Universality for the Toda Algorithm to Compute the Largest Eigenvalue of a Random Matrix

PERCY DEIFT
Courant Institute

THOMAS TROGDON
University of California, Irvine

Abstract

We prove universality for the fluctuations of the halting time for the Toda algorithm to compute the largest eigenvalue of real symmetric and complex Hermitian matrices. 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. © 2017 Wiley Periodicals, Inc.

1 Introduction

In [22] the authors initiated a statistical study of the performance of various standard algorithms $\mathscr{A}$ 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 $\mathscr{A}$, there is, in the discrete case such as QR, a map $\varphi = \varphi_{\mathscr{A}} : \Sigma_N \to \Sigma_N$ with the properties

- isospectrality: $\operatorname{spec}(\varphi_{\mathscr{A}}(H)) = \operatorname{spec}(H)$,
- convergence: the iterates $X_{k+1} = \varphi_{\mathscr{A}}(X_k)$, $k \geq 0$, $X_0 = H$ 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

- isospectrality: $\operatorname{spec}(X(t))$ 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 $\mathscr{O}(\epsilon)$ and hence the diagonal entries of X_m give the eigenvalues of $X_0 = H$ to $\mathscr{O}(\epsilon)$. The situation is similar for continuous algorithms $t \mapsto X(t)$. Rather than running the algorithm until all the off-diagonal entries are $\mathscr{O}(\epsilon)$, it is customary to run the algorithm with *deflations* as follows. For an $N \times N$ matrix Y

Communications on Pure and Applied Mathematics, Vol. LXXI, 0505–0536 (2018)

in block form

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

with Y_{11} of size $k \times k$ and Y_{22} of size $N-k \times N-k$ for some $k \in \{1, \ldots, N-1\}$, the process of projecting $Y \mapsto \operatorname{diag}(Y_{11}, Y_{22})$ is called *deflation*. For a given ϵ, algorithm $\mathscr{A}$, and matrix $H \in \Sigma_N$, define the *k-deflation* time $T^{(k)}(H) = T^{(k)}_{\epsilon,\mathscr{A}}(H)$, $1 \leq k \leq N-1$, to be the smallest value of m such that X_m, the m^{th} iterate of algorithm $\mathscr{A}$ 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 $\|X_{12}^{(k)}\| = \|X_{21}^{(k)}\| \leq \epsilon$.[1] The deflation time $T(H)$ is then defined as

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

If $\hat{k} \in \{1, \ldots, N-1\}$ is such that $T(H) = T^{(\hat{k})}_{\epsilon,\mathscr{A}}(H)$, it follows that the eigenvalues of $H = X_0$ are given by the eigenvalues of the block-diagonal matrix $\operatorname{diag}(X_{11}^{(\hat{k})}, X_{22}^{(\hat{k})})$ to $\mathscr{O}(\epsilon)$. After running the algorithm to time $T_{\epsilon,A}(H)$, the algorithm restarts by applying the basic algorithm $\mathscr{A}$ 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 $\mathscr{O}(\log N)$ to $\mathscr{O}(N)$.

In [22] the authors considered the deflation time $T = T_{\epsilon,\mathscr{A}}$ for $N \times N$ matrices chosen from a given ensemble $\mathscr{E}$. Henceforth in this paper we suppress the dependence on ϵ, N, $\mathscr{A}$, and $\mathscr{E}$, and simply write T with these variables understood. For a given algorithm $\mathscr{A}$ and ensemble $\mathscr{E}$ the authors computed $T(H)$ for 5 000–15 000 samples of matrices H chosen from $\mathscr{E}$ and recorded the *normalized deflation time*

$$\widetilde{T}(H) := \frac{T(H) - \langle T \rangle}{\sigma}, \tag{1.1}$$

where $\langle T \rangle$ and $\sigma^2 = \langle (T - \langle T \rangle)^2 \rangle$ are the sample average and sample variance of $T(H)$, respectively. Surprisingly, the authors found that for the given algorithm $\mathscr{A}$, and ϵ and N in a suitable scaling range with $N \to \infty$, the histogram of $\widetilde{T}$ was

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

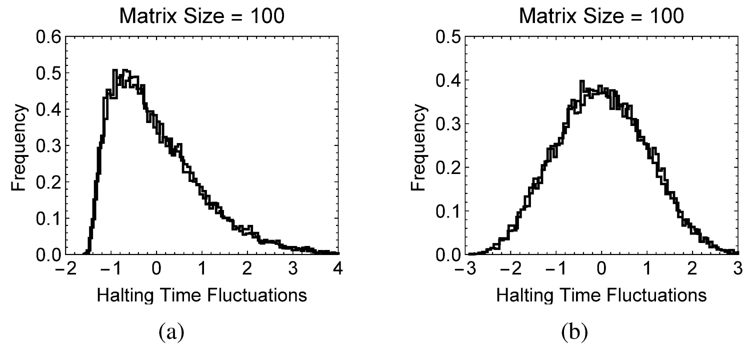

FIGURE 1.1. Universality for $\widetilde{T}$ when (a) $\mathscr{A}$ is the QR eigenvalue algorithm and when (b) $\mathscr{A}$ is the Toda algorithm. Panel (a) displays the overlay of two histograms for $\widetilde{T}$ in the case of QR, one for each of the two ensembles $\mathscr{E} = \text{BE}$, consisting of i.i.d. mean-zero Bernoulli random variables (cf. Definition A.1) and $\mathscr{E} = \text{GOE}$, consisting of i.i.d. mean-zero normal random variables. Here $\epsilon = 10^{-10}$ and $N = 100$. Panel (b) displays the overlay of two histograms for $\widetilde{T}$ in the case of the Toda algorithm, and again $\mathscr{E} = \text{BE}$ or GOE. Here $\epsilon = 10^{-8}$ and $N = 100$.

universal, independent of the ensemble $\mathscr{E}$. In other words, the fluctuations in the deflation time $\widetilde{T}$, suitably scaled, were universal, independent of $\mathscr{E}$. Figure 1.1 displays some of the numerical results from [22]. Figure 1.1(a) displays data for the QR algorithm, which is discrete, and Figure 1.1(b) displays data for the Toda algorithm, which is continuous.

Subsequently, in [9], the authors raised the question of whether the universality results in [22] were limited to eigenvalue algorithms for real symmetric matrices or whether they were present more generally in numerical computation. And indeed the authors in [9] found similar universality results for a wide variety of numerical algorithms, including

- other algorithms such as the QR algorithm with shifts, the Jacobi eigenvalue algorithm, and also algorithms applied to complex Hermitian ensembles with H and b random,
- 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.

