Spectral limit theorems for submatrices and products of projections

Tim Kunisky

Yale University

MIT Probability Seminar
April 24, 2023

I. Introduction

Eigenvalues Under Compression

$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{N\times N}$ orthogonal projections of **linear rank**:

$$\frac{1}{N}\,\mathsf{Tr}(\boldsymbol{A}) \approx \alpha \in (0,1),$$
$$\frac{1}{N}\,\mathsf{Tr}(\boldsymbol{B}) \approx \beta \in (0,1).$$

Eigenvalues Under Compression

$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{N \times N}$ orthogonal projections of **linear rank**:

$$\frac{1}{N}\,\mathsf{Tr}(\boldsymbol{A}) \approx \alpha \in (0,1),$$
$$\frac{1}{N}\,\mathsf{Tr}(\boldsymbol{B}) \approx \beta \in (0,1).$$

This specifies the **empirical spectral distributions**:

$$\text{e.s.d. of } \boldsymbol{A} \approx (1-\alpha)\delta_0 + \alpha\delta_1 = \mathsf{Ber}(\alpha),$$
$$\text{e.s.d. of } \boldsymbol{B} \approx (1-\beta)\delta_0 + \beta\delta_1 = \mathsf{Ber}(\beta).$$

Eigenvalues Under Compression

$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{N\times N}$ orthogonal projections of **linear rank**:

$$\frac{1}{N}\,\mathsf{Tr}(\boldsymbol{A}) \approx \alpha \in (0,1),$$
$$\frac{1}{N}\,\mathsf{Tr}(\boldsymbol{B}) \approx \beta \in (0,1).$$

This specifies the **empirical spectral distributions**:

$$\text{e.s.d. of } \boldsymbol{A} \approx (1-\alpha)\delta_0 + \alpha\delta_1 = \mathsf{Ber}(\alpha),$$
$$\text{e.s.d. of } \boldsymbol{B} \approx (1-\beta)\delta_0 + \beta\delta_1 = \mathsf{Ber}(\beta).$$

How does **compressing** $\boldsymbol{B}$ by $\boldsymbol{A}$ change the eigenvalues?

$$\text{e.s.d. of } \boldsymbol{ABA} \approx ?$$

Geometric Interpretation

Say $\boldsymbol{A}$ projects to $U \subset \mathbb{C}^N$ and $\boldsymbol{B}$ projects to $V \subset \mathbb{C}^N$.

Geometric Interpretation

Say $\boldsymbol{A}$ projects to $U \subset \mathbb{C}^N$ and $\boldsymbol{B}$ projects to $V \subset \mathbb{C}^N$. The **principal angles** between U and V are:

$$\cos(\theta_k) := \max_{\substack{\boldsymbol{a}_k \in U \\ \|\boldsymbol{a}_k\|=1 \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_k \rangle = 0 \text{ for } 1 \le i < k}} \quad \max_{\substack{\boldsymbol{b}_k \in V \\ \|\boldsymbol{b}_k\|=1 \\ \langle \boldsymbol{b}_j, \boldsymbol{b}_k \rangle = 0 \text{ for } 1 \le j < k}} \langle \boldsymbol{a}_k, \boldsymbol{b}_k \rangle.$$

Geometric Interpretation

Say $\boldsymbol{A}$ projects to $U \subset \mathbb{C}^N$ and $\boldsymbol{B}$ projects to $V \subset \mathbb{C}^N$. The **principal angles** between U and V are:

$$\cos(\theta_k) := \max_{\substack{\boldsymbol{a}_k \in U \\ \|\boldsymbol{a}_k\|=1 \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_k\rangle = 0 \text{ for } 1 \le i < k}} \quad \max_{\substack{\boldsymbol{b}_k \in V \\ \|\boldsymbol{b}_k\|=1 \\ \langle \boldsymbol{b}_j, \boldsymbol{b}_k\rangle = 0 \text{ for } 1 \le j < k}} \langle \boldsymbol{a}_k, \boldsymbol{b}_k \rangle.$$

