Return probabilities on groups and large deviations for permuton processes

by

Michał Kotowski

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

Abstract

Return probabilities on groups and large deviations for permuton processes

Michał Kotowski
Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto
2016

The topic of this thesis are random processes on finite and infinite groups. More specifically, we are concerned with random walks on finitely generated amenable groups and stochastic processes which arise as limits of trajectories of the interchange process on a line.

In the first part of the thesis we construct a new class of finitely generated groups, called bubble groups. Analysis of the random walk on such groups shows that they are non-Liouville, but have return probability exponents close to $1/2$. Such behavior was previously unknown for random walks on groups. Our construction is based on permutational wreath products over tree-like Schreier graphs and the analysis of large deviations of inverted orbits on such graphs.

In the second part of we analyze large deviations of the interchange process on a line, which can be thought of as a random walk in the group of all permutations, with adjacent transpositions as generators. This is done in the setting of random permuton processes, which provide a notion of a limit for a permutation-valued stochastic processes. More specifically, we provide bounds on the probability that the trajectory of the interchange process (as a permuton process) is close in distribution to a deterministic permuton process. As an application, we show that short paths joining the identity and the reverse permutation in the Cayley graph of $\mathcal{S}_n$ are typically close to the so-called sine curve process, which is the conjectured limit of random sorting networks. The analysis is done in the framework of interacting particle systems.

Acknowledgements

First of all, I would like to thank all my friends for being an endless source of support and inspiration. Mentioning everyone would take too much space, but I would like to especially thank Piotrek Achinger, Marcin “Grzybek” Grzybowski, Rysiek Kostecki, Paweł Marczewski, Piotrek Migdał, Janek Szejko, Marysia Zimmermann, my sister Magda, my brother Marcin and all people from Polish Children’s Fund.

I would like to thank Bálint for being a wonderful advisor and friend. Without his patience, kindness and eagerness to share mathematical and non-mathematical knowledge this thesis would not have been completed. To me Bálint remains a role model of an ideal advisor if there ever was one.

I am grateful to the Alfréd Rényi Institute of Mathematics in Budapest, where I spent a large part of my PhD, for its hospitality. I would like to thank Miklós Abért and Gábor Pete for interesting mathematical conversations. Gábor’s lecture notes “Probability and Geometry on Groups” were an important inspiration for me to get involved in this field of research.

Special thanks go to Piotr Przytycki. Back in 2011, when I was his student in Warsaw, he sent out an email about Spring School on Limits of Finite Graphs in Leipzig, with a comment ”I don’t know if this is interesting, but maybe”. Not only did it turn out to be very interesting, but it was in Leipzig that I first met Bálint and somehow the atmosphere of mathematical adventure that I remember from the workshop has had a deep impact on me.

Last but not least, I would like to thank Marysia Zimmermann for her vigorous support during the final stages of writing this thesis. Way back when the results of Chapter 2 were still mostly at the level of conjectures and ideas, I told her about Theorem 2.4.1 and she said: “So it means that everyone will eventually end up where they are supposed to be, regardless of whether they are pushed forward or slowed down by their neighbors - sounds rather uplifting!”. All the people mentioned here pushed me forward a lot and I am grateful for their help in getting me to where I am supposed to be not “eventually almost surely”, as is usually the case in probability theory, but in finite time.

Contents

1 Non-Liouville groups with return probability exponent at most $1/2$ **1**

1.1 Introduction 1

1.2 Preliminaries 4

1.3 The bubble group 5

1.4 Bounds on the inverted orbits 5

1.5 Liouville property and transience 8

1.6 Lower bound on return probability 9

2 Limits of random permuton processes and large deviations for the interchange process **13**

2.1 Introduction 13

2.2 Permutons and stochastic processes 16

2.3 Interchange process and stationarity 18

2.4 Law of large numbers 23

2.5 One block estimate 30

2.6 Large deviation lower bound 34

2.7 Large deviation upper bound 37

Bibliography **43**

Chapter 1

Non-Liouville groups with return probability exponent at most $1/2$

1.1 Introduction

One of the basic topics of study in probability and group theory is the behavior of random walks on Cayley graphs of finitely generated groups. Among the interesting parameters of a random walk is the *return probability* $p_{2n}(e,e)$. There are examples for which it decays polynomially in n (like $\mathbb{Z}^d$ or, more generally, groups of polynomial volume growth) or exponentially (which is the case exactly for nonamenable groups). Other, intermediate types of behavior are also possible, which motivates the study of possible exponents γ for which $p_{2n}(e,e) \approx e^{-n^\gamma}$. For example, every group of exponential growth must have $\gamma \geq 1/3$ (see [Var91]).

Another important parameter is the *speed* (or drift) of the random walk. The average distance $\mathbb{E}d(X_0, X_n)$ of the random walk from the origin after n steps may grow linearly with n, in which case we say that the random walk has *positive speed*, or slower, in which case we say that the random walk has *zero speed*. It is thus interesting to ask what exponents $\beta < 1$ such that $\mathbb{E}d(X_0, X_n) \approx n^\beta$ are possible. For example, it is known that for every finitely generated group we have $\beta \geq 1/2$ [LP13], but generally computing speed seems more difficult than computing return probabilities. Note that the exponents γ and β as above need not exist (the return probability and average distance from the origin can oscillate at different scales, see [Bri13]), so in general one should speak about $\liminf$ and $\limsup$ exponents.

Speed of the random walk is closely related to the properties of harmonic functions on groups. Recall that a group has the *Liouville property* (with respect to some generating set) if every bounded harmonic function on its Cayley graph is constant. A classical result (see for example discussion in [Pet14, Chapter 9]) says that for groups (though not for general transitive graphs) having positive speed is equivalent to non-Liouville property. Note, however, that it is not known if this property is independent of the generating set (or, more generally, the step distribution of the random walk), which is in contrast to return probabilities, whose decay rate is stable under quasi-isometries ([PSC00]).

The motivation for this paper is the following remarkable theorem (which is a corollary of a more general result from [SCZ]): if the return probability satisfies $p_{2n}(e,e) \geq Ke^{-cn^\gamma}$ for $\gamma < 1/2$ (and some constants

$K, c > 0$), then the group has the Liouville property [1]. In particular, it has zero speed for every generating set (since, as mentioned above, the property $\gamma < 1/2$ is invariant under quasi-isometries). This is the first known general result connecting return probabilities with speed and showing quasi-isometry invariance of the Liouville property for a broad class of groups. For more discussion of possible relationships between these exponents (and also other quantities like entropy or volume growth) and numerous examples, see ([Gou14, Section 4]).

This result does not characterize the Liouville property, since there exist groups with γ arbitrarily close to 1 which are still Liouville [BE14]. In the other direction, it is natural to ask whether the value 1/2 in the theorem cited above can be improved, i.e. whether there exist groups with γ arbitrarily close to 1/2 from above (or even equal to 1/2) which are non-Liouville. Several examples of groups with $\gamma = 1/2$ are known ([PSC02]), but they all have the Liouville property.

The main result of our paper is the construction of a finitely generated group which has $\gamma \leq 1/2$, but at the same time is non-Liouville. More precisely, consider the *upper return probability exponent*:

$$\overline{\gamma} = \limsup_{n\to\infty} \frac{\log|\log p_{2n}(e,e)|}{\log n}$$

We will prove the following theorem:

Theorem 1.1.1. *There exists a finitely generated group G and a symmetric finitely supported random walk μ on G such that G is non-Liouville with respect to μ and the upper return probability exponent satisfies $\overline{\gamma} \leq 1/2$.*

In other words, the return probability for this random walk satisfies the lower bound $p_{2n}(e,e) \geq Ke^{-n^{1/2+o(1)}}$ for some constant $K > 0$ and the random walk has positive speed. Previously the smallest known return probability exponent for a non-Liouville group was 3/5 for the lamplighter group $\mathbb{Z}_2 \wr \mathbb{Z}^3$ ([PSC02]). Determining a good upper bound for the return probability on G seems to be an interesting problem in its own right.

Idea of the construction

We now sketch the idea of our construction. Among the groups for which one can provide precise asymptotics for the return probabilities are the lamplighter groups $\mathbb{Z}_2 \wr \mathbb{Z}^d$. It is known [PSC02, Theorem 3.5] that in this case we have $\gamma = \frac{d}{d+2}$ - in particular, for $d = 2$ we obtain a group with $\gamma = 1/2$. The group $\mathbb{Z}_2 \wr \mathbb{Z}^2$ is Liouville, but only barely so, as its speed satisfies $\mathbb{E}d(X_0, X_n) \approx \frac{n}{\log n}$. Thus the idea is that if one could in some sense do the lamplighter construction for $d \approx 2 + \varepsilon$ for some small ε, or even $d \approx 2 + o(1)$ (which would correspond to putting the lamps on a graph with volume growth slightly faster than quadratic), one would get a group with γ close to 1/2 and, if the graph grows quickly enough, positive speed.

The problem is of course that there are no "$2+\varepsilon$"-dimensional Cayley graphs. Nevertheless, one can carry out the lamplighter construction over an almost two dimensional graph (this time only a Schreier graph, not a Cayley graph) if we move from ordinary wreath products to *permutational wreath products*. They are a generalization of wreath products to the setting where a finitely generated group acts on a Schreier graph

[1] This theorem was first announced in [Gou14], but the proof there relies on an assumption about off-diagonal heat kernel bounds which has not been proved to hold except for groups of polynomial growth.

(the usual wreath product would correspond to the group acting on itself). They share some similarities with the ordinary lamplighter groups, but there are also important differences (see Section 1.2 for more discussion).

For the construction of the group G we define a tree-like Schreier graph S which grows sufficiently quickly so that the simple random walk on it is transient. The graph naturally defines a group Γ which we call the *bubble group*. The group G is then defined as the permutational wreath product $\mathbb{Z}_2 \wr_S \Gamma$, which corresponds to putting $\mathbb{Z}_2$-valued lamps on S, with Γ acting on lamp configurations. One can show that this product is non-Liouville as soon as S is transient.

In the case of the usual lamplighter group $\mathbb{Z}_2 \wr \mathbb{Z}^d$, providing a lower bound on the return probability requires understanding the range of the simple random walk on the underlying base graph $\mathbb{Z}^d$ (roughly speaking, the dominant contribution to returning to identity in the wreath product comes from switching off all the lamps visited, and the number of visited lamps is governed by the range of the underlying random walk). To obtain a sharp bound we need to know certain large deviation estimates for the range, not only its average size. For permutational wreath products the situation is more complicated, as the size of the lamp configuration on S is governed not by the range of the simple random walk on S, but by the *inverted orbit process*. This is a different random process which is generally not as well understood. In our case the graph S has large parts which locally look like $\mathbb{Z}$, so one can still analyze the inverted orbits using large deviation estimates for $\mathbb{Z}$.

As a closing remark we mention that the idea of using "bubble graphs" comes from looking at orbital Schreier graphs of certain groups of bounded activity acting on trees (used in [AV12] to provide examples of groups with speed exponents between 3/4 and 1), which have somewhat similar branching structure. In particular, Gady Kozma (personal communication, see also [AK]) proposed looking at similar groups permuting vertices of slowly growing trees as examples in group theory. In general it would be desirable to obtain a better understanding of inverted orbits and probabilistic parameters (return probabilities, speed, entropy) on related groups of this type. Some results along these lines can be found for example in [Bri13], where entropy and return probability exponents on groups of directed automorphisms of bounded degree trees are analyzed.

Structure of the paper and notation

The paper is structured as follows. In Section 1.2 we provide the background on permutational wreath products, inverted orbits and switch-walk-switch random walks used for the wreath products. In Section 1.3 we define the family of Schreier graphs and bubble groups used in the main construction. In Section 1.4 we provide estimates on the size of inverted orbits for random walks on the graph. In Section 1.5 we state the theorem used to deduce the non-Liouville property from transience and provide a criterion for checking that the graph defined in the previous section is transient. In Section 1.6 we fix the Schreier graph and the bubble group, prove the graph's transience and provide lower bounds on return probabilities (using results from Section 1.3), thus proving Theorem 1.1.1.

Throughout the paper by c we will denote a positive constant (independent of parameters like m or n) whose exact value is not important and may change from line to line. We will also use the notation $f(n) \lesssim g(n)$ meaning $f(n) \leq Cg(n)$ for some constant $C > 0$.

1.2 Preliminaries

Let us recall the notion of a permutational wreath product. Suppose we have a finitely generated group Γ acting on a set S and a finitely generated group Λ (in our case this group will be finite). For $x \in S$ we will denote the action of $g \in \Gamma$ on x by $x.g$. The graph will usually have a distinguished vertex o called the *root*.

The *permutational wreath product* $\Lambda \wr_S \Gamma$ is the semidirect product $\bigoplus_S \Lambda \rtimes \Gamma$, where Γ acts on the direct sum by permuting the coordinates according to the group action. Elements of the permutational wreath product can be written as pairs (f, g), where $g \in \Gamma$ and $f : S \to \Lambda$ is a function with only finitely many non-identity values. For two such pairs (f, g), (f', g') the multiplication rule is given by:

$$(f, g)(f', g') = (f f'^{g^{-1}}, gg')$$

where $f^{g^{-1}}$ is defined as $f^{g^{-1}}(x) = f(x.g)$. If Γ and Λ are finitely generated, then $\Lambda \wr_S \Gamma$ is also finitely generated.

By $\mathrm{supp} f$ we will denote the set of vertices of S at which $f(s)$ is not identity.

The usual wreath products (with $S = \Gamma$) are often called lamplighter groups - we think of f as being a configuration of lamps on S and g being the position of a lamplighter. A random walk on the lamplighter group corresponds to the lamplighter doing a random walk on Γ and changing values of the lamps along his trajectory.

By analogy with the usual wreath product we will call Γ the *base group* and Λ the *lamp group*. There are however important differences in how random walks on permutational wreath products behave. To see this, consider a symmetric probability distribution μ on Γ and a *switch-walk-switch* random walk $\tilde{X}_n$ on $\Lambda \wr_S \Gamma$:

$$\tilde{X}_n = \prod_{i=1}^{n} (l_i, id_\Gamma)(id_\Lambda, g_i)(l'_i, id_\Gamma)$$

Here g_i are elements of Γ chosen independently according to μ and l_i, l'_i are independent random *switches* of the form:

$$l_i(x) = \begin{cases} id_\Lambda & \text{if } x \neq o \\ L & \text{if } x = o \end{cases}$$

where L is chosen randomly from a fixed symmetric probability distribution on Λ. We can write $\tilde{X}_n = (X_n, Z_n)$, where $Z_n = g_1 \ldots g_n$ is the random walk on Γ corresponding to μ and X_n is a random configuration of lamps on S. We will always assume that the probability distribution on Λ is nontrivial.

Now observe that if we interpret this walk as a lamplighter walking on Γ and switching lamps on S, the switches happen at locations $o, o.g_1^{-1}, o.g_2^{-1}g_1^{-1}, \ldots,\ o.g_n^{-1} \ldots g_2^{-1}g_1^{-1}$. For ordinary wreath products, with o being the identity of the base group, this is the same as the orbit of the left Cayley graph, $o, g_1^{-1}.o, g_2^{-1}g_1^{-1}.o, \ldots,\ g_n^{-1} \ldots g_2^{-1}g_1^{-1}.o$. However, in general the set of locations at which switches happen behaves differently from the usual orbit - for example, it does not even have to be connected.

This phenomenon motivates the definition of the inverted orbit. Suppose that, as above, we have a group Γ, acting from the right on a set S, and a word $w = g_1 \ldots g_n$, where g_i are generators of Γ. Given $o \in S$, its *inverted orbit* under the word w is the set $\mathcal{O}(w) = \{o, o.g_1^{-1}, o.g_2^{-1}g_1^{-1}, \ldots, o.g_n^{-1}g_{n-1}^{-1} \ldots g_1^{-1}\}$.

Likewise, suppose we have a symmetric probability distribution μ on Γ and the corresponding random walk $Z_n = g_1 g_2 \dots g_n$, where each $g_i \in \Gamma$ is chosen independently according to μ. Given $o \in S$, its inverted orbit under the random walk Z_n is the (random) set $\mathcal{O}(Z_n) = \{o, o.g_1^{-1}, o.g_2^{-1} g_1^{-1}, \dots, o.g_n^{-1} g_{n-1}^{-1} \dots g_1^{-1}\}$. We call the set-valued process $\mathcal{O}(Z_n)$ the *inverted orbit process* on S. Abusing the notation slightly we will denote by Z_n both the trajectory of the random walk up to time n and the corresponding group element.

As noted above, this is not the same as the ordinary orbit, which would correspond to the set $\{o, o.g_1, o.g_1 g_2, \dots, o.g_1 g_2 \dots g_n$ In particular, the inverted orbit process is not a reversible Markov process.

There are many examples in which permutational wreath products behave differently from the usual wreath products. For instance, while usual wreath products always have exponential growth if the base group is infinite and the lamp group is nontrivial, permutational wreath products can have intermediate growth. This is directly related to the difference between the behavior of inverted orbits and ordinary orbits (see [BE12] and other work by Bartholdi and Erschler).

1.3 The bubble group

We start by defining the Schreier graph and the group acting on it. Fix a *scaling sequence* $1 \leq \alpha_1 \leq \alpha_2 \leq ....$ The corresponding graph $S(\alpha)$ is constructed as follows. The edges of the graph are labelled by two generators a, b and their inverses. The graph is constructed recursively - the first level consists of the *root* o, followed by a cycle of length $2\alpha_1$. The n-th level is defined in the following way - place a cycle of length 3 (called a *branching cycle*), labelled cyclically by b, in the middle of each cycle from the previous level so that each cycle is split into two paths. Then each of the remaining two vertices on the branching cycle is followed by a cycle of length $2\alpha_n$ (see the picture below). For a given cycle from the n-th level we will denote its starting point by b_n (with $b_1 = o$). We will think of the graph as extending to the right, so the particles most distant from the root are the rightmost ones.

The edges of every path are labelled by a and a^{-1} and every vertex, apart from the vertices on the branching cycles, is mapped by b and b^{-1} to itself.

From this graph we obtain a group in natural way. Each of the generators a, b and their inverses defines a permutation of the vertices of $S(\alpha)$ and we define *the bubble group* $\Gamma(\alpha)$ as the group generated by a and b. $\Gamma(\alpha)$ acts on $S(\alpha)$ from the right and by $x.g$ we will denote the action of $g \in \Gamma(\alpha)$ on a vertex $x \in S(\alpha)$. By $d(x, y)$ we will denote the distance of x and y in S.

1.4 Bounds on the inverted orbits

In what follows we denote $S(\alpha)$ and $\Gamma(\alpha)$ by S and Γ for simplicity.

Consider the simple random walk Z_n on Γ (each of the generators a, b, a^{-1}, b^{-1} is chosen with equal probability) and the corresponding inverted orbit process $\mathcal{O}(Z_n)$ on S. Our goal is to prove that, for a suitably chosen scaling sequence, the inverted orbit process on the Schreier graph S satisfies the same bound on the range as the simple random walk on $\mathbb{Z}$.

