Numerical Computation of the Tracy-Widom Distribution

Zewen Shen, advised by Prof. Kirill Serkh

University of Toronto

July, 2021

Introduction

What is the Tracy-Widom distribution?

- Describes the limiting distribution of the scaled largest eigenvalue of a random matrix belonging to Gaussian ensembles
- A new universal law...
- Appears in various different disciplines (physics, high-dimensional statistics, wireless communications, finance, etc.)

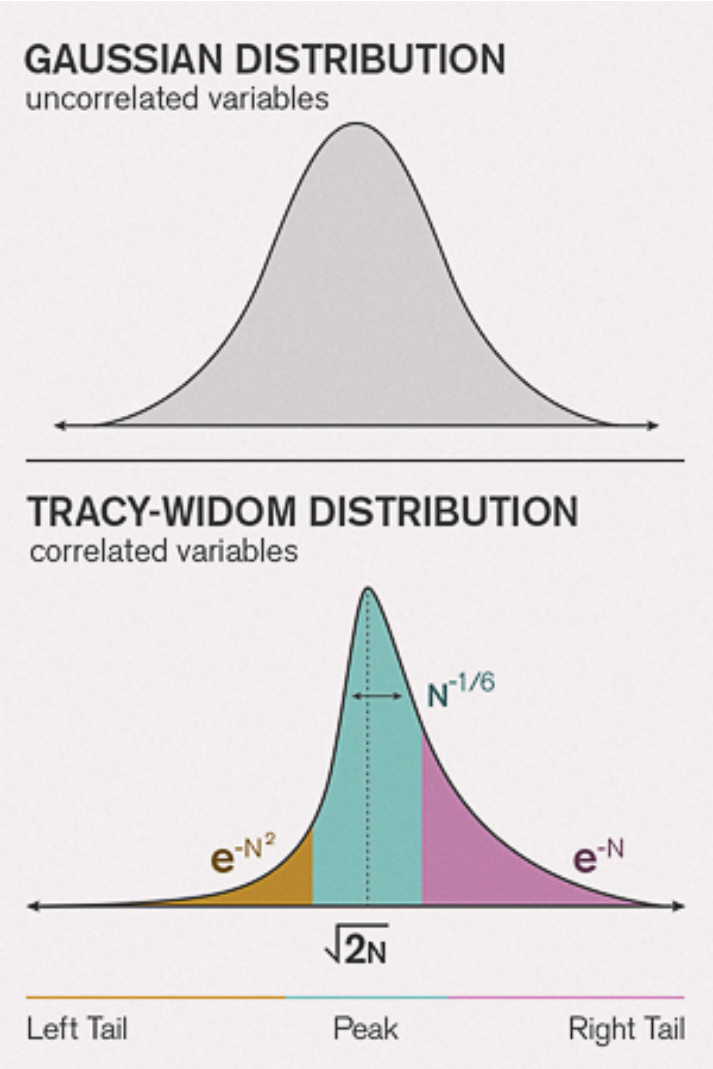
In this talk, we will focus on the Gaussian unitary ensemble case.

