Universality of the Stochastic Airy Operator

Manjunath Krishnapur, Brian Rider
and Bálint Virág

June 25, 2013

Abstract

We introduce a new method for studying universality of random matrices. Let $T_n$ be the Jacobi matrix associated to the Dyson beta ensemble with uniformly convex polynomial potential. We show that after scaling, $T_n$ converges to the Stochastic Airy operator. In particular, the top edge of the Dyson beta ensemble and the corresponding eigenvectors are universal. As a byproduct, our work leads to conjectured operator limits for the entire family of soft edge distributions.
## Contents

1 Introduction
   1.1 Methods of the proof ........................................ 5

2 Tridiagonal models ........................................ 6

3 Local minimizers ........................................ 8

4 Outline of the main argument ............................ 13
   4.1 Limits of random tridiagonal operators ................... 13
   4.2 Application to the Dyson $\beta$ ensembles ............... 16

5 Bounding minimizers ........................................ 17
   5.1 Upper bounds .............................................. 18
   5.2 Lower bounds .............................................. 22

6 Minimizers and boundary conditions ................... 23

7 Conditional and local minimizers are close ............ 26

8 Bounding the field ........................................... 30
   8.1 A Gaussian lemma ............................................ 30
   8.2 Upper and lower bounds on $(A, B)$ ....................... 31

9 Gaussian approximation of the field .................... 33
   9.1 Concentration about the minimizers ....................... 34
   9.2 $(A, B)$ as a mixture of Gaussians ....................... 35

10 A non-universal operator limit in the global scaling . 38

11 Mean and variance of the limiting potential ............. 45

12 Putting it all together ..................................... 48
   12.1 Tightness and Brownian convergence .................... 49
   12.2 Oscillation and growth bounds ........................... 51
   12.3 Proof of the main result ................................ 54

13 The nonregular case ........................................ 55
1 Introduction

The goal of this paper is to introduce a novel approach to universality of random matrices. We consider Dyson’s beta ensembles: these are $n$ random points on the real line with probability density
\[
\frac{1}{Z_{n,\beta}} e^{-\beta n \sum_{k=1}^{n} V(\lambda_k)} \prod_{j<k} |\lambda_j - \lambda_k|^{\beta},
\]
where $V$ is polynomial, and we assume $V'' \geq c_0 > 0$. We show that the distribution of the top points converges to a universal limit that does not depend on $V$. There have been two approaches to universality for such ensembles. The first, classical method (see the book of Deift [9]) is based on asymptotics of the orthogonal polynomials, and is tied to the special values $\beta = 1, 2, 4$. The more recent method, carried out in the bulk by Bourgade, Erdős, and Yau [2, 3], is based on the study the dynamics given by versions of Dyson’s Brownian motion.

Just as in the classical method, our starting point is the theory of orthogonal polynomials. Recall that for a probability measure $\pi$ supported on exactly $n$ points there exists a unique $n \times n$ Jacobi matrix $T$ (i.e., tridiagonal symmetric matrix with positive off-diagonals), so that the spectral measure of $T$ at the first coordinate vector $e_1$ is $\pi$. Consider the Jacobi matrix $T_n = T_n(V, \beta)$ associated to the random measure with support points picked form (1) and with independent Dirichlet $(\beta/2, \ldots, \beta/2)$ distributed weights (8).

We study the structure of the matrix $T_n$ and show that after scaling, it converges as an operator to a unique random limit depending on $\beta$ only. This, in particular, implies the universality of the joint distribution of top eigenvalues: its limit does not depend on $V$.

**Theorem 1.** There exists a coupling of the random matrices $T_n$ on the same probability space and constants $\gamma, \vartheta, \mathcal{E}$ depending on $V$ only (and specified in Remark 4) so that a.s. we have
\[
\gamma n^{2/3} (\mathcal{E} - T_n) \to \text{SAO}_\beta
\]
in the norm-resolvent sense: for every $k$ the bottom $k$th eigenvalue converges the and corresponding eigenvector converges in norm. Here $\mathcal{E} - T_n$ acts on $\mathbb{R}^n \subset L^2(\mathbb{R}_+)$ with coordinate vectors $e_j = (\vartheta n)^{1/6} 1_{|j-1,j|}(\vartheta n)^{-1/3}$.

For the special values of $\beta = 1, 2, 4$, Theorem 1 strengthens (to operator convergence) some previously known results [8]. For those cases, (1) describes the eigenvalue distribution of a random matrix $\Upsilon$ with real, complex or quaternion entries, respectively. Then,
$T_n$ is simply $\Upsilon$ written in the basis given by the Gram-Schmidt procedure with input $e_1, \Upsilon e_1, \ldots, \Upsilon^{n-1} e_1$.

The limiting object $\text{SAO}_\beta$ of Theorem 1 is the Stochastic Airy Operator, a second order differential operator with random potential defined by

$$\text{SAO}_\beta = -\frac{d^2}{dx^2} + x + \frac{2}{\sqrt{\beta}} W'(x). \quad (2)$$

Here $x \mapsto W(x)$ is a standard Brownian motion, and $\text{SAO}_\beta$ acts on a dense subset of $L^2(\mathbb{R}^\uparrow)$ with Dirichlet boundary conditions. $\text{SAO}_\beta$ was introduced in Edelman and Sutton [12] and Ramírez, Rider and Virág [19]. Theorem 1 was proved in [19] for $V = x^2/4$, when $T_n$ has a particularly simple form found by Dumitriu and Edelman [11]. The GOE and GUE are special cases. The paper [19] also establishes some basic criteria for the convergence of random tridiagonal operators. The proof of Theorem 1 relies on [19], but apart from that and a few classical facts about orthogonal polynomials it is self-contained.

The eigenvalue/eigenvector pairs $(\Lambda_k, f_k)$ of $\text{SAO}_\beta$ can also be defined via the variational formalism. We iteratively define

$$\Lambda_k = \inf_{f \in L, f \perp f_{k-1}} \int_0^\infty [(f')^2(x) + x f^2(x)] \, dx + \frac{2}{\sqrt{\beta}} \int_0^\infty f^2(x) \, dW_x,$$

where $L$ is the space of functions satisfying $f(0) = 0$, $\int_0^\infty f^2 = 1$, along with $\int_0^\infty [(f')^2 + x f^2] < \infty$. That working on the space $L$ makes the stochastic integral in (3) sensible as well as the form bounded below (almost surely) is one part of what is proved in [19].

The top eigenvalue of $-\text{SAO}_\beta$ has the so-called Tracy-Widom-$\beta$ distribution. The representation (2) has been been used to study rank-one deformations, to give a quick derivation of the Painlevé formulas for the TW$_\beta$ distribution for $\beta = 2, 4$, [4] and for precise tail bounds [10].

Most of this paper consists of the proof of Theorem 1, and we will conclude the introduction with an outline and motivation. But first, a conjecture, which is supported by further evidence in Section 13.

The empirical distribution of eigenvalues of $T_n$, without scaling, converge to the classical equilibrium measure form potential theory corresponding to $V$ (see Remark 8 below). The convexity and analyticity of $V$ forces this measure to have a density which is decays like $x^{1/2}$ at the edges. As one might guess, this $x^{1/2}$ is crucial for the SAO$_\beta$ limit. When $V$ is analytic, the possible decay rates are $x^{2k+1/2}$ for some integer $k$. We conjecture (see Conjecture 45) that after scaling, $T_n$ in this case converges to the random operator

$$S_{\beta,k} = -\frac{d^2}{dx^2} + x^{\frac{2k+1}{2}} + \frac{2}{\sqrt{\beta}} x^{-\frac{k}{\beta + 1}} W'(x).$$
For $\beta = 2$ the eigenvalue limits have been studied in [6] via the Riemann-Hilbert approach.

1.1 Methods of the proof

To explain, begin with the $\beta$-Hermite ensembles, i.e. the case $V = x^2/4$. For this case, Theorem 1 had been conjectured to hold by Edelman and Sutton [12, 21]. Their reasoning, as well as the rigorous proof in [19], makes essential use of the discovery by Dumitriu and Edelman of a simple tridiagonal matrix model for the $\beta$-Hermite ensembles [11] (see also the earlier work of Trotter [25] for the classical $\beta$ case). Let

$$H_n = \frac{1}{\sqrt{n\beta}} \begin{bmatrix} g_1 & \chi_{(n-1)\beta} \\ \chi_{(n-1)\beta} & g_2 & \chi_{(n-2)\beta} \\ & \ddots & \ddots & \ddots \\ & & \chi_{2\beta} & g_{n-1} & \chi_{\beta} \\ & & & \chi_{\beta} & g_n \end{bmatrix},$$

in which the $g_k$ are Gaussian random variables of mean 0 and variance 2, the $\chi_k$ are $\chi$ random variables indexed by the shape parameter, and all variable are independent save for the condition that $H_n$ is symmetric. The fact is that the eigenvalues of $H_n$ realize the law (1) for $V = x^2/4$ and all $\beta > 0$.

The heuristic behind the Edelman-Sutton conjecture can then be gleaned from the asymptotic assessment: in distribution, $\chi_{n-k} \sim \sqrt{n} + k/n + g$ for a standard Gaussian random variable $g$. One thus sees that, to leading order, the top corner of $H_n - 2I_n$ resembles the discrete second derivative operator. The corrections can be viewed as an additive potential which is of type linear plus Gaussian noise. This heuristic guides the proof of [19] in which the Stochastic Airy Operator is identified by showing that, after “centering” $H_n$ by the appropriate second derivative operator, the running sum of the process of entries (the integrated potential) converges to $\frac{1}{2}x^2 + \frac{2}{\sqrt{\beta}}W(x)$. In this way Tracy-Widom limits are obtained as a consequence of a simple functional central limit theorem.

Here we continue this approach. First, we establish a tridiagonal representation for general $V$. In Proposition 2 we show that the diagonal $(A_1, \ldots, A_n)$ and off-diagonal $(B_1, \ldots, B_{n-1})$ entries of $T_n$ have joint density

$$ce^{-n\beta H}, \quad \text{where } H = H(a, b) = \text{tr}(V(T)) - \sum_{k=1}^{n-1} (1 - k/n - 1/(n\beta)) \log(b_k).$$

(5)
It follows from the path expansion of \( \text{tr}(T) \) that while the entries are not independent any more, they have a certain Markov field property. Indeed, variables with indices that are more than \( \deg V/2 \) apart are conditionally independent given the variables in between. Our goal is to prove the required central limit theorem and tightness conditions for these variables.

The Markov field property suggests that one could understand the distribution of the \((A, B)\) through studying some equilibrium measure of a Markov chain. One issue is that the distribution is not homogeneous in \( k \). However, one expects that mixing happens reasonably fast (in time \( \log n \)). In particular, some local metastable equilibria will develop, and that \((A, B)\) will be close to these local equilibria. In Section 3 we will study the location of these equilibria, and derive some properties of it.

Section 4 contains a far more extensive outline of the proof: we recall the criteria established in [19] and outline how they will be applied. Essentially, we have to establish tightness and a functional CLT for the variables \( A, B \).

The first step is to give rough bounds on the minimizers of \( H \). This is achieved in Section 5. Then, in Section 6 we study the minimizers of versions of the Hamiltonian \( H \)
5, and their dependence on boundary conditions. In the next section we show that they are close to the local minimizers studied in Section 3. In Section 8 we bound the random variables \((A, B)\), and in Section 9 we establish a Gaussian approximation.

One wrinkle is that first \( O(\log n) \) stretch of \((A, B)\) variables do not have a universal behavior. As a \((\deg V/2 - 1)\)-Markov process, \( k \mapsto (A_k, B_k) \) can be expected to require \( O(\log n) \) steps to achieve local equilibrium. Thus, we first study the submatrices \( T_{[c \log n, n]} \).

The effect of the truncation can be controlled by a rank-one perturbation. This is studied in Section 10. In Section 11 we compute the parameters of the CLT, and in Section 12 we complete the proof. Conjectures about the general (nonregular) edge case are discussed in Section 13.

### 2 Tridiagonal models

Let \( T = T(a, b) \) denote the symmetric tridiagonal matrix with \( T_{i,i} = a_i \) for \( i \leq n \) and \( T_{i,i+1} = T_{i+1,i} = b_i \) for \( i \leq n - 1 \).

Recall that for a symmetric matrix \( M \) the spectral measure of \( M \) at a unit vector \( v \) is the probability measure whose \( k \)th moment is \( \langle v, M^k v \rangle \). In the below we take \( v \) and to be the first coordinate vector \( e_1 \).
Proposition 2 (Matrix model). Let \((A,B)\) be sampled from the density
\[
\exp \left\{ -n\beta \left[ \text{tr}(V(T)) - \sum_{k=1}^{n-1} (1 - k/n - 1/(n\beta)) \log(b_k) \right] \right\}.
\]
(6)

Then the eigenvalues of \(T_n = T(A,B)\) have joint density proportional to
\[
\exp\left\{ -n\beta \sum_{j=1}^{n} V(\lambda_j) \right\} \prod_{j<k} |\lambda_j - \lambda_k|^\beta.
\]

Moreover, the weights \(q_i^2\) of the spectral measure \(\mu := \sum_{j=1}^{n} q_j^2 \delta_{\lambda_j}\) of \(T\) are independent with Dirichlet \((\frac{\beta}{2}, \ldots, \frac{\beta}{2})\) distribution.

Henceforth \(T\) will mean this random tridiagonal matrix and we shall assume that \(V\) is a polynomial with even degree and positive leading coefficient. It is not hard to show that in the classical \(\beta = 1, 2, 4\) cases \(T\) has the distribution of the random matrix \(M\) chosen from the probability distribution
\[
Z^{-1} \exp(-n\beta \text{Tr}(M)) \, dM,
\]
written in the orthonormal basis obtained from \(e_1, Me_1, M^2 e_1, \ldots\) via the Gram-Schmidt procedure.

Proof. By definition, \(\langle T^k e_1, e_1 \rangle = \int x^k d\mu(x)\) for any \(k \geq 0\). Consider this equation for each \(0 \leq k \leq 2n - 1\) and write them as
\[
\prod_{j=1}^{\ell} b_j^2 + f_\ell(a_j, b_j; j < \ell) = \sum_{j=1}^{n} q_j^2 \lambda_j^{2\ell} \quad \text{if } k = 2\ell
\]
\[
a_{\ell+1} \prod_{j=1}^{\ell} b_j^2 + g_\ell(a_j, b_j'; j < \ell, j' \leq \ell) = \sum_{j=1}^{n} q_j^2 \lambda_j^{2\ell+1} \quad \text{if } k = 2\ell + 1
\]

Equate the Jacobian determinant with respect to \((a,b)\) of the left side with the Jacobian determinant of the right side with respect to \((\lambda, q^2)\) to find that,
\[
2^{n-1} \prod_{j=1}^{n-1} b_j^{k(n-j) - 1} \, da_1 \cdots da_n \, db_1 \cdots db_{n-1} = \det(M) \, dq_1^2 \cdots dq_{n-1}^2 \, d\lambda_1 \cdots d\lambda_n
\]
in which
\[
M = \begin{bmatrix}
\lambda_1 - \lambda_n & \ldots & \lambda_{n-1} - \lambda_n & q_1^2 & \ldots & q_n^2 \\
\lambda_1^2 - \lambda_n^2 & \ldots & \lambda_{n-1}^2 - \lambda_n^2 & 2q_1^2 \lambda_1 & \ldots & 2q_n^2 \lambda_n \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\lambda_1^{2n-1} - \lambda_n^{2n-1} & \ldots & \lambda_{n-1}^{2n-1} - \lambda_n^{2n-1} & q_1^2 (2n - 1) \lambda_1^{2n-2} & \ldots & q_n^2 (2n - 1) \lambda_n^{2n-2}
\end{bmatrix}.
\]
Note here that $\sum_{k=1}^{n} q_k^2 = 1$. A special case of the confluent Vandermonde determinant identity [17] yields

$$\det(M) = \prod_{i=1}^{n} q_i^2 \prod_{i<j} (\lambda_i - \lambda_j)^4.$$ 

Further, there is the identity

$$\prod_{k=1}^{n-1} b_k^{2(n-k)} = \prod_{i=1}^{n} q_i^2 \prod_{i<j} (\lambda_i - \lambda_j)^2,$$ 

(7) see for example Section 3.1 of [9], expressing the $b_k$ in terms of the spectral measure of $T$.

It follows that, the measure with density $\exp\left\{-n\beta \text{tr}(V(T))\right\} \prod_{k=1}^{n-1} b_k^{\beta(n-k)-1}$ on $(a,b)$ transforms to the measure

$$\left( e^{-n\beta \sum_{k=1}^{n} V(\lambda_k)} \prod_{i<j} |\lambda_i - \lambda_j|^\beta d\lambda_1 \cdots d\lambda_k \right) \left( \prod_{k=1}^{n} q_k^\beta d\eta_1^2 \cdots d\eta_{n-1}^2 \right)$$ 

(8) for $(\lambda, q^2)$. In particular, $\lambda$ and $q$ are independent, $q^2$ has Dirichlet distribution with parameters $(\beta/2, \ldots, \beta/2)$, and $\lambda$ has the desired Coulomb gas distribution. 

\[ \text{Lemma 3 (Uniform Convexity). The function } (a, b) \mapsto \text{tr}(V(a,b)) \text{ is convex or uniformly convex along with } V. \text{ That is, its Hessian is bounded below by } c_u I \text{ for } c_u = \min_x V''(x). \]

\[ \text{Proof. That } \text{tr}(V(a,b)) \text{ is convex follows from Chandler Davis’s theorem [7]: any convex function of the eigenvalues is also a convex function of the entries. For uniform convexity, write } V = \tilde{V}(x) + c_u x^2 / 2, \text{ where } c_u = \min_x V''(x). \]

With now $V(a,b) := \text{tr}(V(a,b))$ it follows that

$$H = H(a, b) = \text{tr}(V(T)) - \sum_{k=1}^{n-1} \left( 1 - k/n - 1/(n\beta) \right) \log(b_k),$$ 

(9) shares its convexity properties with $V$ as long as $\beta \geq 1$. For $\beta < 1$ the same is true for $H$ restricted to the coordinates with indices $k \leq n - 1/\beta$ with the remaining $(a, b)$-values viewed as fixed.

3 Local minimizers

Intuitively, to first order the variables $(A, B)$ should be close to the global minimizers of the Hamiltonian $H$ in (9), which we will denote $(a^\circ, b^\circ)$. (For $\beta < 1$, this minimizer is defined to be subject to the “boundary” condition $b_k = 0$ for $k > n - 1/\beta$). The actual minimizers are difficult to characterize, but, as will show, for indices away from the boundaries
they are locally close to constant functions. The goal of this section is to describe these constants.

Introduce the local Hamiltonian: for fixed $x$,

$$H^{(x)}(a,b) = \text{tr}(V(C)) - \sum_{k=1}^{n} (1 - x) \log(b_k)$$

(10)

where $C$ is the symmetric circulant matrix with main diagonal given by the $a$’s, first off-diagonal given by the $b$’s and zeros elsewhere. This definition aims to mimic the local behavior of $H$ around the index about $k = xn$.

Since $H^{(x)}$ is convex, it has a unique minimizer; since a rotation of the indices does not change $H^{(x)}$, it follows that for the minimizer, all of the $a_i$’s have to be equal, and also all of the $b_i$’s have to be equal. Assuming this, note that as long as $n > \text{deg}(V)$, the expression $H^{(x)}(a,b)/n$ does not depend on $n$. Thus the location of the minimum

$$(a^\dagger = a^\dagger(x), b^\dagger = b^\dagger(x)) \in \mathbb{R}^2$$

(11)

does not depend on $n$. As functions of $x \in [0,1]$ these define the local minimizers. We will also have reason to consider:

$$a_k^\dagger = a^\dagger(k/n + 1/(n\beta)), \quad b_k^\dagger = b^\dagger(k/n + 1/(n\beta)),$$

(12)

the local minimizers corresponding to index $k$.

