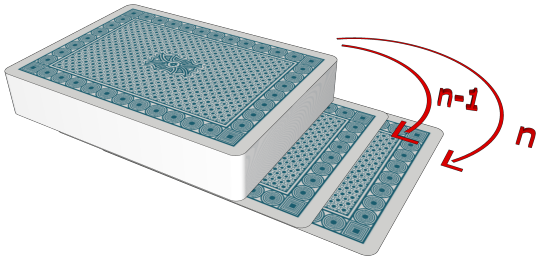
TÉCNICO LISBOA



An analysis of the Rudvalis Shuffle

David Tadeu Lopes Moreira

Thesis to obtain the Master of Science Degree in

Mathematics and Applications

Supervisor(s): Prof. Ana Patrícia Carvalho Gonçalves
Dr. Rodrigo Marinho de Souza

Examination Committee

Chairperson:	Prof. Pedro Simões Cristina de Freitas
Supervisor:	Prof. Ana Patrícia Carvalho Gonçalves
Member of the Committee:	Prof. Dirk Erhard
	Prof. José Manuel Vergueiro Monteiro Cidade Mourão

December, 2021

Agradecimentos

Em primeiro lugar, agradeço à minha família, principalmente aos meus pais e ao meu irmão, pelo seu apoio incondicional.

Agradeço também aos meus amigos que contribuiram para a minha evolução não só profissional mas também como pessoa.

Agradeço a todos os professores que acompanharam o meu percurso académico e se dedicaram e esforçaram para passar os seus conhecimentos. Em especial, agradeço aos meus orientadores nesta dissertação: à professora Patrícia Gonçalves e ao doutor Rodrigo Marinho, que me concederam esta oportunidade, que estiveram sempre disponíveis para ajudar e dedicaram grande parte do seu tempo para constante melhoria da dissertação.

Por último, mas não menos importante, agradeço ao Instituto Superior Técnico como instituição e comunidade.

Resumo

Nesta tese, vamos analisar uma forma específica de baralhar um número finito de cartas. O baralhamento consiste em retirar a primeira carta e lançar ao ar uma moeda: se sair cara, coloca-se a carta na última posição do baralho, mas se sair coroa, coloca-se a carta na penúltima posição do baralho. Este baralhamento foi introduzido por Arunas Rudvalis (cf. [4], página 90), pelo que o denominamos por baralhamento de Rudvalis.

A dissertação encontra-se dividida em dois capítulos. Na primeiro capítulo, começamos por definir propriedades básicas de cadeias de Markov. De seguida, fazemos um estudo do baralhamento de Rudvalis, usando cadeias de Markov e as propriedades vistas anteriormente, tais como o tempo de mistura. O principal objetivo deste capítulo é estudar o problema resolvido por Wilson em [15], i.e., estimar uma cota inferior (em função do número de cartas do baralho) para o número mínimo de vezes que temos de baralhar as cartas se quisermos que o baralho fique bem baralhado.

Na segundo capítulo, aplicamos sucessivamente o baralhamento de Rudvalis e vamos olhando para a configuração das cartas como um sistema de partículas (identificando as cartas vermelhas como sítios vazios e as cartas pretas como sítios ocupados), fazendo uso de cadeias de Markov em tempo contínuo. O resultado principal é o Limite Hidrodinâmico, que caracteriza a evolução da densidade de partículas. Em particular, provamos a existência da solução fraca de uma equação diferencial parcial, a equação do transporte no toro.

Palavras-chave: Cadeias de Markov, equações diferenciais parciais, limite hidrodinâmico, sistema de partículas, tempo de mistura, equação do transporte.

Abstract

In this thesis, we analyze a specific way to shuffle a finite number of cards. The shuffle consists of drawing the first card and flipping a coin: if it lands heads, we place the card at the last position of the deck, while if tails come out, the card is placed at the second-to-last position of the deck. This shuffle was introduced by Arunas Rudvalis (cf. [4], page 90), so we call it the Rudvalis shuffle.

The dissertation is divided into two chapters. In the first chapter, we start by defining basic properties of Markov chains. Next, we study the Rudvalis shuffle, using Markov chains and the properties previously seen, such as the mixing time. The main goal of this chapter is to study the problem solved by Wilson in [15], i.e., to estimate a lower bound (as a function of the number of cards in the deck) for the minimum number of times we have to shuffle the cards if we want the deck to be well shuffled.

In the second chapter, we successively apply the Rudvalis shuffle and look at the configuration of the cards as a particle system (identifying the red cards as empty sites and the black cards as occupied sites), making use of continuous time Markov chains. The main result is the Hydrodynamic Limit, which characterizes the evolution of the particle density. In particular, we prove the existence of the weak solution of a partial differential equation, the transport equation on the torus.

Keywords: Markov chains, partial differential equations, hydrodynamic limit, particle system, mixing time, transport equation.

Contents

Nomenclature

$\|\cdot\|_{TV}$ Given two probability measures μ and ν, we denote by $\|\mu-\nu\|_{TV}$ the total variation distance between these measures.

$d(t)$ We denote by $d(t)$ the total variation distance between the distribution of a Markov chain at time t to its stationary distribution.

t_{mix} We denote by t_{mix} the mixing time of a Markov chain.

$\|\cdot\|_{\infty}$ Given $f : X \to \mathbb{R}$, we denote by $\|f\|_{\infty} = \sup_{x\in X}|f(x)|$ the $L^{\infty}(X)$-norm.

$D_X[0,T]$ Given $T > 0$ and a metric space X, we denote by $D_X[0,T]$ the space of right continuous functions $x : [0,T] \to X$ which have left-hand limits.

$E_{\mu}[\,\cdot\,]$ Given a probability measure μ defined on Ω and a random variable, $Z : \Omega \to \mathbb{R}$, we denote by $E_{\mu}[Z] = \int_{\Omega} Z(\omega)\ \mu(d\omega)$ the expectation of Z with respect μ. Likewise, we denote by $Var_{\mu}(Z)$ the variance of Z with respect to μ.

$\mathbb{E}_{\mu}[\,\cdot\,]$ Given a measure μ on Ω, a measure $\mathbb{P}_{\mu}$ on $D_{\Omega}[0,T]$ starting from μ, we denote by $\mathbb{E}_{\mu}[Z] = \int_{D_{\Omega}[0,T]} Z(x)\mathbb{P}_{\mu}(dx)$ the expectation of Z with respect to $\mathbb{P}_{\mu}$. When referring to a Markov process, $\mathbb{E}_{\mu}$ indicates the expectation given that the initial measure is μ.

$O(g(n))$ Given a complex valued function f and a real valued function g defined in $\mathbb{N}$, we write that $f(n) = O(g(n))$ if there are $M, n_0 \in \mathbb{N}$ such that $|f(n)| \leq Mg(n)$ for all $n \geq n_0$.

$o(g(n))$ Given a complex valued function f and a real valued function g defined in $\mathbb{N}$, we write that $f(n) = o(g(n))$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

$\Re(\,\cdot\,)$ Given a complex number $\lambda = a + bi$ (with $a, b \in \mathbb{R}$), we denote by $\Re(\lambda) = a$ its real part.

$\Im(\,\cdot\,)$ Given a complex number $\lambda = a + bi$ (with $a, b \in \mathbb{R}$), we denote by $\Im(\lambda) = b$ its imaginary part.

$\overline{\lambda}$ Given a complex number $\lambda = a + bi$ (with $a, b \in \mathbb{R}$), we denote by $\overline{\lambda} = a - bi$ its conjugate.

$\mathbb{T}$ We denote by $\mathbb{T} := \mathbb{R}/\mathbb{Z} = [0,1)$ the one-dimensional continuous torus.

$\mathbb{T}_n$ We denote by $\mathbb{T}_n := \mathbb{Z}/n\mathbb{Z} = \{0, 1, \ldots, n-1\}$ the one-dimensional discrete torus with n points.

Ω_n We denote by $\Omega_n := \{0,1\}^{\mathbb{T}_n} = \{\eta : \mathbb{T}_n \to \{0,1\} \mid \eta \text{ is a function}\}$ the space of configurations of particles on the torus.

$Z \sim \mu$ Given a random variable Z and a probability distribution μ, we write that $Z \sim \mu$ if Z has distribution μ.

$\mathcal{M}$ We denote by $\mathcal{M}$ the space of positive measures on $\mathbb{T}$ with total mass bounded by 1 endowed with the weak topology.

$\langle \cdot, f \rangle$ We use this notation for two different operations: if ρ is a function, then we denote by $\langle \rho, f \rangle = \int \rho(u) f(u)\, du$ the inner product in $L^2(\mathbb{T})$, while if π is a measure, then we denote by $\langle \pi, f \rangle = \int f(u)\, \pi(du)$ the integral of f with respect to π.

C^n We denote by C^n the space of functions whose the j^{th} derivative for $j \in \{0, 1, ..., n\}$ exists and it is continuous. In particular, $C^0 := C$ is the space of continuous functions.

$C^{n,m}$ Given $T > 0$, we denote by $C^{n,m}([0,T] \times \mathbb{T})$ the space of functions $f : [0,T] \times \mathbb{T} \to \mathbb{R}$ such that f is $C^n[0,T]$ in the first component and $C^m(\mathbb{T})$ in the second component.

∂_t Given $f : \mathbb{R}^+ \times \mathbb{T} \to \mathbb{R}$, we denote by ∂_t the first derivative on the first component (time).

∇ Given a function $f : \mathbb{R}^+ \times \mathbb{T} \to \mathbb{R}$, we denote by $\nabla f(t,u)$ the first derivative on the second component (space).

$f_t(u)$ Given a function $f := f(t,u)$ that depends on two variables, we write $f_t(u)$ instead of $f(t,u)$ in order to simplify the notation.

$\cdot \lesssim \cdot$ Given two sequences $\{f_n\}_{n \in \mathbb{N}}$ and $\{g_n\}_{n \in \mathbb{N}}$, we write that $f_n \lesssim g_n$ if there exists a positive constant C, independent of n, such that $f_n \leq C\, g_n$ for all $n \in \mathbb{N}$.

$\cdot \ll \cdot$ Given two measures μ and λ, we write that $\mu \ll \lambda$ if μ is absolutely continuous with respect to λ.

$(X, \mathcal{B})$ We denote by X a generic metric space and by $\mathcal{B}$ the σ-algebra generated by the open sets, the Borel σ-algebra on X.

Introduction

Markov chains, named after Andrey Markov, are widely used in several fields, e.g., statistics, economics, machine learning. A Markov chain is a stochastic process, i.e., a collection of random variables indexed on time, which has the property (Markov property) that the future of the process depends only on its present and it is conditionally independent of its past.

There are many real life occurrences of Markov chains but our focus in this dissertation will be a Markov chain generated by a card shuffle, more precisely, a random walk followed by the permutations of the cards when the deck is shuffled. Here the state space is the symmetric group S_n (whose elements are called permutations) and the transitions are given by the specific chosen way to shuffle the cards.

We are interested on the number of shuffles it takes for a deck to be well shuffled. But what does it mean for the deck to be well shuffled? Well, it means that after a certain number of shuffles, all the possible $n!$ permutations are equally likely. In a standard deck of 52 cards, the number of possible states of the deck $(52!)$ is already very large, and it might seem that we need several shuffles to mix the deck. Of course, the number of shuffles we need depends on the shuffle. For example, if we use the classical riffle shuffle (in which half of the deck is held in each hand and then cards are released so that they fall to the table interleaved), after only 7 shuffles, the deck is close to random [2]. We are going to consider instead another shuffle, which was introduced by Arunas Rudvalis [4] and it is very simple to understand. We take the first card of the deck and we insert it in the last position with probability $1/2$ or in the second-to-last position with probability $1/2$. Note that the Markov property is verified because the next permutation of the deck depends only on the previous one and it is conditionally independent of the others.

We shall look for the number of shuffles needed for the deck to be well shuffled after successively applying the previous shuffle. But how can we do this? We will start by defining some basic concepts of discrete time Markov chains, namely, the mixing time, which gives the time it takes for the chain to be arbitrarily close to its stationary distribution. If the deck reaches the stationary state, which is the uniform distribution, this means that at that moment any permutation is equally likely. In other words, when the state of the deck reaches stationarity, the deck is well shuffled. Thus, finding the number of shuffles it takes for the deck to be well shuffled boils down to estimating the mixing time of the Markov chain described by the Rudvalis shuffle. Having realized that, we will use some properties of eigenvectors to estimate the mixing time of the Rudvalis shuffle. The upper bound of the mixing time Rudvalis shuffle is $O(n^3 \log(n))$ [10] (after that number of shuffles the deck is well shuffled). Here we explain the method

used by Wilson in [15], which shows that the lower bound for the shuffle is also $O(n^3 \log(n))$ (we need at least that number of shuffles for the deck to be well shuffled). Therefore we can conclude that the mixing time for the Rudvalis shuffle is $O(n^3 \log(n))$.

For the second part of the dissertation, we will interpret the Rudvalis shuffle as an interacting particle system on n sites, associating black cards with particles and red cards with vacant sites (in a standard 52-card deck of French-suited playing cards, this would correspond to having a system of 52 sites with 26 particles). We want instead to find what happens to the system when n goes to infinity, i.e., we want to know how the particles behave when the number of sites goes to infinity.

This scenario is motivated by the interaction between particles (microscopic) in a real system, such as a gas or a fluid (macroscopic). In this kind of systems, the number of particles is very large and thus it is impossible to study the behaviour of each particle individually in order to understand the collective behaviour of the system, i.e., it is not easy to study the macroscopic space (for example the gas) using the microscopic space (looking at the motion of all individual particles). To overcome this problem, what it is done instead is to consider that the particles move randomly according to some probability law (this is shown to model well real macroscopic systems) and use the results obtained for the microscopic case to draw conclusions for the macroscopic space.

The last strategy is done by considering an initial density of particles at the macroscopic level and a probabilistic law for the microscopic particles (this will be determined by the Rudvalis shuffle). Then we associate the density of the macroscopic system with an initial distribution of particles in the microscopic system, to which we associate a continuous Markov chain, modelling the microscopic system. This system conserves one or more quantities (in our case, the number of red and black cards is the preserved quantity) but the distribution of this quantity changes through time. The question we ask is how can we approximate the distribution of particles through time, and the hydrodynamic limit tells us that the density of particles is described by a solution of a partial differential equation (PDE). In other words, the hydrodynamic limit allows finding a macroscopic law (in general a solution of a PDE) through a microscopic underlying random dynamic.

The hydrodynamic limit is really interesting in the sense that it allows us to relate concepts of probability theory with partial differential equations. This is because the density of particles in the macroscopic space is usually described by a weak solution of a PDE. So if we take a random process (the card shuffle), this random process gives rise, in the limit, to a deterministic process. Furthermore, if we prove the uniqueness of the weak solution of that PDE, the hydrodynamic limit allows concluding the existence of a weak solution to a PDE which a priori we did not know it existed. In other words, we use a random process to prove the existence of a (weak) solution of a PDE. This is exactly what we aim to do in the second part of this dissertation: explore the Rudvalis shuffle (which is a random process) and use it to find the (weak) solution of a PDE, in our case, the transport equation.

Chapter 1

Discrete time Rudvalis shuffle

In this chapter, we start by studying discrete time Markov chains. We define a Markov chain, present basic properties such as stationarity and irreducibility and we also have a brief look at eigenfunctions associated with Markov chains. Then, we define the Rudvalis shuffle and estimate a lower bound for its mixing time by using properties of discrete time Markov chains.

1.1 Discrete time Markov chains

Among the properties of a Markov chain, the irreducibility and the stationarity are of special interest, since, for example, if a Markov chain is irreducible and its state space is finite, then the stationary distribution is unique [13]. For these chains, we can define the mixing time, which is the time the chain takes to be close (to be rigorously specified later on) to its stationary distribution.

1.1.1 Basic properties

Let $(\tilde{\Omega}, \mathcal{F}, \mathbb{P})$ be a probability space. Recall the definition of a (homogeneous) *discrete time Markov chain* [8]:

Definition 1.1.1 (Discrete time Markov chain)**.** *Let P be a $k \times k$ matrix with elements $\{P(i,j) : i,j \in \{1,\cdots,k\}\}$. A random process $\{X_t\}_{t\in\mathbb{N}_0}$ with finite state space $\Omega = \{s_1,\cdots,s_k\}$ is said to be a (homogeneous) Markov chain with transition matrix P, if for all n, all $i,j \in \{1,\cdots,k\}$ and all $i_0,\cdots,i_{n-1} \in \{1,\cdots,k\}$ we have*

$$\begin{aligned}\mathbb{P}(X_{n+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, \cdots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) &= \mathbb{P}(X_{n+1} = s_j | X_n = s_i) \\ &= P(i,j).\end{aligned}$$

The elements of the transition matrix P are called transition probabilities: $P(i,j) = \mathbb{P}(X_{n+1} = s_j | X_n = s_i)$ is the probability of transitioning from state s_i to state s_j in one time step.

The *stationary distribution* of a Markov chain is a probability distribution such that if the chain starts from that distribution, then it will always have that distribution at any time $t \in \mathbb{N}_0$.

Definition 1.1.2 (Stationary distribution)**.** *Consider a (homogeneous) Markov chain $\{X_t\}_{t\in\mathbb{N}_0}$ with transition matrix P and state space $\Omega = \{s_1, \ldots, s_k\}$. A probability distribution $\pi = (\pi_1, \ldots, \pi_k)$ is said to be a stationary distribution (or stationary measure or invariant measure) if $\pi P = \pi$, that is, for $j = 1, \ldots, k$*

$$(\pi P)(j) = \sum_{i=1}^{k} \pi(i)P(i,j) = \pi(j).$$

Remark 1.1.3. *In order to avoid an enumeration of the state space, sometimes we might denote $P(i,j)$ by $P(s_i, s_j)$ and $\pi(i)$ by $\pi(s_i)$.*

Another concept related to Markov chains is the one of irreducibility. A chain is said to be *irreducible* if all states can be reached from every state. More precisely:

Definition 1.1.4 (Irreducibility)**.** *A Markov chain $\{X_t\}_{t\in\mathbb{N}_0}$ with state space $\Omega = \{s_1, \cdots, s_k\}$ and transition matrix P is said to be irreducible if for all $s_i, s_j \in \Omega$ there is a positive probability of ever reaching s_j starting from s_i and ever reaching s_j starting from s_i. In other words, there exist $n, k \in \mathbb{N}_0$ such that*

$$\mathbb{P}(X_{m+n} = s_j | X_m = s_i) > 0,$$
$$\mathbb{P}(X_{m+k} = s_i | X_m = s_j) > 0.$$

Otherwise the chain is said to be reducible.

It is well known that every Markov chain with finite state space has at least one stationary distribution and that if the Markov chain is irreducible, the stationary distribution is unique.

Throughout the rest of this section, we shall denote by $\{X_t\}_{t\in\mathbb{N}_0}$ an irreducible and aperiodic (for any state s, the greatest common divisor of the set of times the chain can return to s is 1) Markov chain with finite state space Ω, transition matrix P, initial measure $\mu_0(\cdot) = \mathbb{P}(X_0 \in \cdot)$ and stationary distribution π. Let $\mathcal{F}_t := \sigma(X_s \mid s \leq t)$ be the σ-algebra generated by $X_1, X_2, \ldots, X_t$ and $\{\mathcal{F}_t\}_{t\in\mathbb{N}_0}$ the filtration associated with $\{X_t\}_{t\in\mathbb{N}_0}$. Moreover, denote by $\mathbb{E}_{\mu_0}$ the expectation starting from μ_0.

One question we might ask is if the distribution of the chain converges (with time $t \in \mathbb{N}_0$) to its stationary distribution. In order to know that, we first need a way to compare two probability distributions.

Definition 1.1.5 (Total variation distance)**.** *Let μ and ν be two probability distributions defined on a finite state space Ω. The total variation distance between these two probability distributions is defined as*

$$\|\mu - \nu\|_{TV} := \max_{A\subseteq\Omega} |\mu(A) - \nu(A)|,$$

where $\mu(A) = \sum_{x\in A} \mu(x)$.

Most often, the set which maximizes the difference of the distributions might not be easy to find, so we propose another way to compute this distance between distributions.

Proposition 1.1.6. *Let μ and ν be two probability distributions defined on the state space Ω. Then*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x\in\Omega} |\mu(x) - \nu(x)|.$$

Proof. Consider the event $B = \{x \in \Omega : \mu(x) \geq \nu(x)\}$ and let $A \subseteq \Omega$.
Let $C := A \cap B^c \subseteq B^c$. Since μ and ν are finite,

$$\mu(C) < \nu(C) \Leftrightarrow \mu(A) - \nu(A) < \mu(A \cap B) - \nu(A \cap B) \tag{1.1}$$

and for $D := A^c \cap B \subseteq B$,

$$\mu(D) \geq \nu(D) \Leftrightarrow \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B). \tag{1.2}$$

Combining (1.1) and (1.2), we obtain

$$\mu(A) - \nu(A) \leq \mu(B) - \nu(B).$$

Similarly, taking $C := A^c \cap B^c$ and $D := A \cap B$,

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c) = \mu(B) - \nu(B).$$

Thus, from the two previous inequalities, we conclude that $|\mu(A) - \nu(A)| \leq \mu(B) - \nu(B)$, so that

$$\begin{aligned}\max_{A \subseteq \Omega} |\mu(A) - \nu(A)| = \mu(B) - \nu(B) &= \frac{1}{2}(\mu(B) - \nu(B)) + \frac{1}{2}(\nu(B^c) - \mu(B^c)) \\ &= \frac{1}{2}\sum_{x \in B} \mu(x) - \nu(x) + \frac{1}{2}\sum_{x \in B^c} |\mu(x) - \nu(x)| = \frac{1}{2}\sum_{x \in \Omega} |\mu(x) - \nu(x)|.\end{aligned}$$

□

Remark 1.1.7. *It is clear that* $0 \leq \|\mu - \nu\|_{TV} \leq 1$ *and if* $\|\mu - \nu\|_{TV} = 0$*, then* $\mu = \nu$*. Furthermore, from the previous result we can also write the total variation distance as*

$$\|\mu - \nu\|_{TV} = \frac{1}{2}(\mu(B) - \nu(B)),$$

where $B = \{x \in \Omega : \mu(x) \geq \nu(x)\}$*.*

Convergence in total variation distance is defined as expected.

Definition 1.1.8 (Convergence in total variation distance)**.** *Let* $\{\mu_t\}_{t \in \mathbb{N}_0}$ *and* μ *be probability distributions on a finite set* Ω*. We say that* $\{\mu_t\}_{t \in \mathbb{N}_0}$ *converges to* μ *in total variation, which we denote by* $\mu_t \xrightarrow[t \to \infty]{TV} \mu$ *if*

$$\lim_{t \to \infty} \|\mu_t - \mu\|_{TV} = 0.$$

In particular, for an irreducible and aperiodic finite chain, the distribution of the chain converges (in total variation distance) to the unique stationary distribution π (Section 4.3 of [13]). But how much time t does it take for the distribution of the chain to get arbitrarily close to the stationary distribution? With that question in mind, we define a function that keeps track of the distance between the distribution of

the chain at time t and the stationary distribution:

$$d(t) = \max_{x \in \Omega} \left\| P^t(x, \cdot) - \pi \right\|_{TV},$$

where $P^t(x, y) := \mathbb{P}(X_t = y | X_0 = x)$ is the probability of reaching state y in t time steps, given that $X_0 = x$, that is, given that the chain started at state x.