$$A_{ij} = x + i\,y,\ x, y \sim N(0,1)$$

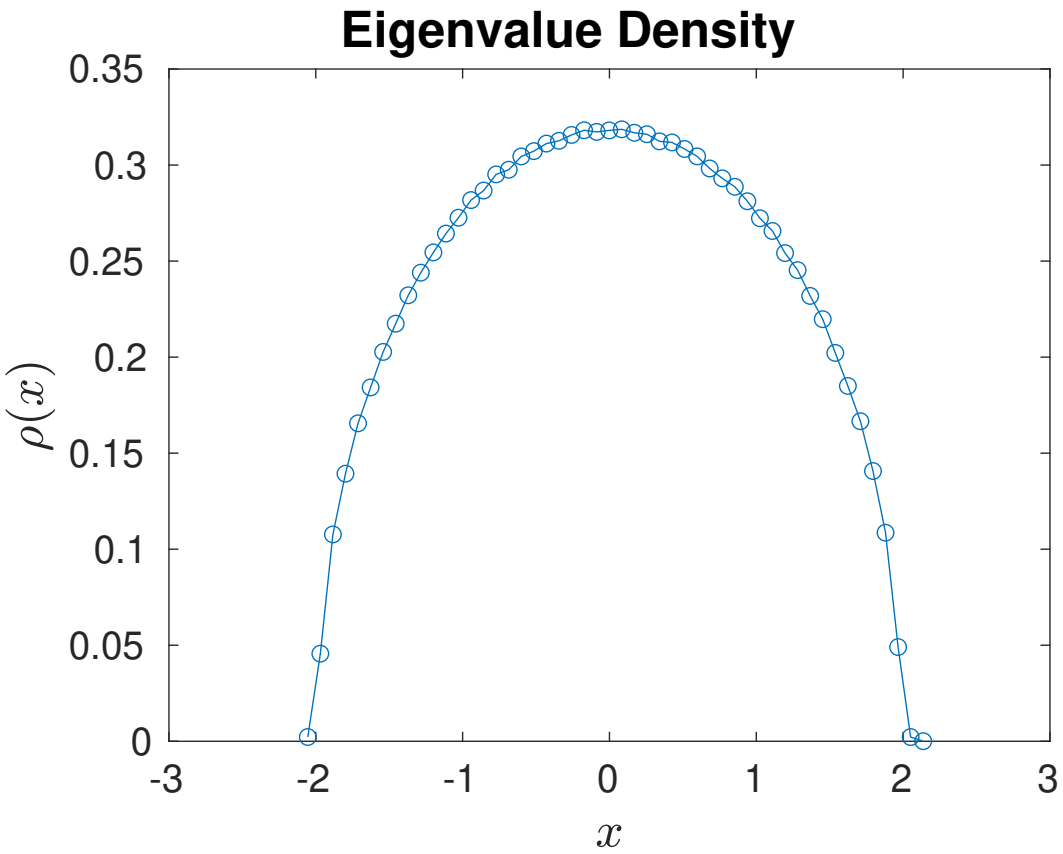
$$X = \frac{A + A^*}{2}$$

Tracy-Widom distribution vs Gaussian distribution

(cited from Quanta Magazine)



Semicircle law: eigenvalue density of $X_n/\sqrt{n}$



- Scale the $n \times n$ GUE X_n by $\frac{1}{\sqrt{n}}$, take $n \to \infty$, ...

Tracy-Widom distribution $F_2(s)$

Definition

The Tracy–Widom distribution $F_2(s)$ is defined as the limit

$$F_2(s) = \lim_{n\to\infty} P\Big(n^{1/6}(\lambda_n^{max} - 2\sqrt{n}) \leq s\Big),$$

where λ_n^{max} represents the largest eigenvalue of a $n \times n$ Gaussian unitary ensemble.

- Goal?
- This definition is unwieldy...

IVP representation of $F_2(s)$

(Slightly modified) IVP representation of $F_2(s)$ [Tracy, Widom (93)]

$$u_{xx} = 2u^3 + xu,$$

$$(\log F_2)_{xx} = -u^2,$$

$$u(x) \sim \mathrm{Ai}(x) \text{ as } x \to \infty$$

- An innocent-looking system of nonlinear IVPs, except that...
- numerical IVP solvers are mature these days...
- F_2 can be computed efficiently and accurately then?

