n}\big)\big\}$$

which is of order n^{-2}, as required. So we see that the gradient condition guarantees the two differences in Lf.

A much-studied example of gradient models is the class of zero-range processes. They have state space $X = \mathbf{Z}_+^{\mathbf{Z}^d}$, and $c(u,v,\eta) = p(u,v)g(\eta(u))$ for a given function g that identifies the process in question. The gradient condition is trivially satisfied for a symmetric jump probability $p(u,v)$. The name *zero-range* stems from the property that the rate of a jump from u to v depends only on the state at u, so the range of interaction is zero. The K-exclusion processes introduced in Chapter 7 do not satisfy the gradient condition (Exercise 8.13).

Comparison of (8.19) and (8.41), and also (8.40), reveal the miracle that made the proof of the hydrodynamic limit for symmetric exclusion unnaturally easy. Namely, the function h in the gradient condition is simply $h(\eta) = \eta(0)$. The consequence of this is that $n^2 Lf(\eta_s)$ can be expressed in terms of the empirical measure π_s^n as in (8.21). That is why (8.25) is a closed equation for $\bar{\pi}_t^n$. (Closed in the sense that no other unknowns appear.)

In the general situation the function h in (8.41) does not disappear, and needs to be dealt with. We shall not pursue this direction, but here is the idea in a nutshell.

Introduce an intermediate scale εn with small positive ε. For $b > 0$ real, let $\Lambda(b) = [-b,b]^d \cap \mathbf{Z}^d$ denote the centered cube with $(2[b]+1)^d$ sites. Use the smoothness of ϕ to write

$$\begin{aligned} Lf(\eta_{n^2 s}) &= \frac{n^{-d}}{2} \sum_u h(\theta_u \eta_{n^2 t}) B_n \phi\big(\tfrac{u}{n}\big) \\ &= \frac{n^{-d}}{2} \sum_u \Big\{ |\Lambda(\varepsilon n)|^{-1} \sum_{v \in u + \Lambda(\varepsilon n)} h(\theta_v \eta_{n^2 t}) \Big\} B_n \phi\big(\tfrac{u}{n}\big) \; + \; [\text{an error of order } \varepsilon]. \end{aligned}$$

As $n \to \infty$, one seeks to show that the average

$$|\Lambda(\varepsilon n)|^{-1} \sum_{v \in u + \Lambda(\varepsilon n)} h(\theta_v \eta_{n^2 t})$$

is approximately $\int h \, d\nu_\rho$, the average of h under an equilibrium measure ν_ρ at density

$$\rho = |\Lambda(\varepsilon n)|^{-1} \sum_{v \in u + \Lambda(\varepsilon n)} \eta_{n^2 t}(v),$$

the local empirical density. Informally speaking, the system behaves as if it were in a *local equilibrium* in the εn-cube around site u. This expresses $Lf(\eta_{n^2 s})$ approximately as a function of the empirical measure, and thereby closes the equation.

Notes

The gradient condition (8.39) can be relaxed by permitting a more complicated linear combination of translates of different functions on the right-hand side. See page 61 in [24]. Hydrodynamic limits for gradient systems appear in the lectures of De Masi and Presutti [7], the monograph of Spohn [44], in Varadhan's lectures [46] and in the monograph of Kipnis and Landim [24]. Hydrodynamic limits for nongradient systems are discussed in [46] and [24].

Exercise 8.2 *Hydrodynamic limit for independent Brownian Motions.* This is an example where no scaling of space or time is needed. For each n, let $\mathbf{x}^n(t) = \big(x_1^n(t), \dots, x_{L_n}^n(t)\big)$ be a vector of independent Brownian motions in $\mathbf{R}^d$. Assume that $L_n \le Cn^d$ for some fixed constant C. Assume that there exists a function $\rho_0 \in L^1(\mathbf{R}^d)$ such that the initial locations of the Brownian motions satisfy

$$\lim_{n\to\infty} E\Big[n^{-d}\sum_i \phi\big(x_i^n(0)\big)\Big] = \int_{\mathbf{R}^d} \phi(x)\rho_0(x)\,dx$$

for all test functions $\phi \in C_c(\mathbf{R}^d)$. Show that for all $t > 0$ and ϕ,

$$\lim_{n\to\infty} n^{-d}\sum_i \phi\big(x_i^n(t)\big) = \int_{\mathbf{R}^d} \phi(x)\rho(x,t)\,dx \quad \text{in } L^2,$$

where ρ is solves the heat equation

$$\rho_t = \tfrac{1}{2}\Delta\rho\,, \quad \rho(x,0) = \rho_0.$$

Hint: Investigate the mean and variance of $n^{-d}\sum\phi\big(x_i^n(t)\big)$.

For a slightly more interesting exercise, consider the case of infinitely many Brownian motions $\{x_i^n(t) : i \in \mathbf{N}\}$ in $\mathbf{R}^d$. Use a hypothesis that prevents the initial particles from accumulating anywhere in space, for example assume there exists a fixed cube B in $\mathbf{R}^d$ and a constant C such that for all $z \in \mathbf{R}^d$,

$$E\Big[n^{-d}\sum_{i\in\mathbf{N}} \mathbf{1}_{z+B}\big(x_i^n(0)\big)\Big] \le C.$$

A third variant of the exercise is this: to avoid problems of unbounded space, hydrodynamic limits are often proved on the torus $\mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$ which can be taken as $[0,1)^d$ with periodic boundary conditions. You can also repeat the exercise above for Brownian motions on the torus.

Exercise 8.3 This exercise features the space scaling without any dynamics. Let $0 \le u(x) \le 1$ be a continuous function on $[0,1]$. Let

$$\{X_{n,i} : 1 \le i \le n, 1 \le n < \infty\}$$

be a triangular array of Bernoulli variables such that, for each fixed n, $X_{n,1}, \dots, X_{n,n}$ are independent with distributions

$$P(X_{n,i} = 1) = u\big(\tfrac{i}{n}\big) \quad \text{and} \quad P(X_{n,i} = 0) = 1 - u\big(\tfrac{i}{n}\big).$$

Prove that, for any continuous function ϕ on $[0,1]$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_{n,i}\, \phi\big(\tfrac{i}{n}\big) = \int_0^1 u(x)\phi(x)\, dx \quad \text{almost surely.}$$

This exercise illustrates the sense in which the function u is the macroscopic profile of the random variables $X_{n,i}$. The proof requires the standard Strong Law of Large Numbers and some simple estimation. Some regularity on ϕ is necessary because the conclusion clearly fails for the indicator function of rationals.

Exercise 8.4 *Hydrodynamic limit for mean zero independent random walks.* This example has both the space and time scaling, but no interaction. For each n, let $\mathbf{x}^n(t) = \big(x_1^n(t), \dots, x_{L_n}^n(t)\big)$ be a vector of independent random walks on $\mathbf{Z}^d$. Each random walk jumps at rate 1, and the common translation-invariant jump kernel $p(u,v) = p(0, v-u)$ has zero mean

$$\sum_{v\in\mathbf{Z}^d} v p(0,v) = 0$$

and a finite covariance matrix $\Gamma = (\gamma_{i,j})_{1\le i,j\le d}$ defined by

$$\gamma_{i,j} = \sum_{v\in\mathbf{Z}^d} v_i v_j p(0,v).$$

Assume that $L_n \le Cn^d$ for some fixed constant C. Assume that there exists a function $\rho_0 \in L^1(\mathbf{R}^d)$ such that the initial locations of the random walks satisfy

$$\lim_{n\to\infty} E\Big[n^{-d} \sum_i \phi\big(n^{-1}x_i^n(0)\big)\Big] = \int_{\mathbf{R}^d} \phi(x)\rho_0(x)\, dx \tag{8.42}$$

for all test functions $\phi \in C_c(\mathbf{R}^d)$. Show that for all $t > 0$ and ϕ,

$$\lim_{n\to\infty} n^{-d} \sum_i \phi\big(n^{-1}x_i^n(n^2 t)\big) = \int_{\mathbf{R}^d} \phi(x)\rho(x,t)\, dx \quad \text{in } L^2,$$

where ρ is defined by $\rho(x,t) = E\rho_0(x - t^{1/2}Z)$ for an $\mathcal{N}(0,\Gamma)$-distributed Gaussian random vector Z. The function ρ is also the solution of a differential equation, see Theorem A.27 in the appendix.

Outline: By the independence of the random walks the variance of

$$U_n = n^{-d} \sum \phi\big(n^{-1} x_i^n(n^2 t)\big)$$

vanishes as $n \to \infty$. With E_0 denoting expectation over the initial locations $\{x_i^n(0)\}$, let

$$g_n(\xi) = E_0\Big[n^{-d} \sum_i \phi\big(n^{-1} x_i^n(0) + \xi\big)\Big].$$

This defines a uniformly bounded and equicontinuous sequence of functions $\{g_n\}$ on $\mathbf{R}^d$. Since initial locations are independent of jumps for a random walk, the mean of U_n can be written as

$$EU_n = E\big[g_n\big(n^{-1}s(n^2t)\big)\big]$$

where $s(\cdot)$ is a random walk starting at the origin. By hypothesis

$$g_n(\xi) \to g(\xi) \equiv \int \phi(x+\xi)\rho_0(x)\, dx.$$

By the multivariate central limit theorem and some properties of weak convergence,

$$E\big[g_n\big(n^{-1}s(n^2t)\big)\big] \to Eg(t^{1/2}Z).$$

The proof shows that the hydrodynamic limit comes from two ingredients: having a large number of particles, each experiencing central limit behavior.

Exercise 8.5 *Hydrodynamic limit for independent random walks with drift.* As a prelude to the asymmetric systems studied in the next chapter, assume that the independent random walks $\mathbf{x}^n(t) = \big(x_1^n(t), \dots, x_{L_n}^n(t)\big)$ have a drift. Precisely, assume that the common translation-invariant jump kernel $p(u,v) = p(0, v-u)$ has nonzero mean

$$\sum_{v \in \mathbf{Z}^d} v p(0,v) = b \neq 0.$$

Assume (8.42) again, and derive a scaling limit for the empirical density. Observe that the correct time scaling now comes from the law of large numbers instead of the central limit theorem. The details in this exercise are easier than in Exercise 8.4 because weak convergence is not needed. What is the partial differential equation satisfied by the limiting density $\rho(x,t)$?

Exercise 8.6 Let η_t^n be a sequence of symmetric exclusion processes defined on a common probability space. Consider a time scale $n^\alpha t$ with $0 \le \alpha < 2$. Without assuming anything further about the processes, show that for any $t > 0$ and $\phi \in C_c^\infty(\mathbf{R}^d)$,

$$\lim_{n\to\infty} \Big\{ n^{-d} \sum_{u \in \mathbf{Z}^d} \eta_{n^\alpha t}^n(u) \phi\big(\tfrac{u}{n}\big) - n^{-d} \sum_{u \in \mathbf{Z}^d} \eta_0^n(u) \phi\big(\tfrac{u}{n}\big) \Big\} = 0 \quad \text{almost surely.}$$

Exercise 8.7 Show that assumption (8.6) is equivalent to requiring that the covariance matrix Σ have strictly positive eigenvalues. *Hint:* Σ is a real symmetric matrix, so it has a basis of orthonormal eigenvectors. Note that

$$\sum_u p(0,u)(u \cdot x)^2 = x \cdot \Sigma x.$$

Exercise 8.8 Show that assumption (8.6) is equivalent to the usual notion of *ellipticity* of the differential operator A defined by (8.9), namely that for some constant $\theta > 0$,

$$\sum_{1 \le i,j \le d} \sigma_{i,j} x_i x_j \ge \theta |x|_2^2 \quad \text{for all } x \in \mathbf{R}^d. \tag{8.43}$$

Exercise* 8.9 For paths $\alpha \in D_{\mathbf{M}}$, let

$$G(\alpha) = \sup_{t \ge 0} e^{-t} d_{\mathbf{M}}(\alpha(t), \alpha(t-)).$$

Show that G is a continuous function on $D_{\mathbf{M}}$. *Hint:* By (A.2), $G(\alpha) > 0$ implies that the supremum in the definition of $G(\alpha)$ is attained at some t. Apply Lemma A.2.

Exercise* 8.10 *Change of variable.* Suppose f is a strictly increasing continuous function from $[a,b]$ onto $[\alpha,\beta]$, with inverse g. Assume g is absolutely continuous. Show that for any bounded Borel function H,

$$\int_a^b H(f(t))\,dt = \int_\alpha^\beta H(s) g'(s)\,ds.$$

Hint: First let H be the indicator function of an interval. Use the π–λ Theorem A.1 to extend to all indicator functions of Borel sets.

Exercise* 8.11 Show that there exists a countable set $\{\psi_j\} \subseteq C_c^\infty(\mathbf{R}^d)$ with this property: given any $\phi \in C_c^\infty(\mathbf{R}^d)$, there exists a compact set K and a subsequence $\{\psi_{j_k}\}$ such that K supports ϕ and all ψ_{j_k}, and the limits $\psi_{j_k} \to \phi$ and $(\psi_{j_k})_{x_i,x_j} \to \phi_{x_i,x_j}$ hold uniformly for all $1 \le i,j \le d$.

Suggestion. Let $\mathcal{H}_0 = \{D^2\phi : \phi \in C_c^\infty(\mathbf{R}^d)\}$ be the space of $d \times d$ matrix-valued functions coming from the Hessian matrices

$$D^2\phi = \begin{bmatrix} \phi_{x_1,x_1} & \cdots & \phi_{x_1,x_d} \\ \vdots & \ddots & \vdots \\ \phi_{x_d,x_1} & \cdots & \phi_{x_d,x_d} \end{bmatrix}$$

of $C_c^\infty(\mathbf{R}^d)$ functions. This is a subspace of the space $\mathcal{H}$ of all compactly supported, $d \times d$ matrix-valued C^∞ functions on $\mathbf{R}^d$. Let $K_N = [-N, N]^d$. First find a countable set $\{G_k\} \subseteq \mathcal{H}$ with the property that for any $G \in \mathcal{H}$ supported on K_N, there is a K_{N+1}-supported subsequence $\{G_{k_m}\}$ that converges to G uniformly (in all matrix entries). Next, for each G_k supported on K_M and $n \in \mathbf{N}$, find a function $\phi_{k,n,M+1}$ supported on K_{M+1} such that $\|D^2\phi_{k,n,M+1} - G_k\|_\infty \leq 1/n$, if such a function exists. The claim is that the collection $\{\phi_{k,n,M+1}\}$ will do the job.

Exercise* 8.12 Let K be a fixed open rectangle in $\mathbf{R}^d$ and C a constant. Show that the set

$$L = \{\alpha \in D_{\mathbf{M}} : \alpha(t, x + K) \leq C \text{ for all } x \in \mathbf{R}^d \text{ and all } t \geq 0\}$$

is closed in Skorokhod topology. *Hint:* Imitate parts of the proof of Lemma 8.7. If $\alpha_j \in L$ and $\alpha_j(\lambda_j(t)) \to \alpha(t)$, consider test functions $f \in C_c(\mathbf{R}^d)$ such that $0 \leq f \leq \mathbf{1}_{x+K}$. The indicator $\mathbf{1}_{x+K}$ is an increasing limit of such functions.

Exercise 8.13 Show that a K-exclusion process with generator (7.1) cannot satisfy the gradient condition (8.39).

9 Variational approach for totally asymmetric systems

In this chapter we address the hydrodynamic limit of a totally asymmetric, nearest-neighbor K-exclusion in one dimension. The approach will utilize a special pathwise variational property of the process. To formulate this property we switch from the particle occupation variables $\eta = (\eta(u) : u \in \mathbf{Z})$ to a representation in terms of a height function $h : \mathbf{Z} \to \mathbf{Z}$. The value $h(u)$ represents the height of an interface over the site u. The occupation variables are the increments of the height function, namely

$$\eta(u) = h(u) - h(u-1).$$

The K-exclusion process η_t and the height process h_t are coupled through Poisson clocks. Whenever a particle jumps from site u to $u+1$, height variable $h(u)$ decreases by one.

The height process possesses the variational property, or envelope property, that we utilize for the proof of the hydrodynamic limit. This property generalizes to a multidimensional totally asymmetric height process $h_t : \mathbf{Z}^d \to \mathbf{Z}$. However, the increment process of a multidimensional height process is not a K-exclusion process in multiple dimensions (see Exercise 9.2). The basic theorems in this chapter are proved for the multidimensional height process. The hydrodynamic limit for the one-dimensional, totally asymmetric, nearest-neighbor K-exclusion process is derived as a corollary.

9.1 An interface model with an envelope property

We consider a model of a randomly evolving interface in $d+1$ space dimensions. The interface is described by an integer-valued *height function* $h : \mathbf{Z}^d \to \mathbf{Z}$ defined on the d-dimensional integer lattice $\mathbf{Z}^d$. The height function moves downward through random jumps.

Let $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $e_d = (0, \ldots, 0, 1)$ be the d standard basis vectors in $\mathbf{Z}^d$. Fix d positive integers $K_1, K_2, \ldots, K_d$. These determine the maximal increments in each coordinate direction of an admissible height function. The state space H of the system is

$$H = \{h : \mathbf{Z}^d \to \mathbf{Z} : \; 0 \le h(u+e_i) - h(u) \le K_i \text{ for } u \in \mathbf{Z}^d \text{ and } 1 \le i \le d\}. \tag{9.1}$$

The dynamics of the interface is determined by a collection $\{\mathcal{T}_u : u \in \mathbf{Z}^d\}$ of mutually independent homogeneous Poisson point processes of rate 1 on the time line $[0, \infty)$. The jump rule is that if $t \in \mathcal{T}_u$, then

$$h_t(u) = h_{t-}(u) - 1$$

provided after the jump the inequalities

$$h_t(u) \ge h_t(u - e_i) \text{ and } h_t(u) \ge h_t(u+e_i) - K_i \text{ for } 1 \le i \le d$$

are true. In other words, whenever Poisson clock $\mathcal{T}_u$ rings, height variable $h(u)$ decreases by one provided this change does not take the state of the system outside the state space H.

The process $h_t = (h_t(u) : u \in \mathbf{Z}^d)$ can be constructed on the probability space $(\Omega, \mathcal{H}, \mathbf{P})$ of the Poisson clocks $\{\mathcal{T}_u : u \in \mathbf{Z}^d\}$ with the percolation argument of Section 2.1, for any initial height profile $h_0 \in H$. Consider only realizations $\{\mathcal{T}_u\}$ of the Poisson clocks such that

$$\begin{aligned}&\text{each } \mathcal{T}_u \text{ has only finitely many jump times in every bounded interval } (0,T],\\ &\text{and no two distinct processes } \mathcal{T}_u \text{ and } \mathcal{T}_v \text{ have a jump time in common.}\end{aligned} \tag{9.2}$$

Define the random graph $\mathcal{G}_{0,t}$ with vertex set $\mathbf{Z}^d$ by declaring that the nearest-neighbor edge $\{u, v\}$ is present iff either $\mathcal{T}_u$ or $\mathcal{T}_v$ has a jump time in $(0, t\,]$. Sites u and v in $\mathbf{Z}^d$ are nearest-neighbor if $u - v = \pm e_i$ for some $1 \le i \le d$. Adapt the proof of Lemma 2.1 to show that if t_0 is fixed small enough but positive, $\mathcal{G}_{0,t_0}$ has finite connected components almost surely. Now the evolution h_t can be defined on each connected component of $\mathcal{G}_{0,t_0}$ for times $0 \le t \le t_0$, and then for all time by iterating this argument.

To have a random initial height h_0 with distribution μ on the state space H, construct the process h_t as a function of the initial height h_0 and the Poisson processes ω on the product space $(H \times \Omega, \mathcal{B}(H) \otimes \mathcal{H}, \mu \otimes \mathbf{P})$. So the clocks are taken independent of the initial state. The distribution of the process is a probability measure P^μ on the path space D_H.

The basic coupling works as before. To couple a family of processes $\{\sigma_t^k\}$ through the Poisson clocks, we need a probability space $(\Sigma, \mathcal{A}, Q)$ on which are defined the initial height functions $\{\sigma_0^k\}$. Each process σ_t^k is then defined on the product space $(\Sigma \times \Omega, \mathcal{A} \otimes \mathcal{H}, Q \otimes \mathbf{P})$ as a function of its initial profile σ_0^k and the Poisson clocks ω.

Notationally, we use h_t to denote a general height process. Height processes with special properties such as particular initial configurations are denoted by other symbols such as σ_t and ξ_t. Despite differences in notation, processes coupled together have the same state space defined by fixed parameters $K_1, \ldots, K_d$.

Two key properties of the height process come in the next lemmas. A coordinatewise ordering is defined among height profiles by

$$h \le \tilde{h} \quad \text{iff} \quad h(u) \le \tilde{h}(u) \quad \text{for all } u \in \mathbf{Z}^d. \tag{9.3}$$

Lemma 9.1 *(Attractivity.) Suppose the processes h_t and $\tilde{h}_t$ are coupled so that they read the same Poisson clocks $\{\mathcal{T}_u\}$. Assume that initially at time zero, $h_0 \le \tilde{h}_0$. Then for almost every realization of the Poisson clocks, $h_t \le \tilde{h}_t$ for all $t \ge 0$.*

We leave the proof of this first lemma to the reader.

Lemma 9.2 *(Envelope property.) Suppose the process h_t and a countable family $\{\sigma_t^k : k \in \mathcal{K}\}$ of height processes are coupled so that they all read the same Poisson clocks $\{\mathcal{T}_u\}$. Assume* *that initially at time zero,*

$$h_0(u) = \sup_{k \in \mathcal{K}} \sigma_0^k(u) \qquad \text{for all } u \in \mathbf{Z}^d. \tag{9.4}$$

Then for almost every realization of the Poisson clocks,

$$h_t(u) = \sup_{k \in \mathcal{K}} \sigma_t^k(u) \qquad \text{for all } u \in \mathbf{Z}^d \text{ and } t \geq 0. \tag{9.5}$$

Proof. Consider realizations $\{\mathcal{T}_u\}$ of the Poisson clocks that satisfy (9.2) and for which the random graphs $\mathcal{G}_{kt_0,(k+1)t_0}$ have finite connected components for all integers $k \geq 0$. Then the processes under consideration can be constructed for all time. It suffices to show that (9.5) holds up to time t_0. Then the argument can be repeated for the restarted processes $\tilde{h}_t = h_{t_0+t}$ and $\tilde{\sigma}_t^k = \sigma_{t_0+t}^k$ that again satisfy the initial assumption (9.4).

By attractivity we already have $h_t(u) \geq \sigma_t^k(u)$ for all $k \in \mathcal{K}$. It remains to show that for each (u,t) there is some k such that $h_t(u) = \sigma_t^k(u)$.

Fix $u_0 \in \mathbf{Z}^d$. Let C be the connected component that contains u_0 in the graph $\mathcal{G}_{0,t_0}$. During time interval $(0,t_0]$ the finitely many Poisson processes $\{\mathcal{T}_u : u \in C\}$ have only finitely many jump times. We prove that immediately after each of these jump times,

$$\text{for all } v \in C \text{ there exists } k \text{ such that } h_t(v) = \sigma_t^k(v). \tag{9.6}$$

The proof proceeds by induction on the jump times.

Suppose $\tau \in \mathcal{T}_u \cap (0,t_0]$ for some $u \in C$, and assume that (9.6) holds for $t < \tau$. Jump time $\tau \in \mathcal{T}_u$ only affects the height values at u, so (9.6) continues to hold at $t = \tau$ for $v \in C \smallsetminus \{u\}$.

Case 1: $h(u)$ jumps at time τ, so $h_\tau(u) = h_{\tau-}(u) - 1$. By the induction assumption $h_{\tau-}(u) = \sigma_{\tau-}^k(u)$ for some k. By attractivity $h_\tau(u) \geq \sigma_\tau^k(u)$ must hold after the jump. So $\sigma^k(u)$ must have jumped too, and we have

$$h_\tau(u) = h_{\tau-}(u) - 1 = \sigma_{\tau-}^k(u) - 1 = \sigma_\tau^k(u).$$

Case 2: $h(u)$ could not jump at time τ, because for some $1 \leq j \leq d$,

$$h_{\tau-}(u) = h_{\tau-}(u - e_j).$$

The edge $\{u, u - e_j\}$ is present in the graph $\mathcal{G}_{0,t_0}$ by virtue of the jump time $\tau \in \mathcal{T}_u \cap (0,t_0]$, so site $u - e_j$ lies in C. By induction,

$$h_{\tau-}(u - e_j) = \sigma_{\tau-}^\ell(u - e_j)$$

for some $\ell \in \mathcal{K}$. By definition (9.1) of the state space,

$$\sigma^{\ell}_{\tau-}(u - e_j) \le \sigma^{\ell}_{\tau-}(u).$$

By attractivity

$$\sigma^{\ell}_{\tau-}(u) \le h_{\tau-}(u).$$

All these together give

$$h_{\tau-}(u) = h_{\tau-}(u - e_j) = \sigma^{\ell}_{\tau-}(u - e_j) \le \sigma^{\ell}_{\tau-}(u) \le h_{\tau-}(u).$$

So they are all equal. Height $\sigma^{\ell}(u)$ could not have jumped at time τ because it was blocked by $\sigma^{\ell}_{\tau-}(u - e_j) = \sigma^{\ell}_{\tau-}(u)$. We have

$$h_{\tau}(u) = h_{\tau-}(u) = \sigma^{\ell}_{\tau-}(u) = \sigma^{\ell}_{\tau}(u)$$

and so (9.6) continues to hold after jump time τ.

Case 3: $h(u)$ could not jump at time τ, because for some $1 \le j \le d$,

$$h_{\tau-}(u) = h_{\tau-}(u + e_j) - K_j.$$

This case follows the same reasoning as the previous one. We leave the details to the reader.

We now know (9.6) holds for times $0 \le t \le t_0$. We can repeat this for all $u_0 \in \mathbf{Z}^d$, and thereby have verified (9.5) for $0 \le t \le t_0$. This completes the proof of the lemma. ∎

To take advantage of this property we need a suitable family $\{\sigma^k\}$. Define an element $w \in H$ by

$$w(u) = \sum_{i=1}^{d} K_i(u_i \wedge 0) \qquad \text{for } u = (u_1, \ldots, u_d) \in \mathbf{Z}^d. \tag{9.7}$$

This wedge profile is the minimal element of H that satisfies $w(0) = 0$, in the sense that $w \le h$ for any $h \in H$ such that $h(0) = 0$.

Let h_t be a height process with deterministic or random initial profile h_0. Define a family $\{\sigma^v_0 : v \in \mathbf{Z}^d\}$ of initial configurations, indexed by sites v, by

$$\sigma^v_0(u) = h_0(v) + w(u - v)\,, \quad u \in \mathbf{Z}^d.$$

Initial profile σ^v_0 is the minimal profile that goes through the point $(v, h_0(v))$ in $\mathbf{Z}^d \times \mathbf{Z}$. It is clear that the property

$$h_0(u) = \sup_{v \in \mathbf{Z}^d} \sigma^v_0(u)$$

holds at time 0. Couple the processes h_t and $\{\sigma_t^v : v \in \mathbf{Z}^d\}$ through the Poisson clocks, so that for each $u \in \mathbf{Z}^d$, $h_t(u)$ and $\sigma_t^v(u)$ obey the clock $\mathcal{T}_u$. By Lemma 9.2 the envelope property is preserved by the coupling, so

$$h_t(u) = \sup_{v \in \mathbf{Z}^d} \sigma_t^v(u) \qquad \text{for all time } t. \tag{9.8}$$

The processes σ_t^v are functions of both the initial profile h_0 and the Poisson clocks. We normalize them to separate the effect of h_0 and the Poisson processes. For each $v \in \mathbf{Z}^d$, define process ξ_t^v by stipulating that initially

$$\xi_0^v(u) = w(u) \quad \text{for} \quad u \in \mathbf{Z}^d, \tag{9.9}$$

and dynamically

$$\text{height variable } \xi^v(u) \text{ attempts to jump at the jump times of } \mathcal{T}_{u+v}. \tag{9.10}$$

These processes are related to the earlier ones via

$$\sigma_t^v(u) = h_0(v) + \xi_t^v(u - v). \tag{9.11}$$

To prove (9.11), note that it is true at $t = 0$ by construction, and then check that the left and right-hand sides of the equation always jump together.

The envelope property (9.8) takes the form

$$h_t(u) = \sup_{v \in \mathbf{Z}^d} \{h_0(v) + \xi_t^v(u - v)\}. \tag{9.12}$$

In (9.12) the family of processes $\{\xi_t^v\}$ is identically distributed. The superscript v translates the index of the Poisson clocks: process ξ_t^v reads the translated Poisson processes $\{\mathcal{T}_{u+v} : u \in \mathbf{Z}^d\}$.

Convention (9.10) is at odds with our original jump rule since variable $\xi^v(u)$ obeys Poisson clock $\mathcal{T}_{u+v}$ instead of $\mathcal{T}_u$. The choice made here matches (9.12) formally with the Hopf-Lax formula for Hamilton-Jacobi equations [see (9.17) below and Section A.12]. Another application of processes that transform the indices of the Poisson clocks occurs in the proof of Proposition 9.20 later in this chapter.

9.2 Hydrodynamic limit for the height process

Now we describe the evolution of the height process on large space and time scales. The first step is a limit for the process started from the wedge profile. Let $\xi_t = \xi_t^0$ be the process defined by (9.9) and (9.10) with $v = 0$, so without translating the Poisson clocks. The ℓ^1

norm on $\mathbf{R}^d$ and $\mathbf{Z}^d$ is denoted by $|x|_1 = |x_1| + \cdots + |x_d|$ for $x = (x_1, \ldots, x_d)$. A function g on $\mathbf{R}^d$ is *concave* if for all $x, y \in \mathbf{R}^d$ and $0 < \alpha < 1$,

$$g(\alpha x + (1-\alpha)y) \geq \alpha g(x) + (1-\alpha)g(y).$$

Theorem 9.3 *There exists a function $g : \mathbf{R}^d \to (-\infty, 0]$ such that the following statement holds for almost every realization of the Poisson clocks:*

$$\lim_{n\to\infty} \frac{1}{n}\xi_{nt}([nx]) = tg\Big(\frac{x}{t}\Big) \qquad \text{for all } x \in \mathbf{R}^d \text{ and } t > 0. \tag{9.13}$$

The deterministic limiting function g is concave, Lipschitz continuous, and satisfies

$$g(x) = \sum_{i=1}^{d} K_i(x_i \wedge 0) \tag{9.14}$$

for all $x \in \mathbf{R}^d$ such that $|x|_1 > 1$.

Note that for $t = 0$ the limit in (9.13) is given by the right-hand side of equation (9.14), by virtue of the initial height ξ_0 defined by (9.9) and (9.7).

Next we consider the general process. Assume that on some probability space is defined a sequence $\{h_0^n : n \in \mathbf{N}\}$ of random initial height profiles. The assumption is that this sequence has a macroscopic profile in the following sense. There exists a deterministic function $\psi_0 : \mathbf{R}^d \to \mathbf{R}$ such that these limits in probability hold: for each $y \in \mathbf{R}^d$ and each $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ \left| n^{-1}h_0^n([ny]) - \psi_0(y) \right| \geq \varepsilon \right\} = 0. \tag{9.15}$$

The notation $[ny]$ above means coordinatewise integer parts. For $x = (x_1, \ldots, x_d) \in \mathbf{R}^d$, $[x] = ([x_1], \ldots, [x_d]) \in \mathbf{Z}^d$ where for a real r, $[r] = \max\{k \in \mathbf{Z} : k \leq r\}$ is the integer part of r.