It can be shown that $d(t)$ is a decreasing function (Exercise 4.2 of [13]) which approaches zero as t goes to infinity, i.e., the more the chain evolves, the closer it is to stationarity. The time t in which the chain gets arbitrarily close to the stationary distribution is called the *mixing time* of the Markov chain and is precisely defined as follows:

Definition 1.1.9 (Mixing time)**.** *For $\varepsilon \in (0, 1)$, we define*

$$t_{mix}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}.$$

It is common to choose $\varepsilon := \frac{1}{4}$ and define $t_{mix} := t_{mix}(1/4)$.

1.1.2 Eigenfunctions

It will be useful to introduce the concept of *eigenfunctions* and *eigenvalues* of the transition matrix of a Markov chain. Moreover, there are properties relating eigenfunctions and the stationary distribution of a Markov chain.

Definition 1.1.10 (Eigenfunction)**.** *Let P be a transition matrix of a Markov chain with finite state space Ω. A function f on Ω is an eigenfunction (or eigenvector) of P associated with the eigenvalue λ if $Pf = \lambda f$.*

Remark 1.1.11. *The eigenfunctions, as well as the eigenvalues, can be complex-valued. Moreover, by abuse of language, we might say "a Markov chain with eigenfunction f" instead of "a Markov chain whose transition matrix has an eigenfunction f".*

The previous definition is equivalent to having

$$(Pf)(z) := \sum_{j \in \Omega} P(z, j) f(j) = \lambda f(z)$$

for every state $z \in \Omega$. From now on, we will omit the first parenthesis and just write $Pf(z)$.

Let f be a function on Ω and μ be a probability distribution on Ω. Then, we can use the *Markov property* to write

$$\mathbb{E}_\mu[f(X_{t+1}) | \mathcal{F}_t] = \mathbb{E}_\mu[f(X_{t+1}) | X_t] = \sum_{x \in \Omega} P(X_t, x) f(x) = Pf(X_t),$$

where the Markov property is used in the first equality. If f is an eigenfunction with eigenvalue λ, then

$$\mathbb{E}_\mu[f(X_{t+1}) | \mathcal{F}_t] = \mathbb{E}_\mu[f(X_{t+1}) | X_t] = \lambda f(X_t). \tag{1.3}$$

The fact that the transition matrix P is right stochastic (i.e., each entry is non-negative and each row adds up to 1) will give rise to an eigenfunction associated with the eigenvalue 1. In fact, it is true that all the eigenvalues have absolute value bounded from above by 1.

Lemma 1.1.12. *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete Markov chain with finite state space Ω, transition matrix P and initial distribution μ_0. Let f be an eigenfunction of P with eigenvalue λ. Then*

1. $\mathbb{E}_{\mu_0}[f(X_t)] = \lambda^t f(X_0)$.

2. $|\lambda| \leq 1$ *and there exists always one eigenvalue equal to* 1.

Proof.

1. If we set $g(t) := \mathbb{E}_{\mu_0}[f(X_t)]$ and take expectations on both sides of $\mathbb{E}_{\mu_0}[f(X_{t+1})|X_t] = \lambda f(X_t)$, we get (by the Tower Law property) that $g(t+1) = \lambda g(t)$, for all $t \in \mathbb{N}_0$. Solving that equation we get $g(t) = \lambda^t g(0)$. Thus, for all $t \in \mathbb{N}_0$, we have that

$$\mathbb{E}_{\mu_0}[f(X_t)] = \lambda^t \mathbb{E}_{\mu_0}[f(X_0)] = \lambda^t f(X_0).$$

2. Let $\|f\|_\infty = \max_{x\in\Omega} |f(x)|$. Since the entries of P are non-negative and the rows add up to 1,

$$\|Pf\|_\infty = \max_{x\in\Omega} |\sum_{y\in\Omega} P(x,y)f(y)| \leq \max_{x\in\Omega} \sum_{y\in\Omega} P(x,y)|f(y)| \leq \|f\|_\infty \max_{x\in\Omega} \sum_{y\in\Omega} P(x,y) = \|f\|_\infty .$$

   Now suppose, by contradiction, that $|\lambda| > 1$. Then $\|\lambda f\|_\infty = |\lambda| \, \|f\|_\infty > \|f\|_\infty$, leading to $\|Pf\|_\infty < \|\lambda f\|_\infty$, which is a contradiction with the definition of the eigenfunction λ.
   In order to see that there is an eigenvalue 1, just note that since the rows of P add up to 1, it is true that $Pf = f$, if $f(x) = 1$ for any $x \in \Omega$, that is, 1 is an eigenvalue of P.

□

We define the expectation of a function $\Psi : \Omega \to \mathbb{C}$ under a probability measure μ on a finite space Ω by

$$E_\mu[\Psi] = \sum_{x\in\Omega} \Psi(x)\mu(x).$$

Lemma 1.1.13. *Let P be a transition matrix of a Markov chain with finite state space Ω and stationary distribution π. If f is an eigenfunction with eigenvalue $\lambda \neq 1$, then $E_\pi[f] = 0$.*

Proof. Note that

$$\lambda \pi f = \lambda \sum_{x\in\Omega} f(x)\pi(x) = \lambda E_\pi[f]$$

and since π is stationary, $\pi P f = \pi f = E_\pi[f]$. Thus, multiplying by π on both sides of $Pf = \lambda f$, we obtain that $E_\pi[f] = \lambda E_\pi[f]$. Since $\lambda \neq 1$, we must have $E_\pi[f] = 0$. □

1.2 Rudvalis shuffle

We will now take a look at a very specific way to shuffle a deck of cards. Since a card shuffle consists of a random walk on the permutations of the deck, this shuffle can be seen as a Markov chain. With that in mind, we will find a lower bound for the mixing time of this chain. In other words, we will determine how many times $g(n)$ we have to shuffle the deck so that after those shuffles, the deck is still not well shuffled (we need at least $g(n)$ shuffles if we want the deck to be well shuffled and we know that shuffling less than that number will not guarantee a well shuffled deck).

Let us consider a deck with n cards numbered from 1 to n (we label the first card with 1, the second with 2, and so on). At each time $t \in \mathbb{N}$, we remove the card on top of the deck and insert that card at the second position from the bottom with probability p and at the bottom of the deck with probability $1-p$, where $p \in (0,1)$.

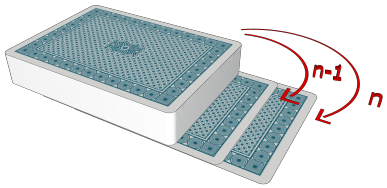


Figure 1.1: Inserting the top card at positions $n-1$ or n with probability p or $1-p$ respectively.

Example 1.2.1. *According to the initial configuration of the deck, we label the cards from 1 to n. We have the following example for $n = 5$. We are seeing the deck from below. The top card, the 3 of hearts*

Figure 1.2: A deck with $n = 5$ cards being labelled.

is labelled with 1. The second card, the king of clubs, is labelled with 2. And so on.

In order to see the dynamics of the shuffle, consider the following example of two consecutive shuffles for a deck with $n = 19$ cards.

Example 1.2.2. *Suppose the deck has $n = 19$ cards. We can use a (not necessarily fair) coin to decide where we place the first card. If it lands "heads", we place the first card at position $n = 19$, if it lands "tails", we place it at position $n - 1 = 18$. The following image shows the four different possibilities after two shuffles. The top card and bottom cards of the initial configuration of the deck are respectively the*

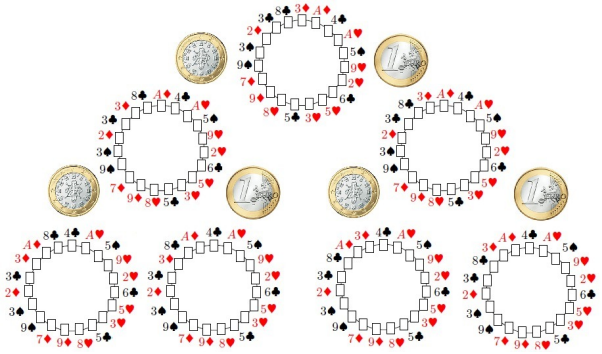


Figure 1.3: Example of the Rudvalis shuffle with $n = 19$.

3 of diamonds and 8 of clubs (we can imagine the deck in a circle where the last and first cards are "connected"). After the first shuffle, the 3 of diamonds either goes to position $n - 1$ or position n. We see that in both cases, all the cards but the 3 of diamonds (and possibly the 8 of clubs) are pushed up one position, i.e., almost all the cards rotate counterclockwise. If the deck keeps being shuffled, the cards keep "rotating" and as we will see later on, this is what makes the Rudvalis shuffle so unique.

The discrete-time Markov chain described by Figure 1.1 was introduced by Arunas Rudvalis (in fact it is a slight generalization of the original shuffle proposed by Rudvalis), which considers $p = \frac{1}{2}$, and for this reason, it is known as the Rudvalis shuffle. Its state space is the symmetric group S_n of all possible permutations of an n-element set. For $\sigma \in S_n$, we write $\sigma(x) = y$ if card y is in position x of the deck. The transition matrix P is given by

$$P(\sigma, \xi) = \begin{cases} p, \text{ if } \xi = \sigma^{n-1} \\ 1 - p, \text{ if } \xi = \sigma^n \\ 0, \text{ otherwise} \end{cases}$$

where σ^x is the permutation obtained from σ after we move the top card to position x. In other words, given $\sigma = (\sigma(1), \sigma(2), \cdots, \sigma(n)) \in S_n$, with probability p, the state of the deck changes to permutation $\sigma^{n-1} = (\sigma(2), \sigma(3), \cdots, \sigma(n-1), \sigma(1), \sigma(n))$ and with probability $1 - p$ to permutation $\sigma^n = (\sigma(2), \sigma(3), \cdots, \sigma(n), \sigma(1))$. Moreover, $\sigma^{n-1}(n-1) = \sigma^n(n) = \sigma(1)$, $\sigma^{n-1}(n) = \sigma^n(n-1) = \sigma(n)$ and for $y \in \{n-1, n\}$, $\sigma^y(x) = \sigma(x+1)$ for any $x \in \{1, \ldots, n-2\}$.

We denote the Rudvalis Markov chain by $\{\sigma_t\}_{t \in \mathbb{N}_0}$. Moreover, we denote by $P^t(\sigma, \cdot)$ the distribution of the deck after t shuffles, given that the initial permutation of the deck is σ. For example, $P^t(\sigma, \xi)$ is the probability of having permutation ξ after t shuffles, given that the starting permutation is σ.

We denote by $\mathbb{E}_\sigma$ the expectation with respect to the law of the process when the initial distribution $\mu_0(\cdot) = \mathbb{P}(\sigma_0 \in \cdot)$ is concentrated on σ, that is, the initial permutation of the deck is σ (we have that

$\mu_0(\sigma) = 1$).

If the Rudvalis shuffle was not an irreducible chain, it would not make much sense to call it a shuffle (we would not be able to achieve some specific configurations). But as expected, the Rudvalis chain is irreducible, as can be seen in the next proposition.

Proposition 1.2.3. *The Rudvalis chain is irreducible.*

Proof. We start by noting that the chain is irreducible if, and only if, we can move any card k from its position to a position between any two consecutive cards without changing the order of any card except k. We then have to show that if the card k is at position i, we have a positive probability of moving it to between the cards at positions $j-1$ and j , without changing the order of the remaining cards. First, note that we can "rotate" the deck by consecutively placing the first card at the bottom of the deck until card k reaches the first position. Let us call this sequence of shuffles an ascension (of card k). An ascension brings card k to the top of the deck without changing the order of the deck. Then, we can place card k at position $n-1$ and so this card "moves up" with respect to all the remaining cards. Let us call this only shuffle, i.e., placing card k at position $n-1$ by increment. Next, we can do another ascension followed

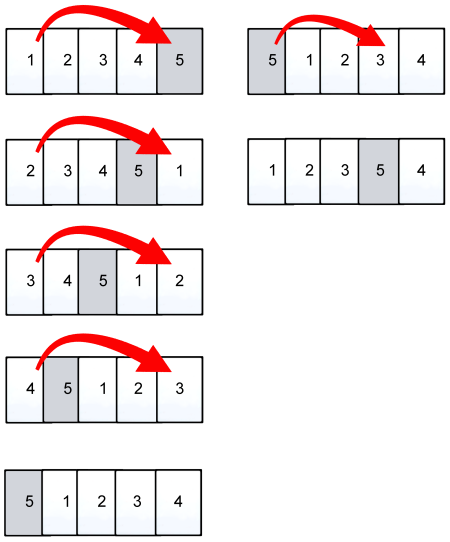


Figure 1.4: Ascension of card $k = 5$ on the left, and its increment on the right.

by an increment, and card k "moves up" one more position. Thus, in order to move card k from position i to between cards at positions $j-1$ and j ("above" the card at position j) , maintaining the order of the remaining cards, we would need $(i-j) \mod n$ ascensions and increments, obtaining a configuration with the cards initially at positions $j-1$ and j at positions $n-2$ and n respectively, and card k at position $n-1$. For example, in Figure 1.4, a card which starts at position $i = 5$, needs $i - j = 1$ ascensions and increments to reach the position between the cards at positions $j - 1 = 3$ and $j = 4$. □

Remark 1.2.4. *As seen in the previous proof, if the Rudvalis shuffle placed the first card at position n with probability 1, we would just be rotating the deck and the Rudvalis chain would be reducible (this shuffle would be equivalent to the usual "cut" of a deck of cards).*

Example 1.2.5. *Card magicians often use reducible card shuffles. For example, it is common to memorize the first (or last) card of the deck and then apply the riffle shuffle but keeping the first card in its place (at the eyes of a spectator, this will look like a normal shuffle). This shuffle is reducible since the first card will always be the same and a configuration where another card is on top is impossible to be reached using this shuffle. Then, the spectator, who thinks that the magician is, in fact, shuffling the deck, does not know that the magician holds that extra information of knowing the first card and that he could use that in his trick.*

Since the chain is irreducible and has a finite state space, it has a unique stationary distribution U_n, which is the uniform distribution on S_n (cf. Proposition 2.16 of [13]). In fact, it is simple to verify that $U_nP = U_n$. For any permutation σ,

$$(U_nP)(\sigma) = \sum_{\xi \in S_n} U_n(\xi)P(\xi,\sigma) = \sum_{\xi \in S_n} \tfrac{1}{n!}P(\xi,\sigma) = \tfrac{1}{n!}(P(\xi_1,\sigma) + P(\xi_2,\sigma)) = \tfrac{1}{n!} = U_n(\sigma),$$

where ξ_1 and ξ_2 are the permutations such that $\sigma = \xi_1^{n-1} = \xi_2^n$.

If we keep shuffling this way, how long must we shuffle the deck until it is well shuffled? But what does it mean for the deck to be well shuffled? It means that the arrangement of the deck is close to random. Thus, another way to ask this question is: How many shuffles does it take for the distribution of the deck to become close to uniform? As we have seen before, this can be done using the mixing time of the Markov chain. While an upper bound of $O(n^3 \log(n))$ for the mixing time was already found (cf. [10]), in this dissertation we focus on finding a lower bound for the mixing time of this chain. The bound for the mixing time translates into a bound on the number of shuffles which tells us that if we shuffle the deck any number of times below that bound, we know for sure that the deck is still not well shuffled. In order to find this bound, we first find a lower bound for the total variation distance between any two probability measures. Then, we consider the case in which one of the measures is the uniform measure on S_n and the other is the distribution of the state of the deck after t shuffles. This will allow finding a lower bound for $d(t)$. Finally, we will use this result to obtain a lower bound for $t_{mix}(\varepsilon)$.

Thus, we need to find $g(n,\varepsilon)$ such that $t_{mix}(\varepsilon) \geq g(n,\varepsilon)$. If we can find such a function, we have just shown that for any initial configuration of the deck, we need to shuffle the deck at least $g(n,\varepsilon)$ times if we want the deck to become well shuffled. However, shuffling $g(n,\varepsilon)$ times does not guarantee that the deck will be well shuffled.

1.2.1 Lower bound for the mixing time of the Rudvalis shuffle

Recall that for $\varepsilon \in (0,1)$, the mixing time of the Rudvalis Markov chain is given by $t_{mix}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}$, where

$$d(t) = \max_{\sigma \in S_n} \left\| P^t(\sigma,\cdot) - U_n \right\|_{TV},$$

where U_n is the uniform measure on S_n.

We introduce the following notation for a probability distribution defined on some other space. For $f : \Omega \to \Gamma$ measurable and μ a probability distribution on Ω, denote by μf^{-1} the probability distribution

on Γ, which is defined for $I \subseteq \Gamma$ as

$$\mu f^{-1}(I) := \mu(f^{-1}(I)).$$

Lemma 1.2.6. *Let μ and ν be probability distributions on Ω and let $f : \Omega \to \Gamma$ be a measurable function, where Γ and Ω are finite sets. Then*

$$\|\mu - \nu\|_{TV} \geq \left\|\mu f^{-1} - \nu f^{-1}\right\|_{TV}.$$

Proof. For a measurable set $J \subseteq \Gamma$, since

$$|\mu f^{-1}(J) - \nu f^{-1}(J)| = |\mu(f^{-1}(J)) - \nu(f^{-1}(J))|,$$

it follows that

$$\|\mu - \nu\|_{TV} = \max_{I \subseteq \Omega} |\mu(I) - \nu(I)| \geq \max_{J \subseteq \Gamma} |\mu f^{-1}(J) - \nu f^{-1}(J)| = \left\|\mu f^{-1} - \nu f^{-1}\right\|_{TV}.$$

□

We will use the following proposition to find a lower bound for the distance between the distribution of the chain at time t and the stationary distribution, i.e., a lower bound for $d(t)$. If we can bound $d(t)$ by a "large" number, i.e., if we show that for small $\delta > 0$, there is t^* (possibly depending on ε) such that $d(t^*) \geq 1 - \delta$ then we can conclude, since d is decreasing, that $t_{mix}(\varepsilon) > t^*$.

Recall that for a complex random variable Z, the expected value and variance with respect to a measure μ are respectively given by $E_\mu[Z] = E_\mu[\Re(Z)] + i\, E_\mu[\Im(Z)]$ and $Var[Z] = E_\mu[|Z|^2] - |E_\mu[Z]|^2 = E_\mu[Z\overline{Z}] - E_\mu[Z]\overline{E_\mu[Z]}$, where the real part, $\Re(Z)$, and the imaginary part, $\Im(Z)$, of Z are real random variables.

Proposition 1.2.7. *Let μ and ν be two probability distributions on a finite state space Ω and let $f : \Omega \to \mathbb{C}$ be a complex-valued function on Ω. Then,*

$$\|\mu - \nu\|_{TV} \geq 1 - 8\, \frac{\max\{Var_\mu(f), Var_\nu(f)\}}{|\,|E_\mu[f]| - |E_\nu[f]|\,|^2}. \tag{1.4}$$

Proof. Assume without loss of generality that $|E_\mu[f]| \geq |E_\nu[f]|$ and let

$$r := \frac{|\,|E_\mu[f]| - |E_\nu[f]|\,|}{\sigma_*},$$

where

$$\sigma_* := \sqrt{\max\{Var_\mu(f), Var_\nu(f)\}}.$$

If $I := (|E_\nu[f]| + \frac{r}{2}\sigma_*, \infty) \subseteq \mathbb{R}$ and $g := |f|$, by Markov's inequality, then (bear in mind that for two

complex numbers z, w, it holds $|z| - |w| \leq |\,|z| - |w|\,| \leq |z - w|$)

$$\begin{aligned}\nu g^{-1}(I) &= \nu(|E_\nu[f]| + \tfrac{r}{2}\sigma_* < g < \infty) = \nu(|f| - |E_\nu[f]| > \tfrac{r}{2}\sigma_*) \leq \nu(|f| - |E_\nu[f]| > \tfrac{r}{2}\sqrt{Var_\nu(f)}) \\ &\leq \nu(|f - E_\nu[f]| > \tfrac{r}{2}\sqrt{Var_\nu(f)}) \leq \frac{4}{r^2}\end{aligned}$$

and

$$\begin{aligned}\mu g^{-1}(I) &= \mu(|E_\nu[f]| + \tfrac{r}{2}\sigma_* < g < \infty) = \mu(|f| > |E_\mu[f]| - |\,|E_\nu[f]| - |E_\mu[f]|\,| + \tfrac{r}{2}\sigma_*) \\ &\geq \mu(|E_\mu[f]| - |f| < \tfrac{r}{2}\sigma_*) \geq \mu(|E_\mu[f]| - |f| < \tfrac{r}{2}\sqrt{Var_\mu(f)}) \\ &\geq \mu(|E_\mu[f] - f| < \tfrac{r}{2}\sqrt{Var_\mu(f)}) \geq 1 - \frac{4}{r^2}.\end{aligned}$$

Thus,

$$\left\|\mu g^{-1} - \nu g^{-1}\right\|_{TV} \geq |\mu g^{-1}(I) - \nu g^{-1}(I)| \geq 1 - \frac{8}{r^2}$$

and by Lemma 1.2.6, since $g : \Omega \to |f(\Omega)|$ with $f(\Omega)$ finite,

$$\|\mu - \nu\|_{TV} \geq \left\|\mu g^{-1} - \nu g^{-1}\right\|_{TV} \geq 1 - \frac{8}{r^2} = 1 - 8\,\frac{\max\{Var_\mu(f), Var_\nu(f)\}}{|\,|E_\mu[f]| - |E_\nu[f]|\,|^2}.$$

□

This result is particularly useful if we consider the Rudvalis Markov chain $\{\sigma_t\}_{t\in\mathbb{N}_0}$ and take $\mu := P^t(\sigma, \cdot)$ (μ is time dependent), the distribution of the state of the deck after t shuffles given that the initial state is $\sigma \in S_n$, and $\nu := U_n$, the uniform measure on S_n (ν does not depend of t). Indeed, by Proposition 1.2.7 we have

$$d(t) \geq \left\|P^t(\sigma, \cdot) - U_n\right\|_{TV} \geq 1 - \frac{8}{r^2} \tag{1.5}$$

where

$$r^2 = \frac{|\,|E_\mu[f(\sigma_t)]| - |E_\nu[f(\sigma_t)]|\,|^2}{\max\{Var_\mu(f(\sigma_t)), Var_\nu(f(\sigma_t))\}},$$

and we would just need to compute $E_\mu[f(\sigma_t)] := \mathbb{E}_\sigma[f(\sigma_t)], E_\nu[f(\sigma_t)], Var_\mu(f(\sigma_t))$ and $Var_\nu(f(\sigma_t))$.

When we look at (1.5), we realize that in order to get $d(t)$ close to 1 (which means that the chain has not mixed by time t), we would like the variances to be small and the difference of expectations to be large. But now the question is, how can we find a function $f : S_n \to \mathbb{C}$ which provides these properties? The main idea, as we will see, is to use the eigenfunctions of the transition matrix of a Markov chain.

