Visualizing and Quantifying Eigenvalues of β–Hermite ensemble in Julia

Rosen Yu

MIT

December 10, 2025

Outline

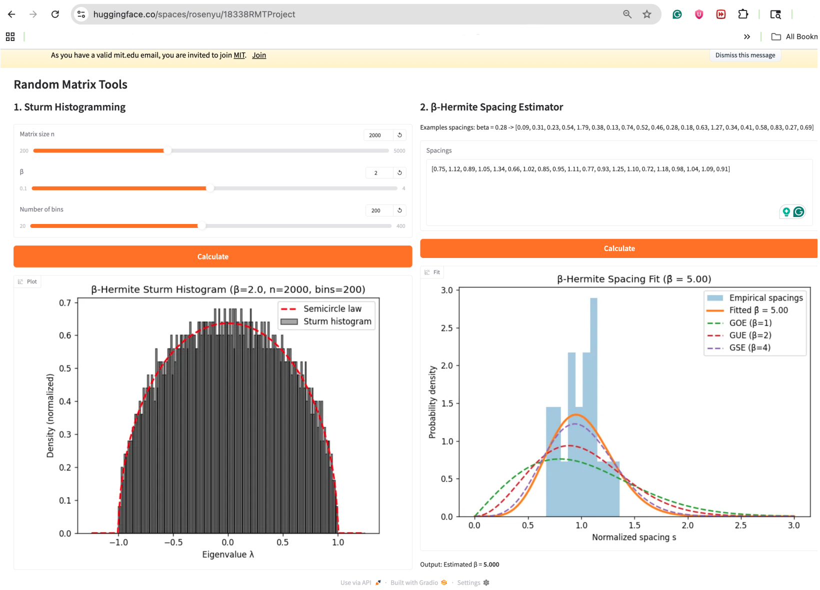
Website: Strum Histogramming + β Estimator

Quantifying “Repulsiveness”

References

Strum Histogramming + β Estimator

https://huggingface.co/spaces/rosenyu/18338RMTProject



Compute Bottleneck with Histogramming

To visualize the spectral density of the β-Hermite ensemble, we need the eigenvalues of the tridiagonal matrix H_β.

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

- **Naive Approach:**
  1. Generate Matrix.
  2. `eig(H)` $\rightarrow$ cost is $\mathcal{O}(n^2)$.
  3. Bin eigenvalues $\rightarrow$ cost is $\mathcal{O}(n)$.
- **The Problem:** For large n (e.g., $n = 1{,}000{,}000$), $\mathcal{O}(n^2)$ is computationally prohibitive for real-time apps.

Solution: Sturm Sequences [Albrecht et al., 2009]

1. Given a symmetric tridiagonal matrix

$$A = \begin{bmatrix} a_1 & b_1 & & \\ b_1 & a_2 & b_2 & \\ & \ddots & \ddots & \ddots \\ & & b_{n-1} & a_n \end{bmatrix},$$

2. Sturm sequence $(d_0, d_1, \ldots, d_n)$: the sequence of determinants of trailing principal minors

$$d_0 = 1, \qquad d_1 = a_1, \qquad d_i = a_i d_{i-1} - b_{i-1}^2 d_{i-2}.$$

3. Sturm ratio sequence: $r_i = \frac{d_i}{d_{i-1}}$

$$r_1 = a_1, \qquad r_i = a_i - \frac{b_{i-1}^2}{r_{i-1}}, \qquad i \geq 2.$$

From Determinants to Ratios

1. Let A_k be the top-left $k \times k$ submatrix.
2. The Sturm sequence is the sequence of determinants: $d_k = \det(A_k)$.
3. **Lemma:** the number of sign changes in the sequence $(1, d_1, d_2, \dots, d_n)$ equals the number of negative eigenvalues.
4. Computing determinants causes numerical overflow. Instead, we compute ratios: $r_k = \frac{d_k}{d_{k-1}}$
   - A sign change occurs between d_{k-1} and d_k if and only if they have opposite signs.
   - If signs are opposite, r_k must be negative.
   - $\therefore$ Count of negative eigenvalues = Count of negative r_k terms.