More concretely, introduce the function

$$W(a,b) = \frac{1}{\text{dim} C} \text{tr} V(C),$$

assuming $\text{dim} C > \text{deg} V$. Then $a^\dagger(x), b^\dagger(x)$, minimize the expression

$$W(a,b) - (1 - x) \log b.$$ 

(13)

The function $W$ may also be written as in

$$W(a,b) = [1]V(a + b(z + 1/z))$$

(14)

where $[1]$ denotes the coefficient of the constant term in the Laurent series in $z$. One way to understand (14) is by counting random walk paths. Another is to note that $C_{0,1}$ corresponds to the sum of the left and right shift operators on the discrete circle. In the Fourier basis it corresponds to by the multiplication operator of $z + 1/z$, and traces of multiplication operators are given by the constant term.
Remark 4 (Constants in the main theorem). We can now specify the constants \( \gamma, \vartheta, \mathcal{E} \) in Theorem 1 for easy reference. With \( a^\dagger(x), b^\dagger(x) \) defined quickly by (13, 14), we have \( \tau = -(a^\dagger)'(0) - 2(b^\dagger)'(0), \gamma = (b^\dagger(0))^{-1/3} \tau^{-2/3} \vartheta = b^\dagger(0)/\tau, \) and \( \mathcal{E} = a^\dagger(0) + 2b^\dagger(0) \).

For the rest of this section, we will drop the \( \dagger \) from the local minimizers. The most basic properties of the local minimizers as function of \( x \) are captured in the following.

Proposition 5. The minimizers \( x \mapsto a(x), b(x) \) of \( W(a, b) - (1 - x) \log b \) are unique and analytic as functions of \( x \in (-\infty, 1) \). Furthermore, they are continuous from the left at \( x = 1 \) with \( b(1) = 0 \).

Proof. Again, uniform convexity implies that \( (a(x), b(x)) \) are well-defined and unique. This is true even as local extrema; there cannot be local maxima and inflection points. Differentiating (13) we get that, at \( (a(x), b(x)) \):

\[
W_1(a(x), b(x)) = 0, \quad W_2(a(x), b(x)) - (1 - x)/b(x) = 0. \tag{15}
\]

Here the indices denote partial derivatives in the \( a \) and \( b \) variables.

Thinking of these as a pair of functions \( f \) of \( a, b, x \), the \( 2 \times 2 \) Jacobian matrix in \( a, b \) is exactly the Hessian of \( W - (1 - x) \log b \). By uniform convexity (Lemma 3) this is bounded away from zero for \( a \in \mathbb{R}, 1 - x, b \geq \epsilon \) and hence also for some complex neighborhood of these sets. The analytic implicit function theorem now implies that \( (a(x), b(x)) \) is an analytic function of \( x \in (-\infty, 1) \). Note that when \( V(x) = x^2 \), we have \( b(x) = \frac{1}{2}\sqrt{1 - x} \), and analyticity breaks down at \( x = 1 \).

Next, differentiating the version

\[
W_1(a(x), b(x)) = 0, \quad b(x)W_2(a(x), b(x)) = 1 - x
\]

of (15) we get

\[
a'W_{11} + b'W_{12} = 0, \quad (a'W_{21} + b'W_{22})b + b'W_2 = -1. \tag{16}
\]

As \( W_{11} > 0 \) by uniform convexity, it follows from the first equation of (16) that \( b'(x) = 0 \) implies \( a'(x) = 0 \), but this would contradict the second equation of (16). From the optimization problem (13) it is also clear that as \( x \to -\infty \), we get \( b(x) \to \infty \), so \( b'(x) < 0 \) for all \( x < 1 \).

Last, since \( W \) is an even function of \( b \), we see that \( b(1) = 0 \). Testing the minimizer against \( a = a(1), b = 1 - x \), we get

\[
W(a(1), 1 - x) - (1 - x) \log(1 - x) \geq W(a(x), b(x)) - (1 - x) \log b(x).
\]
Uniform convexity of $W$ at its minimizer gives the lower bound

$$W(a(1), 0) + c(a(1) - a(x))^2 + cb(x)^2 - (1 - x) \log b(x).$$

Comparing the upper and lower bounds then shows that

$$c(a(1) - a(x))^2 + cb(x)^2 \leq W(a(1), 1 - x) - W(a(1), 0) + (1 - x) \log b(x) - (1 - x) \log(1 - x).$$

Since $b(x)$ is decreasing and $W$ is continuous, the right hand side tends to 0 as $x \downarrow 1$. \qed

Continuing we note that $W$ is not an arbitrary two-variable polynomial. For example, it satisfies

**Lemma 6.**

$$4bW_{11} = bW_{22} + W_2$$

where the indices refer to partial derivatives in the $a$ or $b$ variables.

**Proof.** Using the formulation (14) this reduces to

$$[1]4bV''(a + by) = [1] (by^2V''(a + by) + yV'(a + by))$$

with $y = z + 1/z$. In order to show this, by shifting and scaling $V$, we may assume that $b = 1, a = 0$. Then by linearity, it enough to consider $V'(x) = x^k$. Then we get

$$[1]4ky^{k-1} = [1] (y^2ky^{k-1} + yy^k).$$

which since $[1]y^k = \binom{k}{k/2}$ reduces to the combinatorial identity

$$4k\binom{k-1}{k-1} = (k + 1)\binom{k+1}{k+1}.$$ \qed

We close this section with a formula for the inverse Hessian of $W(a, b) - (1 - x) \log(b)$ at $a(x), b(x)$. Note that by the definition of $W$, this the inverse of the Hessian the local Hamiltonian (10) evaluated at its minimizer, and restricted to the invariant subspace with basis $\xi_a, \xi_b$. Here the $\xi$ are the vectors that are 1 at all $a$ and $b$ variables, respectively, and zero otherwise.

**Proposition 7.** Denote by $\Sigma(x)$ the inverse of the Hessian of $W(a, b) - (1 - x) \log(b)$ at its minimizer $a(x), b(x)$. It holds that

$$\Sigma(x) = -b(x) \begin{bmatrix} 4b'(x) & a'(x) \\ a'(x) & b'(x) \end{bmatrix}.$$ 

(17)
Further, $\Sigma$ is strictly positive and bounded for $x \in [0, 1)$, and it is recorded for later use that

$$b'(x) < 0, \quad a'(x) + 2b'(x) < 0$$

for $x \in [0, 1)$.

**Proof.** Start with the equations

$$a'W_{11} + b'W_{12} = 0, \quad (a'W_{21} + b'W_{22})b + b'W_{2} = -1$$

derived in the course of proving Proposition 5. Now use Lemma 6 to cancel $W_2$ and $W_{22}$ from the second equation, and

$$W_{12} = -\frac{a'}{b'}W_{11}$$

to cancel $W_{12}$. This gives

$$W_{11} = \frac{-1}{bb'(4 - \frac{a'^2}{b'^2})}$$

and by Lemma 6 and the second equation of (15) we have

$$W_{22} = 4W_{11} - \frac{1 - x}{b^2}.$$ 

Finally we can compute

$$\Sigma = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} + \frac{1 - x}{b^2} \end{bmatrix}^{-1} = \frac{1}{W_{11}} \begin{bmatrix} 1 & -\frac{a'}{b'} \\ -\frac{a'}{b'} & 4 \end{bmatrix}^{-1}$$

$$= \frac{1}{W_{11}(4 - \frac{a'^2}{b'^2})} \begin{bmatrix} 4 & \frac{a'}{b'} \\ \frac{a'}{b'} & 1 \end{bmatrix} = -b \begin{bmatrix} 4b' & a' \\ a' & b' \end{bmatrix},$$

as claimed.

That $\Sigma$ is positive is another consequence of uniform convexity along with the continuity of $a(x), b(x)$. That $b'(x) < 0$ was already noted in (the proof of) Proposition 5, but both claims in (18) now follow from testing the quadratic form $\Sigma$ against simple vectors. \qed

**Remark 8.** The equations for $(a(x), b(x))$ can be put in another form. Using the integral formula for the Laurent coefficient, (15) is equivalent to

$$\frac{i}{2\pi} \int_{L_x}^{R_x} \frac{sV_x(s) \, ds}{\sqrt{(s - L_x)(R_x - s)}} = 1, \quad \int_{L_x}^{R_x} \frac{V_x(s) \, ds}{\sqrt{(s - L_x)(R_x - s)}} = 0,$$

where

$$V_x(s) = \frac{1}{1 - x} V(s), \quad L_x = a(x) - 2b(x), \quad R_x = a(x) + 2b(x).$$
This identifies \((L_x, R_x)\) as the left and right endpoints of support for the equilibrium measure \(\mu_V\) associated with the family of potentials \(V_x, -\infty < x < 1\). That is, with

\[
\mu_V = \arg\min \int_{-\infty}^{\infty} V_x(s)\mu(ds) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log \frac{1}{|s-t|} \mu(ds)\mu(dt),
\]

and the (realized) infimum taken over all probability measures \(\mu\), it is the case that \(\text{supp } \mu_V = [L_x, R_x]\). So, with obvious notation,

\[
\mathcal{E}(x) = R_x = a(x) + 2b(x).
\]

From this point of view (19) form the so-called moment conditions used in the determination of \(\mu_V\), see for example [16]. The results there provide another proof that \(a(x), b(x)\) are real analytic with \(b'(x) < 0, L'_x > 0, R'_x = \mathcal{E}'_x < 0\).

4 Outline of the main argument

The starting point of our argument is the paper [19], which provides a set of conditions for a sequence of tridiagonal random matrices to converge (in the norm-resolvent sense) to their natural continuum limit. We begin by repeating the set-up from that paper, along with a needed extension from [4].

4.1 Limits of random tridiagonal operators

Start with a sequence of discrete-time \(\mathbb{R}^2\)-valued random sequences \((y_{n,1,k}, y_{n,2,k})\) for \(1 \leq k \leq n\) with the convention that \(y_{n,i,0} = 0\). Let \(m_n = o(n)\) be a scaling parameter. For each \(n\), build the \(n \times n\) symmetric tridiagonal matrix \(H_n\) with

\[
(2m_n^2 + m_n(y_{n,1,k} - y_{n,1,k-1}), k \geq 1)
\]

on the diagonal and

\[
(-m_n^2 + m_n(y_{n,2,k} - y_{n,2,k-1}), k \geq 1)
\]

below and above the diagonal.

Defining \(y_{n,i}(x) = y_{n,i,\lfloor x m_n \rfloor} 1_{x m_n \in [0,n]}\) and \(\triangle_n\) the discrete Laplacian on the scale \(m_n\) (so \(m_n\) is one over the discretization length), \(H_n\) should be viewed as \(-\triangle_n\) plus integrated potential \(y_{n,1}(x) + 2y_{n,2}(x)\). What is desired is that \(H_n \rightarrow H = -\frac{d^2}{dx^2} + y'(x)\) (in a sense to be made precise) if it holds that \(y_{n,1}(x) + 2y_{n,2}(x)\) converges to a process \(y(x)\) (in a sense to be made precise).
Consider the following:

**Assumption 1 (Tightness/Convergence)** There exists a continuous process \( x \mapsto y(x) \) with \( y(0) = 0 \) such that

\[
\begin{align*}
(y_{n,i}(x); \ x \geq 0) & \quad i = 1, 2 \quad \text{are tight in law}, \\
(y_{n,1}(x) + 2y_{n,2}(x); \ x \geq 0) & \Rightarrow (y(x); \ x \geq 0) \quad \text{in law},
\end{align*}
\]

with respect to the Skorokhod topology of paths, see [13] for definitions.

**Assumption 2 (Growth/Oscillation bound)** There is a decomposition

\[
y_{n,i,k} = m_n^{-1} \sum_{\ell=1}^{k} \eta_{n,i,\ell} + w_{n,i,k},
\]

along with deterministic, unbounded nondecreasing continuous functions \( \eta(x) > 0, \zeta(x) \geq 1 \), and random constants \( \kappa_n(\omega) \geq 1 \) defined on the same probability space which satisfy the following. The \( \kappa_n \) are tight in distribution, and, almost surely,

\[
\begin{align*}
\eta(x)/\kappa_n - \kappa_n \leq \eta_{n,1}(x) + \eta_{n,2}(x) & \leq \kappa_n(1 + \eta(x)), \\
0 & \leq \eta_{n,2}(x) \leq m_n^2, \\
|w_{n,1}(\xi) - w_{n,1}(x)|^2 + |w_{n,2}(\xi) - w_{n,2}(x)|^2 & \leq \kappa_n(1 + \eta(x)/\zeta(x)).
\end{align*}
\]

for all \( n \) and \( x, \xi \in [0, n/m_n] \) with \( |x - \xi| \leq 1 \).

The growth/oscillation bounds in particular imply that the target limit operator \( H \) is almost surely densely defined on \( L^2 \) (of the positive half-line). In particular, it is made sensible through its quadratic form \( \langle f, g \rangle_H = \int f'g' + \int fg'y \) after a suitable integration by parts in the second term. Eigenvalues and eigenvectors of \( H \), which has discrete spectrum with probability one, are also defined through the quadratic form: \( \langle \Lambda, f \rangle \) is an eigenvalue/eigenvector pair if \( \langle f, \varphi \rangle_H = \Lambda \int f \varphi \) for all \( \varphi \in C_0^\infty \). As for convergence of \( H_n \) to \( H \), the needed result from [19], as extended in [4] reads:

**Theorem 9.** (i) [Theorem 5.1 of [19]] Given Assumption 1 and 2 above let \( (\lambda_{n,k}, v_{n,k}) \), \( k = 1, 2, \ldots, \) denote the ordered eigenvalues/eigenvectors of the matrices \( H_n \). Similarly let \( (\Lambda_k, f_k) \) denote the ordered eigenvalues/eigenvectors of the operator \( H \), taken with Dirichlet boundary conditions at the origin. Assume that \( H_n \) acts on \( \mathbb{R}^n \) as a subspace of \( L^2(\mathbb{R}) \) with coordinate vectors \( e_j = \sqrt{m_n} 1_{[j-1,j]}m_n^{-1} \). Then \( H_n, H \) can be coupled on a probability space so that a.s. we have \( \lambda_{n,k} \to \Lambda_k \) and \( v_{n,k} \to L^2 f_k \).
(ii) [Theorem 2.10 of [4]] The result is unchanged for certain perturbations of $H_n$ at its first entry. In particular let $e_{11}$ be the matrix with 11-entry equal to one and otherwise zero and consider the family of matrices $H_n + z_n e_{11}$ with $H_n$ as above and

$$\frac{z_n + m_n^2}{m_n} \to \infty$$

(26)

in probability. Then the eigenvalues/eigenvectors of $H_n$ still converge to those of $H$ (in the manner described) again with Dirichlet conditions at the origin.

In [19] and [4] this was was given as a distributional convergence statement, but by the standard Skorokhod embedding theorem it can be phrased this way (and the actual proof goes through Skorokhod embedding, giving the claimed result).

**Remark 10.** Let us clarify what is meant here by norm-resolvent convergence. Let some operators $H_n$ be defined on a domain $A_n$ of $L^2$, and let $H$ be defined on a subspace $A$. If $H_n$ and $H$ are closed and their spectrum is real, then for any non-real complex $z$ the resolvents $(z - H_n)^{-1}$ and $(z - H)^{-1}$ can be defined on all of $L^2$ and are bounded there by $1/\Im z$. This is very useful since the original operators may have disjoint domains and would be hard to compare.

Norm-resolvent convergence means that these bounded operators converge in norm for some (equivalently, all) such $z$.

It is easy to check that if $H_n, H$ are closed and $H$ has discrete spectrum bounded below with no multiple eigenvalues, then norm-resolvent convergence is equivalent to the following: for every $k$ the $k$th lowest eigenvalue of $H_n$ converges to that of $H$ and the corresponding eigenvector of $H_n$ converges in norm to that of $H$. See [26] for more on convergence of unbounded operators.

The extension in [4] goes beyond subcritical perturbations of the form (26). In particular, there it is proved that for critical perturbations reading $z_n/m_n + m_n \to \omega \in (-\infty, \infty)$ changes the boundary condition for $H$: from Dirichlet, $f(0) = 0$, to Robin $f'(0) = \omega f(0)$.

We should also note that have slightly changed the condition (24) in Assumption 2 from how things were stated in [19]. There it was assumed $\eta_{n,i} \geq 0$ while the proof only actually requires the non-negativity of $\eta_{n,2}$. Also, since it is convenient to define $y_{n,2}$ here to be twice that from [19] we have required $\eta_{n,2} \leq m_n^2$ on the right hand side of (24) (rather than the requirement $\eta_{n,2} \leq 2m_n^2$ of (5.5) of [19]). We will have more to say about this condition at the end of this section.
4.2 Application to the Dyson $\beta$ ensembles

To apply Theorem 9 one anticipates that $\mathcal{E} I - T_n$ has has a Laplacian term after scaling. That is, the main diagonal should be $-2$ times the off-diagonal, to leading order. Recalling the definition of the local minimizers, one would then expect $-2b\dagger(0) = -\mathcal{E} + a\dagger(0)$ — the top of $T$ presumably satisfying $\sim$ trigdiag$(b\dagger(0), a\dagger(0), b\dagger(0))$ to first order. This provides an intuitive understanding of the formula for the edge $\mathcal{E} = a\dagger(0) + 2b\dagger(0)$ derived in Section 3, see (20).

Based on the $\beta$-Hermite case ($V(x) = x^2/4$), we should rescale as in $\gamma n^{2/3}(\mathcal{E} I - T)$ for some $\gamma = \gamma(V, \beta)$. This sets $m_n = \sqrt{\gamma b\dagger(0)} n^{1/3}$, and the appropriate choice is to let

$$\gamma = \frac{1}{\tau^{2/3}(b\dagger(0))^{1/3}}, \text{ with } \tau = -((a\dagger)'(0) + 2(b\dagger)'(0)),$$

so that

$$m_n = (b\dagger(0)n/\tau)^{1/3}. \quad (27)$$

From (18) and (20) we have that $\tau$ is positive and equals the derivative of the edge at zero, $\tau = \mathcal{E}'(0)$. For the $\beta$-Hermite case $\gamma = \tau = b\dagger(0) = 1$.

One wrinkle is that, as alluded to the introduction, the first $O(\log n)$ stretch of $(A, B)$ variables do not have a universal behavior. As a $(\deg V/2 - 1)$-Markov process, $k \mapsto (A_k, B_k)$ can be expected to require $O(\log n)$ steps to achieve local equilibrium. Thus, one can only hope to apply the Theorem 9 as such directly to the submatrices $T_{[c\log n,n]}$ for some $c = c(V, \beta)$. Our strategy will be to show that these truncated matrices satisfy part (i) of the theorem, while the effect of the truncation can be controlled by a rank-one perturbation, to which part (ii) of the theorem applies.

We therefore set for a suitably large constant $c$ and $m_n$ as defined in (27):

$$y_{n,1}(x) = m_n \sum_{k = [c\log n]}^{\lfloor c_n \rfloor} (a\dagger(0) - A_k)b\dagger(0), \quad y_{n,2}(x) = m_n \sum_{k = [c\log n]}^{\lfloor c_n \rfloor} (b\dagger(0) - B_k)b\dagger(0). \quad (28)$$

This choice naturally prompts the further definitions:

$$w_{n,1}(x) = m_n \sum_{k = [c\log n]}^{\lfloor c_n \rfloor} (a\dagger_k - A_k)b\dagger(0), \quad w_{n,2}(x) = m_n \sum_{k = [c\log n]}^{\lfloor c_n \rfloor} (b\dagger_k - B_k)b\dagger(0), \quad (29)$$

and

$$\eta_{n,1}(x) = m_n^2 (a\dagger(0) - a\dagger_{[x_m \rfloor})/b\dagger(0), \quad \eta_{n,2}(x) = m_n^2 (b\dagger(0) - b\dagger_{[x_m \rfloor})/b\dagger(0). \quad (30)$$

In this way we recognize $w_{n,1,i}$’s as noise/oscillation terms and the $\eta_{n,i}$’s as (derivatives of) the drifts.
It is important to note that the tightness/convergence conditions on \( y_{n,i} \) from Assumption 1 only require looking at the variable \( x \) on bounded sets, that is, \( x = O(1) \). On the other hand, the conditions on the noise and drift components set out in Assumption 2 require \( x \)'s that track the entire index set of the matrix, or up to \( x \) of order \( n^{2/3} \). Since our control of the variables at the bottom of the matrix is not so sharp, it is more convenient put everything in the growth terms. More succinctly, what we actually do is to let: for an \( \epsilon = \epsilon(V, \beta) > 0 \) chosen below,

\[
\text{Retain (29), (30) for } x \leq n^{2/3}(1 - \epsilon), \text{ otherwise set:}
\]

\[
\begin{align*}
\eta_{n,1}(x) &= m_n^2(a(0) - A_{[x_{mn}]})/b(0), \\
\eta_{n,2}(x) &= m_n^2(b(0) - B_{[x_{mn}]})/b(0).
\end{align*}
\]

In terms of matrix indices this cutoff occurs at \( k = n(1 - \epsilon) \).

