

LARGE DEVIATION PRINCIPLES FOR PATTERN-AVOIDING PERMUTATIONS, AND LIMIT SHAPES FOR CONSTRAINED MALLOWS PERMUTATIONS

THOMAS BUDZINSKI, VICTOR DUBACH, VALENTIN FÉRAY, MOHAMED SLIM KAMMOUN,
AND MYLÈNE MAÏDA

ABSTRACT. We study Mallows random permutations conditioned to avoid a given pattern α of length 3. When the bias parameter is of the form $e^{\beta/n}$, we prove that these permutations converge to a non-trivial explicit deterministic permuton that depends on the pattern α and on the parameter β . Along the way, we provide parametrizations for α -avoiding permutons, and establish a large deviation principle for uniform α -avoiding permutations. As a byproduct of the proof, we also obtain asymptotic estimates of two versions of q -Catalan numbers in the regime $q = e^{\beta/n}$.

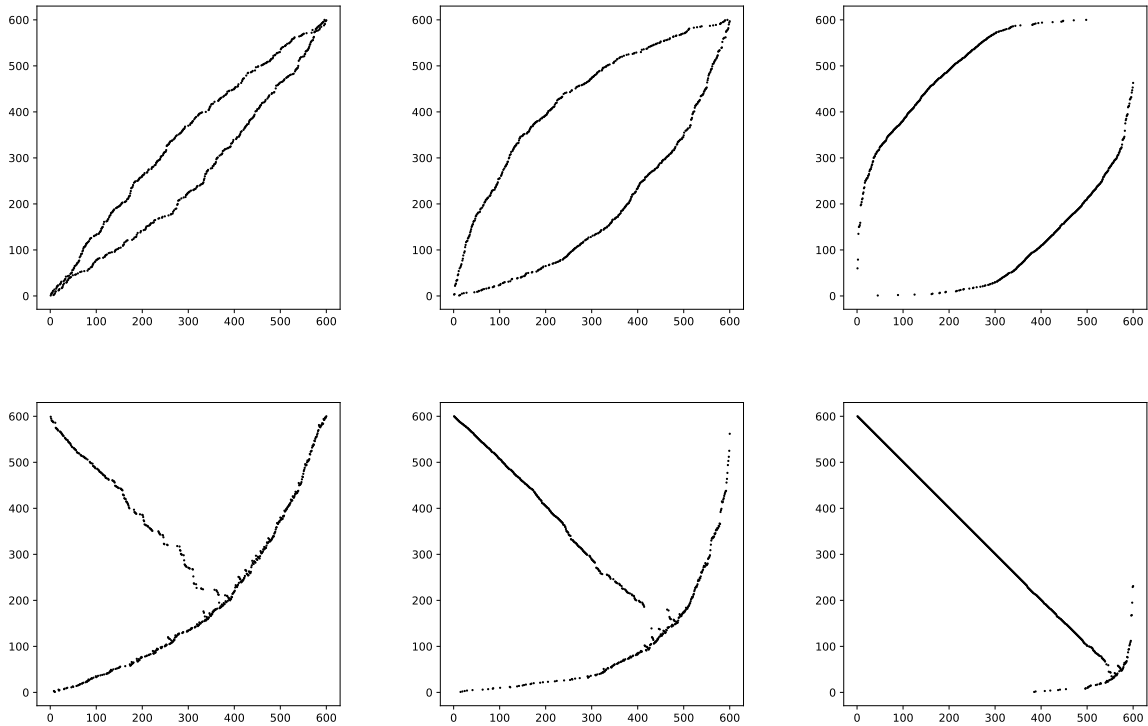


FIGURE 1. Top line: simulations of 321-avoiding Mallows permutations with size 600 and bias parameter $q = e^{\beta/n}$, for $\beta \in \{1, 3, 12\}$. Bottom line: likewise with 231-avoiding Mallows permutations. Here, and throughout the paper, a permutation σ is represented graphically by the set $\{(i, \sigma(i)), 1 \leq i \leq n\}$.

1991 *Mathematics Subject Classification.* 05A05, 05A19, 60C05, 60F10.

Key words and phrases. Random permutations, pattern-avoiding permutations, permutons, large deviation principles, Dyck paths.

1. INTRODUCTION

1.1. Overview. *Mallows/pattern-avoiding random permutations.* Random permutations are among the most standard objects studied in combinatorial theory, with notable applications in statistics, analysis of algorithms, and statistical physics. While earlier works have focused on uniform random permutations of a given size, some non-uniform models have become popular, in particular the so-called Mallows model, introducing a bias depending on the number of inversions of the permutations. This model has been introduced by Mallows in 1957 [Mal57] in the context of statistical rankings, and has been studied under many aspects in the last twenty years, see e.g. [Dub26] for references. In another direction, a number of recent research works analyze uniform random permutations subject to some constraints. The most common constraints consist in fixing the cycle type (see e.g. [FK24, Dub25, HR25]), or in avoiding some patterns. The latter line of research originates from articles of Madras, Pak and co-authors and has grown significantly in recent years; see e.g. [ML10, MP14, HRS19, BBF⁺22, Bor22].

Our first motivation for this work is to combine both points of view, i.e., to study Mallows random permutations constrained to avoid some patterns. In this paper, we focus on avoiding a fixed pattern of length 3, and refer to this as *constrained Mallows permutations*. We focus on 321 (resp. 231)-avoiding permutations, since the study of α -avoiding permutations for other patterns α of size 3 can be reduced to one of these two cases by symmetry. We recall that a permutation σ avoids 321 if it does not contain any decreasing subsequence of length 3, and it avoids 231 if there are no indices $i < j < k$ such that $\sigma(k) < \sigma(i) < \sigma(j)$. Also, we denote by $\text{inv}(\sigma)$ the number of inversions of σ . For $\beta \in \mathbb{R}$, α in $\{231, 321\}$ and $n \geq 1$, we let $\tau_n^{\beta, \alpha}$ be a random permutation of size n such that, for any α -avoiding permutation σ :

$$(1) \quad \mathbb{P}(\tau_n^{\beta, \alpha} = \sigma) = \frac{e^{\frac{\beta}{n} \text{inv}(\sigma)}}{Z_n^{\beta, \alpha}},$$

where $Z_n^{\beta, \alpha} := \sum_{\sigma} e^{\frac{\beta}{n} \text{inv}(\sigma)}$ is the appropriate normalization factor (the sum is taken over all α -avoiding permutation σ). This is the Mallows distribution with parameter $q_n = e^{\beta/n}$ conditioned to avoid α . The reason why we chose this scaling for q_n will be explained after the statements of the main results. Figure 1 shows samples of $\tau_n^{\beta, 231}$ and $\tau_n^{\beta, 321}$ which were obtained by simulating an appropriate Markov chain, as explained in Appendix A. Upon looking at these simulations, one might guess that a non-trivial permuton limit exists for this model; see Section 1.4 for precise statements to that regard.

As far as we are aware of, this question has not yet been investigated, although related models and questions have appeared in the literature. First, let us mention the work [GK25] of Gorin and Kenyon, motivated by the six vertex model of statistical physics, who analyzed random Mallows permutations with a different type of constraints, namely that the diagram of the permutation avoids some designated regions. In another direction, Chelikavada and Pranzo [CP26], and more recently Park and Rizzolo [PR25], studied pattern-avoiding permutations biased by their number of fixed points. Lastly, in [Pin21], Pinsky gives a number of results regarding the probability that a Mallows random permutations avoids a pattern of size 3, but he does not study the resulting conditional model as we do here.

Permutons, and large deviation principles. Among many aspects of random Mallows permutations avoiding some patterns, we focus here on finding the scaling limit of such objects, in the sense of permutons (see Section 2.2 for background on permutons). A good tool for this are large deviation principles (LDPs), which are useful when looking for the limit shapes of random objects distributed according to biased distributions, such as random Mallows permutations.

It indeed transforms the probabilistic problem into some optimization problem on a space of analytic objects (in our case, permutons). See [Sta09, Muk16, KKRW20] for applications of this principle to some random permutation models, in particular to (unconditioned) Mallows random permutations. The space of permutons, which are probability measures on $[0, 1]^2$, is known to be a natural framework to state an LDP for uniform random permutations [Tra08, KKRW20]. However, this “classical” LDP cannot be used to study models of pattern-avoiding permutations, since the rate function is equal to ∞ on all permutons avoiding a given pattern¹.

Our contributions. We overcome this problem by deriving specific LDPs for uniform 321 (resp. 231)-avoiding permutations. A first step in this direction is to find appropriate parametrizations of 321 (resp. 231)-avoiding permutons. Once we have found these parametrizations and established the associated LDPs, we can translate the problem of finding the scaling limit of 321 (resp. 231)-avoiding Mallows permutations to an analytic optimization. It turns out that, in both cases, this optimization problem can be solved explicitly, yielding an explicit description of the limit shapes. In addition to the limit shape results, solving these optimization problems also yields asymptotic estimates for the partition functions $Z_n^{\beta, \alpha}$ (both for $\alpha = 321$ and 231). This complements previous results of Pinsky in the regime where the bias parameter q is constant [Pin21] (recall that for us, $q = \exp(\beta/n)$).

In the remaining part of the introduction, we provide more details on these three steps (parametrization of pattern-avoiding permutations, LDPs, and limit shapes for constrained Mallows permutations), which we believe are all interesting in their own right.

Remark 1. The normalization constant $Z_n^{\beta, \alpha}$ appearing in (1) is the generating polynomial of inversions for 231, resp. 321, avoiding permutations. In the 231-avoiding case, the classical decomposition of 231-avoiding permutations around their maxima yields the following recurrence

$$Z_{n+1}^{\beta, 231} = \sum_{i=0}^n q^i Z_i^{\beta, 231} Z_{n-i}^{\beta, 231},$$

where $q = e^{\beta/n}$ as usual. This is the same recurrence as the q -Catalan numbers defined as the area generating function for Dyck paths; see, e.g., [Hag08, Chapter 1] (see also Proposition 24 for a direct connection between inversions in 231-avoiding permutations and Dyck path area). Moving on now to $\alpha = 321$, the quantity $Z_n^{\beta, 321}$ also satisfies some recurrence equation, though this is harder to prove, see [CEKS13]. We shall not need these recurrences in this paper.

1.2. First main results: parametrizations of 231 or 321-avoiding permutons. We recall that a permuton is a probability measure on the square $[0, 1]^2$ with both marginals equal to the Lebesgue measure on $[0, 1]$. We refer to Section 2.2 below for related precise definitions. While there are many works studying the structure of the set of α -avoiding permutations (see, e.g., [Vat15]), it seems that the question of describing α -avoiding permutons has not been looked up so far. We open the way here by giving natural parametrizations of 231 (resp. 321)-avoiding permutons. Given a pattern α and a permuton μ , we say that μ is α -avoiding if the density of α in μ is equal to 0. Equivalently, μ is α -avoiding if and only if it is the limit of a sequence of α -avoiding permutations.

Before stating our results, let us mention some related literature. In the realm of graphs, a series of papers find natural parametrizations of graphons that are limits of graphs in some specific classes, namely threshold graphs [DHJ08], interval graphs [DHJ13], and string graphs [JU17].

¹This should not come as a surprise, since the number of permutations avoiding a given pattern behaves as $O(C^n)/n!$ by the Stanley–Wilf–Marcus–Tardős theorem (see, e.g., [Bón12, Chapter 4]).

In some sense, we are looking at the analogue problems for permutations. In another direction, in a recent paper [GHKP24], Garbe, Hladký, Kun and Pekárková have looked at some general properties of pattern-avoiding permutations, but were not interested in a complete parametrization of them as we are. Finally, a number of papers consider permuton limits of uniform random permutations avoiding some fixed patterns (see, e.g., [BBF⁺22, Bor22, HS25] and references in these papers). While these limits necessarily are pattern-avoiding permutons, this is of course a different question than trying to describe the set of *all* permutons avoiding given patterns.

We first state our parametrization for 231-avoiding permutons, as it may seem more natural to readers familiar with pattern-avoiding permutations. Indeed, it is standard that 231-avoiding permutations are entirely determined by the positions and values of their right-to-left minima (abbreviated “RLM” from here on out); see, e.g., [Bón12, Chapter 4]. Let us define the RLM curve of a permuton μ by

$$f_\mu : x \in [0, 1] \mapsto \max\{y \in [0, 1] : \mu([x, 1] \times [0, y]) = 0\}.$$

Necessarily, f_μ is a càdlàg nondecreasing function with $f_\mu(x) \leq x$ and $f_\mu(1) = 1$. From there, it will be convenient to define φ_μ as the function whose graph is obtained by rotating the graph of f_μ by $-\pi/4$, flipping it and scaling it down by a factor $\sqrt{2}$ and call it the RLM-excursion of the permuton μ (see Figure 2, left). The function φ_μ belongs to the space of *excursions*, defined as

$$(2) \quad \mathcal{E} := \{\varphi : [0, 1] \rightarrow \mathbb{R}^+ : \varphi \text{ is 1-Lipschitz and } \varphi(0) = \varphi(1) = 0\}.$$

Our first main result for 231 is the following:

Proposition 2. *The map sending a 231-avoiding permuton to its RLM-excursion is one-to-one and bicontinuous.*

We prove this result in Section 6 (see Lemmas 26 and 30) by constructing explicitly the inverse map, i.e., by explaining how to reconstruct a 231-avoiding permuton given its RLM-curve. Interestingly, this reconstruction procedure does not mimic the standard procedure to reconstruct 231-avoiding permutations from their RLM. Indeed, the latter is recursive and difficult to imitate in the continuous setting. Instead we provide a direct formula for the cumulative distribution function of the permuton, see Lemma 25.

We now turn our attention to the pattern 321. Again, it is well-known that 321-avoiding permutations are determined by the positions and values of their RLM (see e.g. [Bón12, Chapter 4]). However, it is not the case for permutons: it turns out that there are multiple 321-avoiding permutons with the same RLM-curve, as illustrated by Figure 5, and that the map sending a 321-permuton to its RLM-curve is not continuous. This prevents from encoding the set of 321-avoiding permutons by the set \mathcal{E} of (2).

Instead, we will show that 321-avoiding permutons can be parametrized by pairs (π_1, π_2) of measures living on the space

$$(3) \quad \mathcal{D} := \left\{ (\pi_1, \pi_2) \in \mathcal{M}([0, 1])^2 : \begin{array}{l} \pi_1 \leq \text{Leb}, \pi_2 \leq \text{Leb}, \pi_1([0, 1]) = \pi_2([0, 1]) \\ \text{and } \forall x \in [0, 1], \pi_1([0, x]) \leq \pi_2([0, x]) \end{array} \right\},$$

where $\mathcal{M}([0, 1])$ is the space of finite measures on $[0, 1]$, equipped with the topology of weak convergence. In other words, \mathcal{D} is the space of pairs of measures on $[0, 1]$ with the same total mass, both subuniform², and such that $\pi_1([0, x]) \leq \pi_2([0, x])$ for all x in $[0, 1]$. Informally, π_1 and π_2 are the projections on the horizontal and vertical axes of the subdiagonal part of the 321-avoiding permuton μ ; this is a continuous analogue of the counting measure of the positions and values of the right-to-left minima of a 321-avoiding permutation σ (see Figure 2, right).

²A measure π on $[0, 1]$ is said to be *subuniform* if, for any measurable $A \subseteq [0, 1]$, one has $\pi(A) \leq \text{Leb}(A)$.

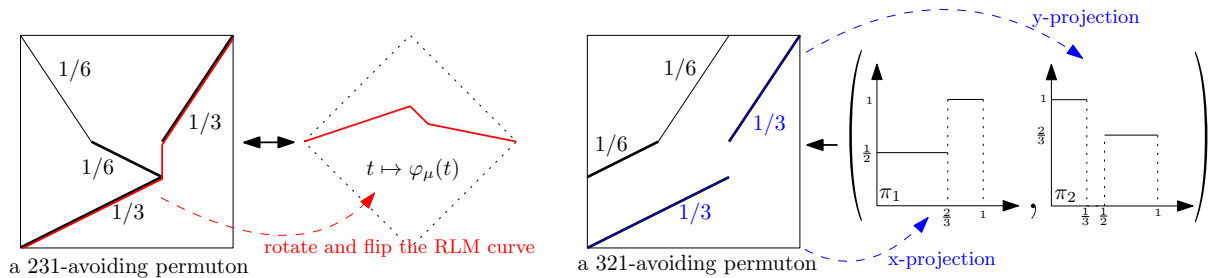


FIGURE 2. Illustration of the parametrizations of 231- and 321-avoiding permutations. 321-avoiding permutations are in bijection with excursions, obtained by rotating and flipping their RLM curves (in red on the figure). On the other hand, there is a continuous surjection from pairs of measures to 231-avoiding permutations (the measures π_1 and π_2 are here represented by the graphs of their densities); informally, these measures are the x - and y -projections of the lower-diagonal part of the permuton (in blue on the figure).

We shall prove the following statement, which is a simplified version of Proposition 18.

Proposition 3. *There exists an explicit continuous surjection Ψ from \mathcal{D} to the set of 321-avoiding permutations.*

The description of Ψ can be found in Section 3. We note that, unlike in the 231-avoiding case, the parametrization by \mathcal{D} is not one-to-one. Nevertheless, we will provide an explicit description of the fibers $\Psi^{-1}(\pi_1, \pi_2)$ (Proposition 18), and show that they are reduced to singletons whenever μ does not put any mass on the main diagonal of the unit square.

The above parametrizations of 231- and 321-avoiding permutations are illustrated on Figure 2. For further discussion on the differences between the two parametrizations, we refer the reader to Section 1.5.

1.3. Second main results: large deviation principles for uniform 231 or 321-avoiding permutations. As explained in Section 1.1, an important step will be to establish LDPs for the random objects we are interested in, in the uniform case.

Roughly speaking, a sequence of random variables (X_n) taking values in a Polish space E satisfies an LDP at speed a_n with rate function I if for any $x \in E$, the probability $\mathbb{P}(X_n \simeq x)$ that X_n is close to x behaves as $\exp(-a_n I(x))$, where (a_n) is a sequence of real numbers going to infinity and I is a lower semicontinuous nonnegative function (see Section 2.1 for proper definitions).

An LDP for uniform random permutations in the space of permutations has been given and used in [Tra08, Muk16, KKRW20]. An LDP for Mallows random permutations follows directly by a standard result of large deviation theory, known as Varadhan's lemma; see, e.g., [AGZ10, Theorem D.8]. More recently, Borga, Das, Mukherjee and Winkler [BDMW24] have given an LDP for a large family of random permutations, obtained by sampling i.i.d. random points from a probability measure in the plane. In the present paper, we provide LDPs at speed n for uniform 231 (resp. 321)-avoiding permutations. These LDPs take place respectively in the spaces \mathcal{E} and \mathcal{D} parametrizing 231 and 321-avoiding permutations introduced in the previous sections, and involve explicit rate functions. An LDP for random Mallows permutations conditioned to avoid 231 (resp. 321) follows immediately (see Corollaries 19 and 31).

Let us first describe the LDP that will be the cornerstone for the analysis of Mallows permutations avoiding 231. In view of Proposition 2, it is natural to establish an LDP in the space

of 1-Lipschitz functions. For a permutation σ of size n , let us denote by f_σ its normalized RLM-curve, defined as follows:

$$f_\sigma(x) := \frac{1}{n} (\min\{\sigma(c) : c > nx\} - 1)$$

and by φ_σ the excursion obtained by rotating and flipping around the graph of f_σ as on Figure 7. By construction, φ_σ is a continuous function on $[0, 1]$, which is differentiable almost everywhere with $\|\varphi'\| \leq 1$. We endow the space \mathcal{E} defined in (2) with the topology of the supremum norm, which makes it a compact space. Our LDP for uniform 231-avoiding permutations reads as follows. For any $y \in [-1, 1]$, we write

$$(4) \quad J(y) := \frac{1}{2}(y+1)\log(y+1) + \frac{1}{2}(1-y)\log(1-y).$$

Proposition 4. *Let $H^{231} : \mathcal{E} \rightarrow [0, \infty]$ defined by*

$$H^{231}(\varphi) := 2 \int_{[0,1]} J(\varphi'(t)) dt.$$

For all $n \geq 1$, let $\sigma_{n,231}$ be a uniform random 231-avoiding permutation of size n . Then the random variables $(\varphi_{\sigma_{n,231}})$ satisfy a large deviation principle at speed n with rate function H^{231} .