The "Shift" and The Histogram

For histogramming, we want eigenvalues less than bin edge k:

$\Rightarrow$ The eigenvalues of matrix $(A - kI)$ are exactly $(\lambda_i - k)$.If $\lambda_i < k$, then $(\lambda_i - k)$ is negative.

The Algorithm:

1. Compute Sturm ratios for shifted matrix: $r_i = (a_i - k) - b_{i-1}^2 / r_{i-1}$.
2. Count negative r_i's $\rightarrow$ gives count of eigenvalues $\lambda < k$.

The Bin Count: To find eigenvalues in bin $(k_{i-1}, k_i]$:

$$\text{Count}_{\text{bin}} = (\#\text{Neg. Ratios of } A - k_i I) - (\#\text{Neg. Ratios of } A - k_{i-1} I)$$

$$H_i(\text{Count}_{\text{bin}}) = \Lambda(k_i) - \Lambda(k_{i-1})$$

Sturm Histogramming Algorithm

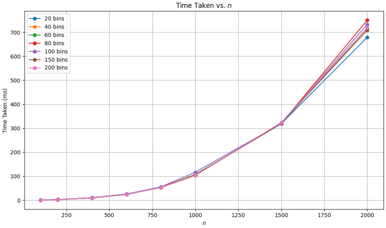
Total Complexity: $\mathcal{O}(mn)$

Algorithm 1 Sturm-Sequence Histogramming of a β–Hermite Matrix

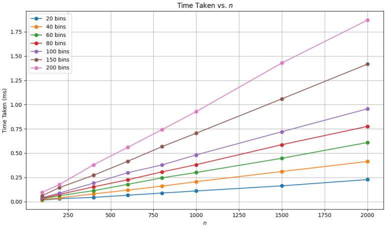
Require: matrix dimension n, histogram bin edges $k_0 < \cdots < k_m$,
β–Hermite diagonal a_i and off-diagonal b_i entries.

for each bin edge k_j **do**
    Initialize $r_1 = a_1 - k_j$;
    count $= 1\{r_1 < 0\}$;
    for $i = 2$ to n **do**
        $r_i = (a_i - k_j) - \frac{b_{i-1}^2}{r_{i-1}}$;
        count $+= 1\{r_i < 0\}$;
    end for
    Store $\Lambda_j =$ count
end for
Compute histogram bins
    $H_1 = \Lambda_1$,
    $H_i = \Lambda_i - \Lambda_{i-1}$,
    $H_m = n - \Lambda_{m-1}$.
return H

Algorihm Runtime Validation



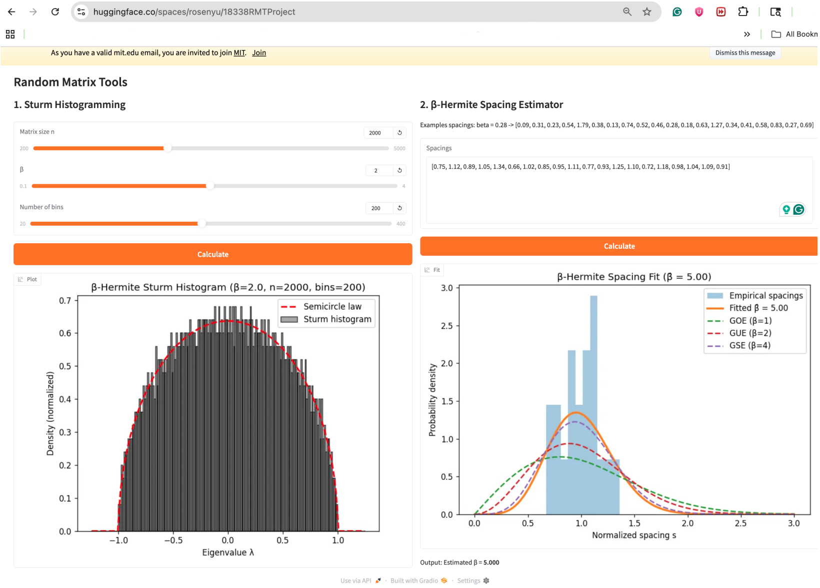
Naive `eig(H)` Histogramming
$O(n^2)$



Sturm Sequence Histogramming
$O(mn)$

Strum Histogramming + β Estimator

https://huggingface.co/spaces/rosenyu/18338RMTProject



β-Estimation via Wigner Surmise

1. Motivation: Beta Estimator (`https://people.csail.mit.edu/cychan/BetaEstimator.html`) is outdated
2. Goal: Given empirical spacings $\{s_i\}$ $((s_i = \lambda_{i+1} - \lambda_i))$ between adjacent eigenvalues, find the β of the β-Hermite ensemble