The approach from [15] uses an eigenvector Φ (with eigenvalue λ) of the transition matrix of a Markov chain $\{X_t\}_{t\in\mathbb{N}_0}$ with initial distribution μ_0 and stationary distribution π. As seen in (1.3), since Φ is an eigenfunction, $\mathbb{E}_{\mu_0}[\Phi(X_{t+1}) \mid X_t] = \Phi(X_t)$. Moreover, we have already seen in Lemma 1.1.12 that by taking expectations on both sides of the last identity, we get $\mathbb{E}_{\mu_0}[\Phi(X_t)] = \lambda^t\Phi(X_0)$. In particular, if the eigenvalue λ verifies $|\lambda| < 1$, then when the chain approaches stationarity, $\mathbb{E}_{\mu_0}[\Phi(X_t)]$ approaches 0 (since $|\lambda|^t \underset{t}{\to} 0$). Furthermore, if the eigenvalue is close to 1, it will take a "long" time before $\mathbb{E}_{\mu_0}[\Phi(X_t)]$ is close to 0, which will make it easier for us to find a time in which the chain is distant from stationarity. If it is still true that $\mathbb{E}_{\mu_0}[\Phi(X_t)]$ is large, then $\mathbb{E}_{\mu_0}[\Phi(X_t)]$ is distinguishable from $E_\pi[\Phi(X_t)] = 0$ (Lemma

1.1.13), that is, the difference of expectations is large. In order to have a small variance, this eigenvector has to be such that $\mathbb{E}_{\mu_0}[\Phi(X_t)]$ is with high probability close to $\Phi(X_t)$.

Another possibility is to find an eigenvector Ψ not for the original chain, but for another chain (X_t, Y_t), where Y_t is obtained by taking into account more information about the shuffle. Wilson [15] refers to the chain $\{(X_t, Y_t)\}_{t\in\mathbb{N}_0}$ as the *lifted chain*. We want the eigenvector of the lifted chain to have the property that for all x, y_1, y_2, the eigenvector verifies $|\Psi(x, y_1)| = |\Psi(x, y_2)|$, so that $|\Psi(X_t, Y_t)|$ is a function of X_t alone. Now let us see how to define Y_t and then obtain an eigenfunction Ψ with the desired properties.

Recall that the state σ_t of the Rudvalis chain is the permutation giving the order of the cards at time t. It will be convenient to look at the position of a particular card. Hence denote the position of card k (with $k \in \{1, \cdots, n\}$) at time t by $X_t(k) := \sigma_t^{-1}(k)$, where $X_t(k) = 1$ if, at time t, card k is on the top of the deck, $X_t(k) = 2$ if, at time t, card k is at the second position and so on. Thus $\{X_t = (X_t(1), \ldots, X_t(n))\}_{t\in\mathbb{N}_0}$ is a Markov chain with state space S_n. In particular $\{X_t(k)\}_{t\in\mathbb{N}_0}$ is the random walk (on $\mathbb{Z}_n$) performed by card k. But we have to consider more information about the Rudvalis chain. We observe that when we place the top card in position n, every card is cyclically shifted counterclockwise (see Figure 1.3) and when we place the top card in position $n-1$, every card except the last is cyclically shifted (counterclockwise). Thus, at each time t, besides keeping track of the position of card k, $X_t(k)$, we will also count, with Y_t, the number of times the deck is shifted counterclockwise (modulo n). We shall consider the lifted chain $\{(X_t, Y_t)\}_{t\in\mathbb{N}_0}$ with state space $(S_n, \mathbb{Z}_n)$, where

$$\begin{cases} X_t(k) = \sigma_t^{-1}(k) \\ Y_t = t \pmod n. \end{cases}$$

With that in mind, given a card k, define

$$\Psi_k(X_t, Y_t) = v(X_t(k))e^{\frac{2\pi i}{n}Z_t(k)} \tag{1.6}$$

where

$$Z_t(k) = (X_t(k) - X_0(k) + Y_t) \mod n$$

and $v : \{1, \ldots, n\} \to \mathbb{C}$ is a function, to be determined later, which will turn, for all $k \in \{1, \cdots, n\}$, Ψ_k into an eigenfunction (for the lifted chain) with the same eigenvalue λ. Furthermore, define

$$\Psi(X_t, Y_t) = \sum_{k=1}^{n} \Psi_k(X_t, Y_t)$$

which is an eigenfunction with eigenvalue λ, since $\{\Psi_k\}_{k=1}^n$ are eigenfunctions with the same eigenvalue λ. This new eigenfunction Ψ will provide a better bound (when compared with a single eigenfunction Ψ_k). The idea is that if we only consider a single Ψ_k, we are only keeping track of a single card and the time it takes to reach the equilibrium. Ideally, we want to follow every card so that we know when the deck has been totally shuffled. For that reason, we consider the sum of the eigenfunctions for each card, obtaining a "better" eigenfunction.

Note that $Z_0(k) = 0$ and $Z_{t+1}(k) - Z_t(k) = X_{t+1}(k) - X_t(k) + 1$. The evolution of Z_t in time happens according to the two possible shuffles:

- if the top card is placed at position n (at time $t+1$),

$$Z_{t+1}(k) - Z_t(k) = -1 + 1 = 0, \text{ for } k \in \{1, \cdots, n\}. \tag{1.7}$$

- if the top card is placed at position $n-1$ (at time $t+1$),

$$Z_{t+1}(k) - Z_t(k) = \begin{cases} -1 & \text{if } X_t(k) = 1 \\ 1 & \text{if } X_t(k) = n \\ 0 & \text{if } 1 < X_t(k) < n. \end{cases} \tag{1.8}$$

Furthermore, note that

$$|\Psi_k(X_t, Y_t)| = |v(X_t(k))| \left| e^{\frac{2\pi i}{n} Z_t(k)} \right| = |v(X_t(k))|,$$

which means that for fixed x, and for y_1, y_2,

$$|\Psi_k(x, y_1)| = |\Psi_k(x, y_2)| = |v(x)|,$$

hence $|\Psi(X_t, Y_t)|$ is a function of X_t only, as we wanted. Finally, the function f to consider in Proposition 1.2.7 is

$$\begin{aligned} f : S_n &\to \mathbb{C} \\ \sigma_t &\mapsto \Psi(X_t, Y_t). \end{aligned}$$

A simple, yet useful, result to keep in mind when dealing with the big O notation is the following.

Lemma 1.2.8. *If $f : \mathbb{N} \to \mathbb{C}$ satisfies $f(n) = O(n^\alpha)$ for $\alpha < 0$, then sufficiently for large n,*

$$\frac{1}{1 + f(n)} = 1 + O(n^\alpha).$$

Proof. For sufficiently large n, $|f(n)| < 1$. Hence, doing a geometric series expansion,

$$\frac{1}{1 + f(n)} = \sum_{k=0}^{\infty} (-1)^k f(n)^k = 1 + O(n^\alpha).$$

□

We shall use this result without mentioning it.

Proposition 1.2.9. *The random walk performed by a card k under the lifted Rudvalis shuffle has an eigenvector of the form*

$$\Psi_k(x, y) = v(x) e^{2\pi i z/n},$$

where $v(x)$ is the x-th number in the list

$$\lambda^{n-2}, \ldots, \lambda^2, \lambda, 1, \chi,$$

the eigenvalue is

$$\lambda = 1 - \frac{p}{1-p}\frac{4\pi^2}{n^3} + O(\tfrac{1}{n^5}),$$

$$\chi = 1 + \frac{p}{1-p}\frac{2\pi i}{n} + O(\tfrac{1}{n^2}),$$

and $z = x - k + y$.

Proof. Let $w = e^{2\pi i/n}$ and note that from (1.6), $\Psi_k(x, y) = v(x)w^z$. If at time t, card k is at position $1 < x < n$, then from (1.7) and (1.8)

$$\Psi_k(X_{t+1}, Y_{t+1}) = v(X_{t+1}(k))w^{Z_{t+1}(k)} = \lambda v(X_t(k))w^{Z_t(k)} = \lambda\Psi_k(X_t, Y_t)$$

deterministically, because at time $t + 1$, card k will be at position $x - 1$. For the remaining cases, recall that the first card is placed at position $n - 1$ with probability p and in position n with probability $1 - p$.
If at time t, card k is at position 1, i.e., if $X_t(k) = 1$, then

$$v(X_t(k)) = \lambda^{n-2} \text{ and } v(X_{t+1}(k)) = \begin{cases} 1, \text{ with probability } p, \\ \chi, \text{ with probability } 1 - p. \end{cases}$$

Moreover, by (1.7) and (1.8),

$$Z_{t+1}(k) - Z_t(k) = \begin{cases} -1, \text{ with probability } p, \\ 0, \text{ with probability } 1 - p. \end{cases}$$

Thus,

$$\frac{E_\mu[\Psi_k(X_{t+1}, Y_{t+1}))|(X_t, Y_t)]}{\lambda\Psi_k(X_t, Y_t)} = \frac{p\ 1\ w^{-1} + (1-p)\ \chi\ w^0}{\lambda\ v(X_t(k))} = \frac{pw^{-1} + (1-p)\chi}{\lambda^{n-1}}.$$

If at time t, card k is at position n, i.e., if $X_t(k) = n$, then

$$v(X_t(k)) = \chi \text{ and } v(X_{t+1}(k)) = \begin{cases} \chi, \text{ with probability } p, \\ 1, \text{ with probability } 1 - p. \end{cases}$$

Moreover, by (1.7) and (1.8),

$$Z_{t+1}(k) - Z_t(k) = \begin{cases} 1, \text{ with probability } p, \\ 0, \text{ with probability } 1 - p. \end{cases}$$

Thus,

$$\frac{E_\mu[\Psi_k(X_{t+1}, Y_{t+1}))|(X_t, Y_t)]}{\lambda\Psi_k(X_t, Y_t)} = \frac{p\ \chi\ w^1 + (1-p)\ 1\ w^0}{\lambda v(X_t(k))} = \frac{pw\chi + (1-p)}{\lambda\chi}.$$

In order to have

$$E_\mu[\Psi_k(X_{t+1}, Y_{t+1})|(X_t, Y_t)] = \lambda\Psi_k(X_t, Y_t),$$

(ensuring that Ψ_k is an eigenfunction) we must have

$$\begin{cases} pw^{-1} + (1-p)\chi = \lambda^{n-1} \\ pw\chi + (1-p) = \lambda\chi \end{cases} \Leftrightarrow \begin{cases} \chi = \frac{\lambda^{n-1} - pw^{-1}}{1-p} \\ \chi = \frac{1-p}{\lambda - pw} \end{cases}$$

and combining both equations for χ, we get $\lambda^n - pw\lambda^{n-1} - pw^{-1}\lambda - 1 + 2p = 0$. Now let

$$q(x) = x^n - pwx^{n-1} - pw^{-1}x - 1 + 2p$$

be a polynomial which has λ as a zero. In order to find λ, we apply Newton's method and start from a root close to 1 (recall that we want λ to be close to 1 so that $\mathbb{E}_{\mu_0}[\Psi_k(X_t, Y_t)]$ decreases "slowly" to 0). Starting with $x_0 = 1$, we iterate for $i \geq 0$,

$$x_{i+1} = x_i - \frac{q(x_i)}{q'(x_i)}. \tag{1.9}$$

By Taylor's Theorem with Lagrange's remainder (cf. [1], page 880),

$$q(x_{i+1}) = q(x_i) + q'(x_i)(x_{i+1} - x_i) + \frac{1}{2}q''(\xi_u)(x_{i+1} - x_i)^2$$

for some $\xi_u = ux_i + (1-u)x_{i+1}$ and some $u \in (0,1)$. Moreover,

$$q(x_i) + q'(x_i)(x_{i+1} - x_i) = q(x_i) + q'(x_i)\left(-\frac{q(x_i)}{q'(x_i)}\right) = 0.$$

Therefore,

$$|q(x_{i+1})| = \frac{1}{2}|q''(\xi_u)|\,|x_{i+1} - x_i|^2 \leq \frac{1}{2}\|q''\|_\infty \left|\frac{q(x_i)}{q'(x_i)}\right|^2, \tag{1.10}$$

where $\|q''\|_\infty = \max_{x\in\mathbb{C}} |q''(x)|$. If $|x-1| = O(\frac{1}{n^2})$ then,

$$q'(x) = nx^{n-1} - (n-1)pwx^{n-2} - pw^{-1} = n - (n-1)p + O(1) = (1-p)n + O(1)$$

and

$$\begin{aligned} q''(x) &= n(n-1)x^{n-2} - (n-1)(n-2)pwx^{n-3} = n(n-1) - (n-1)(n-2)p + O(n) \\ &= (1-p)n^2 + O(n). \end{aligned}$$

By (1.10), if $|x_i - 1| = O(\frac{1}{n^2})$ and $|x_{i+1} - 1| = O(\frac{1}{n^2})$,

$$|q(x_{i+1})| \leq \frac{1}{2}\frac{(1-p)n^2 + O(n)}{(1-p)^2n^2}\,|q(x_i)|^2 = \frac{1 + O\left(\frac{1}{n}\right)}{2(1-p)}\,|q(x_i)|^2. \tag{1.11}$$

Now note that

$$w = 1 + \tfrac{2\pi i}{n} - \tfrac{4\pi^2}{2n^2} - \tfrac{8\pi^3 i}{6n^3} + O(\tfrac{1}{n^4})$$
$$w^{-1} = 1 - \tfrac{2\pi i}{n} - \tfrac{4\pi^2}{2n^2} + \tfrac{8\pi^3 i}{6n^3} + O(\tfrac{1}{n^4})$$

which implies that

$$q(x_0) = p(2 - w - w^{-1}) = p\tfrac{4\pi^2}{n^2} + O(\tfrac{1}{n^4}).$$

We claim that for all $i \in \mathbb{N}_0$ and n sufficiently large,

$$|q(x_i)| \leq (1-p)\left(\tfrac{p}{(1-p)}\tfrac{4\pi^2}{n^2}\right)^{2^i} + O(\tfrac{1}{n^4}) \tag{1.12}$$

and

$$|x_{i+1} - 1| = O(\tfrac{1}{n^2}). \tag{1.13}$$

We can verify both facts by induction on $i \in \mathbb{N}_0$:

- for $i = 0$, the inequality holds on account of the computation of $q(x_0)$ and

$$|x_1 - 1| = |x_1 - x_0| \leq \frac{|q(x_0)|}{|q'(x_0)|} = \frac{p\frac{4\pi^2}{n^2} + O(\frac{1}{n^4})}{(1-p)n + O(1)} = O(\tfrac{1}{n^2}).$$

- Now suppose (1.13) and (1.12) hold for $i \in \mathbb{N}_0$. By (1.11), and for n sufficiently large,

$$\begin{aligned}
|q(x_{i+1})| &\leq \frac{1 + O\left(\frac{1}{n}\right)}{2(1-p)}\,|q(x_i)|^2 \leq \frac{1 + O\left(\frac{1}{n}\right)}{2(1-p)}\left((1-p)\left(\tfrac{p}{(1-p)}\tfrac{4\pi^2}{n^2} + O(\tfrac{1}{n^4})\right)^{2^i}\right)^2 \\
&= \frac{1 + O\left(\frac{1}{n}\right)}{2}(1-p)\left(\tfrac{p}{(1-p)}\tfrac{4\pi^2}{n^2}\right)^{2^{i+1}} + O(\tfrac{1}{n^4}) \\
&\leq (1-p)\left(\tfrac{p}{(1-p)}\tfrac{4\pi^2}{n^2}\right)^{2^{i+1}} + O(\tfrac{1}{n^4}).
\end{aligned}$$

  Moreover,

$$\begin{aligned}
|x_{i+2} - 1| &\leq |x_{i+2} - x_{i+1}| + |x_{i+1} - 1| \leq \frac{|q(x_{i+1})|}{|q(x_i)|} + O(\tfrac{1}{n^2}) \\
&\leq \frac{(1-p)\left(\tfrac{p}{(1-p)}\tfrac{4\pi^2}{n^2}\right)^{2^{i+1}} + O(\tfrac{1}{n^4})}{(1-p)n + O(1)} = O(\tfrac{1}{n^2}).
\end{aligned}$$

Using (1.9) and (1.12), for $i \in \mathbb{N}_0$,

$$\begin{aligned}
|x_{i+1} - x_i| = \left|\frac{q(x_i)}{q'(x_i)}\right| &\leq \left((1-p)\left(\frac{p}{(1-p)}\frac{4\pi^2}{n^2}\right)^{2^i} + O(\tfrac{1}{n^4})\right)\frac{1}{(1-p)n + O(1)} \\
&= \left(\frac{p}{(1-p)}\frac{4\pi^2}{n^2}\right)^{2^i}\frac{1}{n + O(1)} + O(\tfrac{1}{n^4})
\end{aligned}$$

and thus we can conclude that

$$|x_{i+1} - x_0| \leq \sum_{j=0}^{i} |x_{j+1} - x_j| = \frac{1}{n + O(1)} \sum_{j=0}^{i} \left(\frac{p}{(1-p)} \frac{4\pi^2}{n^2} \right)^{2^j} = \frac{1}{n + O(1)} \, O(\tfrac{1}{n^2}) = O(\tfrac{1}{n^3}).$$

Thus, for n sufficiently large, the sequence $\{x_i\}_{i \in \mathbb{N}_0}$ converges to a point λ. Since q is continuous, this limit point must be a root of q. After the first iteration of Newton's method we get

$$x_1 = x_0 - \frac{q(x_0)}{q'(x_0)} = 1 - \frac{\frac{p4\pi^2}{n^2} + O\left(\frac{1}{n^4}\right)}{(1-p)n + O(1)} = 1 - \frac{p}{(1-p)} \frac{4\pi^2}{n^3} + O(\tfrac{1}{n^5}).$$

Since,

$$|\lambda - x_1| \leq \sum_{j=1}^{\infty} |x_{j+1} - x_j| = \frac{1}{n + O(1)} \sum_{j=1}^{\infty} \left(\frac{p}{(1-p)} \frac{4\pi^2}{n^2} \right)^{2^j} = O(\tfrac{1}{n^5})$$

we conclude that the sequence must converge to

$$\lambda = 1 - \frac{p}{(1-p)} \frac{4\pi^2}{n^3} + O(\tfrac{1}{n^5}).$$

Hence we have just found an eigenvalue λ with the properties we wanted. Plugging this value in the first expression for χ, we get that

$$\begin{aligned} \chi &= \frac{\lambda^{n-1} - pw^{-1}}{1-p} = \frac{(1 + O(\frac{1}{n^3}))^{n-1} - p(1 + \frac{2\pi i}{n} + O(\frac{1}{n^2}))}{1-p} = \frac{(1-p) - p\frac{2\pi i}{n} + O(\frac{1}{n^2})}{1-p} \\ &= 1 - \frac{p}{1-p} \frac{2\pi i}{n} + O(\tfrac{1}{n^2}). \end{aligned}$$

□

In order to use Proposition 1.2.7, we need to compute the expected value and variance of $\Psi(X_t, Y_t)$. The expectation can be easily computed, but the variance demands more effort, which is why we compute instead $E_\mu[|\Psi(X_{t+1}) - \Psi(X_t)|^2]$ and use the following Lemma.

Lemma 1.2.10. *Let $\{X_t\}_{t \in \mathbb{N}_0}$ be a Markov chain with finite state space Ω, with eigenfunction Ψ and eigenvalue λ, and let μ be a probability distribution on Ω. If $\Re(\lambda) \geq \frac{1}{2}$ and $|\lambda| < 1$, then*

$$Var_\mu[\Psi(X_t)] \leq \frac{E_\mu[|\Psi(X_{t+1}) - \Psi(X_t)|^2]}{2(1 - \Re(\lambda))}.$$

Proof. Let $\Psi_t = \Psi(X_t)$, $\tilde{\nabla}\Psi_{t+1} = \Psi_{t+1} - \Psi_t$, $V_{t+1} = E_\mu[|\tilde{\nabla}\Psi_{t+1}|^2 | X_t]$ and denote by $\overline{\Psi}$ the conjugate of Ψ. Since Ψ is an eigenvector with eigenvalue λ, by (1.3)

$$E_\mu[\tilde{\nabla}\Psi_{t+1} | X_t] = E_\mu[\Psi_{t+1} | X_t] - \Psi_t = (\lambda - 1)\Psi_t.$$

Since

$$\begin{aligned} \Psi_{t+1}\overline{\Psi}_{t+1} &= \Psi_t\overline{\Psi}_t + \Psi_t(\overline{\Psi}_{t+1} - \overline{\Psi}_{t+1}) + \overline{\Psi}_t(\Psi_{t+1} - \Psi_t) + (\Psi_{t+1} - \Psi_t)(\overline{\Psi}_{t+1} - \overline{\Psi_t}) \\ &= \Psi_t\overline{\Psi}_t + \Psi_t\tilde{\nabla}\overline{\Psi}_{t+1} + \overline{\Psi}_t\tilde{\nabla}\Psi_{t+1} + |\tilde{\nabla}\Psi_{t+1}|^2, \end{aligned}$$

we have

$$\begin{aligned} E_\mu[\Psi_{t+1}\overline{\Psi}_{t+1}|X_t] &= \Psi_t\overline{\Psi}_t + \Psi_t\overline{(\lambda-1)}\Psi_t + \overline{\Psi}_t(\lambda-1)\Psi_t + E_\mu[|\tilde{\nabla}\Psi_{t+1}|^2|X_t] \\ &= (2\Re(\lambda)-1)\Psi_t\overline{\Psi}_t + V_{t+1}. \end{aligned} \tag{1.14}$$

By induction on t, we have that

$$E_\mu[\Psi_t\overline{\Psi}_t] \leq (2\Re(\lambda)-1)^t\Psi_0\overline{\Psi}_0 + \frac{E_\mu[V_{t+1}]}{2-2\Re(\lambda)}. \tag{1.15}$$

Indeed,

- for $t=0$, since $V_1 > 0$, $\Re(\lambda) \leq |\lambda| < 1$ and Ψ_0 is deterministic, we have

$$E_\mu[\Psi_0\overline{\Psi}_0] = \Psi_0\overline{\Psi}_0 \leq \Psi_0\overline{\Psi}_0 + \frac{E_\mu[V_1]}{2-2\Re(\lambda)}.$$

- Now suppose the inequality holds for $t \in \mathbb{N}_0$. Then, by (1.14) and by the Tower Law property,

$$\begin{aligned} E_\mu[\Psi_{t+1}\overline{\Psi}_{t+1}] &= E_\mu[E_\mu[\Psi_{t+1}\overline{\Psi}_{t+1}|X_t]] = E_\mu[(2\Re(\lambda)-1)\Psi_t\overline{\Psi}_t + V_{t+1}] \\ &\leq (2\Re(\lambda)-1)\left((2\Re(\lambda)-1)^t\Psi_0\overline{\Psi}_0 + \frac{E_\mu[V_{t+1}]}{2-2\Re(\lambda)}\right) + E_\mu[V_{t+1}] \\ &\leq (2\Re(\lambda)-1)^{t+1}\Psi_0\overline{\Psi}_0 + \frac{E_\mu[V_{t+1}]}{2-2\Re(\lambda)}. \end{aligned}$$

  By Lemma 1.1.12, $E_\mu[\Psi_t] = \lambda^t\Psi_0$. Subtracting $E_\mu[\Psi_t]\overline{E_\mu[\Psi_t]}$ on both sides of (1.15), we get that for $t \in \mathbb{N}_0$ (recall the definition of the variance of a complex random variable, above Proposition 1.2.7)

$$\begin{aligned} Var_\mu[\Psi_t] &= E_\mu[\Psi_t\overline{\Psi_t}] - E_\mu[\Psi_t]\overline{E_\mu[\Psi_t]} \\ &\leq (2\Re(\lambda)-1)^t\Psi_0\overline{\Psi}_0 + \frac{E_\mu[V_{t+1}]}{2-2\Re(\lambda)} - \lambda^t\Psi_0\overline{(\lambda^t\Psi_0)} \\ &= ((2\Re(\lambda)-1)^t - (\lambda\overline{\lambda})^t)\Psi_0\overline{\Psi}_0 + \frac{E_\mu[V_{t+1}]}{2-2\Re(\lambda)} \\ &\leq \frac{E_\mu[V_{t+1}]}{2(1-\Re(\lambda))} \end{aligned}$$

  where the last inequality comes from the fact that $\Psi_0\overline{\Psi}_0 \geq 0$ and

$$(\lambda\overline{\lambda})^t \geq (2\Re(\lambda)-1)^t \Leftrightarrow (\lambda\overline{\lambda}) \geq (2\Re(\lambda)-1) \Leftrightarrow (1-\lambda)(1-\overline{\lambda}) \geq 0,$$

  which holds for any $t \in \mathbb{N}_0$, since $2\Re(\lambda)-1 \geq 0$ and $(1-\lambda)(1-\overline{\lambda}) \geq 0$ for any complex number λ.

