STURM SEQUENCES AND RANDOM EIGENVALUE DISTRIBUTIONS

JAMES T. ALBRECHT, CY P. CHAN, AND ALAN EDELMAN

ABSTRACT. This paper proposes that the study of Sturm sequences is invaluable in the numerical computation and theoretical derivation of eigenvalue distributions of random matrix ensembles. We first explore the use of Sturm sequences to efficiently compute histograms of eigenvalues for symmetric tridiagonal matrices and apply these ideas to random matrix ensembles such as the β-Hermite ensemble. Using our techniques, we reduce the time to compute a histogram of the eigenvalues of such a matrix from $O(n^2 + m)$ to $O(mn)$ time where n is the dimension of the matrix and m is the number of bins (with arbitrary bin centers and widths) desired in the histogram (m is usually much smaller than n). Second, we derive analytic formulas in terms of iterated multivariate integrals for the eigenvalue distribution and the largest eigenvalue distribution for arbitrary symmetric tridiagonal random matrix models. As an example of the utility of this approach, we give a derivation of both distributions for the β-Hermite random matrix ensemble (for general β). Third, we explore the relationship between the Sturm sequence of a random matrix and its *shooting eigenvectors*. We show using Sturm sequences that, assuming the eigenvector contains no zeros, the number of sign changes in a shooting eigenvector of parameter λ is equal to the number of eigenvalues greater than λ. Finally, we use the techniques presented in the first section to experimentally demonstrate a $O(\log n)$ growth relationship between the variance of histogram bin values and the order of the β-Hermite matrix ensemble.

This paper is dedicated to the fond memory of James T. Albrecht

1. INTRODUCTION

A computational trick can also be a theoretical trick. The two go together probably more often than noticed. Perhaps the constraints of machines come from the same "fabric" that is woven into mathematics itself.

In this case we noticed a cute trick: we can histogram without histogramming. Mathematically, we get new formulas for level densities and largest eigenvalue distributions in terms of iterated integrals. Many numerical experiments in random matrix theory compute the eigenvalues of random matrices and then histogram. What else would one do other than `eig` of as many matrices as possible (perhaps computed from `randn`), followed by `hist`? The answer is that mathematics is kind, and one can compute histogram information without the numerical computation of eigenvalues. One remembers that the Sturm sequence computes the number of eigenvalues less than any "cut" (see Section 2 for elaboration). This paper proposes that this is not only a cute computational trick for efficiency, but a very important theoretical tool in the study of random matrix theory.

We begin the paper by exploring the use of Sturm sequences as a computational tool to efficiently compute histograms of eigenvalues for symmetric tridiagonal matrices. Sturm sequences are currently available in LAPACK when computing eigenvalues of symmetric tridiagonal matrices (via the DLAEBZ routine, which is called by the DSTEBZ routine). Other alternatives for the symmetric tridiagonal problem can also be found with routines starting with the prefix DST, though none explicitly compute a histogram of eigenvalues.

As mentioned in [4], a substantial computational savings to histogramming can be achieved via Sturm sequence methods. Since tridiagonal matrix models exist for certain classical random matrix ensembles [3], the techniques presented here can be applied to these ensembles. Using this method, we can compute a histogram of the eigenvalues of such a matrix in $O(mn)$ time where n is the dimension of the matrix and m is the number of bins (with arbitrary bin centers and widths) desired in the histogram. Using the naive approach of computing the eigenvalues and then histogramming them, computing the histogram would cost $O(n^2 + m)$ time. Our algorithm is a significant improvement because m is usually much smaller than n. For example, we reduced the time to compute a 100 bin histogram of the eigenvalues of a 2000×2000 matrix from 470 ms to 4.2 ms. This algorithm allows us to compute histograms that were computationally infeasible before, such as those for n equal to 1 billion. (As an aside, for those interested in the question of computing the largest eigenvalue of the β-Hermite ensemble quickly, we refer to one of the author's talks [6].)

In the second part of this paper, we describe a theoretical use of Sturm sequences, giving both the eigenvalue distribution (also called the *level density*) and the largest eigenvalue distribution of β-Hermite random matrix ensembles for arbitrary values of β. When normalized properly, the eigenvalue distribution converges to the well known Wigner semicircle [15], and the largest eigenvalue distribution to the Tracy-Widom distribution [13]. We derive analytic formulas *in terms of multivariate integrals* for any n and any β by analyzing the Sturm sequence of the tridiagonal matrix model. The formulas provided here are quite general and can also be generalized beyond the Hermite distribution.

Our main theoretical results are as follows. We show in Section 4.1 that the eigenvalue distribution of a matrix can be expressed as

$$\Pr[\Lambda < \lambda] = \frac{1}{n}\sum_{i=1}^{n}\Pr[r_{i,\lambda} < 0] = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{0} f_{r_i}(s)\ ds, \qquad \text{(see Theorem 4.1)}$$

where $r_{i,\lambda}$ is the ith element of the Sturm ratio sequence $(r_{1,\lambda}, r_{2,\lambda}, \ldots, r_{n,\lambda})$. When applied to the β-Hermite random matrix ensemble in Section 5.1, the marginal densities $f_{r_i}(s)$ are derived by iterated multiplication and integration of the conditional density

$$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{4}[2(s_i+\lambda)^2 - z_i^2]} D_{-p_i}(z_i), \qquad \text{(see Corollary 5.3)}$$

where D is the parabolic cylinder function, $p_i = \frac{1}{2}\beta(i-1)$, and $z_i = \text{sign}(s_{i-1})(s_i + \lambda + s_{i-1})$.

In Section 4.2, we show that the largest eigenvalue distribution of a matrix can be described in terms of the joint density $f_{r_1,r_2,\ldots,r_n}$ of the Sturm ratio sequence:

$$\Pr[\Lambda_{\max} < \lambda] = \int_{-\infty}^{0}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0} f_{r_1,r_2,\ldots,r_n}(s_1, s_2, \ldots, s_n)\ ds_1 ds_2 \ldots ds_n. \quad \text{(see Theorem 4.2)}$$

For the β-Hermite ensemble, the joint density $f_{r_1,r_2,\ldots,r_n}$ can be derived from the conditional density by iterated multiplication as above.

As a guide to the reader, the table below shows where the formal results may be found:

	General Case	Hermite Case (GOE, GUE, GSE, ...)
Level Density	Theorem 4.1	Corollary 5.3 (Corollary 5.7 contains an alternate formulation)
Largest Eigenvalue	Theorem 4.2	Corollary 5.4

This work generalizes a number of formulas, some very well known. The Hermite level density formula for $\beta = 2$ (and under the appropriate normalization) is

$$f_n(x) = \sum_{k=0}^{n-1} C_k e^{-x^2} H_k(x)^2$$

for constants C_k and Hermite polynomials $H_k(x)$ (see Mehta [10, eq. (6.2.10)]). Mehta also covers the $\beta = 1$ [10, eq. (7.2.32)] and $\beta = 4$ [10, second paragraph of page 178] cases. An iterated contour integral formula appears in Forrester and Desrosiers [2]. For even β, Forrester and Baker [1] present a formula using a Hermite polynomial of a matrix argument. In Dumitriu, Edelman, and Shuman [5, page 39], this density appears in the form (using $\alpha \equiv 2/\beta$):

$$f_n(x) = \frac{1}{\sqrt{2\pi}} (-1)^{n/\alpha} \frac{\Gamma(1 + \frac{1}{\alpha})}{\Gamma(1 + \frac{n}{\alpha})} e^{-x^2/2} H^{\alpha}_{[(2/\alpha)^{n-1}]}(xI_n).$$

Software is also provided in [5], and examples are given for the evaluation. Results for Laguerre and other densities are also known in some cases, including those found in the above cited references. There are differential equations that describe the largest eigenvalue statistic for finite n and $\beta = 1, 2, 4$ given in [13].