After several sections of preliminary estimates (to show that the minimizers of \( H \) are indeed well-approximated by the local minimizers, around which the field concentrates well), Section 9 contains in Proposition 29 the basic fluctuation result that allows Assumptions 1 and 2 above to be verified for the truncated matrices \( H_n = \gamma n^{2/3}(E I - T_{[\log n, n]}) \).

Section 10 establishes that the effect of this truncation on the spectrum can be bounded by a suitable perturbation: that the spectrum of \( \gamma n^{2/3}(E I - T_n) \) is bounded in terms of that for an ensemble of the form \( H_n + z_n e_{11} \) with \( z_n \) satisfying requirement (26) of part (ii) of Theorem 9. Our main result is then proved in Section 12.

**Remark 11.** The rather funny condition (24) is automatically satisfied by the \( \eta_{n,2} \) defined in (30): the results of Section 3 show that \( x \mapsto b(0) \) is nonnegative and decreasing.

## 5 Bounding minimizers

A first step toward showing that the global minimizers \((a^\circ, b^\circ)\) are well approximated by the local minimizers \((a^\dagger, b^\dagger)\) defined in Section 3 is to develop some preliminary upper and lower bounds for \((a^\circ, b^\circ)\). Actually, in the course of the proof it will be natural (and necessary) to consider minimizers of \( H \) restricted to some subsets \( J_a, J_b \) of coordinates, the values of \( a, b \) on the complement of \( J_a \) (respectively \( J_b \)) being fixed as boundary conditions. A collection of bounds are therefore established in the context of such “conditional minimizers”.

17
5.1 Upper bounds

We are interested in the minimizer of Hamiltonians similar to $H$ over entries $a_j, j \in J_a$ and $b_j, j \in J_b$. For example, for the overall minimizer we take $J_a = \{1, \ldots, n\}$, $J_b = \{1, \ldots, n - 1\}$. Now let $I$ be the set of integers at most $\deg V/2$ away from $J_a \cup J_b$.

Let $T_I$ be the finite minor of the doubly infinite tridiagonal matrix $T(a, b)$ corresponding to the indices $I$, and we set

$$V_I(a, b) := \text{tr}(V(T_I(a, b))).$$

Fix $y_2, y_3 \in (0, 1]$, and define

$$\mathcal{H}(a, b) = V_I(a, b) - \sum_{j \in J_b} \alpha_j \log b_j, \quad |\alpha_j| \leq y_2 \text{ for } j \in J_b$$

(32)

The variables $a_j, j \in I \setminus J_a, b_j, j \in I \setminus J_b$ are considered fixed, or boundary conditions, and we are interested in the minimizer of $\mathcal{H}$ over $a_j \in \mathbb{R}$ for $j \in J_a$ and

$$b_j \geq \begin{cases} y_3 & \text{for } \alpha_j < 0 \\ 0 & \text{otherwise}, \end{cases}$$

(33)

for $j \in J_b$. We employ the convention that $\log 0 = -\infty$.

The goal of this section is to show Theorem 12, a bound on the minimizers. It will be used in two ways. First, more roughly, if the $\alpha_i$ and the boundary conditions are bounded by a constant, then so are the minimizers. Second, it will be used for indices close to $n$ in our original problem, where the $|\alpha_i|$ are small. In this case, if the boundary conditions are close to $a(1), b(1) = 0$, then we show that so are the minimizers.

**Theorem 12.** There exists a constant $c$ depending on the polynomial $V$ only so that if

$$|a_k - a^\dagger(1)| \leq y_1 \text{ for all } k \in I \setminus J_a, \quad \text{and} \quad b_k \leq y_1 \text{ for all } k \in I \setminus J_b,$$

then for the minimizers $a, b$ of $\mathcal{H}$ satisfying (33) and all $k \in I$ we have

$$|a_k - a^\dagger(1)| + b_k \leq c \max(y_1, y_1^{(\deg V)/2}, y_2 \log(e/y_3), (y_2 \log(e/y_3))^{\deg V/2})^{1/2}.$$

The first step in the strategy is to show that a bound on $\mathcal{H}(a, b)$ gives a bound on the $\ell^2$-norm of $(a, b)$. We will then show that $\mathcal{H}(a, b)$ is small for the minimizers $(a, b)$ by comparing it to an explicit example.
Lemma 13. Let \( V \) be a polynomial satisfying \( V(x) \geq c_1 x^2 \) for \( c_1 > 0 \). Then for \( b_j \) satisfying (33), the Hamiltonian (32) has the bound
\[
\mathcal{H}(a,b) \geq \frac{c_1}{2} \sum_{i \in I} (a_i^2 + b_i^2) - |J_b| y_2 ((\log c_1)^- + \log(1/y_3)).
\]

Proof. By the lower bound on \( V \), the left hand side minus \( \frac{c_1}{2} \) tr\((T_f^2)\) is bounded below by
\[
\frac{c_1}{2} \text{tr}(T_f^2) - \sum_{i \in J_b} \alpha_i \log b_i \geq \sum_{i \in J_b} (-\alpha_i \log b_i + c_1 b_i^2),
\]
where the terms can be minimized individually. When \( \alpha_i < 0 \), and so \( b_i \geq y_3 \) a lower bound is \( y_2 \log y_3 \), and the same lower bound holds for \( \alpha_i = 0 \). When \( \alpha_i > 0 \) the minimal value is
\[
\frac{\alpha_i}{2} (1 - \log \alpha_i + \log(2c_1)) \geq -y_2 (\log c_1)^-.
\]

By comparing minimizers to some specific substitution, we get a bound that depends on the boundary conditions.

Corollary 14. Assume that \( V(a) \) is minimized at \( a = 0 \) and \( V(0) = 0 \). There exists a constant \( c \) depending on \( V \) and so that the conditional minimizing pair \( (a,b) \) of \( \mathcal{H} \) satisfies
\[
\sum_{k \in I} (a_k^2 + b_k^2) \leq c \gamma_4 |I| + c(1 + \gamma_4^{1-\deg V}) \left( \sum_{k \in I \setminus J_a} a_k^{\deg V} + \sum_{k \in I \setminus J_b} b_k^{\deg V} \right),
\]
as long as
\[
\gamma_4 \geq y_2 \log(e/y_3).
\]

Proof. Let \( q \) denote the total of the two sums above, and let \( d = \deg V \). Consider the candidate vector \( a', b' \) in which all \( a'_i \) for \( i \in J_a \) are set identically 0 and all \( b'_i \) for \( i \in J_b \) terms are identically equal to \( y_2 \), and the rest are given by the boundary conditions. For some \( c \) depending on \( V \) only, we have
\[
V_I(a', b') \leq c \sum_{i \in I} \left( |a'_i| + a'^d_i + b'_i + b'^d_i \right)
\]
since in the expansion, each \( a'_i, b'_i \) is contained in a bounded number of monomials and the coefficients are all bounded. Now using the given substitution and the fact that \( y_2 \leq 1 \), we get the upper bound
\[
V_I(a', b') \leq 2cy_2 |I| + c \sum_{i \in I \setminus J_a} \left( |a'_i| + a'^d_i \right) + c \sum_{i \in I \setminus J_b} \left( b'_i + b'^d_i \right).
\]
Note that for any \(x, \eta > 0\) we have \(x + x^d \leq \eta^{1-d} x + \eta\), whence for any \(\eta > 0\)
\[
V_I(a', b') \leq c|I|(2y_2 + \eta) + c(1 + \eta^{1-d})q.
\]
This, together with the bound \(|I|y_2 \log (1/y_3)\) on the logarithmic terms then yields
\[
\mathcal{H}(a', b') \leq c|I|(y_2 \log(e/y_3) + \eta) + c(1 + \eta^{1-d})q. \tag{36}
\]
We have assumed that \(V\) is minimized at 0, and \(V(0) = 0\), so \(V\) is bounded below by \(c_u x^2/2\), where \(c_u\) is the uniform convexity constant form Lemma 3. By Lemma 13 applied to \(T_f\) we have the lower bound for the minimizer \(a, b\)
\[
\mathcal{H}(a, b) \geq c_1/2 \sum_{k \in I} (a_k^2 + b_k^2) - y_2((\log c_2)^- + \log(1/y_3))|I|.
\]
Together with (36) and \(\mathcal{H}(a', b') \geq \mathcal{H}(a, b)\) gives
\[
\sum_{k \in I} (a_k^2 + b_k^2) \leq c|I|(y_2 \log(e/y_3) + \eta) + c(1 + \eta^{1-d})q.
\]
We set \(\eta = y_4\) and the claim follows. \(\square\)

For the proof of Theorem 12 this bound will be iterated. The following lemma isolates what we need for the iteration.

**Lemma 15.** Given \(\gamma_1, \gamma_2 > 0, \alpha \geq 2\), and a nondecreasing positive sequence \(s_n\), assume that for \(n = 1, \ldots, n^* - 1\) we have
\[
s_n \leq (s_{n+1} - s_n)^\alpha/\gamma_1 + \gamma_2 n. \tag{37}
\]
For any positive \(\gamma_3, x_0\) satisfying
\[
\gamma_3 \leq (32\gamma_1)^{1/\alpha}, \quad x_0 \geq \max((\gamma_2/\gamma_3)^{1-1/\alpha} - 1, 0)
\]
set \(f(x) = \gamma_3(x + x_0)^{\alpha/\gamma_3}\).

If \(s_1 \geq f(1)\) then \(s_n \geq f(n)\) for all \(n = 1, \ldots, n^*\).

**Proof.** We show this by induction on \(n\). Note that \(f(x)\) has increasing derivative, so we have \(f(x + 1) - f(x) \leq f'(x + 1)\). With \((\cdot)_+\) denoting the positive part, we write
\[
s_{n+1} = s_n + (s_{n+1} - s_n) \geq s_n + (\gamma_1 s_n - \gamma_1 \gamma_2 n)^{1/2}_+ \geq f(n) + (\gamma_1 f(n) - \gamma_1 \gamma_2 n)^{1/2}_+.
\]
Using the bound on the derivative, we get the lower bound \(f(n + 1) - f(n + 1)' + (\gamma_1 f(n) - \gamma_1 \gamma_2 n)^{1/2}_+\). So it suffices to show that \(f(n + 1)' \leq (\gamma_1 f(n) - \gamma_1 \gamma_2 n)^{1/2}_+\). This reduces to the inequality
\[
\left(\frac{\alpha \gamma_3}{\alpha - 1}\right)^\alpha (n + x_0 + 1)^{\alpha/\gamma_3} \leq \gamma_1 \gamma_3 (n + x_0)^{\alpha/\gamma_3} - \gamma_1 \gamma_2 n.
\]
We first choose \( \gamma_3 \) so that the dominant terms satisfy
\[
\left( \frac{\alpha \gamma_3}{\alpha - 1} \right)^\alpha (n + x_0 + 1)^{\frac{\alpha}{\alpha - 1}} \leq \frac{\gamma_1 \gamma_3}{2} (n + x_0)^{\frac{\alpha}{\alpha - 1}},
\]
which holds for all \( n \geq 1, x_0 \geq 0 \) as long as \( \gamma_3 \leq (32 \gamma_1)^{\frac{1}{\alpha - 1}} \). It suffices to check
\[
\gamma_1 \gamma_2 n \leq \frac{\gamma_1 \gamma_3}{2} (n + x_0)^{\frac{\alpha}{\alpha - 1}},
\]
which follows from our assumptions.

**Proof of Theorem 12.** By shifting the argument and the value of \( V \) by a constant we may assume that its minimizer \( a^*(1) \) is zero and \( V(0) = 0 \). By deleting unnecessary indices, we may assume that \( I \) is an interval. Let \( J = \{ \ell, \ldots, r \} \) be the smallest interval containing \( J_a \cup J_b \), and by symmetry we may assume without loss of generality that the index of interest, \( k \), is closer to the left: \( k - \ell \leq r - k \).

Let \( d = \deg V \), note that \( d \geq 4 \) and even. The iteration will use \( j^* = \lceil 2(r - k)/d \rceil \) steps. Define the nested intervals
\[
J^j = \{ k - jd/2, \ldots, k + jd/2 \} \cap J, \quad j = 1, \ldots, j^* - 1
\]
and let \( J^{j^*} = J \). Let \( s_j = \sum_{k \in J^j} a_k^2 + b_k^2 \). We are interested in \( s_1 \) and we will control the \( s_j \) recursively from \( j^* - 1 \) to 1.

Let \( J_a^j = J_a \cap J^j \), let \( J_b^j = J_b \cap J^j \), and let \( I^j \) be the set of indices at most \( d/2 \) away from \( J_a^j \cup J_b^j \). Corollary 14 applied to \( J_a^j, J_b^j \) and \( I^j \) gives
\[
\sum_{k \in I^j} (a_k^2 + b_k^2) \leq c_{\gamma_4} |I^j| + c(1 + \gamma_4^{1-d}) \left( \sum_{k \in I^j \setminus J_a^j} a_k^2 + \sum_{k \in I^j \setminus J_b^j} b_k^2 \right). \tag{38}
\]
The left hand side is bounded below by \( s_j \). On the right hand side, we have boundary terms
\[
\sum_{k \in I^j \setminus J_a^j} a_k^2 + \sum_{k \in I^j \setminus J_b^j} b_k^2 \leq cy_1^d.
\]
For \( j < j^* \) the rest of the summands have indices from \( I^j \setminus J^j \), which is a subset of \((J_{j+1} \setminus J_j) \cup \partial J \). These can be bounded by \( c ((s_{j+1} - s_j)^{d/2} + y_1^d) \). So for \( j < j^* \), we have
\[
s_j \leq c_{\gamma_4} j + c(1 + \gamma_4^{1-d}) ((s_{j+1} - s_j)^{d/2} + y_1^d). \tag{39}
\]
For \( j = j^* \) only the first kind of boundary terms appear, so we have
\[
s_{j^*} \leq c_{\gamma_4} j^* + c(1 + \gamma_4^{1-d}) y_1^d. \tag{40}
\]
We now proceed to analyze two cases. First assume $y_1 \leq 1$, $y_2 \log(e/y3) \leq 1$, and set $\gamma_4 = y_1 + y_2 \log(e/y3)$. Then (39), (40) simplify to

$$s_j \leq c\gamma_4 j + c\gamma_4^{1-d} (s_{j+1} - s_j)^{d/2}, \quad s_j^* \leq c\gamma_4 j^*.$$ 

We use Lemma 15 with

$$\gamma_2 = c\gamma_4, \quad \gamma_3 = 2\gamma_2, \quad \gamma_1 = c'd^{-1}, \quad x_0 = 0$$

to get that either $s_1 \leq \gamma_3 = 2c\gamma_4$, or we have

$$s_j^* \geq \gamma_3 (j^*)^{1+1/(d/2-1)} \geq c\gamma_4 (j^*)^{1+1/(d/2-1)}.$$ 

Together with $s_j^* < c\gamma_4 j^*$ the latter implies $j^* < c$ and so $s_1 \leq s_j \leq c\gamma_4$. The claim follows.

Now assume $y_1 > 1$ or $y_2 \log(e/y3) > 1$. Set $\gamma_4 = y_1^d + y_2 \log(e/y3)$. Then (39), (40) simplify to

$$s_j \leq c\gamma_4 j + c(s_{j+1} - s_j)^{d/2}, \quad s_j^* \leq c\gamma_4 j^*$$

we use Lemma 15 with

$$\gamma_2 = c\gamma_4, \quad \gamma_3 = c', \quad \gamma_1 = c'', \quad x_0 = c''\gamma_4^{1-2/d}$$

to get that either $s_1 \leq \gamma_3^2 = c\gamma_4^2$, or we have $s_j^* \geq c(j^*)^{1+1/(d/2-1)}$. Together with $s_j < c\gamma_4 j^*$ the latter implies $j^* < c\gamma_4^{d/2-1}$ and so $s_1 \leq s_j \leq c\gamma_4^{d/2}$. The claim follows.

\[ \square \]

### 5.2 Lower bounds

We continue with the setup of the previous subsection with the additional assumption $\alpha_i \geq 0$ and show that:

**Proposition 16.** There exists a constant $c_V$ depending on $V$ only so that if for the minimizers $a^\sharp, b^\sharp$ of $H$ of equation (32) together with the boundary conditions satisfy

$$\max_{j \in J, j \neq J} |a_j^\sharp| + |b_j^\sharp| \leq y$$

for some $y \geq 1$, then for all $k \in J$ we have

$$\log b_k^\sharp \geq -c_V \frac{\alpha_k}{y^{\deg V}}.$$
Proof. Let $a, b$ equal $a^\sharp, b^\sharp$ except let $b_k = 1$. By the minimizer property we have
\[
0 \leq H(a, b) - H(a^\sharp, b^\sharp) \leq -\alpha_k \log b_k + \sum_{k \in \pi} w_\pi(a, b) - w_\pi(a^\sharp, b^\sharp).
\]
The sum is only over path $\pi$ that pass through $k$, and $w_\pi$ is the monomial corresponding to $\pi$ in the path expansion of $\text{tr}(V(T_j))$ (all other paths have the same contribution). Counting such paths we get
\[
\sum_{k \in \pi} |w_\pi(a, b)| \leq c_V (\deg V)^{3 \deg V} y^{\deg V},
\]
and the same holds for $a^\sharp, b^\sharp$. The claim follows. \qed

6 Minimizers and boundary conditions

We continue the study of the conditional minimizers of $H$, demonstrating that they are relatively insensitive to the boundary conditions. The typical setup now is that we fix the values of $a, b$ on a set of indices (which usually will be an interval or two intervals – a one or two sided boundary), and minimize $H$ subject to these conditions.

Again we will consider such minimizers for slightly more general Hamiltonians of the form
\[
H = H(a, b) = \text{tr}(V(T)) - \sum_{k=1}^{n-1} \alpha_k \log(b_k)
\]
with $\alpha_k \in [0, 1]$. Compared with the modified Hamiltonian (32) of the previous section, it is convenient here to assume the nonnegativity of the coefficients $\alpha_k$ as the conditioning we will need to consider will always have the effect of “disconnecting” the final stretch of indices. Recall that in the actual Hamiltonian $H$ it is only the indices $k > n - 1/\beta$ (for $\beta < 1$) for which the analogous coefficients are negative.

From now on, for any set of indices $I$, let $\delta I$ be the set of indices outside $I$ that are at most $\deg V/2$ away from $I$. An important consequence of $V$ being polynomial the values of the conditional minimizer $(a_i, b_i)$ for $i \in I$ depend only on the conditioned values $(a_k, b_k)$ for $k \in \delta I$.

In each of the next two propositions we compare two conditional minimizers of $H$ for same interval $I$, but with respect to different boundary conditions outside $I$. Again, only the conditioned values of the variables in $\partial I$ matter for in problem. To get the desired bounds though, we require the additional assumption that one of the minimizers is in fact a minimizer of for all the variables $I \cup \partial I$ given the variables outside $I \cup \partial I$. 

