Sturm Sequences and the Eigenvalue Distribution of the Beta-Hermite Random Matrix Ensemble

by

Cy P. Chan

A.B. Applied Mathematics, Harvard University, 2002

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2007
[June 2007]

Author ....
Department of Electrical Engineering and Computer Science
May 23, 2007

Certified by ....
Alan Edelman
Professor of Applied Mathematics
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Sturm Sequences and the Eigenvalue Distribution of the Beta-Hermite Random Matrix Ensemble

by

Cy P. Chan

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2007, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

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. Our algorithm is a significant improvement because m is usually much smaller than n. This algorithm allows us to compute histograms that were computationally infeasible before, such as those for n equal to 1 billion.

Second, we give a derivation of the eigenvalue distribution for the β-Hermite random matrix ensemble (for general β). The novelty of the approach presented in this paper is in the use of Sturm sequences to derive the distribution. We derive an analytic formula *in terms of multivariate integrals* for the eigenvalue distribution and the largest eigenvalue distribution for general β by analyzing the Sturm sequence of the tridiagonal matrix model.

Finally, 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.

Thesis Supervisor: Alan Edelman
Title: Professor of Applied Mathematics

Acknowledgments

I owe a large debt of gratitude to James Albrecht for his help with this thesis especially with Section 5.2 and Appendix A. I also thank Brian Rider for his help regarding Chapter 6. Many thanks to Alan, my research advisor, for his continual cheerful guidance and insight. And, of course, I want to thank my family for their tremendous love and support. I especially thank my wife, Jessica, for putting up with me and my strange sleeping habits.

This work was supported by the National Science Foundation under grants DMS 0411962.

Contents

1 Introduction **9**

2 Sturm Sequences **11**
- 2.1 Definition 11
- 2.2 Properties 11
  - 2.2.1 Counting Negative Eigenvalues 11
  - 2.2.2 Sturm Ratio Sequence 12
  - 2.2.3 Recurrence Relation 12
- 2.3 Example: β-Hermite Ensemble 13

3 Application to Eigenvalue Histogramming of Symmetric Tridiagonal Matrices **15**

4 Eigenvalue Distributions in Terms of Sturm Sequence Ratios **19**
- 4.1 The Eigenvalue Distribution 19
- 4.2 The Largest Eigenvalue Distribution 20

5 The Densities of Sturm Sequence Ratios of the β-Hermite Ensemble **21**
- 5.1 The Sturm Ratio Sequence 21
- 5.2 The Shifted Sturm Ratio Sequence 24

6 Connection to the Eigenfunction of a Diffusion Process **27**
- 6.1 Motivation 27
- 6.2 Eigenvector Ratios 27

A Parabolic Cylinder Functions $D_p(z)$ **31**

Chapter 1

Introduction

The Sturm sequence of a matrix is defined as the sequence of determinants of the matrix's principal submatrices (see Chapter 2 for elaboration). It is well known that the eigenvalues of principal submatrices of a matrix interlace [1] [2], thus we can extract information about its eigenvalues by examining the Sturm sequence. 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.

First, we explore the use of Sturm sequences to efficiently compute histograms of eigenvalues for symmetric tridiagonal matrices. Since symmetric tridiagonal matrix models exist for certain classical random matrix ensembles [3], the techniques presented here can be used to analyze the eigenvalues of 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.

Second, we give a derivation of the eigenvalue distribution of β-Hermite random matrix ensembles (for general β) based on the use of Sturm sequences. Previous techniques used to derive the distribution (such as in [4]) have mainly focused on integrating the joint eigenvalue density

$$f_{\lambda_1,\lambda_2,\ldots,\lambda_n} = C \cdot e^{-\frac{1}{2}\beta\sum_{i=1}^{n}\lambda_i^2} \prod_{1\le i<j\le n} |\lambda_i - \lambda_j|^\beta,$$

where C is a normalizing constant. For even β, these distributions have been expressed in terms of multivariate Hermite polynomials and contour integrals [5]. We suspect the techniques in [5] can extend to general β and perhaps be related to the results given here.