In the third section of this paper, we explore the relationship between the Sturm sequence of a tridiagonal random matrix A and its *shooting eigenvectors*. Shooting eigenvectors are those that result from fixing one value (say x_1) of a vector $x = (x_1, x_2, \ldots, x_n)$ and solving for the rest of its values under the equation $(A - \lambda I)x = 0$. We show using Sturm sequences that, assuming the eigenvector contains no zeros, the number of sign changes in the shooting eigenvector is equal the number of eigenvalues of A greater than λ. This connection was inspired by work by Jose Ramirez, Brian Rider, and Balint Virag [11], who proved an analogous result for stochastic differential operators.

In the final part of this paper, we use the histogramming technique presented in the first section to examine the variance of histogram bin values for eigenvalues of the β-Hermite random matrix ensemble. By leveraging our new algorithm, we were able to compute the sample variance for each of 100 histogram bins over 1024 trials for $n = 1$ up to 2^{20} for various values of β. We constructed plots of the mean variance as n increases to illustrate an $O(\log n)$ growth relationship.

As a general comment, some methods described in this paper may refer to symmetric tridiagonal matrices, but they can also be applied to nonsymmetric tridiagonal matrices via a diagonal similarity transform. Indeed, only the pairwise product of the super- and sub-diagonals, in addition to the diagonal itself, matters.

In Section 2, we review the definition of the Sturm sequence of a matrix and describe some of its properties. In Section 3, we describe an efficient algorithm for computing the histogram of eigenvalues of a symmetric tridiagonal matrix and give empirical performance results. In

Section 4, we describe how to derive both the eigenvalue distribution (level density) and the largest eigenvalue distribution in terms of Sturm ratio sequence elements for arbitrary symmetric tridiagonal matrices. Section 5 shows how to apply the results from Section 4 to derive the densities for the β-Hermite random matrix ensemble. Section 6 describes the connection between the sign changes in Sturm sequences and those in shooting eigenvectors, and Section 7 examines the mean variance of histogram bins for the β-Hermite ensemble.

2. Sturm Sequences

2.1. **Definition.** Define $(A_0, A_1, A_2, \ldots, A_n)$ to be the sequence of submatrices of an $n \times n$ matrix A anchored in the lower right corner of A. The Sturm sequence $(d_0, d_1, d_2, \ldots, d_n)_A$ is defined to be the sequence of determinants $(|A_0|, |A_1|, |A_2|, \ldots, |A_n|)$. In other words, d_i is the determinant of the $i \times i$ lower-right submatrix of A. We define d_0, the determinant of the empty matrix, to be 1.

$$
A = A_n = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix},
$$

$$
A_1 = [a_{nn}], \ A_2 = \begin{bmatrix} a_{n-1,n-1} & a_{n-1,n} \\ a_{n,n-1} & a_{n,n} \end{bmatrix}, \ A_3 = \begin{bmatrix} a_{n-2,n-2} & a_{n-2,n-1} & a_{n-2,n} \\ a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\ a_{n,n-2} & a_{n,n-1} & a_{n,n} \end{bmatrix}, \ \text{etc.}
$$

2.2. **Properties.**

2.2.1. *Counting Negative Eigenvalues.* The eigenvalues of principal submatrices of A interlace, yielding the following lemma:

Lemma 2.1. *The number of sign changes in the Sturm sequence $(d_0, d_1, d_2, \ldots, d_n)_A$ is equal to the number of negative eigenvalues of A.*

Proof. See [14, page 228] or [16, pages 103–104]. □

Some extra care has to be taken if zeros are present in the Sturm sequence. In some cases when a short sequence of zeros appears, it can be determined how to assign signs to the zeros such that Lemma 2.1 still holds. However, if enough zeros occur consecutively, the exact number of negative eigenvalues becomes impossible to determine from the Sturm sequence alone. Fortunately, we do not have to worry about zeros in the Sturm sequence for the purposes of this paper because, in the case of the β-Hermite ensemble as well most other random matrix ensembles of interest, the probability of any zeros occurring in the Sturm sequence is zero. Therefore, in the interest of simplicity we assume for the remainder of this paper that no zeros occur in the Sturm sequence.

2.2.2. *Sturm Ratio Sequence.* Since we are mainly interested in the relative sign of consecutive values in the Sturm sequence, we define the *Sturm ratio sequence* $(r_1, r_2, \ldots, r_n)_A$ to be the sequence of ratios of consecutive values in the original Sturm sequence. In other words,

$$
(1) \qquad\qquad r_i = d_i/d_{i-1} \qquad \forall\ i \in \{1, 2, \ldots, n\}.
$$

Lemma 2.2. *The number of negative values in $(r_1, r_2, \ldots, r_n)_A$ equals the number of negative eigenvalues of A.*

Proof. From our definition of the Sturm ratio sequence, the number of negative values in the sequence equals the number of sign changes in $(d_0, d_1, d_2, \ldots, d_n)_A$. From Lemma 2.1, this in turn is equal to the number of negative eigenvalues of A. $\square$

2.2.3. *Recurrence Relation.* Suppose we are given a symmetric tridiagonal matrix, with values $(a_n, a_{n-1}, \ldots, a_1)$ on the diagonal and $(b_{n-1}, b_{n-2}, \ldots, b_1)$ on the super/sub-diagonal (the reason for indexing them from bottom right to upper left will be explained in Section 2.3):

$$
\begin{bmatrix}
a_n & b_{n-1} & & & & \\
b_{n-1} & a_{n-1} & b_{n-2} & & & \\
& \ddots & \ddots & \ddots & & \\
& & b_2 & a_2 & b_1 \\
& & & b_1 & a_1
\end{bmatrix}
$$

Then, by expansion of minors, the terms in the Sturm sequence can be shown to satisfy the recurrence

$$
(2) \qquad d_i = \begin{cases} 1, & \text{if } i = 0; \\ a_1, & \text{if } i = 1; \\ a_i d_{i-1} - b_{i-1}^2 d_{i-2}, & \text{if } i \in \{2, 3, \ldots, n\}, \end{cases}
$$

or equivalently

$$
(3) \qquad r_i = \begin{cases} a_1, & \text{if } i = 1; \\ a_i - \frac{b_{i-1}^2}{r_{i-1}}, & \text{if } i \in \{2, 3, \ldots, n\}. \end{cases}
$$

2.3. **Example: β-Hermite Ensemble.** In the matrix model of the β-Hermite ensemble [3], submatrices anchored in the lower right corner represent smaller instances of the ensemble family. In this case, a subsequence of a larger Sturm sequence is a valid Sturm sequence of a smaller member of the same ensemble family. The β-Hermite matrix model is

$$
H_n^\beta \sim \frac{1}{\sqrt{2}} \begin{bmatrix}
\sqrt{2}G_n & \chi_{(n-1)\beta} & & & & \\
\chi_{(n-1)\beta} & \sqrt{2}G_{n-1} & \chi_{(n-2)\beta} & & & \\
& \ddots & \ddots & \ddots & & \\
& & \chi_{2\beta} & \sqrt{2}G_2 & \chi_\beta \\
& & & \chi_\beta & \sqrt{2}G_1
\end{bmatrix},
$$

where the G_i are Gaussian random variables with mean 0 and variance 1, and the χ are chi-distributed random variables (all are independent of each other), so we clearly see that the lower right square submatrices of H_n^β are just smaller instances of the same model. For this reason, we label the diagonal and subdiagonal elements in order from lower right to upper left.

3. Application to Eigenvalue Histogramming of Tridiagonal Matrices

Given a $n \times n$ tridiagonal matrix A, we can efficiently construct a histogram (given m sorted bins) of its eigenvalues in $O(mn)$ time using Lemma 2.2. Because n is usually much larger than m, this is a significant improvement over the naive approach, which involves first computing the eigenvalues themselves (taking $O(n^2)$ time [7]) and then placing them into bins (taking $O(n+m)$ time since the eigenvalues are presorted). The real-world improvement

5

is striking in cases where n is large: for example, when $n = 2000$ and $m = 100$, our algorithm is over 100 times faster than the naive approach in our empirical tests.

