

# LOWER DEVIATIONS IN $\beta$ -ENSEMBLES AND LAW OF ITERATED LOGARITHM IN LAST PASSAGE PERCOLATION

RIDDHIPRATIM BASU, SHIRSHENDU GANGULY, MILIND HEGDE, AND MANJUNATH KRISHNAPUR

ABSTRACT. For the last passage percolation (LPP) on  $\mathbb{Z}^2$  with exponential passage times, let  $T_n$  denote the passage time from  $(1, 1)$  to  $(n, n)$ . We investigate the law of iterated logarithm of the sequence  $\{T_n\}_{n \geq 1}$ ; we show that  $\liminf_{n \rightarrow \infty} \frac{T_n - 4n}{n^{1/3}(\log \log n)^{1/3}}$  almost surely converges to a deterministic negative constant and obtain some estimates on the same. This settles a conjecture of Ledoux [14] where a related lower bound and similar results for the corresponding upper tail were proved. Our proof relies on a slight shift in perspective from point-to-point passage times to considering point-to-line passage times instead, and exploiting the correspondence of the latter to the largest eigenvalue of the Laguerre Orthogonal Ensemble (LOE). A key technical ingredient, which is of independent interest, is a new lower bound of lower tail deviation probability of the largest eigenvalue of  $\beta$ -Laguerre ensembles, which extends the results proved in the context of the  $\beta$ -Hermite ensembles by Ledoux and Rider [15].

## CONTENTS

1. Introduction and statement of main results	1
2. The Law of Iterated Logarithm: Proof of Theorem 1	5
3. Lower Deviations in $\beta$ -Laguerre ensemble: Proof of Theorem 2	11
References	20

## 1. INTRODUCTION AND STATEMENT OF MAIN RESULTS

Last passage percolation on  $\mathbb{Z}^2$ , where one puts i.i.d. weights on the vertices of  $\mathbb{Z}^2$  and studies the maximum weight of an oriented path between two vertices, is a canonical model believed to be in the (1+1)-dimensional Kardar-Parisi-Zhang (KPZ) universality class. A handful of such models, the so-called exactly solvable models, have been rigorously analysed using some remarkable bijections and connections to random matrix theory, leading to an explosion of activities in the field of integrable probability in recent decades. We shall consider the exponential LPP model on  $\mathbb{Z}^2$  where the field of vertex weights  $\{\xi_v\}_{v \in \mathbb{Z}^2}$  is a family of i.i.d. rate one exponentially distributed random variables.

**Definition 1.1.** *For any up/right path  $\gamma$  in  $\mathbb{Z}^2$ , define the weight of  $\gamma$  as  $\ell(\gamma) := \sum_{v \in \gamma} \xi_v$ , and for  $u, v \in \mathbb{Z}^2$ , with  $u \preceq v$  in the usual partial order, the last passage time  $T_{u,v} = T_{v,u}$  from  $u$  to  $v$  is defined by  $T_{u,v} := \max_{\gamma: u \rightarrow v} \ell(\gamma)$  where the maximum is taken over all oriented paths from  $u$  to  $v$ . For  $n \geq 1$ , we shall denote by  $T_n$  the passage time from  $\mathbf{1}$  to  $\mathbf{n}$  ( $\mathbf{r}$  will denote the point  $(r, r)$  for  $r \in \mathbb{Z}$ ).*

Our primary object of interest will be the family of coupled random variables  $\{T_n\}_{n \geq 1}$ . It is a fact, by now classical, [22] that  $T_n \sim 4n$  and it was shown by Johansson in [12] that  $Z_n := n^{-1/3}(T_n - 4n)$  is a tight sequence of random variables and in particular converges to a scalar multiple of the GUE Tracy-Widom distribution from random matrix theory. Indeed, [12] established the remarkable distributional equality:

$$T_n \stackrel{d}{=} \lambda_n(\text{LUE}_n) \tag{1}$$

where  $\lambda_n(\text{LUE}_n)$  is the largest eigenvalue of the Laguerre Unitary Ensemble (LUE), i.e. the matrix  $X^*X$  where  $X$  is an  $n \times n$  matrix of i.i.d. standard complex Gaussian random variables.

Inspired by a result of Paquette and Zeitouni [20] (see Section 1.1 for details), Ledoux [14] considered the law of iterated logarithm for the sequence  $\{T_n : n \geq 1\}$ , and showed that there exist  $0 < C_1 < C_2 < \infty$  such that almost surely

$$C_1 \leq \limsup_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{2/3}} \leq C_2. \quad (2)$$

Note that the limsup above and the liminf below are almost sure constants by a 0-1 law (see Lemma 2.1). For the liminf, it was shown in [14] that

$$\liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}} > -C_3, \quad (3)$$

almost surely for some  $C_3 < \infty$ , and it was conjectured that  $(\log \log n)^{1/3}$  is indeed the right scale of fluctuation for the lower deviations. The first main result of this paper completes the picture by establishing this conjecture.

**Theorem 1.** *There exists  $C_4 > 0$  such that, almost surely*

$$\liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}} = -C_4. \quad (4)$$

A comparison with the classical law of iterated logarithm for the simple random walk and the results in [20], will be presented in Section 1.1. While we have stated Theorem 1 in the simplest possible form here, a more detailed discussion on the settings considered in [14], the conjectured lim inf value and some possible extensions of Theorem 1 is presented in Section 2.1.

Ledoux's proof for the upper tail [14] is reminiscent of the classical law of iterated logarithm for random walk and uses sub-additivity of  $T_n$  and the moderate deviation estimates for the largest eigenvalue of LUE from [15]. The standard sub-additivity is less useful for the liminf and hence the weaker result in [14]. The starting point in this paper is the observation that the above issue can be circumvented by considering point-to-line LPP and using stochastic ordering between the same and point-to-point LPP.

The main technical ingredient we rely on then is a new lower bound of the lower tail moderate deviation probabilities for the point-to-line last passage time in Exponential LPP. Formally, for any vertex  $v$  and a line  $\mathbb{L}$ , define the point-to-line last passage time  $T_{v,\mathbb{L}} := \sup_{w \in \mathbb{L}} T_{v,w}$ . A particularly canonical case is when  $v = \mathbf{1}$  and  $\mathbb{L}_n = \{x + y = 2n\}$ , in which case we define

$$T_n^* := \sup_{v \in \mathbb{Z}_+^2 : v_1 + v_2 = 2n} T_{\mathbf{1},v}. \quad (5)$$

Similar to how  $n^{-1/3}(T_n - 4n)$  converges to a scalar multiple of the GUE Tracy-Widom distribution, it remarkably turns out and is well-known that  $n^{-1/3}(T_n^* - 4n)$  converges to a scalar multiple of the Gaussian Orthogonal Ensemble (GOE) Tracy-Widom distribution [4]. We shall prove the following new left tail moderate deviation lower bound with optimal exponent for  $T_n^*$ , which will be the key ingredient in the proof of Theorem 1.

**Theorem 1.2.** *There exist  $c_0 > 0$  and  $n_0 \in \mathbb{N}$  such that for all  $n > n_0$  and  $x \in (1, n^{2/3})$  we have*

$$\mathbb{P}(T_n^* \leq 4n - xn^{1/3}) \geq e^{-c_0 x^3}.$$

To prove Theorem 1.2, we shall use the following correspondence from integrable probability between the point-to-line passage time and the largest eigenvalue of the Laguerre Orthogonal Ensemble with certain parameters. Though implicit in the results of [3, 4, 1], we were not able to find an explicit quotable statement in the literature and for completeness provide a proof later in the article using results from [11] and [18].

**Proposition 1.3.** *As defined above,  $T_n^*$  has the same distribution as  $\frac{1}{2}\lambda_{2n-1}$ , where  $\lambda_{2n-1}$  is the largest eigenvalue of  $\text{LOE}_{2n-1}$  (i.e., the largest eigenvalue of  $X^T X$  where  $X$  is a  $2n \times (2n - 1)$  matrix of i.i.d.  $N(0, 1)$  variables).*

Using Proposition 1.3, Theorem 1.2 will follow from a new general lower deviation tail inequality for Laguerre  $\beta$ -ensembles, which is our second main result and is of independent interest (see Theorem 2). In the next section we define  $\beta$ -ensembles, and review in some detail the relevant literature on them and connections to last passage percolation.

**1.1.  $\beta$ -Ensembles, Background and Related Results.** Spectra of classical random matrix ensembles are special cases of a wide class of point processes termed as  $\beta$ -ensembles which are defined through a family of Gibbs measures, with  $\beta$  playing the classical role of inverse temperature. In this framework, the well known Hermite, Laguerre and Jacobi ensembles for parameters  $\beta = 1, 2, 4$  are the ones with the classical random matrix theory representations. For instance, the  $\beta = 2$  case is special as it admits a determinantal structure, for which the Hermite ensemble ( $H_n$ ) corresponds to the eigenvalues of a  $\text{GUE}_n$  matrix, i.e., a hermitian matrix of size  $n$  with i.i.d. standard complex Gaussian entries above the diagonal and independent i.i.d. real Gaussian entries on the diagonal; while for  $m \geq n \geq 1$ , the Laguerre ensemble  $\text{LUE}_{m,n}$  (for  $m = n$  this will simply be denoted by  $\text{LUE}_n$ ) corresponds to the eigenvalues of a complex Wishart matrix, i.e.,  $X^* X$  where  $X$  is an  $m \times n$  matrix of i.i.d. standard complex Gaussians.

For the purposes of this paper we need to define the general  $\beta > 0$  version of the LUE.

**Definition 1.4.** *The Laguerre  $\beta$ -ensemble  $\text{LE}_{m,n}^\beta$ , with parameters  $m \geq n \geq 1$  is a point process on  $\mathbb{R}_+$  whose ordered points  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  have joint density proportional to*

$$\prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n \lambda_i^{\frac{\beta}{2}(m+1-n)-1} e^{-\frac{\beta}{2}\lambda_i}. \tag{6}$$

For  $\beta = 1$  (our one primary case of interest), this is the joint density of eigenvalues of  $X^T X$  where  $X$  is an  $m \times n$  matrix of i.i.d. standard (real) Gaussians. In particular, for the case  $m = n + 1$ , the polynomial term in the density vanishes and the corresponding ensemble will be denoted by  $\text{LOE}_n$ , called the Laguerre Orthogonal Ensemble with parameter  $n$ . For general  $\beta$ , we shall use Tridiagonal random matrix models for  $\beta$ -ensembles that were introduced in the seminal work [10] (see Section 3 for more details).

It is also well-known that the largest eigenvalues of the Hermite and Laguerre  $\beta$ -ensembles are in the same universality class, i.e., both of them (when scaled in a way such that the largest eigenvalue grows linearly in  $n$ ) have fluctuations of the order  $n^{1/3}$ , and after centering and scaling converge to a scalar multiple of  $\text{TW}_\beta$ , the  $\beta$  Tracy-Widom distribution [21]. In particular for  $\beta = 1$  and  $\beta = 2$ , these are the standard GOE and GUE Tracy-Widom distributions respectively.

There are a number of examples of remarkable couplings which furnish distributional equalities of a process of certain eigenvalue statistics (usually across increasing system size) in random matrix models with statistics occurring in other stochastic processes of interest. It is therefore a natural question to study the joint distribution of such processes, including investigating joint convergence and understanding correlation structure.

We mention two concrete couplings of relevance to this paper. The first concerns the GUE minor process, where one starts with an infinite dimensional GUE matrix and considers  $M_n$ , the copy of a  $\text{GUE}_n$  realized as its principal minor of size  $n$ . The process  $\{Y_n\}_{n \geq 0}$  of largest eigenvalues of  $\{M_n\}_{n \geq 0}$  was considered by [20]. Answering a question of Kalai [13], in this case, they showed that

$$\limsup_{n \rightarrow \infty} \frac{\tilde{Y}_n}{(\log n)^{2/3}} = \frac{1}{4^{2/3}} \tag{7}$$

where  $\tilde{Y}_n = (Y_n - \sqrt{2n})\sqrt{2n}^{1/6}$  is the centered and scaled version of  $Y_n$  that converges to the GUE Tracy-Widom distribution. For the lower deviations, they have a weaker result showing

$$-c_1 \leq \liminf_{n \rightarrow \infty} \frac{\tilde{Y}_n}{(\log n)^{1/3}} \leq -c_2 \quad (8)$$

almost surely for some absolute positive constants  $c_1, c_2$ .