The novelty of the approach presented in this paper is in the use of Sturm sequences to derive the distribution. We derive an analytic formula *in terms of multivariate integrals* for the eigenvalue distribution and the largest eigenvalue distribution for general β by analyzing the Sturm sequence of the tridiagonal matrix model (given in [3]).

Finally, we explore the relationship between the Sturm sequence of a random matrix 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 [6], who proved an analogous result for stochastic differential operators (the continuous version of a random matrix).

The rest of the paper is laid out as follows: In Chapter 2, we introduce the concept of the Sturm sequence of a matrix and describe some of its properties. In Chapter 3, we describe our new algorithm for computing the histogram of eigenvalues of a symmetric tridiagonal matrix and give empirical performance results. In Chapter 4, we describe how to derive both the eigenvalue distribution and the largest eigenvalue distribution in terms of the Sturm ratio sequence elements. Chapter 5 shows how to derive the densities of the Sturm sequence elements themselves, which, when combined with Chapter 4, yields the second result of the paper. Finally, Chapter 6 describes the connection between the sign changes in Sturm sequences and those in shooting eigenvectors.

Chapter 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.

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

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

2.2 Properties

2.2.1 Counting Negative Eigenvalues

The eigenvalues of principal submatrices of A interlace [1] [2], thus we have the following lemma:

Lemma 2.2.1. *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. Assume for the moment that no zeros occur in the Sturm sequence $(d_0, d_1, d_2, \ldots, d_n)_A$. Since the eigenvalues interlace, for every negative eigenvalue of A_{i-1}, we can pair it with the largest eigenvalue of A_i less than or equal to it. Similarly, for every positive eigenvalue of A_{i-1}, we can pair it with the smallest eigenvalue of A_i greater than or equal to it. Each eigenvalue pair shares the same sign. A_i has one additional eigenvalue greater than the largest negative eigenvalue and less than the smallest positive eigenvalue of A_{i-1}. If this eigenvalue is negative, then A_i has one more negative eigenvalue than A_{i-1}, thus $sign(d_i) = -sign(d_{i-1})$. If this eigenvalue is positive, then A_i has the same number of negative eigenvalues as A_{i-1}, thus $sign(d_i) = sign(d_{i-1})$. Therefore, as i increases, every time a sign change occurs in $(d_0, d_1, d_2, \ldots, d_n)_A$, the number of negative eigenvalues in A_i increases by one. The number of sign changes in the resulting Sturm sequence equals the number of negative eigenvalues present in A.

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 2.2.1 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,

$$r_i = d_i / d_{i-1} \qquad \forall\, i \in \{1, 2, \ldots, n\}. \tag{2.1}$$

Lemma 2.2.2. *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 2.2.1, this in turn is equal to the number of negative eigenvalues of A.

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

$$d_i = \begin{cases} 1, & \text{if } i = 0; \\ a_1, & \text{if } i = 1; \\ a_i d_{i-1} - b_{i-1}^2 d_{i-2}, & \text{if } i \in \{2, 3, \dots, n\}, \end{cases} \tag{2.2}$$

or equivalently

$$r_i = \begin{cases} a_1, & \text{if } i = 1; \\ a_i - \frac{b_{i-1}^2}{r_{i-1}}, & \text{if } i \in \{2, 3, \dots, n\}. \end{cases} \tag{2.3}$$

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 standard normal random variables, and the χ are chi-distributed random variables (all are independent of each other), so we clearly see that the lower right 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.

Chapter 3

Application to Eigenvalue Histogramming of Symmetric Tridiagonal Matrices

Given a $n \times n$ symmetric tridiagonal matrix A, we can efficiently construct a histogram (given m sorted bins) of its eigenvalues in $O(mn)$ time using Lemma 2.2.2. 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 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 (2.3), 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 $A - k_i I$. 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 $O(mn)$.