We now sketch our algorithm and its time complexity. Let the sequence $(k_1, k_2, \ldots, k_{m-1})$ be the sequence of separators between histogram bins. For convenience, define k_0 to be $-\infty$ and k_m to be ∞. Then the output is the histogram sequence $(H_1, H_2, \ldots, H_m)$, where H_i is the number of eigenvalues between k_{i-1} and k_i for $1 \leq i \leq m$.

If we let $\Lambda(M)$ be the number of negative eigenvalues of a matrix M, then the number of A's eigenvalues between k_1 and k_2 (where $k_1 < k_2$) equals $\Lambda(A - k_2 I) - \Lambda(A - k_1 I)$. Our histogramming algorithm first computes $\Lambda(A - k_i I)$ for each k_i. Using (3), we can compute the Sturm ratio sequence, counting negative values along the way, to yield $\Lambda(A - k_i I)$ in $O(n)$ time for each $A - k_i I$ (note that if the matrix is not symmetric, we can still use (3), replacing each b_{i-1}^2 with the product of the corresponding values on the super- and sub-diagonal). This step thus takes $O(mn)$ time in total. We then compute the histogram values:

$$
\begin{aligned}
H_1 &= \Lambda(A - k_1 I), \\
H_2 &= \Lambda(A - k_2 I) - \Lambda(A - k_1 I), \\
H_3 &= \Lambda(A - k_3 I) - \Lambda(A - k_2 I), \\
&\vdots \\
H_{m-1} &= \Lambda(A - k_{m-1} I) - \Lambda(A - k_{m-2} I), \\
H_m &= n - \Lambda(A - k_{m-1} I),
\end{aligned}
$$

in $O(m)$ time. The total running time of our algorithm is thus $O(mn)$.

In comparison, directly computing the eigenvalues takes $O(n^2)$ time using a standard LAPACK algorithm DSTEQR [7] for computing the eigenvalues of a symmetric tridiagonal matrix. Histogramming those values (they are returned in sorted order) then takes $O(n+m)$ time, yielding a total runtime of $O(m + n^2)$. Therefore, our algorithm is asymptotically superior for cases where $n > m$, which encompasses most practical situations.

Figures 1 and 2 show comparisons of the runtime of the two algorithms for the β-Hermite ensemble for $n = \{100, 200, \ldots, 1000\}$ and for $m = \{20, 40, \ldots, 100\}$. Computations were run using compiled C code (via MATLAB mex files) on a 2.4 GHz Intel Xeon Server with 2 GB of RAM. The times were taken by running 100 trials for each data point and computing the average time to complete for each trial.

From Figure 1, it is clear that the number of bins is of little relevance to the running time of the naive algorithm because the computation is completely dominated by the $O(n^2)$ time to compute the eigenvalues. Although our algorithm has a linear time dependence on the number of bins, that parameter does not usually scale with the problem size, so it is the linear dependence on n that leads to the drastic improvement over existing methods.

The real-world advantage of our algorithm is greater than the asymptotic runtimes might suggest because its simplicity yields a very small constant factor on current architectures. Figure 3 shows the remarkable speedup (defined as the number of times one algorithm is faster than the other) achieved for $n = \{100, 200, \ldots, 2000\}$ and $m = 100$.

4. Eigenvalue Distributions in Terms of the Sturm Ratio Sequence

In this section, we describe how the eigenvalue distribution and maximum eigenvalue distributions can be expressed in terms of the distribution of the Sturm ratio sequence.
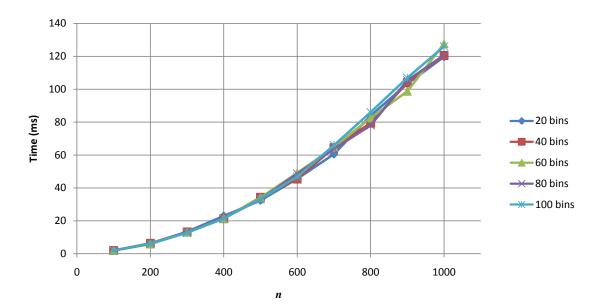
FIGURE 1. Performance of naive histogramming algorithm. This figure makes readily apparent the dominance of the eigenvalue computation (which takes $O(n^2)$ time): the number of histogram bins makes no significant difference in performance.
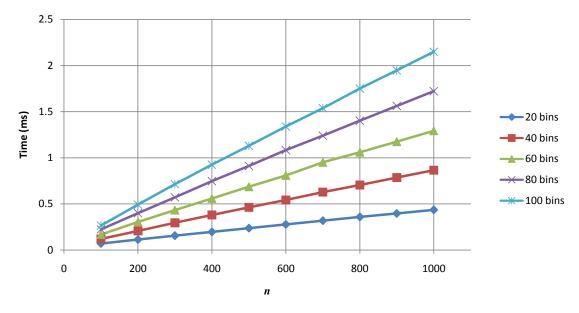

FIGURE 2. Performance of Sturm sequence based histogramming algorithm. This figure clearly demonstrates the bilinear dependency from the $O(mn)$ computation time.

4.1. **The Eigenvalue Distribution (or Level Density).** Given any matrix distribution D, its eigenvalue density $f(t)$ is equivalent to the distribution that would result from the following two step process:

- Draw a random $n \times n$ matrix A from our matrix distribution D.
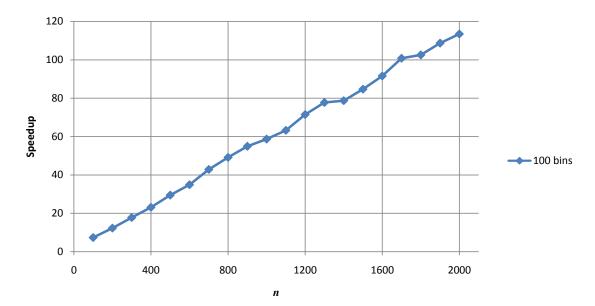
7

FIGURE 3. Speedup of the Sturm-based histogramming algorithm compared to the naive histogramming algorithm. Speedup is defined as the number of times our algorithm is faster than the naive approach. For large n, the speedup is quite remarkable.

- Uniformly draw an eigenvalue from all of A's eigenvalues.

Then, if random variable Λ follows the density $f(t)$, $\Pr[\Lambda < \lambda]$ is equal to the expected proportion of eigenvalues of $A - \lambda I$ that are negative.

Theorem 4.1. *If random variable Λ is drawn from the eigenvalue distribution of matrices following distribution D,*

$$(4) \qquad \Pr[\Lambda < \lambda] = \frac{1}{n} \sum_{i=1}^{n} \Pr[r_{i,\lambda} < 0] = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{0} f_{r_{i,\lambda}}(s) \ ds,$$

where $r_{i,\lambda}$ is the ith element of the Sturm ratio sequence $(r_{1,\lambda}, r_{2,\lambda}, \ldots, r_{n,\lambda})$ of the matrix $A - \lambda I$, where A is drawn from D. $f_{r_{i,\lambda}}(s)$ is the probability density function of $r_{i,\lambda}$.