β-Estimation via Wigner Surmise

The Wigner Surmise: While the exact Mehta–Gaudin distribution is complex, the Wigner Surmise provides an approximation for the spacing distribution $P(s)$:

$$P_\beta(s) = a_\beta s^\beta e^{-b_\beta s^2}.$$

- $\beta = 1$: GOE, $\beta = 2$: GUE, $\beta = 4$: GSE

My Method: Given empirical spacings $\{s_i\}$, we solve for $\hat{\beta}$ via Maximum Likelihood Estimation (MLE):

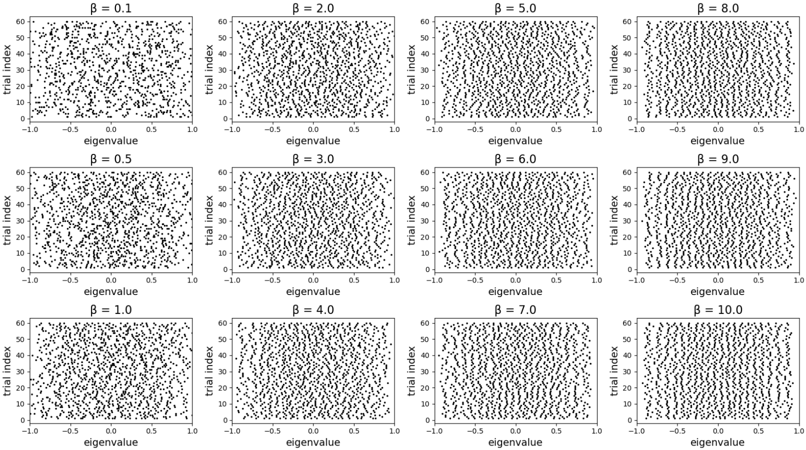
$$\hat{\beta} = \arg\max_{\beta>0} \sum_i \left(\ln(a_\beta) + \beta \ln(s_i) - b_\beta s_i^2\right).$$

Website Demo

https://huggingface.co/spaces/rosenyu/18338RMTProject

What is "Repulsiveness"?

Observation: As β increases, eigenvalues repel more strongly.



Literature that discussed quantifying DPP

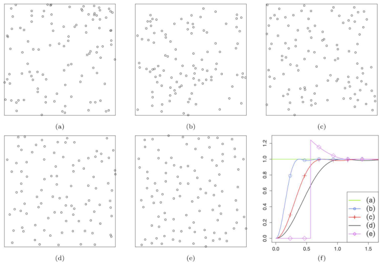


Figure 1. Realizations on $[-5,5]^2$ of (a) the stationary Poisson process, (b)–(d) DPPs with kernels (5.1) where $\sigma = 0$ and $\alpha = 0.2, 0.4, \frac{1}{\sqrt{\pi}}$, (e) the type II Matérn hardcore process with hardcore radius $\frac{1}{\sqrt{\pi}}$. (f) Their associated theoretical p.c.f.s. The intensity is $\rho = 1$ for all models and (d) represents the most repulsive stationary DPP in this case.

[Biscio and Lavancier, 2016]

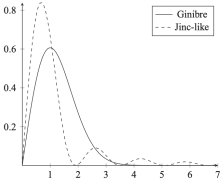


FIGURE 1: The densities of $|Z_0|$ for $Z_0 := X \setminus X^0$ for two globally most repulsive DPPs.

[Møller and O'Reilly, 2021]

Pair Correlation Function [Biscio and Lavancier, 2016]

Definition: $g(r)$ describes the probability of finding two particles separated by distance r, normalized by the average density $\rho 3$.

$$g(r) = \frac{\rho^{(2)}(x, x + r)}{\rho^2}$$

- $\rho^{(2)}$: The joint intensity (probability of particles at both x and $x + r$).
- ρ^2: The probability if the points were completely independent.