Let $s_k = \alpha_1 + \ldots + \alpha_k + k$ be the total distance from o to the branching point b_{k+1}, with $s_0 = 0$.

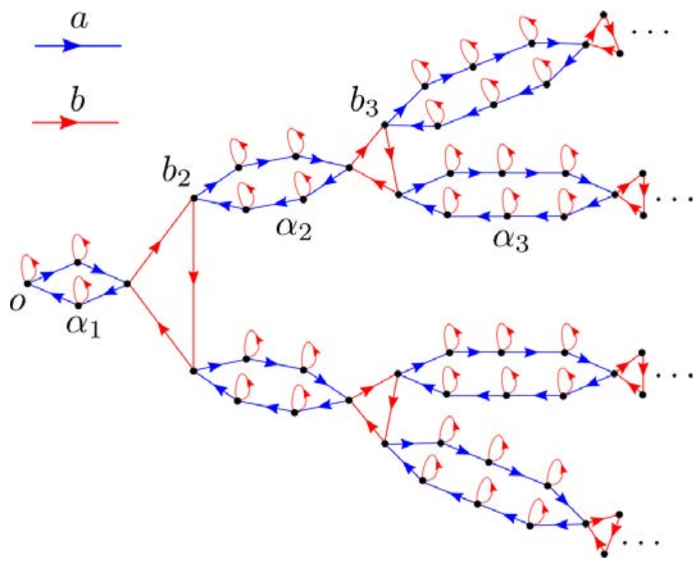

Figure 1.1: First three levels of the Schreier graph $S(\alpha)$ for $\alpha_1 = 2$, $\alpha_2 = 3$, $\alpha_3 = 4$.

Assumption 1. From now on we will assume that the scaling sequence satisfies:

$$ds_{k-1} \leq \alpha_k$$

for all $k \geq 2$ and some constant $d > 0$.

In other words, we require each level to be of length comparable to the sum of all previous levels, so that the graph S is like a tree with branches of length growing at least exponentially.

We want to reduce bounding the inverted orbit of Z_n to analyzing a one-dimensional random walk. To any given word w in a, b, a^{-1}, b^{-1} we can naturally associate a path on $\mathbb{Z}$ - a corresponds to moving right, a^{-1} corresponds to moving left and b, b^{-1} both correspond to staying put. As a, b, a^{-1}, b^{-1} appear with equal probability as steps of Z_n, we get that the random walk $Z_n = g_1 \dots g_n$ projects to a *lazy random walk* $\hat{Z}_n = \hat{g}_1 \dots \hat{g}_n$ on $\mathbb{Z}$ (started at the origin), which moves right with probability 1/4, moves left with probability 1/4 and stays put with probability 1/2.

Let R_n denote the range of $\hat{Z}_n$, i.e. the set of all vertices visited by $\hat{Z}_n$ up to time n. Let $A_{n,m}$ denote the event that the range of $\hat{Z}_n$ is contained in a small ball, $A_{n,m} = \{R_n \subseteq [-m, m]\}$. We have the following lemma on large deviations of a lazy random walk:

Lemma 1.4.1. *For every n, $m \geq 1$ we have:*

$$\mathbb{P}(A_{n,m}) = \mathbb{P}(R_n \subseteq [-m, m]) \gtrsim e^{-c\frac{n}{m^2}}$$

Proof. See [Ale92, Lemma 1.2] (or [PSC02, Theorem 3.12] for a more general case). ☐

The following simple observation will be useful: if the trajectory $\hat{g}_1 \dots \hat{g}_n$ has its range bounded between $-m$ and m, then for any subword $w = g_k g_{k+1} \dots g_l$ the trajectory $\hat{g}_k \hat{g}_{k+1} \dots \hat{g}_l$ (started at the origin) has its range bounded between $-2m$ and $2m$. Furthermore w has range bounded between $-2m$ and $2m$ if and only if $w^{-1} = g_l^{-1} \dots g_{k+1}^{-1} g_k^{-1}$ satisfies the same bound.

Now consider a particle moving on the graph according to the action of a word w or its inverse, starting at some vertex x. For two vertices y, z we will say that y is *to the right* (resp. *to the left*) of z if $d(o,z) < d(o,y)$ (resp. $d(o,z) > d(o,y)$).

We will repeatedly use the following lemma (which is a direct consequence of the observation above and the assumption $A_{n,m}$):

Lemma 1.4.2. *Suppose that $A_{n,m}$ holds for a word w. Let v be a vertex visited by the particle at some sequence of times and consider any subword w' of w corresponding to the minimal part of the trajectory between two subsequent visits to v (or after the last visit, if v is not visited after certain time). Whenever the particle visits v, if there is no branching cycle within distance $2m$ to the right (resp. to the left) of v, then w' will move the particle no further away than $2m$ to the right (resp. to the left) from v.*

Theorem 1.4.3. *Suppose that the scaling sequence satisfies Assumption 1. If $A_{n,m}$ holds for the trajectory $Z_n = g_1 \dots g_n$, then for each $x \in S$ and every subword $w = g_k g_{k+1} \dots g_l$ or its inverse we have $d(x, x.w) \leq Km$ (for some $K \geq 1$).*

Proof. The idea of the proof is that due to the assumption on exponential-like growth, the largest level contained in $B_m(x)$ is roughly of the same size as the whole ball, so we can bound the particle's position by looking only at its behavior at the last level (or levels of comparable size), where it behaves like a walk on $\mathbb{Z}$.

We consider three types of vertices: such that $B_{2m}(x)$ intersects only one level, intersects two levels or intersects at least three levels.

(1) In the first case there is no branching cycle within distance $2m$ from x, so the ball $B_{2m}(x)$ is isomorphic to a ball in $\mathbb{Z}$ and we can directly use the assumption $A_{n,m}$ to conclude that the particle stays within distance at most $2m$ from x.

(2) In the second case, assume that x belongs to the k-th level and the ball intersects also the $k+1$-st level (the case when the ball intersects the $k-1$-st level is analogous). To the left the ball doesn't intersect any branching cycle, so we can again directly use the property $A_{n,m}$. To the right, either the particle doesn't hit any b_{k+1}, in which case it is within distance $2m$ to the right of x, or it hits b_{k+1} (for one of the two cycles from the $k+1$-st level) - then we can apply Lemma 1.4.2 with $v = b_{k+1}$ to conclude that it never goes further than $2m$ to the right of b_{k+1}. This implies that we always stay within distance at most $4m$ from x.

(3) In the third case x must be close to the origin. Namely, if x belongs to the k-th level, then at least one of $\alpha_{k-1}, \alpha_k, \alpha_{k+1}$ is smaller than $4m$ (since $B_{2m}(x)$ intersects at least three levels). Since $\alpha_{k-1} \leq \alpha_k \leq \alpha_{k+1}$, we have $\alpha_{k-1} < 4m$. As $\alpha_{k-1} \geq d s_{k-2}$ by Assumption 1, we have $(1+d)\alpha_{k-1} \geq d(s_{k-1} - 1)$. Now $B_{2m}(x)$ intersects the $k-1$-st level (otherwise we would have $2m \leq \alpha_k \leq \alpha_{k+1}$ and the ball would intersect only two levels), so $d(o,x) \leq s_{k-1} + 2m$. This gives us:

$$d(o,x) \leq \frac{1+d}{d}\alpha_{k-1} + 1 + 2m \leq \left(2 + \frac{4(1+d)}{d}\right) m + 1 \leq \left(3 + \frac{4(1+d)}{d}\right) m$$

Thus x belongs to a ball $B_{c_1 m}(o)$, where c_1 is the constant on the right hand side of the inequality above.

Now take the first level l which has $\alpha_l \geq 4m$. Then b_l is to the right of x and $\alpha_{l-1} < 4m$. We have $d(o, b_l) = s_{l-1}$ and $ds_{l-2} \leq \alpha_{l-1}$, so $d(s_{l-1} - 1) \leq (1+d)\alpha_{l-1} < (1+d)4m$. Thus:

$$d(o, b_l) < \frac{4(1+d)}{d} m + 1 \leq \left(\frac{4(1+d)}{d} + 1\right) m$$

Let c_2 be the constant multiplying m in the inequality above. If the particle stays to the left of b_l, it is within distance at most $c_2 m$ from the origin and thus within distance at most $(c_1 + c_2)m$ from x. If it hits b_l at some point, then, as $\alpha_l \geq 4m$, for each visit we can apply Lemma 1.4.2 with $v = b_l$ to conclude that the particle stays within distance $4m$ to the right from b_l, so it is within distance $(4 + c_2)m$ from the origin and thus within distance $(4 + c_1 + c_2)m$ from x.

Thus the theorem holds with $K = 4 + c_1 + c_2$. $\square$

Corollary 1.4.4. *Under the assumption of the previous theorem, if $A_{n,m}$ holds, the inverted orbit process $\mathcal{O}(Z_n)$ on S satisfies $\mathcal{O}(Z_n) \subseteq B_{Km}(o)$, where $B_{Km}(o)$ denotes the ball of radius Km and center o in S (with K as in the previous theorem).*

Proof. Recall that $\mathcal{O}(Z_n) = \{o, o.g_1^{-1}, o.g_2^{-1}g_1^{-1}, \ldots, o.g_n^{-1}g_{n-1}^{-1} \ldots g_1^{-1}\}$. We can apply the previous theorem to words of the form $g_k^{-1} \ldots g_2^{-1} g_1^{-1}$ for $k = 1, \ldots, n$. We get that
$d(o, o.g_k^{-1} \ldots g_2^{-1} g_1^{-1}) \leq Km$, which proves $\mathcal{O}(Z_n) \subseteq B_{Km}(o)$. $\square$

Thus with probability at least a constant times $e^{-c\frac{n}{m^2}}$ no vertex is moved by Z_n further than Km from itself and the inverted orbit of o is small (contained in a ball of radius Km around o).

1.5 Liouville property and transience

We briefly recall the notions related to the Liouville property and harmonic functions. Given a measure μ on a group G, a function $f : G \to \mathbb{R}$ is said to be *harmonic* (with respect to μ) if we have $f(g) = \sum_{h \in G} f(gh)\mu(h)$. G is said to have the *Liouville property* if every bounded harmonic function on G is constant. As mentioned in the introduction, this is equivalent to the random walk associated to μ having zero asymptotic speed. This property a priori depends on the choice of μ (in the case when μ is a simple random walk - on the choice of the generating set of G).

We want to construct a group which is non-Liouville, i.e. supports nonconstant bounded harmonic functions. For permutational wreath products one can ensure this by requiring that the Schreier graph used in the wreath product is transient:

Theorem 1.5.1. *Let Γ and F be nontrivial finitely generated groups and let μ be a finitely supported symmetric measure on Γ whose support generates the whole group. Let $\tilde{\mu}$ be the measure associated to the corresponding switch-walk-switch random walk on the permutational wreath product $F \wr_S \Gamma$. If the induced random walk on S is transient, then the group $F \wr_S \Gamma$ has nontrivial Poisson boundary, i.e. supports nonconstant bounded harmonic functions (with respect to $\tilde{\mu}$).*

Related results appear in several places [AV14]. The formulation we use here comes from [BE11, Proposition 3.5]. We briefly sketch the idea of the construction here.

To construct a nonconstant harmonic function on the group, consider the state of the lamp at o. Since the walk on S is transient, with probability 1 this vertex will be visited only finitely many times, so after a certain point the value of the lamp will not change anymore and thus the eventual state L of this lamp is well-defined as $n \to \infty$. Now one can show that for any vertex x the mapping $x \mapsto \mathbb{P}_x(L = e)$ (where $\mathbb{P}_x$ denotes the probability with respect to a random walk started at x) defines a nonconstant bounded harmonic function on the group.

A useful criterion for establishing transience is based on electrical flows (we formulate it for simple random walks). Given a graph S, a *flow* I from a vertex o is a nonnegative real function on the set of directed edges of S which satisfies Kirchhoff's law: for each vertex except o the sum of incoming values of I is equal to the sum of outgoing values. A *unit flow* is a flow for which the outgoing values from o sum up to 1. The *energy* of the flow is given by $\mathcal{E}(I) = \frac{1}{2}\sum_e I(e)^2$, where the sum is over the set of all directed edges.

Proposition 1.5.2 ([LP14, Theorem 2.11])**.** *If a graph S admits a unit flow with finite energy, then S is transient.*

1.6 Lower bound on return probability

Consider the Schreier graph $S(\alpha)$ and the bubble group $\Gamma(\alpha)$, depending on a scaling sequence $\alpha = (\alpha_1, \alpha_2, \ldots)$, as described in Section 1.3. As mentioned in the introduction, we would like the graph $S(\alpha)$ to be transient and have "$2 + o(1)$"-dimensional volume growth, and also satisfy the Assumption 1 on exponential-like growth.

To analyze volume growth, consider n such that $s_{k-1} \le n < s_k$ (following the notation of Section 1.4). Because of the branching structure of $S(\alpha)$, the size of the ball $B_n(o)$ of radius n around o satisfies:

$$|B_n(o)| \le 2\left(\alpha_1 + 1 + 2(\alpha_2 + 1) + \ldots + 2^{k-1}(\alpha_k + 1)\right)$$

For a scaling sequence satisfying $\alpha_k = \alpha^{k+o(k)}$, with $\alpha > 1$, it is easy to see that the volume of the ball will satisfy:

$$|B_n(o)| \le n^{1+\frac{\log 2}{\log \alpha}+o(1)}$$

as $n \to \infty$. In particular if we take $\alpha_k = \frac{2^k}{f(k)}$ for some positive and sufficiently slowly increasing function $f(k)$, then:

$$|B_n(o)| \le n^{2+\varepsilon(n)} \tag{1.1}$$

for some nonnegative function $\varepsilon(n) \to 0$ as $n \to \infty$. How slowly $f(k)$ should grow will be determined by the transience requirement.

Consider the graph $S(\alpha)$ and the group $\Gamma(\alpha)$ defined by taking a scaling sequence α_k satisfying:

$$\sum_{k=1}^{\infty} \frac{\alpha_k}{2^k} < \infty$$

Proposition 1.6.1. *For α_k as above the graph $S(\alpha)$ is transient.*

Proof. We use the flow criterion from Proposition 1.5.2. Consider any cycle on the k-th level of the graph. If the edge e is on the upper half of the cycle and is labelled by a, or is on the lower half of the cycle and is labelled by a^{-1}, we take the value of $I(e)$ to be $1/2^k$. The two edges labelled by b and b^{-1} adjacent to the rightmost point of the cycle also get the value $1/2^k$ and all other edges have values 0. One readily checks that this function satisfies Kirchoff's law and its energy is given by:

$$\mathcal{E}(I) = \frac{1}{2}\sum_e I(e)^2 = \sum_{k=1}^{\infty} 2^{k-1}\alpha_1 \left(\frac{1}{2^k}\right)^2 = \frac{1}{2}\sum_{k=1}^{\infty} \frac{\alpha_k}{2^k}$$

which is finite by the assumption on the scaling sequence. □

An example of a scaling sequence satisfying this assumption is $\alpha_k = \lceil \frac{2^k}{k^2} \rceil$ and from now on we denote by S and Γ the graph and the group corresponding to this choice of α. One can easily check (by induction) that this scaling sequence satisfies Assumption 1 on exponential-like growth.

The graph S satisfies the volume growth condition $|B_n(o)| \leq n^{2+\varepsilon(n)}$ described above for $\varepsilon(n) \lesssim \frac{\log \log n}{\log n}$ (so that $|B_n(o)| \approx n^2 \log^\delta n$ for some $\delta > 0$). We will use the graph S and the group Γ to construct a group with the desired behavior of return probabilities.

Consider the permutational wreath product $G = \mathbb{Z}_2 \wr_S \Gamma$. Let Z_n be the simple random walk on Γ and denote by $\tilde{X}_n = (X_n, Z_n)$ the associated switch-walk-switch random walk on G (with the uniform distribution on the lamp group $\mathbb{Z}_2$).

Denote by $p_n(g,h)$ the probability that $\tilde{X}_n = h$ given $\tilde{X}_0 = g$, where $g, h \in G$. To bound the return probability $p_{2n}(e,e)$, for any finite set $A \subseteq G$ we can write, using the symmetry of the random walk and Cauchy-Schwarz inequality:

$$p_{2n}(e,e) = \sum_{g\in G} p_n(e,g)p_n(g,e) = \sum_{g \in G} p_n(e,g)^2 \geq \sum_{g\in A} p_n(e,g)^2 \geq \frac{p_n(A)^2}{|A|}$$

where $p_n(A) = \sum\limits_{g\in A} p_n(e,g)$ is the probability that $\tilde{X}_n$ is in the set A after n steps.

For the usual lamplighter $\mathbb{Z}_2 \wr \mathbb{Z}^d$ we would take A to be the set of all elements with lamp configurations contained in a ball of radius n^α (with α to be optimized later) and lower bound $p_n(A)$ by the probability of the simple random walk on $\mathbb{Z}^d$ to be actually confined to a ball of radius n^α. Since the base group has polynomial growth, the main contribution to $|A|$ comes from the number of lamp configurations, which is of the order of $e^{n^{d\alpha}}$ (as balls in $\mathbb{Z}^d$ have volume growth $\approx n^d$). The probability that the range of a simple random walk on $\mathbb{Z}^d$ is contained in a ball of radius n^α can be shown to be of the order of $e^{-n^{1-2\alpha}}$. We want these two terms to be of the same order - optimizing for α gives that one should consider balls of radius $n^{\frac{1}{d+2}}$, which gives the correct return probability exponent of $\frac{d}{d+2}$.

We use the same approach for the permutational wreath product $\mathbb{Z}_2 \wr_S \Gamma$, the difference being that we are dealing with inverted orbits instead of ordinary random walks and we have to be more careful with estimating the possible positions of the random walker on the base group.

Let $B_{Km}(o)$ be a ball of radius Km around o in S (with K as in Theorem 1.4.3 and m to be optimized later). We will say that a word w has *small inverted orbits* if $\mathcal{O}(w) \subseteq B_{Km}(o)$. Consider the set C of group elements with the following property: each element of C can be represented by a word w of length n such that w has small inverted orbits and $d(x, x.w) \leq Km$ for every $x \in S$.

Following the same approach as for the ordinary lamplighter group, in the bound above we take $A = \{(f, \gamma) \in G \,|\, \mathrm{supp} f \subseteq B_{Km}(o),\, \gamma \in C\}$.

We have to provide a lower bound on $p_n(A)$ and an upper bound on the size of A.