Proof. We can express $f(t)$ as a sum of delta functions $\delta(t)$ whose locations $\lambda_i(A)$ are distributed as the eigenvalues of matrices A drawn from D:

$$f(t) = \int_{A \in D} \left[\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A)}(t) \right] \cdot P_D(A) \ dA = \mathrm{E} \left[\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A)}(t) \right].$$

We then have:

$$\Pr[\Lambda < \lambda] = \int_{-\infty}^{\lambda} f(t) \ dt = \mathrm{E} \left[\frac{1}{n} \sum_{i=1}^{n} \mathrm{I}[\lambda_i(A) < \lambda] \right],$$

where I is the indicator function. The quantity $\sum_{i=1}^{n} \mathrm{I}[\lambda_i(A) < \lambda]$ is just the number of eigenvalues of A less than λ, which we showed in Lemma 2.2 to be equal to number of

8

negative values in the Sturm ratio sequence of $A - \lambda I$. By linearity of expectation, we have

$$\Pr[\Lambda < \lambda] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}[r_{i,\lambda} < 0]\right] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\mathrm{I}[r_{i,\lambda} < 0]] = \frac{1}{n}\sum_{i=1}^{n} \Pr[r_{i,\lambda} < 0].$$

We can also express this quantity in terms of the marginal densities $f_{r_{i,\lambda}}(s)$ of the Sturm ratio sequence variables:

$$\Pr[\Lambda < \lambda] = \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{0} f_{r_{i,\lambda}}(s)\ ds.$$

$\square$

4.2. **The Largest Eigenvalue Distribution.** As shown in Lemma 2.2, the number of negative values in the Sturm ratio sequence $(r_1, r_2, \ldots, r_n)$ equals the number of A's negative eigenvalues. We can therefore express the probability that the largest eigenvalue of a matrix is negative simply as the probability that all terms in $(r_1, r_2, \ldots, r_n)$ are negative.

Theorem 4.2. *If random variable Λ_{max} is drawn from the largest eigenvalue distribution of matrices following distribution D,*

$$(5) \quad \Pr[\Lambda_{max} < \lambda] \;=\; \Pr[(r_{i,\lambda} < 0)\ \forall\ i \in \{1, 2, \ldots, n\}]$$

$$= \int_{-\infty}^{0}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0} f_{r_{1,\lambda},r_{2,\lambda},\ldots,r_{n,\lambda}}(s_1, s_2, \ldots, s_n)\ ds_1 ds_2 \ldots ds_n,$$

where $r_{i,\lambda}$ is the ith element of the Sturm ratio sequence $(r_{1,\lambda}, r_{2,\lambda}, \ldots, r_{n,\lambda})$ of the matrix $A - \lambda I$, where A is is drawn from D. $f_{r_{1,\lambda},r_{2,\lambda},\ldots,r_{n,\lambda}}$ refers to the joint density function of the Sturm ratio sequence $(r_{1,\lambda}, r_{2,\lambda}, \ldots, r_{n,\lambda})$.

Proof. From Lemma 2.2, the matrix $A - \lambda I$ has all negative eigenvalues exactly when its Sturm ratio sequence has all negative elements. Therefore, the probabilities of those events are identical. $\square$

Remark 4.3. Note we cannot break up $\Pr[(r_{i,\lambda} < 0)\ \forall\ i \in \{1, 2, \ldots, n\}]$ into the product $\prod_{i=1}^{n}\Pr[(r_{i,\lambda} < 0)]$ since the $r_{i,\lambda}$'s are not independent of one another.

5. Sturm Ratios of the β-Hermite Ensemble

In this section, we apply Theorems 4.1 and 4.2 to the β-Hermite random matrix ensemble using two different methods. In the first, we derive the conditional densities of Sturm ratios directly, yielding Corollaries 5.3 and 5.4. The second method utilizes *shifted Sturm ratios* (defined later), resulting in different but similar expressions for the distributions. The authors feel both derivations are illustrative and different enough to warrant inclusion in this paper.

5.1. **Conditional Densities of Sturm Ratios.** We begin by deriving an analytic formula in terms of an integral for the *conditional* density of Sturm ratios for the β-Hermite ensemble. The β-Hermite matrix model introduced in Section 2.3 and displayed again here has been

shown to have the same eigenvalue distribution as the β-Hermite ensemble [3].

$$H_n^\beta \sim \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2}G_n & \chi_{(n-1)\beta} & & & & \\ \chi_{(n-1)\beta} & \sqrt{2}G_{n-1} & \chi_{(n-2)\beta} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \chi_{2\beta} & \sqrt{2}G_2 & \chi_\beta \\ & & & \chi_\beta & \sqrt{2}G_1 \end{bmatrix}$$

Using recurrence (3), we can derive the following generative model for the Sturm ratio sequence of $H_n^\beta - \lambda I$:

$$(6) \qquad r_i = \begin{cases} G(-\lambda, 1), & \text{if } i = 1; \\ G(-\lambda, 1) - \frac{\chi^2_{\beta(i-1)}}{2r_{i-1}}, & \text{if } i \in \{2, 3, \ldots, n\}. \end{cases}$$

Note that in this section, we drop the λ subscript from the $r_{i,\lambda}$ variables used in previous sections to make the notation clearer. The λ parameter is implicit in the r_i notation.

In our derivation of the density of r_i, we make use of the following statistical property:

Lemma 5.1. *Let X, Y, Z, and W be random variables (X and Y independent) such that $Z = X + Y$ and $W = \frac{X}{k}$, where k is a constant. If f_X, f_Y, f_Z, and f_W are their respective probability densities, the following two identities hold:*

$$f_Z(s) = \int_{-\infty}^{\infty} f_X(s - t) f_Y(t) \, dt,$$

and

$$f_W(s) = |k| f_X(ks).$$

Proof. See Rice [12, pages 92–95]. □

Lemma 5.2. *For $i \geq 2$, the density of r_i conditioned on r_{i-1} is:*

$$(7) \qquad f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{4}[2(s_i+\lambda)^2 - z_i^2]} D_{-p_i}(z_i),$$

where D is the parabolic cylinder function, $p_i = \frac{\beta(i-1)}{2}$, and $z_i = \text{sign}(s_{i-1})(s_i + \lambda + s_{i-1})$.

Proof. For $i \geq 2$, if we let $f_G(t)$ be the density of $G(-\lambda, 1)$, $f_{\chi^2}(t)$ be the density of $\chi^2_{\beta(i-1)}$, and $f_{r_i|r_{i-1}}(s_i|s_{i-1})$ be the density of r_i given r_{i-1}, then we can combine (6) and Lemma 5.1 to yield

$$(8) \qquad f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \int_{-\infty}^{\infty} f_G(s_i - t) |2s_{i-1}| f_{\chi^2}(-2s_{i-1}t) \, dt.$$

Substituting the densities of the standard Gaussian random variable

$$f_G(s_i - t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(s_i - t + \lambda)^2}$$

and Chi-square random variable

$$f_{\chi^2}(-2s_{i-1}t) = \begin{cases} \frac{(-s_{i-1}t)^{p_i-1} e^{s_{i-1}t}}{2\Gamma(p_i)}, & \text{if } s_{i-1}t < 0; \\ 0, & \text{otherwise,} \end{cases}$$

10

where $p_i = \frac{1}{2}\beta(i-1)$, yields

$$(9) \quad f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \begin{cases} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[t-(s_i+\lambda)]^2} |2s_{i-1}| \frac{(-s_{i-1}t)^{p_i-1}e^{s_{i-1}t}}{2\Gamma(p_i)} \, dt, & \text{if } s_{i-1} < 0; \\ \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[t+(s_i+\lambda)]^2} |2s_{i-1}| \frac{(s_{i-1}t)^{p_i-1}e^{-s_{i-1}t}}{2\Gamma(p_i)} \, dt, & \text{otherwise.} \end{cases}$$

This can be simplified to

$$(10) \quad f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\Gamma(p_i)\sqrt{2\pi}} \int_0^\infty t^{p_i-1} e^{-\frac{1}{2}[t+sign(s_{i-1})(s_i+\lambda)]^2 - sign(s_{i-1})s_{i-1}t} \, dt$$

$$(11) \quad = \frac{|s_{i-1}|^{p_i} e^{-\frac{1}{2}(s_i+\lambda)^2}}{\Gamma(p_i)\sqrt{2\pi}} \int_0^\infty t^{p_i-1} e^{-\frac{1}{2}[t^2 + 2sign(s_{i-1})(s_i+\lambda+s_{i-1})t]} \, dt.$$