23
Introducing the following notation for $\ell^2$-distance over a set of indices
\[ \| (a, b) - (a', b') \|^2_I = \sum_{i \in I} (a_i - a'_i)^2 + (b_i - b'_i)^2, \]
the warm-up bound reads:

**Proposition 17.** Consider any two minimizers $(a, b)$ and $(a', b')$ of $\mathcal{H}$ over the variables with indices in a set $I$. The variables outside $I$ serve as boundary conditions and are generally different. Assume further that $(a', b')$ is also a minimizer over variables with indices in $I \cup \partial I$. Then,
\[ \| (a, b) - (a', b') \|^2_I \leq c \| (a, b) - (a', b') \|^2_{\partial I} \]
as long as $|a_k|, |a'_k|, b_k, b'_k \leq b^*$ for all $k$, and $b_k, b'_k \geq b^*$ for $k \in I$. The constant $c$ depends on $V$, and $b^*$ only.

**Proof.** Without loss of generality, we can assume that $(a, b)_i = (a', b')_i$ for indices outside $I \cup \partial I$. Let $(a^\circ, b^\circ)$ equal $(a', b')$ on $I$ and $(a, b)$ elsewhere. We consider the problem of minimizing $\mathcal{H}$ over the indices in $I \cap \partial I$ with the variables with different indices fixed. This problem is solved by $(a', b')$, so by uniform convexity (Lemma 3), the candidate $(a, b)$ satisfies
\[ c_1 \| (a, b) - (a', b') \|^2 \leq \mathcal{H}(a, b) - \mathcal{H}(a', b'). \]

Now consider the problem of minimizing $\mathcal{H}$ over the indices in $I$, with variables with indices outside $I$ fixed. For the boundary conditions given by $(a, b)$ on $\partial I$, the variables $(a, b)$ solve this problem, and the other candidate $(a^\circ, b^\circ)$ satisfies $\mathcal{H}(a, b) \leq \mathcal{H}(a^\circ, b^\circ)$. Thus we get
\[ \mathcal{H}(a, b) - \mathcal{H}(a', b') \leq \mathcal{H}(a^\circ, b^\circ) - \mathcal{H}(a', b') \leq c_2 \| (a^\circ, b^\circ) - (a', b') \|^2 \]
\[ = c_2 \| (a, b) - (a', b') \|^2_{\partial I}. \]
The last inequality needs Taylor expansion around $(a', b')$ and a bound on the second derivative. Note that since $(a', b')$ solves the minimization problem for the variables with indices in $I \cup \partial I$, the linear terms vanish, and we have
\[ \mathcal{H}(a, b) - \mathcal{H}(a', b') \leq \frac{1}{2} \sup \| \text{Hess}_H \|_{2 \rightarrow 2} \| (a, b) - (a', b') \|^2 \]
where the supremum is over $a$’s and $b$’s on the line segment between $(a, b)$ and $(a', b')$. The second partial derivatives of the polynomial part of $\mathcal{H}$ are bounded above by constant
times a power of $b^*$, and that of the logarithmic part are bounded above by a constant times $b^*_2$.

Since $H_{\text{ess}}$ has at most $2 \deg(V)$ nonzero entries in each row, and each one is bounded, it follows that as a quadratic form it is bounded by a constant times the identity matrix by the Gershgorin Circle Theorem.

Iterating the above produces the estimate we will use going forward.

**Proposition 18.** Consider now any two minimizers $(a, b)$ and $(a', b')$ of $H$ over the variables with indices in an interval $I$. The variables outside $I$ serve as boundary conditions and are generally different. Assume further that $(a', b')$ is also a minimizer over variables with indices in $I \cup \partial I$.

Then, for all $k \in I$,

$$|a_k - a'_k| + |b_k - b'_k| \leq c \exp(-c' \text{dist}(k, \partial I)) \| (a, b) - (a', b') \|_{\partial I}$$

as long as $|a_k|, |a'_k|, b_k, b'_k \leq b^*$ and $\alpha_k \geq \alpha_\ast > 0$ for all $k \in \partial J$. The constants $c, c' > 0$ depend on $V$ and $\alpha_\ast, b^*$ only.

**Proof.** By Theorem 12 and Proposition 16 the upper bound $b^*$ on $\partial I$ coupled with the lower bound on the $\alpha$’s imply upper bounds on $|a_k|, b_k, |a'_k|, b'_k \leq b^*$ as well as lower bounds on $b_k, b'_k \geq b_\ast$ for $k \in I$. This fact will be required in order to employ Proposition 17.

Fix $k$, and divide the interval $I \cup \partial I$ into consecutive blocks $I_{-m}, \ldots, I_m$, so that

- $\partial I = I_{-m} \cup I_m$,
- for $1 \leq |j| \leq m - 1$ the intervals $I_j$ have length $\deg V / 2$,
- $k \in I_0$, and $\text{dist}(I_0^c, k) \leq \deg V / 2$.

This forces $m \geq \text{dist}(k, \partial I) \times 2 / \deg V$. For $j = 0, \ldots, m$ set

$$x_j = \| (a, b) - (a', b') \|_{I_j \cup \partial \l_j}^2.$$

The value of a conditional minimizer at an index $i$ only depends on any conditioned values for indices within $\deg V / 2$ of $i$. This implies that, for any subinterval $I'$ of $I$, the values of $(a, b)$ (or $(a', b')$ are the minimizers of $H$ conditioned on the respective values of $(a, b)$ (or $(a', b')$ on $\partial I'$. Therefore, applying Proposition 17 to the interval $I_{1-j} \cup \cdots \cup I_{j-1}$ we find that

$$\sum_{i=0}^{j-1} x_i \leq cx_j,$$
with the same constant $c$ for each $j$. Applying the Gronwall-type inequality of Lemma 19 below yields
\[ x_0 \leq cx_m e^{-c'(m+1)} \]
for some $c' > 0$. This implies the statement of the proposition.

**Lemma 19.** Let $x_i \geq 0$ satisfy $x_1 + \ldots + x_k \leq cx_k$ for some $c > 0$. Then, $x_1 \leq cx_k e^{-c'k}$ with another constant $c' > 0$.

**Proof.** Define block sums $y_j = x_{jq} + \ldots + x_{(j+1)q-1}$ where $q = 1/c$. Then $y_1 + \ldots + y_\ell \leq y_{\ell+1}$. Inductively it is easy to see that
\[ y_\ell \geq 2^\ell y_1, \]
and it follows that
\[ cx_k \geq x_1 + \ldots + x_k \]
\[ \geq y_1 + \ldots + y_{[k/q]} \]
\[ \geq y_1(1 + 2 + \ldots + 2^{[k/q] - 2}) \geq x_1 e^{c'k}, \]
as desired.

---

**7 Conditional and local minimizers are close**

We are finally in position to the global (and conditional) minimizers are well approximated by the the local minimizers. Recall the local Hamiltonian 10 introduced in Section 3,
\[ H(x) = H(x)(a, b) = \text{tr}(V(C)) - \sum_{k=1}^{n}(1 - x) \log(b_k), \]
where again $C$ is a symmetric circulant matrix with main diagonal given by the $a$’s, first off-diagonal given by the $b$’s and zeros elsewhere. Denote again by $(a^\dagger = a^\dagger(x), b^\dagger = b^\dagger(x))$ its (independent of $n$, unique) minimizer, Proposition 5. Recall as well the local minimizers corresponding to index $k$:
\[ a^\dagger_k = a^\dagger(k/n + 1/(n\beta)), \quad b^\dagger_k = b^\dagger(k/n + 1/(n\beta)). \]

**Proposition 20.** Let $H, H'$ be two Hamiltonians given by $\alpha_i, \alpha'_i$, respectively as in (41). Let $J$ be a subinterval of $1, \ldots, n$, let $(a, b) = (a', b')$ outside $J$, and let $(a, b) = (a', b')$ be the minimizers of the corresponding Hamiltonians inside $J$. Then with $q = \sum_{j \in J}(a_j - a'_j)^2 + (b_j - b'_j)^2$ we have
\[ q \leq \left| \sum_{k \in J}(\alpha_k - \alpha'_k) \log b_k \right| + \left| \sum_{k \in J}(\alpha_k - \alpha'_k) \log b'_k \right| \leq c \log(b^*/b_*) \max_{k \in J} |\alpha_k - \alpha'_k|. \quad (42) \]
where the second inequality assumes that

\[ b_k, b'_k \in [b_*, b^*] \quad \text{for } k \in J. \quad \text{(43)} \]

Moreover, for every \( y > 0 \) there exists \( b_*, b^* > 0 \) so that (43) holds as long as \( \alpha_k > 1/y \) for \( k \in J \) and \( |a'_k|, b'_k \leq y \) for \( k \) with \( 0 < \text{dist}(J, k) \leq \deg V/2 \).

**Proof.** By the definition of the minimizer, we have

\[ \mathcal{H}(a, b) \leq \mathcal{H}(a', b') \quad \mathcal{H}'(a', b') \leq \mathcal{H}'(a, b) \]

This implies that the two intervals, defined by their endpoints (whose order we do not specify)

\[ \{\mathcal{H}(a, b), \mathcal{H}'(a, b)\} \quad \text{and} \quad \{\mathcal{H}(a', b'), \mathcal{H}'(a', b')\} \]

overlap, and therefore the difference between any two of these four values is bounded above by the sum of the lengths of these intervals.

\[ \mathcal{H}(a', b') - \mathcal{H}(a, b) \leq |\mathcal{H}(a, b) - \mathcal{H}'(a, b)| + |\mathcal{H}(a', b') - \mathcal{H}'(a', b')| \]

by convexity, the left hand side is bounded below by a constant multiple of \( q \). For the right hand side, all polynomial terms vanish. The remaining terms are bounded by the right hand side of (42).

**Corollary 21.** Let \( J = \{\kappa - \ell, \ldots, \kappa + \ell\} \) be a subinterval of \( 1, \ldots, n \), and consider the minimizer \( (a, b) \) of \( H \) given the values outside \( J \). Consider also the modified hamiltonian

\[ H_n(a, b) = V(a, b) - \sum_{k=1}^{n-1} (1 - \kappa/n - 1/(n\beta)) \log b_k \quad \text{(44)} \]

which differs from \( H \) in that the coefficients of the \( \log \) terms do not change with \( k \). Fix \( (a', b') = (a, b) \) outside \( J \), and let \( (a', b') \) be the minimizer of \( H_n \) inside \( J \). Assume that \( b_k \geq b_* \) for \( k \in J \).

Let

\[ q = \sum_{j \in J} (a_j - a'_j)^2 + (b_j - b'_j)^2, \quad y = \max_{i=1, \ldots, \ell} |b'_{\kappa-i} - b'_{\kappa+i}| \]

Then we have

\[ q \leq c \frac{\ell^2}{n} \max \left( y, \frac{\ell^2}{n} \right) \]

for a constant \( c = c(b_*) \).
Proof. We use the first bound in (42). The first sum is bounded above by
\[ \left| \sum_{k \in J} \frac{k - \ell}{n} \log b_k \right| = \left| \sum_{i=1}^{\ell} \frac{i}{n} \log(b_{\kappa+i} - \log b_{\kappa-i}) \right| \leq \frac{1}{b_*n} \sum_{i=1}^{\ell} i |b_{\kappa+i} - b_{\kappa-i}|, \]
and likewise for \((a', b')\), having used the simple inequality \(|\log x - \log y| \leq \frac{|x - y|}{x \land y}|.
Combining the above gives
\[ q \leq \frac{2}{b\gamma n} \sum_{i=1}^{\ell} i |b'_{\kappa+i} - b'_{\kappa-i}| + \frac{1}{b\gamma n} \sum_{i=1}^{\ell} i (|b_{\kappa+i} - b'_{\kappa+i}| + |b_{\kappa-i} - b'_{\kappa-i}|) \]
\[ \leq \frac{c\ell^2}{n} y + \frac{c\ell^{3/2}}{n} \sqrt{q}, \]
after an application of Cauchy-Schwarz. In the case \(y \geq \sqrt{q}\) we get \(q \leq 2c\ell^2 n y\). When \(y < \sqrt{q}\) we get \(q \leq 2c\ell^2 n \sqrt{q}\), and so
\[ q \leq 4c^2 \ell^4 \leq 4c^2 \ell^2 \max \left( y, \frac{\ell^2}{n} \right). \]

Proposition 22. Let \(J\) be a subinterval of \(1 \ldots n - \lfloor \epsilon n \rfloor\), and let \(\kappa\) be its midpoint. For any set of indices \(I\), define
\[ \delta_I = \max_{i \in I} |a_j - a_{\kappa}^1| + |b_j - b_{\kappa}^1|. \]
Let \((a, b)\) denote the minimizers of \(H\) for the variables with indices in \(J\) with some boundary conditions on \(\partial J\). Then for any \(j \in J\) we have
\[ |a_j - a_{\kappa}^1| + |b_j - b_{\kappa}^1| \leq c \max \left( \delta_{\partial J} \exp(-c_1 \text{dist}(j, \partial J)), \frac{\ell^2}{n} \right), \]
where \(c, c_1\) depend on \(V, \beta\) and \(\epsilon\) only.

Proof. First note that Theorem 12 and Proposition 16 give constant upper and lower bounds on minimizers in terms of \(\epsilon, V, \beta\) and the boundary values. When we invoke Corollary 23, we will implicitly use these bounds. It is important that in the repeated use of Corollary 23 we can use the same bounds, and these don’t have to be iterated.

By induction, it suffices to prove that \(\delta_J \leq c \max(\delta_{\partial J}/2, \ell^2 / n)\) and that there exist a constant \(\ell'\), so that if \(J'\) is interval \(J\) reduced by \(\ell'\) on both sides, then \(\delta_{J'} \leq \max(\delta_{\partial J}/2, \ell c^2 / n)\).
Corollary 21 (in its notation) says that with
\[ q = \sum_{j \in J} (a_j - a_j')^2 + (b_j - b_j')^2, \quad y = \max_{i=1, \ldots, \ell} |b'_{\kappa-i} - b'_{\kappa+i}|, \]
\[ q \leq c \frac{\ell^2}{n} \max \left( y, \frac{\ell^2}{n} \right) \] \quad (45)

The values \( a', b' \) are minimizers of the local Hamiltonian (10) with some fixed boundary conditions. Note also that the constant function \( a_{\kappa}^{\dagger}, b_{\kappa}^{\dagger} \) is also a minimizer of the same Hamiltonian on any interval where boundary conditions are constant also given by \( a_{\kappa}^{\dagger}, b_{\kappa}^{\dagger} \) (this follows from the definition (12) that says that they are minimizers of the periodic problem). So we can apply Proposition 18 to get

\[ |a_j' - a_{\kappa}^{\dagger}| + |b_j' - b_{\kappa}^{\dagger}| \leq c \exp(-c_1 \text{dist}(j, J^c)) \left( \max_{i \in \partial J} |a_i' - a_{\kappa}^{\dagger}| + |b_i' - b_{\kappa}^{\dagger}| \right) \] \quad (46)

the last equality follows since \( a', b' \) agree with \( a, b \) on \( \partial J_k \). From here, since \( (a_{\kappa}^{\dagger}, b_{\kappa}^{\dagger}) \) is constant we have that

\[ y \leq 2c \delta_{\partial J} \] \quad (47)

Next, by the triangle inequality and (46), we also have that

\[ |a_j - a_{\kappa}^{\dagger}| + |b_j - b_{\kappa}^{\dagger}| \leq c \exp(-c_1 \text{dist}(j, J^c)) \delta_{\partial J} + 2\sqrt{q}. \]

Now using (47) in (45) produces

\[ 2\sqrt{q} \leq 2 \sqrt{c \frac{\ell^2}{n} \max(\delta_{\partial J}, \frac{\ell^2}{n})} = 1 \sqrt{64c \frac{\ell^2}{n} \max(\delta_{\partial J}, \frac{\ell^2}{n})} \leq 1 \max(\delta_{\partial J}, 64c \frac{\ell^2}{n}), \]

which when substituted in the previous display gives

\[ |a_j - a_{\kappa}^{\dagger}| + |b_j - b_{\kappa}^{\dagger}| \leq \max(\delta_{\partial J}, \frac{\ell' c \ell^2}{n}) \left( c \exp(-c_1 \text{dist}(j, J^c)) + 1/4 \right). \]

We choose \( \ell' \) so that \( c \exp(-c_1 \ell') < 1/4 \), and the two claims follow. \( \square \)

**Corollary 23.** Let again \( J \) be any subinterval of \( 1 \ldots n \). Let \( (a, b) \) denote the minimizers of \( H \) for the variables with indices in \( J \) with some boundary conditions on \( \partial J \). For any \( j \in J \) it holds that

\[ |a_j - a_{\kappa}^{\dagger}| + |b_j - b_{\kappa}^{\dagger}| \leq c \max(\delta_{\partial J}, \frac{\ell' c \ell^2}{n}) \left( \frac{\log n}{n} \right)^2. \]

where again \( \delta_{\partial J} = \max_{j \in \partial J} |a_j - a_{\kappa}^{\dagger}| + |b_j - b_{\kappa}^{\dagger}| \).

**Proof.** First note that if \( |J| \leq c \log n \) one simply restates Proposition 22 using that

\[ \| (a_{\kappa}^{\dagger}, b_{\kappa}^{\dagger}) - (a_j^{\dagger}, b_j^{\dagger}) \|_J \leq \ell' \frac{\log n}{n}, \]
which holds since by Proposition 5, the functions \( x \mapsto (a^\dagger(x), b^\dagger(x)) \) are analytic for \( x \in (-\infty, 1) \).

If now \(|J| > c \log n \) and \( \text{dist}(j, \partial J) \leq \log n \) we can again apply Proposition 22 by taking \( j \) in the role of \( \kappa \). That is, \( j \) can be made the midpoint of \( J' \subset J \) with \(|J'| = O(\log n) \) and the boundary conditions on \( \partial J' \) just equal to the values of \((a, b)\) there. The point being that, as noted before, \((a, b)\) is still the conditional minimizer on \( J' \) subject to these boundary conditions.

Finally, consider the case that \(|J| > c \log n \), but \( j \) is within \( \log n \) distance of \( \partial J \). Assume say that \( j \) is closer to the left edge of \( J \). By moving in order \( \log n \) steps from the left boundary we will find \( \deg(V)/2 \) stretch of \((a, b)\) which already satisfies \(|a_j - a^\dagger_j| + |b_j - b^\dagger_j| \leq c(\log n)^2 \). The Proposition 22 can then be applied yet again to the subinterval \( J' \) defined by the shared boundary of \( J \) to its left and the just identified good stretch of coordinates to its right. This will produce the type of statement desired for the midpoint \( \kappa \) of \( J' \) and with \( \delta_{\partial J'} \) in place of \( \delta_{\partial J} \). For the first issue use again that \((a^\dagger_j, b^\dagger_j)\) and \((a^\dagger_\kappa, b^\dagger_\kappa)\) are close throughout \( J' \). For the second, easily \( \delta_{\partial J'} \leq \delta_{\partial J} \). \( \Box \)

8 Bounding the field

As in the case of the minimizers, good concentration properties of the field of random variable \((A, B)\) begins by showing some simple upper and lower bounds hold, now with high probability. The convexity of the Hamiltonian \( H \) plays a key role again here: the law \( e^{-n\beta H} \) expected to place most of its mass in a neighborhood of minimizers, for which we have sharp bounds. We are able to get by with a relatively simple Gaussian domination.

8.1 A Gaussian lemma

The following may be viewed as instance of the Brascamp-Lieb inequality [5]. The short proof is included as we were unable to locate a statement in precisely the form required, allowing the “center” of the log-concave measure \( e^{-f} \) below to take place on the boundary of the domain in question.

Lemma 24. Let \( A \) be a convex subset of \( \mathbb{R}^n \). Let \( f : A \to \mathbb{R} \) be a convex with its minimum on \( A \) achieved at \( y \in A \) (this may be on the boundary). Assume that \( f \) has Hessian satisfying \( Hf(x) \geq uI \) for all \( x \in A \). Let \( X \) be a random variable with density proportional to \( e^{-f} \). Let
$G$ be the Gaussian vector with density proportional to $e^{-u\|x\|^2/2}$. Then $\|X - y\|$ is stochastically dominated by $\|G\|$.

