Universality for the Toda algorithm to compute the eigenvalues of a random matrix

Percy Deift and Thomas Trogdon[0]

Courant Institute of Mathematical Sciences
New York University
251 Mercer St.
New York, NY 10012, USA

April 25, 2016

Abstract

We prove universality for the fluctuations of the halting time $T^{(1)}$ (see Introduction) for the Toda algorithm to compute the eigenvalues of real symmetric and Hermitian matrices H. $T^{(1)}$ controls the computation time for the largest eigenvalue of H to a prescribed tolerance. The proof relies on recent results on the statistics of the eigenvalues and eigenvectors of random matrices (such as delocalization, rigidity and edge universality) in a crucial way.

Keywords: Universality, Toda Lattice, Random matrix theory

MSC Classification: 15B52, 65L15, 70H06

1 Introduction

In [Pfrang et al., 2014] the authors initiated a statistical study of the performance of various standard algorithms $\mathcal{A}$ to compute the eigenvalues of random real symmetric matrices H. Let Σ_N denote the set of real $N \times N$ symmetric matrices. Associated with each algorithm $\mathcal{A}$, there is, in the discrete case such as QR, a map $\varphi = \varphi_{\mathcal{A}} : \Sigma_N \to \Sigma_N$, with the properties

- (isospectral) $\mathrm{spec}(\varphi_{\mathcal{A}}(H)) = \mathrm{spec}(\mathrm{H})$,
- (convergence) the iterates $X_{k+1} = \varphi_{\mathcal{A}}(X_k)$, $k \geq 0$, $X_0 = H$ given, converge to a diagonal matrix X_∞, $X_k \to X_\infty$ as $k \to \infty$,

and in the continuum case, such as Toda, there is a flow $t \mapsto X(t) \in \Sigma_N$ with the properties

- (isospectral) $\mathrm{spec}(X(t))$ is constant,
- (convergence) the flow $X(t)$, $t \geq 0$, $X(0) = H$ given, converges to a diagonal matrix X_∞, $X(t) \to X_\infty$ as $t \to \infty$.

In both cases, necessarily, the (diagonal) entries of X_∞ are the eigenvalues of the given matrix H.

Given $\epsilon > 0$, it follows, in the discrete case, that for some m the off-diagonal entries of X_m are $\mathcal{O}(\epsilon)$ and hence the diagonal entries of X_m give the eigenvalues of $X_0 = H$ to $\mathcal{O}(\epsilon)$. The situation is similar for continuous algorithms $t \mapsto X(t)$. Rather than running the algorithm until all the off-diagonal entries are $\mathcal{O}(\epsilon)$, it is customary to run the algorithm with *deflations* as follows. For an $N \times N$ matrix Y in block form

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix},$$

[0]Email: deift@cims.nyu.edu, trogdon@cims.nyu.edu (corresponding author).

with Y_{11} of size $k \times k$ and Y_{22} of size $N-k \times N-k$ for some $k \in \{1, 2, \ldots, N-1\}$, the process of projecting $Y \mapsto \mathrm{diag}(Y_{11}, Y_{22})$ is called deflation. For a given ϵ, algorithm $\mathcal{A}$ and matrix $H \in \Sigma_N$, define the *k-deflation* time $T^{(k)}(H) = T^{(k)}_{\epsilon,\mathcal{A}}(H)$, $1 \le k \le N-1$, to be the smallest value of m such that X_m, the mth iterate of algorithm $\mathcal{A}$ with $X_0 = H$, has block form

$$X_m = \begin{bmatrix} X_{11}^{(k)} & X_{12}^{(k)} \\ X_{21}^{(k)} & X_{22}^{(k)} \end{bmatrix},$$

with $X_{11}^{(k)}$ of size $k \times k$ and $X_{22}^{(k)}$ of size $N-k \times N-k$ and[1] $\|X_{12}^{(k)}\| = \|X_{21}^{(k)}\| \le \epsilon$. The defation time $T(H)$ is then defined as

$$T(H) = T_{\epsilon,\mathcal{A}}(H) = \min_{1 \le k \le N-1} T^{(k)}_{\epsilon,\mathcal{A}}(H).$$

If $\hat{k} \in \{1, \ldots, N-1\}$ is such that $T(H) = T^{(\hat{k})}_{\epsilon,\mathcal{A}}(H)$, it follows that the eigenvalues of $H = X_0$ are given by the eigenvalues of the block-diagonal matrix $\mathrm{diag}(X_{11}^{(\hat{k})}, X_{22}^{(\hat{k})})$ to $\mathcal{O}(\epsilon)$. After, running the algorithm to time $T = T_{\epsilon,A}(H)$, the algorithm restarts by applying $\mathcal{A}$ separately to the smaller matrices $X_{11}^{(\hat{k})}$ and $X_{22}^{(\hat{k})}$ until the next deflation time, and so on. There are again similar considerations for continuous algorithms.

As the algorithm proceeds, the number of matrices after each deflation doubles. This is counterbalanced by the fact that the matrices are smaller and smaller in size, and the calculations are clearly parallelizable. Allowing for parallel computation, the number of deflations to compute all the eigenvalues of a given matrix H to a given accuracy ϵ, will vary from $\mathcal{O}(\log N)$ to $\mathcal{O}(N)$.

In [Pfrang et al., 2014] the authors considered the deflation time T for $N \times N$ matrices chosen from a given ensemble $\mathcal{E}$. To make the dependence of T on $\epsilon, N, \mathcal{A}$ and $\mathcal{E}$ explicit, we write $T^{(k)}_{\epsilon,\mathcal{A}}(H) = T^{(k)}_{\epsilon,N,\mathcal{A},\mathcal{E}}$ and $T(H) = T_{\epsilon,\mathcal{A}}(H) = T_{\epsilon,N,\mathcal{A},\mathcal{E}}(H)$. For a given algorithm $\mathcal{A}$ and ensemble $\mathcal{E}$ the authors computed $T_{\epsilon,N,\mathcal{A},\mathcal{E}}(H)$ for 5,000-15,000 samples of matrices H chosen from $\mathcal{E}$, and recorded the *normalized deflation time*

$$\tilde{T}_{\epsilon,N,\mathcal{A},\mathcal{E}}(H) := \frac{T_{\epsilon,N,\mathcal{A},\mathcal{E}}(H) - \langle T_{\epsilon,N,\mathcal{A},\mathcal{E}} \rangle}{\sigma_{\epsilon,N,\mathcal{A},\mathcal{E}}}, \tag{1.1}$$

where $\langle T_{\epsilon,N,\mathcal{A},\mathcal{E}} \rangle$ and $\sigma^2_{\epsilon,N,\mathcal{A},\mathcal{E}} = \langle (T_{\epsilon,N,\mathcal{A},\mathcal{E}} - \langle T_{\epsilon,N,\mathcal{A},\mathcal{E}} \rangle)^2 \rangle$ are the sample average and sample variance of $T_{\epsilon,N,\mathcal{A},\mathcal{E}}(H)$, respectively. Surprisingly, the authors found that for the given algorithm $\mathcal{A}$, and ϵ and N in a suitable scaling range with $N \to \infty$, the histogram of $\tilde{T}_{\epsilon,N,\mathcal{A},\mathcal{E}}$ was universal, independent of the ensemble $\mathcal{E}$. In other words, the fluctuations in the deflation time $\tilde{T}_{\epsilon,N,\mathcal{A},\mathcal{E}}$, suitably scaled, were universal, independent of $\mathcal{E}$. Figure 1 displays some of the numerical results from [Pfrang et al., 2014]. Figure 1(a) displays data for the QR algorithm, which is discrete, and Figure 1(b) displays data for the Toda algorithm, which is continuous.

Subsequently, in [Deift et al., 2014], the authors raised the question of whether the universality results in [Pfrang et al., 2014] were limited to eigenvalue algorithms, or whether they were present more generally in numerical computation. And indeed the authors in [Deift et al., 2014] found similar universality results for a wide variety of numerical algorithms, including

- more general eigenvalue algorithms such as the Jacobi eigenvalue algorithm, and also algorithms for Hermitian ensembles,
- the conjugate gradient and GMRES algorithms to solve linear $N \times N$ systems $Hx = b$,
- an iterative algorithm to solve the Dirichlet problem $\Delta u = 0$ on a random star-shaped region $\Omega \subset \mathbb{R}^2$ with random boundary data f on $\partial\Omega$, and
- a genetic algorithm to compute the equilibrium measure for orthogonal polynomials on the line.

[1] Here we use $\|\cdot\|$ to denote the Frobenius norm $\|X\|^2 = \sum_{i,j} |X_{ij}|^2$ for $X = (X_{ij})$.

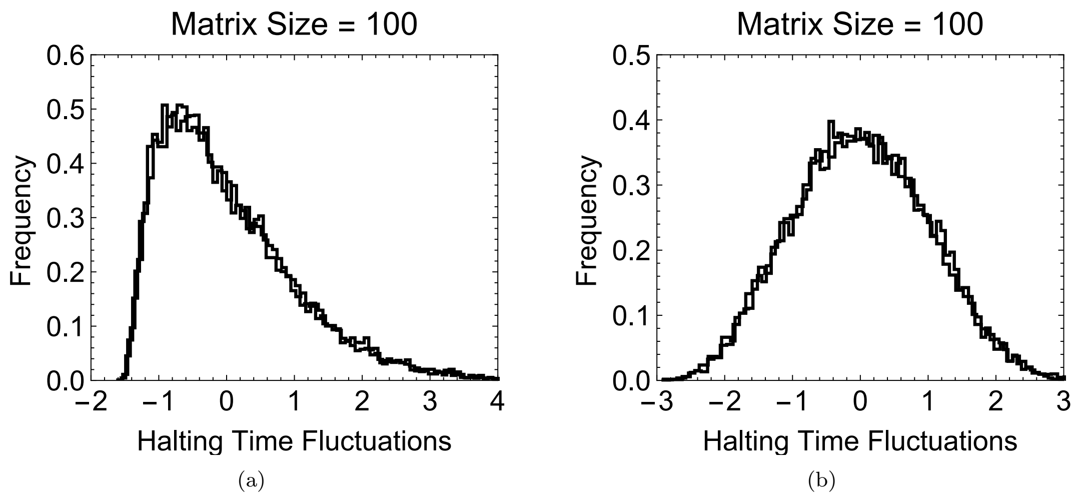


Figure 1: Universality for $\tilde{T}_{\epsilon,N,\mathcal{A},\mathcal{E}}$ when (a) $\mathcal{A}$ is the QR eigenvalue algorithm and when (b) $\mathcal{A}$ is the Toda algorithm. Panel (a) displays the overlay of two histograms for $\tilde{T}_{\epsilon,N,\mathcal{A},\mathcal{E}}$ in the case of QR, one for each of the two ensembles $\mathcal{E} = \text{BE}$, consisting of iid mean-zero Bernoulli random variables (see Definition 1.1) and $\mathcal{E} = \text{GOE}$, consisting of iid mean-zero normal random variables. Here $\epsilon = 10^{-10}$ and $N = 100$. Panel (b) displays the overlay of two histograms for $\tilde{T}_{\epsilon,N,\mathcal{A},\mathcal{E}}$ in the case of the Toda algorithm, and again $\mathcal{E} = \text{BE}$ or GOE. And here $\epsilon = 10^{-8}$ and $N = 100$.

An example from [Deift et al., 2014] is the solution of the linear system $Hx = b$ using the conjugate gradient algorithm where H is chosen randomly from an ensemble $\mathcal{E}$ of positive definite matrices, and b has iid components. The algorithm is iterative, $b \to x_0 \to x_1 \to \cdots \to x_m \to \cdots$ and halts, for a given ϵ, when

$$\|Hx_m - b\| \leq \epsilon. \tag{1.2}$$

The smallest value m for which (1.2) holds is called the *halting time* $h = h_{\epsilon,N,\mathcal{E}}(H, b)$: h is the analog for the conjugate gradient method of the deflation time T for eigenvalue algorithms. Figure 2, taken from [Deift et al., 2014], displays universality in the halting time for the conjugate gradient algorithm for various ensembles for H and b. Note that all the algorithms considered above, except for the genetic algorithm, have the character of deterministic dynamical systems with random initial data. On the other hand, for the genetic algorithm, not only is the initial data random, but the algorithm itself is stochastic.

In [Deift et al., 2014] the authors also considered recent laboratory experiments of Bakhtin and Correll [Bakhtin and Correll, 2012] in which participants were required to make a sequence of decisions comparing geometrical properties of figures projected on a screen. Bakhtin and Correll recorded the decision times τ and the plots of the histograms for the normalized times $\tilde{\tau}$ as in (1.1) for each participant, strongly suggest universality in the decision process. Furthermore, using a Curie–Weiss spin model, Bakhtin and Correll derive an explicit formula

$$\tilde{f}_{\mathrm{BC}}(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2}\mathrm{e}^{-2x} - x\right) \tag{1.3}$$

for the histogram for the decision process. Let $f_{\mathrm{BC}}(x) = \tilde{\gamma}\tilde{f}_{\mathrm{BC}}(\tilde{\gamma}(x - \tilde{a}))$ where $\tilde{\gamma}$ and $\tilde{a}$ are chosen so that $f_{\mathrm{BC}}(x)$ has mean zero and variance one. It is an interesting fact, observed recently by [Sagun et al., 2015], that the fluctuations of search times on `Google` for randomly chosen English and Turkish words, in particular, also appear to follow the law f_{BC}, see Figure 3.

The above calculations and experiments suggest strongly that calculation in the presence of random data in a wide variety of situations obeys two-component universality, that is, once the mean and variance are known, the remaining statistics are universal. So far, however, all the evidence is numerical and experimental.