Theorem 1.6.2. *$p_n(A) \gtrsim e^{-c\frac{n}{m^2}}$ for all $n \geq 1$.*

Proof. We have $p_n(A) = \mathbb{P}\left(\tilde{X}_n \in A\right) = \mathbb{P}\left(\mathcal{O}(Z_n) \subseteq B_{Km}(o), Z_n \in C\right)$. By Lemma 1.4.1, Theorem 1.4.3 and Corollary 1.4.4 with probability at least $e^{-c\frac{n}{m^2}}$ (up to a multiplicative constant) the random element Z_n simultaneously has small inverted orbits, so $\mathcal{O}(Z_n) \subseteq B_{Km}(o)$, and does not move any vertex further than Km from itself, which implies that $Z_n \in C$. □

Theorem 1.6.3. *$|A| \lesssim e^{cm^{2+\eta(m)}}$ for some sequence $\eta(m) \to 0$ as $m \to \infty$.*

Proof. The size of A is at most the number of all lamp configurations with support in $B_{Km}(o)$ times the size of C. The number of configurations can be bounded above by $2^{|B_{Km}|}$, which by the growth condition (1.1) is at most $e^{cm^{2+\varepsilon(m)}}$.

To bound the size of C, we use the property that words with small inverted orbits admit a concise description. Every element $\gamma \in \Gamma$ can be described by specifying for each vertex its image under the action of γ. Now suppose γ can be represented by a word w with the property that $d(x, x.w) \leq Km$ for every vertex x. Since every vertex $x \in S$ is mapped under the action of w into some other vertex from the ball $B_{Km}(x)$, for a fixed vertex x we have at most $|B_{Km}(x)|$ possible choices.

Now, for a fixed m we have only finitely many types of vertices for which we have to specify their images in order to describe γ (since the image of a vertex x under w depends only on the isomorphism type of the ball of radius at most Km around x). We distinguish three types of vertices: 1) vertices such that $B_{Km}(x)$ intersects only one level in S, 2) $B_{Km}(x)$ intersects two levels in S, 3) $B_{Km}(x)$ intersects at least three levels in S.

For vertices of the first kind, the ball $B_{Km}(x)$ does not intersect any branching cycle, which means that it looks like a ball in $\mathbb{Z}$ and all vertices of this kind are mapped by γ in the same way. Thus we have at most $2Km$ choices for vertices of this kind.

For vertices of the second kind, each of them must be in a ball of radius Km around a branching point which does not intersect any other branching cycle. Such a ball can have at most $6Km$ vertices and each of them is mapped into a ball of radius at most $2Km$ around a branching point, which can have at most cm vertices (for some c). This give us at most $(cm)^{6Km}$ possibilities.

For vertices of the third kind, we observe that if $B_{Km}(x)$ intersects at least three levels and x belongs to the k-th level, then at least one of α_{k-1}, α_k, α_{k+1} is smaller than $2Km$. From this and Assumption 1 it follows that $d(o, x) \leq cm$ for some $c > 0$ (like in the proof of Theorem 1.4.3). Thus we have at most $|B_{cm}(o)|$ vertices of this kind. Since $B_{Km}(x) \subseteq B_{(c+K)m}(o)$, we have at most $|B_{(c+K)m}(o)|$ choices for each vertex. As $\varepsilon(m) \to$, this gives us at most $|B_{(c+K)m}(o)|^{|B_{cm}(o)|} \leq e^{cm^{2+o(1)}\log m}$ choices for vertices of this kind.

Thus there at most a constant times $2m \cdot (8m)^{cm} \cdot e^{cm^{2+o(1)}\log m}$ possible choices determining an element γ which can be represented by a word which has small inverted orbits. This gives us $|C| \leq e^{cm^{2+o(1)}\log m}$ and $|A| \leq e^{cm^{2+o(1)}} \cdot |C| \leq e^{cm^{2+o(1)}\log m}$, so the theorem holds with $\eta(m) \lesssim \frac{\log\log m}{\log m}$. □

Corollary 1.6.4. *The return probability for the random walk $\tilde{X}_n$ on $G = \mathbb{Z}_2 \wr_S \Gamma$ satisfies for all $n \geq 1$:*

$$p_{2n}(e,e) \gtrsim e^{-cn^{1/2+o(1)}}$$

Proof. By combining Theorem 1.6.2 and Theorem 1.6.3 we obtain the bound:

$$p_{2n}(e,e) \gtrsim e^{-cm^{2+\eta(m)}} e^{-c\frac{n}{m^2}}$$

To make this bound optimal we want both terms on the right hand side to be of the same order, which corresponds to taking m such that $\frac{n}{m^2} = m^{2+\eta(m)}$. This means that $m = n^{1/4-\varepsilon'(n)}$ for some $\varepsilon'(n) \geq 0$, $\varepsilon'(n) \to 0$. Inserting this back into the lower bound gives us:

$$p_{2n}(e,e) \gtrsim e^{-cn^{1/2+f(n)}}$$

with $f(n) \lesssim \frac{\log \log n}{\log n} = o(1)$ as $n \to \infty$. $\square$

Remark 1.6.1. One can do a similar calculation for a more general scaling sequence satisfying $\alpha_n = \alpha^{n+o(n)}$, with $\alpha > 1$, which then gives:

$$|A| \lesssim e^{cm^{d+o(1)}}$$

and

$$p_{2n}(e,e) \gtrsim e^{-cn^{\frac{d}{d+2}}}$$

with $d = 1 + \frac{\log 2}{\log \alpha} + o(1)$ as $m \to \infty$.

We can now prove the main theorem:

Proof of Theorem 1.1.1. Take $G = \mathbb{Z}_2 \wr_S \Gamma$ for S and Γ as above. By Corollary 1.6.4 the return probability for the switch-walk-switch random walk μ on G, induced from the simple random walk on Γ and a uniform distribution on $\mathbb{Z}_2$, satisfies:

$$p_{2n}(e,e) \gtrsim e^{-cn^{1/2+o(1)}}$$

which gives the return probability exponent $\overline{\gamma} \leq 1/2$. The induced random walk on S is the simple random walk, which by Proposition 1.6.1 is transient, so by Theorem 1.5.1 the group G supports nonconstant bounded harmonic functions. Thus G has both $\overline{\gamma} \leq 1/2$ and the non-Liouville property. $\square$

Acknowledgements

We would like to thank Gady Kozma for bringing our attention to groups permuting vertices of slowly growing trees. We would also like to thank anonymous referees for their valuable comments.

Chapter 2

Limits of random permuton processes and large deviations for the interchange process

2.1 Introduction

In this work we study large deviations for stochastic processes originating from certain models of random permutations.

As a first motivation, let us consider *sorting networks*. These combinatorial objects are perhaps simplest to describe in terms of paths in the symmetric group $\mathcal{S}_N$. A sorting network on N elements is a sequence of $M = \binom{N}{2}$ transpositions $(\tau_1, \tau_2, \ldots, \tau_M)$ such that each τ_i is a transposition of adjacent elements and $\tau_1 \ldots \tau_M = \rho$, where $\rho = (N\, N-1\, \ldots\, 2\, 1)$ is the reverse permutation. It is easy to see that any sequence of adjacent transpositions giving the reverse permutation must have length at least $\binom{N}{2}$, hence sorting networks can be thought of as shortest paths joining the identity permutation and the reverse permutation in the Cayley graph of S_N generated by adjacent transpositions.

A rich probabilistic structure is revealed if we consider the set of all sorting networks of size N and choose one uniformly at random. If we write $\sigma_t = \tau_1 \ldots \tau_t$, such a random sorting network can be thought of as a stochastic process $(\sigma_t : t = 1, \ldots, M)$ with values in S_N. The study of random sorting networks was initiated in [AHRV07] (see also [AH10], [AGH12]), where remarkable conjectures regarding asymptotic behavior of such processes were made. Computer simulations (see `https://www.math.ubc.ca/~holroyd/sort/` for a beautiful gallery of pictures) strongly suggest that for large N they exhibit the following behavior. If we pick $i = 1, \ldots, N$ at random and then follow the trajectory $\sigma_t(i)$ of i for $t = 1, \ldots, M$ in a random sorting network, the particle seems to move along a sine curve with a randomly chosen amplitude and initial phase (a distribution called the *sine curve process*). Even more, trajectories of all particles simulatenously behave like sine curves, moving at the same speed and with each particle having its initial amplitude and phase chosen independently. If we fix time t and look at the distribution of the permutation σ_t at time t, another remarkable picture appears - the distribution of 0s and 1 in the permutation matrix of the halfway permutation $\sigma_{M/2}$ is close, after rescaling, to the *Archimedean law*, which is the measure obtained by projecting the surface area of the 2-dimensional half-sphere onto the unit square. More generally, at each time tM the permutation matrix will look like at the distribution of the *Archimedean path* at time t, given by the density concentrated on an ellipse:

$$\mathcal{A}_t dx\, dy = \frac{1}{2\pi} \frac{1}{\sqrt{\sin^2(\pi t) + 2xy\cos(\pi t) - x^2 - y^2}} dx\, dy$$

It is natural to ask whether, as $N \to \infty$ and after proper rescaling, the permutation seen at fixed time t or the trajectory of a random particle converge in distribution to some sort of limiting process. Such limits have been recently studied under the name of *permutons* and *permuton processes* (for example in [GGKK15], [HKM$^+$13], [RVb], [RVa]). The theory of asymptotic limits of permutations has been inspired by the theory of graph limits ([Lov12]), where the analogous notion of a *graphon* as a limit of a sequence of dense graphs appears.

In this work we study limits of analogous random processes related to the *interchange process*. The interchange process is a process which can be roughly described as follows. We have N particles on a line $\{1, \dots, N\}$, labelled from 1 to N, and at each time step an edge is chosen at random and the adjacent particles are swapped. By looking at the configuration of labels obtained after a certain time we obtain a random permutation. As in the random sorting network, we can also pick a particle at random and follow its trajectory as the permutation changes.

There are many questions regarding the structure of permutations obtained in this way. If we pick a random particle and run the process for a long time, its trajectory will look like a symmetric random walk and hence the limit of such a process will be a Brownian motion ([RVa]). For very short times most particle will have displacement asymptotically much smaller than N, so we expect all the permutations to be close to the identity permutation (in an appropriate limiting sense). Nevertheless, one can still ask about the probability of the unlikely event that the resulting permutation is close to a fixed non-trivial permuton or, more generally, that the trajectory of a random particle behaves as in a non-trivial permutation-valued process. For example, we can ask questions like "what is the probability of getting from the identity to the reverse permutation in time t?" or "what is the probability that a randomly chosen particle moves like the trajectory of the sine curve process?".

A natural setting for such problems is the *large deviations theory*. Roughly speaking, after appropriate space and time scaling the probabilities of rare events will decay exponentially quickly in some power of N. The rate at which this happens is determined by what is called a *rate function* and it will turn out to be closely related to similar analysis appearing in the study of random walks.

Here we state loosely (in terms of permutations rather than general permuton processes) one of our results about the asymptotic rate of decay for the probability that the sequence of permutations obtained in the interchange process belongs to a given set (a precise formulation is given in Section 2.7):

Theorem 2.1.1. *If η^N is a sequence of $N^{2+\varepsilon}$ random permutations (with $\varepsilon \in (0,1)$) obtained by multiplying random adjacent transpositions, then the probability that η^N belongs to a given set $\mathcal{K}$ of sequences of permutations satisfies asymptotically:*

$$\frac{1}{N^{2-\varepsilon}} \log \mathbb{P}(\eta^N \in \mathcal{K}) \lesssim - \inf_{\pi \in \mathcal{K}} I(\pi)$$

where the rate function $I(\pi)$ is approximately the energy of the path of a randomly chosen particle in π.

To illustrate applications of this framework, let us come back to sorting networks. A remarkable formula due to Stanley ([Sta84]) says that the number of all sorting networks on N elements is equal to:

$$\frac{\binom{N}{2}!}{1^{N-1}3^{N-2}\dots(2N-3)^1}$$

which is asymptotic to $\exp\left\{\frac{N^2-N}{2}\log N + (\frac{1}{4}-\log 2)N^2 + O(N)\right\}$. This can be interpreted as a formula for the number of shortest paths (of length $\binom{N}{2}$) from the identity to the reverse permutation in $\mathcal{S}_N$ with adjacent transpositions as generators. Using our large deviation bounds we can obtain a similar asymptotic formula for the number of slightly longer paths joining these two permutations:

Theorem 2.1.2. *The number $\mathcal{P}_N^\varepsilon$ of paths of length $\frac{1}{2}N^{2+\varepsilon}$ (with $\varepsilon > 0$) joining the identity and the reverse permutation in the Cayley graph of $\mathcal{S}_n$ with adjacent transpositions as generators is given asymptotically by:*

$$\mathcal{P}_N^\varepsilon \sim \exp\left\{\frac{1}{2}N^{2+\varepsilon}\log N - \frac{\pi^2}{6}N^{2-\varepsilon} + o(N^{2-\varepsilon})\right\}$$

Furthermore, if we choose one of such paths uniformly at random, with high probability it will be close (in permuton topology) to the Archimedean path.

The main techniques used here come from the field of interacting particle systems. A comprehensive introduction to the subject can be found in [KL99]. The novelty in our approach is in applying tools usually used to study hydrodynamic limits to a setting which is in some respects more involved, since the limiting objects we consider, permuton processes, are stochastic processes instead of deterministic objects like solutions of PDEs apearing, for example, for exclusion processes.

Here we describe briefly the main ideas of the proofs. To lower bound the probability of a permutation sequence behaving, for example, like the sine curve process, we consider a perturbed process in which particles have additional parameters (corresponding to random initial speeds) and asymmetric jump rates (defined by means of an ODE which trajectories of the sine curve process have to satisfy). For the perturbed process we are able to prove a law of large numbers, stating that the distribution of the path of a random particle converges to a deterministic limit, which is the distribution of the the sine curve process.

The main idea here is that it is possible to make the perturbed process exactly stationary, which simplifies the analysis, since one does not have to estimate relative entropy with respect to the equilibrium distribution. To prove convergence of path distributions to a deterministic limit, one has to prove that in the interchange process particles behave approximately like independent random walks (with independently chosen random speeds). This requires showing that their speeds are on average uncorrelated and is accomplished by means of a local ergodicity result called the one-block estimate.

For the large deviation upper bound we use a family of exponential martingales and again use local mixing to show that the large deviation behavior of the system is similar to a collection of independent random walks. The result on permutations joining the identity with the reverse permutation then follows easily by using a recent result from [RVb], which enables one to explicitly compute the value of the rate function (the average energy of a random path) for the relevant set of permutations.

The paper is organized as follows. In Section 2.2 we introduce the notion of limit for permutations and permutation-valued stochastic processes. In Section 2.3 we introduce the interchange process and its perturbed version. In Section 2.4 we prove the law of large numbers for the perturbed process, postponing the proof of the necessary one-block estimate to Section 2.5. In Section 2.6 we prove the large deviation lower bound. In Section 2.7 we prove the matching upper bound and Theorem 2.1.2.

2.2 Permutons and stochastic processes

We start by introducing the concept of a permuton and a random permuton process.

Consider the space $\mathcal{M}([0,1]^2)$ of all probability measures on the unit square $[0,1]^2$, endowed with the weak topology. A *permuton* is a probability measure $\mu \in \mathcal{M}([0,1]^2)$ with uniform marginals. In other words μ is the joint distribution of a pair of random variables (X,Y), with X, Y taking values in $[0,1]$ and having marginal distribution $X, Y \sim \mathcal{U}[0,1]$. A few simple examples of permutons are the *identity permuton* (X,X), the *uniform permuton* (the distribution of two independent copies of X, which is the uniform measure on the square) or the *reverse permuton* $(X, 1-X)$.

Permutons can be thought of as continuous limits of permutations in the following sense. Given a permutation σ on N elements, we associate to it the empirical measure:

$$\mu_\sigma = \frac{1}{N}\sum_{i=1}^{N} \delta_{\left(\frac{i}{N}, \frac{\sigma(i)}{N}\right)}$$

which is an element of $\mathcal{M}([0,1]^2)$. Since every such measure has uniform marginals on $\left\{\frac{1}{N}, \frac{2}{N}, \ldots, 1\right\}$, it is not difficult to see that if a sequence of empirical measures converges weakly, the limiting measure will be a permuton. Conversely, every permuton can be realized as a limit of finite permutations, in the sense of weak convergence of empirical measures (see [HKM$^+$13]).

A *permuton process* is a stochastic process $X = (X_t, 0 \le t \le T)$ taking values in $[0,1]$, with continuous paths and such that for every time t the marginal distibution X_t is uniform on $[0,1]$. The name is justified by observing that for any fixed t we can define a permuton by looking at the joint distribution (X_0, X_t). More generally, for every permuton process there is an associated path $\mu = (\mu_t, 0 \le t \le T)$ with values in the permuton space, where μ_t is the distribution of (X_0, X_t). For a process X its finite dimensional distribution $(X_{t_1}, \ldots, X_{t_k})$ will be denoted by $X_{t_1,\ldots,t_k}$.

A *random permuton process* is a permuton process chosen from some probability distribution on the space of all permuton processes, i.e. a random variable $X : \Omega \to \mathcal{P}$, defined for a probability space Ω, such that $X(\omega)$ is a permuton process for $\omega \in \Omega$. By identifying the random variable with its distribution we can also think of a random permuton process as an element of $\mathcal{M}(\mathcal{P})$. In this setting, with weak topology on $\mathcal{M}(\mathcal{P})$, one can consider convergence in distribution of random permuton processes X_n to a (possibly also random) permuton process X. For clarity of notation we will denote random processes by bold letters like $\boldsymbol{X}$ and deterministic processes by capital letters like X.

Now consider a permutation-valued path $\eta^N = (\eta_t^N, 0 \le t \le T)$, with η_t^N taking values in the symmetric group $\mathcal{S}_N$. Let $\eta^N(i) = \left(\eta_t^N(i), 0 \le t \le T\right)$ be the trajectory of i. We define the *permutation process* $X^{\eta^N} = (X_t^{\eta^N}, 0 \le t \le T)$ in the following way: choose $i = 1, \ldots, N$ uniformly at random and follow the rescaled trajectory $\frac{1}{N}\eta^N(i)$. In this way X^{η^N} can be considered as a collection of N (not necessarily continuous) paths with marginals uniform on $\left\{\frac{1}{N}, \frac{2}{N}, \ldots, 1\right\}$ at each time. If η^N is random, then X^{η^N} itself will be a random permutation process.

Since every permutation process has marginals uniform on $\{1, \ldots, N\}$, we can call it an *approximate permuton process*. We will denote the space of all permuton processes and approximate permuton processes by $\mathcal{P}$. From now on we will usually omit the word "approximate".

One can prove (see [RVb]) that if a sequence of random permutation processes $\mathbf{X}^{\eta^N}$ converges in distribution, then the limit is a permuton process (possibly also random). Of particular interest will be sequences of random permutation-valued paths η^N (coming for example from the interchange process) such that the corresponding permutation processes $\boldsymbol{X}^{\eta^N}$ converge in distribution to a deterministic permuton process (for example the sine curve process described below).