Proof. Without loss of generality, assume $u = 1$ and $y = 0$. Fix any $0 \leq s < t$ and $\omega \in S^{n-1}$ so that $\omega t \in A$. Then $rac{d}{dt}f(t\omega) = \omega^t H f(t\omega)\omega \geq 1$, from which it follows that $\frac{d}{dt}f(t\omega) \geq t$, since the derivative is nonnegative at $t = 0$, because 0 is a minimum of $f$ (it may be positive if $y$ is on the boundary). Thus,

$$f(t\omega) = f(s\omega) + \int_s^t \frac{d}{du}f(u\omega)du \geq f(s\omega) + \frac{t^2}{2} - \frac{s^2}{2}.$$ 

Hence,

$$e^{-f(s\omega)}e^{-t^2/2} \geq e^{-f(t\omega)}e^{-s^2/2}.$$ 

This also holds for $s\omega$ or $t\omega$ outside $A$ with the convention that in that case $e^{-f(s\omega)} = 0$ or $e^{-f(t\omega)} = 0$, respectively. Fix $r > 0$. Multiply by $s^{n-1}t^{n-1}$ and integrate over $t > r$, $s < r$ and over $\omega \in S^{n-1}$ to get

$$\int_{\|x\| > r} e^{-f(x)}dx \int_0^t e^{-s^2/2}s^{n-1}ds \leq \int_{\|x\| < r} e^{-f(x)}dx \int_r^\infty e^{-t^2/2}t^{n-1}dt.$$ 

Multiplication by the appropriate normalization constants yields

$$P[\|X\| > r] P[\|G\| < r] \leq P[\|X\| < r] P[\|G\| > r].$$

We can rewrite this as

$$P[\|X\| > r] (1 - P[\|G\| > r]) \leq (1 - P[\|X\| > r]) P[\|G\| > r],$$

from which we get the desired domination $P[\|G\| > r] \geq P[\|X\| > r].$ \hfill \Box

### 8.2 Upper and lower bounds on $(A, B)$

The Gaussian Lemma (Lemma 24) coupled with the bounds on the minimizers from Section 5 imply that, with high probability, the random variables $(|A|, B)$ themselves are bounded above and the $B$’s are bounded below (with a small caveats depending on $\beta$).

Here we return to the original tridiagonal matrix entries for the Dyson $\beta$-ensemble with Hamiltonian

$$H(a, b) = V(a, b) - \sum_{k=1}^{n-1} (1 - k/n - 1/(n\beta)) \log b_k.$$ 

The upper bound reads:
Proposition 25 (Upper bound on the field). Let $A, B$ denote the random tridiagonal matrix entries for the Dyson $\beta$-ensemble. There exists constants $c_1, c_2, c_3 > 0$ depending on $V$ and $\beta$ so that for all $\beta \geq 1$ and all $n$ we have

$$P \left( \sum_{k=1}^{n} (A_k - a_k^o)^2 + \sum_{k=1}^{n-1} (B_k - b_k^o)^2 > c_1 \right) \leq e^{-c_2 n} \tag{48}$$

here $a^o, b^o$ are the minimizers of the corresponding Hamiltonian $H$.

For all $\beta > 0$ we also have

$$P( |A_1|, \ldots, |A_n| \leq c_3, \text{ and } B_1, \ldots, B_{n-1} \leq c_3) \geq 1 - e^{-c_2 n}. \tag{49}$$

Moreover, for all $\beta > 0$ and $\epsilon > 0$ there is $\delta > 0$ so that for all $n$ and $s \geq 1 - \delta$ we have

$$P \left( |A_k - a^\dagger(1)|, B_k \leq \epsilon \text{ for all } k \geq sn \mid |A_k - a^\dagger(1)|, B_k \leq \delta \text{ for all } k \in [sn, sn + \deg V] \right) \geq 1 - e^{-c_2 n}. \tag{50}$$

Proof. We first treat the $\beta \geq 1$ case.

The Hamiltonian corresponding to the distribution is uniformly convex with Hessian bounded below by $c_u I$ by Lemma 3. Thus the Gaussian Lemma (Lemma 24) implies that $\| (A - a^o, B - b^o) \|^2$ is stochastically dominated by the norm an i.i.d. Gaussian vector $G$ in dimension $n$ and entry variance $c_2^2 = 1/(2\beta c_u n)$.

Noting that $E\|G\|^2 = c_2^2$ and that $G/c_2$ is the average of $n$ independent $\chi_1^2$ random variables $P(G/c_2 > c_1)$ for any $c_1 > 1$ and $I(\cdot)$ the rate function for the $\chi^2$-distribution. For the second claim (49) recall that Theorem 12 proves that (conditional) minimizers satisfy a uniform bound. For (50), use the fact that this bound tends to zero as the coefficients of the logarithmic terms do, and apply previous argument to the conditional distribution.

In the $\beta < 1$ case the the log terms with positive coefficients destroy the convexity of the Hamiltonian $H$. Fix a value $y$ so that for $b \geq y$ the second derivative $-1/\beta \log b$ is at most half the uniform convexity constant $c_u$ from Lemma 3.

We therefore condition on the set of indices $k > n - 1/\beta$ so that $B_k < y$, as well as the values of these $B_k$. The entries minimizer of the conditional Hamiltonian are then bounded by a constant (depending on $V$ only) by Theorem 12 (this requires the smallest coefficient $-1/\beta n \geq -1$, so it holds for large enough $n$; for small values of $n$ any bound works).

The conditional Hamiltonian is uniformly convex: the log terms with $b_k < y$ are removed; for $b_k \geq y$ the second derivatives of $\log b_k$ terms are dominated by the uniform
convexity constant $c_n$. We can then bound the conditional distribution of the rest of the $A$, $B$ the the same way as in the $\beta > 1$ case. Averaging over the conditioning gives (49).

To get (50), we can let $y \to 0$ with $\epsilon$, and apply this previous argument with the quantitative version of Theorem 12.

For a lower bound we have:

**Proposition 26** (Lower bound on the field). For any $k \leq n - 1/\beta$, there are constants $c_1, c_2, c_3$ such that

$$
P(B_k > e^{-c_1 \frac{n}{\beta}}, |A_1|, \ldots, |A_n| \leq c_3 \text{ and } B_1, \ldots, B_{n-1} \leq c_3) \geq 1 - e^{-c_2 n}.
$$

Here $c_1$ must be chosen to depend on $c_3$. Otherwise the $c_1, c_2, c_3$ depend only on $V$ and $\beta$.

**Proof.** Consider $(a, b)$ and $(a', b')$ which agree everywhere except at a given index $k \leq n - 1/\beta$ where it holds $b_k < b'_k$. By the same type of considerations used in Proposition 16 we have that, assuming all $|a_j|, b_j \leq b^*$,

$$
\exp\{-n\beta H(a, b) + n\beta H(a', b')\} \leq \exp\{nc + (\beta(n - k) - 1) \log(b_k / b'_k)\},
$$

with $c = c(b^*, \beta, V)$. We tacitly assume $b^* \geq 1$. The inequality is written in this way to emphasize that the left hand side is the ratio of the densities at $(a, b)$ and $(a', b')$. If then $\log(b_k / b'_k) < \rho := -\frac{2}{1 - k/n - 1/\beta} < 0$ this ratio is bounded by $e^{-nc}$, still granted the overall bound on the $|a|$’s and the $b$’s. Choosing for example $b_k < e^\rho$ and $b'_k = b_k + 1$ implies that

$$
P(b_k < e^\rho \mid a_j \text{ and } b_j \text{ for } j \neq k) 1_{|a_j|, |b_j| \leq b^*} \leq e^{-nc}.
$$

Taking expectations and slightly adjusting $\rho$ yields the claim. \qed

## 9 Gaussian approximation of the field

We now show that, for suitable ranges of the indices, the random variables $(A, B)$ are well approximated by a Gaussian process, or really by a mixture of Gaussian processes. The precise statement is given in Proposition 29 below. This is a first pass at the functional central limit theorem which identifies the Brownian potential in the Stochastic Airy Operator.

First we record two more concentration estimates for the field about various minimizers of the Hamiltonian $H$. The second is a refinement of the first. Both are consequences of the Gaussian Lemma (Lemma 24).
9.1 Concentration about the minimizers

Fix $0 < b_* < b^*$. We say that $G_i(a,b)$ holds if we have $|a_i|, b_i \leq b^*$ and $b_i > b_*$. For a set of indices $J \subset \{1, \ldots, n\}$, we say that $G_J(a,b)$ holds if $G_i(a,b)$ holds for all $i \in J$; we will say $G(a,b)$ holds if $G_i(a,b)$ holds for all indices.

For an interval $I$, let $\partial I$ denote union of the stretches of length $$(\deg V) / 2$$ immediately before and after $I$.

Lemma 27 (Concentration in short intervals). Fix $\delta > 0$. Let $I$ be a subinterval of $\{1, \ldots, n\}$ of length at most $c_1 \log n$.

Consider random variables $(A_i, B_i)_{i \in I}$ picked from the distribution (6), where the values $a, b$ on $\partial I$ are fixed. Let $(a^i, b^i)$ denote the minimizer of the density given the boundary conditions $\partial I$. Then there is a $c = c(\delta, b_*, b^*)$ so that for all $n$ we have

$$P \left( \sum_{i \in I} (A_i - a^i)^2 + (B_i - b^i)^2 > n^{2\delta - 1} \right) \leq ce^{-n^\delta}. \tag{51}$$

Proof. The log density of the conditional distribution is uniformly convex by Lemma 3. Thus the Gaussian Lemma (Lemma 24) implies that $|B - b_k|$ is stochastically dominated by the norm an i.i.d. Gaussian vector in dimension $2c_1 \log n$ and entry variance $c_2/n$, where $c_2$ depends on the uniform convexity constant only.

The square of a 2-dimensional Gaussian has exponential distribution. This gives the (standard) large deviation bound that for the sum of squares of independent standard Gaussians $X_i$

$$P(X_1^2 + \ldots + X_m^2 > a) \leq mP(X_1^2 + X_2^2 > a/(m/2)) = m \exp(-a/m)$$

Finally, Gaussian scaling gives for (51) the upper bound

$$(2 \log n) \exp(-n^{2\delta} / (2c_2 c_1 \log n)) \leq ce^{-n^\delta}. \quad \square$$

Lemma 28 (Concentration everywhere). Fix $\delta > 0$. Let $I$ be any subinterval of $\{1 \ldots (1-\epsilon)n\}$.

Consider random variables $(A_i, B_i)_{i \in I}$ picked from the distribution (6), where the values $a, b$ on $\partial I$ are fixed, and assume that

$$|a_i - a_i^\dagger| + |b_i - b_i^\dagger| \leq n^{\delta - 1/2} \quad \text{for all } i \in \partial I.$$

where $(a_i^\dagger, b_i^\dagger)$ are the local minimizers defined in (12). Then there is a $c = c(\epsilon, \delta, b_*, b^*)$ so that for all $n$ and $I$ we have

$$P \left( \exists k \in I : |A_k - a_k^\dagger| + |B_k - b_k^\dagger| > n^{\delta - 1/2} \right) \leq ce^{-n^\delta}. \tag{52}$$
Proof. **Case 1.** First we show that with high probability
\[ |A_k - a^*_k| + |B_k - b^*_k| \leq n^{\delta - 1/2} \]
holds for indices \( k \) satisfying \( \text{dist}(k, \delta I) \geq c \log n \). Let \( J \) be a subinterval of \( I \) of length \( 1 + 2[c_1 \log n] \) centered at \( k \). Fix the values of \( A, B \) to be \( (a, b) \), \( \partial J \) (if \( \partial J \) intersects \( \partial I \), then these values are already fixed by the boundary condition). Let \( (A^{(k)}_i, B^{(k)}_i)_{i \in J_k} \) denote the minimizer of the conditional density of the variables given these (random) boundary conditions. Since \( k \) is far enough from the boundary of \( J \), this minimizer is close to the local minimizers. Indeed, Corollary 23 applied to the interval \( J \) gives that on \( G(A, B)_{\partial J} \) we have
\[ |A^{(k)}_k - a^*_k| + |B^{(k)}_k - b^*_k| \leq n^{\delta - 1/2} \]
now Lemma 27 applied to the conditional distribution of \( (A^{(k)}_i, B^{(k)}_i)_{i \in J_k} \) gives that with probability at least \( 1 - ce^{-n^d} \) we have
\[ |A^{(k)}_k - A_k| + |B^{(k)}_k - B_k| \leq n^{\delta - 1/2} \]
together with the union bound, after increasing \( \delta \) a bit, this gives the first case.

**Case 2.** Now consider indices \( k \) satisfying \( \text{dist}(k, \delta I) \geq c \log n \). When \( |I| < 2c_1 \log n \), all indices are like that, so we consider this case first. Here, Corollary 23 shows that the maximizers \( a^\sharp, b^\sharp \) of the conditional density of \( A, B \) are \( cn^{\delta - 1/2} \)-close to the local minimizers, and the Gaussian Lemma 24 shows that the values are concentrated around \( a^\sharp, b^\sharp \), and the claim follows after increasing \( \delta \) a bit.

**Case 3.** Finally, assume that \( |I| > 2c_1 \log n \). Assume that \( k \) is closer than \( c_1 \log n \) to the left boundary \( i_1 \) of \( I \). Consider the \( \deg(V)/2 \) indices starting at \( i_1 + c_1 \log n \). From the first case, it follows that \( A, B \) at these indices are with high probability \( cn^{\delta - 1/2} \)-close to the local minimizers. Now apply the second case to the conditional distribution of \( A, B \) on the interval \( i_1, \ldots, i_1 + c_1 \log n \) to get the desired claim. 

**9.2 (A,B) as a mixture of Gaussians**

Now let \( (A, B) \) are picked from the Dyson \( \beta \) ensemble distribution (6), and let \( I \subset 1, \ldots, (1-\epsilon)n \) be an interval. The distribution of \( A_i, B_i, i \in I \) depends only on the values (denote by \( q \)) of \( A, B \) with indices in \( \partial I \). Let \( \mu_q \) denote this conditional distribution. Let \( a^{(q)}_i, b^{(q)}_i \) denote maximizers of the density of \( \mu_q \) on \( I \), and let them correspond to \( q \) on \( \partial I \).

Let \( \nu_q \) denote the Gaussian measure on \( a_i, b_i, i \in I \) with mean \( (a^{(q)}_i, b^{(q)}_i) \) and inverse covariance matrix \( n\beta \text{Hess}_H(a^{(q)}_i, b^{(q)}_i) \).
Let \( \mu_I \) denote the averaging of \( \mu_Q \) with respect to the random \( Q \) picked from its marginal distribution. Note that \( \mu_I \) is just the marginal distribution of \((A_i, B_i; i \in I)\). Let \( Q \) denote the set of boundary conditions \((a_i, b_i), i \in \partial I \) so that

\[
|a_i - a_i^I| + |b_i - b_i^I| \leq cn^{\delta - 1/2} \quad \text{for all } i \in \partial I.
\]

Pick \( q_0 \in Q \), and let \( Q' = Q \) when \( Q \in Q \) and \( Q' = q_0 \) otherwise. Let \( \nu_I = \text{Ev}_{Q'} \).

**Proposition 29.** Let \( I \subset \{c \log n, \ldots, (1 - \epsilon)n\} \) be an interval of length at most \( n^{1/2 - \epsilon} \). Then as \( n \to \infty \) we have

\[
\text{dist}_{TV}(\mu_I, \nu_I) = o(1).
\]

More precisely, we can choose the constant \( c \) so that the event

\[
S = \{|A_i| \leq c, \frac{1}{c} \leq B_i \leq c \text{ for } i \in 1, \ldots n(1 - \epsilon) \cap \max_{i \in I} |A_i| - a_i^{(Q)}| + |B_i - b_i^{(Q)}| < cn^{\delta - \frac{1}{2}}\}
\]
satisfies \( \mu(S) \geq 1 - ce^{-n^{\delta'/c}} \) as well as \( \frac{d\nu_I}{d\mu_I} |_S = 1 + o(1) \) so long as \( \delta < \epsilon/3 \). Moreover, on the event \( S \),

\[
|a_j^{(q)} - a_j^I| + |b_j^{(q)} - b_j^I| \leq c \max \left(n^{\delta - 1/2} \exp(-c_1 \text{dist}(j, \partial I)), \frac{(\log n)^2}{n}\right),
\]

for all \( j \in I \).

**Proof.** Let \( H_q \) be the Hamiltonian corresponding to \( \mu_q \), and that

\[
\mu_q = \frac{1}{Z_{\mu_q}} \exp(-n\beta H_q) \, da \, db.
\]

Let \( \text{Hess}_H \) be the Hessian of \( H_q \) evaluated at its minimizer \( a^{(q)}, b^{(q)} \). We compare the probability measures \( \mu_q \) and \( \nu_q \), which can be written as

\[
\nu_q = \frac{1}{Z_{\nu_q}} \exp \left(-\frac{n\beta}{2} \left< (a - a^{(q)}, b - b^{(q)}), \text{Hess}_H(a - a^{(q)}, b - b^{(q)}) \right> \right) \, da \, db.
\]

Note that by adjusting the \( Z_{\nu_q} \) we can assume that \( H_q(a^{(q)}, b^{(q)}) = 0 \). Then by Taylor expansion

\[
|H - \left< (a - a^{(q)}, b - b^{(q)}), \frac{1}{2} \text{Hess}_H(a - a^{(q)}, b - b^{(q)}) \right> | \leq \xi \sum_{i=1}^k |a_i - a_i^{(q)}|^3 + |b_i - b_i^{(q)}|^3
\]

where

\[
\xi(a, b) = c \max_{i \in I, t \in [0, 1]} \left( \left| \frac{\partial^3 H_q}{\partial a_i^3} \right|_{a_i(t)} + \left| \frac{\partial^3 H_q}{\partial b_i^3} \right|_{b_i(t)} \right)
\]

36
and \( a_i(t) = ta_i + (1 - t)a_i^{(q)} \), and \( b_i(t) = tb_i + (1 - t)b_i^{(q)} \). Note that the bounded-range interaction implies that there are only a linear number of non-zero mixed third partial derivatives, and they can be bounded this way.

Next define

\[
S_\gamma = \left\{ (a,b) : n \xi(a,b) \left( \sum_{i=1}^k |a_i - a_i^{(q)}|^3 + |b_i - b_i^{(q)}|^3 \right) \leq \gamma \right\}. 
\]

By (54) and the preceding display, on the event \( S_\gamma \), the Radon-Nikodym derivatives between the measures \( \mu_q \) and \( \nu_q \) lie in the interval \( \left[ e^{-\gamma} - \kappa, e^\gamma \right] \). Further setting,

\[
e^{-\kappa} := \min(\mu_q(S_\gamma), \nu_q(S_\gamma)),
\]

we have that,

\[
e^{-\kappa} \leq \nu_q(S_\gamma) \leq \frac{Z_{\mu_q}}{Z_{\nu_q}} \int_{S_\gamma} e^\gamma d\mu_q,
\]

from which it follows that

\[
\frac{Z_{\mu_q}}{Z_{\nu_q}} \geq e^{-\gamma - \kappa}, \quad \text{and by symmetry} \quad \frac{Z_{\nu_q}}{Z_{\mu_q}} \geq e^{-\gamma - \kappa}.
\]

Thus \( \frac{d\mu_q}{d\nu_q} \) is in fact in the interval \( [e^{-2\gamma - \kappa}, e^{2\gamma + \kappa}] \). This implies that for any event \( D \)

\[
\mu_q(D) - \nu_q(D) \leq \mu_q(D) - e^{-2\gamma - \kappa} \mu_q(D) + \nu_q(S_\gamma^c) \leq 2(\gamma + \mu_q(S_\gamma^c) + \nu_q(S_\gamma^c)),
\]

using simply that \( 1 - e^{-2\gamma} \leq 2\gamma \) and the definition of \( e^{-\kappa} \).

