Report 2000-045

Moderate Deviations for Longest Increasing Subsequences: The Upper Tail

Franz Merkl and Matthias Löwe

ISSN: 1389–2355

MODERATE DEVIATIONS FOR LONGEST INCREASING SUBSEQUENCES: THE UPPER TAIL

Matthias Löwe[1] and Franz Merkl[2]

Abstract

We derive the upper tail moderate deviations for the length of a longest increasing subsequence in a random permutation. This concerns the regime between the upper tail large deviation regime and the central limit regime. Our proof uses a formula to describe the relevant probabilities in terms of the solution of a rank 2 Riemann-Hilbert problem (RHP); this formula was invented by Baik, Deift, and Johansson [3] to find the central limit asymptotics of the same quantities. In contrast to the work of these authors, who apply a third order (nonstandard) steepest descend approximation at an inflection point of the transition matrix elements of the RHP, our approach is based on a (more classical) second order (Gaussian) saddle point approximation at the stationary points of the transition function matrix elements.

2000 Mathematics Subject Classification: Primary 60F10, secondary 05A05, 45E05, 60C05, 14H60.
Key words: Ulam's problem, random permutations, moderate deviations, Riemann-Hilbert problems, saddle point approximation.

1 Introduction

Often combinatorics provides problems with non-standard and surprising limit theorems in probability theory. An example, that has attracted much attention especially in the past five years, is *Ulam's problem* : Consider the permutation group S_n on $\{1, \ldots, n\}$. We say that $1 \leq i_1 < \ldots < i_k \leq n$ is an increasing subsequence of length k of $\pi \in S_n$ iff $\pi(i_1) < \ldots < \pi(i_k)$. The length of a longest increasing subsequence of a permutation π will be denoted by $L_n := L_n(\pi)$; such a subsequence in general is not unique. Ulam's problem asks for the typical asymptotic behaviour of L_n as $n \to \infty$, if π is chosen from S_n with uniform probability $1/n!$.

There is an alternative version of this problem: Take a Poisson process with intensity one in the plane. For a fixed realisation ω of this point process an up/right ω-path from $(0,0)$ to (t,t) is a polygonal path starting in $(0,0)$, ending in (t,t), and connecting points from ω in such a way that it only moves upwards and to the right. Denote by $\mathcal{L}_t := \mathcal{L}_t(\omega)$ the maximal number of Poisson points in a up/right ω-path from $(0,0)$ to (t,t). Ordering

[1]Department of Mathematics, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands. loewe@sci.kun.nl

[2]Universität Bielefeld, Fakultät für Mathematik, Postfach 100131, D-33501 Bielefeld, Germany, and Eurandom, PO Box 513, NL-5600 MB Eindhoven, The Netherlands. merkl@eurandom.tue.nl

the x- and y-coordinates of ω induces a label (n_1, n_2) to every point, where n_1, resp. n_2 denote the order-number of the x-coordinate, resp. y-coordinate of the corresponding point. These labels are almost surely well defined. They induce a permutation π via $\pi(n_1) = n_2$. Conditioned on the number of points in ω, π is selected with uniform probability from all possible permutations. Hence $\mathcal{L}_{\sqrt{\lambda}}$ has the same distribution as L_N, where N is a Poisson random variable with expected value $\lambda = t^2$.

Ulam's problem also has connections to various other mathematical topics. For example, by the *Schensted correspondence* there is a bijection between permutations $\pi \in S_n$ and pairs of n-Young tableaux of equal shape with length $L_n(\pi)$ of the first row. Other connections are to Ulam's metric, patience sorting, random matrices, and to the *Hammersley process*. For a survey over recent developments in Ulam's problem and explanations of the cross-connections mentioned above, the reader is referred to a recent article by Deift [4].

Already Erdös and Szekeres [6] proved that for all n one has $\mathbb{E}_n[L_n] \geq \frac{1}{2}\sqrt{n-1}$ (where $\mathbb{E}_n$ denotes the expectation with respect to the uniform distribution on S_n). Ulam [17], on the basis of numerical simulations, found for $1 \leq n \leq 10$ that $\mathbb{E}[L_n] \approx 1.7\sqrt{n}$ and conjectured that

$$c := \lim_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}_n[L_n] \tag{1.1}$$

exists. This conjecture was proved by Hammersley in 1972 [10] by an application of the subadditive ergodic theorem. While rigorously establishing the existence of c, Hammersley did not give a numerical value for c. After approximating steps by Kingman [13], Logan and Shepp [14] and independently Kerov and Vershik [12] showed in 1977 that $c = 2$. Following ideas of Logan and Shepp [14], Deuschel and Zeitouni [5] determined the following lower tail large deviation principle for L_n: For $0 < x < 2$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[L_n \leq x\sqrt{n}] = -1 + \frac{x^2}{4} + 2\log\frac{x}{2} - 2\left(1 + \frac{x^2}{4}\right)\log\left(\frac{2x^2}{4+x^2}\right). \tag{1.2}$$

The result (1.2) was derived using an analysis of Young diagrams.

The combinatorial work in the above papers could be replaced by a "hydrodynamical argument" to show the same result $c = 2$, in two papers by Aldous and Diaconis [1] and Seppäläinen [15]. This argument was presented in a pure way in a recent paper by Groeneboom [9], who again proved that $c = 2$. Yet a different proof of the same result was given by Johansson [11].

Based on the paper by Seppäläinen [15], Deuschel and Zeitouni [5] also derived the following upper tail large deviations: For $x > 2$:

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \log \mathbb{P}\left[L_n > x\sqrt{n}\right] = -2x \operatorname{arcosh}\frac{x}{2} + 2\sqrt{x^2-4}. \tag{1.3}$$

One observes the following asymptotics for the "lower end" of the upper tail:

$$\lim_{t\searrow 0} \lim_{n\to\infty} \frac{\log \mathbb{P}\left[L_n > (2+t)\sqrt{n}\right]}{t^{3/2} n^{1/2}} = -\frac{4}{3}. \tag{1.4}$$

While all these methods could compute the value of c and thus solve Ulam's original problem, none of them were appropriate to analyse other statistics of L_n such as the variance: until the mid 1990's there was only the conjecture that $\mathrm{Var}[L_n]$ asymptotically behaves like n^α with different values for α (among them the correct $\alpha = 1/3$ given by Kesten on the basis of arguments from first passage percolation).

Only in 1999 Baik, Deift, and Johansson [3] came up with a method based on the theory of integrable systems to prove a non-standard Central Limit Theorem (CLT) for the quantity L_n. Their result (Theorem 1.1 in [3]) can be stated as follows: Scale L_n as

$$\chi_n(\pi) := \frac{L_n(\pi) - 2\sqrt{n}}{n^{1/6}}. \tag{1.5}$$

Then χ_n converges in distribution as $n \to \infty$ to the Tracy-Widom distribution, introduced by Tracy and Widom in [16]. This distribution can be defined as follows: Let $u(x)$ be the solution to the Painlevé II equation

$$u_{xx} = 2u^3 + xu \quad \text{with} \quad u(x) \sim -\mathrm{Ai}(x) \sim -\frac{e^{-(2/3)x^{3/2}}}{2\sqrt{\pi}x^{1/4}} \quad \text{as } x \to \infty; \tag{1.6}$$

the notation $a \sim b$ means that the quotient of both sides converges to 1, and Ai denotes the Airy function. Then the Tracy-Widom distribution has the distribution function

$$F(t) = \exp\left(\int_t^\infty (x-t)u^2(x)dx\right). \tag{1.7}$$

Interestingly, the Tracy-Widom distribution first appeared in the context of eigenvalue statistics of the Gaussian Unitary ensemble.

One observes the following upper tail of the Tracy-Widom distribution:

$$1 - F(t) \sim \frac{e^{-(4/3)t^{3/2}}}{16\pi t^{3/2}} \quad \text{as } t \to \infty. \tag{1.8}$$

Hence the the following upper-tail asymptotics of the central limit regime holds:

$$\lim_{t\to\infty}\lim_{n\to\infty} \frac{\log \mathbb{P}\left[L_n > (2 + tn^{-1/3})\sqrt{n}\right]}{t^{3/2}} = -\frac{4}{3}. \tag{1.9}$$

In order to show that the moments of χ_N converge to the corresponding moments of the Tracy-Widom distribution, Baik, Deift and Johansson also derived the following rough upper bound for the upper tail probabilities; see formula (1.8) in [3]: For $M > 0$ sufficiently large, there are constants $c > 0$ and $C(M) > 0$ such that if $M \le t \le n^{5/6} - 2n^{1/3}$, then

$$\mathbb{P}[\chi_n > t] \le C(M)e^{-ct^{3/5}}. \tag{1.10}$$

1.1 Results

As the starting point for the present article, we observe the similarity between the upper end asymptotics of the central limit (CL) regime (1.9) and the lower end asymptotics of the upper tail large deviations (LD) regime (1.4), although these two results were proved using completely different methods. In fact a similar asymptotics holds in the upper tail moderate deviations (MD) regime, i.e. the intermediate regime between the CL and the upper tail LD regime:

Theorem 1.1 *For all* $0 < \eta < 1/3$ *and* $t > 0$,

$$\lim_{n\to\infty} \frac{\log \mathbb{P}\left[L_n > (2 + tn^{-\eta})\sqrt{n}\right]}{n^{(1-3\eta)/2} t^{3/2}} = -\frac{4}{3}. \tag{1.11}$$

This is a simplified version of the more detailed Theorem 1.2 below: On the one hand, (1.11) contains no uniformity information in η whatsoever, and it does not catch the cases $\eta \searrow 0$ or $\eta \nearrow 1/3$. On the other hand, one can improve (1.11) and also (1.3) by finding the asymptotic behaviour of $\mathbb{P}[L_n > l]$ with error terms on a non-logarithmic scale. In order to describe a refined result, we introduce a convenient parametrisation for (n, l), which is well suited for an easy description of the MD regime, the asymptotic CL regime, and the lower-end asymptotics of the LD regime: The moderate deviations regime is characterised by $1 \gg 1 - 2n^{1/2}/l \gg l^{-2/3}$; thus we set

$$\gamma_{l,n} := \frac{2\sqrt{n}}{l}, \quad M_{l,n} := \frac{l - 2\sqrt{n}}{l^{1/3}} = (1 - \gamma_{l,n}) l^{2/3}. \tag{1.12}$$

Using these new parameters $\gamma_{l,n}$ and $M_{l,n}$, the different upper tail asymptotic regimes are characterised as follows:

CL:	$\gamma_{l,n} \to 1$ with $M_{l,n}$ being fixed.
upper end asymptotics of the CL:	first $\gamma_{l,n} \nearrow 1$, second $M_{l,n} \to \infty$.
upper tail MD:	$\gamma_{l,n} \nearrow 1$ and $M_{l,n} \to \infty$ simultaneously.
lower end asymptotics of the upper tail LD:	first $M_{l,n} \to \infty$, second $\gamma_{l,n} \nearrow 1$.
upper tail LD:	$M_{l,n} \to \infty$ with $\gamma_{l,n}$ being fixed.

We set

$$w_0(\gamma) := \sqrt{1 - \gamma^2} - \operatorname{arcosh} \frac{1}{\gamma}. \tag{1.13}$$

Then the following refinement of the moderate deviations result Theorem 1.1 and the large deviations result (1.3), proved by Deuschel and Zeitouni [5], holds:

Theorem 1.2 *1.* **Moderate deviations.** *The following asymptotics hold uniformly as* $\gamma_{l,n}$ *converges to* 1 *from below and* $M_{l,n}$ *diverges to* ∞ *(independently of each other):*

$$\mathbb{P}[L_n > l] \sim \frac{e^{2l w_0(\gamma_{l,n})}}{8\pi l (1 - \gamma_{l,n}^2)^{3/2}}, \tag{1.14}$$

and (more roughly):

$$\log \mathbb{P}[L_n > l] \sim -\frac{4\sqrt{2}}{3}\frac{(l-2\sqrt{n})^{3/2}}{\sqrt{l}}. \tag{1.15}$$

2. **Large deviations.** *There are continuous functions $f_1, f_\pm :]0,1[\to]0,\infty[$, such that for all l, n with $l > 2\sqrt{n}$ and $M_{l,n} \geq f_1(\gamma_{l,n})$ we have*

$$f_-(\gamma_{l,n}) \leq \frac{\mathbb{P}[L_n > l]}{l^{-1}e^{2lw_0(\gamma_{l,n})}} \leq f_+(\gamma_{l,n}). \tag{1.16}$$

A version of (1.14) with quantitative error bounds is described in Lemma 4.2 below. We remark that (1.14) holds in the the full moderate deviations regime, in the asymptotic central limit regime (in consistency with (1.8/1.9)), and in the asymptotic large deviations regime (being consistent with (1.4), too).

For the poissonised quantity, i.e. the for the random variable $\mathcal{L}_t$ introduced above, one gets even finer asymptotics: we set

$$\phi_l(\lambda) := \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^N}{N!}\mathbb{P}[L_n \leq l] = \mathbb{P}\left[\mathcal{L}_{\sqrt{\lambda}} \leq l\right]. \tag{1.17}$$

Then one obtains:

Theorem 1.3 *For every $\lambda > 0$, $l \in \mathbb{N}$ with $2\sqrt{\lambda} < l$ we have the following asymptotics, uniformly in $\gamma_{l,\lambda}$:*

$$1 - \phi_l(\lambda) \sim \frac{\left(\gamma_{l,\lambda} + 2\sqrt{1-\gamma_{l,\lambda}^2}\right)\gamma_{l,\lambda}^2 e^{2lw_0(\gamma_{l,\lambda})}}{8\pi l(1-\gamma_{l,\lambda}^2)^{3/2}\left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)} \quad \text{as } M_{l,\lambda} \to \infty. \tag{1.18}$$

A quantitative bound for the error term in (1.18) is described in Theorem 3.1 below. Let us compare Theorem 1.2 to Theorem 1.3: On the one hand, the asymptotics of the *poissonised* probabilities $\phi_l(\lambda)$ is explicitly known both in the LD regime and in the MD regime up to error terms which converge to zero. Compared to Seppäläinen's result [15] we not only cover the moderate deviations regime, but also in the regime of large deviations we derive a finer asymptotics. Also note that the *depoissonised* quantities $\mathbb{P}[L_n > l]$ are known in the LD regime (on a non-logarithmic scale) only up to bounded factors, while in the MD regime the error terms still vanish asymptotically: During the "depoissonisation" step in the proof, which may be compared with a "deconvolution procedure", we have to take some loss of precision into account.