Populate columns of $\boldsymbol{U}, \boldsymbol{V}$ with orthonormal bases of U, V, so $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{U}^*$, $\boldsymbol{B} = \boldsymbol{V}\boldsymbol{V}^*$. Then:

$$\begin{aligned} \cos(\theta_k) &= k\text{th singular value of } \boldsymbol{U}^*\boldsymbol{V} \\ &= (k\text{th eigenvalue of } \boldsymbol{U}^*\boldsymbol{V}\boldsymbol{V}^*\boldsymbol{U})^{1/2} \\ &= (k\text{th eigenvalue of } \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{V}\boldsymbol{V}^*\boldsymbol{U}\boldsymbol{U}^*)^{1/2} \\ &= (k\text{th eigenvalue of } \underbrace{\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}}_{\text{“angle operator”}})^{1/2}. \end{aligned}$$

Application: Submatrices

Special case: $\boldsymbol{A}$ is diagonal, a **coordinate projection**:

$$\boldsymbol{A} = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 1 \end{bmatrix}$$

Application: Submatrices

Special case: $\boldsymbol{A}$ is diagonal, a **coordinate projection**:

$$\boldsymbol{A} = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 1 \end{bmatrix}$$

Then, $\boldsymbol{ABA}$ extracts a **submatrix**:

$$\boldsymbol{A} \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \boldsymbol{A} \quad = \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & 0 & \cdot \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \cdot & \cdot & 0 & \cdot \end{bmatrix}$$

Submatrices ⇝ Induced Subgraphs

A **strongly d-regular graph** has only three eigenvalues: the "trivial" d eigenvalue, and two with large multiplicity:

$$\begin{aligned} \boldsymbol{G} &= \frac{d}{N}\mathbf{1}\mathbf{1}^* + \lambda_1 \boldsymbol{B}_1 + \lambda_2 \boldsymbol{B}_2 \\ &= \frac{d}{N}\mathbf{1}\mathbf{1}^* + \lambda_1 \boldsymbol{B}_1 + \lambda_2 \left(\boldsymbol{I} - \boldsymbol{B}_1 - \frac{1}{N}\mathbf{1}\mathbf{1}^* \right) \\ &= (\lambda_1 - \lambda_2)\boldsymbol{B}_1 + \text{simple adjustment.} \end{aligned}$$

Submatrices $\rightsquigarrow$ Induced Subgraphs

A **strongly d-regular graph** has only three eigenvalues: the "trivial" d eigenvalue, and two with large multiplicity:

$$\begin{aligned} \boldsymbol{G} &= \frac{d}{N}\mathbf{1}\mathbf{1}^* + \lambda_1 \boldsymbol{B}_1 + \lambda_2 \boldsymbol{B}_2 \\ &= \frac{d}{N}\mathbf{1}\mathbf{1}^* + \lambda_1 \boldsymbol{B}_1 + \lambda_2 \left(\boldsymbol{I} - \boldsymbol{B}_1 - \frac{1}{N}\mathbf{1}\mathbf{1}^*\right) \\ &= (\lambda_1 - \lambda_2)\boldsymbol{B}_1 + \text{simple adjustment.} \end{aligned}$$

If $(\boldsymbol{A}_S)_{ii} = \mathbb{1}\{i \in S\}$ for $S \subseteq [N]$, then

$$\boldsymbol{A}_S \boldsymbol{G} \boldsymbol{A}_S = \text{adjacency matrix of induced subgraph on } S,$$

and we can understand the spectrum via $\boldsymbol{A}_S \boldsymbol{B}_1 \boldsymbol{A}_S$.

Submatrices $\rightsquigarrow$ Restricted Isometry Property

In **compressed sensing**, want $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N \in \mathbb{C}^M$ with $N \gg M$ so that any small subset is close to orthonormal.