All of the above results are numerical. The goal of this paper is to establish universality as a bona fide phenomenon in numerical analysis and not just an artifact suggested, however strongly, by certain computations as above. To this end we

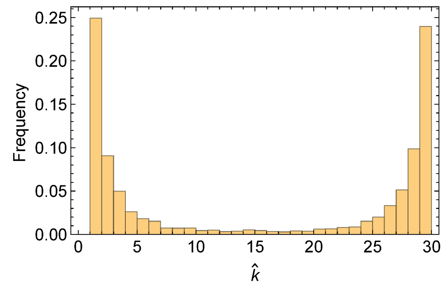

FIGURE 1.2. The distribution of $\widehat{k}$ for GOE when $N = 30$, $\epsilon = 10^{-8}$ for the Toda algorithm: $\widehat{k} = 1, N-1$ are equally likely.

seek out and prove universality for an algorithm of interest. We focus, in particular, on eigenvalue algorithms. To analyze eigenvalue algorithms with deflation, one must first analyze $T^{(k)}$ for $1 \le k \le N-1$, and then compute the minimum of these $N-1$ dependent variables. The analysis of $T^{(k)}$ for $1 \le k \le N-1$ requires very detailed information on the eigenvalues and eigenvectors of random matrices that, at this time, has only been established for $T^{(1)}$ (see below). Computing the minimum requires knowledge of the distribution of $\widehat{k}$ such that $T(H) = T^{(\widehat{k})}(H)$, which is an analytical problem that is still untouched. In Figure 1.2 we show the statistics of $\widehat{k}$ obtained numerically for the Toda algorithm.[2] In view of the above issues, a comprehensive analysis of the algorithms with deflation currently seems to be out of reach. In this paper we restrict our attention to the Toda algorithm, and as a *first step* towards understanding $T(H)$ we prove universality for the fluctuations of $T^{(1)}(H)$, the 1-deflation time for Toda; see Theorem 1.2. As we see from Proposition 1.3, 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. Theorem 1.2 and Proposition 1.3 are the main results in this paper. Much of the detailed statistical information on the eigenvalues and eigenvectors of H needed to analyze $T^{(1)}(H)$ was only established in the last three or four years.

In this paper we always order the eigenvalues $\lambda_n \le \lambda_{n+1}$, $n = 1, \dots, N$. 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 1.2. Note that Figure 1.3 for $T^{(1)}(H)$ is very different from Figure 1.1(b) for $T(H)$. In Sections 2 and 3 we will prove universality for $T^{(1)}$ for matrices from generalized Wigner ensembles and also

[2] A similar histogram for the QR algorithm has an asymmetry that reflects the fact that typically H has an eigenvalue near 0: For QR, a simple argument shows that eigenvalues near 0 favor $\widehat{k} = N-1$.

from invariant ensembles. See the Appendix (p. 534) for a full description of these random matrix ensembles. The techniques in this paper can also be used to prove universality for the fluctuations in the halting times for other eigenvalue algorithms, in particular, QR (without shifts)—see Remark 1.5 below.

1.1 Main Result

The Toda algorithm is an example of the generalized continuous eigenvalue algorithms 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]

$$\dot{X} = [X, B(X)], \quad B(X) = X_- - (X_-)^*, \quad X(0) = H = H^*, \tag{1.2}$$

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 $X_\infty = \operatorname{diag}(\lambda_N, \ldots, \lambda_1)$; see, for example, [8]. As noted above, necessarily, the diagonal elements of X_∞ are the eigenvalues of H. By the *Toda algorithm* to compute the eigenvalues of a Hermitian matrix H we mean solving (1.2) 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 [28] and describes the motion of N particles x_i, $i = 1, \ldots, N$, on the line under the Hamiltonian

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

In 1974, Flaschka [15] (see also [18]) showed that Hamilton's equations

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

can be written in the Lax pair form (1.2) 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 $B(X) = X_- - (X_-)^{\mathsf{T}}$ as in (1.2). 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 0 as $t \to \infty$ [20]. Inspired by this result, and also related work of Symes [26] on the QR algorithm, the authors in [10] suggested that the Toda lattice be viewed as an eigenvalue algorithm, the Toda algorithm. The Lax equations (1.2) clearly give rise to a global flow not only

[3] In the real symmetric case $*$ should be replaced with T.

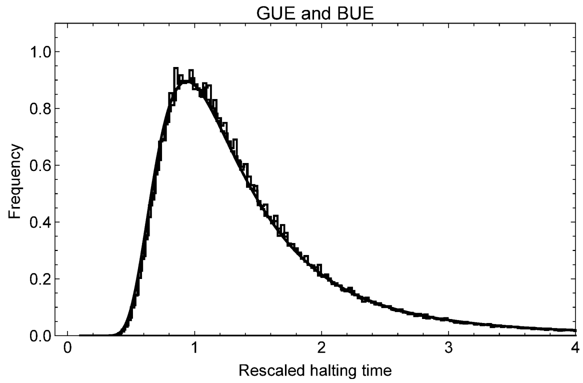

FIGURE 1.3. The simulated rescaled histogram for $\widetilde{T}^{(1)}$ for both BUE and GUE. Here $\epsilon = 10^{-14}$ and $N = 500$ with 250 000 samples. The solid curve is the rescaled density $f_2^{\text{gap}}(t) = \mathrm{d}/\mathrm{d}t F_2^{\text{gap}}(t)$. The density $f_2^{\text{gap}}(t) = \frac{1}{\sigma t^2} A^{\text{soft}}(\frac{1}{\sigma t})$, where $A^{\text{soft}}(s)$ is shown in [31, fig. 1]: In order to match the scale in [31], our choice of distributions (BUE and GUE) we must take $\sigma = 2^{-7/6}$. This is a numerical demonstration of Theorem 1.2. See also Section 1.2, "Numerical Demonstration," for further discussion of this figure and also Figure 1.4.

on tridiagonal matrices but also on general real symmetric matrices. It turns out that in this generality (1.2) is also Hamiltonian [1, 17] and, in fact, integrable [7]. From that point on, by the Toda algorithm one means the action of (1.2) 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 $T^{(1)}$, 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 $\mathscr{O}(\epsilon)$. Define

$$E(t) = \sum_{n=2}^{N} |X_{1n}(t)|^2, \tag{1.3}$$

so that if $E(t) = 0$, then $X_{11}(t)$ is an eigenvalue of H. Thus, with $E(t)$ as in (1.3), the halting time (or 1-deflation time) for the Toda algorithm is given by

$$T^{(1)}(H) = \inf\{t : E(t) \leq \epsilon^2\}. \tag{1.4}$$

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 $X(0)$.