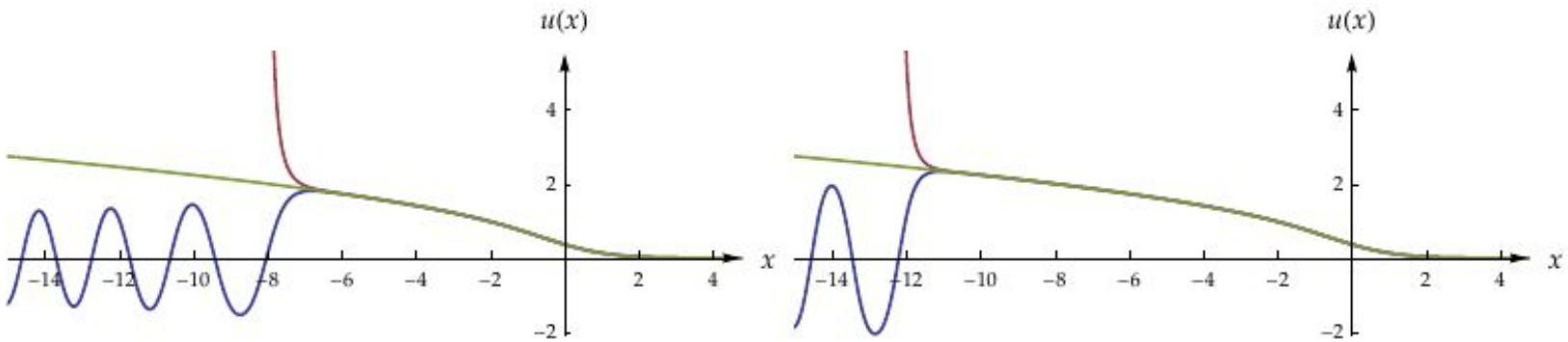
No!

The ODE $u_{xx} = 2u^3 + xu$ is very sensitive to its initial condition $u(x) \sim \mathrm{Ai}(x)$ $(x \to \infty)$.

Instability of the IVP representation

Consider a slightly perturbed initial condition
$u(x) \sim (1+\epsilon)\mathrm{Ai}(x) \ \ (x \to \infty)$,

(cited from [Bornemann (10)])



a. $\epsilon = -10^{-8}$ (blue), 0 (green), 10^{-8} (red)

b. $\epsilon = -10^{-16}$ (blue), 0 (green), 10^{-16} (red)

Bifurcation happens as x decreases, even for a tiny perturbation (which is unavoidable due to finite precision arithmetic).

Instability of the IVP representation: remedy

This can be fixed by applying some asymptotic connection formulas to the solution as $x \to -\infty$, which results in a BVP.

$$u(x) = \sqrt{-\frac{x}{2}}\left(1 + \frac{1}{8}x^{-3} - \frac{73}{128}x^{-6} + \mathcal{O}(x^{-9})\right) \quad \text{as } x \to -\infty$$

However, it

- requires considerable analytic skills
- requires parameter fine-tuning (truncation of the domain, initial guess, etc.)
- only achieves absolute accuracy in computing $F_2(s)$ and $\frac{d}{ds}F_2(s)$

Fredholm determinant representation of $F_2(s)$

Fredholm determinant representation of $F_2(s)$ [Tracy, Widom (1993)]

$$F_2(s) = \det\left(I - \mathcal{K}_s|_{L^2[s,\infty)}\right),$$

where

$$\mathcal{K}_s[f] = \int_s^\infty K_{Ai}(x,y)f(y)\,\mathrm{d}y,$$

and

$$K_{Ai}(x,y) = \int_s^\infty \mathrm{Ai}(x+z-s)\mathrm{Ai}(y+z-s)\,\mathrm{d}z. \tag{1}$$

- Observe that $\mathcal{K}_s[f] = \mathcal{G}_s^2[f]$, where $\mathcal{G}_s[f] = \int_s^\infty \mathrm{Ai}(x+y-s)f(y)\,\mathrm{d}y$.
- $\mathcal{T}_s[f](x) = \int_0^\infty \mathrm{Ai}(x+y+s)f(y)\,\mathrm{d}y$ shares the same eigenvalues and eigenfunctions (up to a translation) with $\mathcal{G}_s$. Airy integral operator!

Fredholm determinant representation of $F_2(s)$

$$\mathcal{T}_s[f](x) = \int_0^\infty \mathrm{Ai}(x+y+s) f(y)\,\mathrm{d}y$$

Since the operator $\mathcal{K}_s$ is a trace-class operator, after some derivation, we found that