□

We need to compute the eigenvector at time $t=0$ in order to get its expected value. The next lemma illustrates that simple computation. In order to find a lower bound for $t_{mix}(\varepsilon)$, we can find a lower bound for $d(t)$. By the definition of $d(t)$, if we pick an initial configuration σ, a lower bound for

$\|P^t(\sigma,\cdot) - U_n\|_{TV}$ gives a lower bound for $d(t)$. Looking at (1.4), we want to start from a configuration which maximizes the expectation of $\Psi(X_0, Y_0)$. In our case, we show that this value does not depend on the initial configuration of the deck, which means that starting from any given configuration of the deck, we will obtain the same lower bound.

Lemma 1.2.11. *Let $\{(X_t, Y_t)\}_{t\geq 0}$ be the lifted Markov chain described above and let Ψ be the eigenvector with eigenvalue λ. Then,*

$$\Psi(X_0, Y_0) = n + O(\tfrac{1}{n}).$$

In particular, this value is independent from the initial configuration of the deck.

Proof. Let $X_0 = \sigma_0^{-1}$ for some initial configuration σ_0 of the deck. Then,

$$\begin{aligned}\Psi(X_0, Y_0) &= \sum_{k=1}^{n} \Psi_k(X_0, Y_0) = \sum_{k=1}^{n} v(X_0(k)) w^{Z_0(k)} = \sum_{k=1}^{n} v(X_0(k)) \\ &= 1 + \chi + \lambda + ... + \lambda^{n-2} = 1 + (1 + O(\tfrac{1}{n})) + (1 + O(\tfrac{1}{n^3})) + ... + (1 + O(\tfrac{1}{n^3})) \\ &= n + O(\tfrac{1}{n}).\end{aligned}$$

In particular, since $X_0(k)$ is the position of card k in the initial configuration, it ranges over $\{1, ..., n\}$ and thus the value of $\Psi(X_0, Y_0)$ does not depend on the initial configuration of the deck. □

Now we can compute the expected value and the variance of $\Psi(X_t, Y_t)$.

Lemma 1.2.12. *Let (X_t, Y_t) be the lifted Markov chain described above, $\Psi = \sum_{k=1}^{n} \Psi_k$ the eigenvector (with eigenvalue λ) obtained from Proposition 1.2.9 and $\mu_t^\sigma := P^t(\sigma,\cdot)$. Then, for any $t \in \mathbb{N}_0$, the expectation and the variance of the complex random variable $\Psi(X_t, Y_t)$ under the probability distribution μ_t^σ are given, respectively, by*

$$E_{\mu_t^\sigma}[\Psi(X_t, Y_t)] = \lambda^t \Psi(X_0, Y_0) = \lambda^t n + O(\tfrac{1}{n}) \quad \textit{and} \quad Var_{\mu_t^\sigma}[\Psi(X_t, Y_t)] = O(n).$$

Proof. By Lemma 1.1.12, $E_{\mu_t^\sigma}[\Psi(X_t, Y_t)] = \lambda^t \Psi(X_0, Y_0)$. By the previous lemma, the first result follows. Let $\Phi_{t,k}$ be the (complex) random variable given by

$$\Phi_{t,k} = \frac{\Psi_k(X_{t+1}, Y_{t+1}) - \Psi_k(X_t, Y_t)}{w^{Z_t(k)}} = v(X_{t+1}(k)) w^{Z_{t+1}(k) - Z_t(k)} - v(X_t(k)).$$

Given the position of card k (at time t) and where we place the top card, we can express $\Phi_{t,k}$ deterministically:

- if the top card is placed at position n, then

$$\Phi_{t,k} = \begin{cases} \chi w^0 - \lambda^{n-2} = O(\tfrac{1}{n}) & \text{if } X_t(k) = 1, \\ w^0 - \chi = O(\tfrac{1}{n}) & \text{if } X_t(k) = n, \\ (\lambda - 1)\lambda^{X_t(k)} = O(\tfrac{1}{n^3}) & \text{if } 1 < X_t(k) < n. \end{cases}$$

- if the top card is placed at position $n-1$, then

$$\Phi_{t,k} = \begin{cases} w^{-1} - \lambda^{n-2} = O(\frac{1}{n}) & \text{if } X_t(k) = 1, \\ w^1\chi - \chi = \chi(w-1) = O(\frac{1}{n}) & \text{if } X_t(k) = n, \\ (\lambda-1)\lambda^{X_t(k)} = O(\frac{1}{n^3}) & \text{if } 1 < X_t(k) < n. \end{cases}$$

Since $|w| = 1$,

$$\begin{aligned} |\Psi(X_{t+1}, Y_{t+1}) - \Psi(X_t, Y_t)| &\leq \sum_{k=1}^{n} |\Psi_k(X_{t+1}, Y_{t+1}) - \Psi_k(X_t, Y_t)| = \sum_{k=1}^{n} |\Phi_{t,k}| \, |w^{Z_t(k)}| \\ &= 2\, O(\tfrac{1}{n}) + (n-2) O(\tfrac{1}{n^3}) = O(\tfrac{1}{n}) \end{aligned}$$

and consequently

$$E_{\mu_t^\sigma}[|\Psi(X_{t+1}, Y_{t+1}) - \Psi(X_t, Y_t)|^2] = O(\tfrac{1}{n^2}).$$

By Lemma 1.2.10,

$$Var_{\mu_t^\sigma}[\Psi(X_t, Y_t)] \leq \frac{O(\frac{1}{n^2})}{2(1 - \Re(\lambda))} = \frac{O(\frac{1}{n^2})}{2\frac{p}{1-p}\frac{4\pi^2}{n^3} + O(\frac{1}{n^5})} = O(n).$$

□

We can finally find a lower bound for the mixing time of the Rudvalis shuffle.

Theorem 1.2.13 (Lower bound on the mixing time of the Rudvalis shuffle)**.** *For the Rudvalis shuffle with n cards,*

$$d(t) \geq 1 - 8e^{-2t \log|\lambda|} \frac{O(1)}{n + O(\frac{1}{n})},$$

where

$$\lambda = 1 - \frac{p}{1-p}\frac{4\pi^2}{n^3} + O(\tfrac{1}{n^5}).$$

Furthermore, for $\varepsilon \in (0,1)$,

$$t_{mix}(\varepsilon) \geq \frac{1-p}{p}\frac{1}{8\pi^2} n^3 \log(n) + O(n^3).$$

Proof. Let $\mu_t^\sigma := P^t(\sigma, \cdot)$ be the distribution of the deck at time t given that the initial configuration of the deck is σ, and let U_n be the uniform distribution on S_n. Let $\Psi_t := \Psi(X_t, Y_t)$ be the eigenfunction with eigenvalue λ determined in Proposition 1.2.9. If we take $f(\sigma_t) := \Psi_t$ in Proposition 1.2.7,

$$\|\mu_t^\sigma - U_n\|_{TV} \geq 1 - 8\, \frac{\max\{Var_{\mu_t^\sigma}(\Psi_t), Var_{U_n}(\Psi_t)\}}{|\,|E_{\mu_t^\sigma}[\Psi_t]| - |E_{U_n}[\Psi_t]|\,|^2}. \tag{1.16}$$

We have seen in the proof of Lemma 1.2.12 that $|E_{\mu_t^\sigma}[\Psi_t]| = |\lambda|^t |\Psi_0|$ and $Var_{\mu_t^\sigma}[\Psi_t] = O(n)$. In particular, the variance does not depend on t. Hence $\max\{Var_{\mu_t^\sigma}(\Psi_t), Var_{U_n}(\Psi_t)\} = O(n)$. Furthermore, by

Lemma 1.1.13 , $E_{U_n}[\Psi_t] = E_{U_n}[\Psi] = 0$. Plugging these values in (1.16), we obtain

$$\|\mu_t^\sigma - U_n\|_{TV} \geq 1 - 8\,\frac{O(n)}{|\,|\lambda|^t|\Psi_0| - 0|^2} = 1 - 8|\lambda|^{-2t}\,\frac{O(n)}{|\Psi_0|^2} = 1 - 8e^{-2t\log|\lambda|}\,\frac{O(1)}{n + O(\frac{1}{n})}.$$

Hence

$$d(t) = \max_{\sigma\in S_n} \|\mu_t^\sigma - U_n\|_{TV} \geq 1 - 8e^{-2t\log|\lambda|}\,\frac{O(1)}{n + O(\frac{1}{n})}.$$

Since $-\log|\lambda| > 0$, we can see that for fixed n, for "small" t the distance $d(t)$ stays close to 1 whereas when t increases, the lower bound keeps getting worse. For sufficiently large n, there is a constant k such that

$$d(t) \geq 1 - \frac{8k}{n}e^{-2t\log|\lambda|}.$$

For $\varepsilon \in (0,1)$, solving $1 - \frac{8k}{n}e^{-2t\log|\lambda|} = \varepsilon$, we get

$$t = \frac{\log(n)}{-2\log|\lambda|} + \frac{\log\left(\frac{1-\varepsilon}{8k}\right)}{-2\log|\lambda|}.$$

Owing to the fact that $|\lambda| = 1 - \gamma + O(\frac{1}{n^5})$ with $\gamma = \frac{p}{1-p}\frac{4\pi^2}{n^3}$, we have that for n sufficiently large, $-\log|\lambda| = \gamma + O(\frac{1}{n^5}) > 0$ and consequently

$$t = \frac{1-p}{p}\frac{1}{8\pi^2}n^3\log(n) + O(n^3).$$

Above, note that ε is inside of $O(n^3)$. Since $d(t)$ is decreasing, this implies that by doing

$$t^*(\varepsilon) = \frac{1-p}{p}\frac{1}{8\pi^2}n^3\log(n) + O(n^3)$$

number of shuffles, the total variation distance from uniformity is greater or equal than ε, i.e., $t_{mix}(\varepsilon) \geq t^*(\varepsilon)$. □

Remark 1.2.14. *A generalization of the Rudvalis shuffle is the top to bottom-k shuffle which places the first card uniformly at random in one of the bottom k_n positions (for $k_n = 2$, we have the Rudvalis shuffle). In [6] we can see that if k_n is a constant, we would still have a lower bound (and upper bound) of order $n^3\log(n)$.*

Chapter 2

Hydrodynamic limit for the continuous time Rudvalis shuffle

Now that we have studied the discrete time Rudvalis Markov chain, it is time to extend it to continuous time. In this chapter, we focus on the case $p = \frac{1}{2}$ (to simplify the computations and notation). Anyhow, a general $0 < p < 1$ would lead to the same result (see the remark after Lemma 2.6.3). After defining the process, we will convert it into a particle system and study the hydrodynamic limit of this system. Moreover, we shall prove the existence and uniqueness of a (weak) solution of a PDE, the transport equation, and this is the main goal of this chapter.

As we will see, the hydrodynamic limit can be seen as the convergence (in probability) of a random measure to an absolutely continuous measure (with respect to the Lebesgue measure) whose density is the unique weak solution of a PDE. If we prove this convergence together with the uniqueness of the solution, we will have proved that there exists a unique weak solution of that PDE [7].

To prove this convergence, the entropy method is employed. First, we prove that the sequence of probability measures (associated with the random measures of the density) is relatively compact, which tells us that every subsequence has a weakly convergent subsequence. Then, we characterize their limit points by showing that they are unique. Since all the (weakly) convergent subsequences converge to the same limit point, the whole sequence converges to that unique limit point.

2.1 Rudvalis process

Let us define a continuous-time version of the Rudvalis Markov chain. Let us consider a Poisson process $\{\mathcal{T}_k : k \in \mathbb{N}\}$ of rate 1, that is, $\mathcal{T}_0 = 0$ and $\{(\mathcal{T}_k - \mathcal{T}_{k-1}) : k \in \mathbb{N}\}$ is a sequence of i.i.d. exponential random variables with mean 1 which is responsible for giving our process the Markov property. The process $\{\mathcal{T}_k : k \in \mathbb{N}\}$ is known as the *Harris process* or the *clock process* (see [9] for instance). It is important to observe that in Harris' construction, the probability that two clocks ring simultaneously is equal to 0. At each time a clock rings, we toss a fair coin (since we fixed $p = \frac{1}{2}$). If the coin lands heads up, then we remove the top card from the deck (the card at position 1) and insert it at position n. If the

coin lands tails up, then we remove the top card from the deck and insert it at position $n-1$. In particular, every card performs a continuous-time random walk on the deck.

The continuous-time Rudvalis shuffle is the Markov process $\{\sigma_t : t \geq 0\}$ with state space S_n and *infinitesimal generator* (cf. Chapter 2 of [14]) given on $f : S_n \to \mathbb{R}$, by

$$\mathcal{L}_n^R f(\sigma) = \frac{1}{2} \sum_{x=n-1}^{n} \Theta_x f(\sigma),$$

where

$$\Theta_x f(\sigma) = f(\sigma^x) - f(\sigma),$$

and σ^x is the permutation obtained from σ by placing the first card at position x.

Denote by $\mathbb{T} = \mathbb{R}/\mathbb{Z} = [0, 1)$ the one-dimensional continuous torus (macroscopic space), by $\mathbb{T}_n = \mathbb{Z}/n\mathbb{Z} = \{0, 1, \ldots, n-1\}$ the one-dimensional discrete torus with n points (microscopic space) and by $\Omega_n = \{0, 1\}^{\mathbb{T}_n}$ the space of functions from $\mathbb{T}_n$ to $\{0, 1\}$.

Recall that $\sigma(x) = y$ if, and only if, card y is at position x in permutation σ. Now we will color the cards, so, in order to identify the color of the card at a position x, we define $\eta(x) = \mathbb{1}_{\{\sigma(x) \text{ is black}\}}$. Thus, the projection $\sigma \mapsto P(\sigma) = \eta$ induces a Markov process with state space Ω_n, which we shall call the *space of configurations*, and with infinitesimal generator given on $f : \Omega_n \to \mathbb{R}$, by

$$\mathcal{L}_n f(\eta) = \frac{1}{2} \sum_{x=n-1}^{n} \Theta_x f(\eta),$$

where

$$\Theta_x f(\eta) = f(\eta^x) - f(\eta),$$

and $\eta^x = P(\sigma^x)$.

Now we can think of a simple, yet remarkable, mapping. Taking each position of the deck as a site, the black cards as particles and the red cards as empty sites, a particle system emerges!

Remark 2.1.1. *For $x \in \mathbb{T}_n$ and $\eta \in \Omega_n$, $\eta(x)$ is the number of particles at site x:*

$$\eta(x) = \begin{cases} 1, & \text{if card } x \text{ is black (site } x \text{ is occupied)} \\ 0, & \text{if card } x \text{ is red (site } x \text{ is empty)} \end{cases}.$$

Example 2.1.2. *Consider a simple example of a deck of $n = 5$ cards.*

Figure 2.1: The configuration of the particle system corresponding to a deck with 5 cards.

The top card of the deck, the 3 of hearts, is red, so we do not assign a particle to site 1, the king of

clubs is black, so we assign a particle to site 2*. We keep doing this procedure until there are no more cards. Moreover, taking the shuffle of Figure 1.3 as an example, we can see a particle system in motion.*

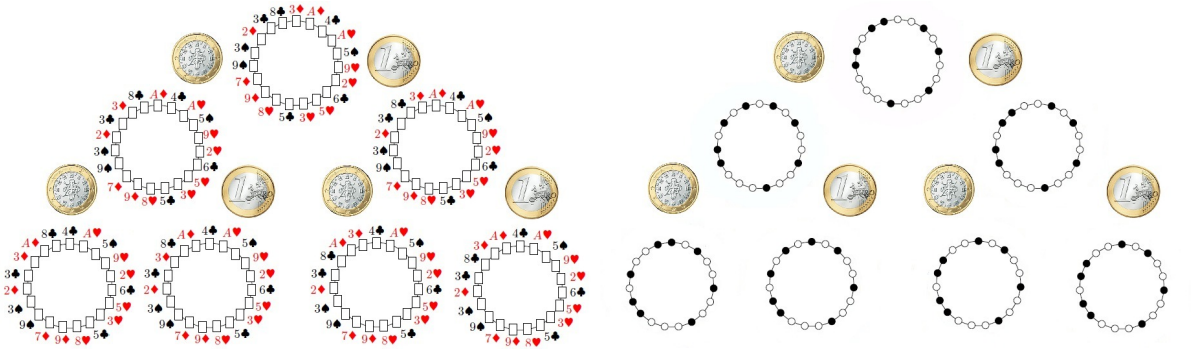


Figure 2.2: Particle system in motion obtained from the Rudvalis shuffle.

Our goal is to study the evolution in space and time of this system. In order to restrict the system to $\mathbb{T}$, we shall consider the process evolving in $\mathbb{T}_n/n$ and then we will take $n \to \infty$. We can identify the microscopic space with $\mathbb{T}_n$ (discrete) and the macroscopic space with $\mathbb{T}$ (continuous). A site x in the microscopic space $\mathbb{T}_n$, can be identified with $\frac{x}{n}$ in the macroscopic space $\mathbb{T}$. This means that the particles move between the sites $0, \frac{1}{n}, \frac{2}{n}, ..., 1$. If we let n go to ∞, then we can think of an initial density of particles, denoted by $\gamma : \mathbb{T} \to [0,1]$. Moreover, if we let the system evolve through time, we can wonder what function describes the density of particles at each time, given that the initial density is γ. As we will see, this function will be given by a weak solution (proved to be unique) of some PDE. This kind of behaviour of the density is the so-called hydrodynamic limit.

Observe that applying the generator $\mathcal{L}_n^R$ to functions that only depend on the colors of the cards is equivalent to applying the generator $\mathcal{L}_n$. But if we want to see a non trivial evolution of the system when $n \to \infty$, speeding up the process is absolutely necessary. This is justified by the fact that we want to divide macroscopic space $\mathbb{T}$ into n intervals of size $\frac{1}{n}$, that is, $\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}\} = \mathbb{T}_n/n$, and then let $n \to \infty$. For this reason, if we want to see an evolution of the initial density of particles $\gamma : \mathbb{T} \to [0,1]$ in the macroscopic space, we will need to accelerate the process by a function of n. For example, consider the process with generator $\mathcal{L}_n$ and imagine that a particle moves from site k_1 to site k_2 with $k_1, k_2 \in \mathbb{T}_n$. If k_1 and k_2 do not depend on n then, although we see an evolution in the microscopic system (a particle moves from one site to another), when we look at the macroscopic space (by letting $n \to \infty$), it appears as if the particle stayed at the same place, since $\frac{k_1}{n}, \frac{k_2}{n} \to 0$. According to the Rudvalis shuffle, in the microscopic system a particle will take $O(n)$ shuffles (time) to return to the same place. Let us go back to the deck of cards and see how much time it takes for a card to return to its original position. This value will tell us the factor by which we need to scale the time. Consider for example the top card (position 1). That card goes to position $n-1$ with probability $\frac{1}{2}$ and to position n with probability $\frac{1}{2}$. If the card goes to position n, then, in the next coin flip, it stays there with probability $\frac{1}{2}$ (if the first card is inserted at position $n-1$) and goes to position $n-1$ with probability $\frac{1}{2}$ (if the first card is inserted at position n). Hence, at each shuffle, the card has probability $\frac{1}{2}$ to go to position $n-1$ (if that position was not yet reached). It

follows that the number of shuffles in which the top card reaches position $n-1$ is a geometric random variable with parameter $\frac{1}{2}$. Consequently, the average number of shuffles in which the top card reaches the second-to-last position is 2. Once the card is in position $n-2$, it reaches the top of the deck in $n-2$ shuffles (deterministically). We conclude that the top card takes on average n shuffles to return to its initial position. Similarly, every card takes on average n shuffles to return to its original position. For this reason, from now on we will speed up the process in the hyperbolic time scale tn.

We will still denote by $\{\eta_t\}_{t\geq 0}$ the Markov process with generator $n\mathcal{L}_n$ and denote the initial measure by μ_n. Owing to the fact that it was derived from the Rudvalis card shuffle, we call the process $\{\eta_t\}_{t\geq 0}$ the Rudvalis process.

Remark 2.1.3. *Note that $\eta_t \in \Omega_n$ is already in the hyperbolic time scale, i.e., it is already accelerated by n. For this reason, these processes are often denoted by η_{tn}. However, to simplify the notation, we just write η_t but it should always be kept in mind that the Rudvalis process depends on n, which is the number of sites (cards), and that this process is already accelerated by n.*

In order to study $\{\eta_t\}_{t\geq 0}$, we start by defining the space where our process lives.

Definition 2.1.4 (Space $D_X[0,T]$)**.** *Let $T>0$ and X be a metric space. Denote by $D_X[0,T]$ the space of functions $x:[0,T]\to X$ that are right-continuous and have left-hand limits:*

(i) for $0\leq t<T$, $x(t^+)=\lim_{s\downarrow t} x(s)$ exists and $x(t^+)=x(t)$;

(ii) for $0<t\leq T$, $x(t^-)=\lim_{s\uparrow t} x(s)$ exists;

and which is endowed with the Skorohod topology (A.3).

Remark 2.1.5. *Functions satisfying the two properties above are called càdlàg (French: "continue à droite, limite à gauche", meaning "continuous on the right, limit on the left") functions. We call the elements of $D_X[0,T]$ trajectories and denote them by $x_\cdot$, i.e., for each $t\in[0,T]$, we denote $x(t)$ by x_t.*

The trajectories of the Rudvalis process $\{\eta_t\}_{t\geq 0}$ belong to $D_{\Omega_n}[0,T]$, i.e., for each time t, η_t is an element of Ω_n. Moreover, the function $t\mapsto\eta_t$ is càdlàg (it has jumps).

Let $\mathbb{P}_{\mu_n}$ be the probability measure in $D_{\Omega_n}[0,T]$ induced by the Rudvalis process with initial measure μ_n. Denote by $\mathbb{E}_{\mu_n}$ the expectation with respect to $\mathbb{P}_{\mu_n}$.

2.2 Initial measures

We have an initial profile, that we assume to be measurable, denoted by $\gamma:\mathbb{T}\to[0,1]$, i.e., γ is a function which assigns mass to points on the torus. As an example, if the density of a gas was initially described by $\gamma(x)=\mathbb{1}_{[0,\frac{1}{2}]}(x)$, this would mean that the density of particles is concentrated "on the left".