We now state our result for 321. As explained above and illustrated in more details in Section 1.5 below, we cannot use the excursion φ_σ to characterize 321-avoiding permutons. At the level of LDPs, this will require establishing a different LDP in this case, although the underlying discrete objects are related to Dyck paths in both cases. For a 321-avoiding permutation σ , we let $(\pi_1^\sigma, \pi_2^\sigma)$ be the rescaled occupation measures of positions and values of its strict right-to-left minima, see Section 3.1 for a precise definition. The pair $(\pi_1^\sigma, \pi_2^\sigma)$ is an element of the space \mathcal{D} parametrizing 321-avoiding permutons, which we defined in (3). Then, the following holds.

Proposition 5. *Let $H^{321} : \mathcal{D} \rightarrow [0, \infty]$ be defined as follows:*

$$H^{321}(\pi_1, \pi_2) := \int_0^1 (\rho_1 \log \rho_1 + (1-\rho_1) \log(1-\rho_1) + \rho_2 \log \rho_2 + (1-\rho_2) \log(1-\rho_2) + 2 \log 2) d \text{Leb},$$

where ρ_1 and ρ_2 are the densities of π_1 and π_2 with respect to the Lebesgue measure³.

For all $n \geq 1$, let $\sigma_{n,321}$ be a random permutation chosen uniformly among 321-avoiding permutations of size n . Then, the associated pairs $(\pi_1^{\sigma_{n,321}}, \pi_2^{\sigma_{n,321}})_{n \geq 1}$ of measures satisfy a large deviation principle on \mathcal{D} at speed n with rate function H^{321} .

Note that in both cases, the rate functions H^{231} and H^{321} are explicit and simple, so that their analysis will be possible. In particular, H^{231} (resp. H^{321}) attains its unique minimum at $\varphi^* = 0$ (resp. $\pi_1^* = \pi_2^* = \frac{1}{2} \text{Leb}|_{[0,1]}$), corresponding in both cases to the diagonal permuton. We retrieve the already known convergence of the uniform 231 (resp. 321)-avoiding permutations to this diagonal permuton. The fluctuations around this limit have been described in [HRS17] and the LDPs that we provide here complement these results naturally.

Also, the decay rate of single points probabilities, namely, $\mathbb{P}(\sigma_n(i) = j)$ when i and j are both of order n , have been studied in depth; see [MP14, AM14]. Such results are neither weaker nor stronger than our LDPs. On the one hand, we look at deviation probabilities of the whole object rather than the image of a single element, so it might seem that our results contain more information. However, on the other hand, the topology on the space of permutons is too weak for the image of a specific element to be a continuous functional, so one can not deduce deviation of single point probabilities from our LDPs.

³These densities do exist since elements of \mathcal{D} are pairs of subuniform measures.

To conclude this section, let us say a few words of the proof. As pointed above, we do need to establish two LDPs respectively in the spaces \mathcal{E} and \mathcal{D} . In both cases, the underlying random objects are Dyck paths, that is random excursions conditioned to stay nonnegative. LDPs for random walks have been essentially known since the work of Mogulskii in the 70's [Mog76]. This involves a rate function which coincides with H^{231} when it is finite and we need to prove that the result can be transferred to the conditioned models. Note that LDPs for continuous-time processes conditioned to return near zero have been obtained in the literature by [MPY14, DGM24], but up to our knowledge, there are no known LDPs for Dyck paths. The scheme of proof is similar for Propositions 4 and 5, respectively. The upper bound is easy as it follows immediately from the LDP for the unconditioned model, and the fact that the conditioning event has a polynomially small probability (and not an exponentially small one). More work is needed on the lower bound: we need to perform small transformations on our objects that force the conditioning event to hold, while not changing the probability too much. This kind of transformation arguments is standard in the large deviation literature.

1.4. Third main results: explicit limit shapes and asymptotic partition functions for Mallows random permutations conditioned to avoid 321 or 231. We now come back to our original motivation of studying constrained Mallows permutations. Recall that for $\alpha \in \{321, 231\}$ and $n \geq 1$, we let $\tau_n^{\beta, \alpha}$ denote a Mallows permutation of size n with bias parameter $q_n = e^{\beta/n}$ conditioned to avoid α ; see (1).

Our main theorems say that the random permutations $\tau_n^{\beta, \alpha}$ converge to a deterministic permuton μ_β^α which depends on β . To describe the limit permutons, it is useful to use the language of push-forward measures. If $g : E \rightarrow F$ is a measurable function and μ a measure on E , then the push-forward measure $g_\# \mu$ is the measure defined by $g_\# \mu(A) = \mu(g^{-1}(A))$ for all measurable $A \subseteq F$. In the following, we implicitly use X as the argument of g , i.e. we write $g(X)$ for the function $x \mapsto g(x)$. Lastly, $g_\#$ is seen as an operator on measures so that $(g_\# + h_\#)\mu$ is simply equal to $g_\# \mu + h_\# \mu$. In the results below, almost sure convergences hold for any coupling between the objects $\tau_n^{\beta, \alpha}$ for $n \geq 1$.

Theorem 6. *For $\beta > 0$, we let*

$$f_\beta^{231}(x) := 1 - \frac{1}{\beta} \log \left(1 + e^\beta - e^{\beta x} \right).$$

Then, the random permutations $\tau_n^{\beta, 231}$ converge almost surely in the sense of permutons to the deterministic permuton

$$\mu_\beta^{231} := (X, f_\beta^{231}(X))_\# \min(1, (f_\beta^{231})'(x)) dx + (X, 1 - X)_\# \max(0, 1 - (f_\beta^{231})'(x)) dx$$

if $\beta > 0$, and to the diagonal permuton if $\beta \leq 0$.

Theorem 7. *For $\beta > 0$, we let*

$$f_\beta^{321}(x) := \frac{1}{2} + \frac{1}{\beta} \log \left(\frac{e^{\beta x} + e^{\beta/2}}{1 + e^{\beta/2} + e^\beta - e^{\beta x}} \right).$$

Then, the random permutations $\tau_n^{\beta, 321}$ converge almost surely in the sense of permutons to the deterministic permuton

$$\mu_\beta^{321} := ((X, f_\beta^{321}(X))_\# + (f_\beta^{321}(X), X)_\#) \frac{dx}{1 + \exp(\beta(\frac{1}{2} - x))}$$

if $\beta > 0$, and to the diagonal permuton if $\beta \leq 0$.

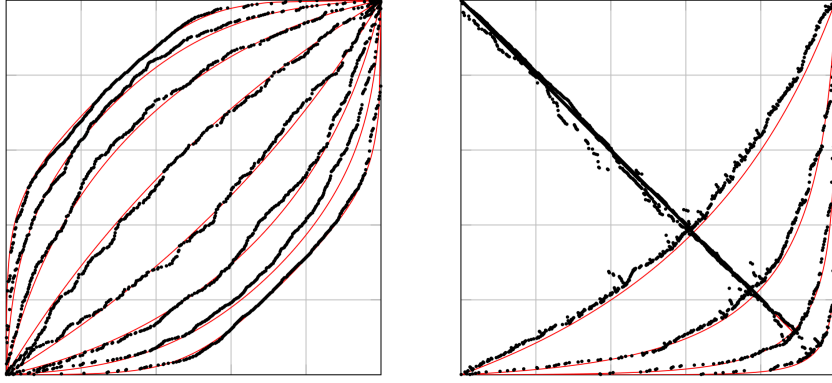


FIGURE 3. Black dots: simulations of $\tau_n^{\beta,\alpha}$ with $n = 600$ and $\beta \in \{1, 3, 6, 12\}$. Red curves: support of the limit permutons μ_β^α , again for $\beta \in \{1, 3, 6, 12\}$. Left: $\alpha = 321$. Right: $\alpha = 231$.

In particular, we see that μ_β^{231} is supported on the union of an increasing curve and a part of the decreasing diagonal, namely $\{(x, 1-x) : 0 \leq x \leq x_\beta^*\}$, where x_β^* fulfills $(f_\beta^{231})'(x_\beta^*) = 1$ (explicitly, $x_\beta^* = (\log(1+e^\beta) - \log(2))/\beta$). On the other hand, μ_β^{321} is supported on two increasing curves, symmetric of each other. Figure 3 shows realizations of $\tau_n^{\beta,321}$ and $\tau_n^{\beta,231}$ for various values of β , together with the support of their limit permutons.

To finish this section, let us briefly discuss the proof strategy. These results are obtained by using our LDPs for uniform α -avoiding permutations, and then solving optimization problems. Informally speaking, our LDPs state that the probability that a uniform α -avoiding permutation is close to a point x in the permuton space is close to $\exp(-nI^\alpha(x))$. Since all these permutations have roughly $\frac{1}{2}n^2 \text{dens}(21, x)$ inversions (where $\text{dens}(21, x)$ is the density of pattern 21 in x , see Section 2.2), the probability that the Mallows random permutation $\tau_n^{\beta,\alpha}$ is close to x is roughly proportional to

$$q_n^{n^2 \text{dens}(21, x)/2} \exp(-nI^\alpha(x)) = e^{n(\beta \text{dens}(21, x)/2 - I^\alpha(x))}.$$

The limit permuton will be the one that makes this quantity as large as possible, i.e., that maximizes $\beta \text{dens}(21, x)/2 - I^\alpha(x)$. Note that these heuristics explain why $q_n = e^{\beta/n}$ was the right order of magnitude to have a nontrivial permuton limit. Our proofs follow this heuristic, except that we formulate and solve the optimization problems in the parameter spaces (namely, on excursions in the 231-avoiding case and on pairs of measures in the 321-avoiding case). Remarkably, the density of inversions has a very simple expression in terms of these parameters, which eases the analysis. In particular, the functions that we need to maximize are strictly convex, and finding the optimum is a standard computation (even though the objects live in an infinite-dimensional space). This demonstrates the applicability of our LDPs.

As mentioned earlier, solving the above optimization problems also yield asymptotic estimates for the underlying partition functions. We now state these results.

Proposition 8. *For any $\beta > 0$, we have*

$$(5) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log(Z_n^{\beta,231}) = 2 \log(1 + e^\beta) - \beta - \frac{1}{\beta} \int_{e^{-\beta}}^{e^\beta} \frac{\log y}{1+y} dy;$$

$$(6) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log(Z_n^{\beta,321}) = \frac{2}{\beta} \int_{e^{-\beta/2}}^{e^{\beta/2}} \frac{\log(1+y)}{y} dy.$$

For $\beta \leq 0$, we have $\lim_{n \rightarrow \infty} \frac{1}{n} \log(Z_n^{\beta,231}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(Z_n^{\beta,321}) = \log 4$.

We conclude this section with several comments.

Remark 9. The last claim of Proposition 8, in the case $\beta \leq 0$, is easy to prove. Indeed, let $C_n = \frac{1}{n+1} \binom{2n}{n}$ be the n -th Catalan number. Since most of the C_n many α -avoiding permutations of size n have $o(n^2)$ inversions, we have

$$C_n (1 - o(1)) e^{-|\beta|o(n)} \leq Z_n^{\beta, 231} \leq C_n,$$

implying $\lim_{n \rightarrow \infty} \frac{1}{n} \log(Z_n^{\beta, 321}) = \log 4$.

Remark 10. When $\beta \rightarrow 0$, the permutons μ_β^α converge to the diagonal permuton which is the limit of uniform α -avoiding permutations as recalled above. On the other hand, when $\beta \rightarrow \infty$, the permutons μ_β^{231} and μ_β^{321} converge respectively to the anti-diagonal permuton (in the 231-avoiding case), and to the union of two increasing segments from $(0, 1/2)$ to $(1/2, 1)$ and from $(1/2, 0)$ to $(1, 1/2)$ (in the 321-avoiding case). In both cases, it is easy to see that these are the α -avoiding permutons maximizing the density of inversions, as expected.

Remark 11. Our results extend to all other patterns of size 3 by symmetry. For instance, 123-avoiding permutations are in one-to-one correspondance with 321-avoiding permutations via $\sigma \mapsto \sigma^{\text{rev}} := \sigma(n), \dots, \sigma(1)$. Through this map, right-to-left minima become left-to-right minima and inversions are swapped with non-inversions. Furthermore, the random permutations $\tau_n^{\beta, 123}$ converge to the permuton μ_β^{123} obtained by flipping the permuton μ_β^{321} horizontally. In a similar way, 213-avoiding permutations are in one-to-one correspondance with 231-avoiding permutations via $\sigma \mapsto \sigma^{\text{comp}} := n+1 - \sigma(1), \dots, n+1 - \sigma(n)$, and thus, the random permutations $\tau_n^{\beta, 213}$ converge to the permuton μ_β^{213} obtained by flipping the permuton μ_β^{231} vertically. That way, we get limit shapes for $\tau_n^{\beta, \alpha}$ for all patterns α of size 3.

1.5. Comparing 231 and 321-avoiding permutations with given RLM. In this section, we provide a few examples illustrating the differences between the 321-avoiding and 231-avoiding cases.

In our first family of examples, we consider two sets of right-to-left minima which differ by a single point. The associated pairs (π_1, π_2) of measures are then very close to each other, but the difference in the sets of right-to-left minima have been chosen so that the associated RLM curves are far from each other. In this case, the corresponding 321-avoiding permutations converge to the same permuton, while the corresponding 231-avoiding have different permuton limits — see Figure 4.

Our second family of examples illustrates the opposite situation. Here, we choose two sets of right-to-left minima such that the associated RLM curves are close to each other, but the associated pairs of measures are very different (because one contains twice as many RLMs as the other). In this case, the corresponding 231-avoiding permutations converge to the same permuton, but not the corresponding 321-avoiding permutations — see Figure 5.

These two families of examples illustrate that the RLM curve is a continuous parametrization for 231-avoiding permutons, while the description via pair of measures is more appropriate for 321-avoiding objects.

1.6. Outline of the paper. We end this introduction with a brief description of the organization of the paper. In the next section, we gather some notations and preliminaries on the various objects (LDPs, permutons, etc) that we use. The rest of the paper, from Section 3 on, will be split in two parts, respectively devoted to the study of 321 and of 231-avoiding permutations. Unlike in the introduction, in the core of the paper, we shall start with the 321-avoiding case, since the proof of our parametrization result for α -avoiding permutation is more involved in the 231-avoiding permutation.

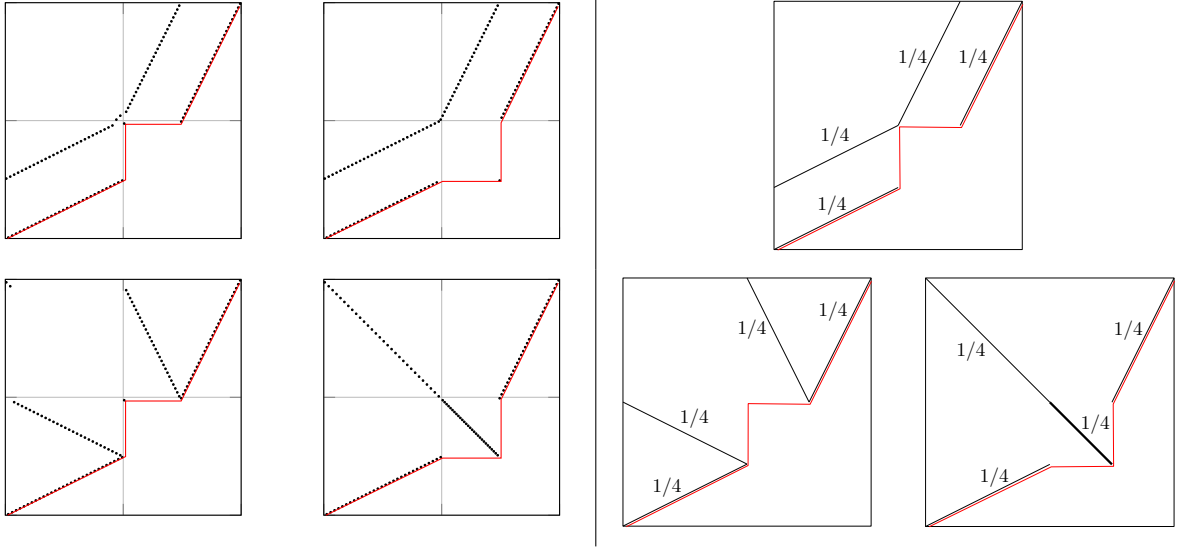


FIGURE 4. Examples of α -avoiding permutations of size 120 with given right-to-left minima on the left, and their limit permutons on the right. In red, we plot the RLM curves of the permutations/permutons (slightly shifted so that they do not hide the points). In the top line $\alpha = 321$, while in the bottom line $\alpha = 231$. First column: RLMs are the points $(2i, i)$ for $i < \frac{1}{4}n$, $(\frac{1}{2}n, \frac{1}{2}n)$, and $(n-i, n-2i)$ for $i < \frac{1}{4}n$. Second column: we have the same RLMs except that $(\frac{1}{2}n, \frac{1}{2}n)$ is replaced by $(\frac{3}{4}n, \frac{1}{4}n)$

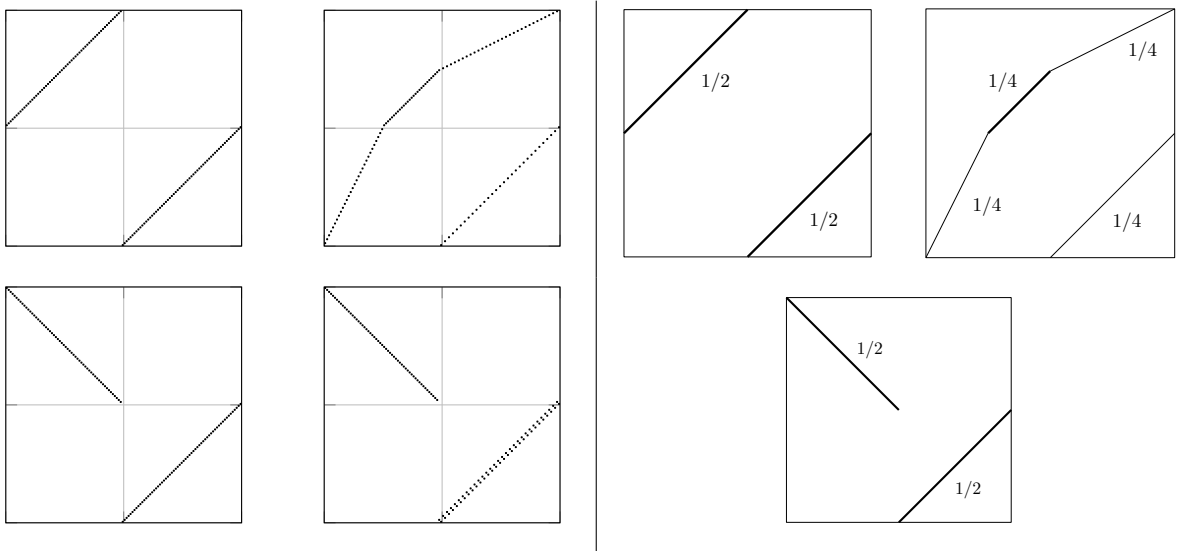


FIGURE 5. Further examples of α -avoiding permutations of size 120 with given right-to-left minima on the left, and their limit permutons on the right. In the top line $\alpha = 321$, while in the bottom line $\alpha = 231$. First column: RLMs are the points $(\frac{1}{2}n + i, i)$ for $i \leq \frac{1}{2}n$. Second column: we keep only half of the RLMs, namely the RLMs are the points $(\frac{1}{2}n + 2i, 2i)$ for $i \leq \frac{1}{4}n$.