Augment the probability space of the initial profiles $\{h_0^n\}$ so it also supports the Poisson clocks $\{\mathcal{T}_x\}$, independent of $\{h_0^n\}$. For each n, construct the process h_t^n with these Poisson clocks and initial state h_0^n. The objective is to prove that the random evolution h_t^n converges under suitable space and time scaling.

By the restrictions (9.1) on admissible profiles the function ψ_0 in (9.15) must satisfy the Lipschitz property

$$0 \leq \psi_0(x + re_i) - \psi_0(x) \leq K_i r \qquad \text{for } 1 \leq i \leq d \text{ and } r > 0. \tag{9.16}$$

Define a function $\psi(x,t)$ for $(x,t) \in \mathbf{R}^d \times [0,\infty)$ by $\psi(x,0) = \psi_0(x)$ and for $t > 0$

$$\psi(x,t) = \sup_{y\in\mathbf{R}^d} \left\{ \psi_0(y) + tg\Big(\frac{x-y}{t}\Big) \right\}. \tag{9.17}$$

The function g inside the braces is the one defined by the limit (9.13). It is determined by the rules of the process, and does not depend on the particular initial conditions h_0^n or ψ_0. Continuity, (9.16) and (9.14) together imply that the supremum in (9.17) is attained at some y such that $|x-y|_1 \le t$ (Lemma 9.14 in Section 9.5.2). As a function of x, $\psi(x,t)$ satisfies the Lipschitz bounds (9.16) for each fixed t.

Theorem 9.4 *Assume* (9.15). *For each* $x \in \mathbf{R}^d$ *and* $t \ge 0$, *we have the following limit in probability. For any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ \left| n^{-1} h_{nt}^n([nx]) - \psi(x,t) \right| \ge \varepsilon \right\} = 0. \tag{9.18}$$

The pathwise variational formula (9.12) certainly suggests (9.17) as a natural description of a limiting macroscopic evolution. We can also describe the limit with a partial differential equation. From Section A.12 in the Appendix, we see that (9.17) is an example of a Hopf-Lax formula which gives the solution of a Hamilton-Jacobi equation. Let

$$f(\rho) = \inf_{x\in\mathbf{R}^d} \{\rho \cdot x - g(x)\} \quad \text{for} \quad \rho = (\rho_1, \dots, \rho_d) \in \mathbf{R}^d \tag{9.19}$$

be the concave conjugate of g. Let

$$V = \prod_{i=1}^d [0, K_i]$$

be the set of possible gradients of macroscopic height profiles that satisfy Lipschitz condition (9.16). Partial derivative with respect to time is $\psi_t = \partial\psi/\partial t$, and the spatial gradient is $\nabla\psi = (\psi_{x_1}, \dots, \psi_{x_d})$.

Theorem 9.5 *The limiting height function* ψ *is Lipschitz continuous on* $\mathbf{R}^d \times [0,\infty)$, *and therefore differentiable Lebesgue almost everywhere. At every point* (x,t) *of differentiability in* $\mathbf{R}^d \times (0,\infty)$, $\nabla\psi(x,t) \in V$ *and*

$$\frac{\partial\psi}{\partial t}(x,t) = -f\big(\nabla\psi(x,t)\big). \tag{9.20}$$

The velocity function f *defined by* (9.19) *is concave and continuous on* V. *It satisfies* $f = 0$ *on the boundary of* V, *and* $0 < f \le 1$ *in the interior of* V. *Outside* V *the function* f *is identically* $-\infty$.

Let I *be any subset of* $\{1, \dots, d\}$. *Suppose* $\rho, \widetilde{\rho} \in V$ *satisfy*

$$\widetilde{\rho}_i = \rho_i \text{ for } i \notin I, \text{ and } \widetilde{\rho}_i = K_i - \rho_i \text{ for } i \in I. \tag{9.21}$$

Then $f(\rho) = f(\widetilde{\rho})$.

At the current level of generality we cannot say much more about the functions f and g. Only in the case $d = 1$ and $K_1 = 1$ do we have an explicit expression: $f(\rho) = \rho(1-\rho)$ for $0 \le \rho \le 1$. This case is the totally asymmetric nearest-neighbor exclusion process in one dimension. In Theorem 9.10 below we calculate f and g explicitly from the equilibrium Bernoulli measures. A way to calculate f from the distributions of the state of the process is given in the next theorem.

Define the event

$$B = \{h \in H : h(0) \ge h(-e_i) + 1 \text{ and } h(0) \ge h(e_i) - K_i + 1 \text{ for } 1 \le i \le d\}. \tag{9.22}$$

A jump time $t \in \mathcal{T}_0$ causes a jump $h_t(0) = h_{t-}(0) - 1$ iff $h_{t-} \in B$.

Theorem 9.6 *Fix a vector $\rho \in V$, the set of admissible gradients. Let h_0 be a deterministic or random initial height function such that $E|h_0(0)|^2 < \infty$ and for every $x \in \mathbf{R}^d$,*

$$\lim_{n\to\infty} n^{-1} h_0([nx]) = \rho \cdot x \quad \textit{in probability.} \tag{9.23}$$

Let h_t be the process constructed from initial condition h_0. Then

$$f(\rho) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}\{h_s \in B\}\, ds = \lim_{t\to\infty} \frac{1}{t} \int_0^t P\{h_s \in B\}\, ds, \tag{9.24}$$

where the first limit is in probability.

Example 9.7 Here is an initial height function h_0 that satisfies (9.23) and has spatially ergodic increments $\eta_0(u,i) = h_0(u) - h_0(u - e_i)$. Let

$$\{\eta^i(\ell) : \ell \in \mathbf{Z}, 1 \le i \le d\}$$

be mutually independent random variables such that for each i, $\{\eta^i(\ell) : \ell \in \mathbf{Z}\}$ are i.i.d. with values in $\{0, 1, \dots, K_i\}$ and common mean $E\eta^i(\ell) = \rho_i$. For each $1 \le i \le d$ define a one-dimensional height function on $\mathbf{Z}$ by

$$\sigma^i(m) = \begin{cases} -\sum_{\ell=m+1}^{0} \eta^i(\ell), & m < 0 \\ 0, & m = 0 \\ \sum_{\ell=1}^{m} \eta^i(\ell), & m > 0. \end{cases}$$

With these, define the height function h_0 by

$$h_0(u) = \sum_{i=1}^{d} \sigma^i(u_i) \qquad \text{for } u = (u_1, \dots, u_d) \in \mathbf{Z}^d.$$

The convergence in Theorems 9.4 and 9.6 can be made almost sure simply by making the convergence in the hypotheses almost sure. The additional technical tool needed for the proof is a summable deviation estimate for the convergence in Theorem 9.3. Such estimates can be obtained conveniently through the last-passage representation in Section 9.4.

Next we address a point of partial differential equations theory. Equation (9.20) is a Hamilton-Jacobi equation. Even with smooth initial data, its solutions can develop points of nondifferentiability, or *shocks* (Exercise 9.4). Consequently it is not enough to consider classical solutions that are differentiable everywhere. One is forced to deal with weak solutions. But with weak solutions comes the possibility of nonuniqueness. What is needed then is the correct definition to capture the physically relevant solution among the many weak solutions. The appropriate notion of weak solution for a Hamilton-Jacobi equation is the so-called viscosity solution. Its rather abstract intrinsic definition is the following.

Suppose F is a continuous function defined on $\mathbf{R}^d$. A continuous function $u(x,t)$ on $\mathbf{R}^d \times [0,\infty)$ is a *viscosity solution* of

$$u_t + F(\nabla u) = 0\,, \quad u|_{t=0} = u_0$$

if

(i) $u(x,0) = u_0(x)$ for all $x \in \mathbf{R}^d$, and

(ii) if the following holds for all continuously differentiable functions ϕ on $\mathbf{R}^d \times (0,\infty)$: if $u - \phi$ has a local maximum at (x_0,t_0), then

$$\phi_t(x_0,t_0) + F\big(\nabla\phi(x_0,t_0)\big) \le 0\,, \tag{9.25}$$

and if $u - \phi$ has a local minimum at (x_0,t_0), then

$$\phi_t(x_0,t_0) + F\big(\nabla\phi(x_0,t_0)\big) \ge 0\,. \tag{9.26}$$

As always with notions of weak solutions, the point is to move the derivatives onto a smooth test function so that no differentiability requirements are imposed on the solution itself. Chapter 10 in Evans's textbook [14] discusses this notion of viscosity solution and explains the motivation behind it. See also Section A.12 in the Appendix.

To apply the definition to our setting, first extend f to a continuous function $\bar f$ on all of $\mathbf{R}^d$. The simplest way to do this is to set $\bar f \equiv 0$ outside V. The exact nature of the extension of f turns out not to influence the definition. For as we show later in the proof, if $u = \psi$ defined by the Hopf-Lax formula (9.17), ϕ is continuously differentiable, and $u - \phi$ has a local maximum or minimum at (x_0,t_0), then $-1 \le \phi_t(x_0,t_0) \le 0$ and $\nabla\phi(x_0,t_0) \in V$.

Theorem 9.8 *Let ψ_0 be the initial height function in assumption* (9.15)*, and ψ the limiting profile in* (9.18)*, defined by the Hopf-Lax formula* (9.17)*. Then for any $0 < T < \infty$, ψ is the* *unique uniformly continuous viscosity solution of the Hamilton-Jacobi equation*

$$\psi_t + \bar{f}(\nabla\psi) = 0\,, \quad \psi|_{t=0} = \psi_0 \tag{9.27}$$

on $\mathbf{R}^d \times [0,T]$.

Before turning to the proofs of the theorems we derive as a corollary a hydrodynamic limit for the K-exclusion process.

9.3 Hydrodynamic limit for totally asymmetric nearest-neighbor K-exclusion in one dimension

Fix an integer $1 \le K < \infty$. Consider the special case of the K-exclusion process in one dimension that permits only nearest-neighbor jumps to the right. The state of the process at time t is the sequence $\eta_t = (\eta_t(u) : u \in \mathbf{Z})$ of occupation numbers $\eta_t(u) \in \{0, 1, \ldots, K\}$, and the state space is the compact space $X = \{0, 1, \ldots, K\}^{\mathbf{Z}}$. The process is constructed with an i.i.d. family $\{\mathcal{T}_u : u \in \mathbf{Z}\}$ of rate one Poisson point processes on the time line $[0,\infty)$. The jump rule is this: at jump times $t \in \mathcal{T}_u$ one particle is moved from site u to $u+1$, provided immediately before the jump there is at least one particle present at u and at most $K-1$ particles present at $u+1$. The generator given in (7.1) specializes to

$$Lf(\eta) = \sum_{u\in\mathbf{Z}} \mathbf{1}\{\eta(u) \ge 1,\, \eta(u+1) \le K-1\}[f(\eta^{u,u+1}) - f(\eta)] \tag{9.28}$$

where

$$\eta^{u,u+1}(v) = \begin{cases} \eta(u) - 1, & v = u \\ \eta(u+1) + 1, & v = u+1 \\ \eta(v), & v \notin \{u, u+1\} \end{cases} \tag{9.29}$$

is the configuration that results from moving a single particle from u to $u+1$.

One motivation for introducing the interface process h_t was that it represents the K-exclusion process. Specialize the process h_t described in Section 9.1 to one dimension, with state space

$$H = \{h : \mathbf{Z} \to \mathbf{Z} :\; 0 \le h(u) - h(u-1) \le K \text{ for } u \in \mathbf{Z}\}. \tag{9.30}$$

Given an initial configuration η_0 for the K-exclusion process, define an initial height function $h_0 \in H$ by

$$h_0(0) = 0\,, \qquad h_0(u) - h_0(u-1) = \eta_0(u) \quad \text{for all } u \in \mathbf{Z}. \tag{9.31}$$

When processes η_t and h_t read the same Poisson clocks, the relation

$$h_t(u) - h_t(u-1) = \eta_t(u) \quad \text{for all } u \in \mathbf{Z} \tag{9.32}$$

is preserved for all time. The jumps of h_t keep track of particle current for η_t. The number of particles that have jumped across edge $\{u, u+1\}$ during time interval $(s, t\,]$ equals $h_s(u) - h_t(u)$.

We get a hydrodynamic limit for the K-exclusion process as an immediate corollary of Theorem 9.4. Assume we have a sequence of random or deterministic initial configurations η_0^n, $n = 1, 2, 3, \ldots$, defined on some probability space, together with i.i.d. Poisson clocks $\{\mathcal{T}_u\}$ independent of $\{\eta_0^n\}$. Construct on this probability space the totally asymmetric, nearest-neighbor K-exclusion processes η_t^n as described above.

The only hypothesis needed for the hydrodynamic limit is the existence of a macroscopic density profile at time 0. Let $0 \le \rho_0(x) \le K$ be a bounded measurable function on $\mathbf{R}$. Assume that for all $-\infty < a < b < \infty$ and all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ \left| n^{-1} \sum_{u=[na]+1}^{[nb]} \eta_0^n(u) - \int_a^b \rho_0(x)\,dx \right| > \varepsilon \right\} = 0. \tag{9.33}$$

To produce the macroscopic limiting density function $\rho(x, t)$ for the K-exclusion process, first define ψ_0 as the antiderivative of ρ_0 given by

$$\psi_0(0) = 0\,, \qquad \int_a^b \rho_0(x)\,dx = \psi_0(b) - \psi_0(a) \quad \text{for all } -\infty < a < b < \infty. \tag{9.34}$$

Let $\psi(x, t)$ be the viscosity solution of the Hamilton-Jacobi equation (9.27), defined by the Hopf-Lax formula (9.17). The function g in (9.17) is now defined by the limit

$$g(x) = \lim_{n\to\infty} n^{-1} \xi_n([nx])$$

of the one-dimensional process started from the wedge

$$\xi_0(u) = \begin{cases} 0, & u \ge 0 \\ Ku, & u < 0 \end{cases}$$

given by Theorem 9.3. The function $\psi(x, t)$ is Lipschitz continuous, and so the partial x-derivative exists Lebesgue almost everywhere:

$$\rho(x, t) = \frac{\partial}{\partial x} \psi(x, t). \tag{9.35}$$

This function $\rho(x, t)$ is a weak solution of the scalar conservation law

$$\rho_t + f(\rho)_x = 0\,, \quad \rho(x, 0) = \rho_0(x). \tag{9.36}$$

The flux f of the conservation law is the concave conjugate of g.

Theorem 9.9 *Under assumption* (9.33) *this limit in probability holds: for all* $-\infty < a < b < \infty$ *and* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ \left| n^{-1} \sum_{u=[na]+1}^{[nb]} \eta^n_{nt}(u) - \int_a^b \rho(x,t)\,dx \right| > \varepsilon \right\} = 0. \tag{9.37}$$

To derive this theorem from Theorem 9.4, just observe that by (9.32)

$$\sum_{u=[na]+1}^{[nb]} \eta^n_{nt}(u) = h^n_{nt}([nb]) - h^n_{nt}([na])$$

and by the Lipschitz property of ψ

$$\int_a^b \rho(x,t)\,dx = \psi(b,t) - \psi(a,t).$$

Now (9.37) follows from (9.18).

When we specialize to the case $K = 1$, we can finally get an explicit formula for the flux f.

Theorem 9.10 *For the totally asymmetric nearest-neighbor exclusion process in one dimension (the case* $K = 1$*), the flux function is*

$$f(\rho) = \rho(1-\rho)\,, \quad 0 \le \rho \le 1. \tag{9.38}$$

The limiting shape g *for the wedge growth is the concave conjugate of* f *given by*

$$g(x) = \begin{cases} x, & x < -1 \\ -\frac{1}{4}(1-x)^2, & -1 \le x \le 1 \\ 0, & x > 1. \end{cases} \tag{9.39}$$

Proof. Now $V = [0,1]$. On the boundary $f(0) = f(1) = 0$, either by considering the process with no particles for $\rho = 0$ and the process with no empty sites for $\rho = 1$, or by Theorem 9.5.

Let $\rho \in (0,1)$. We apply Theorem 9.6. Let η_t be a process in equilibrium with Bernoulli distribution ν_ρ. Define the initial height function h_0 by (9.31). Then the hypotheses of Theorem 9.6 are satisfied. The event B defined by (9.22) is the same as

$$B = \{\eta(0) = 1, \eta(1) = 0\}.$$

By stationarity and the definition of the Bernoulli measure,

$$f(\rho) = \lim_{t\to\infty} \frac{1}{t} \int_0^t P\{\eta_s(0) = 1, \eta_s(1) = 0\}\, ds = \nu^\rho\{\eta(0) = 1, \eta(1) = 0\} = \rho(1-\rho).$$

Since g is concave and upper semicontinuous (in fact, continuous by Theorem 9.3), it is its own double conjugate, so

$$g(x) = \inf_{0\le\rho\le 1} \{x\rho - f(\rho)\}.$$

We refer to Section 12 in Rockafellar's monograph [31] for justification of this, and further basics of convex (and concave) duality. From this equation formula (9.39) can be derived. ∎

9.4 The last-passage percolation model

In this section we reformulate the process ξ_t with the wedge initial condition as a last-passage percolation model. The last-passage approach is convenient for certain tasks such as probability bounds, large deviations, and distributional limits. See the notes of this chapter for references.

Recall again the construction of the wedge process. Initially

$$\xi_0(u) = \sum_{i=1}^d K_i(u_i \wedge 0) \quad \text{ for } u = (u_1, \dots, u_d) \in \mathbf{Z}^d.$$

At jump times $t \in \mathcal{T}_u$, $\xi_t(u) = \xi_{t-}(u) - 1$, provided just before time t the inequalities

$$\xi_{t-}(u) \ge \xi_{t-}(u - e_i) + 1 \text{ and } \xi_{t-}(u) \ge \xi_{t-}(u + e_i) - K_i + 1 \text{ for } 1 \le i \le d$$

were valid.

As the height function ξ_t marches downward, the set A_t between the graphs of ξ_0 and ξ_t describes a randomly growing finite set of points or *cells* in $\mathbf{Z}^{d+1}$. To describe this precisely, define the set

$$\mathbf{L} = \left\{(u, h) \in \mathbf{Z}^d \times \mathbf{Z} : h < \sum_{i=1}^d K_i(u_i \wedge 0)\right\}$$

with outer boundary

$$\partial_e \mathbf{L} = \left\{(u, h) \in \mathbf{Z}^d \times \mathbf{Z} : h = \sum_{i=1}^d K_i(u_i \wedge 0)\right\}.$$

Note the notation: a cell in $\mathbf{Z}^{d+1}$ is denoted by a pair (u, h) where the first component u is a site in $\mathbf{Z}^d$ and the second component h is a $\mathbf{Z}$-valued height.

The boundary $\partial_e \mathbf{L}$ is the graph of the initial height profile ξ_0. The growing set

$$A_t = \{(u, h) : \xi_0(u) > h \geq \xi_t(u)\}$$

is a finite subset of $\mathbf{L}$, and adds a new cell whenever ξ_t experiences a jump. For $(u, h) \in \mathbf{L}$ let

$$T(u, h) = \inf\{t \geq 0 : (u, h) \in A_t\} = \inf\{t \geq 0 : \xi_t(u) \leq h\}$$

be the time when cell (u, h) joins the set A_t. The initial values are $T(u, h) = 0$ for all $(u, h) \notin \mathbf{L}$. The random times $T(u, h)$ are stopping times for the filtration

$$\mathcal{H}_t = \sigma\left\{N_u(s) : u \in \mathbf{Z}^d, 0 \leq s \leq t\right\}$$

of the Poisson processes.

The rules of the process ξ_t tell us that before (u, h) can join the growing set A_t, each of the $2d+1$ cells $(u, h+1)$, $(u - e_i, h)$, and $(u + e_i, h + K_i)$ for $1 \leq i \leq d$ must have already joined A_t or must lie outside $\mathbf{L}$. After this (u, h) must wait a mean one exponential random time to join, and this random time is independent of the past. Let $Y_{u,h}$ denote this waiting time. It is defined for $(u, h) \in \mathbf{L}$ by

$$\begin{aligned} Y_{u,h} &= T(u, h) - \max\{T(u, h+1), T(u - e_i, h), \\ &\qquad\qquad T(u + e_i, h + K_i) : 1 \leq i \leq d\}. \end{aligned} \tag{9.40}$$

The *immediate predecessors* of (u, h) are those of the $2d+1$ cells $(u, h+1)$, $(u - e_i, h)$, and $(u+e_i, h+K_i)$ that lie in $\mathbf{L}$. Any cell that can be reached from (u, h) by taking successive steps from a cell to one of its immediate predecessors is called a *predecessor* of (u, h). Starting with any $(u, h) \in \mathbf{L}$ and following backwards any path of immediate predecessors in $\mathbf{L}$ ultimately leads to cell $(0, -1)$.

Lemma 9.11 *The random variables $\{Y_{u,h} : (u, h) \in \mathbf{L}\}$ are i.i.d. mean one exponentials.*

Proof. We can arrange the cells of $\mathbf{L}$ in a sequence

$$\mathbf{L} = \{(u_1, h_1), (u_2, h_2), (u_3, h_3), \dots\}$$

with the property that for each n the immediate predecessors of (u_n, h_n) are among the (u_1, h_1), (u_2, h_2), $\dots$, (u_{n-1}, h_{n-1}). Necessarily $(u_1, h_1) = (0, -1)$. After that the sequence can be constructed through the following inductive steps.

Set $Q_1 = \{(u_1, h_1)\} = \{(0, -1)\}$.

Assume $Q_m = \{(u_1, h_1), (u_2, h_2), \ldots, (u_m, h_m)\}$ has been constructed, and that it has the property mentioned above. Call $(v, k) \in \mathbf{L}$ a *growth cell* for Q_m if $(v, k) \notin Q_m$ but all the immediate predecessors of (v, k) are in Q_m. Let (u_{m+1}, h_{m+1}) be an arbitrary growth cell of Q_m and define

$$Q_{m+1} = \{(u_1, h_1), (u_2, h_2), \ldots, (u_m, h_m), (u_{m+1}, h_{m+1})\}.$$

Continue in this manner.

Let $(u, h) = (u_n, h_n)$ and $Q = Q_{n-1}$. We show that $Y_{u,h}$ is a mean one exponential independent of $\{T(v, k) : (v, k) \in Q\}$. The waiting times $\{Y_{v,k} : (v, k) \in Q\}$ are a function of the passage times $\{T(v, k) : (v, k) \in Q\}$, so we can conclude inductively that Y_{u_1,h_1}, Y_{u_2,h_2}, $\ldots$, Y_{u_n,h_n} are i.i.d. mean one exponentials for any n.

Set

$$S(u, h) = \max\{T(u, h + 1), T(u - e_i, h), T(u + e_i, h + K_i) : 1 \le i \le d\}. \tag{9.41}$$

$S(u, h)$ is a stopping time for the Poisson processes. It is the time at which the jump of $\xi(u)$ from $h + 1$ to h becomes admissible. The remaining time $Y_{u,h}$ till this jump happens is the first jump time in the restarted Poisson process $\theta_{S(u,h)}\mathcal{T}_u$.

By the strong Markov property, the restarted Poisson processes $\{\theta_{S(u,h)}\mathcal{T}_w : w \in \mathbf{Z}^d\}$ are again i.i.d. rate one Poisson processes, independent of the σ-algebra $\mathcal{H}_{S(u,h)}$. Thus we find that the restarted process $\theta_{S(u,h)}\mathcal{T}_u$ is independent of the past $\mathcal{H}_{S(u,h)}$ and of the other restarted processes $\{\theta_{S(u,h)}\mathcal{T}_w : w \ne u\}$.

The passage times $\{T(v, k) : (v, k) \in Q\}$ can be constructed from the information in $\mathcal{H}_{S(u,h)}$ and $\{\theta_{S(u,h)}\mathcal{T}_w : w \ne u\}$. This information contains the entire Poisson prosesses $\mathcal{T}_w$ for $w \ne u$, and $\mathcal{T}_u$ up to time $S(u, h)$. Any (v, k) that lies on a backward path of predecessors started from (u, h) must have $T(v, k) \le S(u, h)$. Such a $T(v, k)$ is measurable with respect to $\mathcal{H}_{S(u,h)}$. For any other $(v, k) \in Q$, $T(v, k)$ can be constructed from the passage times of the predecessors of (u, h) and the Poisson processes $\mathcal{T}_w$, $w \ne u$.

To summarize, we have shown that $Y_{u,h}$ is the first jump time in the restarted Poisson process $\theta_{S(u,h)}\mathcal{T}_u$ which is independent of $\{T(v, k) : (v, k) \in Q\}$. In particular, $Y_{u,h}$ is a mean one exponential independent of $\{T(v, k) : (v, k) \in Q\}$. ∎

Let $\Pi(u, h)$ be the collection of paths

$$\pi = \{(0, -1) = (v_1, k_1), (v_2, k_2), \ldots, (v_n, k_n) = (u, h)\} \tag{9.42}$$

in $\mathbf{L}$ that lead from $(0, -1)$ to (u, h), and each step of which is one of $2d + 1$ admissible steps: for each $1 \le \ell < n$,

$$(v_{\ell+1}, k_{\ell+1}) - (v_\ell, k_\ell) = (0, -1),\ (e_i, 0),\ \text{or}\ (-e_i, -K_i)\ \text{for some}\ 1 \le i \le d. \tag{9.43}$$

Write (9.40) in the form

$$T(u,h) = \max\{T(u,h+1), T(u-e_i,h), T(u+e_i,h+K_i) : 1 \le i \le d\} + Y_{u,h}.$$

Iterate this backwards until all paths have reached $(0,-1)$. Then we have the last-passage representation

$$T(u,h) = \max_{\pi \in \Pi(u,h)} \sum_{(v,k)\in\pi} Y_{v,k}. \tag{9.44}$$

The term "last-passage" is used because the *slowest* path determines the passage time $T(u,h)$, in contrast with first-passage percolation where a passage time is determined by the fastest path.

Generalize the definitions to paths and passage times between arbitrary cells. For $(z,\ell) \in \mathbf{L} \cup \partial_e \mathbf{L}$ and $(u,h) \in \mathbf{L}$, let $\Pi((z,\ell),(u,h))$ be the collection of paths

$$\pi = \{(v_1,k_1),(v_2,k_2),\ldots,(v_n,k_n)\}$$

such that all steps are admissible as in (9.43), $(v_1,k_1)-(z,\ell)$ is also an admissible step, and $(v_n,k_n)=(u,h)$. Define the passage time between (z,ℓ) and (u,h) by

$$T((z,\ell),(u,h)) = \max_{\pi \in \Pi((z,\ell),(u,h))} \sum_{(v,k)\in\pi} Y_{v,k}. \tag{9.45}$$

Then we have the superadditivity

$$T((z,\ell),(u,h)) \ge T((z,\ell),(v,k)) + T((v,k),(u,h)) \tag{9.46}$$

whenever admissible paths are possible between (z,ℓ) and (v,k) and between (v,k) and (u,h). The starting cell (z,ℓ) was not included in a path π in $\Pi((z,\ell),(u,h))$ so that the cost $Y_{v,k}$ would not be counted twice in (9.46), once as the last cell in $T((z,\ell),(v,k))$ and again as the first cell in $T((v,k),(u,h))$. The earlier definition (9.44) relates to (9.45) via $T(u,h) = T((0,0),(u,h))$.

We finish this section with an estimate for later use.

Lemma 9.12 *For $(u,h) \in \mathbf{L}$, let $M = 2|h| + \sum_{i=1}^d u_i$. The definition of $\mathbf{L}$ guarantees that M is a strictly positive integer. Then for $s > 1$,*

$$P\left[T(u,h) \ge Ms\right] \le (2d+1)^M \exp\bigl(-M \cdot I(s)\bigr) \tag{9.47}$$

where $I(s) = s - 1 - \log s$ is the Cramér rate function for large deviations of mean one exponential random variables.

Proof. Estimating in the crudest way,

$$P\big[T(u,h) \geq Ms\big] \leq \sum_{\pi \in \Pi(u,h)} P\left[S^1_{|\pi|} \geq Ms\right]$$

where $|\pi|$ is the number of cells in path π and S^1_m represents a sum of m i.i.d. mean one exponential random variables.

We claim that M is an upper bound on the number of cells in a path π that satisfies (9.42) and (9.43). If we let $(v_0, k_0) = (0, 0)$, we can write

$$(u,h) = \sum_{\ell=1}^{|\pi|} \big((v_\ell, k_\ell) - (v_{\ell-1}, k_{\ell-1})\big) = a(0,-1) + \sum_{i=1}^{d} b_i(e_i, 0) + \sum_{i=1}^{d} c_i(-e_i, -K_i)$$

where a, $b_1, \ldots, b_d$ and $c_1, \ldots, c_d$ count how many times the $2d+1$ different steps appear in π. From the coordinate by coordinate equations above,

$$\begin{aligned}
|\pi| &= a + \sum_{i=1}^{d} b_i + \sum_{i=1}^{d} c_i \\
&= 2\Big(a + \sum_{i=1}^{d} c_i K_i\Big) + \sum_{i=1}^{d} (b_i - c_i) + \sum_{i=1}^{d} c_i\,(2 - 2K_i) - a \\
&= -2h + \sum_{i=1}^{d} u_i + \sum_{i=1}^{d} c_i\,(2 - 2K_i) - a \\
&\leq 2|h| + \sum_{i=1}^{d} u_i = M.
\end{aligned}$$

For the last inequality we used $h < 0$, $c_i \geq 0$, $K_i \geq 1$ and $a \geq 0$. From this,

$$P\left[S^1_{|\pi|} \geq Ms\right] \leq P\left[S^1_M \geq Ms\right] \leq \exp\big(-M \cdot I(s)\big)$$

by the large deviation bounds in Proposition A.18. The assumption $s > 1$ is used here. The bound would be trivial for $s = 1$, and not true for $s < 1$ because then the infimum of $I(x)$ over $[s, \infty)$ would be 0 and not $I(s)$.

It remains to observe that the total number of paths in $\Pi(u,h)$ cannot exceed $(2d+1)^M$ because each path has at most M steps, and each step is chosen from the set of $2d+1$ admissible steps. ∎

9.5 Proofs

We prove the theorems discussed in Section 9.2.

9.5.1 Proof of Theorem 9.3

To express random variables as functions of the Poisson processes, we write $\omega = (\mathcal{T}_u : u \in \mathbf{Z}^d)$ for the entire collection of Poisson clocks, and also $\omega_u = \mathcal{T}_u$ for the Poisson clock at site u. The proof of the wedge limit is an application of the subadditive ergodic theorem.

We prove the almost sure convergence

$$\lim_{n\to\infty} \frac{1}{n}\xi_{n\tau}([nx]) = \tau\, g\Big(\frac{x}{\tau}\Big) \tag{9.48}$$

for successively more general points x, all $\tau > 0$, and simultaneously develop properties of the limit. The main tool is the subadditive ergodic theorem stated as Theorem A.12 in the appendix.