Definition 2.2.1 (Sequence of measures associated with a profile)**.** *A sequence of probability measures $\{\mu_n\}_{n\in\mathbb{N}}$ on Ω_n is said to be associated with a measurable profile $\gamma:\mathbb{T}\to[0,1]$ if for every $\delta>0$ and*

every continuous function $G : \mathbb{T} \to \mathbb{R}$*,*

$$\lim_{n\to\infty} \mu_n \left(\eta \in \Omega_n : \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) \eta(x) - \int_{\mathbb{T}} G(u)\gamma(u)\, du \right| > \delta \right) = 0.$$

We might ask if such sequence of measures exists for the Rudvalis process defined above. In order to see that in fact it does, consider the following example. Throughout the rest of the text, for some space X and $G : X \to \mathbb{R}$, denote the $L^\infty(X)$-norm by $\|G\|_\infty = \sup_{x\in X} |G(x)|$.

Example 2.2.2. *Let* $\gamma : \mathbb{T} \to [0, 1]$ *be continuous. An example of a sequence of measures satisfying the assumption above is* $\{\nu^n_{\gamma(\cdot)}\}_{n\in\mathbb{N}}$*, where* $\nu^n_{\gamma(\cdot)}$ *is the Bernoulli product measure associated with* γ*, that is, it is the measure defined on the state space* Ω_n *such that:*

- *the random variables* $\{\eta(x)\}_{x\in\mathbb{T}_n}$ *are independent;*
- $\eta(x) \sim$ *Bernoulli* $(\gamma(\frac{x}{n}))$*.*

Thus, the Bernoulli product measures verifies

$$\nu^n_{\gamma(\cdot)}(\eta) = \prod_{x\in\mathbb{T}_n} [\eta(x)\gamma(\tfrac{x}{n}) + (1 - \eta(x))(1 - \gamma(\tfrac{x}{n}))]$$

which means that this measure assigns a particle independently to each site x *with probability* $\gamma(\frac{x}{n})$*. In order to see that this sequence of measures satisfies the condition of Definition 2.2.1, we do the following:*

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) \eta(x) - \int_{\mathbb{T}} G(u)\gamma(u)\, du \right| \\
&= \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) (\eta(x) - \gamma(\tfrac{x}{n})) - \left(\int_{\mathbb{T}} G(u)\gamma(u)\, du - \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) \gamma(\tfrac{x}{n}) \right) \right| \\
&\leq \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) (\eta(x) - \gamma(\tfrac{x}{n})) \right| + \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) \gamma(\tfrac{x}{n}) - \int_{\mathbb{T}} G(u)\gamma(u)\, du \right|.
\end{aligned}$$

It is then enough to prove that for any positive δ*,*

$$\lim_{n\to\infty} \nu^n_{\gamma(\cdot)} \left(\eta \in \Omega_n : \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) (\eta(x) - \gamma(\tfrac{x}{n})) \right| > \frac{\delta}{2} \right) = 0$$

and

$$\lim_{n\to\infty} \nu^n_{\gamma(\cdot)} \left(\eta \in \Omega_n : \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) \gamma(\tfrac{x}{n}) - \int_{\mathbb{T}} G(u)\gamma(u)\, du \right| > \frac{\delta}{2} \right) = 0.$$

For the first limit, since $\{\eta(x)\}_{x\in\mathbb{T}_n}$ *are independent with respect to* $\nu^n_{\gamma(\cdot)}$*,* $\{G(\frac{x}{n})(\eta(x) - \gamma(\frac{x}{n}))\}_{x\in\mathbb{T}_n}$ *are also independent random variables with respect to* $\nu^n_{\gamma(\cdot)}$ *but this time with mean zero. Thus,*

$$E_{\nu^n_{\gamma(\cdot)}} \left[\left(\frac{1}{n} \sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right) (\eta(x) - \gamma(\tfrac{x}{n})) \right)^2 \right] = \frac{1}{n^2} \sum_{x\in\mathbb{T}_n} E_{\nu^n_{\gamma(\cdot)}} \left[\left(G\left(\tfrac{x}{n}\right) (\eta(x) - \gamma(\tfrac{x}{n})) \right)^2 \right] \leq \frac{\|G\|^2_\infty}{n}.$$

From Markov's inequality,

$$\nu^n_{\gamma(\cdot)}\left(\left|\frac{1}{n}\sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right)\left(\eta(x)-\gamma(\tfrac{x}{n})\right)\right|>\frac{\delta}{2}\right)\leq\frac{4}{\delta^2}E_{\nu^n_{\gamma(\cdot)}}\left[\left(\frac{1}{n}\sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right)\left(\eta(x)-\gamma(\tfrac{x}{n})\right)\right)^2\right]\leq\frac{4}{\delta^2}\frac{\|G\|^2_\infty}{n}.$$

For the second limit, since G and γ are continuous, $f := G\gamma$ is continuous and whence there exists $N\in\mathbb{N}$ such that for $n\geq N$,

$$\left|\frac{1}{n}\sum_{x\in\mathbb{T}_n} G(\tfrac{x}{n})\gamma(\tfrac{x}{n})-\int_{\mathbb{T}} G(u)\gamma(u)\,du\right|\leq\frac{\delta}{2} \tag{2.1}$$

and the limit follows.

Example 2.2.3. *As an example, for $\gamma(x)=\mathbb{1}_{[0,\frac{1}{2}]}(x)$ (γ is not continuous, but the previous example can be adapted so that* (2.1) *holds), the associated initial measure would set particles with probability 1 on the sites $0,1,...,\frac{n}{2}$ and with probability 0 on the remaining sites.*

Since the Bernoulli product measures are quite simple, we can take them as initial distributions when one wants to obtain scaling limits. In terms of the generator, a measure ν is invariant if

$$E_\nu[\mathcal{L}f]=\int_\Omega \mathcal{L}f(\eta)\,\nu(d\eta)=0.$$

Proposition 2.2.4. *Let $\{\eta_t\}_{t\geq 0}$ denote the Rudvalis process. Then, for $\alpha\in[0,1]$, the Bernoulli product measure with parameter α, ν^n_α, is an invariant measure.*

Proof. Let $\nu^n_\alpha:\Omega_n\to[0,1]$ be the Bernoulli measure with parameter α, with $\alpha\in[0,1]$. We need to show that for any $f:\Omega_n\to\mathbb{R}$,

$$\int_{\Omega_n}\mathcal{L}_n f(\eta)\nu^n_\alpha(d\eta)=0.$$

It is then enough to show that

$$\int_{\Omega_n} f(\eta^n)\nu^n_\alpha(d\eta)=\int_{\Omega_n} f(\eta^{n-1})\nu^n_\alpha(d\eta)=\int_{\Omega_n} f(\eta)\nu^n_\alpha(d\eta).$$

Since ν^n_α does not depend of the position of the particles (it only depends on the number of particles), if ξ and ξ' have the same number of particles, then $\nu^n_\alpha(\xi)=\nu^n_\alpha(\xi')=\alpha^l(1-\alpha)^{n-l}$, where l is the number of particles in ξ. Thus, by a change of variables,

$$\int_{\Omega_n} f(\eta^n)\nu^n_\alpha(d\eta)=\int_{\Omega_n} f(\xi)\nu^n_\alpha(d\xi)=\int_{\Omega_n} f(\eta^{n-1})\nu^n_\alpha(d\eta).$$

□

2.3 Empirical measure

We have already seen how to obtain a particle system from the Rudvalis process. Now it is time to study the evolution of the Rudvalis process $\{\eta_t\}_{t\geq 0}$ with infinitesimal generator $n\mathcal{L}_n$ and state space Ω_n.

Let $\mathcal{M}$ be the set of positive measures on $\mathbb{T}$ with mass bounded by 1 endowed with the weak topology. If $\{\pi^n\}_{n\in\mathbb{N}}$, $\pi \in \mathcal{M}$, we say that π^n converges weakly to π, which we denote by

$$\pi^n \xrightarrow[n\uparrow\infty]{w} \pi,$$

if for all $G \in C(\mathbb{T})$

$$\int_{\mathbb{T}} G\, \pi^n(du) \xrightarrow[n\uparrow\infty]{} \int_{\mathbb{T}} G\, \pi(du).$$

In order to obtain a scaling limit of the Rudvalis process, we introduce a measure which gives weight $\frac{1}{n}$ to each occupied site of the configuration η.

Definition 2.3.1 (Empirical Measure)**.** *For each configuration $\eta \in \Omega_n$, we define the empirical measure $\pi^n(\eta, du)$ in $[0,1]$ by*

$$\pi^n(\eta, du) = \frac{1}{n}\sum_{x\in\mathbb{T}_n} \eta(x)\delta_{\frac{x}{n}}(du),$$

where δ_y is the Dirac measure concentrated on $y \in \mathbb{T}$.

Since we are interested in studying the behaviour of this measure with respect to the Rudvalis process $\{\eta_t\}_{t\geq 0}$, define

$$\pi^n_t(du) := \pi^n(\eta_t, du) = \frac{1}{n}\sum_{x\in\mathbb{T}_n} \eta_t(x)\delta_{\frac{x}{n}}(du) \in \mathcal{M}.$$

Given a function $G : \mathbb{T} \to \mathbb{R}$, we denote the integral of G with respect to a measure μ (on $\mathbb{T}$) by

$$\langle \mu, G\rangle := \int G(u)\, \mu(du).$$

In particular, for the empirical measure π^n_t, we have

$$\langle \pi^n_t, G\rangle := \int G(u)\, \pi^n_t(du) = \frac{1}{n}\sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right)\eta_t(x).$$

Remark 2.3.2. *This definition should not be confused with the inner product in $L^2(\mathbb{T})$, i.e., for $\rho, G \in L^2(\mathbb{T})$,*

$$\langle \rho, G\rangle := \int \rho(u)G(u)\, du.$$

For this reason, if a measure $\pi_t(du)$ has a density $\rho_t(u)$, i.e., $\pi_t(du) = \rho(t,u)\, du$, then

$$\langle \pi_t, G\rangle = \langle \rho_t, G\rangle = \int G(u)\rho(t,u)\, du.$$

Using this notation, the convergence in $\mathcal{M}$ can be restated as follows: $\pi^n \xrightarrow[n\uparrow\infty]{w} \pi$ if, and only if, for all

$G \in C(\mathbb{T})$,

$$|\langle \pi^n, G\rangle - \langle \pi, G\rangle| \xrightarrow[n\uparrow\infty]{} 0.$$

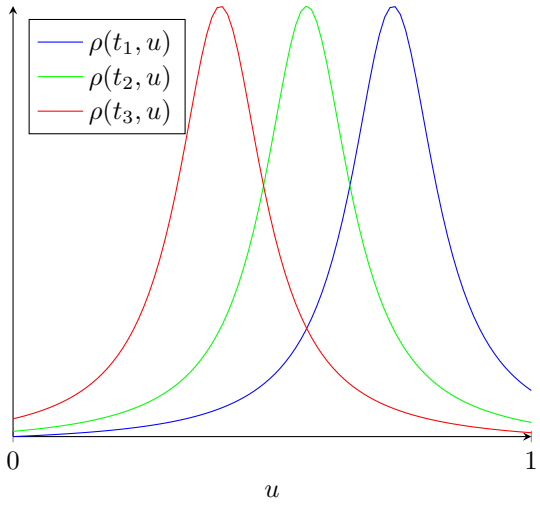
2.4 Transport equation

From now on, fix $T > 0$. When we think about the Rudvalis shuffle, the behaviour of the cards is quite similar to a wave: we take the top card and put it in the bottom (or almost at the bottom), pushing all the cards one position up. We shuffle the deck over and over again, that is, the cards are being pushed up continuously (Figure 1.3). With that picture in mind, we analyze the following PDE: the *transport equation* (sometimes called the one-way wave equation), given formally by

$$\begin{cases} \partial_t \rho(t,u) = \nabla \rho(t,u), & \text{for } t \in [0,T], u \in \mathbb{T}, \\ \rho(0,u) = \gamma(u), & \text{for } u \in \mathbb{T}, \end{cases} \tag{2.2}$$

where ∇ is the space derivative and $\gamma : \mathbb{T} \to [0,1]$ is a measurable function.

Observe that the solutions of this equation are functions that behave like "waves" moving to the left, i.e., if γ is C^1 in space, then the function $\rho(t,u) = \gamma(u+t)$ is a (strong) solution of (2.2). In particular, if $\gamma(u) = \alpha \in [0,1]$, then $\rho(t,u) = \alpha$ for any (t,u). For example, we would have the following graph for a solution of (2.2) captured at three distinct times $0 < t_1 < t_2 < t_3 < T$.



2.5 Motivation for the weak solution

When we look at equation (2.2), we see that a necessary condition for a function to be its solution is that it must be differentiable in time and space. If γ is not differentiable, a function satisfying (2.2) (*strong solution*) might not be easy to find and we will instead try to find a weaker solution, i.e., a function for which the derivatives may not all exist but which satisfies the equation in some sense that we are going to define below.

Define $C^{2,1}([0,T]\times\mathbb{T})$ as the space of functions $f:[0,T]\times\mathbb{T}\to\mathbb{R}$ with first derivative in time and second derivative in space, both being continuous. Let $f\in C^{2,1}([0,T]\times\mathbb{T})$. We multiply both sides of the equation $\partial_s\rho(s,u)=\partial_u\rho(s,u)$ by $f(s,u)$ and integrate it on time $[0,t]$, for $t<T$, and on space $\mathbb{T}$. Then, applying Fubini's theorem, we obtain

$$\int_{\mathbb{T}}\int_0^t \partial_s\rho(s,u)f(s,u)\,ds\,du=\int_0^t\int_{\mathbb{T}}\partial_u\rho(s,u)f(s,u)\,du\,ds.$$

Applying integration by parts on the left hand side,

$$\int_{\mathbb{T}}\int_0^t \partial_s\rho(s,u)f(s,u)\,ds\,du=\int_{\mathbb{T}}\rho(t,u)f(t,u)-\rho(0,u)f(0,u)\,du-\int_{\mathbb{T}}\int_0^t\rho(s,u)\partial_s f(s,u)\,ds\,du.$$

Integrating by parts again, but this time on the right hand side and using the fact that f is periodic in $\mathbb{T}$, we obtain

$$\int_0^t\int_{\mathbb{T}}\partial_u\rho(s,u)f(s,u)\,du\,ds=-\int_0^t\int_{\mathbb{T}}\rho(s,u)\partial_u f(s,u)\,du\,ds.$$

Using the initial condition $\rho(0,u)=\gamma(u)$ and combining both equations, we get

$$\int_{\mathbb{T}}\rho(t,u)f(t,u)-\gamma(u)f(0,u)\,du-\int_0^t\int_{\mathbb{T}}\rho(s,u)\partial_s f(s,u)\,du\,ds+\int_0^t\int_{\mathbb{T}}\rho(s,u)\partial_u f(s,u)\,du\,ds=0.$$

Having come to this, we can define the concept of *weak solution* for the transport equation (2.2):

Definition 2.5.1 (Weak solution of the transport equation on $\mathbb{T}$)**.** *Let $\gamma:\mathbb{T}\to[0,1]$ be a measurable function and $T>0$. A measurable function $\rho:[0,T]\times\mathbb{T}\to[0,1]$ is said to be a weak solution of equation* (2.2) *if for all $f\in C^{2,1}([0,T]\times\mathbb{T})$ and all $t\in[0,T]$, ρ satisfies*

$$\int_{\mathbb{T}}\rho(t,u)f(t,u)-\gamma(u)f(0,u)\,du-\int_0^t\int_{\mathbb{T}}\rho(s,u)\partial_s f(s,u)\,du\,ds+\int_0^t\int_{\mathbb{T}}\rho(s,u)\partial_u f(s,u)\,du\,ds=0.$$

2.6 Heuristic argument for weak solution

As stated above, our goal is to prove that the density of particles is ruled by the weak solution of a PDE, in our case, the transport equation. In order to do that, we work with the microscopic space $\mathbb{T}_n$ and then take the limit on the number of sites, n. For that reason, it is useful to define the notion of derivative in the discrete case.

Definition 2.6.1 (Discrete left derivative)**.** *Given $G:\mathbb{T}\to\mathbb{R}$, $n\in\mathbb{N}$ and $x\in\mathbb{T}_n$, the discrete left derivative at $\frac{x}{n}$ is defined by*

$$\nabla_n^- G\left(\tfrac{x}{n}\right)=n\left(G\left(\tfrac{x}{n}\right)-G\left(\tfrac{x-1}{n}\right)\right).$$

Remark 2.6.2. *Note that, if G is C^1,*

$$\lim_{n\to\infty}|\nabla_n^- G(\tfrac{x}{n})-\nabla G(\tfrac{x}{n})|=0.$$

Computing the generator on $\langle \pi_t^n, G\rangle$ is one of the first steps when one wants to find how the empirical measure evolves in space/time. The following lemma shows that computation.

Lemma 2.6.3. *Let $\{\eta_t\}_{t\geq 0}$ be the Markov process with generator $n\mathcal{L}_n$ (Rudvalis process). Then, for any $t \geq 0$ and $G : \mathbb{T} \to \mathbb{R}$*

$$n\mathcal{L}_n\langle \pi_t^n, G\rangle = -\langle \pi_t^n, \nabla_n^- G\rangle - \frac{1}{2n}(\eta_t(1) - \eta_t(0))\nabla_n^- G(0).$$

Proof. We start by computing $\mathcal{L}_n\eta_t(x)$, for $x \in \mathbb{T}_n$.

$$\mathcal{L}_n\eta_t(x) = \frac{1}{2}\sum_{y=n-1}^{n}\Theta_y\eta_t(x) = \frac{1}{2}(\eta_t^n(x)+\eta_t^{n-1}(x)-2\eta_t(x)) = \frac{1}{2}\begin{cases} 2\eta_t(x+1) - 2\eta_t(x), \text{ if } x \notin \{n-1, n\} \\ \eta_t(n) + \eta_t(1) - 2\eta_t(n-1), \text{ if } x = n-1 \\ \eta_t(1) - \eta_t(n), \text{ if } x = n \end{cases}.$$

Thus,

$$\begin{aligned} n\mathcal{L}_n\langle \pi_t^n, G\rangle &= \sum_{x=1}^{n-2} G(\tfrac{x}{n})[\eta_t(x+1) - \eta_t(x)] + \frac{1}{2}G(\tfrac{n-1}{n})[\eta_t(n) + \eta_t(1) - 2\eta_t(n-1)] + \frac{1}{2}G(\tfrac{n}{n})[\eta_t(1) - \eta_t(n)] \\ &= -\sum_{x\in\mathbb{T}_n}\eta_t(x)[G(\tfrac{x}{n}) - G(\tfrac{x-1}{n})] - \frac{1}{2}\eta_t(1)\left[G\left(\tfrac{n}{n}\right) - G\left(\tfrac{n-1}{n}\right)\right] + \frac{1}{2}\eta_t(n)\left[G\left(\tfrac{n}{n}\right) - G\left(\tfrac{n-1}{n}\right)\right] \\ &= -\frac{1}{n}\sum_{x\in\mathbb{T}_n}\eta_t(x)\nabla_n^- G(\tfrac{x}{n}) - \frac{1}{2n}(\eta_t(1) - \eta_t(0))\nabla_n^- G(0) \\ &= -\langle \pi_t^n, \nabla_n^- G\rangle - \frac{1}{2n}(\eta_t(1) - \eta_t(0))\nabla_n^- G(0). \end{aligned}$$

□

Remark 2.6.4. *If we had considered the general Rudvalis process with generator given by $\mathcal{L}_n f(\eta) = p\,\Theta_{n-1}f(\eta) + (1-p)\,\Theta_n f(\eta)$, similar computations would lead to*

$$n\mathcal{L}_n\langle \pi_t^n, G\rangle = -\langle \pi_t^n, \nabla_n^- G\rangle - \frac{p}{n}(\eta_t(1) - \eta_t(0))\nabla_n^- G(0).$$

As we will see next, what matters is that the rightmost term of the above expression is $O(\frac{1}{n})$ and thus taking a general $p \in (0,1)$ would lead to the same result.

Remark 2.6.5. *Since there is at most one particle per site, $|\eta_t(1) - \eta_t(0)| \leq 1$. If we take $G \in C(\mathbb{T})$, then, since G is continuous on the compact set $\mathbb{T}$,*

$$|n\mathcal{L}_n\langle \pi_t^n, G\rangle + \langle \pi_t^n, \nabla_n^- G\rangle| = O(\tfrac{1}{n}).$$

Now fix $f \in C^{2,1}([0,T] \times \mathbb{T})$ and apply Dynkin's formula (A.1.2) with $F(t, \eta_t) = \langle \pi_t^n, f_t \rangle$. Then

$$M_t^n(f) := \langle \pi_t^n, f_t \rangle - \langle \pi_0^n, f_0 \rangle - \int_0^t (\partial_s + n\mathcal{L}_n)\langle \pi_s^n, f_s \rangle \, ds,$$
$$N_t^n(f) := (M_t^n(f))^2 - \int_0^t \Gamma^n(\langle \pi_s^n, f_s \rangle) \, ds, \text{ where } \Gamma^n(\langle \pi_s^n, f_s \rangle) := n\mathcal{L}_n\langle \pi_s^n, f_s \rangle^2 - 2\langle \pi_s^n, f_s \rangle n\mathcal{L}_n\langle \pi_s^n, f_s \rangle, \tag{2.3}$$

are martingales with respect to the natural filtration $\mathcal{F}_t = \sigma(\eta_s : s \leq t)$.

Using Lemma 2.6.3, we can rewrite the first martingale as

$$M_t^n(f) = \langle \pi_t^n, f_t \rangle - \langle \pi_0^n, f_0 \rangle - \int_0^t \langle \pi_s^n, \partial_s f_s \rangle \, ds + \int_0^t \langle \pi_s^n, \nabla_n^- f_s \rangle \, ds + \frac{1}{2n} \int_0^t (\eta_s(1) - \eta_s(0)) \nabla_n^- f_s(0) \, ds. \tag{2.4}$$

Since the expected value of a martingale remains constant and $M_0^n(f) = 0$, we have for all t that

$$\mathbb{E}_{\mu_n} [M_t^n(f)] = 0.$$

Consider the function $\rho_t^n(x) := \mathbb{E}_{\mu_n}[\eta_t(x)]$. The martingale gives us an idea (in the discrete case) of the condition satisfied by the solution of the PDE which we are expecting. Taking the expectation with respect to $\mathbb{P}_{\mu_n}$ in (2.4), we get the following equation for ρ_t^n:

$$\begin{aligned} 0 =& \frac{1}{n} \sum_{x \in \mathbb{T}_n} f_t\left(\tfrac{x}{n}\right) \rho_t^n(x) - \frac{1}{n} \sum_{x \in \mathbb{T}_n} f_0\left(\tfrac{x}{n}\right) \rho_0^n(x) \\ & - \int_0^t \frac{1}{n} \sum_{x \in \mathbb{T}_n} \partial_s f_s\left(\tfrac{x}{n}\right) \rho_s^n(x) \, ds + \int_0^t \frac{1}{n} \sum_{x \in \mathbb{T}_n} \nabla_n^- f_s\left(\tfrac{x}{n}\right) \rho_s^n(x) \, ds + O(\tfrac{1}{n}). \end{aligned}$$

If we look carefully and assume that $|\rho_t^n(x) - \rho_t(\frac{x}{n})| \xrightarrow[n \uparrow \infty]{} 0$, these are Riemann sums for the integrals in Definition 2.5.1.