For LUE, a geometric coupling is offered via the LPP representation as discussed in (1), and Ledoux [14] observed a contrasting behaviour where the lim sup (and conjecturally also, lim inf) scaled as a power of  $\log \log n$  like the classical law of iterated logarithm for the simple random walk and unlike the polylog behaviour for GUE alluded to above. The difference in the two models lies in the rate of decay of the correlation functions. In the minor process,  $\text{Corr}(Y_n, Y_{n+k})$  starts decaying when  $k = \Theta(n^{2/3})$ , whereas in the LPP coupling,  $\text{Corr}(T_n, T_{n+k})$  starts decaying only when  $k = \Theta(n)$ . However for the GUE, there is another coupling of  $Y_n$  via Brownian LPP, [6, 19], an LPP model where the underlying noise is defined by a system of two sided Brownian motion. Since the correlation structure under such a coupling is expected to behave as in the Exponential LPP model considered by [14] and in this article, one can speculate that the law of iterated logarithm replaces the fractional logarithm behaviour under such a coupling. However, although the techniques of this paper are probably enough to establish such claims, we shall not pursue them in this paper and focus only on proving Theorem 1.

**Moderate deviation estimates in  $\beta$ -ensembles:** Key aspects of the proof in [14] rely on moderate deviation inequalities, the latter topic for  $\beta$ -ensembles being a subject of independent interest. It is known [21] that the upper tail of  $\beta$  Tracy-Widom distribution decays as  $e^{-cx^{3/2}(1+o(1))}$  for  $c = c(\beta) = \frac{2}{3}\beta$  whereas the lower tail decays as  $e^{-c'|x|^3(1+o(1))}$  for  $c' = \frac{\beta}{24}$ . Hence one might expect similar tail decays for the largest eigenvalues of Hermite and Laguerre ensembles. This was considered, for the case  $\beta \geq 1$  in the important work [15] by Ledoux and Rider. For clarity of exposition as well to maintain context, let us only describe in detail their results in the Laguerre case, although similar (and in some cases stronger, see below) results were proved for the Hermite case as well. Let, as before,  $\lambda_n$  (we shall always suppress the dependence on  $m$  and  $\beta$  to reduce notational overhead) denote the largest eigenvalue of  $\text{LE}_{m,n}^\beta$ . It was showed in [15, Theorem 2] that for all  $\beta \geq 1$ ,  $0 < \varepsilon \leq 1$  and  $m \geq n$  we have for some absolute constants  $C, c > 0$

$$\mathbb{P}(\lambda_n \geq (\sqrt{m} + \sqrt{n})^2(1 + \varepsilon)) \leq C e^{-c\beta\varepsilon^{3/2}(mn)^{1/2}(\frac{1}{\varepsilon} \wedge (\frac{m}{n})^{1/4})} \quad (9)$$

$$\mathbb{P}(\lambda_n \leq (\sqrt{m} + \sqrt{n})^2(1 - \varepsilon)) \leq C^\beta e^{-c\beta\varepsilon^3 mn(\frac{1}{\varepsilon} \wedge (\frac{m}{n})^{1/2})} \quad (10)$$

In particular, observe that, when  $m = n$ , and  $\varepsilon \approx n^{-2/3}$ , this gives the optimal exponents as predicted from the tails of the Tracy-Widom distribution.

For the Hermite case, [15] also proved lower bounds of the the deviation probabilities with matching exponents, see [15, Theorem 4] for the precise statement. It was remarked there that the lower bound for the upper tail deviation probability in the Laguerre case can be proved using their methods but the lower bound for the lower tail would require a different argument. Our second main result in this article is to complement the results of [15], by providing the corresponding lower bound with matching exponents for the lower tail probability in the Laguerre case.

**Theorem 2.** *There exists absolute constants  $C_0, c, c', c'' > 0$  such that for any  $0 < \varepsilon < c'$ , and for all integers  $m \geq n \geq 1$  and  $\beta \geq 1$ , we have*

$$\mathbb{P}\{\lambda_n \leq (\sqrt{m} + \sqrt{n})^2(1 - \varepsilon)\} \geq \begin{cases} \exp\{-c\beta(\varepsilon\sqrt{mn})^2\} & \text{if } \varepsilon \geq c'' \frac{\sqrt{n}}{\sqrt{m}}, \\ (C_0)^\beta \cdot \exp\{-c\beta(\varepsilon^{\frac{3}{2}} m^{\frac{3}{4}} n^{\frac{1}{4}})^2\} & \text{if } 0 < \varepsilon \leq c'' \frac{\sqrt{n}}{\sqrt{m}}. \end{cases} \quad (11)$$

In the square case, (by taking  $c' = c''$ ) we have the simpler looking

$$\mathbb{P}\{\lambda_n \leq 4n(1 - \varepsilon)\} \geq (C_0)^\beta \cdot \exp\{-c\beta(n\varepsilon^{\frac{3}{2}})^2\}. \tag{12}$$

**Remark 1.5.** Observe that the exponents in Theorem 2 are optimal as they match the corresponding upper bounds in (10). Typically  $\lambda_n$  fluctuates on the scale  $\sigma_{m,n} = n^{-1/6}m^{1/2}$ , and Theorem 2 (together with (10)) shows that if  $\frac{m}{n}$  is bounded, then  $\mathbb{P}(\lambda_n \leq (\sqrt{m} + \sqrt{n})^2 - x\sigma_{m,n})$  decays like  $e^{-cx^3}$  for  $x$  large, as expected. It is worthwhile to notice the following interesting transition in the regime  $\frac{m}{n} \rightarrow \infty$ . If  $n$  is bounded one can observe that  $\mathbb{P}(\lambda_n \leq (\sqrt{m} + \sqrt{n})^2 - x\sigma_{m,n})$  decays as  $e^{-cx^2}$  as  $x$  large. On the other hand if  $n \rightarrow \infty$  then for each large but fixed  $x$ ,  $\mathbb{P}(\lambda_n \leq (\sqrt{m} + \sqrt{n})^2 - x\sigma_{m,n})$  decays as  $e^{-cx^3}$ . This transition from Gaussian to Tracy-Widom tail behaviour is not surprising and is understood at the level of Wishart matrices ( $\beta = 1, 2$ ). For  $\beta = 2$ , there is also an interpretation in terms of the fluctuation of last passage times across a thin rectangle in exponential LPP, which, via a coupling (or an invariance principle) can also be extended to more general LPP models [7, 5, 24].

The proof of Theorem 2, at a high level, follows the general program of [15]. To obtain the upper bounds for tails in Laguerre and Hermite  $\beta$ -ensembles [15] used the the bi-diagonal and tri-diagonal models respectively for these ensembles. For the Hermite case, the proof of lower bound of the left tail in [15] relied on the independence in the tridiagonal model and Gaussianity of the diagonal entries in a crucial way, which unfortunately is not available for the bi-diagonal models, and hence could not be extended to the Laguerre case, as pointed out in [15]. We circumvent this issue by using the idea of linearisation, elaborated in Section 3, which lets us write  $\lambda_n = s_n^2$  where  $s_n$  is the largest eigenvalue of a certain  $2n \times 2n$  tridiagonal matrix with independent entries. Note that Theorem 2 covers  $\varepsilon$  only up to some small constant. When  $\frac{n}{m}$  is bounded away from 0, one can prove similar tail bounds for  $c' \leq \varepsilon < 1$ , by a much simpler argument presented in Section 3.4.

**1.2. Organization of the article.** The rest of this paper is organised as follows. In Section 2 we provide a proof of Proposition 1.3, prove Theorem 1.2 using Theorem 2, and complete the proof of Theorem 1 using Theorem 1.2. In Section 3, we provide the proof of Theorem 2.

**Acknowledgements.** The authors thank Ofer Zeitouni for bringing the law of iterated logarithm question to their attention and Ivan Corwin for pointing out the LOE connection. RB is partially supported by a Ramanujan Fellowship (SB/S2/RJN-097/2017) from the Government of India and an ICTS-Simons Junior Faculty Fellowship. SG is partially supported by a Sloan Research Fellowship in Mathematics and NSF Award DMS-1855688. MH is supported by a summer grant of the UC Berkeley Mathematics department. MK is partially supported by UGC Centre for Advanced Study and the SERB-MATRICES grant MTR2017/000292.

## 2. THE LAW OF ITERATED LOGARITHM: PROOF OF THEOREM 1

We start by proving Proposition 1.3. As explained before, this result is implicitly known, following works of Baik and Rains in early 2000s, but we could not find a precise reference in the literature and hence for completeness provide a short proof using the recent works [18, 11], borrowing their notations where convenient.

*Proof of Proposition 1.3.* Theorem 1.2 of [18] says that

$$\lambda_{2n-1} \stackrel{d}{=} 4 \left( \sup_{t \in [0,1]} B_{2n-1}(t) \right)^2,$$

where  $B_1 < \dots < B_{2n-1}$  is a collection of  $2n - 1$  non-intersecting Brownian bridges on  $[0, 1]$  and  $\lambda_1 < \dots < \lambda_{2n-1}$  are the eigenvalues of a LOE matrix  $X^T X$ , where  $X$  is a  $2n \times (2n - 1)$  matrix with i.i.d.  $N(0, 1)$  random variables.

Now the calculation immediately preceding equation (5) in [11] shows that

$$\left( \sup_{t \in [0,1]} B_{2n-1}(t) \right)^2 \stackrel{d}{=} \sup_{t \geq 0} \lambda_{\max}(H(t) - tI),$$

where  $H(t)$  is a  $(2n-1) \times (2n-1)$  Hermitian Brownian motion, i.e. a  $(2n-1) \times (2n-1)$  Hermitian matrix with i.i.d. standard complex Brownian motions below the diagonal and i.i.d. standard real Brownian motions along the diagonal. Finally, Theorem 1 of [11] says that

$$\sup_{t \geq 0} \lambda_{\max}(H(t) - tI) \stackrel{d}{=} \max_{\pi \in \Pi_n^{\text{flat}}} \sum_{(i,j) \in \pi} \xi'_{ij},$$

where  $\xi'_{ij}$  are i.i.d. rate 2 exponential random variables, and  $\Pi_n^{\text{flat}}$  is the collection of up-right paths from  $(1, 1)$  to the line  $i + j = 2n$ . Combining these and using the scaling property of exponentials then yields that

$$T_n^* = \max_{\pi \in \Pi_n^{\text{flat}}} \sum_{(i,j) \in \pi} \xi_{ij} \stackrel{d}{=} \frac{1}{2} \lambda_{2n-1},$$

where  $\xi_{ij}$  are i.i.d. rate 1 exponential random variables and the first equality is by definition (5).  $\square$

We next prove Theorem 1.2, which is an almost immediate consequence of Proposition 1.3 and Theorem 2.

*Proof of Theorem 1.2.* Using Proposition 1.3 and setting the parameters  $(m, n) = (2n, 2n-1)$ ,  $\beta = 1$ , and  $\varepsilon = \frac{1}{4}xn^{-2/3}$  it follows that

$$\mathbb{P}\left(T_n^* \leq 4n - xn^{1/3}\right) = \mathbb{P}\left(\lambda_{2n-1} \leq 8n - 2xn^{1/3}\right) \geq \mathbb{P}\left(\lambda_{2n-1} \leq (\sqrt{2n} + \sqrt{2n-1})^2(1 - \varepsilon)\right)$$

and the proof is completed by invoking Theorem 2.  $\square$

Note that Theorem 1 states that  $\liminf \frac{Z_n}{(\log \log n)^{1/3}}$  is almost surely a constant. This is proved in the next lemma, which then reduces proving Theorem 1 to showing that it is bounded below from 0 with positive probability.