---

[4] The Toda flow (1.2) also generates a completely integrable Hamiltonian system on real (not necessarily symmetric) $N \times N$ matrices; see [8]. The Toda flow (1.2) on Hermitian matrices was first investigated by Watkins [30].

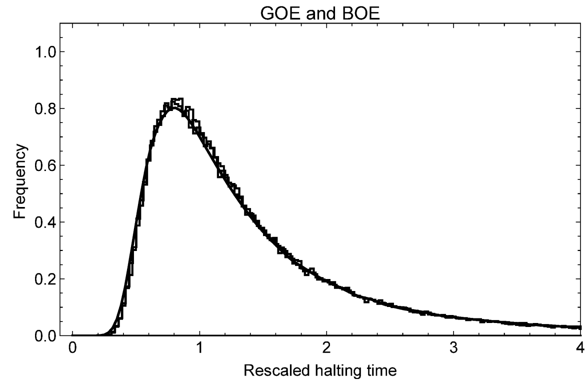

FIGURE 1.4. The simulated rescaled histogram for $\widetilde{T}^{(1)}$ for both BOE and GOE demonstrating Theorem 1.2. Here $\epsilon = 10^{-14}$ and $N = 500$ with 250 000 samples. The solid curve is an approximation to the density $f_1^{\mathrm{gap}}(t) = \mathrm{d}/\mathrm{d}t F_1^{\mathrm{gap}}(t)$. 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.

For invariant ensembles and generalized Wigner random matrix ensembles (IEs and WEs; see the Appendix) there is a constant c_V, which depends on the ensemble, such that the following limit exists ($\beta = 1$ for the real symmetric case, $\beta = 2$ for the complex Hermitian case):

$$F_\beta^{\mathrm{gap}}(t) = \lim_{N\to\infty} \mathbb{P}\left(\frac{1}{c_V^{2/3} 2^{-2/3} N^{2/3}(\lambda_N - \lambda_{N-1})} \le t\right), \quad t \ge 0. \tag{1.5}$$

The precise value of c_V is described in Theorem 1.7, and this limit is discussed further in Definition 1.10. For fixed β, the limit is independent of the choice of ensemble. This is the rescaled distribution of the inverse of the top gap in the spectrum of the random matrix. This distribution is the universal limit of $T^{(1)}$, capturing the fact that the rate of convergence is asymptotically governed by the gap $\lambda_N - \lambda_{N-1}$.

DEFINITION 1.1 (Scaling region). Fix $0 < \sigma < 1$. The *scaling region* for (ϵ, N) is given by $\frac{\log \epsilon^{-1}}{\log N} \ge \frac{5}{3} + \frac{\sigma}{2}$.[5]

Note that for $\epsilon = 10^{-15}$, a relevant value for double-precision arithmetic, (ϵ, N) is in the scaling region for all values of N less than 10^9.

THEOREM 1.2 (Universality for $T^{(1)}$). *Let* $0 < \sigma < 1$ *be fixed and let* (ϵ, N) *be in the scaling region* $\frac{\log \epsilon^{-1}}{\log N} \ge \frac{5}{3} + \frac{\sigma}{2}$. *Then if* H *is distributed according to any real*

[5] From the statement of the theorem, it is reasonable to ask if $\frac{5}{3}$ can be replaced with $\frac{2}{3}$ in the definition of the scaling region. Also, one should expect different limits for larger values of ϵ as other eigenvalues will contribute. These questions have yet to be explored.

($\beta = 1$) or complex ($\beta = 2$) invariant or Wigner ensemble, we have

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{T^{(1)}}{c_V^{2/3} 2^{-2/3} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N)} \leq t\right) = F_\beta^{\mathrm{gap}}(t). \tag{1.6}$$

Here c_V is the same constant as in (1.5).

Example 1.1. Consider the case of real symmetric 2×2 matrices. For $X(0) = H$, it follows that as $t \to \infty$, $X_{11}(t) \to \lambda_2$, the largest eigenvalue, while $X_{22}(t) \to \lambda_1$, the second-largest eigenvalue. And so, one should expect $T^{(1)}$ to be larger for

$$X(0) = H_+ := \begin{bmatrix} -1 & \delta \\ \delta & 1 \end{bmatrix} \quad \text{than for} \quad X(0) = H_- := \begin{bmatrix} 1 & \delta \\ \delta & -1 \end{bmatrix},$$

despite the fact that these matrices have the same eigenvalues. Said differently, it is surprising that the fluctuations of $T^{(1)}$ in Theorem 1.2 depend only on the eigenvalues and are independent of the eigenvectors of H.

Let $U = (U_{ij})_{1\leq i,j\leq 2}$ be the matrix of normalized eigenvectors of $X(0)$. It then follows from the calculations in Section 2 that

$$|X_{12}(t)|^2 = (\lambda_2 - \lambda_1)^2 \frac{|U_{11}(0)|^2 \mathrm{e}^{2\lambda_1 t}}{|U_{11}(0)|^2 \mathrm{e}^{2\lambda_1 t} + |U_{12}(0)|^2 \mathrm{e}^{2\lambda_2 t}}.$$

It is then clear that

$$|X_{12}(t)|^2 \sim (\lambda_2 - \lambda_1)^2 \frac{|U_{11}(0)|^2}{|U_{12}(0)|^2} \mathrm{e}^{-2(\lambda_2 - \lambda_1)t} \quad \text{as } t \to \infty.$$

First, note that this, roughly speaking, explains the appearance of $\lambda_N - \lambda_{N-1}$ in the definition of the universal limit $F_\beta^{\mathrm{gap}}(t)$. Second, a simple calculation shows that as $\delta \downarrow 0$, $|U_{12}(0)| \sim \delta$ for H_+ while $|U_{12}(0)| \sim 1$ for H_-, explaining why $T^{(1)}(H_+) \geq T^{(1)}(H_-)$. However, the matrices H_+ and H_- are not "typical." With high probability, the eigenvectors of random matrices in the ensembles under consideration are delocalized, so that U_{1j}, $j = 1, \ldots, N$, are all of the same order. For general N, we then have $\sum_{k=2}^N |X_{1k}|^2 \asymp (\lambda_N - \lambda_{N-1})^2 \mathrm{e}^{-2(\lambda_{N-1} - \lambda_N)t}$ and the dependence on the eigenvectors is effectively removed as $\epsilon \downarrow 0$.