2.7 Hydrodynamic limit

Now that we have seen an heuristic argument to obtain a weak solution for equation (2.2), let us see how to prove rigorously the hydrodynamic limit for the dynamics of the Rudvalis process. Recall Definition 2.1.4 and the space $D_{\Omega_n}[0,T]$, where $\mathbb{P}_{\mu_n}$ is defined. The process $\{\pi_t^n\}_{t \geq 0}$, induced by the Rudvalis process, can be seen as a Markov process, but this time in $D_{\mathcal{M}}[0,T]$ endowed with the Skorohod topology.

Let $\{\mathbb{Q}^n\}_{n \in \mathbb{N}}$ be the sequence of probability measures in $D_{\mathcal{M}}[0,T]$ induced by the process $\{\pi_t^n\}_{t \geq 0}$ and by the probability measure $\mathbb{P}_{\mu_n}$. Equivalently, $\mathbb{Q}^n$ is induced by the application

$$\begin{aligned} \pi^n : (D_{\Omega_n}[0,T], \mathbb{P}_{\mu_n}) &\to (D_{\mathcal{M}}[0,T], \mathbb{Q}^n) \\ \eta_\cdot &\to \pi^n(\eta_\cdot, du). \end{aligned}$$

In particular, given a measurable set $A \in D_{\mathcal{M}}[0,T]$,

$$\mathbb{Q}^n(A) = \mathbb{P}_{\mu_n}((\pi^n)^{-1}(A)) = \mathbb{P}_{\mu_n}(\eta_\cdot \in D_{\Omega_n}[0,T] : \pi^n(\eta_\cdot) \in A) = \mathbb{P}_{\mu_n}(\eta_\cdot \in D_{\Omega_n}[0,T] : \pi^n_\cdot \in A) \tag{2.5}$$

As we will see later, the sequence $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ of probability measures converges weakly to a limit point. The concept of *relative compactness* will help us to assert this fact. In what follows, $(X, \mathcal{B})$ denotes a metric space X with σ-algebra $\mathcal{B}$ generated by the open sets, the Borel σ-algebra.

Definition 2.7.1 (Relatively Compact)**.** *Let Π be a family of probability measures on $(X, \mathcal{B})$. We say that Π is relatively compact if every sequence of elements of Π contains a weakly convergent subsequence, that is, for every sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ in Π there exists a subsequence $\{\mathbb{P}_{n_i}\}_{i\in\mathbb{N}}$ and a probability measure $\mathbb{P}$ (defined on $(X, \mathcal{B})$ but not necessarily an element of Π) such that $\mathbb{P}_{n_i} \xrightarrow[i\uparrow\infty]{w} \mathbb{P}$.*

Recall the concept of a sequence of measures associated with a profile, given in Definition 2.2.1. We now present the main result of this chapter.

Theorem 2.7.2 (Hydrodynamic Limit)**.** *Consider the Rudvalis process $\{\eta_t\}_{t\geq 0}$. Fix a measurable initial profile $\gamma : \mathbb{T} \to [0,1]$ and let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures associated with γ. Then, for any $t \in [0,T]$, for any $\delta > 0$ and any $G \in C(\mathbb{T})$, it holds*

$$\lim_{n\to\infty} \mathbb{P}_{\mu_n}\left(\eta_\cdot \in D_{\Omega_n}[0,T] : \left|\frac{1}{n}\sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right)\eta_t(x) - \int_{\mathbb{T}} G(u)\rho(t,u)\,du\right| > \delta\right) = 0 \tag{2.6}$$

where $\rho(t,\cdot)$ is the unique weak solution of (2.2) *with initial condition $\rho(0,\cdot) := \gamma(\cdot)$.*

Remark 2.7.3. *We call equation* (2.2) *the hydrodynamic equation of the Rudvalis process.*

But what is the hydrodynamic limit really saying? We start by noting that we can rewrite what is inside of $\mathbb{P}_{\mu_n}$ in (2.6) as

$$\left|\frac{1}{n}\sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right)\eta_t(x) - \int_{\mathbb{T}} G(u)\rho(t,u)\,du\right| = \left|\langle \pi^n_t, G\rangle - \langle \pi_t, G\rangle\right|,$$

where $\pi_t(du) = \rho(t,u)\,du$, i.e., $\pi_t(du)$ is an absolutely continuous measure (with respect to the Lebesgue measure) whose density $\rho_t(u)$ is the unique weak solution of the transport equation.

If all the measures involved were deterministic, we would just say that π^n_t converges weakly to π_t. However, this is not the case, since $\pi^n_t := \pi^n(\eta_t)$ is a random measure (which depends on the Rudvalis process) we have to express in which sense this convergence holds. With respect to $\mathbb{P}_{\mu_n}$, the measure induced by the Rudvalis process with infinitesimal generator $n\mathcal{L}_n$ and initial measure μ_n, verifies $\pi^n_t \xrightarrow{w} \pi_t$ so another way to express the limit in (2.6) is to say that π^n_t converges in probability to π_t (because the convergence is only verified under $\mathbb{P}_{\mu_n}$). Another way to see this condition, is that the Law of Large Numbers holds at time $t \in [0,T]$ for π^n_t.

Similarly, the definition of probability measures associated with γ can be rewritten as the sequence

of measures for which given any integrable function $G : \mathbb{T} \to [0,1]$ and $\delta > 0$, verifies

$$\lim_{n\to\infty} \mu_n \left(\eta \in \Omega_n : |\langle \pi^n(\eta), G\rangle - \langle \gamma, G\rangle| > \delta\right) = 0.$$

In conclusion, what we are really doing when proving the hydrodynamic limit is showing that if the Law of Large Numbers for the empirical measure holds at the initial time $t = 0$, then it also holds for any subsequent time t. The hydrodynamic limit is a consequence of the following proposition.

Proposition 2.7.4. *Let $\gamma : \mathbb{T} \to [0,1]$ be measurable and consider a sequence of measures $\{\mu_n\}_{n\in\mathbb{N}}$ on Ω_n associated with γ. Let $\mathbb{Q}^*$ be the probability measure concentrated on a trajectory $\pi_\cdot$ of $D_{\mathcal{M}}[0,T]$ consisting of absolutely continuous measures with respect to the Lebesgue measure, i.e., $\pi_t(du) = \rho_t(u)du$ where the density $\rho_\cdot$ is the unique weak solution of* (2.2)*. Then*

$$\mathbb{Q}^n \xrightarrow{w} \mathbb{Q}^*.$$

Since $\mathbb{Q}^n$ is the measure induced by $\mathbb{P}_{\mu_n}$ and the application π^n, the weak convergence of $\mathbb{Q}^n$ to $\mathbb{Q}^*$ is the same as the convergence in distribution of π_t^n to π_t, for any $t \in [0,T]$. Since $\pi_t(du) = \rho_t(u)du$ is a deterministic measure, the previous convergence is also valid in probability (with respect to $\mathbb{P}_{\mu_n}$) and whence, by the previous observations, Proposition 2.7.4 implies Theorem 2.7.2. For this reason, when we want to prove the hydrodynamic limit, we often follow the sequence of steps given below:

1. The sequence $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is relatively compact with respect to the Skorohod's topology (A.3.4).
2. The limit points of the subsequences of $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ are concentrated on trajectories of measures which are absolutely continuous with respect to the Lebesgue measure and whose densities $\rho(t,u)$ are weak solutions of the hydrodynamic equation.
3. The hydrodynamic equation has a unique weak solution.

Note that proving these 3 steps is exactly what we need to establish Proposition 2.7.4: if $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is relatively compact, then there exists a subsequence of $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ which converges weakly to a limit point $\mathbb{Q}^*$. By 2., the limit points of every convergent subsequence are concentrated on a trajectory $\pi_\cdot$ such that for $t \in [0,T]$, π_t is absolutely continuous with respect to the Lebesgue measure, i.e., $\pi_t(du) = \rho_t(u)\, du$. Moreover, the density $\rho(t,u)$ is a weak solution of the hydrodynamic equation. In 3., we are going to see that the hydrodynamic equation has a unique weak solution $\rho(t,u)$, and this all together implies the existence and uniqueness of the limit point $\mathbb{Q}^*$. Finally, we conclude (since the limit point is unique) that the whole sequence $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ converges weakly to $\mathbb{Q}^*$, which is a measure concentrated on a trajectory $\pi_\cdot$ of absolutely continuous measures with respect to the Lebesgue measure and whose density is the unique weak solution of the hydrodynamic equation.

2.7.1 Relative compactness

The concepts of tightness and relatively compactness are related in a well know theorem by Prohorov (Theorem A.2.1). In that regard, we define the concept of *tightness*:

Definition 2.7.5 (Tightness)**.** *Let Π be a family of probability measures on $(X, \mathcal{B})$. We say that Π is tight if for every $\varepsilon > 0$, there exists a compact set K such that $\mathbb{P}(K) > 1 - \varepsilon$ for every probability measure $\mathbb{P}$ in Π.*

Put into words, a family of probability measures is tight (and by Prohorov's Theorem, is relatively compact as well) if there exists a compact set that accumulates mass for every measure of that family. The notion of a compact, which is summarized in A.3, is not so clear when we are working with the Skorohod topology and we refer the reader to [3] (Chapter 3).

When proving the tightness of $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$, we will use Dynkin's formula as in Section 2.6. Moreover, the following lemma will be useful. But before, let us just introduce the next notation (similar to the big O). For two sequences $\{f_k\}_{k\in\mathbb{N}}$ and $\{g_k\}_{k\in\mathbb{N}}$, we say that $f_n \lesssim g_n$ if there exists a positive constant C, independent of n, such that $f_n \leq C\, g_n$.

Lemma 2.7.6. *Recall the operator Γ^n defined in* (2.3)*. Let $G \in C^1(\mathbb{T})$. Then,*

$$\Gamma^n(\langle \pi_s^n, G\rangle) \lesssim \frac{\|\nabla G\|_\infty^2}{n}.$$

Proof. Recall the operator Θ_x defined in Section 2.1. Note that for $\eta_s \in \Omega_n$, if $h(\eta_s) = \langle \pi_s^n, G\rangle$ then $\Theta_x(h(\eta_s)^2) - 2h(\eta_s)\Theta_x h(\eta_s) = (\Theta_x h(\eta_s))^2$. Using this fact and the mean value theorem, we get

$$\begin{aligned}
\Gamma^n(\langle \pi_s^n, G\rangle) &:= n\mathcal{L}_n\langle \pi_s^n, G\rangle^2 - 2\langle \pi_s^n, G\rangle n\mathcal{L}_n\langle \pi_s^n, G\rangle \\
&= n\sum_{x=n-1}^{n} \tfrac{1}{2}\Theta_x(\langle \pi_s^n, G\rangle^2) - 2n\langle \pi_s^n, G\rangle \sum_{x=n-1}^{n} \tfrac{1}{2}\Theta_x\langle \pi_s^n, G\rangle \\
&= \tfrac{n}{2}\sum_{x=n-1}^{n} \left(\Theta_x\langle \pi_s^n, G\rangle\right)^2 = \tfrac{n}{2}\sum_{x=n-1}^{n} \left(\tfrac{1}{n}\sum_{y\in\mathbb{T}_n} G\left(\tfrac{y}{n}\right)\Theta_x\eta_s(y)\right)^2 \\
&= \tfrac{n}{2}\sum_{x=n-1}^{n} \left(\tfrac{1}{n}\sum_{y\in\mathbb{T}_n} G\left(\tfrac{y}{n}\right)\left(\eta_s^x(y) - \eta_s(y)\right)\right)^2 = \tfrac{1}{2n}\left(\sum_{y\in\mathbb{T}_n} G\left(\tfrac{y}{n}\right)\left(\eta_s(y+1) - \eta_s(y)\right)\right)^2 \\
&\quad + \tfrac{1}{2n}\left(\sum_{y\notin\{n-1,0\}} G\left(\tfrac{y}{n}\right)\left(\eta_s(y+1) - \eta_s(y)\right) + G\left(\tfrac{n-1}{n}\right)\left(\eta_s(1) - \eta_s(n-1)\right)\right)^2 \\
&= \tfrac{1}{2n}\left(-\sum_{y\in\mathbb{T}_n} \eta_s(y)\left[G\left(\tfrac{y}{n}\right) - G\left(\tfrac{y-1}{n}\right)\right]\right)^2 \\
&\quad + \tfrac{1}{2n}\left(-\sum_{y\in\mathbb{T}_n} \eta_s(y)\left[G\left(\tfrac{y}{n}\right) - G\left(\tfrac{y-1}{n}\right)\right] - \left(\eta_s(1) - \eta_s(0)\right)\left[G(0) - G\left(\tfrac{n-1}{n}\right)\right]\right)^2 \\
&\leq \tfrac{1}{2n}\left(\sum_{y\in\mathbb{T}_n} \left|G\left(\tfrac{y}{n}\right) - G\left(\tfrac{y-1}{n}\right)\right|\right)^2 + \tfrac{1}{2n}\left(\sum_{y\in\mathbb{T}_n} \left|G\left(\tfrac{y}{n}\right) - G\left(\tfrac{y-1}{n}\right)\right| + \left|G(0) - G\left(\tfrac{n-1}{n}\right)\right|\right)^2 \\
&\leq \tfrac{1}{n}\left(\sum_{y\in\mathbb{T}_n} \left|G\left(\tfrac{y}{n}\right) - G\left(\tfrac{y-1}{n}\right)\right| + \left|G(0) - G\left(\tfrac{n-1}{n}\right)\right|\right)^2 \leq \tfrac{1}{n}\left(\|\nabla G\|_\infty + \tfrac{1}{n}\|\nabla G\|_\infty\right)^2 \\
&\lesssim \frac{\|\nabla G\|_\infty^2}{n}.
\end{aligned}$$

□

Proposition 2.7.7. *The sequence $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is relatively compact.*

Proof. Recall that $\mathbb{Q}^n$ is the probability measure defined in $D_{\mathcal{M}}[0,T]$ which is induced by the process $\{\pi_t^n\}_{t\geq 0}$ and $\mathbb{P}_{\mu_n}$ and is defined in $D_{\mathcal{M}}[0,T]$. We will show that the sequence $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is tight and use that fact together with Prohorov's Theorem to conclude that the sequence is relatively compact.

We start by noting that $C^1(\mathbb{T})$ is dense in $C(\mathbb{T})$ under the uniform topology. Let $G \in C^1(\mathbb{T})$. By Lemma A.2.2, it is enough to show that $\{\mathbb{Q}^{n,G}\}_{n\in\mathbb{N}}$ is tight, where $\mathbb{Q}^{n,G}$ is the probability measure induced by

$$\begin{aligned}\mathcal{G} : (D_{\mathcal{M}}[0,T], \mathbb{Q}^n) &\to (D_{\mathbb{R}}[0,T], \mathbb{Q}^{n,G})\\ \{\pi_t^n\}_{t\geq 0} &\to \{\langle \pi_t^n, G\rangle\}_{t\geq 0}.\end{aligned}$$

So if $A \in D_{\mathbb{R}}[0,T]$ is measurable, then, by (2.5)

$$\begin{aligned}\mathbb{Q}^{n,G}(A) &= \mathbb{Q}^n(\mathcal{G}^{-1}(A)) = \mathbb{Q}^n(\pi_\cdot^n \in D_{\mathcal{M}}[0,T] : \mathcal{G}(\pi_\cdot^n) \in A) = \mathbb{Q}^n(\pi_\cdot^n \in D_{\mathcal{M}}[0,T] : \langle \pi_\cdot^n, G\rangle \in A)\\ &= \mathbb{P}_{\mu_n}(\eta_\cdot \in D_{\Omega_n}[0,T] : \langle \pi^n(\eta_\cdot), G\rangle \in A).\end{aligned}$$

We thus have to show the tightness of the family $\{\langle \pi_t^n, G\rangle : 0 \leq t \leq T\}_{n\in\mathbb{N}}$ for $G \in C(\mathbb{T})$, which will be done by applying Aldous' criterion (Proposition A.2.4):

- Condition (i) of Aldous' criterion.
  Let $t \in [0,T]$ and $\varepsilon > 0$. Note that

  $$|\langle \pi_t^n, G\rangle| = \left|\frac{1}{n}\sum_{x\in\mathbb{T}_n} G\left(\tfrac{x}{n}\right)\eta_t(x)\right| \leq \|G\|_\infty < \infty$$

  where the last inequality comes from the fact that $|\eta_t(x)| \leq 1$ and that G is a continuous function defined on a compact set and therefore it is bounded. Thus we can take $K := \overline{B_r(0)}$ with $r > \|G\|_\infty$ (closed ball with radius r centred in 0) as the compact we need for condition (i) of Proposition A.2.4 since, for this set,

  $$\mathbb{Q}^{n,G}(\langle \pi_\cdot^n, G\rangle \in D_{\mathbb{R}}[0,T] : \langle \pi_t^n, G\rangle \notin K) = 0 < \varepsilon$$

  and the first condition of Aldous' criterion is proved.

- Condition (ii) of Aldous' criterion.
  Let $\varepsilon > 0$ and $\mathcal{T}_T$ be the set of stopping times bounded by T. We have to show that

  $$\lim_{\gamma\to 0}\limsup_{n\to\infty}\sup_{\tau\in\mathcal{T}_T,\theta\leq\gamma}\mathbb{Q}^{n,G}(\langle \pi_\cdot^n, G\rangle \in D_{\mathbb{R}}[0,T] : |\langle \pi_{\tau+\theta}^n, G\rangle - \langle \pi_\tau^n, G\rangle| > \varepsilon) = 0. \tag{2.7}$$

  Using Dynkin's formula with $\langle \pi_t^n, G\rangle$ as in (2.3), we have that

  $$\begin{aligned}M_t^n(G) &:= \langle \pi_t^n, G\rangle - \langle \pi_0^n, G\rangle - \int_0^t (\partial_s + n\mathcal{L}_n)\langle \pi_s^n, G\rangle\, ds,\\ N_t^n(G) &:= (M_t^n(G))^2 - \int_0^t \Gamma^n(\langle \pi_s^n, G\rangle)\, ds,\end{aligned}$$

where

$$\Gamma^n(\langle \pi_s^n, G\rangle) := n\mathcal{L}_n\langle \pi_s^n, G\rangle^2 - 2\langle \pi_s^n, G\rangle n\mathcal{L}_n\langle \pi_s^n, G\rangle,$$

are martingales with respect to the natural filtration $\mathcal{F}_t = \sigma(\eta_s : s \leq t)$. We can bound the probability inside (2.7) as follows:

$$\begin{aligned}
&\mathbb{Q}^{n,G}(\langle \pi_\cdot^n, G\rangle \in D_{\mathbb{R}}[0,T] : |\langle \pi_{\tau+\theta}^n, G\rangle - \langle \pi_\tau^n, G\rangle| > \varepsilon) \\
&= \mathbb{Q}^n(\pi_\cdot^n \in D_{\mathcal{M}}[0,T] : |\langle \pi_{\tau+\theta}^n, G\rangle - \langle \pi_\tau^n, G\rangle| > \varepsilon) \\
&= \mathbb{P}_{\mu_n}\left(\eta_\cdot \in D_{\Omega_n}[0,T] : \left|M_{\tau+\theta}^n(G) - M_\tau^n(G) + \int_\tau^{\tau+\theta} n\mathcal{L}_n\langle \pi_s^n, G\rangle \, ds\right| > \varepsilon\right) \\
&\leq \mathbb{P}_{\mu_n}\left(\eta_\cdot \in D_{\Omega_n}[0,T] : |M_{\tau+\theta}^n(G) - M_\tau^n(G)| + \left|\int_\tau^{\tau+\theta} n\mathcal{L}_n\langle \pi_s^n, G\rangle \, ds\right| > \varepsilon\right) \\
&\leq \mathbb{P}_{\mu_n}\left(\eta_\cdot \in D_{\Omega_n}[0,T] : |M_{\tau+\theta}^n(G) - M_\tau^n(G)| > \frac{\varepsilon}{2}\right) \\
&\quad + \mathbb{P}_{\mu_n}\left(\eta_\cdot \in D_{\Omega_n}[0,T] : \left|\int_\tau^{\tau+\theta} n\mathcal{L}_n\langle \pi_s^n, G\rangle \, ds\right| > \frac{\varepsilon}{2}\right) \\
&\leq \frac{4}{\varepsilon^2}\mathbb{E}_{\mu_n}\left[|M_{\tau+\theta}^n(G) - M_\tau^n(G)|^2\right] + \frac{2}{\varepsilon}\mathbb{E}_{\mu_n}\left[\left|\int_\tau^{\tau+\theta} n\mathcal{L}_n\langle \pi_s^n, G\rangle \, ds\right|\right]
\end{aligned}$$

where the last inequality comes from applying Markov's inequality. Consequently, in order to achieve our goal, it is enough to prove the following facts:

$$\lim_{\gamma\to 0}\limsup_{n\to\infty}\sup_{\tau\in\mathcal{T}_T,\theta\leq\gamma}\mathbb{E}_{\mu_n}\left[|M_{\tau+\theta}^n(G) - M_\tau^n(G)|^2\right] = 0; \tag{2.8}$$

$$\lim_{\gamma\to 0}\limsup_{n\to\infty}\sup_{\tau\in\mathcal{T}_T,\theta\leq\gamma}\mathbb{E}_{\mu_n}\left[\left|\int_\tau^{\tau+\theta} n\mathcal{L}_n\langle \pi_s^n, G\rangle \, ds\right|\right] = 0. \tag{2.9}$$

a) *Proof of* (2.8)

We start by noting that for $t, s > 0$, the quadratic variation of the martingale verifies

$$\begin{aligned}
\mathbb{E}_{\mu_n}[(M_{t+s}^n(G) - M_t^n(G))^2] &= \mathbb{E}_{\mu_n}[M_{t+s}^n(G)^2] - 2\mathbb{E}_{\mu_n}[M_{t+s}^n(G)M_t^n(G)] + \mathbb{E}_{\mu_n}[M_t^n(G)^2] \\
&= \mathbb{E}_{\mu_n}[M_{t+s}^n(G)^2] - \mathbb{E}_{\mu_n}[M_t^n(G)^2] \\
&= \mathbb{E}_{\mu_n}\left[\int_t^{t+s}\Gamma^n(\langle \pi_s^n, G\rangle) \, ds\right] + \mathbb{E}_{\mu_n}[N_{t+s}^n(G)] + \mathbb{E}_{\mu_n}[N_t^n(G)] \\
&= \mathbb{E}_{\mu_n}\left[\int_t^{t+s}\Gamma^n(\langle \pi_s^n, G\rangle) \, ds\right],
\end{aligned}$$

where the second equality comes from martingale's properties, namely

$$\begin{aligned}
\mathbb{E}_{\mu_n}[M_{t+s}^n(G)M_t^n(G)] = \mathbb{E}_{\mu_n}[\mathbb{E}_{\mu_n}[M_{t+s}^n(G)M_t^n(G)|\mathcal{F}_t]] &= \mathbb{E}_{\mu_n}[M_t^n(G)\mathbb{E}_{\mu_n}[M_{t+s}^n(G)|\mathcal{F}_t]] \\
&= \mathbb{E}_{\mu_n}[M_t^n(G)^2].
\end{aligned}$$

By the previous observation, by the Optional Stopping Theorem and by Lemma 2.7.6,

$$\mathbb{E}_{\mu_n}\left[|M^n_{\tau+\theta}(G) - M^n_\tau(G)|^2\right] = \mathbb{E}_{\mu_n}\left[\int_\tau^{\tau+\theta} \Gamma^n(\langle \pi^n_s, G\rangle)\, ds\right] \lesssim \mathbb{E}_{\mu_n}\left[\int_\tau^{\tau+\theta} \frac{\|\nabla G\|^2_\infty}{n}\, ds\right]$$
$$= \frac{\|\nabla G\|^2_\infty \theta}{n},$$

and to conclude (2.8) it is enough to send n to infinity.

b) *Proof of* (2.9)

By the mean value theorem, we have that

$$|\langle \pi^n_s, \nabla^-_n G\rangle| \leq \sum_{x\in\mathbb{T}_n} |G(\tfrac{x}{n}) - G(\tfrac{x-1}{n})|\eta_s(x) \leq \frac{1}{n}\sum_{x\in\mathbb{T}_n} \|\nabla G\|_\infty = \|\nabla G\|_\infty$$

and

$$\frac{1}{2n}|(\eta_s(1) - \eta_s(0))\nabla^-_n G(0)| \leq \frac{1}{n}\|\nabla G\|_\infty .$$

By Lemma 2.6.3,

$$|n\mathcal{L}_n\langle \pi^n_s, G\rangle| \leq |\langle \pi^n_s, \nabla^-_n G\rangle| + \frac{1}{2n}|(\eta_s(1) - \eta_s(0))\nabla^-_n G(0)| \leq \left(1 + \frac{1}{n}\right)\|\nabla G\|_\infty .$$

We can use this relation to get an upper bound for the expectation in (2.9)

$$\mathbb{E}_{\mu_n}\left[\left|\int_\tau^{\tau+\theta} n\mathcal{L}_n\langle \pi^n_s, G\rangle\, ds\right|\right] \leq \mathbb{E}_{\mu_n}\left[\int_\tau^{\tau+\theta} |n\mathcal{L}_n\langle \pi^n_s, G\rangle|\, ds\right] \leq \int_\tau^{\tau+\theta}\left(1 + \frac{1}{n}\right)\|\nabla G\|_\infty\, ds$$
$$= \theta\left(1 + \frac{1}{n}\right)\|\nabla G\|_\infty ,$$

and to conclude (2.9) it is enough to take the limit as $n \to \infty$, and then take the limit as $\gamma \to 0$ (since $\theta \leq \gamma$).