Using the following property of the parabolic cylinder function (whose properties are further discussed in the Appendix):

$$(12) \quad D_{-p}(z) = \frac{e^{-\frac{z^2}{4}}}{\Gamma(p)} \int_0^\infty t^{p-1} e^{-\frac{1}{2}(t^2+2zt)} \, dt, \quad \text{for } \mathrm{Re}(p) > 0,$$

by letting $z_i = sign(s_{i-1})(s_i + \lambda + s_{i-1})$, we can rewrite (11) as

$$\begin{aligned} f_{r_i|r_{i-1}}(s_i|s_{i-1}) &= \frac{|s_{i-1}|^{p_i} e^{-\frac{1}{2}(s_i+\lambda)^2}}{\Gamma(p_i)\sqrt{2\pi}} \cdot \frac{\Gamma(p_i)}{e^{-\frac{z_i^2}{4}}} D_{-p_i}(z_i) \\ &= \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{4}[2(s_i+\lambda)^2 - z_i^2]} D_{-p_i}(z_i), \end{aligned}$$

thus concluding our proof. $\qquad\square$

Figure 4 shows a comparison between our conditional density for r_2 (drawn using (7)) and a histogram of 10 million randomly generated conditional r_2 values (generated using (6)). The parameters used for the density function and histogram are: $\beta = 2$, $\lambda = 0$, and $r_1 = 1$.

By taking the product of conditional densities and integrating, we can derive the joint and marginal densities of the Sturm ratios. We can then build the eigenvalue density and largest eigenvalue density as discussed in Sections 4.1 and 4.2. Specifically,

$$(13) \quad f_{r_1,r_2,\ldots,r_n}(s_1, s_2, \ldots, s_n) = f_{r_1}(s_1) \prod_{i=2}^n f_{r_i|r_{i-1}}(s_i|s_{i-1}),$$

where $f_{r_1} = f_{G_{-\lambda}}$ is the Gaussian density with mean $-\lambda$ and variance 1, and

$$(14) \quad f_{r_i}(s_i) = \int_{-\infty}^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty f_{r_1,r_2,\ldots,r_i}(s_1, s_2, \ldots, s_i) \, ds_1 ds_2 \ldots ds_{i-1}.$$

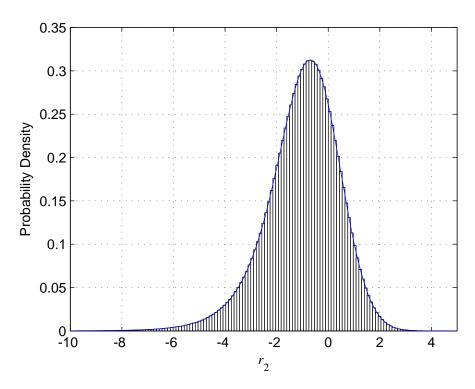These formulas can then be substituted into (4) and (5) as in the Corollaries below.

11

FIGURE 4. Analytic (blue line) and empirical (black histogram) conditional densities of r_2 given $\beta = 2$, $\lambda = 0$, and $r_1 = 1$.

Corollary 5.3. ("Eigenvalue distribution" or "Level density") *Let $f_{G_{-\lambda}}$ be the Gaussian density with mean $-\lambda$ and variance 1, and let $D_p(z)$ be the parabolic cylinder function. Define $z_i \equiv sign(s_{i-1})(s_i + \lambda + s_{i-1})$, and $p_i \equiv \frac{1}{2}\beta(i-1)$. If Λ is drawn from the β-Hermite eigenvalue density, then*

$$\Pr[\Lambda < \lambda] = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{0} f_{r_i}(s_i) \, ds_i,$$

where

$$f_{r_i}(s_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{r_1, r_2, \ldots, r_i}(s_1, s_2, \ldots, s_i) \, ds_1 ds_2 \ldots ds_{i-1},$$

$$f_{r_1, r_2, \ldots, r_n}(s_1, s_2, \ldots, s_n) = f_{G_{-\lambda}}(s_1) \prod_{i=2}^{n} f_{r_i|r_{i-1}}(s_i|s_{i-1}), \text{ and}$$

$$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{4}[2(s_i+\lambda)^2 - z_i^2]} D_{-p_i}(z_i).$$

Proof. Substitution of (7) into (13), (14), and (4). □

> **Corollary 5.4. (Largest eigenvalue distribution)** *As in Corollary 5.3, let $f_{G_{-\lambda}}$ be the Gaussian density with mean $-\lambda$ and variance 1, $D_p(z)$ be the parabolic cylinder function, $z_i \equiv sign(s_{i-1})(s_i + \lambda + s_{i-1})$, and $p_i \equiv \frac{1}{2}\beta(i-1)$. If Λ_{max} is drawn from the β-Hermite largest eigenvalue density, then*
>
> $$\Pr[\Lambda_{max} < \lambda] = \int_{-\infty}^{0} \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f_{r_1,r_2,\ldots,r_n}(s_1, s_2, \ldots, s_n) \; ds_1 ds_2 \ldots ds_n,$$
>
> *where*
>
> $$f_{r_1,r_2,\ldots,r_n}(s_1, s_2, \ldots, s_n) = f_{G_{-\lambda}}(s_1) \prod_{i=2}^{n} f_{r_i|r_{i-1}}(s_i|s_{i-1}), \quad and$$
>
> $$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{4}[2(s_i+\lambda)^2 - z_i^2]} D_{-p_i}(z_i).$$

Proof. Substitution of (7) into (13) and (5). □

5.2. Conditional Densities of Shifted Sturm Ratios.

We can also derive the eigenvalue distribution by describing the densities of a sequence of shifted Sturm ratios. We will define this sequence, derive its density, and show how they can also be used to derive the eigenvalue distribution. The structure of this section is very similar to that in the preceding section.

If we define the shifted Sturm ratio sequence $(x_1, x_2, \ldots, x_n)$ such that

$$x_i = a_i - r_i \qquad \forall i \in \{1, 2, \ldots, n\},$$

then, analogous to (3), we have the recurrence relation

$$x_i = \begin{cases} 0, & \text{if } i = 1; \\ \frac{b_{i-1}^2}{a_{i-1} - x_{i-1}}, & \text{if } i \in \{2, 3, \ldots, n\}, \end{cases}$$

and, analogous to (6), we have the generative model

$$(15) \qquad x_i = \begin{cases} 0, & \text{if } i = 1; \\ \frac{\chi_{\beta(i-1)}^2}{2G(-\lambda - x_{i-1}, 1)}, & \text{if } i \in \{2, 3, \ldots, n\}. \end{cases}$$

In our derivation of the conditional density of x_i given x_{i-1}, we make use of the following Lemma:

Lemma 5.5. *The density of a random variable distributed as $\frac{\chi_m^2}{G(n,4)}$ is*

$$J_{m,n}(t) \equiv \frac{m|t|^{p-2} e^{-\frac{1}{8}(n^2 - 2z^2)}}{2\sqrt{2\pi}} D_{-p}(z),$$

where D is the parabolic cylinder function (12), $p = \frac{m}{2} + 1$, and $z = sign(t)\left(t - \frac{n}{2}\right)$.