First suppose $x = z \in \mathbf{Z}^d$ in (9.48). Keep (z, τ) fixed for the moment. Define

$$X_{0,n} = \xi_{n\tau}(nz).$$

For positive integers m, define a height process $\sigma_t^{(m)}$ by stipulating that initially

$$\sigma_0^{(m)}(u) = \xi_{m\tau}(mz) + w(u), \quad u \in \mathbf{Z}^d, \tag{9.49}$$

and that the height value $\sigma^{(m)}(u)$ attempts jumps at the jump times of Poisson clock $\theta_{m\tau}\mathcal{T}_{mz+u}$. Here θ_s is a time-shift on the Poisson process that moves the time origin to point s. In terms of the set of jump times,

$$\theta_s\mathcal{T}_u = \{t - s : t \in \mathcal{T}_u, t > s\}.$$

So $\sigma_t^{(m)}$ is a new wedge process, with time origin at $m\tau$ and space-height origin at $(mz, \xi_{m\tau}(mz))$.

Let us also introduce a space-time shift $\theta_{u,s}$, $(u, s) \in \mathbf{Z}^d \times [0,\infty)$, of the Poisson processes. $\theta_{u,s}$ translates the spatial index by u and restarts the clocks at time s, $(\theta_{u,s}\omega)_v = \theta_s\mathcal{T}_{u+v}$. The probability measure $\mathbf{P}$ on the Poisson processes is stationary and ergodic under these space-time shifts. In terms of the space-time shift, process $\sigma_t^{(m)}$ reads Poisson clocks $\theta_{mz,m\tau}\omega$.

The process $h'_t(u) = \xi_{m\tau+t}(mz+u)$ reads exactly the same clocks as $\sigma_t^{(m)}$. By the minimality property of the wedge profile w,

$$\sigma_0^{(m)}(u) = \xi_{m\tau}(mz) + w(u) \le \xi_{m\tau}(mz+u) = h'_0(u) \quad \text{for all } u \in \mathbf{Z}^d.$$

By attractivity, the inequality $\sigma_t^{(m)} \le h'_t$ is preserved for all time t. In particular for $0 < m < n$,

$$\sigma^{(m)}_{(n-m)\tau}((n-m)z) \le h'_{(n-m)\tau}((n-m)z) = \xi_{n\tau}(nz), \tag{9.50}$$

which is the same as

$$\sigma^{(m)}_{(n-m)\tau}((n-m)z) - \sigma_0^{(m)}(0) + \xi_{m\tau}(mz) \le \xi_{n\tau}(nz).$$

If we define a space-time increment in the $\sigma_t^{(m)}$ process by

$$X_{m,n} = \sigma^{(m)}_{(n-m)\tau}((n-m)z) - \sigma_0^{(m)}(0),$$

we get the superadditivity

$$X_{0,m} + X_{m,n} \le X_{0,n} \qquad \text{for } 0 \le m < n. \tag{9.51}$$

We check the other conditions of the subadditive ergodic theorem for the variables $-X_{m,n}$. Let $X_{0,n} = F_n(\omega)$ represent $X_{0,n} = \xi_{n\tau}(nz)$ as a function of the Poisson processes ω. Then $X_{m,n} = F_{n-m}(\theta_{mz,m\tau}\omega)$. In particular, it follows that the sequence

$$\{X_{n\ell,(n+1)\ell} : n \ge 1\} = \{F_\ell(\theta_{n\ell z, n\ell\tau}\omega) : n \ge 1\}$$

is stationary and ergodic for a fixed ℓ. The distribution of the sequence

$$\{X_{m,m+n} : n \ge 1\} = \{F_n(\theta_{mz,m\tau}\omega) : n \ge 1\}$$

does not depend on m because $\theta_{mz,m\tau}\omega$ has the same distribution as ω.

Finally, $-X_{0,n}$ is nonnegative, and bounded above by the total number of jump attempts of $\xi(nz)$ up to time $n\tau$, which is a Poisson random variable with mean $n\tau$. This gives the moment conditions needed for the Subadditive Ergodic Theorem A.12. We conclude that for each $(z,\tau) \in \mathbf{Z}^d \times [0,\infty)$, there exists a finite deterministic number $g_0(z,\tau)$ such that

$$\lim_{n\to\infty} \frac{1}{n}\xi_{n\tau}(nz) = g_0(z,\tau) \qquad \text{almost surely.} \tag{9.52}$$

By the time monotonicity of the process, for $s < t$ and fixed n,

$$\xi_{ns}(nz) \ge \xi_{nt}(nz) \ge \xi_{ns}(nz) - N_{nz}(ns,nt]. \tag{9.53}$$

$N_{nz}(ns,nt] = |\mathcal{T}_{nz} \cap (ns,nt]\,|$ is the number of jump attempts height variable $\xi(nz)$ experiences during time interval $(ns,nt]$. $N_{nz}(ns,nt]$ has Poisson$(n(t-s))$ distribution, and so can be thought of as a sum of n i.i.d. Poisson$(t-s)$ variables. Large deviation estimates such as those in Proposition A.18 (or in Exercise A.8) and the Borel-Cantelli lemma imply that

$$\lim_{n\to\infty} \frac{1}{n} N_{nz}(ns,nt] = t - s \qquad \text{almost surely.}$$

Letting $n \to \infty$ in (9.53) gives

$$g_0(z,s) \ge g_0(z,t) \ge g_0(z,s) - (t-s) \qquad \text{for } s < t. \tag{9.54}$$

The function g_0 has also the following homogeneity property. For any positive integer m,

$$\begin{aligned} g_0(mz, mt) &= \lim_{n\to\infty} \frac{1}{n}\xi_{nmt}([nmz]) = \lim_{N\to\infty} \frac{m}{N}\xi_{Nt}([Nz]) \\ &= mg_0(z, t). \end{aligned}$$

Let $q \in \mathbf{Q}^d$ be a vector with rational coordinates. The homogeneity implies that we can unambiguously define

$$g_0(q, t) = \frac{1}{m} g_0(mq, mt) \tag{9.55}$$

for any positive integer m such that $mq \in \mathbf{Z}^d$. And then homogeneity extends: for any $q \in \mathbf{Q}^d$, $t > 0$, and positive rational r,

$$g_0(rq, rt) = rg_0(q, t). \tag{9.56}$$

Now we argue that the almost sure limit

$$\lim_{n\to\infty} \frac{1}{n}\xi_{nt}([nq]) = g_0(q, t) \tag{9.57}$$

holds for any $q \in \mathbf{Q}^d$ and $t \geq 0$. Fix m such that $mq \in \mathbf{Z}^d$. By limit (9.52) and definition (9.55), limit (9.57) holds along the subsequence $n = \ell m$, $\ell \to \infty$. To fill in the remaining terms, note that by (9.1),

$$|h(u) - h(v)| \leq \sum_{i=1}^{d} K_i |u_i - v_i| \quad \text{for all } u, v \in \mathbf{Z}^d \tag{9.58}$$

for any height profile $h \in H$. Consequently if we write $n = \ell m + j$ for $\ell = \ell(n)$ and $0 \leq j < m$,

$$\begin{aligned} &\left| n^{-1}\xi_{nt}([nq]) - g_0(q,t) \right| \\ &\quad \leq \left| n^{-1}\xi_{nt}([nq]) - n^{-1}\xi_{nt}(\ell mq) \right| + \left| n^{-1}\xi_{nt}(\ell mq) - n^{-1}\xi_{\ell mt}(\ell mq) \right| \\ &\qquad\quad + \left| n^{-1}\xi_{\ell mt}(\ell mq) - g_0(q,t) \right| \\ &\quad \leq n^{-1} \sum_{i=1}^{d} \big(K_i |mq_i| + 1 \big) + n^{-1} N_{\ell mq}(\ell mt, nt] + \left| n^{-1}\xi_{\ell mt}(\ell mq) - g_0(q,t) \right| . \end{aligned}$$

Letting $n \to \infty$ gives (9.57).

The function g_0 thus far defined on $\mathbf{Q}^d \times [0,\infty)$ is Lipschitz continuous also in the space variable. This follows directly from (9.58). For $p, q \in \mathbf{Q}^d$,

$$\begin{aligned} |g_0(q,t) - g_0(p,t)| &= \lim_{n\to\infty} \left| n^{-1}\xi_{nt}([np]) - n^{-1}\xi_{nt}([nq]) \right| \\ &\le \lim_{n\to\infty} n^{-1} \sum_{i=1}^{d} K_i \left| \, [np_i] - [nq_i] \, \right| \\ &= \sum_{i=1}^{d} K_i |p_i - q_i|. \end{aligned} \tag{9.59}$$

Lipschitz continuity on $\mathbf{Q}^d \times [0,\infty)$ allows us to extend g_0 uniquely to a Lipschitz function on $\mathbf{R}^d \times [0,\infty)$. This extension satisfies the homogeneity

$$g_0(rx, rt) = r g_0(x,t) \qquad \text{for all } x \in \mathbf{R}^d,\ t > 0, \text{ and } r > 0. \tag{9.60}$$

Define $g(x) = g_0(x,1)$. Then the limit in (9.57) has the scaling form $g_0(x,t) = tg(x/t)$.

The limit in Theorem 9.3 was claimed to hold for all $(x,t) \in \mathbf{R}^d \times [0,\infty)$ outside a single exceptional event of probability zero. To achieve this, let Ω_0 be the event on which limit (9.57) holds for all $q \in \mathbf{Q}^d$ and rational $t \ge 0$. Ω_0^c has probability zero. To show that on the event Ω_0 the limit (9.13) holds for all (x,t), approximate x with a rational point $q \in \mathbf{Q}^d$ and approximate t with a positive rational τ. Use properties (9.53) and (9.58) of the random height profiles, and the Lipschitz properties (9.54) and (9.59) of the limiting function. We leave the details of this last step to the reader.

We have verified the limit (9.13) in Theorem 9.3. It remains to check the properties of g. Lipschitz continuity has already been developed along the way. To see the concavity of g, return to the restarted processes defined by (9.49). The subadditivity derived there gives

$$\sigma_{ns}^{(n)}(nv) - \sigma_0^{(n)}(0) + \xi_{nt}(nz) \le \xi_{nt+ns}(nz + nv).$$

The increment $\sigma_{ns}^{(n)}(nv) - \sigma_0^{(n)}(0)$ is equal in distribution to $\xi_{ns}(nv)$. Thus dividing by n and letting $n \to \infty$ gives

$$g_0(v,s) + g_0(z,t) \le g_0(v+z, s+t) \qquad \text{for } v, z \in \mathbf{Z}^d \text{ and } s, t > 0.$$

This subadditivity of g_0 extends first to $\mathbf{Q}^d \times [0,\infty)$ by (9.55), and then to $\mathbf{R}^d \times [0,\infty)$ via the extension by Lipschitz continuity. By subadditivity and homogeneity,

$$\begin{aligned} \alpha g_0(x,1) + (1-\alpha) g_0(y,1) &= g_0(\alpha x, \alpha) + g_0((1-\alpha)y, (1-\alpha)) \\ &\le g_0(\alpha x + (1-\alpha)y, 1). \end{aligned}$$

for $x, y \in \mathbf{R}^d$ and $\alpha \in (0,1)$. Thus we have concavity for g.

To obtain (9.14) we shall show that, if $|x|_1 > 1$, then almost surely

$$\xi_n([nx]) = \xi_0([nx]) \qquad \text{for all large enough } n. \tag{9.61}$$

This implies (9.14) because then by the definition of the initial height function,

$$g(x) = \lim_{n\to\infty} n^{-1}\xi_0([nx]) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{d} K_i([nx_i]\wedge 0) = \sum_{i=1}^{d} K_i(x_i \wedge 0).$$

Without much additional work we prove a stronger result.

Lemma 9.13 *For any $t > 0$ and $\varepsilon > 0$, there exists a constant $C = C(t,\varepsilon) > 0$ such that*

$$P\left\{\xi_{nt}(u) < \xi_0(u) \text{ for some } u \text{ such that } |u|_1 \geq n(t+\varepsilon)\right\} \leq e^{-Cn}$$

for all n.

This lemma implies (9.61) because for $\varepsilon = (|x|_1 - 1)/3 > 0$,

$$\begin{aligned} |\,[nx]\,|_1 &= |\,[nx_1]\,| + \cdots + |\,[nx_d]\,| \geq n|x_1| + \cdots + n|x_d| - d \\ &= n(|x|_1 - d/n) \geq n(|x|_1 - \varepsilon) > n(1+\varepsilon) \end{aligned}$$

for large enough n.

Proof. For $u \in \mathbf{Z}^d$, let $T_u = \inf\{t > 0 : \xi_t(u) < \xi_0(u)\}$ be the time of first jump for height variable $\xi(u)$. Let S_m^1 denote a sum of m i.i.d. exponential mean one random variables. We show that T_u is a stochastically larger than S_m^1 with $m = |u|_1$.

Let $0 = z(0), z(1), z(2), \ldots, z(m) = u$ be a non-intersecting nearest-neighbor path from the origin to u that moves in each coordinate direction in turn, with this property: if $u_i \geq 0$ then the path is nondecreasing in direction e_i, while if $u_i \leq 0$ then the path is nonincreasing in direction e_i. If $u_i = 0$ then $z_i(k) = 0$ for all $0 \leq k \leq m$. Due to the restrictions (9.1) on admissible profiles, $\xi(z(k+1))$ cannot jump before $\xi(z(k))$ has jumped at least once. Define inductively the following random times.

Let $R_{z(0)}$ be the time of the first ring in $\mathcal{T}_{z(0)}$. Then $R_{z(0)} = T_{z(0)}$ is the time of first jump for $\xi(0)$.

Let $R_{z(1)}$ be the time of the first ring in $\mathcal{T}_{z(1)}$ after time $R_{z(0)}$. $R_{z(1)} - R_{z(0)}$ is an exponential variable with mean one, independent of $R_{z(0)}$ by the strong Markov property of the Poisson processes. $R_{z(1)} \leq T_{z(1)}$ because $\xi(z(1))$ can jump only after $\xi(z(0))$ has made its first jump.

Inductively, let $R_{z(k)}$ be the time of the first ring in $\mathcal{T}_{z(k)}$ after time $R_{z(k-1)}$. $R_{z(k)} - R_{z(k-1)}$ is an exponential variable with mean one, independent of $R_{z(0)}, \ldots, R_{z(k-1)}$. $R_{z(k)} \leq T_{z(k)}$

because by induction $R_{z(k-1)} \le T_{z(k-1)}$, and $\xi(z(k))$ can jump only after $T_{z(k-1)}$ when $\xi(z(k-1))$ took its first jump.

So we find that

$$T_{z(m)} \ge R_{z(m)} = \sum_{k=1}^{m} (R_{z(k)} - R_{z(k-1)}) + R_{z(0)}.$$

The sum has i.i.d. mean one exponential terms. We can now bound

$$\begin{aligned} &P\{\xi_{nt}(u) < \xi_0(u) \text{ for some } u \text{ such that } |u|_1 \ge n(t+\varepsilon)\} \\ &\quad \le \sum_{u:|u|_1 \ge n(t+\varepsilon)} P\{T_u \le nt\} \le \sum_{m \ge n(t+\varepsilon)} \sum_{u:|u|_1=m} P\{T_u \le nt\} \\ &\quad \le \sum_{m \ge n(t+\varepsilon)} C_0 m^{d-1} P\left\{S_m^1 \le nt\right\} \le \sum_{m \ge n(t+\varepsilon)} C_0 m^{d-1} \exp\bigl\{-mI\bigl(\tfrac{nt}{m}\bigr)\bigr\}. \end{aligned}$$

Above we first picked a constant C_0 so that for all $m \ge 1$, $C_0 m^{d-1}$ is an upper bound on the number of sites $u \in \mathbf{Z}^d$ such that $|u|_1 = m$. $I(x) = x - 1 - \log x$ is the large deviations rate function for mean one exponential random variables (Exercise A.9), and we used Proposition A.18. $I'(x) = 1 - x^{-1} < 0$ for $0 < x < 1$, so for $m \ge n(t+\varepsilon)$,

$$I\bigl(\tfrac{nt}{m}\bigr) \ge I\bigl(\tfrac{t}{t+\varepsilon}\bigr) > 0.$$

The final upper bound is

$$\sum_{m \ge n(t+\varepsilon)} C_0 m^{d-1} \exp\bigl\{-mI\bigl(\tfrac{t}{t+\varepsilon}\bigr)\bigr\} \le \exp(-Cn)$$

where the last equality is true for all n, if $C > 0$ is small enough. ∎

We have verified the properties of g and thereby proved Theorem 9.3.

An alternative approach. A proof of Theorem 9.3 could have been based on (9.46) and the subadditive ergodic theorem, followed by reasoning similar to that used above. This would give a $(0, \infty)$-valued, homogeneous, concave function γ, defined on the open set

$$\mathcal{L} = \left\{ (x, r) \in \mathbf{R}^d \times (-\infty, 0) : r < \sum_{i=1}^{d} K_i (x_i \wedge 0) \right\}$$

such that the limit

$$\lim_{n \to \infty} \frac{1}{n} T([nx], [nr]) = \gamma(x, r) \tag{9.62}$$

holds for all $(x,r) \in \mathcal{L}$, outside an exceptional event of probability zero. The finiteness of the limit $\gamma(x,r)$ follows from Lemma 9.12. The limiting height function g in Theorem 9.3 represents the level curve of γ, defined by

$$g(x) = \sup\{r : (x,r) \in \mathcal{L}, \gamma(x,r) \geq 1\}. \tag{9.63}$$

This connection follows because

$$\xi_{nt}([nx]) \leq [nr] \quad \text{iff} \quad T([nx],[nr]) \leq nt.$$

We let the reader work out the details of the connection of the limit (9.62) and the properties of γ with Theorem 9.3.

9.5.2 Proof of Theorem 9.4

The setting is such that the sequence of height processes h_t^n and the Poisson clocks $\{\mathcal{T}_u\}$ are defined on a common probability space. The height processes ξ_t^v are defined as functions of the Poisson processes, as described by (9.9) and (9.10).

Without technicalities, we can summarize the proof of Theorem 9.4 in a few lines. Write the variational equation (9.12) for the nth process in the form

$$n^{-1}h_{nt}^n([nx]) = \sup_{y\in\mathbf{R}^d} \left\{ n^{-1}h_0^n([ny]) + n^{-1}\xi_{nt}^{[ny]}([nx]-[ny]) \right\}. \tag{9.64}$$

The terms inside the braces on the right converge, the first one by hypothesis (9.15) to $\psi_0(y)$, and the second one by (9.13) to $tg((x-y)/t)$. Assuming we can pass the limit through the supremum over y, we conclude that

$$n^{-1}h_{nt}^n([nx]) \to \sup_y \{\psi_0(y) + tg((x-y)/t)\}.$$

Now the details. The paragraph following (9.17) claimed that the supremum is attained in a certain compact set. We begin by checking this.

Lemma 9.14 *Let g be the shape function defined by the limit* (9.13) *in Theorem 9.3. Suppose ψ_0 satisfies the Lipschitz bounds* (9.16). *Let $(x,t) \in \mathbf{R}^d \times (0,\infty)$. Then in the Hopf-Lax formula*

$$\psi(x,t) = \sup_{y\in\mathbf{R}^d} \left\{ \psi_0(y) + tg\left(\frac{x-y}{t}\right) \right\} \tag{9.65}$$

the supremum is attained at some y such that $|x-y|_1 \leq t$.

Proof. The proof is based on the property

$$g(x) = \sum_{i=1}^{d} K_i(x_i \wedge 0) \quad \text{for all } x \in \mathbf{R}^d \text{ such that } |x|_1 > 1,$$

derived as part of Theorem 9.3. Let

$$A = \{y \in \mathbf{R}^d : |x - y|_1 \le t\}.$$

It suffices to show that for any $\widetilde{y} \notin A$ there exists $\bar{y} \in A$ such that

$$\psi_0(\bar{y}) + tg\left(\frac{x - \bar{y}}{t}\right) \ge \psi_0(\widetilde{y}) + tg\left(\frac{x - \widetilde{y}}{t}\right). \tag{9.66}$$

For then the supremum over the entire space is the same as the supremum over A, and by continuity and compactness, this supremum is attained at some point.

Given $\widetilde{y} \notin A$, let $I = \{1 \le i \le d : \widetilde{y}_i > x_i\}$. Then

$$tg\left(\frac{x - \widetilde{y}}{t}\right) = \sum_{i \in I} K_i(x_i - \widetilde{y}_i).$$

Let $\beta = t \cdot |x - \widetilde{y}|_1^{-1}$ and set $\bar{y} = x + \beta(\widetilde{y} - x)$. Then $|x - \bar{y}|_1 = t$ and in particular $\bar{y} \in A$. Since $0 < \beta < 1$ and $x_i - \bar{y}_i = \beta(x_i - \widetilde{y}_i)$,

$$tg\left(\frac{x - \bar{y}}{t}\right) = \beta \sum_{i \in I} K_i(x_i - \widetilde{y}_i) = tg\left(\frac{x - \widetilde{y}}{t}\right) + (1 - \beta) \sum_{i \in I} K_i(\widetilde{y}_i - x_i).$$

On the other hand, by Lipschitz bounds (9.16),

$$\psi_0(\bar{y}) \ge \psi_0(\widetilde{y}) - \sum_{i=1}^{d} K_i(\widetilde{y}_i - \bar{y}_i)^+ = \psi_0(\widetilde{y}) - (1 - \beta) \sum_{i \in I} K_i(\widetilde{y}_i - x_i).$$

Combining these gives (9.66). ∎

Fix (x, t). One half of Theorem 9.4 is immediate. Pick a point y at which the supremum in (9.17) is attained. Then by (9.64),

$$\begin{aligned} P&\left\{n^{-1} h^n_{nt}([nx]) \le \psi(x, t) - \varepsilon\right\} \\ &\le P\left\{n^{-1} h^n_0([ny]) \le \psi_0(y) - \varepsilon/2\right\} \\ &\quad + P\left\{n^{-1} \xi^{[ny]}_{nt}([nx] - [ny]) \le tg((x - y)/t) - \varepsilon/2\right\}. \end{aligned} \tag{9.67}$$

The first term after the inequality vanishes as $n \to \infty$ by hypothesis (9.15). The last term equals

$$P\left\{n^{-1}\xi_{nt}([nx]-[ny]) \le tg((x-y)/t) - \varepsilon/2\right\}$$

because the random variable $\xi_{nt}^{[ny]}([nx]-[ny])$ has the same distribution as the random variable $\xi_{nt}([nx]-[ny])$ without the superscript. This probability vanishes as $n \to \infty$ by (9.13). The difference between $n^{-1}\xi_{nt}([nx]-[ny])$ and $n^{-1}\xi_{nt}([n(x-y)])$ is controlled by (9.58) and vanishes as $n \to \infty$.

We have shown

$$\lim_{n\to\infty} P\left\{n^{-1}h_{nt}^n([nx]) \le \psi(x,t) - \varepsilon\right\} = 0. \tag{9.68}$$

To prove the other direction, we restrict the potential maximizers that need to be considered in the variational formula. Let the boundary ∂B of a rectangle $B \subseteq \mathbf{Z}^d$ be the set of those sites $v \in B$ that have a neighbor outside B.

Lemma 9.15 *Let $u \in \mathbf{Z}^d$ lie in a rectangle $B \subseteq \mathbf{Z}^d$. Suppose that $\xi_t^v(u-v) = \xi_0^v(u-v)$ for $v \in \partial B$. Then*

$$h_t(u) = \max_{v\in B}\{h_0(v) + \xi_t^v(u-v)\}.$$

Proof. For $v \notin B$, let v' be the projection of v to the boundary of B. In other words, if

$$B = \{a_1,\ldots,b_1\} \times \{a_2,\ldots,b_2\} \times \cdots \times \{a_d,\ldots,b_d\}$$

for two points $a = (a_1,\ldots,a_d), b = (b_1,\ldots,b_d) \in \mathbf{Z}^d$, then $v' = (v_1',\ldots,v_d')$ is defined by

$$v_i' = \begin{cases} a_i & \text{if } v_i < a_i \\ v_i & \text{if } a_i \le v_i \le b_i \\ b_i & \text{if } v_i > b_i. \end{cases}$$

We show that v' dominates v in the variational formula (9.12).

$$\begin{aligned} h_0(v) + \xi_t^v(u-v) &\le h_0(v) + \xi_0^v(u-v) \\ &\le h_0(v') + \sum_{i=1}^d K_i(v_i - v_i')^+ + \sum_{i=1}^d K_i\left((u_i - v_i) \wedge 0\right) \\ &= h_0(v') + \sum_{i=1}^d K_i\left((u_i - v_i') \wedge 0\right) \\ &= h_0(v') + \xi_0^{v'}(u-v') = h_0(v') + \xi_t^{v'}(u-v'). \end{aligned}$$

Height profiles decrease with time. This gives the first inequality above. Then substitute in the definition of the initial profile ξ_0^v. The second inequality comes from the state space restrictions: for any height profile $h \in H$,

$$h(v) \le h(v \vee v') \le h(v') + \sum_{i=1}^{d} K_i(v_i \vee v'_i - v'_i) = h(v') + \sum_{i=1}^{d} K_i(v_i - v'_i)^+.$$

Since v'_i is between u_i and v_i,

$$(v_i - v'_i)^+ \; + \; ((u_i - v_i) \wedge 0) = ((u_i - v'_i) \wedge 0)\,.$$

Lastly we use the assumption that $\xi^{v'}(u - v')$ has not jumped by time t. This shows that all $v \notin B$ can be ignored in (9.12). ∎

Define the cube

$$A = \prod_{i=1}^{d} [x_i - t - 1, x_i + t + 1] \;\subseteq \mathbf{R}^d$$

and set

$$Y_n = \max_{v \in nA} \{h_0^n(v) + \xi_{nt}^v([nx] - v)\}\,. \tag{9.69}$$

Lemma 9.16 *There exist finite constants $C_0, C_1 > 0$ such that for all n,*

$$P\{h_{nt}^n([nx]) \ne Y_n\} \le C_0 e^{-C_1 n}.$$

Proof. By the previous lemma, it suffices to show that for large enough n there exists a cube $B_n \subseteq \mathbf{Z}^d$ centered at $[nx]$ such that $B_n \subseteq nA$ and

$$P\{\xi_{nt}^v([nx] - v) < \xi_0^v([nx] - v) \text{ for some } v \in \partial B\} \le C_0 e^{-C_1 n}. \tag{9.70}$$

When n is large enough, we can pick $B_n \subseteq nA$ so that

$$\prod_{i=1}^{d} \big[n(x_i - t - 1/2), n(x_i + t + 1/2)\big] \;\subseteq B_n.$$

Then for $v \in \partial B_n$ and large enough n,

$$|\,[nx] - v\,|_1 = \sum_{i=1}^{d} |\,[nx_i] - v_i\,| \ge dn(t + 1/4).$$

The number of sites in ∂B_n is bounded by $C_2 n^{d-1}$ for some constant C_2. Each process ξ_t^v is distributed as process ξ_t, so we can apply Lemma 9.13 to each process ξ_t^v. The probability in (9.70) is then bounded by $C_2 n^{d-1} e^{-Cn}$. This bound is at most $C_0 e^{-C_1 n}$ for a sufficiently large $C_0 < \infty$ and sufficiently small $C_1 > 0$. ∎

Given the $\varepsilon > 0$ in the statement (9.18) we are working to prove, pick $\delta > 0$ so that

$$\delta \sum_{i=1}^{d} K_i < \frac{\varepsilon}{2}. \tag{9.71}$$

Partition the cube A into a finite collection of subcubes $\{D^k : 1 \le k \le M\}$ with sidelength at most δ. Let y^k be the lower left corner and $\widetilde{y}^k$ the upper right corner of D^k. In other words, $D^k = \prod_{i=1}^{d} [y_i^k, \widetilde{y}_i^k]$ with $y_i^k < \widetilde{y}_i^k \le y_i^k + \delta$ for $1 \le i \le d$ and $1 \le k \le M$.

The variation inside a subcube will be controlled by this monotonicity lemma.

Lemma 9.17 *For any $u, v, w \in \mathbf{Z}^d$ such that $v \le w$ (coordinatewise order),*

$$\xi_t^w(u - w) \le \xi_t^v(u - v).$$

Proof. Since $u_i - w_i \le u_i - v_i$ for each i, at time zero

$$\xi_0^w(u - w) = \sum_{i=1}^{d} K_i \left((u_i - w_i) \wedge 0\right) \le \sum_{i=1}^{d} K_i \left((u_i - v_i) \wedge 0\right) = \xi_0^v(u - v).$$

Use attractivity to compare the processes $h_t'(u) = \xi_t^w(u - w)$ and $h_t''(u) = \xi_t^v(u - v)$. ∎

Starting with definition (9.69) of Y_n, use spatial monotonicity of h_0^n and Lemma 9.17 to write

$$\begin{aligned} Y_n &= \max_{1 \le k \le M} \max_{v \in nD^k} \{h_0^n(v) + \xi_{nt}^v([nx] - v)\} \\ &\le \max_{1 \le k \le M} \left\{ h_0^n([n\widetilde{y}^k]) + \xi_{nt}^{[ny^k]}([nx] - [ny^k]) \right\}. \end{aligned}$$

We can finish the proof of Theorem 9.4. First observe from the Hopf-Lax formula (9.17), the Lipschitz property (9.16), and the choice (9.71) of δ,

$$\psi_0(\widetilde{y}^k) + tg((x - y^k)/t) \le \psi_0(y^k) + tg((x - y^k)/t) + \delta \sum_{i=1}^{d} K_i \le \psi(x, t) + \frac{\varepsilon}{2}.$$

By Lemma 9.16,

$$
\begin{aligned}
&P\left\{n^{-1}h^n_{nt}([nx]) \geq \psi(x,t)+\varepsilon\right\} \\
&\quad \leq P\left\{h^n_{nt}([nx]) \neq Y_n\right\} + P\left\{n^{-1}Y_n \geq \psi(x,t)+\varepsilon\right\} \\
&\quad \leq C_0 e^{-C_1 n} + \sum_{1\leq k\leq M} P\Bigg\{ n^{-1}h^n_0([n\widetilde{y}^k]) + n^{-1}\xi^{[ny^k]}_{nt}([nx]-[ny^k]) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \geq \psi_0(\widetilde{y}^k) + tg\left(\frac{x-y^k}{t}\right) + \frac{\varepsilon}{2}\Bigg\} \\
&\quad \leq C_0 e^{-C_1 n} + \sum_{1\leq k\leq M} P\left\{n^{-1}h^n_0([n\widetilde{y}^k]) \geq \psi_0(\widetilde{y}^k)+\varepsilon/4\right\} \\
&\qquad\qquad + \sum_{1\leq k\leq M} P\left\{ n^{-1}\xi^{[ny^k]}_{nt}([nx]-[ny^k]) \geq tg\left(\frac{x-y^k}{t}\right) + \frac{\varepsilon}{4}\right\}.
\end{aligned}
$$

In the final bound, all terms vanish as $n\to\infty$ by (9.15) and (9.13), and the number of terms is fixed. We have shown

$$
\lim_{n\to\infty} P\left\{n^{-1}h^n_{nt}([nx]) \geq \psi(x,t)+\varepsilon\right\} = 0. \tag{9.72}
$$

Together with (9.68) this concludes the proof of Theorem 9.4.

9.5.3 Proof of Theorem 9.5

The Lipschitz continuity of the function ψ on $\mathbf{R}^d\times[0,\infty)$ can be seen either from the Hopf-Lax formula (9.17), or from the properties of the interface process. To take the latter route, bounds (9.58) on the spatial variation, and the temporal bounds

$$
h_s(u) \geq h_t(u) \geq h_s(u) - N_u(s,t\,] \qquad \text{for } s<t \text{ and } u\in\mathbf{Z}^d \tag{9.73}
$$

imply

$$
|\psi(x,t)-\psi(y,s)| \leq \sum_{i=1}^d K_i|x_i-y_i| + |t-s|. \tag{9.74}
$$

By Rademacher's Theorem (Section 3.1 in [15]), ψ is differentiable Lebesgue almost everywhere on $\mathbf{R}^d\times[0,\infty)$. Then by Lemma A.32 the partial differential equation (9.20) is satisfied at points of differentiability.