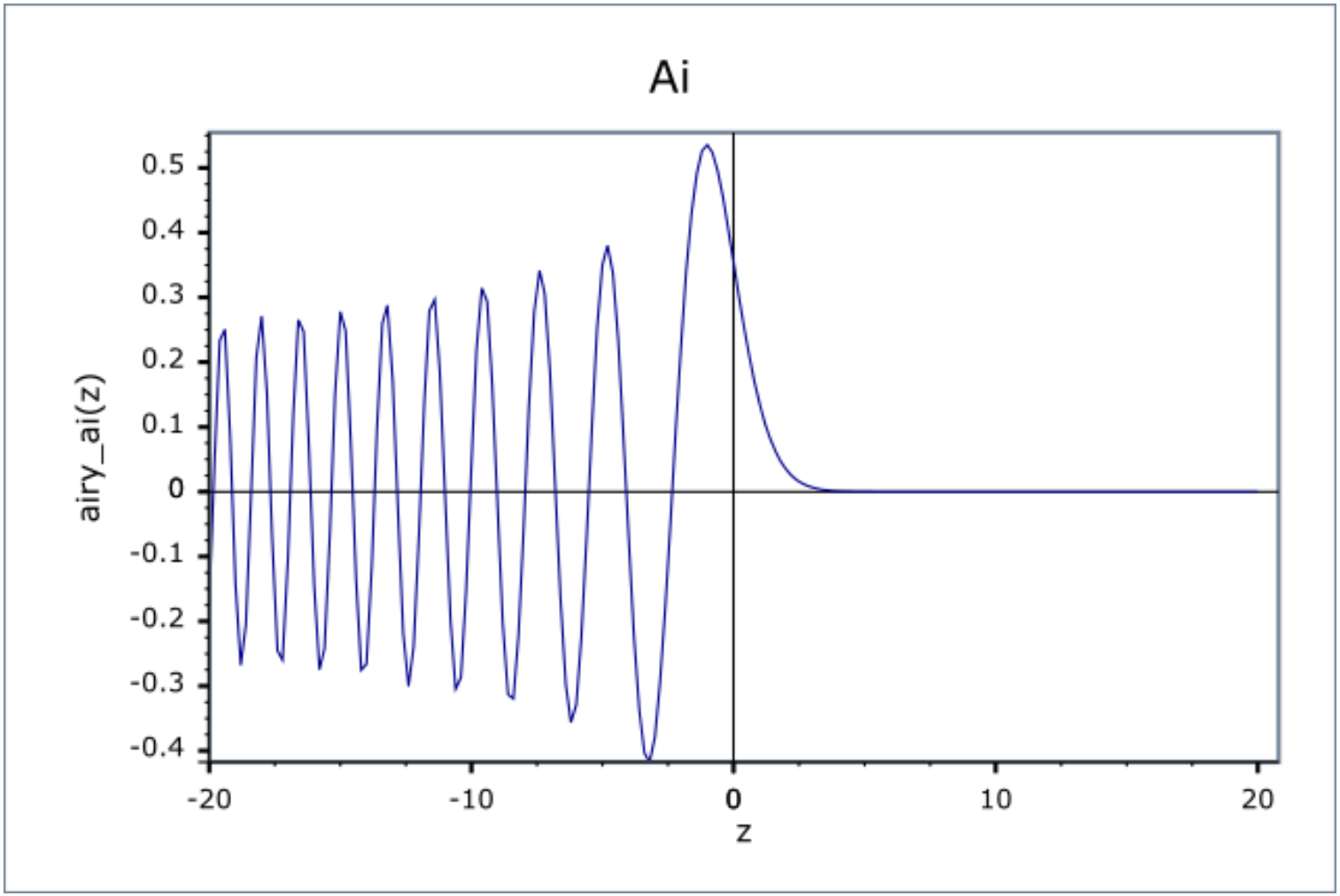
$$F_2(s) = \prod_{i=0}^{\infty}(1-\lambda_{i,s}^2),$$

$$\frac{d}{ds}F_2(s) = \sum_{i=0}^{\infty}\lambda_{i,s}^2\left(\psi_{i,s}(0)\right)^2 \prod_{j=0,j\neq i}^{\infty}(1-\lambda_{j,s}^2),$$

where λ_i and $\psi_{i,s}$ is the $(i+1)$-th eigenvalue and eigenfunction of $\mathcal{T}_s$. This leads us to study the eigendecomposition of the operator $\mathcal{T}_s$.