1.2 Review of some methods in the proof of the Baik / Deift / Johansson theorem

Our proof of the theorems starts with a representation of $\phi_l(\lambda)$ in terms of the solution of a certain noncommutative, rank 2 Riemann-Hilbert problem, which was derived by

Baik/Deift/Johansson. In order to explain this starting point, we briefly review parts of the proof of the Baik/Deift/Johansson theorem. For a detailed description of the steps, we refer the reader to Baik/Deift/Johansson's article [3] and the references therein.

The first step consists of a Poissonisation, i.e. instead of considering the quantity L_n we consider L_N, where N is a random Poisson number with parameter λ (this step is only necessary if we start with L_n, since $\mathcal{L}_{\sqrt{\lambda}}$ already carries the desired random structure). A concentration result for Poisson random variables reduces the problem of studying the asymptotics of $\mathbb{P}[L_n > l]$ to the study of the asymptotics of $\phi_l(\lambda)$ with $\lambda \sim n \to \infty$. The reason why this Poissonisation step helps at all is the following beautiful identity derived by Gessel [8] in 1990: $\phi_l(\lambda) = e^{-\lambda} D_{l-1}(\lambda)$. Here $D_{l-1}(\lambda)$ is a $l \times l$ Toeplitz determinant:

$$D_{l-1}(\lambda) = \det\left(\int_{-\pi}^{\pi} e^{-i(k-j)\theta} e^{2\sqrt{\lambda}\cos\theta} \frac{d\theta}{2\pi}\right)_{0\le k,j\le l-1}. \tag{1.19}$$

The problem is thus reduced to analysing the asymptotics of the above Toeplitz determinants when $\lambda \to \infty$ and $l \sim 2\sqrt{\lambda}$. It turns out that the above Toeplitz determinants are intrinsically related to certain orthogonal polynomials. More precisely, let

$$p_{l,\lambda}(z) = \sum_{j=0}^{l} \kappa_{l,j}(\lambda) z^j, \qquad \kappa_l(\lambda) := \kappa_{l,l}(\lambda) > 0 \tag{1.20}$$

be the l'th orthonormal polynomial with respect to the weight function $e^{2\sqrt{\lambda}\cos\theta}\frac{d\theta}{2\pi}$ on the unit circle, i.e.,

$$\int_{-\pi}^{\pi} \overline{p_{l,\lambda}(e^{i\theta})} p_{k,\lambda}(e^{i\theta}) e^{2\sqrt{\lambda}\cos\theta} \frac{d\theta}{2\pi} = \delta_{l,k}, \quad l,k \ge 0. \tag{1.21}$$

Then one can show (see (1.24) in [3]):

$$\kappa_l^2(\lambda) = \frac{D_{l-1}(\lambda)}{D_l(\lambda)}, \tag{1.22}$$

which leads to (see (1.25) in [3]):

$$\log \phi_l(\lambda) = \sum_{k=l}^{\infty} \log \kappa_k^2(\lambda). \tag{1.23}$$

At this stage Riemann-Hilbert problems (RHP's) enter the field. There are several equivalent versions to describe Riemann-Hilbert problems; here we describe them in terms of *open coverings*: The basic ingredients to a rank k RHP are an open covering $(U_i)_{i\in J}$ of the Riemann sphere $\mathbb{C} \cup \{\infty\}$ and holomorphic maps ("transition functions") $H_{i,j} : U_i \cap U_j \to \mathrm{Gl}(k,\mathbb{C})$, $i,j \in J$, which satisfy the consistency condition ("cocycle relation") $H_{i,j}H_{j,k} = H_{i,k}$ over $U_i \cap U_j \cap U_k$ for all $i,j,k \in J$. Then the RHP with data $(U_i)_i, (H_{i,j})_{i,j}$ consists of the following: Find $k \times k$-matrix valued holomorphic functions $A_i : U_i \to \mathrm{Gl}(k,\mathbb{C})$, $i \in J$, such that $A_j = A_i H_{i,j}$ over $U_i \cap U_j$, with the normalisation

condition $A_{j_0}(z_0) = I$ for a fixed $j_0 \in J$ and a fixed $z_0 \in U_{j_0}$. (I denotes the $k \times k$ identity matrix; sometimes other normalisation conditions than the identity matrix are useful, too.) We will always use $z_0 = \infty$. We remark that the solution of a Riemann Hilbert problem over the Riemann sphere (or more generally over compact Riemann surfaces) is always unique, provided a solution exists. For the particular case of interest, this is proven in [3], Lemma 4.1; but the same proof applies to the general case as well.

The following transformation procedure is frequently used in Baik/Deift/Johansson's article as well as in the present article: Given arbitrary holomorphic maps ("base changes") $B_i : U_i \to \mathrm{Gl}(k, \mathbb{C})$, one can pass to an equivalent RHP with the same open covering $(U_i)_i$ and new transition functions $\tilde{H}_{i,j} = B_i^{-1} H_{i,j} B_j$; the solutions $\tilde{A}_j$ of the transformed RHP and A_j of the original RHP are connected by $\tilde{A}_i = B_{j_0}^{-1}(z_0) A_i B_i$; here the left factor $B_{j_0}^{-1}(z_0)$ has the only purpose to guarantee the normalisation condition $\tilde{A}_{j_0}(z_0) = I$ for the transformed problem as well.

Sometimes - especially when using Cauchy's integral formula - it is technically easier to work with *closed* refinements $(\tilde{U}_j)_j$ of the open covering $(U_j)_j$, with piecewise smooth curves as intersections of different $\tilde{U}_j$. The boundary curves $\partial\tilde{U}_j \cap \partial\tilde{U}_i$ can be conveniently adapted by changing the choice of $(\tilde{U}_j)_j$ as long as they do not leave the domain $U_i \cap U_j$ of the transition functions; we will use this freedom below to choose specific curves which run through saddle points of the transition functions.

Fokas, Its, and Kitaev [7] discovered the following key fact: The orthonormal polynomials introduced in (1.20/1.21) above can be described in terms of the solution of a certain RHP. Baik/Deift/Johansson ([3], sections 4 and 5) then transformed this RHP several times according to the general transformation procedure for RHP's, which was sketched above. They end up with a version of the RHP (see [3], formulas (5.9/5.10)), which in our language reads as follows: we consider the open covering $(U_+ = \mathbb{C}, U_* = \mathbb{C}^*, U_- = \mathbb{C}^* \cup \{\infty\})$ of the Riemann sphere, and the transition functions

$$H_{-,*} = \begin{pmatrix} 1 & 0 \\ \frac{1}{f} & 1 \end{pmatrix}, \quad H_{*,+} = \begin{pmatrix} 1 & -f \\ 0 & 1 \end{pmatrix}, \quad H_{-,+} = H_{-,*} H_{*,+} \tag{1.24}$$

with

$$f(z) := (-z)^q \exp\left(\frac{q\gamma}{2}(z - z^{-1})\right), \qquad 0 < \gamma < 1, \qquad q \in \mathbb{N}. \tag{1.25}$$

Of course, U_+ and U_- alone would already suffice to cover the whole Riemann sphere, but the factorisation of $H_{-,+}$ into triangular matrices is technically very convenient.

According to [3], the RHP specified by $A_j = A_i H_{i,j}$ over $U_i \cap U_j$, with the normalisation condition $A_-(\infty) = I$, has a unique solution $A_i : U_i \to \mathrm{Gl}(k, \mathbb{C})$, $i \in \{+, -, *\}$, and the 22-entry of this solution yields the following important connection between longest increasing subsequences of random permutations and RHPs: With κ_{q-1} from (1.20,1.22,1.23):

$$\kappa_{q-1}^2(\lambda) = (A_+)_{22}(0), \tag{1.26}$$

where $\gamma = \gamma_{q,\lambda}$ is given by (1.12).

1.3 Intuitive ideas for the proof

Before starting the proofs formally, we describe roughly the intuitive ideas underlying our method, and we compare the method with Baik/Deift/Johansson's approach: The first step consists in estimating the solution of an auxiliary RHP; the solution of this auxiliary RHP serves to construct a base change of our original problem: The auxiliary problem is specified by the artificial modification of one of the transition matrices in (1.24): The auxiliary transition matrices are defined as $H'_{-,*} := I$, $H'_{*,+} = H'_{-,+} := H_{*,+}$. Here and in the following, an index $i \in \{+, *, -\}$ (and $i, j \in \{+, *, -\}$) for a matrix valued function stands for the region U_i (and $U_i \cap U_j$, respectively) where the corresponding function is defined. This auxiliary specification of a RHP consists only of upper triangular 2×2-matrices with 1's in the diagonal; its solution can be written explicitly in terms of a Cauchy integral, and the solution again consists of triangular matrix valued functions G_-, G_*, and G_+ with 1's in the diagonal: $G_+ = G_- H'_{-,+} = G_* H'_{*,+}$. Using the solution $(G_j)_{j=+,*,-}$ for a base change $\tilde{H}_{i,j} = G_i H_{i,j} G_j^{-1}$, we observe that the transformed RHP has a simpler structure: the transformed transition matrices are $\tilde{H}_{*,+} = I$, $\tilde{H}_{-,+} = \tilde{H}_{-,*} = G_- H_{-,*} G_*^{-1} = G_- H_{-,*} G_-^{-1}$. However, conjugation of the *lower* triangular matrix $H_{-,*}$ with the *upper* triangular matrix G_- destroys the triangular structure: $\tilde{H}_{-,+}$ is not a triangular matrix, and we cannot solve the transformed RHP as simply as we solved the auxiliary problem. To overcome this complication, one observes that on a certain circle C_- centered at the origin (to be described in more details below), either G_- is very close to a constant matrix G_0 (this occurs on an arc $C_{-,1} \subseteq C_-$) or $H_{-,*}$ is very close to the identity matrix (this occurs on the complementary arc $C_{-,2} \subseteq C_-$). In both cases one has $\tilde{H}_{-,+} \approx G_0 H_{-,*} G_0^{-1}$. Since $H_{-,*}$ is lower triangular with 1's in the diagonal, the second auxiliary RHP with transition matrices $F_{-,+} = F_{-,*} := H_{-,*}$, $F_{*,+} := I$ can be explicitly solved in terms of a Cauchy integral, similarly to the auxiliary RHP above: Let $(P_j)_{j=+,*,-}$ denote the solution of the second auxiliary RHP: $P_+ = P_- F_{-,+}$. (For technical reasons, we work with a small modification $\tilde{F}_{-,+}$ of $F_{-,+}$ in the construction of P_j in the formal proof below, but we ignore this technical detail in this informal explanation.) Then $G_0 P_+ G_0^{-1} = (G_0 P_- G_0^{-1})(G_0 F_{-,+} G_0^{-1})$; hence $(G_0 P_j G_0^{-1})_j$ solves *approximately* the transformed RHP with transition matrices $\tilde{H}_{i,j}$. Again, the factor G_0 on the left hand side appears because of normalisation. Taking this approximate solution to transform the transformed RHP again, we end up with a RHP very close to the trivial RHP, i.e. the RHP with the identity matrix I as transition matrix. The solution of such an approximation of the trivial RHP is close to the identity matrix; a quantitative version of this well-known statement (see e.g. [3], section 2) is given in the appendix, Lemma A.2, below.

Next we discuss how to find approximate solutions to the above auxiliary RHPs via a saddle point approximation. First we investigate the function f given by (1.25): f is wildly oscillating on circles centered at the origin, unless the circle hits a saddle point of $\log f$. $\log f$ has precisely two saddle points $z_\pm \in \mathbb{R}$ with $0 > z_+ > -1 > z_-$. We solve our first (and second) auxiliary RHP using a Cauchy integral over a circle C_+ (and C_-)

through z_+ (and z_-), respectively: We define

$$g_\pm := \frac{1}{2\pi i} \oint_C \frac{f(s)}{s-z}\, ds \tag{1.27}$$

where C is a closed curve in $\mathbb{C}^* \setminus \{z\}$ with winding number 1 around the origin and winding number 1 around z (for g_+) and winding number 0 around z (for g_-), respectively. Then $g_+ - g_- = f$, and the solution of our first auxiliary RHP is indeed given by

$$G_\pm = \begin{pmatrix} 1 & -g_\pm \\ 0 & 1 \end{pmatrix}. \tag{1.28}$$

One sees that f can be well approximated on $C = C_+$ by a Gaussian centered at the saddle point z_+, at least for large q: this can be derived by a second order Taylor expansion of $\log f$ at z_+ in a neighbourhood $C_{+,1}$ of z_+ in C_+; in the complement $C_{+,2}$ of the arc $C_{+,1}$ the function f is negligible. The **key fact** is that in the LD and MD regime the length scale r of $C_{+,1}$ is asymptotically much smaller than the distance $|z_+ - z_-|$. This is to be contrasted to the CL regime, where these two length scales are of the same order: This is important, because the ratio of the two length scales determines the error term of the Gaussian approximation. This is why a *second order* Taylor expansion at the saddle points is insufficient to catch the CL behaviour. Indeed, Baik, Deift and Johansson [3] use a *third order* Taylor approximation near $z = -1$: this point is (in appropriate coordinates) an inflection point of $\log f$. On the other hand, this third order approximation is not well suited to describe the correct MD and LD behaviour.