The goal of this paper is prove universality rigorously for an algorithm of interest and we focus, in particular, on eigenvalue algorithms. To analyze eigenvalue algorithms with deflation, in particular, one

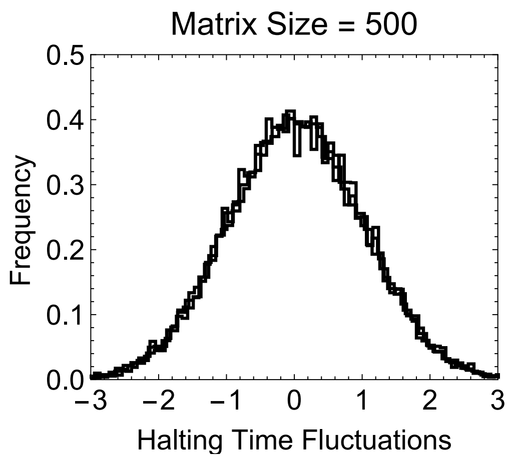


Figure 2: Universality in the halting time for the conjugate gradient algorithm. This plot shows three histograms of $\tilde{h}_{\epsilon,N,\mathrm{CG}}(H,b) = \dfrac{h_{\epsilon,N,\mathrm{CG}}(H,b) - \langle h_{\epsilon,N,\mathrm{CG}}\rangle}{\sigma_{\epsilon,N,\mathrm{CG}}}$ corresponding to different ensembles. Here $\epsilon = 10^{-14}$, $N = 500$, b has iid entries, uniform on $[-1,1]$, and $H = ZZ^*$ where Z is $N \times (N + \lfloor\sqrt{N}\rfloor)$ with standard normal (real and complex) or ± 1 Bernoulli entries. See [Deift et al., 2014, Menon et al., 2016] for more details on these computations.

faces the challenge of analyzing the minimum of $N-1$ dependent variables $T^{(k)}_{\epsilon,N,\mathcal{A},\mathcal{E}}$, $1 \leq k \leq N-1$ as $N \to \infty$. Clearly, the distribution of the $\hat{k}$'s such that $T_{\epsilon,N,\mathcal{A},\mathcal{E}}(H) = T^{(\hat{k})}_{\epsilon,N,\mathcal{A},\mathcal{E}}(H)$, plays a central role and in Figure 4 we show the statistics of $\hat{k}$ for both QR and the Toda algorithm.[2] In this paper we further restrict our attention to the Toda algorithm, and as a first step towards understanding $T_{\epsilon,N,\mathrm{Toda},\mathcal{E}}$, we prove universality for $T^{(1)}_{\epsilon,N,\mathrm{Toda},\mathcal{E}}$, the 1-deflation time for Toda — see Theorem 1.1. This is the main result in the paper.

Running Toda until time $T = T^{(1)}(H)$, is, in addition, an algorithm of independent interest. Indeed, as we see from Proposition 1.1, with high probability $X_{11}(T^{(1)}) \sim \lambda_N$, the largest eigenvalue of $X(0) = H$. In other words, $T^{(1)}(H)$ controls the computation of the largest eigenvalue of H via the Toda algorithm. In this paper we always order the eigenvalues $\{\lambda_n\}$ of a real symmetric or Hermitian matrix H as $\lambda_n \leq \lambda_{n+1}$, $n = 1, \ldots, N$. As we will see, analyzing $T^{(1)}(H)$ requires very detailed statistical information on the eigenvalues and eigenvectors of H: Much of this information was only established in the last three or four years. In Sections 1.1 and 1.3 we will describe some of the properties of the Toda algorithm and some results from random matrix theory. In Section 1.2 we describe some numerical results demonstrating Theorem 1.1. Note that Figure 5 for $T^{(1)}(H)$ is very different from Figure 1(b) for $T(H)$. In Sections 2 and 3 we will prove universality for $T^{(1)}_{\epsilon,N,\mathrm{Toda},\mathcal{E}}$ for matrices from generalized Wigner ensembles and also from invariant ensembles.

1.1 Main result

The Toda algorithm is an example of the generalized continuous eigenvalue algorithm described above. For an $N \times N$ real symmetric or Hermitian matrix $X(t) = (X_{ij}(t))_{i,j=1}^N$, the Toda equations are given by[3]

$$\dot{X} = [X, B(X)], \quad B(X) = X_- - (X_-)^*, \quad X(0) = H = H^*, \tag{1.4}$$

where X_- is the strictly lower-triangular part of X and $[A,B]$ is the standard matrix commutator. It is well known that this flow is isospectral and converges as $t \to \infty$ to a diagonal matrix $X_\infty = \mathrm{diag}(\lambda_N, \ldots, \lambda_1)$; see for example [Deift et al., 1985]. As noted above, necessarily, the diagonal elements of X_∞ are the eigenvalues

[2]The asymmetry for the QR histogram reflects the fact that typically H has an eigenvalue near zero: For QR, a simple argument shows that eigenvalues near zero favor $\hat{k} = N-1$.

[3]In the real symmetric case * should be replaced with ^{T}.

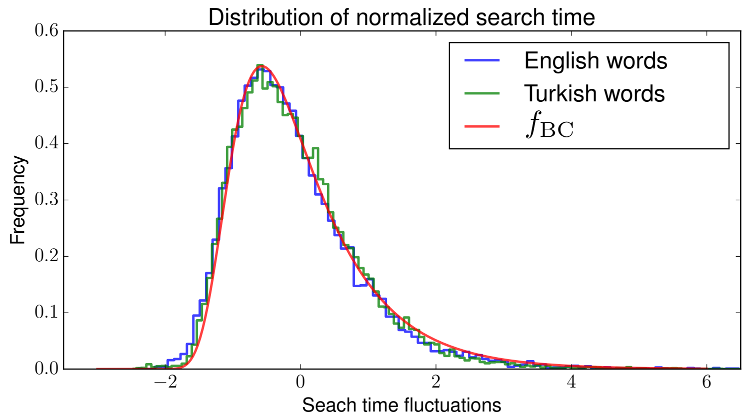
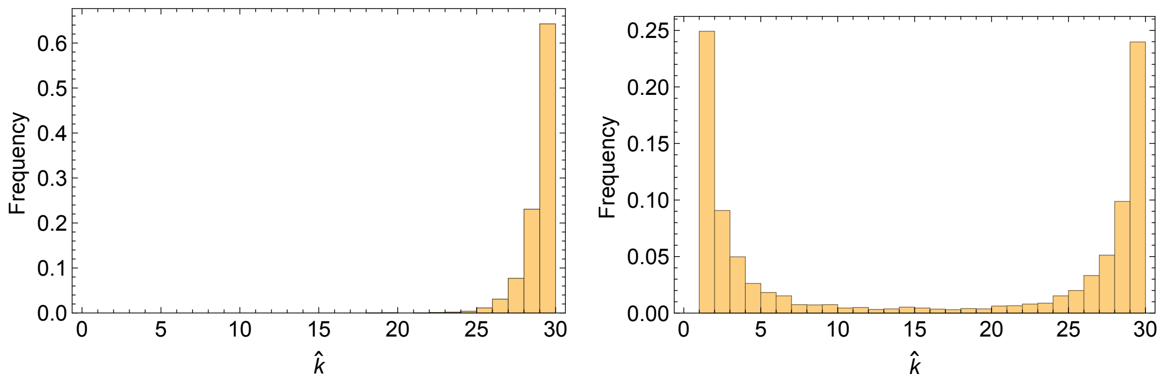


Figure 3: Histograms of the normalized `Google` search times obtained for English and Turkish words obtained in [Sagun et al., 2015]. The solid curve is the normalized distribution f_{BC} of Bakhtin and Correll for the decision times in [Bakhtin and Correll, 2012].



(a) (b)

Figure 4: The distribution of $\hat{k}$ for GOE when $N = 30$, $\epsilon = 10^{-8}$ and (a) $\mathcal{A} = \mathrm{QR}$ and (b) $\mathcal{A} = \mathrm{Toda}$. For QR, $\hat{k} = N - 1$ is most likely, indicating that the last column decays most quickly. For Toda, $\hat{k} = 1, N - 1$ are equally likely.

of H. By the **Toda algorithm** (without deflations) to compute the eigenvalues of a Hermitian matrix H we mean solving (1.4) with $X(0) = H$ until such time t' that the off-diagonal elements in the matrix $X(t)$ are of order ϵ. The eigenvalues of $X(t')$ then give the eigenvalues of H to $\mathcal{O}(\epsilon)$.

The history of the Toda algorithm is as follows. The Toda Lattice was introduced by M. Toda in 1967 [Toda, 1967] and describes the motion of N particles x_i, $i = 1, \ldots, N$, on the line under the Hamiltonian

$$H_{\mathrm{Toda}}(x, y) = \frac{1}{2} \sum_{i=1}^{N} y_i^2 + \frac{1}{2} \sum_{i=1}^{N} \mathrm{e}^{x_i - x_{i+1}}.$$

In 1974, Flaschka [Flaschka, 1974] (see also [Manakov, 1975]) showed that Hamilton's equations

$$\dot{x} = \frac{\partial H_{\mathrm{Toda}}}{\partial y}, \quad \dot{y} = -\frac{\partial H_{\mathrm{Toda}}}{\partial x},$$

can be written in the Lax pair form (1.4) where X is tridiagonal

$$X_{ii} = -y_i/2, \quad 1 \le i \le N,$$
$$X_{i,i+1} = X_{i+1,i} = \frac{1}{2}\mathrm{e}^{\frac{1}{2}(x_i - x_{i+1})}, \quad 1 \le i \le N-1,$$

and $B(X)$ is the tridiagonal skew-symmetric matrix $B(X) = X_- - (X_-)^T$ as in (1.4). As noted above, the flow $t \mapsto X(t)$ is isospectral. But more is true: The flow is completely integrable in the sense of Liouville with the eigenvalues of $X(0) = H$ providing N Poisson commuting integrals for the flow. In 1975, Moser showed that the off-diagonal elements $X_{i,i+1}(t)$ converge to zero as $t \to \infty$ [Moser, 1975]. Inspired by this result, and also related work of Symes [Symes, 1982] on the QR algorithm, the authors in [Deift et al., 1983] suggested that the Toda Lattice be viewed as an eigenvalue algorithm, the Toda algorithm. The Lax equations (1.4) clearly give rise to a global flow not only on tridiagonal matrices but also on general real symmetric matrices. It turns out that in this generality (1.4) is also Hamiltonian [Kostant, 1979, Adler, 1978] and, in fact, integrable [Deift et al., 1986]. From that point on, by the Toda algorithm one means the action of (1.4) on full real symmetric matrices, or by extension, on complex Hermitian matrices.[4]

As noted in the Introduction, in this paper we consider running the Toda algorithm only until time $T = T^{(1)}$, the deflation time with block decomposition $k = 1$ fixed, when the norm of the off-diagonal elements in the first row, and hence the first column, is $\mathcal{O}(\epsilon)$. Define

$$E(t) = \sum_{n=2}^{N} |X_{1n}(t)|^2, \tag{1.5}$$

so that if $E(t) = 0$ then $X_{11}(t)$ is an eigenvalue of H. Thus, with $E(t)$ as in (1.5), the halting time (or 1-deflation time) for the Toda algorithm is given by

$$T^{(1)}(H) = T^{(1)}_{\epsilon,N,\mathcal{A},\mathcal{E}}(H) = \inf\{t : E(t) \le \epsilon^2\}. \tag{1.6}$$

Note that by the min-max principle if $E(t) < \epsilon^2$ then $|X_{11}(t) - \lambda_j| < \epsilon$ for some eigenvalue λ_j of $X(0)$.

In order to state our main result, we refer the reader to Definitions 1.1 and 1.2 for the definition of invariant ensembles (IE) and Wigner ensembles (WE) and we reference Definition 1.4 for the definition of $F_\beta^{\mathrm{gap}}(t)$ ($\beta = 1$ or 2). The constants c_V and b_V are discussed in Theorem 1.2. Our main result is a restated version of Theorem 3.1.

Theorem 1.1 (Universality for the Toda algorithm)**.** *Let $0 < \sigma < 1$ be fixed and let (ϵ, N) be in the scaling region $\frac{\log \epsilon^{-1}}{\log N} \ge \frac{5}{3} + \frac{\sigma}{2}$, see Definition 1.3. Then if H is distributed according to any real ($\beta = 1$) or complex ($\beta = 2$) invariant or Wigner ensemble*

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{T^{(1)}}{c_V^{2/3} 2^{-2/3} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N)} \le t \right) = F_\beta^{\mathrm{gap}}(t). \tag{1.7}$$

To see that the algorithm computes the top eigenvalue, to an accuracy beyond its fluctuations, we have the following proposition which is a restatement of Proposition 3.1 that shows our error is $\mathcal{O}(\epsilon)$ with high probability.

Proposition 1.1 (Computing the largest eigenvalue)**.** *Let (ϵ, N) be in the scaling region. Then if H is distributed according to any real or complex invariant or Wigner ensemble*

$$\epsilon^{-1}|\lambda_N - X_{11}(T^{(1)})|$$

converges to zero in probability as $N \to \infty$. Furthermore, both

$$\epsilon^{-1}|b_V - X_{11}(T^{(1)})|, \quad \epsilon^{-1}|\lambda_j - X_{11}(T^{(1)})|$$

converge to ∞ in probability for any $j = j(N) < N$ as $N \to \infty$.

[4]The Toda flow (1.4) also generates a completely integrable Hamiltonian system on real (not necessarily symmetric) $N \times N$ matrices, see [Deift et al., 1985]. The Toda flow (1.4) on Hermitian matrices was first investigated by Watkins [Watkins, 1984].