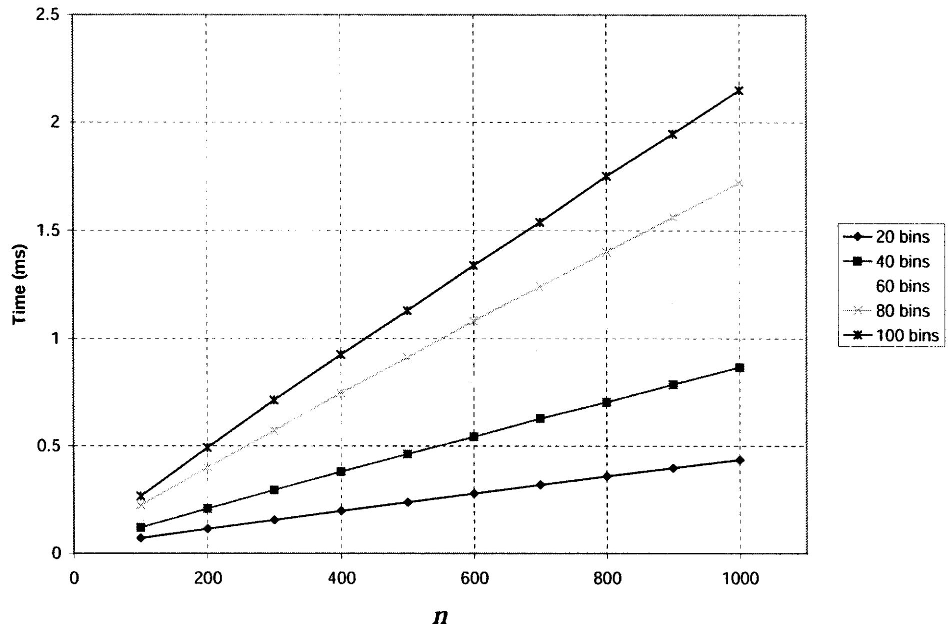

Figure 3-1: Performance of Sturm sequence-based histogramming algorithm.

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 3-1 and 3-2 show comparisons of the runtime of the two algorithms for the β-Hermite ensemle 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 averaging the results.

From Figure 3-2, 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 3-3 shows a comparison of the two algorithms for $n = \{100, 200, \ldots, 2000\}$ and $m = 100$.

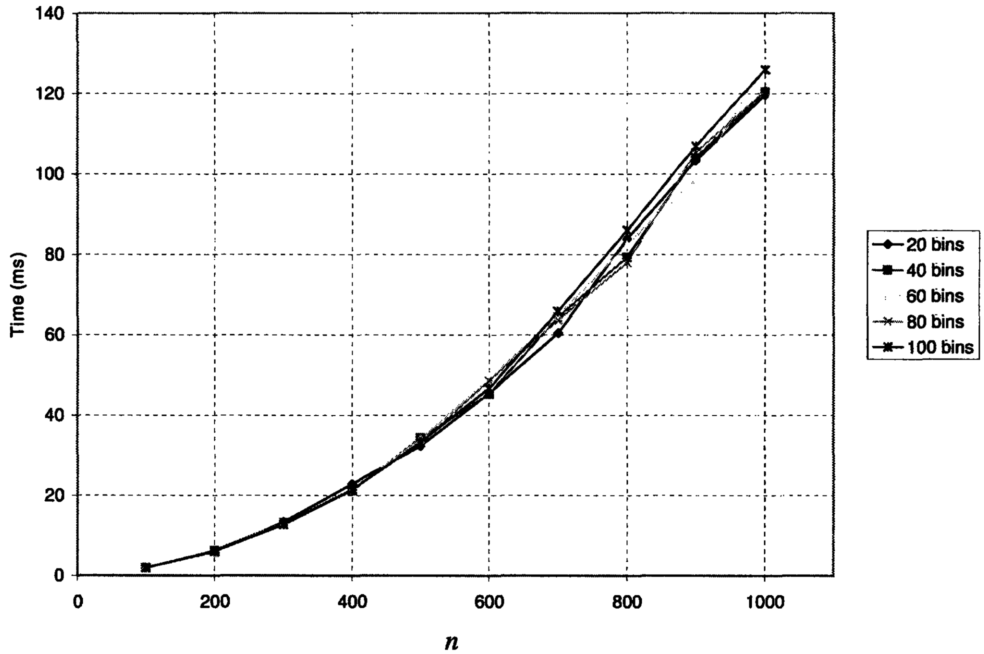

Figure 3-2: Performance of naive histogramming algorithm.