By symmetry and the last formula we have that

\[
\text{dist}_{TV}(\mu_{I}, \nu_{I}) \leq 2(\gamma + \mu_q(S_\gamma^c) + \nu_q(S_\gamma^c)).
\]

Now recall the measure \( \nu_I \): it is the mixture of \( \nu_q \) given by \( E\nu_Q' \). The final form of the total variation bound the reads,

\[
\text{dist}_{TV}(\mu_I, \nu_I) \leq E \text{dist}_{TV}(\mu_Q, \nu_Q') \leq 2(\gamma + \mu_q(S_\gamma^c) + \nu_q(S_\gamma^c)).
\]

(55)

where the expectation is with respect to the random values assigned to \( Q \).

We wish to show the last three quantities in (55) tend to zero for a \( \gamma = o(1) \). To begin, introduce the event

\[
E = \{|a_i|, b_i \leq b^* \text{ and } |b_i| > b^* \text{ for all } i \in 1, \ldots, (1 - \epsilon)n\}.
\]
By Proposition 25 we have that the indicated upper bounds on the field hold with high probability. Lemma 26 gives the desired conclusion for the lower bound as we have restricted to indices less than \((1 - \epsilon) n\).

Working on \(E\) we can show that the left portion \(L\) of \(\partial I\) satisfies \(Q\) with high probability (the argument for the right part of \(\partial I\) is the same). Consider an interval of \(I_L\) length \(c \log n\) containing \(L\). Again on \(E\), the events

\[
\max_{k \in L} |A_k - A^1_k| + |B_k - B^2_k| \leq cn^{\delta - \frac{1}{2}},
\]

and

\[
\max_{k \in L} |A_k^1 - a^1_k| + |B_k^2 - b^2_k| \leq c (\log n)^2 n^{-1/2},
\]

both hold with high probability. Here \(\{A^1_k, B^2_k\}\) are the conditional minimizers of \(H\) subject to the (random) boundary conditions on \(\partial I_L\) (drawn from the basic law \(\mu\)). The first of these holds by Lemma 27. The second follows from Corollary 23. The point is that, save for an event of exponentially small probability, \((A_k, B_k)\) are within \(O(n^{-1/2 + \delta})\) of the local minimizers throughout \(L\). That is, \(|P(Q / \in Q)| \to 0\).

To control \(E_{\mu_Q}(S^c_{\gamma})\) it is enough to consider \(\mu_q(S^c_{\gamma} E)\) for \(q \in Q\). We have only just shown that on \(E\) the boundaries conditions may be assumed to satisfy \(Q\). Now, also on \(E\), we have that \(\xi(A, B) < c',\) recall (54). The upper bounds on the field and conditional minimizers (Proposition 25 and Theorem 12) bound the third derivatives of the polynomial part; the lower bounds from Propositions 26 and 16 will bound the derivatives of the logarithms. Finally, Lemma 28 implies that \(\mu_q(\max_{k \in I} |A_k - a^{(q)}_k| + |B_k - b^{(q)}_k| > n^{\delta - 1/2}) \leq ce^{-n^{\delta}/c}\) where \(c\) is independent of \(q\). But this estimate entails a similar bound on \(\mu_q(S^c_{\gamma} E)\) by adjusting \(\delta\) so that \(\delta < \epsilon/3\).

The Gaussian calculation \(\sup_{q \in Q} \nu_q(S^c_{\gamma}) \to 0\) is more immediate as under this law the centered variables each have variance bounded by \(c/n\) (with again \(c\) independent of \(q\)).

The more refined statement \(\frac{d\mu}{dq}|_S = 1 + o(1)\) (with \(\mu(S) = 1 - o(1)\)) follows readily after noticing that the advertised \(S\) is contained in \(E \cap S_{\gamma}(Q)\) for a \(\gamma\) chosen to be \(o(1)\). The estimate (53) simply puts together the outcome of Corollary 23 with the assumption that one is working on the event \(S\).

\[\square\]

10 A non-universal operator limit in the global scaling

With the results of the last section everything is in place to prove our main result, at least as applied to the truncated operator \(T_{[c \log n, n]}\). Adapting this truncation is not a matter of
convenience: the first order $\log n$ entries of the matrix $T_n$ behave very differently from the quadratic case. Even their first-order behavior is different from the entries that come after $\log n$ steps. We will show that at one hand the top of $T_n$, on a global scale, encodes the equilibrium measure $\mu_V$ and is thus non-universal. On the other hand, Theorem 36 below shows that the effect of this global non-universal behavior for the first stretch of matrix entries can be absorbed into a suitably small perturbation that does not harm the universality of the edge scaling limit.

Recall that for a bounded self-adjoint operator $J$ on a Hilbert space, the spectral measure of $J$ at a unit vector $v$ is the measure whose $k$th moment is $\langle v, J^k v \rangle$. We will omit mentioning the vector $v$, and by default take it to be the first coordinate vector. The spectral measure of a matrix $M$ is then the weighted sum of delta masses at the eigenvalues with weights given by the squared first entries of the corresponding normalized eigenvectors.

Let $J$ denote the Jacobi operator, or semi-infinite symmetric tridiagonal matrix, whose entries $a_i, b_i > 0$ solve the optimization problem (41) with $\alpha_i = 1$, $n = \infty$, and left boundary conditions $0$.

More precisely, its entries are the unique bounded sequence so that for any finite subinterval of indices $I$, the variables with indices of $I$ solve the (now finite) optimization given the rest as boundary conditions.

**Lemma 30.** The operator $J$ defined above exists, is unique, and satisfies that for some $c > 0$, as $k \to \infty$ we have

$$ (J_{k,k}, J_{k,k+1}) = (a^\dagger(0), b^\dagger(0)) + O(e^{-ck}). \quad (56) $$

**Proof.** The solutions for the optimization problem with zero boundary conditions on the interval $[1, n]$ have to converge to a limit $J$ by Proposition 18, which says that the values of two solutions with different boundary conditions are close away from the boundary. The same Proposition implies that the limit $J$ will solve the optimization problem restricted to subintervals, as in the definition.

Uniqueness also follows from Proposition 18: $J$ and any different solution has to be arbitrarily close in any subinterval.

Finally, Proposition 18 also implies (56). Indeed, the vector with entries $a^\dagger(0), b^\dagger(0)$ minimizes the same Hamiltonian as the on- and off-diagonals of $J$ with different boundary conditions.

The first step toward Theorem 36 shows that the leading minors of $T_n$ are well approximated by those of $J$. For clarity, throughout the rest of this section we use the notation
Proposition 31. For the matrix $T_n^\circ$ of global minimizers for $\beta \in [1, \infty]$ and for every $c \log n < m < c'n$ we have
\[
\|T_n^\circ[1,m] - J[1,m]\| \leq c\frac{\max(m, \log^2 n)}{n}.
\]
Moreover, for all $\beta \in (0, \infty)$ with probability at least $1 - ce^{-c\log^2 n}$ we have
\[
\|T_n[1,m] - J[1,m]\| \leq c\frac{\max(m, \log^2 n)}{n}.
\]

Corollary 32. The spectral measure $\mu$ of $J$ equals

(i) the limit of the eigenvalue distribution of $T_n$ for every $\beta$ and

(ii) the limit of the empirical distribution of the Fekete points, the $n$-point minimizers of the density $ce^{-\sum V(\lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|}$.

The second characterization (ii) identifies $\mu_V$ as the classical equilibrium measure of potential theory associated with the potential $V$. In particular, $\mu_V$ and $J$ are non-universal. The fact that (i) equals $\mu_V$ regardless of $\beta$ is classical; we include it to clarify how it fits into our framework. The proof a simple consequence of the steps we need to prove our main result.

Proof of Corollary 32. For claim (ii), note that the $\beta = \infty$ energy associated with the tridiagonal matrix $T_n(a,b)$ can be expressed in terms of $q_i, \lambda_i$ using the expression (7):
\[
H(a, b) = \text{tr}(V(T_n(a, b))) + \sum_{k=1}^{n-1} (1 - k/n) \log(b_k) = \log \left( \prod_{i=1}^{n} q_i e^{-V(\lambda_i)} \prod_{i<j} |\lambda_i - \lambda_j| \right) + c
\]
so $H(a, b)$ is minimal exactly when the $\lambda_i$ are the Fekete points, and all the $q_i$ are equal. In particular, for $\beta = \infty$, the spectral measure of the tridiagonal matrix $T_n^\circ = T_n(a^\circ, b^\circ)$ built from the minimizers of $H(a, b)$ is the same as the empirical distribution of the Fekete points.

By the Proposition, the $k$-th moment of this measure is given by the $\beta = \infty$ case
\[
((T_n^\circ)^k)_{1,1} = ((T_n^\circ[1,k])^k)_{1,1} \to ((J[1,k])^k)_{1,1} = (J^k)_{1,1},
\]

A[$\ell, k$] for the minor of a matrix $A$ drawn from the rows and columns with indices in the set {$\ell, k$}.
and so these measures converge weakly to the spectral measure of \( J \), showing (ii).

The same argument, using \( T_n \) in the finite \( \beta \) case, shows that the spectral measures converge to that of \( J_k \). Finally, recall Proposition 2 that for the \( \beta \)-ensemble, the spectral measure is just a reweighted version of the empirical distributions, where the weights are Dirichlet(\( \beta/2, \ldots, \beta/2 \)), independent from the eigenvalues. If we realize the Dirichlet distribution as independent Gamma variables normalized by their sum, (i) follows easily from the law of large numbers.

**Proof of Proposition 31.** We first handle the \( \beta \in [1, \infty) \) case.

Let \( T'_{n,m} \) be the \( m \times m \) matrix built from the minimizer of the main Hamiltonian \( H = H_n \) of (9) with boundary conditions given by the corresponding entries of \( J \) on \( m - \deg V/2, \ldots, m \) on the right, and zero on the left. The entries of \( T'_{n,m} \) and \( J[1, m] \) minimize different Hamiltonians with the same boundary conditions. Proposition 20 gives the bound

\[
\| J[1, m] - T'_{n,m} \| \leq cm/n, \tag{59}
\]

in a slightly altered form: for the maximum of entries of the difference matrix. The operator norm bound follows from this and close and the Greshgorin circle theorem (the matrix is diagonally dominant).

To compare \( T'_{n,m} \) to \( T^\circ \), Corollary 23 gives that for \( j = m - \deg V/2, \ldots, m \) the entries \( a^\circ, b^\circ \) of \( T^\circ \) satisfy

\[
\|(a_j^\circ, b_j^\circ) - (a_j^\dagger, b_j^\dagger)\| \leq c \frac{\max(m, \log^2 n)}{n},
\]

by analyticity of \( a^\dagger, b^\dagger \) we have

\[
\|(a_j^\dagger, b_j^\dagger) - (a_j^\dagger(0), b_j^\dagger(0))\| \leq c \frac{m}{n}.
\]

Now (56) gives that for \( m \geq c \log n \) we have

\[
\|(a_j^\circ, b_j^\circ) - (a_j^\dagger(0), b_j^\dagger(0))\| \leq c \frac{m}{n}.
\]

So the boundary conditions for \( T'_{n,m} \) and \( T^\circ \) are indeed very close, and they minimize the same Hamiltonian. This implies that the minimizers are also close by Proposition 17:

\[
\| T'_{n,m} - T^\circ[1, m] \| \leq c \frac{m}{n}
\]

which, together with (59) shows the first claim (57).

Now assume \( \beta \in [1, \infty) \). Proposition 29 gives the entries \( A, B \) of the random tridiagonal matrix \( T_n \) satisfy

\[
\mathbb{P}(\|(A_k, B_k) - (a_k^\dagger, b_k^\dagger)\| < cn^{-1/2+\delta} \text{ for all } m - \deg V/2 \leq k \leq m) \geq 1 - e^{-cn^\delta}
\]

41
now given these entries, the $A_k, B_k$ for $k < m - \deg V/2$ are picked from the conditional Hamiltonian. These conditional minimizers are now $n^{-1/2+\delta}$-close to $T^\circ$ by Proposition 17, since the boundary conditions are so close (with $m' = \max(m, \log^2 n)$). The Gaussian Lemma 24 shows that the deviations from the minimizer are bounded above by constant times the norm-squared of an i.i.d. Gaussian vector of dimension $m$ and variance $c/n$. So it is exponentially (in $m$) unlikely to be more than $cm/n$. In summary, we get that with probability at least $1 - e^{c\log^2 n}$, we get the desired conclusion

$$\|T_n[1, m] - T_n^\circ[1, m]\| \leq \frac{m'}{n}.$$  

The $\beta < 1$ case can be treated by conditioning on the $b$ terms with positive log coefficient the same way as in Proposition 25.

Next, denote by $j_{0,0}$ the first zero of the Bessel function $J_0$ of the first kind. It will be crucial for the proof that we have the strict inequality $j_{0,0} \sim 2.40482 > \pi/2$. Let also $\lambda_{\max}(A)$ denote the largest eigenvalue of a matrix $A$. The point of Proposition 31 is that top corner of $T_n$ looks more and more like $J_k$, and next we show that $J_k$ has no large eigenvalues.

**Proposition 33.** We have that $\lambda_{\max}(J_k) \leq a^\dagger(0) + b^\dagger(0)(2 - (j_{00}/k)^2) + o(k^{-2})$.

For the proof of this proposition, we recall a few facts from both orthogonal polynomial theory and potential theory.

**Fact 34.** The following hold.

- For any probability measure $\mu$ of bounded support there exists a unique Jacobi operator $J$ with spectral measure $\mu$. If the support of $\mu$ consists of $m$ points, then $J$ reduces to an $m \times m$ Jacobi matrix.
- The top $k \times k$ minor $J_k$ of $J$ is uniquely determined by the first $2k - 1$ moments of $\mu$.
- For $k \leq m$ the spectral measure $\mu_k$ of $J_k$ satisfies
  $$\text{essup}(\mu_k) = \lambda_{\max}(J_k) \leq \text{essup}(\mu).$$  
  (60)
- For the case when $\mu$ is the uniform measure on $[-1, 1]$, then $\mu_k$ is supported on the zeros of the $k$th Legendre polynomial. The Bessel asymptotics for the Legendre polynomials imply that
  $$\text{essup}(\mu_k) = 1 - (j_{00}/k)^2/2 + o(k^{-2}).$$  
  (61)
• When \( V \) is convex, the measure \( \mu = \mu_V \) has continuous positive density on the interior of its support \([-2b^\dagger(0)+a^\dagger(0), 2b^\dagger(0)+a^\dagger(0)]\). The density is monotone decreasing in a neighborhood of the top endpoint \( 2b^\dagger(0)+a^\dagger(0) \).

The first item is typically referred to as Favard’s Theorem [1]. The second is implied by Heine’s formulas, see again Section 3.1 of [9]. The estimate on the largest zero of the \( k \)th Legendre polynomial may be found in Chapter 6 of [22]. For the last fact, use [20] Theorem IV.1.10 (b) (the support is an interval) and [16] (in particular, that the density is the root of the positive part of a real analytic function), whence the desired monotonicity.

Assembling the above items leads to the next lemma, from which Proposition 33 quickly follows.

**Lemma 35.** Let \( \mu \) be a probability measure with support \([-2, 2]\) and having concave density. Then for any \( m \geq 1 \), there is a measure that has the same first \( 2m - 1 \) moments as \( \mu \) and is supported on a subset of \([-2, 2 - m^{-2}\eta_m]\), where

\[
\eta_m \to \eta = j_{0,0}^2.
\]

**Proof.** For a sufficiently small \( \epsilon \) by the last item of Fact 34 we may write \( \mu \) as a convex combination of \( \nu_t, t \leq \epsilon \) and \( \nu \), where, \( \nu_t \) is the uniform measure on \([-2 + \epsilon, 2 - t]\) and \( \nu \) is supported on \([-2, 2 - \epsilon]\).

For each \( \nu_t \), by (61) there are \( \gamma_m \to j_{0,0}^2/2 \) so that we may find a measure supported on \([-2, 2 - (2 - \epsilon/2)m^{-2}\gamma_m]\) and having the same first \( 2m - 1 \) moments as \( \nu_t \).

Thus we may find a measure supported on

\[
[-2, 2 - (\epsilon \wedge (2 - \epsilon/2)m^{-2}\gamma_m)]
\]

and having the same first \( 2m - 1 \) moments as \( \mu \). Choose \( \epsilon_m = 1/m \) and set \( \eta_m = (2 - 1/m)\gamma_m \) to get the desired claim. \( \square \)

**Proof of Proposition 33.** By shifting and scaling we may assume \( a^\dagger(0) = 0, b^\dagger(0) = 1 \). Lemma 35 shows that there is a measure \( \mu' \) that has the same first \( 2k - 1 \) moments as \( \mu \) and satisfies \( \operatorname{essup}(\mu') \leq 2 - (j_{00}/k)^2 + o(k^{-2}) \). By Fact 34, the matrix \( J' \) corresponding to \( \mu' \) has the same \( k \times k \) top minor as \( J \), and so by (60) \( \lambda_{\max}(J_k) \leq \operatorname{essup}(\mu') \), as required. \( \square \)

Finally, the following theorem will be used to show that for the sake of edge asymptotics we can safely ignore the first \( m \times m \) block of the matrix \( T_n \) as long as \( m \) is not too large. For this theorem, let \( T_{|[a,b]} \) denote the matrix \( T \) where all entries except the ones with both indices in \([a, b]\) are set to zero.
**Theorem 36.** There exists \( c, c_0, \delta, \kappa > 0 \) so that with probability at least \( 1 - e^{-n^\delta} \) for every \( m \in [c \log n, c_0 n^{1/3}] \), and large enough \( n \) the following holds in the positive definite order:

\[
T \leq (b^\dagger(0) - \frac{\kappa}{m^2}) e_{mm} + T_{[1,m]} + (2b^\dagger(0) + a^\dagger(0) - \frac{\kappa}{m^2}) I_{[1,m-1]}.
\]

We will need the following simple lemma. Let \( \lambda_{\text{max}} \) denote the top eigenvalue, and let \( e_{ij} \) denote the elementary matrix with zeros everywhere except for a one at entry \( i, j \); its dimension will be clear from the context.

**Lemma 37.** Let \( T \) be an \( n \times n \) tridiagonal matrix with positive off-diagonals, let \( m \in \{1, \ldots, n - 1\} \), and let \( q = T_{m,m+1} \). We have

\[
\lambda_{\text{max}}(T) = \max_{r > 0} \min \left( \lambda_{\text{max}}(T[1,m] + q r e_{mm}), \lambda_{\text{max}}(T[m+1,n] + \frac{q}{r} e_{11}) \right).
\]

**Proof.** The two \( \lambda_{\text{max}} \) expressions on the right are nondecreasing (respectively, nonincreasing) as functions of \( r \), so it suffices to show that for some \( r > 0 \) we have

\[
\lambda_{\text{max}}(T) = \lambda_{\text{max}}(T[1,m] + q r e_{mm}) = \lambda_{\text{max}}(T[m+1,n] + \frac{q}{r} e_{11}).
\]

By adding a sufficiently large constant times identity to \( T \) we may assume without loss of generality that it has nonnegative entries. By the Perron-Frobenius theorem, there exists an eigenvector \( \varphi \) of \( T \) with nonnegative entries for the positive eigenvalue \( \lambda_{\text{max}}(T) \). If there were neighboring coordinates \( i, j \) so that \( \varphi_i = 0 \) and \( \varphi_j \neq 0 \), then the eigenvalue equation at coordinate \( i \) would fail. Thus \( \varphi \) is strictly positive.