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 3.7 that shows our error is $\mathscr{O}(\epsilon)$ with high probability.

PROPOSITION 1.3 (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,*

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

converges to 0 *in probability as $N \to \infty$. Furthermore, both*

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

converge to ∞ in probability for any $j = j(N) < N$ as $N \to \infty$, where b_V is the supremum of the support of the equilibrium measure for the ensemble.

The relation of this theorem to two-component universality as discussed in [9] is the following. Let $\xi = \xi_\beta$ be the random variable with distribution $F_\beta^{\text{gap}}(t)$, $\beta = 1$ or 2. For $\beta = 2$ IEs one can prove that[6]

$$\mathbb{E}[T^{(1)}] = c_V^{2/3} 2^{-\frac{2}{3}} N^{\frac{2}{3}} \left(\log \epsilon^{-1} - \frac{2}{3} \log N \right) \mathbb{E}[\xi](1 + o(1)), \tag{1.7}$$

$$\sqrt{\operatorname{Var}(T^{(1)})} = \kappa c_V^{2/3} 2^{-\frac{2}{3}} N^{\frac{2}{3}} \left(\log \epsilon^{-1} - \frac{2}{3} \log N \right)(1 + o(1)), \quad \kappa > 0. \tag{1.8}$$

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

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

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 1.4. If one constructs matrices $H = U \Lambda U^*$, $\Lambda = \operatorname{diag}(\lambda_N, \lambda_{N-1}, \ldots, \lambda_1)$ 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 1.2 holds for any $\beta \geq 1$. Here V should satisfy the hypotheses in Definition A.2.

Remark 1.5. To compute the largest eigenvalue of H, one can alternatively consider the flow

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

It follows that

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

---

[6] We can also prove (1.7) for $\beta = 1$ IEs. The proofs of these facts require an extension of the level repulsion estimates in [3, theorem 3.2] to the case $K = 1$. When $\beta = 2$, again with this extension of [3, theorem 3.2] to the case $K = 1$, we can prove that $\kappa = \operatorname{Var}(\xi)$. This extension is known to be true [2]. The calculations in Table 1.1 below are consistent with (1.7) and (1.8) (even for WEs) and lead us to believe that (1.8) 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 [21]. In other words, we face the unusual situation where the variance seems to converge, but not to the variance of the limiting distribution.

So, define

$$T_{\mathrm{ODE}}(H) = \inf\left\{t : \left|\log \frac{\|X(t+1)\|}{\|X(t)\|} - \lambda_N\right| \le \epsilon\right\}.$$

Using the proof technique we present here, one can show that Theorem 1.2 also holds with $T^{(1)}$ replaced with T_{ODE}. The same is true for the power method, the inverse power method, and the QR algorithm without shifts on positive-definite random matrices (see [11]).

1.2 A Numerical Demonstration

We can demonstrate Theorem 1.2 numerically using the following WEs defined by letting X_{ij} for $i \le j$ be i.i.d. with distributions:

- GUE: Mean zero standard complex normal.
- BUE: $\xi + \mathrm{i}\eta$ where ξ and η 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.

In Figure 1.3, for $\beta = 2$, we show how the histogram of $T^{(1)}$ (more precisely, $\widetilde{T}^{(1)}$; see (1.9) below), after rescaling, matches the density $\mathrm{d}/\mathrm{d}t F_2^{\mathrm{gap}}(t)$, which was computed numerically in [31].[7] In Figure 1.4, for $\beta = 1$, we show the histogram for $T^{(1)}$ (again, $\widetilde{T}^{(1)}$), after rescaling, matches the density $\mathrm{d}/\mathrm{d}t F_1^{\mathrm{gap}}(t)$. To the best of our knowledge, a computationally viable formula for $\mathrm{d}/\mathrm{d}t F_1^{\mathrm{gap}}(t)$, analogous to $\mathrm{d}/\mathrm{d}t F_2^{\mathrm{gap}}(t)$ in [31], is not yet known, and so we estimate the density $\mathrm{d}/\mathrm{d}t F_1^{\mathrm{gap}}(t)$ 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 1.2 that the convergence of the left-hand side in (1.6) to F_β^{gap} is slow. In fact, we expect a rate proportional to $1/\log N$. This means that in order to demonstrate (1.6) numerically with convincing accuracy one would have to consider very large values of N. In order to display the convergence in (1.6) for more reasonable values of N, we observe, using a simple calculation, that for any fixed $\gamma \ne 0$ the limiting distribution of

$$\widetilde{T}^{(1)} = \widetilde{T}^{(1)}_\gamma := \frac{T^{(1)}}{c_V^{2/3} 2^{-2/3} N^{2/3} (\log \epsilon^{-1} - \frac{2}{3}\log N + \gamma)} \tag{1.9}$$

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 $\widehat{T}$ in (3.1):

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

---

[7] 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 1.1. A numerical demonstration of (1.11). The third row of the table confirms that (ϵ, N) is in the scaling region for, say, $\sigma = \frac{1}{2}$. The last four rows demonstrate that the ratio of the sample mean to the sample standard deviation is order 1.

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

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

where γ is chosen to make

$$T^* - \widehat{T}_\gamma = \frac{\log N^{2/3}(\lambda_N - \lambda_{N-1}) + \frac{1}{2}\log \nu_{N-1} - \gamma}{\lambda_N - \lambda_{N-1}} \tag{1.10}$$

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 (1.10) with the expectation of their respective limiting distributions as $N \to \infty$ (see Theorem 1.9: 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 1.3 and 1.4 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 1.1 using a million samples and $\epsilon = 10^{-5}$ that the sample standard deviation is on the same order as the sample mean:

$$\sigma_{T^{(1)}} \sim \langle T^{(1)}\rangle \sim N^{2/3}\left(\log \epsilon^{-1} - \frac{2}{3}\log N\right). \tag{1.11}$$

Remark 1.6. The ideas that allow us to establish (1.7) for IEs requires the convergence of

$$\mathbb{E}\left[\frac{1}{N^{2/3}(\lambda_N - \lambda_{N-1})}\right]. \tag{1.12}$$

For BUE, (1.12) must be infinite for all N as there is a nonzero 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 (1.7) still holds in this case of discrete WEs even though (1.12) is infinite. Specifically, the convergence in the definition of ξ (Definition 1.10) for discrete WEs cannot take place in expectation. Hence $T^{(1)}$ acts as a mollified version of the inverse of the top gap—it is always finite.