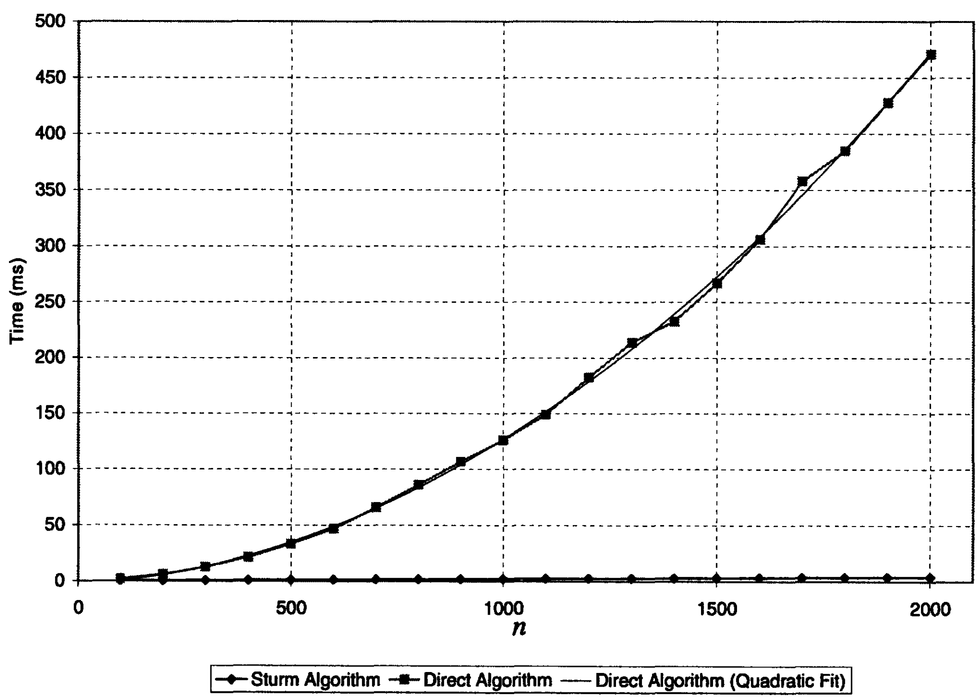

Figure 3-3: Comparison of performances of the Sturm sequence-based and naive histogramming algorithms. The number of bins used was $m = 100$.

Chapter 4

Eigenvalue Distributions in Terms of Sturm Sequence Ratios

4.1 The Eigenvalue Distribution

Given any matrix distribution D, the eigenvalue density $f(t)$ represents the distribution that would result from the following two-step process:

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

2. 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.

Lemma 4.1.1. *If random variable Λ is drawn from the eigenvalue distribution of matrices following distribution D,*

$$\Pr[\Lambda < \lambda] = \frac{1}{n}\sum_{i=1}^{n} \Pr[r_{i,\lambda} < 0],$$

where $r_{i,\lambda}$ is the i^{th} 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.

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

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

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

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

4.2 The Largest Eigenvalue Distribution

As shown in Lemma 2.2.2, 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.

Lemma 4.2.1. *If random variable Λ is drawn from the largest eigenvalue distribution of matrices following distribution D,*

$$\Pr[\Lambda < \lambda] = \Pr[(r_{i,\lambda} < 0)\ \forall\, i \in \{1, 2, \ldots, n\}],$$

where $r_{i,\lambda}$ is the i^{th} element of the Sturm ratio sequence $(r_1, r_2, \ldots, r_n)$ of the matrix $A - \lambda I$, where A is is drawn from D.

Proof. From Lemma 2.2.2, 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.