**Lemma 2.1.**  $\liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}}$  is a constant, almost surely.

*Proof.* This is a straightforward consequence of the Kolmogorov 0-1 law since the random variable in question is a tail random variable. To see this, fix  $r$  and let  $\mathbb{L}_r$  be the line  $x + y = 2r$ . Then, since all the variables  $\xi_v$  are non-negative, by definition, for  $n \geq r$ ,

$$\sup_{v \in \mathbb{L}_r} T_{v, \mathbf{n}} \leq T_n \leq \sup_{v \in \mathbb{L}_r} T_{v, \mathbf{n}} + \sup_{v \in \mathbb{L}_r} T_{0, v}.$$

Now clearly  $\sup_{v \in \mathbb{L}_r} T_{v, \mathbf{n}}$  is a function of only the independent field including and above the line  $\mathbb{L}_r$ , while

$$\liminf_{n \rightarrow \infty} \frac{\sup_{v \in \mathbb{L}_r} T_{0, v} - 4r}{n^{1/3}(\log \log n)^{1/3}} = 0,$$

as  $\sup_{v \in \mathbb{L}_r} T_{0, v}$  is finite a.s. Thus, for every  $r$ ,

$$\liminf_{n \rightarrow \infty} \frac{T_n - 4n}{n^{1/3}(\log \log n)^{1/3}} = \liminf_{n \rightarrow \infty} \frac{\sup_{v \in \mathbb{L}_r} T_{v, \mathbf{n}} - 4(n-r)}{n^{1/3}(\log \log n)^{1/3}},$$

and hence it is a tail random variable.  $\square$

Observe that the same argument would also show that  $\limsup_{n \rightarrow \infty} Z_n / (\log \log n)^{2/3}$  is constant almost surely.

We are now ready to prove the following intermediate result, which in conjunction with Lemma 2.1 finishes the proof of Theorem 1.

**Proposition 2.2.** *For  $c_0$  as in Theorem 1.2, there exists  $\delta > 0$  such that*

$$\mathbb{P} \left( \liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}} \leq -(2c_0)^{-1/3} \right) \geq \delta.$$

The choice of parameters in the proof can be slightly tweaked allowing us to replace  $(2c_0)^{-1/3}$  by  $c_0^{-1/3}$ , as pointed out in Remark 2.3.

*Proof.* Define  $n_j = 2^j$  and fix  $k > 1$  even and recall that for any  $r \in \mathbb{Z}_+$ ,  $\mathbb{L}_r$  denotes the line  $x + y = 2r$ . Our objective is to establish that for  $\tilde{c} = (2c_0)^{-1/3} > 0$ , as in the statement of the proposition, and for every  $k$  sufficiently large, with probability  $\delta > 0$  there exists a  $j$  such that  $k/2 \leq j \leq k$  and

$$T_{n_j} \leq 4n_j - \tilde{c}n_j^{1/3}(\log \log n_j)^{1/3}.$$

Clearly this will suffice.

To achieve this, we will divide up the region from  $(1, 1)$  to  $(n_k, n_k)$  in dyadic scales (so the  $j$ th region is between  $\mathbb{L}_{n_{j-1}}$  and  $\mathbb{L}_{n_j}$ ), and proceed by examining these regions sequentially, starting from the  $k^{\text{th}}$  region and decreasing the index, till we find the first  $j$  such that  $T_{\mathbf{n}_j, \mathbb{L}_{n_{j-1}}}$  (weight of the largest weight path from the line  $\mathbb{L}_{n_{j-1}}$  to  $(n_j, n_j)$ ) is sufficiently low. Suppose this region is the one between  $\mathbb{L}_{n_{j-1}}$  and  $\mathbb{L}_{n_j}$ . Since, we know that the regions are disjoint and hence independent, with high probability the point-to-line weight from  $(1, 1)$  to  $\mathbb{L}_{n_{j-1}}$  does not have too high a weight and hence on the intersection of the above events, one can conclude that  $T_{n_j}$  is low, finishing the proof (see Figure 1 for an illustration). The rigorous argument will require some minor tweaks to the above high level description to ensure the necessary independence.

Moving now to make the above argument precise, let us, for notational convenience, denote by  $T_{(j)}^*$ , the weight of the line-to-point polymer from the anti-diagonal line  $\mathbb{L}_{n_{j-1}}$  to the point  $(n_j - 1, n_j - 1)$ . We note that, since  $2n_{j-1} = n_j$ , by the symmetry of the random environment we get

$$T_{(j)}^* \stackrel{d}{=} T_{n_{j-1}}^*,$$

and that  $T_{(j)}^*$  are independent as  $j$  varies. Also throughout the proof,  $c_0$  is as in Theorem 1.2.

Let us fix  $\varepsilon > 0$  sufficiently small, and set  $C_\varepsilon := (1 - \varepsilon)^{1/3} c_0^{-1/3}$ . We need to define a number of events. For  $j \in \mathbb{N}$  let us define

$$A_j := \left\{ T_{(j)}^* \leq 4n_{j-1} - C_\varepsilon n_{j-1}^{1/3} (\log \log n_{j-1})^{1/3} \right\},$$

and for a fixed  $k$  sufficiently large let us set  $A := \bigcup_{j=k/2}^k A_j$ . Now define  $\tau = \max\{j \leq k : A_j \text{ occurs}\}$  so that on  $A$ ,  $\tau > k/2$ . For  $v = (x, y) \in \mathbb{Z}^2$  we shall denote the random variable  $\xi_v$  by  $\xi_{x,y}$  for notational convenience. Define the  $\sigma$ -algebras

$$\mathcal{F}_j := \sigma(\{\xi_{x,y} : x + y < 2n_{j-1}\}) \quad \text{and} \quad \mathcal{F}_j^c := \sigma(\{\xi_{x,y} : x + y \geq 2n_{j-1}\}).$$

Notice that the event  $\{\tau = j\}$  is  $\mathcal{F}_j^c$ -measurable. Define the events

$$B_j = \left\{ \tilde{T}_{n_{j-1}}^* < 4n_{j-1} \right\} \quad \text{and} \quad B_\tau = \left\{ \tilde{T}_{n_{\tau-1}}^* < 4n_{\tau-1} \right\},$$

where  $\tilde{T}_n^*$  is the weight of the best path from  $(1, 1)$  to the line  $x + y = 2n$  with the weight of the final vertex excluded, i.e.,

$$\tilde{T}_n^* = \sup_{v \in \mathbb{L}_n} (T_{1,v} - \xi_v).$$

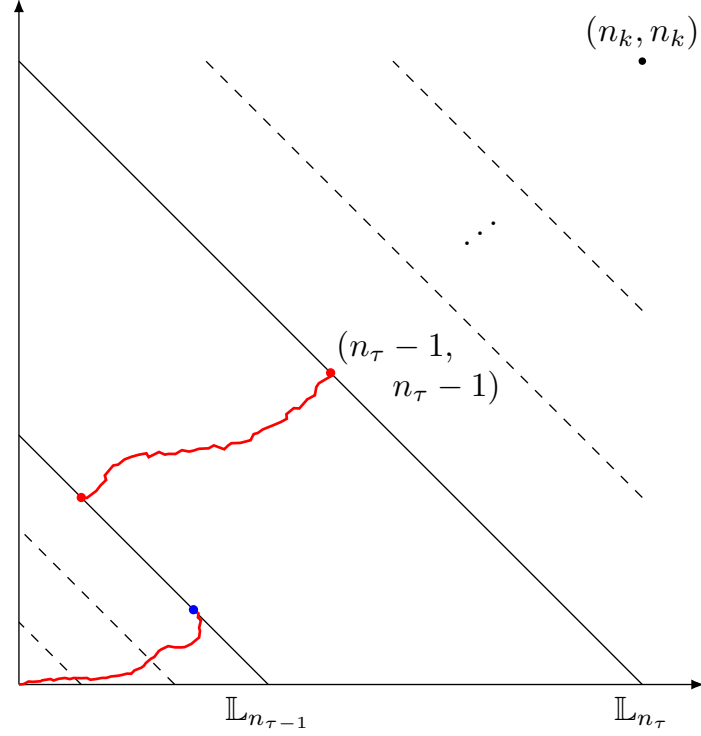


FIGURE 1. The argument of Proposition 2.2. The dotted lines to the right of  $(n_\tau - 1, n_\tau - 1)$  are the boundaries of regions where the event  $A_j$  did not occur, i.e. the line-to-point weight was not sufficiently low. The region bounded by black lines is the first region, when traveling southwest, that  $A_j$  occurs, and the corresponding line-to-point polymer is marked in red. Since the red point-to-line polymer from  $(1, 1)$  to  $\mathbb{L}_{n_{\tau-1}}$  has weight less than  $4n_{\tau-1}$ , the weight of the point-to-point polymer from  $(1, 1)$  to  $(n_\tau - 1, n_\tau - 1)$  must have weight less than  $4n_\tau - (1 - \varepsilon)^{1/3} c_0^{-1/3} n_{\tau-1}^{1/3} (\log \log n_{\tau-1})^{1/3}$ . The red points are endpoints of polymers whose weight is included in the weight of the polymer, while the blue point is the endpoint whose weight is not included.

We want to show that  $A \cap B_\tau$  occurs with uniformly positive probability. To this end, note that using Theorem 1.2, we have  $\mathbb{P}\left(T_{(j)}^* \leq 4n_{j-1} - xn_{j-1}^{1/3}\right) \geq e^{-c_0 x^3}$  for  $x \in (1, n^{2/3})$  and setting  $x = C_\varepsilon (\log \log n_{j-1})^{1/3}$  we get

$$\mathbb{P}(A_j) \geq e^{-(1-\varepsilon) \log \log n_{j-1}} \geq \frac{1}{j^{1-\varepsilon}},$$

since  $n_j = 2^j$  and  $\log 2 < 1$ . Then, using independence of the  $A_j$ 's we have for any  $\delta > 0$ ,

$$\mathbb{P}(A) \geq 1 - \left(1 - \frac{1}{k^{1-\varepsilon}}\right)^{k/2} \geq 1 - \delta \quad (13)$$

for  $k > k_0(\varepsilon, \delta)$ .

Note that  $T_n^*$  is  $\mathcal{F}_j$  measurable. Hence we have

$$\mathbb{P}(B_\tau \mid \tau = j) = \mathbb{P}(B_j), \quad (14)$$

since  $B_j \in \mathcal{F}_j$  and  $\{\tau = j\} \in \mathcal{F}_j^c$ , which are independent  $\sigma$ -algebras (it is for this independence that we removed the weight of the last vertex in the definition of  $\tilde{T}_n^*$ ). Since  $\tilde{T}_{n_j}^* \leq T_{n_j}^*$  for every



$j$ , from Theorem 1.2 there exists  $\delta > 0$  such that  $\mathbb{P}(B_j) \geq 2\delta$  for all  $k/2 \leq j \leq k$  and all  $k$  large enough<sup>1</sup>, and so (13) and (14) together imply that,  $\mathbb{P}(A \cap B_\tau) \geq \delta$ , for large enough  $k$ .