For any random permuton process $\boldsymbol{X}$ with values in a compact space one can define its associated *random particle process* $\bar{X} = \mathbb{E}_\omega \mathbf{X}(\omega)$, which is a deterministic process obtained by first sampling a permuton process X from $\boldsymbol{X}$ and then sampling a path from X.

To elucidate the difference between random and deterministic permuton processes, consider a random permuton process $\mathbf{X}$ and its associated random particle process $\bar{X}$. If we sample an outcome X from $\boldsymbol{X}$ and then a path from X, then obviously the distribution of paths will be the same as for $\bar{X}$. However, consider now sampling an outcome X from $\boldsymbol{X}$ and then sampling independently two paths from X. The distribution of a pair of paths obtained in this way will not in general be the same as the distribution of two independent copies sampled from $\bar{X}$, since the paths might be correlated within the outcome X. The following general lemma will be useful later for showing that limits of certain random permutation processes are in fact deterministic ([RVa]):

Lemma 2.2.1. *Let K compact metric space and let $\boldsymbol{\mu}$ be a random probability measure on K. Let X and Y be two independent samples from an outcome of μ and let Z be a sample from an outcome of an independent copy of $\boldsymbol{\mu}$. If (X, Y), as a K^2-valued random variable, has the same distribution as (X, Z), then $\boldsymbol{\mu}$ is in fact deterministic, i.e. there exists μ such that $\boldsymbol{\mu} = \mu$ almost surely.*

Given a continuous path $\gamma : [0, T] \to [0, 1]$ its *Dirichlet energy* is defined by:

$$\mathcal{E}(\gamma) = \sup_{\Pi} \frac{1}{2} \sum_{i=1}^{k} \frac{|\gamma(t_i) - \gamma(t_{i-1})|^2}{t_i - t_{i-1}}$$

where the supremum is over all finite partitions $\Pi = \{0 = t_0 < t_1 < \ldots < t_k = T\}$. For a path which is not absolutely continuous the supremum is equal to ∞. If a path γ is differentiable, its energy is equal to:

$$\frac{1}{2} \int_0^T \dot{\gamma}(s)^2 \, ds$$

If Π is a partition we will write $\mathcal{E}^{\Pi}(\gamma)$ for the energy with respect to Π, i.e. the sum under the supremum in the definition of $\mathcal{E}$. Note that we can always assume that there is a sequence of nested partitions Π_n such that $\mathcal{E}^{\Pi_n}(\gamma)$ is increasing and $\mathcal{E}^{\Pi_n}(\gamma) \to \mathcal{E}(\gamma)$ monotonically.

For a permuton $\mu \sim (X, Y)$ its energy is defined by:

$$I(\mu) = \frac{1}{2}\mathbb{E}|X - Y|^2$$

If $\mu = \mu_\sigma$ for a permutation σ, then it is simply equal to:

$$I(\mu_\sigma) = \frac{1}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sigma(i) - i}{N}\right)^2\right)$$

Note that I is a continuous function of a permuton in the weak topology.

We also define the energy of a permuton process π:

$$I(\pi) = \mathbb{E}_{\gamma\sim\pi}\mathcal{E}(\gamma)$$

where the expectation is over paths γ sampled from π. For a partition Π we will write $I^{\Pi}(\pi) = \mathbb{E}_{\gamma\sim\pi}\mathcal{E}^{\Pi}(\gamma)$. This function will turn out to correspond to the rate function in large deviation bounds for random permuton process. It can be checked that I is lower semicontinuous (in the weak topology on $\mathcal{P}$) and its level sets $\{\pi \in \mathcal{P} : I(\pi) \leq C\}$ are compact.

Now we describe the sine curve process mentioned in the introduction (see [AHRV07]). Here it will be more convenient to consider the square $[-1, 1]^2$ and processes with values in $[-1, 1]$. The *Archimedean law* is the measure on $[-1, 1]^2$ obtained by projecting the normalized surface area of a 2-dimensional half-sphere. One can also describe it by its density, given by $1/(2\pi\sqrt{1 - x^2 - y^2})dx\, dy$. Observe that thanks to the well-known plank property each strip $[a, b] \times [-1, 1]$ has measure proportional to $b - a$, hence the Archimedean law defines a permuton.

The *sine curve process* is the permuton process $\mathcal{A} = (\mathcal{A}_t, 0 \leq t \leq 1)$ with the following distribution: we choose (X, Y) from the Archimedean law and then follow the path $\mathcal{A}(t) = X\cos\pi t + Y\sin\pi t$. Observe that $(\mathcal{A}_0, \mathcal{A}_0) = (X, X)$ and $(\mathcal{A}_0, \mathcal{A}_1) = (X, -X)$, hence the sine curve process can be thought of as a path between the identity permuton and the reverse permuton (which explains the connection to random sorting networks mentioned in the introduction). The trajectories of this process are sine curves with random initial phase and amplitude. One can also describe the sine curve process as choosing a pair (R, ϕ) at random (where the angle ϕ is uniform on $[0, 2\pi]$ and R has density $R/2\pi\sqrt{1 - R^2})$ and following the path $A(t) = R\cos(\pi t + \phi)$. The intuition that a particle in the sine curve process chooses its initial speed $-R\sin\phi$ at random (from a mean zero distribution depending on its initial location), with the angle ϕ having uniform distribution, will be useful in the next section where we investigate perturbations of the interchange processes. The *Archimedean path* is the permuton-valued path whose value at time t is the distribution of $(\mathcal{A}_0, \mathcal{A}_t)$ (with the density given by the formula from the introduction).

2.3 Interchange process and stationarity

The main object of our study will be the *interchange process* on the interval $\{1, \ldots, N\}$. This is a Markov process in continuous time defined in the following way. Consider particles labelled from 1 to N on a line with N vertices. Each edge has an independent exponential clock that fires at rate 1. Whenever a clock fires, the particles at the endpoints of the corresponding edge swap places. By comparing the initial position of each particle with its position after time t we obtain a random permutation of $\{1,\dots,N\}$.

Since we will be interested in taking the limit $N\to\infty$, we rescale the time by N^α, with $\alpha\in(1,2)$. Formally we define the state space as consisting of permutations $\eta=(x_1,\dots,x_N)$, where x_i is the position of particle with label i, and the dynamics is given by the generator:

$$(\mathcal{L}f)(\eta)=\frac{1}{2}N^\alpha\sum_{x=1}^{N-1}\left(f(\eta^{x,x+1})-f(\eta)\right)$$

where $\eta^{x,x+1}$ is the configuration η with particles at locations x and $x+1$ swapped.

From now on the time scale $\alpha\in(1,2)$ will be fixed. Note that if we pick a particle uniformly at random and follow its trajectory in the interchange process, its position will be distributed as the stationary simple random walk (in continuous time) on the line $\{1,\dots,N\}$. If we look at time scales much shorter than N^2, typically each particle will have distance $o(N)$ from its origin, so the permutation obtained at time t such that $tN^\alpha\ll N^2$ will be close (in the sense of permutons) to the identity permutation. As mentioned in the introduction, we will be interested in large deviations for rare events such as seeing a nontrivial permutation after a short time.

For the sake of proving a large deviation lower bound, we will need to perturb the interchange process to obtain dynamics which typically exhibits behavior of a fixed permuton process. Consider the following *biased interchange process*: the configuration space consists of $\eta=((x_i,\phi_i),\,i=1,\dots,N)$, where $(x_1,\dots,x_N)$ is a permutation of $(1,\dots,N)$ and ϕ_i has N possible values, $1,\dots,N$ (taken mod N). Here x_i will be the position of the particle number i and ϕ_i will be its *color*. The configuration at time t will be denoted by η_t^N (or simply η_t), and likewise by $x_i(\eta_t^N)$ and $\phi_i(\eta_t^N)$ we denote the position and the color of the particle number i at time t. We will use notation $X_i(\eta_t^N)=\frac{1}{N}x_i(\eta_t^N)$, $\Phi_i(\eta_t^N)=\frac{1}{N}\phi_i(\eta_t^N)$ for the rescaled positions and colors. Also, let $\eta^{-1}(y)$ be the label (number) of the particle at position y (so that $\eta^{-1}(x_i)=i$). For a position x we will often write ϕ_x as a shorthand for $\phi_{\eta^{-1}(x)}$ and likewise for Φ_x (the positions will be always denoted by x or y and labels by i, so there is no risk of ambiguity). If we are not interested in colors, we will also often refer to η_t^N itself as a permutation.

Let $\varepsilon=N^{1-\alpha}$. The dynamics is given by the generator $\widetilde{\mathcal{L}}$:

$$(\widetilde{\mathcal{L}}f)(\eta)=\frac{1}{2}N^\alpha\sum_{y=1}^{N-1}\left(1+\varepsilon\left[s(y,\phi_y(\eta))-s(y+1,\phi_{y+1}(\eta))\right]\right)\left(f(\eta^{y,y+1})-f(\eta)\right)+$$

$$+\frac{1}{2}N^\alpha\sum_{x=1}^{N}\left[\left(1+\varepsilon r(x,\phi_x(\eta))\right)\left(f(\eta^{x,+})-f(\eta)\right)+\left(1-\varepsilon r(x,\phi_x(\eta))\right)\left(f(\eta^{x,-})-f(\eta)\right)\right]$$

for some functions $s(x,\phi)$, $r(x,\phi)$. Here $\eta^{y,y+1}$ is the configuration η with particles at locations y and $y+1$ exchanged, and $\eta^{x,\pm}$ is the configuration η with ϕ_x changed by ±1.

In other words, at each time neighboring particles make a swap at rate close to 1, with bias proportional to the difference of their speeds $s(x,\phi_x)$, and each particle independently changes its color by ±1, also at rate close to 1 with bias proportional to $\pm r(x,\phi_x)$. The parameter ε has been chosen so that we expect particles to have displacement of order N at macroscopic times.

To make the connection with deterministic permuton processes and the sine curve process, consider first the following general setup. Suppose we have a system of ODEs:

$$\begin{cases} \frac{dX}{dt}(t) = S(X(t), \Phi(t)) \\ \frac{d\Phi}{dt}(t) = R(X(t), \Phi(t)) \end{cases} \tag{2.1}$$

for $t \in [0, T]$, with $(X(t), \Phi(t)) \in [0,1] \times [0,1]$ (Φ is taken mod 1) and a boundary condition $S(0, \Phi) = S(1, \Phi) = 0$.

Let $P = ((X_t, \Phi_t), 0 \leq t \leq T)$ be the stochastic process with values in $[0,1] \times [0,1]$ with the following distribution: choose (X_0, Φ_0) uniformly at random from $[0,1] \times [0,1]$ and then follow the the solution $(X(t), \Phi(t))$ of the ODE above with initial conditions (X_0, Φ_0). Assume that S and R are such that:

$$S(X, \Phi) = -\frac{\partial F}{\partial \Phi}(X, \Phi)$$
$$R(X, \Phi) = \frac{\partial F}{\partial X}(X, \Phi)$$

for some function $F(X, \Phi)$. It is easy to see that the process P will be stationary. i.e. for each time t the marginal distribution of $P_t = (X_t, \Phi_t)$ is uniform on $[0,1] \times [0,1]$, since the corresponding vector field (S, R) is divergence-free. In particular $X = (X_t, 0 \leq t \leq T)$ will be a permuton process.

We expect that a trajectory of a random particle in the biased interchange process will approximately follow the solution to (2.1) if the rates s and r are chosen to be $s(x, \phi) = S\left(\frac{x}{N}, \frac{\phi}{N}\right)$, $r(x, \phi) = R\left(\frac{x}{N}, \frac{\phi}{N}\right)$. However, we also need to require that the uniform distribution on positions and colors is be stationary for the dynamics.

More precisely, consider the uniform distribution on configurations of the biased interchange process, i.e. a distribution in which the labelling of particles is a uniformly random permutation and each particle has a uniformly random color, chosen indepedently for each of them. We want to find a condition on rates $s(x, \phi)$ and $r(x, \phi)$ such that this measure will be stationary for the dynamics of $\widetilde{\mathcal{L}}$.

The stationarity condition for the uniform measure means that for each state the sums of outgoing and incoming jump rates have to be equal. We write down the stationarity condition as follows. For a given configuration η and each location x (with particle at x having color ϕ_x) there are the following possible outgoing jumps:

- the particle at x swaps with particle at $x - 1$, at rate $1 + \varepsilon \left[s(x-1, \phi_{x-1}) - s(x, \phi_x)\right]$
- the particle at x swaps with particle at $x + 1$, at rate $1 + \varepsilon \left[s(x, \phi_x) - s(x+1, \phi_{x+1})\right]$
- the particle changes its color from ϕ to $\phi + 1$, at rate $1 + \varepsilon r(x, \phi_x)$
- the particle changes its color from ϕ to $\phi - 1$, at rate $1 - \varepsilon r(x, \phi_x)$

and incoming jumps:

- the particle at x swaps with particle at $x - 1$, at rate $1 + \varepsilon \left[s(x-1, \phi_x) - s(x, \phi_{x-1})\right]$
- the particle at x swaps with particle at $x + 1$, at rate $1 + \varepsilon \left[s(x, \phi_{x+1}) - s(x+1, \phi_x)\right]$

- the particle changes its color from $\phi+1$ to ϕ, at rate $1-\varepsilon r(x,\phi_x+1)$
- the particle changes its color from $\phi-1$ to ϕ, at rate $1+\varepsilon r(x,\phi_x-1)$

This gives us the following equation:

$$\begin{aligned}&\sum_{x=1}^{N-1}(s(x,\phi_x)-s(x+1,\phi_{x+1}))=\\&=\sum_{x=1}^{N-1}(s(x,\phi_{x+1})-s(x+1,\phi_x))+\sum_{x=1}^{N}(r(x,\phi_x-1)-r(x,\phi_x+1))\end{aligned}$$

so:

$$\begin{aligned}&\sum_{x=2}^{N-1}(s(x-1,\phi_x)-s(x+1,\phi_x)+[r(x,\phi_x-1)-r(x,\phi_x+1)])+\\&+s(N-1,\phi_N)+s(N,\phi_N)-s(1,\phi_1)-s(2,\phi_1)+\\&+[r(1,\phi_1-1)-r(1,\phi_1+1)]+[r(N,\phi_N-1)-r(N,\phi_N+1)]=0\end{aligned}$$

Since we want this equation to be satisfied for each configuration (for any choice of ϕ_x and for each x), we want each term in th sum and each of the boundary terms to vanish. This gives us a set of equations:

$$\begin{cases}s(1,\phi)+s(2,\phi)=r(1,\phi-1)-r(1,\phi+1)\\s(x+1,\phi)-s(x-1,\phi)=r(x,\phi-1)-r(x,\phi+1),\;\; x=2,\dots,N-1\\s(N-1,\phi)+s(N,\phi)=r(N,\phi+1)-r(N,\phi-1)\end{cases}$$

which have to be satisfed for every ϕ.

By analogy with the ODE case one can check that for any function $F(x,\phi)$ the rates given by:

$$\begin{cases}s(x,\phi)=\frac{1}{N}\left[F(x,\phi-1)-F(x,\phi+1)\right]\\r(x,\phi)=\frac{1}{N}\left[F(x+1,\phi)-F(x-1,\phi)\right]\end{cases}$$

solve the equations for stationarity, provided that $F(0,\phi)+F(1,\phi)=F(N,\phi)+F(N+1,\phi)=0$. For simplicity we can take $F(0,\phi)=F(1,\phi)=F(N,\phi)=F(N+1,\phi)=0$ for all ϕ.

Consider now the sine curve process $\mathcal{A}_t$. In this framework one can check that for the system of equations:

$$\begin{cases}\frac{dX}{dt}(t)=\frac{1}{\sqrt{\pi}}\sqrt{1-(2X-1)^2}\sin 2\pi\Phi\\\frac{d\Phi}{dt}(t)=-\frac{1}{\sqrt{\pi}}\frac{X}{\sqrt{1-(2X-1)^2}}\cos 2\pi\Phi\end{cases}$$

the position process X_t is exactly the sine curve process $\mathcal{A}_t$ (as the position satisfies the harmonic oscillator equation $\frac{d^2X}{dt^2}=-X$).

For technical reasons we will require the rates to be smooth functions of X and Φ, so we modify the sine curve equation as follows. Let $G_\varepsilon(X)$ be a smooth approximation of the function $\sqrt{1-(2X-1)^2}$ which satisfies the boundary conditions and is equal to $\sqrt{1-(2X-1)^2}$ on $[\varepsilon, 1-\varepsilon]$. Consider the system of ODEs:

$$\begin{cases} \frac{dX}{dt}(t) = \frac{1}{\sqrt{\pi}} G_\varepsilon(X(t)) \sin(2\pi\Phi(t)) \\ \frac{d\Phi}{dt}(t) = \frac{1}{\sqrt{\pi}} \left(\frac{d}{dX} G_\varepsilon\right)(X(t)) \cos(2\pi\Phi(t)) \end{cases} \tag{2.2}$$

This is an approximation of the equation for the sine curve and the corresponding trajectory process $P^\varepsilon = (X^\varepsilon, \Phi^\varepsilon)$ is stationary, since the stationarity condition is satisfied with $F(X, \Phi) = -\frac{1}{\sqrt{\pi}} G_\varepsilon(X) \cos(2\pi\Phi)$.

Consider the biased interchange process with rates given by:

$$s(x, \phi) = \frac{1}{N}\left[F(x, \phi - 1) - F(x, \phi + 1)\right]$$
$$r(x, \phi) = \frac{1}{N}\left[F(x + 1, \phi) - F(x - 1, \phi)\right]$$

for $F(X, \Phi)$ as above. To simplify notation we will sometimes use rescaled variables:

$$S(X, \Phi) = s(NX, N\Phi)$$
$$R(X, \Phi) = r(NX, N\Phi)$$

Note in particular that due to the smoothness assumption the rates are bounded (in case of s by ± 1).

By the discussion above, with these rates the uniform distribution is stationary. From now on by biased interchange process we will always mean the process started from the stationary distribution. The particular form of the rates will not be important apart from the fact that they are bounded and the easily verified (but crucial) property that for any x we have $\mathbb{E}\, s(x, \phi) = 0$ when ϕ is drawn from the uniform distribution.

If η^N is the trajectory of the biased interchange process, then by analogy with the permutation process X^{η^N} we can define the trajectory process $\boldsymbol{P}^{\eta^N} = (\boldsymbol{X}^{\eta^N}, \boldsymbol{\Phi}^{\eta^N})$ obtained by choosing a particle i at random and following the path $(X_i(\eta_t^N), \Phi_i(\eta_t^N))$ (so we keep track both of the position and the color of a random particle and the process itself is random, since η^N is random).