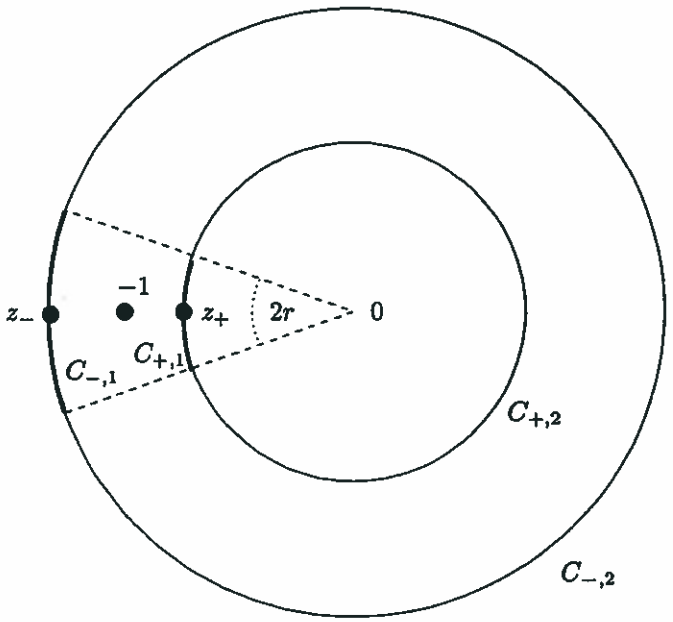

Figure 1: The saddle points $z_\pm$ and the arcs $C_{\pm,1}$, $C_{\pm,2}$ in a complex plane.

The ideas described above are carried out in Section 2 below. In Section 3, we estimate the sum (1.23). Roughly speaking, the logarithm of the summands is linearly

approximated and the sum is compared with a geometric series. Section 4 contains the depoissonisation estimates, which use concentration bounds for Poissonian random variables. The spirit of these depoissonisation results is roughly similar to the depoissonisation lemmas in Section 8 of [3]. However, those depoissonisation lemmas in the reference do not yield small enough error terms in the MD regime for our purposes.

2 Saddle point approximation

In this section, we formally estimate the solution $(A_j)_j$ of the RHP specified by (1.24). Let $0 < \gamma < 1$ and $q \in \mathbb{N}$. As in (1.12), we use the abbreviation $M := (1-\gamma)q^{2/3}$; M and $1-\gamma$ serve as reference parameters. Positive constants are denoted by c_n, where n is a counting index. If c_n depends on additional "fixed" parameters, then this is denoted explicitly.

We prove:

Theorem 2.1 *Given a positive number $\alpha < 1/4$, there are positive numbers $c_1(\alpha)$ and $b_1(\alpha) \geq 1$, such that for all $\gamma \in]0,1[$, $q \in \mathbb{N}$ with $M = (1-\gamma)q^{2/3} \geq b_1(\alpha)$ the following holds:*

$$1 - (A_+)_{22}(0) = \frac{\gamma + 2\sqrt{1-\gamma^2}}{4\pi q(1-\gamma^2)} e^{2qw_0(\gamma)}(1 + R(M,\gamma)) \tag{2.1}$$

with an error term $R(M,\gamma)$ that fulfills the bound

$$|R(M,\gamma)| \leq c_1(\alpha) M^{-3/4+3\alpha}. \tag{2.2}$$

Remark: The rate function $2w_0$ in the exponential in (2.1) (with q removed) has the asymptotics

$$2w_0(\gamma) = -\frac{4\sqrt{2}}{3}(1-\gamma)^{3/2} + O((1-\gamma)^{5/2}) \qquad \text{as } 1-\gamma \searrow 0. \tag{2.3}$$

One may compare this with the exponent in the bound in part 2 of Lemma 5.1 in [3]: this reference tells us for large M, some constant c_2 and $\frac{1}{2} \leq \gamma \leq 1 - Mq^{-2/3}$:

$$|1 - (A_+)_{22}(0)| \leq c_2 q^{-1/3} \exp\left\{-q\frac{2\sqrt{2}}{3}(1-\gamma)^{3/2}\right\}, \tag{2.4}$$

which is in the MD regime roughly on the same scale as the square root of the estimate (2.1).

Proof of Theorem 2.1: The equation $f'(z_\pm) = 0$ yields the saddle points

$$z_\pm = -\gamma^{-1} \pm \sqrt{\gamma^{-2} - 1} = -e^{\pm u_0} \quad \text{with} \quad u_0 = -\operatorname{arcosh}\frac{1}{\gamma}, \tag{2.5}$$

and we get

$$f(z_+e^{i\theta}) = f(z_-e^{-i\theta})^{-1} = \exp\left\{q\left(-\operatorname{arcosh}\frac{1}{\gamma} + \sqrt{1-\gamma^2}\cos\theta + i(\theta - \sin\theta)\right)\right\}. \quad (2.6)$$

Expanding around $\theta = 0$ gives

$$f(z_+e^{i\theta}) = \exp\left\{q\left(w_0 - w_2\theta^2 + \epsilon_\gamma(\theta)\right)\right\} \quad (2.7)$$

with $w_0 = w_0(\gamma)$ given by (1.13), $w_2 := \sqrt{1-\gamma^2}/2$, and an error term ϵ_γ, which is bounded for real θ and some real x with $|x| \le |\theta|$ by

$$\begin{aligned} |\epsilon_\gamma(\theta)| &\le \frac{|\theta|^3}{6}\left|\frac{d^3}{dx^3}\left(\sqrt{1-\gamma^2}\cos x + i(x - \sin x)\right)\right| \quad (2.8)\\ &= \frac{|\theta|^3}{6}\left|\sqrt{1-\gamma^2}\sin x + i\cos x\right| \le \frac{|\theta|^3}{6}|ie^{-ix}| = \frac{|\theta|^3}{6}. \end{aligned}$$

(Having our goal (2.1) and Deuschel/Zeitouni's (resp. Seppäläinen's) result (1.3) in mind, we observe that the exponential rate function e^{2qw_0} is determined - up to a square - by the value of the transition matrix entry f at the saddle points.) For $-\pi < \theta < \pi$, the simple estimate $\cos\theta \le 1 - 2\pi^{-2}\theta^2$ implies the the bound

$$|f(z_+e^{i\theta})| \le \exp\left\{q\left(w_0 - c_3w_2\theta^2\right)\right\} \quad (2.9)$$

with $c_3 := 4\pi^{-2}$. We define the length scale

$$r := q^{-1/2+2\alpha/3}(1-\gamma)^{\alpha-1/4} = M^{-3/4+\alpha}\sqrt{1-\gamma}. \quad (2.10)$$

(Intuitively, the choice of r arises as a compromise: on a disk of radius r around the saddle points, f should be approximated by a Gaussian function well enough, and outside this disk but on a circle through the saddle point z_+, f should be small enough.) It is instructive to compare r with the distance between the saddle points:

$$z_+ - z_- = 2\sqrt{\gamma^{-2}-1} = 2\sqrt{2}\sqrt{1-\gamma}(1+o(1)) \quad \text{as } \gamma \nearrow 1. \quad (2.11)$$

We estimate for $M \ge 1$: For some positive constant c_4,

$$\sup_{-r<\theta<r}\left|\frac{f(z_+e^{i\theta})}{e^{q(w_0-w_2\theta^2)}} - 1\right| = \sup_{-r<\theta<r}|e^{q\epsilon_\gamma(\theta)} - 1| \overset{(2.8)}{\le} c_4qr^3 = c_4M^{-3/4+3\alpha}; \quad (2.12)$$

fo the second step we used that $|q\epsilon_\gamma(\theta)| \le qr^3/6 \le M^{-3/4+3\alpha}/6$ is bounded and that exp is uniformly Lipschitz continuous on bounded domains. Let $C_\pm$ denote the circle through $z_\pm$, centered at the origin. We parametrise $C_\pm$ by $z = z_\pm e^{i\theta}$, $-\pi < \theta < \pi$. We split $C_\pm$ further into the two arcs $C_{\pm,1}$ and $C_{\pm,2}$, parametrised by $|\theta| \le r$ and $r < |\theta| < \pi$ respectively. (One observes $r < \pi$ for $M \ge 1$.)

Next we examine the following auxiliary RHP; recall the definition (1.24) of $H_{*,+} = H'_{-,+}$ from Section 1.3:
Find holomorphic $G_\pm : U_\pm \to \mathrm{Gl}_n(\mathbb{C})$ *with*

$$G_+ = G_- H'_{-,+}, \qquad G_-(\infty) = I. \tag{2.13}$$

We introduce the abbreviation

$$g_0 := \frac{|z_+|e^{qw_0}}{2|z_+ - z_-|\sqrt{\pi q w_2}} = \frac{|z_+|}{2\pi|z_+ - z_-|}\int_{\mathbb{R}} e^{q(w_0 - w_2\theta^2)}\, d\theta > 0. \tag{2.14}$$

Here are the estimates that we need for our auxiliary RHP:

Lemma 2.2 *The solution* $G_\pm$ *of the auxiliary RHP (2.13) is given by formula (1.28). Given* $0 < \alpha < 1/4$, *there are positive constants* $c_5(\alpha)$, c_6, $c_7(\alpha)$ *and* $b_2(\alpha) \geq 1$ *such that if* $M \geq b_2(\alpha)$ *and* $0 < \gamma < 1$, *then the following three bounds hold:*

$$\sup_{z \in C_{-,1}} |g_-(z) + g_0| \leq c_5(\alpha) M^{-3/4+3\alpha} g_0, \tag{2.15}$$

$$\sup_{z \in C_-} |g_-(z)| \leq c_6 g_0, \tag{2.16}$$

$$\left| g_+(0) - \frac{|z_+|e^{qw_0}}{2\sqrt{\pi q w_2}} \right| \leq c_7(\alpha) M^{-3/4+3\alpha} \frac{|z_+|e^{qw_0}}{2\sqrt{\pi q w_2}}. \tag{2.17}$$

Proof of Lemma 2.2. We define $g_\pm$ as in (1.27) and obtain the solution $G_\pm$ as in (1.28) of the RHP (2.13) as a consequence of the residue theorem. Using $C = C_+$, we estimate g_- over C_-: We estimate the Cauchy integral (1.27) at first for $z \in C_{-,1}$: By splitting the integration path, we get $g_-(z) = I_1 + I_2$, where

$$I_j := \frac{1}{2\pi i}\int_{C_{+,j}} \frac{f(s)}{s - z}\, ds, \qquad j = 1, 2. \tag{2.18}$$

We compare I_1 with

$$\tilde{I}_1 := \frac{1}{2\pi i}\int_{C_{+,1}} \frac{e^{q(w_0 + w_2 \log^2(s/z_+))}}{z_+ - z_-}\frac{z_+\, ds}{s} = -\frac{e^{qw_0}|z_+|}{2\pi|z_+ - z_-|}\int_{-r}^{r} e^{-qw_2\theta^2}\, d\theta < 0 \tag{2.19}$$

for z close to z_-; here log denotes the principal branch of the logarithm, and we used the substitution $s = z_+ e^{i\theta}$. For $s \in C_{+,1}$, $z \in C_{-,1}$ we have $|s - z_+| \leq |z_+|r$, $|z - z_-| \leq |z_-|r$, hence

$$\left|\frac{z_+}{s} - 1\right| \leq r \leq M^{-3/4+\alpha}, \tag{2.20}$$

$$\left|\frac{s - z}{z_+ - z_-} - 1\right| \leq \frac{(|z_+| + |z_-|)r}{|z_+ - z_-|} \leq M^{-3/4+\alpha}; \tag{2.21}$$

we used (2.10), $|z_+| + |z_-| = 2\gamma^{-1}$, and $|z_+ - z_-|^{-1} = (2\sqrt{\gamma^{-2}-1})^{-1} \le \frac{\gamma}{2}(1-\gamma)^{-1/2}$. In the estimate (2.22) below, we make use of Lipschitz continuity of multiplication and division at 1, more precisely of the following fact: There are positive constants c_8, c_9 such that for every ϵ with $0 < \epsilon \le c_8$ and every $x, y \in \mathbb{C}$ with $|x-1| \le \epsilon$ and $|y-1| \le \epsilon$ we have the bounds $|xy-1| \le c_9\epsilon$ and $|x/y-1| \le c_9\epsilon$. This fact and the estimates (2.12), (2.20), and (2.21) together imply

$$\left| \frac{f(s)}{s-z} - \frac{e^{q(w_0 - w_2 \log^2(s/z_+))}}{z_+ - z_-} \frac{z_+}{s} \right| \le c_{10} M^{-3/4+3\alpha} \left| \frac{e^{q(w_0 - w_2 \log^2(s/z_+))}}{z_+ - z_-} \frac{z_+}{s} \right| \tag{2.22}$$

for $s \in C_{+,1}$, $z \in C_{-,1}$, some constant $c_{10} > 0$, $0 < \gamma < 1$, and M being large enough (say $M \ge b_2(\alpha) \ge 1$). Hence (using that the second integrand in the definition of $\tilde{I}_1$ is positive):

$$|I_1 - \tilde{I}_1| \le c_{10} M^{-3/4+3\epsilon_3} |\tilde{I}_1|. \tag{2.23}$$

We compare $-\tilde{I}_1$ with a Gaussian integral over $\mathbb{R}$: The bounds

$$\int\limits_{|x|>r} e^{-ax^2}\,dx \;\le\; \int\limits_{|x|>r} \frac{|x|}{r} e^{-ax^2}\,dx = (ar)^{-1} e^{-ar^2} \qquad (a, r > 0), \tag{2.24}$$

$$q w_2 r^2 \;=\; \frac{\sqrt{1+\gamma}}{2} M^{2\alpha} \ge \frac{1}{2} M^{2\alpha} \tag{2.25}$$

imply

$$\begin{aligned} \left| \tilde{I}_1 + g_0 \right| \;&=\; \frac{|z_+|}{2\pi |z_+ - z_-|} \int_{|\theta|>r} e^{q(w_0 - w_2\theta^2)}\,d\theta \\ &\overset{(2.24)}{\le} \frac{|z_+|}{2\pi |z_+ - z_-|} \frac{e^{q(w_0 - w_2 r^2)}}{q w_2 r} \overset{(2.14)}{=} \frac{e^{-q w_2 r^2} g_0}{\sqrt{\pi q w_2 r^2}} \overset{(2.25)}{\le} \sqrt{\frac{2}{\pi}} M^{-\alpha} e^{-\frac{1}{2}M^{2\alpha}} g_0. \end{aligned} \tag{2.26}$$

Next we estimate I_2 for $z \in C_-$: Using (2.9), $\operatorname{dist}(C_+, C_-) = |z_+ - z_-|$, and $|z-s| \ge |z_+ - z_-|$ for $z \in C_-$, $s \in C_+$, we obtain similarly to (2.26):

$$\begin{aligned} |I_2| = \left| \frac{1}{2\pi} \int_{r<|\theta|<\pi} \frac{f(z_+ e^{i\theta})}{z_+ e^{i\theta} - z} z_+ e^{i\theta}\,d\theta \right| &\overset{(2.9)}{\le} \frac{|z_+|}{2\pi |z_+ - z_-|} \int_{|\theta|>r} e^{q(w_0 - c_3 w_2 \theta^2)}\,d\theta \\ \overset{(2.24)}{\le} \frac{|z_+|}{2\pi |z_+ - z_-|} \frac{e^{q(w_0 - c_3 w_2 r^2)}}{q c_3 w_2 r} &\overset{(2.14)}{=} \frac{e^{-q c_3 w_2 r^2} g_0}{c_3 \sqrt{\pi q w_2 r^2}} \overset{(2.25)}{\le} c_{11} M^{-\alpha} e^{-\frac{c_3}{2} M^{2\alpha}} g_0 \end{aligned} \tag{2.27}$$

with some constant $c_{11} > 0$. In the next step we substitute $g_-(z) = I_1 + I_2$ for $z \in C_{-,1}$ and combine (2.23), (2.26), and (2.27):

$$\begin{aligned} |g_-(z) + g_0| = |I_1 + g_0 + I_2| &\le |I_1 - \tilde{I}_1| + |\tilde{I}_1 + g_0| + |I_2| \\ &\le c_{10} M^{-3/4+3\alpha} (g_0 + |\tilde{I}_1 + g_0|) + |\tilde{I}_1 + g_0| + |I_2| \le c_5(\alpha) M^{-3/4+3\alpha} g_0 \end{aligned} \tag{2.28}$$

for $M \geq b_2(\alpha)$ and some sufficiently large constant $c_5(\alpha)$. This proves (2.15).