Now on  $A \cap B_\tau$ , we have  $k/2 \leq \tau \leq k$  and

$$\begin{aligned} T_{n_{\tau-1}} &\leq \tilde{T}_{n_{\tau-1}}^* + T_{(\tau)}^* < 4n_{\tau-1} + 4(n_\tau - n_{\tau-1}) - C_\varepsilon n_{\tau-1}^{1/3} (\log \log n_{\tau-1})^{1/3} \\ &= 4n_\tau - C_\varepsilon n_{\tau-1}^{1/3} (\log \log n_{\tau-1})^{1/3}. \end{aligned}$$

Since  $n_{\tau-1} = n_\tau/2$ , replacing  $\varepsilon$  by  $2\varepsilon$ , for all large enough  $k$ , we get,

$$\mathbb{P} \left( \bigcup_{j=k/2}^k \left\{ \frac{T_{n_j-1} - 4(n_j - 1)}{(n_j - 1)^{1/3} (\log \log (n_j - 1))^{1/3}} < -2^{-1/3} C_{2\varepsilon} \right\} \right) \geq \mathbb{P}(A \cap B_\tau) \geq \delta, \quad (15)$$

This shows that with probability at least  $\delta$ ,

$$\liminf_{n \rightarrow \infty} \frac{T_n - 4n}{n^{1/3} (\log \log n)^{1/3}} < -2^{-1/3} C_{2\varepsilon} = (1 - 2\varepsilon)^{1/3} (2c_0)^{-1/3}.$$

Indeed, we have

$$\begin{aligned} \mathbb{P} \left( \liminf_{n \rightarrow \infty} \frac{T_n - 4n}{n^{1/3} (\log \log n)^{1/3}} < -2^{-1/3} C_{2\varepsilon} \right) &= \mathbb{P} \left( \bigcap_{k=0}^{\infty} \bigcup_{j=k}^{\infty} \left\{ \frac{T_j - 4j}{j^{1/3} (\log \log j)^{1/3}} < -2^{-1/3} C_\varepsilon \right\} \right) \\ &= \lim_{k \rightarrow \infty} \mathbb{P} \left( \bigcup_{j=k}^{\infty} \left\{ \frac{T_j - 4j}{j^{1/3} (\log \log j)^{1/3}} < -2^{-1/3} C_{2\varepsilon} \right\} \right) \\ &\geq \lim_{k \rightarrow \infty} \mathbb{P} \left( \bigcup_{j=2^{k/2-1}}^{2^k-1} \left\{ \frac{T_j - 4j}{j^{1/3} (\log \log j)^{1/3}} < -2^{-1/3} C_{2\varepsilon} \right\} \right) \\ &\geq \delta. \end{aligned}$$

where the last inequality follows from (15). As this is true for every  $\varepsilon > 0$ , sending  $\varepsilon$  to 0 we get that

$$\mathbb{P} \left( \liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}} \leq -(2c_0)^{-1/3} \right) = \mathbb{P} \left( \liminf_{n \rightarrow \infty} \frac{T_n - 4n}{n^{1/3} (\log \log n)^{1/3}} \leq -(2c_0)^{-1/3} \right) \geq \delta,$$

completing the proof.  $\square$

**Remark 2.3.** *If instead of a dyadic breakup with  $n_j = 2^j$  as in the proof, choosing  $n_j = \lceil (1+\eta)^j \rceil$  for some  $\eta > 0$ , allows us to replace  $(2c_0)^{-1/3}$  in the statement of the proposition by  $((1+\eta^{-1})c_0)^{-1/3}$ , which, on taking  $\eta \rightarrow \infty$ , converges to  $c_0^{-1/3}$ .*

**2.1. Sharpness and possible extensions.** We wrap up this section with a discussion on the sharpness of our argument and several possible extensions. The first natural question to determine the limiting value of  $\liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}}$ . It is widely believed that for  $T_n$ , a stronger moderate deviation estimate than is given by the results of [15] holds. As is well known, it was shown in [12] that  $2^{-4/3} n^{-1/3} (T_n - 4n)$  converges weakly to the GUE Tracy-Widom distribution ( $\beta = 2$ ). Comparing with the tails of Tracy-Widom distribution from [21], one can make predictions about the optimal constants for the tail estimates in (9) and (10) which are indeed conjectured to be true. In particular, it is believed that

$$\log \mathbb{P}(T_n \geq 4n + xn^{1/3}) = -\frac{4}{3} (2^{-4/3} x)^{3/2} + O(x^2 n^{-1/3}) + O(\log x) \quad (16)$$

<sup>1</sup>This in fact is just a straightforward consequence of the weak convergence result alluded to right after (5).

for all large  $x \leq \delta n^{2/3}$  and for all  $n$  sufficiently large. Similarly, for the lower tail, it is believed that

$$\log \mathbb{P}(T_n \leq 4n - xn^{1/3}) = -\frac{1}{12}(2^{-4/3}x)^3 + O(x^4 n^{-4/3}) + O(\log x) \quad (17)$$

for all large  $x \leq \delta n^{2/3}$  and for all  $n$  sufficiently large. Such sharp results have indeed been proved in cases of Poissonian and Geometric last passage percolation using the Riemann-Hilbert approach in [16, 17, 2, 9]. Under the stronger hypothesis (16), Ledoux in [14] showed that

$$\limsup_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{2/3}} = 3^{2/3}, \text{ almost surely.}$$

He also conjectured based on the believed bound (17) that

$$\liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}} = -192^{1/3}. \quad (18)$$

(The statistic considered in [14] is  $2^{-4/3}n^{-1/3}(T_n - 4n)$  and so the numerical values there differ from the ones above by a factor of  $2^{4/3}$ .) Furthermore, the Borel-Cantelli lemma based argument in [14] in conjunction with the conjectured bound (17), does yield a lower bound of  $-192^{1/3}$  for the LHS in (18).

On the other hand in the point-to-line case, as indicated before, it is known (see [4, 23, 8]) that  $2^{-2/3}n^{-1/3}(T_n^* - 4n)$  converges weakly to GOE Tracy-Widom distribution ( $\beta = 1$ ). In analogy with (17), comparing with the left tail of GOE Tracy-Widom distribution, the optimal estimate in this case is predicted to be

$$\mathbb{P}(T_n^* \leq 4n - xn^{1/3}) = -\frac{x^3}{96}(1 + o(1)), \quad (19)$$

for  $x \ll n^{2/3}$ . Even though we are unaware of such a sharp estimate in the literature, (19), along with our arguments will indeed imply that

$$\liminf_{n \rightarrow \infty} \frac{Z_n}{(\log \log n)^{1/3}} \leq -96^{1/3}$$

almost surely (see Remark 2.3), which is still far from the conjectured value in [14], indicating, not surprisingly, that dominating the point-to-point passage times by point-to-line counterparts incurs a loss in the constant.

Going beyond Exponential LPP, Ledoux points out in [14] that his results hold also for LPP on  $\mathbb{Z}^2$  with geometrically distributed weights. This is because, the upper bounds for the moderate deviation probabilities (for both the left and right tails), i.e., analogues of (9), (10) are available also for the geometric case (see e.g. [2, 9]) which are the only inputs needed for the argument of [14]. On the other hand, for our argument, we rely on the lower tail bounds for the point-to-line last passage times and while there does exist an explicit distributional formula for the latter for the geometric case as well (see [3]), the random matrix connection, as far as we understand, exists only in the Laguerre limit. While it is possible that using the formula of [3] one can obtain a result analogous to Theorem 1.2 for geometric LPP, we are unaware of any such result, rendering our current arguments inapplicable in the geometric case.

Finally, passage times in more general non-axial directions other than along the diagonal were also considered in [14]. That is, for any fixed  $\gamma \in (0, \infty)$ , a similar law of iterated logarithm was proved for the (properly centered) sequence  $T_n^\gamma := T_{\mathbf{1}, (n, \lfloor \gamma n \rfloor)}$ . Since our proof relies on point-to-line estimates, it is not hard to see that the same proof verbatim also yields Theorem 1 in this more general case. However we do not attempt to provide any details.

All that is left to be done now is prove Theorem 2 which is accomplished in the following section.

3. LOWER DEVIATIONS IN  $\beta$ -LAGUERRE ENSEMBLE: PROOF OF THEOREM 2

As mentioned earlier, for the proof of Theorem 2, we shall rely on a tridiagonal matrix model for the  $\beta$ -Laguerre ensemble [10]. To define the tridiagonal matrix, we start by introducing some notation. We write  $\chi_r^2$  for the Chi-square distribution with parameter  $r$ , and by an abuse of notation, also for a random variable having this distribution. Its density is proportional to  $x^{r/2-1}e^{-x/2}$  on  $\mathbb{R}_+$  and it has expectation  $r$ . Similarly, we write  $\chi_r$  for the random variable (and the distribution) which is the positive square root of a  $\chi_r^2$  variable. It has density proportional to  $x^{r-1}e^{-x^2/2}$  and its expectation is equal to

$$2^{1/2} \frac{\Gamma(r/2 + 1/2)}{\Gamma(r/2)}. \tag{20}$$

We also recall the well-known facts (see e.g. [15] for a reference) that  $\mathbb{E}\chi_r$  is increasing in  $r$  for all  $r > 0$ ,  $\mathbb{E}\chi_r \leq r^{1/2}$  (Jensen's inequality) and  $\mathbb{E}\chi_r \geq \sqrt{r - 1/2}$  for all  $r \geq 1$ .

Now fixing  $m \geq n$  and  $\beta \geq 1$ , let

$$X_{2k-1} = a_k, \text{ for } 1 \leq k \leq n, \text{ and } X_{2k} = b_k, \text{ for } 1 \leq k \leq n - 1 \tag{21}$$

be independent random variables where  $\beta a_k^2 \sim \chi_{\beta(m+1-k)}^2$  and  $\beta b_k^2 \sim \chi_{\beta(n-k)}^2$ .

Given the above, we define the following matrices.

- (1) Let  $B = B_\beta$  be a  $n \times n$  bi-diagonal matrix with  $B_{k,k} = a_k$  and  $B_{k+1,k} = b_k$ .
- (2) Let  $L = L_\beta = BB^T$ , an  $n \times n$  positive semi-definite matrix, where  $B^T$  as usual denotes the transpose of  $B$ .
- (3) Let  $M = \begin{bmatrix} 0 & B^T \\ B & 0 \end{bmatrix}$ , a  $2n \times 2n$  symmetric matrix.
- (4) Let  $T$  be a  $2n \times 2n$  symmetric tridiagonal matrix with zeros on the diagonal and  $X = (X_1, \dots, X_{2n-1})$  on the super-diagonal and sub-diagonal, i.e. for each  $i = 1, 2, \dots, 2n - 1$ ,

$$T_{i,i+1} = T_{i+1,i} = X_i. \tag{22}$$

**Fact:** The joint density of eigenvalues of  $L$  is given by (6), the  $\beta$ -Laguerre ensemble,  $LE_{m,n}^\beta$ . This is by now well known (see for example [10, Theorem 3.1]). We shall, however, not be relying on this joint density.

We next state and prove a simple lemma relating the eigenvalues of  $L$  and  $M$  and  $T$ .

**Lemma 3.1.** *If  $L$  has eigenvalues  $s_1^2 \leq \dots \leq s_n^2$  (since  $L$  is positive semi-definite), then  $M$  and  $T$  have eigenvalues  $\pm s_1, \pm s_2, \dots, \pm s_n$ .*

*Proof.* The characteristic polynomial of  $M$  is  $\det(zI_n) \det(zI_n - \frac{1}{z}BB^T) = \det(z^2I_n - L)$ . This shows that  $M$  has eigenvalues  $\pm s_k$ . One can check easily that if we permute the rows and columns of  $M$  in the order  $n + 1, 1, n + 2, 2, n + 3, 3, \dots$ , then we get the matrix  $T$ . Thus  $T$  has the same eigenvalues as  $M$ .  $\square$

Now note that using Lemma 3.1, our Theorem 2 reduces to proving the following result. For some  $c, c', c'' > 0$  and any  $0 < \varepsilon < c'$  and all  $\beta \geq 1$ ,

$$\mathbb{P}\{s_n \leq (\sqrt{m} + \sqrt{n})(1 - \varepsilon)\} \geq \begin{cases} \exp\{-c\beta(\varepsilon\sqrt{mn})^2\} & \text{if } \varepsilon \geq c'' \frac{\sqrt{n}}{\sqrt{m}}, \\ C_0^\beta \cdot \exp\{-c\beta(\varepsilon^{\frac{3}{2}} m^{\frac{3}{4}} n^{\frac{1}{4}})^2\} & \text{if } 0 < \varepsilon \leq c'' \frac{\sqrt{n}}{\sqrt{m}}. \end{cases} \tag{23}$$

Note that Theorem 2 is a statement about  $s_n^2$ , while (23) concerns  $s_n$ . There is no issue in making this change, except that  $\varepsilon$  changes by a factor of 2; this is safely absorbed in the constant  $c$ .