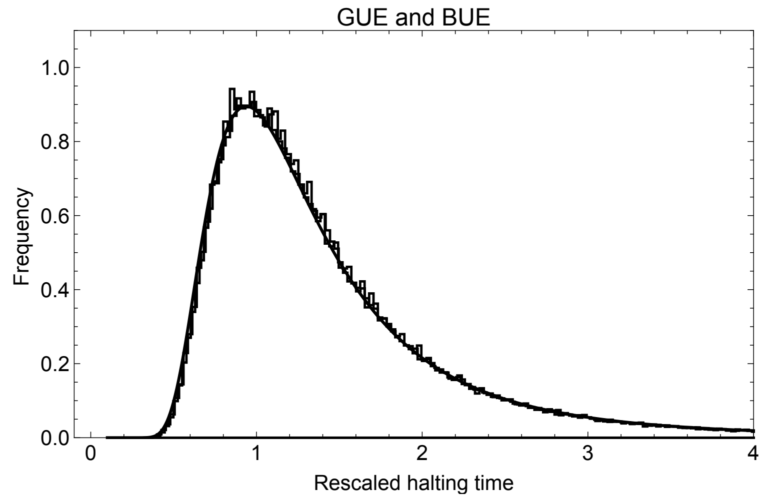


Figure 5: The simulated rescaled histogram for $\tilde{T}^{(1)}$ for both BUE and GUE. Here $\epsilon = 10^{-14}$ and $N = 500$ with 250,000 samples. The solid curve is the rescaled density $f_2^{\mathrm{gap}}(t) = \mathrm{d}/\mathrm{d}tF_2^{\mathrm{gap}}(t)$. The density $f_2^{\mathrm{gap}}(t) = \frac{1}{\sigma t^2} A^{\mathrm{soft}}\left(\frac{1}{\sigma t}\right)$, where $A^{\mathrm{soft}}(s)$ is shown in [Witte et al., 2013, Figure 1]: In order to match the scale in [Witte et al., 2013] our choice of distributions (BUE and GUE) we must take $\sigma = 2^{-7/6}$. This is a numerical demonstration of Theorem 1.1.

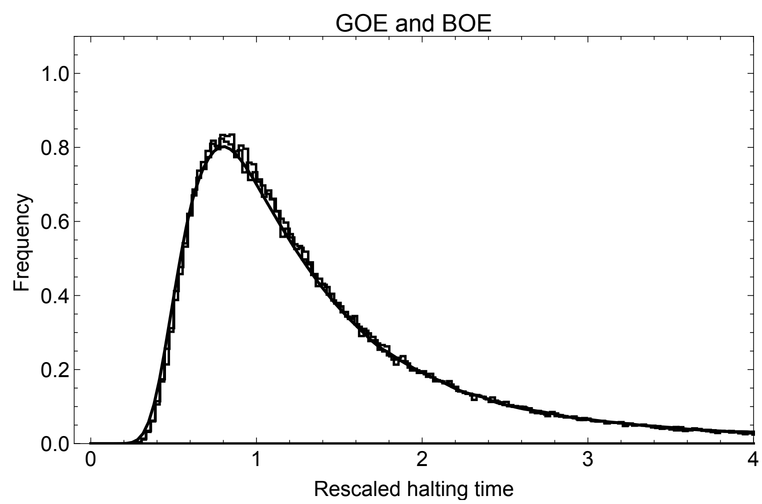


Figure 6: The simulated rescaled histogram for $\tilde{T}^{(1)}$ for both BOE and GOE demonstrating Theorem 1.1. Here $\epsilon = 10^{-14}$ and $N = 500$ with 250,000 samples. The solid curve is an approximation to the density $f_1^{\mathrm{gap}}(t) = \mathrm{d}/\mathrm{d}tF_1^{\mathrm{gap}}(t)$. We compute $f_1^{\mathrm{gap}}(t)$ by smoothing the histogram for $c_V^{-2/3}2^{-2/3}N^{-2/3}(\lambda_N - \lambda_{N-1})$ when $N = 800$ with 500,000 samples.

The relation of this theorem to two-component universality is the following. Let $\xi = \xi_\beta$ be the random variable with distribution $F_\beta^{\mathrm{gap}}(t)$, $\beta = 1$ or 2. For $\beta = 2$ IEs one can prove that[5]

$$\mathbb{E}[T^{(1)}] = c_V^{2/3} 2^{-2/3} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N) \mathbb{E}[\xi](1 + o(1)), \tag{1.8}$$

$$\sqrt{\mathrm{Var}(T^{(1)})} = \kappa c_V^{2/3} 2^{-2/3} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N)(1 + o(1)), \quad \kappa > 0. \tag{1.9}$$

By the Law of Large Numbers, if the number of samples is sufficiently large for any fixed, but sufficiently large N, we can restate the result as

$$\mathbb{P}\left(\frac{T^{(1)} - \langle T^{(1)} \rangle}{\sigma_{T^{(1)}}} \leq t \right) \approx F_\beta^{\mathrm{gap}}(\kappa t + \mathbb{E}[\xi]).$$

This is a universality theorem for the halting time $T^{(1)}$ as the limiting distribution does not depend on the distribution of the individual entries of the matrix ensemble, just whether it is real or complex.

Remark 1.1. *If one constructs matrices $H = U \Lambda U^*$, $\Lambda = \mathrm{diag}(\lambda_N, \lambda_{N-1}, \ldots, \lambda_1)$ where the joint distribution of $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ is given by*

$$\propto \prod_{j=1}^N \mathrm{e}^{-N \frac{\beta}{2} V(\lambda_j)} \prod_{j<n} |\lambda_j - \lambda_n|^\beta,$$

and U is distributed (independently) according to Haar measure on either the orthogonal or unitary group then Theorem 1.1 holds for any $\beta \geq 1$. Here V should satisfy the hypotheses in Definition 1.2.

Remark 1.2. *To compute the largest eigenvalue of H, one can alternatively consider the flow*

$$\dot{X}(t) = H X(t), \quad X(0) = [1, 0, \ldots, 0]^T.$$

It follows that

$$\log \frac{\|X(t+1)\|}{\|X(t)\|} \to \lambda_N, \quad t \to \infty.$$

So, define $T_{\mathrm{ODE}}(H) = \inf \left\{ t : \left| \log \frac{\|X(t+1)\|}{\|X(t)\|} - \lambda_N \right| \leq \epsilon \right\}$. Using the proof technique we present here, one can show that Theorem 1.1 also holds with $T^{(1)}$ replaced with T_{ODE}.

1.2 A numerical demonstration

We can demonstrate Theorem 1.1 numerically using the following WEs defined by letting X_{ij} for $i \leq j$ be iid with distributions:

GUE Mean zero standard complex normal.

BUE $X + \mathrm{i}Y$ where X and Y are each the sum of independent mean zero Bernoulli random variables, *i.e.* binomial random variables.

GOE Mean zero standard (real) normal.

BOE Mean zero Bernoulli random variable

[5] We can also prove (1.8) for $\beta = 1$ IEs. These proofs of these facts require an extension of the level repulsion estimates in [Bourgade et al., 2014, Theorem 3.2] to the case '$K = 1$'. When $\beta = 2$, again with this extension of [Bourgade et al., 2014, Theorem 3.2] to the case '$K = 1$', we can prove that $\kappa = \mathrm{Var}(\xi)$. This extension is known to be true [Bourgade, 2016]. The calculations in Table 1 below are consistent with (1.8) and (1.9) (even for WEs) and lead us to believe that (1.9) also holds for $\beta = 1$. Note that for $\beta = 2$, $\mathbb{E}[\xi^2] < \infty$, but it is believed that $\mathbb{E}[\xi^2] = \infty$ for $\beta = 1$, see [Perret and Schehr, 2014]. In other words, we face the unusual situation where the variance seems to converge, but not to the variance of the limiting distribution.

In Figure 5, for $\beta = 2$, we show how the histogram of $T^{(1)}$ (more precisely, $\tilde{T}^{(1)}$, see (1.10) below), after rescaling, matches the density $\mathrm{d}/\mathrm{d}tF_2^{\mathrm{gap}}(t)$ which was computed numerically[6] in [Witte et al., 2013]. In Figure 6, for $\beta = 1$, we show the histogram for $T^{(1)}$ (again, $\tilde{T}^{(1)}$), after rescaling, matches the density $\mathrm{d}/\mathrm{d}tF_1^{\mathrm{gap}}(t)$. To the best of our knowledge, a computationally viable formula for $\mathrm{d}/\mathrm{d}tF_1^{\mathrm{gap}}(t)$, analogous to $\mathrm{d}/\mathrm{d}tF_2^{\mathrm{gap}}(t)$ in [Witte et al., 2013], is not yet known and so we estimate the density $\mathrm{d}/\mathrm{d}tF_1^{\mathrm{gap}}(t)$ using Monte Carlo simulations with N large. For convenience, we choose the variance for the above ensembles so that $[a_V, b_V] = [-2\sqrt{2}, 2\sqrt{2}]$ which, in turn, implies $c_V = 2^{-3/2}$.

It is clear from the proof of Theorem 1.1 that the convergence of the left-hand side in (1.7) to F_β^{gap} is slow. In fact, we expect a rate proportional to $1/\log N$. This means that in order to demonstrate (1.7) numerically with convincing accuracy one would have to consider very large values of N. In order to display the convergence in (1.7) for more reasonable values of N, we observe, using a simple calculation, that for any fixed $\gamma \neq 0$ the limiting distribution of

$$\tilde{T}^{(1)} = \tilde{T}_\gamma^{(1)} := \frac{T^{(1)}}{c_V^{2/3} 2^{-2/3} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N + \gamma)} \tag{1.10}$$

as $N \to \infty$ is the same as for $\gamma = 0$. A "good" choice for γ is obtained in the following way. To analyze the $T^{(1)}$ in Sections 2 and 3 below we utilize two approximations to $T^{(1)}$, viz. T^* in (2.9) and $\hat{T}$ in (3.1):

$$T^{(1)} = \hat{T} + (T^{(1)} - T^*) + (T^* - \hat{T}).$$

The parameter γ can be inserted into the calculation by replacing $\hat{T}$ with $\hat{T}_\gamma$

$$\hat{T} \to \hat{T}_\gamma := \frac{(\alpha - 4/3)\log N + 2\gamma}{\delta_{N-1}}$$

where γ is chosen to make

$$T^* - \hat{T}_\gamma = \frac{\log N^{2/3}(\lambda_N - \lambda_{N-1}) + \frac{1}{2}\log \nu_{N-1} - \gamma}{\lambda_N - \lambda_{N-1}}, \tag{1.11}$$

as small as possible. Here ν_{N-1} and δ_{N-1} are defined at the beginning of Section 1.4. Replacing $\log N^{2/3}(\lambda_N - \lambda_{N-1})$ and $\log \nu_N$ in (1.11) with the expectation of their respective limiting distributions as $N \to \infty$ (see Theorem 1.3: note that ν_{N-1} is asymptotically distributed as ζ^2 where ζ is Cauchy distributed) we choose $\gamma_2 = -\mathbb{E}(\log(c_V^{2/3} 2^{-5/3}\xi_2)) + \frac{1}{2}\mathbb{E}[\log|\zeta|] \approx 0.883$ when $\beta = 2$ and $\gamma_1 = -\mathbb{E}(\log(c_V^{2/3} 2^{-5/3}\xi_1)) + \frac{1}{2}\mathbb{E}[\log|\zeta|] \approx 0.89$ when $\beta = 1$. Figures 5 and 6 are plotted using γ_1 and γ_2, respectively.

We can also examine the growth of the mean and standard deviation. We see from Table 1 using a million samples and $\epsilon = 10^{-5}$, that the sample standard deviation is on the same order as the sample mean:

$$\sigma_{T^{(1)}} \sim \langle T^{(1)} \rangle \sim N^{2/3}(\log \epsilon^{-1} - 2/3 \log N). \tag{1.12}$$

Remark 1.3. *The ideas that allow us to establish* (1.8) *for IEs requires the convergence of*

$$\mathbb{E}\left[\frac{1}{N^{2/3}(\lambda_N - \lambda_{N-1})}\right]. \tag{1.13}$$

For BUE, (1.13) *must be infinite for all N as there is a non-zero probability that the top two eigenvalues coincide owing to the fact that the matrix entries are discrete random variables. Nevertheless, the sample mean and sample standard deviation of $T^{(1)}$ are observed to converge, after rescaling. It is an interesting open problem to show that convergence in* (1.8) *still holds in this case of discrete WEs even though* (1.13) *is infinite. Specifically, the convergence in the definition of ξ (Definition 1.4) for discrete WEs cannot take place in expectation. Hence $T^{(1)}$ acts as a molified version of the inverse of the top gap — it is always finite.*

[6]Technically, the distribution of the first gap was computed , and then F_2^{gap} can be computed by a change of variables. We thank Folkmar Bornemann for the data to plot F_2^{gap}.

N	50	100	150	200	250	300
$\log \epsilon^{-1}/\log N - 5/3$	1.28	0.833	0.631	0.506	0.418	0.352
$\langle T^{(1)}\rangle \sigma_{T^{(1)}}^{-1}$ for GUE	1.58	1.62	1.59	1.63	1.6	1.58
$\langle T^{(1)}\rangle \sigma_{T^{(1)}}^{-1}$ for BUE	1.6	1.57	1.6	1.62	1.62	1.58
$\langle T^{(1)}\rangle \sigma_{T^{(1)}}^{-1}$ for GOE	0.506	0.701	0.612	0.475	0.705	0.619
$\langle T^{(1)}\rangle \sigma_{T^{(1)}}^{-1}$ for BOE	0.717	0.649	0.663	0.747	0.63	0.708

Table 1: A numerical demonstration of (1.12). The second row of the table confirms that (ϵ, N) is in the scaling region for, say, $\sigma = 1/2$. The last four rows demonstrate that the ratio of the sample mean to the sample standard deviation is order one.