Submatrices $\rightsquigarrow$ Restricted Isometry Property

In **compressed sensing**, want $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N \in \mathbb{C}^M$ with $N \gg M$ so that any small subset is close to orthonormal.

$$\boldsymbol{V} = \begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_N \\ | & & | \end{bmatrix},$$

$$\begin{aligned} k\text{-RIP constant} &= \max_{i_1,\ldots,i_k \in [N]} \left\| (\langle \boldsymbol{v}_{i_a}, \boldsymbol{v}_{i_b} \rangle)_{a,b=1}^k - \boldsymbol{I}_k \right\| \\ &= \max_{S \subset [N], |S|=k} \left\| \boldsymbol{A}_S \boldsymbol{V}^* \boldsymbol{V} \boldsymbol{A}_S - \boldsymbol{I}_k \oplus \boldsymbol{0} \right\|. \end{aligned}$$

Submatrices $\rightsquigarrow$ Restricted Isometry Property

In **compressed sensing**, want $\boldsymbol{v}_1, \dots, \boldsymbol{v}_N \in \mathbb{C}^M$ with $N \gg M$ so that any small subset is close to orthonormal.

$$\boldsymbol{V} = \begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_N \\ | & & | \end{bmatrix},$$

$$\begin{aligned} k\text{-RIP constant} &= \max_{i_1,\dots,i_k \in [N]} \left\| (\langle \boldsymbol{v}_{i_a}, \boldsymbol{v}_{i_b} \rangle)_{a,b=1}^k - \boldsymbol{I}_k \right\| \\ &= \max_{S \subset [N], |S| = k} \left\| \boldsymbol{A}_S \boldsymbol{V}^* \boldsymbol{V} \boldsymbol{A}_S - \boldsymbol{I}_k \oplus \boldsymbol{0} \right\|. \end{aligned}$$

The $\boldsymbol{v}_i$ are a **tight frame** if $\sum_{i=1}^N \boldsymbol{v}_i \boldsymbol{v}_i^* = \boldsymbol{V}\boldsymbol{V}^* = c\boldsymbol{I}_M$.

If so, $\boldsymbol{V}^*\boldsymbol{V} = c\boldsymbol{B}$ is a rescaled projection, so this is a question about the eigenvalues (over all S) of $\boldsymbol{A}_S \boldsymbol{B} \boldsymbol{A}_S$.

MANOVA Universality Class

MANOVA Universality Class

Summary: If U and V are in "sufficiently general position," then the eigenvalues of $\boldsymbol{ABA}$ follow the universal **Wachter MANOVA distribution** with density:

$$
\begin{aligned}
d\mu_{\alpha,\beta}(x) &= \frac{\sqrt{(r_+ - x)(x - r_-)}}{2\pi x(1-x)} \mathbb{1}_{[r_-, r_+]}(x)\,dx \\
&\quad + \max\{1-\beta, 1-\alpha\}\,\delta_0(x) \\
&\quad + \max\{\beta - (1-\alpha), 0\}\,\delta_1(x),
\end{aligned}
$$

$$
\begin{aligned}
r_\pm &= \alpha + \beta - 2\alpha\beta \pm 2\sqrt{\alpha(1-\alpha)\beta(1-\beta)} \\
&= \left(\sqrt{\alpha(1-\beta)} \pm \sqrt{\beta(1-\alpha)}\right)^2 \in (0,1).
\end{aligned}
$$

MANOVA Universality Class

Summary: If U and V are in "sufficiently general position," then the eigenvalues of $\boldsymbol{ABA}$ follow the universal **Wachter MANOVA distribution** with density:

$$d\mu_{\alpha,\beta}(x) = \frac{\sqrt{(r_+ - x)(x - r_-)}}{2\pi x(1-x)}\mathbb{1}_{[r_-,r_+]}(x)dx + \max\{1-\beta, 1-\alpha\}\,\delta_0(x) + \max\{\beta-(1-\alpha), 0\}\,\delta_1(x),$$

$$r_\pm = \alpha + \beta - 2\alpha\beta \pm 2\sqrt{\alpha(1-\alpha)\beta(1-\beta)} = \left(\sqrt{\alpha(1-\beta)} \pm \sqrt{\beta(1-\alpha)}\right)^2 \in (0,1).$$

Interpretation: Eigenvalues $(1-\beta)\delta_0 + \beta\delta_1$ of $\boldsymbol{B}$ are **smoothed**, **0 atom increases**, **1 atom decreases**.