To derive properties of the velocity f, we start with an upper bound on the limiting shape $g(x)$ from Lemma 9.12.

Lemma 9.18 *Let $s > 1$ be defined by $I(s) = \log(2d+1)$. Then $g(0) \le -1/(2s)$.*

Proof. Let $s_1 > s$ and $r = 1/(2s_1)$. Then $n = 2rns_1 \ge 2[rn]s_1$. Take $(u, h) = (0, -[nr])$ in Lemma 9.12 so that $M = 2[nr]$. Then

$$P\left\{\xi_n(0) > -[nr]\right\} \le P\left\{T(0, -[nr]) > n\right\} \le \exp\left\{-2[nr]\left(I(s_1) - \log(2d+1)\right)\right\}.$$

By Borel-Cantelli $n^{-1}\xi_n(0) \le -n^{-1}[nr]$ for all large enough n, almost surely. Consequently $g(0) \le -r$. Let $r \nearrow 1/(2s)$. ∎

We prove some properties of the concave conjugate of g defined by

$$f(\rho) = \inf_{x \in \mathbf{R}^d}\{\rho \cdot x - g(x)\} \quad \text{for} \quad \rho = (\rho_1, \ldots, \rho_d) \in \mathbf{R}^d. \tag{9.75}$$

Recall that $V = \prod_{i=1}^d [0, K_i]$ is the set of admissible gradients of macroscopic, deterministic height functions.

Proposition 9.19 *Outside V, f is identically $-\infty$. On V, f is continuous and concave. $f = 0$ on the boundary of V, and $0 < f \le 1$ in the interior of V.*

Proof. Let $\rho \notin V$. For some i, either $\rho_i < 0$ or $\rho_i > K_i$. In the former case take $x = \alpha e_i$ in (9.75), in the latter take $x = -\alpha e_i$, and let $\alpha \nearrow \infty$. This shows $f(\rho) = -\infty$.

Since $g(x)$ is bounded above by the initial interface,

$$g(x) \le \sum_{i=1}^d K_i(x_i \wedge 0) \le \rho \cdot x$$

for all x and any $\rho \in V$. This gives $f(\rho) \ge 0$ for $\rho \in V$. To get $f = 0$ on the boundary of V, take $x = e_i$ in (9.75) if $\rho_i = 0$, and $x = -e_i$ if $\rho_i = K_i$. In both cases $f(\rho) \le 0$ results.

Concavity of f follows from its definition as an infimum of affine functions.

We argue the continuity of f on V. Suppose $\rho^j \to \rho$ in V. The infimum in (9.75) is achieved by Lemma 9.14, so we may pick x so that $f(\rho) = \rho \cdot x - g(x)$. Then

$$\limsup_{j\to\infty} f(\rho^j) \le \limsup_{j\to\infty}\{\rho^j \cdot x - g(x)\} = \rho \cdot x - g(x) = f(\rho).$$

Suppose some sequence $\rho^j \to \rho$ satisfies $f(\rho^j) \to f(\rho) - \delta$ for some $\delta > 0$. Then ρ cannot be on the boundary of V, and neither can ρ^j for large enough j. Find λ^j on the boundary of V such that

$$\rho^j = \alpha^j \rho + (1 - \alpha^j)\lambda^j$$

for some $\alpha^j \in (0,1)$. Then $\rho^j - \lambda^j = \alpha^j(\rho - \lambda^j)$, and by the triangle inequality (for any norm $|\cdot|$ on $\mathbf{R}^d$)

$$|\rho - \lambda^j| \le |\rho - \rho^j| + |\rho^j - \lambda^j| = |\rho - \rho^j| + \alpha^j|\rho - \lambda^j|.$$

As $j \to \infty$, $|\rho - \rho^j| \to 0$ while $|\rho - \lambda^j|$ is bounded below by the positive distance of ρ to the boundary of V. Then necessarily $\alpha^j \to 1$. By the concavity of f,

$$f(\rho^j) \ge \alpha^j f(\rho) + (1 - \alpha^j) f(\lambda^j) = \alpha^j f(\rho).$$

From this and $\alpha^j \to 1$ follows $\liminf_{j\to\infty} f(\rho^j) \ge f(\rho)$. We have shown that f is continuous on V.

The height variable $\xi(0)$ advances downward at rate at most 1, hence $g(0) \ge -1$. This gives $f(\rho) \le -g(0) \le 1$.

Finally we show $f > 0$ on the interior of V. It suffices to show the existence of one point ρ such that $f(\rho) > 0$. Such a point ρ must lie in the interior of V because elsewhere $f \le 0$. If ρ' is any other interior point we can find a point ρ'' on the boundary of V so that $\rho' = \alpha\rho + (1-\alpha)\rho''$ for some $\alpha \in (0,1)$. By concavity

$$f(\rho') \ge \alpha f(\rho) + (1-\alpha) f(\rho'') = \alpha f(\rho) > 0.$$

For the final argument we use a separation theorem from functional analysis. By the concavity and continuity of g,

$$G = \{(x, r) \in \mathbf{R}^d \times \mathbf{R} : r < g(x)\}$$

is a convex, open set. The point $(0, g(0))$ lies outside G, so there exists a separating linear functional on $\mathbf{R}^{d+1}$. In other words, there exists a pair $(y, c) \in \mathbf{R}^d \times \mathbf{R}$ such that

$$y \cdot 0 + cg(0) > y \cdot x + cr$$

for all $(x, r) \in G$. (See for example Theorem 3.4 in [32].) Taking $x = 0$ and $r < g(0)$ shows that $c > 0$. Letting $r \nearrow g(x)$ gives

$$cg(0) \ge y \cdot x + cg(x)$$

for all x. Taking $x = \alpha e_i$ for $\alpha > 1$ gives $0 > cg(0) \ge \alpha y_i$ which shows $y_i < 0$. Set $\rho = -c^{-1}y$. Then for all x,

$$\rho \cdot x - g(x) \ge -g(0),$$

which gives $f(\rho) \ge -g(0)$. So in fact for this ρ we have $f(\rho) = -g(0)$ which is strictly positive by Lemma 9.18.

We have checked all the properties. ∎

Instead of appealing to a separation theorem in the last proof, we could have appealed to some convex analysis. By Theorem 23.4 in [31], g possesses at least one subgradient ρ at $x = 0$. This means that for all x,

$$g(x) \leq g(0) + \rho \cdot x,$$

which is exactly what we found above.

Next we prove the symmetry property of the velocity. This completes the proof of Theorem 9.5.

Proposition 9.20 *Let I be a subset of $\{1, \ldots, d\}$, and assume $\rho, \widetilde{\rho} \in V$ satisfy*

$$\widetilde{\rho}_i = \rho_i \text{ for } i \notin I \text{ and } \widetilde{\rho}_i = K_i - \rho_i \text{ for } i \in I.$$

Then $f(\rho) = f(\widetilde{\rho})$.

Proof. Define a bijection S on $\mathbf{Z}^d$ by

$$(Su)_i = \begin{cases} -u_i, & i \in I \\ u_i, & i \notin I. \end{cases}$$

Define a random initial height function h_0 as in Example 9.7 for slope ρ. Define $\widetilde{h}_0$ by

$$\widetilde{h}_0(u) = h_0(Su) + \sum_{j \in I} K_j u_j.$$

Check that $\widetilde{h}_0 \in H$. Let the process $\widetilde{h}_t$ obey Poisson clocks $\{\widetilde{\mathcal{T}}_u\}$ where $\widetilde{\mathcal{T}}_u = \mathcal{T}_{Su}$.

We claim that

$$\widetilde{h}_t(u) = h_t(Su) + \sum_{j \in I} K_j u_j \tag{9.76}$$

for all $t \geq 0$. We have arranged for it to hold at time $t = 0$. By the percolation argument mentioned in Section 9.1 it suffices to observe that height variables $\widetilde{h}_t(u)$ and $h_t(Su)$ always jump together. Their jump attempts are synchronized and happen at the jump times of $\widetilde{\mathcal{T}}_u = \mathcal{T}_{Su}$. The claim is then proved by checking that if (9.76) holds, then $h(u)$ is allowed to jump iff $\widetilde{h}(Su)$ is allowed to jump. In other words, that

$$\widetilde{h}(u) \geq \widetilde{h}(u - e_i) + 1 \text{ and } \widetilde{h}(u) \geq \widetilde{h}(u + e_i) - K_i + 1 \text{ for } 1 \leq i \leq d$$

iff

$$h(Su) \geq h(Su - e_i) + 1 \text{ and } h(Su) \geq h(Su + e_i) - K_i + 1 \text{ for } 1 \leq i \leq d.$$

We leave these details for the reader.

The initial heights satisfy

$$n^{-1}h_0([nx]) \to \rho \cdot x \qquad \text{and} \qquad n^{-1}\widetilde{h}_0([nx]) \to \widetilde{\rho} \cdot x.$$

For $\psi_0(x) = \rho \cdot x$, the Hopf-Lax formula gives

$$\psi(0,1) = \sup_y \{\rho \cdot y + g(-y)\} = -\inf_x \{\rho \cdot x - g(x)\} = -f(\rho). \tag{9.77}$$

Similarly for $\widetilde{\rho}$. Divide by t in (9.76) and let $t \to \infty$. Theorem 9.4 gives the limits

$$-f(\rho) = \lim_{t\to\infty} t^{-1}h_t(0) = \lim_{t\to\infty} t^{-1}\widetilde{h}_t(0) = -f(\widetilde{\rho}).$$

∎

9.5.4 Proof of Theorem 9.6

Recall the definition (9.22) of the event B. Let $\mathcal{F}_t$ be the filtration generated by the initial profile h_0 and the Poisson processes up to time t.

Lemma 9.21 *Assume $h_0(0)$ has finite mean. Then the process*

$$M_t = h_t(0) + \int_0^t \mathbf{1}\{h_s \in B\}\, ds$$

is a martingale with respect to the filtration $\mathcal{F}_t$.

Proof. Integrability of M_t is guaranteed because $|h_t(0) - h_0(0)|$ is stochastically dominated by a mean t Poisson random variable.

As before, $N_u(s,t]$ denotes the number of jump times in Poisson process $\mathcal{T}_u$ during time interval $(s,t]$. Consider first a small time interval $(s, s+\delta]$. The height variable $h(0)$ jumps downward at some time $t \in (s, s+\delta]$ if $t \in \mathcal{T}_0$ and $h_{t-} \in B$. If there are no other jumps in $\mathcal{T}_0 \cup \mathcal{T}_{\pm e_i}$ during $(s, s+\delta]$, then $h_{t-} \in B$ implies $h_s \in B$. Define the event

$$A = \Big\{ \sum_{u\in\{0,\pm e_i\}} N_u(s, s+\delta] \geq 2 \Big\}.$$

Write first

$$\begin{aligned}
&h_{s+\delta}(0) - h_s(0) \\
&\quad = \big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}\{N_0(s,s+\delta] = 1,\ N_{\pm e_i}(s,s+\delta] = 0 \text{ for } 1 \leq i \leq d\} \\
&\qquad\quad + \big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}_A \\
&\quad = -\mathbf{1}\{h_s \in B\} \cdot \mathbf{1}\{N_0(s,s+\delta] = 1,\ N_{\pm e_i}(s,s+\delta] = 0 \text{ for } 1 \leq i \leq d\} \\
&\qquad\quad + \big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}_A
\end{aligned}$$

The total number of jumps in the rate $2d+1$ Poisson process $\mathcal{T}_0 \cup \mathcal{T}_{\pm e_i}$ during $(s, s+\delta]$ is an upper bound on $|h_{s+\delta}(0) - h_s(0)|$, and consequently

$$\begin{aligned} &\Big| E\big[\big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}_A \big| \mathcal{F}_s\big] \Big| \\ &\qquad \le \sum_{k=2}^{\infty} k P\Big\{ \sum_{u \in \{0, \pm e_i\}} N_u(s, s+\delta] = k \Big\} \\ &\qquad = \sum_{k=2}^{\infty} k \cdot \frac{e^{-(2d+1)\delta}(2d+1)^k \delta^k}{k!} \le (2d+1)^2 \delta^2. \end{aligned}$$

As

$$P\{N_0(s, s+\delta] = 1,\ N_{\pm e_i}(s, s+\delta] = 0 \text{ for } 1 \le i \le d \ |\mathcal{F}_s\} = \delta e^{-(2d+1)\delta},$$

the random variable

$$R_{s,s+\delta} = E\big[h_{s+\delta}(0) - h_s(0) \big| \mathcal{F}_s\big] + \delta e^{-(2d+1)\delta} \mathbf{1}\{h_s \in B\}$$

satisfies the inequality $|R_{s,s+\delta}| \le (2d+1)^2 \delta^2$.

Given $s < t$, let m be a positive integer, $\delta = \frac{t-s}{m}$, and $s_i = s + i\delta$ for $i = 0, \ldots, m$.

$$\begin{aligned} E\big[h_t(0) - h_s(0) \big| \mathcal{F}_s\big] &= E\Big[\sum_{i=0}^{m-1} E\big(h_{s_{i+1}}(0) - h_{s_i}(0) \big| \mathcal{F}_{s_i}\big) \Big| \mathcal{F}_s \Big] \\ &= E\Big[-e^{-(2d+1)\delta} \delta \sum_{i=0}^{m-1} \mathbf{1}\{h_{s_i} \in B\} \Big| \mathcal{F}_s \Big] + E\Big[\sum_{i=0}^{m-1} R_{s_i, s_{i+1}} \Big| \mathcal{F}_s \Big] \\ &= E\Big[-e^{-(2d+1)\delta} \int_s^t \sum_{i=0}^{m-1} \mathbf{1}\{h_{s_{i+1}} \in B\} \mathbf{1}_{(s_i, s_{i+1}]}(s)\, ds \Big| \mathcal{F}_s \Big] \\ &\qquad - E\big[e^{-(2d+1)\delta} \delta \big(\mathbf{1}\{h_{s_0} \in B\} - \mathbf{1}\{h_{s_m} \in B\}\big) \big| \mathcal{F}_s\big] + E\Big[\sum_{i=0}^{m-1} R_{s_i, s_{i+1}} \Big| \mathcal{F}_s \Big]. \end{aligned}$$

The last sum satisfies

$$\sum_{i=0}^{m-1} R_{s_i, s_{i+1}} \le m(2d+1)^2 \delta^2 = (2d+1)^2 (t-s)^2 m^{-1}.$$

Let $m \nearrow \infty$, so that simultaneously $\delta \searrow 0$. The error terms vanish in the limit. By the right-continuity of paths,

$$\int_s^t \sum_{i=0}^{m-1} \mathbf{1}\{h_{s_{i+1}} \in B\} \mathbf{1}_{(s_i, s_{i+1}]}(s)\, ds \to \int_s^t \mathbf{1}\{h_s \in B\}\, ds$$

almost surely. Thus we obtain

$$E\left[h_t(0) - h_s(0) + \int_s^t \mathbf{1}\{h_s \in B\}\, ds \,\middle|\, \mathcal{F}_s\right] = 0.$$

This is the same as the conclusion of the lemma. ∎

Lemma 9.22 *Assume $E[h_0(0)^2] < \infty$, and let M_t be the martingale of Lemma 9.21. Then the process*

$$L_t = M_t^2 - \int_0^t \mathbf{1}\{h_s \in B\}\, ds$$

is a martingale with respect to the filtration $\mathcal{F}_t$.

Proof. The square integrability of M_t is guaranteed by the hypothesis $E[h_0(0)^2] < \infty$. As in the previous proof, we begin by considering a small time increment. Abbreviate $\gamma(s) = \mathbf{1}\{h_s \in B\}$.

$$\begin{aligned}
E\big[M_{s+\delta}^2 - M_s^2 \big| \mathcal{F}_s\big] &= E\big[\big(M_{s+\delta} - M_s\big)^2 \big| \mathcal{F}_s\big] \\
&= E\left[\left(h_{s+\delta}(0) - h_s(0) - \int_s^{s+\delta} \gamma(r)\, dr\right)^2 \middle| \mathcal{F}_s\right] \\
&= E\big[\big(h_{s+\delta}(0) - h_s(0)\big)^2 \big| \mathcal{F}_s\big] - 2E\left[\big(h_{s+\delta}(0) - h_s(0)\big) \int_s^{s+\delta} \gamma(r)\, dr \middle| \mathcal{F}_s\right] \\
&\quad + E\left[\left(\int_s^{s+\delta} \gamma(r)\, dr\right)^2 \middle| \mathcal{F}_s\right].
\end{aligned}$$

Consider the last three terms. Since $0 \le \gamma(r) \le 1$, the last term is at most δ^2 in absolute value. Same is true for the second, since $h(0)$ is nonincreasing with time:

$$\begin{aligned}
0 &\le -2E\left[\big(h_{s+\delta}(0) - h_s(0)\big) \int_s^{s+\delta} \gamma(r)\, dr \middle| \mathcal{F}_s\right] \\
&\le -2\delta E\big[h_{s+\delta}(0) - h_s(0) \big| \mathcal{F}_s\big] = 2\delta E\left[\int_s^{s+\delta} \gamma(r)\, dr \middle| \mathcal{F}_s\right] \le 2\delta^2.
\end{aligned}$$

All the contribution comes from the first of the three.

$$\begin{aligned}
&E\big[\big(h_{s+\delta}(0) - h_s(0)\big)^2 \big| \mathcal{F}_s\big] \\
&\quad = E\big[\big(h_{s+\delta}(0) - h_s(0)\big)^2 \cdot \mathbf{1}\{N_0(s, s+\delta] \le 1\} \big| \mathcal{F}_s\big] \\
&\qquad + E\big[\big(h_{s+\delta}(0) - h_s(0)\big)^2 \cdot \mathbf{1}\{N_0(s, s+\delta] \ge 2\} \big| \mathcal{F}_s\big] \\
&\quad = -E\big[\big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}\{N_0(s, s+\delta] \le 1\} \big| \mathcal{F}_s\big] \qquad (9.78) \\
&\qquad + E\big[\big(h_{s+\delta}(0) - h_s(0)\big)^2 \cdot \mathbf{1}\{N_0(s, s+\delta] \ge 2\} \big| \mathcal{F}_s\big]. \qquad (9.79)
\end{aligned}$$

In the last equality we simply noted that $m^2 = -m$ if $m = 0$ or -1. Term (9.79) is an error term, bounded as follows:

$$E\big[\big(h_{s+\delta}(0) - h_s(0)\big)^2 \cdot \mathbf{1}\{N_0(s, s+\delta] \ge 2\}\big|\mathcal{F}_s\big] \le \sum_{k=2}^{\infty} k^2 \cdot \frac{e^{-\delta}\delta^k}{k!} \le 2\delta^2.$$

The main term (9.78) satisfies

$$\begin{aligned}
&- E\big[\big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}\{N_0(s, s+\delta] \le 1\}\big|\mathcal{F}_s\big] \\
&\quad = -E\big[h_{s+\delta}(0) - h_s(0)\big|\mathcal{F}_s\big] + E\big[\big(h_{s+\delta}(0) - h_s(0)\big) \cdot \mathbf{1}\{N_0(s, s+\delta] \ge 2\}\big|\mathcal{F}_s\big] \\
&\quad = E\left[\int_s^{s+\delta} \gamma(r)\, dr \middle| \mathcal{F}_s\right] + (\text{a term at most } \delta^2 \text{ in absolute value}).
\end{aligned}$$

Putting all these estimates together gives

$$\left| E\left[M_{s+\delta}^2 - M_s^2 - \int_s^{s+\delta} \gamma(r)\, dr \middle| \mathcal{F}_s \right] \right| \le 6\delta^2.$$

Given $s < t$, let again $t - s = m\delta$ and add up the error terms over the m subintervals of length δ, to get the estimate

$$\left| E\left[M_t^2 - M_s^2 - \int_s^{t} \gamma(r)\, dr \middle| \mathcal{F}_s \right] \right| \le 6m\delta^2 \le \frac{6(t-s)^2}{m}.$$

Letting $m \to \infty$ completes the proof. ∎

Lemma 9.23 *Under the assumptions of Lemmas 9.21 and 9.22, $t^{-1}M_t \to 0$ almost surely.*

Proof. Let $t_k = k^{3/2}$. First we use Borel-Cantelli to show convergence along the subsequence t_k. Let $\varepsilon > 0$. As before, $\gamma(s) = \mathbf{1}\{h_s \in B\}$.

$$\begin{aligned}
P\big(|M_{t_k}| \ge \varepsilon t_k\big) &\le \varepsilon^{-2} t_k^{-2} E\big[M_{t_k}^2\big] = \varepsilon^{-2} t_k^{-2} E\left[L_{t_k} + \int_0^{t_k} \gamma(s)\, ds\right] \\
&= \varepsilon^{-2} t_k^{-2} E L_0 + \varepsilon^{-2} t_k^{-2} \int_0^{t_k} E\gamma(s)\, ds \le C t_k^{-1},
\end{aligned}$$

for a constant C. Above we used the martingale property of L_t, and then

$$EL_0 = E[M_0^2] = E[h_0(0)^2] < \infty.$$

Since $\sum t_k^{-1} < \infty$ and this argument works for any $\varepsilon > 0$, Borel-Cantelli implies $t_k^{-1} M_{t_k} \to 0$.

It remains to fill in the t-values between the points t_k. Suppose $t_k \le t \le t_{k+1}$. By the monotonicity of $t \mapsto h_t(0)$ and the bounds $0 \le \gamma(s) \le 1$,

$$h_{t_{k+1}}(0) + \int_0^{t_{k+1}} \gamma(s)\,ds - (t_{k+1} - t_k) \le h_t(0) + \int_0^t \gamma(s)\,ds = M_t$$
$$\le h_{t_k}(0) + \int_0^{t_k} \gamma(s)\,ds + (t_{k+1} - t_k).$$

The choice $t_k = k^{3/2}$ has the properties $t_{k+1}/t_k \to 1$ and $(t_{k+1} - t_k)/t_k \to 0$. Hence dividing by t above and letting $t \to \infty$ gives the conclusion. ∎

We complete the proof of Theorem 9.6. For $\psi_0(x) = \rho \cdot x$, the Hopf-Lax formula gives $\psi(0, 1) = -f(\rho)$ as already seen in (9.77). By Lemma 9.23,

$$t^{-1}h_t(0) + t^{-1} \int_0^t \mathbf{1}\{h_s \in B\}\,ds \to 0 \qquad \text{in probability, as } t \to \infty.$$

By Theorem 9.4 and the assumptions made in Theorem 9.6,

$$t^{-1}h_t(0) \to \psi(0, 1) = -f(\rho)$$

in probability. Consequently

$$t^{-1} \int_0^t \mathbf{1}\{h_s \in B\}\,ds \to f(\rho)$$

in probability. The random variable $t^{-1} \int_0^t \mathbf{1}\{h_s \in B\}\,ds$ is bounded uniformly over t, so by dominated convergence we also have

$$t^{-1} \int_0^t P\{h_s \in B\}\,ds \to f(\rho).$$

This completes the proof of Theorem 9.6.

9.5.5 Proof of Theorem 9.8

Let $u = \psi$ be the function defined by the Hopf-Lax formula (9.17). Let $\bar{f}$ be an extension of f to a continuous function on all of $\mathbf{R}^d$. We apply Ishii's uniqueness result stated as Theorem A.33 in Section A.12 in the Appendix. It suffices to show that for any continuously differentiable ϕ on $\mathbf{R}^d \times (0, \infty)$, inequality (9.25) holds if $u - \phi$ has a local maximum at (x_0, t_0), and inequality (9.26) holds if $u - \phi$ has a local minimum at (x_0, t_0), with $F = \bar{f}$.

We begin by showing that such a ϕ has $\nabla\phi(x_0, t_0) \in V$, so $\bar{f}(\nabla\phi(x_0, t_0)) = f(\nabla\phi(x_0, t_0))$.

Lemma 9.24 *Suppose $u - \phi$ has either a local maximum or a local minimum at $(x_0, t_0) \in \mathbf{R}^d \times (0, \infty)$. Then*

$$-1 \le \phi_t(x_0, t_0) \le 0 \quad \textit{and} \quad \nabla\phi(x_0, t_0) \in V.$$

Proof. We go through the case of a local maximum and leave the other case to the reader. The assumption is that for some neighborhood B of (x_0, t_0) in $\mathbf{R}^d \times (0, \infty)$,

$$u(x_0, t_0) - \phi(x_0, t_0) \ge u(x, t) - \phi(x, t)$$

for $(x, t) \in B$. We use this in the form

$$u(x, t) - u(x_0, t_0) \le \phi(x, t) - \phi(x_0, t_0).$$

Thus for small enough $h > 0$, by the Lipschitz bounds on the increments of u,

$$0 \le u(x_0 + he_i, t_0) - u(x_0, t_0) \le \phi(x_0 + he_i, t_0) - \phi(x_0, t_0)$$

and

$$-K_i h \le u(x_0 - he_i, t_0) - u(x_0, t_0) \le \phi(x_0 - he_i, t_0) - \phi(x_0, t_0).$$

These inequalities imply $0 \le \phi_{x_i}(x_0, t_0) \le K_i$. For the time increment,

$$0 \le u(x_0, t_0 - he_i) - u(x_0, t_0) \le \phi(x_0, t_0 - he_i) - \phi(x_0, t_0)$$

and

$$-h \le u(x_0, t_0 + he_i) - u(x_0, t_0) \le \phi(x_0, t_0 + he_i) - \phi(x_0, t_0).$$

These inequalities imply $0 \ge \phi_t(x_0, t_0) \ge -1$. ∎

The statement $0 \ge \phi_t(x_0, t_0) \ge -1$ is not needed. It was included only to make the lemma complete. In the next two lemmas we check the defining inequalities of the viscosity solution for u.

Lemma 9.25 *Suppose $u - \phi$ has a local maximum at (x_0, t_0). Then*

$$\phi_t(x_0, t_0) + f\big(\nabla\phi(x_0, t_0)\big) \le 0.$$

Proof. For small $0 < h < t_0$, pick a point $z = z(h) \in \mathbf{R}^d$ such that $|z|_1 \le 1$ and

$$u(x_0, t_0) = u(x_0 - hz, t_0 - h) + hg(z).$$

The existence of such z follows from the semigroup property of the Hopf-Lax formula (Lemma A.31 in Section A.12 in the Appendix) and the fact that in our case the supremum in the Hopf-Lax formula is always attained by a suitably bounded z (Lemma 9.14). Combining this with the local maximum,

$$\begin{aligned} u(x_0 - hz, t_0 - h) + hg(z) - \phi(x_0, t_0) &= u(x_0, t_0) - \phi(x_0, t_0) \\ &\ge u(x_0 - hz, t_0 - h) - \phi(x_0 - hz, t_0 - h) \end{aligned}$$

for small enough h. From this

$$0 \ge \phi(x_0, t_0) - \phi(x_0 - hz, t_0 - h) - hg(z).$$

Define a path $r : [0,1] \to \mathbf{R}^d \times (0, \infty)$ by

$$r(s) = \bigl(x_0 - (1-s)hz, t_0 - (1-s)h\bigr). \tag{9.80}$$

Then by the presumed continuous differentiablity of ϕ,

$$\begin{aligned} 0 &\ge \int_0^1 \Bigl\{ \frac{d}{ds}\phi(r(s)) - hg(z) \Bigr\}\, ds = h \int_0^1 \bigl\{ \phi_t(r(s)) + \nabla\phi(r(s)) \cdot z - g(z) \bigr\}\, ds \\ &\ge h \int_0^1 \bigl\{ \phi_t(r(s)) + f\bigl(\nabla\phi(r(s))\bigr) \bigr\}\, ds. \end{aligned}$$

For the last inequality we used the concave duality (9.19) of f and g. Divide away the factor h from the front. Let $h \to 0$, and note that $r(s) \to (x_0, t_0)$ for all $0 \le s \le 1$. Use continuity of the integrand and dominated converge to conclude that

$$0 \ge \phi_t(x_0, t_0) + f\bigl(\nabla\phi(x_0, t_0)\bigr).$$

∎

Lemma 9.26 *Suppose $u - \phi$ has a local minimum at (x_0, t_0). Then*

$$\phi_t(x_0, t_0) + f\bigl(\nabla\phi(x_0, t_0)\bigr) \ge 0.$$

Proof. To get a contradiction, suppose there exists $\beta > 0$ such that

$$\phi_t(x_0, t_0) + f\big(\nabla\phi(x_0, t_0)\big) \le -3\beta < 0.$$

By the concave duality (9.19), it is possible to find $z \in \mathbf{R}^d$ such that

$$\phi_t(x_0, t_0) + \nabla\phi(x_0, t_0) \cdot z - g(z) \le -2\beta.$$

By the continuous differentiability, there exists a neighborhood B of (x_0, t_0) in $\mathbf{R}^d \times (0, \infty)$ such that

$$\phi_t(x, t) + \nabla\phi(x, t) \cdot z - g(z) \le -\beta$$

for $(x, t) \in B$.

Define the path $r(s)$ as in (9.80) in the previous proof, and repeat the integration step. Take h small enough so that $r(s) \in B$, and note that by the semigroup property of the Hopf-Lax formula,

$$u(x_0, t_0) \ge u(x_0 - hz, t_0 - h) + hg(z).$$

Putting all this together gives

$$\begin{aligned}\phi(x_0, t_0) - \phi(x_0 - hz, t_0 - h) &= h \int_0^1 \big\{\phi_t(r(s)) + \nabla\phi(r(s)) \cdot z\big\}\, ds \\ &\le hg(z) - \beta h \le u(x_0, t_0) - u(x_0 - hz, t_0 - h) - \beta h.\end{aligned}$$

Rearranging this gives

$$u(x_0, t_0) - \phi(x_0, t_0) \ge u(x_0 - hz, t_0 - h) - \phi(x_0 - hz, t_0 - h) + \beta h$$

for all small enough $h > 0$. Thus (x_0, t_0) cannot be a local minimum. ∎

We have shown that u satisfies the properties of a uniformly continuous viscosity solution of (9.27). By Ishii's uniqueness theorem A.33, Theorem 9.8 is proved.

Notes

Variational approach

The variational approach to exclusion processes and the related height processes was introduced in articles [35], [34], and [37]. The approach works for admissible height functions somewhat more complicated than the one described in Section 9.1. For example, one can take a function $\varrho : \mathbf{Z}^d \to \mathbf{Z}_+$ such that $\varrho(0) = 0$, and define the state space H of height functions by

$$H = \{h : \mathbf{Z}^d \to \mathbf{Z} : h(v) - h(u) \le \varrho(v - u)\}.$$

This type of example was treated in [30]. It is also possible to permit the height values, or the exclusion particles, to jump more than one step at a time. Such an example appears in [38]. Examples where the rates of the Poisson clocks are inhomogeneous in space or random appear in [30], [37], [42], and [39]. The approach does not seem to work when jumps in both directions are permitted, or when the rate of jumping is a more general function of the local configuration.

The pathwise variational property can be formulated as a linear mapping in max–plus algebra (unpublished work of I. Grigorescu and M. Kang).