Difficulties

- The eigendecomposition of a compact integral operator is usually difficult to compute to high relative accuracy!
- The kernel function $\mathrm{Ai}(x)$ is bizarre!



Key observation: bi-spectral property

$$\mathcal{T}_s[f](x) = \int_0^\infty \mathrm{Ai}(x+y+s) f(y)\,\mathrm{d}y$$

- The integral operator $\mathcal{T}_s$ admits a commuting differential operator

$$\mathcal{L}_s[f](x) = -\frac{d}{dx}\left(x\frac{d}{dx}f\right) + x(x+s)f,$$

  and the spectra of them are both non-degenerate. Thus, they share the same eigenfunctions.
- The computation of the eigendecomposition of a differential operator is much easier, because ...
- Finally, we recover the eigenvalues of $\mathcal{T}_s$ using the computed eigenfunctions.

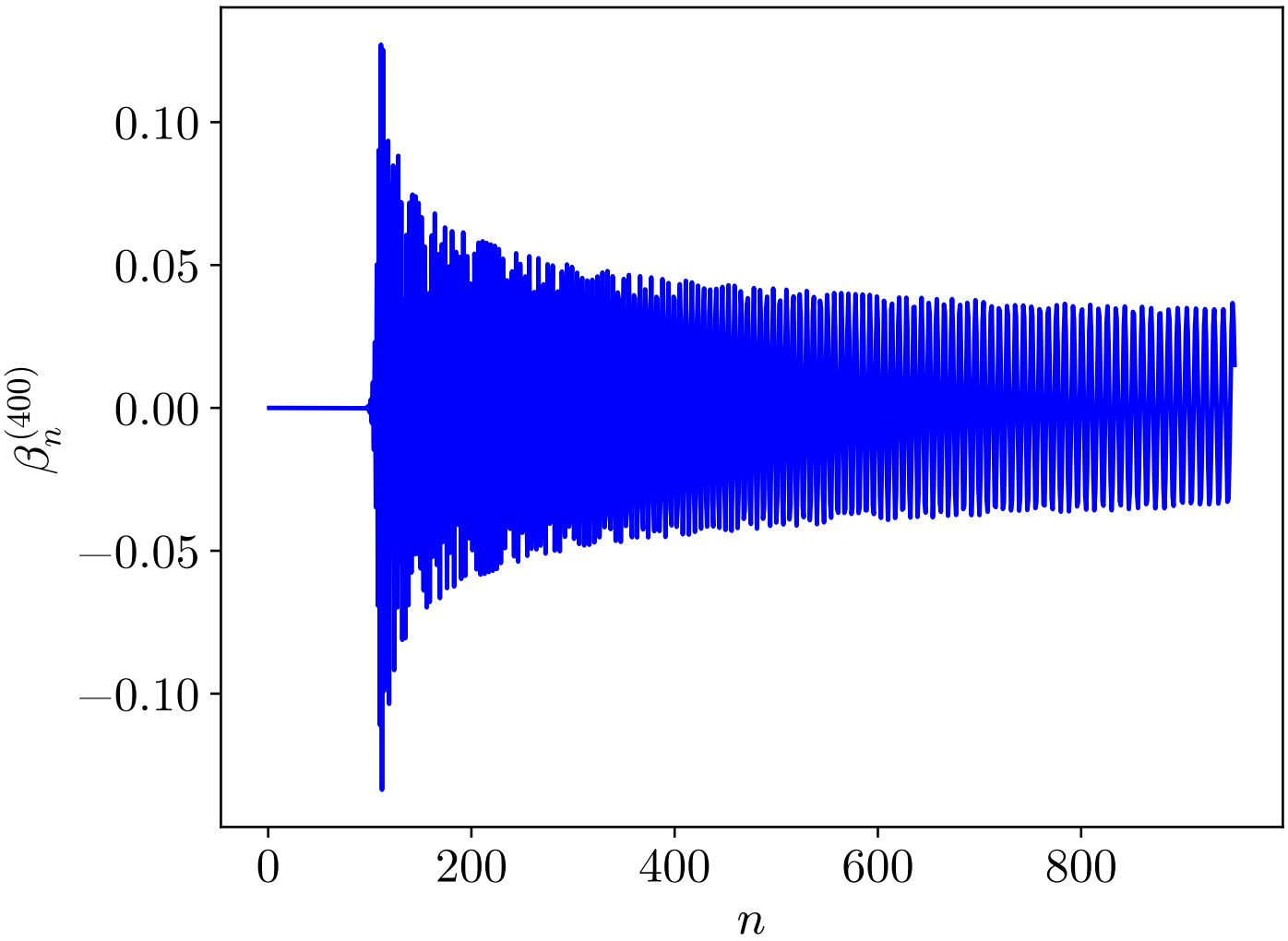
Hidden difficulties I: choice of basis functions