II. Empirical Spectral Distribution

Free Probability Perspective [Voiculescu '90s]

Consider **sequences** of **random** orthogonal projections $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)} \in \mathbb{C}^N$ with

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{A}^{(N)} = \alpha, \quad \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{B}^{(N)} = \beta.$$

Free Probability Perspective [Voiculescu '90s]

Consider **sequences** of **random** orthogonal projections $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)} \in \mathbb{C}^N$ with

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{A}^{(N)} = \alpha, \quad \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{B}^{(N)} = \beta.$$

Equivalent to convergence of e.s.d.'s to $\mathsf{Ber}(\alpha), \mathsf{Ber}(\beta)$.

Free Probability Perspective [Voiculescu '90s]

Consider **sequences** of **random** orthogonal projections $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)} \in \mathbb{C}^N$ with

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E} \operatorname{Tr} \boldsymbol{A}^{(N)} = \alpha, \quad \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \operatorname{Tr} \boldsymbol{B}^{(N)} = \beta.$$

Equivalent to convergence of e.s.d.'s to $\mathsf{Ber}(\alpha), \mathsf{Ber}(\beta)$.

Free probability $\rightsquigarrow$ if $(\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)})$ **asymptotically free**, have convergence (in moments) of e.s.d. of $\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)}$ to

$$\mathsf{Ber}(\alpha) \boxtimes \mathsf{Ber}(\beta) = \mu_{\alpha,\beta}.$$

To establish weak convergence, suffices to establish asymptotic freeness.

Asymptotic Freeness for Projections

Usual definition: for all $s_1, t_1, \ldots, s_k, t_k \geq 1$, let

$$a_i := \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathrm{Tr}\, \boldsymbol{A}^{(N)^{s_i}}, \qquad b_i := \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathrm{Tr}\, \boldsymbol{B}^{(N)^{t_i}}.$$

Then, asymptotic freeness $\Leftrightarrow$ for any such choice,

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathrm{Tr} \prod_{i=1}^{k} (\boldsymbol{A}^{(N)^{s_i}} - a_i \boldsymbol{I}_N)(\boldsymbol{B}^{(N)^{t_i}} - b_i \boldsymbol{I}_N) = 0.$$

Asymptotic Freeness for Projections

Usual definition: for all $s_1, t_1, \ldots, s_k, t_k \geq 1$, let

$$a_i := \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \operatorname{Tr} \boldsymbol{A}^{(N)^{s_i}}, \quad b_i := \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \operatorname{Tr} \boldsymbol{B}^{(N)^{t_i}}.$$

Then, asymptotic freeness $\Leftrightarrow$ for any such choice,

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \operatorname{Tr} \prod_{i=1}^{k} (\boldsymbol{A}^{(N)^{s_i}} - a_i \boldsymbol{I}_N)(\boldsymbol{B}^{(N)^{t_i}} - b_i \boldsymbol{I}_N) = 0.$$

But for projections, by idempotence, enough to analyze **one-parameter family of traces** $s_1 = t_1 = \cdots = s_k = t_k = 1$:

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \operatorname{Tr} \left((\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N)(\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N) \right)^k = 0.$$

Main Theorem 1 [K '23]

$\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)} \in \mathbb{C}^{N \times N}$ random orthogonal projections with

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{A}^{(N)} = \alpha, \quad \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{B}^{(N)} = \beta,$$

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\left((\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N)(\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N)\right)^k = 0 \text{ for all } k \geq 1.$$

Main Theorem 1 [K '23]

$\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)} \in \mathbb{C}^{N\times N}$ random orthogonal projections with

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{A}^{(N)} = \alpha, \quad \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\,\boldsymbol{B}^{(N)} = \beta,$$