Since the interchange process is a pure jump Markov process, for each particle the trajectories $X_i(\eta^N)$ will be a càdlàg path from $[0, T]$ to $[0, 1]$. The space of such paths will be denoted by $\mathcal{D}([0, T], [0, 1])$ and endowed with the standard J_1 Skorokhod topology (see [KL99]). The space $\mathcal{P}$ of permuton processes is embedded in $\mathcal{D}([0, T], [0, 1])$ in a natural way. In the same way we can consider joint trajectories $(X_i(\eta^N), \Phi_i(\eta^N))$ as paths in $\mathcal{D}([0, T], [0, 1]^2)$. By $\mathcal{D}$ we will denote both spaces of stochastic processes on $\mathcal{D}([0, T], [0, 1])$ and $\mathcal{D}([0, T], [0, 1]^2)$ (there will be no ambiguity), endowed with the weak topology (metrizable by a complete metric d).

For the large deviation lower bound we will want to compare the unbiased interchange process with the biased one. Since they have different configuration spaces, for convenience we introduce the *unbiased interchange process with colors*, which has the same configuration space as the biased process and the generator obtained by putting all speeds s to 0:

$$(\mathcal{L}f)(\eta) = \frac{1}{2} N^\alpha \sum_{y=1}^{N-1} \left(f(\eta^{y,y+1}) - f(\eta)\right) + \frac{1}{2} N^\alpha \sum_{x=1}^{N} \left[1 \pm \varepsilon r(x, \phi_x(\eta))\right] \left(f(\eta^{x,\pm}) - f(\eta)\right)$$

Since here the colors do not influence the dynamics of swaps, the corresponding path process $\boldsymbol{X}^{\eta^N}$ will be the same as for the ordinary unbiased process (an we will never be interested in thte distribution of $\boldsymbol{\Phi}^{\eta^N}$ for this process).

2.4 Law of large numbers

Throughout this section $\widetilde{\mathbb{P}}^N$ will denote the probability distribution of the stationary biased interchange process on N particles and $\mathbb{P}^N$ will denote the disribution of the unbiased process with colors.

Let $P = P^\varepsilon$ be the trajectory process associated to the equation (2.2). We will prove the following theorem:

Theorem 2.4.1. *Let η^N be the trajectory of the stationary biased interchange process. The random process $\boldsymbol{P}^{\eta^N}$ converges in distribution to the deterministic process P as $N \to \infty$.*

The theorem above can be thought of as a law of large numbers for random permuton processes and it will be useful for establishing large deviation lower bound.

To prove Theorem 2.4.1, we will show that typically trajectories of most particles approximately follow the same ODE (2.2) as trajectories of the limiting process. In other words, we would like a particle at x to move on average according to its jump rate $s(x, \phi_x)$. However, because of swaps between particles it will be influenced by speeds of its neighbors. Nevertheless, since speed at each site has mean 0 in stationarity, we will be able to show that on average the contribution from speeds of the particle's neighbors cancels out - this will be the one block estimate proved in the next section.

Note that to prove that the random processes converge indeed to a deterministic process, it is not enough to look only at single path distributions, as explained in the previous section. Nevertheless, we will show that in the interchange process typically any two particles (in fact almost all of them) behave like independent random walks, which by Lemma 2.2.1 will be enough to establish a deterministic limit.

For now we we will simply write $\eta = \eta^N$. If $\mathcal{O}_N$ is an event concerning a single particle, we will say that it holds for *almost all particles* if it holds simultaneously for all particles except for $o(N)$ of them as $N \to \infty$.

We start by introducing some useful martingales (see [KL99] for a comprehensive treatment of the techniques used here). Recall that:

$$X_i(\eta_t) = \frac{1}{N} x_i(\eta_t)$$

is the rescaled position of the particle with label i. By the martingale formula for Markov processes we can write:

$$X_i(\eta_t) - X_i(\eta_0) = M_t^i + \int_0^t \widetilde{\mathcal{L}} X_i(\eta_s)\, ds$$

where M_t^i is a martingale with $M_0^i = 0$. We have:

$$\begin{aligned}
&\widetilde{\mathcal{L}}X_i(\eta_s) = \frac{1}{N}\widetilde{\mathcal{L}}(x_i(\eta_s)) = \frac{1}{2}N^{\alpha-1}\sum_{x=1}^{N-1}(1+\varepsilon\left[s(x,\phi_x(\eta_s)) - s(x+1,\phi_{x+1}(\eta_s))\right])\left(x_i(\eta_s^{x,x+1}) - x_i(\eta_s)\right) = \\
&= \frac{1}{2}N^{\alpha-1}\varepsilon\Big[-\left[s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) - s(x_i(\eta_s),\phi_i(\eta_s))\right] + \\
&+ \left[s(x_i(\eta_s),\phi_i(\eta_s)) - s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\right]\Big] = \\
&= \frac{1}{2}\left[2s(x_i(\eta_s),\phi_i(\eta_s)) - s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) - s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\right]
\end{aligned}$$

since the position of the particle i changes by ± 1 depending on whether it makes a swap with its left or right neighbor.

We get:

$$\begin{aligned}
&X_i(\eta_t) - X_i(\eta_0) = \\
&= M_t^i + \int_0^t s(x_i(\eta_s),\phi_i(\eta_s))\,ds + \frac{1}{2}\int_0^t \left[s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) + s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\right]\,ds
\end{aligned}$$

so:

$$\begin{aligned}
&X_i(\eta_t) - X_i(\eta_0) - \int_0^t s(x_i(\eta_s),\phi_i(\eta_s))\,ds = \\
&M_t^i + \frac{1}{2}\int_0^t \left[s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) + s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\right]\,ds
\end{aligned} \tag{2.3}$$

We want to show that this difference is small with high probability for a random particle:

$$\frac{1}{N}\sum_{i=1}^{N}\Big(X_i(\eta_t) - X_i(\eta_0) - \int_0^t s(x_i(\eta_s),\phi_i(\eta_s))\,ds\Big)^2 \to 0$$

It is enough to bound the expectation of the sum. For each particle i the martingale term M_t^i will be small with high probability. For:

$$Q_t^i = \widetilde{\mathcal{L}}X_i(\eta_s)^2 - 2X_i(\eta_t)\widetilde{\mathcal{L}}X_i(\eta_t)$$

we have that:

$$N_t^i = (M_t^i)^2 - \int_0^t Q_s^i\,ds \tag{2.4}$$

is a mean 0 martingale. A quick calculation gives:

$$\widetilde{\mathcal{L}}X_i(\eta_s)^2 = \frac{1}{2}\Big[\big(s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) - s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\big)\left(\frac{-2x_i(\eta_s)+1}{N}\right) +$$

$$+ \big(s(x_i(\eta_s),\phi_{x_i(\eta_s)}(\eta_s)) - s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\big)\left(\frac{2x_i(\eta_s)+1}{N}\right)\Big]$$

$$2F_i(\eta_s)\widetilde{\mathcal{L}}F_i(\eta_s) = \frac{x_i(\eta_s)}{N}\big(2s(x_i(\eta_s),\phi_i(\eta_s)) - s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) - s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\big)$$

so these two terms are the same up to order $\frac{1}{N}$. This gives us:

$$\mathbb{E}(M_t^i)^2 = O\left(\frac{1}{N}\right)$$

and by Doob's inequality these terms will be small with high probability for all t.

Exactly the same calculation (only simpler, since it does not involve correlations between adjacent particles) and the martingale argument gives us that for $\Phi_i(\eta_t) = \frac{1}{N}\phi_i(\eta_t)$:

$$\Phi_i(\eta_t) - \Phi_i(\eta_0) - \int_0^t r(x_i(\eta_s),\phi_i(\eta_s))\,ds = o(1)$$

for every particle. So we only need to show that the following term is small for a random particle:

$$Y_i^t = \int_0^t \left[s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s)) + s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\right]\,ds$$

This will give us:

Proposition 2.4.2. *For any fixed $t>0$ and ε we have for almost all particles i:*

$$\widetilde{\mathbb{P}}^N\left(\left|X_i(\eta_t) - X_i(\eta_0) - \int_0^t s(x_i(\eta_s),\phi_i(\eta_s))\,ds\right| > \varepsilon\right) \to 0$$

$$\widetilde{\mathbb{P}}^N\left(\left|\Phi_i(\eta_t) - \Phi_i(\eta_0) - \int_0^t r(x_i(\eta_s),\phi_i(\eta_s))\,ds\right| > \varepsilon\right) \to 0$$

as $N\to\infty$.

To prove the proposition we need to show that:

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left(Y_i^t\right)^2 \to 0$$

as $N \to \infty$. We have:

$$\left(Y_i^t\right)^2 = \left(\int_0^t \left[s(x_i(\eta_s) - 1, \phi_{x_i(\eta_s)-1}(\eta_s)) + s(x_i(\eta_s) + 1, \phi_{x_i(\eta_s)+1}(\eta_s)) \right] \, ds \right)^2$$

We will have four cross-terms here, it is enough to show that each of them is small in expectation. The argument is similar in all cases, so let us focus on:

$$\mathbb{E} \left(\int_0^t s(x_i(\eta_s) - 1, \phi_{x_i(\eta_s)-1}(\eta_s)) \, ds \right) \left(\int_0^t s(x_i(\eta_s) - 1, \phi_{x_i(\eta_s)-1}(\eta_s)) \, ds \right) =$$

$$= \mathbb{E} \int_0^t \int_0^t s(x_i(\eta_{t_1}) - 1, \phi_{x_i(\eta_{t_1})-1}(\eta_{t_1})) s(x_i(\eta_{t_2}) - 1, \phi_{x_i(\eta_{t_2})-1}(\eta_{t_2})) \, dt_1 \, dt_2$$

For each particle we are looking at the correlation of the speed of its left neighbor at time t_1 with the speed of its left neighbor at time t_2. By averaging over particles $i = 1, \ldots, N$ and using the symmetry between t_1 and t_2 we can write the contribution to the second moment of Y_i^t as:

$$\frac{2}{N} \sum_{i=1}^{N} \mathbb{E} \int_0^t \int_{t_1}^t s(x_i(\eta_{t_1}) - 1, \phi_{x_i(\eta_{t_1})-1}(\eta_{t_1})) s(x_i(\eta_{t_2}) - 1, \phi_{x_i(\eta_{t_2})-1}(\eta_{t_2})) \, dt_2 \, dt_1 =$$

$$= 2 \int_0^t dt_1 \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \int_{t_1}^t s(x_i(\eta_{t_1}) - 1, \phi_{x_i(\eta_{t_1})-1}(\eta_{t_1})) s(x_i(\eta_{t_2}) - 1, \phi_{x_i(\eta_{t_2})-1}(\eta_{t_2})) \, dt_2 \right)$$

It is enough to show that for each fixed $t_1 > 0$ the expression inside the bracket is close to 0, so let us look at:

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \int_{t_1}^t s(x_i(\eta_{t_1}) - 1, \phi_{x_i(\eta_{t_1})-1}(\eta_{t_1})) s(x_i(\eta_{t_2}) - 1, \phi_{x_i(\eta_{t_2})-1}(\eta_{t_2})) \, dt_2$$

Since the average here depends only on the configuration at time t_1 and its evolution from that point on (and not otherwise on the trajectory of the process before time t_1), by stationarity it will be the same as:

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \int_0^{t-t_1} s(x_i(\eta_0) - 1, \phi_{x_i(\eta_0)-1}(\eta_0)) s(x_i(\eta_s) - 1, \phi_{x_i(\eta_s)-1}(\eta_s)) \, ds \tag{2.5}$$

since the dynamics of the process is also time-homogeneous.

To prove that for a random particle the initial speed of its left neighbor is uncorrelated (when averaged over time) with the current speed of its left neighbor, we introduce the following setup. We can rewrite the average above in terms of a sum over sites (for $x = x_i(\eta_s)$) instead over particles:

$$\frac{1}{N} \sum_{x=1}^{N} \mathbb{E} \int_0^{t-t_1} s(x_{\eta_s^{-1}(x)}(\eta_0) - 1, \phi_{\eta_0^{-1}(x_{\eta_s^{-1}(x)}(\eta_0)-1)}(\eta_0)) s(x - 1, \phi_{x-1}(\eta_s)) \, ds$$

Consider the extended configuration in which each particle, in addition to its color ϕ_i, also has an additional color L_i in which remember the speed of its left neighbor at time 0, that is:

$$L_i = s(x_i(\eta_0) - 1, \phi_{x_i(\eta_0)-1}(\eta_0))$$

The dynamics stays the same (i.e. colors and labels are exchanged by swaps of adjacent particles and ϕ has its own evolution). For a site x let $L_x(\eta)$ be the additional label at site x in configuration η. We can now treat η as a function which assigns to each site x a pair (L_x, ϕ_x).

In this setup we have:

$$\frac{1}{N}\sum_{x=1}^{N}\mathbb{E}\int_0^{t-t_1} L_x(\eta_s)s(x-1, \phi_{x-1}(\eta_s))\,ds$$

Let $f_x(\eta) = L_x(\eta)s(x-1, \phi_{x-1}(\eta))$. Denote by $\Lambda_{x,l}$ a box of size l around x and let $\mu_{x,l}(\eta)$ be the empirical distribution of colors in $\Lambda_{x,l}$, i.e. a product measure over configurations restricted to $\Lambda_{x,l}$ such that the probability of color (L, ϕ) is proportional to the number of sites in $\Lambda_{x,l}$ with color (L, ϕ).

The *superexponential one-block estimate* says that for a local function f_x (i.e. depending only on a bounded neighborhood of x) in the time average above we can replace $f_x(\eta_s)$ by its average $\mathbb{E}_{\mu_{x,l}(\eta_s)}f$ with respect to the local empirical distribution over a sufficiently large box. In other words, due to local mixing the distribution of colors in a microscopic box becomes almost exchangeable.

Lemma 2.4.3. *Let $V_{x,l}(\eta) = f_x(\eta) - \mathbb{E}_{\mu_{x,l}(\eta)}f$. For any $t \in (0,T]$ and $\delta > 0$ we have:*

$$\limsup_{l\to\infty}\limsup_{N\to\infty} N^{-\gamma}\log\widetilde{\mathbb{P}}^N\left(\left|\int_0^t \frac{1}{N}\sum_{x=1}^{N} V_{x,l}(\eta_s)\,ds\right| > \delta\right) = -\infty$$

where $\gamma = 3 - \alpha$.

The lemma is proved in the next section. The superexponential decay of probability will be important for the large deviation upper bound, where γ will turn out to be the large deviation exponent. For the purpose of the law of large numbers it would be enough to have just probability going to 0.

Let us see how it enables us to finish the proof of Proposition 2.4.2. By the one block estimate we can replace:

$$\frac{1}{N}\sum_{x=1}^{N}\int_0^{t-t_1} L_x(\eta_s)s(x-1, \phi_{x-1}(\eta_s))\,ds$$

by:

$$\frac{1}{N}\sum_{x=1}^{N}\int_0^{t-t_1} \mathbb{E}_{\mu_{x,l}(\eta_s)}\left[L_x(\eta_s)s(x-1, \phi_{x-1}(\eta_s))\right]\,ds$$

with high probability as N and then $l \to \infty$. Since the measure $\mu_{x,l}(\eta_s)$ is product and L_x, ϕ_{x-1} depend on different sites, the expectation above is just a product and we have:

$$\frac{1}{N}\sum_{x=1}^{N}\int_{0}^{t-t_1}\left(\mathbb{E}_{\sigma\sim\mu_{x,l}(\eta_s)}L_x(\sigma)\right)\left(\mathbb{E}_{\sigma\sim\mu_{x,l}(\eta_s)}s(x-1,\phi_{x-1}(\sigma))\right)\,ds$$

Since the distribution of η_s is stationary, at fixed time s the distribution of the average $\mathbb{E}_{\sigma\sim\mu_{x,l}(\eta_s)}s(x-1,\phi_{x-1}(\sigma))$ does not depend on s. So we only need to show that $\mathbb{E}_{\sigma\sim\mu_{x,l}(\eta_0)}s(x-1,\phi_{x-1}(\sigma))$ is small, since L_x is bounded. Since in stationarity ϕ_x has uniform distribution, the average with respect to $\mu_{x,l}(\eta_0)$ is simply:

$$\frac{1}{2l+1}\sum_{k=1}^{2l+1}s(x-1,\phi_k)$$

where ϕ_k are independent and uniformly distributed. As $s(x,\phi_k)$ has mean 0 and is bounded by 1, by any concentration inequality for i.i.d. variables we get that this average goes to 0 as $l\to\infty$.

This finishes the proof of Proposition 2.4.2.

We can now prove the law of large numbers.

Proof of Theorem 2.4.1. We will first show that the random particle process $\bar{P}^N=(\bar{X}^N,\bar{\Phi}^N)$ defined by $\bar{P}^N=\mathbb{E}_{\eta^N}\boldsymbol{P}^{\eta^N}$ converges in distribution to P.

First we will show that the estimate from Proposition 2.4.2 holds not only at each time $t>0$, but also for supremum over all times $t\le T$. Consider the deterministic process (A^N,B^N) given by:

$$A_t^N=X_i(\eta_t)-X_i(\eta_0)-\int_0^t s(x_i(\eta_s),\phi_i(\eta_s))\,ds$$

$$B_t^N=\Phi_i(\eta_t)-\Phi_i(\eta_0)-\int_0^t r(x_i(\eta_s),\phi_i(\eta_s))\,ds$$

where i is a random particle in a random configuration $\eta=\eta^N$. Proposition 2.4.2 implies that all finite dimensional distributions of (A^N,B^N) converge to 0. To obtain convergence to 0 for the whole process we only need to check tightness. We will use the following condition stopping time criterion ([KL99, Chapter 1]). Let X^N be a family of stochastic processes on $D([0,T],[0,1]^2)$ whose one dimensional marginals at each time are tight. If for every $\varepsilon>0$:

$$\lim_{\gamma\to 0}\limsup_{N\to\infty}\sup_{\substack{\tau\\ \theta\le\gamma}}\mathbb{P}\left(|X^N_{\tau+\theta}-X^N_\tau|>\varepsilon\right)=0$$

where the supremum is over all stopping times τ bounded by T, then the family X^N is tight. We have from formula 2.3:

$$A^\eta_{\tau+\theta}-A^\eta_\tau=M^i_{\tau+\theta}-M^i_\tau-\frac{1}{2}\int_\tau^{\tau+\theta}\left[s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s))+s(x_i(\eta_s)+1,\phi_{x_i(\eta_s)+1}(\eta_s))\right]\,ds$$

Since s is bounded the integral is bounded by $C\theta$ for some constant $C > 0$, regardless of τ, so goes to 0 as θ (deterministically and for every i). So it only remains to bound the martingale term. By formula (2.4) we have (as τ is a stopping time):

$$\mathbb{E}\left|M^i_{\tau+\theta} - M^i_\tau\right|^2 = \mathbb{E}\left(\int\limits_\tau^{\tau+\theta} Q_s \, ds\right)$$

As in the calculation of $\mathbb{E}(M^i_t)^2$ we have that the right hand side is $O\left(\frac{1}{N}\right)$ for fixed θ. In particular with high probability we have $\left|M^i_{\tau+\theta} - M^i_\tau\right| \to 0$. The calculation for B^η is analogous. This shows that the family (A^N, B^N) is tight, so it converges to 0 in supremum norm. This proves that for any ε for a random particle i:

$$\widetilde{\mathbb{P}}^N\left(\sup_{0\le t\le T}\left|X_i(\eta_t) - X_i(\eta_0) - \int\limits_0^t s(x_i(\eta_s), \phi_i(\eta_s))\, ds\right| > \varepsilon\right) \to 0$$
$$\widetilde{\mathbb{P}}^N\left(\sup_{0\le t\le T}\left|\Phi_i(\eta_t) - \Phi_i(\eta_0) - \int\limits_0^t r(x_i(\eta_s), \phi_i(\eta_s))\, ds\right| > \varepsilon\right) \to 0$$

as $N \to \infty$.