As mentioned earlier, the lower bound for the lower tail in the Laguerre case was not addressed in [15]. The main reason why we are able to analyze it is that we do not use  $BB^T$  (in which the entries are not independent and are sums of products of  $\chi$  random variables with different parameters) but the matrix  $T$  which has independent entries. Further, the matrix  $T$  is very similar in appearance to the tridiagonal matrix for the Hermite model. However the proof in [15] for the Hermite model uses in an essential way the Gaussians on the diagonal, while  $T$  has zeros on the diagonal. Hence some modification is needed in the proof of the lower bound. This idea of *linearization* is often useful when working with the Laguerre ensembles, and has been used before (see for example, the appendix to [25]).

As the following proof is rather technical, before delving into it we provide a brief high level overview. The most natural idea would be to condition  $X_{k,s}$  to be small, and indeed, that together with a simple approximation for the largest eigenvalue works for  $\varepsilon$  bounded away from 0 (see Section 3.4). To treat all values of  $\varepsilon$  down to the fluctuation scale, one needs to estimate the eigenvalue more accurately.

We achieve this by an appropriate tilting argument. We do a change of measure changing  $X_{k,s}$  to  $Y_{k,s}$  (defined in (32)), so that for the tridiagonal matrix obtained in (22) by replacing the  $X_{k,s}$  by the  $Y_{k,s}$ , the largest eigenvalue is typically smaller than  $(\sqrt{m} + \sqrt{n})(1 - \varepsilon)$ .

The sought lower bound is then obtained by lower bounding the Radon-Nikodym derivative of  $X_{k,s}$  with respect to  $Y_{k,s}$ . To achieve the first step, recalling  $\lambda_n = \max_{\|w\|=1} Q(w)$ , we shall define a quadratic form  $Q_b$  (that arises naturally by approximating  $\mathbb{E}Q$  and completing squares, see below) with  $Q_b \geq Q$  and estimate  $\max_{\|w\|=1} Q_b(w)$  instead. It turns out (Lemma 3.6) there exists a change of measure from  $X_{k,s}$  to  $Y_{k,s}$  which is simply scaling a number of  $X_{k,s}$  by a factor of  $\sqrt{1 - \varepsilon}$  that gives the lower bound with the right exponents. We now start with the details.

**3.1. The quadratic forms and their comparison.** The quadratic form corresponding to  $T - (\sqrt{m} + \sqrt{n})I_{2n}$  is

$$Q(w) = 2 \sum_{k=1}^{2n-1} X_k w_k w_{k+1} - (\sqrt{m} + \sqrt{n}) \sum_{k=1}^{2n} w_k^2. \quad (24)$$

where  $w = (w_1, w_2, \dots, w_{2n})$ . Let  $\hat{X}_k = X_k - \mathbb{E}[X_k]$ , and for  $b > 0$  define the idealized quadratic forms

$$Q_b(w) = 2 \sum_{k=1}^{2n-1} \hat{X}_k w_k w_{k+1} - b\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2 - b\sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 - \frac{b}{\sqrt{n}} \sum_{k=1}^{2n} k w_k^2, \quad (25)$$

where  $w_0 = w_{2n+1} = 0$  by convention. If  $Z_k$  are centered random variables, define  $Q_b(w; Z)$  by the same expression as  $Q_b$ , except that  $\hat{X}_k$  is replaced by  $Z_k$ . In particular,  $Q_b(w) = Q_b(w, \hat{X})$ .

We briefly describe the motivation for defining  $Q_b(w)$  as above, which naturally arises from completing squares in  $\mathbb{E}Q(w)$ . Observe that we have

$$\mathbb{E}Q(w) := 2 \sum_{k=0}^n \mathbb{E}X_{2k-1} w_{2k-1} w_{2k} + 2 \sum_{k=1}^{n-1} \mathbb{E}X_{2k} w_{2k} w_{2k+1} - (\sqrt{m} + \sqrt{n}) \sum_{k=1}^{2n} w_k^2.$$

Now, using the approximation (coming from (20))  $\mathbb{E}X_{2k-1} \approx \sqrt{m-k} \approx \sqrt{m} - \frac{k}{2\sqrt{m}}$  and  $\mathbb{E}X_{2k} \approx \sqrt{n-k} \approx \sqrt{n} - \frac{k}{2\sqrt{n}}$  and the identities  $2w_k w_{k+1} = w_k^2 + w_{k+1}^2 - (w_k - w_{k+1})^2$  we get

$$\mathbb{E}Q(w) \approx -\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2 - \sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 - \left( \frac{1}{2\sqrt{n}} + \frac{1}{2\sqrt{m}} \right) \sum_{k=1}^{2n} k w_k^2.$$

It is now easy to see that (at least for  $m = n$ ),

$$Q_b(w) \approx (Q(w) - \mathbb{E}Q(w)) + b\mathbb{E}Q(w)$$

and for  $m > n$ ,  $Q_b(w)$  is even larger. As the approximant of  $\mathbb{E}Q(w)$  is negative, one could reasonably expect that for small  $b$ ,  $Q_b(w)$  would be larger than  $Q(w)$ . This is the content of the next lemma.

**Lemma 3.2.** *If  $b$  is sufficiently small ( $b < \frac{1}{4}$  suffices), then  $Q \leq Q_b$ .*

*Proof.* Observe that

$$Q_b(w) - Q(w) = \sum_{k=1}^{2n} V_{k,k} w_k^2 + 2 \sum_{k=1}^{2n-1} V_{k,k+1} w_k w_{k+1}$$

where  $V_{k,k} = \sqrt{m} + \sqrt{n} - b\sqrt{m} - b\sqrt{n} - \frac{bk}{\sqrt{n}}$  and  $V_{k,k+1} = -\mathbb{E}[X_k] + b\sqrt{m}$  for  $k$  odd and  $V_{k,k+1} = -\mathbb{E}[X_k] + b\sqrt{n}$  for  $k$  even.

That is, if we form the symmetric tridiagonal matrix  $V$  with these entries, then

$$(Q_b - Q)(w) = \langle Vw, w \rangle.$$

Thus, to show that  $Q_b \geq Q$ , it suffices to show that

$$V_{k,k} \geq |V_{k,k-1}| + |V_{k,k+1}|$$

for all  $1 \leq k \leq 2n$  (with the interpretation that  $V_{1,0} = 0$  and  $V_{2n,2n+1} = 0$ ). In fact, considering the different combinations of signs of  $V_{k,k-1}$  and  $V_{k,k+1}$ , it is sufficient if we have

$$\sqrt{m} + \sqrt{n} - b\sqrt{m} - b\sqrt{n} - \frac{bk}{\sqrt{n}} \geq \begin{cases} b\sqrt{m} + b\sqrt{n} - \mathbb{E}[X_{k-1}] - \mathbb{E}[X_k]; \\ \mathbb{E}[X_{k-1}] + \mathbb{E}[X_k] - b\sqrt{m} - b\sqrt{n}; \\ \mathbb{E}[X_{k-1}] - \mathbb{E}[X_k] - b\sqrt{m} + b\sqrt{n} & \text{if } k \text{ is even} \\ \mathbb{E}[X_{k-1}] - \mathbb{E}[X_k] - b\sqrt{n} + b\sqrt{m} & \text{if } k \text{ is odd.} \\ -\mathbb{E}[X_{k-1}] + \mathbb{E}[X_k] + b\sqrt{m} - b\sqrt{n} & \text{if } k \text{ is even} \\ -\mathbb{E}[X_{k-1}] + \mathbb{E}[X_k] + b\sqrt{n} - b\sqrt{m} & \text{if } k \text{ is odd.} \end{cases}$$

Notice that the left hand side is at least  $\sqrt{m}(1-b) + \sqrt{n}(1-3b)$ . By using the fact that the expectation of  $\chi$  variables increases with its parameter, it follows that the contribution of the expectation terms is negative in the right hand side of the first, fourth and the fifth inequality. If we ignore these negative terms, what remains is at most  $b\sqrt{m} + b\sqrt{n}$ . Therefore, all three inequalities now follow by choosing  $b \leq \frac{1}{4}$ . Using  $\mathbb{E}\chi_\alpha \leq \sqrt{\alpha}$  it also follows that the right hand sides of the third and the sixth inequalities are at most  $(1-b)\sqrt{m} + b\sqrt{n}$ , and these inequalities also follow by taking  $b \leq \frac{1}{4}$ .

It remains to prove the second inequality. For this, assume that  $k = 2\ell$  (similar reasoning works for odd  $k$ ) in which case, invoking the facts following (20) again, the right hand side is at most  $\sqrt{m+1-\ell} + \sqrt{n-\ell} - b\sqrt{m} - b\sqrt{n}$ . Since  $\sqrt{m+1-\ell} \leq \sqrt{m}$ , all we need is that

$$\frac{2\ell b}{\sqrt{n}} \leq \sqrt{n} - \sqrt{n-\ell} = \frac{\ell}{\sqrt{n} + \sqrt{n-\ell}}.$$

The right hand side is at least  $\frac{\ell}{2\sqrt{n}}$ , hence the desired inequality is valid if  $b < \frac{1}{4}$ .  $\square$

**3.2. Exponential tail bound to the right.** We now prove a deviation inequality for the upper tail of  $Q_b(\cdot, \cdot)$ , which, via Lemma 3.2 also provides a bound on the upper deviation of  $Q(\cdot, \cdot)$ . A similar result was proved in [15] for a different but related quadratic form on the way to prove an upper tail deviation inequality for the largest eigenvalue for Laguerre  $\beta$ -ensemble (see [15, Section 3.2]).

**Lemma 3.3.** *Assume that  $Z_k$  are independent random variables with zero mean and satisfying  $\mathbb{E}[e^{\lambda Z_k}] \leq e^{c\lambda^2}$  for all  $k \leq 2n$  and some  $c > 0$  and all  $\lambda \in \mathbb{R}$ . Then for any  $0 < \varepsilon < 1$ ,*

$$\mathbb{P} \left\{ \max_{\|w\|=1} Q_b(w; Z) \geq \varepsilon \sqrt{m} \right\} \leq C e^{-c' \varepsilon^{3/2} \sqrt{mn} (\frac{1}{\sqrt{\varepsilon}} \wedge (\frac{m}{n})^{1/4})} \quad (26)$$

where  $c', C > 0$  depends only on  $c$  and  $b$ .

*Proof.* Define  $S_k = Z_1 + \dots + Z_k$  for  $1 \leq k \leq 2n$  (and  $S_0 = 0$  and  $S_k = S_{2n}$  for  $k > 2n$ ). For  $p \geq 1$  define

$$\Delta_p(k) = \max\{|S_{k+j} - S_k| : 1 \leq j \leq p\}.$$

By the summation by parts formula ([15, Lemma 8]), we have for any unit vector  $w$ ,

$$\begin{aligned} \sum_{k=1}^{2n-1} Z_k w_k w_{k+1} &= \sum_{k=0}^{2n-2} \frac{1}{p} [S_{k+p} - S_k] w_{k+1} w_{k+2} + \sum_{k=0}^{2n-1} \left( \frac{1}{p} \sum_{\ell=k}^{k+p-1} [S_\ell - S_k] \right) w_{k+1} (w_{k+2} - w_k) \\ &\leq \frac{1}{2p} \sum_{k=0}^{2n-2} \Delta_p(k) (w_{k+1}^2 + w_{k+2}^2) + \frac{2}{b\sqrt{n}} \sum_{k=0}^{2n-1} \Delta_p(k)^2 w_{k+1}^2 + \frac{b\sqrt{n}}{8} \sum_{k=0}^{2n} (w_{k+2} - w_k)^2 \\ &\leq \frac{1}{2p} \sum_{k=0}^{2n-2} \Delta_p(k) (w_{k+1}^2 + w_{k+2}^2) + \frac{2}{b\sqrt{n}} \sum_{k=0}^{2n-1} \Delta_p(k)^2 w_{k+1}^2 + \frac{1}{2} b\sqrt{n} \sum_{k=0}^{2n} (w_{k+1} - w_k)^2 \end{aligned} \quad (27)$$

where for the first inequality we bound all  $S_\ell - S_k$  terms by  $\Delta_p(k)$  and then used Cauchy-Schwarz in the form

$$|\Delta_p(k) w_{k+1} (w_{k+2} - w_{k+1})| \leq \frac{1}{4\lambda} \Delta_p(k)^2 w_{k+1}^2 + \lambda (w_{k+2} - w_{k+1})^2, \quad \text{with } \lambda = \frac{1}{8} b\sqrt{n}. \quad (28)$$

To see the second inequality in (27), write

$$(w_{k+2} - w_k)^2 \leq 2(w_{k+1} - w_k)^2 + 2(w_{k+2} - w_{k+1})^2,$$

to see that the last term is at most  $\frac{1}{2} b\sqrt{n} \sum_{k=0}^{2n} (w_{k+1} - w_k)^2$ .