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\left((\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N)(\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N)\right)^k = 0 \text{ for all } k \geq 1.$$

Then, we have **convergence in moments**: for all $k \geq 1$,

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}(\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)})^k = \mathop{\mathbb{E}}_{\lambda\sim\mu_{\alpha,\beta}} \lambda^k.$$

Main Theorem 1 [K '23]

$\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)} \in \mathbb{C}^{N\times N}$ random orthogonal projections with

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\, \mathsf{Tr}\, \boldsymbol{A}^{(N)} = \alpha, \quad \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\, \mathsf{Tr}\, \boldsymbol{B}^{(N)} = \beta,$$

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\, \mathsf{Tr}\left((\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N)(\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N)\right)^k = 0 \text{ for all } k \geq 1.$$

Then, we have **convergence in moments**: for all $k \geq 1$,

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\, \mathsf{Tr}(\boldsymbol{A}^{(N)} \boldsymbol{B}^{(N)} \boldsymbol{A}^{(N)})^k = \mathop{\mathbb{E}}_{\lambda \sim \mu_{\alpha,\beta}} \lambda^k.$$

Remarks:

- Straightforward application of free probability tools.
- Extended to weak convergence in probability or a.s.

Application: Random Subsets of Frames

Main Theorem 1 applies with:

- $\boldsymbol{A}^{(N)}$ diagonal, $A_{ii} \overset{\text{iid}}{\sim} \mathsf{Ber}(\alpha)$,
- $\boldsymbol{B}^{(N)} = \frac{M}{N}\boldsymbol{V}^{(N)*}\boldsymbol{V}^{(N)}$ for $\boldsymbol{V}^{(N)} = [\, \boldsymbol{v}_1 \cdots \boldsymbol{v}_N \,] \in \mathbb{C}^{M \times N}$ (deterministic!) tight frames having $\frac{M}{N} \to \beta \in (0,1)$ and

$$\max_{i,j \in [N]} \left| \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle - \mathbb{1}\{i = j\} \right| \leq N^{-1/2+o(1)}.$$

Application: Random Subsets of Frames

Main Theorem 1 applies with:

- $\boldsymbol{A}^{(N)}$ diagonal, $A_{ii} \overset{\mathsf{iid}}{\sim} \mathsf{Ber}(\alpha)$,
- $\boldsymbol{B}^{(N)} = \frac{M}{N}\boldsymbol{V}^{(N)*}\boldsymbol{V}^{(N)}$ for $\boldsymbol{V}^{(N)} = [\,\boldsymbol{v}_1 \cdots \boldsymbol{v}_N\,] \in \mathbb{C}^{M\times N}$ (deterministic!) tight frames having $\frac{M}{N} \to \beta \in (0,1)$ and

$$\max_{i,j\in[N]} \left|\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle - \mathbb{1}\{i=j\}\right| \leq N^{-1/2+o(1)}.$$

Answers signal processing and combinatorics questions:

- Proves conjecture of [Haikin, Zamir, Gavish '17]
- Simplifies [Mixon, Magsino, Parshall '21] (Paley frames)
- Simplifies [Farrell '11] (Fourier frames)

Application: Random Subsets of Frames

Main Theorem 1 applies with:

- $\boldsymbol{A}^{(N)}$ diagonal, $A_{ii} \overset{\text{iid}}{\sim} \mathsf{Ber}(\alpha)$,
- $\boldsymbol{B}^{(N)} = \frac{M}{N} \boldsymbol{V}^{(N)*} \boldsymbol{V}^{(N)}$ for $\boldsymbol{V}^{(N)} = [\, \boldsymbol{v}_1 \cdots \boldsymbol{v}_N \,] \in \mathbb{C}^{M \times N}$ (deterministic!) tight frames having $\frac{M}{N} \to \beta \in (0, 1)$ and

$$\max_{i,j \in [N]} \left| \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle - \mathbb{1}\{i = j\} \right| \leq N^{-1/2 + o(1)}.$$

Answers signal processing and combinatorics questions:

- Proves conjecture of [Haikin, Zamir, Gavish '17]
- Simplifies [Mixon, Magsino, Parshall '21] (Paley frames)
- Simplifies [Farrell '11] (Fourier frames)

...with easy proof!