$$\mathcal{L}_s[f](x) = -\frac{d}{dx}\left(x\frac{d}{dx}f\right) + x(x+s)f$$

Once we find the appropriate expansion basis for the eigenfunctions, we can express the differential operator $\mathcal{L}_s$ in a matrix form, and compute the expansion coefficients (a.k.a eigenvectors) to relative precision.

- Intuitively, one would use the Laguerre functions $\{e^{-x/2}L_n(x)\}_{n=0,1,\ldots}$ to expand L^2 functions over $[0,\infty)$.
- However, we found that the Laguerre function expansion approximates $\psi_{i,s}$ poorly, especially when i is moderately large.

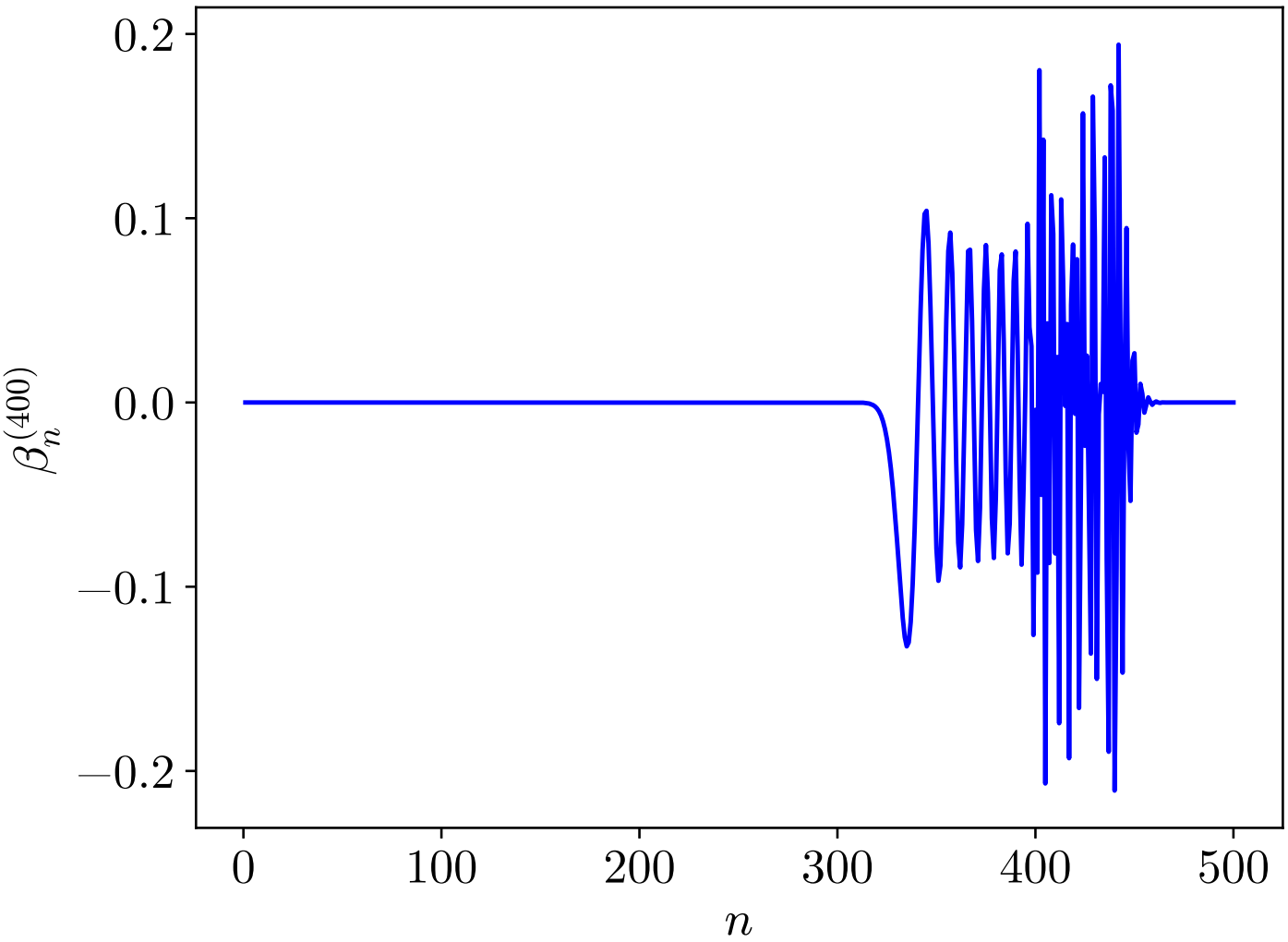
Approximating $\psi_{400,10}$ w/ Laguerre functions