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 each other. We can, however, express this quantity in terms of the joint density $f_\lambda(s_1, s_2, \ldots, s_n)$ of the Sturm ratio sequence:

$$\Pr[\Lambda < \lambda] = \int_{-\infty}^{0}\int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f_\lambda(s_1, s_2, \ldots, s_n)\, ds_1 ds_2 \ldots ds_n. \tag{4.2}$$

Chapter 5

The Densities of Sturm Sequence Ratios of the β-Hermite Ensemble

5.1 The Sturm Ratio Sequence

We now derive an analytic formula in terms of an integral for the conditional density of Sturm sequence ratios for the β-Hermite ensemble. The β-Hermite matrix model introduced in Section 2.3 and displayed again here has been shown [3] to have the same eigenvalue distribution as the β-Hermite ensemble.

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

Using recurrence (2.3), we can derive the following generative model for the Sturm ratio sequence of $H_n^\beta - \lambda I$:

$$r_i = \begin{cases} G(-\lambda, 1), & \text{if } i = 1; \\ G(-\lambda, 1) - \frac{\chi^2_{\beta(i-1)}}{2r_{i-1}}, & \text{if } i \in \{2, 3, \ldots, n\}. \end{cases} \tag{5.1}$$

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 Lemma:

Lemma 5.1.1. *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 [8].

Theorem 5.1.2. *For $i \geq 2$, the density of r_i conditioned on r_{i-1} is:*

$$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \frac{|s_{i-1}|^{p_i}}{\sqrt{2\pi}} e^{-\frac{1}{4}[2(s_i+\lambda)^2 - z_i^2]} D_{-p_i}(z_i), \tag{5.2}$$

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

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 (5.1) and Lemma 5.1.1 to yield

$$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \int_{-\infty}^{\infty} f_G(s_i - t) |2s_{i-1}| f_{\chi^2}(-2s_{i-1}t)\, dt. \tag{5.3}$$

Substituting the densities of the 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

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

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

$$f_{r_i|r_{i-1}}(s_i|s_{i-1}) = \begin{cases} \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[t-(s_i+\lambda)]^2} |2s_{i-1}| \frac{(-s_{i-1}t)^{p_i-1} e^{s_{i-1}t}}{2\Gamma(p_i)}\, dt, & \text{if } s_{i-1} < 0; \\ \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[t+(s_i+\lambda)]^2} |2s_{i-1}| \frac{(s_{i-1}t)^{p_i-1} e^{-s_{i-1}t}}{2\Gamma(p_i)}\, dt, & \text{otherwise.} \end{cases} \tag{5.4}$$

This can be simplified to

$$\begin{aligned} f_{r_i|r_{i-1}}(s_i|s_{i-1}) &= \frac{|s_{i-1}|^{p_i}}{\Gamma(p_i)\sqrt{2\pi}} \int_0^{\infty} t^{p_i-1} e^{-\frac{1}{2}[t + sign(s_{i-1})(s_i+\lambda)]^2 - sign(s_{i-1}) s_{i-1} t}\, dt & (5.5) \\ &= \frac{|s_{i-1}|^{p_i} e^{-\frac{1}{2}(s_i+\lambda)^2}}{\Gamma(p_i)\sqrt{2\pi}} \int_0^{\infty} t^{p_i-1} e^{-\frac{1}{2}[t^2 + 2 sign(s_{i-1})(s_i+\lambda+s_{i-1})t]}\, dt. & (5.6) \end{aligned}$$