1.3 Estimates from Random Matrix Theory

We now introduce the results from random matrix theory that are needed to prove Theorem 1.2 and Proposition 1.3. 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. We assume the entries of H are distributed according to an invariant or generalized Wigner ensemble (see the Appendix). 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.

THEOREM 1.7 (Equilibrium measure [3]). *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, for suitable constants C_μ and c_V, has a density ρ that satisfies $\rho(x) \leq C_\mu \sqrt{b_V - x}\,\chi_{(-\infty, b_V]}(x)$ and $\rho(x) = \frac{2^{3/4} c_V}{\pi} \sqrt{b_V - x}(1 + \mathscr{O}(b_V - x))$ 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 $c_V = 1$ [3]. 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 1.11), 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 $0 < p < \frac{1}{3}$, then we choose $s < \min\{\frac{\sigma}{44}, \frac{p}{8}\}$, and then finally $0 < c \leq \frac{10}{\sigma}$ will be a constant that will allow us to estimate the size of various sums. The specific meanings of the first three parameters are given below. Also, C denotes 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 $N > N^*$. For convenience in what follows, we use the notation $\epsilon = N^{-\alpha/2}$, so

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

and $\alpha = \alpha_N$ is allowed to vary with N. Our calculations that follow involve first deterministic estimates and then probabilistic estimates. The following conditions provide the setting for the deterministic estimates.

Condition 1. For $0 < p < \frac{\sigma}{4}$,

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

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

Condition 2. For any fixed $0 < s < \min\{\frac{\sigma}{44}, \frac{p}{8}\}$

(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 1.8. 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 [16, 25], for example.

The following theorem has its roots in the pursuit of proving universality in random matrix theory. See [29] for the seminal result when $V(x) = x^2$ and $\beta = 2$. Further extensions include the works of Soshnikov [24] and Tao and Vu [27] for Wigner ensembles and [6] for invariant ensembles.

THEOREM 1.9. *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 i.i.d. 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* 1.

PROOF. The first claim follows from [4, theorem 1.2]. The second claim follows from [3, cor. 2.2, theorem 2.7]. The last claim follows from [23, theorem 1.1]. □

DEFINITION 1.10. 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 [19, 21, 31], 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. These results show that the conditions described above hold with arbitrarily high probability.

THEOREM 1.11. *For WEs or IEs Condition* 2 *holds with high probability as $N \to \infty$; that is, for any $s > 0$*

$$\mathbb{P}(R_{N,s}) = 1 + o(1) \quad \text{as } N \to \infty.$$

PROOF. We first consider WEs. The fact that the probability of Condition 2(1) tends to unity follows from [14, theorem 2.1] using estimates on the (1,1)-entry of the Green's function. See [12, sec. 2.1] for a discussion of using these estimates. The fact that the probability of each of Condition 2(2)–(3) tends to unity follows from Theorem 1.9 using Corollary 3.3. Finally, the statement that the probability Condition 2(4) tends to unity as $N \to \infty$ is the statement of the rigidity of eigenvalues, the main result of [14]. Following Remark 1.8, we then have that the probability of Condition 2(1)-(2) tends to unity for IEs.

For IEs, the fact that the probability of Condition 2(4) tends to unity follows from [4, theorem 2.4]. Again, the fact that the probability of Condition 2(3) tends to unity follows from Theorem 1.9 using Corollary 3.2. □

THEOREM 1.12. *For both WEs and IEs*

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

PROOF. It follows from Theorem 1.9 that

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

Then

$$\begin{aligned}\lim_{p\downarrow 0} &\mathbb{P}(\Lambda_{3,\beta} - \Lambda_{2,\beta} < p(\Lambda_{2,\beta} - \Lambda_{1,\beta})) \\ &= \mathbb{P}\Big(\bigcap_{p>0}\{\Lambda_{3,\beta} - \Lambda_{2,\beta} < p(\Lambda_{2,\beta} - \Lambda_{1,\beta})\}\Big) \\ &= \mathbb{P}(\Lambda_{3,\beta} = \Lambda_{2,\beta}).\end{aligned}$$

But from [23, 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 1.13. *Let* $0 < c < 10/\sigma$. *Given Condition* 2

$$|I_c^{\mathsf{c}}| \le N^{2s} \quad \textit{for } N \textit{ sufficiently large,}$$

where $^{\mathsf{c}}$ *denotes the complement relative to* $\{1, 2, \dots, N-1\}$.

PROOF. We use rigidity of the eigenvalues, Condition 2(4). So, $|\lambda_n - \gamma_n| \le N^{-2/3+s}(\widehat{n})^{-1/3}$ where $\widehat{n} = \min\{n, N-n+1\}$. Recall

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

Define

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

If $n \in I_c$, then

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

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

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

Then using Definition 1.7, $\gamma_N = b_V$, and

$$n^* \ge b_V - (2 + c + (\widehat{n})^{-1/3})N^{-2/3+s} \ge b_V - (3+c)N^{-2/3+s},$$

we see that

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

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

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

LEMMA 1.14. *Given Condition 2, $0 < c \leq 10/\sigma$, and $j \leq 3$, there exists an absolute constant C such that*

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

for N sufficiently large.

PROOF. For $j \leq 3$

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

It also follows that $\lambda_N - \lambda_1 \leq b_V - a_V + 1$ so that by Lemma 1.13 for sufficiently large N

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

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

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

2 Estimates for the Toda Algorithm

Remarkably, (1.2) can be solved explicitly by a QR factorization procedure; see, for example, [26]. For $X(0) = H$ we have for $t \geq 0$

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

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 applying Gram-Schmidt to the columns of e^{tH}. We claim that $X(t) = Q^*(t)HQ(t)$

is the solution of (1.2). Indeed, by differentiating, we obtain

$$He^{tH} = HQ(t)R(t) = \dot{Q}(t)R(t) + Q(t)\dot{R}(t),$$

$$X(t) = Q^*(t)\dot{Q}(t) + \dot{R}(t)R^{-1}(t). \tag{2.1}$$

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 (1.2) 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 (1.2).