1.3 Estimates from random matrix theory

We now introduce the ideas and results from random matrix theory that are needed to prove Theorem 1.1 and Proposition 1.1. Let H be an $N \times N$ Hermitian (or just real symmetric) matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ and let $\beta_1, \beta_2, \ldots, \beta_N$ denote the absolute value of the first components of the normalized eigenvectors. The following definitions are taken from [Bourgade and Yau, 2013, Deift, 2000].

Definition 1.1 (Generalized Wigner Ensemble (WE))**.** *A generalized Wigner matrix (ensemble) is a real symmetric ($\beta = 1$) or Hermitian ($\beta = 2$) matrix $H = (H_{ij})_{i,j=1}^N$ such that H_{ij} are independent random variables for $i \leq j$ given by a probability measure ν_{ij} with*

$$\mathbb{E}H_{ij} = 0, \quad \sigma_{ij}^2 := \mathbb{E}H_{ij}^2.$$

Next, assume there is a fixed constant v (independent of N, i, j) such that

$$\mathbb{P}(|H_{ij}| > x\sigma_{ij}) \leq v^{-1} \exp(-x^v), \quad x > 0.$$

Finally, assume there exists $C_1, C_2 > 0$ such that for all i, j

$$\sum_{i=1}^N \sigma_{ij}^2 = 1, \quad \frac{C_1}{N} \leq \sigma_{ij}^2 \leq \frac{C_2}{N},$$

and for $\beta = 2$ the matrix

$$\Sigma_{ij} = \begin{bmatrix} \mathbb{E}(\operatorname{Re} H_{ij})^2 & \mathbb{E}(\operatorname{Re} H_{ij})(\operatorname{Im} H_{ij}) \\ \mathbb{E}(\operatorname{Re} H_{ij})(\operatorname{Im} H_{ij}) & \mathbb{E}(\operatorname{Im} H_{ij})^2 \end{bmatrix}$$

has its smallest eigenvalue $\lambda_{\min}$ satisfy $\lambda_{\min} \geq C_1 N^{-1}$.

Definition 1.2 (Invariant Ensemble (IE))**.** *Let $V : \mathbb{R} \to \mathbb{R}$ satisfy $V \in C^4(\mathbb{R})$, $\inf_{x\in\mathbb{R}} V''(x) > 0$ and $V(x) > (2+\delta)\log(1+|x|)$ for sufficiently large x and some fixed $\delta > 0$. Then we define an invariant ensemble[7] to be the set of all $N \times N$ symmetric ($\beta = 1$) or Hermitian ($\beta = 2$) matrices $H = (H_{ij})_{i,j=1}^N$ with probability density*

$$\frac{1}{Z_N} \mathrm{e}^{-N\frac{\beta}{2}\operatorname{tr} V(H)} \mathrm{d}H$$

Here $\mathrm{d}H = \prod_{i\leq j} \mathrm{d}H_{ij}$ if $\beta = 1$ and $\mathrm{d}H = \prod_{i=1}^N \mathrm{d}H_{ii} \prod_{i\leq j} \mathrm{d}\operatorname{Re} H_{ij} \mathrm{d}\operatorname{Im} H_{ij}$ if $\beta = 2$.

Define the averaged empirical spectral measure

$$\mu_N(z) = \mathbb{E}\frac{1}{N}\sum_{i=1}^N \delta(\lambda_i - z),$$

where the expectation is taken with respect to the given ensemble.

[7]This is not the most general class of V but these assumptions simplify the analysis.

Theorem 1.2 (Equilibrium measure, [Bourgade et al., 2014]). *For any WE or IE the measure μ_N converges weakly to a measure μ, called the equilibrium measure, which has support on a single interval $[a_V, b_V]$ and has a density ρ that satisfies $\rho(x) \leq C_\mu \sqrt{b_V - x}\chi_{(-\infty,b_V]}(x)$ and $\rho(x) = \frac{2^{3/4}c_V}{\pi}\sqrt{b_V - x}(1 + \mathcal{O}(b_V - x))$ as $x \to b_V$.*

With the chosen normalization for WEs, $\sum_{i=1}^N \sigma_{ij}^2 = 1$, $[a_v, b_V] = [-2, 2]$ and $c_V = 1$ [Bourgade et al., 2014]. One can vary the support as desired by shifting and scaling, $H \to aH + bI$: the constant c_V then changes accordingly. When the entries of H are distributed according to a WE or an IE with high probability (see Theorem 1.4) the top three eigenvalues are distinct and $\beta_j \neq 0$ for $j = N, N-1, N-2$. Next, let $d\mu$ denote the limiting spectral density or equilibrium measure for the ensemble as $N \to \infty$. Then define γ_n to be the smallest value of t such that

$$\frac{n}{N} = \int_{-\infty}^{t} \mathrm{d}\mu.$$

Thus $\{\gamma_n\}$ represent the quantiles of the equilibrium measure.

There are four fundamental parameters involved in our calculations. First we fix $0 < \sigma < 1$ once and for all, then we fix $0 < p < 1/3$, then we choose $s < \min\{\sigma/44, p/8\}$ and then finally $0 < c \leq 10/\sigma$ will be a constant that will allow us to estimate the size of various sums. The specific meanings of the first three parameters is given below. Also, C will a generic constant that can depend on σ or p but not on s or N. We also make statements that will be "true for N sufficiently large". This should be taken to mean that there exists $N^* = N^*(\mu, \sigma, s, p)$ such that the statement is true for $N > N^*$.

Definition 1.3 (Scaling region). *Fix $0 < \sigma < 1$. The scaling region for ϵ is given by $\dfrac{\log \epsilon^{-1}}{\log N} \geq 5/3 + \sigma/2$.*

For convenience in what follows we use the notation $\epsilon = N^{-\alpha/2}$, so

$$(\epsilon, N) \text{ are in the scaling region if and only if } \alpha - 10/3 \geq \sigma > 0$$

and $\alpha = \alpha_N$ is allowed to vary with N.

Condition 1.1. *For $0 < p < 1/3$,*

- $\lambda_{N-1} - \lambda_{N-2} \geq p(\lambda_N - \lambda_{N-1})$.

Let $G_{N,p}$ denote the set of matrices that satisfy this condition.

Condition 1.2. *For any fixed $0 < s < \min\{\sigma/44, p/8\}$*

1. *$\beta_n \leq N^{-1/2+s/2}$ for all n*
2. *$N^{-1/2-s/2} \leq \beta_n$ for $n = N, N-1$,*
3. *$N^{-2/3-s} \leq \lambda_N - \lambda_{n-1} \leq N^{-2/3+s}$, for $n = N, N-1$, and*
4. *$|\lambda_n - \gamma_n| \leq N^{2/3+s}(\min\{n, N-n+1\})^{-1/3}$ for all n.*

Let $R_{N,s}$ denote the set of matrices that satisfy these conditions.

Remark 1.4. *It is known that the distribution (Haar measure on the unitary or orthogonal group) on the eigenvectors for IEs depends only on $\beta = 1, 2$. And, if $V(x) = x^2$ the IE is also a WE. Therefore, if one can prove a general statement about the eigenvectors for WEs then it must also hold for IEs. But, it should be noted that stronger results can be proved for the eigenvectors for IEs, see [Stam, 1982] and [Jiang, 2006] for example.*

The following theorem has it roots in the pursuit of proving universality in random matrix theory. See [Tracy and Widom, 1994] for the seminal result when $V(x) = x^2$ and $\beta = 2$. Further extensions include the works of Soshnikov [Soshnikov, 1999] and Tao and Vu [Tao and Vu, 2010] for Wigner ensembles and [Deift and Gioev, 2007] for invariant ensembles.

Theorem 1.3. *For both IEs and WEs*

$$N^{1/2}(|\beta_N|, |\beta_{N-1}|, |\beta_{N-2}|)$$

converges jointly in distribution to $(|X_1|, |X_2|, |X_3|)$ *where* $\{X_1, X_2, X_3\}$ *are iid real* ($\beta = 1$) *or complex* ($\beta = 2$) *standard normal random variables. Additionally, for IEs and WEs*

$$2^{-2/3}N^{2/3}(b_V - \lambda_N, b_V - \lambda_{N-1}, b_V - \lambda_{N-2})$$

converges jointly in distribution to random variables $(\Lambda_{1,\beta}, \Lambda_{2,\beta}, \Lambda_{3,\beta})$ *which are the smallest three eigenvalues of the so-called stochastic Airy operator. Furthermore,* $(\Lambda_{1,\beta}, \Lambda_{2,\beta}, \Lambda_{3,\beta})$ *are distinct with probability one.*

Proof. The first claim follows from [Bourgade et al., 2014, Theorem 1.2]. The second claim follows from [Bourgade and Yau, 2013, Corollary 2.2 & Theorem 2.7]. The last claim follows from [Ramírez et al., 2011, Theorem 1.1]. □

Definition 1.4. *The distribution function* $F_\beta^{\text{gap}}(t)$ *for* $\beta = 1, 2$ *is given by*

$$F_\beta^{\text{gap}}(t) = \mathbb{P}\left(\frac{1}{\Lambda_{2,\beta} - \Lambda_{1,\beta}} \le t\right) = \lim_{N\to\infty} \mathbb{P}\left(\frac{1}{c_V^{2/3} 2^{-2/3} N^{2/3}(\lambda_N - \lambda_{N-1})} \le t\right), \quad t \ge 0.$$

Properties of $G_\beta(t) := 1 - F_\beta^{\text{gap}}(1/t)$, the distribution function for the first gap, are examined in [Perret and Schehr, 2014, Witte et al., 2013, Monthus and Garel, 2013] including the behavior of $G_\beta(t)$ near $t = 0$ which is critical for understanding which moments of $F'_\beta(t)$ exist.

The remaining theorems in this section are compiled from results that have been obtained recently in the literature.

Theorem 1.4. *For WEs or IEs Condition 1.2 holds with high probability as* $N \to \infty$*, that is, for any* $s > 0$

$$\mathbb{P}(R_{N,s}) = 1 + o(1),$$

as $N \to \infty$.

Proof. We first consider WEs. The fact that the probability of Condition 1.2.1 tends to unity follows from [Erdős et al., 2012, Theorem 2.1] using estimates on the (1,1) entry of the Green's function. See [Erdős, 2012, Section 2.1] for a discussion of using these estimates. The fact that the probability of each of Condition 1.2.2-3 tends to unity follows from Theorem 1.3 using Corollary 3.1. And finally, the statement that the probability Condition 1.2.4 tends to unity as $N \to \infty$ is the statement of the rigidity of eigenvalues, the main result of [Erdős et al., 2012]. Following Remark 1.4, we then have that the probability of Condition 1.2.1-2 tends to unity for IEs.

For IEs, the fact that the probability of Condition 1.2.4 tends to unity follows from [Bourgade and Yau, 2013, Theorem 2.4]. Again, the fact that the probability of Condition 1.2.3 tends to unity follows from Theorem 1.3 using Corollary 3.2. □

Theorem 1.5. *For both WEs and IEs*

$$\lim_{p\downarrow 0} \limsup_{N\to\infty} \mathbb{P}(G_{N,p}^c) = 0.$$

Proof. It follows from Theorem 1.3 that

$$\limsup_{N\to\infty} \mathbb{P}(G_{N,p}^c) = \lim_{N\to\infty} \mathbb{P}(\lambda_{N-1} - \lambda_N < p(\lambda_N - \lambda_{N-1})) = \mathbb{P}(\Lambda_{3,\beta} - \Lambda_{2,\beta} < p(\Lambda_{2,\beta} - \Lambda_{1,\beta})).$$

Then

$$\lim_{p\downarrow 0} \mathbb{P}(\Lambda_{3,\beta} - \Lambda_{2,\beta} < p(\Lambda_{2,\beta} - \Lambda_{1,\beta})) = \mathbb{P}\left(\bigcap_{p>0} \{\Lambda_{3,\beta} - \Lambda_{2,\beta} < p(\Lambda_{2,\beta} - \Lambda_{1,\beta})\}\right) = \mathbb{P}(\Lambda_{3,\beta} = \Lambda_{2,\beta}).$$

But from [Ramírez et al., 2011, Theorem 1.1] $\mathbb{P}(\Lambda_{3,\beta} = \Lambda_{2,\beta}) = 0$. □

Throughout what follows we assume we are given a WE or an IE.

1.4 Technical lemmas

Define $\delta_j = 2(\lambda_N - \lambda_j)$ and $I_c = \{1 \le n \le N-1 : \delta_n/\delta_{N-1} \ge 1+c\}$ for $c > 0$.

Lemma 1.1. *Let $0 < c < 10/\sigma$. Given Condition 1.2*

$$|I_c^c| \le N^{2s}$$

for N sufficiently large, where ^{c} denotes the compliment relative to $\{1, \dots, N-1\}$.