Proof. Let $X \sim \chi_m^2$, and $Y \sim G(n,4)$ independent of X. Then

$$\Pr\left[\frac{X}{Y} < t\right] = \int_{\frac{x}{y} < t} f_Y(y) \cdot f_X(x) \; dy dx,$$

13

where f_X and f_Y are the density functions of X and Y respectively. Making the change of variables $x = a$ and $y = \frac{a}{b}$ (with Jacobian $\frac{a}{b^2}$), we have:

$$
\begin{aligned}
\Pr\left[\frac{X}{Y} < t\right] &= \int_{b<t} f_Y\left(\frac{a}{b}\right) \cdot f_X(a)\left(\frac{a}{b^2}\ dbda\right) \\
&= \int_0^\infty \int_{-\infty}^t \frac{1}{2\sqrt{2\pi}}e^{-\frac{\left(\frac{a}{b}-n\right)^2}{8}} \cdot \frac{a^{\frac{m}{2}-1}e^{-\frac{a}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}}\left(\frac{a}{b^2}\ dbda\right).
\end{aligned}
$$

We can then take the derivative with respect to t to get the probability density function of $\frac{X}{Y}$:

$$
\begin{aligned}
J_{m,n}(t) &= \int_0^\infty \frac{1}{2\sqrt{2\pi}}e^{-\frac{\left(\frac{a}{t}-n\right)^2}{8}} \cdot \frac{a^{\frac{m}{2}-1}e^{-\frac{a}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \cdot \frac{a}{t^2}\ da. \\
(16) \qquad &= \frac{1}{2\sqrt{2\pi}\cdot\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \cdot \frac{1}{t^2}\int_0^\infty a^{\frac{m}{2}}e^{-\frac{1}{8}\left(\frac{a}{t}-n\right)^2-\frac{a}{2}}\ da.
\end{aligned}
$$

Noting the similarity between the integral in (16) and the one in the parabolic cylinder function (12), we make the substitution $a = 2|t|y$ to yield:

$$
\begin{aligned}
J_{m,n}(t) &= \frac{1}{2\sqrt{2\pi}\cdot\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \cdot \frac{1}{t^2}\int_0^\infty (2|t|y)^{\frac{m}{2}}e^{-\frac{1}{8}\left(\frac{2|t|y}{t}-n\right)^2-\frac{2|t|y}{2}}(2|t|\ dy) \\
&= \frac{|t|^{\frac{m}{2}-1}}{\sqrt{2\pi}\cdot\Gamma\left(\frac{m}{2}\right)}\int_0^\infty y^{\frac{m}{2}}e^{-\frac{1}{8}(4y^2-4sign(t)ny+n^2)-|t|y}\ dy \\
&= \frac{|t|^{\frac{m}{2}-1}e^{-\frac{n^2}{8}}}{\sqrt{2\pi}\cdot\Gamma\left(\frac{m}{2}\right)}\int_0^\infty y^{\frac{m}{2}}e^{-\frac{1}{2}\left(y^2+2sign(t)\left(t-\frac{n}{2}\right)y\right)}\ dy.
\end{aligned}
$$

Finally, letting $p = \frac{m}{2} + 1$ and $z = sign(t)\left(t - \frac{n}{2}\right)$, we have

$$
\begin{aligned}
J_{m,n}(t) &= \frac{|t|^{p-2}e^{-\frac{n^2}{8}+\frac{z^2}{4}}}{\sqrt{2\pi}} \cdot \frac{\Gamma\left(\frac{m}{2}+1\right)}{\Gamma\left(\frac{m}{2}\right)} \cdot D_{-p}(z) \\
&= \frac{m|t|^{p-2}e^{-\frac{1}{8}(n^2+2z^2)}}{2\sqrt{2\pi}} \cdot D_{-p}(z),
\end{aligned}
$$

thus, concluding our proof. □

Lemma 5.6. *For $i \geq 2$, the density of x_i conditioned on x_{i-1} is:*

$$
(17) \qquad\qquad f_{x_i|x_{i-1}}(y_i|y_{i-1}) = J_{\beta(i-1),-\lambda-y_{i-1}}(y_i).
$$

Proof. Follows directly from (15) and Lemma 5.5. □

We can derive the joint and marginal densities of shifted Sturm ratios using equations analogous to (13) and (14) to build both the eigenvalue and the largest eigenvalue densities (as in the following Corollaries).

14

Corollary 5.7. ("Eigenvalue distribution" or "Level density") *Let $f_{G_{-\lambda}}$ be the Gaussian density with mean $-\lambda$ and variance 1, δ be the Dirac delta function, and $D_p(z)$ be the parabolic cylinder function. Define $p \equiv \frac{m}{2} + 1$ and $z \equiv sign(t)\left(t - \frac{n}{2}\right)$. If Λ is drawn from the β-Hermite eigenvalue density, then*

$$\Pr[\Lambda < \lambda] = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{y_i} f_{G_{-\lambda}}(c_i)f_{x_i}(y_i)\ dc_i dy_i,$$

where

$$f_{x_i}(y_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{x_1,x_2,\dots,x_i}(y_1,y_2,\dots,y_i)\ dy_1 dy_2\dots dy_{i-1},$$

$$f_{x_1,x_2,\dots,x_n}(y_1,y_2,\dots,y_n) = \delta(y_1)\prod_{i=2}^{n} J_{\beta(i-1),-\lambda-y_{i-1}}(y_i),\quad and$$

$$J_{m,n}(t) = \frac{m|t|^{p-2}e^{-\frac{1}{8}(n^2-2z^2)}}{2\sqrt{2\pi}}D_{-p}(z).$$

Proof. In order to use Theorem 4.1, we first derive

$$
\begin{aligned}
\Pr[r_i < 0] = \Pr[a_i < x_i] \ &= \ \int_{-\infty}^{\infty}\int_{-\infty}^{y_i} f_{a_i,x_i}(c_i,y_i)\ dc_i dy_i,\\
&= \ \int_{-\infty}^{\infty}\int_{-\infty}^{y_i} f_{a_i}(c_i)f_{x_i}(y_i)\ dc_i dy_i,\\
&= \ \int_{-\infty}^{\infty}\int_{-\infty}^{y_i} f_{G_{-\lambda}}(c_i)f_{x_i}(y_i)\ dc_i dy_i.
\end{aligned}
$$

We can factor the joint density of a_i and x_i into their marginal distributions because the a_i and x_i are pairwise independent of one another even though x_i is dependent on a_{i-1} and x_{i-1}. Substituting into (4) yields

$$\Pr[\Lambda < \lambda] = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{y_i} f_{G_{-\lambda}}(c_i)f_{x_i}(y_i)\ dc_i dy_i.$$

We then use equations analogous to (13) and (14) with the substitutions $f_{x_1}(y_1) = \delta(y_1)$ and (17) to arrive at the result. $\square$

6. Connection to the Eigenfunction of a Diffusion Process

6.1. Motivation. Edelman and Sutton showed in [8] that the tridiagonal model of the β-Hermite ensemble, when taken to the continuous limit, can be expressed as a stochastic differential operator H. Ramírez, Rider, and Virag [11] then used this connection to show that the number of roots in an eigenfunction ψ of H is equal to the number of eigenvalues of H greater than the ψ's corresponding eigenvalue λ. In this section, we discretize the continuous quantities found in [11] and show that Theorem 1.2 in [11] may be viewed as a recasting of Sturm theory. Specifically, we show that Lemma 2.2 is the discrete analogue of this theorem.

As before, we ignore cases where zeros occur in the Sturm sequence or eigenvector since the probability of such an event is zero for random matrix ensembles of interest.

6.2. Eigenvector Ratios.

6.2. Eigenvector Ratios. Consider the $n \times n$ symmetric tridiagonal matrix A, having eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and eigenvectors $\vec{\mathbf{x_1}}, \vec{\mathbf{x_2}}, \ldots, \vec{\mathbf{x_n}}$. For each $k \in \{1, 2, \ldots, n\}$, define $T_{\lambda_k} = A - \lambda_k I$. Since $\vec{\mathbf{x_k}}$ is the eigenvector of A corresponding to λ_k, we have

$$(18) \qquad (A - \lambda_k I)\vec{\mathbf{x_k}} = T_{\lambda_k}\vec{\mathbf{x_k}} = 0.$$

Given T_{λ_k}, we may solve equation (18) to find the particular eigenvector $\vec{\mathbf{x_k}}$ of A corresponding to λ_k. Let $\vec{\mathbf{x_k}} = (x_n, x_{n-1}, \ldots, x_1)^T$. (Note the use of variables x_i here is different from those used in Section 5.2. In that section the x_i were used as shifted Sturm ratio values.) The reason for labeling the elements last-to-first is to align the labeling with the bottom-to-top labeling of our tridiagonal matrix model.