Now recalling the definition

$$Q_b(w; Z) = 2 \sum_{k=1}^{2n-1} Z_k w_k w_{k+1} - b\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2 - b\sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 - \frac{b}{\sqrt{n}} \sum_{k=1}^{2n} k w_k^2,$$

and plugging in the above upper bound for  $\sum_{k=1}^{2n-1} Z_k w_k w_{k+1}$ , together with  $m \geq n$  we obtain

$$\begin{aligned} Q_b(w; Z) &\leq \frac{1}{p} \sum_{k=0}^{2n-2} \Delta_p(k) (w_{k+1}^2 + w_{k+2}^2) + \frac{4}{b\sqrt{n}} \sum_{k=0}^{2n-1} \Delta_p(k)^2 w_{k+1}^2 - \frac{b}{\sqrt{n}} \sum_{k=1}^{2n} k w_k^2 \\ &= \sum_{k=0}^{2n-1} w_{k+1}^2 \left[ \frac{1}{p} (\Delta_p(k) + \Delta_p(k-1)) + \frac{4}{b\sqrt{n}} \Delta_p(k)^2 - \frac{b(k+1)}{\sqrt{n}} \right] \\ &\leq \max_{0 \leq k \leq 2n-1} \left[ \frac{1}{p} (\Delta_p(k) + \Delta_p(k-1)) + \frac{4}{b\sqrt{n}} \Delta_p(k)^2 - \frac{bk}{\sqrt{n}} \right] \end{aligned}$$

since  $\sum_k w_k^2 = 1$  and we define  $\Delta_p(-1) = 0$ . Now, for  $k \in [(j-1)p+1, jp]$  and any  $i$ , since

$$|S_{k+i} - S_k| \leq |S_k - S_{(j-1)p}| + |S_{k+i} - S_{(j-1)p}|$$

as in [15] we can write,

$$\Delta_p(k) \vee \Delta_p(k-1) \leq 2\Delta_{2p}((j-1)p),$$

and it can separately be verified that the above inequality also holds for the case  $k = 0$  and  $j = 1$ . Therefore,

$$Q_b(w; Z) \leq \max_{1 \leq j \leq \lceil 2n/p \rceil} \left[ \frac{4}{p} \Delta_{2p}((j-1)p) + \frac{16}{b\sqrt{n}} \Delta_{2p}((j-1)p)^2 - \frac{b(j-1)p}{\sqrt{n}} \right]. \quad (29)$$

Thus it follows that,

$$\begin{aligned} \mathbb{P} \left( \max_{\|w\|=1} Q_b(w; Z) \geq \varepsilon \sqrt{m} \right) &\leq \sum_{j=1}^{2n/p} \mathbb{P} \left( \frac{4}{p} \Delta_{2p}((j-1)p) - \frac{b(j-1)p}{2\sqrt{n}} \geq \frac{\varepsilon \sqrt{m}}{2} \right) \\ &\quad + \sum_{j=1}^{2n/p} \mathbb{P} \left( \frac{16}{b\sqrt{n}} \Delta_{2p}((j-1)p)^2 - \frac{b(j-1)p}{2\sqrt{n}} \geq \frac{\varepsilon \sqrt{m}}{2} \right), \\ &\leq \sum_{j=1}^{2n/p} \mathbb{P} \left( \Delta_{2p}((j-1)p) \geq \frac{1}{8} \varepsilon p \sqrt{m} + \frac{b(j-1)p^2}{8\sqrt{n}} \right) \\ &\quad + \sum_{j=1}^{2n/p} \mathbb{P} \left( \Delta_{2p}((j-1)p)^2 \geq \frac{1}{32} b \varepsilon \sqrt{mn} + \frac{b^2(j-1)p}{32} \right). \end{aligned}$$

Using the assumption about the exponential moments of the  $Z_k$ , and applying Doob's maximal inequality, we see that for any  $k \geq 1$  and any  $t > 0$ , by choosing  $\lambda = \frac{t}{p}$

$$\mathbb{P}\{\Delta_{2p}(k) \geq t\} \leq e^{-c't^2/p}.$$

Thus we get

$$\mathbb{P} \left( \Delta_{2p}((j-1)p) \geq \frac{1}{8} \varepsilon p \sqrt{m} + \frac{b(j-1)p^2}{8\sqrt{n}} \right) \leq \exp \left( -c' \varepsilon^2 p m - \frac{c' b^2 (j-1)^2 p^3}{n} \right)$$

and

$$\mathbb{P} \left( \Delta_{2p}((j-1)p)^2 \geq \frac{1}{32} b \varepsilon \sqrt{mn} + \frac{b^2(j-1)p}{32} \right) \leq \exp \left( -\frac{c' b \varepsilon \sqrt{mn}}{p} - c' b^2 (j-1) \right).$$

The sum of the first bound over  $j = 1, \dots, \lceil 2n/p \rceil$  is upper bounded by

$$C \frac{n^{1/2}}{p^{3/2}} e^{-c' \varepsilon^2 p m},$$

while the sum of the second bound is upper bounded by

$$C e^{-c' b \varepsilon \sqrt{mn}/p},$$

using that  $\exp(-c' b^2 (j-1))$  is summable and bounded independent of  $m, n$  and  $p$ . We now set  $p = \max(\lfloor \varepsilon^{-1/2} m^{-1/4} n^{1/4} \rfloor, 1)$ . Observe that if  $p > 1$ , this yields an overall bound of

$$C (\varepsilon^{3/2} m^{3/4} n^{1/4})^{1/2} e^{-c' \varepsilon^{3/2} m^{3/4} n^{1/4}} + C e^{-c' \varepsilon^{3/2} m^{3/4} n^{1/4}} \leq C e^{-c' \varepsilon^{3/2} m^{3/4} n^{1/4}}$$

by increasing the constant  $C$  and reducing the constant  $c'$  suitably. On the other hand, if  $p = 1$ , i.e., if  $\varepsilon \geq \frac{\sqrt{n}}{\sqrt{m}}$ , we get, using  $\varepsilon^2 m \geq n$ , an overall bound of

$$C e^{-c' \varepsilon^{3/2} m^{3/4} n^{1/4}} + C e^{-c' \varepsilon \sqrt{mn}}.$$

Combining the two cases we get

$$\mathbb{P} \left( \max_{\|w\|=1} Q_b(w; Z) \geq \varepsilon \sqrt{m} \right) \leq C e^{-c' \varepsilon^{3/2} \sqrt{mn} (\frac{1}{\sqrt{\varepsilon}} \wedge (\frac{m}{n})^{1/4})};$$

as desired.  $\square$

**Remark 3.4.** For later purposes, we note that if  $Y \sim \chi_\nu$  and  $\hat{Y} = Y - \mathbb{E}[Y]$  then  $\mathbb{E}[e^{\lambda \hat{Y}}] \leq e^{\frac{1}{2}\lambda^2}$  for all  $\lambda > 0$  and for all  $\nu > 0$ . This is the content of [15, Lemma 9]. Hence, if  $Z_k$ s are independent centred  $\chi$ -random variables (not necessarily identically distributed), then Lemma 3.3 is applicable.

**Remark 3.5.** It can be checked using Remark 3.4 and definition of  $X_k$ s that  $Z_k = \hat{X}_k$  we have  $\mathbb{E}[e^{\lambda Z_k}] \leq e^{c\lambda^2/\beta}$  for all  $\lambda \in \mathbb{R}$  and all  $k$ . Tracking the dependence of  $\beta$  throughout the calculations, it follows that

$$\mathbb{P}\left(\max_{\|w\|=1} Q_b(w; Z) \geq \varepsilon\sqrt{m}\right) \leq C e^{-c'\beta\varepsilon^{3/2}\sqrt{mn}(\frac{1}{\sqrt{\varepsilon}} \wedge (\frac{m}{n})^{1/4})},$$

where  $C, c'$  do not depend on  $\beta$ . Clearly, since  $\beta \geq 1$  the  $\beta$ -term can simply be dropped from the exponent to get a uniform upper bound. Using Lemma 3.2, one also gets an upper bound for  $\mathbb{P}(\lambda_n \geq (1+\varepsilon)(\sqrt{m} + \sqrt{n})^2)$  of the form  $C e^{-c'\beta\varepsilon^{3/2}\sqrt{mn}(\frac{1}{\sqrt{\varepsilon}} \wedge (\frac{m}{n})^{1/4})}$ . This recovers the first item of [15, Theorem 2] with possibly different absolute constants.

**3.3. Lower bound for the left tail.** We are now in a position to prove (23). Since  $s_n - (\sqrt{m} + \sqrt{n}) = \max_{\|w\|=1} Q(w)$  and  $Q \leq Q_b$  for small enough  $b$  (by Lemma 3.2), we have that

$$\mathbb{P}\left(s_n < (\sqrt{m} + \sqrt{n})(1 - \varepsilon)\right) \geq \mathbb{P}\left(\max_{\|w\|=1} Q_b(w) \leq -\varepsilon(\sqrt{m} + \sqrt{n})\right), \quad (30)$$

and so the following result implies (23). Note that we can replace  $\varepsilon(\sqrt{m} + \sqrt{n})$  by  $\varepsilon\sqrt{m}$  as our probability estimates are not sharp enough to differentiate  $\varepsilon$  from  $2\varepsilon$ .

**Lemma 3.6.** Fix  $0 < b < 1/4$ . Then there are constants constant  $c, C_0 > 0$  and  $m_0 \in \mathbb{N}$  such that for all  $m \geq m_0 \vee n$  and all  $\beta \geq 1$

$$\mathbb{P}\left\{\max_{\|w\|=1} Q_b(w) \leq -\varepsilon\sqrt{m}\right\} \geq \begin{cases} \exp\{-c\beta(\varepsilon\sqrt{mn})^2\} & \text{if } \frac{b\sqrt{n}}{20\sqrt{m}} \leq \varepsilon \leq \frac{b}{2}, \\ C_0^\beta \cdot \exp\{-c\beta(\varepsilon^{\frac{3}{2}} m^{\frac{3}{4}} n^{\frac{1}{4}})^2\} & \text{if } 0 < \varepsilon \leq \frac{2b\sqrt{n}}{\sqrt{m}} \wedge \frac{b}{2}. \end{cases} \quad (31)$$

Observe that Lemma 3.6 completes the proof of Theorem 2 except for the case  $m \leq m_0$ . However, for  $m \leq m_0$ ,  $s_n < \sqrt{m} + \sqrt{n} - \varepsilon\sqrt{m}$  can simply be ensured by making all (non-zero) entries of the matrix  $T$  sufficiently small. Considering the density of  $\chi$  random variables, it is easy to check that the probability of such an event is lower bounded by  $C_0^{\beta mn} \geq C_0^{\beta m_0^2}$  for some (possibly different) constant  $C_0$  (depending on  $b$  and  $m_0$ ). This takes care of the remaining case and completes the proof of Theorem 2.

*Proof of Lemma 3.6.* We first assume  $\varepsilon \geq \varepsilon_0 := C_1 m^{-1/2} n^{-1/6}$ , for a  $C_1$  which will be taken to be a sufficiently large absolute constant, chosen appropriately later. We shall choose  $m$  sufficiently large so that  $\varepsilon_0$  will be much smaller than  $b/2$ . We further divide into two cases depending on the range of  $\varepsilon$ .