We turn to the proof of (2.16): Substituting the bound (2.9) into the Cauchy integral (1.27) yields

$$\sup_{z \in C_-} |g_-(z)| \leq \frac{|z_+|}{2\pi |z_+ - z_-|} \int_{-\pi}^{\pi} e^{q(w_0 - c_3 w_2 \theta^2)} \, d\theta \leq c_6 g_0 \tag{2.29}$$

for the constant $c_6 := c_3^{-1/2}$; this proves (2.16).

Next we estimate $g_+(0)$: We split the Cauchy integral:

$$g_+(0) = \frac{1}{2\pi i} \oint_{C_+} f(s) \frac{ds}{s} = \frac{1}{2\pi i} \int_{C_{+,1}} f(s) \frac{ds}{s} + \frac{1}{2\pi i} \int_{C_{+,2}} f(s) \frac{ds}{s}. \tag{2.30}$$

Using (2.12) and (2.9) again, we obtain for $M \geq 1$:

$$\left| g_+(0) - \frac{|z_+| e^{q w_0}}{2\sqrt{\pi q w_2}} \right| = \left| g_+(0) - \frac{|z_+|}{2\pi} \int_{-\infty}^{\infty} e^{q(w_0 - w_2 \theta^2)} \, d\theta \right| \tag{2.31}$$

$$\begin{aligned}
&\leq \frac{c_4}{2\pi} |z_+| M^{-3/4+3\alpha} \int_{-r}^{r} e^{q(w_0 - w_2\theta^2)} \, d\theta + \frac{|z_+|}{2\pi} \int_{|\theta|>r} e^{q(w_0 - c_3 w_2 \theta^2)} \, d\theta + \frac{|z_+|}{2\pi} \int_{|\theta|>r} e^{q(w_0 - w_2 \theta^2)} \, d\theta \\
&\leq |z_+| e^{q w_0} \left(\frac{c_4 M^{-3/4+3\alpha}}{2\sqrt{\pi q w_2}} + \frac{e^{-c_3 q w_2 r^2}}{2\pi c_3 q w_2 r} + \frac{e^{-q w_2 r^2}}{2\pi q w_2 r} \right) \\
&\leq \frac{|z_+| e^{q w_0}}{2\sqrt{\pi q w_2}} \left(c_4 M^{-3/4+3\alpha} + \frac{e^{-\frac{c_3}{2} M^{2\alpha}}}{\sqrt{\frac{\pi}{2}} c_3 M^{\alpha}} + \frac{e^{-\frac{1}{2} M^{2\alpha}}}{\sqrt{\frac{\pi}{2}} M^{\alpha}} \right) \leq c_7(\alpha) M^{-3/4+3\alpha} \frac{|z_+| e^{q w_0}}{2\sqrt{\pi q w_2}}
\end{aligned}$$

for some positive constant $c_7(\alpha)$; we several times used the bounds (2.24,2.25). This finishes the proof of (2.17) and also of Lemma 2.2.

□

We continue the proof of Theorem 2.1: We introduce the following approximations f_0 to $1/f$ on C_-, $\tilde{F}_{-,+}$ to $H_{-,*}$ (recall definition (1.24)), and G_0 to G_-:

$$f_0(z_- e^{i\theta}) := e^{q(w_0 - 2w_2(\cos\theta - 1))}, \qquad \tilde{F}_{-,+} := \begin{pmatrix} 1 & 0 \\ f_0 & 1 \end{pmatrix}, \tag{2.32}$$

$$G_0 := \begin{pmatrix} 1 & g_0 \\ 0 & 1 \end{pmatrix}. \tag{2.33}$$

The function f_0 fulfills the following bounds, which are analogous to (2.9) and (2.12): For $-\pi < \theta < \pi$,

$$|f_0(z_- e^{i\theta})| \leq e^{q(w_0 - c_3 w_2 \theta^2)}, \tag{2.34}$$

$$\sup_{-r<\theta<r} \left| \frac{f_0(z_- e^{i\theta})}{e^{q(w_0 - w_2\theta^2)}} - 1 \right| \leq c_4 q r^3 = c_4 M^{-3/4+3\alpha}; \tag{2.35}$$

the last estimate holds for $M \geq 1$. Analogously to the definition of $g_\pm$ and $G_\pm$ we define

$$p_\pm(z) := \frac{1}{2\pi i} \oint_{C_-} \frac{f_0(s)}{s-z}\, ds, \qquad P_\pm := \begin{pmatrix} 1 & 0 \\ p_\pm & 1 \end{pmatrix}. \tag{2.36}$$

Here p_+ (p_-) is defined inside (outside) the disk with boundary C_-; it is continuously extended to the curve C_-. By Cauchy's integral formula, we have again on C_-:

$$f_0 = p_+ - p_-, \qquad \tilde{F}_{-,+} = P_-^{-1} P_+. \tag{2.37}$$

We apply Lemma A.1 in the appendix with $k = |z_-|(qw_2)^{-1/2}$ to f_0, assuming $M \geq 1$ and using the bounds

$$\|f_0\|_{L^1(C_-)} \leq c_{12} k e^{qw_0}, \quad \|f_0'\|_{L^\infty(C_-)} \leq c_{13} k^{-1} e^{qw_0}, \quad \|f_0\|_{L^\infty(C_-)} \leq e^{qw_0}, \tag{2.38}$$

to see:

$$\|p_\pm\|_{L^\infty(C_-)} \leq c_{14} e^{qw_0}. \tag{2.39}$$

Just as in (2.31) one obtains

$$\left| p_+(0) - \frac{|z_-| e^{qw_0}}{2\sqrt{\pi q w_2}} \right| \leq c_7(\alpha) M^{-3/4+3\alpha} \frac{|z_-| e^{qw_0}}{2\sqrt{\pi q w_2}}. \tag{2.40}$$

We introduce the scaling matrix

$$S := \begin{pmatrix} g_0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.41}$$

Conjugation of a matrix X with S is abbreviated by $X^S := S^{-1} X S$, or explicitly:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}^S = \begin{pmatrix} x_{11} & x_{12}/g_0 \\ x_{21} g_0 & x_{22} \end{pmatrix}. \tag{2.42}$$

We employ this operation below to "precondition" the RHP: nondiagonal terms which have very different orders of magnitude will be transformed to terms of comparable order.

To deal with small perturbations of the identity matrix as transition functions, we work with estimates in $L^p(C)$, $1 \leq p \leq \infty$: if X is a matrix-valued function on a curve C and $p < \infty$, let $\|X\|_{L^p(C)} := (\int_C |X(z)|^p\, |dz|)^{1/p}$; here $|\cdot|$ denotes any fixed submultiplicative matrix norm. Similarly $\|X\|_{L^\infty(C)}$ is defined using the same submultiplicative matrix norm $|\cdot|$. In the case $p = \infty$ the estimates below should be interpreted as the limits as $p \to \infty$; especially factors $(qw_2)^{-1/(2p)}$ simply drop out in this limit. We estimate for $M \geq b_2(\alpha)$, using $p^{-1/(2p)} \leq 1$ several times:

$$\left\| (H_{-,*} - \tilde{F}_{-,+})^S \right\|_{L^p(C_{-,1})} = \left\| \begin{pmatrix} 0 & 0 \\ (\frac{1}{f} - f_0) g_0 & 0 \end{pmatrix} \right\|_{L^p(C_{-,1})} \tag{2.43}$$

$$\overset{\substack{(2.6,\\ 2.12,2.35)}}{\leq} c_{15} |z_-|^{1/p} M^{-3/4+3\alpha} g_0 \left(\int_{\mathbb{R}} e^{pq(w_0 - w_2\theta^2)}\, d\theta \right)^{1/p} \tag{2.44}$$

$$\leq c_{16} |z_-|^{1/p} M^{-3/4+3\alpha} (qw_2)^{-1/(2p)} e^{qw_0} g_0,$$

$$\left\|(H_{-,*}-I)^S\right\|_{L^p(C_{-,1})} \overset{(2.9)}{\leq} c_{17}|z_-|^{1/p}(qw_2)^{-1/(2p)}e^{qw_0}g_0, \tag{2.45}$$

$$\left\|(\tilde{F}_{-,+}-I)^S\right\|_{L^p(C_{-,1})} \overset{(2.34)}{\leq} c_{17}|z_-|^{1/p}(qw_2)^{-1/(2p)}e^{qw_0}g_0 \tag{2.46}$$

$$\left\|((G_-^{-1}G_0)^{\pm 1}-I)^S\right\|_{L^\infty(C_{-,1})} = \left\|\begin{pmatrix} 0 & \frac{g_-+g_0}{g_0} \\ 0 & 0 \end{pmatrix}\right\|_{L^\infty(C_{-,1})} \overset{(2.15)}{\leq} c_{18}(\alpha)M^{-3/4+3\alpha}, \tag{2.47}$$

$$\left\|(H_{-,*}-I)^S\right\|_{L^p(C_{-,2})} \overset{(2.9)}{\leq} c_{19}|z_-|^{1/p}g_0\left(\int_{|\theta|>r} e^{pq(w_0-c_3w_2\theta^2)}\,d\theta\right)^{1/p} \tag{2.48}$$

$$\begin{aligned} &\overset{(2.24)}{\leq} c_{20}|z_-|^{1/p}(qw_2r)^{-1/p}e^{-qc_3w_2r^2}e^{qw_0}g_0 \overset{(2.25)}{\leq} c_{21}|z_-|^{1/p}M^{-\alpha/p}e^{-\frac{c_3}{2}M^{2\alpha}}(qw_2)^{-1/(2p)}e^{qw_0}g_0 \\ &\leq c_{22}(\alpha)|z_-|^{1/p}M^{-3/4+3\alpha}(qw_2)^{-1/(2p)}e^{qw_0}g_0, \end{aligned}$$

$$\left\|(\tilde{F}_{-,+}-I)^S\right\|_{L^p(C_{-,2})} \overset{(2.34,2.25)}{\leq} c_{22}(\alpha)|z_-|^{1/p}M^{-3/4+3\alpha}(qw_2)^{-1/(2p)}e^{qw_0}g_0, \tag{2.49}$$

$$\left\|((G_-^{-1}G_0)^{\pm 1})^S\right\|_{L^\infty(C_-)} \overset{(2.16)}{\leq} c_{23}, \tag{2.50}$$

hence

$$\begin{aligned} &\left\|(G_0^{-1}G_-H_{-,*}G_-^{-1}G_0-\tilde{F}_{-,+})^S\right\|_{L^p(C_-)} \qquad (2.51)\\ &= \left\|(G_0^{-1}G_-(H_{-,*}-I)G_-^{-1}G_0-(\tilde{F}_{-,+}-I))^S\right\|_{L^p(C_-)} \\ &\leq \left\|(G_0^{-1}G_--I)^S\right\|_{L^\infty(C_{-,1})}\left\|(H_{-,*}-I)^S\right\|_{L^p(C_{-,1})}\left\|(G_-^{-1}G_0)^S\right\|_{L^\infty(C_{-,1})} \\ &\quad + \left\|(H_{-,*}-I)^S\right\|_{L^p(C_{-,1})}\left\|(G_-^{-1}G_0-I)^S\right\|_{L^\infty(C_{-,1})} + \left\|(H_{-,*}-\tilde{F}_{-,+})^S\right\|_{L^p(C_{-,1})} \\ &\quad + \left\|(G_0^{-1}G_-)^S\right\|_{L^\infty(C_{-,2})}\left\|(H_{-,*}-I)^S\right\|_{L^p(C_{-,2})}\left\|(G_-^{-1}G_0)^S\right\|_{L^\infty(C_{-,2})} \\ &\quad + \left\|(\tilde{F}_{-,+}-I)^S\right\|_{L^p(C_{-,2})} \\ &\leq c_{24}(\alpha)|z_-|^{1/p}M^{-3/4+3\alpha}(qw_2)^{-1/(2p)}e^{qw_0}g_0. \end{aligned}$$