Proof. We use rigidity of the eigenvalues, Condition 1.2.4. So, $|\lambda_n - \gamma_n| \le N^{-2/3+s}(\hat{n})^{-1/3}$ where $\hat{n} = \min\{n, N-n+1\}$. Recall

$$I_c^c \subset \{1 \le n \le N-1 : \lambda_N - \lambda_n < (1+c)(\lambda_N - \lambda_{N-1})\}.$$

Define

$$J_c = \{1 \le n \le N-1 : \gamma_N - \gamma_n \le (2 + c + (\hat{n})^{-1/3})N^{-2/3+s}\}.$$

If $n \in I_c$ then

$$\begin{aligned} \lambda_N - \lambda_n &\le (1+c)N^{-2/3+s}, \\ \gamma_N - N^{-2/3+s} - (\gamma + (\hat{n})^{-1/3}N^{-2/3+s}) \le \lambda_N - \lambda_n &\le (1+c)N^{-2/3+s}, \\ \gamma_N - \gamma_n &\le (2 + c + (\hat{n})^{-1/3})N^{-2/3+s}, \end{aligned}$$

and hence $n \in J_c$. Then compute the asymptotic size of the set J_c let n^* be the smallest element of J_c. Then $|J^*| = N - n^*$ so that

$$\frac{n^*}{N} = \int_{-\infty}^{\gamma_{n^*}} \mathrm{d}\mu, \quad |I_c^c| \le |J_c| = N - n^* = N \int_{n^*}^{\infty} \mathrm{d}\mu.$$

The using Definition 1.2, $\gamma_N = b_V$ and $n* \ge b_V - (2 + c + (\hat{n})^{-1/3})N^{-2/3+s} \ge b_V - (3+c)N^{-2/3+s}$ to see

$$|I_c^c| \le N \int_{n^*}^{\infty} \mathrm{d}\mu \le C_\mu N \int_{b_V - (3+c)N^{-2/3+s}}^{b_V} \sqrt{b_V - x}\, \mathrm{d}x = \frac{2C_\mu}{3}(3+c)^{3/2} N^{3s/2}.$$

and then because σ is fixed hence c has an upper bound and $s > 0$, $|I_c^c| \le N^{2s}$ for sufficiently large N. ◻

We use the notation $\nu_n = \beta_n^2/\beta_N^2$ and note that for a matrix in $R_{N,s}$ we have $\nu_n \le N^{2s}$ and $\sum_n \nu_n = \beta_N^{-2} \le N^{1+s}$. One of the main tasks that will follow is estimating the following sums.

Lemma 1.2. *Given Condition 1.2, $0 < c \le 10/\sigma$ and $j \le 3$ there exists an absolute constant C such that*

$$N^{-2s}\delta_{N-1}^j \mathrm{e}^{-\delta_{N-1}t} \le \sum_{n=1}^{N-1} \nu_n \delta_n^j \mathrm{e}^{-\delta_n t} \le C\mathrm{e}^{-\delta_{N-1}t}\left(N^{4s}\delta_{N-1}^j + N^{1+s}\mathrm{e}^{-c\delta_{N-1}t}\right),$$

for N sufficiently large.

Proof. For $j \le 3$

$$\begin{aligned} \sum_{n=1}^{N-1} \nu_n \delta_n^j \mathrm{e}^{-\delta_n t} &= \left(\sum_{n \in I_c} + \sum_{n \in I_c^c}\right) \nu_n \delta_n^j \mathrm{e}^{-\delta_n t} \\ &\le \sum_{n \in I_c^c} \nu_n (1+c)^j \delta_{N-1}^j \mathrm{e}^{-\delta_{N-1}t} + 2^j \sum_{n \in I_c} \nu_n |\lambda_1 - \lambda_N|^j \mathrm{e}^{-(1+c)\delta_{N-1}t}. \end{aligned}$$

It also follows that $\lambda_N - \lambda_1 \le b_V - a_V + 1$ so that by Lemma 1.1 for sufficiently large N

$$\sum_{n=1}^{N-1} \nu_n \delta_n^j \mathrm{e}^{-\delta_n t} \le C \mathrm{e}^{-\delta_{N-1} t} \left(N^{4s} \delta_{N-1}^j + N^{1+s} \mathrm{e}^{-c\delta_{N-1} t} \right).$$

To find a lower bound, we just keep the first term, as that should be the largest

$$\sum_{n=1}^{N-1} \nu_n \delta_n^j \mathrm{e}^{-\delta_n t} \ge \nu_{N-1} \delta_{N-1}^j \mathrm{e}^{-\delta_{N-1} t} \ge N^{-2s} \delta_{N-1}^j \mathrm{e}^{-\delta_{N-1} t}.$$

□

2 Estimates for the Toda algorithm

Remarkably, (1.4) can be solved explicitly by a QR factorization procedure, see for example [Symes, 1982]. For $X(0) = H$ we have for $t \ge 0$

$$\mathrm{e}^{tH} = Q(t)R(t),$$

where Q is orthogonal ($\beta = 1$) or unitary ($\beta = 2$) and R has positive diagonal entries. This *QR factorization* for e^{tH} is unique: Note that $Q(t)$ is obtained by apply Gram–Schmidt to the columns of e^{tH}. We claim that $X(t) = Q^*(t)HQ(t)$ is the solution of (1.4). Indeed, by differentiating, we obtain

$$\begin{aligned} H\mathrm{e}^{tH} &= HQ(t)R(t) = \dot{Q}(t)R(t) + Q(t)\dot{R}(t), \\ X(t) &= Q^*(t)\dot{Q}(t) + \dot{R}(t)R^{-1}(t). \end{aligned} \tag{2.1}$$

Then because $\dot{R}(t)R^{-1}(t)$ is upper triangular

$$(X(t))_- = (Q^*(t)\dot{Q}(t))_-.$$

Furthermore, from $Q^*(t)Q(t) = I$ we have $Q^*(t)\dot{Q}(t) = -\dot{Q}^*(t)Q(t)$ so that $Q^*(t)\dot{Q}(t)$ is skew Hermitian. Thus, $B(X(t)) = Q^*(t)\dot{Q}(t) - [Q^*(t)\dot{Q}(t)]_D$, where $[\cdot]_D$ gives the diagonal part of the matrix. However, as $Q^*(t)\dot{Q}(t)$ is skew Hermitian $[Q^*(t)\dot{Q}(t)]_D$ is purely imaginary. On the other hand, we see from (2.1) that the diagonal is real. It follows that $[Q^*(t)\dot{Q}(t)]_D = 0$ and $B(X(t)) = Q^*(t)\dot{Q}(t)$. Using (1.4) we have

$$\dot{X}(t) = \dot{Q}^*(t)HQ(t) + Q^*(t)H\dot{Q}(t),$$

and so

$$\dot{X}(t) = X(t)B(X(t)) - B(X(t))X(t).$$

When $t = 0$, $Q(0) = I$ so that $X(0) = H$ and by uniqueness for ODEs this shows $X(t)$ is indeed the solution of (1.4).

As the eigenvalues of $X(0) = H$ are not necessarily simple (indeed for BOE there is a non-zero probability for a matrix to have repeated eigenvalues), it is not clear a priori that the eigenvectors of $X(t)$ can be chosen to be smooth functions of t. However, for the case at hand we can proceed in the following way. For $X(0) = H$ there exists a (not necessarily unique) unitary matrix U_0 such that $X(0) = U_0 \Lambda U_0^*$ where $\Lambda = \mathrm{diag}(\lambda_N, \ldots, \lambda_1)$. Then $X(t) = Q^*(t)HQ(t) = U(t)\Lambda U^*(t)$ where $U(t) = Q^*(t)U_0$. Then the jth column $u_j(t)$ of $U(t)$ is a smooth eigenvector of $X(t)$ corresponding to eigenvalue λ_j. From the eigenvalue equation

$$(X(t) - \lambda_j)u_j(t) = 0,$$

we obtain (following Moser [Moser, 1975])

$$
\begin{aligned}
\dot{X}(t)u_j(t) + (X(t) - \lambda_j)\dot{u}_j(t) = 0,\\
(X(t)B(X(t)) - B(X(t))X(t))u_j(t) + (X(t) - \lambda_j)\dot{u}_j(t) = 0,\\
(X(t) - \lambda_j)[\dot{u}_j(t) + B(X(t))u_j(t)] = 0.
\end{aligned}
$$

This last equation implies $\dot{u}_j(t) + B(X(t))u_j$ must be a (possibly time-dependent) linear combination of the eigenvectors corresponding to λ_j. Let $U_j(t) = [u_{j_1}(t), \ldots, u_{j_m}(t)]$ be eigenvectors corresponding to a repeated eigenvalue λ_j so that for $i = 1, \ldots, m$

$$
\dot{u}_{j_i}(t) + B(X(t))u_{j_i}(t) = \sum_{k=1}^{m} d_{ki}(t)u_{j_k}(t),
$$

and so

$$
\left[\frac{d}{dt} + B(X(t))\right] U_j(t) = U_j(t)D(t), \quad D(t) = (d_{ki}(t))_{k,i=1}^{m}. \tag{2.2}
$$

Note that $U_j^*(t)U_j(t) = I_m$, the $m \times m$ identity matrix. Then multiplying (2.2) on the left by $U_j^*(t)$ and then multiplying the conjugate transpose of (2.2) on the right by $U_j(t)$, we obtain

$$
\begin{aligned}
U_j^*(t)\dot{U}_j(t) + U_j^*(t)B(X(t))U_j(t) = D(t),\\
\dot{U}_j^*(t)U_j(t) + U_j^*(t)[B(X(t))]^*U_j(t) = D^*(t).
\end{aligned}
$$

Because $\mathrm{d}/\mathrm{d}t[U_j^*(t)U_j(t)] = 0$ and $B(X(t))$ is skew Hermitian, the addition of these two equations gives $D(t) = -D^*(t)$. Let $S(t)$ be the solution of $\dot{S}(t) = -D(t)S(t)$ with $S(0) = I_m$. Then $\mathrm{d}/\mathrm{d}t[S^*(t)S(t)] = -S^*(t)D(t)S(t) + S^*(t)D(t)S(t) = 0$ and hence $S^*(t)S(t) = C = I_m$, *i.e.*, $S(t)$ is unitary. In particular, $\tilde{U}_j(t) := U_j(t)S(t)$ has orthonormal columns and we find

$$
\left[\frac{d}{dt} + B(X(t))\right] \tilde{U}_j(t) = U_j(t)D(t)S(t) - U_j(t)D(t)S(s) = 0.
$$

We see that a smooth normalization for the eigenvectors of $X(t)$ can always be chosen so that $D(t) = 0$. Without loss of generality, we can assume that $U(t)$ solves (2.2) with $D(t) = 0$. Then for $U(t) = (U_{ij}(t))_{i,j=1}^{N}$

$$
\begin{aligned}
\dot{U}_{1j}(t) &= -e_1^*B(X(t))u_j(t) = (B(X(T))e_1)^*u_j(t) = (X(t)e_1 - X_{11}^*(t)e_1)^*u_j(t)\\
&= e_1^*(X(t) - X_{11}(t))u_j(t) = (\lambda_j - X_{11}(t))U_{1j}(t).
\end{aligned}
$$

A direct calculation using

$$
X_{11}(t) = e_1^*X(t)e_1 = \sum_{j=1}^{N} \lambda_j |U_{1j}(t)|^2,
$$

shows that

$$
U_{1j}(t) = \frac{U_{1j}(0)\mathrm{e}^{\lambda_j t}}{\left(\displaystyle\sum_{j=1}^{N} |U_{1j}(0)|^2 \mathrm{e}^{2\lambda_j t}\right)^{1/2}}, \quad 1 \le j \le N.
$$

Also

$$
X_{1k}(t) = \sum_{j=1}^{N} \lambda_j U_{1j}^*(t)U_{kj}(t),
$$

and hence

$$\sum_{k=2}^{N}|X_{1k}(t)|^2 = \sum_{k=2}^{N}X_{1k}(t)X_{k1}(t) = [X^2(t)]_{11} - X_{11}^2(t)$$
$$= \sum_{k=1}^{N}\lambda_k^2|U_{1k}(t)|^2 - \left(\sum_{k=1}^{N}\lambda_k|U_{1k}(t)|^2\right)^2 = \sum_{k=1}^{N}(\lambda_k - X_{11}(t))^2|U_{1k}(t)|^2.$$

Thus

$$E(t) := \sum_{k=2}^{N}|X_{1k}(t)|^2 = \sum_{j=1}^{N}(\lambda_j - X_{11}(t))^2|U_{1j}(t)|^2.$$

We also note that

$$\lambda_N - X_{11}(t) = \sum_{j=1}^{N}(\lambda_N - \lambda_j)|U_{1j}(t)|^2.$$

From these calculations, if $U_{11}(0) \neq 0$, it follows that

$$\frac{X_{11}(t) - \lambda_N}{E(t)} \to 0, \quad N \to \infty.$$

While $X_{11}(t) - \lambda_N$ is of course the true error in computing λ_N we use $E(t)$ to determine a convergence criterion as it is easily observable: Indeed, as noted above, if $E(t) < \epsilon$ then $|X_{11}(t) - \lambda_j| < \epsilon$, for some j. With high probability, $\lambda_j = \lambda_N$.

Note that, in particular, from the above formulae, $E(t)$ and $\lambda_N - X_{11}(t)$ depend **only** on the eigenvalues and the moduli of the first components of the eigenvectors of $X(0) = H$. This fact is critical to our analysis. With the notation $\beta_j = |U_{1j}(0)|$ we have that

$$|U_{1j}(t)| = \frac{\beta_j \mathrm{e}^{\lambda_j t}}{\left(\sum_{n=1}^{N}\beta_n^2\mathrm{e}^{2\lambda_n t}\right)^{1/2}}.$$

A direct calculation shows that

$$E(t) = E_0(t) + E_1(t),$$

where

$$E_0(t) = \frac{1}{4}\frac{\sum_{n=1}^{N-1}\delta_n^2\nu_n\mathrm{e}^{-\delta_n t}}{\left(1 + \sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right)^2}.$$

$$E_1(t) = \frac{\left(\sum_{n=1}^{N-1}\lambda_n^2\nu_n\mathrm{e}^{-\delta_n t}\right)\left(\sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right) - \left(\sum_{n=1}^{N-1}\lambda_n\nu_n\mathrm{e}^{-\delta_n t}\right)^2}{\left(1 + \sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right)^2}.$$

Note that $E_1(t) \geq 0$ by the Cauchy–Schwarz inequality; of course, $E_0(t)$ is trivially positive. It follows that $E(t)$ is small if and only if both $E_0(t)$ and $E_1(t)$ are small, a fact that is extremely useful in our analysis.