In addition to exclusion processes and related height processes, the variational approach has been successfully used in connection with Hammersley's process. This process describes the totally asymmetric evolution of point particles on the real line. Its variational representation involves the increasing sequences model on planar Poisson points. The idea for this process appeared in Hammersley's classic paper [20]. Aldous and Diaconis [1] defined the process on the infinite line. Properties of this process have been investigated in papers [33], [36], [40], and [41]. A review of the connection between Hammersley's process and increasing sequences appeared in [19], and of the wider mathematical context in [2].

In addition to Evans's textbook [14], another excellent reference on conservation laws is [25].

Exercise 9.1 Prove the attractivity lemma 9.1.

Exercise 9.2 *Multidimensional increment process.* Formulate an increment process for the multidimensional height process $h_t : \mathbf{Z}^d \to \mathbf{Z}$ for $d \geq 2$. For example, one could set $\eta(u, i) = h(u) - h(u - e_i)$, and define the state as $\eta = (\eta(u, i) : u \in \mathbf{Z}^d, 1 \leq i \leq d)$. What is the state space? Observe that not all configurations η can represent increments because

$$\begin{aligned} &\big(h(u) - h(u - e_i)\big) + \big(h(u - e_i) - h(u - e_i - e_j)\big) \\ &\qquad = \big(h(u) - h(u - e_j)\big) + \big(h(u - e_j) - h(u - e_j - e_i)\big). \end{aligned}$$

What is the jump rule for the η_t process?

No examples of translation-invariant equilibrium distributions for such increment processes are known in more than one dimension.

Exercise 9.3 *Riemann solution.* Compute the evolution of the simplest shock profile. Take $d = 1$ and $K = 1$. Let $0 < \lambda < \rho < 1$, and

$$\psi_0(x) = \begin{cases} \lambda x, & x \leq 0 \\ \rho x, & x > 0. \end{cases}$$

Compute $\psi(x,t)$ from the Hopf-Lax formula (9.17) with g given by (9.39). Differentiate to find the macroscopic particle density $\rho(x,t)$. You will find that the shock travels at speed

$$v = \frac{f(\lambda) - f(\rho)}{\lambda - \rho}.$$

Exercise 9.4 *Shock profiles from smooth profiles.* With $0 < \lambda < \rho < 1$ as in Exercise 9.3, let u_0 be an arbitrary smooth function such that $u_0(x) = \lambda x$ for $x \le -1$ and $u_0(x) = \rho x$ for $x \ge 1$. Let $u(x,t)$ be the solution from the Hopf-Lax formula. Show that after some time T_0, $u(x,t) = \psi(x,t)$ where ψ is the solution calculated in Exercise 9.3.

Exercise 9.5 *Rarefaction fan.* Repeat Exercise 9.3 with $\lambda > \rho$. What you see now is called a rarefaction fan.

A Appendix

This section is a collection of material from analysis and probability. Some definitions and terminology are included as a reminder to the reader. We state some important theorems whose proof is outside the scope of the text, such as Choquet's and de Finetti's theorems. And we prove some technical lemmas that are used in the text.

A.1 Basic measure theory and probability

The fundamental mathematical object in probability theory is the *probability space* $(\Omega, \mathcal{F}, P)$, which models a random experiment or collection of experiments. The *sample space* Ω is the set of all possible outcomes of the experiment, $\mathcal{F}$ is a σ-algebra of *events*, and P is a *probability measure* on $\mathcal{F}$. A σ-algebra is a collection of subsets of the space that satisfies these axioms:

(1) $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.
(2) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
(3) If $A_j \in \mathcal{F}$ for $j = 1, 2, 3, \ldots$ then $\cup_{j=1}^{\infty} A_j \in \mathcal{F}$.

The axioms for the probability measure P are these:

(1) $0 \le P(A) \le 1$ for all $A \in \mathcal{F}$, $P(\emptyset) = 0$ and $P(\Omega) = 1$.
(2) If $A_j \in \mathcal{F}$ for $j = 1, 2, 3, \ldots$, $A_i \cap A_j = \emptyset$ whenever $i \neq j$, and $A = \cup_{j=1}^{\infty} A_j$, then

$$P(A) = \sum_{j=1}^{\infty} P(A_j).$$

More generally, a *measure* μ is a $[0, \infty]$-valued function on a σ-algebra that satisfies $\mu(\emptyset) = 0$ and the countable additivity axiom (2) above.

For any collection $\mathcal{A}$ of subsets of a space Ω, the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$ is by definition the intersection of all the σ-algebras that contain $\mathcal{A}$. It is the smallest σ-algebra that contains $\mathcal{A}$.

Let Ω be an arbitrary space, and $\mathcal{L}$ and $\mathcal{P}$ collections of subsets of Ω. $\mathcal{P}$ is a π-system if it is closed under intersections, in other words if $A, B \in \mathcal{P}$, then $A \cap B \in \mathcal{P}$. $\mathcal{L}$ is a λ-system if it has the following three properties:

(1) $\Omega \in \mathcal{L}$.
(2) If $A, B \in \mathcal{L}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{L}$.
(3) If $\{A_n : 1 \le n < \infty\} \subseteq \mathcal{L}$ and $A_n \nearrow A$ then $A \in \mathcal{L}$.

Theorem A.1 *(Dynkin's π-λ-theorem) If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then $\mathcal{L}$ contains the σ-algebra $\sigma(\mathcal{P})$ generated by $\mathcal{P}$.*

For a proof, see the Appendix in [11].

Exercise A.1 A collection $\mathcal{A}$ of subsets of Ω is an *algebra* if $\Omega \in \mathcal{A}$, $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$, and $A \cup B \in \mathcal{A}$ whenever $A \in \mathcal{A}$ and $B \in \mathcal{A}$. Suppose P is a probability measure on the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$. Show that for every $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$ there exists $A \in \mathcal{A}$ such that $P(A \triangle B) < \varepsilon$. The operation $\triangle$ is the symmetric difference defined by $A \triangle B = (A \smallsetminus B) \cup (B \smallsetminus A)$.

A.2 Metric spaces

A *metric* on a space Y is a distance function ρ that has these properties for all $x, y, z \in Y$:

$$\begin{aligned}
&(1) \qquad 0 \le \rho(x,y) < \infty, \text{ and } \rho(x,y) = 0 \text{ iff } x = y \\
&(2) \qquad \rho(x,y) = \rho(y,x) \qquad \text{(symmetry)} \\
&(3) \qquad \rho(x,y) \le \rho(x,z) + \rho(z,y) \qquad \text{(triangle inequality).}
\end{aligned}$$

Convergence of a sequence $\{x_n\}$ to a point x in Y means that the distance vanishes in the limit: $x_n \to x$ if $\rho(x_n, x) \to 0$. $\{x_n\}$ is a *Cauchy sequence* if $\sup_{m>n} \rho(x_m, x_n) \to 0$ as $n \to \infty$. *Completeness* of a metric space means that every Cauchy sequence in the space has a limit in the space. A countable set $\{y_k\}$ is *dense* in Y if for every $x \in Y$ and every $\varepsilon > 0$, there exists a k such that $\rho(x, y_k) < \varepsilon$. Y is a *separable* metric space if it has a countable dense set. A complete, separable metric space is called a *Polish space.*

The open ball of radius r centered at x is

$$B(x,r) = \{y \in Y : \rho(x,y) < r\}.$$

A set $G \subseteq Y$ is *open* if every point in G has an open ball around it that lies in G. The collection of open sets is called a *topology.* A topology is a more fundamental notion than a metric. A metric is just one of many ways of specifying a topology (in other words, a class of open sets) on a space.

If two metrics ρ_1 and ρ_2 on Y determine the same open sets, they share many properties. For example, they have the same convergent sequences, and the same dense sets. Completeness is an important counterexample to this, for it is a property of the metric and not of the topology. For example, on $Y = [1, \infty)$ the metrics $\rho_1(x,y) = |x-y|$ and $\rho_2(x,y) = |x^{-1} - y^{-1}|$ have the same open sets, but ρ_1 is complete while ρ_2 is not.

The *Borel σ-algebra* $\mathcal{B}(Y)$ is the smallest σ-algebra on Y that contains all the open sets. Elements of $\mathcal{B}(Y)$ are Borel sets, and measures defined on $\mathcal{B}(Y)$ are Borel measures.

The Cartesian product $Y^n = Y \times Y \times \cdots \times Y$ is a metric space with metric $\overline{\rho}(\mathbf{x}, \mathbf{y}) = \sum_i \rho(x_i, y_i)$ defined for vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in Y^n. Y^n has its own Borel σ-algebra $\mathcal{B}(Y^n)$, but also the product σ-algebra $\mathcal{B}(Y)^{\otimes n}$. Since the projections $\mathbf{x} \mapsto x_i$ are continuous, $\mathcal{B}(Y^n)$ always contains $\mathcal{B}(Y)^{\otimes n}$. If Y is separable, then we have equality $\mathcal{B}(Y^n) = \mathcal{B}(Y)^{\otimes n}$. This fact extends to the countably infinite product space $Y^{\mathbf{N}}$ of sequences $\mathbf{x} = (x_i)_{1 \le i < \infty}$, metrized by

$$\overline{\rho}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\infty} 2^{-i} \left(\rho(x_i, y_i) \wedge 1 \right).$$

If Y is compact, then so is $Y^{\mathbf{N}}$. This is a consequence of the much more general theorem of Tychonoff (Section 4.6 in [17]). In metric spaces compactness is equivalent to sequential compactness, which requires that every sequence has a convergent subsequence. This latter property can be verified relatively easily for $Y^{\mathbf{N}}$. Given a sequence $\{\mathbf{x}^n\}$ in $Y^{\mathbf{N}}$, each coordinate sequence $\{x_i^n\}$ lies in the compact space Y. The familiar diagonal argument constructs a subsequence $\{\mathbf{x}^{n_k}\}$ along which each coordinate sequence converges: $x_i^{n_k} \to x_i$ as $k \to \infty$. This coordinatewise convergence is equivalent to $\overline{\rho}(\mathbf{x}^{n_k}, \mathbf{x}) \to 0$.

A.2.1 Weak topology on probability measures

Let $\mathcal{M}_1(Y)$ the space of Borel probability measures on Y, and $C_b(Y)$ the space of bounded continuous functions on Y. The ε-neighborhood of a set $A \subseteq Y$ is by definition

$$A^{(\varepsilon)} = \{x \in Y : \rho(x, y) < \varepsilon \text{ for some } y \in A\}.$$

The *Prohorov metric* on $\mathcal{M}_1(Y)$ is defined by

$$r(\mu, \nu) = \inf\{\varepsilon > 0 : \nu(F) \le \mu(F^{(\varepsilon)}) + \varepsilon \text{ for every closed set } F \subseteq Y\}. \tag{A.1}$$

Convergence under the Prohorov metric is the familiar *weak convergence* of probability measures:

$$r(\mu_n, \mu) \to 0 \quad \text{iff} \quad \int f \, d\mu_n \to \int f \, d\mu \quad \text{for all } f \in C_b(Y).$$

It is important to know that if Y is a complete separable metric space, then so is $\mathcal{M}_1(Y)$, and if Y is compact, so is $\mathcal{M}_1(Y)$. The Borel σ-algebra $\mathcal{B}(\mathcal{M}_1(Y))$ on the space of probability measures is the same as the σ-algebra generated by the maps $\mu \mapsto \mu(A)$ as A varies over Borel subsets of Y.

A key fact about the weak topology of probability measures on a Polish space Y is that a set $U \subseteq \mathcal{M}_1(Y)$ is relatively compact iff it is *tight*, which means that for every $\varepsilon > 0$ there exists a compact set $K \subseteq Y$ such that $\inf_{\mu \in U} \mu(K) \ge 1 - \varepsilon$.

A.2.2 Skorokhod topology on path space

Let D_Y be the space of functions $\omega : [0, \infty) \to Y$ that are right-continuous and have left limits everywhere. Right-continuity at t means that $\omega(s) \to \omega(t)$ as s approaches t from above. The existence of a left limit at t means that a point $\omega(t-) \in Y$ exists such that $\omega(s) \to \omega(t-)$ as s approaches t from below. These properties are required to hold at each $t \geq 0$. Path ω has a jump at t if $\rho(\omega(t), \omega(t-)) > 0$, and then this quantity is the magnitude of the jump. There cannot be too many large jumps in any bounded interval. Namely, for any $\varepsilon > 0$ and $0 < T < \infty$,

$$\begin{aligned}&\text{a path } \omega \text{ can have only finitely many jumps}\\ &\text{of magnitude at least } \varepsilon \text{ in time interval } [0, T].\end{aligned} \tag{A.2}$$

The reason is that any accumulation point t of such jumps would fail either right-continuity or the existence of the left limit.

This is the path space for stochastic processes whose paths are not continuous in time, but are right-continuous, so for example for continuous-time Markov chains and for interacting particle systems.

D_Y has a metric defined as follows. Assume that the metric on Y satisfies $\rho(x, y) \leq 1$. This is not a restriction because we can replace the metric ρ with $\rho(x, y) \wedge 1$. Let Λ be the collection of strictly increasing, bijective Lipschitz functions $\lambda : [0, \infty) \to [0, \infty)$ that satisfy

$$\gamma(\lambda) = \sup_{s > t \geq 0} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right| < \infty. \tag{A.3}$$

For $\omega, \zeta \in D_Y$, $\lambda \in \Lambda$ and $0 < u < \infty$ set

$$s(\omega, \zeta, \lambda, u) = \sup_{t \geq 0} \rho\left(\omega(t \wedge u), \zeta(\lambda(t) \wedge u)\right),$$

and finally the Skorokhod distance between ω and ζ is

$$s(\omega, \zeta) = \inf_{\lambda \in \Lambda} \left\{ \gamma(\lambda) \vee \int_0^\infty e^{-u} s(\omega, \zeta, \lambda, u)\, du \right\}. \tag{A.4}$$

Two paths are close in this topology if, on any bounded time interval, they are uniformly close after a small distortion λ of the time axis aligns their large jumps. On the subspace of continuous functions, convergence in the s-metric is the same as uniform convergence on compact intervals. Here is a useful characterization of convergence under this metric.

Lemma A.2 *Convergence $s(\omega_n, \omega) \to 0$ is equivalent to the existence of $\lambda_n \in \Lambda$ such that $\gamma(\lambda_n) \to 0$ and*

$$\sup_{0 \leq t \leq T} \rho\big(\omega_n(t), \omega(\lambda_n(t))\big) \to 0 \tag{A.5}$$

for all $0 < T < \infty$.

Condition (A.5) can be equivalently replaced by

$$\sup_{0\le t\le T} \rho\big(\omega_n(\lambda_n(t)), \omega(t)\big) \to 0. \tag{A.6}$$

The condition $\gamma(\lambda_n) \to 0$ implies that the derivative λ_n' converges to 1 uniformly (λ_n is a.e. differentiable by Lipschitz continuity), and that λ_n converges to the identity function uniformly on compact intervals. Note that $\lambda_n \in \Lambda$ entails $\lambda_n(0) = 0$.

The coordinate mappings $\omega \mapsto \omega(t)$ are not continuous on D_Y but they are Borel measurable. If Y is a separable metric space, the Borel σ-algebra $\mathcal{B}(D_Y)$ is the same as the σ-algebra $\mathcal{F}$ generated by the coordinate mappings. If Y is a Polish space, then so is D_Y. Exercise A.2 below implies that the function $(\omega, t) \mapsto \omega(t)$ is jointly measurable on $D_Y \times [0,\infty)$. This is useful for example for concluding that integrals of the type $\int_0^t g(\omega(s))\,ds$ are measurable functions of a path ω.

Next we state a compactness criterion for a sequence of probability measures on D_Y. This comes in terms of the following modulus of continuity. For $\zeta \in D_Y$, $\delta > 0$, and $0 < T < \infty$,

$$w'(\zeta, \delta, T) = \inf_{\{t_i\}} \sup\{\rho(\zeta(s), \zeta(t)) : s, t \in [t_{i-1}, t_i) \text{ for some } i\}$$

where the infimum is over finite partitions $0 = t_0 < t_1 < \cdots < t_{n-1} < T \le t_n$ that satisfy $\min_{1\le i\le n}(t_i - t_{i-1}) > \delta$. Note the elementary but important property that

$$\lim_{\delta\to 0} w'(\zeta, \delta, T) = 0 \quad \text{for any fixed } \zeta \in D_Y \text{ and } 0 < T < \infty. \tag{A.7}$$

To check this, fix $N > 0$, define $\tau_0^N = 0$, and inductively

$$\tau_k^N = \inf\big\{s > \tau_{k-1}^N : \rho\big(\zeta(s), \zeta(\tau_{k-1}^N)\big) > 1/N\big\}.$$

By right-continuity $\tau_k^N > \tau_{k-1}^N$, and by the existence of left limits, $\tau_k^N \nearrow \infty$ as $k \nearrow \infty$. Pick n so that $\tau_n^N > T$, and let

$$\delta < \min_{1\le k\le n} (\tau_k^N - \tau_{k-1}^N).$$

Then $\{\tau_k^N\}_{0\le k\le n}$ is an admissible partition in the infimum in the definition of $w'(\zeta, \delta, T)$, and we conclude that $w'(\zeta, \delta, T) \le N^{-1}$.

As a byproduct we get the useful fact that

$$\text{a path } \zeta \in D_Y \text{ is bounded in a bounded time interval } [0, T]. \tag{A.8}$$

Boundedness in an abstract metric space means that the path $\{\zeta(t) : 0 \le t \le T\}$ lies inside a large enough ball. This follows because by the choice of n above,

$$\sup_{0\le t\le T} \rho\big(\zeta(0), \zeta(t)\big) \le \max_{0\le k<n} \rho\big(\zeta(0), \zeta(\tau_k^N)\big) + 1/N.$$

Here is a compactness criterion.

Theorem A.3 *Let (Y,ρ) be a complete, separable metric space, and let $\{Q^n\}$ be a sequence of probability measures on D_Y. Then $\{Q^n\}$ is tight iff these two conditions hold.*

(i) *For every $\varepsilon > 0$ and $t \ge 0$, there exists a compact set $K \subseteq Y$ such that*

$$\limsup_{n\to\infty} Q^n\{\zeta : \zeta(t) \in K^c\} \le \varepsilon.$$

(ii) *For every $\varepsilon > 0$ and $0 < T < \infty$ there exists a $\delta > 0$ such that*

$$\limsup_{n\to\infty} Q^n\{\zeta : w'(\zeta,\delta,T) \ge \varepsilon\} \le \varepsilon.$$

Much used references for the weak topology and D-space are [3] and [13]. Lemma A.2 is part of Proposition 5.2 on p. 119 in [13], and Theorem A.3 is Corollary 7.4 on p. 129 in [13].

Exercise A.2 Suppose Z is a measurable space, and a function $f : Z\times[0,\infty) \to \mathbf{R}$ has these properties: $x \mapsto f(x,t)$ is measurable for each fixed t, and $t \mapsto f(x,t)$ is right-continuous for each fixed x. Show that f is jointly measurable, by considering

$$f_n(x,t) = f(x,0)\cdot \mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{\infty} f(x,k2^{-n})\cdot \mathbf{1}_{((k-1)2^{-n},k2^{-n}]}(t).$$

A.3 Ordering among configurations and measures

Let W be a compact subset of $\mathbf{R}$ and S a countable set. A partial order between configurations $\eta,\zeta \in X = W^S$ is defined by $\eta \ge \zeta$ iff $\eta(x) \ge \zeta(x)$ for all $x \in S$. A continuous function f on X is increasing if $f(\eta) \ge f(\zeta)$ whenever $\eta \ge \zeta$. In terms of such functions we can define an order among probability measures. For probability measures μ,ν on X let us say $\mu \ge \nu$ if

$$\int f\,d\mu \ge \int f\,d\nu$$

for all increasing continuous functions f. Ordering between measures turns out to be equivalent to a coupling property.

Theorem A.4 *(Strassen's Theorem) Let μ and ν be probability measures on X. Then $\mu \ge \nu$ iff there exists a probability measure Q on $X \times X$ with these properties: $Q(A\times X) = \mu(A)$ and $Q(X\times B) = \nu(B)$ for all measurable sets $A,B \subseteq X$, and $Q\{(\eta,\zeta) : \eta \ge \zeta\} = 1$.*

The 'if' part of the theorem is immediate, but the other direction is not so easy. We actually do not need the hard part of the theorem if we take as definition of $\mu \geq \nu$ the existence of the coupling measure Q. A proof of the theorem can be found in Section II.2 of [27].

Next some properties of the order relation.

Lemma A.5 *Suppose $\mu \geq \nu$ and for all $x \in S$,*

$$\int \eta(x)\, \mu(d\eta) = \int \eta(x)\, \nu(d\eta).$$

Then $\mu = \nu$.

In particular, this conclusion follows from having both $\mu \geq \nu$ and $\nu \geq \mu$.

Proof. Let Q be the coupling measure. Then for all x, $Q\{\eta(x) \geq \zeta(x)\} = 1$ but

$$\int (\eta(x) - \zeta(x))\, Q(d\eta, d\zeta) = \int \eta(x)\, \mu(d\eta) - \int \zeta(x)\, \nu(d\zeta) = 0.$$

It follows that $\eta(x) = \zeta(x)$ Q-almost surely. Since S is countable, then $Q\{\eta = \zeta\} = 1$ and equality of the marginal distributions μ and ν follows.

To verify the last statement of the lemma, just observe that $\mu \geq \nu$ and $\nu \geq \mu$ together imply that $\int \eta(x)\, \mu(d\eta) = \int \eta(x)\, \nu(d\eta)$. ∎

Lemma A.6 *Suppose $\{\mu_n\}$ is a monotone sequence of probability measures on X. Then the weak limit $\mu_n \to \mu$ exists.*

Proof. By compactness of X, it suffices to show that any two weak limit points of the sequence $\{\mu_n\}$ agree. So let μ' and μ'' be two limit points. Fix a finite set $\{x_1, x_2, \dots, x_m\} \subseteq S$ of sites, and let

$$F_n(u_1, u_2, \dots, u_m) = \mu_n\{\eta(x_1) \leq u_1, \eta(x_2) \leq u_2, \dots, \eta(x_m) \leq u_m\}$$

be the multivariate distribution functions of the marginal distributions of the μ_n's on the $\eta(x_1), \eta(x_2), \dots, \eta(x_m)$ coordinates. The function

$$f(\eta) = -\mathbf{1}\{\eta(x_1) \leq u_1, \eta(x_2) \leq u_2, \dots, \eta(x_m) \leq u_m\}$$

is increasing, so by the monotonicity assumption the limits

$$\widetilde{F}(u_1, u_2, \dots, u_m) = \lim_{n\to\infty} F_n(u_1, u_2, \dots, u_m)$$

exist. For all but countably many vectors $(u_1, u_2, \ldots, u_m)$,

$$\begin{aligned}\mu'\{\eta(x_1) = u_1\} &= \mu'\{\eta(x_2) = u_2\} = \cdots = \mu'\{\eta(x_m) = u_m\} \\ &= \mu''\{\eta(x_1) = u_1\} = \mu''\{\eta(x_2) = u_2\} = \cdots = \mu''\{\eta(x_m) = u_m\} = 0.\end{aligned}$$

It is a basic property of weak convergence $\nu_j \to \nu$ of probablity measures that $\nu_j(A) \to \nu(A)$ for any Borel set A whose boundary ∂A satisfies $\nu(\partial A) = 0$. Hence for all vectors that satisfy the above condition,

$$\begin{aligned}\mu'\{\eta(x_1) \le u_1, \eta(x_2) \le u_2, \ldots, \eta(x_m) \le u_m\} &= \widetilde{F}(u_1, u_2, \ldots, u_m) \\ &= \mu''\{\eta(x_1) \le u_1, \eta(x_2) \le u_2, \ldots, \eta(x_m) \le u_m\}.\end{aligned}$$

Since such vectors are dense, we have shown that μ' and μ'' have identical marginal distributions on any finite set of coordinates. Consequently $\mu' = \mu''$. ∎

Recall the definition of the Bernoulli measures ν_ρ from (4.15).

Lemma A.7 *Let μ be a probability measure on $X = \{0,1\}^S$. Suppose there exists a number $\rho_0 \in [0,1]$ such that $\nu_\rho \le \mu$ for $\rho < \rho_0$ and $\nu_\lambda \ge \mu$ for $\lambda > \rho_0$. Then $\mu = \nu_{\rho_0}$.*

Proof. For any finite set $A \subseteq S$, $f(\eta) = \mathbf{1}\{\eta = 1 \text{ on } A\}$ is an increasing function. Thus for $\rho < \rho_0 < \lambda$,

$$\rho^{|A|} = \nu_\rho\{\eta = 1 \text{ on } A\} \le \mu\{\eta = 1 \text{ on } A\} \le \nu_\lambda\{\eta = 1 \text{ on } A\} = \lambda^{|A|}.$$

Letting $\rho \nearrow \rho_0$ and $\lambda \searrow \rho_0$ gives $\mu\{\eta = 1 \text{ on } A\} = \rho_0^{|A|}$. It is an exercise to verify that the measure on sets of this type determines the entire measure on X. ∎

A.4 Translation invariance and ergodicity

The basic formulation of the pointwise ergodic theorem (Birkhoff's ergodic theorem) is the following. Let $(\Omega, \mathcal{H}, P)$ be a probability space, and $T : \Omega \to \Omega$ a measure-preserving transformation. In other words, T is a measurable map on Ω, and $P(T^{-1}A) = P(A)$ for all $A \in \mathcal{H}$. Let $\mathcal{J}_T$ be the σ-algebra of T-invariant events:

$$\mathcal{J}_T = \{A \in \mathcal{H} : T^{-1}A = A\}.$$

$T^k = T \circ T \circ \cdots \circ T$ denotes k-fold composition of T with itself.

Theorem A.8 *(Ergodic Theorem) Let $f \in L^1(P)$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k = E[\, f \,|\, \mathcal{J}_T \,]$$

P-almost surely and in $L^1(P)$.

If T is invertible with a measurable inverse, then the limits for T and T^{-1} are the same:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^{-k} = E[\, f \,|\, \mathcal{J}_{T^{-1}} \,] = E[\, f \,|\, \mathcal{J}_T \,] = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k. \tag{A.9}$$

The reason is that $\mathcal{J}_T = \mathcal{J}_{T^{-1}}$, as can be checked directly from the definition.

The setting where we apply these notions is that of spatial translations or shifts. Let W be a Polish space, $d \geq 1$ an integer, and $X = W^{\mathbf{Z}^d}$ the product space of configurations or functions $\eta : \mathbf{Z}^d \to W$. X is a Polish space under the product metric, and $\mathcal{B}(X)$ is its Borel σ-algebra, also the product σ-algebra generated by coordinate mappings. Translations are invertible, continuous maps θ_x defined on X by $\theta_x \eta(y) = \eta(x+y)$ for all $x, y \in \mathbf{Z}^d$ and $\eta \in X$. They form a group $\Theta = \{\theta_x : x \in \mathbf{Z}^d\}$. A probability measure μ on X is translation invariant if $\mu(\theta_x^{-1} A) = \mu(A)$ for all $x \in \mathbf{Z}^d$ and $A \in \mathcal{B}(X)$. We write $\mathcal{S}$ for the space of translation invariant probability measures on X. It is a closed subset of the space $\mathcal{M}_1$ of probability measures on X in the weak topology. If X is compact, then so are $\mathcal{M}_1$ and $\mathcal{S}$.

Let $\mathcal{J}_\Theta$ be the sub-σ-algebra of translation invariant events:

$$\mathcal{J}_\Theta = \{A \in \mathcal{B}(X) : \theta_x^{-1} A = A \text{ for all } x \in \mathbf{Z}^d\}.$$

A measure $\mu \in \mathcal{S}$ is *ergodic* if $\mu(A) \in \{0, 1\}$ for all $A \in \mathcal{J}_\Theta$. These definitions also apply to the coordinate process $\{\eta(x) : x \in \mathbf{Z}^d\}$ defined on the probability space $(X, \mathcal{B}(X), \mu)$. The process $\{\eta(x)\}$ is *stationary* if the underlying measure is translation invariant, and ergodic if the measure is ergodic.

Let $\Lambda_k = [-k, k]^d \cap \mathbf{Z}^d$ be the cube with $(2k+1)^d$ sites, centered at the origin. The multiparameter ergodic theorem states that for any $\mu \in \mathcal{S}$ and $f \in L^1(\mu)$,

$$\lim_{k\to\infty} \frac{1}{(2k+1)^d} \sum_{x \in \Lambda_k} f \circ \theta_x = E^\mu[\, f \,|\, \mathcal{J}_\Theta \,] \quad \text{both } \mu\text{-almost surely and in } L^1(\mu). \tag{A.10}$$

If μ is ergodic, the limit is the constant $\int f \, d\mu$. A proof of this can be found in Chapter 14 of [18].

One can use the ergodic theorem to show that two translation invariant probability measures coincide iff they agree on $\mathcal{J}_\Theta$. From this one deduces that a translation invariant probability measure is ergodic iff it is an extreme point of $\mathcal{S}$. A measure $\mu \in \mathcal{S}$ is an extreme point if it cannot be written nontrivially as a convex combination of other elements of $\mathcal{S}$. $\mathcal{S}_e$ denotes the set of extreme points of $\mathcal{S}$, or equivalently, the set of ergodic probability measures.

Next some basic facts used in the text.

Lemma A.9 *Suppose U and V are Polish spaces, $Y = U^{\mathbf{Z}^d}$, $Z = V^{\mathbf{Z}^d}$, μ is an i.i.d. product measure on Y and ν is an ergodic measure on Z. Let $W = U \times V$. Then $\bar{\mu} = \mu \otimes \nu$ is an ergodic measure on $X = Y \times Z = W^{\mathbf{Z}^d}$.*

Proof. The ζ-section B^ζ of a measurable set $B \subseteq X$ is by definition

$$B^\zeta = \{\eta \in Y : (\eta, \zeta) \in B\} \qquad \text{for } \zeta \in Z.$$

Check that translations operate as follows:

$$(\theta_x^{-1} B)^\zeta = \theta_x^{-1}(B^{\theta_x \zeta}).$$

Let A be a translation invariant event on X. We need to show $\bar{\mu}(A) = 0$ or 1. Let $c = \bar{\mu}(A)$. By Fubini's theorem,

$$\int_Z \mu(A^\zeta)\, \nu(d\zeta) = c.$$

By the translation invariance of μ and A,

$$\mu(A^{\theta_x \zeta}) = \mu(\theta_x^{-1}[A^{\theta_x \zeta}]) = \mu([\theta_x^{-1} A]^\zeta) = \mu(A^\zeta).$$

Thus the function $\zeta \mapsto \mu(A^\zeta)$ is invariant, and so by the ergodicity of ν, $\mu(A^\zeta) = c$ for ν-almost every ζ.