Now with \( r = \frac{\varphi_{n+1}}{\varphi_n} \), \( \varphi \) restricted to the first \( m \) coordinates is an eigenvector of \( T[1,m] + q r e_{mm} \) with eigenvalue \( \lambda_{\text{max}}(T) \). Moreover, it must be a top eigenvector: by the Perron-Frobenius theorem, there is a nonnegative top eigenvector which cannot be orthogonal to the positive \( \varphi \). Similarly, \( \varphi \) restricted to the last \( n - m \) coordinates is a top eigenvector of \( T[m+1,n] + \frac{q}{r} e_{11} \) with eigenvalue \( \lambda_{\text{max}}(T) \), as required. \( \square \)

**Proof of Theorem 36.** By shifting and scaling we may assume that \( a^\dagger(0) = 0, b^\dagger(0) = 1 \). We decompose \( T \) as follows:

\[
T = \left( T_{[1,m]} + (1 - T_{mm}) e_{mm} - (2 - \frac{\kappa}{m^2}) I_{[1,m]} \right) + \left( T_{[m,n]} + (2 - \frac{\kappa}{m^2}) I_{[1,m-1]} + (1 - \frac{\kappa}{m^2}) e_{mm} \right)
\]

It suffices to show that the nontrivial part \( T[1,m] + (1 - T_{mm}) e_{mm} - (2 - \frac{\kappa}{m^2}) I_m \) of the first matrix is negative definite with high probability. In light of Proposition 33, and (56), we have

\[
|T_{mm}|, ||T[1,m] - J[1,m]|| < c \max(m, \log^2 n)/n < cc_0/m^2
\]
and so (by decreasing $\kappa$ a bit and choosing $c_0$ small enough) it suffices to show that
\[
\lambda_{\max}(J[1, m] + e_{mm}) < 2 - \frac{\kappa}{m^2}.
\] (62)

Now Proposition 33 gives us control of $\lambda_{\max}(J[1, m])$, but here we clearly need a little bit more. For this, we let $p > 1$ so that $pm$ is an integer, and study the matrix $J[1, pm]$. By (56) the matrix $J[m + 1, pm]$ is exponentially close in norm to the $\ell \times \ell$ matrix $D_{\ell}$ with 1-s on the first off-diagonals and zeros elsewhere (and $\ell = pm - m + 1$). It is easy to check that
\[
\lambda_{\max}(D_{\ell} + e_{11}) = \lambda_{\max}(D_{2\ell}) = 2 \cos\left(\frac{\pi}{2\ell + 1}\right) = 2 - \frac{\pi^2}{4\ell^2} + O(\ell^{-3}).
\]

Since $q = J_{m,m+1}$ is exponentially close to 1 by (56), we have
\[
\lambda_{\max}(J[m + 1, pm] + q^2e_{11}) = 2 - \frac{\pi^2}{4(p-1)^2m^2} + O(m^{-3}).
\]

If $p$ is large enough so that $\pi/(2j_{00}) > 1 - 1/p$, then for large $m$ this is bounded below by
\[
2 - \frac{j_{00}^2}{p^2m^2} + o(m^{-2}) \geq \lambda_{\max}(J[1, pm]),
\]
the last inequality coming from Proposition 33. Lemma 37 applied to $J[1, pm]$ with $r = 1/q$ now gives
\[
\lambda_{\max}(J[1, m] + e_{mm}) \leq \lambda_{\max}(J[1, pm]) = 2 - \frac{j_{00}^2}{p^2m^2} + o(m^{-2})
\]
showing (62) and thus the Theorem. In fact, carefully following the constants shows that we may use $\kappa = 1/2$.

\begin{flushright}
\Box
\end{flushright}

11 Mean and variance of the limiting potential

The final ingredient needed to establish our main result is the computation of mean and variance of the limiting (Brownian) potential of the operator identified by Theorem 9 (the limit of the processes defined in (28)).

Proposition 29 shows that the field over the required range is close to a mixture of Gaussians, centered at certain conditional minimizers and with variance given in terms of the inverse Hessian of the associated conditional Hamiltonian (evaluated at those conditional minimizers). Further, with the event $S$ in that statement holding with high probability, any of these conditional minimizers are close to the corresponding local minimizers, recall (53). Thus we expect the limiting mean and variance to be given only in terms of the local quantities.

45
Return to the given (by the Dyson beta ensemble) conditional Hamiltonian: for a set of indices $I$,

$$H_I(a, b) = V_I(a, b) - \sum_{k \in I \cap \{1, \ldots, n-1\}} \left(1 - \frac{k}{n} - \frac{1}{n^2} \right) \log b_k,$$

where again $V_I(a, b) := \text{tr}(V(T_I(a, b))$ for the minor $T_I$ tied to the index set $I$.

For the mean, the following provides the basic ingredient. The condition (64) below distills what is required out of Proposition 29. It is a deterministic condition that implies the bound (53) as a direct consequence of Corollary 23.

**Lemma 38.** Fix $\delta > 0$. Let $I = \{\lfloor c \log n \rfloor, \ldots, \lfloor n^{1/3} - \delta \rfloor\}$ and assume that for some $a_i, b_i$ we have

$$|a_i - a_i^\dagger| + |b_i - b_i^\dagger| \leq n^{\delta-1/2} \quad \text{for all } i \in \partial I,$$

where $(a_i^\dagger, b_i^\dagger)$ are the local minimizers defined in (12). In $I$, let $(a, b)$ be the minimizers of the conditional Hamiltonian $H_I$ with the above boundary conditions. It holds that

$$n^{1/3} \sum_{k = e \log n}^{\lfloor xn^{1/3} \rfloor} (a_k - a_k^\dagger) \rightarrow \frac{1}{2} (a^\dagger)'(0) x^2, \quad n^{1/3} \sum_{k = e \log n}^{\lfloor xn^{1/3} \rfloor} (b_k - b_k^\dagger) \rightarrow \frac{1}{2} (b^\dagger)'(0) x^2$$

as $n \rightarrow \infty$.

**Proof.** Consider the expression over the $a$-variables (the treatment for the $b$-variables is identical). Re-centering to arrive at

$$n^{1/3} \sum_{k = e \log n}^{\lfloor xn^{1/3} \rfloor} (a_k - a_k^\dagger) + n^{1/3} \sum_{k = e \log n}^{\lfloor xn^{1/3} \rfloor} (a_k^\dagger - a_0^\dagger),$$

the claimed evaluation comes from the second sum. By the analyticity of the local minimizer $x \mapsto a^\dagger(x)$ this sum may be replaced by $n^{1/3} \sum_{k=1}^{\lfloor xn^{1/3} \rfloor} (a^\dagger(k/n) - a^\dagger(0))$ which, again by analyticity, is asymptotic to $(a^\dagger(0))'$ times $n^{1/3} \sum_{k=1}^{\lfloor xn^{1/3} \rfloor} (k/n + O(k^2/n^2)) = x^2/2 + O(n^{-2/3})$.

As for the first sum, Corollary 23 implies that, given that (64) is in place, we have the bound

$$|a_k - a_k^\dagger| \leq c \max \left(n^{\delta-1/2} \exp(-c_1 \text{dist}(k, \partial I)), \frac{(\log n)^2}{n} \right).$$

It follows that the sum of $n^{1/3}|a_k - a_k^\dagger|$ up to $xn^{1/3}$ is bounded above by $c \times n^{1/3} n^{\delta-1/2} \log n \rightarrow 0$, as required.

For the variance we record:
Lemma 39. Keep the setup of the previous lemma: given $\delta > 0$ and $I = \{\lfloor c \log n \rfloor, \ldots, \lfloor n^{1/2-\delta} \rfloor\}$ assume that for some $a_i, b_i$ we have

$$|a_i - a_i^\dagger| + |b_i - b_i^\dagger| \leq n^{\delta-1/2} \quad \text{for all } i \in \partial I.$$  

Here again $(a_i^\dagger, b_i^\dagger)$ are the local minimizers, and in $I$, let $(a, b)$ are the minimizers of the conditional Hamiltonian $H_I$ with the above boundary conditions. Fix $0 \leq x < y$, and let and let $\xi_a$ be the indicator of an subinterval $[\max(c \log n, xn^{1/3})], \ldots, \lfloor yn^{1/3} \rfloor$ of $I$, of the $a$ variables. Define $\xi_b$ analogously. Let $H(a, b) = \text{Hess}_{H_I}(a, b)$. Then as $n \to \infty$ we have

$$n^{-1/3} (\xi_a^I, H^{-1} \xi_b^I) \to (y - x) \Sigma(0)$$  \hspace{1cm} (66)$$

where $i, j \in \{a, b\}$ and the covariance matrix $\Sigma(x)$ is defined in (17).

Proof. By polarization and linearity, it suffices to show that for any normalized eigenvector $(\alpha_a, \alpha_b)$ of $\Sigma(0)$ with eigenvalue $\lambda$, and $w = w_n = \alpha_a \xi_a^I + \alpha_b \xi_b^I$ we have

$$n^{-1/3} \langle w, H^{-1} w \rangle \to (y - x) \lambda.$$  

Consider the Hessian of the local Hamiltonian (10) evaluated at its minimizer. This is invariant under the rotation of the indices, so the space spanned by the indicator $\xi_a$ of all variables $a$ and the indicator $\xi_b$ of all $b$ variables form an invariant subspace. Then by the discussion before Proposition 7 it holds that $v = \alpha_a \xi_a + \alpha_b \xi_b$ is an eigenvector with eigenvalue $\lambda^{-1}$.

Note that the matrix $C$ is a block circulant with $2 \times 2$ blocks. It is also a band matrix – any blocks entry whose two block indices are further than $\deg V/2$ apart mod $|I|$ are zero. This implies that if two vectors $v$ agree on a long interval (of blocks), then this will hold after applying $\mathbb{H}$, apart from a short section of $\deg_V$ blocks that could have changed. Note that the vector $v$ agrees with $v$ on a long stretch and then with 0 on another long stretch. Thus we have

$$\|Cw - \lambda^{-1}w\| \leq c$$

the constant comes from the boundary error terms, which are bounded and are of a bounded number. Moreover, the same consideration shows that if $C'$ is $C$ with the periodic boundary removed (i.e. all block entries with indices farther apart than $\deg V/2$ changed to zero), we have

$$\|C'w - \lambda^{-1}w\| \leq c$$

47
Note that $C'$ is Hessian for the non-periodic Hamiltonian (41), with all coefficients of the log terms there given by 1. The Hessian is for the variables with indices in $I$, evaluated when all variables (including the boundary conditions with indices in $\partial I$) are set to $a^\dagger(0), b^\dagger(0)$, respectively. In particular, $C'$ is bounded below by a constant times the identity, which implies
\[
\|\lambda w - C^{-1}w\| \leq \lambda \|C'^{-1}\| \|C'w - \lambda^{-1}w\| \leq c.
\]
We can now use the assumption on the boundaries along with Corollary 23 to compare the entries of $C'$ and $H$. These are given by polynomials of $a_i, b_i$, plus a constant times $b_i^{-2}$, for $H$, and exactly the same expressions with all $a_i, b_i$ replaced by $a^\dagger(0), b^\dagger(0)$ for $C'$. By the proposition the $a_i$ are $cn^{\delta-1/2}$-close to $a_i^{\dagger}$, and $a_i^{\dagger} = a^\dagger(i/n + o(1/n))$ are at most $cn^{\delta-1/2}$-close to $a(0)$. This is because the function $a$ is analytic and so Lipschitz in a neighborhood of 0 (5), and we are on a stretch where $i/n \leq cn^{1/2-\delta}$. The same considerations hold for the $b$ variables.

Since $H$ and $C'$ are band matrices, the Greshgorin circle theorem implies that for the operator norm $\|H - C'\| \leq cn^{\delta-1/2}$, and so
\[
\|H^{-1} - C'^{-1}\| = \|H^{-1}(C' - H)C'^{-1}\| \leq \|H^{-1}\| \|C' - H\| \|C'^{-1}\| \leq cn^{\delta-1/2}.
\]
Thus we get the desired conclusion
\[
\langle w, H^{-1}w \rangle = \lambda^{-1}\|w\|^2 + O(n^{1/3}n^{\delta-1/2}) + O(1).
\]

\section{Putting it all together}

Everything is now in place to prove our main result, which state more completely as follows.

\textbf{Corollary 40}. Let $\gamma = (b^\dagger(0))^{-1/3}\tau^{-2/3}$ for $\tau = -(a^\dagger)'(0) - 2(b^\dagger)'(0)$, let $\vartheta = b^\dagger(0)/\tau$, and let $\mathcal{E} = a^\dagger)'(0) + 2(b^\dagger)'(0)$. There exists a coupling of the random matrices $T_n$ on the same probability space so that a.s. we have
\[
\gamma n^{2/3}(\mathcal{E} - T_n) \rightarrow \text{SAO}_\beta
\]
in the norm-resolvent sense: for every $k$ the bottom $k$th eigenvalue converges the and corresponding eigenvector converges in norm. Here $\mathcal{E} - T_n$ acts on $\mathbb{R}^n \subset L^2(\mathbb{R}_+)$ with coordinate vectors
\[
e_j = (\vartheta n)^{1/6}1_{[j-1,j]}(\vartheta n)^{-1/3}.
\]
The next two subsections check the conditions of Theorem 9, in particular how part (i) of that theorem applies to the (centered and scaled) truncated matrix $H_n = \gamma_n^{2/3} (EI - T_{[c, \log n, n]})$. After this, the results of Section 10 are incorporated to bypass the truncation and prove the corollary.

12.1 Tightness and Brownian convergence

The following establishes Assumption 1 of Theorem 9, in particular the tightness and convergence of the processes $x \mapsto y_{n,1}(x), y_{n,2}(x)$ defined in (28). Giving the limit of the (integrated) potential, this result already identifies the Stochastic Airy Operator.

Here we can assume that the conclusions of Proposition 29 are in place. Said another way we can assume we are on the event $S$ defined there.

**Proposition 41.** Consider the measures on paths

\[
X_n(x) = n^{1/3} \sum_{k = c \log n}^{\lfloor n^{1/3}x \rfloor} (b_k^I - B_k),
\]

\[
Y_n(x) = n^{1/3} \sum_{k = c \log n}^{\lfloor n^{1/3}x \rfloor} (a_k^I - A_k).
\]

For $c$ chosen large enough, $X_n$ and $Y_n$ from a tight family with respect to the uniform convergence on compact sets. Moreover, we have the following convergence in distribution with respect to the uniform-compact topology:

\[
X_n + 2Y_n \Rightarrow \sigma W_x + \frac{\tau}{2} x^2. \tag{67}
\]

Here $x \mapsto W_x$ is a standard Brownian motion and

\[
\tau = -E'(0) = -\left((a^I)'(0) + 2(b^I)'(0)\right), \quad \sigma^2 = \frac{4}{\beta} b^I(0) \tau. \tag{68}
\]

Recall from (18) that $\tau > 0$.

**Proof.** We first switch to variables picked from the distribution $\nu_I$ introduced before Proposition 29. We will still denote these random variables $A_k, B_k$. The $c$ in the statement should simply be large enough that we can assume the outcome of Proposition 29.

To prove tightness, note that it suffices to do this for the components of the mixture of $\nu_q$’s $\nu_I$, since mixtures from tight families are also tight. Thus let

\[
Z_n(x) = n^{1/3} \sum_{k = c \log n}^{\lfloor n^{1/3}x \rfloor} (B_k - b_k^q).
\]
be the partial sums as above from the sample from the normal distribution corresponding to the variables $B_k$.

By Kallenberg [15], Corollary 16.9, in order to establish tightness for the sequence $Z_n$ it suffices to show that the Kolmogorov-Chentsov criterion holds $Z_n$ and uniformly for all $n$ and boundary conditions $q$. In our case, this comes down to showing that

$$
E|Z_n(x) - Z_n(y)|^2 \leq c|x - y| \text{ for all } x, y \text{ such that } |x - y| < 1.
$$

(69)

Since the increments are normally distributed, this automatically implies a 4th moment bound that works as an input for the Kolmogorov-Chentsov criterion, Lemma 43.

Indeed, Lemma 3 implies that the Hessian $\text{Hess} \geq c_u I$, whence $\text{Hess}^{-1} \leq c_u^{-1} I$ uniformly over $q$ and $n$. Let $v$ be the indicator of the coordinates of Hess corresponding to $B_k, \ldots B_l$. Then we have that

$$
E[(B_k - b_k^{(q)}) + \ldots + (B_{l-1} - b_{l-1}^{(q)})]^2 = \langle v, (n \text{ Hess})^{-1} v \rangle \leq \frac{c}{n} \langle v, v \rangle = \frac{c(l - k)}{n}
$$

which, after appropriate scaling, gives precisely (69).

We also need to prove tightness of the integrated drifts,

$$
n^{1/3} \sum_{k = c \log n}^{[n^{1/3}x]} (b_k^1 - b_k^{(q)}).
$$

However, this along with the convergence to the desired value of $-\frac{1}{2}(a^\dagger)'(0) x^2$ is the content of Lemma 38. Note again what is required from the condition (64) in that lemma is just the final statement of Proposition 29. In the same way Lemma 39 shows that $\text{Var}[Z_n(x)]$ converges.

Since each $Z_n$ is a Gaussian process, the convergence of the means and variances (plus tightness) is enough to yield unique limit in the uniform-on-compacts topology. That the limit process is the appropriately shifted and scaled Brownian motion (to match the given mean and variance) is implicit in the covariance structure described in Lemma 39.

This establishes tightness and convergence for $Z_n$. Again, that for $Y_n$ follows from Proposition 29 which shows that the processes are close in total variation. The same procedure gives the desired conclusions for the $X_n$ process. The integrated drifts in that case converging to $-\frac{1}{2}(a^\dagger)'(0) x^2$, thus giving the formula for $\tau$.

To verify the formula for $\sigma^2$, return to the formula given in Proposition 7 and compute

$$
\beta \sigma^2 := \Sigma(0)_{11} + 4 \Sigma(0)_{22} + 4 \Sigma(0)_{12} = -4b^1(0) \left((a^\dagger)'(0) + 2(b^\dagger)'(0)\right).
$$

\[\square\]
12.2 Oscillation and growth bounds

We establish the non-immediate conditions of Assumption 2, the bounds (23) and (25).
Recall from Section 4 that the definition of growth and oscillation terms differs depending
on the range of indices.

On the interval \([c \log n, (1 - \epsilon) n]\)

The \(\epsilon\) is as defined in Proposition 29, and again we can assume that we are on the event
\(S\) on which the various bounds described in that proposition all hold. On this stretch, the
oscillation (or noise) terms are defined, up to constants, by

\[ w_{n,1}(x) = n^{1/3} \sum_{k=[c \log n]}^{\lfloor x n^{1/3} \rfloor} (a_k - \hat{A}_k), \]

and similarly for \(w_{n,2}\) in the \(B\)-variables. The growth terms, \(\eta_{n,i}(x)\), then read, again
up to constants,

\[ n^{2/3}(a_0 - a_{\lfloor n^{1/3} x \rfloor}) \text{ and } n^{2/3}(b_0 - b_{\lfloor n^{1/3} x \rfloor}). \]

Compare (29) and (30).

From the established differentiability of 

\[ x \mapsto a^\dagger(x), b^\dagger(x) \]

we see that 

\[ x/c \leq \eta_{n,i}(x) \leq cx \]

with a (deterministic) constant, so that (23) is satisfied with 

\[ \eta(x) = x \].

The following then shows (25) is satisfied with

\[ \zeta(x) = x^\delta \] for some \(\delta > 0\) small.

Proposition 42 (Oscillation bounds). Let \(k_0 = c \log n\), and for \(0 \leq x \leq (1 - \epsilon)n\) and some
sequence \(\alpha_i\) denote

\[ \text{osc}_{n,x}(\alpha) = \max_{0 \leq \ell < n^{1/3}} \left( n^{1/3} \sum_{i=k_0+\lfloor x n^{1/3} \rfloor}^{k_0+\lfloor x n^{1/3} + \ell \rfloor} \alpha_i \right)^2 \leq C_n(1 + x^{1-\delta}) \]

then for some \(\delta > 0\) and a tight sequence of random constants \(C_n\) we have for all \(n\) and \(0 \leq x \leq (1 - \epsilon)n\)

\[ \text{osc}_{n,x}(A_n - a_n^\dagger) + \text{osc}_{n,x}(B_n - b_n^\dagger) \leq C_n(1 + x^{1-\delta}). \]

Proof. We will show the claim for \(\text{osc}_{n,x}(A_n - a_n^\dagger)\), the proof for the \(a\) variables is identical.