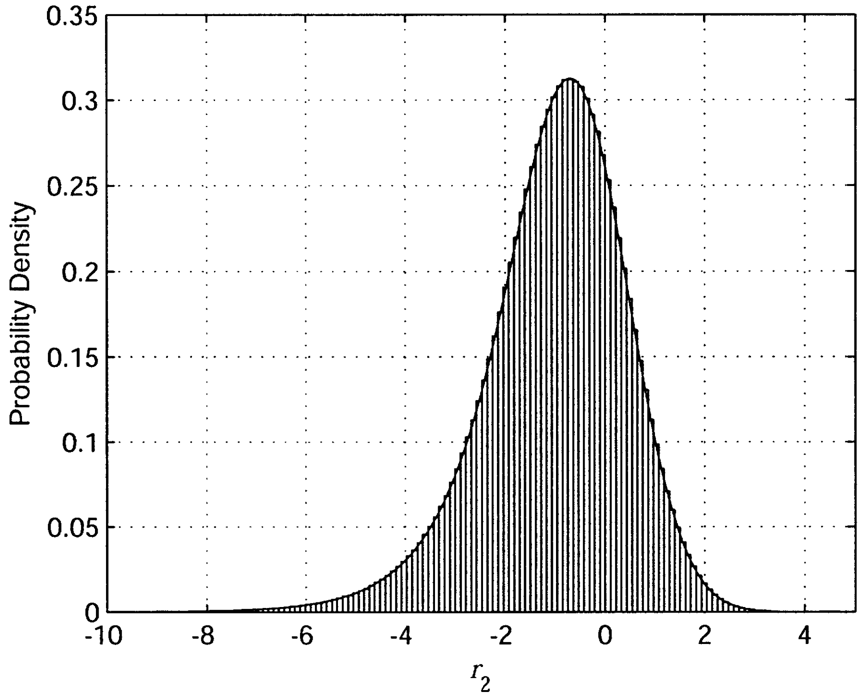

Figure 5-1: Analytic (blue line) and empirical (black histogram) conditional densities of r_2 given $\beta = 2$, $\lambda = 0$, and $r_1 = 1$.

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

$$D_{-p}(z) = \frac{e^{-\frac{z^2}{4}}}{\Gamma(p)} \int_0^\infty t^{p-1} e^{-\frac{1}{2}(t^2+2zt)}\, dt, \quad \text{for } \mathrm{Re}(p) > 0, \tag{5.7}$$

by letting $z_i = sign(s_{i-1})(s_i + \lambda + s_{i-1})$, we can rewrite (5.6) 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.

Figure 5-1 shows a comparison between our conditional density for r_2 (derived from (5.2)) and a histogram of 10 million randomly generated conditional r_2 values (generated using (5.1)). 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, we can derive the joint and marginal densities of Sturm sequence elements. We can then build the eigenvalue density and largest eigenvalue density as discussed in Sections 4.1 and 4.2. Specifically,

$$f_{r_1,r_2,\ldots,r_n}(s_1, s_2, \ldots, s_n) = f_{r_1}(s_1) \prod_{i=2}^{n} f_{r_i|r_{i-1}}(s_i|s_{i-1}), \tag{5.8}$$

where $f_{r_1} = f_G$, and

$$f_{r_i}(s_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{r_1,r_2,\ldots,r_n}(s_1, s_2, \ldots, s_i)\; ds_1 ds_2 \ldots ds_{i-1}. \tag{5.9}$$

This can be substituted directly into (4.1) and (4.2) to yield the desired eigenvalue distribution formulas. Thus, the densities of the Sturm ratio random variables are expressed as multivariate iterated integrals, and the eigenvalue density is expressed as a sum of iterated integrals.

5.2 The Shifted Sturm Ratio Sequence

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 derivations given in this and the previous section are equally valid, though different enough to merit inclusion of both in this paper. 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 (2.3), 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 (5.1), we have the generative model

$$x_i = \begin{cases} 0, & \text{if } i = 1; \\ \frac{\chi^2_{\beta(i-1)}}{2G(-\lambda - x_{i-1}, 1)}, & \text{if } i \in \{2, 3, \ldots, n\}. \end{cases} \tag{5.10}$$

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

Lemma 5.2.1. *The density of a random variable distributed as* $\frac{\chi^2_m}{G(n,4)}$ *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 (5.7),* $p = \frac{m}{2}+1$, *and* $z = sign(t)\left(t - \frac{n}{2}\right)$.

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

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

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 $\frac{X}{Y}$:

$$J_{m,n}(t) = \int_0^\infty \frac{1}{2\sqrt{2\pi}} e^{-\frac{\left(\frac{a}{t}-n\right)^2}{8}} \cdot \frac{a^{\frac{m}{2}-1}e^{-\frac{a}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \cdot \frac{a}{t^2}\, da. \tag{5.11}$$

$$= \frac{1}{2\sqrt{2\pi}\cdot\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \cdot \frac{1}{t^2}\int_0^\infty a^{\frac{m}{2}} e^{-\frac{1}{8}\left(\frac{a}{t}-n\right)^2 - \frac{a}{2}}\, da. \tag{5.12}$$

Noting the similarity between the integral in (5.12) and the one in the parabolic cylinder function (5.7), 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

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

thus, concluding our proof.