2. NOTATION AND PRELIMINARIES

For a compact Polish space X equipped with its Borel σ -algebra, we let $\mathcal{M}(X)$ be the space of measures with finite mass on X , endowed with the topology of weak convergence.

For two measures μ_1 and μ_2 , we write $\mu_1 \leq \mu_2$ to mean that $\mu_1(A) \leq \mu_2(A)$ for any Borel set A . Also, we write $f_{\#}\mu$ for the push-forward measure of μ by f , defined by $f_{\#}\mu(A) = \mu(f^{-1}(A))$ for any Borel set A .

2.1. Large deviation principles. In this subsection, we give a short introduction to large deviation principles (LDP). This is standard material. We refer the reader to Appendix D in [AGZ10] for a concise introduction of useful tools on LDPs, and to [DZ98] for a thorough reference.

Let E be a Polish space, (X_n) a sequence of E -valued random variables, and (a_n) a sequence of positive real numbers tending to ∞ . Let I be a nonnegative lower semicontinuous function on E , possibly taking value ∞ . We say that the sequence (X_n) satisfies an LDP⁴ at speed a_n with rate function I if, for any closed set $F \subseteq E$, one has

$$(7) \quad \limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in F) \leq - \inf_{x \in F} I(x),$$

and for any open set $O \subseteq E$, one has

$$(8) \quad \liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in O) \geq - \inf_{x \in O} I(x).$$

Moreover, we say that the sequence (X_n) satisfies a weak LDP if (8) holds for open sets and if the upper bound (7) holds for compact sets [AGZ10, Definition D.2]. Note that if E is a compact set, weak and full LDP are equivalent. If moreover d is a metric on E (giving the required topology), and $B(x, \delta) := \{y \in E : d(x, y) \leq \delta\}$, then weak LDPs follow from the two following bounds (see [AGZ10, Corollary D.6]):

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in B(x, \delta)) \leq -I(x),$$

and

$$\lim_{\delta \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in B(x, \delta)) \geq -I(x).$$

The following property, known as the contraction principle, is a useful tool to transfer LDPs, see e.g. [DZ98, Theorem 4.2.1].

Proposition 12. *Assume that (X_n) satisfies an LDP at speed a_n with good⁵ rate function I on some space E . Let $F : E \rightarrow G$ be a bounded continuous function. Then, the sequence of random variables $Y_n = F(X_n)$ satisfies an LDP at speed a_n with good rate function J given, for any $y \in G$, by*

$$J(y) = \inf\{I(x) : F(x) = y\}.$$

We will also use the following lemma, that can be seen as a corollary of Varadhan's lemma (see [DZ98, Theorem 4.3.1] and [AGZ10, Theorem D.8]).

⁴In the standard references [DZ98] and [AGZ10], LDPs are defined for sequences of probability measures and not of random variables. However, as often in probability theory, we shall identify random variables with their distribution, and say that a sequence of random variables satisfies an LDP when their distributions do.

⁵A rate function is said to be *good* if its level sets $\{x \in E : I(x) \leq M\}$ are compact.

Lemma 13. For any $n \geq 1$, let μ_n denote the distribution of X_n and assume that (X_n) satisfies an LDP at speed a_n with good rate function I . Let $F : E \rightarrow \mathbb{R}$ be a bounded continuous function. Let the distribution ν_n be defined as follows:

$$d\nu_n(x) := \frac{1}{Z_n} e^{a_n F(x)} d\mu_n(x),$$

where Z_n is a normalizing constant such that ν_n is a probability measure. Then we have

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \log Z_n = - \inf_{y \in E} (I(y) - F(y)).$$

Moreover, if we let Y_n have distribution ν_n , the sequence (Y_n) satisfies an LDP at speed a_n with good rate function J given by

$$\forall x \in E, \quad J(x) = I(x) - F(x) - \inf_{y \in E} (I(y) - F(y)).$$

The last point explains how to deduce almost sure convergence from an LDP. It is an easy consequence of the Borel–Cantelli lemma.

Lemma 14. Assume that (X_n) satisfies an LDP at speed n with good rate function I and that I has a unique minimizer x^* . Then (X_n) converges almost surely to x^* .

2.2. Permutons. Let \mathcal{P} be the set of permutons, that is, of probability measures μ on $[0, 1]^2$ satisfying

$$\mu([0, x] \times [0, 1]) = \mu([0, 1] \times [0, x]) = x$$

for every x in $[0, 1]$ (i.e., μ has uniform marginals). We endow the set of permutons with the topology of weak convergence. If σ is a permutation of size n , we associate with it a permuton μ_σ which attributes mass $1/n$ uniformly to each cell

$$C_i^\sigma := [(i-1)/n, i/n] \times [(\sigma(i)-1)/n, \sigma(i)/n], \quad 1 \leq i \leq n.$$

We say that a sequence (σ_n) of permutations, converges to a permuton μ if μ_{σ_n} converges to μ weakly. It is sometimes useful to define slight variations of μ_σ : we let μ_σ^\swarrow (resp. μ_σ^\searrow) be the permuton which attributes mass $1/n$ uniformly on the *diagonal* (resp. *antidiagonal*) of each cell C_i^σ . Since the Kolmogorov–Smirnov distances between these three permutons are bounded by $\mathcal{O}(1/n)$, convergence results and large deviation principles hold for all three simultaneously.

If σ is a permutation of size n and α a permutation of size $k \leq n$, the number of occurrences of α in σ is

$$\text{occ}(\alpha, \sigma) := \left| \left\{ (i_1, \dots, i_k) : 1 \leq i_1 < \dots < i_k \leq n \text{ and } \sigma(i_{\alpha^{-1}(1)}) < \dots < \sigma(i_{\alpha^{-1}(k)}) \right\} \right|$$

and the density of the pattern α in σ is

$$\text{dens}(\alpha, \sigma) := \binom{n}{k}^{-1} \text{occ}(\alpha, \sigma).$$

We say that the permutation σ is α -avoiding if $\text{dens}(\alpha, \sigma) = 0$.

Similarly, if α is again a permutation of given size k , we define the density of the pattern α in a permuton μ as

$$\text{dens}(\alpha, \mu) := \int k! \mathbf{1}_{x_1 < \dots < x_k} \mathbf{1}_{y_{\alpha^{-1}(1)} < \dots < y_{\alpha^{-1}(k)}} d\mu(x_1, y_1) \dots d\mu(x_k, y_k),$$

and we say that the permuton μ is α -avoiding if $\text{dens}(\alpha, \mu) = 0$. Note that if μ_σ is the permuton associated to the permutation σ , we have in general $\text{dens}(\alpha, \sigma) \neq \text{dens}(\alpha, \mu_\sigma)$. Nonetheless, it is easy to see that $\text{dens}(321, \sigma) = \text{dens}(321, \mu_\sigma^\swarrow)$ and that $\text{dens}(231, \sigma) = \text{dens}(231, \mu_\sigma^\searrow)$.

In particular, a permutation σ is 321-avoiding if and only if μ_σ^\swarrow is 321-avoiding, and σ is 231-avoiding if and only if μ_σ^\searrow is 231-avoiding. Moreover, we know from [HKM⁺13] that the weak

convergence of permutons is equivalent to the convergence of all pattern densities. Standard arguments show that a permuton μ is α -avoiding if and only if there exists a sequence (σ_n) of α -avoiding permutations which converges to μ .

2.3. Nondecreasing subsets and coupling. A subset S of \mathbb{R}^2 will be called nondecreasing if one has $(x_1 - x_2)(y_1 - y_2) \geq 0$ for any (x_1, y_1) and (x_2, y_2) in S . A coupling of two real-valued random variables X and Y is called nondecreasing if the support of (X, Y) is nondecreasing. Given distributions μ and ν on \mathbb{R} , there is a unique nondecreasing coupling (X, Y) of X with law μ and Y with law ν . Let us denote by $\mu \nearrow \nu$ the corresponding distribution of (X, Y) .

More precisely, if we define

$$(9) \quad G_\mu : u \in (0, 1) \mapsto \sup\{x \in \mathbb{R} : \mu([0, x]) \leq u\}$$

the quantile function of μ and similarly G_ν the quantile function of ν , we can write

$$\mu \nearrow \nu = (G_\mu, G_\nu)_\# \text{Leb}_{[0,1]}.$$

We extend the notation to finite measures with the same total mass M by setting

$$\mu \nearrow \nu := M \cdot \left(\left(\frac{\mu}{M} \right) \nearrow \left(\frac{\nu}{M} \right) \right).$$

Finally if μ is a measure on $[0, 1]^2$, we denote by $\text{proj}(\mu)$ the pair (μ_1, μ_2) , where μ_1 and μ_2 are the projections of μ on the horizontal and vertical axes, respectively. By construction it holds that $\text{proj}(\mu_1 \nearrow \mu_2) = (\mu_1, \mu_2)$ and conversely if $\text{proj}(\pi) = (\mu_1, \mu_2)$ and if π has a nondecreasing support, then necessarily $\pi = \mu_1 \nearrow \mu_2$.

Part 1. Avoiding 321

The goal of this part is to study the limit shape of 321-avoiding random Mallows permutations with parameter $q_n = e^{\frac{\beta}{n}}$. As explained in the introduction, we will successively establish Proposition 3, Proposition 5, Equation 6 and Theorem 7. To ease the notations, we will drop the superscript 321 in all the notation in this part.

3. PERMUTATIONS AND PERMUTONS AVOIDING 321

3.1. A classical bijection for finite objects. We say that a pair (a, b) is a strict right-to-left (RL) minimum of a permutation σ if $b = \sigma(a) < a$ and for any $c > a$, we have $\sigma(c) > b$. In the literature, RL minimum sometimes refers to the position a , or to the value b ; here we chose to call RL minimum the pair (a, b) . In 321-avoiding permutations, a pair (a, b) is a strict RL minimum if and only if $b = \sigma(a) < a$.

For any permutation σ of size n , we denote by $((a_1, b_1), \dots, (a_k, b_k))$ the set of its RL minima in increasing order, that is

$$(10) \quad 1 < a_1 < \dots < a_k \leq n, \quad 1 \leq b_1 < \dots < b_k < n, \quad \text{and } a_i > b_i \text{ for all } 1 \leq i \leq k.$$

We also define

$$\text{Dyck}_n := \{(A, B) \in \mathcal{P}_n^2 : |A| = |B| \text{ and } \forall i \leq |A|, a_i > b_i\},$$

where \mathcal{P}_n is the set of all subsets of $[n] = \{1, \dots, n\}$ and a_i (resp. b_i) is the i th smallest element of A (resp. B). Note that the choice of the notation Dyck_n is related to the fact that this set is also in bijection with Dyck paths with $2n$ steps.

The following lemma holds, see e.g. [Cal07, Section 4].

Lemma 15. *The application $\sigma \mapsto ((a_1, b_1), \dots, (a_k, b_k))$ which maps a permutation σ to its list of strict RL minima in increasing order is a one-to-one correspondence between the set of 321-avoiding permutations of size n and the set Dyck_n .*

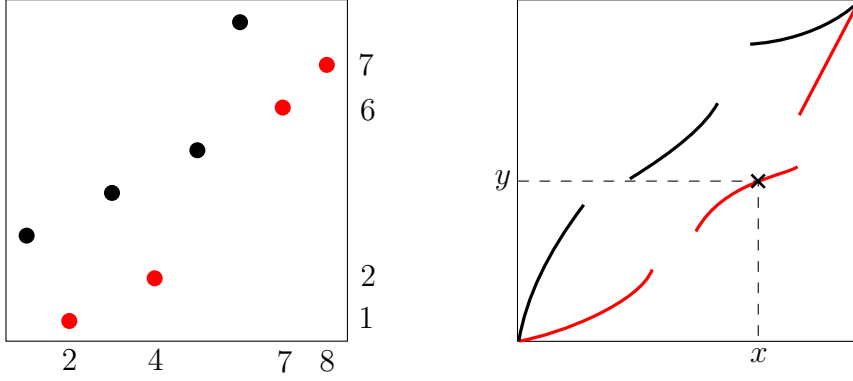


FIGURE 6. Left: a 321-avoiding permutation σ , with its strict RL minima in red. The corresponding list encoding is $((2, 1), (4, 2), (7, 6), (8, 7))$. Note that the element $(5, 5)$ is a weak RL minimum (in the sense that it is a fixed point of the permutation) and does not appear in the list. Right: the support of a 321-avoiding permutation $\mu = \Psi(\pi_1, \pi_2)$. The support of $\pi_1 \nearrow \pi_2$ is in red. A point (x, y) on the red curve should satisfy $\pi_1([0, x]) = \pi_2([0, y])$. For $(\pi_1, \pi_2) \in \mathcal{D}$, we assume $\pi_1([0, x]) \leq \pi_2([0, x])$, which forces $x \geq y$, i.e. the support of $\pi_1 \nearrow \pi_2$ lies in the closed lower-right triangle $\{(x, y) : x \geq y\}$.

Indeed, given the positions and values of the strict RL minima, which need to form an increasing subsequence below the diagonal, one can complete the permutation in a unique way by a second increasing subsequence weakly above the diagonal. See the left part of Figure 6.

For our purposes, it will be useful to encode these two lists as two measures, namely, if $((a_1, b_1), \dots, (a_k, b_k))$ is the list of RL minima of the permutation σ of size n , we define

$$\pi_1^\sigma := \sum_{j=1}^k \text{Leb} \left[\frac{a_{j-1}}{n}, \frac{a_j}{n} \right], \quad \pi_2^\sigma := \sum_{j=1}^k \text{Leb} \left[\frac{b_{j-1}}{n}, \frac{b_j}{n} \right].$$

This encoding turns out to be handy to compute the number of inversions of the permutation. In the following, we write $\text{inv}(\sigma)$ for the number of inversions of a permutation σ .

Lemma 16. *Let σ be a 321-avoiding permutation, and $((a_1, b_1), \dots, (a_k, b_k))$ be its list of strict RL minima. Let also π_1^σ and π_2^σ be the associated measures on $[0, 1]$, and ρ_1^σ and ρ_2^σ be their densities. We have:*

$$\text{inv}(\sigma) = \sum_{i=1}^k (a_i - b_i) = n^2 \int_0^1 x (\rho_1^\sigma(x) - \rho_2^\sigma(x)) dx.$$

Proof. Since σ is 321-avoiding, the elements in σ that are not RL minima necessarily form an increasing subsequence. It follows that, in each inversion of σ , the lower right element is an RL minimum. We claim that the number of inversions whose lower right element is a given RL minimum (a_i, b_i) is $(a_i - b_i)$. Indeed, since (a_i, b_i) is an RL minimum, the number of elements above and on the left of it is equal to the number $a_i - 1$ of elements on its left, minus the number $b_i - 1$ of elements below it. Hence the total number of inversions in σ is $\sum (a_i - b_i)$, where the sum is taken over RL minima of σ , or equivalently over strict RL minima (since fixed points contribute 0 to the sum). This justifies the first equality.

For the second equality, we simply observe that

$$\int_0^1 x \rho_1^\sigma(x) dx = \int_0^1 x \pi_1^\sigma(dx) = \frac{1}{n} \sum_{i=1}^k \frac{a_i - 1/2}{n},$$

and a similar equality holds replacing ρ_1^σ and π_1^σ by ρ_2^σ and π_2^σ , and a_i by b_i . \square

3.2. Parametrization of 321-avoiding permutons. The goal of this section is to give a concrete parametrization of 321-avoiding permutons, extending in some sense the bijection of Lemma 15. For this it is useful to decompose permutons as follows: given a permuton μ in \mathcal{P} , we denote by μ_{\sqcup} , μ_{\sqsupset} and $\mu_{/}$ the restrictions of μ respectively to the lower-right triangle $\{(x, y) : x > y\}$, to the upper-left triangle $\{(x, y) : x < y\}$, and to the diagonal $\{(x, y) : x = y\}$. We start with a lemma.

Lemma 17. *Let μ be a 321-avoiding permuton. The supports of $\mu_{\sqcup} + \mu_{/}$ and $\mu_{\sqsupset} + \mu_{/}$ are nondecreasing subsets of $[0, 1]^2$.*

Proof. Assume for the sake of contradiction that there exist (x_1, y_1) and (x_2, y_2) in the support of $\mu_{\sqcup} + \mu_{/}$ that satisfy $x_1 < x_2$, but $y_1 > y_2$. By construction, this support is a subset of the closed triangle $\{(x, y) : x \geq y\}$, so we also have $x_1 \geq y_1$ (and $x_2 \geq y_2$). Since μ is a permuton, it holds that

$$\mu([0, x_1] \times [0, 1]) = x_1 \geq y_1 = \mu([0, 1] \times [0, y_1]).$$

Subtracting $\mu([0, x_1] \times [0, y_1])$ to both sides, we have

$$\mu([0, x_1] \times [y_1, 1]) \geq \mu([x_1, 1] \times [0, y_1]).$$

But the right-hand side is nonzero, since (x_2, y_2) is in the support of μ . Therefore there exists a point (x_0, y_0) satisfying $x_0 < x_1$ and $y_0 > y_1$ in the support of μ . The existence of three points (x_0, y_0) , (x_1, y_1) and (x_2, y_2) in decreasing position in the support of μ contradicts the fact that μ is 321-avoiding. This proves that the support of $\mu_{\sqcup} + \mu_{/}$ is nondecreasing. The proof that the support of $\mu_{\sqsupset} + \mu_{/}$ is nondecreasing is similar. \square

We recall that the space \mathcal{D} has been defined in (3). Note that \mathcal{D} is a compact and convex subset of $\mathcal{M}([0, 1]^2)$. It is straightforward to check that, for any 321-avoiding permutation σ , the pair $(\pi_1^\sigma, \pi_2^\sigma)$ defined above is in \mathcal{D} . For any (π_1, π_2) in \mathcal{D} , the measure

$$\Psi(\pi_1, \pi_2) := \pi_1 \nearrow \pi_2 + (\text{Leb}_{[0,1]} - \pi_1) \nearrow (\text{Leb}_{[0,1]} - \pi_2),$$

where the operation \nearrow has been defined in Section 2.3, is a permuton supported on a union of two nondecreasing sets and is thus 321-avoiding. See the right part of Figure 6.

Proposition 18. *The map Ψ is a continuous surjective map from \mathcal{D} to the set $\mathcal{P}_{\text{Av}(321)}$ of 321-avoiding permutons. Moreover, considering a 321-avoiding permuton μ and its decomposition $\mu = \mu_{\sqcup} + \mu_{\sqsupset} + \mu_{/}$ as above, we have*

$$(11) \quad \Psi^{-1}(\mu) = \{ \text{proj}(\mu_{\sqcup} + \nu) : 0 \leq \nu \leq \mu_{/} \},$$

where we recall that $\text{proj}(\pi)$ is the pair of measures given by the projections of π on the horizontal and vertical axes respectively.

Proof. We first prove that Ψ is a continuous mapping. Consider a sequence (π_1^n, π_2^n) converging weakly to (π_1, π_2) . Formally, we can write

$$\pi_1^n \nearrow \pi_2^n = (G_1^n, G_2^n) \# \text{Leb}_{[0, M_n]},$$

where G_1^n and G_2^n are the quantile functions of π_1^n and π_2^n , and $\text{Leb}([0, M_n])$ is the Lebesgue measure on the interval $[0, M_n]$ (of total mass M_n). If $M = \pi_1([0, 1]) = \pi_2([0, 1])$, the weak convergence of π_1^n to π_1 implies that $M_n = \pi_1^n([0, 1])$ tends to M , and consequently $\text{Leb}_{[0, M_n]}$ converges in total variation distance to $\text{Leb}_{[0, M]}$. Moreover, we have the convergence of the quantile functions G_1^n, G_2^n to the quantile functions G_1 and G_2 of π_1 and π_2 , at least at continuity

points of G_1, G_2 . Since G_1, G_2 have at most countably many discontinuity points, Lemma 34 applies, and $\pi_1^n \nearrow \pi_2^n$ converges weakly to $\pi_1 \nearrow \pi_2$. Similarly, $(\text{Leb} - \pi_1^n) \nearrow (\text{Leb} - \pi_2^n)$ converges weakly to $(\text{Leb} - \pi_1) \nearrow (\text{Leb} - \pi_2)$, which proves continuity.