Since the scaling of the eigenvector is arbitrary, we may set x_1 to any value and then solve (18) for the remaining values. This yields

$$(19) \qquad x_i = \begin{cases} -\frac{1}{b_1}(a_1 x_1), & \text{if } i = 2; \\ -\frac{1}{b_{i-1}}(a_{i-1}x_{i-1} + b_{i-2}x_{i-2}), & \text{if } i \in \{3, 4, \ldots, n\}, \end{cases}$$

where the a_i and b_i denote the diagonal and super/sub-diagonal values of T_{λ_k} respectively (as in Section 2.2.3).

In general, we can solve equation (19) to derive $\vec{\mathbf{x}}$ for matrices $T_\lambda = A - \lambda I$, where λ is not necessarily an eigenvalue. If λ is indeed an eigenvalue of A, then $T_\lambda \vec{\mathbf{x}} = 0$. If not, there will be a residue ω present in the first element of $T_\lambda \vec{\mathbf{x}}$:

$$T_\lambda \vec{\mathbf{x}} = (\omega, 0, 0, \ldots, 0)^T.$$

$$\omega = a_n x_n + b_{n-1}x_{n-1}$$

Note that the process of setting x_1 to an arbitrary value and solving $T_\lambda \vec{\mathbf{x}} = 0$ for the rest of vector $\vec{\mathbf{x}}$ is similar to the shooting process used by Ramirez and Rider. We call the resulting $\vec{\mathbf{x}}$ a *shooting eigenvector*.

Define $x_{n+1} = -\omega$ and $b_n = 1$. If we let $s_i = x_i/x_{i-1}$ for $i \in \{2, 3, \ldots, n+1\}$, we have

$$(20) \qquad s_i = \begin{cases} -\frac{a_1}{b_1}, & \text{if } i = 2; \\ -\frac{1}{b_{i-1}}\left(a_{i-1} + \frac{b_{i-2}}{s_{i-1}}\right), & \text{if } i \in \{3, 4, \ldots, n+1\}. \end{cases}$$

(Note the use of variables s_i here is different from those used in Section 5.1. In that section the s_i were used as dummy variables.)

Theorem 6.1. *The sequence of ratios of eigenvector elements $S = (s_2, s_3, \ldots, s_{n+1})$ is related to the Sturm ratio sequence $R = (r_1, r_2, \ldots, r_n)$ by the following:*

$$(21) \qquad s_i = -\frac{r_{i-1}}{b_{i-1}} \quad \text{for } i \in \{2, 3, \ldots, n+1\}.$$

Proof. By induction. We have $s_2 = -\frac{a_1}{b_1} = -\frac{r_1}{b_1}$ from the definitions of s_2 and r_1. Now assume $s_j = -\frac{r_{j-1}}{b_{j-1}}$ for some $j \in \{2, 3, \ldots, n\}$. Using (20) and (3) we get:

$$s_{j+1} = -\frac{1}{b_j}\left(a_j + \frac{b_{j-1}}{s_j}\right) = -\frac{1}{b_j}\left(a_j - \frac{b_{j-1}^2}{r_{j-1}}\right) = -\frac{r_j}{b_j}.$$

$\square$

Since each of the elements b_{i-1} in (21) are positive with probability 1, $sign(s_i) = -sign(r_{i-1})$ for $i \in \{2, 3, \ldots, n+1\}$. Thus, the number of negative values in S equals the number of positive values in R. This in turn equals the number of positive eigenvalues of T_λ, or equivalently, the number of eigenvalues of A greater than λ. Since a negative value in S indicates a sign change in the underlying $\vec{x}$, we have shown the number of sign changes in $\vec{x}$ equals the number of eigenvalues of A that are greater than λ.

7. Analysis of the Variance of Histogram Bin Values for the β-Hermite Ensemble

In the following, we use the term *bin value* to denote the number of items (e.g. eigenvalues) that fall in a particular histogram bin.

Using the method described in Section 3, we conducted Monte Carlo simulations to examine the variance of bin values in a histogram of eigenvalues drawn from the β-Hermite ensemble. In a related work by Dumitriu and Edelman [4], the *smoothed fluctuation*, given by taking the inner product of the vector of bin differences (the difference between bin values and their expected values) and a smoothing function, was analytically shown to behave as a Gaussian random variable. Our experiment differs in that we are interested in the mean variance of the bin values, rather than a smoothed function of bin differences.

To be more precise, for each particular n and β, we generated p histograms

$$H^i = (h_1^i, h_2^i, \ldots, h_m^i)$$

of eigenvalues taken from the $n \times n$ scaled β-Hermite ensemble distribution, for $1 \leq i \leq p$, where h_j^i represents the number of eigenvalues that falls in the jth bin on the ith trial. We use the scaled distribution so that the eigenvalues fall in the interval $[-1, 1]$ for all values of n and β, though our results would be the same if we had used the unscaled ensembles and varied the histogram bin locations and widths. We computed the sample variance of the set $\{h_j^1, h_j^2, \ldots, h_j^p\}$ for each bin j, and finally computed their mean over all bins. By varying the order n of the matrices, we observed from our numerical experiments that the mean variance of the bin values grows asymptotically as $O(\log n)$.

The logarithm is perhaps predicted intuitively from Johansson [9, page 158] in that the Fourier coefficients of the box function do not quickly decrease, so the a_k in [9, eq. (2.12)] will very likely go as $1/k$. Then [9, eq. (2.11)] becomes a harmonic series, which makes us think of going to infinity as a logarithm. Future work must sort this out more carefully. We do not claim any specific result here other than the numerical observation.

Figures 5 and 6 show logarithmic plots of the mean sample variance (over $p = 1024$ trials) of histogram bin values for a $m = 100$ bin histogram as n varies from 1 to 2^{20} in powers of 2. The different series represent different values of β, which vary from 1/32 to 1024 in powers of two.

Two properties of Figures 5 and 6 are of interest. First, linear growth (manifest as an exponential curve on the log-plot) is evident for small values of n, which then transitions into logarithmic growth (manifest as a linear curve) for high values of n. The clearest instance of this transitional property is seen for $\beta = 8$ (the third curve from the top in Figure 6). Second, this transition occurs at smaller n for higher β values.

Both properties can be understood by thinking of β as a distance dependent repulsion factor. In the case of $\beta = 0$, the eigenvalues are independent identically distributed (i.i.d.)
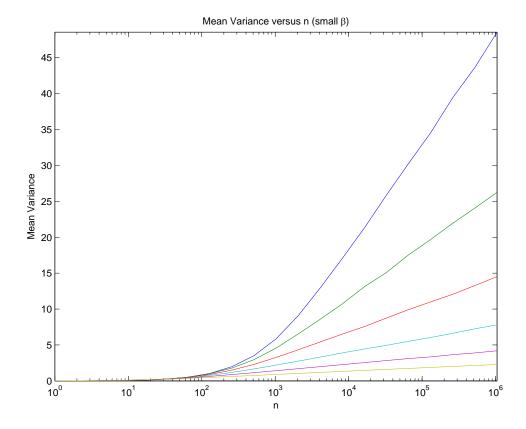
FIGURE 5. (**Mean Variance for Small** β) This figure illustrates that for large n, the variance appears to grow as $O(\log n)$. Presented is a logarithmic plot of the mean sample variance of histogram bin values for a 100 bin histogram as n varies from 1 to 2^{20} over 1024 trials. Different series are given for $\beta = 1/32$ to 1 in powers of two (β increases from top to bottom).