**Case 1:**  $\varepsilon \leq \frac{2b\sqrt{n}}{\sqrt{m}} \wedge \frac{b}{2}$ . In this case we fix  $K = \lceil \varepsilon\sqrt{mn}/4b \rceil \leq \frac{n+1}{2}$ . Let  $Y_k$  be independent random variables such that

$$Y_{2k-1}^2 \stackrel{d}{=} (1 - \varepsilon) X_{2k-1}^2 \sim \text{Gam}\left(\frac{\beta(m+1-k)}{2}, \frac{\beta}{2(1-\varepsilon)}\right)^2 \quad (32)$$

for  $1 \leq k \leq K$ , while  $Y_k^2 \stackrel{d}{=} X_k^2$  for all other  $k \leq 2n-1$ . The constraint on  $\varepsilon$  was used to ensure that  $K < n$  (otherwise the definition of  $Y_k$  does not make sense). In fact, we have that  $K \leq (n+1)/2$ , and so recalling (21), parameters of the  $\chi_{\beta(m+1-k)}^2$  distributions of  $X_k$  are all at least  $\beta m/2$  for  $k \leq K$ .

<sup>2</sup>Gam( $\alpha, \kappa$ ) denotes the Gamma distribution with density proportional to  $x^{\alpha-1} e^{-\kappa x}$ .



We adopt the following perturbative strategy: Using  $Y_k$  in place of  $X_k$  in the definition of  $Q_b$ , implies that  $Q_b(w) \leq -\varepsilon\sqrt{m}$  for all unit vectors  $w$ , with probability close to 1. This is because the first  $K$  of the  $Y_{2k-1}$  random variables have a slightly reduced mean (by about  $\varepsilon\sqrt{m}$ ), as compared to  $X_{2k-1}$ . Then we compute the Radon-Nikodym derivatives of  $X$  with respect to  $Y$  to get a lower bound for the probability of the same event under the  $X_k$ .

So replacing  $X$  by  $Y$ , in (25), consider

$$\tilde{Q}_b(w, Y) := 2 \sum_{k=1}^{2n-1} (Y_k - \mathbb{E}[X_k])w_k w_{k+1} \quad (33)$$

$$\begin{aligned} & - b\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2 - b\sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 - \frac{b}{\sqrt{n}} \sum_{k=1}^{2n} k w_k^2 \\ & = Q_{b/2}(w, Z) + 2 \sum_{k=1}^K (\mathbb{E}[Y_{2k-1}] - \mathbb{E}[X_{2k-1}])w_{2k-1}w_{2k} \\ & \quad - \frac{b}{2}\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2 - \frac{b}{2}\sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 - \frac{b}{2\sqrt{n}} \sum_{k=1}^{2n} k w_k^2 \end{aligned} \quad (34)$$

where  $Z_k = Y_k - \mathbb{E}[Y_k]$ .

Using Remark 3.5 (it is applicable since  $Y_k$ s are all scalar multiples of  $X_k$ s in distribution with a multiplication factor uniformly bounded from above and below), recalling the upper bound of  $e^{-c'\varepsilon^{3/2}\sqrt{m\bar{m}}(\frac{1}{\sqrt{\varepsilon}} \wedge (\frac{m}{n})^{1/4})}$  from there, and our choice of  $\varepsilon_0 := C_1 m^{-1/2} n^{-1/6}$ , setting  $C_1$  large enough (independent of  $\beta$ ) allows us to ensure that for all  $\varepsilon \geq \varepsilon_0$ ,

$$\mathbb{P}(Q_{b/2}(w, Z) \leq \frac{1}{8}\varepsilon\sqrt{m}) \geq \mathbb{P}(Q_{b/2}(w, Z) \leq \frac{1}{8}\varepsilon_0\sqrt{m}) \geq 0.9. \quad (35)$$

To estimate  $\tilde{Q}_b(Y)$ , in (34) we will use

$$2w_{2k-1}w_{2k} = w_{2k}^2 + w_{2k-1}^2 - (w_{2k-1} - w_{2k})^2 \quad (36)$$

Furthermore, using the distributional equality in (32), it follows from (20) that

$$-\varepsilon\sqrt{m} \leq \mathbb{E}[Y_{2k-1}] - \mathbb{E}[X_{2k-1}] \leq -\frac{1}{4}\varepsilon\sqrt{m} \quad (37)$$

for all  $1 \leq k \leq K$ . Putting these together, we have that with probability 0.9 or more for all  $w$ ,

$$\tilde{Q}_b(w, Y) \leq \frac{1}{8}\varepsilon\sqrt{m} - \frac{1}{4}\varepsilon\sqrt{m} \sum_{k=1}^{2K} w_k^2 - \left(\frac{b}{2} - \varepsilon\right) \sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 - \frac{2bK}{2\sqrt{n}} \sum_{k=2K+1}^{2n} w_k^2 \quad (38)$$

$$\leq \frac{1}{8}\varepsilon\sqrt{m} - \frac{1}{4}\varepsilon\sqrt{m} \sum_{k=1}^{2K} w_k^2 - \frac{1}{4}\varepsilon\sqrt{m} \sum_{k=2K+1}^{2n} w_k^2 = -\frac{1}{8}\varepsilon\sqrt{m}. \quad (39)$$

The first inequality follows by considering (34), and:

- (1) Using (35) to bound  $Q_{b/2}(w, Z)$ .
- (2) Using (36), and (37) to the term  $2 \sum_{k=1}^K (\mathbb{E}[Y_{2k-1}] - \mathbb{E}[X_{2k-1}])w_{2k-1}w_{2k}$  to obtain the term  $-\frac{1}{4}\varepsilon\sqrt{m} \sum_{k=1}^{2K} w_k^2$ .
- (3) Reduce the coefficient in the third term to  $\frac{b}{2} - \varepsilon$  for all  $k$ .
- (4) Dropping the term  $-\frac{b}{2}\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2$ .
- (5) Dropping the terms with  $k \leq 2K$  in the last sum on the RHS and lower bounded  $k \geq 2K+1$  by  $2K$ .

For the second inequality (39), we have dropped the third term (which can be done since  $2\varepsilon < b$  by hypothesis on  $\varepsilon$ ). In the final term we use that by choice,  $\frac{bK}{\sqrt{n}} \geq \frac{1}{4}\varepsilon\sqrt{m}$ . Finally using  $\sum_j w_j^2 = 1$ , yields the final equality.

**Bounding the Radon-Nikodym derivative.** For notational convenience we start by defining the following nice set.

$$\mathcal{A} = \left\{ t \in \mathbb{R}_+^{2n-1} : \forall w \text{ with } \|w\| = 1, 2 \sum_{k=1}^{2n-1} (\sqrt{t_k} - \mathbb{E}[X_k])w_k w_{k+1} - b\sqrt{n} \sum_{k=0}^n (w_{2k} - w_{2k+1})^2 - b\sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 x - \frac{b}{\sqrt{n}} \sum_{k=1}^{2n} kw_k^2 < -\frac{1}{8}\varepsilon\sqrt{m} \right\}.$$

Through the discussion so far leading to (39), we have shown that

$$\mathbb{P}\{Y^2 \in \mathcal{A}\} \geq 0.9,$$

where  $Y^2 := (Y_1^2, \dots, Y_{2n-1}^2)$ . Similarly we will use  $X^2$  to denote  $(X_1^2, \dots, X_{2n-1}^2)$ .

Now let  $g_k$  denote the density of  $X_k^2$  and let  $f_k$  denote the density of  $Y_k^2$ . Of course  $f_k = g_k$  except for  $k = 1, 3, \dots, 2K - 1$ . Recalling from (32), for convenience, we list here the exact forms of  $f_{2k-1}$  and  $g_{2k-1}$ , for  $t \geq 0$  and  $1 \leq k \leq K$ :

$$f_{2k-1}(t) = \frac{t^{\frac{1}{2}\beta(m+1-k)-1} \exp\left(-\frac{\beta t}{2(1-\varepsilon)}\right)}{\Gamma\left(\frac{\beta(m+1-k)}{2}\right) (2(1-\varepsilon))^{\frac{1}{2}\beta(m+1-k)}},$$

$$g_{2k-1}(t) = \frac{t^{\frac{1}{2}\beta(m+1-k)-1} \exp\left(-\frac{\beta t}{2}\right)}{\Gamma\left(\frac{\beta(m+1-k)}{2}\right) 2^{\frac{1}{2}\beta(m+1-k)}}.$$

Next we substitute these expressions to get,

$$\begin{aligned} \mathbb{P}\{X^2 \in \mathcal{A}\} &= \int_{\mathcal{A}} \prod_{k=1}^K \frac{g_{2k-1}(t_{2k-1})}{f_{2k-1}(t_{2k-1})} \prod_{k=1}^{2n-1} f_k(t_k) dt_1 \dots dt_{2n-1} \\ &= \int_{\mathcal{A}} \exp\left\{ \frac{\beta\varepsilon}{2(1-\varepsilon)} \sum_{k=1}^K t_{2k-1} \right\} (1-\varepsilon)^{\frac{\beta}{2} \sum_{k=1}^K (m+1-k)} \prod_{k=1}^{2n-1} f_k(t_k) dt_1 \dots dt_{2n-1}. \end{aligned}$$

Now let  $\mathcal{B} = \left\{ t \in \mathbb{R}_+^{2n-1} : \sum_{k=1}^K t_{2k-1} > (1-\varepsilon) \sum_{k=1}^K (m+1-k) \right\}$ . Since  $\mathbb{E}[Y_{2k-1}^2] = (1-\varepsilon)(m+1-k)$ , it follows that

$$\mathbb{P}\{Y^2 \in \mathcal{B}\} \geq 0.4. \tag{40}$$

A quick way to see this is to use the normal approximation for Gamma variables. More precisely by (32): For any  $K$  as above,

$$\sum_{k=1}^K Y_{2k-1}^2 \sim \text{Gam}\left(\frac{\sum_{k=1}^K \beta(m+1-k)}{2}, \frac{\beta}{2(1-\varepsilon)}\right).$$

Now as  $\varepsilon$  is bounded away from 1 and  $\beta$  is a given constant, uniformly for all  $K \leq \frac{m}{2}$ , as  $m \rightarrow \infty$  since  $\sum_{k=1}^K \frac{\beta(m+1-k)}{2}$  goes to infinity, and  $\frac{\beta}{2(1-\varepsilon)}$  stays fixed, by the Central Limit Theorem for

gamma variables with increasing shape parameters and a given scale parameter,<sup>3</sup> we have,

$$\mathbb{P} \left[ \text{Gam} \left( \frac{\sum_{k=1}^K \beta(m+1-k)}{2}, \frac{\beta}{2(1-\varepsilon)} \right) \geq \mathbb{E} \left( \text{Gam} \left( \frac{\sum_{k=1}^K \beta(m+1-k)}{2}, \frac{\beta}{2(1-\varepsilon)} \right) \right) \right] \rightarrow \frac{1}{2},$$

and hence in particular (using  $\beta \geq 1$ ) for any large enough  $m \geq m_0$ , (40) is satisfied. Thus,

$$\begin{aligned} \mathbb{P}\{X^2 \in \mathcal{A}\} &\geq \int_{\mathcal{A} \cap \mathcal{B}} \exp \left\{ \frac{\beta\varepsilon}{2(1-\varepsilon)} \sum_{k=1}^K t_{2k-1} \right\} (1-\varepsilon)^{\frac{\beta}{2} \sum_{k=1}^K (m+1-k)} \prod_{k=1}^{2n-1} f_k(t_k) dt_1 \dots dt_{2n-1}, \\ &\geq \exp \left\{ \frac{\beta\varepsilon}{2(1-\varepsilon)} (1-\varepsilon) \sum_{k=1}^K (m+1-k) \right\} (1-\varepsilon)^{\frac{\beta}{2} \sum_{k=1}^K (m+1-k)} \mathbb{P}\{Y^2 \in \mathcal{A} \cap \mathcal{B}\} \\ &\geq \frac{1}{4} \exp \left( \frac{\beta}{2} \left( \sum_{k=1}^K (m+1-k) \right) (\varepsilon + \log(1-\varepsilon)) \right) \geq \frac{1}{4} e^{-\frac{1}{2}\varepsilon^2 \beta K m}. \end{aligned} \quad (41)$$

In the last line we used  $\mathbb{P}\{Y^2 \in \mathcal{A} \cap \mathcal{B}\} \geq 0.9 + 0.4 - 1 > \frac{1}{4}$  and  $\sum_{k=1}^K (m+1-k) \leq Km$  together with  $\varepsilon + \log(1-\varepsilon) \in (-\varepsilon^2, 0)$  (if  $0 < \varepsilon < \frac{1}{2}$ ). By our choice of  $K$  we get that this is further lower bounded by  $\frac{1}{4} e^{-c\beta\varepsilon^3 m^{3/2} n^{1/2}}$ .