We now prove (11), which will imply the surjectivity of Ψ , since the right-hand side is always nonempty. Let μ be a 321-avoiding permuton, and (π_1, π_2) a pair in \mathcal{D} such that $\Psi(\pi_1, \pi_2) = \mu$. The condition “ $\pi_1([0, x]) \leq \pi_2([0, x])$ for all x ” implies that $\pi_1 \nearrow \pi_2$ is supported on the *closed* lower-right triangle $\{(x, y) : x \geq y\}$ (see Figure 6), while $(\text{Leb} - \pi_1) \nearrow (\text{Leb} - \pi_2)$ is supported on the closed upper-left triangle $\{(x, y) : x \leq y\}$. Hence we have

$$\mu_{\lrcorner} \leq \pi_1 \nearrow \pi_2 \leq \mu_{\lrcorner} + \mu_{\swarrow}.$$

Hence, there is a measure $\nu \leq \mu$ such that $\pi_1 \nearrow \pi_2 = \mu_{\lrcorner} + \nu$, so $(\pi_1, \pi_2) = \text{proj}(\mu_{\lrcorner} + \nu)$.

It remains to prove the reverse inclusion. Let $\nu \leq \mu_{\swarrow}$, and set $(\pi_1, \pi_2) = \text{proj}(\mu_{\lrcorner} + \nu)$. By Lemma 17, the support of $\mu_{\lrcorner} + \mu_{\swarrow}$ is nondecreasing, and this also holds for $\mu_{\lrcorner} + \nu$. By uniqueness of the nondecreasing coupling, we have

$$(12) \quad \mu_{\lrcorner} + \nu = \pi_1 \nearrow \pi_2.$$

Similarly, we have $(\text{Leb} - \pi_1, \text{Leb} - \pi_2) = \text{proj}(\mu_{\ulcorner} + (\mu_{\swarrow} - \nu))$, and $\mu_{\ulcorner} + (\mu_{\swarrow} - \nu)$ has a nondecreasing support, so

$$(13) \quad \mu_{\ulcorner} + (\mu_{\swarrow} - \nu) = (\text{Leb} - \pi_1) \nearrow (\text{Leb} - \pi_2).$$

Combining equations (12) and (13) gives $\mu = \Psi(\pi_1, \pi_2)$, as wanted. \square

4. LARGE DEVIATION PRINCIPLE FOR 321-AVOIDING PERMUTATIONS

We recall that

$$\mathcal{D} := \left\{ (\pi_1, \pi_2) \in \mathcal{M}([0, 1])^2 : \begin{array}{l} \pi_1 \leq \text{Leb}, \pi_2 \leq \text{Leb}, \pi_1([0, 1]) = \pi_2([0, 1]) \\ \text{and } \forall x \in [0, 1], \pi_1([0, x]) \leq \pi_2([0, x]) \end{array} \right\}$$

and define

$$\mathcal{S} := \{ \pi \in \mathcal{M}([0, 1]) : \pi \leq \text{Leb} \}$$

the space of subuniform measures on $[0, 1]$. The space \mathcal{S} can be endowed with the Kolmogorov distance

$$d(\pi_1, \nu_1) := \sup_{x \in [0, 1]} |\pi_1([0, x]) - \nu_1([0, x])|,$$

which induces the topology of weak convergence (since measures in \mathcal{S} have no atoms). We extend this to \mathcal{S}^2 (and thus \mathcal{D}) by taking the maximum of the distances between coordinates. For $\pi \in \mathcal{S}^2$ and $\varepsilon > 0$, we denote by

$$B(\pi, \varepsilon) := \{ \nu \in \mathcal{S}^2 : d(\pi, \nu) \leq \varepsilon \}$$

the ball of center π and radius ε .

Our goal is to prove the LDP for 321-avoiding permutations, that is, Proposition 5. We start by introducing some notation. In the following, if (σ_n) is a sequence of permutations, we write (π_1^n, π_2^n) instead of $(\pi_1^{\sigma_n}, \pi_2^{\sigma_n})$. Let \mathcal{P}_n be the set of subsets of $\{1, \dots, n\}$. For $A \in \mathcal{P}_n$, we let ν_n^A be the measure $\sum_{a \in A} \text{Leb}_{[\frac{a-1}{n}, \frac{a}{n}]}$, which lives in \mathcal{S} .

By Lemma 15, the set Dyck_n is in bijection with 321-avoiding permutations of size n . Therefore, if σ_n is a uniform random 321-avoiding permutation of size n , then (π_1^n, π_2^n) has the same distribution as (ν_n^A, ν_n^B) , where (A, B) is taken uniformly at random in Dyck_n . We will compare this model with a related but simpler model: let A' and B' be two uniform random subsets of $\{1, \dots, n\}$, independent from each other. Note that $(\nu_n^{A'}, \nu_n^{B'})$ belongs to \mathcal{S}^2 but not necessarily to \mathcal{D} .

We define the continuous function $F_n^{A'}$ which is affine on each interval $[\frac{a-1}{n}, \frac{a}{n}]$ and such that

$$F_n^{A'}(x) = \frac{1}{n} \sum_{i=1}^{nx} \mathbf{1}[i \in A']$$

when xn is an integer. Then $F_n^{A'}$ is a normalized nondecreasing random walk with Bernoulli(1/2) steps. From [DZ98, Theorem 5.1.2], it follows that $F_n^{A'}$ satisfies a large deviation principle⁶ at speed n with rate function $H_1(f) = \int_0^1 I(f'(x))dx$, where

$$I(y) := \begin{cases} y \log(y) + (1-y) \log(1-y) + \log(2) & \text{if } 0 \leq y \leq 1; \\ \infty & \text{otherwise} \end{cases}$$

is the Fenchel–Legendre transform of the moment generating function of a Bernoulli(1/2) random variable. Note that $H_1(f) = \infty$ unless f' takes its values in $[0, 1]$ almost everywhere. But a function f with $0 \leq f' \leq 1$ a.e. can be identified with the measure $f'd\text{Leb}$, and the L^∞ topology on such functions coincides with the Kolmogorov distance on the set \mathcal{S} of subuniform measures. Hence the above large deviation principle can be rephrased as a large deviation principle on measures: if A' is chosen uniformly at random in \mathcal{P}_n , then the sequence $(\nu_n^{A'})$ satisfies a large deviation principle on the set \mathcal{S} at speed n with rate function H_1 where

$$H_1(\pi) = \int_0^1 \rho \log(\rho) + (1-\rho) \log(1-\rho) + \log(2)$$

if π has density ρ .

Since $\nu_n^{B'}$ is an independent copy of $\nu_n^{A'}$, it follows that the pair $(\nu_n^{A'}, \nu_n^{B'})_{n \geq 1}$ satisfies a large deviation principle on \mathcal{S}^2 at speed n with rate function $H(\pi_1', \pi_2') = H_1(\pi_1') + H_1(\pi_2')$, which coincides with the definition given in Proposition 5. We note for later use that the same large deviation principle holds if we take A' and B' to be uniform random subsets of $\{1, \dots, n-1\}$ (indeed, the two models are exponentially equivalent in the sense of [DZ98, Definition 4.2.12]).

Proof of Proposition 5. As explained in Section 2.1, as \mathcal{D} is compact, it is enough to show that, for any $\pi \in \mathcal{D}$,

$$(14) \quad \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\pi_1^n, \pi_2^n) \in B(\pi, \varepsilon)) \leq -H(\pi),$$

and

$$(15) \quad \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\pi_1^n, \pi_2^n) \in B(\pi, \varepsilon)) \geq -H(\pi).$$

We start with the upper bound (14), which is easy to establish. The number of Dyck paths of length $2n$ is the Catalan number $C_n = \frac{1}{n+1} \binom{2n}{n}$. From the Stirling formula, it is easy to show that $C_n = \frac{1}{\sqrt{\pi}} n^{-3/2} 4^n (1 + o(1))$. Consequently, for any $\pi \in \mathcal{D}$, we have

$$\begin{aligned} P((\pi_1^n, \pi_2^n) \in B(\pi, \varepsilon)) &= \frac{1}{C_n} \# \{(A, B) \in \text{Dyck}_n : (\nu_n^A, \nu_n^B) \in B(\pi, \varepsilon)\} \\ &\leq \frac{\# \{(A', B') \in \mathcal{P}_n^2 : (\nu_n^{A'}, \nu_n^{B'}) \in B(\pi, \varepsilon)\}}{4^n} \frac{4^n}{C_n} \\ &\leq \mathbb{P}((\nu_n^{A'}, \nu_n^{B'}) \in B(\pi, \varepsilon)) n^{3/2} (\sqrt{\pi} + o(1)). \end{aligned}$$

⁶[DZ98, Theorem 5.1.2] is in fact stated for a càdlàg version of F_n , but both versions are exponentially equivalent [DZ98, Lemma 5.1.4]. Hence, large deviation principles for both versions are equivalent [DZ98, Theorem 4.2.13].

Therefore,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\pi_1^n, \pi_2^n) \in B(\pi, \varepsilon)) \leq \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\left(\nu_n^{A'}, \nu_n^{B'}\right) \in B(\pi, \varepsilon)\right) \leq -H(\pi),$$

where the second inequality is given by the large deviation principle for $(\nu_n^{A'}, \nu_n^{B'})$. If $\pi \in \mathcal{S}^2 \setminus \mathcal{D}$, for ε small enough $B(\pi, \varepsilon) \cap \mathcal{D} = \emptyset$ so that $\mathbb{P}((\pi_1^n, \pi_2^n) \in B(\pi, \varepsilon)) = 0$.

We now prove the lower bound. Let $\pi = (\pi_1, \pi_2) \in \mathcal{D}$. We denote by $F_1(x) = \pi_1([0, x])$ and $F_2(x) = \pi_2([0, x])$ the respective cumulative distribution functions of π_1 and π_2 . By definition of \mathcal{D} , for all $x \in [0, 1]$, we have $0 \leq F_1(x) \leq F_2(x) \leq x$ and $F_1(1) = F_2(1)$. Consider a pair (A', B') in \mathcal{P}_{n-1}^2 (note the shift of index, in particular $n \notin B'$) such that $(\nu_n^{A'}, \nu_n^{B'})$ is in $B(\pi, \varepsilon)$. The goal is to associate with it a pair (A, B) in Dyck_n such that (ν_n^A, ν_n^B) is also close to π .

First observe that $(\nu_n^{A'}, \nu_n^{B'}) \in B(\pi, \varepsilon)$ implies that, for all $1 \leq k \leq n$,

$$\left| \frac{1}{n} \sum_{a \in A'} \mathbf{1}[a \leq k] - F_1\left(\frac{k}{n}\right) \right| \leq \varepsilon,$$

and

$$\left| \frac{1}{n} \sum_{b \in B'} \mathbf{1}[b \leq k] - F_2\left(\frac{k}{n}\right) \right| \leq \varepsilon.$$

Hence, using $F_1\left(\frac{k}{n}\right) \leq F_2\left(\frac{k}{n}\right)$, we get that

$$(16) \quad \min_{1 \leq k \leq n} \left(\sum_{b \in B'} \mathbf{1}[b < k] - \sum_{a \in A'} \mathbf{1}[a < k] \right) \geq -2\varepsilon n.$$

Also, since $F_1(1) = F_2(1)$, setting $\delta = |B'| - |A'|$, we have $|\delta| \leq 2\varepsilon n$. We now define A as follows:

- If $|A'| > 2\varepsilon n$, we first let A'' be obtained from A' by removing its $\lfloor 2\varepsilon n \rfloor + 1$ smallest elements. Then we let A be obtained from A'' by adding the $\lfloor 2\varepsilon n \rfloor + \delta + 1$ largest elements of $\{1, \dots, n\}$ which are not in A'' .
- If $|A'| \leq 2\varepsilon n$, we let $A = \{n - |B'| + 1, \dots, n\}$.

Finally simply set $B = B'$. We claim that (A, B) is in Dyck_n . Let us prove it in the first case above, the second case being easier. That $|A| = |B|$ holds by construction. To check that $a_i > b_i$ for all $i \leq |A|$, we verify the equivalent assertion: for all $k \leq n$

$$(17) \quad \sum_{b \in B} \mathbf{1}[b < k] \geq \sum_{a \in A} \mathbf{1}[a < k],$$

with a strict inequality whenever $k \in A$. We set $\bar{a} = \max(A' \setminus A'')$ and $\tilde{a} = \min(A \setminus A'')$, i.e. these are respectively the largest element that has been removed and the smallest element that has been added to go from A' to A . If $\bar{a} \geq \tilde{a}$, then A consists of the $|B|$ largest elements in $\{1, \dots, n\}$, and $a_i > b_i$ holds trivially for all $i \leq |A|$ (recall that $n \notin B$ by construction). So assume $\bar{a} < \tilde{a}$, and let us prove (17).

- If $k \leq \bar{a}$, then $\sum_{a \in A} \mathbf{1}[a \leq k] = 0$ and (17) holds trivially.
- If $k \geq \tilde{a}$, then A contains every element larger than k . Consequently, we have

$$\sum_{a \in A} \mathbf{1}[a \geq k] = n - k + 1 > \sum_{b \in B} \mathbf{1}[b \geq k],$$

since $n \notin B$. Since $|A| = |B|$, this implies (17) with a strict inequality.

- Finally, if $\bar{a} < k < \tilde{a}$, then we have

$$\sum_{a \in A} \mathbf{1}[a \leq k] = \sum_{a \in A'} \mathbf{1}[a \leq k] - \lfloor 2\varepsilon n \rfloor - 1.$$

Indeed, the $[2\varepsilon n] + 1$ elements of the set difference $A' \setminus A$ are all smaller than k , while the elements of the set difference $A \setminus A'$ are all larger than k . The strict version of inequality (17) then follows from (16).

Thus (17) holds for all k (with a strict inequality when $k \in A$), and (A, B) lies in Dyck_n as claimed.

Moreover, since we have transformed A' into A by removing at most $2\varepsilon n + 1$ and adding at most $2\varepsilon n + \delta + 1$ elements, we have

$$d(\nu_n^A, \nu_n^{A'}) \leq 4\varepsilon + \frac{\delta+2}{n} \leq 7\varepsilon$$

for n sufficiently large since $\delta \leq 2\varepsilon n$. Let us call T the map associating (A, B) with (A', B') . A preimage (A', B') of (A, B) is determined by the set of $k \leq 7\varepsilon n$ elements that have been added to or removed from A' to obtain A . Choose ε small enough so that $7\varepsilon < 1/2$. Then, for a given $(A, B) \in \text{Dyck}_n$, the number of its preimages by T is at most

$$\sum_{k \leq 7\varepsilon n} \binom{n}{k} \leq 7\varepsilon n \binom{n}{7\varepsilon n} \leq \frac{e\sqrt{7\varepsilon n}}{2\pi\sqrt{1-7\varepsilon}} ((7\varepsilon)^{7\varepsilon} (1-7\varepsilon)^{1-7\varepsilon})^{-n},$$

where we used that $k \mapsto \binom{n}{k}$ is increasing if $k < n/2$ and $k!e^k k^{-k-1/2} \in [\sqrt{2\pi}, e]$ for all $k \geq 1$.

Since T maps any (A', B') in \mathcal{P}_{n-1}^2 with $(\nu_n^{A'}, \nu_n^{B'}) \in B(\pi, \varepsilon)$ to a pair (A, B) in Dyck_n with $(\nu_n^A, \nu_n^B) \in B(\pi, 8\varepsilon)$, we have

$$(18) \quad \#\{(A, B) \in \text{Dyck}_n : (\nu_n^A, \nu_n^B) \in B(\pi, 8\varepsilon)\} \geq \frac{2\pi\sqrt{1-7\varepsilon}}{e\sqrt{7\varepsilon n}} ((7\varepsilon)^{7\varepsilon} (1-7\varepsilon)^{1-7\varepsilon})^n \cdot \#\{(A', B') \in \mathcal{P}_{n-1}^2 : (\nu_n^{A'}, \nu_n^{B'}) \in B(\pi, \varepsilon)\}.$$

Therefore, if (A, B) and (A', B') are uniformly distributed in Dyck_n and \mathcal{P}_{n-1}^2 respectively,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\nu_n^A, \nu_n^B) \in B(\pi, 8\varepsilon)) \geq 7\varepsilon \log(7\varepsilon) + (1-7\varepsilon) \log(1-7\varepsilon) + \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\nu_n^{A'}, \nu_n^{B'}) \in B(\pi, \varepsilon)).$$

Letting ε go to 0 and using the large deviation principle for $(\nu_n^{A'}, \nu_n^{B'})$ give (15). This concludes the proof of Proposition 5. \square

From there, it is straightforward to deduce the corresponding LDP for the pair of measures associated with a random Mallows permutation, conditioned to avoid 321. Omitting the superscript 321, let us denote by τ_n^β a 321-avoiding Mallows random permutation with parameters $n \geq 1$ and $q_n = e^{\frac{\beta}{n}}$. To lighten notation, we write $(\pi_1^{[n, \beta]}, \pi_2^{[n, \beta]}) := (\pi_1^{\tau_n^\beta}, \pi_2^{\tau_n^\beta})$. Finally, for $\beta \in \mathbb{R}$, we define the action $A_\beta(\pi_1, \pi_2)$ on \mathcal{D} by

$$(19) \quad A_\beta(\pi_1, \pi_2) := H(\pi_1, \pi_2) - \beta \int_0^1 x (\rho_1(x) - \rho_2(x)) dx,$$

where ρ_1 and ρ_2 are the respective densities of π_1 and π_2 . Then, we have the following LDP.

Corollary 19. *Let $\beta \in \mathbb{R}$. The pair $(\pi_1^{[n, \beta]}, \pi_2^{[n, \beta]})$ satisfies a large deviation principle on \mathcal{D} at speed n with rate function*

$$A_\beta(\pi_1, \pi_2) - \inf(A_\beta).$$

Proof. By Lemma 16, the law of $(\pi_1^{[n, \beta]}, \pi_2^{[n, \beta]})$ is absolutely continuous with respect to the law of $(\pi_1^{[n, 0]}, \pi_2^{[n, 0]})$, with Radon–Nikodym derivative given by

$$\frac{Z_n^0}{Z_n^\beta} \exp\left(\frac{\beta}{n} \text{inv}(\sigma)\right) = \frac{Z_n^0}{Z_n^\beta} \exp\left(\beta n \int_0^1 x (\rho_1(x) - \rho_2(x)) dx\right),$$

where ρ_1 and ρ_2 are the respective densities of π_1 and π_2 with respect to the Lebesgue measure. The corollary now follows from Proposition 5 and Varadhan's Lemma 13. \square

The limit shape will be determined by minimizing the function A_β ; this is detailed in the next section. Before getting there, we make a remark on these LDPs. For this, we observe that, by construction, for any 321-avoiding permutation σ , we have $\mu_\sigma^\swarrow = \Psi(\pi_1^\sigma, \pi_2^\sigma)$. Since Ψ is a continuous mapping, using the contraction principle for large deviations (Proposition 12), it is possible to deduce from Proposition 5 an LDP in the space of permutons. Although we won't use it in the sequel, we think it may be interesting *per se*.

Corollary 20. *For all $n \geq 1$, let σ_n be a uniform random 321-avoiding permutation of size n . Then, the permutons $\mu_{\sigma_n}^\swarrow$ satisfy a large deviation principle at speed n with rate function*

$$\overline{H}(\mu) := \begin{cases} \inf_{0 \leq \nu \leq \mu^\swarrow} H(\text{proj}(\mu, \cdot) + \nu) & \text{if } \mu \in \mathcal{P}_{\text{Av}(321)}, \\ \infty & \text{otherwise.} \end{cases}$$

The same LDP holds for the permutons μ_{σ_n} .