Furthermore for $M \geq 1$:

$$\left\|(P_\pm^{\mp 1})^S\right\|_{L^\infty(C_-)} \leq 1 + \left\|(P_\pm^{\mp 1}-I)^S\right\|_{L^\infty(C_-)} \overset{(2.39)}{\leq} 1 + c_{25}e^{qw_0}g_0 \leq c_{26}; \tag{2.52}$$

the last estimate follows from

$$e^{qw_0} \leq 1 \quad \text{and} \quad g_0 \leq c_{27}M^{-3/4}e^{qw_0} \leq c_{27}; \tag{2.53}$$

where we used that qw_0 is negative:

$$w_0 = -\int_\gamma^1 \frac{\sqrt{1-x^2}}{x}\,dx < 0. \tag{2.54}$$

Using (2.37), (2.51), and (2.52) we get

$$\begin{aligned}
&\left\|(P_-G_0^{-1}G_-H_{-,*}G_-^{-1}G_0P_+^{-1} - I)^S\right\|_{L^p(C_-)} \\
&\quad\le \left\|P_-^S\right\|_{L^\infty(C_-)} \left\|(G_0^{-1}G_-H_{-,*}G_-^{-1}G_0 - \tilde F_{-,+})^S\right\|_{L^p(C_-)} \left\|(P_+^{-1})^S\right\|_{L^\infty(C_-)} \\
&\quad\le c_{28}(\alpha)|z_-|^{1/p}M^{-3/4+3\alpha}(qw_2)^{-1/(2p)}e^{qw_0}g_0.
\end{aligned} \tag{2.55}$$

Note that also for $p=\infty$

$$\left\|(P_-G_0^{-1}G_-H_{-,*}G_-^{-1}G_0P_+^{-1} - I)^S\right\|_{L^\infty(C_-)} \le c_{28}(\alpha)M^{-3/4+3\alpha}e^{qw_0}g_0. \tag{2.56}$$

We consider the following "error-term" Riemann-Hilbert problem: *Find holomorphic matrix valued functions $L_\pm$, with $L_+(z)$ being defined inside the disk $|z|\le|z_-|$, and $L_-(z)$ being defined outside this disk (its boundary and ∞ included), such that $L_-(\infty)=I$ and*

$$L_+ = L_-P_-G_0^{-1}G_-H_{-,*}G_-^{-1}G_0P_+^{-1} \quad \text{on } C_-. \tag{2.57}$$

Let $b_1(\alpha)\ge b_2(\alpha)$ be so large that $c_{28}(\alpha)c_{27}b_1(\alpha)^{-3/2+3\alpha}\le 1/2$; then the condition $M\ge b_1(\alpha)$ implies

$$\left\|(P_-G_0^{-1}G_-H_{-,*}G_-^{-1}G_0P_+^{-1} - I)^S\right\|_{L^\infty(C_-)} \le \frac{1}{2}; \tag{2.58}$$

to see this, one uses (2.56) and (2.53). (1/2 was just chosen to have some definite number between 0 and 1.) Using Lemma A.2 in the appendix and (2.55), we see that the "error term" RHP (2.57) has a solution for $M\ge b_1(\alpha)$ with

$$\begin{aligned}
&|(L_+(0)-I)^S| \\
&\quad\le c_{28}(\alpha)|z_-|M^{-3/4+3\alpha}(qw_2)^{-1/2}e^{qw_0}g_0 + 2\left[c_{28}(\alpha)\sqrt{|z_-|}M^{-3/4+3\alpha}(qw_2)^{-1/4}e^{qw_0}g_0\right]^2 \\
&\quad\overset{(2.53)}{\le} c_{29}(\alpha)|z_-|M^{-3/4+3\alpha}(qw_2)^{-1/2}e^{qw_0}g_0.
\end{aligned} \tag{2.59}$$

The solution of the original RHP with transition functions specified by (1.24) can be written in terms of the solution (2.57):

$$A_+ = G_0L_+P_+G_0^{-1}G_+, \qquad A_- = G_0L_-P_-G_0^{-1}G_-, \tag{2.60}$$

since $A_-(\infty)=I$ and

$$\begin{aligned}
A_+ &\overset{(2.60,2.57)}{=} G_0L_-P_-G_0^{-1}G_-H_{-,*}G_-^{-1}G_+ \\
&\overset{(2.60,2.13)}{=} A_-H_{-,*}H_{*,+} = A_-H_{-,+} \quad \text{on } C_-.
\end{aligned} \tag{2.61}$$

We prepare some estimates, which are applied below: First we observe

$$|z_-| \leq |z_+| + |z_+ - z_-| \leq 1 + |z_+ - z_-| \leq 2|z_-|. \tag{2.62}$$

Second, we use the bound (2.52) together with the maximum principle for holomorphic functions to see

$$|P_+^S(0)| \leq c_{26}. \tag{2.63}$$

We also need the following bounds below: as a consequence of

$$|g_+(0)| \overset{(2.17)}{\leq} \frac{c_{30}|z_+|e^{qw_0}}{\sqrt{qw_2}}, \tag{2.64}$$

$$\frac{|g_+(0)|}{g_0} \overset{(2.64,2.14)}{\leq} c_{31}|z_+ - z_-| \tag{2.65}$$

we get

$$|G_+^S(0)| = \left|\begin{pmatrix} 1 & -g_+(0)/g_0 \\ 0 & 1 \end{pmatrix}\right| \overset{(2.65)}{\leq} \frac{c_{32}(\alpha)}{2}(1 + |z_+ - z_-|) \overset{(2.62)}{\leq} c_{32}(\alpha)|z_-|, \tag{2.66}$$

$$|(G_0^{\pm 1})^S| = \left|\begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}\right| \leq c_{33}. \tag{2.67}$$

We also observe the simple fact $z_+z_- \overset{(2.5)}{=} 1$. We estimate $(A_+)_{22}(0)$:

$$\begin{aligned} |(A_+ - G_0P_+G_0^{-1}G_+)_{22}(0)| &= |((A_+ - G_0P_+G_0^{-1}G_+)^S)_{22}(0)| \qquad (2.68)\\ &\leq |(A_+ - G_0P_+G_0^{-1}G_+)^S(0)| \leq |G_0^S|\,|(L_+(0) - I)^S|\,|P_+^S(0)|\,|(G_0^{-1})^S|\,|G_+^S(0)| \\ &\overset{(2.59,2.63,\ 2.66,2.67)}{\leq} c_{34}(\alpha)|z_-|^2M^{-3/4+3\alpha}(qw_2)^{-1/2}e^{qw_0}g_0 \overset{(2.14)}{=} c_{35}(\alpha)\frac{|z_-|^2|z_+|}{|z_+ - z_-|}M^{-3/4+3\alpha}\frac{e^{2qw_0}}{4\pi qw_2} \\ &\overset{(2.5)}{=} c_{35}(\alpha)\frac{|z_-|}{|z_+ - z_-|}M^{-3/4+3\alpha}\frac{e^{2qw_0}}{4\pi qw_2} \overset{(2.62)}{\leq} c_{35}(\alpha)(1 + |z_+ - z_-|^{-1})M^{-3/4+3\alpha}\frac{e^{2qw_0}}{4\pi qw_2}. \end{aligned}$$

We explicitly determine the 22-entry of the matrix product in (2.68): Using the bounds (2.14), (2.17), (2.40), and the fact $z_+z_- = 1$, there exists a constant $c_{36}(\alpha) > 0$ and an error term $|\delta_1(M,\gamma)| \leq 1$ such that the following holds:

$$\begin{aligned} (G_0P_+G_0^{-1}G_+)_{22} &= 1 - (g_0 + g_+(0))p_+(0) \qquad (2.69)\\ &= 1 - \frac{1 + |z_+ - z_-|^{-1}}{4\pi qw_2}e^{2qw_0}(1 + c_{36}(\alpha)M^{-3/4+3\alpha}\delta_1(M,\gamma)). \end{aligned}$$

Combining (2.69), (2.68), and using (2.5) respectively (2.11), we obtain for some positive constant $c_1(\alpha)$ and some $|\delta_2(M,\gamma)| \leq 1$:

$$\begin{aligned} (A_+)_{22}(0) &= 1 - \frac{1 + |z_+ - z_-|^{-1}}{4\pi qw_2}e^{2qw_0}(1 + c_1(\alpha)M^{-3/4+3\alpha}\delta_2(M,\gamma)) \qquad (2.70)\\ &= 1 - \frac{\gamma + 2\sqrt{1-\gamma^2}}{4\pi q(1-\gamma^2)}e^{2qw_0}(1 + c_1(\alpha)M^{-3/4+3\alpha}\delta_2(M,\gamma)). \end{aligned}$$

This proves (2.1) and therefore Theorem 2.1.

□

3 Summation

With the notation introduced in Section 1, we prove in this section:

Theorem 3.1 *For every fixed α with $0 < \alpha < 1/4$ there are constants $b_3(\alpha) \geq 1$ and $c_{37}(\alpha) > 0$ such that for every $\lambda > 0$ and every $l \in \mathbb{N}$ with $M_{l,\lambda} \geq b_3(\alpha)$ we have the estimate*

$$1 - \phi_l(\lambda) = \left(1 + \delta_7(l,\lambda,\alpha) M_{l,\lambda}^{-3/4+3\alpha}\right) \frac{\left(\gamma_{l,\lambda} + 2\sqrt{1-\gamma_{l,\lambda}^2}\right) \gamma_{l,\lambda}^2 e^{2lw_0(\gamma_{l,\lambda})}}{8\pi l (1-\gamma_{l,\lambda}^2)^{3/2}\left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)} \tag{3.1}$$

with a bounded error term $|\delta_7(l,\lambda,\alpha)| \leq c_{37}(\alpha)$.

Proof of Theorem 3.1. The first step in our derivation is to expand the logarithm on the right hand side of (1.23): By (2.1/2.2), we know for $q \geq 1$:

$$|1 - \kappa_{q-1}^2(\lambda)| \leq \frac{c_{38} e^{2qw_0(\gamma_{q,\lambda})}}{q(1-\gamma_{q,\lambda}^2)} \leq c_{38} e^{2qw_0(\gamma_{q,\lambda})} M_{q,\lambda}^{-1} \tag{3.2}$$

if only $M_{q,\lambda} = (1-\gamma_{q,\lambda}) q^{2/3}$ is large enough, say $M_{q,\lambda} \geq b_4 \geq 1$. (One takes e.g. $\alpha = 1/8$ in Theorem 2.1 and observes $\gamma_{q,\lambda} + 2\sqrt{1-\gamma_{q,\lambda}^2} \leq 3$ for the numerator in (2.1).) We estimate

$$-w_0(\gamma) = \operatorname{arcosh} \frac{1}{\gamma} - \sqrt{1-\gamma^2} = \int_\gamma^1 \frac{\sqrt{1-x^2}}{x}\, dx \geq \int_\gamma^1 \sqrt{1-x}\, dx = \frac{2}{3}(1-\gamma)^{3/2}, \tag{3.3}$$

hence with $c_{39} := 4/3$ and $M_{q,\lambda}$ being large enough, say $M_{q,\lambda} \geq b_5 \geq b_4$:

$$e^{2qw_0(\gamma_{q,\lambda})} \leq e^{-c_{39} M_{q,\lambda}^{3/2}}, \tag{3.4}$$

$$|1 - \kappa_{q-1}^2(\lambda)| \overset{(3.2)}{\leq} c_{38} e^{-c_{39} M_{q,\lambda}^{3/2}} M_{q,\lambda}^{-1} \leq c_{38} e^{-c_{39} M_{q,\lambda}^{3/2}}. \tag{3.5}$$

Consequently we have for fixed positive $\alpha < 1/4$ if $M_{q,\lambda}$ is large enough (say $M_{q,\lambda} \geq b_6(\alpha)$):

$$\begin{aligned} -\log \kappa_{q-1}^2(\lambda) &= (1 - \kappa_{q-1}^2(\lambda))(1 + e^{-c_{39} M_{q,\lambda}^{3/2}} \delta_3(q,\lambda)) \\ &\overset{(2.1)}{=} \frac{\gamma_{q,\lambda} + 2\sqrt{1-\gamma_{q,\lambda}^2}}{4\pi q (1-\gamma_{q,\lambda}^2)} e^{2qw_0(\gamma_{q,\lambda})} (1 + M_{q,\lambda}^{-3/4+3\alpha} \delta_4(q,\lambda,\alpha)) \end{aligned} \tag{3.6}$$

with some bounded error terms $|\delta_3(q,\lambda)| \leq c_{40}$, $|\delta_4(q,\lambda,\alpha)| \leq c_{41}(\alpha)$. We sum over these approximations: To bound the error term, we observe that $M_{q,\lambda}$ is monotonically increasing in the argument q for fixed λ. If $M_{l,\lambda} \geq b_6(\alpha)$, a combination of (3.6) and (1.23) yields for some bounded error term $|\delta_5(l,\lambda,\alpha)| \leq c_{41}(\alpha)$:

$$-\log \phi_l(\lambda) = \left(1 + M_{l,\lambda}^{-3/4+3\alpha} \delta_5(l,\lambda,\alpha)\right) \sum_{q=l+1}^{\infty} \frac{\gamma_{q,\lambda} + 2\sqrt{1-\gamma_{q,\lambda}^2}}{4\pi q (1-\gamma_{q,\lambda}^2)} e^{2qw_0(\gamma_{q,\lambda})}. \tag{3.7}$$

We derive upper and lower bounds for the sum in (3.7), starting with the upper bound: First we observe that for $0 < \gamma < 1$ the map $\gamma \mapsto (\gamma+2\sqrt{1-\gamma^2})/(1-\gamma^2)$ is monotonically increasing, and for fixed λ, $\gamma_{q,\lambda} = 2\sqrt{\lambda}/q$ is monotonically decreasing in q; hence

$$\sum_{q=l+1}^{\infty} \frac{\gamma_{q,\lambda} + 2\sqrt{1-\gamma_{q,\lambda}^2}}{4\pi q(1-\gamma_{q,\lambda}^2)} e^{2qw_0(\gamma_{q,\lambda})} \leq \frac{\gamma_{l,\lambda} + 2\sqrt{1-\gamma_{l,\lambda}^2}}{4\pi l(1-\gamma_{l,\lambda}^2)} \sum_{q=l+1}^{\infty} e^{2qw_0(\gamma_{q,\lambda})}. \tag{3.8}$$

We use a linear bound for the exponent in (3.8): An explicit calculation shows

$$\frac{\partial}{\partial q}(qw_0(\gamma_{q,\lambda})) = -\operatorname{arcosh}\frac{1}{\gamma_{q,\lambda}}, \quad \frac{\partial^2}{\partial q^2}(qw_0(\gamma_{q,\lambda})) = -\frac{\gamma_{q,\lambda}}{2\sqrt{\lambda}\sqrt{1-\gamma_{q,\lambda}^2}} < 0; \tag{3.9}$$

hence

$$qw_0(\gamma_{q,\lambda}) \leq lw_0(\gamma_{l,\lambda}) - (q-l)\operatorname{arcosh}\frac{1}{\gamma_{l,\lambda}}. \tag{3.10}$$