In terms of the probability ρ_N measure on $\{1, 2, \ldots, N\}$ defined by

$$\rho_N(E) = \left(\sum_{n=1}^{N} \nu_n \mathrm{e}^{-\delta_n t} \right)^{-1} \sum_{n \in E} \nu_n \mathrm{e}^{-\delta_n t},$$

and a function $\lambda(j) = \lambda_j$

$$E(t) = \mathrm{Var}_{\rho_N}(\lambda).$$

We will also use the alternate expression

$$E_1(t) = \left(\frac{\sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n t}}{1 + \sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n t}} \right)^2 \mathrm{Var}_{\rho_{N-1}}(\lambda). \tag{2.3}$$

Additionally,

$$\lambda_N - X_{11}(t) = \frac{1}{2} \frac{\sum_{n=1}^{N} \delta_n \beta_n^2 \mathrm{e}^{2\lambda_n t}}{1 + \sum_{n=1}^{N-1} \beta_n^2 \mathrm{e}^{2\lambda_n t}}. \tag{2.4}$$

2.1 The halting time and its approximation

Here we present a list of useful inequalities and definitions that are used extensively throughout what follows:

1. $0 < \sigma < 1$ is fixed,
2. $0 < p < 1/3$,
3. $\alpha \geq 10/3 + \sigma$,
4. $s \leq \min\{\sigma/44, p/8\}$,
5. $\alpha - 4/3 - ns \geq 2$ if $n \leq 44$,
6. $c \leq 10/\sigma$ can be chosen for convenience line by line when estimating sums with Lemma 1.2,
7. $\delta_n = 2(\lambda_N - \lambda_n)$,
8. $\nu_n = \beta_n^2/\beta_N^2$,
9. given Condition 1.2
   - $2N^{-2/3-s} \leq \delta_{N-1} \leq 2N^{-2/3+s}$,
   - $N^{-2s} \leq \nu_n \leq N^{2s}$,
   - $\sum_{n=1}^{j} \nu_n \leq \sum_{n=1}^{N} \nu_n = \beta_N^{-2} \leq N^{1+s}$, for $1 \leq j \leq N$, and
10. $C > 0$ is a generic constant.

Definition 2.1. *The halting time (or the* 1*-deflation time) for the Toda lattice (compare with* (1.6)*) is defined to be*

$$T^{(1)} = \inf\{t : E(t) \leq \epsilon^2\}.$$

We find bounds on the halting time.

Lemma 2.1. *Given Condition 1.2, the halting time T for the Toda lattice satisfies*

$$(\alpha - 4/3 - 5s)\log N/\delta_{N-1} \le T^{(1)} \le (\alpha - 4/3 + 7s)\log N/\delta_{N-1},$$

for sufficiently large N.

Proof. We use that $E(t) \ge E_0(t)$ so if $E_0(t) > N^{-\alpha}$ then $T^{(1)} \ge t$. First, we show that $E_0(t) > \epsilon^2$, $0 \le t \le \sigma/2 \log N/\delta_{N-1}$ and sufficiently large N and then we use this to show that $E_0(t) > \epsilon^2$, $t \le (\alpha - 4/3 - 5s)\log N/\delta_{N-1}$ and sufficiently large N.

Indeed, assume $t = a \log N/\delta_{N-1}$ for $0 \le a \le \sigma/2$. Using Lemma 1.2

$$1 + \sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n t} \le 1 + C\mathrm{e}^{-\delta_{N-1}t}\left(N^{4s} + N^{1+s}\mathrm{e}^{-c\delta_{N-1}t}\right). \tag{2.5}$$

Then using Lemma 1.2 we have

$$E_0(t) \ge N^{-2s}\delta_{N-1}^2 \mathrm{e}^{-\delta_{N-1}t}\left(1 + C\mathrm{e}^{-\delta_{N-1}t}\left(N^{4s} + N^{1+s}\mathrm{e}^{-c\delta_{N-1}t}\right)\right)^{-2}.$$

Since $a \le \sigma/2$ and we find

$$E_0(t) \ge N^{-4s-4/3-\sigma/2}\left(1 + C(N^{4s} + N^{1+s})\right)^{-2} \ge CN^{-8s-10/3-\sigma/2},$$

for some new constant $C > 0$. This last inequality follows because $N^{4s} \le N^{1+s}$ as $s \le 1/44$ (see Condition 1.2). But then from Definition 1.3 this right-hand side is larger than $\epsilon^2 = N^{-\alpha}$ for sufficiently large N. Now, assume $t = a \log N/\delta_{N-1}$ for $\sigma/2 \le a \le (\alpha - 4/3 - 5s)\log N/\delta N - 1$. We choose $c = 2(2+s)/\sigma \le 10/\sigma$

$$\begin{aligned} E_0(t) &\ge \frac{1}{4}N^{-4s-4/3-a}\left(1 + C(N^{4s-a} + N^{1+s-ca})\right)^{-2} \\ &\ge N^{-\alpha+s}(1 + C(N^{4s-\sigma/2} + N^{-1})) > N^{-\alpha} \end{aligned}$$

for sufficiently large N. Here we used that $s \le \sigma/44$. This shows $(\alpha - 4/3 - 5s)\log N/\delta_{N-1} \le T^{(1)}$ for N sufficiently large.

Now, we work on the upper bound. Let $t = a\log N/\delta_{N-1}$ for $a \ge (\alpha - 4/3 + 7s)$ and we find using Lemma 1.2

$$E_0(t) \le CN^{-a}\left(N^{-4/3+6s} + N^{1+s-ca}\right).$$

Then using the minimum value for a

$$E_0(t) \le N^{-\alpha}\left(C(N^{-s} + CN^{1+7s-ca+4/3})\right).$$

It follows from Definition 1.3 that $a \ge 10/3 + \sigma - 4/3 + 7s > 2$. If we set $c = 2$ and use $s \le 1/44$ then $1 + 7s - ca + 4/3 \le -3 + 4/3 + 7s \le -2$

$$E_0(t) \le N^{-\alpha}\left(C(N^{-s} + CN^{-2})\right) < CN^{-\alpha-s}$$

for sufficiently large N.

Next, we must estimate $E_1(t)$ when $a \ge (\alpha - 4/3 + 7s)$. We use (2.3) and $\mathrm{Var}_{N-1}(\lambda) \le C$. Then by (2.5)

$$E_1(t) \le CN^{-2a}(N^{4s} + N^{1+s-ca})^2.$$

Again, using $c = 1$ and the fact that $a > 2$ we have

$$E_1(t) \le CN^{-\alpha}N^{8s-\alpha+8/3-14s} \le CN^{-\alpha}N^{-\alpha+8/3} \le N^{-\alpha} \tag{2.6}$$

for N sufficiently large. This shows $T^{(1)} \le (\alpha - 4/3 + 7s)\log N/\delta_{N-1}$ for sufficiently large N as $E(t) = E_0(t) + E_1(t) \le \epsilon^2$ if $t < (\alpha - 4/3 + 7s)\log N/\delta_{N-1}$ and N is sufficiently large. $\square$

In light of this lemma we define

$$I_\alpha = [(\alpha - 4/3 - 5s)\log N/\delta_{N-1}, (\alpha - 4/3 + 7s)\log N/\delta_{N-1}].$$

Next, we estimate the derivative of $E_0(t)$. We find

$$E_0'(t) = \frac{-\left(\sum_{n=1}^{N-1}\delta_n^3\nu_n\mathrm{e}^{-\delta_n t}\right)\left(1+\sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right) + 2\left(\sum_{n=1}^{N-1}\delta_n^2\nu_n\mathrm{e}^{-\delta_n t}\right)\left(\sum_{n=1}^{N-1}\delta_n\nu_n\mathrm{e}^{-\delta_n t}\right)}{\left(1+\sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right)^3}. \tag{2.7}$$

Lemma 2.2. *Given Condition 1.2 and $t \in I_\alpha$*

$$-E_0'(t) \geq CN^{-12s-\alpha-2/3},$$

for sufficiently large N.

Proof. We use (2.7). The denominator is bounded below by unity so we estimate the numerator. By Lemma 1.2

$$\left(\sum_{n=1}^{N-1}\delta_n^3\nu_n\mathrm{e}^{-\delta_n t}\right)\left(1+\sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right) \geq \sum_{n=1}^{N-1}\delta_n^3\nu_n\mathrm{e}^{-\delta_n t} \geq N^{-2s}\delta_{N-1}^3\mathrm{e}^{-\delta_{N-1}t}.$$

For $t \in I_\alpha$

$$N^{-2s}\delta_{N-1}^3\mathrm{e}^{-\delta_{N-1}t} \geq N^{-12s-2/3-\alpha}.$$

Next, again by Lemma 1.2

$$\left(\sum_{n=1}^{N-1}\delta_n\nu_n\mathrm{e}^{-\delta_n t}\right)\left(\sum_{n=1}^{N-1}\delta_n^2\nu_n\mathrm{e}^{-\delta_n t}\right) \leq C\mathrm{e}^{-2\delta_{N-1}t}\left(N^{4s}\delta_{N-1}^2 + N^s\mathrm{e}^{-c\delta_{N-1}t}\right)\left(N^{4s}\delta_{N-1} + N^{1+s}\mathrm{e}^{-c\delta_{N-1}t}\right).$$

Then estimate with $c = 2$,

$$\begin{aligned} N^{4s}\delta_{N-1}^2 + N^s\mathrm{e}^{-c\delta_{N-1}t} &\leq 4N^{6s-4/3} + N^{s-4} \leq CN^{6s-4/3}, \\ N^{4s}\delta_{N-1} + N^s\mathrm{e}^{-c\delta_{N-1}t} &\leq 2N^{4s-2/3} + N^{s-4} \leq CN^{4s-2/3}, \end{aligned}$$

where we used $t \geq 2\log N/\delta_{N-1}$ and $s \leq 1/44$. Further, $\mathrm{e}^{-2\delta_{N-1}t} \leq N^{-\alpha}N^{8/3-\alpha+10s} \leq N^{-\alpha-2/3-\sigma+10s}$ as $s \leq \sigma/44$. Then

$$-E_0(t) \geq N^{-12s-2/3-\alpha} - CN^{-\alpha-2/3-\sigma+10s},$$

provided that this is positive. Indeed,

$$-E_0(t) \geq N^{-12s-2/3-\alpha}(1 - CN^{-\sigma+22s}) \geq 0,$$

for N sufficiently large as $s \leq \sigma/44$. ☐

Now we look at the leading-order behavior of $E_0(t)$:

$$E_0(t) = \frac{1}{4}\delta_{N-1}^2\nu_{N-1}\mathrm{e}^{-\delta_{N-1}t}\frac{1+\sum_{n=1}^{N-2}\frac{\delta_n^2}{\delta_{N-1}^2}\frac{\nu_n}{\nu_{N-1}}\mathrm{e}^{-(\delta_n-\delta_{N-1})t}}{\left(1+\sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}\right)^2}. \tag{2.8}$$

Define T^* by

$$\begin{aligned} \frac{1}{4}\delta_{N-1}^2\nu_{N-1}\mathrm{e}^{-\delta_{N-1}T^*} &= N^{-\alpha}, \\ T^* &= \frac{\alpha\log N + 2\log\delta_{N-1} + \log\nu_{N-1} - 2\log 2}{\delta_{N-1}}. \end{aligned} \tag{2.9}$$

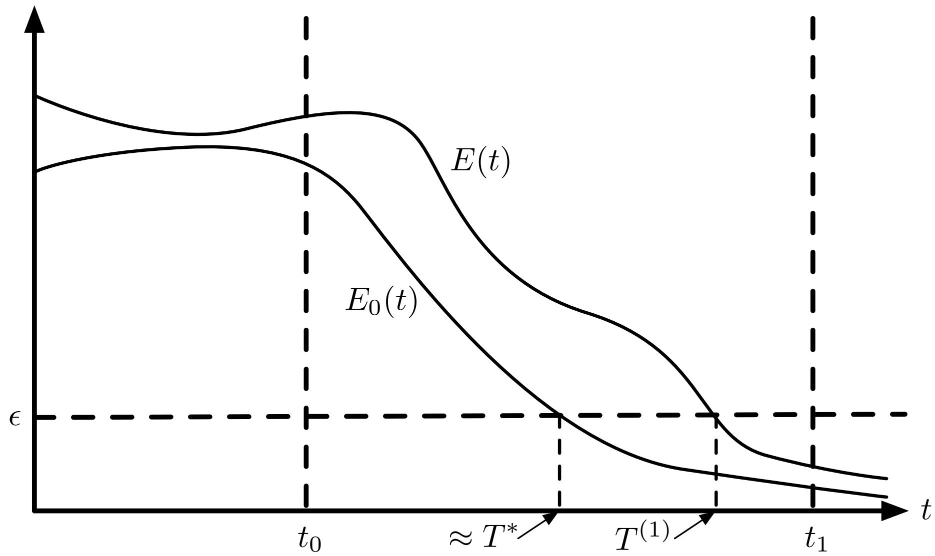


Figure 7: A schematic for the relationship between the functions $E_0(t)$, $E(t)$ and the times $T^{(1)}$ and T^*. Here $t_0 = (\alpha - 4/3 - 5s)\log N/\delta_{N-1}$ and $t_1 = (\alpha - 4/3 + 7s)\log N/\delta_{N-1}$. Note that E_0 is monotone on $[t_0, t_1]$.