Gaussian random variables, so each histogram bin value is simply a binomial random variable with parameter p equal to some fractional area under the Gaussian PDF corresponding to the histogram bin's support. Thus, for the $\beta = 0$ case, the mean variance grows as $O(n)$. We see a departure from that behavior as β increases, and the eigenvalues more strongly repel one another. The correlation becomes more evident as n increases because the eigenvalues become more densly packed. For larger values of β (stronger repulsion), the transition from linear to logarithmic growth occurs at smaller n since the eigenvalues repel each other even at lower densities.

We conjecture the noisy behavior of the curves for higher β is due to the positions of clusters (illustrated in Figure 7) in the eigenvalue density function either falling mostly within a bin (contributing little to bin variance) or straddling bins (contributing more to bin variance). Since some values of n will have more clusters land on bin edges than others, this leads to the noisy character of the observed mean variance curves in Figure 6.
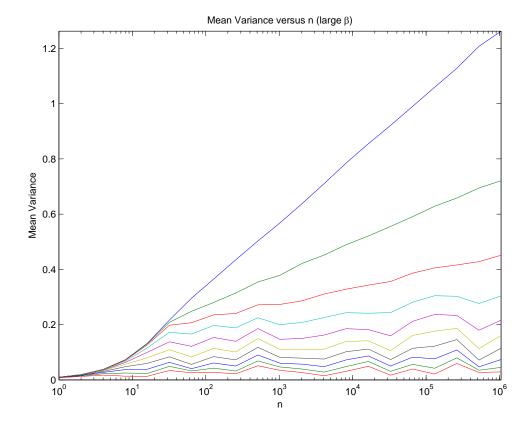
FIGURE 6. (**Mean Variance for Large** β) This figure illustrates that, for large n, the variance appears to grow as $O(\log n)$. Presented is a logarithmic plot of the mean sample variance of histogram bin values for a 100 bin histogram as n varies from 1 to 2^{20} over 1024 trials. Different series are given for $\beta = 2$ to 1024 in powers of two (β increases from top to bottom).

APPENDIX: PARABOLIC CYLINDER FUNCTIONS $D_p(z)$

In this appendix, we describe some basic properties of parabolic cylinder functions for readers unfamiliar with them. For our paper, we only need the following property:

$$D_{-p}(z) = \frac{e^{-\frac{z^2}{4}}}{\Gamma(p)} \int_0^\infty t^{p-1} e^{-\frac{1}{2}(t^2 + 2zt)} \, dt, \quad \text{for } \mathrm{Re}(p) > 0.$$

They also satisfy the interesting recurrences:

$$
\begin{aligned}
D_{p+1}(z) - z D_p(z) + p D_{p-1}(z) &= 0, \\
\frac{d}{dz} D_p(z) + \frac{z}{2} D_p(z) - p D_{p-1}(z) &= 0, \\
\frac{d}{dz} D_p(z) - \frac{z}{2} D_p(z) + D_{p+1}(z) &= 0.
\end{aligned}
$$

For positive integers n, we have:

$$D_n(z) = 2^{-\frac{n}{2}} e^{-\frac{z^2}{4}} H_n\left(\frac{z}{\sqrt{2}}\right),$$
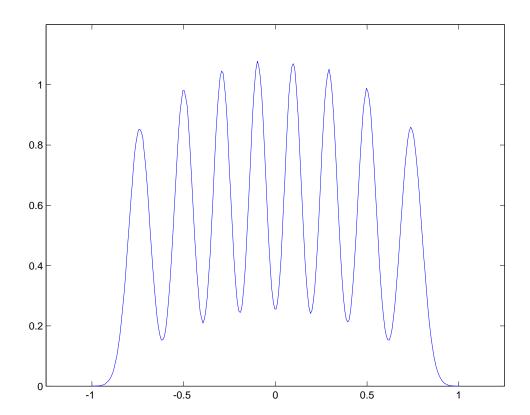
19

FIGURE 7. Probability density function of eigenvalues of the β-Hermite ensemble for $n = 8$ and $\beta = 8$. This density was generated using 2^{20} randomly generated histograms with $m = 256$ bins. Note the clustering behavior of the eigenvalues due to the repulsion effect of a large β value.

where H_n is a Hermite polynomial. This formula looks quite promising at first; however, in our analysis we are looking at D_{-p} with $p > 0$. In the cases where p is equal to 1, 2, and 3, we have:

$$
\begin{aligned}
D_{-1}(z) &= e^{\frac{z^2}{4}}\sqrt{2\pi}\Phi(-z), \\
D_{-2}(z) &= e^{-\frac{z^2}{4}} - ze^{\frac{z^2}{4}}\sqrt{2\pi}\Phi(-z), \\
D_{-3}(z) &= -\frac{z}{2}e^{-\frac{z^2}{4}} + \frac{1+z^2}{2}e^{\frac{z^2}{4}}\sqrt{2\pi}\Phi(-z),
\end{aligned}
$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$ is the cumulative distribution function of the standard normal density. Via the first recurrence above, one can easily verify that for positive integers p, $D_{-p}(z)$ is of the form

$$
D_{-p}(z) = \frac{e^{-\frac{z^2}{4}}}{\Gamma(p)}(A_p(z) + B_p(z)e^{\frac{z^2}{2}}\sqrt{2\pi}\Phi(-z)),
$$

where $A_1(z) = 0$, $B_1 = 1$, $A_2(z) = 1$, $B_2 = -z$, and both A and B satisfy

$$
\begin{aligned}
A_p(z) &= -zA_{p-1}(z) + (p-2)A_{p-2}(z), \\
B_p(z) &= -zB_{p-1}(z) + (p-2)B_{p-2}(z).
\end{aligned}
$$

For other values of p, $D_{-p}(z)$ can only be described in terms of Whittaker functions and confluent hypergeometric functions, thus a concise description has been elusive.

Acknowledgments

References

[1] T. Baker and P. J. Forrester. The Calogero-Sutherland model and generalized classical polynomials. *Communications in Mathematical Physics*, 188:175–216, 1997.

[2] P. Desrosiers and P. J. Forrester. Hermite and Laguerre β-ensembles: Asymptotic corrections to the eigenvalue density. *Nuclear Physics B*, 743(3):307–332, 2006.

[3] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847, 2002.

[4] I. Dumitriu and A. Edelman. Global spectrum fluctuations for the β-hermite and β-laguerre ensembles via matrix models. *Journal of Mathematical Physics*, 47(6), 2006.

[5] I. Dumitriu, A. Edelman, and G. Shuman. MOPS: Multivariate orthogonal polynomials (symbolically). arXiv:math-ph/0409066v1.

[6] A. Edelman. Stochastic differential equations and random matrices. `http://www-math.mit.edu/~edelman/homepage/talks/siam2003.ppt`.

[7] A. Edelman and N. R. Rao. Random matrix theory. *Acta Numerica*, 14:233–297, 2005.

[8] A. Edelman and B. Sutton. From random matrices to stochastic operators. arXiv:math-ph/0607038v2.

[9] K. Johansson. On fluctuations of eigenvalues of random hermitian matrices. *Duke Mathematical Journal*, 91(1):151–204, 1998.

[10] M. L. Mehta. *Random Matrices*. Elsevier, 1991.

[11] J. Ramírez, B. Rider, and B. Virag. Beta ensembles, stochastic airy spectrum, and a diffusion. arXiv:math/0607331v2.

[12] J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, second edition, 1995.

[13] C. A. Tracy and H. Widom. The distribution of the largest eigenvalue in the gaussian ensembles, in Calogero-Moser-Sutherland models. *CRM Series in Mathematical Physics*, 4:461–472, 2000.

[14] L. N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, 1997.

[15] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67:325–328, 1958.

[16] J. H. Wilkinson. *The algebraic eigenvalue problem*. Clarendon Press, 1965.

Department of Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.
  E-mail address: `cychan@mit.edu`

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.
  E-mail address: `edelman@math.mit.edu`