The last statement holds because the laws of μ_{σ_n} and $\mu_{\sigma_n}^\swarrow$ are exponentially equivalent to each other in the sense of [DZ98, Definition 4.2.10].

5. AN OPTIMIZATION PROBLEM FOR PAIRS OF MEASURES

Our goal is now to minimize the action A_β appearing in Corollary 19.

Proposition 21. *For any $\beta \in \mathbb{R}$, the action A_β has a unique minimizer $(\pi_1^{(\beta)}, \pi_2^{(\beta)})$ in \mathcal{D} . For $\beta \leq 0$, we have $\pi_1^{(\beta)} = \pi_2^{(\beta)} = \frac{1}{2} \text{Leb}$. For $\beta > 0$, the densities of $\pi_1^{(\beta)}$ and $\pi_2^{(\beta)}$ on $[0, 1]$ are given by*

$$(20) \quad \rho_1^{(\beta)}(x) := \frac{1}{1 + \exp(\beta(\frac{1}{2} - x))}, \quad \rho_2^{(\beta)}(x) := \frac{1}{1 + \exp(\beta(x - \frac{1}{2}))}.$$

In particular, both $\pi_1^{(\beta)}$ and $\pi_2^{(\beta)}$ have total mass $\frac{1}{2}$.

Proof of Proposition 21. We will rely on differential calculus on the space \mathcal{D} (see Proposition 36 in Appendix B for a precise statement). We first note that the domain \mathcal{D} is convex and compact for the topology of weak convergence of measures. Moreover, the action A_β is a lower semicontinuous function of (π_1, π_2) . Finally, the entropy H is a strictly convex function of (π_1, π_2) and the additional term in the action is affine, so A_β is strictly convex, which shows the existence and uniqueness of the minimizer. For $\beta \leq 0$, the term H has a unique minimum at $\pi_1 = \pi_2 = \frac{1}{2} \text{Leb}$, and the second term of (19) is always nonnegative and is 0 for $\pi_1 = \pi_2$, which identifies the minimum.

We now assume $\beta > 0$. By Proposition 36, to ensure that the minimizer of A_β is given by the $(\pi_1^{(\beta)}, \pi_2^{(\beta)})$ of (20), it is now sufficient to show that the differential of A_β vanishes at $(\pi_1^{(\beta)}, \pi_2^{(\beta)})$.

To this end, we define the set of directions

$$V_{(\pi_1^{(\beta)}, \pi_2^{(\beta)})}(\mathcal{D}) := \{(\mu_1, \mu_2) : (\pi_1^{(\beta)} + t\mu_1, \pi_2^{(\beta)} + t\mu_2) \in \mathcal{D} \text{ for } t > 0 \text{ small enough}\}$$

in which the derivative could make sense. We first check that any $(\mu_1, \mu_2) \in V_{(\pi_1^{(\beta)}, \pi_2^{(\beta)})}(\mathcal{D})$ is a pair of signed measures with $\mu_1([0, 1]) = \mu_2([0, 1])$ and such that there is a constant C such that $|\mu_1|, |\mu_2| \leq C \text{Leb}$. Indeed, one must have $\mu_1([0, 1]) = \mu_2([0, 1])$ to ensure that $\pi_1^{(\beta)}([0, 1]) + t\mu_1([0, 1]) = \pi_2^{(\beta)}([0, 1]) + t\mu_2([0, 1])$ for t small enough. Moreover, for $i \in \{1, 2\}$, the condition $|\mu_i| \leq C \text{Leb}$ is necessary to ensure that $0 \leq \pi_i^{(\beta)} + t\mu_i \leq \text{Leb}$ for t small enough.

We now differentiate the action A_β . For any direction $(\mu_1, \mu_2) \in V_{(\pi_1^{(\beta)}, \pi_2^{(\beta)})}(\mathcal{D})$, writing α_1, α_2 for the densities of μ_1, μ_2 , we have

$$\left. \frac{d}{dt} \right|_{t=0^+} H\left(\pi_1^{(\beta)} + t\mu_1, \pi_2^{(\beta)} + t\mu_2\right) = \int_0^1 \alpha_1 \log \frac{\rho_1^{(\beta)}}{1 - \rho_1^{(\beta)}} + \alpha_2 \log \frac{\rho_2^{(\beta)}}{1 - \rho_2^{(\beta)}}.$$

Note that differentiating inside the integral sign is indeed legit since the densities $\rho_1^{(\beta)}$ and $\rho_2^{(\beta)}$ are bounded away from 0 and 1, and hence everything is uniformly bounded for t small enough. On the other hand, the second term of the action A_β is a linear function so we can easily differentiate it. We obtain:

$$(21) \quad \left. \frac{d}{dt} \right|_{t=0^+} A_\beta\left(\pi_1^{(\beta)} + t\mu_1, \pi_2^{(\beta)} + t\mu_2\right) \\ = \int_0^1 \alpha_1 \log \frac{\rho_1^{(\beta)}}{1 - \rho_1^{(\beta)}} + \alpha_2 \log \frac{\rho_2^{(\beta)}}{1 - \rho_2^{(\beta)}} - \beta \int_0^1 x(\alpha_1(x) - \alpha_2(x)) dx.$$

Replacing $\rho_1^{(\beta)}$ and $\rho_2^{(\beta)}$ by their explicit expressions, this becomes

$$\int_0^1 \left(\alpha_1(x) \beta \left(x - \frac{1}{2} \right) + \alpha_2(x) \beta \left(\frac{1}{2} - x \right) - \beta x \alpha_1(x) + \beta x \alpha_2(x) \right) dx \\ = \frac{\beta}{2} \int_0^1 (\alpha_2(x) - \alpha_1(x)) dx = 0,$$

since $\mu_1([0, 1]) = \mu_2([0, 1])$.

Finally, the fact that $\pi_1^{(\beta)}$ and $\pi_2^{(\beta)}$ both have mass $\frac{1}{2}$ follows from $\rho_1^{(\beta)} + \rho_2^{(\beta)} = 1$. \square

Remark 22. The above proof may look like a “guess-and-check”, but we could also *find* the formulas (20) from the condition in Proposition 36. Indeed, we look for densities ρ_1 and ρ_2 such that the right-hand side of (21) vanishes for any α_1, α_2 with $\int \alpha_1 = \int \alpha_2$. In particular, the case $\alpha_2 = 0$ implies that one should have $\int \alpha_1 \left(\log \frac{\rho_1}{1 - \rho_1} - \beta x \right) = 0$ whenever $\int \alpha_1 = 0$. Therefore $\log \frac{\rho_1}{1 - \rho_1} - \beta x$ must be equal a.e. to a constant c_1 , which means that $\rho_1(x) = \frac{1}{1 + e^{-\beta x - c_1}}$ a.e., and similarly for ρ_2 with another constant c_2 . Then taking $\alpha_1(x) = \alpha_2(x) = 1$ for all x in $[0, 1]$ in (21) shows $c_1 = c_2$. Finally, for the minimizer to live in \mathcal{D} , we need to have $\int_0^1 \rho_1 = \int_0^1 \rho_2$, which happens only for $c_1 = c_2 = -\beta/2$.

Before going to the proof of Theorem 7, let us use the minimisation result that we have just obtained to study the partition function of the model.

Proof of the estimate (6). By the first estimate in Lemma 13, Corollary 19 and Proposition 21, we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{Z_n^\beta}{Z_n^0} = -\inf A_\beta = -A_\beta(\pi_1^{(\beta)}, \pi_2^{(\beta)}).$$

Recalling that $\lim_{n \rightarrow \infty} \frac{1}{n} \log(Z_n^0) = 2 \log(2)$ (since Z_n^0 is the n -th Catalan number C_n), it suffices to evaluate $A_\beta(\pi_1^{(\beta)}, \pi_2^{(\beta)})$. The latter can be (almost) explicitly computed: using the change of variable $y = e^{\beta(x - \frac{1}{2})}$, we have $dy = \beta y dx$, and $\rho_1^{(\beta)} = y/(1 + y)$, so that

$$\int_0^1 \rho_1^{(\beta)}(x) \log \left(\frac{\rho_1^{(\beta)}(x)}{1 - \rho_1^{(\beta)}(x)} \right) dx = \beta^{-1} \int_{e^{-\beta/2}}^{e^{\beta/2}} \frac{\log y}{1 + y} dy,$$

while

$$\int_0^1 \log \left(1 - \rho_1^{(\beta)}(x) \right) dx = -\beta^{-1} \int_{e^{-\beta/2}}^{e^{\beta/2}} \log(1 + y) \frac{dy}{y}.$$

By symmetry the integral with $\rho_2^{(\beta)}$ have the same values and we get

$$(22) \quad H(\pi_1^{(\beta)}, \pi_2^{(\beta)}) = \frac{2}{\beta} \int_{e^{-\beta/2}}^{e^{\beta/2}} \left(\frac{\log y}{1+y} - \frac{\log(1+y)}{y} \right) dy + 2 \log 2.$$

We now compute, using that $\rho_1^{(\beta)}$ and $\rho_2^{(\beta)}$ both have integral 1/2,

$$\beta \int_0^1 x \left(\rho_1^{(\beta)}(x) - \rho_2^{(\beta)}(x) \right) dx = \beta \int_0^1 \left(x - \frac{1}{2} \right) \left(\rho_1^{(\beta)}(x) - \rho_2^{(\beta)}(x) \right) dx = \int_{e^{-\beta/2}}^{e^{\beta/2}} \left(\frac{1}{\beta} \log(y) \right) \frac{y-1}{y+1} \frac{dy}{y}.$$

Now, rewriting $(y-1)/(y+1)$ as $1 - 2/(y+1)$ and observing that

$$\int_{e^{-\beta/2}}^{e^{\beta/2}} \frac{\log(y)}{y} dy = \frac{1}{2} (\log(e^{\beta/2})^2 - \log(e^{-\beta/2})^2) = 0,$$

we get

$$\begin{aligned} \beta \int_0^1 x \left(\rho_1^{(\beta)}(x) - \rho_2^{(\beta)}(x) \right) dx &= -2\beta^{-1} \int_{e^{-\beta/2}}^{e^{\beta/2}} \frac{\log(y)}{y(y+1)} dy \\ &= 2\beta^{-1} \int_{e^{-\beta/2}}^{e^{\beta/2}} \log(y) \left(\frac{1}{y+1} - \frac{1}{y} \right) dy = 2\beta^{-1} \int_{e^{-\beta/2}}^{e^{\beta/2}} \frac{\log(y)}{y+1} dy. \end{aligned}$$

Removing this quantity from (22) leads to the asymptotic formula announced in Equation (6). \square

We can now conclude the proof of Theorem 7.

Proof of Theorem 7. We recall from Corollary 19 that $(\pi_1^{[n,\beta]}, \pi_2^{[n,\beta]})$ satisfies an LDP at speed n with rate function $A_\beta(\pi_1, \pi_2) - \inf(A_\beta)$. In particular, since A_β has a unique minimizer $(\pi_1^{(\beta)}, \pi_2^{(\beta)})$ by Proposition 21, the sequence $(\pi_1^{[n,\beta]}, \pi_2^{[n,\beta]})$ converges a.s. to $(\pi_1^{(\beta)}, \pi_2^{(\beta)})$. Since the mapping $\Psi : \mathcal{D} \rightarrow \mathcal{P}_{\text{Av}(321)}$ is continuous by Proposition 18, it follows that the sequence $\tau_n^\beta = \Psi(\pi_1^{[n,\beta]}, \pi_2^{[n,\beta]})$ converges a.s. to the permuton $\Psi(\pi_1^{(\beta)}, \pi_2^{(\beta)})$.

To conclude the proof, we only need to check that $\Psi(\pi_1^{(\beta)}, \pi_2^{(\beta)})$ is indeed the permuton described by Theorem 7. We will drop the dependency in β for ease of notation. We denote by $G_1 = G_{2\pi_1}$ and $G_2 = G_{2\pi_2} : [0, 1] \rightarrow [0, 1]$ the respective quantile functions of the probability measures $2\pi_1$ and $2\pi_2$ as defined in (9). Note that π_1, π_2 have positive densities, so G_1 and G_2 are both increasing bijections. We can write

$$(23) \quad \begin{aligned} \pi_1 \nearrow \pi_2 &= \frac{1}{2} (G_1, G_2)_{\#} \text{Leb}_{[0,1]} \\ &= \frac{1}{2} (X, G_2 \circ G_1^{-1}(X))_{\#} ((G_1)_{\#} \text{Leb}_{[0,1]}) \\ &= (X, G_2 \circ G_1^{-1}(X))_{\#} \pi_1. \end{aligned}$$

We now write $f = G_2 \circ G_1^{-1}$ and check that this coincides with the function f_β appearing in Theorem 7. By definition of quantile functions, for any $u \in [0, 1]$, we have

$$2\pi_1([0, G_1(u)]) = u = 2\pi_2([0, G_2(u)]),$$

so $\pi_1([0, x]) = \pi_2([0, f(x)])$. Since the density ρ_2 is continuous and positive, this identity implies that f is \mathcal{C}^1 as well. Differentiating yields $\rho_1(x) = f'(x)\rho_2(f(x))$, and therefore:

$$f'(x) = \frac{\rho_1(x)}{\rho_2(f(x))} = \frac{1 + e^{\beta(f(x)-1/2)}}{1 + e^{\beta(1/2-x)}}.$$

Substituting $g := e^{\beta(f-1/2)}$, we find that

$$g' = qg(1+g)$$

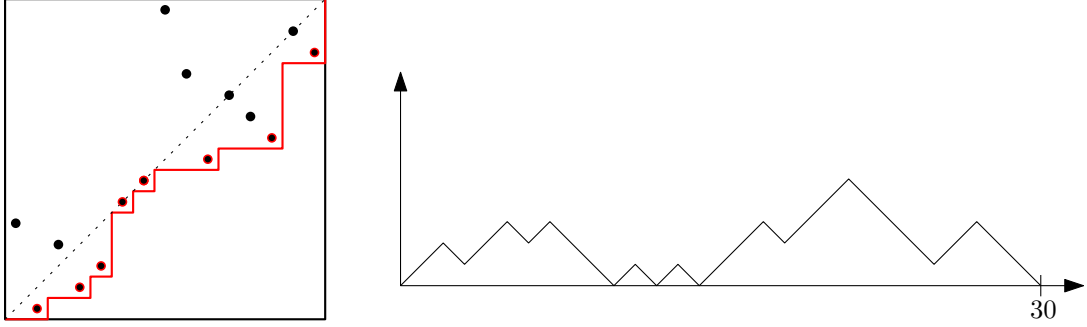


FIGURE 7. Left: a 231-avoiding permutation σ of size 15. Its set of RL minima, and the associated curve F_σ , are shown in red. Right: the Dyck path Φ_σ of length 30 obtained by rotating and flipping the RLM curve.

where $q(x) := \frac{\beta}{1+e^{\beta(1/2-x)}}$. This is a simple Riccati equation, which can be solved⁷ by

$$g(x) = -\frac{e^{\beta x} + e^{\beta/2}}{e^{\beta x} + c}$$

for $c \in \mathbb{R}$ (the only other solution is $g = 0$). Using $g(0) = e^{-\beta/2}$ yields $c = -(1 + e^{\beta/2} + e^\beta)$, and the identification of f follows. Combined with (23) and the same computation for $(\text{Leb} - \pi_1) \nearrow (\text{Leb} - \pi_2)$, this shows that $\Psi(\pi_1, \pi_2)$ is indeed as described by Theorem 7. \square

Part 2. Avoiding 231

We now consider permutations and permutons avoiding 231. Again, in this part we will omit the superscript 231 in all the notations.

6. 231-AVOIDING PERMUTATIONS AND PERMUTONS

6.1. A classical bijection for 231-avoiding permutations. If σ avoids 231, let

$$F_\sigma : x \in [0, n) \mapsto \min \{ \sigma(c) : c > x \} - 1.$$

This defines a càdlàg, piecewise constant, nondecreasing function⁸. We can let $F_\sigma(n) := n$, so that F_σ starts at 0 and ends at n . Also define Φ_σ as the Dyck path of size $2n$ obtained by rotating F_σ by $-\pi/4$, flipping it, and scaling it up by $\sqrt{2}$. See Figure 7 for an illustration. The following lemma can e.g. be found in [Bón12, Proof of Lemma 4.3]:

Lemma 23. *The application $\sigma \mapsto \Phi_\sigma$ is one-to-one, from 231-avoiding permutations to Dyck paths of length $2n$.*

In particular, if σ_n is a uniformly random 231-avoiding permutation with size n then Φ_{σ_n} is a uniformly random Dyck path of length $2n$. Finally, we define the normalized RLM curve of σ as $f_\sigma := \frac{1}{n} F_\sigma(n \cdot)$ and its normalized excursion as $\varphi_\sigma := \frac{1}{2n} \Phi_\sigma(2n \cdot)$ on $[0, 1]$.

⁷To solve $g' = qg(1+g)$, consider $u := ae^{\beta x} + b$ for arbitrary $a \neq 0$ and $b \in \mathbb{R}$. The latter solves $u'' = Ru'$ where $R := q + \frac{q'}{q} = \beta$. It can then be checked that $g := \frac{-u'}{qu}$ solves the desired equation, and finally, c is defined as b/a .

⁸For readers familiar with the notion of right-to-left minima of a permutation, F_σ is the largest integer-valued function whose graph stays below the right-to-left minima of σ . In particular, it is completely determined by, and completely determines, the positions and values of the right-to-left minima of σ .

Proposition 24. *If σ is a 231-avoiding permutation then*

$$\text{inv}(\sigma) = \binom{n}{2} - \sum_{x=0}^{n-1} F_\sigma(x).$$

Therefore,

$$\text{inv}(\sigma) = \binom{n}{2} - n^2 \int_0^1 f_\sigma(x) dx = \frac{-n}{2} + 2n^2 \int_0^1 \varphi_\sigma(x) dx.$$

Proof. Let $\text{noninv}(\sigma) := |\{(x, y) \in [n]^2 : x < y, \sigma(x) < \sigma(y)\}|$ be the number of noninversions of σ . Since $\text{inv}(\sigma) + \text{noninv}(\sigma) = \binom{n}{2}$, we may as well compute $\text{noninv}(\sigma)$. For this, we enumerate noninversions by their north-east point. Let $c \in [n]$ be fixed. We claim that

$$(24) \quad \{x \in [n] : x < c, \sigma(x) < \sigma(c)\} = \{\sigma^{-1}(y) : y \leq F_\sigma(c-1)\},$$

and thus

$$|\{x \in [n] : x < c, \sigma(x) < \sigma(c)\}| = F_\sigma(c-1),$$

so that $\text{noninv}(\sigma) = \sum_{x=0}^{n-1} F_\sigma(x)$.

Indeed, for $c \in [n]$, by definition of F_σ , for any $y \leq F_\sigma(c-1)$, we have $\sigma^{-1}(y) < c$ since there is no point under the RLM curve.

Moreover, by construction $\sigma(c) \geq F_\sigma(c-1) + 1$. If $\sigma(c) = F_\sigma(c-1) + 1$, then the previous remark directly gives (24). Now, if $\sigma(c) > F_\sigma(c-1) + 1$, then there exists $c' > c$ such that $\sigma(c') = F_\sigma(c-1) + 1$ (intuitively, c' is the first RL minimum at the right of c). Let $F_\sigma(c-1) < y < \sigma(c)$. If $\sigma^{-1}(y) < c$, then the permutation σ is not 231-avoiding because we have $\sigma^{-1}(y) < c < c'$ and $\sigma(c') < y < \sigma(c)$. Therefore, (24) holds in both cases and this concludes the proof of the first formula.