Given $\varepsilon > 0$, pick an event $B \subseteq X$ such that $\bar{\mu}(A \triangle B) \le \varepsilon$ and B depends on only finitely many coordinates (Exercise A.1). So for some finite set $\Lambda \subseteq \mathbf{Z}^d$ and Borel set $\widetilde{B} \subseteq W^\Lambda$,

$$B = \{(\eta, \zeta) : (\eta_\Lambda, \zeta_\Lambda) \in \widetilde{B}\}$$

where $\eta_\Lambda = (\eta(x) : x \in \Lambda)$ denotes the configuration restricted to Λ, and similarly for ζ_Λ. Fix $x \in \mathbf{Z}^d$ so that $(\Lambda + x) \cap \Lambda = \emptyset$. Then B and $\theta_x^{-1} B$ depend on disjoint sets of coordinates. First observe that

$$\begin{aligned} |\bar{\mu}(A) - \bar{\mu}(B \cap \theta_x^{-1} B)| \;&=\; |\bar{\mu}(A \cap \theta_x^{-1} A) - \bar{\mu}(B \cap \theta_x^{-1} B)| \\ \le\; \bar{\mu}\left([A \cap \theta_x^{-1} A] \triangle [B \cap \theta_x^{-1} B]\right) \;&\le\; \bar{\mu}(A \triangle B) + \bar{\mu}(\theta_x^{-1} A \triangle \theta_x^{-1} B) \;\le\; 2\varepsilon. \end{aligned} \tag{A.11}$$

For the steps above, the reader needs to check that in general for a measure ρ and any events,

$$|\rho(G) - \rho(H)| \leq \rho(G \triangle H)$$

and also that

$$(G_1 \cap G_2) \triangle (H_1 \cap H_2) \subseteq (G_1 \triangle H_1) \cup (G_2 \triangle H_2).$$

Next, by the product form and translation invariance of μ,

$$\begin{aligned} \bar{\mu}(B \cap \theta_x^{-1} B) = \int \mu(B^\zeta \cap [\theta_x^{-1} B]^\zeta)\, \nu(d\zeta) = \int \mu(B^\zeta \cap \theta_x^{-1}[B^{\theta_x \zeta}])\, \nu(d\zeta) \\ = \int \mu(B^\zeta)\mu(B^{\theta_x \zeta})\, \nu(d\zeta). \end{aligned}$$

We already observed that $\mu(A^\zeta) = \mu(A^{\theta_x \zeta})$, so

$$\begin{aligned} & |\mu(B^\zeta)\mu(B^{\theta_x\zeta}) - \mu(A^\zeta)^2| = |\mu(B^\zeta)\mu(B^{\theta_x\zeta}) - \mu(A^\zeta)\mu(A^{\theta_x\zeta})| \\ \leq\ & |\mu(B^\zeta) - \mu(A^\zeta)|\mu(B^{\theta_x\zeta}) + |\mu(B^{\theta_x\zeta}) - \mu(A^{\theta_x\zeta})|\mu(A^\zeta) \\ \leq\ & \mu(A^\zeta \triangle B^\zeta) + \mu(A^{\theta_x\zeta} \triangle B^{\theta_x\zeta}) = \mu([A\triangle B]^\zeta) + \mu([A\triangle B]^{\theta_x\zeta}). \end{aligned}$$

For the last equality, the reader needs to check another property of the symmetric difference operation $\triangle$. By integrating over the above inequality, we get

$$\begin{aligned} & \left| \bar{\mu}(B \cap \theta_x^{-1} B) - \int \mu(A^\zeta)^2\, \nu(d\zeta) \right| \\ =\ & \left| \int \mu(B^\zeta)\mu(B^{\theta_x\zeta})\, \nu(d\zeta) - \int \mu(A^\zeta)^2\, \nu(d\zeta) \right| \\ \leq\ & \int \left| \mu(B^\zeta)\mu(B^{\theta_x\zeta}) - \mu(A^\zeta)^2 \right|\, \nu(d\zeta) \\ \leq\ & \int \left[\mu([A\triangle B]^\zeta) + \mu([A\triangle B]^{\theta_x\zeta}) \right]\, \nu(d\zeta) \\ =\ & 2\bar{\mu}(A \triangle B) \leq 2\varepsilon. \end{aligned}$$

Combine (A.11) with above to get

$$\left| \bar{\mu}(A) - \int \mu(A^\zeta)^2\, \nu(d\zeta) \right| \leq 4\varepsilon.$$

This says $|c - c^2| \leq 4\varepsilon$, and since $\varepsilon > 0$ was arbitrary, $c = 0$ or 1. ∎

Products of ergodic measures are not always ergodic. Here is an example.

Exercise A.3 On $X = \{-1,1\}^{\mathbf{Z}}$, define the configuration $\zeta = (\zeta(x) : x \in \mathbf{Z})$ by $\zeta(x) = (-1)^x$. Let $\mu = \frac{1}{2}(\delta_\zeta + \delta_{\theta_1\zeta})$. Show that μ is ergodic but $\mu \otimes \mu$ is not.

Lemma A.10 *Let μ be a translation invariant probability measure on $X = W^{\mathbf{Z}^d}$. Let g be a measurable function from X into a Polish space U. Let $Y = U^{\mathbf{Z}^d}$. Define a measurable map G from X into Y by $G(\eta)(x) = g(\theta_x\eta)$, and a measure ν on Y by $\nu = \mu \circ G^{-1}$. Then ν is translation invariant. If μ is also ergodic, then so is ν.*

Proof. Check that G commutes with translations: $\theta_x \circ G = G \circ \theta_x$. Then both conclusions follow. First translation invariance of ν:

$$\nu(\theta_x^{-1}A) = \mu(G^{-1}\theta_x^{-1}A) = \mu(\theta_x^{-1}G^{-1}A) = \mu(G^{-1}A) = \nu(A).$$

Then ergodicity: suppose A is an invariant event on Y. Then so is $G^{-1}A$ on X, and if μ is ergodic, $\nu(A) = \mu(G^{-1}A) \in \{0,1\}$. ∎

This lemma is perhaps clearer in stochastic process terms. Define a U-valued process $\{Y_x\}$ by $Y_x = g(\theta_x\eta)$. Then $\{Y_x\}$ inherits stationarity (and ergodicity) from $\{\eta(x)\}$.

Lemma A.11 *Suppose P is a translation invariant probability measure on $X = \{0,1\}^{\mathbf{Z}}$ such that the all zero configuration $\eta \equiv 0$ has zero P-probability. Let*

$$Y = \inf\{x \geq 1 : \eta(x) = 1\}$$

be the position of the next value 1 to the right of the origin. Then $E[\eta(0)Y] = 1$.

Proof. A computation:

$$\begin{aligned}
E[\eta(0)Y] &= \sum_{k=1}^{\infty} kP[\eta(0)=1,\, \eta(i)=0\ (0<i<k),\, \eta(k)=1] \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{k} P[\eta(0)=1,\, \eta(i)=0\ (0<i<k),\, \eta(k)=1] \\
&= \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} P[\eta(0)=1,\, \eta(i)=0\ (0<i<k),\, \eta(k)=1] \\
&= \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} P[\eta(-j)=1,\, \eta(i)=0\ (-j<i<k-j),\, \eta(k-j)=1] \\
&= \sum_{j=1}^{\infty}\sum_{\ell=0}^{\infty} P[\eta(-j)=1,\, \eta(i)=0\ (-j<i<\ell),\, \eta(\ell)=1] \\
&= P[\text{for some } m < 0 \leq n,\, \eta(m)=\eta(n)=1] \\
&= 1.
\end{aligned}$$

The last equality comes from the assumption that there is a value 1 somewhere with probability 1. For then

$$1 = \lim_{m\to\infty} P[\eta(x) = 1 \text{ for some } x \le m] = \lim_{n\to-\infty} P[\eta(x) = 1 \text{ for some } x \ge n]$$

by the convergence of probability along monotone sequences of events. But by stationarity the probabilities above are equal for different n (and different m), hence in particular

$$P[\eta(x) = 1 \text{ for some } x \le -1] = P[\eta(x) = 1 \text{ for some } x \ge 0] = 1. \qquad \blacksquare$$

Weak convergence preserves invariance under continuous mappings. For example, the spatial shift $\nu \mapsto \nu \circ \theta_x^{-1}$ of a probability measure is weakly continuous. Consequently a weak limit of translation invariant probability measures is itself translation invariant. But ergodic measures do not form a weakly closed subset of probability measures, and consequently a weak limit of ergodic processes may fail to be ergodic. Here is an example.

Exercise A.4 For $0 \le \alpha < 1$, let P_α be the distribution on $\{0,1\}^{\mathbf{Z}}$ of the stationary Markov chain with transition matrix

$$\begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}.$$

Show the weak convergence $P_\alpha \to P_0$ as $\alpha \to 0$, and that here ergodic measures converge to a nonergodic one.

The next theorem is Liggett's version of Kingman's subadditive ergodic theorem. A proof can be found in Section 6.6 of [11].

Theorem A.12 *Suppose a stochastic process $\{X_{m,n} : 0 \le m < n\}$ satisfies these properties.*

(a) $X_{0,m} + X_{m,n} \ge X_{0,n}$.

(b) *For each fixed ℓ, the process $\{X_{n\ell,(n+1)\ell} : n \ge 1\}$ is stationary and ergodic.*

(c) *The distribution of the sequence $\{X_{m,m+k} : k \ge 1\}$ is the same for all values of m.*

(d) $EX_{0,1}^+ < \infty$, *and* $\gamma = \inf_n n^{-1} EX_{0,n} > -\infty$.

Then

$$\lim_{n\to\infty} \frac{1}{n} X_{0,n} = \gamma \qquad \textit{almost surely and in } L^1.$$

Exercise A.5 Use a truncation argument to show that a process $X_{m,n}$ that satisfies $X_{m,n} \ge 0$, $X_{0,n} \ge X_{0,m} + X_{m,n}$, and assumptions (b) and (c) of Theorem A.12 satisfies $\lim_{n\to\infty} n^{-1} X_{0,n} = \gamma = \sup_n n^{-1} EX_{0,n}$ without any moment assumptions, even if $\gamma = \infty$.

A.5 Integral representations

Some functional analytic terminology used in this section does not appear anywhere else in the lectures. The reader who is not familiar with this area can simply accept Corollary A.14.

For any convex set K in a vector space, a point $x \in K$ is an *extreme point* if x cannot be expressed as a convex combination of points from K in a nontrivial fashion. In other words, $x = \beta x' + (1-\beta)x''$ for some $0 < \beta < 1$ and $x', x'' \in K$ forces $x' = x'' = x$. We write K_e for the set of extreme points of K.

Theorem A.13 *(Choquet's Theorem) Suppose K is a metrizable compact convex subset of a locally convex topological vector space $\mathcal{X}$, and $x_0 \in K$. Then there exists a Borel probability measure γ on K_e that represents x_0 in the following sense: for every continuous linear functional φ on $\mathcal{X}$,*

$$\varphi(x_0) = \int_{K_e} \varphi(x)\,\gamma(dx). \tag{A.12}$$

In the setting described in the theorem K_e is a G_δ-set (a countable intersection of open sets), so in particular a Borel set. Thus it is not problematic to integrate over the set K_e. A short proof of Theorem A.13 can be found in [28]. Reading the proof requires knowledge of the Hahn-Banach Theorem and the Rietz Representation Theorem.

The following corollary of Choquet's theorem is used in several places in the text. Let Y be a metric space, and $\mathcal{M}_1(Y)$ the space of probability measures on Y, endowed with its weak topology.

Corollary A.14 *Let $\mathcal{K}$ be a compact convex subset of $\mathcal{M}_1(Y)$, and $\mathcal{K}_e$ the set of extreme points of $\mathcal{K}$. Then $\mu \in \mathcal{K}$ iff there exists a probability measure Γ on $\mathcal{K}_e$ such that*

$$\mu = \int_{\mathcal{K}_e} \nu\,\Gamma(d\nu). \tag{A.13}$$

Remark. Interpret (A.13) in the sense that

$$\int f\,d\mu = \int_{\mathcal{K}_e} \left\{ \int f\,d\nu \right\} \Gamma(d\nu) \quad \text{for bounded Borel functions } f \text{ on } Y. \tag{A.14}$$

Proof. Let $\mathcal{M}$ be the vector space of finite signed Borel measures on Y, topologized by the weak topology defined by $C_b(Y)$, the space of bounded continuous functions on Y. This space is a locally convex topological vector space, and $C_b(Y)$ is the dual space $\mathcal{M}^*$. This topology on $\mathcal{M}_1(Y)$ is metrizable because it is the familiar weak topology, hence $\mathcal{K}$ satisfies the hypotheses of Choquet's theorem.

Conclusion (A.12) from Choquet's theorem gives (A.14) for $f \in C_b(Y)$. By taking bounded limits, we obtain (A.14) for $f = \mathbf{1}_A$ for closed sets A. An application of the π-λ-theorem A.1 extends this to $f = \mathbf{1}_A$ for all Borel sets A. A final round of bounded pointwise limits gives (A.14) as it stands.

Suppose μ satisfies (A.14). If $\mu \notin \mathcal{K}$, by the separation theorem (for example, Theorem 3.4 in [32]) there would have to exist $f \in C_b(Y)$ such that

$$\int f \, d\mu > \sup_{\lambda \in \mathcal{K}} \int f \, d\lambda.$$

This contradicts (A.14). ∎

Note that if Y is compact to begin with, then $\mathcal{M}_1(Y)$ is compact also, and in Corollary A.14 it suffices to assume that $\mathcal{K}$ is closed.

A particular case of this theorem is the ergodic decomposition of translation invariant measures. In this situation we also have uniqueness of the representation, which we use in the text. Recall the setting of Section A.4 where $X = W^{\mathbf{Z}^d}$, and assume W is compact. $\mathcal{S}$ is the space of translation invariant probability measures on X. $\mathcal{S}_e$ is the subset of ergodic measures, which is also the set of extreme points of $\mathcal{S}$.

Theorem A.15 *(Ergodic decomposition) For $\mu \in \mathcal{S}$ there is a unique probability measure Γ on $\mathcal{S}_e$ such that*

$$\mu = \int_{\mathcal{S}_e} \lambda \, \Gamma(d\lambda).$$

Proof. W is compact by assumption, hence so is X, hence so is the space $\mathcal{M}_1$ of probability measures on X, and hence so is $\mathcal{S}$. The existence of Γ follows from Corollary A.14.

Let $\Lambda_k = [-k, k]^d \cap \mathbf{Z}^d$ be the cube with $(2k+1)^d$ sites, centered at the origin. For any bounded measurable function f on X, define

$$\bar{f}(\eta) = \begin{cases} \lim_{k\to\infty} \frac{1}{(2k+1)^d} \sum_{x \in \Lambda_k} f(\theta_x \eta) & \text{if this limit exists,} \\ 0 & \text{otherwise.} \end{cases} \tag{A.15}$$

Let us abbreviate $\nu(f) = \int f \, d\nu$ for the integral. For an ergodic measure λ, $\lambda\{\bar{f} = \lambda(f)\} = 1$. Pick bounded measurable functions $f_1, f_2, \ldots, f_m$ on X and Borel subsets $A_1, A_2, \ldots, A_m$ of

R. Then

$$\begin{aligned}
&\mu\{\bar{f}_1 \in A_1, \bar{f}_2 \in A_2, \dots, \bar{f}_m \in A_m\} \\
&\quad = \int_{\mathcal{S}_e} \lambda\{\bar{f}_1 \in A_1, \bar{f}_2 \in A_2, \dots, \bar{f}_m \in A_m\}\, \Gamma(d\lambda) \\
&\quad = \int_{\mathcal{S}_e} \left\{ \prod_{i=1}^{m} \mathbf{1}_{A_i}(\lambda(f_i)) \right\} \Gamma(d\lambda) \\
&\quad = \Gamma\{\lambda : \lambda(f_1) \in A_1, \lambda(f_2) \in A_2, \dots, \lambda(f_m) \in A_m\}.
\end{aligned}$$

As the functions f_i and the sets A_i vary, the class of sets

$$\{\lambda \in \mathcal{M}_1 : \lambda(f_1) \in A_1, \lambda(f_2) \in A_2, \dots, \lambda(f_m) \in A_m\}$$

form a π-system that generates the Borel σ-algebra on the space $\mathcal{M}_1$. The above computation shows that the Γ-measures of these sets are determined by μ, and so Γ itself is uniquely determined. ∎

A.6 Exchangeable measures and de Finetti's theorem

Suppose S is an arbitrary countable set. A probability measure μ on $X = \{0,1\}^S$ is *exchangeable* if the occupation variables $\eta(x)$ can be permuted without affecting the joint distributions under μ. In other words, for any two sets $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\}$ of n sites, and any choice of numbers $k_1, \dots, k_n \in \{0,1\}$,

$$\begin{aligned}
&\mu\{\eta : \eta(x_1) = k_1, \eta(x_2) = k_2, \dots, \eta(x_n) = k_n\} \\
&\qquad = \mu\{\eta : \eta(y_1) = k_1, \eta(y_2) = k_2, \dots, \eta(y_n) = k_n\}.
\end{aligned}$$

Exercise A.6 Show that μ is exchangeable iff for all finite sets $A \subseteq S$, $\mu\{\eta = 1 \text{ on } A\}$ depends only on the size $|A|$ of A.

In general, de Finetti's theorem says that when the index set S is infinite, exchangeable measures are mixtures of i.i.d. measures. In the special case of $X = \{0,1\}^S$, i.i.d. measures are precisely the Bernoulli measures ν_ρ, indexed by the density $\rho \in [0,1]$, defined by (4.15). So de Finetti's theorem specializes to this statement:

Theorem A.16 *(de Finetti) A probability measure μ on X is exchangeable iff there exists a probability measure γ on $[0,1]$ such that*

$$\mu = \int_{[0,1]} \nu_\rho\, \gamma(d\rho).$$

A martingale proof of De Finetti's theorem can be found in Chapter 4 of [11].

A.7 Large deviations

Suppose $\{X_i\}$ are independent and identically distributed real valued random variables, and $S_n = X_1 + \cdots + X_n$. Assume that the logarithmic moment generating function

$$\Lambda(t) = \log E[e^{tX}]$$

is finite in some neighborhood of the origin. This guarantees that X_1 has all moments. Let

$$I(x) = \sup_{t\in\mathbf{R}}\{xt - \Lambda(t)\} \tag{A.16}$$

be the convex conjugate of Λ. I is $[0,\infty]$-valued, and under the finiteness assumption on Λ, $I(x) = 0$ iff $x = EX_1$. For $x < EX_1$ the supremum in (A.16) can be restricted to $t < 0$, and I is strictly decreasing to the left of EX_1. Conversely, for $x > EX_1$ the supremum in (A.16) can be restricted to $t > 0$, and I is strictly increasing to the right of EX_1. I is called the *Cramér rate function* for large deviations, a term explained by the next theorem.

Theorem A.17 *(Cramér's Theorem) Let H be a Borel subset of $\mathbf{R}$, with interior H^o and closure $\overline{H}$. Then*

$$\begin{aligned} -\inf_{x\in H^o} I(x) &\le \liminf_{n\to\infty}\frac{1}{n}\log P\left\{n^{-1}S_n \in H\right\} \\ &\le \limsup_{n\to\infty}\frac{1}{n}\log P\left\{n^{-1}S_n \in H\right\} \le -\inf_{x\in\overline{H}} I(x). \end{aligned}$$

The bounds in the theorem give the rate of exponential decay of the probability that $n^{-1}S_n$ deviates from its limit EX_1. Proofs of Cramér's theorem can be found in all books on large deviations, such as [8] and [9].

When the set H in question is an interval, the upper large deviation bound is valid already for finite n.

Proposition A.18 *Suppose H is an interval such that $P[X_1 \in H] > 0$. Then for all n,*

$$P\left\{n^{-1}S_n \in H\right\} \le \exp\left\{-\inf_{x\in\overline{H}} I(x)\right\}.$$

Proof. Let $S_{m,n} = \sum_{i=m+1}^{n} X_i$ so that $S_{0,n} = S_n$. Then by convexity and independence,

$$\begin{aligned} P\left\{(m+n)^{-1}S_{m+n} \in H\right\} &\ge P\left\{m^{-1}S_{0,m} \in H \text{ and } n^{-1}S_{m,m+n} \in H\right\} \\ &= P\left\{m^{-1}S_m \in H\right\} \cdot P\left\{n^{-1}S_n \in H\right\}. \end{aligned}$$

Thus the sequence

$$a_n = \log P\left\{n^{-1}S_n \in H\right\}$$

has the superadditivity property

$$a_{m+n} \geq a_m + a_n.$$

The hypothesis $P[X_1 \in H] > 0$ implies that $a_n > -\infty$ for all n. Then by Exercise A.7,

$$\lim_{n\to\infty} \frac{1}{n} \log P\left\{n^{-1}S_n \in H\right\} = \sup_{n\geq 1} \frac{1}{n} \log P\left\{n^{-1}S_n \in H\right\},$$

and the conclusion follows from Cramér's theorem. ∎

Exercise A.7 Suppose a_n is a $[-\infty, \infty)$-valued sequence such that $a_{m+n} \geq a_m + a_n$ for all m, n, and $a_n > -\infty$ for large enough n. Show that

$$\lim_{n\to\infty} \frac{a_n}{n} = \sup_n \frac{a_n}{n},$$

regardless of whether the supremum is finite or infinite. Show by example that the conclusion may fail if $a_n = -\infty$ for arbitrarily large n.

Exercise A.8 Prove directly the conclusion of Proposition A.18 by an application of Chebychev's inequality. For example, if $EX_1 < a < b$, then for $t \geq 0$,

$$P\left\{n^{-1}S_n \in [a,b]\right\} \leq P\left\{e^{tS_n} \geq e^{nta}\right\} \leq e^{-nta}E[e^{tS_n}] = \exp\left\{-n[ta - \Lambda(t)]\right\}.$$

Choose $t \geq 0$ to minimize the last expression.

Exercise A.9 Derive some basic Cramér rate functions.

(a) For Bernoulli variables with $P(X_i = 1) = p$ and $P(X_i = 0) = q = 1 - p$, $I(x) = x\log x + (1-x)\log(1-x) - x\log p - (1-x)\log q$ for $0 \leq x \leq 1$.

(b) For the rate α exponential distribution $I(x) = \alpha x - 1 - \log \alpha x$.

(c) For the mean λ Poisson distribution $I(x) = x\log(x/\lambda) - x + \lambda$.

A.8 Laplace and Fourier transforms

For a measurable function u on $[0, \infty)$, its *Laplace transform* is defined by

$$\phi(\lambda) = \int_0^\infty e^{-\lambda t} u(t)\, dt \quad \text{for } \lambda \geq 0.$$

Lemma A.19 *A bounded measurable function u on $[0,\infty)$ is determined by its Laplace transform up to Lebesgue-null sets. In particular, a bounded right-continuous function on $[0,\infty)$ is uniquely determined by its Laplace transform.*

Proof. For a bounded u, all derivatives of ϕ exist and are given by

$$\phi^{(k)}(\lambda) = (-1)^k \int_0^\infty e^{-\lambda t} t^k u(t)\, dt.$$

Then for integers $n > 0$ and real $x > 0$,

$$\begin{aligned} \sum_{0 \le k \le nx} \frac{1}{k!}(-1)^k n^k \phi^{(k)}(n) &= \int_0^\infty \sum_{0 \le k \le nx} \frac{1}{k!} e^{-nt} (nt)^k u(t)\, dt \\ &= \int_0^\infty P[Y_{nt} \le nx] u(t)\, dt \end{aligned}$$

where Y_{nt} is a Poisson(nt) distributed random variable. By the weak law of large numbers, $P[Y_{nt} \le nx] \to \mathbf{1}_{t \le x}$ as $n \to \infty$ for $t \ne x$, in other words for almost every t. Letting $n \to \infty$ above gives

$$\lim_{n\to\infty} \sum_{0 \le k \le nx} \frac{1}{k!}(-1)^k n^k \phi^{(k)}(n) = \int_0^x u(t)\, dt.$$

Thus these integrals are determined by the Laplace transform. If

$$\int_0^x u(t)\, dt = \int_0^x v(t)\, dt$$

for all x, then $u = v$ Lebesgue almost everywhere. And if u and v are right-continuous, they agree everywhere. ∎

The *Fourier coefficients* of a finite measure μ on $\mathbf{T}^d = (-\pi, \pi]^d$ are defined by

$$\alpha_x = \int_{\mathbf{T}^d} e^{-i\langle x,t\rangle} \mu(d\theta) \qquad \text{for } x \in \mathbf{Z}^d. \tag{A.17}$$

A sequence $\{\alpha_x\}_{x\in\mathbf{Z}^d}$ of complex numbers is *positive definite* if

$$\sum_{x,y} \alpha_{x-y} w_x \overline{w}_y \ge 0 \tag{A.18}$$

for every finite collection $w_1, \dots, w_n$ of complex numbers.

Theorem A.20 *(Herglotz' Theorem) Let $\{\alpha_x\}_{x\in\mathbf{Z}^d}$ be a sequence of complex numbers. They are the Fourier coefficients of a finite nonnegative measure on $\mathbf{T}^d$ iff the sequence is positive definite.*

The necessity of positive definiteness follows from substituting (A.17) into (A.18) and noting that the sum in (A.18) becomes $\int \left|\sum e^{-i\langle x,t\rangle}\right|^2 \mu(dt)$.

To show sufficiency, we start by checking that a positive definite sequence is bounded.

Lemma A.21 *For any positive definite sequence* $\{\alpha_x\}$*,* $\alpha_0 \geq 0$*,* $\alpha_{-x} = \overline{\alpha_x}$*, and* $|\alpha_x| \leq \alpha_0$*.*

Proof. Fix x, and in (A.18) set $w_x = w$, $w_0 = 1$, and all other w's zero. Then

$$\alpha_0(|w|^2 + 1) + \alpha_x w + \alpha_x \overline{w} \geq 0.$$

$w = 0$ gives $\alpha_0 \geq 0$. Taking $w > 0$ forces $\alpha_x + \alpha_{-x}$ real, and taking $w = i\theta$ for $\theta > 0$ forces $i\alpha_x - i\alpha_{-x}$ real. This implies $\alpha_{-x} = \overline{\alpha_x}$. And if $\alpha_0 = 0$, taking $w = -\overline{\alpha_x}$ gives $0 \geq -2|\alpha_x|^2$, so $\alpha_0 = 0$ implies that the sequence is identically zero.

In case $\alpha_0 > 0$, setting $w = \overline{\alpha_x}/\alpha_0$ gives $|\alpha_x|^2 \leq \alpha_0^2$. ∎

We can dispose of the trivial case immediately: if $\alpha_0 = 0$, take the measure identically zero. From now on, assume $\alpha_0 > 0$. We claim that for $0 < r < 1$, the function

$$f_r(s) = \sum_x \alpha_x r^{|x|_1} e^{i\langle x,s\rangle}$$

defined for $s = (s_1, \ldots, s_d)$ is the density of a bounded nonnegative measure on $\mathbf{T}^d$. To show $f_r(s) \geq 0$, set $w_x = r^{|x|_1} e^{i\langle x,s\rangle}$ for $x \geq 0$, and $w_x = 0$ otherwise. Take a limit in (A.18) to a sum over all x, y, justified for $0 < r < 1$.

$$0 \leq \sum_{x,y\geq 0} \alpha_{x-y} r^{|x|_1+|y|_1} e^{i\langle x-y,s\rangle} = \sum_z \alpha_z e^{i\langle z,s\rangle} \sum_{x\geq 0\vee z} r^{2\sum_{j=1}^d x_j - \sum_{j=1}^d z_j} = \left(\sum_{y\geq 0} r^{2|y|_1}\right) f_r(s).$$

Next,

$$(2\pi)^{-d} \int_{\mathbf{T}^d} f_r(s) e^{-\langle x,s\rangle} ds = \alpha_x r^{|x|_1}$$

shows that

$$\mu_r(ds) = (2\pi)^{-d} \alpha_0^{-1} f_r(s) ds$$

is a probability measure on $\mathbf{T}^d$, with Fourier coefficients $(\alpha_x/\alpha_0) r^{|x|_1}$. Let μ be any weak limit of μ_r as $r \nearrow 1$ along some subsequence. Such a limit exists along some subsequence by the compactness of $\mathbf{T}^d$. Then the measure $\alpha_0\mu$ has Fourier coefficients $\{\alpha_x\}$. This completes the proof of Theorem A.20. This proof is from Feller's Volume II [16].

A.9 Sampling an ergodic random field with an irreducible random walk

Suppose μ is a translation invariant, ergodic probability measure on the space $\{0,1\}^{\mathbf{Z}^d}$. Let $\rho = \mu\{\eta : \eta(x) = 1\}$, independent of x by the translation invariance assumption.

Let $p(x,y) = p(0, y-x)$ be a random walk transition on $\mathbf{Z}^d$, and let X_t be the corresponding continuous time random walk with transition probabilities

$$p_t(x,y) = \sum_{n=0}^{\infty} \frac{e^{-t}t^n}{n!} p^{(n)}(x,y).$$

The next proposition shows that the ergodic average ρ of $\{\eta(x)\}$ is produced by the average of the values $\eta(X_t)$ sampled by the random walk.

Proposition A.22 *Assume that all bounded harmonic functions for $p(x,y)$ are constant. Then for every $x \in \mathbf{Z}^d$, we have the following limit in $L^2(\mu)$:*

$$\lim_{t\to\infty} \int |E^x[\eta(X_t)] - \rho|^2 \, \mu(d\eta) = 0.$$

In particular, the conclusion is valid if $p(x,y)$ is irreducible in the sense of definition (1.26).

Proof. For $x \in \mathbf{Z}^d$, $t \geq 0$, and $\eta \in \{0,1\}^{\mathbf{Z}^d}$, set

$$g_t(x,\eta) = E^x[\eta(X_t)] = \sum_y p_t(x,y)\eta(y).$$

Let

$$\phi(\theta) = \sum_x p(0,x)e^{i\langle x,\theta\rangle}, \quad \theta = (\theta_1,\ldots,\theta_d) \in \mathbf{R}^d,$$

be the characteristic function of the jump distribution of X_t. Then

$$\begin{aligned} E^x e^{i\langle X_t,\theta\rangle} &= \sum_y p_t(0,y)e^{i\langle x+y,\theta\rangle} = e^{i\langle x,\theta\rangle} \sum_{n=0}^{\infty} \frac{e^{-t}t^n}{n!} \sum_y p^{(n)}(0,y)e^{i\langle y,\theta\rangle} \\ &= e^{i\langle x,\theta\rangle} \sum_{n=0}^{\infty} \frac{e^{-t}t^n\phi(\theta)^n}{n!} = e^{i\langle x,\theta\rangle} e^{-t(1-\phi(\theta))}. \end{aligned}$$

The next to last equality above followed from the fact that the iterated transition probability $p^{(n)}(0,y)$ is the distribution of a sum of n i.i.d. steps each with characteristic function ϕ.