Theorem 5.2.2. *For $i \geq 2$, the density of x_i conditioned on x_{i-1} is:*

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

Proof. Follows directly from (5.10) and Lemma 5.2.1.

We can then write the joint density of $(x_1, x_2, \ldots, x_k)$ as

$$f_{x_1,x_2,\ldots,x_n}(y_1, y_2, \ldots, y_n) = f_{x_1}(y_1) \prod_{i=2}^{n} f_{x_i|x_{i-1}}(y_i|y_{i-1}),$$

and the marginal density of x_i as

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

To write the eigenvalue distribution (Section 4.1), we compute

$$\Pr[r_i < 0] = \Pr[a_i < x_i] = \int_{-\infty}^{\infty} \int_{-\infty}^{y_i} f_G(t) f_{x_i}(y_i)\ dt dy_i,$$

where f_G is the probability density function for $a_i \sim G(-\lambda, 1)$, which yields

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

Chapter 6

Connection to the Eigenfunction of a Diffusion Process

6.1 Motivation

Edelman and Sutton showed in [9] 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 [6] 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 [6] and show that Theorem 1.2 in [6] may be viewed as a recasting of Sturm theory. Specifically, we show that Lemma 2.2.2 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

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

$$(A - \lambda_k I)\vec{\mathbf{x_k}} = T_{\lambda_k}\vec{\mathbf{x_k}} = 0. \tag{6.1}$$

As shown by Lemma 2.2.2, the number of eigenvalues of A less than λ_k equals the number of sign changes in the Sturm ratio sequence $(r_1, r_2, \ldots, r_n)_{T_{\lambda_k}}$. Given T_{λ_k}, we may solve equation (6.1) 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 (6.1) for the remaining values. This yields

$$x_i = \begin{cases} -\frac{1}{b_1}(a_1 x_1), & \text{if } i = 2; \\ -\frac{1}{b_{i-1}}(a_{i-1}x_{i-1} + b_{i-2}x_{i-2}), & \text{if } i \in \{3, 4, \ldots, n\}, \end{cases} \tag{6.2}$$

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 (6.2) 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

$$s_i = \begin{cases} -\frac{a_1}{b_1}, & \text{if } i = 2; \\ -\frac{1}{b_{i-1}}\left(a_{i-1} + \frac{b_{i-2}}{s_{i-1}}\right), & \text{if } i \in \{3, 4, \ldots, n+1\}. \end{cases} \tag{6.3}$$

(Note the use of variables s_i here is different from those used in Chapter 5. In that chapter the s_i were used as dummy variables.)

Theorem 6.2.1. *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:*

$$s_i = -\frac{r_{i-1}}{b_{i-1}} \quad \textit{for } i \in \{2, 3, \ldots, n+1\}. \tag{6.4}$$

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 (6.3) and (2.3) we get:

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

This completes the induction.

Since each of the elements b_{i-1} in (6.4) 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 $\vec{\mathbf{x}}$, we have shown the number of sign changes in $\vec{\mathbf{x}}$ equals the number of eigenvalues of A that are greater than λ.

Appendix A

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:

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

For positive integers n, we have:

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

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:

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

Via the first recurrence above, one can easily verify that for positive integers p, $D_{-p}(z)$ is going to be of the form:

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

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.

Bibliography

[1] Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra*, pages 227–228. SIAM, 1997.

[2] James Hardy Wilkinson. *The algebraic eigenvalue problem*, pages 103–104. Clarendon Press, 1965.

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

[4] Madan Lal Mehta. *Random Matrices.* Elsevier, 1991.

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

[6] José Ramírez, Brian Rider, and Balint Virag. Beta ensembles, stochastic airy spectrum, and a diffusion. arXiv:math/0607331v2.

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

[8] John A. Rice. *Mathematical Statistics and Data Analysis*, pages 92–95. Duxbury Press, second edition, 1995.

[9] Alan Edelman and Brian Sutton. From random matrices to stochastic operators. arXiv:math-ph/0607038v2.