Proof Sketch

Just expand and bound naively:

$$\frac{1}{N}\left|\mathbb{E}\,\mathrm{Tr}\left((\boldsymbol{A}-\alpha\boldsymbol{I})(\boldsymbol{B}-\beta\boldsymbol{I})\right)^k\right|$$

$$\leq \frac{1}{N}\sum_{i_1,\ldots,i_k=1}^{N} \left|\mathbb{E}[(A_{i_1 i_1}-\alpha)\cdots(A_{i_k i_k}-\alpha)]\right|$$

$$|B_{i_1 i_2}-\beta\mathbb{1}\{i_1=i_2\}|\cdots|B_{i_k i_1}-\beta\mathbb{1}\{i_k=i_1\}|$$

$$\lesssim_k N^{-1}\cdot\underbrace{N^{k/2}}_{\#\text{ non-zero terms}}\cdot N^{-k/2+o(1)}$$

$$\to 0.$$

Proof Sketch

Just expand and bound naively:

$$\frac{1}{N}|\mathbb{E}\operatorname{Tr}\left((\boldsymbol{A}-\alpha\boldsymbol{I})(\boldsymbol{B}-\beta\boldsymbol{I})\right)^k|$$

$$\leq \frac{1}{N}\sum_{i_1,\dots,i_k=1}^{N} |\mathbb{E}[(A_{i_1 i_1}-\alpha)\cdots(A_{i_k i_k}-\alpha)]|$$

$$|B_{i_1 i_2}-\beta\mathbb{1}\{i_1=i_2\}|\cdots|B_{i_k i_1}-\beta\mathbb{1}\{i_k=i_1\}|$$

$$\lesssim_k N^{-1} \cdot \underbrace{N^{k/2}}_{\text{\# non-zero terms}} \cdot N^{-k/2+o(1)}$$

$$\to 0.$$

Previous work finds a “main term” and “error term” in $\frac{1}{N}\mathbb{E}\operatorname{Tr}(\boldsymbol{ABA})^k$ directly, redoing free probability by hand.

Aside: Free Probability with "Less Randomness"

Classical results (Voiculescu et al.) treat **"very random"** models: $\boldsymbol{A}, \boldsymbol{B}$ **unitarily invariant** projections, i.e., to uniformly random subspaces of $\mathbb{C}^N$.

Aside: Free Probability with "Less Randomness"

Classical results (Voiculescu et al.) treat **"very random"** models: $\boldsymbol{A}, \boldsymbol{B}$ **unitarily invariant** projections, i.e., to uniformly random subspaces of $\mathbb{C}^N$.

This line of work: the same freeness and spectral limits hold for **much less random** models (e.g., $\boldsymbol{B}$ deterministic and $\boldsymbol{A}$ diagonal).

Aside: Free Probability with “Less Randomness”

Classical results (Voiculescu et al.) treat **“very random”** models: $\boldsymbol{A}, \boldsymbol{B}$ **unitarily invariant** projections, i.e., to uniformly random subspaces of $\mathbb{C}^N$.

This line of work: the same freeness and spectral limits hold for **much less random** models (e.g., $\boldsymbol{B}$ deterministic and $\boldsymbol{A}$ diagonal).

Companion work [K ’23]: some cases where the same limits hold for **completely deterministic** models, e.g., built on the Paley frames and Paley graph of number theory.

Asymptotic Freeness in Paley Graphs

G_p a graph on vertices $\mathbb{Z}/p\mathbb{Z}$ (with $p \equiv 1 \bmod 4$) with $i \sim j$ iff $j - i$ is a **square** mod p (for some $x \neq 0$, $j - i \equiv x^2$).

G_p is strongly regular; fits in previous framework.