Now we can prove that $\bar{P}^N$ converges to P in distribution.

A natural way to couple these two processes is as follows: let $\bar{P}^N_t$ be a path sampled from $\bar{P}^N$ starting at $(\bar{X}^N_0, \bar{\Phi}^N_0)$. Let $P(t) = (X(t), \Phi(t))$ be the solution of the ODE (2.2) started from an initial condition $(X(0), \Phi(0))$ chosen uniformly at random from $\left[\bar{X}^N_0 - \frac{1}{N}, \bar{X}^N_0\right] \times \left[\bar{\Phi}^N_0 - \frac{1}{N}, \bar{\Phi}^N_0\right]$ (so the two processes start close to each other). Because the initial condition is uniformly distributed, the path $P(t)$ will be distributed according to P.

Since $P(t) = (X(t), \Phi(t))$ is the solution of (2.2), we have at each time $t \le T$:

$$X(t) - X(0) = \int\limits_0^t S(X(s), \Phi(s))\, ds$$
$$\Phi(t) - \Phi(0) = \int\limits_0^t R(X(s), \Phi(s))\, ds$$

By the bound above with high probability the path $P^N(t) = (X^N(t), \Phi^N(t))$ will satisfy for all times $t \le T$ (after replacing s by S and r by R):

$$\bar{X}^N(t) - \bar{X}^N(0) = \int\limits_0^t S(\bar{X}^N(s), \bar{\Phi}^N(s))\, ds + \varepsilon^1_t$$
$$\bar{\Phi}^N(t) - \bar{\Phi}^N(0) = \int\limits_0^t R(\bar{X}^N(s), \bar{\Phi}^N(s))\, ds + \varepsilon^2_t$$

for some ε^1_t, $\varepsilon^2_t \to 0$ as $N \to \infty$. So $(\bar{X}^N, \bar{\Phi}^N)$ approximately satisfies the same ODE as (X, Φ) and an application of Gronwall's inequality gives that for any ε with high probability we have:

$$\left\|P - \bar{P}^N\right\| \le C \max\{|\bar{X}^N(0) - X(0)| + \varepsilon, |\bar{\Phi}^N(0) - \Phi(0)| + \varepsilon\} e^{KT}$$

for some $C > 0$, where $K > 0$ depends only on the Lipschitz constants of S an R.

By definition of processes $\bar{P}^N$ and P the initial conditions $\bar{X}^N(0)$, $X(0)$ and $\bar{\Phi}^N(0)$, $\Phi(0)$ differ by at most $\frac{1}{N}$, which implies that with high probability $\bar{P}^N \to P$ in the supremum norm.

Now we can show that the random processes $\boldsymbol{P}^{\eta^N}$ converge to a deterministic process P. It is easy to check that since the tightness criterion used above holds with high probability with respect to the choice of η^N, the family $\boldsymbol{P}^{\eta^N}$ is itself tight, so we only need to check uniqueness of limits. Since the limit must have the associated random particle process distributed according to P, it is enough to show that the limit is deterministic.

Consider an outcome of $\boldsymbol{P}^{\eta^N}$ and sample independently two paths X, Y from it. This corresponds to choosing uniformly at random a pair of particles i, j and following their trajectories in η^N. By convergence $\bar{P}^N \to P$ each of X and Y has marginal distribution converging to the distribution of P. Moreover, observe that in η^N the initial colors for any two particles are chosen uniformly at random, in particular they are independent, so the joint distribution of (X, Y) converges to the distribution of two independent paths sampled from P (as a path in P is uniquely determined by its initial condition). Since we already have convergence to a limit, applying Lemma 2.2.1 gives that the limit has to be deterministic, which finishes the proof.

□

2.5 One block estimate

Here we prove the one-block estimate from Lemma 2.4.3.

We will first show that it holds for the unbiased process with colors, but with all rates equal to 1, i.e. the process with the generator:

$$(\mathcal{L}_0 f)(\eta) = \frac{1}{2} N^{\alpha} \sum_{x=1}^{N-1} (f(\eta^{x,x+1}) - f(\eta)) + \frac{1}{2} N^{\alpha} \sum_{x=1}^{N} \left[(f(\eta^{x,+}) - f(\eta)) + (f(\eta^{x,-}) - f(\eta)) \right]$$

and then transfer the result to the biased process by estimating its Radon-Nikodym derivative.

Lemma 2.5.1. *Let $\mathbb{P}_0^N$ be the distribution of the unbiased process with rates 1 described above. With the notation from the previous section, for any $t \in (0, T]$ and $\delta > 0$ we have:*

$$\limsup_{l \to \infty} \limsup_{N \to \infty} N^{-\gamma} \log \mathbb{P}_0^N \left(\left| \int_0^t \frac{1}{N} \sum_{x=1}^{N} V_{x,l}(\eta_t)\, dt \right| > \delta \right) = -\infty$$

Proof. First we show that we can disregard the evolution of colors. Under the dynamics $\mathcal{L}_0$ the color of each particle performs a symmetric random walk at rate 1 and the colors do not influence the swaps between sites. Take a fixed particle i and consider the the average:

$$\mathbb{E}\left(s(x_i(\eta_0)-1,\phi_{x_i(\eta_0)-1}(\eta_0))s(x_i(\eta_s)-1,\phi_{x_i(\eta_s)-1}(\eta_s))\right)$$

from equation (2.5). Suppose that the left neighbor of i at time s is j. Conditioned on all swaps and color jumps of all particles except j (up to time s), its color $\phi_j(\eta_s)$ is a non-random function $h(\phi_j(\eta_0),s)$ of s and its initial color $\phi_j(\eta_0)$ (since the color jumps of j are independent of everything else). If in the expectation above we first condition on all swaps and color jumps of all particles except j and then observe that the speed of the initial left neighbor of i does not change in time (hence can be pulled out of the conditional expectation), we can replace this expectation by:

$$\mathbb{E}\left(s(x_i(\eta_0)-1,\phi_{x_i(\eta_0)-1}(\eta_0))s(x_i(\eta_s)-1,h(\phi_{x_i(\eta_s)-1}(\eta_0),s))\right)$$

and this expression does not depend on how the colors evolve. So we only need to prove the one-block estimate for the unbiased process without color evolution.

Since we can approximate general smooth functions of x, ϕ and t by functions which have a product form, we can without loss of generality assume that $s(x,h(\phi_x(\eta_s),s))=\psi(x,s)g(\phi_x(\eta_s))$ for some smooth function $\psi(x,s)$. We want to show that in the sum (using the notation from the previous section):

$$\frac{1}{N}\sum_{x=1}^{N}\int_0^{t-t_1}\psi(x,s)L_x(\eta_s)g(\phi_x(\eta_s))\,ds$$

we can replace the value of $L_x(\eta_s)g(\phi_x(\eta_s))$ by its local average. We can reduce this to the case of only four possible colors as follows. We have:

$$L_x(\eta_s)g(\phi_x(\eta_s))=\left(L_x^+(\eta_s)-L_x^-(\eta_s)\right)\left(g^+(\phi_x(\eta_s))-g^-(\phi_x(\eta_s))\right)$$

It is enough to prove the lemma separately for each of the four cross terms. We can write:

$$L_x^+(\eta_s)=\int_0^1 \mathbb{1}_{\{L_x^+(\eta_s)>\lambda\}}\,d\lambda$$

$$g^+(\phi_x(\eta_s))=\int_0^1 \mathbb{1}_{\{g^+(\phi_x(\eta_s))>\theta\}}\,d\theta$$

and:

$$\frac{1}{N}\sum_{x=1}^{N}\int_0^{t-t_1}\psi(x)L_x^+(\eta_s)g^+(\phi_x(\eta_s))\,ds=\int_0^1\int_0^1 d\,\theta d\lambda\left(\frac{1}{N}\sum_{x=1}^{N}\int_0^{t-t_1}\psi(x)\mathbb{1}_{\{L_x^+(\eta_s)>\lambda\}}\mathbb{1}_{\{g^+(\phi_x(\eta_s))>\theta\}}\right)$$

so it is enough to prove the lemma for any fixed $\lambda,\theta\geq 0$. Now we can think that each particle has just four possible colors, corresponding to possible values of the pair $(\mathbb{1}_{\{L_x^+(\eta_s)>\lambda\}}\mathbb{1}_{\{g^+(\phi_x(\eta_s))>\theta\}})$, and these colors do not change in time for a given particle. In this setting the statement of the lemma follows easily from the usual one block estimate for the simple exclusion process, see [KL99] or [HV99]. □

We can now prove the superexponential estimate for the biased process.

Proof of Lemma 2.4.3. Since we have the superexponential estimate for the process $\mathbb{P}_0^N$, to transfer the superexponential estimate to the biased process we will need to estimate the Radon-Nikodym derivative of the two processes.

For a Markov process $\mathbb{P}$ with jump rates $\lambda(x)p(x,y)$ and another process $\widetilde{\mathbb{P}}$ with rates $\widetilde{\lambda}(x)\widetilde{p}(x,y)$ the Radon-Nikodym derivative up to time t is given by:

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}(t) = \exp\left\{ -\int_0^t \left(\widetilde{\lambda}(X_s) - \lambda(X_s)\right)\, ds + \sum_{s\le t} \log \frac{\widetilde{\lambda}(X_{s-})\widetilde{p}(X_{s-}, X_s)}{\lambda(X_{s-})p(X_{s-}, X_s)} \right\}$$

where the sum is over jump times s.

Let us look at $\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}_0^N}$. In our case the intensities λ, $\widetilde{\lambda}$ cancel, since the sum of differences of rates $\varepsilon(s_x - s_{x+1})$ telescopes, the rates s at the boundaries are 0 and the rates for the color change also sum up to 1. We get:

$$\begin{aligned}\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}_0^N}(t) =& \exp\Big\{ \sum_{s\le t} \log\big(1 + \varepsilon\left[s(x_{j_s}, \phi_{x_{j_s}}(\eta_s)) - s(x_{j_s}+1, \phi_{x_{j_s}+1}(\eta_s))\right]\big) + \\ &+ \sum_{s_+\le t} \log\Big(1 + \varepsilon\Big[r(x_{j_{s_+}}, \phi_{x_{j_{s_+}}}(\eta_{s_+}))\Big]\Big) + \sum_{s_-\le t} \log\Big(1 - \varepsilon\Big[r(x_{j_{s_-}}, \phi_{x_{j_{s_-}}}(\eta_{s_-}))\Big]\Big) \Big\}\end{aligned}$$

where j_s is the label of the particle which makes a swap at time s and $j_{s_\pm}$ is the label of the particle that changes its color at times $s_\pm$ by ± 1.

Now we use the fact that empirical currents across edges can be approximated by their averages modulo a small martingale. Denote for simplicity $(\nabla_x s)(\eta) = s(x, \phi_x(\eta)) - s(x+1, \phi_{x+1}(\eta))$. Consider any functions $h(x,\eta)$, $h^\pm(x,\eta)$ and the extended state space consisting of pairs (η, J), $J \in \mathbb{R}$, with the generator acting by (with the convention that the $x = N$ term for swaps is 0):

$$\begin{aligned}(\mathcal{L}f)(\eta, J) =& \frac{1}{2}N^\alpha \Big[\sum_{x=1}^N \big(f(\eta^{x,x+1}, J + h(x, x+1, \eta)) - f(\eta, J)\big) + \\ &+ \sum_{x=1}^N \big(f(\eta^{x,\pm}, J + h^\pm(x,\eta)) - f(\eta, J)\big) \Big]\end{aligned}$$

For the test function $f(\eta, J) = J$ we get $(\mathcal{L}f)(\eta, J) = \frac{1}{2}N^\alpha \sum\limits_{x=1}^N (h(x,\eta) + h^\pm(x,\eta))$ (assuming $h(N,\eta) = 0$ for simplicity of notation from this poin ton). With this notation and $h(x,\eta) = \log\left[1 + \varepsilon(\nabla_x s)(\eta_s)\right]$, $h^\pm(x,\eta) = \log\left[1 \pm \varepsilon r(x, \phi_x(\eta))\right]$ we have that J at time t is equal to the sum over jumps in the exponent above. By the martingale formula we can write:

$$J_t = M_t + \frac{1}{2}N^\alpha \Big[\sum_{x=1}^N \int_0^t \log\left[1 + \varepsilon(\nabla_x s)(\eta_s)\right]\, ds + \sum_{x=1}^N \int_0^t \log\Big[1 \pm \varepsilon r(x, \phi_x(\eta_s)\Big]\, ds \Big]$$

where M_t is a mean zero martingale with respect to $\mathbb{P}^N$. Expanding all terms up to order ε^2 allows us to write (with the ε term with rates r vanishing simply because they are $\pm r(x, \phi_x)$):

$$J_t = M_t + \frac{1}{2}N^\alpha \int_0^t \sum_{x=1}^N \left[\varepsilon(\nabla_x s)(\eta_s) - \frac{\varepsilon^2}{2}\left[(\nabla_x s)(\eta_s)\right]^2 - \varepsilon^2 r(x, \phi_x(\eta_s))^2\right] ds + O(N^\alpha \varepsilon^3)$$

Note that the term linear in ε also vanishes and all the other terms are bounded by $aTN^{\alpha+1}\varepsilon^2$ for some $a > 0$, since the speeds s are bounded. Recalling that $\varepsilon = N^{1-\alpha}$ and $\gamma = 3 - \alpha$ we have:

$$J_t \leq M_t + aTN^\gamma$$

so we only need to bound the martingale term. This is done by a standard use of exponential martingales and we only sketch the argument. For any $\lambda > 0$ the following function:

$$Z_t = \exp\left\{\lambda J_t - \int_0^t e^{-\lambda J_t}\mathcal{L}e^{\lambda J_t}\right\}$$

is a mean one martingale. By performing a similar calculation as above we get:

$$Z_t = \exp\left\{\lambda J_t - \frac{1}{2}N^\alpha \int_0^t \sum_{x=1}^N \left[\left(e^{\lambda \log(1+\varepsilon \nabla_x s(\eta_s))} - 1\right) + \left(e^{\lambda \log(1 \pm \varepsilon r(x, \phi_x(\eta_s))} - 1\right)\right] ds\right\}$$

By expanding in powers of λ and ε, disregarding terms which are $o(N^\gamma)$ and replacing J_t back by M_t we arrive at:

$$Z_t = \exp\left\{\lambda M_t - N^\gamma \frac{\lambda^2}{4}\int_0^t \frac{1}{N}\sum_{x=1}^{N-1}(\nabla_x(s(\eta_s))^2 + 2r(x, \phi_x(\eta_s)^2))\, ds + O(\lambda^3 N^\gamma \varepsilon)\right\}$$

Since the coefficient quadratic in λ is bounded by a constant and M_t has mean 0 (as the rates are bounded), an application of Chebyshev exponential inequality and optimization over λ gives:

$$\mathbb{P}^N\left(M_t \geq CN^\gamma\right) \leq e^{-N^\gamma KC^2}$$

for some $K > 0$ and for any $C > 0$. Now it is easy to transfer the superexponential bound from $\mathbb{P}^N$ to $\widetilde{\mathbb{P}}^N$. Let $\mathcal{O}_{N,l}$ be the event from the statement of the lemma. We have:

$$\widetilde{\mathbb{P}}^N\left(\mathcal{O}_{N,l}\right) = \widetilde{\mathbb{E}}\left(\mathbb{1}_{\mathcal{O}_{N,l}}\right) = \mathbb{E}\left(\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}^N}\mathbb{1}_{\mathcal{O}_{N,l}}\right)$$

Let $U_{N,C} = \{M_t \leq CN^\gamma\}$. On this set the derivative is bounded by $e^{(aT+C)N^\gamma}$. By the concentration inequality above we have $\mathbb{P}^N(U^c_{N,C}) \leq e^{-N^\gamma K C^2}$, so we can write:

$$\widetilde{\mathbb{P}}^N\left(\mathcal{O}_{N,l}\right) \leq \mathbb{E}\left(\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}^N}\mathbb{1}_{\mathcal{O}_{N,l}}\right) = \mathbb{E}\left(\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}^N}\mathbb{1}_{\mathcal{O}_{N,l}}\mathbb{1}_{U_{N,C}}\right) + \mathbb{E}\left(\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}^N}\mathbb{1}_{\mathcal{O}_{N,l}}\mathbb{1}_{U^c_{N,C}}\right) \leq$$

$$\leq e^{(aT+C)N^\gamma}\mathbb{P}^N(\mathcal{O}_{N,l}) + \mathbb{E}\left(\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}^N}\mathbb{1}_{U^c_{N,C}}\right)$$

For the last term we have:

$$\mathbb{E}\left(\frac{\mathrm{d}\widetilde{\mathbb{P}}^N}{\mathrm{d}\mathbb{P}^N}\mathbb{1}_{U^c_{N,C}}\right) = \mathbb{E}\left(e^{J_t}\mathbb{1}_{U^c_{N,C}}\right) \leq e^{aTN^\gamma}\mathbb{E}\left(e^{M_t}\mathbb{1}_{\{M_t > CN^\gamma\}}\right)$$

Given the Gaussian tail bound for M_t, it is easy to see that for large enough C the second expression will be bounded by $e^{-bC^2N^\gamma}$ for some $b > 0$. Since the probability $\mathbb{P}^{\mathcal{O}_{N,l}}$ also decays superexponentially, this proves the superexponential estimate. □

2.6 Large deviation lower bound

In this section we prove a large deviation lower bound for being close to a specified deterministic permuton process.

