Spectral Estimation with Free Decompression

Siavash Ameli [1,2] Chris van der Heide [3] Liam Hodgkinson [4]
Michael W. Mahoney [1,2,5]

[1] *Department of Statistics, UC Berkeley*
[2] *International Computer Science Institute*
[3] *Dept. of Electrical and Electronic Eng., University of Melbourne*
[4] *School of Mathematics and Statistics, University of Melbourne*
[5] *Lawrence Berkeley National Laboratory*

NeurIPS 2025

Overview

Motivation:

- **Eigenvalues** encode essential matrix information. The empirical spectral distribution is useful for diagnostics.
- **Spectral Functions** are particularly useful, including

$$\mathrm{logdet}(\mathbf{A}) = \sum_i \log \lambda_i(\mathbf{A}), \qquad \mathrm{trace}(\mathbf{A}^{-k}) = \sum_i \lambda_i(\mathbf{A})^{-k}, \qquad \mathrm{cond}(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}.$$

Challenges:

- These quantities are important, e.g., for Gaussian processes, but require the **entire range** of eigenvalues.
- **Standard eigenvalue solvers** have $\mathcal{O}(n^3)$ complexity; expensive for large matrices!
- **Matrix formation** can also bottleneck at the $\mathcal{O}(n^2)$ **memory wall**.

Outline

I. Background

- Free probability
- Stieltjes transform

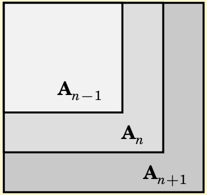
II. Method

- Free decompression
- Challenges

III. Results

- Synthetic data
- Real datasets

I. Background



Suppose matrix of interest is embedded in an infinite sequence of nested matrices

$$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots \quad \text{with } \mathbf{A}_n \in \mathbb{R}^{n \times n} \quad \text{such that} \quad (\mathbf{A}_n)_{ij} = (\mathbf{A}_{n+1})_{ij}.$$

Objective: Find the eigenspectrum of $\mathbf{A}_n$ using only knowledge of $\mathbf{A}_{n_0}$, where $n_0 \ll n$.

How to ensure the eigenvalues of submatrices represent the whole matrix?

Free Probability

- An important topic in random matrix theory involving random matrix with uniformly random eigenvectors.
- Properties of the matrix (including submatrices) depend only on the eigenvalues and **not on eigenvectors**.

THEOREM (NICA, 1993)

Any sequence of matrices can be turned into an (asymptotically) free sequence of random matrices by applying random permutations σ to the rows and columns:

$$\tilde{\mathbf{A}}_{ij} = \mathbf{A}_{\sigma(i)\sigma(j)}.$$

- The spectral density of $\mathbf{A} \in \mathbb{R}^{n \times n}$ is encoded in its **Stieltjes transform** $m : \mathbb{C} \to \mathbb{C}$:

$$m_n(z) = \frac{1}{n} \mathsf{trace}(\mathbf{A} - z\mathbf{I})^{-1}$$

- In the large matrix limit, $m_n \to m$ and eigenvalues are drawn from the **density** ρ.
- Stieltjes transform is the **Cauchy integral** of density:

$$m(z) = \int_{-\infty}^{\infty} \frac{\rho(x)}{x - z} \, \mathrm{d}x.$$

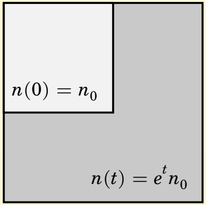
- Density can be retrieved back by the **inverse transform**:

$$\rho(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \Im[m(x + i\varepsilon)].$$

- There is a **one-to-one** correspondence between ρ and m.

II. Method

Free Decompression



- Suppose a matrix $\mathbf{A}_n$ of variable size $n(t) = e^t n_0$.
- Let $m(t, \cdot)$ be the Stiletjes transform of $\mathbf{A}_{n(t)}$.
- Under the large matrix limit, $m(t, \cdot)$ satisfies the **partial differential equation**

$$\frac{\partial m}{\partial t} = -m + \frac{1}{m}\frac{\partial m}{\partial z}$$

- This operation has always been considered in reverse (as **free compression**) to find eigenspectra of submatrices of **A**.
- We are the first to attempt ***free decompression***.