Hidden difficulties I: Remedy

- The issue is that, when the turning points of the Laguerre functions $\{e^{-x/2}L_n(x)\}_{n=0,1,\ldots}$ and the eigenfunctions $\{\psi_{n,s}\}$ are far away from each other, the approximation quality deteriorates.
- We cannot easily modify the turning point of $\psi_{n,s}$, but we can translate the turning point of Laguerre functions by scaling it! Define $h_n^a(x) = \sqrt{a}e^{-ax/2}L_n(ax)$. Its turning point equals $\frac{4n+2}{a}$.

Approximating $\psi_{400,10}$ w/ optimal h_n^a



Hidden difficulties II: computation of expansion coefficients

$$\mathcal{L}_s[f](x) = -\frac{d}{dx}\left(x\frac{d}{dx}f\right) + x(x+s)f$$

Therefore, once we find the appropriate expansion basis for the eigenfunctions, **we can express the differential operator $\mathcal{L}_s$ in a matrix form, and compute the expansion coefficients (a.k.a eigenvectors) to relative precision**.

- To compute the eigenfunctions, we express the differential operator $\mathcal{L}_s$ in the basis spanned by $\{h_n^a(x)\}_{n=0,1,\ldots}$, which results in an infinite dimensional five-diagonal matrix.
- It's fine to truncate the infinite dimensional matrix, because ...
- How to compute eigenvectors to entry-wise relative precision?

Hidden difficulties III: recover integral operator's spectrum w/ its eigenfunctions

$$\mathcal{T}_s[f](x) = \int_0^\infty \mathrm{Ai}(x+y+s) f(y)\,\mathrm{d}y$$

The eigenvalues of $\mathcal{T}_s$ can be recovered once its eigenfunctions are solved accurately, by first computing $\lambda_{0,s}$ via

$$\lambda_{0,s} = \frac{\mathcal{T}_s[\psi_{0,s}](x)}{\psi_{0,s}(x)},$$

where x is chosen sophisticatedly, then compute $\lambda_{i,s}$ by

$$\lambda_{i,s} = \frac{\lambda_{1,s}}{\lambda_{0,s}} \cdot \frac{\lambda_{2,s}}{\lambda_{1,s}} \cdots \frac{\lambda_{i,s}}{\lambda_{i-1,s}}.$$

- Why don't we just compute $\lambda_{i,s} = \frac{\mathcal{T}_s[\psi_{i,s}](x)}{\psi_{i,s}(x)}$ directly?
- How to compute $\frac{\lambda_{j,s}}{\lambda_{j-1,s}}$?