First we derive the formula for the Radon-Nikodym derivative of the unbiased process with respect to the biased one. The calculation is the same as in proof of Lemma 2.4.3, with the only difference that we are using generator $\widetilde{\mathcal{L}}$ instead of $\mathcal{L}$ (and the sign in the exponent is reversed). By writing:

$$\frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N}(t) = \exp\left\{-\sum_{s\leq t}\log\left(1+\varepsilon\left[s(x_{j_s},\phi_{x_{j_s}}(\eta_s)) - s(x_{j_s}+1,\phi_{x_{j_s}+1}(\eta_s))\right]\right)\right\}$$

and denoting the sum in the exponent by J_t we obtain:

$$J_t = M_t + \frac{1}{2}N^\alpha\sum_{x=1}^{N-1}\int_0^t\left[1+\varepsilon(\nabla_x s)(\eta_s)\right]\log\left[1+\varepsilon(\nabla_x s)(\eta_s)\right]\,ds$$

where M_t is a mean zero martingale with respect to $\widetilde{\mathbb{P}}^N$. Expanding all terms up to order ε^2 allows us to write:

$$\frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N}(t) = \exp\left\{-M_t - \frac{1}{2}N^\alpha\int_0^t\sum_{x=1}^{N-1}\left[\varepsilon(\nabla_x s)(\eta_s) + \frac{\varepsilon^2}{2}\left[(\nabla_x s)(\eta_s)\right]^2\right]\,ds + O(N^\alpha\varepsilon^3)\right\}$$

The term linear in ε vanishes. Recalling that $\varepsilon = N^{1-\alpha}$ and $\gamma = 3-\alpha$ we have:

$$\frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N}(t) = \exp\left\{-M_t - \frac{1}{4}N^\gamma\int_0^t\frac{1}{N}\sum_{x=1}^{N}\left[(\nabla_x s)(\eta_s)\right]^2\,ds + o(N^\gamma)\right\}$$

Expanding $(\nabla_x s)^2$ leads us to:

$$\frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N}(t) = \exp\Bigg\{ - M_t - \frac{1}{2}N^\gamma \int_0^t \frac{1}{N}\sum_{x=1}^N s(x,\phi_x(\eta_s))^2\, ds +$$

$$+ \frac{1}{2}N^\gamma \int_0^t \frac{1}{N}\sum_{x=1}^N s(x,\phi_x(\eta_s))s(x+1,\phi_{x+1}(\eta_s))\, ds + o(N^\gamma)\Bigg\}$$

The martingale term will be typically $o(N^\gamma)$ - one can actually show the same exponential bound as in proof of Lemma 2.4.3, but here we only need that it is small with high probabiltiy. By the formula for quadratic variation and a calculation similar to the one above we get:

$$N_t = M_t^2 - \frac{1}{2}N^\alpha \int_0^t \sum_{x=1}^N \left[1+\varepsilon(\nabla_x s)(\eta_s)\right] \log\left[1+\varepsilon(\nabla_x s)(\eta_s)\right]^2\, ds$$

is a mean 0 martingale. By expanding log up to ε^2 we see that quadratic variation is bounded by $CN^{\alpha+1}\varepsilon^2 = CN^\gamma$ for some $C > 0$ and now an application of Doob's inequality gives us that $M_t = o(N^\gamma)$ with high probability as $N \to \infty$.

The second sum in the exponent will be small by stationarity. At fixed time s for each x the correlation term $s(x,\phi_x(\eta_s))s(x+1,\phi_{x+1}(\eta_s))$ has mean 0, since η_s has stationary distribution. Moreover these terms are independent, so application of any concentration inequality for independent bounded random variables will give us that for fixed s any $\varepsilon > 0$:

$$\widetilde{\mathbb{P}}^N\left(\frac{1}{N}\sum_{x=1}^N s(x,\phi_x(\eta_s))s(x+1,\phi_{x+1}(\eta_s)) > \varepsilon\right) \to 0$$

as $N \to \infty$. Since this holds for fixed s, we get:

$$\widetilde{\mathbb{P}}^N\left(\int_0^t \frac{1}{N}\sum_{x=1}^N s(x,\phi_x(\eta_s))s(x+1,\phi_{x+1}(\eta_s))\, ds > \varepsilon\right) \to 0$$

which proves that the correlation term is $o(N^\gamma)$ with high probability.

These bounds imply that there exist sets U_N such that $\widetilde{\mathbb{P}}^N(U_N) \to 1$ as $N \to \infty$ and on the set U_N the Radon-Nikodym derivative is equal to:

$$\exp\Bigg\{ - \frac{1}{2}N^\gamma \int_0^t \frac{1}{N}\sum_{x=1}^N s(x,\phi_x(\eta_s))^2\, ds + o(N^\gamma)\Bigg\}$$

We can now use this formula and the law of large numbers proved in the previous section to establish a large deviations lower bound for the interchange process. Recall that for any permuton process π its energy was defined by:

$$I(\pi) = \frac{1}{2}\mathbb{E}_{\gamma\sim\pi}\mathcal{E}(\gamma)$$

Let $X = X^\varepsilon$ be the deterministic permuton process associated to $P^\varepsilon = (X^\varepsilon, \Phi^\varepsilon)$. We have the following large deviation principle:

Theorem 2.6.1. *Let $\mathbb{P}^N$ be the distribution of the unbiased interchange process and let $\boldsymbol{X}^{\eta^N}$ be the corresponding random path process. For any open set $\mathcal{O} \subseteq \mathcal{P}$ containing X we have:*

$$\liminf_{N\to\infty} N^{-\gamma} \log \mathbb{P}^N \left(\boldsymbol{X}^{\eta^N} \in \mathcal{O} \right) \geq -I(X)$$

Proof. It will be enough to show the bound above for $\mathcal{O}$ being any open ball $B(X, \varepsilon)$ around X. In fact we will show the bound for the probability:

$$\liminf_{N\to\infty} N^{-\gamma} \log \mathbb{P}^N \left(\boldsymbol{P}^{\eta^N} \in B(P, \varepsilon) \right)$$

where we have denoted the distribution of the unbiased process with colors also by $\mathbb{P}^N$. Since $\mathbb{P}^N \left(\boldsymbol{X}^{\eta^N} \in B(X, \varepsilon) \right) \geq \mathbb{P}^N \left(\boldsymbol{P}^{\eta^N} \in B(P, \varepsilon) \right)$, this will prove the large deviation bound.

Let $\widetilde{\mathbb{P}}_N$ be the distribution of the biased process associated to P and consider the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N}(t)$. By considering sets U_N introduced in the preceding calculation we get that on U_N the derivative is equal to:

$$\exp\left\{ -\frac{1}{2} N^\gamma \int_0^t \frac{1}{N} \sum_{x=1}^N s(x, \phi_x(\eta_s))^2 \, ds + o(N^\gamma) \right\}$$

Now by the law of large numbers from Theorem 2.4.1 random processes $\boldsymbol{P}^{\eta^N}$ converge in distribution to P when η^N is distributed according to $\widetilde{\mathbb{P}}^N$, so there exists a sequence of balls $B_n = B_n(P, \varepsilon_n)$ around P, with $\varepsilon_n \to 0$, such that $\widetilde{\mathbb{P}}^N \left(\boldsymbol{P}^{\eta^N} \in B_n \right) \to 1$. Let $V_n = U_n \cap \{ \boldsymbol{P}^{\eta^N} \in B_n \}$. With $\mathbb{E}$ denoting the expectation with respect to $\mathbb{P}^N$ and $\widetilde{\mathbb{E}}$ with respect to $\widetilde{\mathbb{P}}^N$ we have for sufficiently large N:

$$\mathbb{P}^N \left(\boldsymbol{P}^{\eta^N} \in B(T, \varepsilon) \right) = \mathbb{E} \left(\mathbb{1}_{\{ \boldsymbol{P}^{\eta^N} \in B(T, \varepsilon) \}} \right) \geq \mathbb{E}(\mathbb{1}_{V_n}) = \widetilde{\mathbb{E}} \left(\frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N} \mathbb{1}_{V_n} \right) \geq \widetilde{\mathbb{P}}^N(V_n) \left(\inf_{\eta \in V_n} \frac{\mathrm{d}\mathbb{P}^N}{\mathrm{d}\widetilde{\mathbb{P}}^N}(\eta) \right)$$

Since on the set V_n the Radon-Nikodym derivative is given by the formula above and $\widetilde{\mathbb{P}}^N(V_n) \to 1$, by taking log and dividing by N^γ we get:

$$N^{-\gamma} \log \mathbb{P}^N \left(\boldsymbol{P}^{\eta^N} \in B(P, \varepsilon) \right) \geq - \inf_{\eta \in V_n} I_N(\eta) + o(1)$$

where:

$$I_N(\eta) = \frac{1}{2} \left(\frac{1}{N} \sum_{x=1}^N \int_0^t s(x, \phi_x(\eta_s))^2 \, ds \right)$$

Now it is not difficult to see that the infimum on the right hand side converges to $I(X)$ as $N \to \infty$ and then $\varepsilon \to 0$. Recalling the notation $S(X, \Phi) = s(NX, N\Phi)$, we need to show that:

$$\frac{1}{N} \sum_{i=1}^N \int_0^t S(X_i(\eta_s^N), \Phi_i(\eta_s^N))^2 \, ds \to \mathbb{E} \left(\int_0^t S(X(s), \Phi(s))^2 \, ds \right) \tag{2.6}$$

where $(X_i(\eta_t^N), \Phi_i(\eta_t^N))$ is the rescaled trajectory of particle i in η^N and on the right hand side we have the expectation over trajectories $(X(t), \Phi(t))$ of the process P. Since $\{\boldsymbol{P}^{\eta^N} \in B_n\} \subseteq V_n$, for any $\delta > 0$ we have:

$$\widetilde{\mathbb{P}}^N \left(\max\{\sup_{t \leq T} \left| X_i(\eta_t^N) - X_i(t) \right|, \sup_{t \leq T} \left| \Phi_i(\eta_t^N) - \Phi_i(t) \right| \} > \delta \right) \to 0$$

for almost all particles i, where $(X_i(t), \Phi_i(t))$ is the solution of (2.2) corresponding to the initial condition $(X_i(\eta_0^N), \Phi_i(\eta_0^N))$. Since $S(X, \Phi)$ is bounded by 1 we get for any i:

$$\left| \int_0^t \left[S(X_i(\eta_s^N), \Phi_i(\eta_s^N)^2 - S(X_i(s), \Phi_i(s))^2 \right] \, ds \right| \leq 2 \int_0^t \left| S(X_i(\eta_s^N), \Phi_i(\eta_s^N)) - S(X_i(s), \Phi_i(s)) \right| ds$$
$$\leq KT \max \left\{ \sup_{t \leq T} \left| X_i(\eta_t^N) - X_i(t) \right|, \sup_{t \leq T} \left| \Phi_i(\eta_t^N) - \Phi_i(t) \right| \right\}$$

for some $K > 0$ depending on the Lipschitz constant of S. Since the right hand side goes to 0 in probability for almost all particles, we get that with high probability as $N \to \infty$ the left hand side of (2.6) is close to:

$$\frac{1}{N} \sum_{i=1}^{N} \int_0^t S(X_i(s), \Phi_i(s))^2 \, ds$$

Since $(X_i(s), \Phi_i(s))$ is a solution of (2.2) and S is the derivative of X, the integral is equal simply to the energy of the path $X_i(t)$. Since for each i the initial condition $\Phi_i(\eta_0^N)$ has uniform distribution on $\left\{ \frac{1}{N}, \ldots, 1 \right\}$, independently for all i, it follows easily that this expression converges with high probability to the expected energy on the right hand side of (2.6). This implies $\inf_{\eta \in V_N} I_N(\eta) \to I(X)$ as $N \to \infty$ and finishes the proof of the lower bound. □

Corollary 2.6.2. *The statement of Theorem 2.6.1 holds also for the sine curve process $\mathcal{A}_t$.*

Proof. The proofs of the law of large numbers and the lower bound required the rates of the biased process to be bounded, so we cannot directly apply the theorem to the sine curve process. However, it is enough to prove that processes P^ε converge in distribution as $\varepsilon \to 0$ to the sine curve process together with its associated color process Φ. One can show that the solutions of the equation 2.2 together with their derivatives will be close to the corresponding sine curves in the uniform norm. This can be done by directly analyzing the singularity of the ODE for the sine curve process. We skip the proof. □

2.7 Large deviation upper bound

In this section we derive large deviation upper bound for the distribution of the unbiased interchange process to be close to a specified permuton process. As a first step we will bound the probability that after a short time we see a fixed permutation in the interchange process. This is done by means of an exponential martingale. The idea is as follows - if $J_S(\eta)$ is a function of the process (depending on some set of parameters S) such that $e^{J_S(\eta)}$ is a nonnegative mean one martingale, then for any permutation σ we can write (with (η_0, η_t)

denoting the permutation $\eta_0^{-1}\eta_t$):

$$\mathbb{P}^N((\eta_0,\eta_t)=\sigma)=\mathbb{E}(\mathbb{1}_{\{(\eta_0,\eta_t)=\sigma\}})=\mathbb{E}\left(e^{J_S}e^{-J_S}\mathbb{1}_{\{(\eta_0,\eta_t)=\sigma\}}\right)\leq$$
$$\inf_{\{(\eta_0,\eta_t)=\sigma\}}e^{-J_S(\eta)}\mathbb{E}\left(e^{J_S}\mathbb{1}_{\{(\eta_0,\eta_t)=\sigma\}}\right)\leq\inf_{\{(\eta_0,\eta_t)=\sigma\}}e^{-J_S(\eta)}$$

where the last inequality comes from the fact that e^{J_S} is a nonnegative mean one martingale. If J_S depends only on η_0 and η_t we obtain a particularly simple expression:

$$\mathbb{P}^N((\eta_0,\eta_t)=\sigma)\leq e^{J_S(\sigma)}$$

We can then optimize over all S to obtain a large deviation upper bound. The family of martingales we will use is similar to the one used in analyzing large deviations for a simple random walk.

Fix $\delta>0$ and a sequence $S=(s_1,\ldots,s_N)$, with $s_i\in\left\{\frac{-1}{\delta},\frac{-1+\frac{1}{N}}{\delta},\ldots,\frac{1-\frac{1}{N}}{\delta},\frac{1}{\delta}\right\}$. Consider the function:

$$F_S(\eta_t)=\varepsilon\sum_{i=1}^{N}s_ix_i(\eta_t)$$

where $x_i(\eta_t)$ is the position of the particle i in the configuration η_t. We will write $s_x(\eta_t)$ for $s_{\eta_t^{-1}(x)}$. If $\mathcal{L}$ is the generator of the unbiased interchange process, then by the formula for exponential martingales we obtain that:

$$M_t^S=\exp\left\{F_S(\eta_t)-F_S(\eta_0)-\int_0^t e^{-F_S(\eta_s)}\mathcal{L}e^{F_S(\eta_s)}\,ds\right\}$$

is a mean one positive martingale with respect to $\mathbb{P}^N$. We have:

$$\mathcal{L}e^{F_S(\eta)}=\frac{1}{2}N^\alpha\sum_{x=1}^{N-1}\left(e^{F_S(\eta^{x,x+1})}-e^{F_S(\eta)}\right)=\frac{1}{2}N^\alpha\sum_{x=1}^{N-1}\left(e^{F_S(\eta)+\varepsilon[s_x(\eta)-s_{x+1}(\eta)]}-e^{F_S(\eta)}\right)$$

so:

$$M_t^S=\exp\left\{\varepsilon\sum_{i=1}^{N}s_i\left(x_i(\eta_t)-x_i(\eta_0)\right)-\frac{1}{2}N^\alpha\int_0^t\sum_{x=1}^{N-1}\left(e^{\varepsilon[s_x(\eta_s)-s_{x+1}(\eta_s)]}-1\right)\,ds\right\}$$

Expanding up to order ε^2 (with the constants in $O(\cdot)$ depending only on T and δ, as s are bounded by $1/\delta$):

$$M_t^S=\exp\left\{\varepsilon\sum_{i=1}^{N}s_i\left(x_i(\eta_t)-x_i(\eta_0)\right)-\frac{1}{2}N^\alpha\varepsilon\int_0^t\sum_{x=1}^{N-1}[s_x(\eta_s)-s_{x+1}(\eta_s)]\,ds-\right.$$
$$\left.-\frac{1}{4}N^\alpha\varepsilon^2\int_0^t\sum_{x=1}^{N-1}\left(s_x(\eta_s)-s_{x+1}(\eta_s)\right)^2\,ds+O(N^{\alpha+1}\varepsilon^3)\right\}\tag{2.7}$$

Rescaling by appropriate powers of N and expressing the exponents in terms of the large deviation exponent

γ (observe that the sum of $s_x - s_{x+1}$ cancels up to a term which is $O(N^\alpha \varepsilon) = o(N^\gamma)$) we get:

$$M_t^S = \exp\left\{N^\gamma \left[\frac{1}{N}\sum_{i=1}^N s_i \left(\frac{x_i(\eta_t) - x_i(\eta_0)}{N}\right) - \frac{1}{4}\int_0^t \frac{1}{N}\sum_{x=1}^{N-1} \left(s_x(\eta_s) - s_{x+1}(\eta_s)\right)^2 \, ds\right] + o(N^\gamma)\right\}$$

Since the sum of s_x^2 does not depend on time we can write:

$$M_t^S = \exp\left\{N^\gamma \left[\frac{1}{N}\sum_{i=1}^N s_i \left(\frac{x_i(\eta_t) - x_i(\eta_0)}{N}\right) - \frac{t}{2}\left(\frac{1}{N}\sum_{i=1}^N s_i^2\right) + \frac{1}{2}\int_0^t \frac{1}{N}\sum_{x=1}^{N-1} s_x(\eta_s)s_{x+1}(\eta_s)\, ds + o(1)\right]\right\}$$

As before we want to use the one block estimate to get rid of the sum involving correlations between s_x for adjacent x. This time the correlation term might not be small, since s_i are arbitrary, but it will be always nonnegative, so we can neglect it for the sake of the upper bound. More precisely, consider the interchange process in which each particle also has an additional color s_i (which does not influence the dynamics). Let $\Lambda_{x,l}$ be a box of size l around x and let $\mu_{x,l}(\eta)$ be the empirical measure of colors in $\Lambda_{x,l}$ in configuration η, i.e. a product measure over configurations restricted to $\Lambda_{x,l}$ such that the probability of s_i is proportional to the number of particles with color s_i in $\Lambda_{x,l}$. By the superexponential estimate for the unbiased interchange process (2.5.1 or actually its simpler version without color evolution) and a local function $f(s_x, s_{x+1}) = s_x s_{x+1}$ in the integral above we can replace:

$$s_x(\eta_s)s_{x+1}(\eta_s)$$

with

$$\mathbb{E}_{\mu_{x,l}(\eta_s)}\left(s_x s_{x+1}\right)$$

on a set of superexponentially high probability as N and then $l \to \infty$. Since $\mu_{x,l}(\eta_s)$ is a product measure, $\mathbb{E}_{\mu_{x,l}(\eta_s)}\left(s_x s_{x+1}\right) = \left(\mathbb{E}_{\mu_{x,l}(\eta_s)} s_x\right)^2 \geq 0$.