Free decompression of a random submatrix $\mathbf{A}_{n_0}$ to a larger matrix $\mathbf{A}_n$ requires:

1. **estimation** of the Stieltjes transform $m_{\mathbf{A}_{n_0}}$, giving the initial condition $m(0, \cdot)$;
2. **evolution** of $m(t, \cdot)$ in t via the PDE from $t = 0$ to $T = \log(n/n_0)$;
3. **evaluation** of the spectral distribution of $\mathbf{A}_n$ from $m(T, \cdot)$.

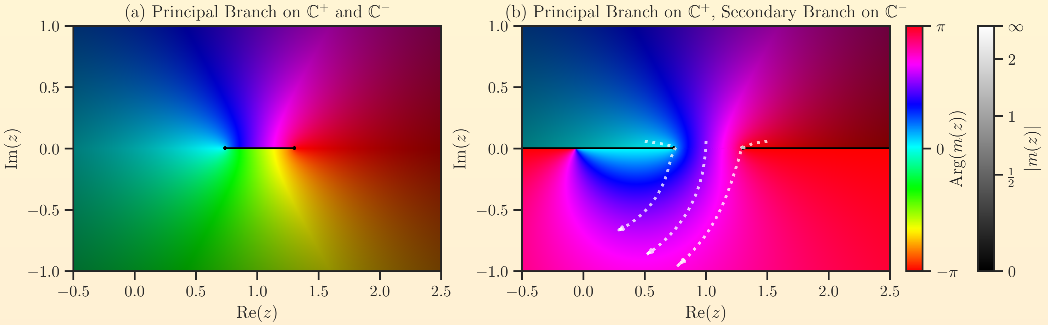
An Engineering Challenge

We solve the PDE using method of characteristics in the complex plane. But, this is a ***difficult problem to solve!***

Proposition

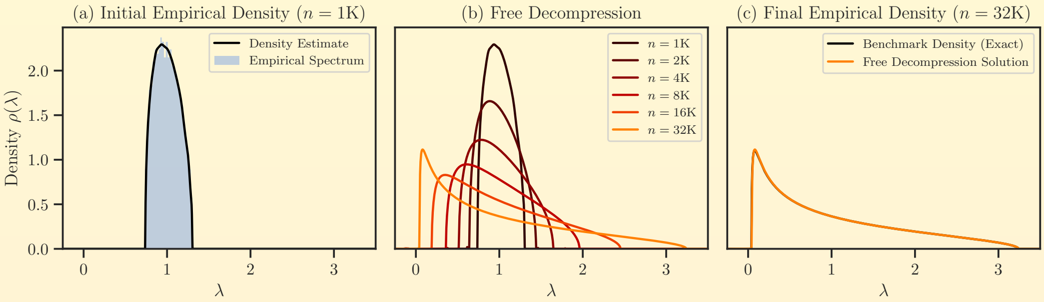
*All characteristic curves pass through the (**discontinuous**) branch cut for the principal branch of the Stieltjes transform.*

- To solve the characteristic equations, a **secondary branch** is required
- This is tantamount to analytic continuation (which is **ill-posed**)
- Naïvely solving the PDE fails: we need to directly solve the analytic continuation problem