Finally, since the non-normalized RLM curve F_σ is locally constant and jumps at integer positions, $\sum_{x=0}^{n-1} F_\sigma(x) = \int_0^n F_\sigma(x) dx$. Using $\int_0^1 f_\sigma(x) dx = \frac{1}{n^2} \int_0^n F_\sigma(x) dx$ and $\frac{1}{2} = \int_0^1 f_\sigma(x) dx + 2 \int_0^1 \varphi_\sigma(x) dx$, the rest of the Proposition follows. \square

6.2. 231-avoiding permutons. We write $\mathcal{P}_{\text{Av}(231)}$ for the set of 231-avoiding permutons. Given $\mu \in \mathcal{P}_{\text{Av}(231)}$, we define its RLM curve as

$$f_\mu : x \in [0, 1] \mapsto \max \{y \in [0, 1] : \mu([x, 1] \times [0, y]) = 0\}.$$

The function f_μ is càdlàg, nondecreasing and satisfies $f_\mu(1) = 1$ and $f_\mu(x) \leq x$ for x in $[0, 1]$ (using the fact that permutons have uniform marginals).

Conversely, let f be a nondecreasing càdlàg function from $[0, 1]$ to $[0, 1]$ satisfying $f(x) \leq x$ and $f(1) = 1$. We let \mathcal{F} denote the space of such functions. We want to construct a 231-avoiding permuton μ_f whose RLM-curve is f . The standard way to proceed for 231-avoiding permutations is a recursive construction from right to left, which is hard to transpose to continuous objects. Instead, we construct μ_f by specifying the area of lower-right rectangles $[x, 1] \times [0, y]$.

For this, we let $H_f := \{(x, y) \in [0, 1]^2 : y \geq f(x)\}$ be the region above the graph of f , $G_f = \{(x, y) : f(x^-) \leq y \leq f(x)\}$ be the completed graph of f , and for (x, y) in $[0, 1]^2$ we let $G_f(x, y) = G_f \cap ([x, 1] \times [0, y])$ be the part of the graph which lies below and to the right of (x, y) . Finally, we define $P_f : [0, 1]^2 \rightarrow [0, 1]$ as follows:

$$(25) \quad P_f(x, y) = \begin{cases} y - x + \min_{(x', y') \in G_f(x, y)} (x' - y') & \text{if } (x, y) \in H_f; \\ 0 & \text{if } y < f(x). \end{cases}$$

Note that $P_f(x, y) = \min_{(x', y') \in G_f} \|(x, y) - (x', y')\|_1$ if $(x, y) \in H_f$. In particular, $P_f(x, y) = 0$ if $(x, y) \in G_f$, and P_f is continuous. See Figure 8 for an illustration of (25).

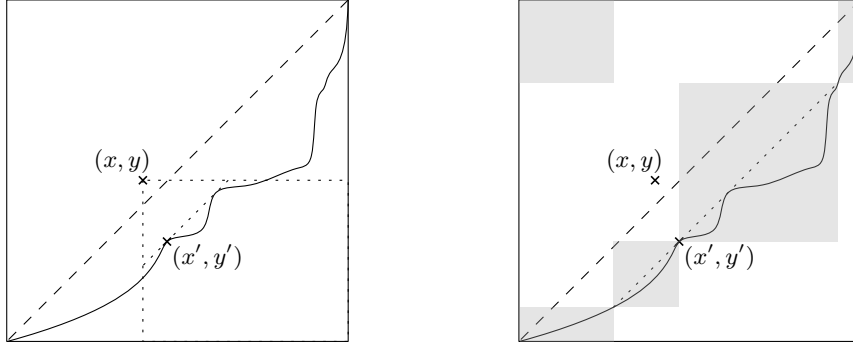


FIGURE 8. Illustration of (25). Left: the point (x', y') is one of the closest to the diagonal of $[0, 1]^2$ in $G_f(x, y)$. Right: It can actually be shown that the permuton with RLM-curve f has support contained in the shaded area (see the proof of Lemma 26), explaining the formula (25).

Lemma 25. *Let $f \in \mathcal{F}$. There exists a unique 231-avoiding permuton μ_f such that*

$$(26) \quad \text{for all } (x, y) \in [0, 1]^2, \quad \mu_f([x, 1] \times [0, y]) = P_f(x, y).$$

Moreover, the RLM-curve of μ_f is f .

Proof. The uniqueness is straightforward. It remains to construct a measure μ_f satisfying (26) and to verify that it is a permuton, that it is 231-avoiding, and that its RLM-curve is f .

We first check that for any pair $(x_1, y_1), (x_2, y_2)$ in H_f with $x_1 < x_2$ and $y_1 > y_2$, we have

$$(27) \quad P_f(x_1, y_1) - P_f(x_1, y_2) - P_f(x_2, y_1) + P_f(x_2, y_2) \geq 0.$$

It is equivalent to check that

$$\min_{(x', y') \in G_f(x_1, y_1)} (x' - y') - \min_{(x', y') \in G_f(x_1, y_2)} (x' - y') - \min_{(x', y') \in G_f(x_2, y_1)} (x' - y') + \min_{(x', y') \in G_f(x_2, y_2)} (x' - y') \geq 0.$$

But the above follows from Lemma 35 since $G_f(x_1, y_2) \cap G_f(x_2, y_1) = G_f(x_2, y_2)$ and $G_f(x_1, y_2) \cup G_f(x_2, y_1) = G_f(x_1, y_1)$. For the latter equality, it is important to note that $(x_2, y_2) \in H_f$. Hence (27) is proved.

We now define a pre-measure μ_f on finite unions of semi-open rectangles included in H_f by

$$\mu_f([x_1, x_2] \times (y_2, y_1]) = P_f(x_1, y_1) - P_f(x_1, y_2) - P_f(x_2, y_1) + P_f(x_2, y_2).$$

By Carathéodory's extension theorem, we can extend μ_f to a Borel measure on the set H_f that we still denote μ_f . This measure μ_f can be further extended to a measure on $[0, 1]^2$ by putting no mass on the complement set $\{(x, y) \in [0, 1]^2 : y < f(x)\}$. Note that μ_f indeed satisfies (26). Let us prove that it is a permuton: for $y \in [0, 1]$, we have

$$\mu_f([0, 1] \times [0, y]) = P_f(0, y) = y - \min_{(x', y') \in G_f(0, y)} (x' - y').$$

But since $f(0) = 0$ and $f(x) \leq x$, the minimum is reached for $x' = y' = 0$, and we get $\mu_f([0, 1] \times [0, y]) = y$. Similarly, $\mu_f([x, 1] \times [0, 1]) = 1 - x$ (in this case, the minimum is reached for $x' = y' = 1$), proving that μ_f is a permuton.

The next step is to prove that μ_f is 231-avoiding. The proof will be illustrated by Figure 9. Assume for the sake of contradiction that we have three points $(x_1, y_1), (x_2, y_2)$ and (x_3, y_3) in the support of μ_f such that $x_1 < x_2 < x_3$ and $y_3 < y_1 < y_2$. Since (x_3, y_3) is both the lowermost and rightmost point of the three, the points $(x_1, y_1), (x_2, y_2)$ cannot lie on the curve G_f , and we

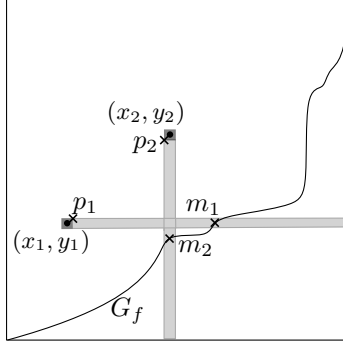


FIGURE 9. Illustration of the proof of Lemma 25. Left: we consider two hypothetical points (x_1, y_1) and (x_2, y_2) with $y_1 < y_2$ and $x_1 < x_2 < f^{-1}(y_1)$. The minimum of $x' - y'$ on the curve G_f below and on the right of $p_1 := (x_1 + \varepsilon, y_1 + \varepsilon)$ is reached somewhere in the horizontal light gray band, namely in m_1 . On the other hand, the minimum of $x' - y'$ on the curve G_f below and on the right of $p_2 := (x_2 - \varepsilon, y_2 - \varepsilon)$ is reached in the vertical light gray band, resp. m_2 . This is impossible, since m_1 would need to be at the same time closer and further from the main diagonal than m_2 .

necessarily have $y_i > f(x_i)$ for $i \in \{1, 2\}$. Moreover, since (x_3, y_3) lies weakly above G_f , we can write

$$(28) \quad x_2 < x_3 \leq f^{-1}(y_3) \leq f^{-1}(y_1).$$

For small $\varepsilon > 0$, we have $\mu_f([x_1 - \varepsilon, x_1 + \varepsilon] \times [y_1 - \varepsilon, y_1 + \varepsilon]) > 0$. Looking at the strict inequality case in (27) and in Lemma 35, this implies that the minimum of $x' - y'$ over $G_f(x_1 + \varepsilon, y_1 + \varepsilon)$ is reached only on the horizontal band

$$[x_1 + \varepsilon, 1] \times [y_1 - \varepsilon, y_1 + \varepsilon].$$

Let m_1 be a point in this band where the minimum is attained, as on Figure 9. Similarly, the minimum of $x' - y'$ over $G_f(x_2 - \varepsilon, y_2 - \varepsilon)$ is reached only on the vertical band

$$[x_2 - \varepsilon, x_2 + \varepsilon] \times [0, y_2 - \varepsilon]$$

and it is reached at a point m_2 . By (28), we have $x_1 < x_2 < f^{-1}(y_1)$ so for ε small enough, the intersection of G_f with the band $[x_1 + \varepsilon, 1] \times [y_1 - \varepsilon, y_1 + \varepsilon]$ is included in $G_f(x_2 - \varepsilon, y_2 - \varepsilon)$, so $m_1 \in G_f(x_2 - \varepsilon, y_2 - \varepsilon)$. But the definition of m_2 implies that m_2 is strictly closer than m_1 to the main diagonal. Vice-versa, we also have $m_2 \in G_f(x_1 + \varepsilon, y_1 + \varepsilon)$, so by definition m_1 is strictly closer than m_2 to the main diagonal, and we get a contradiction (see Figure 9). This proves that μ_f is 231-avoiding.

Finally, we verify that the RLM-curve of μ_f is f . Let us call g this RLM-curve. That $g(x) \geq f(x)$ is trivial since for any $y < f(x)$, one has $\mu_f([x, 1] \times [0, y]) = 0$. Let us prove the reverse inequality. Fix x in $[0, 1]$, and consider $y > f(x)$. Since f is càdlàg, there exists $\varepsilon > 0$ such that for $x' \in [x, x + \varepsilon]$, we have $f(x') \leq y - \varepsilon$. This implies that all points (x', y') on $G_f(x, y)$ are at L^1 -distance at least ε from (x, y) (either they satisfy $x' \geq x + \varepsilon$ or $y' \leq y - \varepsilon$), so that $\mu_f([x, 1] \times [0, y]) \geq \varepsilon > 0$, so $g(x) \leq y$. We conclude by letting $y \rightarrow f(x)$. \square

Our next goal is to prove that the map $\mu \mapsto f_\mu$ is one-to-one on the space of 231-avoiding permutations.

Lemma 26. *Let $f \in \mathcal{F}$. Then, the above constructed permutation μ_f is the unique 231-avoiding permutation whose RLM-curve is equal to f .*

Proof. Let ν be a 231-avoiding permuton whose RLM-curve is equal to f . A point (x, y) in $[0, 1]^2$ with $y \geq f(x)$ will be called an excursion corner of f if the minimum of $x' - y'$ on $G_f(x, y)$ is attained both at $(x, f(x))$ and at $(f^{-1}(y), y)$. We let $S_{x,y}$ be the corresponding square $[x, f^{-1}(y)] \times [f(x), y]$. We claim that necessarily $\nu(S_{x,y}) = f^{-1}(y) - x = y - f(x)$, i.e. that there is no mass on the left, on the right, below or above this square.

That there is no mass below and on the right of $S_{x,y}$ is trivial since f is the RLM-curve of ν . Assume that there is some mass on the left, then we can find a point (x_1, y_1) in the support of ν , verifying $x_1 < x$ and $f(x) < y_1 < y$ and $\nu([0, x] \times [f(x), y_1]) > 0$. Consider the square $S = [x, x + y_1 - f(x)] \times [f(x), y_1]$; see Figure 10, left. Since ν is a permuton with $\nu([0, x] \times [f(x), y_1]) > 0$, we have

$$\nu(S) < \nu([0, 1] \times [f(x), y_1]) = y_1 - f(x).$$

Thus there is some mass of ν above S (since f is the RLM curve of ν , there cannot be any mass of ν below S), i.e. there exists a point (x_2, y_2) in the support of ν with $x < x_2 < x + y_1 - f(x)$ and $y_2 > y_1$. Since (x, y) is an excursion corner, the restriction of the curve G_f to $[x, x + y_1 - f(x)]$ lies inside S and we can find (x_3, y_3) on $G_f \cap (x_2, x + y_1 - f(x)] \times [f(x), y_1]$. The next step is to let x'_3 be maximal such that (x'_3, y_3) is on G_f ; note that (x'_3, y_3) might be outside S , but this ensures that (x'_3, y_3) is in the support of ν (while (x_3, y_3) might not be). Finally, we have constructed three points (x_1, y_1) , (x_2, y_2) and (x'_3, y_3) which are in the support of ν and satisfy $x_1 < x < x_2 < x_3 \leq x'_3$ and $y_3 < y_1 < y_2$, i.e. form a 231 pattern. This is in contradiction with the assumption that ν is 231-avoiding. We have proved that, for any excursion corner (x, y) there is no mass on the left of $S_{x,y}$. A similar argument ensures that there is no mass above $S_{x,y}$ either.

We now consider any point $(x, y) \in (0, 1)^2$ (not necessarily an excursion corner) weakly above the RLM-curve f and we let (x', y') minimizing the quantity $x' - y'$ on the set $G_f(x, y)$. We denote by (x_l, y_l) the rightmost point on G_f for which $x_l < x$ and $x_l - y_l = x' - y'$. By construction (x_l, y_l) is an excursion corner of f . See Figure 10, right. Similarly, let (x_r, y_r) be the leftmost point on G_f satisfying $y_r \geq y$ and $x_r - y_r = x' - y'$. Again (x', y_r) is an excursion corner of f . Using that ν has no mass below G_f , nor on the left of or above excursion squares, we can write

$$\nu([x, 1] \times [0, y]) = \nu([x, x'] \times [y_l, y']) + \nu([x', x_r] \times [y', y]) = x' - x + y - y'.$$

Hence ν coincides with the permuton μ_f introduced in Lemma 25. \square

We end this section with a more satisfying description of the permuton μ_f in a special case: if the RLM curve is strictly convex and symmetric, then the support of μ_f is the union of its graph with a piece of the antidiagonal, with explicit densities. This special case will be the case of interest for us later.

Proposition 27. *Let $f \in \mathcal{F}$, and let μ_f be the unique permuton with RLM curve f . Assume the following:*

- (1) f is increasing and continuous on $[0, 1]$. In particular, there is a unique $x^* \in (0, 1)$ such that $f(x^*) = 1 - x^*$.
- (2) The graph G_f of f is invariant under the symmetry $(x, y) \mapsto (1 - y, 1 - x)$, that is:

$$\text{for all } x \in [0, 1], \quad f(1 - f(x)) = 1 - x.$$

- (3) f is strictly convex.

Then μ_f equals $\mu_{\nearrow} + \mu_{\searrow}$ where:

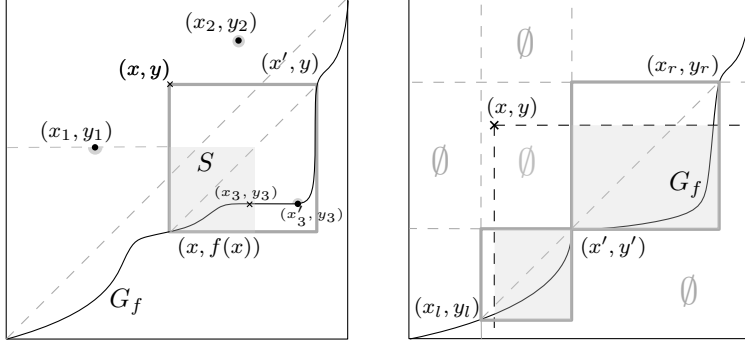


FIGURE 10. Illustration of the proof of Lemma 26. The black curve is the graph G_f of f , and gray fat squares represent excursions of f . The black disks with some gray zones around are points of the support of ν with some zone having a positive mass around them. The crosses are points that are not necessarily in the support. *Left*: we assume by contradiction that there is some mass below a point (x_1, y_1) on the left of the excursion square. Then there must be some mass around (x_2, y_2) which is above the light-gray square S (notation from the proof), and around a point (x'_3, y_3) on G_f , which creates the forbidden pattern 231. *Right*: We want to compute the $\nu([x, 1] \times [0, y])$. Empty set symbols indicate zones where ν does not have any mass. The mass $\nu([x, 1] \times [0, y])$ is obtained by summing the contributions of the two light gray rectangles.

- μ_{\nearrow} is the pushforward of the measure with cumulative distribution function (CDF)

$$F_{\nearrow} : x \in [0, 1] \mapsto \begin{cases} f(x) & \text{if } x \leq x^*; \\ 1 + x - 2x^* & \text{if } x \geq x^*; \end{cases}$$

via $x \mapsto (x, f(x))$;

- μ_{\searrow} is the pushforward of the measure with CDF

$$F_{\searrow} : x \in [0, 1] \mapsto \begin{cases} x - f(x) & \text{if } x \leq x^*; \\ 2x^* - 1 & \text{if } x \geq x^*; \end{cases}$$

via $x \mapsto (x, 1 - x)$.

The support of μ_{\nearrow} is the graph G_f of f , and the support of μ_{\searrow} is the antidiagonal line from $(1, 0)$ to $(x^*, f(x^*))$.

Proof. Let $\mu := \mu_{\nearrow} + \mu_{\searrow}$ be the measure defined by the proposed formulas. First, let us check that these measures are well-defined. By (1), the functions F_{\nearrow} and F_{\searrow} are continuous and the function F_{\nearrow} is increasing. Then, by (2) and (3), the function $x \mapsto x - f(x)$ is increasing on $[0, x^*]$ and decreasing on $[x^*, 1]$, and so F_{\searrow} is monotonous. Thus, these functions are indeed the CDFs of two measures on $[0, 1]$.

The claim about the supports of μ_{\nearrow} and μ_{\searrow} is a direct consequence of the assumptions made on f . Note that the support of μ_{\searrow} can be described as

$$\{(x, 1 - x) : x \leq x^*\} = \{(x, 1 - x) : 1 - x \geq f(x^*)\}.$$

In particular, it is immediate from the description of the support of μ that this measure is 231-avoiding and that its RLM curve is f .

Now, let us check that μ is a permuton. Since $F_{\nearrow}(x) + F_{\searrow}(x) = x$ for all $x \in [0, 1]$, the first marginal of μ is indeed uniform. It remains to check that the second marginal is uniform as

well. Fix $0 \leq y \leq z \leq 1$, and let us check that $\mu([0, 1] \times [y, z]) = z - y$. It suffices to check this in the two cases $z \leq f(x^*)$ and $y \geq f(x^*)$.

First, assume $z \leq f(x^*)$. The support of μ_{\searrow} does not intersect the band $[0, 1] \times [y, z]$. Also, $f^{-1}(z) \leq x^*$. Hence:

$$\mu([0, 1] \times [y, z]) = F_{\nearrow}(f^{-1}(z)) - F_{\nearrow}(f^{-1}(y)) = z - y.$$

as desired.

Now, assume $y \geq f(x^*)$. Since $f^{-1}(y) \geq x^*$, we have that:

$$\begin{aligned} \mu([0, 1] \times [y, z]) &= F_{\searrow}(1 - y) - F_{\searrow}(1 - z) + F_{\nearrow}(f^{-1}(z)) - F_{\nearrow}(f^{-1}(y)) \\ &= (1 - y - f(1 - y)) - (1 - z - f(1 - z)) + (f^{-1}(z) - f^{-1}(y)). \end{aligned}$$

By the symmetry of f , it holds that $f^{-1}(x) = 1 - f(1 - x)$ for all $x \in [0, 1]$. Therefore, $\mu([0, 1] \times [y, z]) = z - y$ as desired.