This proves that there exist sets $\mathcal{O}_{N,l}$ such that on $\mathcal{O}_{N,l}$ we have:

$$M_t^S(\eta) \geq \exp\left\{N^\gamma \left[\frac{1}{N}\sum_{i=1}^N s_i \left(\frac{x_i(\eta_t) - x_i(\eta_0)}{N}\right) - \frac{t}{2}\left(\frac{1}{N}\sum_{i=1}^N s_i^2\right) + o(1)\right]\right\}$$

and $\mathbb{P}^N(\mathcal{O}_{N,l}^c) \to 0$ (as N and then $l \to \infty$) superexponentially fast:

$$\limsup_{l\to\infty}\limsup_{N\to\infty} N^{-\gamma}\log \mathbb{P}^N(\mathcal{O}_{N,l}^c) = -\infty$$

Now we can use the strategy outlined earlier with $e^{J_S(\eta)} = M_t^S(\eta)$. We write:

$$\mathbb{P}^N((\eta_0, \eta_t) = \sigma) = \mathbb{E}(\mathbb{1}_\sigma) = \mathbb{E}\left(e^{-J_S(\eta)}e^{J_S(\eta)}\mathbb{1}_{\{(\eta_0,\eta_t)=\sigma\}}\right) =$$
$$\mathbb{E}\left(e^{-J_S(\eta)}e^{J_S(\eta)}\mathbb{1}_{\{(\eta_0,\eta_t)=\sigma\}}\mathbb{1}_{\mathcal{O}_{N,l}}\right) + \mathbb{E}\left(\mathbb{1}_{\{(\eta_0,\eta_t)=\sigma\}}\mathbb{1}_{\mathcal{O}_{N,l}^c}\right)$$

On $\mathcal{O}_{N,l}$ we can use the bound obtained from the one block estimate and we get:

$$\mathbb{P}^N((\eta_0, \eta_t) = \sigma) \leq e^{-N^\gamma(I_S(\sigma)+o(1))} + \mathbb{P}^N(\mathcal{O}_{N,l}^c)$$

where:

$$I_S(\sigma) = \frac{1}{N}\sum_{i=1}^{N} s_i \left(\frac{\sigma(i)-i}{N}\right) - \frac{t}{2}\left(\frac{1}{N}\sum_{i=1}^{N} s_i^2\right)$$

The probability on the right hand side goes to 0 superexponentially in N^γ, so we can neglect it after taking log and dividing by N^γ.

Now let $t = \delta$. To optimize over choice of S, we observe that $I_S(\sigma)$ is quadratic in s_i, so we should take:

$$s_i = \frac{\sigma(i)-i}{\delta N}$$

which is possible, since we assumed s_i were bounded by $1/\delta$. This gives the optimal value equal to:

$$\frac{1}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\delta}\left(\frac{\sigma(i)-i}{N}\right)^2\right)$$

which just the permuton energy $I(\mu_\sigma)$ rescaled by the time δ. So as $N \to \infty$ we have:

$$N^{-\gamma}\log \mathbb{P}^N((\eta_0, \eta_\delta) = \sigma) \leq -\frac{1}{\delta} I(\mu_\sigma) + o(1) \tag{2.8}$$

In other words, the large deviation rate of seeing a permutation σ at (possibly short) time δ in the interchange process is proportional to the energy of permutation σ. Armed with this result we can now prove a general large deviation upper bound.

Theorem 2.7.1. *Let $\mathbb{P}^N$ be the distribution of the unbiased interchange process and let $\boldsymbol{X}^{\eta^N}$ be the corresponding random permutation process. For any closed set $\mathcal{K} \subseteq \mathcal{P}$ we have:*

$$\limsup_{N\to\infty} N^{-\gamma}\log \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in \mathcal{K}\right) \leq -\inf_{\pi\in\mathcal{K}} I(\pi)$$

where $I(\pi)$ is the energy of the process π.

Proof. We will first prove the bound for the case when $\mathcal{K}$ is compact and then extend it to closed sets by proving exponential tightness.

Fix $\varepsilon > 0$ and consider a covering of $\mathcal{K}$ with balls $B(\pi, \varepsilon)$ for $\pi \in \mathcal{K}$. By choosing a finite subcover we obtain a finite set $\pi_1, \ldots, \pi_{K(\varepsilon)}$ for some $K(\varepsilon)$ and by using a union bound over balls from the subcover we obtain:

$$\mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in \mathcal{K}\right) \leq K(\varepsilon)\sup_i \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in B(\pi_i, \varepsilon)\right)$$

so it is enough to prove the large deviation bound for $\mathcal{K}$ equal to $B(\pi, \varepsilon)$ for any permuton process π. Furthermore, since $I_(\pi)$ is lower semicontinous, by adjusting ε we can always assume that it attains its infimum on this ball at π.

Fix a finite set of times $0 \leq t_1 \leq \ldots \leq t_k \leq T$. Since $\boldsymbol{X}^{\eta^N}$ being close to π implies that their finite dimensional distributions must be close, it is enough to bound the probability of $\boldsymbol{X}^{\eta^N}_{t_1,\ldots,t_k} = (\boldsymbol{X}^{\eta^N}_{t_1}, \ldots, \boldsymbol{X}^{\eta^N}_{t_k})$ being close to $\pi_{t_1,\ldots,t_k} = (\pi_{t_1}, \ldots, \pi_{t_k})$. Now note that the process $\boldsymbol{X}^{\eta^N}$ has independent increments, i.e. the permutations $(\eta^N_{t_i}, \eta^N_{t_{i+1}})$ for any family non-overlapping intervals $[t_i, t_{i+1}]$ are independent. Thus we can write:

$$\mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in B(\pi,\varepsilon)\right) \le \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{t_1,\ldots,t_k} \in B(\pi_{t_1,\ldots,t_k},\varepsilon)\right) \le$$
$$\mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{t_1,t_2} \in B(\pi_{t_1,t_2},\varepsilon) \wedge \ldots \wedge \boldsymbol{X}^{\eta^N}_{t_1,t_2} \in B(\pi_{t_{k-1},t_k}\right) = \prod_{i=1}^{k-1} \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{t_i,t_{i+1}} \in B(\pi_{t_i,t_{i+1}},\varepsilon)\right)$$

In this way we have reduced the problem to bounding the probability that the distribution of $(\boldsymbol{X}_{t_i}, \boldsymbol{X}_{t_{i+1}})$ is close to a fixed permuton $(\pi_{t_i}, \pi_{t_{i+1}})$. By stationarity it is enough to consider $(\boldsymbol{X}_0, \boldsymbol{X}_{t_{i+1}-t_i})$.

Let $t = t_{i+1} - t_i$ and let $P = \pi_{t_i,t_{i+1}}$ be a fixed permuton. Consider all permutations σ on N elements such that the empirical measure μ_σ is in $B(P,\varepsilon)$. Since there are at most $N!$ such permutations and this is subexponential in N^γ, we can perform a union bound over all such permutations and write:

$$\mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{0,t} \in B(P,\varepsilon)\right) \le N! \sup_{\sigma\in B(P,\varepsilon)} \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{0,t} \sim \mu_\sigma\right)$$

where on the right hand side we have the probability that the distribution of $(\boldsymbol{X}^{\eta^N}_0, \boldsymbol{X}^{\eta^N}_t)$ is equal to μ_σ. This is simply equal to $\mathbb{P}^N\left((\eta^N_0, \eta^N_t) = \sigma)\right)$. By the bound (2.8) and $N! = e^{o(N^\gamma)}$ we get:

$$N^{-\gamma} \log \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{0,t} \in B(P,\varepsilon)\right) \le \sup_{\sigma\in B(P,\varepsilon)} \left(-\frac{1}{\delta} I(\mu_\sigma)\right) + o(1)$$

Now observe that for any $\sigma \in B(P,\varepsilon)$ the energy $I\mu_\sigma$ has to be close to the energy $I(P)$, since I is continuous in the permuton topology. So there exists some $\varepsilon' > 0$ going to 0 as $\varepsilon \to 0$ such that:

$$N^{-\gamma} \log \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N}_{0,t} \in B(P,\varepsilon)\right) \le -\frac{1}{\delta} I(P) + \frac{\varepsilon'}{\delta} + o(1)$$

This chain of estimates give us the following bound:

$$N^{-\gamma} \log \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in B(\pi,\varepsilon)\right) \le -\sum_{i=1}^{k-1} \frac{1}{t_{i+1}-t_i} I(\pi_{t_i,t_{i+1}}) + \frac{k\varepsilon'}{\delta} + o(1)$$

Since ε was arbitrary, we obtain for any fixed partition $t_1, \ldots, t_k$:

$$\limsup_{N\to\infty} N^{-\gamma} \log \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in \mathcal{K}\right) \le -\inf_{\pi\in\mathcal{K}\cap\mathcal{P}} \sum_{i=1}^{k-1} \frac{1}{t_{i+1}-t_i} I(\pi_{t_i,t_{i+1}})$$

By optimizing over all finite partitions we obtain:

$$\limsup_{N\to\infty} N^{-\gamma} \log \mathbb{P}^N\left(\boldsymbol{X}^{\eta^N} \in \mathcal{K}\right) \le -\sup_{\Pi} \inf_{\pi\in\mathcal{K}\cap\mathcal{P}} \sum_{i=1}^{k-1} \frac{1}{t_{i+1}-t_i} I(\pi_{t_i,t_{i+1}})$$

where the supremum is over all finite partitions $\Pi = \{0 = t_0 < t_1 < \ldots < t_k = T\}$. As $I(\pi_{t_i,t_{i+1}})$ is also lower semicontinuous as a function of π and $\mathcal{K}$ is compact, we can exchanging the infimum and the supremum:

$$\limsup_{N\to\infty} N^{-\gamma} \log \mathbb{P}^N \left(X^{\eta^N} \in \mathcal{K} \right) \leq - \inf_{\pi\in\mathcal{K}\cap\mathcal{P}} \sup_{\Pi} \sum_{i=1}^{k-1} \frac{1}{t_{i+1}-t_i} I(\pi_{t_i,t_{i+1}}) = - \inf_{\pi\in\mathcal{K}\cap\mathcal{P}} \sup_{\Pi} I^{\Pi}(\pi)$$

Now it is easy to see that the supremum on the right hand side is equal to the energy $I(\pi)$. To show that:

$$\sup_{\Pi} \mathbb{E}_{\gamma\sim\pi} \mathcal{E}^{\Pi}(\gamma) = \mathbb{E}_{\gamma\sim\pi} \sup_{\Pi} \mathcal{E}^{\Pi}(\gamma)$$

we observe that for any γ by taking a nested sequence of partitions Π_n we can always assume that the sup is in fact a limit and $\mathcal{E}^{\Pi^n}(\gamma) \to \mathcal{E}^{\Pi}(\gamma)$ monotonically. Now we apply monotone convergence to get the same result with $\mathbb{E}_{\gamma\sim\pi}$ and this finishes the proof for compact sets.

Now we extend the large deviation upper bound to closed sets. It is enough to prove that the family of processes X^{η^N} is exponentially tight, i.e. that there exists a sequence of compact sets K_n such that:

$$\limsup_{N\to\infty} N^{-\gamma} \log \mathbb{P}^N(X^{\eta^N} \notin K_n) \leq -n$$

Since the technique is standard (see [KL99, Chapter 10]), we only sketch the proof. Fix $\delta > 0$ and divide $[0, T]$ into subintervals of length δ. Using the exponential martingale (2.7) one can show that for any interval $[k\delta, (k+1)\delta]$ and for any $\varepsilon > 0$ we have:

$$\mathbb{P}^N \left(\sup_{s,t\in[k\delta,(k+1)\delta]} \frac{1}{N} \sum_{i=1}^{N} \left| X_i^{\eta^N}(t) - X_i^{\eta^N}(s) \right| > \varepsilon \right) \leq e^{-N^{\gamma} K \frac{\varepsilon^2}{\delta} + o(1)}$$

for some $K > 0$. Since increments of each $X_i^{\eta^N}$ are independent between subintervals, one easily obtains a bound on the total sum J of increments over all particles and all subintervals. This in turn implies implies that for any $c > 0$ the probability that there are more than $N\frac{J}{c}$ particles which have an increment greater than c on any of the subintervals is at most:

$$e^{-CN^{\gamma}J^2 + o(1)}$$

for some $C > 0$. By adjusting δ, c and J we obtain that there are functions $f(\varepsilon)$, $g(\varepsilon)$, $h(\varepsilon)$ such that $f(\varepsilon), g(\varepsilon) \to 0$, $h(\varepsilon) \to \infty$ as $\varepsilon \to 0$ and the following holds: the probability that there exist more than $f(\varepsilon)N$ particles which have increment more than ε on some interval of length $g(\varepsilon)$ is smaller than $\exp\{-N^{\gamma}h(\varepsilon) + o(1)\}$. By taking an appropriate sequence of ε_m we can assure that this probability is smaller than $\exp\{-N^{\gamma}nm\}$. By intersecting over $m \geq 1$ sets on which this event does not hold, we obtain that with probability asymptotically at least:

$$1 - \sum_{m\geq 1} e^{-N^{\gamma}nm} \geq 1 - e^{-N^{\gamma}n}$$

trajectories of almost all particles in η^N will have modulus of continuity determined by the function $g(\varepsilon)$, which easily implies compactness. □

As an application of the large deviation bounds from the last sections we sketch the proof of Theorem 2.1.2.

Proof of Theorem 2.1.2. Let $\mathcal{K}$ be the set of all permuton processes (and permutation processes for finite permutations) which are equal to the reverse permuton (or the reverse permutation for finite permutations) at time 1. If we run the interchange process η^N for time $\frac{1}{2}N^{\alpha} = \frac{1}{2}N^{1+\varepsilon}$ (note that we rescale time by N, since the interchange process is in continuous time), the number of swaps will not be exactly equal to $N^{1+\alpha}$, but will be tightly concentrated around this value. Therefore asymptotically the probability of η^N belonging to $\mathcal{K}$ will be equal to the number of all paths from the identity to the reverse permutation divided by the number of all paths.

Since the distribution of the sine curve process $\mathcal{A}$ at time 1 is equal to the reverse permuton, by the large deviation upper bound of Theorem 2.7.1 the infimum of the rate function over $\mathcal{K}$ will be at most equal to the energy $I(\mathcal{A})$. By results of [RVb] the sine curve process is the unique minimizer of energy on $\mathcal{K}$. Together with the lower bound of Theorem 2.6.1 and Corollary 2.6.2 this implies that the fraction of paths of length $N^{2+\varepsilon}$ which lie in $\mathcal{K}$ is asymptotically:

$$\sim \exp\{-N^{2-\varepsilon}I(\mathcal{A}) + o(N^{2-\varepsilon})\}$$

The energy of $\mathcal{A}$ is easily computed to be $\frac{\pi^2}{6}$. Since the number of all paths of length $\frac{1}{2}N^{2+\varepsilon}$ is $(N-1)^{\frac{1}{2}N^{2+\varepsilon}}$, we obtain the desired formula.

Note that this also implies the second part of the theorem. All paths with energy asymptotically smaller than $I(\mathcal{A})$ will occur with negligible probability (as its exponential rate of decay will have a larger exponent), so we need to consider only paths close to permuton processes with energy $I(\mathcal{A})$. Since the Archimiedean path is the unique minimizer of energy among such paths, the claim follows. □

Bibliography

[AGH12] Omer Angel, Vadim Gorin, and Alexander E. Holroyd, *A pattern theorem for random sorting networks*, Electron. J. Probab. **17** (2012), no. 99, 16.

[AH10] Omer Angel and Alexander E. Holroyd, *Random subnetworks of random sorting networks*, Electron. J. Combin. **17** (2010), no. 1, Note 23, 7.

[AHRV07] Omer Angel, Alexander E. Holroyd, Dan Romik, and Bálint Virág, *Random sorting networks*, Adv. Math. **215** (2007), no. 2, 839–868.

[AK] Gideon Amir and Gady Kozma, *Minimal harmonic functions III: the sublogarithmic regime*, in preparation.

[Ale92] G. Alexopoulos, *A lower estimate for central probabilities on polycyclic groups*, Canad. J. Math. **44** (1992), no. 5, 897–910.

[AV12] Gideon Amir and Bálint Virág, *Speed exponents of random walks on groups*, arXiv:1203.6226v3, math.PR.

[AV14] ———, *Positive speed for high-degree automaton groups*, Groups Geom. Dyn. **8** (2014), no. 1, 23–38.

[BE11] Laurent Bartholdi and Anna Erschler, *Poisson-Furstenberg boundary and growth of groups*, arXiv:1107.5499v1, math.GR.

[BE12] ———, *Growth of permutational extensions*, Invent. Math. **189** (2012), no. 2, 431–455.

[BE14] ———, *Imbeddings into groups of intermediate growth*, arXiv:1403.5584v2, math.GR.

[Bri13] Jérémie Brieussel, *Behaviors of entropy on finitely generated groups*, Ann. Probab. **41** (2013), no. 6, 4116–4161.

[GGKK15] Roman Glebov, Andrzej Grzesik, Tereza Klimošová, and Daniel Král', *Finitely forcible graphons and permutons*, J. Combin. Theory Ser. B **110** (2015), 112–135.

[Gou14] Antoine Gournay, *The Liouville property via Hilbertian compression*, arXiv:1403.1195v4, math.GR.

[HKM+13] Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio, *Limits of permutation sequences*, J. Combin. Theory Ser. B **103** (2013), no. 1, 93–113.

[HV99] E.P. Hsu and S.R.S. Varadhan, *Probability theory and applications*, IAS/Park City mathematics series, American Mathematical Soc., 1999.

[KL99] Claude Kipnis and Claudio Landim, *Scaling limits of interacting particle systems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 320, Springer-Verlag, Berlin, 1999.

[Lov12] László Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012.

[LP13] James R. Lee and Yuval Peres, *Harmonic maps on amenable groups and a diffusive lower bound for random walks*, Ann. Probab. **41** (2013), no. 5, 3392–3419.

[LP14] Russell Lyons and Yuval Peres, *Probability on trees and networks*, `http://http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html`, 2014.

[Pet14] Gábor Pete, *Probability and geometry on groups*, `http://www.math.bme.hu/~gabor/PGG.pdf`, 2014.

[PSC00] Ch. Pittet and L. Saloff-Coste, *On the stability of the behavior of random walks on groups*, J. Geom. Anal. **10** (2000), no. 4, 713–737.

[PSC02] C. Pittet and L. Saloff-Coste, *On random walks on wreath products*, Ann. Probab. **30** (2002), no. 2, 948–977.

[RVa] Mustazee Rahman and Bálint Virág, *Limit of the interchange process, in preparation.*

[RVb] ______, *Limits of permutation processes, in preparation.*

[SCZ] Laurent Saloff-Coste and Tianyi Zheng, *Random walks and isoperimetric profiles under moment conditions*, in preparation.

[Sta84] Richard P. Stanley, *On the number of reduced decompositions of elements of Coxeter groups*, European J. Combin. **5** (1984), no. 4, 359–372.

[Var91] N. Th. Varopoulos, *Groups of superpolynomial growth*, Harmonic analysis (Sendai, 1990), ICM-90 Satell. Conf. Proc., Springer, Tokyo, 1991, pp. 194–200.