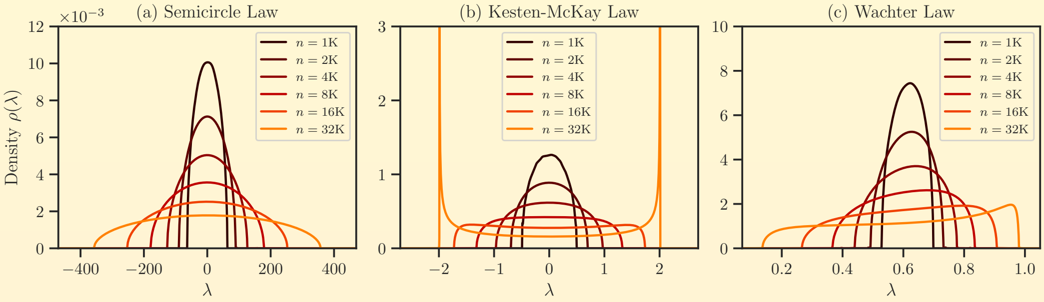
III. Results

Experiments with Random Matrix Ensembles I

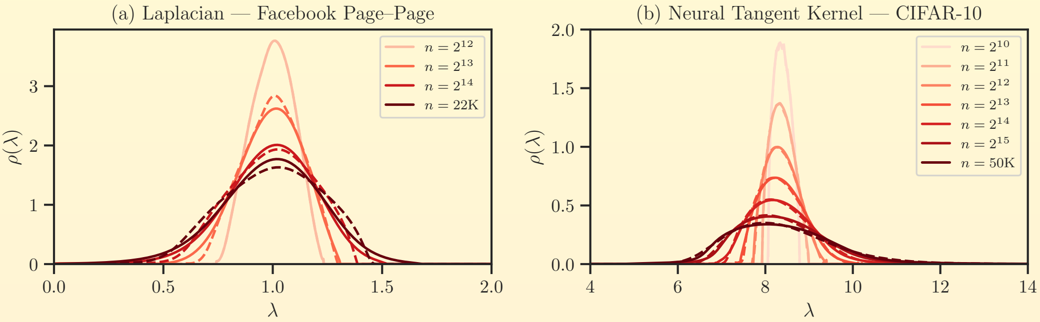


- Synthetic examples act as convenient baselines since the expected shape of the eigenspectrum is **known in advance**.
- Under Wishart initial data, we expand $n_0 = 1000 \xrightarrow{\textit{free decompression}} n = 32,000$ (**Marchenko-Pastur law**).

Experiments with Random Matrix Ensembles II



- More examples: **Semicircle** (a), **Kesten–McKay** (b), and **Wachter** (c) laws
- All distributors accurately match benchmark (not shown here, see paper).



1. **Facebook SNAP Dataset**
2. 22,470 × 22,470 adjacency matrix
3. Perturbed by an Erdős-Rényi regularization.

1. Log-**Neural Tangent Kernel** (ResNet-50)
2. 50,000 × 50,000 dense matrix
3. Low-rank components removed

Large real data covariance and kernel matrices with **disconnected spectral densities** ***remain challenging***.

Size	Process Time (sec)		Distributional Distances			Moments Rel. Error	
n	Direct	FD (ours)	TV	JS	KS	$\Delta\mu_1/\mu_1$	$\Delta\mu_2/\mu_2$
2^{10}	3.6	$\mathbf{3.6 + 0.0}$	0.00%	0.00%	0.00%	0.00%	0.00%
2^{11}	10.2	$\mathbf{3.6 + 0.6}$	1.72%	7.60%	0.48%	0.05%	0.09%
2^{12}	50.9	$\mathbf{3.6 + 0.6}$	2.06%	4.67%	0.70%	0.01%	0.02%
2^{13}	358.9	$\mathbf{3.6 + 0.6}$	3.24%	6.30%	1.18%	0.01%	0.02%
2^{14}	2820.2	$\mathbf{3.6 + 0.7}$	4.33%	7.55%	1.76%	0.01%	0.03%
2^{15}	20451.2	$\mathbf{3.6 + 0.8}$	5.16%	7.96%	2.51%	0.02%	0.05%
50K	67331.1	$\mathbf{3.6 + 0.8}$	5.94%	8.33%	3.02%	0.17%	0.49%

- Showing runtime and accuracy results for **NTK** data (previous slide)
- Runtime with direct method increase by $\mathcal{O}(n^3)$.
- Runtime with **free decompression (FD)** is an initial overhead $\mathcal{O}(n_0^3)$ plus $\mathcal{O}(1)$.

Resources

Reference

Ameli, S., van der Heide, C., Hodgkinson, L., Mahoney, M.W. (2025). Spectral Estimation with Free Decompression. *The Thirty-ninth Annual Conference on Neural Information Processing Systems*

Related Work

Ameli, S., van der Heide, C., Hodgkinson, L., Roosta, F., Mahoney, M.W. (2025). Determinant Estimation under Memory Constraints and Neural Scaling Laws, *The 42nd International Conference on Machine Learning*.

Software

Package	Documentation	Install	Implements
freealg	`ameli.github.io/freealg`	`pip install freealg`	This work
detkit	`ameli.github.io/detkit`	`pip install detkit`	Related work
imate	`ameli.github.io/imate`	`pip install imate`	SLQ