In conclusion, we have shown that μ is a 231-avoiding permuton with RLM curve f . By Lemma 26, this shows that $\mu_f = \mu$. \square

6.3. Bicontinuity of the RLM curve parametrization. Recall that \mathcal{F} is the space of non-decreasing càdlàg functions from $[0, 1]$ to $[0, 1]$ such that $f(1) = 1$ and $f(x) \leq x$ for all $x \in [0, 1]$. For a permuton μ , its RLM-curve f_μ lives in \mathcal{F} . To discuss continuity properties (and for large deviation principles later), it is convenient to introduce also φ_μ as the 1-Lipschitz function on $[0, 1]$ whose graph is obtained by rotating G_{f_μ} by $-\pi/4$, flipping it, and scaling it down by $\sqrt{2}$ (as on Figures 2 and 7). The function φ_μ is 1-Lipschitz, nonnegative, and satisfies $\varphi_\mu(0) = \varphi_\mu(1) = 0$. We recall that we denote by \mathcal{E} the space of such functions, endowed with the uniform norm.

Finally, we will also consider the (completed) graph G_f of f as living in the space of compact subsets of $[0, 1]^2$, endowed with the Hausdorff distance d_H . Namely for compact sets G_1 and G_2 ,

$$d_H(G_1, G_2) = \inf\{\varepsilon > 0 : G_1 \subseteq G_2^\varepsilon \text{ and } G_2 \subseteq G_1^\varepsilon\},$$

where for a compact set G and $\varepsilon > 0$, the ε -halo G^ε of G is the set of points at distance at most ε from G . This notion depends on an underlying metric on the ambient space, here $[0, 1]^2$, and we will use the supremum norm on $[0, 1]^2$ in the sequel.

Lemma 28. *Let f and f_n , $n \geq 1$, be functions in \mathcal{F} . Let φ and φ_n , $n \geq 1$, be the corresponding functions in \mathcal{E} . The following assertions are equivalent:*

- (i) $f_n(x) \rightarrow f(x)$ for all continuity points x of f ;
- (ii) $G_{f_n} \rightarrow G_f$ for the Hausdorff distance;
- (iii) $\varphi_n \rightarrow \varphi$ for the uniform distance.

Proof. (i) \implies (ii). Assume that $f_n \rightarrow f$ pointwise at continuity points of f . Fix $\varepsilon > 0$. Since f is nondecreasing, almost every point is a continuity point and we can construct a sequence $0 = x_0 < x_1 < \dots < x_{\ell-1} < x_\ell = 1$ of continuity points (except maybe x_ℓ , but this does not matter in the following) which is a subdivision of $[0, 1]$ with steps bounded by ε . For each $k < \ell$, $G_f \cap ((x_k, x_{k+1}] \times [0, 1])$ is a continuous nondecreasing curve from $(x_k, f(x_k))$ to $(x_{k+1}, f(x_{k+1}))$, and thus its ε -halo covers the rectangle $[x_k, x_{k+1}] \times [f(x_k) - \varepsilon, f(x_{k+1}) + \varepsilon]$.

For large enough n , we have $|f_n(x_k) - f(x_k)| < \varepsilon$ for all $k \in [0, \ell]$. Therefore $G_{f_n} \cap ((x_k, x_{k+1}] \times [0, 1])$ is contained in the rectangle $[x_k, x_{k+1}] \times [f(x_k) - \varepsilon, f(x_{k+1}) + \varepsilon]$. Since this holds for each $k < \ell$, this proves that for large enough n , the graph G_{f_n} is contained within the ε -halo of G_f . The reverse inclusion is proved with the same reasoning. This proves the Hausdorff convergence of completed graphs.

(ii) \implies (iii). Assume that $G_{f_n} \rightarrow G_f$ for the Hausdorff distance, and let G_φ and G_{φ_n} , $n \geq 1$, be their rotated-flipped-rescaled analogs. Obviously, $G_{\varphi_n} \rightarrow G_\varphi$ for the Hausdorff distance as

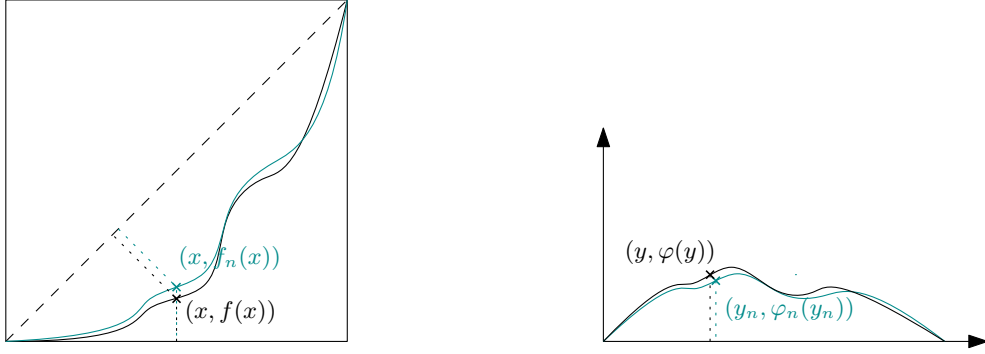


FIGURE 11. Construction of the sequence (y_n) in the proof of Lemma 28, (iii) \implies (i).

well. Now let $\varepsilon > 0$. For large enough n , the graph G_φ is contained within the ε -halo of G_{φ_n} . Consequently, for any x in $[0, 1]$, there exists $x_n \in [0, 1]$ such that $|x_n - x| < \varepsilon$ and $|\varphi_n(x_n) - \varphi(x)| < \varepsilon$. Since φ_n is 1-Lipschitz, we deduce that $|\varphi_n(x) - \varphi(x)| < 2\varepsilon$. This proves that $\varphi_n \rightarrow \varphi$ uniformly.

(iii) \implies (i). Assume that $\varphi_n \rightarrow \varphi$ uniformly, and let $x \in [0, 1]$ be a continuity point of f . Let $y \in [0, 1]$ be such that the point $(y, \varphi(y))$ corresponds to the point $(x, f(x))$ after rotation, flip, and rescaling, see Figure 11. Likewise, let $y_n \in [0, 1]$ be such that the point $(y_n, \varphi_n(y_n))$ corresponds to the point $(x, f_n(x))$.

Let us consider a subsequence y_{n_k} of y_n which converges to a limit point y_∞ . Since $\varphi_n \rightarrow \varphi$ uniformly, we get that $(y_{n_k}, \varphi_{n_k}(y_{n_k})) \rightarrow (y_\infty, \varphi(y_\infty))$. Let $(x_\infty, z_\infty) \in G_f$ be the point corresponding to $(y_\infty, \varphi(y_\infty))$, and note that $f(x_\infty^-) \leq z_\infty \leq f(x_\infty)$. By continuity of scaling operations, we have $(x, f_{n_k}(x)) \rightarrow (x_\infty, z_\infty)$. Therefore $x_\infty = x$, and since f is continuous at x , we deduce that $z_\infty = f(x)$. In particular, $y_\infty = y$, i.e. y is the only possible limit point of the sequence y_n . This implies $y_n \rightarrow y$ so $(y_n, \varphi_n(y_n)) \rightarrow (y, \varphi(y))$ and finally $(x, f_n(x))$ converges to $(x, f(x))$. \square

Remark 29. The equivalence $(G_{\varphi_n} \rightarrow G_\varphi) \iff (\varphi_n \rightarrow \varphi)$ is a special case of a general result relating convergence of functions and convergence of their graphs, see e.g. [Wat76].

Lemma 30. *The correspondence $\mu \in \mathcal{P}_{\text{Av}(231)} \mapsto \varphi_\mu \in \mathcal{E}$ is bicontinuous.*

Proof. Since \mathcal{E} is compact, it is sufficient to check the continuity of the reverse mapping. Therefore, let (φ_n) be a sequence in \mathcal{E} converging towards φ , and let (μ_n) and μ be the corresponding 231-avoiding permutons. Let (G_n) and G be the completed graphs of their RLM curves, and let (H_n) and H be the parts of $[0, 1]^2$ weakly above (G_n) and G . Thanks to Lemma 28, it holds that $G_n \rightarrow G$ for the Hausdorff distance, and therefore $\overline{[0, 1]^2 \setminus H_n} \rightarrow \overline{[0, 1]^2 \setminus H}$ as well. For any $x, y \in [0, 1]$, let $P_n(x, y) = \mu_n([x, 1] \times [0, y])$. We recall from (25) that

$$(29) \quad P_n(x, y) = \text{dist}_{L^1}((x, y), [0, 1]^2 \setminus H_n)$$

Since the map $\text{dist}_{L^1}((x, y), \cdot)$ is continuous on closed subsets of $[0, 1]^2$ for the Hausdorff distance, we get that

$$P_n(x, y) = \text{dist}_{L^1}((x, y), \overline{[0, 1]^2 \setminus H_n}) \rightarrow \text{dist}_{L^1}((x, y), \overline{[0, 1]^2 \setminus H}) = P(x, y).$$

Since this holds for all $(x, y) \in [0, 1]^2$, this proves that $\mu_n \rightarrow \mu$. \square

7. LARGE DEVIATION PRINCIPLE FOR 231-AVOIDING PERMUTATIONS

The goal of this section is to provide a proof of Proposition 4, that is, to establish a large deviation principle for φ_{σ_n} when σ_n is a uniform random 231-avoiding permutation of size n . As pointed out in the previous section, φ_{σ_n} is a uniform Dyck path of length $2n$. The proof will be similar to the proof of Proposition 5 but we again emphasize that, although the underlying discrete objects are the same, the topology in which we work is different in both cases. As in Part I, our strategy will consist in comparing our random objects with a simpler model.

Let $(\varepsilon_i)_{i \geq 1}$ be i.i.d. Rademacher random variables, $S_k := \sum_{i=1}^k \varepsilon_i$,

$$Z_n : t \in [0, 1] \mapsto \frac{1}{2n} S_{\lfloor 2nt \rfloor},$$

and W_n its continuous polygonal interpolation. We recall that the function J is defined in (4). We will write $\mathcal{C} = \mathcal{C}([0, 1])$, and we recall from (2) the definition of the compact subset $\mathcal{E} \subset \mathcal{C}$. By Mogulskii's theorem (see [Mog76], or [DZ98, Theorem 5.1.2]), $(W_n)_{n \geq 1}$ satisfies an LDP on \mathcal{C} at speed n with good rate function

$$H(\varphi) = \begin{cases} 2 \int_0^1 J(\varphi'(t)) dt & \text{if } \varphi \text{ is 1-Lipschitz and } \varphi(0) = 0; \\ \infty & \text{otherwise.} \end{cases}$$

The random variables φ_{σ_n} can be written as W_n conditioned on $\{S_{2n} = 0, \min_{k \leq 2n} S_k \geq 0\}$.

Proof of Proposition 4. For $\varphi \in \mathcal{C}$ and $\varepsilon > 0$, we denote by

$$B(\varphi, \varepsilon) := \{f \in \mathcal{C} : \|\varphi - f\|_\infty \leq \varepsilon\}$$

the ball of radius ε for the uniform norm. By construction, for $n \geq 1$, both φ_{σ_n} and W_n belong to the ball $B(0, 1) \subset \mathcal{C}$, which is compact, so it suffices to establish a weak LDP in $B(0, 1)$. As $\varphi_{\sigma_n} \in \mathcal{E}$, which is closed, Proposition 4 is obtained by restriction. More precisely, we have to show the following upper and lower bounds:

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\varphi_{\sigma_n} \in B(\varphi, \varepsilon)) \leq -H(\varphi)$$

and

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\varphi_{\sigma_n} \in B(\varphi, \varepsilon)) \geq -H(\varphi),$$

for any $\varphi \in \mathcal{E}$. Indeed, if $\varphi \notin \mathcal{E}$, for ε small enough and n large enough, the event $\{\varphi_{\sigma_n} \in B(\varphi, \varepsilon)\}$ is empty and the bounds are trivial.

The upper bound is easy to show. For any Borel set B in $\mathcal{C}([0, 1])$, counting paths gives

$$\begin{aligned} \mathbb{P}[\varphi_{\sigma_n} \in B] &= \frac{\#\{\text{Dyck paths of length } 2n \text{ in } B\}}{C_n} \\ &\leq \frac{\#\{\text{walks of length } 2n \text{ in } B\}}{C_n} = \frac{4^n}{C_n} \mathbb{P}[W_n \in B], \end{aligned}$$

where we recall that C_n is the Catalan number and $4^n/C_n = \Theta(n^{3/2})$ is subexponential. This is the same counting argument as in the proof of Proposition 5. Now, if we choose $B = B(\varphi, \varepsilon) \cap \mathcal{E}$, which is a closed subset of \mathcal{C} , and we use the LDP for (W_n) , we get the upper bound.

We now go to the proof of the lower bound. Fix $\varphi \in \mathcal{E}$ with $H^{231}(\varphi) < \infty$ and $\varepsilon > 0$. Note that φ is nonnegative and that $\varphi(0) = \varphi(1) = 0$.

For any possible path W_n , we denote by $(S_k)_{0 \leq k \leq 2n}$ the corresponding walk. We introduce $m_n := -\min_{k \leq 2n} S_k \geq 0$ and $a_n := S_{2n}$ and note that $a_n + 2m_n \geq 0$. If $W_n \in B(\varphi, \varepsilon)$, we have $m_n \leq 2\varepsilon n$ and $|a_n| \leq 2\varepsilon n$ since φ is nonnegative on $[0, 1]$ and $\varphi(1) = 0$. The scheme of the proof

of the lower bound is very similar to the proof of Proposition 5. We define a map T from paths to Dyck paths as follows:

- Flip the first m_n negative steps of the path to $+1$ (this lifts the minimum to 0).
- If $a_n + 2m_n > 0$, flip the last $\frac{a_n + 2m_n}{2}$ positive steps to -1 ; if $a_n + 2m_n = 0$, do nothing.

After the first item, the path is nonnegative and ends at height $a_n + 2m_n$, and the quantity $\frac{|a_n + 2m_n|}{2}$ is an integer since a_n is even. The second item then brings the endpoint to 0 without creating negative heights. The resulting path $T(W_n)$ is therefore a Dyck path of length $2n$. Moreover, since at most $m_n + \frac{|a_n + 2m_n|}{2} \leq 5\varepsilon n$ steps are flipped, the sup norm changes by at most 5ε .

Let us evaluate the number of preimages of $T(W_n)$. The choice of flipped steps is determined by which m_n negative steps are turned positive and which $\frac{|a_n + 2m_n|}{2}$ steps are flipped by the second item. As in the proof of Proposition 5, for ε such that $5\varepsilon < 1/2$, the number of possibilities is bounded by

$$\sum_{k \leq 5\varepsilon n} \binom{2n}{k} \leq 5\varepsilon n \binom{2n}{5\varepsilon n} \leq \frac{e\sqrt{10\varepsilon n}}{2\pi\sqrt{1-10\varepsilon}} ((10\varepsilon)^{10\varepsilon} (1-10\varepsilon)^{1-10\varepsilon})^{-n},$$

which is subexponential in n . Therefore

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\varphi_{\sigma_n} \in B(\varphi, 5\varepsilon)) \geq \lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_n \in B(\varphi, \varepsilon)) \geq -H(\varphi) = -H^{231}(\varphi),$$

where we used the lower bound of the LDP for (W_n) and the fact that $H(\varphi) = H^{231}(\varphi)$ whenever they are both finite. \square

As in Part I, the last step is to apply Lemma 13 to deduce an LDP for Mallows random permutations. We define the action A_β on \mathcal{E} by

$$A_\beta(\varphi) = H(\varphi) - 2\beta \int_0^1 \varphi(t) dt.$$

By the exact same argument as Corollary 19, we obtain the following.

Corollary 31. *Let $\beta \in \mathbb{R}$. Let τ_n^β be a Mallows random permutation with parameter $q_n := e^{\beta/n}$, conditioned to avoid 231. Then the function $\varphi_{\tau_n^\beta}$ satisfies a large deviation principle on \mathcal{E} at speed n with rate function*

$$\varphi \mapsto A_\beta(\varphi) - \inf A_\beta.$$

8. AN OPTIMIZATION PROBLEM FOR EXCURSIONS

We fix $\beta \in \mathbb{R}$.

Proposition 32. *The action A_β has a unique minimizer φ_β in \mathcal{E} . Moreover, for $\beta \leq 0$, we have $\varphi_\beta(t) = 0$ for all $t \in [0, 1]$. For $\beta > 0$, it is given by the formula*

$$(30) \quad \varphi_\beta(t) = \frac{1}{\beta} \log \left(\frac{1 + e^\beta}{1 + e^{\beta(1-2t)}} \right) - t.$$

Proof. We first note that the case $\beta \leq 0$ is immediate since $H^{231}(\varphi) \geq 0$ for all $\varphi \in \mathcal{E}$, with equality if and only if $\varphi \equiv 0$. We now assume $\beta > 0$.

As in Part 1, we will rely on Proposition 36. We first note that \mathcal{E} is compact for the topology of uniform convergence, and that the action A_β is strictly convex as the sum of a strictly convex term and a linear one. We now define the subset

$$\mathcal{E}' := \left\{ \varphi \in \mathcal{C}^1([0, 1]) : \begin{array}{l} \varphi(0) = \varphi(1) = 0, \varphi(t) > 0 \text{ for all } 0 < t < 1, \\ \varphi'(0) > 0, \varphi'(1) < 0, \|\varphi'\|_\infty < 1 \end{array} \right\},$$

and the set of bridges

$$\mathcal{B} := \{\alpha \in \mathcal{C}^1([0, 1]) : \alpha(0) = \alpha(1) = 0\} .$$

It is straightforward to check that $\varphi_\beta \in \mathcal{E}'$ and that if $\varphi \in \mathcal{E}'$ then $V_\varphi(\mathcal{E}) = \mathcal{B}$, where we recall that $V_\varphi(K)$ is the set of directions α such that $\varphi + t\alpha \in K$, for t small enough. Therefore, by Proposition 36, it is sufficient to check, for all $\alpha \in \mathcal{B}$,

$$\begin{aligned} 0 &= \frac{d}{ds} \Big|_{s=0^+} A_\beta(\varphi + s\alpha) \\ &= \int_0^1 (2\alpha'(t)J'(\varphi'_\beta(t)) - 2\beta\alpha(t)) dt \\ &= \int_0^1 (2J'(\varphi'_\beta(t)) + 2\beta t) \alpha'(t) dt . \end{aligned}$$

Since $\int \alpha' = 0$, it is sufficient to check that $J'(\varphi'_\beta(t)) + \beta t$ is constant, which is straightforward⁹. \square

We can now prove the asymptotic estimate (5) for the partition function of the model.

Proof of (5). The arguments are essentially the same as for equation (6). By Proposition 4 and the first estimate in Lemma 13, we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{Z_n^\beta}{Z_n^0} = -\inf A_\beta = -A_\beta(\varphi_\beta).$$

The latter can be (almost) explicitly computed. We have that

$$\varphi'_\beta(t) = \frac{e^{\beta(1-2t)} - 1}{e^{\beta(1-2t)} + 1}$$

and using the change of variable $y = e^{\beta(1-2t)}$, we get that

$$H(\varphi_\beta) = 2 \int_0^1 J(\varphi'_\beta(t)) dt = 2 \log 2 + \frac{1}{\beta} \int_{e^{-\beta}}^{e^\beta} \left(\frac{\log y}{1+y} - \frac{\log(1+y)}{y} \right) dy.$$

On the other hand,

$$2\beta \int_0^1 \varphi_\beta(t) dt = 2 \log(1 + e^\beta) - \beta - \frac{1}{\beta} \int_{e^{-\beta}}^{e^\beta} \frac{\log(1+y)}{y} dy,$$

leading to the asymptotics announced in equation (5). \square

We now conclude by proving the limit shape for 231-avoiding Mallows permutations.