Lemma 2.3. *Given Condition 1.2*

$$(\alpha - 4/3 - 4s)\log N/\delta_{N-1} \le T^* \le (\alpha - 4/3 + 4s)\log N/\delta_{N-1}.$$

Proof. This follows immediately from the statements

$$\begin{aligned} N^{-2s} &\le \nu_{N-1} \le N^{2s}, \\ 2N^{-2/3-s} &\le \delta_{N-1} \le 2N^{-2/3+s}. \end{aligned}$$

□

Thus, given Condition 1.2, $T^* \in I_\alpha$. The quantity that we want to estimate is $N^{-2/3}|T - T^*|$. And we do this by considering the formula

$$E_0(T^{(1)}) - E_0(T^*) = E_0'(\eta)(T^{(1)} - T^*), \quad \text{for some } \eta \in I_\alpha.$$

And because E_0 is monotone in I_α, $E_0(T^{(1)}) = E(T^{(1)}) - E_1(T^{(1)}) = N^{-\alpha} - E_1(T^{(1)})$ we have

$$|T^{(1)} - T^*| \le \frac{|N^{-\alpha} - E_0(T^*) - E_1(T^{(1)})|}{\min\limits_{\eta \in I_\alpha} |E_0'(\eta)|} \le \frac{|N^{-\alpha} - E_0(T^*)| + \max_{\eta \in I_\alpha} |E_1(\eta)|}{\min\limits_{\eta \in I_\alpha} |E_0'(\eta)|}. \tag{2.10}$$

See Figure 7 for a schematic of $E_0, E, T^{(1)}$ and T^*.

Since we already have an adequate estimate on $E_1(T)$ in (2.6), it remains to estimate $|N^{-\alpha} - E_0(T^*)|$.

Lemma 2.4. *Given Conditions 1.1 and 1.2*

$$|E_0(T^*) - N^{-\alpha}| \le CN^{-\alpha-2p+4s}.$$

Proof. From (2.8) and (2.9) we obtain

$$|E_0(T^*) - N^{-\alpha}| = N^{-\alpha} \frac{\left| \sum_{n=1}^{N-2} \frac{\delta_n^2}{\delta_{N-1}^2} \frac{\nu_n}{\nu_{N-1}} \mathrm{e}^{-(\delta_n - \delta_{N-1})T^*} - 2\sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n T^*} - \left(\sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n T^*} \right)^2 \right|}{\left(1 + \sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n T^*} \right)^2}.$$

We estimate the terms in the numerator individually using the bounds on T^*. For $c = 1$, we use that $\alpha - 4/3 - 4s > 2$ and Lemma 1.2 to find

$$\sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n T^*} \leq CN^{-\alpha+4/3+4s}\left(N^{4s} + N^{1+s-2c}\right) \leq CN^{-2-\sigma+8s} \leq N^{-2}$$

for sufficiently large N. Then we consider the first term the numerator using the index set I_c and Condition 1.1. Since our sum is now up to $N-2$ we define $\hat{I}_c = I_c \cap \{1, \ldots, N-2\}$ and $\hat{I}_c^c$ to denote the compliment relative to $\{1, \ldots, N-2\}$. Continuing,

$$\ell(T^*) := \sum_{n=1}^{N-2} \frac{\delta_n^2}{\delta_{N-1}^2} \frac{\nu_n}{\nu_{N-1}} \mathrm{e}^{-(\delta_n - \delta_{N-1})T^*} = \left(\sum_{n \in \hat{I}_c^c} + \sum_{n \in \hat{I}_c}\right) \frac{\delta_n^2}{\delta_{N-1}^2} \frac{\nu_n}{\nu_{N-1}} \mathrm{e}^{-(\delta_n - \delta_{N-1})T^*}$$

For $n \in \hat{I}_c^c$, $\delta_n^2/\delta_{N-1}^2 \leq (1+c)^2$ and

$$\delta_n - \delta_{N-1} = 2(\lambda_{N-1} - \lambda_n) \geq 2(\lambda_{N-1} - \lambda_{N-2}) \geq p\delta_{N-1},$$

using Condition 1.1. On the other hand for $n \in \hat{I}_c$, $\delta_n > (1+c)\delta_{N-1}$, and if $c = 3$

$$\delta_n - \delta_{N-1} > c\delta_{N-1} = p\delta_{N-1} + (c-p)\delta_{N-1} \geq p\delta_{N-1} + 2\delta_{N-1},$$

as $p < 1/3$ and hence $c > 2 + p$. Using Lemma 1.1 to estimate $|\hat{I}_c^c|$

$$\sum_{n \in \hat{I}_c^c} \frac{\delta_n^2}{\delta_{N-1}^2} \frac{\nu_n}{\nu_{N-1}} \mathrm{e}^{-(\delta_n - \delta_{N-1})T^*} \leq (1+c)^2 N^{4s} \mathrm{e}^{-p\delta_{N-1}T^*},$$

$$\sum_{n \in \hat{I}_c} \frac{\delta_n^2}{\delta_{N-1}^2} \frac{\nu_n}{\nu_{N-1}} \mathrm{e}^{-(\delta_n - \delta_{N-1})T^*} \leq [\max_n \delta_n^2] N^{7/3+3s} \mathrm{e}^{-(p+2)\delta_{N-1}T^*}.$$

Given Condition 1.2 $[\max_n \delta_n^2] \leq 4(b_V - a_V + 1)^2$ and hence for some $C > 0$, using that $\alpha - 4/3 - 4s > 2$

$$\begin{aligned}
\ell(T^*) &\leq C\mathrm{e}^{-p\delta_{N-1}T^*}\left(N^{4s} + N^{7/3+3s}\mathrm{e}^{-2\delta_{N-1}T^*}\right) \\
&\leq CN^{-p(\alpha-4/3-4s)}(N^{4s} + N^{7/3+3s-2(\alpha-4/3-4s)}) \\
&\leq CN^{-p(\alpha-4/3-4s)}(N^{4s} + N^{-5/3+3s}) \\
&\leq CN^{-2p+4s}(1 + N^{-5/3-s}).
\end{aligned}$$

Thus

$$\ell(T^*) \leq CN^{-2p+4s}.$$

From this it follows that

$$|E_0(T^*) - N^{-\alpha}| \leq CN^{-\alpha-2p+4s}.$$

□

Lemma 2.5. *Given Conditions 1.1 and 1.2, σ and p fixed and $s < \min\{\sigma/44, p/8\}$*

$$N^{-2/3}|T^{(1)} - T^*| \leq CN^{-2p+16s} \to 0, \quad \text{as } N \to \infty.$$

Proof. Combining Lemmas 2.2 and 2.4 with (2.6) and (2.10) we have for sufficiently large N

$$N^{-2/3}|T^{(1)} - T^*| \leq CN^{-2/3}N^{\alpha+12s+2/3}\left(N^{-\alpha-2p+4s} + N^{-\alpha}N^{-\alpha+8/3}\right) \leq C\left(N^{-2p+16s} + N^{-2/3+12s}\right),$$

where we used $\alpha - 8/3 > 2/3$. Since $p < 1/3$ the right-hand side is bounded by $CN^{-2p+16s}$ which goes to zero as $N \to \infty$ provided that $s < p/8$. □

From (2.4), we have

$$|\lambda_N - X_{11}(t)| = \frac{1}{2}\frac{\sum_{n=1}^{N-1}\delta_n\nu_n\mathrm{e}^{-\delta_n t}}{1+\sum_{n=1}^{N-1}\nu_n\mathrm{e}^{-\delta_n t}} \leq \frac{1}{2}\sum_{n=1}^{N-1}\delta_n\nu_n\mathrm{e}^{-\delta_n t}.$$

Lemma 2.6. *Given Condition 1.2, σ and p fixed and $s < \min\{\sigma/44, p/8\}$*

$$\epsilon^{-1}|\lambda_N - X_{11}(T^{(1)})| = N^{\alpha/2}|\lambda_N - X_{11}(T^{(1)})| \leq CN^{-1}$$

for sufficiently large N.

Proof. We use Lemma 1.2 with $c = 1$. By 2.1 we have

$$|\lambda_N - X_{11}(T^{(1)})| \leq CN^{-\alpha+4/3+5s}(N^{-2/3+5s} + N^{-1+s}) \leq CN^{-\alpha/2}N^{-1}$$

because $\alpha - 4/3 - 5s \geq 2$.

□

3 Adding probability

We now use the probabilistic facts about Conditions 1.2 and 1.1 as stated in Theorems 1.4 and 1.5 to understand T and T^* as random variables.

Lemma 3.1. *For $\alpha \geq 10/3 + \sigma$ and $\sigma > 0$*

$$\frac{|T^{(1)} - T^*|}{N^{2/3}}$$

converges to zero in probability as $N \to \infty$.

Proof. Let $\eta > 0$. Then

$$\mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta\right) = \mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta, G_{N,p}\cap R_{N,s}\right) + \mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta, G_{N,p}^c \cup R_{N,s}^c\right).$$

If s satisfies the hypotheses in Lemma 2.5, $s < \min\{\sigma/44, p/8\}$, then on the set $G_{N,p}\cap R_{N,s}$, $N^{-2/3}|T - T^*| < \eta$ for N sufficiently large, and hence

$$\mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta, G_{N,p}\cap R_{N,s}\right) \to 0,$$

as $N \to \infty$. We then estimate

$$\mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta, G_{N,p}^c \cup R_{N,s}^c\right) \leq \mathbb{P}(G_{N,p}^c) + \mathbb{P}(R_{N,s}^c),$$

and by Theorem 1.4

$$\limsup_{N\to\infty}\mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta, G_{N,p}^c \cup R_{N,s}^c\right) \leq \limsup_{N\to\infty}\mathbb{P}(G_{N,p}^c).$$

This is true for any $0 < p < 1/3$ and we use Theorem 1.5. So, as $p \downarrow 0$, we find

$$\lim_{N\to\infty}\mathbb{P}\left(\frac{|T^{(1)} - T^*|}{N^{2/3}} > \eta\right) = 0.$$

□

Define

$$\hat{T} = \frac{(\alpha - 4/3)\log N}{\delta_{N-1}}. \tag{3.1}$$

We need the following simple lemmas in what follows

Lemma 3.2. *If $X_N \to X$ in distribution*[8] *as $N \to \infty$ then*

$$\mathbb{P}\left(|X_N/a_N| < 1\right) = 1 + o(1)$$

as $N \to \infty$ provided that $a_N \to \infty$.

Proof. For two points of continuity a, b of $F(t) = \mathbb{P}(X \leq t)$ we have

$$\mathbb{P}\left(a < X_N \leq b\right) \to \mathbb{P}(a < X \leq b).$$

Let $M > 0$ such that $\pm M$ is a point of continuity of F. Then for sufficiently large N, $a_N > M$ and

$$\liminf_{N\to\infty} \mathbb{P}(-a_N < X_N < a_N) \geq \liminf_{N\to\infty} \mathbb{P}(-M < X_N \leq M) = \mathbb{P}(-M < X \leq M).$$

Letting $M \to \infty$ we see that $\mathbb{P}(-a_N \leq X_N \leq a_N) = 1 + o(1)$ as $N \to \infty$. □

Letting $a_N \to \eta a_N$, $\eta > 0$, we see that the following is true.

Corollary 3.1. *If $X_N \to X$ in distribution as $N \to \infty$ then*

$$|X_N/a_N|$$

converges to zero in probability provided $a_N \to \infty$.

Lemma 3.3. *If as $N \to \infty$, $X_N \to X$ in distribution and $|X_N - Y_N| \to 0$ in probability then $Y_N \to X$ in distribution.*

Proof. Let t be a point of continuity for $\mathbb{P}(X \leq t)$, then for $\eta > 0$

$$\begin{aligned}
\mathbb{P}(Y_N \leq t) &= \mathbb{P}(Y_N \leq t, X_N \leq t + \eta) + \mathbb{P}(Y_N \leq t, X_N > t + \eta) \\
&\leq \mathbb{P}(X_N \leq t + \eta) + \mathbb{P}(Y_N - X_N \leq t - X_N, t - X_N < -\eta) \\
&\leq \mathbb{P}(X_N \leq t + \eta) + \mathbb{P}(|Y_N - X_N| > \eta).
\end{aligned}$$

Interchanging the roles of X_N and Y_N and replacing t with $t - \eta$ we find

$$\mathbb{P}(X_N \leq t - \eta) \leq \mathbb{P}(Y_N \leq t) + \mathbb{P}(|Y_N - X_N| > \eta) \leq \mathbb{P}(X_N \leq t + \eta) + 2\mathbb{P}(|Y_N - X_N| > \eta).$$

From this we find that for any η such that $t \pm \eta$ are points of continuity

$$\mathbb{P}(X \leq t - \eta) \leq \liminf_{N\to\infty} \mathbb{P}(Y_N \leq t) \leq \limsup_{N\to\infty} \mathbb{P}(Y_N \leq t) \leq \mathbb{P}(X \leq t + \eta).$$

By sending $\eta \downarrow 0$ the result follows. □

Now, we compare T^* to $\hat{T}$.

Lemma 3.4. *For $\alpha \geq 10/3 + \sigma$*

$$\frac{|T^* - \hat{T}|}{N^{2/3}\log N}$$

converges to zero in probability as $N \to \infty$.