This concludes the proof that $\{\mathbb{Q}^{n,G}\}_{n\in\mathbb{N}}$ is tight. Using Lemma A.2.2, $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is tight and finally by Prohorov's Theorem, we conclude that $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is relatively compact. □

The next step is to characterize all possible limit points, i.e., to find their properties.

2.7.2 Characterization of limit points

Having proven that $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ is relatively compact, it is time to characterize the limit points of the subsequences. We divide this proof into two parts. First, we show that any limit point $\mathbb{Q}^*$ is concentrated on trajectories (in $D_{\mathcal{M}}[0,T]$) which are absolutely continuous with respect to the Lebesgue measure. This means that the trajectories $\pi_\cdot$ satisfy $\pi_t(du) = \rho(t,u)du$ for some function ρ (the density). Then, we prove that the density ρ is a weak solution of (2.2). We start by recalling the concept of *absolutely continuous measure*.

Definition 2.7.8 (Absolutely continuous measure)**.** *A measure μ on Borel sets of $\mathbb{R}$ is absolutely continuous with respect to the Lebesgue measure λ, which we denote by $\mu \ll \lambda$ if, for every measurable set A, $\lambda(A) = 0$ implies $\mu(A) = 0$.*

The following lemma gives a sufficient condition for a measure to be absolutely continuous with respect to the Lebesgue measure. If we can prove that the sufficient condition of the lemma holds for trajectories in which $\mathbb{Q}^*$ is concentrated on, then we can conclude the first part.

Lemma 2.7.9. *Let μ be a measure which satisfies $|\langle \mu, G\rangle| \leq \int_0^1 |G(u)|\ \lambda(du)$ for all $G \in C([0,1])$. Then, $\mu \ll \lambda$, where λ is the Lebesgue measure.*

Proof. Let $F \subseteq [0,1]$ be closed and define for $n \in \mathbb{N}$, $F_n = \{x \in [0,1] : d(x,F) \leq \frac{1}{n}\}$, where d is the usual metric in $\mathbb{R}$. Let $\{g_n\}_{n\in\mathbb{N}}$ be a sequence of functions such that for each $n \in \mathbb{N}$, $0 \leq g_n \leq 1$ is a continuous function taking value 1 inside of F and 0 outside of F_n. In other words, g_n is a continuous function satisfying $\mathbb{1}_F \leq g_n \leq \mathbb{1}_{F_n}$. We get that for $n \in \mathbb{N}$

$$\mu(F) = \langle \mu, \mathbb{1}_F\rangle \leq \langle \mu, g_n\rangle \leq \int_0^1 g_n(u)\ \lambda(du) \leq \int_0^1 \mathbb{1}_{F_n}(u)\ \lambda(du) = \langle \lambda, \mathbb{1}_{F_n}\rangle = \lambda(F_n).$$

Since $\bigcap\limits_{n\in\mathbb{N}} F_n = F$ and $\lambda(F_1) < \infty$, by continuity from above,

$$\mu(F) \leq \lim_{n\to\infty} \lambda(F_n) = \lambda(F).$$

Now assume that $A \subseteq [0,1]$ is a measurable set such that $\lambda(A) = 0$. Hence, (by definition of set with measure zero) for $\varepsilon > 0$, there are closed intervals $I_1, I_2, ...$ such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} \lambda(I_n) < \varepsilon$. Thus

$$\mu(A) \leq \mu\Big(\bigcup_{n=1}^{\infty} I_n\Big) = \lim_{n\to\infty} \mu\Big(\bigcup_{k=1}^{n} I_k\Big) \leq \lim_{n\to\infty} \lambda\Big(\bigcup_{k=1}^{n} I_k\Big) \leq \lim_{n\to\infty} \sum_{k=1}^{n} \lambda(I_k) < \varepsilon.$$

Since ε is arbitrary, we conclude that $\mu(A) = 0$. □

Proposition 2.7.10. *Let $\mathbb{Q}^*$ be a limit point of a subsequence of the sequence $\{\mathbb{Q}^n\}_{n\geq 1}$. Then, $\mathbb{Q}^*$ is concentrated on trajectories of measures absolutely continuous with respect to the Lebesgue measure.*

Proof. Let $\{\mathbb{Q}^{n_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{\mathbb{Q}^n\}_{n\geq 1}$ converging to $\mathbb{Q}^*$. We will show that the condition of Lemma 2.7.9 holds for the trajectories in which $\mathbb{Q}^*$ is concentrated on. Let $G \in C(\mathbb{T})$ and denote by $\Theta(\pi^n_\cdot) := \sup\limits_{0\leq t\leq T} |\langle \pi^n_t, G\rangle|$ a real random variable which depends on the process $\{\eta_t\}_{t\geq 0}$ (recall that this process is indexed on n). Since we have at most one particle per site,

$$\Theta(\pi^n_\cdot) = \sup_{0\leq t\leq T} |\langle \pi^n_t, G\rangle| = \sup_{0\leq t\leq T} \left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} G(\tfrac{x}{n})\eta_t(x) \right| \leq \frac{1}{n} \sum_{x\in\mathbb{T}_n} |G(\tfrac{x}{n})|. \tag{2.10}$$

Now let $\varepsilon > 0$. Since G is continuous, there exists $N := N(\varepsilon) \in \mathbb{N}$ such that for any $n \geq N$,

$$\left| \frac{1}{n} \sum_{x\in\mathbb{T}_n} \left| G(\tfrac{x}{n}) \right| - \int_0^1 |G(u)|\ du \right| < \varepsilon. \tag{2.11}$$

From (2.10) and (2.11), for sufficiently large n, we get that $\mathbb{Q}^n(F_\varepsilon) = 1$, where

$$F_\varepsilon = \left\{ \pi_\cdot \in D_{\mathcal{M}}[0,T] : \Theta(\pi_\cdot) \leq \int_0^1 |G(u)|\, du + \varepsilon \right\}$$

is the set of trajectories of measures in which $\mathbb{Q}^n$ is concentrated on , for $n \in \mathbb{N}$. Now we proceed to show that F_ε is closed in the Skorohod topology $D_{\mathcal{M}}[0,T]$. Let $\{\pi_\cdot^n\} \in F_\varepsilon$ and $\pi_\cdot \in D_{\mathcal{M}}[0,T]$ such that $\pi_\cdot^n \to \pi_\cdot$ in Skorohod's topology. We aim to show that $\pi_\cdot \in F_\varepsilon$. By Lemma A.3.8, for almost every s in $[0,T]$ (including $s = 0$ and $s = T$), $\pi_s^n \xrightarrow[n\uparrow\infty]{w} \pi_s$. We claim that for every $t < T$, we have a sequence $\{t_k\}_{k\in\mathbb{N}}$ with $t_k \downarrow t$ such that $\pi_{t_k}^n \xrightarrow[n\uparrow\infty]{w} \pi_{t_k}$ for all $k \in \mathbb{N}$.

Let $S = \{s \in [0,T] : \pi_s^n \xrightarrow[n\to\infty]{w} \pi_s\}$ with $\lambda(S) = T$, where λ is the Lebesgue measure. We want to prove that for every $0 \leq t < T$ and every $\delta > 0$ such that $t + \delta < T$, we have $(t, t+\delta) \cap S \neq \emptyset$, which is always true because

$$\lambda((t,t+\delta) \cap S) = \lambda((t,t+\delta) \cap [0,T]) = \delta > 0.$$

By definition of weak convergence and since $\pi_\cdot^n \in F_\varepsilon$,

$$|\langle \pi_{t_k}, G\rangle| = \lim_{n\to\infty} |\langle \pi_{t_k}^n, G\rangle| \leq \lim_{n\to\infty} \Theta(\pi_\cdot^n) \leq \int_0^1 |G(u)|\, du + \varepsilon, \text{ for all } k \in \mathbb{N}.$$

Since $\pi_\cdot$ is right continuous, $t \mapsto |\langle \pi_t, G\rangle|$ is right continuous, which leads to

$$|\langle \pi_t, G\rangle| = \lim_{k\to\infty} |\langle \pi_{t_k}, G\rangle| \leq \int_0^1 |G(u)|\, du + \varepsilon.$$

Since $\pi_T^n \xrightarrow[n\uparrow\infty]{w} \pi_T$, we have that $|\langle \pi_T, G\rangle| = \lim_{n\to\infty} |\langle \pi_T^n, G\rangle| \leq \int_0^1 |G(u)|\, du + \varepsilon$, which concludes that $\pi_\cdot \in F_\varepsilon$, that is, F_ε is closed. Since, by hypothesis, $\{\mathbb{Q}^{n_k}\}_{n\in\mathbb{N}}$ converges weakly to $\mathbb{Q}^*$, by Portmanteau's Theorem, given a closed set F,

$$\limsup_{k\to\infty} \mathbb{Q}^{n_k}(F) \leq \mathbb{Q}^*(F).$$

Taking $F := F_\varepsilon$,

$$\mathbb{Q}^*(F_\varepsilon) \geq \limsup_{k\to\infty} \mathbb{Q}^{n_k}(F_\varepsilon) = 1,$$

which means that $\mathbb{Q}^*$ is concentrated on F_ε. Since ε is arbitrary, we have that $\mathbb{Q}^*$ is concentrated on $F_{\frac{1}{m}}$, for $m \in \mathbb{N}$. Moreover, since $F_{\frac{1}{m+1}} \subseteq F_{\frac{1}{m}}$ (and $\mathbb{Q}^*(F_1) < \infty$), we have by the property of continuity from above of measures that

$$\mathbb{Q}^*(F_0) = \mathbb{Q}^*\Big(\bigcap_{m=1}^{\infty} F_{\frac{1}{m}}\Big) = \lim_{m\to\infty} \mathbb{Q}^*(F_{\frac{1}{m}}) = 1,$$

where $F_0 := \bigcap_{m=1}^{\infty} F_{\frac{1}{m}} = \{\pi_\cdot \in D_{\mathcal{M}}[0,T] : \sup_{0\leq t\leq T} |\langle \pi_t^n, G\rangle| \leq \int_0^1 |G(u)|\, du\}$ is a set of trajectories whose measures satisfy the condition of Lemma 2.7.9. Hence, given a trajectory $\pi_\cdot$ on which $\mathbb{Q}^*$ is concentrated, we have that for all $t \in [0,T]$, π_t satisfies the condition of Lemma 2.7.9 which means that π_t is absolutely continuous with respect to the Lebesgue measure. □

We conclude from this result that $\mathbb{Q}^*$ is concentrated in absolutely continuous trajectories, i.e., for each t, $\pi_t(du) = \rho(t,u)du$ where ρ is the density of π. Note that in order to achieve this result, we do not

need to consider anything about the dynamics of the process (the process may have any generator), we just need η_t to be bounded.

For the second part, we prove rigorously the ideas explored before in the heuristic argument (Section 2.6). This is the main ingredient for the hydrodynamic limit since we are now characterizing the limit of the sequence, proving that it is concentrated on a set of trajectories whose density is a weak solution of the transport equation.

For the next proposition, in order to simplify the notation, denote a subsequence of $\{\mathbb{Q}^n\}_{n\geq 1}$ just by $\{\mathbb{Q}^n\}_{n\geq 1}$ and assume (without loss of generality) that it converges weakly to $\mathbb{Q}^*$.

Proposition 2.7.11. *Let $\mathbb{Q}^*$ be a limit point of a subsequence of $\{\mathbb{Q}^n\}_{n\geq 1}$ and assume without loss of generality that $\mathbb{Q}^n \xrightarrow{w} \mathbb{Q}^*$. Let $f \in C^{2,1}([0,T]\times\mathbb{T})$ and for $t\in[0,T]$, define*

$$A_t = \left\{\pi_\cdot \in D_{\mathcal{M}}[0,T] : \langle \rho_t, f_t\rangle - \langle \gamma, f_0\rangle - \int_0^t \langle \rho_s, \partial_s f_s\rangle\, ds + \int_0^t \langle \rho_s, \nabla f_s\rangle\, ds = 0\right\}. \tag{2.12}$$

Then, for $t\in[0,T]$, $\mathbb{Q}^*$ is concentrated on A_t, that is, $\mathbb{Q}^*(A_t)=1$.

Proof. Let $\delta>0$ and recall that $\pi_t(du)=\rho(t,u)du$. Fix $f\in C^{2,1}([0,T]\times\mathbb{T})$. Define $\Phi_f : D_{\mathcal{M}}[0,T]\to\mathbb{R}$ as

$$\Phi_f(\pi_\cdot) = \sup_{0\leq t\leq T}\left|\langle \pi_t, g_1\rangle - \langle \pi_0, g_2\rangle - \int_0^t \langle \pi_s, g_3\rangle\, ds\right|,$$

where $g_1 := f_t : \mathbb{T}\to[0,1]$, $g_2 := f_0 : \mathbb{T}\to[0,1]$, $g_3 := (\partial_s - \nabla)f_s : \mathbb{T}\to[0,1]$ are continuous for all $s,t\in[0,T]$. Since for any $t\in[0,T]$,

$$\Phi_f(\pi_\cdot) \geq \left|\langle \rho_t, f_t\rangle - \langle \gamma, f_0\rangle - \int_0^t \langle \rho_s, \partial_s f_s\rangle\, ds + \int_0^t \langle \rho_s, \nabla f_s\rangle\, ds\right|,$$

it is sufficient to show that

$$\mathbb{Q}^*(\pi_\cdot \in D_{\mathcal{M}}[0,T] : \Phi_f(\pi_\cdot) > \delta) = 0. \tag{2.13}$$

By Lemma A.3.9, since g_1, g_2 and g_3 are continuous, $\Phi_f(\pi_\cdot)$ is continuous as well. This means that $\{\pi_\cdot \in D_{\mathcal{M}}[0,T] : \Phi_f(\pi_\cdot)>\delta\}$ is an open set (because it is the inverse image of the open set $\{y\in\mathbb{R} : y>\delta\}$ by a continuous function) and by Portmanteau's Theorem, it is enough to show that

$$\liminf_{n\to\infty} \mathbb{Q}^n(\pi_\cdot \in D_{\mathcal{M}}[0,T] : \Phi_f(\pi_\cdot)>\delta) = 0, \tag{2.14}$$

as the probability in (2.13) is bounded from below by the limit above. Using the relationship $\mathbb{P}_{\mu_n}\circ(\pi^n)^{-1} := \mathbb{Q}^n$ between $\mathbb{Q}^n$ and $\mathbb{P}_{\mu_n}$, we get

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{Q}^n(\pi_\cdot \in D_{\mathcal{M}}[0,T] : \Phi_f(\pi_\cdot)>\delta) = \\
&\liminf_{n\to\infty} \mathbb{P}_{\mu_n}\Big(\eta_\cdot \in D_{\Omega_n}[0,T] : \sup_{0\leq t\leq T}\left|\langle \pi_t^n, f_t\rangle - \langle \pi_0^n, f_0\rangle - \int_0^t \langle \pi_s^n, (\partial_s-\nabla)f_s\rangle\, ds\right| > \delta\Big).
\end{aligned}$$

Summing and subtracting $n\mathcal{L}\langle \pi_s^n, f_s\rangle$ inside of the time integral, the previous limit is bounded by the sum

of

$$\liminf_{n\to\infty} \mathbb{P}_{\mu_n}\Big(\eta_\cdot \in D_{\Omega_n}[0,T] : \sup_{0\le t\le T} \Big|\langle \pi_t^n, f_t\rangle - \langle \pi_0^n, f_0\rangle - \int_0^t \langle \pi_s^n, \partial_s f_s\rangle \, ds - \int_0^t n\mathcal{L}_n\langle \pi_s^n, f_s\rangle \, ds\Big| > \frac{\delta}{2}\Big) \quad (2.15)$$

and

$$\liminf_{n\to\infty} \mathbb{P}_{\mu_n}\Big(\eta_\cdot \in D_{\Omega_n}[0,T] : \sup_{0\le t\le T} \Big| \int_0^t \langle \pi_s^n, \nabla f_s\rangle \, ds + \int_0^t n\mathcal{L}_n\langle \pi_s^n, f_s\rangle \, ds\Big| > \frac{\delta}{2}\Big). \quad (2.16)$$

So it is sufficient to show that (2.15) and (2.16) are zero.

- For (2.15):

  Applying Dynkin's formula as in (2.3), we have that $M_t^n(f)$ is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma(\eta_s : s \le t)$. Moreover, note that the martingale is equal to the term inside the absolute value in (2.15). By Doob's inequality (A.1.5) with $p := 2$, the probability in (2.15) satisfies

  $$\mathbb{P}_{\mu_n}\Big(\sup_{0\le t\le T} |M_t^n(f)| > \frac{\delta}{2}\Big) \le \frac{4}{\delta^2}\mathbb{E}\Big[|M_T^n(f)|^2\Big] = \frac{4}{\delta^2}\mathbb{E}\Big[\int_0^T \Gamma^n(\langle \pi_s^n, f_s\rangle) \, ds\Big] \lesssim \frac{1}{n\delta^2}\int_0^T \|\nabla f_s\|_\infty^2 \, ds,$$

  where the last inequality comes from Lemma 2.7.6. Therefore we conclude that

  $$\lim_{n\to\infty} \mathbb{P}_{\mu_n}\Big(\sup_{0\le t\le T} |M_t^n(f)| > \frac{\delta}{2}\Big) = 0.$$

- For (2.16):

  Looking at the action of the generator, we see that in order to prove that the limit is zero, it is sufficient to show that the following limits are zero:

  $$\lim_{n\to\infty} \mathbb{P}_{\mu_n}\Big(\eta_\cdot \in D_{\Omega_n}[0,T] : \sup_{0\le t\le T} \Big| \int_0^t (\langle \pi_s^n, \nabla_n^- f_s\rangle - \langle \pi_s^n, \nabla f_s\rangle) \, ds\Big| > \frac{\delta}{4}\Big) = 0;$$

  $$\lim_{n\to\infty} \mathbb{P}_{\mu_n}\Big(\eta_\cdot \in D_{\Omega_n}[0,T] : \sup_{0\le t\le T} \Big| \int_0^t \frac{1}{2n}(\eta_s(1) - \eta_s(0))\nabla_n^- f_s(0) \, ds\Big| > \frac{\delta}{4}\Big) = 0.$$

  For the first limit, note that

  $$|\langle \pi_s^n, \nabla_n^- f_s\rangle - \langle \pi_s^n, \nabla f_s\rangle| = \tfrac{1}{n}\sum_{x\in\mathbb{T}_n} \eta_t(x)|\nabla_n^- f_s(\tfrac{x}{n}) - \nabla f_s(\tfrac{x}{n})| \le \tfrac{1}{n}\sum_{x\in\mathbb{T}_n} |\nabla_n^- f_s(\tfrac{x}{n}) - \nabla f_s(\tfrac{x}{n})|.$$

  We can use Taylor's Theorem to simplify the difference of the derivatives:

  $$\begin{aligned}|\nabla_n^- f_s(\tfrac{x}{n}) - \nabla f_s(\tfrac{x}{n})| &= |n(f_s(\tfrac{x}{n}) - f_s(\tfrac{x-1}{n})) - \nabla f_s(\tfrac{x}{n})| = |n(\tfrac{1}{n}\nabla f_s(\tfrac{x}{n}) + o(\tfrac{1}{n})) - \nabla f_s(\tfrac{x}{n})| \\ &= n\, o(\tfrac{1}{n}).\end{aligned}$$

  It follows that for any t,

  $$\Big| \int_0^t (\langle \pi_s^n, \nabla_n^- f_s\rangle - \langle \pi_s^n, \nabla f_s\rangle) \, ds\Big| = n\, o(\tfrac{1}{n})$$

  converges to zero (with n). For the second limit, $\nabla_n^- f_s(0)$ converges to $\nabla f_s(0)$, which is bounded.

Thus, by a similar argument as the previous one, we conclude that the limit is zero.

□

Until now we have seen that $\mathbb{Q}^n$ converges to $\mathbb{Q}^*$, where $\mathbb{Q}^*$ is a Dirac delta on a trajectory $\pi_\cdot$ which is absolutely continuous with respect to the Lebesgue measure and whose density is a weak solution of (2.2). So we might have different subsequences of $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ converging to limit points concentrated on trajectories whose densities might not be the same. It is thus left to show the uniqueness of the weak solution of (2.2).