This is exactly what was asked for in Lemma 3.6, for the case when  $\varepsilon \leq c'' \sqrt{n/m}$  for a small enough  $c''$ .

**Case 2:**  $\frac{b\sqrt{n}}{20\sqrt{m}} < \varepsilon < \frac{b}{2}$ . In this case, we take  $K = n$  and define  $Y_k$  as in (32). Thus all the odd  $Y_k$  are different from all the odd  $X_k$ . Proceeding exactly as before, in the analysis of (38) (the first place where the definition of  $K$  was used) the last term is an empty sum and hence dropped yielding the following instead of (39).

$$\tilde{Q}_b(w, Y) \leq \frac{1}{8} \varepsilon \sqrt{m} - \frac{1}{4} \varepsilon \sqrt{m} \sum_{k=1}^{2K} w_k^2 - \left( \frac{b}{2} - \varepsilon \right) \sqrt{m} \sum_{k=1}^n (w_{2k-1} - w_{2k})^2 \leq -\frac{1}{8} \varepsilon \sqrt{m}. \quad (42)$$

Continuing, we get the same probability bound (41) as before, except that  $K = n$ . Thus we arrive at

$$\mathbb{P}\{X^2 \in \mathcal{A}\} \geq \frac{1}{4} e^{-c\varepsilon^2 \beta mn} \geq e^{-c\varepsilon^2 \beta mn}$$

by increasing the constant  $c$ , and using  $\beta \geq 1$  together with  $\varepsilon^2 mn \geq \frac{b^2 n}{400} \geq \frac{b^2}{400}$ . This is what the lemma claims.

**Finally we address the case:**  $\varepsilon \leq \varepsilon_0 = C_1 m^{-1/2} n^{-1/6}$ .

$$\mathbb{P} \left\{ \max_{\|w\|=1} Q_b(w) \leq -\varepsilon \sqrt{m} \right\} \geq \mathbb{P} \left\{ \max_{\|w\|=1} Q_b(w) \leq -\varepsilon_0 \sqrt{m} \right\} \geq \frac{1}{4} e^{-c\beta C_1^3} \geq (C_0)^\beta e^{-c\beta \varepsilon^3 m^{3/2} n^{1/2}},$$

for a suitably increased constant  $c$  (depending on  $b$ ) where  $C_0 = \frac{1}{4} e^{-cC_1^3}$ .  $\square$

**3.4. Deviation bounds in the large deviation regime.** We finish with a discussion of lower bounds of the large deviation probabilities in the left tail: i.e.,  $\mathbb{P}(\lambda_n \leq (\sqrt{m} + \sqrt{n})^2 (1-\varepsilon))$  where  $c' \leq \varepsilon < 1$ . Notice that this case is not covered by Theorem 2. For  $\beta = 2$  case, Johansson [12] obtained that if  $m = \gamma n$ , then

$$-n^{-2} \log \mathbb{P}(\lambda_n \leq (\sqrt{m} + \sqrt{n})^2 (1-\varepsilon)) \rightarrow J_\gamma(\varepsilon)$$

<sup>3</sup>Since  $\text{Gam}(\alpha, \kappa) = \text{Gam}([\alpha], \kappa) * \text{Gam}(\alpha - [\alpha], \kappa)$ , and the first term is a sum of  $[\alpha]$  many i.i.d.  $\text{Gam}(1, \kappa)$  and  $\text{Gam}(\alpha - [\alpha], \kappa)$  is a tight random variable.

for a large deviation rate function  $J_\gamma(\cdot)$ . We shall briefly describe below how to show a corresponding lower bound for finite  $n$  in this regime with a rather simple argument (the upper bound was covered in [15]).

While for this discussion we shall restrict ourselves to the case  $m = n$  sufficiently large, and also to  $\beta = 1$ , one can use the same argument for all  $\beta \geq 1$  and  $m, n$  with  $\frac{m}{n}$  is bounded away from infinity.

Recall the  $2n \times 2n$  tridiagonal matrix  $T$  with largest eigenvalue  $s_n$  where  $s_n^2 = \lambda_n$ . We use the following well known and easy to prove bound (Gershgorin theorem)

$$s_n \leq \max_{0 \leq i \leq 2n-1} (X_i + X_{i+1})$$

where  $X_i$ s are the independent  $\chi$  variables defined in (22) with the convention that  $X_0 = X_{2n} = 0$ . It now follows easily that

$$\mathbb{P}(s_n \leq 2\sqrt{n}(1 - \varepsilon)) \geq \prod_{i=1}^{2n-1} \mathbb{P}(X_i \leq (1 - \varepsilon)\sqrt{n}).$$

Now each term in the product can be lower bounded by a constant (say,  $\frac{1}{4}$ ) if  $\mathbb{E}X_i^2 \leq (1 - \varepsilon)^2 n$ . Using the left tail of  $\chi^2$  distribution one can lower bound each of the other terms by  $e^{-c(\varepsilon)n}$ , leading to a lower bound of the form  $e^{-c'(\varepsilon)n^2}$  for  $n$  sufficiently large. A slightly more careful version of the above calculation yields  $c'(\varepsilon) \approx \varepsilon^3$  if  $\varepsilon$  is sufficiently small (but still bounded away from 0) thus matching the lower bound in Theorem 2.

Note however, that this approach cannot be carried over to get optimal tails bounds all the way up to the moderate deviation tails, i.e.,  $\varepsilon \approx n^{-2/3}$ . In this case, observe that at least the first  $\Theta(n^{1/3})$  many of the terms in the product  $\prod_{i=1}^{2n-1} \mathbb{P}(X_i \leq (1 - \varepsilon)\sqrt{n})$  are bounded away from 1. Hence this approach gives us a lower bound that decays to 0 at least as fast as  $e^{-cn^{1/3}}$ , much worse than the constant order lower bound obtained in Theorem 2.

## REFERENCES

- [1] Jinho Baik. Painlevé expressions for LOE, LSE and interpolating ensembles. *International Mathematics Research Notices*, 2002(33):1739–1789, 2002.
- [2] Jinho Baik, Percy Deift, Ken T.-R. McLaughlin, Peter Miller, and Xin Zhou. Optimal tail estimates for directed last passage site percolation with geometric random variables. *Adv. Theor. Math. Phys.*, 5(6):1207–1250, 2001.
- [3] Jinho Baik and Eric M. Rains. Algebraic aspects of increasing subsequences. *Duke Math. J.*, 109(1):1–65, 2001.
- [4] Jinho Baik and Eric M. Rains. Symmetrized random permutations. In *Random Matrix Models and Their Applications, volume 40 of Mathematical Sciences Research Institute Publications*, pages 1–19, 2001.
- [5] Jinho Baik and Toufic M. Suidan. A GUE central limit theorem and universality of directed first and last passage site percolation. *International Mathematics Research Notices*, 2005(6):325–337, 2005.
- [6] Yu Baryshnikov. GUEs and queues. *Probability Theory and Related Fields*, 119(2):256–274, 2001.
- [7] Thierry Bodineau and James Martin. A Universality Property for Last-Passage Percolation Paths Close to the Axis. *Electron. Commun. Probab.*, 10:105–112, 2005.
- [8] Alexei Borodin, Patrik L. Ferrari, Michael Prähofer, and Tomohiro Sasamoto. Fluctuation Properties of the TASEP with Periodic Initial Configuration. *Journal of Statistical Physics*, 129(5):1055–1080, 2007.
- [9] Ivan Corwin, Zhipeng Liu, and Dong Wang. Fluctuations of TASEP and LPP with general initial data. *Ann. Appl. Probab.*, 26(4):2030–2082, 08 2016.
- [10] Ioana Dumitriu and Alan Edelman. Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847, 2002.
- [11] Will FitzGerald and Jon Warren. Point-to-line last passage percolation and the invariant measure of a system of reflecting Brownian motions. *arXiv e-prints*, page arXiv:1904.03253, Apr 2019.
- [12] Kurt Johansson. Shape fluctuations and random matrices. *Communications in Mathematical Physics*, 209(2):437–476, 2000.
- [13] G. Kalai. Laws of iterated logarithm for random matrices and random permutation. <http://mathoverflow.net/questions/142371/Laws-of-iterated-logarithm-for-random-matrices-and-random-permutation>, 2013.

- [14] Michel Ledoux. A law of the iterated logarithm for directed last passage percolation. *J. Theor. Probab.*, 31(4):2366–2375, 2018.
- [15] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15:1319–1343, 2010.
- [16] Matthias Löwe and Franz Merkl. Moderate deviations for longest increasing subsequences: The upper tail. *Comm. Pure Appl. Math.*, 54:1488–1519, 2001.
- [17] Matthias Löwe, Franz Merkl, and Silke Rolles. Moderate deviations for longest increasing subsequences: The lower tail. *J. Theor. Probab.*, 15(4):1031–1047, 2002.
- [18] Gia Bao Nguyen and Daniel Remenik. Non-intersecting Brownian bridges and the Laguerre Orthogonal Ensemble. *Ann. Inst. H. Poincaré Probab. Statist.*, 53(4):2005–2029, 2017.
- [19] Neil O’Connell and Marc Yor. A representation for non-colliding random walks. *Electronic communications in probability*, 7:1–12, 2002.
- [20] Elliot Paquette and Ofer Zeitouni. Extremal eigenvalue correlations in the GUE minor process and a law of fractional logarithm. *Ann. Probab.*, 45(6A):4112–4166, 2017.
- [21] José A. Ramirez, Brian Rider, and Bálint Virág. Beta ensembles, stochastic Airy spectrum, and a diffusion. *Journal of the American Mathematical Society*, 24(4):919–944, 2011.
- [22] H. Rost. Nonequilibrium behaviour of a many particle process: Density profile and local equilibria. *Zeitschrift f. Warsch. Verw. Gebiete*, 58(1):41–53, 1981.
- [23] Tomohiro Sasamoto. Spatial correlations of the 1d KPZ surface on a flat substrate. *Journal of Physics A: Mathematical and General*, 38(33):L549–L556, 2005.
- [24] Toufic Suidan. A remark on a theorem of Chatterjee and last passage percolation. *Journal of Physics A: Mathematical and General*, 39(28):8977–8981, jun 2006.
- [25] Terence Tao and Van Vu. Random matrices: Universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an Appendix by Manjunath Krishnapur.

RIDDHIPRATIM BASU, INTERNATIONAL CENTRE FOR THEORETICAL SCIENCES, TATA INSTITUTE OF FUNDAMENTAL RESEARCH, BANGALORE, INDIA

*E-mail address:* `rbasu@icts.res.in`

SHIRSHENDU GANGULY, DEPARTMENT OF STATISTICS, UC BERKELEY, CA, USA

*E-mail address:* `sganguly@berkeley.edu`

MILIND HEGDE, DEPARTMENT OF MATHEMATICS, UC BERKELEY, CA, USA

*E-mail address:* `mhegde@math.berkeley.edu`

MANJUNATH KRISHNAPUR, DEPARTMENT OF MATHEMATICS, INDIAN INSTITUTE OF SCIENCE, BANGALORE, INDIA.

*E-mail address:* `manju@iisc.ac.in`