Therefore

$$\begin{aligned} \sum_{q=l+1}^{\infty} e^{2qw_0(\gamma_{q,\lambda})} &\leq e^{2lw_0(\gamma_{l,\lambda})} \sum_{k=1}^{\infty} e^{-2k\operatorname{arcosh}\gamma_{l,\lambda}^{-1}} \\ &= \frac{e^{2lw_0(\gamma_{l,\lambda})}}{e^{2\operatorname{arcosh}\gamma_{l,\lambda}^{-1}} - 1} = \frac{\gamma_{l,\lambda}^2}{2\sqrt{1-\gamma_{l,\lambda}^2}\left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)} e^{2lw_0(\gamma_{l,\lambda})}, \end{aligned} \tag{3.11}$$

and hence, using (3.7) and (3.8):

$$-\log\phi_l(\lambda) \leq \left(1 + M_{l,\lambda}^{-3/4+3\alpha}\delta_5(l,\lambda,\alpha)\right) \frac{\left(\gamma_{l,\lambda} + 2\sqrt{1-\gamma_{l,\lambda}^2}\right)\gamma_{l,\lambda}^2 e^{2lw_0(\gamma_{l,\lambda})}}{8\pi l(1-\gamma_{l,\lambda}^2)^{3/2}\left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)}. \tag{3.12}$$

Next we derive a lower bound for the sum in (3.7): We choose a fixed number α_1 with $0 < \alpha_1 \leq 3/8$, e.g. $\alpha_1 = 1/4$, and define

$$m_{l,\lambda} := \left\lceil \frac{M_{l,\lambda}^{\alpha_1}}{\operatorname{arcosh}\gamma_{l,\lambda}^{-1}} \right\rceil, \tag{3.13}$$

where $\lceil x \rceil$ denotes the smallest integer j with $j \geq x$. It will turn out that it suffices to consider only $m_{l,\lambda}$ summands in (3.7) to derive a good lower bound. Observe that for $0 < x < 1$ the bound $\operatorname{arcosh} x^{-1} \geq 2^{1/2}(1-x)^{1/2}$ holds true. Therefore we get the following estimates for k with $l \leq k \leq l + m_{l,\lambda}$ and $M_{l,\lambda} \geq 1$:

$$m_{l,\lambda} \leq 1 + \frac{M_{l,\lambda}^{\alpha_1}}{\operatorname{arcosh}\gamma_{l,\lambda}^{-1}} \leq \frac{2M_{l,\lambda}^{\alpha_1}}{\sqrt{1-\gamma_{l,\lambda}}}, \tag{3.14}$$

$$\left|\frac{1-\gamma_{k,\lambda}}{1-\gamma_{l,\lambda}}-1\right| = \gamma_{k,\lambda}\frac{k-l}{l-2\sqrt{\lambda}} \le \gamma_{k,\lambda}\frac{m_{l,\lambda}}{l-2\sqrt{\lambda}} \le \frac{m_{l,\lambda}}{l-2\sqrt{\lambda}} \tag{3.15}$$

$$\le \frac{2M_{l,\lambda}^{\alpha_1}}{l(1-\gamma_{l,\lambda})^{3/2}} = 2M_{l,\lambda}^{-3/2+\alpha_1},$$

$$\left|\frac{\gamma_{l,\lambda}}{\gamma_{k,\lambda}}-1\right| = \frac{k}{l}-1 \le \frac{m_{l,\lambda}}{l} \le \frac{m_{l,\lambda}}{l-2\sqrt{\lambda}} \le 2M_{l,\lambda}^{-3/2+\alpha_1}. \tag{3.16}$$

In the calculations below, we use the following Lipschitz-continuity arguments: There are positive constants c_{42} and c_{43}, such that for all ϵ with $0 < \epsilon \le c_{42}$ and for all $x_1, x_2, y_1, y_2 > 0$ with $|x_1/x_2 - 1| \le \epsilon$ and $|y_1/y_2 - 1| \le \epsilon$ the following holds: $|(x_1+y_1)/(x_2+y_2) - 1| \le \epsilon$, $|(x_1y_1)/(x_2y_2) - 1| \le c_{43}\epsilon$, $|(x_1/y_1)/(x_2/y_2) - 1| \le c_{43}\epsilon$, $|\sqrt{x_1}/\sqrt{x_2} - 1| \le c_{43}\epsilon$. We get successively for $M_{l,\lambda}$ being large enough, say $M_{l,\lambda} \ge b_7(\alpha_1) \ge 1$:

$$\left|\frac{1+\gamma_{l,\lambda}}{1+\gamma_{q,\lambda}}-1\right| \overset{(3.16)}{\le} 2M_{l,\lambda}^{-3/2+\alpha_1}, \qquad \left|\frac{1-\gamma_{l,\lambda}^2}{1-\gamma_{q,\lambda}^2}-1\right| \overset{(3.15)}{\le} c_{44}M_{l,\lambda}^{-3/2+\alpha_1}, \tag{3.17}$$

$$\left|\frac{\sqrt{1-\gamma_{l,\lambda}^2}}{\sqrt{1-\gamma_{q,\lambda}^2}}-1\right| \le c_{45}M_{l,\lambda}^{-3/2+\alpha_1}, \qquad \left|\frac{\gamma_{q,\lambda}(1-\gamma_{l,\lambda}^2)}{(1-\gamma_{q,\lambda}^2)\gamma_{l,\lambda}}-1\right| \overset{(3.16,3.17)}{\le} c_{46}M_{l,\lambda}^{-3/2+\alpha_1}, \tag{3.18}$$

$$\left|\frac{\frac{\gamma_{q,\lambda}}{1-\gamma_{q,\lambda}^2}+\frac{2}{\sqrt{1-\gamma_{q,\lambda}^2}}}{\frac{\gamma_{l,\lambda}}{1-\gamma_{l,\lambda}^2}+\frac{2}{\sqrt{1-\gamma_{l,\lambda}^2}}}-1\right| \overset{(3.18)}{\le} c_{47}M_{l,\lambda}^{-3/2+\alpha_1}, \tag{3.19}$$

$$\left|\frac{\gamma_{q,\lambda}+2\sqrt{1-\gamma_{q,\lambda}^2}}{q(1-\gamma_{q,\lambda}^2)}\frac{l(1-\gamma_{l,\lambda}^2)}{\gamma_{l,\lambda}+2\sqrt{1-\gamma_{l,\lambda}^2}}-1\right| \overset{(3.19,3.16)}{\le} c_{48}M_{l,\lambda}^{-3/2+\alpha_1}. \tag{3.20}$$

We expand $qw_0(\gamma_{q,\lambda})$ for $l \le q \le l+m_{l,\lambda}$: We assume again that $M_{l,\lambda}$ is large enough, say $M_{l,\lambda} \ge b_8(\alpha_1) \ge 1$:

$$\left|qw_0(\gamma_{q,\lambda}) - lw_0(\gamma_{l,\lambda}) - (q-l)\frac{\partial}{\partial l}(lw_0(\gamma_{l,\lambda}))\right| \le \frac{m_{l,\lambda}^2}{2}\sup_{k\in[l,l+m_{l,\lambda}]}\left|\frac{\partial^2}{\partial k^2}(kw_0(\gamma_{k,\lambda}))\right|$$

$$\overset{(3.14,3.9)}{\le} \frac{2M_{l,\lambda}^{2\alpha_1}}{1-\gamma_{l,\lambda}}\sup_{k\in[l,l+m_{l,\lambda}]}\frac{1}{k\sqrt{1-\gamma_{k,\lambda}^2}} \le \frac{2M_{l,\lambda}^{2\alpha_1}}{l(1-\gamma_{l,\lambda})}\sup_{k\in[l,l+m_{l,\lambda}]}\frac{1}{\sqrt{1-\gamma_{k,\lambda}}} \tag{3.21}$$

$$\overset{(3.15)}{\le} \frac{2M_{l,\lambda}^{2\alpha_1}}{l(1-\gamma_{l,\lambda})^{3/2}}(1+c_{49}M_{l,\lambda}^{-3/2+\alpha_1}) \le 3M_{l,\lambda}^{-3/2+2\alpha_1}.$$

Using the estimates (3.20) and (3.21), we obtain:

$$\sum_{q=l+1}^{\infty} \frac{\gamma_{q,\lambda}+2\sqrt{1-\gamma_{q,\lambda}^2}}{4\pi q(1-\gamma_{q,\lambda}^2)} e^{2qw_0(\gamma_{q,\lambda})} \geq \sum_{q=l+1}^{l+m_{l,\lambda}} \frac{\gamma_{q,\lambda}+2\sqrt{1-\gamma_{q,\lambda}^2}}{4\pi q(1-\gamma_{q,\lambda}^2)} e^{2qw_0(\gamma_{q,\lambda})} \tag{3.22}$$

$$\geq (1-c_{48}M_{l,\lambda}^{-3/2+\alpha_1}) \frac{\gamma_{l,\lambda}+2\sqrt{1-\gamma_{l,\lambda}^2}}{4\pi l(1-\gamma_{l,\lambda}^2)} \sum_{q=l+1}^{l+m_{l,\lambda}} e^{2lw_0(\gamma_{l,\lambda})+2(q-l)\frac{\partial}{\partial l}(lw_0(\gamma_{l,\lambda}))-6M_{l,\lambda}^{-3/2+2\alpha_1}}.$$

The last sum is bounded from below by

$$(1-c_{50}M_{l,\lambda}^{-3/2+2\alpha_1}) e^{2lw_0(\gamma_{l,\lambda})} \sum_{k=1}^{m_{l,\lambda}} e^{-2k \operatorname{arcosh} \gamma_{l,\lambda}^{-1}} \tag{3.23}$$

$$= (1-c_{50}M_{l,\lambda}^{-3/2+2\alpha_1})(1-e^{-2m_{l,\lambda} \operatorname{arcosh} \gamma_{l,\lambda}^{-1}}) \frac{e^{2lw_0(\gamma_{l,\lambda})}}{e^{2 \operatorname{arcosh} \gamma_{l,\lambda}^{-1}}-1}$$

$$\overset{(3.13)}{\geq} (1-c_{50}M_{l,\lambda}^{-3/2+2\alpha_1})(1-e^{-2M_{l,\lambda}^{\alpha_1}}) \frac{e^{2lw_0(\gamma_{l,\lambda})}}{e^{2 \operatorname{arcosh} \gamma_{l,\lambda}^{-1}}-1}$$

$$\geq (1-c_{51}M_{l,\lambda}^{-3/2+2\alpha_1}) \frac{e^{2lw_0(\gamma_{l,\lambda})}}{e^{2 \operatorname{arcosh} \gamma_{l,\lambda}^{-1}}-1}.$$

We define an error term $\delta_6(l,\lambda,\alpha)$ implicitly by the following equation:

$$-\log \phi_l(\lambda) = \left(1+\delta_6(l,\lambda,\alpha)M_{l,\lambda}^{-3/4+3\alpha}\right) \frac{\left(\gamma_{l,\lambda}+2\sqrt{1-\gamma_{l,\lambda}^2}\right)\gamma_{l,\lambda}^2 e^{2lw_0(\gamma_{l,\lambda})}}{8\pi l(1-\gamma_{l,\lambda}^2)^{3/2}\left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)}. \tag{3.24}$$

We combine (3.7), (3.22), (3.23), and the last step in (3.11) to obtain a lower bound $\delta_6(l,\lambda,\alpha) \geq -c_{52}(\alpha)$ for the error term. The upper bound (3.12) tells us that $\delta_6(l,\lambda,\alpha)$ is bounded form above, too, hence $|\delta_6(l,\lambda,\alpha)| \leq c_{53}(\alpha)$. We also need the following rough bound for (3.24):

$$|\log \phi_l(\lambda)| \leq c_{54} \frac{e^{2lw_0(\gamma_{l,\lambda})}}{l(1-\gamma_{l,\lambda}^2)^{3/2}} \overset{(3.4)}{\leq} c_{54} M_{l,\lambda}^{-3/2} e^{-c_{39}M_{l,\lambda}^{3/2}}, \tag{3.25}$$

$$\text{hence} \quad \left|\frac{-\log \phi_l(\lambda)}{1-\phi_l(\lambda)} - 1\right| \leq c_{55} M_{l,\lambda}^{-3/2} e^{-c_{39}M_{l,\lambda}^{3/2}}. \tag{3.26}$$

The estimates (3.26) and (3.24) together yield (3.1); this finishes the proof of Theorem 3.1.