Hidden difficulties III: recover integral operator's spectrum w/ its eigenfunctions

$$\mathcal{T}_s[f](x) = \int_0^\infty \mathrm{Ai}(x+y+s) f(y)\,\mathrm{d}y, \quad h_n^a(x) = \sqrt{a} e^{-ax/2} L_n(ax)$$

$$\lambda_{0,s} = \frac{\mathcal{T}_s[\psi_{0,s}](x)}{\psi_{0,s}(x)} = \frac{1}{\psi_{0,s}(x)} \sum_{k=0}^{\infty} \alpha_k \mathcal{T}_s[h_k^a](x)$$

- We cannot just evaluate $\mathcal{T}_s[h_k^a](x) = \int_0^\infty \mathrm{Ai}(x+y+s)\sqrt{a}e^{-ay/2}L_k(ay)\,\mathrm{d}y$ via numerical integration, because ...

Hidden difficulties III: Remedy

$$\mathcal{T}_s[h_k^a](x) = \int_0^\infty \mathrm{Ai}(x+y+s)\sqrt{a}e^{-ay/2}L_k(ay)\,\mathrm{d}y$$

- We derived a five-term linear homogeneous recurrence relation involving $\mathcal{T}_s[h_{k-2}^a](x), \mathcal{T}_s[h_{k-1}^a](x), \ldots, \mathcal{T}_s[h_{k+2}^a](x)$ that satisfies one linear constraint.
- Express it as an infinite dimensional matrix. The matrix must has a non-trivial null space, so if we view it as an eigenproblem ...
- Truncate the matrix, solve for the eigenvector that corresponds to the zero eigenvalue by inverse power iterations. The resulting eigenvector equals the value of $\{\mathcal{T}_s[h_k^a](x)\}_{k=0,1,\ldots,N}$ up to a scaling.
- Scale it back by computing $\mathcal{T}_s[h_0^a](x)$.

Distribution of the k-th largest eigenvalue of Gaussian ensembles in the scaling limit

Once the spectrum and expansion coefficients of $\mathcal{T}_s$ are computed to relative precision, we are also able to evaluate the distribution of k-th largest eigenvalue of GUE/GOE/GSE at the soft edge scaling limit to relative precision everywhere except in the left tail!

$$F_2(k;s) = \sum_{j=0}^{k-1} \frac{1}{j!} \sum_{i_1=0}^{\infty} \lambda_{i_1,s}^2 \sum_{\substack{i_2=0,\\ i_2\neq i_1}}^{\infty} \lambda_{i_2,s}^2 \quad \cdots \quad \sum_{\substack{i_j=0,\\ i_j\neq i_1,\ldots,i_{j-1}}}^{\infty} \lambda_{i_j,s}^2 \prod_{\substack{i=0,\\ i\neq i_1,\ldots,i_j}}^{\infty} (1-\lambda_{i,s}^2),$$

$$\frac{d}{ds} F_2(k;s) = \frac{1}{(k-1)!} \sum_{i_1=0}^{\infty} \lambda_{i_1,s}^2 \sum_{\substack{i_2=0,\\ i_2\neq i_1}}^{\infty} \lambda_{i_2,s}^2 \quad \cdots \quad \sum_{\substack{i_k=0,\\ i_k\neq i_1,\ldots,i_{k-1}}}^{\infty} \lambda_{i_k,s}^2 \left(\psi_{i_k,s}(0)\right)^2 \prod_{\substack{i=0,\\ i\neq i_1,\ldots,i_k}}^{\infty} (1-\lambda_{i,s}^2)$$

Conclusions and remarks

- The key contribution of our work ("On the Evaluation of the Eigendecomposition of the Airy Integral Operator", joint with Prof. Kirill Serkh) is the algorithm for evaluating the spectra and eigenfunctions of the Airy integral operator to relative accuracy in $\mathcal{O}(n^2)$ time.
- The so-called Airy integral operator has application not only in random matrix theory, but also in optics, etc. Check it out!
- In [Bornemann 10], the author describes an algorithm for computing the spectrum by directly discretizing the integral operator. Although the algorithm is simple to implement, the eigenvalues are solved only to absolute precision. Therefore, the computation of $F_2(k;s)$, $\frac{d}{ds}F_2(k;s)$ becomes less and less accurate as k increases.