2.7.3 Uniqueness of weak solutions of the transport equation

Recall Definition 2.5.1. Assume that $\rho^l : [0,T]\times\mathbb{T}\to[0,1]$, $l=1,2$, are two weak solutions of (2.2) starting from the same initial condition. Define $\hat\rho = \rho^1 - \rho^2$ and observe that $\hat\rho(0,u)=0$ for any $u\in\mathbb{T}$. Thus, from the definition of weak solution, for all $f\in C^{2,1}([0,T]\times\mathbb{T})$ and $t\in[0,T]$

$$\int_{\mathbb{T}}\hat\rho(t,u)f_t(u)du = \int_0^t\int_{\mathbb{T}}\hat\rho(s,u)(\partial_s f_s(u))\,du\,ds - \int_0^t\int_{\mathbb{T}}\hat\rho(s,u)\,(\nabla f_s(u))\,du\,ds. \tag{2.17}$$

Now, for $t\in[0,T]$ and $u\in\mathbb{T}$, let $\phi_m(t)=\frac{1}{\sqrt{T}}e^{2\pi imt/T}$ and $\xi_k(u)=e^{2\pi iku}$ and recall that $\{\phi_m : m\in\mathbb{N}_0\}$ and $\{\xi_k : k\in\mathbb{N}_0\}$ are orthonormal basis of $L^2([0,T])$ and $L^2(\mathbb{T})$ respectively, for the inner products $\langle f,g\rangle=\int_0^T f(t)\overline{g(t)}\,dt$ and $\langle f,g\rangle=\int_{\mathbb{T}}f(u)\overline{g(u)}\,du$, where $\overline{g}$ is the conjugate of g. Thus, if $\psi_{m,k}(t,u)=\phi_m(t)\xi_k(u)=\frac{1}{\sqrt{T}}e^{2\pi i(mt/T+ku)}$, then $\{\psi_{m,k} : m,k\in\mathbb{N}_0\}$ is an orthonormal basis of $L^2([0,T]\times\mathbb{T})$ for the inner product $\langle f,g\rangle=\int_0^T\int_{\mathbb{T}}f(t,u)\overline{g(t,u)}\,du\,dt$. It is easy to verify that the functions are orthonormal

$$\begin{aligned}\langle\psi_{m,k},\psi_{m',k'}\rangle &= \int_0^T\int_{\mathbb{T}}\tfrac{1}{\sqrt{T}}e^{2\pi i(mt/T+ku)}\tfrac{1}{\sqrt{T}}e^{-2\pi i(m't/T+k'u)}\,du\,dt\\ &= \tfrac{1}{T}\int_0^T\int_{\mathbb{T}}e^{2\pi i((m-m')t/T+(k-k')u)}\,du\,dt = \begin{cases}1 & \text{if } m=m' \text{ and } k=k',\\ 0 & \text{if } m\neq m' \text{ or } k\neq k'.\end{cases}\end{aligned}$$

To see that the functions form a basis of $L^2([0,T]\times\mathbb{T})$, check Chapter 7 of [11]. Moreover, for any $m,k\in\mathbb{N}_0$ we have $\partial_s\psi_{m,k}(s,u)=2\pi i\frac{m}{T}\psi_{m,k}(s,u)$ and $\nabla\psi_{m,k}(s,u)=2\pi ik\psi_{m,k}(s,u)$. Therefore, replacing f_s by $\psi_{m,k}(s,\cdot)$ in (2.17) for fixed m and k, we obtain

$$\int_{\mathbb{T}}\hat\rho(t,u)\psi_{m,k}(t,u)du = 2\pi i\left(\tfrac{m}{T}-k\right)\int_0^t\int_{\mathbb{T}}\hat\rho(s,u)\psi_{m,k}(s,u)\,du\,ds.$$

Taking a time derivative on both sides of the above expression, we conclude that

$$\partial_t\left(\int_{\mathbb{T}}\hat\rho(t,u)\psi_{m,k}(t,u)\,du\right) = 2\pi i\left(\tfrac{m}{T}-k\right)\int_{\mathbb{T}}\hat\rho(t,u)\psi_{m,k}(t,u)\,du.$$

Hence, $\int_{\mathbb{T}} \hat{\rho}(t,u)\psi_{m,k}(t,u)\,du$ is a solution of the first order linear Ordinary differential equation $y'(t) - 2\pi i\left(\frac{m}{T} - k\right)\, y(t) = 0$, that is,

$$\int_{\mathbb{T}} \hat{\rho}(t,u)\psi_{m,k}(t,u)\,du = \left(\int_{\mathbb{T}} \hat{\rho}(0,u)\psi_{m,k}(0,u)\,du\right) e^{2\pi i(m/T-k)t} = 0.$$

Recall that we can write $\hat{\rho}$ as a linear combination of the functions $\{\psi_{m,k} : m,k \in \mathbb{N}_0\}$:

$$\hat{\rho}(t,u) = \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} c_{m,k}\,\psi_{m,k}(t,u)$$

with ($\hat{\rho}$ is a real function)

$$c_{m,k} = \langle \hat{\rho}, \psi_{m,k}\rangle = \int_{\mathbb{T}} \hat{\rho}(t,u)\psi_{m,k}(t,u)\,du = 0.$$

Therefore we conclude that $\hat{\rho}(t,u) = 0$ a.s. (almost surely), i.e., $\rho^1 = \rho^2$ a.s., which proves the uniqueness of weak solution of (2.2).

Chapter 3

Conclusions

In this thesis, we analyzed a card shuffle known as *the Rudvalis shuffle*. In the first chapter, we proved, by using tools from the theory of Markov chains, that performing this shuffle on a deck of n cards, we would need at least $O(n^3 \log(n))$ shuffles in order to have a well shuffled deck. As expected, the Rudvalis shuffle would not be a very practical way to shuffle a deck of cards but although the shuffle is "slow", we were able to use it to prove some interesting properties regarding the hydrodynamic limit.

In the second chapter, we considered the continuous version of the Rudvalis chain and we did the a mapping with a particle system, namely, we made a correspondence between the colors of the cards with particles in the following way: a black card corresponds to a particle and a red card corresponds to a hole. Therefore, from a deck of n cards, we obtained a particle system with particles and holes evolving on a discrete set with n sites. By shuffling the deck according to the Rudvalis shuffle, and by using the previous mapping, we were able to describe the behavior of the density in the particle system. More precisely, we showed that the space/time evolution of the density of the system is given by the unique weak solution of the transport equation. In conclusion, we were able to prove the existence (and uniqueness) of a weak solution of a partial differential equation by means of a random process.

Another interesting problem we could have considered, but for a matter of time we did not, was, instead of the Law of Large Numbers, the Central Limit Theorem for the empirical measure, which is often called density fluctuations. In this case, the equation we expect to obtain is no longer a deterministic one (like the transport equation), but instead a stochastic partial differential equation.

Bibliography

[1] Abramowitz, M., Stegun, I. A., et al. (1972). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. National bureau of standards Washington, DC.

[2] Bayer, D., Diaconis, P., et al. (1992). Trailing the dovetail shuffle to its lair. *Annals of applied probability*, 2(2):294–313.

[3] Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Interscience.

[4] Diaconis, P. (1988). Group representations in probability and statistics. *Lecture notes-monograph series*.

[5] Franco, T., Gonçalves, P., and Neumann, A. (2013). Hydrodynamical behavior of symmetric exclusion with slow bonds. *Annales de l'IHP Probabilités et statistiques*, 49(2):402–427.

[6] Goel, S. (2006). Analysis of top to bottom-k shuffles. *The Annals of Applied Probability*, 16:30–55.

[7] Guo, M., Papanicolaou, G., and Varadhan, S. (1988). Nonlinear diffusion limit for a system with nearest neighbor interactions. *Communications in Mathematical Physics*, 118(1):31–59.

[8] Häggström, O. (2002). *Finite Markov chains and algorithmic applications*. Cambridge University Press.

[9] Harris, T. (1972). Nearest-neighbor markov interaction processes on multidimensional lattices. *Advances in Mathematics*, 9:66 – 89.

[10] Hildebrand, M. V. (1990). *Rates of convergence of some random processes on finite groups*. Harvard University.

[11] Hunter, J. K. and Nachtergaele, B. (2001). *Applied analysis*. World Scientific Publishing Company.

[12] Kipnis, C. and Landim, C. (1999). *Scaling limits of interacting particle systems*. Springer-Verlag, Berlin.

[13] Levin, D. A., Peres, Y., and Wilmer, E. L. (2017). *Markov chains and mixing times*. American Mathematical Soc.

[14] Liggett, T. M. (2010). *Continuous time Markov processes: an introduction*. American Mathematical Soc.

[15] Wilson, D. (2003). Mixing time of the rudvalis shuffle. *Electronic Communications in Probability*, 8:77–85.

Appendix A

Auxiliary results

In the next results, denote by X a generic metric space and by $\mathcal{B}$ the σ-algebra generated by the open sets, the Borel σ-algebra on X. We will mainly follow Chapter 4.1 of [12] and the Chapter 3 of [3].

A.1 Probability theory

Proposition A.1.1 (Markov's inequality)**.** *Let Z be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}$. Assume that $\mathbb{E}[|Z|]^t < \infty$ for some $t \in \mathbb{N}$. Then, for any $\delta > 0$,*

$$P(|Z| \geq \delta) \leq \frac{\mathbb{E}[|Z|^t]}{\delta^t}$$

Theorem A.1.2 (Dynkin's formula)**.** *Let $\{\eta_t\}_{t\geq 0}$ be a Markov process with infinitesimal generator $\mathcal{L}$ and with countable state space J. Take a bounded function $F : \mathbb{R}^+ \times J \to \mathbb{R}$ such that:*

1. *for all $x \in J,\ F(\cdot, x) \in C^2(\mathbb{R}^+)$;*
2. *there is a constant C such that for $k = 1, 2$ we have $\sup_{(s,x)} |\partial_s^k F(s, x)| \leq C$.*

Define

$$M_t(F) := F(t, \eta_t) - F(0, \eta_0) - \int_0^t (\partial_s + \mathcal{L})F(s, \eta_s)\, ds,$$

$$N_t(F) := (M_t(F))^2 - \int_0^t \Gamma(F(s, \eta_s))\, ds, \text{ where } \Gamma(f) = \mathcal{L}(f^2) - 2f\mathcal{L}f \text{ is the carré du champ operator.}$$

Then, the sequences $\{M_t(F)\}_{t\geq 0}$ and $\{N_t(F)\}_{t\geq 0}$ are martingales with respect to the natural filtration $\mathcal{F}_t = \sigma(\eta_s : s \leq t)$.

Proof. The proof can be found in Appendix 1.5 of [12]. □

Remark A.1.3. *The integral term $\int_0^t \Gamma(F(s, \eta_s))\, ds$ is the quadratic variation of $M_t(F)$.*

Theorem A.1.4 (Portmanteau)**.** *Let $\{\mathbb{P}_n\}_{n\in\mathbb{N}}, \mathbb{P}$ be probability measures on $(X, \mathcal{B})$. Then the following conditions are equivalent:*

(i) $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$*;*

(ii) $\limsup_{n\to\infty} \mathbb{P}_n(F) \leq \mathbb{P}(F)$*, for all closed* $F \subseteq X$*;*

(iii) $\liminf_{n\to\infty} \mathbb{P}_n(G) \geq \mathbb{P}(G)$*, for all open* $G \subseteq X$*;*

Theorem A.1.5 (Doob's inequality)**.** *Let* $\{M_t\}_{0\leq t\leq T}$ *be a submartingale with right continuous trajectories. Let* $\delta > 0$ *and* $p \geq 1$*. Then*

$$\mathbb{P}\Big(\sup_{0\leq t\leq T} |M_t| > \delta\Big) \leq \frac{1}{\delta^p}\mathbb{E}[|M_T|^p]$$

and

$$\mathbb{E}\Big[\sup_{0\leq t\leq T} |M_t|\Big] \leq \Big(\frac{p}{1-p}\Big)^p \mathbb{E}[|M_T|^p].$$

Proposition A.1.6 (Tower Law)**.** *If* Z *is integrable and* $\mathcal{F}_1 \subseteq \mathcal{F}_2$ *are two* σ*-algebras, then*

$$\mathbb{E}[\mathbb{E}[Z|\mathcal{F}_2]|\mathcal{F}_1] = \mathbb{E}[Z|\mathcal{F}_1] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_1]|\mathcal{F}_2].$$

A.2 Tightness criteria

Theorem A.2.1 (Prohorov's theorem)**.** *Let* Π *be a family of probability measures on* $(X, \mathcal{B})$*.*

1. *If* Π *is tight, then* Π *is relatively compact;*
2. *If* X *is separable and complete, then* Π *is tight if and only if* Π *is relatively compact.*

Proof. The proof of this theorem can be found in Theorem 5.1 and Theorem 5.2 (Section 1.5) of [3]. □

Lemma A.2.2 (Proposition 1.7 of [12])**.** *Let* $\{g_k\}_{k\in\mathbb{N}}$ *be a dense (with respect to the uniform topology) sequence in* $C(\mathbb{T})$*. Let* $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ *be a sequence of probability measures on* $D_{\mathcal{M}}[0,T]$ *and define for each* k *the sequence* $\{\mathbb{Q}^{n,g_k}\}_{n\in\mathbb{N}}$ *of probability measures on* $D_{\mathbb{R}}[0,T]$ *where* $\mathbb{Q}^{n,g_k}$ *is the probability measure induced by* $\mathbb{Q}^n$ *and by the application*

$$\begin{aligned} \mathcal{G} : (D_{\mathcal{M}}[0,T], \mathbb{Q}^n) &\to (D_{\mathbb{R}}[0,T], \mathbb{Q}^{n,g_k}) \\ \{\pi_t^n\}_{t\geq 0} &\mapsto \{\langle \pi_t^n, g_k\rangle\}_{t\geq 0}. \end{aligned}$$

If for every k *the sequence* $\{\mathbb{Q}^{n,g_k}\}_{n\in\mathbb{N}}$ *is tight in* $D_{\mathbb{R}}[0,T]$ *then the sequence* $\{\mathbb{Q}^n\}_{n\in\mathbb{N}}$ *is tight in* $D_{\mathcal{M}}[0,T]$*.*

Remark A.2.3. *The above result tells us that in order to prove the tightness in* $D_{\mathcal{M}}[0,T]$*, we just need to check the tightness in* $D_{\mathbb{R}}[0,T]$ *(which is easier in general).*

Proposition A.2.4 (Aldous' criterion)**.** *Let* $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ *be a sequence of probability measures on* $D_{\mathbb{R}}[0,T]$*. Then* $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ *is tight in the Skorohod topology of* $D_{\mathbb{R}}[0,T]$ *if the next two conditions hold:*

(i) for each $t \in [0,T]$ *and* $\varepsilon > 0$*, there is a compact* $K \subseteq \mathbb{R}$ *such that*

$$\sup_{n\in\mathbb{N}} \mathbb{P}^n(x_\cdot \in D_{\mathbb{R}}[0,T] : x_t \notin K) < \varepsilon;$$

(ii) for each $\varepsilon > 0$*,*

$$\lim_{\gamma \to 0} \limsup_{n \to \infty} \sup_{\tau \in \mathcal{T}_T, \theta \leq \gamma} \mathbb{P}^n(x_\cdot \in D_{\mathbb{R}}[0,T] : |x_{\tau+\theta} - x_\tau| > \varepsilon) = 0,$$

where $\mathcal{T}_T$ *is the set of stopping times bounded by* T*.*

The Aldous' criterion is obtained from Theorem A.3.5 and Proposition A.3.7 which are stated in Section A.3.

A.3 Skorohod topology

Let (X, δ) be a metric space and $\{\mathbb{P}^n\}_{n \in \mathbb{N}}$ be a sequence of probability measures defined on $D_X[0,T]$, the space of right continuous functions on $[0,T]$ with left limits taking values in X. To endow this space with a reasonable topology (cf. Chapter 3 of [3]), consider the following definitions:

$$\Lambda = \{\lambda : [0,T] \to [0,T] \mid \lambda \text{ is a strictly increasing function}\},$$

$$\|\lambda\| = \sup_{s \neq t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|$$

and for two trajectories $x_\cdot, y_\cdot \in D_X[0,T]$,

$$d(x_\cdot, y_\cdot) := \inf_{\lambda \in \Lambda} \max \left\{ \|\lambda\|, \sup_{0 \leq t \leq T} \delta(x_t, y_{\lambda(t)}) \right\}.$$

Proposition A.3.1 (Theorem 12.2 of [3])**.** $D_X[0,T]$*, endowed with the metric* d*, is a complete separable metric space.*

Definition A.3.2 (Convergence in the Skorohod topology)**.** *We say that a sequence* $\{x_\cdot^n\}_{n \in \mathbb{N}}$ *of elements of* $D_X[0,T]$ *converges to a limit* $x_\cdot$ *in the Skorohod topology if there exist functions* λ_n *in* Λ *such that*

1. $\lim_{n \to \infty} x^n_{\lambda_n(t)} = x_t$ *uniformly in* t*;*
2. $\lim_{n \to \infty} \lambda_n(t) = t$ *uniformly in* t*.*

Consider the following definitions:

$$\omega_x(\gamma) := \sup_{|s-t| \leq \gamma} \delta(x_s, x_t)$$

$$\omega'_x(\gamma) := \inf_{\{t_i\}_{0 \leq i \leq r}} \max_{0 \leq i < r} \sup_{t_i \leq s < t < t_{i+1}} \delta(x_s, x_t)$$

where the infimum is taken over the partition $0 = t_0 < t_1 < \cdots < t_r = T$ with $t_i - t_{i-1} > \gamma$ for $i = 1, \ldots, r$.

Remark A.3.3. *A function* $x : [0,T] \to X$ *is in* $D_X[0,T]$ *if and only if* $\lim_{\gamma \to 0} \omega'_x(\gamma) = 0$ (cf. page 123 of [3])*.*

Proposition A.3.4 (Proposition 1.2 of [12])**.** $A \subseteq D_{\mathcal{M}}[0,T]$ *is relatively compact if and only if the next two conditions hold.*

1. $\{x_t \in X : x_\cdot \in A, t \in [0,T]\}$ *is relatively compact in* $\mathcal{M}$.

2. $\lim_{\gamma \to 0} \sup_{x \in A} \omega'_x(\gamma) = 0$.

Theorem A.3.5 (Theorem 1.3 of [12]). *Let* $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ *be a sequence of probability measures on* $D_X[0,T]$. *This sequence is relatively compact if and only if the next two conditions hold.*

1. *For* $t \in [0,T]$ *and* $\varepsilon > 0$, *there is a compact* $K(t,\varepsilon) \subseteq X$ *such that* $\sup_{n\in\mathbb{N}} \mathbb{P}^n(x_t \notin K(t,\varepsilon)) \leq \varepsilon$.

2. *For* $\varepsilon > 0$, $\lim_{\gamma\to 0} \limsup_{n\to\infty} \mathbb{P}^n(x_\cdot \in D_X[0,T] : \omega'_x(\gamma) > \varepsilon) = 0$.

Remark A.3.6. *Since* $\omega'_x(\gamma) \leq \omega_x(2\gamma)$, *we can replace the second condition of the previous theorem by:*

$$\textit{For } \varepsilon > 0,\ \lim_{\gamma\to 0} \limsup_{n\to\infty} \mathbb{P}^n(x_\cdot \in D_X[0,T] : \omega_x(\gamma) > \varepsilon) = 0.$$

Proposition A.3.7 (Proposition 1.6 of [12]). *Let* $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ *be a sequence of probability measures on* $D_X[0,T]$. *This sequence satisfies the second property of Theorem A.3.5 if for every* $\varepsilon > 0$,

$$\lim_{\gamma} \limsup_{n\to\infty} \sup_{\tau\in\mathcal{T}_T, \theta\leq\gamma} \mathbb{P}^n(x_\cdot \in D_X[0,T] : \delta(x_{\tau+\theta}, \mu_\tau) > \varepsilon) = 0.$$

In the previous results, two spaces are of our interest, $X := \mathcal{M}$ and $X := \mathbb{R}$. Recall that $\mathcal{M}$ is the space of positive measures on $\mathbb{T}$ with total mass bounded by 1, endowed with the weak topology. Furthermore, the space for the evolution of the empirical measure π^n_t is $D_{\mathcal{M}}[0,T]$, the set of right continuous functions with left limits taking values in $\mathcal{M}\cdot$ In order to define a distance in $\mathcal{M}$, consider a dense (with respect to the uniform topology) countable family $\{f_k\}_{k\in\mathbb{N}}$ of continuous functions on $\mathbb{T}$ and define $\hat{\delta}$ (which is a metric). For $\mu, \nu \in \mathcal{M}$,

$$\hat{\delta}(\mu,\nu) = \sum_{k\in\mathbb{N}} \frac{1}{2^k} \frac{|\langle \mu, f_k\rangle - \langle \nu, f_k\rangle|}{1 + |\langle \mu, f_k\rangle - \langle \nu, f_k\rangle|}. \tag{A.1}$$

In the case $X := \mathbb{R}$, we can take $\hat{\delta}$ as the usual distance distance in $\mathbb{R}$, $\hat{\delta}(x,y) = |x-y|$.

Lemma A.3.8. *Let* $\{x^n_\cdot\}_{n\in\mathbb{N}}, x_\cdot$ *be trajectories on* $D_X[0,T]$.

1. *If* $x^n_t \xrightarrow[n\uparrow\infty]{} x_t$ *uniformly in* t, *then the convergence is also valid in the Skorohod topology;*

2. *If* $X := \mathcal{M}$ *and* $x^n_\cdot \to x_\cdot$ *in the Skorohod topology then, for every almost every* $t \in [0,T]$ *including* $t = 0$ *and* $t = T$, $x^n_t \xrightarrow{w} x_t$.

Proof. By Definition A.3.2, there is a sequence of elements $\{\lambda_n\}_{n\in\mathbb{N}} \in \Lambda$ such that for $t \in [0,T]$, $|\lambda_n(t) - t| \xrightarrow[n\uparrow\infty]{} 0$ uniformly in t, and $x^n_{\lambda_n(t)} \xrightarrow[n\to\infty]{w} x_t$. For the first statement, just take $\lambda_n(t) := t$. For the second one, first note that since $\lambda_n(0) = 0$ and $\lambda_n(T) = T$, we have that $x^n_s \xrightarrow[n\uparrow\infty]{w} x_s$ for $s \in \{0,T\}$. Now let $A = \{t \in [0,T] : x_t \text{ is continuous in } t\}$ and let $t \in A$. Note that the Lebesgue measure of A is T because $x_\cdot$ is càdlàg (it has at most countably many discontinuities). If d is the metric on X defined above,

$$\lim_{n\to\infty} d(x_t, x^n_t) = 0.$$

By hypothesis and by symmetry of $\lambda_n(t)$ and $\lambda_n^{-1}(t)$ with t,

$$\sup_{0\leq t\leq T}|t-\lambda_n^{-1}(t)| = \sup_{0\leq t\leq T}|t-\lambda_n(t)| \xrightarrow[n\uparrow\infty]{} 0$$

which implies that $\lim_{n\to\infty}|t-\lambda_n^{-1}(t)|=0$. Since $x_\cdot$ is continuous at t, we have that

$$\lim_{n\to\infty} d(x_t, x_{\lambda_n^{-1}(t)}) = 0.$$

Furthermore,

$$d(x_{\lambda_n^{-1}(t)}, x_t^n) \leq \sup_{0\leq t\leq T} d(x_t, x^n_{\lambda_n(t)}) \xrightarrow[n\uparrow\infty]{} 0.$$

Finally, by the triangle inequality

$$d(x_t, x_t^n) \leq d(x_t, x_{\lambda_n^{-1}(t)}) + d(x_{\lambda_n^{-1}(t)}, x_t^n) \xrightarrow[n\uparrow\infty]{} 0.$$

□

Lemma A.3.9. *Let $g_1, g_2, g_3 \in C(\mathbb{T})$ and $\Phi : D_{\mathcal{M}}[0,T]\to\mathbb{R}$ be defined as follows:*

$$\Phi(\pi_\cdot) = \sup_{0\leq t\leq T}\left|\langle\pi_t, g_1\rangle - \langle\pi_0, g_2\rangle + \int_0^t \langle\pi_s, g_3\rangle\, ds\right|.$$

Then, Φ is continuous for the Skorohod metric in $D_{\mathcal{M}}[0,T]$.

Proof. Check Proposition 8.3 of [5]. □

Alphabetical Index