□

4 Depoissonisation

We start with a quantitative continuity consideration for ϕ_l:

Lemma 4.1 *There are constants $c_{56} \in]0, 1/2]$ and $c_{57} > 0$, and for every fixed $\alpha \in]0, 1/4[$ there are positive constants $b_9(\alpha)$ and $c_{58}(\alpha)$, such that for all $l, n \in \mathbb{N}$ with $l > 2\sqrt{n}$ and $M_{l,n} \geq b_9(\alpha)$, for all $\xi \in \mathbb{R}$ with*

$$\frac{|\xi|}{1-\gamma_{l,n}} \leq c_{56}, \tag{4.1}$$

and for $\lambda := n(1+\xi)$ the following bound holds:

$$\left|\log \frac{1-\phi_l(\lambda)}{1-\phi_l(n)}\right| \leq c_{58}(\alpha) M_{l,n}^{-3/4+3\alpha} + \frac{c_{57}|\xi| M_{l,n}^{3/2}}{1-\gamma_{l,n}}. \tag{4.2}$$

Proof of Lemma 4.1. Choose $c_{56} \in]0, 1/2]$ small enough, take $\alpha \in]0, 1/4[$ fixed, set $b_9(\alpha) := 2b_3(\alpha)$, where $b_3(\alpha)$ is taken as in Theorem 3.1. Then for some positive constants c_{59}, c_{60}, c_{61}, c_{62}, c_{57}, and $c_{58}(\alpha)$ the considerations below hold true: Let $l, n \in \mathbb{N}$, $\xi \in \mathbb{R}$, and $\lambda > 0$ fulfill the hypotheses of Lemma 4.1. We compare $\gamma_{l,n}$ with $\gamma_{l,\lambda}$:

$$\left|\frac{\gamma_{l,\lambda}}{\gamma_{l,n}} - 1\right| = \left|\sqrt{\frac{\lambda}{n}} - 1\right| = \left|\sqrt{1+\xi} - 1\right| \overset{(4.1)}{\leq} |\xi|, \tag{4.3}$$

$$\left|\frac{1-\gamma_{l,\lambda}}{1-\gamma_{l,n}} - 1\right| = \frac{\gamma_{l,n}}{1-\gamma_{l,n}}\left|\frac{\gamma_{l,\lambda}}{\gamma_{l,n}} - 1\right| \overset{(4.3)}{\leq} \frac{\gamma_{l,n}|\xi|}{1-\gamma_{l,n}}. \tag{4.4}$$

We combine these two estimates in the form

$$\max\left\{\left|\frac{\gamma_{l,\lambda}}{\gamma_{l,n}} - 1\right|, \left|\frac{1-\gamma_{l,\lambda}}{1-\gamma_{l,n}} - 1\right|\right\} \leq \frac{|\xi|}{1-\gamma_{l,n}}. \tag{4.5}$$

As a consequence,

$$\left|\frac{M_{l,\lambda}}{M_{l,n}} - 1\right| \overset{(1.12)}{=} \left|\frac{1-\gamma_{l,\lambda}}{1-\gamma_{l,n}} - 1\right| \leq \frac{|\xi|}{1-\gamma_{l,n}} \overset{(4.1)}{\leq} \frac{1}{2}, \tag{4.6}$$

$$M_{l,\lambda} \overset{(4.6)}{\geq} \frac{M_{l,n}}{2}. \tag{4.7}$$

Using (4.5), Lipschitz estimates similar to (3.17–3.20) yield:

$$\left|\frac{\left(\gamma_{l,\lambda} + 2\sqrt{1-\gamma_{l,\lambda}^2}\right)\gamma_{l,\lambda}^2}{(1-\gamma_{l,\lambda}^2)^{3/2}\left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)} \frac{(1-\gamma_{l,n}^2)^{3/2}\left(1+\sqrt{1-\gamma_{l,n}^2}\right)}{\left(\gamma_{l,n} + 2\sqrt{1-\gamma_{l,n}^2}\right)\gamma_{l,n}^2} - 1\right| \leq \frac{c_{59}|\xi|}{1-\gamma_{l,n}}. \tag{4.8}$$

Furthermore,

$$|w_0(\gamma_{l,\lambda}) - w_0(\gamma_{l,n})| \overset{(3.3)}{=} \left| \int_{\gamma_{l,\lambda}}^{\gamma_{l,n}} \frac{\sqrt{1-x^2}}{x}\, dx \right| \leq |\gamma_{l,n} - \gamma_{l,\lambda}| \max_{x \in \{\gamma_{l,\lambda}, \gamma_{l,n}\}} \frac{\sqrt{1-x^2}}{x} \tag{4.9}$$

$$\overset{(4.5)}{\leq} |\gamma_{l,n} - \gamma_{l,\lambda}| \frac{\sqrt{1-\gamma_{l,n}^2}}{\gamma_{l,n}} (1 + c_{60}|\xi|(1-\gamma_{l,n})^{-1}) \overset{(4.3,4.1)}{\leq} c_{61}|\xi|(1-\gamma_{l,n})^{1/2},$$

then

$$l|w_0(\gamma_{l,\lambda}) - w_0(\gamma_{l,n})| \overset{(4.9,1.12)}{\leq} \frac{c_{61}|\xi| M_{l,n}^{3/2}}{1-\gamma_{l,n}}, \tag{4.10}$$

hence we get for $M_{l,n}$ being large enough, say $M_{l,n} \geq b_9(\alpha)$:

$$\begin{aligned} \left| \log \frac{1-\phi_l(\lambda)}{1-\phi_l(n)} \right| &\overset{(3.1)}{\leq} \left| \log \left(\frac{1+\delta_7(l,\lambda,\alpha) M_{l,\lambda}^{-3/4+3\alpha}}{1+\delta_7(l,n,\alpha) M_{l,n}^{-3/4+3\alpha}} \right) \right| \\ &+ \left| \log \left(\frac{\left(\gamma_{l,\lambda} + 2\sqrt{1-\gamma_{l,\lambda}^2}\right) \gamma_{l,\lambda}^2}{(1-\gamma_{l,\lambda}^2)^{3/2} \left(1+\sqrt{1-\gamma_{l,\lambda}^2}\right)} \frac{(1-\gamma_{l,n}^2)^{3/2} \left(1+\sqrt{1-\gamma_{l,n}^2}\right)}{\left(\gamma_{l,n} + 2\sqrt{1-\gamma_{l,n}^2}\right) \gamma_{l,n}^2} \right) \right| \\ &+ 2l|w_0(\gamma_{l,\lambda}) - w_0(\gamma_{l,n})| \\ &\overset{(3.1,4.7,\ 4.8,4.10)}{\leq} c_{58}(\alpha) M_{l,n}^{-3/4+3\alpha} + \frac{c_{62}|\xi|}{1-\gamma_{l,n}} + \frac{2c_{61}|\xi| M_{l,n}^{3/2}}{1-\gamma_{l,n}} \\ &\leq c_{58}(\alpha) M_{l,n}^{-3/4+3\alpha} + \frac{c_{57}|\xi| M_{l,n}^{3/2}}{1-\gamma_{l,n}}. \end{aligned} \tag{4.11}$$

This proves Lemma 4.1.

□

For $l \in \mathbb{N}_0$ and $n \in \mathbb{N}$ let $q_{l,n}$ denote the probability that the longest increasing subsequence in a random permutation of $\{1, \ldots, n\}$ (with the uniform distribution) has a length $L_n \leq l$.

For $\lambda > 0$, let $\mathbb{P}_\lambda$ denote the Poisson distribution with parameter λ over $\mathbb{N}_0$, and let N denote the identity function on $\mathbb{N}_0$, thus N is a Poissonian random variable with parameter λ with respect to $\mathbb{P}_\lambda$. The expectation operator corresponding to $\mathbb{P}_\lambda$ is denoted by $\mathbb{E}_\lambda$. We know $\phi_l(\lambda) = \mathbb{E}_\lambda[q_{l,N}]$; see formula (1.11) in [3]. Furthermore we have the monotonicity $q_{l,n+1} \leq q_{l,n}$ for all $n \in \mathbb{N}_0$; see Lemma 8.1 in [3]. We state the following quantitative version of (1.14):

Lemma 4.2 *There are constants $c_{63} \in]0,1[$, $c_{64} > 0$, and for all fixed $\beta \in]0, 3/4[$ there are constants $b_{12}(\beta) > 1$ and $c_{65}(\beta) > 0$, such that for all l, n with $\gamma_{l,n} \in [c_{63}, 1[$ and $M_{l,n} \geq b_{12}(\beta)$ we have*

$$1 - q_{l,n} = \frac{e^{2lw_0(\gamma_{l,n})}}{8\pi l (1-\gamma_{l,n}^2)^{3/2}} \delta_9(l,n) \tag{4.12}$$

with an error term δ_9 that fulfills the bound

$$|\log \delta_9(l,n)| \le c_{65}(\beta) M_{l,n}^{-\beta} + c_{64}\sqrt{(1-\gamma_{l,n})|\log(1-\gamma_{l,n})|}. \tag{4.13}$$

Proof of part 2 of Theorem 1.2 and of Lemma 4.2. Consider $\xi \in \mathbb{R}$, $\lambda = n(1+\xi) > 0$ such that (4.1) holds. (Specific choices for ξ and λ will be given below.) We observe that $d\mathbb{P}_\lambda = e^{n-\lambda}\left(\frac{\lambda}{n}\right)^N d\mathbb{P}_n$. If $\xi < 0$, i.e. if $\lambda < n$, then $d\mathbb{P}_n/d\mathbb{P}_\lambda$ is monotonically increasing, and if $\xi > 0$, then $d\mathbb{P}_n/d\mathbb{P}_\lambda$ is monotonically decreasing. Abbreviating

$$v := \frac{d\mathbb{P}_\lambda}{d\mathbb{P}_n}(n) = e^{n-\lambda}\left(\frac{\lambda}{n}\right)^n = \exp\{n(\log(1+\xi)-\xi)\} \tag{4.14}$$

and $s_a := 1$ for $a \ge 0$, $s_a := -1$ for $a < 0$, we get

$$s_{N-n}s_\xi\left(1 - v\frac{d\mathbb{P}_n}{d\mathbb{P}_\lambda}\right) \ge 0. \tag{4.15}$$

Furthermore the fact that $n \mapsto q_{l,n}$ is monotonically decreasing implies

$$s_{N-n}(1-q_{l,n}) \le s_{N-n}(1-q_{l,N}). \tag{4.16}$$

Using the term on the right hand side in (4.2), we abbreviate:

$$\epsilon_{l,n,\alpha}(\xi) := \exp\left\{s_\xi\left(c_{58}(\alpha)M_{l,n}^{-3/4+3\alpha} + \frac{c_{57}|\xi|M_{l,n}^{3/2}}{1-\gamma_{l,n}}\right)\right\}, \tag{4.17}$$

thus we can rewrite (4.2) as

$$s_\xi(1-\phi_l(\lambda)) \le s_\xi \epsilon_{l,n,\alpha}(\xi)(1-\phi_l(n)). \tag{4.18}$$

Using the inequality $s_\xi(\log(1+\xi)-\xi) \ge -s_\xi\xi^2/2$ (recall $|\xi| < 1$), we obtain

$$s_\xi v \overset{(4.14)}{\ge} s_\xi e^{-n\xi^2/2}. \tag{4.19}$$

Then

$$\begin{aligned}
s_\xi(1-q_{l,n})(1-v) &= s_\xi \mathbb{E}_\lambda\left[(1-q_{l,n})\left(1-v\frac{d\mathbb{P}_n}{d\mathbb{P}_\lambda}\right)\right] \qquad (4.20)\\
&\overset{(4.15,4.16)}{\le} s_\xi \mathbb{E}_\lambda\left[(1-q_{l,N})\left(1-v\frac{d\mathbb{P}_n}{d\mathbb{P}_\lambda}\right)\right] = s_\xi(\mathbb{E}_\lambda[1-q_{l,N}] - v\mathbb{E}_n[1-q_{l,N}])\\
&= s_\xi(1-\phi_l(\lambda) - v(1-\phi_l(n))) \overset{(4.18)}{\le} s_\xi(\epsilon_{l,n,\alpha}(\xi)-v)(1-\phi_l(n))\\
&\overset{(4.19)}{\le} s_\xi(\epsilon_{l,n,\alpha}(\xi) - e^{-n\xi^2/2})(1-\phi_l(n)).
\end{aligned}$$

In the case $1 > \xi \geq 0$, the bound $\log(1+\xi) - \xi \leq -\xi^2/4$ implies the upper bound $v \leq e^{-n\xi^2/4}$ for v in addition to the lower bound (4.19). Consequently we get

$$\epsilon_{l,n,\alpha}(-|\xi|) - e^{-n\xi^2/2} \overset{(4.20)}{\leq} \frac{1-q_{l,n}}{1-\phi_l(n)} \overset{(4.20)}{\leq} \frac{\epsilon_{l,n,\alpha}(|\xi|)}{1-e^{-n\xi^2/4}}, \tag{4.21}$$

and thus, a little rougher, but more symmetrically:

$$\epsilon_{l,n,\alpha}(-|\xi|) - e^{-n\xi^2/4} \leq \frac{1-q_{l,n}}{1-\phi_l(n)} \leq \left(\epsilon_{l,n,\alpha}(-|\xi|) - e^{-n\xi^2/4}\right)^{-1}, \tag{4.22}$$

provided the bounds are positive. An exact maximisation of $\epsilon_{l,n,\alpha}(-|\xi|) - e^{-n\xi^2/4}$ would lead to a transcendental equation, but for our purposes, rougher bounds suffice. In fact we derive two different lower bounds for the maximum, the first one being more adapted to the LD regime, the second one being more useful in the MD regime. For the first bound, we choose positive constants c_{66} so large and c_{67} so small that $c_{68} := c_{66}^2/16 > c_{67} + c_{66}c_{57} =: c_{69}$. Then we set

$$\xi := c_{66}\gamma_{l,n}^{-2}(1-\gamma_{l,n})^2 M_{l,n}^{-3/2}, \tag{4.23}$$

and for $M_{l,n}$ being so large that

$$c_{58}(\alpha) M_{l,n}^{-3/4+3\alpha} \leq c_{67}\gamma_{l,n}^{-2}(1-\gamma_{l,n}) \tag{4.24}$$

we obtain

$$\frac{n\xi^2}{4} \overset{(1.12)}{=} c_{68}\gamma_{l,n}^{-2}(1-\gamma_{l,n}), \tag{4.25}$$

$$c_{58}(\alpha) M_{l,n}^{-3/4+3\alpha} + \frac{c_{57}\xi M_{l,n}^{3/2}}{1-\gamma_{l,n}} \leq c_{69}\gamma_{l,n}^{-2}(1-\gamma_{l,n}). \tag{4.26}$$

Thus we get the following lower bound for the term in (4.22):

$$\epsilon_{l,n,\alpha}(-|\xi|) - e^{-n\xi^2/4} \geq g(\gamma_{l,n}) \tag{4.27}$$

with

$$g(\gamma) := e^{-c_{69}\gamma^{-2}(1-\gamma)} - e^{-c_{68}\gamma^{-2}(1-\gamma)} > 0 \quad \text{for } 0 < \gamma < 1, \tag{4.28}$$

hence by (4.22):

$$g(\gamma_{l,n}) \leq \frac{1-q_{l,n}}{1-\phi_l(n)} \leq g(\gamma_{l,n})^{-1}. \tag{4.29}$$

Take a fixed $\alpha \in]0, 1/4[$. We choose a continuous function $f_1 :]0,1[\to]0,\infty[$ so large that the assumption $M_{l,n} \geq f_1(\gamma_{l,n})$ implies (4.24), $M_{l,n} \geq b_9(\alpha)$, $c_{37}(\alpha) M_{l,n}^{-3/4+3\alpha} \leq 1/2$, and $c_{66}\gamma_{l,n}^{-2}(1-\gamma_{l,n}) M_{l,n}^{-3/2} \leq c_{56}$. (The last condition is just (4.1) combined with the choice

(4.23) of ξ.) Then Theorem 3.1 is applicable: we apply the bounds (4.29) and (3.1); then we get with the definitions

$$f_\pm(\gamma) := \frac{\left(\gamma + 2\sqrt{1-\gamma^2}\right)\gamma^2}{8\pi(1-\gamma^2)^{3/2}\left(1+\sqrt{1-\gamma^2}\right)} 2^{\pm 1} g(\gamma)^{\mp 1} \tag{4.30}$$

the desired bounds (1.16) for $M_{l,n} \geq f_1(\gamma_{l,n})$; thus we proved part 2. of Theorem 1.2.