The covariance

$$\alpha(x) = \mu\{\eta(0) = 1, \eta(x) = 1\} - \rho^2$$

is a positive definite sequence, so by Herglotz's theorem there exists a bounded measure γ on $\mathbf{T}^d = [-\pi, \pi)^d$ such that

$$\alpha(x) = \int e^{-i\langle x,\theta\rangle}\gamma(d\theta).$$

In the next computation, let Y_t be an independent copy of X_t, and write $E^{(x,x)}$ for expectation over the pair process (X_t, Y_t).

$$\begin{aligned}
\int g_t(x,\eta)g_s(x,\eta)\mu(d\eta) &= \int E^x[\eta(X_t)] \cdot E^x[\eta(X_s)]\mu(d\eta) \\
= \quad E^{(x,x)} \int \eta(X_t)\eta(Y_s)\mu(d\eta) &= E^{(x,x)}[\alpha(Y_s - X_t)] + \rho^2 \\
= \quad E^{(x,x)} \int e^{-i\langle Y_s - X_t,\theta\rangle}\gamma(d\theta) + \rho^2 &= \int \overline{E^x e^{i\langle Y_s,\theta\rangle}} \cdot E^x e^{i\langle X_t,\theta\rangle}\gamma(d\theta) + \rho^2 \\
= \quad \int \overline{e^{-s(1-\phi(\theta))}} \cdot e^{-t(1-\phi(\theta))}\gamma(d\theta) + \rho^2.
\end{aligned}$$

Now apply this calculation to the three terms inside the second integral below, and note that the ρ^2 terms all cancel.

$$\begin{aligned}
\int |g_t(x,\eta) - g_s(x,\eta)|^2\mu(d\eta) &= \int \left\{g_t(x,\eta)^2 - 2g_t(x,\eta)g_s(x,\eta) + g_s(x,\eta)^2\right\}\mu(d\eta) \\
= \quad \int \left|e^{-s(1-\phi(\theta))} - e^{-t(1-\phi(\theta))}\right|^2 \gamma(d\theta).
\end{aligned}$$

The integrand in the last integral above is bounded, vanishes for θ such that $\phi(\theta) = 1$, and converges to zero as $s, t \to \infty$ for other θ. (Note that $|\phi(\theta)| \le 1$.) We conclude that for each fixed x, $g_t(x, \cdot)$ is Cauchy in $L^2(\mu)$ as $t \to \infty$. Thus there exists an $L^2(\mu)$ limit

$$g(x,\eta) = \lim_{t\to\infty} g_t(x,\eta).$$

The Chapman-Kolmogorov equations for the random walk imply that

$$g_{s+t}(x,\eta) = \sum_y p_t(x,y)g_s(y,\eta)$$

for each $s, t \ge 0$. Consequently

$$\begin{aligned}
&\left\| g(x,\eta) - \sum_y p_t(x,y)g(y,\eta) \right\|_{L^2(\mu)} \\
\le \quad &\|g(x,\eta) - g_{s+t}(x,\eta)\|_{L^2(\mu)} + \sum_y p_t(x,y)\,\|g_s(y,\eta) - g(y,\eta)\|_{L^2(\mu)}
\end{aligned}$$

which vanishes as $s \to \infty$. This can be repeated for the countably many states $x \in \mathbf{Z}^d$ and a countable dense set of times t. We conclude that for μ-almost every η, $g(\cdot, \eta)$ is a harmonic function of the random walk, and so by the assumption, for μ-almost all η and all x, $g(x, \eta) = g(0, \eta)$.

By the translation property of the random walk transition and after a change in the summation index,

$$
\begin{aligned}
g_t(x,\eta) &= \sum_y p_t(x,y)\eta(y) = \sum_y p_t(0, y-x)\eta(y) \\
&= \sum_w p_t(0,w)\eta(w+x) = g_t(0, \theta_x\eta).
\end{aligned}
$$

Passing to the $t \to \infty$ limit gives $g(x,\eta) = g(0, \theta_x \eta)$ μ-almost surely. Combining with the previous paragraph, we get

$$g(0,\eta) = g(0, \theta_x\eta) \quad \text{for } \mu\text{-almost every } \eta.$$

Ergodicity implies that $g(0,\eta)$ is μ-a.s. equal to its mean, and so almost surely

$$g(0,\eta) = \int g(0,\eta)\mu(d\eta) = \lim_{t\to\infty} \int g_t(0,\eta)\mu(d\eta) = \rho.$$

We have proved that $g_t(x,\eta) \to \rho$ in $L^2(\mu)$. ∎

Corollary A.23 *Let $(X_1(t), \dots, X_n(t))$ be a vector of independent random walks of the kind considered in Proposition A.22, and $\mathbf{x} = (x_1, \dots, x_n) \in S^n$. Then*

$$\lim_{t\to\infty} \int \left| E^{\mathbf{x}}[\eta(X_1(t))\,\eta(X_2(t)) \,\cdots\, \eta(X_n(t))] \,-\, \rho^n \right|^2 \mu(d\eta) = 0.$$

Proof. Since the random walks are independent and have identical transitions, the integral equals

$$\int \left| \prod_{i=1}^n E^{x_i}[\eta(X(t))] \,-\, \rho^n \right|^2 \mu(d\eta).$$

Proceed by induction on n to show that this vanishes. The case $n = 1$ is Proposition A.22. The basic step is this: Suppose $f_t(\eta)$ and $g_t(\eta)$ are $L^2(\mu)$ functions such that $f_t \to b$ and $g_t \to c$ in $L^2(\mu)$ with constant limits, and $\|g_t\|_{L^\infty(\mu)}$ is bounded uniformly in t. Then by the triangle inequality

$$
\begin{aligned}
\|f_t g_t - bc\|_{L^2(\mu)} &= \|(f_t - b)g_t + b(g_t - c)\|_{L^2(\mu)} \\
&\le \|f_t - b\|_{L^2(\mu)} \|g_t\|_{L^\infty(\mu)} + |b| \, \|g_t - c\|_{L^2(\mu)}
\end{aligned}
$$

which vanishes as $t \to \infty$. ∎

The proof of Proposition A.22 is from Section 2.2 in [26].

A.10 The vague topology of Radon measures

A $[0, \infty]$-valued measure μ on the Borel sets of $\mathbf{R}^d$ is a *Radon measure* if $\mu(B) < \infty$ for all bounded Borel sets. The abstract definition of Radon measures on locally compact Hausdorff spaces sometimes requires some regularity of μ (see for example Chapter 7 in [17]). On $\mathbf{R}^d$ such properties are automatically satisfied, so we ignore the point. The space of Radon measures on $\mathbf{R}^d$ is denoted by $\mathbf{M}$. We go through the basic properties of the vague topology of $\mathbf{M}$ here, assuming that the reader is familiar with the usual weak topology of probability measures on Polish spaces.

Vague convergence $\mu_n \to \mu$ of Radon measures is defined by requiring that

$$\int_{\mathbf{R}^d} f\, d\mu_n \to \int_{\mathbf{R}^d} f\, d\mu$$

for all compactly supported, continuous functions f. The space of such functions is denoted by $C_c(\mathbf{R}^d)$.

Vague convergence can be defined by a metric that we next construct. First choose an increasing sequence $\{K_\ell : \ell \geq 1\}$ of compact sets such that K_ℓ lies in the interior of $K_{\ell+1}$, $\mathbf{R}^d = \bigcup K_\ell$, and for every compact set $H \subseteq \mathbf{R}^d$ there is some ℓ such that $H \subseteq K_\ell$.

For each ℓ, choose a sequence $\phi_{\ell,m} \in C_c(\mathbf{R}^d)$ of functions that are supported on $K_{\ell+1}$ and uniformly dense among the $C_c(\mathbf{R}^d)$-functions supported on K_ℓ. This can be done for example as follows. Let

$$g(x) = 1 \wedge \frac{\operatorname{dist}(x, K_{\ell+1}^c)}{\operatorname{dist}(K_\ell, K_{\ell+1}^c)}. \tag{A.19}$$

This defines a $C_c(\mathbf{R}^d)$ function $0 \leq g \leq 1$ that is identically 1 on K_ℓ and supported on $K_{\ell+1}$. The distance function used above is defined on any metric space (X, r) as follows: between a point x and a set A,

$$\operatorname{dist}(x, A) = \inf\{r(x, y) : y \in A\},$$

and between two sets A and B,

$$\operatorname{dist}(A, B) = \inf\{r(x, y) : x \in A, y \in B\}.$$

For the collection $\{\phi_{\ell,m} : m \geq 1\}$ we take all products gp where p ranges over real polynomials on $\mathbf{R}^d$ with rational coefficients. This creates a countable set. We check that the functions gp are dense among K_ℓ-supported $C_c(\mathbf{R}^d)$-functions. For a given $f \in C_c(\mathbf{R}^d)$ supported on

K_ℓ, by the Stone-Weierstrass Approximation Theorem (Corollary 4.50 in [17]) there exists a polynomial p with rational coefficients such that

$$\sup_{x\in K_{\ell+1}} |f(x)-p(x)| \le \varepsilon.$$

Since $g(x)\equiv 1$ on K_ℓ, we get

$$|f(x)-g(x)p(x)| = |f(x)-p(x)| \le \varepsilon \quad \text{for } x\in K_\ell.$$

On $K_{\ell+1}\setminus K_\ell$, $f(x)\equiv 0$ and $0\le g(x)\le 1$, so

$$|f(x)-g(x)p(x)| = |g(x)p(x)| \le |p(x)| = |f(x)-p(x)| \le \varepsilon \quad \text{for } x\in K_{\ell+1}\setminus K_\ell.$$

Finally on $K_{\ell+1}^c$ both f and gp vanish. These steps show that $\|f-gp\|_\infty \le \varepsilon$. We have shown that the class of gp is dense among K_ℓ-supported $C_c(\mathbf{R}^d)$-functions.

Note also that the sequence $\{\phi_{\ell,m}\}$ includes the function g itself (the case $p\equiv 1$), so some $\phi_{\ell,m}$ satisfies $\phi_{\ell,m}\ge \mathbf{1}_{K_\ell}$.

Once this has been done for each ℓ, let $\{\phi_j : j\ge 1\}$ be a relabeling of the entire collection $\{\phi_{\ell,m} : \ell, m\ge 1\}$. Define the metric $d_{\mathbf{M}}$ on $\mathbf{M}$ by

$$d_{\mathbf{M}}(\mu,\nu) = \sum_{j=1}^\infty 2^{-j}\Big(1\wedge \Big|\int \phi_j\,d\mu - \int \phi_j\,d\nu\Big|\Big). \tag{A.20}$$

The open sets in $\mathbf{M}$ determined by the metric $d_{\mathbf{M}}$ form the *vague topology* of $\mathbf{M}$.

If necessary, the functions ϕ_j in the metric can be assumed infinitely differentiable. Simply replace the cutoff function g of (A.19) by a suitable convolution

$$\mathbf{1}_H * \psi(x) = \int_H \psi(x-y)\,dy.$$

Take H to be a set such that $K_\ell\subseteq H\subseteq K_{\ell+1}$, $\text{dist}(K_\ell, H^c) > \delta$ and $\text{dist}(H, K_{\ell+1}^c) > \delta$ for some $\delta > 0$. Let ψ be an infinitely differentiable nonnegative function supported on the δ-ball $B(0,\delta)$ centered at the origin, with integral $\int\psi(x)\,dx = 1$. Then the convolution $\mathbf{1}_H*\psi$ is supported on $K_{\ell+1}$, identically one on K_ℓ, and infinitely differentiable.

It is clear that vague convergence $\mu_n\to\mu$ implies $d_{\mathbf{M}}(\mu_n,\mu)\to 0$. Let us show that $d_{\mathbf{M}}(\mu_n,\mu)\to 0$ implies $\int f\,d\mu_n\to\int f\,d\mu$ for an arbitrary $f\in C_c(\mathbf{R}^d)$.

Lemma A.24 *Let $\mu\in\mathbf{M}$, $f\in C_c(\mathbf{R}^d)$ and $\varepsilon>0$. Then there are finite constants $C = C(\mu,f,\varepsilon)$ and $h = h(\mu,f,\varepsilon)$ such that*

$$\Big|\int f\,d\mu - \int f\,d\nu\Big| \le \varepsilon + C d_{\mathbf{M}}(\mu,\nu)$$

for all $\nu\in\mathbf{M}$ such that $d_{\mathbf{M}}(\mu,\nu) < 2^{-h}$.

Proof. Fix ℓ such that f is supported on K_ℓ. Then fix m such that $\phi_m \ge \mathbf{1}_{K_{\ell+1}}$. Given $\varepsilon > 0$, let

$$\delta = \frac{\varepsilon}{2} \cdot \Big(1 + \textstyle\int \phi_m \, d\mu\Big)^{-1}$$

and pick ϕ_j supported on $K_{\ell+1}$ so that $\|\phi_j - f\|_\infty < \delta$. By the triangle inequality

$$\begin{aligned}
\Big|\int f\,d\mu - \int f\,d\nu\Big| &\le \int \big|f - \phi_j\big|\,d\mu + \Big|\int \phi_j\,d\mu - \int \phi_j\,d\nu\Big| \\
&\qquad + \int \big|\phi_j - f\big|\,d\nu \\
&\le \delta\mu(K_{\ell+1}) + \Big|\int \phi_j\,d\mu - \int \phi_j\,d\nu\Big| + \delta\nu(K_{\ell+1}) \\
&\le \delta \int \phi_m\,d\mu + \Big|\int \phi_j\,d\mu - \int \phi_j\,d\nu\Big| + \delta \int \phi_m\,d\nu \\
&\le 2\delta \int \phi_m\,d\mu + \Big|\int \phi_j\,d\mu - \int \phi_j\,d\nu\Big| + \delta\Big|\int \phi_m\,d\mu - \int \phi_m\,d\nu\Big|.
\end{aligned}$$

Let $h = j + m$. Then if $d_{\mathbf{M}}(\mu, \nu) < 2^{-h}$, the last line above is bounded by

$$\varepsilon + (2^j + 2^m\delta)d_{\mathbf{M}}(\mu, \nu).$$

This completes the proof of the Lemma. ∎

This lemma shows that $d_{\mathbf{M}}(\mu_n, \mu) \to 0$ implies $\int f\,d\mu_n \to \int f\,d\mu$. Thereby we have shown that convergence under the metric $d_{\mathbf{M}}$ is the same as the earlier defined vague convergence.

For separate use we retain one point from the proof of the lemma. Given $f \in C_c(\mathbf{R}^d)$, we found ϕ_m such that $|f| \le \|f\|_\infty \cdot \phi_m$, and then

$$\Big|\int f\,d\mu\Big| \le \|f\|_\infty \cdot \Big\{ \int \phi_m\,d\nu + \Big|\int \phi_m\,d\mu - \int \phi_m\,d\nu\Big| \Big\}.$$

In particular, given $f \in C_c(\mathbf{R}^d)$, there exists an $m = m(f)$ with this property: for any $\mu, \nu \in \mathbf{M}$ such that $d_{\mathbf{M}}(\mu, \nu) < 2^{-m}$,

$$\Big|\int f\,d\mu\Big| \le \|f\|_\infty \cdot \Big\{ \int \phi_m\,d\nu + 2^m d_{\mathbf{M}}(\mu, \nu) \Big\}. \tag{A.21}$$

In the remainder of this section we treat two points: (i) a compactness criterion for $\mathbf{M}$, and (ii) the completeness and separability of $d_{\mathbf{M}}$, in other words that $(\mathbf{M}, d_{\mathbf{M}})$ is a Polish space. A set is *precompact*, also called *relatively compact*, if its closure is compact.

Proposition A.25 *Let $U \subseteq \mathbf{M}$. U is precompact iff*

$$\sup_{\mu \in U} \mu(K) < \infty \quad \text{for all compact } K \subseteq \mathbf{R}^d. \tag{A.22}$$

Proof. Assume first that the closure $\bar{U}$ is compact. Given a compact set $K \subseteq \mathbf{R}^d$, pick $f \in C_c(\mathbf{R}^d)$ such that $f \geq \mathbf{1}_K$. The function $\mu \mapsto \int f\,d\mu$ is continuous on $\mathbf{M}$, and consequently bounded on $\bar{U}$. Then

$$\sup_{\mu \in U} \mu(K) \leq \sup_{\mu \in \bar{U}} \int f\,d\mu < \infty.$$

For the converse part, suppose U has the boundedness property (A.22). We first show that an arbitrary sequence in U has a convergent subsequence, although the limit does not have to lie in U since U is not assumed closed. Let $\{\mu_n\}$ be a sequence in U. For each ℓ, the sequence $\{\mu_n(K_\ell)\}$ is bounded by assumption. Use a diagonal argument to pick a subsequence, again denoted by $\{\mu_n\}$, along which

$$\mu_n(K_\ell) \xrightarrow[n\to\infty]{} c_\ell$$

for a finite number c_ℓ, for each ℓ. The sequence c_ℓ is nondecreasing. If $c_\ell = 0$ for all ℓ, the subsequence μ_n converges to the identically zero measure.

Otherwise there exists an index $\bar{\ell}$ such that $c_\ell > 0$ for $\ell \geq \bar{\ell}$. By dropping finitely many terms from the subsequence μ_n, we may assume $\mu_n(K_\ell) > 0$ for all n and $\ell \geq \bar{\ell}$.

Let first $\ell = \bar{\ell}$. Define a probability measure $\bar{\nu}_{\ell,n}$ on K_ℓ by

$$\bar{\nu}_{\ell,n}(B) = \frac{\mu_n(B \cap K_\ell)}{\mu_n(K_\ell)} \quad \text{for Borel sets } B \subseteq K_\ell.$$

The weak topology of probability measures on a compact set is compact. Consequently there exists a subsequence $\{\bar{\nu}_{\ell,n_k}\}$ and a probability measure $\bar{\nu}_\ell$ on K_ℓ such that

$$\int h\,d\bar{\nu}_{\ell,n_k} \xrightarrow[k\to\infty]{} \int h\,d\bar{\nu}_\ell$$

for all continuous functions h on K_ℓ. Define a Borel measure ν_ℓ on $\mathbf{R}^d$ by

$$\nu_\ell(A) = c_\ell \bar{\nu}_\ell(A \cap K_\ell) \quad \text{for Borel } A \subseteq \mathbf{R}^d.$$

Then for any $f \in C_c(\mathbf{R}^d)$ supported on K_ℓ,

$$\int f\,d\mu_{n_k} = \mu_{n_k}(K_\ell) \int_{K_\ell} f\,d\bar{\nu}_{\ell,n_k} \xrightarrow[k\to\infty]{} c_\ell \int_{K_\ell} f\,d\bar{\nu}_\ell = \int f\,d\nu_\ell.$$

Starting with the subsequence μ_{n_k} thus constructed for ℓ, repeat the step for $\ell + 1$, and so on, inductively for all $\ell \geq \bar{\ell}$. Then a diagonal argument gives a single subsequence $\{\mu_{n_k}\}$ and a collection of Borel measures $\{\nu_\ell : \ell \geq \bar{\ell}\}$ on $\mathbf{R}^d$ such that

$$\int f \, d\mu_{n_k} \xrightarrow[k\to\infty]{} \int f \, d\nu_\ell \tag{A.23}$$

for any $f \in C_c(\mathbf{R}^d)$ supported on K_ℓ. The measures $\{\nu_\ell\}$ are consistent in the sense that if $f \in C_c(\mathbf{R}^d)$ is supported on K_{ℓ_0}, then for all $\ell \geq \ell_0$,

$$\int f \, d\nu_\ell = \lim_{k\to\infty} \int f \, d\mu_{n_k} = \int f \, d\nu_{\ell_0}.$$

Then we can uniquely define a measure $\mu \in \mathbf{M}$ by setting, for $f \in C_c(\mathbf{R}^d)$,

$$\int f \, d\mu = \int f \, d\nu_\ell$$

for any ℓ such that f is supported on K_ℓ. The vague convergence $\mu_{n_k} \to \mu$ is already contained in the limits established above.

We have shown that any sequence in U has a subsequence that converges vaguely to some measure in $\mathbf{M}$.

In a metric space, a set A is compact iff it is sequentially compact, which means that every sequence in A has a subsequence that converges to a limit in A. Given a sequence $\{\mu_n\}$ in $\bar{U}$, pick a sequence $\{\mu'_n\}$ in U such that $d_{\mathbf{M}}(\mu_n, \mu'_n) < n^{-1}$. By the above argument, there is a convergent subsequence $\mu'_{n_k} \to \mu$. The limit μ of a sequence in U lies in the closure $\bar{U}$. Since $d_{\mathbf{M}}(\mu_{n_k}, \mu) \leq d_{\mathbf{M}}(\mu'_{n_k}, \mu) + n_k^{-1}$, also $\mu_{n_k} \to \mu$. We have shown that $\bar{U}$ is sequentially compact. ∎

Proposition A.26 $(\mathbf{M}, d_{\mathbf{M}})$ *is a complete separable metric space.*

Proof. We leave it to the reader to check that a countable dense set in $\mathbf{M}$ is given by measures

$$\nu = \sum_{i=1}^{m} b_i \delta_{x^i}$$

where $m < \infty$, $b_1, \dots, b_m$ are positive rationals, and $x^1, \dots, x^m$ are points in $\mathbf{R}^d$ with rational coordinates.

To show completeness, let $\{\mu_n\}$ be a Cauchy sequence in the $d_{\mathbf{M}}$ metric. Then $\{\int \phi_j \, d\mu_n\}$ is a Cauchy sequence for each ϕ_j. For any compact set K, $\mathbf{1}_K$ is dominated by some ϕ_j by the original choice of the functions $\{\phi_j\}$. Consequently the sequence $\{\mu_n(K)\}$ is bounded. By

Proposition A.25 the set $\{\mu_n\}$ is precompact, and so there is a $d_{\mathbf{M}}$-convergent subsequence $\mu_{n_k} \to \mu$. Then by the Cauchy property the full sequence μ_n converges to μ. This establishes the completeness of the metric $d_{\mathbf{M}}$. ∎

A.11 Heat equation

We derive here existence and uniqueness theorems for the linear partial differential equations that arise in the hydrodynamic limits of symmetric processes.

Let $\Gamma = (\gamma_{i,j})_{1\le i,j\le d}$ be a real symmetric matrix with nonnegative eigenvalues. Define a differential operator A by

$$Av = \sum_{1\le i,j\le d} \gamma_{i,j} v_{x_i,x_j}. \tag{A.24}$$

Given a bounded measurable function v_0 on $\mathbf{R}^d$, consider the initial value problem

$$v_t = \tfrac{1}{2}Av \quad \text{in } \mathbf{R}^d \times (0,T), \quad v(x,0) = v_0(x) \text{ for } x \in \mathbf{R}^d. \tag{A.25}$$

We say a bounded measurable function v on $\mathbf{R}^d \times [0,\infty)$ is a *weak solution* of (A.25) if

$$\int_{\mathbf{R}^d} \phi(x)v(x,t)\,dx - \int_{\mathbf{R}^d} \phi(x)v_0(x)\,dx - \tfrac{1}{2}\int_0^t \int_{\mathbf{R}^d} A\phi(x)v(x,s)\,dx\,ds = 0 \tag{A.26}$$

for all compactly supported, infinitely differentiable test functions ϕ. Let $C_c^\infty(\mathbf{R}^d)$ denote the space of such functions.

We can find a weak solution in terms of a multidimensional Gaussian distribution. Let us represent elements of $\mathbf{R}^d$ as column vectors and let a prime denote transposition. An $\mathbf{R}^d$-valued random vector $X = [X_1, \dots, X_d]'$ has the $\mathcal{N}(0,\Gamma)$ distribution if its characteristic function is given by

$$Ee^{i\langle\theta,X\rangle} = \exp\bigl\{-\tfrac{1}{2}\langle\theta,\Gamma\theta\rangle\bigr\} \qquad \text{for } \theta \in \mathbf{R}^d.$$

Here

$$\langle\theta, X\rangle = \sum_{i=1}^d \theta_i X_i = \theta' X$$

is the Euclidean inner product on $\mathbf{R}^d$. An $\mathcal{N}(0,\Gamma)$ random vector can be manufactured from independent standard normal random variables, see Exercise A.10.

Theorem A.27 *Suppose v_0 is a bounded measurable function on $\mathbf{R}^d$ and let X have $\mathcal{N}(0,\Gamma)$ distribution. Then*

$$v(x,t) = Ev_0(x - t^{1/2}X) \tag{A.27}$$

is a weak solution of (A.25) *in the sense* (A.26).

Proof. Noting that

$$\begin{aligned}\int_{\mathbf{R}^d} A\phi(x)v(x,s)\,dx &= \sum_{i,j}\gamma_{i,j}\,E\int_{\mathbf{R}^d} v_0(x-s^{1/2}X)\phi_{x_i,x_j}(x)\,dx\\ &= \sum_{i,j}\gamma_{i,j}\,E\int_{\mathbf{R}^d} v_0(x)\phi_{x_i,x_j}(x+s^{1/2}X)\,dx\\ &= \int_{\mathbf{R}^d} v_0(x)\Big\{\sum_{i,j}\gamma_{i,j}\,E\phi_{x_i,x_j}(x+s^{1/2}X)\Big\}\,dx\end{aligned}$$

and by doing a simple reorganization in the first integral of (A.26), we see that the requirement is to show

$$\int_{\mathbf{R}^d} dx\, v_0(x)\Big\{E\phi(x+t^{1/2}X)-\phi(x)-\tfrac12\int_0^t\sum_{i,j}\gamma_{i,j}E\phi_{x_i,x_j}(x+s^{1/2}X)\,ds\Big\}=0.$$

In other words, we need to check that

$$E\phi(x+t^{1/2}X)-\phi(x)-\tfrac12\int_0^t\sum_{i,j}\gamma_{i,j}E\phi_{x_i,x_j}(x+s^{1/2}X)\,ds=0$$

for $\phi\in C_c^\infty(\mathbf{R}^d)$, $x\in\mathbf{R}^d$ and $t\ge 0$. We leave this to the reader. See Exercise A.11. ∎

The weak solution found here may or may not be differentiable, depending on v_0. See Exercise A.12.

We turn to the uniqueness theorem. For this we assume Γ nonsingular. Then the $\mathcal{N}(0,\Gamma)$ distribution has a density on $\mathbf{R}^d$ (Exercise A.10). The function $v(x,t)=Ev_0(x+t^{1/2}X)$ is given by

$$v(x,0)=v_0(x) \tag{A.28}$$

and for $t>0$

$$v(x,t)=(2\pi t)^{-d/2}(\det\Gamma)^{-1/2}\int_{\mathbf{R}^d} v_0(y)\exp\big\{-\tfrac{1}{2t}\langle x-y,\Gamma^{-1}(x-y)\rangle\big\}\,dy \tag{A.29}$$

We generalize the type of solution from a function to a measure-valued path. Suppose $\alpha:[0,\infty)\to\mathbf{M}$ is a vaguely continuous path in the space $\mathbf{M}$ of Radon measures on $\mathbf{R}^d$. Write $\alpha(t,dx)$ for the measure on $\mathbf{R}^d$ that is the value of α at time t, and

$$\alpha(t,\phi)=\int_{\mathbf{R}^d}\phi(x)\,\alpha(t,dx)$$

for the intergral. Let us say that such a path α is a weak solution of (A.25) if the initial condition

$$\alpha(0, dx) = v_0(x)dx \tag{A.30}$$

is satisfied, and

$$\alpha(t, \phi) - \int_{\mathbf{R}^d} \phi(x) v_0(x)\, dx - \tfrac{1}{2} \int_0^t \alpha(s, A\phi)\, ds = 0 \tag{A.31}$$

for all $\phi \in C_c^\infty(\mathbf{R}^d)$.

Here is the uniqueness theorem for weak solutions that we need for the hydrodynamic limit of symmetric exclusion. We make a boundedness assumption on the measures $\alpha(t)$ in terms of open Euclidean balls $B(x, r)$ of radius r centered at x in $\mathbf{R}^d$. This assumption is stronger than needed but easily satisfied in our application.

Theorem A.28 *Suppose v_0 is a bounded measurable function on $\mathbf{R}^d$. Suppose α is a vaguely continuous $\mathbf{M}$-valued path that satisfies* (A.30) *and* (A.31) *for all test functions $\phi \in C_c^\infty(\mathbf{R}^d)$. Assume that for some constants $0 < C, r_0 < \infty$,*

$$\alpha(t, B(x, r)) \le C r^d \tag{A.32}$$

for all $r \ge r_0$, $x \in \mathbf{R}^d$, and $t \ge 0$. Define $v(x, t)$ by (A.28) *and* (A.29)*. Then $\alpha(t, dx) = v(x, t)dx$ for all $t \ge 0$.*

This theorem will be proved after several steps. We begin with a textbook result for the heat equation. For $0 < T < \infty$, let $C_1^2(\mathbf{R}^d \times (0, T])$ denote the class of functions u defined for $(x, t) \in \mathbf{R}^d \times (0, T]$ such that the partial derivatives u_t, u_{x_i}, and u_{x_i, x_j} $(1 \le i, j \le d)$ exist and are continuous on $\mathbf{R}^d \times (0, T)$, and can be extended continuously to $\mathbf{R}^d \times (0, T]$. For $x = (x_1, \ldots, x_d) \in \mathbf{R}^d$, $|x|$ denotes the Euclidean norm $|x| = \left(x_1^2 + \cdots + x_d^2\right)^{1/2}$. The Laplacian is the operator $\Delta u = u_{x_1,x_1} + u_{x_2,x_2} + \cdots + u_{x_d,x_d}$.

Let u_0 be a given bounded continuous function on $\mathbf{R}^d$. Consider the initial value problem for the heat equation

$$u_t = \tfrac{1}{2}\Delta u \quad \text{in} \quad \mathbf{R}^d \times (0, T), \tag{A.33}$$

$$u(x, 0) = u_0(x) \quad \text{for} \quad x \in \mathbf{R}^d. \tag{A.34}$$

The relevant distribution is now the standard d-dimensional Gaussian, so we define a function u on $\mathbf{R}^d \times [0, \infty)$ by $u(x, 0) = u_0(x)$, and

$$u(x, t) = (2\pi t)^{-d/2} \int_{\mathbf{R}^d} e^{-\frac{1}{2t}|x-y|^2} u_0(y)\, dy \quad \text{for } (x, t) \in \mathbf{R}^d \times (0, \infty). \tag{A.35}$$

Theorem A.29 *The function u defined above is continuous on $\mathbf{R}^d \times [0,\infty)$, infinitely differentiable on $\mathbf{R}^d \times (0,\infty)$, and satisfies $u_t = \frac{1}{2}\Delta u$ on $\mathbf{R}^d \times (0,\infty)$.*

Let $0 < T < \infty$. There is no other solution u of (A.33)–(A.34) *of class $C_1^2(\mathbf{R}^d \times (0,T])$, continuous on $\mathbf{R}^d \times [0,T]$, and satisfying the bound*

$$|u(x,t)| \le Ce^{B|x|^2} \quad \textit{for } (x,t) \in \mathbf{R}^d \times [0,T]$$

for some finite constants B, C.

This theorem is a combination of Theorems 1 and 7 of Section 2.3 in [14]. The uniqueness part is proved there via the maximum principle.

Next we extend Theorem A.29 to the operator A defined in (A.24), so consider the initial value problem (A.25). From now on we assume that the matrix Γ is nonsingular. Since we already assumed Γ has nonnegative eigenvalues, it is equivalent to require strictly positive eigenvalues. Yet another equivalent statement is that there exists a constant $\theta > 0$ such that

$$\sum_{i,j} \gamma_{i,j} x_i x_j \ge \theta |x|^2$$

for all $x \in \mathbf{R}^d$. Equation (A.25) is then called *uniformly parabolic.*

Theorem A.30 *Suppose v_0 is a bounded continuous function on $\mathbf{R}^d$. Then the function v defined in* (A.28)–(A.29) *is continuous on $\mathbf{R}^d \times [0,\infty)$, infinitely differentiable on $\mathbf{R}^d \times (0,\infty)$, and satisfies $v_t = \frac{1}{2}Av$ on $\mathbf{R}^d \times (0,\infty)$.*

Let $0 < T < \infty$. There is no other solution v of (A.25) *of class $C_1^2(\mathbf{R}^d \times (0,T])$, continuous on $\mathbf{R}^d \times [0,T]$, and satisfying the bound*

$$|v(x,t)| \le Ce^{B|x|^2} \quad \textit{for } (x,t) \in \mathbf{R}^d \times [0,T] \tag{A.36}$$

for some finite constants B, C.