As the eigenvalues of $X(0) = H$ are not necessarily simple (indeed for BOE there is a nonzero 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 $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$. Then $X(t) = Q^*(t)HQ(t) = U(t)\Lambda U^*(t)$ where $U(t) = Q^*(t)U_0$. Then the j^{th} 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 [20])

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

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 $i = 1, 2, \ldots, m$,

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

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

$$\frac{\mathrm{d}}{\mathrm{d}t}[S^*(t)S(t)] = -S^*(t)D(t)S(t) + S^*(t)D(t)S(t) = 0$$

and hence $S^*(t)S(t) = C = I_m$; i.e., $S(t)$ is unitary. In particular, $\widetilde{U}_j(t) := U_j(t)S(t)$ has orthonormal columns and we find

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

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

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

Also

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

and hence

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

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

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

A direct calculation shows that

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

where

$$\begin{aligned}E_0(t) &= \frac{1}{4} \frac{\sum_{n=1}^{N-1} \delta_n^2 \nu_n e^{-\delta_n t}}{s\left(1 + \sum_{n=1}^{N-1} \nu_n e^{-\delta_n t}\right)^2}, \\ E_1(t) &= \frac{\left(\sum_{n=1}^{N-1} \lambda_n^2 \nu_n e^{-\delta_n t}\right)\left(\sum_{n=1}^{N-1} \nu_n e^{-\delta_n t}\right) - \left(\sum_{n=1}^{N-1} \lambda_n \nu_n e^{-\delta_n t}\right)^2}{\left(1 + \sum_{n=1}^{N-1} \nu_n e^{-\delta_n t}\right)^2}.\end{aligned}$$

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

To aid the reader we provide a glossary to summarize inequalities for parameters and quantities that have previously appeared:

(1) $0 < \sigma < 1$ is fixed,
(2) $0 < p < \frac{1}{3}$,
(3) $\alpha \geq \frac{10}{3} + \sigma$,
(4) $s \leq \min\{\frac{\sigma}{44}, \frac{p}{8}\}$,
(5) $\alpha - \frac{4}{3} - 44s \geq 2$,
(6) $c \leq \frac{10}{\sigma}$ can be chosen for convenience line by line when estimating sums with Lemma 1.14,
(7) $\delta_n = 2(\lambda_N - \lambda_n)$,
(8) $\nu_n = \beta_n^2 / \beta_N^2$,
(9) given Condition 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 (1.4)) is defined to be

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

We find bounds on the halting time.

LEMMA 2.2. *Given Condition* 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 $E_0(t) > \epsilon^2$, for $0 \le t \le \frac{\sigma}{2}\log N/\delta_{N-1}$ 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 1.14

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

Then using Lemma 1.14 we have

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

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

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

for some new constant $C > 0$. This last inequality follows because $N^{4s} \le N^{1+s}$ as $s \le 1/44$ (see Condition 2). But then from Definition 1.1 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 $c = 2(2+s)/\sigma \le 10/\sigma$

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

for sufficiently large N. Here we used that $s \le \sigma/44$. This shows that

$$(\alpha - 4/3 - 5s)\log N/\delta_{N-1} \le T^{(1)} \quad \text{for } N \text{ sufficiently large.}$$

Now, we work on the upper bound. Letting $t = a\log N/\delta_{N-1}$ for $a \ge (\alpha - 4/3 + 7s)$, we find using Lemma 1.14 that

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

Then using the minimum value for a, we obtain

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

It follows from Definition 1.1 that $a \ge 10/3 + \sigma - 4/3 + 7s > 2$. If we set $c = 2$ and use $s \le 1/44$, then $1 + 7s - ca + 4/3 \le -3 + 4/3 + 7s \le -2$ and

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

for sufficiently large N.

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

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

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

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

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

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

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

LEMMA 2.3. *Given Condition* 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 1.14

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

For $t \in I_\alpha$,

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

Next, again by Lemma 1.14

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

Then estimate with $c = 2$,

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

where we used $t \geq 2\log N/\delta_{N-1}$ and $s \leq 1/44$. Furthermore,

$$e^{-2\delta_{N-1}t} \leq N^{-\alpha}N^{8/3-\alpha+10s} \leq N^{-\alpha-2/3-\sigma+10s} \quad \text{as } s \leq \sigma/44.$$

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

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

LEMMA 2.4. *Given Condition* 2

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

PROOF. This follows immediately from the statements

$$N^{-2s} \leq \nu_{N-1} \leq N^{2s},$$
$$2N^{-2/3-s} \leq \delta_{N-1} \leq 2N^{-2/3+s}.$$ □

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

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

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

$$\begin{aligned} |T^{(1)} - T^*| &\leq \frac{|N^{-\alpha} - E_0(T^*) - E_1(T^{(1)})|}{\min_{\eta \in I_\alpha} |E_0'(\eta)|} \\ &\leq \frac{|N^{-\alpha} - E_0(T^*)| + \max_{\eta \in I_\alpha} |E_1(\eta)|}{\min_{\eta \in I_\alpha} |E_0'(\eta)|}. \end{aligned} \tag{2.10}$$

See Figure 2.1 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 $|N^{-\alpha} - E_0(T^*)|$.

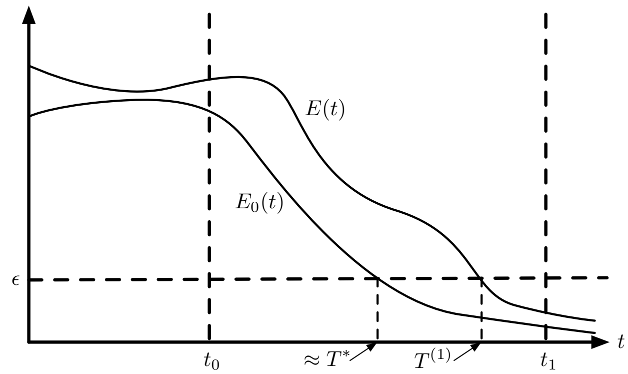

FIGURE 2.1. A schematic for the relationship between the functions $E_0(t)$ and $E(t)$ and the times $T^{(1)}$ and T^*. Here $t_0 = (\alpha - 4/3 - 5s)\log N/\delta_{N-1}$ and $t_1 = (\alpha - 4/3 + 7s)\log N/\delta_{N-1}$. Note that E_0 is monotone on $[t_0, t_1]$.