We derive a second bound for (4.22) next, which is well adapted for the MD regime: We set

$$\xi := \sqrt{2n^{-1}|\log(1-\gamma_{l,n})|} \overset{(1.12)}{=} 2^{3/2}\gamma_{l,n}^{-1}(1-\gamma_{l,n})^{3/2}|\log(1-\gamma_{l,n})|^{1/2} M_{l,n}^{-3/2}. \tag{4.31}$$

Then

$$e^{-n\xi^2/4} = (1-\gamma_{l,n})^{1/2}, \tag{4.32}$$

$$\frac{c_{57}\xi M_{l,n}^{3/2}}{1-\gamma_{l,n}} = 2^{3/2}c_{57}\gamma_{l,n}^{-1}(1-\gamma_{l,n})^{1/2}|\log(1-\gamma_{l,n})|^{1/2}. \tag{4.33}$$

Let $b_9(\alpha)$ be taken as in Lemma 4.1, where α is given by $\beta = 3/4 - 3\alpha$. We choose constants $c_{63} \in]0,1[$ so close to 1 and $b_{12}(\beta) \geq b_9(\alpha)$ so large that the assumptions $\gamma_{l,n} \in [c_{63}, 1[$ and $M_{l,n} \geq b_{12}(\beta)$ imply, say, $\epsilon_{l,n,\alpha}(-\xi) \geq 1/2$, $c_{37}(\alpha)M^{-\beta} \leq 1/2$, $(1-\gamma_{l,n})^{1/2} \leq 1/4$, and (4.1); see (4.17) and (4.33). Then we get for some positive constants c_{70}, c_{71}, c_{72}:

$$\begin{aligned} \epsilon_{l,n,\alpha}(-\xi) - e^{-n\xi^2/4} &\geq \epsilon_{l,n,\alpha}(-\xi)(1-2e^{-n\xi^2/4}) \\ &\overset{(4.32)}{\geq} \epsilon_{l,n,\alpha}(-\xi)\exp\{-c_{70}(1-\gamma_{l,n})^{1/2}\} \\ &\overset{(4.17,4.33)}{\geq} \exp\left\{-c_{58}(\alpha)M_{l,n}^{-\beta} - c_{71}(1-\gamma_{l,n})^{1/2}|\log(1-\gamma_{l,n})|^{1/2}\right\}, \end{aligned} \tag{4.34}$$

and by Lipschitz arguments

$$\left|\log\frac{\gamma_{l,n}+2\sqrt{1-\gamma_{l,n}^2}}{1+\sqrt{1-\gamma_{l,n}^2}}\right| \leq c_{72}(1-\gamma_{l,n})^2 \leq c_{72}(1-\gamma_{l,n})^2|\log(1-\gamma_{l,n})|^{1/2}, \tag{4.35}$$

and hence (4.12/4.13) follows by (3.1) and (4.22). Thus Lemma 4.2 is proved.

□

Proof of part 1. in Theorem 1.2 and of Theorem 1.1. The statement (1.14) is an immediate consequence of Lemma 4.2. We observe

$$2lw_0(\gamma_{l,n}) \sim -\frac{4\sqrt{2}}{3}l(1-\gamma_{l,n})^{3/2} = -\frac{4\sqrt{2}}{3}\frac{(l-2\sqrt{n})^{3/2}}{\sqrt{l}} = \frac{4\sqrt{2}}{3}M_{l,n}^{3/2} \tag{4.36}$$

as $\gamma_{l,n} \nearrow 1$ by (2.3) and (1.12). In the estimate (4.37) below, the notation $a \ll b$ means that a/b converges to 0. Furthermore,

$$|\log(8\pi l(1-\gamma_{l,n}^2)^{3/2})| \overset{(1.12)}{=} \left|\log(8\pi(1+\gamma_{l,n})^{3/2}) + \frac{3}{2}\log M_{l,n}\right| \ll M_{l,n}^{3/2} \tag{4.37}$$

as $M_{l,n} \to \infty$. The combination of (1.14), (4.36) and (4.37) yields the claim (1.15); this finishes the proof of Theorem 1.2.

To derive Theorem 1.1 we take $l = l(n) := \lceil 2\sqrt{n} + tn^{1/2-\eta} \rceil$: We have as $n \to \infty$: $\gamma_{l(n),n} \nearrow 1$, i.e. $l(n) \sim 2\sqrt{n}$ from $\eta > 0$, and

$$M_{l(n),n} \sim tn^{1/2-\eta} l^{-1/3} \sim 2^{-1/3} t n^{1/3-\eta} \to \infty \tag{4.38}$$

from $\eta < 1/3$. Hence

$$n^{(1-3\eta)/2} t^{3/2} \sim (l - 2\sqrt{n})^{3/2} n^{-1/4} \sim (l - 2\sqrt{n})^{3/2} (2/l)^{1/2}. \tag{4.39}$$

Theorem 1.1 now follows from formula (1.15).

□

A Appendix

Lemma A.1 *There is a constant $c_{73} > 0$ such that for every circle C in the complex plane, every $k > 0$, and every $f \in C^1(C)$ the following bound holds:*

$$\sup_{z \in \mathbb{C} \setminus C} \left| \int_C \frac{f(s)}{s-z}\, ds \right| \le k^{-1} \|f\|_{L^1(C)} + c_{73} k \|f'\|_{L^\infty(C)} + 2\pi \|f\|_{L^\infty(C)}. \tag{A.1}$$

Proof of Lemma A.1. By scaling the circle and scaling k proportional to the radius of the circle, we may assume without loss of generality $k = 1$. Given $z \in \mathbb{C} \setminus C$, let C_1 and C_2 be the arcs of points $s \in C$ with $|s - z| > 1$ and $|s - z| \le 1$, respectively; C_2 may be empty, or it may be the whole circle. We have

$$\left| \int_{C_1} \frac{f(s)}{s-z}\, ds \right| \le \|f\|_{L^1(C_1)} \sup_{s \in C_1} |s-z|^{-1} \le \|f\|_{L^1(C_1)}. \tag{A.2}$$

If C_2 is empty, we are done. Else let x, y denote the start and end point of C_2, respectively. (If C_2 is the whole circle C, we take a point $x = y \in C$ with maximal distance $|x - z|$ from z as start point and end point of C_2.) In the calculations below with $s \in C_2$, $\log((s-z)/(x-z))$ means $\int_x^s dw/(w-z)$, integrated on C_2. By partial integration:

$$\int_{C_2} \frac{f(s)}{s-z}\, ds = -\int_{C_2} f'(s) \log \frac{s-z}{x-z}\, ds + f(y) \log \frac{y-z}{x-z}, \tag{A.3}$$

hence (using $|x-z| = |y-z|$):

$$\left| \int_{C_2} \frac{f(s)}{s-z}\, ds \right| \le \|f'\|_{L^\infty(C_2)} \int_{C_2} \left| \log \frac{s-z}{x-z} \right| |ds| + 2\pi |f(y)|. \tag{A.4}$$

The integral on the right hand side of (A.4) is bounded by a constant c_{73}. Indeed: We split C_2 into pieces $C_{2,j} := \{s \in C_2 : \ e^{-j} < |s-z| \le e^{-j+1}\}$, $j \in \mathbb{N}$. Then the length of $C_{2,j}$ is bounded by $2\pi e^{-j+1}$, and $|\log((s-z)(x-z))| \le j + 2\pi$ for $s \in C_{2,j}$; the last statement follows from $|\mathrm{Im} \log((s-z)(x-z))| \le 2\pi$ and from $|\mathrm{Re} \log((s-z)(x-z))| = -\log(|s-z|/|x-z|) \le -\log|s-z| \le j$; recall that x is a point on C_2 with maximal distance from z and that $|x-z| \le 1$. Hence the integral on the right hand side of (A.4) is bounded by $c_{73} := \sum_{j \in \mathbb{N}} (j + 2\pi) 2\pi e^{-j+1} < \infty$.

Combining (A.2) and (A.4) we obtain (A.1).

□

The next lemma considers the (matrix-valued) Riemann-Hilbert problems with a transition function $I + W$ which is a small perturbation of the identity. Assume that W is defined on a circle C centered at the origin, and the RHP is specified by

$$L_+ = L_-(I+W) \text{ on } C, \qquad L_-(\infty) = I, \tag{A.5}$$

L_+ to be defined inside the disk with boundary C and L_- to be defined outside the disk.

Lemma A.2 *Assume that $\|W\|_{L^\infty(C)} < 1$. Then the RHP (A.5) has a solution $L_\pm$ which fulfills*

$$|L_+(0) - I| \le \frac{1}{|C|} \left(\|W\|_{L^1(C)} + \frac{\|W\|^2_{L^2(C)}}{1 - \|W\|_{L^\infty(C)}} \right). \tag{A.6}$$

Proof of Lemma A.2. By scaling, we may assume that C is the unit circle. Let $C_\pm : L^2(C) \to L^2(C)$ denote the Cauchy operator

$$C_\pm f(z) = \lim_{w \to z} \frac{1}{2\pi i} \oint_C \frac{f(s)}{s-w}\, ds, \tag{A.7}$$

where w is taken in the interior of the unit disk for C_+ and in the exterior of the unit disk for C_-. Then $\pm C_\pm$ are orthogonal projectors with $C_+ - C_- = \mathrm{id}$; this can be seen using a Fourier series of f. (To be precise, $C_\pm$ is first defined for smooth f only and then can be continuously extended to $L^2(C)$, since it is a bounded operator.) We write (A.5) with the substitution $\Lambda_\pm = L_\pm - I$ as

$$\Lambda_+ - \Lambda_- = W + \Lambda_- W \text{ on } C, \qquad \Lambda_-(\infty) = 0, \tag{A.8}$$

and

$$\Lambda_\pm = C_\pm(W + \Lambda_- W) \text{ on } C. \tag{A.9}$$

The map $K : \Lambda \mapsto C_-(\Lambda W)$ is a bounded linear operator in $L^2(C)$ with operator norm $\|K\|_{L^2(C)\to L^2(C)} \leq \|W\|_{L^\infty(C)} < 1$; hence equation (A.9) has a solution Λ_- with

$$\|\Lambda_-\|_{L^2(C)} \leq \frac{\|C_-W\|_{L^2(C)}}{1-\|W\|_{L^\infty(C)}} \leq \frac{\|W\|_{L^2(C)}}{1-\|W\|_{L^\infty(C)}} \tag{A.10}$$

and therefore

$$|\Lambda_+(0)| = \left|\frac{1}{2\pi i}\oint_C (W(s)-\Lambda_-(s)W(s))\,\frac{ds}{s}\right| \leq \frac{1}{2\pi}\left(\|W\|_{L^1(C)} + \|\Lambda_-\|_{L^2(C)}\,\|W\|_{L^2(C)}\right). \tag{A.11}$$

The estimates (A.10) and (A.11) together yield (A.6).

□

Acknowledgement. For F.M.: This work is part of the research programme of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'.

References

[1] D. Aldous, P. Diaconis, *Hammersley's interacting particle process and longest increasing subsequences.* Probab. Theory Related Fields 103 (1995), 199–213.

[2] D. Aldous, P. Diaconis, *Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem.* Bull. Amer. Math. Soc. (N.S.) 36 (1999), 413–432.

[3] J. Baik, P. Deift, K. Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations.* J. Amer. Math. Soc. 12 (1999), 1119–1178.

[4] P. Deift, *Integrable systems and combinatorial theory.* Notices Amer. Math. Soc. 47 (2000), 631–640.

[5] J.D. Deuschel, O. Zeitouni, *On increasing subsequences of I.I.D. samples.*, Combin. Probab. Comput. (1999), 247–263.

[6] P. Erdös, G. Szekeres, *A combinatorial theorem in geometry.*, Compositio Math. 2 (1935), 463–470.

[7] A.S. Fokas, A.R. Its, V.E. Kitaev, *Discrete Painlevé equations and their appearance in quantum gravity.* Comm. Math. Phys. 142 (1991), 313–344.

[8] I.M. Gessel, *Symmetric functions and P-recursiveness.*, J. Combin. Theory Ser. A 53 (1990), 257–285.

[9] P. Groeneboom, *Ulam's problem and Hammersley's process.*, Preprint, Universiteit Delft (2000).

[10] J.M. Hammersley, *A few seedlings of research.* Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics, pp. 345–394. Univ. California Press, Berkeley, Calif., 1972.

[11] K. Johansson, *The longest increasing subsequence in a random permutation and a unitary random matrix model.*, Math. Res. Lett. 5 (1998), no. 1-2, 63–82.

[12] S.V. Kerov, A.M. Vershik, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables.*, Soviet Math. Dokl. 18 (1977), 527–531.

[13] J.F.C. Kingman, *Subadditive processes.* École d'Été de Probabilités de Saint-Flour, V–1975, pp. 167–223. Lecture Notes in Math., Vol. 539, Springer, Berlin, 1976.

[14] B.F. Logan, L.A. Shepp, *A variational problem for random Young tableaux.* Advances in Math. 26 (1977), no. 2, 206–222.

[15] T. Seppäläinen, *Large deviations for increasing sequences on the plane.* Probab. Theory Related Fields 112 (1998), 221–244.

[16] C.A. Tracy, H. Widom, *Level-Spacing distributions and the Airy kernel.* Comm. Math. Phys. 159 (1994), 151-174.

[17] S. Ulam, *Monte Carlo calculations in problems of mathematical physics.* Modern mathematics for the engineer: Second series pp. 261–281. McGraw-Hill, New York (1961)