Proof. That v has the continuity and smoothness claimed, and satisfies $v_t = \frac{1}{2}Av$ on $\mathbf{R}^d \times (0,\infty)$, can be verified from the definition (A.29). These properties come also from a direct relation of (A.29) to the heat semigroup (A.35), which will give us the uniqueness.

As a real symmetric matrix, Γ can be diagonalized as $\Gamma = H\Lambda H'$ where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix}$$

is the diagonal matrix of eigenvalues, and the columns of $H = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_d]$ are an orthonormal set of eigenvectors of Γ. H is an orthogonal matrix, which means that $H'H = HH' = I$.

Since the eigenvalues of Γ are assumed strictly positive, the matrices

$$\Lambda^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & 0 & \cdots & 0 \\ 0 & \lambda_2^{1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d^{1/2} \end{bmatrix} \quad \text{and} \quad \Lambda^{-1/2} = \begin{bmatrix} \lambda_1^{-1/2} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d^{-1/2} \end{bmatrix},$$

are nonsingular and inverses of each other. From

$$\Gamma^{-1} = H\Lambda^{-1}H' = H\Lambda^{-1/2}\Lambda^{-1/2}H' = (\Lambda^{-1/2}H')'(\Lambda^{-1/2}H')$$

the quadratic form can be re-expressed as

$$(x-y)'\Gamma^{-1}(x-y) = (x-y)'(\Lambda^{-1/2}H')'(\Lambda^{-1/2}H')(x-y) = |\Lambda^{-1/2}H'(x-y)|^2.$$

Thereby the definition of $v(x,t)$ can be rewritten as

$$v(x,t) = (2\pi t)^{-d/2}(\det\Gamma)^{-1/2}\int_{\mathbf{R}^d} v_0(y)\exp\bigl\{-\tfrac{1}{2t}|\Lambda^{-1/2}H'(x-y)|^2\bigr\}\,dy.$$

After a linear change of variable in the integral,

$$v(x,t) = (2\pi t)^{-d/2}\int_{\mathbf{R}^d} v_0(H\Lambda^{1/2}y)\exp\bigl\{-\tfrac{1}{2t}|\Lambda^{-1/2}H'x - y|^2\bigr\}\,dy. \tag{A.37}$$

Set $u(x,t) = v(H\Lambda^{1/2}x, t)$. Then (A.37) becomes a special case of (A.35) with $u_0(x) = v_0(H\Lambda^{1/2}x)$.

By Theorem A.29, u is continuous on $\mathbf{R}^d \times [0,\infty)$, infinitely differentiable on $\mathbf{R}^d \times (0,\infty)$, and satisfies $u_t = \frac{1}{2}\Delta u$. Since $v(x,t) = u(\Lambda^{-1/2}H'x, t)$, v has the regularity properties of u, and the chain rule shows $v_t = \frac{1}{2}Av$.

Let $\widetilde{v}$ be an arbitrary solution of (A.25) of class $C_1^2(\mathbf{R}^d \times (0,T]) \cap C(\mathbf{R}^d \times [0,T])$ and exponentially bounded as in (A.36). Define $\widetilde{u}(x,t) = \widetilde{v}(H\Lambda^{1/2}x, t)$. Then

$$\widetilde{u}(x,0) = \widetilde{v}(H\Lambda^{1/2}x, 0) = v_0(H\Lambda^{1/2}x) = u_0(x).$$

Differentiation shows that $\widetilde{u}_t = \frac{1}{2}\Delta\widetilde{u}$. The uniqueness part of Theorem A.29 implies that $\widetilde{u} = u$, from which follows $\widetilde{v} = v$.

Note that we used several times in the proof the nonsingularity of the matrix $H\Lambda^{1/2}$, which is a consequence of the assumption $\lambda_i > 0$. ∎

Proof of Theorem A.28. To apply Theorem A.30 we need some regularity for the solution, which we get from convolution with a smooth approximate identity. Let $f \in C_c^\infty(\mathbf{R}^d)$ be nonnegative, symmetric [meaning $f(x) = f(-x)$], supported on the unit ball $B(0,1)$, and have integral $\int f\,dx = 1$. For $\varepsilon > 0$, let $f^\varepsilon(x) = \varepsilon^{-d} f(x/\varepsilon)$. If we think of f^ε as a probability distribution on $\mathbf{R}^d$, f^ε converges weakly to a point mass at the origin as $\varepsilon \to 0$.

Define

$$v^\varepsilon(x,t) = [f^\varepsilon * \alpha(t)](x) = \int_{\mathbf{R}^d} f^\varepsilon(x-y)\,\alpha(t,dy).$$

First we argue that $v^\varepsilon \in C_1^2(\mathbf{R}^d \times (0,T]) \cap C(\mathbf{R}^d \times [0,T])$ for any T.

Suppose for the moment g is any compactly supported, continuous function on $\mathbf{R}^d$, and set

$$\bar{g}(x,t) = [g * \alpha(t)](x) = \int_{\mathbf{R}^d} g(x-y)\,\alpha(t,dy). \tag{A.38}$$

Fix (x,t) and consider a nearby point (y,s).

$$\begin{aligned}
&|\bar{g}(x,t) - \bar{g}(y,s)| \le |\bar{g}(x,t) - \bar{g}(x,s)| + |\bar{g}(x,s) - \bar{g}(y,s)| \\
&\le \left| \int g(x-z)\,\alpha(t,dz) - \int g(x-z)\,\alpha(s,dz) \right| + \int |g(x-z) - g(y-z)|\,\alpha(s,dz).
\end{aligned}$$

The last line above tends to zero as $(y,s) \to (x,t)$, for the following reasons. By the assumption of continuity of $t \mapsto \alpha(t)$ and the definition of vague convergence,

$$\int g(x-z)\,\alpha(s,dz) \underset{s\to t}{\longrightarrow} \int g(x-z)\,\alpha(t,dz).$$

Suppose $y \in B(x,1)$. There is a fixed ball $B(0,r)$ such that both $z \mapsto g(x-z)$ and $z \mapsto g(y-z)$ are supported on $B(0,r)$ for all $y \in B(x,1)$. By assumption, $\alpha(s, B(0,r)) \le Cr^d$ for all s (increase r if necessary for this). Since g is uniformly continuous, we get

$$\lim_{\delta\to 0} \sup_{y\in B(x,\delta)} \sup_{s\ge 0} \int |g(x-z) - g(y-z)|\,\alpha(s,dz) \le \lim_{\delta\to 0} \sup_{|z-w|\le\delta} |g(z) - g(w)| \cdot Cr^d = 0.$$

This shows $\bar{g}$ continuous on $\mathbf{R}^d \times [0,\infty)$.

By the smoothness assumption on f and the bound on $\alpha(t)$, spatial derivatives of v^ε can be taken inside the integral. For example,

$$v^\varepsilon_{x_i,x_j}(x,t) = \int (f^\varepsilon)_{x_i,x_j}(x-y)\,\alpha(t,dy).$$

By taking $g = (f^\varepsilon)_{x_i,x_j}$ in (A.38), the continuity of $\bar{g}$ is the same as the continuity of $v^\varepsilon_{x_i,x_j}$. This gives the continuity in (x,t) of all spatial partial derivatives of v^ε.

To get the continuity of the time derivative v_t we use the equation. Replace the test function ϕ in (A.31) by $f^\varepsilon * \phi$, and change the order of integration. Note that $A(f^\varepsilon * \phi) = f^\varepsilon * A\phi$. This gives

$$\int \phi(x) v^\varepsilon(x,t)\,dx - \int \phi(x) v^\varepsilon(x,0)\,dx - \tfrac{1}{2}\int_0^t ds \int dx\, A\phi(x)\, v^\varepsilon(x,s) = 0.$$

Integrate by parts in the last term to get

$$\int \phi(x) v^\varepsilon(x,t)\,dx - \int \phi(x) v^\varepsilon(x,0)\,dx - \tfrac{1}{2}\int dx\,\phi(x) \int_0^t ds\, Av^\varepsilon(x,s) = 0.$$

Since this is true for arbitrary compactly supported smooth ϕ, it follows that

$$v^\varepsilon(x,t) - v^\varepsilon(x,0) - \tfrac{1}{2}\int_0^t ds\, Av^\varepsilon(x,s) = 0 \tag{A.39}$$

for Lebesgue almost every x. By continuity, this is true for all x.

$$Av^\varepsilon(x,s) = \int Af^\varepsilon(x-y)\,\alpha(s,dy)$$

is continuous in s by the vague continuity of $\alpha(s)$. Differentiating (A.39) gives $v_t^\varepsilon(x,t) = Av^\varepsilon(x,t)$, and we see that v^ε has a continuous time derivative. We have shown that $v^\varepsilon \in C_1^2(\mathbf{R}^d \times (0,\infty))$. The continuity of v^ε down to the $t=0$ boundary was already part of the continuity argument above.

The equation $v_t^\varepsilon(x,t) = \frac{1}{2}Av^\varepsilon(x,t)$ was derived as part of the previous paragraph. Boundedness of v^ε follows from the assumption on $\alpha(t)$:

$$|v^\varepsilon(x,t)| \le \|f^\varepsilon\|_\infty\, \alpha(t, B(x,\varepsilon)) \le C\varepsilon^{-d}\|f\|_\infty. \tag{A.40}$$

Thus Theorem A.30 applies to v^ε. Abbreviate

$$q_t(z) = (2\pi t)^{-d/2}(\det\Gamma)^{-1/2}\exp\bigl\{-\tfrac{1}{2t}|\Lambda^{-1/2}H'z|^2\bigr\}.$$

By the definition of v^ε and the initial condition on $\alpha(0)$,

$$v^\varepsilon(y,0) = \int f^\varepsilon(y-z)\,\alpha(0,dz) = \int f^\varepsilon(y-z)\,v_0(z)\,dz.$$

By the uniqueness part of Theorem A.30 v^ε must be given by formula (A.29) with initial function $v^\varepsilon(y,0)$. Use this and symmetry $f(y-z) = f(z-y)$ to write

$$\begin{aligned} v^\varepsilon(x,t) &= \int v^\varepsilon(y,0) q_t(x-y)\,dy = \int dz\, v_0(z) \int dy\, f^\varepsilon(y-z) q_t(x-y) \\ &= \int dz\, v_0(z) \int dy\, f^\varepsilon(z-y) q_t(x-y). \end{aligned}$$

This already implies that $v^\varepsilon(x,t) \to v(x,t)$ as $\varepsilon \to 0$ because by standard results for approximate identities (Theorem 8.14(a) in [17]), for a fixed (x,t)

$$\big[f^\varepsilon * q_t(x-\cdot)\big](z) \underset{\varepsilon\to 0}{\longrightarrow} q_t(x-z) \quad \text{in } L^1(\mathbf{R}^d).$$

Our goal is to derive the equality

$$\int \phi(x)\,\alpha(t,dx) = \int \phi(x) v(x,t)\,dx$$

by showing that both sides are limits of $\int \phi(x) v^\varepsilon(x,t)\,dx$ as $\varepsilon \to 0$. Hence we need to integrate over the difference $v^\varepsilon(x,t) - v(x,t)$ multiplied by a test function $\phi(x)$. The crude bound (A.40) is useless as $\varepsilon \to 0$, so we derive a bound for the difference $v^\varepsilon(x,t) - v(x,t)$.

Consider $t > 0$ fixed for the remainder of the proof. In the next calculation, use the boundedness of v_0 and $\int dy\, f^\varepsilon(z-y) = 1$, do a change of variable $y = z - \varepsilon w$ in the inner dy-integral, note that f is supported on $B(0,1)$ and again $\int f(w)\,dw = 1$, and finally one more change of variable.

$$\begin{aligned}
&|v^\varepsilon(x,t) - v(x,t)| \\
&\quad = \left| \int dz\, v_0(z) \int dy\, f^\varepsilon(z-y) q_t(x-y) - \int dz\, v_0(z) q_t(x-z) \right| \\
&\quad \le \|v_0\|_\infty \cdot \int dz \left| \int dy\, f^\varepsilon(z-y)\big(q_t(x-y) - q_t(x-z)\big) \right| \\
&\quad = \|v_0\|_\infty \cdot \int dz \left| \int dw\, f(w)\big(q_t(x-z+\varepsilon w) - q_t(x-z)\big) \right| \\
&\quad \le \|v_0\|_\infty \cdot \int dz \sup_{w\in B(0,1)} \big| q_t(x-z+\varepsilon w) - q_t(x-z) \big| \\
&\quad \le \|v_0\|_\infty \cdot \int dz \sup_{w\in B(0,1)} \big| q_t(z+\varepsilon w) - q_t(z) \big| \\
&\quad \le C\varepsilon.
\end{aligned} \tag{A.41}$$

The last inequality follows for example by applying the mean value theorem to the integrand (Exercise A.13).

From above we get the uniform estimate

$$\sup_{x\in\mathbf{R}^d} |v^\varepsilon(x,t) - v(x,t)| \le C\varepsilon. \tag{A.42}$$

For $\phi \in C_c^\infty(\mathbf{R}^d)$, this implies

$$\int \phi(x) v^\varepsilon(x,t)\,dx \longrightarrow \int \phi(x) v(x,t)\,dx \qquad \text{as } \varepsilon \to 0.$$

Note that the integrals are actually restricted to a fixed compact set that supports ϕ, and so (A.42) is sufficient for the convergence. On the other hand,

$$\int \phi(x)v^\varepsilon(x,t)\,dx = \int \alpha(t,dy)\int dx\, f^\varepsilon(x-y)\phi(x) \underset{\varepsilon\to 0}{\longrightarrow} \int \alpha(t,dy)\phi(y).$$

We used the convergence $f^\varepsilon * \phi \to \phi$ that happens uniformly and supported by a fixed compact set (Theorem 8.14(b) in [17]). Comparing these two limits for all test functions implies that $\alpha(t,dx) = v(x,t)dx$. ∎

Exercise A.10 *Multivariate Gaussian distributions.* Let Γ be a real symmetric nonnegative definite matrix. Let $\Lambda = \mathrm{diag}[\lambda_1,\dots,\lambda_d]$ be the diagonal matrix of eigenvalues of Γ, and $\Gamma = H\Lambda H'$ an orthogonal diagonalization of Γ. Let $Z_1,\dots,Z_d$ be i.i.d. standard normal random variables, in other words $EZ_i = 0$ and $E[Z_i^2] = 1$. Set $Y_i = \lambda_i Z_i$. In particular if $\lambda_i = 0$ then Y_i is identically zero. Check that $Y = [Y_1,\dots,Y_d]'$ has $\mathcal{N}(0,\Lambda)$-distribution. Define X by the matrix product $X = HY$. Then X has $\mathcal{N}(0,\Gamma)$-distribution.

If Γ is nonsingular, X has density

$$f(x) = (2\pi)^{-d/2}(\det\Gamma)^{-1/2}\exp\bigl\{-\tfrac12\langle x,\Gamma^{-1}x\rangle\bigr\}\,.$$

Exercise* A.11 *Completion of the proof of Theorem A.27.* Check that for $X \sim \mathcal{N}(0,\Gamma)$ and $\phi \in C_c^\infty(\mathbf{R}^d)$,

$$E\phi(x+t^{1/2}X) - \phi(x) - \tfrac12\int_0^t \sum_{i,j}\gamma_{i,j}E\phi_{x_i,x_j}(x+s^{1/2}X)\,ds = 0 \tag{A.43}$$

for all $(x,t)\in\mathbf{R}^d\times[0,\infty)$.

Here is one approach. First check (A.43) by explicit computation for $\phi^\theta(x) = \exp\{i\langle\theta,x\rangle\}$. Then consider the Fourier transform

$$\hat f(\xi) = \int_{\mathbf{R}^d} f(y)e^{-2\pi i\langle y,\xi\rangle}\,dy = \int_{\mathbf{R}^d} f(y)\phi^{-2\pi y}(\xi)\,dy$$

of a function f in Schwartz space $\mathcal{S}$. (We follow here the conventions of Folland [17]). By the case of ϕ^θ,

$$\begin{aligned} E\hat f(x+t^{1/2}X) &= \int f(y)E\phi^{-2\pi y}(x+t^{1/2}X)\,dy \\ &= \int f(y)\phi^{-2\pi y}(x)\,dy + \int dy\, f(y)\,\tfrac12\int_0^t\sum_{i,j}\gamma_{i,j}E\phi^{-2\pi y}_{x_i,x_j}(x+s^{1/2}X)\,ds. \end{aligned}$$

The last line above can be transformed into

$$\hat{f}(x) + \tfrac{1}{2}\int_0^t \sum_{i,j} \gamma_{i,j} E\Big[(\hat{f})_{x_i,x_j}(x + s^{1/2}X)\Big]\, ds$$

and this shows that (A.43) holds for $\hat{f}$ if $f \in \mathcal{S}$. It remains only to note that this covers all of $C_c^\infty(\mathbf{R}^d)$ because the Fourier transform is an isomorphism on $\mathcal{S}$ [17, Corollary 8.28].

Exercise A.12 Let $a > 0$, and

$$\Gamma = \begin{bmatrix} 1 & a \\ a & a^2 \end{bmatrix}.$$

Use Theorem A.27 to solve the initial value problem (A.25) on $\mathbf{R}^2 \times [0, \infty)$ with initial data $v_0(y_1, y_2) = g(y_1 - a^{-1}y_2)$ for a given bounded function g on $\mathbf{R}$. [Answer: $v((x_1, x_2), t) = g(x_1 - a^{-1}x_2)$.] Check that if $g \in C^2(\mathbf{R})$, then the solution is classical, in other words the derivatives v_t and v_{x_i,x_j} exist and satisfy $v_t = \frac{1}{2}Av$.

Exercise* A.13 *Completion of the proof of inequality (A.41).* Show that

$$\int \sup_{w \in B(0,1)} \big| q_t(z + \varepsilon w) - q_t(z) \big|\, dz \le C\varepsilon$$

for a constant C that depends on t and the matrix Γ but not on ε. To simplify the integral, start with a change of variable that converts the kernel $q_t(z)$ into a standard normal density.

A.12 Hamilton-Jacobi equations

We discuss here a first-order partial differential equation of the type

$$u_t + f(\nabla u) = 0\,, \qquad u|_{t=0} = u_0, \tag{A.44}$$

where a real-valued initial function u_0 on $\mathbf{R}^d$ is given, and the solution $u(x,t)$ on $\mathbf{R}^d \times [0, \infty)$ is sought. $u_t = \partial u / \partial t$ is the time derivative of u, and $\nabla u = (u_{x_1}, \dots, u_{x_d})$ is the gradient with respect to spatial variables. Equation (A.44) is of Hamilton-Jacobi type. To understand its solutions one is forced to consider weak solutions u for which the derivatives do not exist at every point. This type of equation appeared as the hydrodynamic limit for the interface model studied in Chapter 9. Here we prove results that are not particular to the case in Chapter 9 but are needed there.

We assume that the function f is $[-\infty, \infty)$-valued, upper semicontinuous, and concave on $\mathbf{R}^d$. Upper semicontinuity means that the sets $\{f \geq s\}$ are closed for all real s, or equivalently

$$\limsup_{\rho \to \lambda} f(\rho) \leq f(\lambda) \qquad \text{for all } \lambda \in \mathbf{R}^d.$$

Concavity means that

$$f(s\rho + (1-s)\lambda) \geq sf(\rho) + (1-s)f(\lambda) \qquad \text{for all } \rho, \lambda \in \mathbf{R}^d \text{ and } 0 < s < 1.$$

The concave conjugate f^* of f is defined by

$$f^*(x) = \inf_{\rho \in \mathbf{R}^d} \{x \cdot \rho - f(\rho)\} \qquad \text{for } x \in \mathbf{R}^d.$$

Then f^* is again concave and upper semicontinuous, and f is its own double dual:

$$f(\rho) = \inf_{x \in \mathbf{R}^d} \{x \cdot \rho - f^*(x)\} \qquad \text{for } \rho \in \mathbf{R}^d.$$

Let u_0 be a real-valued function on $\mathbf{R}^d$, and define

$$u(x,t) = \sup_{y \in \mathbf{R}^d} \left\{ u_0(y) + tf^*\left(\frac{x-y}{t}\right) \right\} \tag{A.45}$$

for $(x,t) \in \mathbf{R}^d \times (0, \infty)$. For $t = 0$ set $u(x,0) = u_0(x)$.

Lemma A.31 *For all* $0 < s < t$ *and* $x \in \mathbf{R}^d$,

$$u(x,t) = \sup_{y \in \mathbf{R}^d} \left\{ u(y,s) + (t-s)f^*\left(\frac{x-y}{t-s}\right) \right\} \tag{A.46}$$

Proof. Let β be the quantity on the right-hand side (A.46). By (A.45) concavity,

$$\begin{aligned} \beta &= \sup_{y,z} \left\{ u_0(z) + sf^*\left(\frac{y-z}{s}\right) + (t-s)f^*\left(\frac{x-y}{t-s}\right) \right\} \\ &\leq \sup_{y,z} \left\{ u_0(z) + tf^*\left(\frac{x-z}{t}\right) \right\} = u(x,t). \end{aligned}$$

Let $c < u(x,t)$. Pick y so that

$$u_0(y) + tf^*\left(\frac{x-y}{t}\right) \geq c.$$

Let

$$z = \frac{s}{t}x + \left(1 - \frac{s}{t}\right) y.$$

Then

$$\frac{x-y}{t} = \frac{x-z}{t-s} = \frac{z-y}{s},$$

and by (A.45) applied to $u(z,s)$,

$$\begin{aligned} \beta &\geq u(z,s) + (t-s)f^*\left(\frac{x-z}{t-s}\right) \geq u_0(y) + sf^*\left(\frac{z-y}{s}\right) + (t-s)f^*\left(\frac{x-z}{t-s}\right) \\ &= u_0(y) + tf^*\left(\frac{x-y}{t}\right) \geq c. \end{aligned}$$

Let $c \nearrow u(x,t)$. ∎

Lemma A.32 *Assume that for each x the supremum in* (A.45) *is attained at some y. Suppose u is differentiable at $(x,t) \in \mathbf{R}^d \times (0,\infty)$. Then*

$$u_t(x,t) + f(\nabla u(x,t)) = 0.$$

Proof. For $z \in \mathbf{R}^d$ and $\delta > 0$,

$$\begin{aligned} u(x+\delta z, t+\delta) &= \sup_y \left\{ u(y,t) + \delta f^*\left(\frac{x+\delta z - y}{\delta}\right)\right\} \\ &\geq u(x,t) + \delta f^*(z). \end{aligned}$$

From this

$$\delta^{-1}\left\{u(x+\delta z, t+\delta) - u(x,t)\right\} \geq f^*(z),$$

and then after $\delta \searrow 0$ by the differentiability assumption,

$$u_t(x,t) + z \cdot \nabla u(x,t) - f^*(z) \geq 0.$$

Since this was valid for all z,

$$u_t(x,t) + f(\nabla u(x,t)) = u_t(x,t) + \inf_z\{z \cdot \nabla u(x,t) - f^*(z)\} \geq 0.$$

To get the opposite inequality, choose y so that $u(x,t) = u_0(y) + tf^*((x-y)/t)$. For $\delta > 0$, let

$$z = \frac{t-\delta}{t}x + \frac{\delta}{t}y = x - \delta \cdot \frac{x-y}{t}.$$

Then

$$\frac{x-y}{t} = \frac{z-y}{t-\delta}\,,$$

and

$$\begin{aligned} u(x,t) - u(z,t-\delta) &\le \left\{ u_0(y) + tf^*\left(\frac{x-y}{t}\right)\right\} - \left\{ u_0(y) + (t-\delta)f^*\left(\frac{z-y}{t-\delta}\right)\right\} \\ &= \delta f^*\left(\frac{x-y}{t}\right). \end{aligned}$$

Consequently

$$u_t(x,t) + \frac{x-y}{t}\cdot\nabla u(x,t) = \lim_{\delta\to 0} \delta^{-1}\left\{u(x,t) - u(z,t-\delta)\right\} \le f^*\left(\frac{x-y}{t}\right).$$

And finally by concave duality,

$$\begin{aligned} u_t(x,t) + f(\nabla u(x,t)) &= u_t(x,t) + \inf_{w\in\mathbf{R}^d}\{w\cdot\nabla u(x,t) - f^*(w)\} \\ &\le u_t(x,t) + \frac{x-y}{t}\cdot\nabla u(x,t) - f^*\left(\frac{x-y}{t}\right) \le 0. \end{aligned}$$

The lemma is proved. ∎

The correct notion of weak solution for Hamilton-Jacobi equations is the viscosity solution, developed in references in [5] and [4]. Suppose now the function f in the equation (A.44) is continuous and real-valued on all of $\mathbf{R}^d$. Concavity is not relevant for this definition.

A continuous function $u(x,t)$ on $\mathbf{R}^d \times [0,\infty)$ that satisfies the initial condition $u(x,0) = u_0(x)$ is a *viscosity solution* of (A.44) if the following holds for all continuously differentiable functions ϕ on $\mathbf{R}^d \times (0,\infty)$:

if $u-\phi$ has a local maximum at (x_0,t_0), then

$$\phi_t(x_0,t_0) + f\big(\nabla\phi(x_0,t_0)\big) \le 0\,,$$

and if $u-\phi$ has a local minimum at (x_0,t_0), then

$$\phi_t(x_0,t_0) + f\big(\nabla\phi(x_0,t_0)\big) \ge 0\,.$$

We refer to Chapter 10 in Evans [14] for a general discussion of viscosity solutions of Hamilton-Jacobi equations. From our point of view, a small drawback in Evans's treatment is that only bounded viscosity solutions are considered. The height functions in Chapter 9 for which we need this theory are typically unbounded. Partly for this reason we prove the viscosity solution property for our case in Section 9.5.5. Another reason for giving this proof is that our Hamiltonian (or velocity) f also fails the assumptions used in [14].

The following uniqueness theorem of Ishii [21] for unbounded viscosity solutions applies to our situation in Chapter 9.

Theorem A.33 *Assume the Hamiltonian* f *in* (A.44) *is a continuous function on* $\mathbf{R}^d$*. Suppose* u *and* v *are uniformly continuous functions on* $\mathbf{R}^d \times [0,T]$ *and both are viscosity solutions of* (A.44)*. Then* $u = v$*.*

References

[1] D. Aldous and P. Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Probab. Theory Related Fields*, 103(2):199–213, 1995.

[2] D. Aldous and P. Diaconis. Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc. (N.S.)*, 36(4):413–432, 1999.

[3] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[4] M. G. Crandall, L. C. Evans, and P.-L. Lions. Some properties of viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 282(2):487–502, 1984.

[5] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 277(1):1–42, 1983.

[6] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes.* Springer Series in Statistics. Springer-Verlag, New York, 1988.

[7] A. De Masi and E. Presutti. *Mathematical methods for hydrodynamic limits*, volume 1501 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1991.

[8] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics.* Springer-Verlag, New York, second edition, 1998.

[9] J.-D. Deuschel and D. W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics.* Academic Press Inc., Boston, MA, 1989.

[10] R. M. Dudley. *Real analysis and probability.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[11] R. Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.

[12] M. Ekhaus and L. Gray. Convergence to equilibrium and a strong law for the motion of restricted interfaces. Unpublished manuscript, 1994.

[13] S. N. Ethier and T. G. Kurtz. *Markov processes: Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[14] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[15] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[16] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[17] G. B. Folland. *Real analysis: Modern techniques and their applications*. Pure and Applied Mathematics. John Wiley & Sons Inc., New York, second edition, 1999.

[18] H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988.

[19] P. Groeneboom. Hydrodynamical methods for analyzing longest increasing subsequences. *J. Comput. Appl. Math.*, 142(1):83–105, 2002. Probabilistic methods in combinatorics and combinatorial optimization.

[20] J. M. Hammersley. A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, pages 345–394, Berkeley, Calif., 1972. Univ. California Press.

[21] H. Ishii. Uniqueness of unbounded viscosity solution of Hamilton-Jacobi equations. *Indiana Univ. Math. J.*, 33(5):721–748, 1984.

[22] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[23] J. D. Keisling. An ergodic theorem for the symmetric generalized exclusion process. *Markov Process. Related Fields*, 4(3):351–379, 1998.

[24] C. Kipnis and C. Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[25] R. J. LeVeque. *Numerical methods for conservation laws*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 1992.

[26] T. Liggett. Interacting particle systems–an introduction. Lectures at the School on Probability Theory, Trieste, May 2002.

[27] T. M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[28] R. R. Phelps. *Lectures on Choquet's theorem*, volume 1757 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2001.

[29] S. Resnick. *Adventures in stochastic processes*. Birkhäuser Boston Inc., Boston, MA, 1992.

[30] F. Rezakhanlou. Continuum limit for some growth models. *Stochastic Process. Appl.*, 101(1):1–41, 2002.

[31] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[32] W. Rudin. *Functional analysis*. McGraw-Hill Book Co., New York, 1973. McGraw-Hill Series in Higher Mathematics.

[33] T. Seppäläinen. A microscopic model for the Burgers equation and longest increasing subsequences. *Electron. J. Probab.*, 1:no. 5, approx. 51 pp. (electronic), 1996.

[34] T. Seppäläinen. Coupling the totally asymmetric simple exclusion process with a moving interface. *Markov Process. Related Fields*, 4(4):593–628, 1998. I Brazilian School in Probability (Rio de Janeiro, 1997).

[35] T. Seppäläinen. Hydrodynamic scaling, convex duality and asymptotic shapes of growth models. *Markov Process. Related Fields*, 4(1):1–26, 1998.

[36] T. Seppäläinen. Large deviations for increasing sequences on the plane. *Probab. Theory Related Fields*, 112(2):221–244, 1998.

[37] T. Seppäläinen. Existence of hydrodynamics for the totally asymmetric simple K-exclusion process. *Ann. Probab.*, 27(1):361–415, 1999.

[38] T. Seppäläinen. A variational coupling for a totally asymmetric exclusion process with long jumps but no passing. In *Hydrodynamic limits and related topics (Toronto, ON, 1998)*, volume 27 of *Fields Inst. Commun.*, pages 117–130. Amer. Math. Soc., Providence, RI, 2000.

[39] T. Seppäläinen. Hydrodynamic profiles for the totally asymmetric exclusion process with a slow bond. *J. Statist. Phys.*, 102(1-2):69–96, 2001.

[40] T. Seppäläinen. Perturbation of the equilibrium for a totally asymmetric stick process in one dimension. *Ann. Probab.*, 29(1):176–204, 2001.

[41] T. Seppäläinen. Diffusive fluctuations for one-dimensional totally asymmetric interacting random dynamics. *Comm. Math. Phys.*, 229(1):141–182, 2002.

[42] T. Seppäläinen and J. Krug. Hydrodynamics and platoon formation for a totally asymmetric exclusion model with particlewise disorder. *J. Statist. Phys.*, 95(3-4):525–567, 1999.

[43] F. Spitzer. *Principles of random walks.* Springer-Verlag, New York, second edition, 1976. Graduate Texts in Mathematics, Vol. 34.

[44] H. Spohn. *Large scale dynamics of interacting particles.* Springer-Verlag, Berlin, 1991.

[45] K. R. Stromberg. *Introduction to classical real analysis.* Wadsworth International, Belmont, Calif., 1981. Wadsworth International Mathematics Series.

[46] S. R. S. Varadhan. Lectures on hydrodynamic scaling. In *Hydrodynamic limits and related topics (Toronto, ON, 1998)*, volume 27 of *Fields Inst. Commun.*, pages 3–40. Amer. Math. Soc., Providence, RI, 2000.

Index