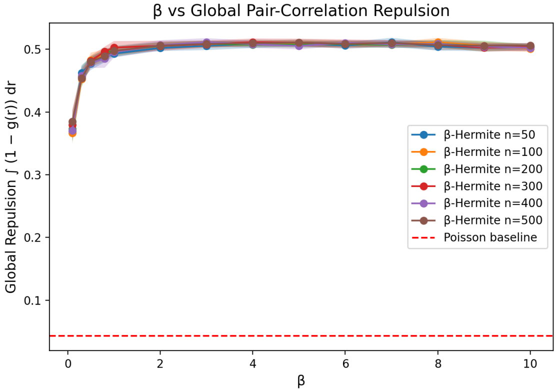
Statistical Meaning:

- $g(r) = 1$: The Poisson Baseline. Knowing a particle exists at x gives no information about $x + r$. Random placement.
- $g(r) < 1$: Repulsion. Finding a particle at x makes it less likely to find one at $x + r$.
- $g(r) > 1$: Attraction.

Metric 1: Global Repulsion

Definition: $R_{global} = \int_0^\infty (1 - g_\beta(r))dr$

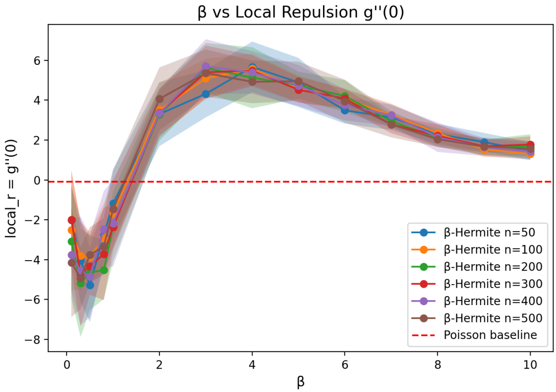
- Measures the total "area" of the repulsion hole.
- Increases sharply from $\beta = 0$ to $\beta = 1$. Saturates for $\beta \geq 1$.
- *Interpretation*: This confirms the Sum Rule. The system is hyperuniform; a particle creates a total deficit of exactly one particle around it, regardless of how high β gets.



Metric 2: Local Repulsion

Definition: $R_{local} = g''_\beta(0)$ (Curvature at origin)

- Measures how flat the repulsion hole is near $r = 0$.
- Peaks around $\beta \approx 3$.
- *Interpretation*: For high β, the repulsion becomes "hard-core" (flat bottom basin). The curvature actually decreases as the hole widens and flattens (r^β becomes flat at 0 for large β).

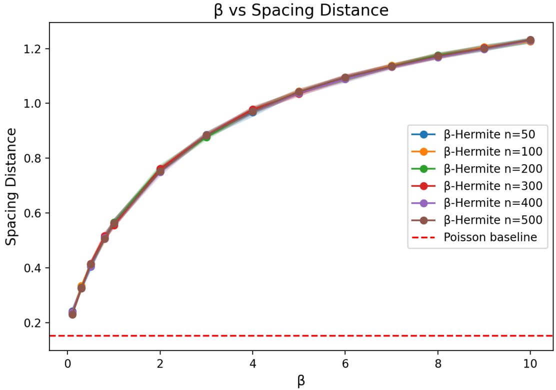


Metric 3: Spacing Distance

Definition: L_1 distance between the empirical spacing distribution and the Poisson exponential distribution (e^{-s}).

$$\mathcal{D}_{spacing} = \int_0^\infty |p_\beta(s) - e^{-s}| ds$$

- Monotonic increase with β.



References I

Albrecht, J. T., Chan, C. P., and Edelman, A. (2009).
Sturm sequences and random eigenvalue distributions.
Foundations of Computational Mathematics, 9(4):461–483.

Biscio, C. A. N. and Lavancier, F. (2016).
Quantifying repulsiveness of determinantal point processes.

Møller, J. and O'Reilly, E. (2021).
Couplings for determinantal point processes and their reduced palm distributions with a view to quantifying repulsiveness.
Journal of Applied Probability, 58(2):469–483.