Partition the interval \(k_0, \ldots, (1 - \epsilon)n\) into with subintervals \(I_k\) of size \(n^{1/2-\delta}\). Assume
that the indices \(k_0 + x n^{1/3}\) and \(k_0 + (x + 1)n^{1/3}\) fall in the same interval \(I_k\) (otherwise we
can split the sup and bound it in two steps).

Let

\[ X_{\ell} = X_{n,x,\ell} = n^{1/3} \sum_{i=k_0+\lfloor x n^{1/3} \rfloor}^{k_0+\lfloor x n^{1/3} + \ell \rfloor} A_{n,i} - A_{n,i}^\dagger \]

where \(A^\dagger\) is the (random) minimizer of the conditional Hamiltonian of the variables in \(I_k\),
given the values on \(\partial I_k\). Let \(S_{n,k}\) denote the event \(S\) in Proposition 29 applied to \(I_k\). Let
\( S_n = \bigcap_k S_{n,k} \). Note that by the Proposition \( \mathbb{P} S_n \to 1 \). Let \( R'_{n,x} = \text{osc}_{n,x}(A_n - A_n^\sharp) \) and \( R''_{n,x} = \text{osc}_{n,x}(a_n^\dagger - A_n^\sharp) \). It suffices to show that
\[
\left( \max_{0 \leq x \leq (1-\epsilon)n^{2/3}} R''_{n,x} \right) 1_{S_n} \to 0,
\]
in probability, and that we have
\[
\sum_{x=0}^{(1-\epsilon)n^{2/3}} \mathbb{E}\left[ R''_{n,x}^2 ; S_n \right] \leq c. \tag{70}
\]
For the first, note that by Proposition 29 on \( S_n \) we have
\[
|A_j^\sharp - a_j^\dagger| \leq c \max \left( n^{\delta - 1/2} \exp(-c_1 \text{dist}(j, \partial J)), \frac{\log n}{n} \right).
\]
It follows that the sum of \( n^{1/3}|A_j^\sharp - a_j^\dagger| \) over any range of length \( n^{1/3} \) is bounded above by \( cn^{1/3}n^{\delta - 1/2} \log n \to 0 \), as required.

By Proposition 29 on the event \( S_n \) and conditionally on \( Q \) the random vector \( (X_k, 0 \leq k \leq n^{1/3}) \) is close to a Gaussian process with mean zero and \( \mathbb{E}(X_i - X_j)^2 \leq c|i - j|/n^{1/3} \) in the sense of Radon-Nikodym derivatives being close to 1. By Lemma 43 below we have that
\[
\mathbb{E}\left[ (R_{n,x} - R'_{n,x})^2 ; S_n | Q \right] \leq 2c(n^{-1/6})^4(n^{1/3})^{4/2} = 2c
\]
where the factor 2 is to control the ratio of Radon-Nikodym derivatives. Taking expectations (70) follows.

**Lemma 43.** Let \( (X(k), k = 0, \ldots, n - 1) \) be a centered Gaussian vector. Assume that \( \mathbb{E}(X_i - X_j)^2 \leq \sigma^2|i - j| \) for all \( i, j \). Then for any \( q > 0 \) we have
\[
\mathbb{E} \sup_i |X_i - X_0|^q \leq c_q \sigma^q n^{q/2}.
\]

**Proof.** We prove this by showing a quantitative version of the Kolmogorov-Chentsov criterion – the proof is standard. We have
\[
\sup_{i,j} \frac{|X_i - X_j|}{|i - j|^r} \leq \frac{2^r}{2^r - 1} \sup_{\ell,k} \frac{|X(k^2\ell) - X((k+1)^2\ell)|}{2^{\ell r}},
\]
as can be seen by repeated use of the triangle inequality. Indeed, consider a shortest sequence \( i_1 = i, \ldots, i_m = j \) so that consecutive \( i \)'s differ by exactly one bit in their binary expansion. Then by the triangle inequality we get
\[
|X_i - X_j| \leq \sum_{\ell=0}^{\log_2 |i-j|} 2^{\ell r} s = \frac{2^r |i - j|^r - 1}{2^r - 1}.
\]
where $s$ is the supremum on the right hand side. The claim follows. Bounding the sup by the sum gives

$$
\left( \sup_{\ell,k} \frac{|X(k2^\ell) - X((k+1)2^\ell)|}{2^{r\ell}} \right)^q \leq \sum_{\ell=1}^{[\log_2 n]} \sum_{k=1}^{n/2^\ell} \frac{(X(k2^\ell) - X((k+1)2^\ell))^q}{2^{r\ell q}}
$$

The expectation of the right hand side is bounded above by

$$
q!! \sum_{\ell=1}^{[\log_2 n]} n2^{-\ell} \frac{(a^{2/2q})^{q/2}}{2^{r\ell q}} = q!! \sum_{\ell=1}^{[\log_2 n]} \frac{(2^{[\log_2 n]}q/2-rq-1)^{q/2-rq-1}}{2^{q/2-rq-1}} \leq c_{q,r} \sigma^q (n^{q/2-rq} \lor 1)
$$

and we get

$$
E \left( \sup_{i,j} \left| X_i - X_j \right| \right)^q \leq c_{q,r} \sigma^q (n^{q/2-rq} \lor 1)
$$

for $r < 1/2$ this gives, in particular

$$
E \sup_i |X_i - X_0|^q \leq c_{q,r} \sigma^q n^{q/2}.
$$

On the interval $[(1-\epsilon)n, n]$

Here we take the definitions in (31): $w_{n,i}(x) \equiv 0$ and $\eta_{n,1}(x) = m_n^2 (a_0^\dagger - A_{[x_m]})/b^\dagger(0)$, $\eta_{n,2}(x) = m_n^2 (b_0^\dagger - B_{[x_m]})/b^\dagger(0)$.

By the proposition that follows we have a tight upper bounds on the $A$’s and $B$’s over this span of indices. This is sufficient for (23) to hold as then $\eta_{n,i}(x) = O(n^{2/3})$ which is $O(x)$ throughout this span. The condition (24) is satisfied as the bound of $B_k$, $k \in [(1-\epsilon)n, n]$ can be made small: $B_k \leq \eta$ (with high probability) where $\eta < b_0^\dagger/2$, for example.

**Proposition 44** (Last entries are small). For every $\eta > 0$ there exists $\delta > 0$, with $\delta < 1$, and $c_1, c_2, c_3 > 0$ so that we have

$$
P \left( |A_k - a^\dagger(1)|, B_k \leq \eta \text{ for all } k \geq \delta n \right) \geq 1 - c_1 \exp(-c_2 n^{c_3}).
$$

**Proof.** This is an immediate corollary of Proposition 25 which shows there is $\delta' > 0$ so that for all $n$ and $s \geq 1 - \delta'$ we have

$$
P \left( |A_k - a^\dagger(1)|, B_k \leq \eta \text{ for all } k \geq sn \text{ and } |A_k - a^\dagger(1)|, B_k \leq \delta' \text{ for all } k \in [sn, sn + \deg V] \right) \geq 1 - e^{-c_2 n}.
$$

By Proposition 5 we have $a^\dagger(x), b^\dagger(x) \to (a^\dagger(1), 0)$ as $x \to 1$, and so Proposition 29 shows that for some $\delta, s$ the conditioning event above holds with the claimed probability. 

53
12.3 Proof of the main result

**Proof of Corollary 40.** Due to the interlacing inequalities, the lowest eigenvalues $\lambda_{n,1}, \lambda_{n,2}, \ldots$ of the matrix

$$\gamma n^{2/3} (\mathcal{E} I - T_n)$$

are bounded above by the lowest eigenvalues of any minor; we use the minor

$$H_n = \gamma n^{2/3} (\mathcal{E} I - T_{[c \log n, n]}),$$

and denote its lowest eigenvalues by $\bar{\lambda}_{n,1}, \bar{\lambda}_{n,2}, \ldots$

By Theorem 36, for $n$ large enough, the lowest eigenvalues of (71) are also bounded below by $\min(\lambda_{n,i}, y_n)$ where $y_n = cn^{2/3}/\log n$ and $\lambda_{n,i}$ are eigenvalues of a rank-1 perturbation of the same minor (72) at its first coordinate. The perturbation adds the matrix

$$\gamma n^{2/3} \left( -b^\dagger(0) + \frac{c}{\log^2 n} \right) e_{11} = \left( -m_n^2 + \frac{c'}{\log^2 n} m_n^2 \right) e_{11} := z_n e_{11}.$$ 

Recall from (27) the definition $(\gamma/n^{1}(0)) n^{2/3} = m_n^2$. Since $y_n \to \infty$, it suffices to show that the $\bar{\lambda}_{n,i}$ and the $\lambda_{n,i}$, the eigenvalues of $H_n$ and its perturbation $H_n + z_n e_{11}$, converge to the same limit.

By Theorem 9 (ii), in particular the condition (26), the perturbation here is subcritical: both sets of eigenvalues will converge to the same limit so long as $H_n$ converges in the manner set out in the first part of Theorem 9. But this has just been verified in the preceding subsections: Theorem 41 establishing convergence of the random potentials, or Assumption 1 (of Theorem 9), and Propositions 42 and 44 combined verifying the growth/oscillation conditions, Assumptions 2.

To identify the Stochastic Airy Operator $\text{SAO}_\beta$ only requires replacing the simple scalings on the $X_n(x)$ and $Y_n(x)$ sums in Theorem 41 with what is actually given via (27) and (28). With $a = a^\dagger(0)$ and $b = b^\dagger(0)$ for short, the convergence,

$$\left( \frac{n}{\tau b^2} \right)^{1/3} \frac{|x(bn/\tau)^{1/3}|}{k=c \log n} (a - A_k) + 2(b - B_k) \to \frac{1}{2} x^2 + \frac{2}{\sqrt{3}} (b\tau)^{1/2} \cdot \frac{1}{\tau^{1/3} \sqrt{2/3}} W((b/\tau)^{1/3} x),$$

may be read off from (67)-(68). By Brownian scaling the right hand side is equivalent in law to $\frac{1}{2} x^2 + \frac{2}{\sqrt{7}} W_x$ as is required.

For convergence of eigenvectors, note that we do have from Theorem 9 that the eigenvectors of $H_n$ and $H_n + z_n e_{11}$ converge to the same limit, the eigenvectors of $\text{SAO}_\beta$. Now
denote by $T_n$ the full matrix (71) and by $\tilde{T}_n$ its (positive definite) lower bound from Theorem 36, scaled to our setting. That is, $\tilde{T}_n = y_n I_{[1, c \log n]} + [H_n + z_n e_{11}]$, in which the previous perturbed minor is padded by zeros in the first $c \log n$ rows and columns.

We now show, by induction on $k$ that $|\langle \varphi_{nk}, \varphi_{nk} \rangle| \to 1$ i.e. the normalized eigenvectors are close in norm (up to sign).

Let $A_{n,k}$ denote the $k$-dimensional subspace spanned by the lowest eigenvectors of $\tilde{T}_n$. Assume that $n$ is large enough so that $y_n > \lambda_{n,k+1} \geq \lambda_{n,k}$, and the elements of $A_{n,k}$ vanish on the first $c \log n$ coordinates.

Write $\varphi_{n,k} = \varphi' + \varphi'' + \varphi'''$, where $\varphi'$ is the orthogonal projection to $A_{n,k-1}$, and $\varphi''$ is the orthogonal projection to the 1-dimensional space $\varphi_{n,k}$. Then, granted that $y_n \geq \Lambda_{n,k+1}$, we have

$$\lambda_{n,k} = \langle \varphi', T_n \varphi' \rangle \geq \langle \varphi', \tilde{T}_n \varphi'' \rangle \geq \lambda_{n,1} \|\varphi'\|^2 + \lambda_{n,k} \|\varphi''\|^2 + \lambda_{n,k+1} \|\varphi'''\|^2$$

where the last inequality follows from the eigenvector decomposition of $\tilde{T}_n$.

Writing the inequality between the leftmost and rightmost expressions as in

$$(\lambda_{n,k} - \lambda_{n,1}) \|\varphi'\|^2 + (\lambda_{n,k} - \lambda_{n,k}) \|\varphi''\|^2 \geq (\lambda_{n,k+1} - \lambda_{n,k}) \|\varphi'''\|^2,$$

we claim that the left hand side converges to zero. Indeed, by the inductive hypothesis $A_{n,k-1}$ is close to the subspace spanned by the first $k-1$ eigenvectors of $T_n$, so $\|\varphi'\| \to 0$, and the coefficient stays bounded. For the second term, we have established that $\lambda_{n,k} - \lambda_{n,k} \to 0$.

Now SAO$_{\beta}$ has almost surely discrete eigenvalues, and so $\lambda_{n,k+1} - \lambda_{n,k} > C > 0$ for a random constant $C$ and all $n$ large enough. Thus for the right hand side to converge to zero it has to be the case that $|\varphi'''| \to 0$. Hence $|\varphi'''| \to 1$, as claimed.

Thus the lower eigenvectors of $T_n$ and $\tilde{T}_n$ are close in norm. The latter are vanishingly small shifts of the eigenvectors of $H_n + z_n e_{11}$, which are close to the eigenvectors of SAO$_{\beta}$, as required.

13 The nonregular case

Even for the classical values of the parameter $\beta = 1, 2$ and $4$ there exist external fields $V$ for which one does not see Tracy-Widom limits at the edge of the spectrum. The relevant condition is understood through the behavior of the corresponding equilibrium measure $\mu_V$, again defined through the minimizing

$$I(\mu) = \int V(s) d\mu(s) + \int \int \log \frac{1}{|s-t|} d\mu(s) d\mu(t)$$

55
in the space of probability measures. If \( \frac{V(s)}{\log |s|} \to \infty \) as \( s \to \pm \infty \) the minimizer \((\mu_V)\) exists and is unique. If it is further assumed that \( V \) is real analytic, then it is the case that has a density of the form
\[
\psi_V(s) = \frac{d\mu_V(s)}{ds} = \sqrt{(Q_V(s))_+}
\]
for a real analytic function \( Q_V(x) \). Generically \( Q_V \) has simple zeros and \( \psi_V \) vanishes like a square root at the edge of its support [16]. In this case \( V \) is called regular and the rescaled maximal eigenvalue has the appropriate Tracy-Widom law (see [8], again for \( \beta = 1, 2, 4 \)).

In general though it is possible for \( Q_V \) to have a zero of order \( 4k' + 1, k' = 0, 1, \ldots \) (but not of order \( 4k' + 3 \)). The resulting “higher-order Tracy-Widom” laws which can result from such non regular \( V \)'s have been studied in [6] at \( \beta = 2 \). The eigenvalue fluctuations in these settings is of order \( n^{2/(4k+3)} \).

Throughout the above we have made essential use of the assumption that \( V \) is (uniformly) convex, specifically to control the regularity of the minimizers as well as the concentration of the field about the minimizers. In fact, the only simple geometric condition for \( V \) to be regular is convexity. Nonetheless, our framework provides conjectural operator limits, of type Laplacian plus random potential, describing the limiting spectral edge for any non regular \( V \).

To explain, one first notices that the shape of the potential (linear plus a constant multiple of white noise) that characterizes the Stochastic Airy Operator is tied directly to the differentiability of the local minimizers \( a^\dagger(x), b^\dagger(x) \) at zero. Really what is important is the shape of the edge \( E(x) = a^\dagger(x) + 2b^\dagger(x) \) tied to the family of potentials \( x \mapsto \frac{1}{1-x} V \), again about zero. Recall Remark 8 at the end of Section 4. It has already been noted in [16] that the behavior of \( \psi_V \) at \( E = E(0) \) determines that of \( E(x) \). Lemma 8.1 of that paper implies:
\[
\text{If } \psi_V(t) \sim (E - t)^{\frac{4k+1}{2}} \text{ as } t \uparrow E \text{ then } E - E(0) \sim \epsilon^{\frac{1}{4k+3}} \text{ as } \epsilon \downarrow 0.
\] (73)

Again, square-root vanishing of the density \((k = 0) \) produces linear/differentiable behavior of the edge \( E(x) \sim E(0) + E'(0)x \).

After centering the tridiagonal operator by \( E(0) \), and so again factoring \( \frac{b^\dagger(0)}{x} \) the discrete second-derivative operator, the summed potential once more takes the form
\[
y_n(x) = m_n \sum_{k=q_n}^{\lfloor m_n x \rfloor} (E(0) - (A_k + 2B_k)),
\]
compare (28). Here a variable lower limit \( q_n \) has been inserted assuming there will again be a cutoff required. Before of course \( q_n = c \log n \), what is important is that \( q_n = o(m_n) \).
Now assuming that $A_k, B_k$ concentrate about appropriate local minimizes, at least in the vicinity of the edge, one has up to constants that

$$E[y_n(x)] \sim m_n \sum_{k=q_n}^{[m_n x]} (k/n)^{1/k+1} \sim m_n \frac{n^{4k+3}}{2k+1} x \frac{2k+2}{2k+1},$$

(74)

granted (73). This implies that $m_n$ must be chosen so that $m_n \sim n^{1/4k+3}$, which corresponds to the general fluctuation exponent known to hold in the $\beta = 2$ case (the fluctuation is $O(m_n^2)$), as it should. Note that this can be cast in terms of the concentration of the random “edge” $A_k + 2B_k$, without discussion of separate $(a, b)$-minimizers.

For the variance, the formula in the regular case (see Proposition 41) can be written in an intuitive manner as in

$$-2\beta \mathcal{E}(0) \mathcal{E}'(0)x = -\beta (\mathcal{E}^2)'(0)x \sim \beta n^{2/3} \sum_{k \leq n^{1/3}x} -(\Delta \mathcal{E}^2)(k/n),$$

where $\Delta f(t)$ is short for $f(t + 1/n) - f(t)$. On the left hand side, one recognizes the formula given previously, with constant $-4\beta b^1(0) \mathcal{E}'(0)$, noting that it can be assumed that $a^1(0) = 0$ by a simple shift in which case $\mathcal{E}(0) = 2b^1(0)$. The right hand side captures that the (Gaussian) fluctuation a single step of the walk $k \rightarrow A_k + 2B_k$ should scale with the difference-square of its mean profile, $\mathcal{E}(k/n)$. This prompts

$$\text{Var}[y_n(x)] \sim m_n^2 (\mathcal{E}^2(0) - \mathcal{E}^2(xm_n/n)) \sim m_n^2 (x_{m_n/n})^{1/k+1},$$

(75)

as the proposed variance, neglecting constants. Note that the exponents of $m_n$ and $n$ match those in (74).

Restoring the appropriate constant factors, (74) and (75) allow us to formulate the following.

**Conjecture 45.** Assume $V$ is nonregular: with $k \geq 1$, let $\psi_V(t) \sim (\mathcal{E} - t)^{4k+1}$ as $t \uparrow \mathcal{E} = \mathcal{E}(0)$, the rightmost edge of the support of $\psi_V$. In line with (73) define the constant $c$ by

$$\lim_{\epsilon \downarrow 0} \frac{\mathcal{E} - \mathcal{E}(\epsilon)}{\epsilon^{1/k+1}} = c,$$

and set $\gamma = c^{-2/3}(\mathcal{E}/2)^{-1/3}$. Then we have that

$$H_{n,k} = \gamma n^{2/4k+3} (\mathcal{E} I - T_n)$$

converges in the sense of Theorem 9 (i) to the operator

$$S_{\beta,k} = -\frac{d^2}{dx^2} + x^{1/k+1} + \frac{2}{\sqrt{\beta}} x^{-1/k+1} W'(x),$$

on the half-line with Dirichlet conditions at the origin. In particular, the ordered eigenvalues/eigenvectors of $H_{n,k}$ converges jointly in law to those of $S_{\beta,k}$.
References