LEMMA 2.5. *Given Conditions* 1 *and* 2

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

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 1.14 to find

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

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

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

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

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

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

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

as $p < 1/3$ and hence $c > 2 + p$. Using Lemma 1.13 to estimate $|\widehat{I}_c^{\mathrm{c}}|$

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

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

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

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

Thus

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

From this it follows that

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

LEMMA 2.6. *Given Conditions* 1 *and* 2*,* $s < \min\{\sigma/44, p/8\}$*, and* σ *and* p *fixed,*

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

PROOF. Combining Lemmas 2.3 and 2.5 with (2.6), which can be extended to give $E_1(t) \le N^{-\alpha-2/3-\sigma/2}$, and (2.10), we have for sufficiently large N

$$\begin{aligned}
N^{-2/3}|T^{(1)} - T^*| &\le CN^{-2/3}N^{\alpha+12s+2/3}(N^{-\alpha-2p+4s} + N^{-\alpha}N^{-2/3-\sigma/2}) \\
&\le C(N^{-2p+16s} + N^{-\sigma/2+12s}),
\end{aligned}$$

where we used $\alpha - \frac{8}{3} > \frac{2}{3}$. Since $p < \frac{1}{3}$ the right-hand side is bounded by $CN^{-2p+16s}$, which goes to 0 as $N \to \infty$ provided that $s < \frac{p}{8}$, $p < \frac{\sigma}{4}$. $\square$

From (2.4), we have

$$|\lambda_N - X_{11}(t)| = \frac{1}{2} \frac{\sum_{n=1}^{N-1} \delta_n \nu_n \mathrm{e}^{-\delta_n t}}{s1 + \sum_{n=1}^{N-1} \nu_n \mathrm{e}^{-\delta_n t}} \le \frac{1}{2} \sum_{n=1}^{N-1} \delta_n \nu_n \mathrm{e}^{-\delta_n t}.$$

LEMMA 2.7. *Given Condition* 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 1.14 with $c = 1$. By 2.2 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 2 and 1 as stated in Theorems 1.11 and 1.12 to understand $T^{(1)}$ 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 0 *in probability as* $N \to \infty$.

PROOF. Let $\eta > 0$. Then

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

If s satisfies the hypotheses in Lemma 2.6, $s < \min\{\sigma/44, p/8\}$, 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

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

and by Theorem 1.11

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

This is true for any $0 < p < \frac{1}{3}$, and we use Theorem 1.12. So, as $p \downarrow 0$, we find

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

Define

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

We need the following simple lemmas in what follows.

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

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

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

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

$$\mathbb{P}(a < X_N \le b) \to \mathbb{P}(a < X \le b).$$

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

$$\begin{aligned}\liminf_{N\to\infty} \mathbb{P}(-a_N < X_N < a_N) &\ge \liminf_{N\to\infty} \mathbb{P}(-M < X_N \le M)\\ &= \mathbb{P}(-M < X \le M).\end{aligned}$$

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

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

COROLLARY 3.3. *If $X_N \to X$ in distribution as $N \to \infty$, then*

$$|X_N/a_N|$$

converges to 0 in probability provided $a_N \to \infty$.

LEMMA 3.4. *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 $\mathbb{P}(X \le t)$; then for $\eta > 0$,

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

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

$$\begin{aligned}\mathbb{P}(X_N \le t - \eta) &\le \mathbb{P}(Y_N \le t) + \mathbb{P}(|Y_N - X_N| > \eta)\\ &\le \mathbb{P}(X_N \le t + \eta) + 2\mathbb{P}(|Y_N - X_N| > \eta).\end{aligned}$$

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

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

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

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

Now, we compare T^* with $\widehat{T}$.

LEMMA 3.5. *For* $\alpha \geq 10/3 + \sigma$

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

converges to 0 *in probability as* $N \to \infty$.

PROOF. Consider

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

For

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

we have $\mathbb{P}(L_N^{\mathsf{c}}) + \mathbb{P}(U_N^{\mathsf{c}}) + \mathbb{P}(P_n^{\mathsf{c}}) \to 0$ as $N \to \infty$ by Lemma 3.2 and Theorem 1.9. For these calculations it is important that the limiting distribution function for $N^{2/3}\delta_{N-1}$ be continous at 0; see Theorem 1.9. Then for $\eta > 0$

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

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

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

Hence the first term on the right-hand side of (3.2) is 0 for sufficiently large N and the second term is bounded by $\mathbb{P}(U_N^{\mathsf{c}}) + \mathbb{P}(L_N^{\mathsf{c}}) + \mathbb{P}(P_N^{\mathsf{c}})$, which tends to 0. This shows convergence in probability. □

We now arrive at our main result.

THEOREM 3.6. *If* $\alpha \geq \frac{10}{3} + \sigma$ *and* $\sigma > 0$, *then*

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

PROOF. Combining Lemma 3.1 and Lemma 3.5, we have that

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

converges to 0 in probability. Then by Lemma 3.4 and Theorem 1.9 the result follows as

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

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

PROPOSITION 3.7. *For $\alpha \ge \frac{10}{3}+\sigma$, $\sigma>0$, and any $q<1$,*

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

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

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

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

PROOF. We recall that $R_{N,s}$ is the set on which Condition 2 holds. Then for any $\eta>0$

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

Using Lemma 2.7, the first term on the right-hand side is 0 for sufficiently large N and the second term vanishes from Theorem 1.11. 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

$$\begin{aligned}
|\lambda_j - X_{11}(T^{(1)})| &\ge |\lambda_j-\lambda_N|-|\lambda_N - X_{11}(T^{(1)})|\\
&\ge |\lambda_{N-1}-\lambda_N|-|\lambda_N - X_{11}(T^{(1)})|,
\end{aligned}$$

and for sufficiently large N (see Lemma 2.7)

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

This tends to ∞ as $s < 1/3$ and $s < r$. Hence for any $K > 0$, again using the arguments of Theorem 3.6,

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

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 $\mathbb{P}(R^{\mathsf{c}}_{N,s})$ and hence

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

Next, under the same assumption (Condition 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 3.3 and Theorem 1.9 by using $\gamma_N = b_V$

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

converges to 0 in probability (with no point mass at 0), implying its inverse converges to ∞ in probability. This shows $N^\alpha|b_V - X_{11}(T^{(1)})|$ converges to ∞ in probability. □

Appendix: Invariant and Wigner Ensembles

The following definitions are taken from [4, 5, 13]. The first definition appeared initially in [14] and was made more explicit in [13]. These are the two classes of random matrices to which our results apply.