Proof of Theorem 6. It follows immediately from Corollary 31 and Proposition 32 that $\varphi_{\tau_n^\beta}$ converges almost surely (for any coupling) to φ_β . By Lemma 30, this implies the almost sure convergence of τ_n^β to the unique permuton μ_β^{231} such that $\varphi_{\mu_\beta^{231}} = \varphi_\beta$. The only thing left to check is that μ_β^{231} corresponds to the description given by Theorem 6. We first notice that for $\beta \leq 0$, we have $\varphi_\beta = 0$, and the diagonal permuton μ_{\diagdown} indeed satisfies $f_{\mu_{\diagdown}}(t) = t$ for $0 \leq t \leq 1$, so $\varphi_{\mu_{\diagdown}} \equiv 0$.

We now assume $\beta > 0$. Let $f_\beta \in \mathcal{F}$ be obtained from $\varphi_\beta \in \mathcal{E}$ by flipping, rotating and rescaling as described in Section 6.3. The graph of f_β is characterized by

$$\{(x, f_\beta(x)) : x \in [0, 1]\} = \{(t + \varphi_\beta(t), t - \varphi_\beta(t)) : t \in [0, 1]\} .$$

⁹Alternatively, as in Remark 22, we could recover the formula for φ_β by solving the differential equation $\log\left(\frac{1+\varphi'}{1-\varphi'}\right) = -2\beta t + c$.

Plugging in the definition (30) of φ_β , we obtain

$$f_\beta(x) = 1 - \frac{1}{\beta} \log \left(1 + e^\beta - e^{\beta x} \right).$$

From here, we deduce an explicit description of μ_β using Proposition 27. Indeed, $\varphi_\beta(1-t) = \varphi_\beta(t)$ for all $t \in [0, 1]$, and thus the graph of f_β is symmetric with respect to the antidiagonal. The other assumptions of Proposition 27 are easy to check, and we have

$$x_\beta^* = \frac{1}{\beta} \log \left(\frac{1 + e^\beta}{2} \right), \quad f_\beta(x_\beta^*) = 1 - \frac{1}{\beta} \log \left(\frac{1 + e^\beta}{2} \right).$$

Furthermore, for all $x \in [0, 1]$, we have

$$f'_\beta(x) = \frac{e^{\beta x}}{1 + e^\beta - e^{\beta x}}.$$

Therefore, it follows from Proposition 27 that μ_β is the permuton with densities $1 \wedge \frac{e^{\beta x}}{1 + e^\beta - e^{\beta x}}$ on $\{(x, f_\beta(x)) : x \in [0, 1]\}$ and $0 \vee \frac{1 + e^\beta - 2e^{\beta x}}{1 + e^\beta - e^{\beta x}}$ on $\{(x, 1-x) : x \in [0, 1]\}$. Note that the first density (resp. the second) is equal to 1 (resp. to 0) if and only if $x \geq x_\beta^*$. This fits the formula given in Theorem 6. \square

Part 3. Appendix

APPENDIX A. MARKOV CHAIN SIMULATION

In this appendix, we explain how to generate the random permutations $\tau_n^{\beta, \alpha}$ for $\alpha \in \{321, 231\}$, explaining how we got the simulations of Figures 1 and 3. This relies on interpreting the law of $\tau_n^{\beta, \alpha}$ as the stationary measure of an adequate Markov chain on the set $\text{Av}_n(\alpha)$ of α -avoiding permutations. Specifically, we use a reversible Markov chain and the following standard lemma.

Lemma 33. *Let X be a finite state space, $q \geq 1$, and $f : X \rightarrow \mathbb{R}_+$. Write μ^f for the (q, f) -tilted measure on X , that is, $\mu^{q, f}(x) \propto q^{f(x)}$ for all $x \in X$. If $p : X^2 \rightarrow [0, 1]$ is a transition kernel, denote by \tilde{p} the transition kernel characterized by:*

$$\forall (x, y) \in X^2, \quad \tilde{p}(x, y) = \begin{cases} p(x, y) & \text{if } x \neq y \text{ and } f(y) \geq f(x); \\ p(x, y)q^{f(y)-f(x)} & \text{if } x \neq y \text{ and } f(y) < f(x); \end{cases}$$

and $\tilde{p}(x, x)$ is adjusted so that this defines a transition kernel. If p is reversible for the uniform measure on X , then \tilde{p} is reversible for the measure $\mu^{q, f}$.

In words, the tilted transition kernel \tilde{p} is constructed as follows: from $x \in X$, use p to find a transition towards some $y \in X$. If $f(y) \geq f(x)$, accept this transition. If $f(y) < f(x)$, accept this transition with probability $q^{f(y)-f(x)}$; in case of reject, stay at x .

Proof. This is a straightforward formal check. Let $(x, y) \in X$, and assume w.l.o.g. that $x \neq y$ and $f(y) \leq f(x)$. Therefore $\tilde{p}(x, y) = p(x, y)q^{f(y)-f(x)}$ and $\tilde{p}(y, x) = p(y, x)$, and thus $q^{f(x)}\tilde{p}(x, y) = q^{f(y)}p(x, y) = q^{f(y)}p(y, x) = q^{f(y)}\tilde{p}(y, x)$, where we used the reversibility of p in the second equality. \square

Finding a reversible Markov chain for $\tau_n^{\beta, \alpha}$ then boils down to finding a reversible Markov chain for the uniform measure on $\text{Av}_n(\alpha)$, for which the inversion differential $\text{inv}(\tau) - \text{inv}(\sigma)$ can be efficiently computed on the transitions $\sigma \mapsto \tau$.

Recall that $\text{Av}_n(\alpha)$ is in one-to-one correspondence with the set Dyck_n of Dyck paths of length $2n$, both for $\alpha = 321$ and $\alpha = 231$. If $d : [0, 2n] \rightarrow \mathbb{R}_+$ is a Dyck path and $i \in [1, 2n-1]$, transform d into $d^{(i)} \in \text{Dyck}_n$ as follows:

- If i is a peak index for d , i.e. if $d(i-1) = d(i) - 1 = d(i+1)$, and if $d(i) \geq 2$, then transform this peak into a valley in $d^{(i)}$. That is, let $d^{(i)}(j) := d(j)$ for all indices $j \neq i$, and let $d^{(i)}(i) := d(i) - 2$.
- If i is a valley index for d , i.e. if $d(i-1) = d(i) + 1 = d(i+1)$, then transform this valley into a peak in $d^{(i)}$. That is, let $d^{(i)}(j) := d(j)$ for all indices $j \neq i$, and let $d^{(i)}(i) := d(i) + 2$.
- In all other cases, let $d^{(i)} := d$.

Then define a transition kernel on Dyck_n via $p(d, \cdot) = \frac{1}{2^{n-1}} \sum_{i=1}^{2^{n-1}} \delta_{d^{(i)}}$, that is, apply the above transform at a uniformly random index. It is straightforward to check that this defines an irreducible Markov chain on Dyck_n that is reversible for the uniform measure.

Hence we can approximate the uniform measure on $\text{Av}_n(\alpha)$ by running this Markov chain on Dyck_n for “a long time”, and then applying the one-to-one correspondence from Dyck_n to $\text{Av}_n(\alpha)$. To approximate $\tau_n^{\beta, \alpha}$ instead, it suffices to use the transition kernel \tilde{p} given by Lemma 33. In the case $\alpha = 321$, $\text{inv}(d^{(i)}) - \text{inv}(d)$ can be computed in constant time thanks to Lemma 16. This is even simpler for $\alpha = 231$ thanks to Proposition 24: $\text{inv}(d^{(i)}) - \text{inv}(d)$ is equal to $+1$ if a valley was transformed into a peak, and to -1 if a peak was transformed into a valley.

Since we do not know the mixing time of the above Markov chains¹⁰, we use the following standard heuristics. We start from two very different initial conditions (e.g., the Dyck paths with minimal and maximal global heights respectively; the first one corresponds to the permutations id , whereas the second one corresponds to $23 \dots (n-1)1$ for $\alpha = 321$ and to $n(n-1) \dots 1$ for $\alpha = 231$), perform the Markov chain steps on both (using the same random index i on both at each step), and compare the resulting pictures. If the two pictures look similar, it means that, heuristically, the number of steps was large enough so that the chain forgets its initial condition. In practice, we get satisfying results after 100 million steps; to be on the safe side, we used 300 million steps to generate the pictures of Figures 1 and 3.

APPENDIX B. USEFUL LEMMAS

Lemma 34. *Let μ_n be a sequence of finite measures on a space Ω converging in total variation to a finite measure μ , and f_n a sequence of measurable functions on Ω converging μ -a.e. to a function f . Then $(f_n)_{\#} \mu_n$ converges weakly to $f_{\#} \mu$.*

Proof. The distance $d_{\text{TV}}((f_n)_{\#} \mu_n, (f_n)_{\#} \mu)$ is at most $d_{\text{TV}}(\mu_n, \mu)$, which tends to 0. Hence it is enough to prove that $(f_n)_{\#} \mu$ converges weakly to $f_{\#} \mu$. Letting X be a random variable with law $\mu/\mu(\Omega)$, it holds by assumption that $f_n(X)$ converges almost surely to $f(X)$. Since almost sure convergence implies convergence in distribution, the law $(f_n)_{\#} \mu/\mu(\Omega)$ of $f_n(X)$ converges weakly to that of $f(X)$, which is $f_{\#} \mu/\mu(\Omega)$. \square

Lemma 35. *Let us consider two sets A and B and a function h on $A \cup B$. It holds that*

$$(31) \quad \inf_{z \in A \cup B} h(z) + \inf_{z \in A \cap B} h(z) \geq \inf_{z \in A} h(z) + \inf_{z \in B} h(z).$$

Moreover, if the inequality is strict, then

$$(32) \quad \inf_{z \in A} h(z) = \inf_{z \in A \setminus B} h(z) < \inf_{z \in A \cap B} h(z),$$

and the same holds if we switch the roles of A and B .

¹⁰It is known that the Markov chain p on Dyck_n has mixing time $O(n^3 \log(n))$, see e.g. [MT99, Wil04], but this does not give us an estimate for the mixing time of \tilde{p} .

Proof. We can assume without loss of generality that $\inf_{z \in A \cup B} h(z) = \inf_{z \in B} h(z)$. Then (31) follows from the trivial inequality $\inf_{z \in A \cap B} h(z) \geq \inf_{z \in A} h(z)$. This inequality being strict implies

$$\inf_{z \in A \cap B} h(z) > \inf_{z \in A} h(z) \geq \inf_{z \in A \cup B} h(z) = \inf_{z \in B} h(z).$$

The first inequality implies (32), and the strict inequality between the first and last terms implies (32) with A and B switched. \square

Optimization on convex spaces. Let E be a topological vector space and $K \subset E$ convex. We say that $f : K \rightarrow \mathbb{R}$ is strictly convex if for any $x \in K$ and $y \in E$, the function $t \rightarrow f(x + ty)$ is strictly convex on the interval where it is defined (with the convention that a function defined on a singleton of \mathbb{R} is strictly convex). For any $x \in K$, we denote by $V_x(K)$ the set of directions $y \in E$ such that $x + ty \in K$ for $t > 0$ small enough.

Proposition 36. *Let E be a topological vector space and $K \subset E$ compact and convex, and let $f : K \rightarrow \mathbb{R}$ be lower semicontinuous and strictly convex. Then f has a minimum which is attained at exactly one point $x_0 \in K$. Moreover, if $x \in K$ is such that for any $y \in V_x(K)$, we have*

$$\left. \frac{d}{dt} \right|_{t=0^+} f(x + ty) = 0,$$

then $x = x_0$.

Proof. The existence of the minimum is immediate by lower semicontinuity. Moreover, if $f(x_1) = f(x_2)$ with $x_1 \neq x_2$, then the strict convexity of f implies $f\left(\frac{x_1+x_2}{2}\right) < f(x_1)$, which gives the uniqueness of the minimum. Finally, if x satisfies the given condition and $x' \in K \setminus \{x\}$, then let $y = x' - x \in V_x(K)$. The function $h : t \mapsto f(x + ty)$ has right-derivative 0 at $t = 0$ and is strictly convex, so $f(x') = h(1) > h(0) = f(x)$. We conclude that x must be the unique minimum x_0 of f on K , as claimed. \square

ACKNOWLEDGEMENTS

This work was initiated during an open problem workshop organized as part of the ANR project LOUCCCOUM (Large Objects Under Combinatorial Constraints and Outside Uniform Models, ANR-24-CE40-7809). All authors are partially supported by this grant.

REFERENCES

- [AGZ10] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Camb. Stud. Adv. Math.* Cambridge: Cambridge University Press, 2010.
- [AM14] M. Atapour and N. Madras. Large deviations and ratio limit theorems for pattern-avoiding permutations. *Comb. Probab. Comput.*, 23(2):161–200, 2014.
- [BBF⁺22] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, and A. Pierrot. Scaling limits of permutation classes with a finite specification: a dichotomy. *Advances in Mathematics*, 405:108513, 2022.
- [BDMW24] J. Borga, S. Das, S. Mukherjee, and P. Winkler. Large deviation principle for random permutations. *Int. Math. Res. Not.*, 2024(3):2138–2191, 2024.
- [Bón12] M. Bóna. *Combinatorics of permutations*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, second edition, 2012.
- [Bor22] J. Borga. The permuton limit of strong-Baxter and semi-Baxter permutations is the skew Brownian permuton. *Electron. J. Probab.*, 27:53, 2022. Id/No 158.
- [Cal07] D. Callan. Bijections from Dyck paths to 321-avoiding permutations revisited, 2007. Preprint arXiv:0711.2684.
- [CEKS13] S.-E. Cheng, S. Elizalde, A. Kasraoui, and B. E. Sagan. Inversion polynomials for 321-avoiding permutations. *Discrete Math.*, 313(22):2552–2565, 2013.

- [CP26] A. Chelikavada and H. Panzo. Limit theorems for fixed point biased permutations avoiding a pattern of length three. *Discrete Mathematics & Theoretical Computer Science*, vol. 28:2, Mar 2026.
- [DGM24] L. Dort, C. Goldschmidt, and G. Miermont. A large deviation principle for the normalized excursion of an α -stable Lévy process without negative jumps. *ALEA Lat. Am. J. Probab. Math. Stat.*, 21(2):1625–1653, 2024.
- [DHJ08] P. Diaconis, S. Holmes, and S. Janson. Threshold graph limits and random threshold graphs. *Internet Math.*, 5(3):267–320, 2008.
- [DHJ13] P. Diaconis, S. Holmes, and S. Janson. Interval graph limits. *Ann. Comb.*, 17(1):27–52, 2013.
- [Dub25] V. Dubach. A geometric approach to conjugation-invariant random permutations. *Probab. Theory Relat. Fields*, 194:245–298, 2025.
- [Dub26] V. Dubach. Classical patterns in Mallows permutations. *Electronic Journal of Probability*, 31:1–39, 2026.
- [DZ98] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.*, volume 38 of *Appl. Math. (N. Y.)*. New York, NY: Springer, 2nd ed. edition, 1998.
- [FK24] V. Féray and M. S. Kammoun. Asymptotic normality of pattern counts in conjugacy classes. *Electron. J. Probab.*, 29:22, 2024. Id/No 56.
- [GHKP24] F. Garbe, J. Hladký, G. Kun, and K. Pekárková. On pattern-avoiding permutons. *Random Struct. Algorithms*, 65(1):46–60, 2024.
- [GK25] V. Gorin and R. Kenyon. Six-vertex model with rare corners and random restricted permutations. *Probability Theory and Related Fields*, 2025.
- [Hag08] J. Haglund. *The q, t -Catalan numbers and the space of diagonal harmonics. With an appendix on the combinatorics of Macdonald polynomials*, volume 41 of *Univ. Lect. Ser.* Providence, RI: American Mathematical Society (AMS), 2008.
- [HKM⁺13] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth, and R. Menezes Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory. Series B*, 103(1):93–113, 2013.
- [HR25] Z. Hamaker and B. Rhoades. Moments of permutation statistics by cycle type. Preprint, 2025.
- [HRS17] C. Hoffman, D. Rizzolo, and E. Slivken. Pattern-avoiding permutations and Brownian excursion, Part I: shapes and fluctuations. *Random Structures & Algorithms*, 50(3):394–419, 2017.
- [HRS19] C. Hoffman, D. Rizzolo, and E. Slivken. Scaling limits of permutations avoiding long decreasing sequences. Preprint, 2019.
- [HS25] K. Hohmeier and E. Slivken. Permuton limits for some permutations avoiding a single pattern. Preprint, 2025.
- [JU17] S. Janson and A. J. Uzzell. On string graph limits and the structure of a typical string graph. *J. Graph Theory*, 84(4):386–407, 2017.
- [KKRW20] R. Kenyon, D. Král’, C. Radin, and P. Winkler. Permutations with fixed pattern densities. *Random Struct. Algorithms*, 56(1):220–250, 2020.
- [Mal57] C. L. Mallows. Non null ranking models. I. *Biometrika*, 44:114–130, 1957.
- [ML10] N. Madras and H. Liu. Random pattern-avoiding permutations. In *Algorithmic probability and combinatorics. Papers from the AMS special sessions, Chicago, IL, USA, October 5–6, 2007 and Vancouver, BC, Canada, October 4–5, 2008*, pages 173–194. Providence, RI: American Mathematical Society (AMS), 2010.
- [Mog76] A. A. Mogul’skij. Large deviations for trajectories of multi-dimensional random walks. *Theory Probab. Appl.*, 21:300–315, 1976.
- [MP14] S. Miner and I. Pak. The shape of random pattern-avoiding permutations. *Adv. Appl. Math.*, 55:86–130, 2014.
- [MPY14] A. Mogulskii, E. Pechersky, and A. Yambartsev. Large deviations for excursions of non-homogeneous Markov processes. *Electron. Commun. Probab.*, 19:no. 37, 8, 2014.
- [MT99] L. McShine and P. Tetali. On the mixing time of the triangulation walk and other Catalan structures. In *Randomization methods in algorithm design. DIMACS workshop, Princeton Univ., NJ, USA, December 12–14, 1997*, pages 147–160. Providence, RI: AMS, American Mathematical Society, 1999.
- [Muk16] S. Mukherjee. Estimation in exponential families on permutations. *Ann. Stat.*, 44(2):853–875, 2016.

- [Pin21] R. G. Pinsky. Permutations avoiding a pattern of length three under Mallows distributions. *Random Struct. Algorithms*, 58(4):676–690, 2021.
- [PR25] J. Park and D. Rizzolo. Limit theorems for fixed point biased pattern avoiding involutions, 2025. preprint arXiv:2512.25006.
- [Sta09] S. Starr. Thermodynamic limit for the Mallows model on S_n . *J. Math. Phys.*, 50(9):nb. 095208, 15 pp., 2009.
- [Tra08] J. Trashorras. Large deviations for symmetrised empirical measures. *J. Theor. Probab.*, 21(2):397–412, 2008.
- [Vat15] V. Vatter. Permutation classes. In *Handbook of enumerative combinatorics*, pages 753–833. Boca Raton, FL: CRC Press, 2015.
- [Wat76] W. C. Waterhouse. Uniform convergence and graph convergence. *Am. Math. Mon.*, 83:641–643, 1976.
- [Wil04] D. B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *The Annals of Applied Probability*, 14(1):274–325, 2004.

ENS DE LYON, CNRS, UMPA

Email address: `thomas.budzinski@ens-lyon.fr`

DEPARTMENT OF MATHEMATICS, UPPSALA UNIVERSITY, SWEDEN

Email address: `victor.dubach@math.uu.se`

UNIVERSITÉ DE LORRAINE, CNRS, IECL, F-54000 NANCY, FRANCE

Email address: `valentin.feray@univ-lorraine.fr`

UNIVERSITÉ DE POITIERS, LMA, FRANCE

Email address: `skammoun@math.univ-poitiers.fr`

UNIVERSITÉ DE LILLE, CNRS, INRIA, UMR 8524 - LABORATOIRE PAUL PAINLEVÉ, F-59000 LILLE, FRANCE

Email address: `mylene.maida@univ-lille.fr`