[8] For convergence in distribution, we require that the limiting random variable X satisfies $\mathbb{P}(|X| < \infty) = 1$.

Proof. Consider

$$\begin{aligned}\frac{T^* - \hat{T}}{N^{2/3}\log N} &= \frac{1}{\log N}\frac{\log \nu_{N-1} + 2\log N^{2/3}\delta_{N-1}}{N^{2/3}\delta_{N-1}} \\ &= \frac{1}{\sqrt{\log N}}\left(\frac{1}{(\log N)^{1/4}}|N^{2/3}\delta_{N-1}|^{-1}\right)\left(\frac{2}{(\log N)^{1/4}}\log \nu_{N-1} + \frac{1}{(\log N)^{1/4}}\log N^{2/3}\delta_{N-1}\right).\end{aligned}$$

For

$$\begin{aligned}L_N &= \left\{\frac{1}{(\log N)^{1/4}}|N^{2/3}\delta_{N-1}|^{-1} \leq 1\right\}, \\ U_N &= \left\{\frac{1}{(\log N)^{1/4}}|\log \nu_{N-1}| \leq 1\right\}, \\ P_N &= \left\{\frac{1}{(\log N)^{1/4}}|\log N^{2/3}\delta_{N-1}| \leq 1\right\},\end{aligned}$$

we have $\mathbb{P}(L_N^c) + \mathbb{P}(U_N^c) + \mathbb{P}(P_n^c) \to 0$ as $N \to \infty$ by Lemma 3.2 and Theorem 1.3. For these calculations it is important that the limiting distribution function for $N^{2/3}\delta_{N-1}$ is continous at zero, see Theorem 1.3. Then for $\eta > 0$

$$\begin{aligned}\mathbb{P}\left(\left|\frac{T^* - \hat{T}}{N^{2/3}\log N}\right| > \eta\right) &= \mathbb{P}\left(\left|\frac{T^* - \hat{T}}{N^{2/3}\log N}\right| > \eta, L_N \cap U_N \cap P_N\right) \\ &\quad + \mathbb{P}\left(\left|\frac{T^* - \hat{T}}{N^{2/3}\log N}\right| > \eta, L_N^c \cup U_N^c \cup P_N^c\right).\end{aligned} \tag{3.2}$$

On the set $L_N \cap U_N \cap P_N$ we estimate

$$\left|\frac{T^* - \hat{T}}{N^{2/3}\log N}\right| \leq \frac{3}{\sqrt{\log N}}.$$

Hence first term on the right-hand side of (3.2) is zero for sufficiently large N and the second term is bounded by $\mathbb{P}(U_N^c) + \mathbb{P}(L_N^c) + \mathbb{P}(P_N^c)$ which tends to zero. This shows convergence in probability. ☐

We now arrive at our main result.

Theorem 3.1. *If $\alpha \geq 10/3 + \sigma$ and $\sigma > 0$ then*

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{2^{2/3}T^{(1)}}{c_V^{2/3}(\alpha - 4/3)N^{2/3}\log N} \leq t\right) = F_\beta^{\mathrm{gap}}(t).$$

Proof. Combining Lemma 3.1 and Lemma 3.4 we have that

$$\left|2^{2/3}\frac{T^{(1)} - \hat{T}}{c_V^{2/3}(\alpha - 4/3)N^{2/3}\log N}\right|$$

converges to zero in probability. Then by Lemma 3.3 and Theorem 1.3 the result follows as

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{2^{2/3}\hat{T}}{c_V^{2/3}(\alpha - 4/3)N^{2/3}\log N}\right) = \lim_{N\to\infty} \mathbb{P}(c_V^{-2/3}2^{2/3}N^{-2/3}(\lambda_N - \lambda_{N-1})^{-1} \leq t) = F_\beta^{\mathrm{gap}}(t).$$

☐

We also prove a result concerning the true error $|\lambda_N - X_{11}(T^{(1)})|$:

Proposition 3.1. *For $\alpha \geq 10/3 + \sigma$ and $\sigma > 0$ and any $q < 1$*

$$N^{\alpha/2+q}|\lambda_N - X_{11}(T^{(1)})|$$

converges to zero in probability as $N \to \infty$. Furthermore, for any $r > 0$

$$N^{2/3+r}|\gamma_N - X_{11}(T^{(1)})|, \quad N^{2/3+r}|\lambda_j - X_{11}(T^{(1)})|,$$

converges to ∞ in probability, if $j = j(N) < N$.

Proof. We recall that $R_{N,s}$ is the set on which Condition 1.2 holds. Then for any $\eta > 0$

$$\begin{aligned}
&\mathbb{P}(N^{\alpha/2+q}|\lambda_N - X_{11}(T^{(1)})| > \eta) \\
&\qquad = \mathbb{P}(N^{\alpha/2+q}|\lambda_N - X_{11}(T^{(1)})| > \eta, R_{N,s}) + \mathbb{P}(N^{\alpha/2+q}|\lambda_N - X_{11}(T^{(1)})| > \eta, R_{N,s}^c) \\
&\qquad \leq \mathbb{P}(N^{\alpha/2+q}|\lambda_N - X_{11}(T^{(1)})| > \eta, R_{N,s}) + \mathbb{P}(R_{N,s}^c).
\end{aligned}$$

Using Lemma 2.6, the first term on the right-hand side is zero for sufficiently large N and the second term vanishes from Theorem 1.4. This shows the first statement, *i.e.*,

$$\lim_{N\to\infty} \mathbb{P}(N^{\alpha/2+q}|\lambda_N - X_{11}(T^{(1)})| > \eta) = 0.$$

For the second statement, on the set $R_{N,s}$ with $s < \min\{r, \sigma/44, p/8\}$ we have

$$|\lambda_j - X_{11}(T^{(1)})| \geq |\lambda_j - \lambda_N| - |\lambda_N - X_{11}(T^{(1)})| \geq |\lambda_{N-1} - \lambda_N| - |\lambda_N - X_{11}(T^{(1)})|,$$

and for sufficiently large N (see Lemma 2.6)

$$N^{2/3+r}|\lambda_j - X_{11}(T^{(1)})| \geq N^r(N^{2/3}|\lambda_{N-1} - \lambda_N| - N^{-1/3-\alpha/2}) \geq N^{r-s}(1 - CN^{-1/3-\alpha/2+s}).$$

This tends to ∞ as $s < 1/3$ and $s < r$. Hence for any $K > 0$, again using the arguments of Theorem 3.1,

$$\begin{aligned}
&\mathbb{P}\left(N^{2/3+r}|\lambda_j - X_{11}(T^{(1)})| > K\right) \\
&\quad = \mathbb{P}\left(N^{2/3+r}|\lambda_j - X_{11}(T^{(1)})| > K, R_{N,s}\right) + \mathbb{P}\left(N^{2/3+r}|\lambda_j - X_{11}(T^{(1)})| > K, R_{N,s}^c\right).
\end{aligned}$$

For sufficiently large N, the first term on the right-hand side is equal to $\mathbb{P}(R_{N,s})$ and the second term is bounded by $\mathbb{P}(R_{N,s}^c)$ and hence

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3+r}|\lambda_j - X_{11}(T^{(1)})| > K\right) = 1.$$

Next, under the same assumption (Condition 1.2)

$$N^{2/3+r}|b_V - X_{11}(T^{(1)})| \geq N^r(N^{2/3}|b_V - \lambda_N| - CN^{-1/3-\alpha/2}).$$

From Corollary 3.1 and Theorem 1.3 using $\gamma_N = b_V$

$$N^{-r}(N^{2/3}|b_V - \lambda_N| - CN^{-1/3-\alpha/2})^{-1}$$

converges to zero in probability (with no point mass at zero), implying its inverse converges to ∞ in probability. This shows $N^{\alpha}|b_V - X_{11}(T^{(1)})|$ converges to ∞ in probability. ◻

Acknowledgments

The authors would like to thank Yuri Bakhtin and Paul Bourgade for many useful conversations. We also thank Folkmar Bornemann for the data for F_2^{gap}. This work was supported in part by grants NSF-DMS-1303018 (TT) and NSF-DMS-1300965 (PD).

References

[Adler, 1978] Adler, M. (1978). On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteweg-devries type equations. *Invent. Math.*, 50(3):219–248.

[Bakhtin and Correll, 2012] Bakhtin, Y. and Correll, J. (2012). A neural computation model for decision-making times. *J. Math. Psychol.*, 56(5):333–340.

[Bourgade, 2016] Bourgade, P. (2016). Personal communication.

[Bourgade et al., 2014] Bourgade, P., Erdős, L., and Yau, H.-T. (2014). Edge Universality of Beta Ensembles. *Commun. Math. Phys.*, 332(1):261–353.

[Bourgade and Yau, 2013] Bourgade, P. and Yau, H.-T. (2013). The Eigenvector Moment Flow and local Quantum Unique Ergodicity. pages 1–35.

[Deift, 2000] Deift, P. (2000). *Orthogonal Polynomials and Random Matrices: a Riemann-Hilbert Approach*. Amer. Math. Soc., Providence, RI.

[Deift and Gioev, 2007] Deift, P. and Gioev, D. (2007). Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices. *Commun. Pure Appl. Math.*, 60(6):867–910.

[Deift et al., 1986] Deift, P., Li, L. C., Nanda, T., and Tomei, C. (1986). The Toda flow on a generic orbit is integrable. *Commun. Pure Appl. Math.*, 39(2):183–232.

[Deift et al., 1985] Deift, P., Li, L. C., and Tomei, C. (1985). Toda flows with infinitely many variables. *J. Funct. Anal.*, 64(3):358–402.

[Deift et al., 1983] Deift, P., Nanda, T., and Tomei, C. (1983). Ordinary differential equations and the symmetric eigenvalue problem. *SIAM J. Numer. Anal.*, 20:1–22.

[Deift et al., 2014] Deift, P. A., Menon, G., Olver, S., and Trogdon, T. (2014). Universality in numerical computations with random data. *Proc. Natl. Acad. Sci. U. S. A.*, 111(42):14973–8.

[Erdős, 2012] Erdős, L. (2012). Universality for random matrices and log-gases. In Jerison, D., Kisin, M., Mrowka, T., Stanley, R., Yau, H.-T., and Yau, S.-T., editors, *Curr. Dev. Math.*, number 2010.

[Erdős et al., 2012] Erdős, L., Yau, H.-T., and Yin, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math. (N. Y).*, 229(3):1435–1515.

[Flaschka, 1974] Flaschka, H. (1974). The Toda lattice. II. Existence of integrals. *Phys. Rev. B*, 9(4):1924–1925.

[Jiang, 2006] Jiang, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.*, 34(4):1497–1529.

[Kostant, 1979] Kostant, B. (1979). The solution to a generalized Toda lattice and representation theory. *Adv. Math. (N. Y).*, 34(3):195–338.

[Manakov, 1975] Manakov, S. V. (1975). Complete integrability and stochastization of discrete dynamical systems. *Sov. Phys. JETP*, 40(2):269–274.

[Menon et al., 2016] Menon, G., Trogdon, T., and Deift, P. A. (2016). On the condition number of the critically-scaled Laguerre Unitary Ensemble. *Discret. Contin. Dyn. Syst.*, 36(8):4287–4347.

[Monthus and Garel, 2013] Monthus, C. and Garel, T. (2013). Typical versus averaged overlap distribution in spin glasses: Evidence for droplet scaling theory. *Phys. Rev. B*, 88(13):134204.

[Moser, 1975] Moser, J. (1975). Three integrable Hamiltonian systems connected with isospectral deformations. *Adv. Math. (N. Y).*, 16(2):197–220.

[Perret and Schehr, 2014] Perret, A. and Schehr, G. (2014). Near-Extreme Eigenvalues and the First Gap of Hermitian Random Matrices. *J. Stat. Phys.*, 156(5):843–876.

[Pfrang et al., 2014] Pfrang, C. W., Deift, P., and Menon, G. (2014). How long does it take to compute the eigenvalues of a random symmetric matrix? *Random matrix theory, Interact. Part. Syst. Integr. Syst. MSRI Publ.*, 65:411–442.

[Ramírez et al., 2011] Ramírez, J. A., Rider, B., and Virág, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Am. Math. Soc.*, 24(4):919–944.

[Sagun et al., 2015] Sagun, L., Trogdon, T., and LeCun, Y. (2015). Universality in halting time and its applications in optimization. *arXiv Prepr. arXiv1511.06444*.

[Soshnikov, 1999] Soshnikov, A. (1999). Universality at the Edge of the Spectrum in Wigner Random Matrices. *Commun. Math. Phys.*, 207(3):697–733.

[Stam, 1982] Stam, A. J. (1982). Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *J. Appl. Probab.*, 19(1):221–228.

[Symes, 1982] Symes, W. W. (1982). The QR algorithm and scattering for the finite nonperiodic Toda lattice. *Phys. D Nonlinear Phenom.*, 4(2):275–280.

[Tao and Vu, 2010] Tao, T. and Vu, V. (2010). Random Matrices: Universality of Local Eigenvalue Statistics up to the Edge. *Commun. Math. Phys.*, 298(2):549–572.

[Toda, 1967] Toda, M. (1967). Vibration of a chain with nonlinear interaction. *J. Phys. Soc. Japan*, 22(2):431–436.

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

[Watkins, 1984] Watkins, D. S. (1984). Isospectral flows. *SIAM Rev.*, 26(3):379–391.

[Witte et al., 2013] Witte, N. S., Bornemann, F., and Forrester, P. J. (2013). Joint distribution of the first and second eigenvalues at the soft edge of unitary ensembles. *Nonlinearity*, 26(6):1799–1822.