DEFINITION A.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 A.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[9] 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 $\mathrm{d}H = \prod_{i\le j}\mathrm{d}H_{ij}$ if $\beta = 1$ and $\mathrm{d}H = \prod_{i=1}^N \mathrm{d}H_{ii}\prod_{i<j}\mathrm{d}\operatorname{Re} H_{ij}\,\mathrm{d}\operatorname{Im} H_{ij}$ if $\beta = 2$.

Acknowledgments. The authors would like to thank Yuri Bakhtin and Paul Bourgade for many useful conversations and the anonymous referee for valuable suggestions. 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).

Bibliography

[1] Adler, M. On a trace functional for formal pseudo differential operators and the symplectic structure of the Korteweg–de Vries type equations. *Invent. Math.* **50** (1978/79), no. 3, 219–248. doi:10.1007/BF01410079

[2] Bourgade, P. Personal communication, 2016.

[3] Bourgade, P.; Erdős, L.; Yau, H.-T. Edge universality of beta ensembles. *Comm. Math. Phys.* **332** (2014), no. 1, 261–353. doi:10.1007/s00220-014-2120-z

[4] Bourgade, P.; Yau, H.-T. The eigenvector moment flow and local quantum unique ergodicity. Preprint, 2013. arXiv:1312.1301 [math.PR]

[5] Deift, P. A. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach.* Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, R.I., 1999.

[6] Deift, P.; Gioev, D. Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices. *Comm. Pure Appl. Math.* **60** (2007), no. 6, 867–910. doi:10.1002/cpa.20164

[7] Deift, P.; Li, L. C.; Nanda, T.; Tomei, C. The Toda flow on a generic orbit is integrable. *Comm. Pure Appl. Math.* **39** (1986), no. 2, 183–232. doi:10.1002/cpa.3160390203

[8] Deift, P.; Li, L. C.; Tomei, C. Toda flows with infinitely many variables. *J. Funct. Anal.* **64** (1985), no. 3, 358–402. doi:10.1016/0022-1236(85)90065-5

[9] Deift, P. A.; Menon, G.; Olver, S.; Trogdon, T. Universality in numerical computations with random data. *Proc. Natl. Acad. Sci. USA* **111** (2014), no. 42, 14973–14978. doi:10.1073/pnas.1413446111

[10] Deift, P.; Nanda, T.; Tomei, C. Ordinary differential equations and the symmetric eigenvalue problem. *SIAM J. Numer. Anal.* **20** (1983), no. 1, 1–22. doi:10.1137/0720001

[11] Deift, P.; Trogdon, T. Universality for eigenvalue algorithms on sample covariance matrices. Preprint, 2017. arXiv:1701.01896 [math.NA]

[12] Erdős, L. Universality for random matrices and log-gases. *Current developments in mathematics 2012*, 59–132. International Press, Somerville, Mass., 2013.

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

[13] Erdős, L.; Knowles, A.; Yau, H.-T.; Yin, J. The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** (2013), no. 59, 58 pp. doi:10.1214/EJP.v18-2473

[14] Erdős, L.; Yau, H.-T.; Yin, J. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** (2012), no. 3, 1435–1515. doi:10.1016/j.aim.2011.12.010

[15] Flaschka, H. The Toda lattice. I. Existence of integrals. *Phys. Rev. B (3)* **9** (1974), 1924–1925. doi:10.1103/PhysRevB.9.1924

[16] Jiang, T. How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.* **34** (2006), no. 4, 1497–1529. doi:10.1214/009117906000000205

[17] Kostant, B. The solution to a generalized Toda lattice and representation theory. *Adv. in Math.* **34** (1979), no. 3, 195–338. doi:10.1016/0001-8708(79)90057-4

[18] Manakov, S. V. Complete integrability and stochastization of discrete dynamical systems. *Soviet Physics JETP* **40** (1974), no. 2, 269–274 (1975); translated from *Ž. Èksper. Teoret. Fiz.* **67** (1974), no. 2, 543–555.

[19] Monthus, C.; Garel, T. Typical versus averaged overlap distribution in spin glasses: Evidence for droplet scaling theory. *Phys. Rev. B* **88** (2013), no. 13, 134204. doi:10.1103/PhysRevB.88.134204

[20] Moser, J. Three integrable Hamiltonian systems connected with isospectral deformations. *Advances in Math.* **16** (1975), 197–220. doi:10.1016/0001-8708(75)90151-6

[21] Perret, A.; Schehr, G. Near-extreme eigenvalues and the first gap of Hermitian random matrices. *J. Stat. Phys.* **156** (2014), no. 5, 843–876. doi:10.1007/s10955-014-1044-5

[22] Pfrang, C. W.; Deift, P.; Menon, G. How long does it take to compute the eigenvalues of a random symmetric matrix? *Random matrix theory, interacting particle systems, and integrable systems*, 411–442. Mathematical Sciences Research Institute Publications, 65. Cambridge University Press, New York, 2014.

[23] Ramírez, J. A.; Rider, B.; Virág, B. Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.* **24** (2011), no. 4, 919–944. doi:10.1090/S0894-0347-2011-00703-0

[24] Soshnikov, A. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (1999), no. 3, 697–733. doi:10.1007/s002200050743

[25] Stam, A. J. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *J. Appl. Probab.* **19** (1982), no. 1, 221–228.

[26] Symes, W. W. The QR algorithm and scattering for the finite nonperiodic Toda lattice. *Phys. D* **4** (1981/82), no. 2, 275–280. doi:10.1016/0167-2789(82)90069-0

[27] Tao, T.; Vu, V. Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** (2010), no. 2, 549–572.

[28] Toda, M. Vibration of a chain with nonlinear interaction. *J. Phys. Soc. Japan* **22** (1967), no. 2, 431–436. doi:10.1143/JPSJ.22.431

[29] Tracy, C. A.; Widom, H. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** (1994), no. 1, 151–174.

[30] Watkins, D. S. Isospectral flows. *SIAM Rev.* **26** (1984), no. 3, 379–391. doi:10.1137/1026075

[31] Witte, N. S.; Bornemann, F.; Forrester, P. J. Joint distribution of the first and second eigenvalues at the soft edge of unitary ensembles. *Nonlinearity* **26** (2013), no. 6, 1799–1822. doi:10.1088/0951-7715/26/6/1799

PERCY DEIFT
Courant Institute
251 Mercer St.
New York, NY 10012
USA
E-mail: deift@cims.nyu.edu

THOMAS TROGDON
University of California, Irvine
Rowland Hall
Irvine, CA 92697-3875
USA
E-mail: ttrogdon@math.uci.edu

Received May 2016.