Asymptotic Freeness in Paley Graphs

G_p a graph on vertices $\mathbb{Z}/p\mathbb{Z}$ (with $p \equiv 1 \bmod 4$) with $i \sim j$ iff $j - i$ is a **square** mod p (for some $x \neq 0$, $j - i \equiv x^2$).

G_p is strongly regular; fits in previous framework.

Theorem: [K '23] The associated projection $\boldsymbol{B}$ is asymptotically free of coordinate projections $\boldsymbol{A}_S$ for any

$$S = \{j : j \sim i \text{ in } G_p\} = \text{“neighborhood of } i\text{”}.$$

So, e.s.d. $\to \mu_{1/2,1/2}$ = **arcsine law**. (Partial generalization to other “low-degree” sets S.)

Asymptotic Freeness in Paley Graphs

G_p a graph on vertices $\mathbb{Z}/p\mathbb{Z}$ (with $p \equiv 1 \bmod 4$) with $i \sim j$ iff $j - i$ is a **square** mod p (for some $x \neq 0$, $j - i \equiv x^2$).

G_p is strongly regular; fits in previous framework.

Theorem: [K '23] The associated projection $\boldsymbol{B}$ is asymptotically free of coordinate projections $\boldsymbol{A}_S$ for any

$$S = \{j : j \sim i \text{ in } G_p\} = \text{"neighborhood of } i\text{"}.$$

So, e.s.d. $\to \mu_{1/2,1/2}$ = **arcsine law**. (Partial generalization to other "low-degree" sets S.)

Moral: Freeness does not require probability!

III. The Largest Eigenvalue

Conjecture and Difficulties

Natural extension of previous results:

Edge Conjecture: For many $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ with e.s.d. of $\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)}$ converging weakly to $\mu_{\alpha,\beta}$,

$$\lambda_{\mathsf{max}}(\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)}) \xrightarrow{\text{(p)}} \mathsf{edge}(\alpha,\beta) := \text{right edge of } \mu_{\alpha,\beta}.$$

Conjecture and Difficulties

Natural extension of previous results:

Edge Conjecture: For many $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ with e.s.d. of $\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)}$ converging weakly to $\mu_{\alpha,\beta}$,

$$\lambda_{\mathsf{max}}(\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)}) \xrightarrow{(\mathsf{p})} \mathsf{edge}(\alpha,\beta) := \text{right edge of } \mu_{\alpha,\beta}.$$

As in models from classical random matrix theory (e.g., Wigner and Wishart), the difficulty is in controlling moments of diverging order:

$$\mathbb{E}\,\mathsf{Tr}(\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)})^k \overset{?}{\leq} (\mathsf{edge}(\alpha,\beta) + o(1))^k$$

for $k \gg \log(N)$.

Main Theorem 2

For $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ as in Main Theorem 1, suppose additionally that one of α or β is $\frac{1}{2}$ and that for $k = k(N) \gg \log(N)$,

$$\max_{1 \le a \le k} \left| \mathbb{E} \operatorname{Tr} \left(\frac{\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N}{\sqrt{\alpha(1-\alpha)}} \frac{\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N}{\sqrt{\beta(1-\beta)}} \right)^a \right| \le \exp(o(k)).$$

Then,

$$\lambda_{\max}(\boldsymbol{A}^{(N)} \boldsymbol{B}^{(N)} \boldsymbol{A}^{(N)}) \xrightarrow{(\mathrm{p})} \mathrm{edge}(\alpha, \beta).$$

Main Theorem 2

For $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ as in Main Theorem 1, suppose additionally that one of α or β is $\frac{1}{2}$ and that for $k = k(N) \gg \log(N)$,

$$\max_{1 \le a \le k} \left| \mathbb{E}\,\mathrm{Tr}\left(\frac{\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N}{\sqrt{\alpha(1-\alpha)}} \frac{\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N}{\sqrt{\beta(1-\beta)}} \right)^a \right| \le \exp(o(k)).$$

Then,

$$\lambda_{\max}(\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)}) \xrightarrow{(p)} \mathrm{edge}(\alpha, \beta).$$

Similar idea to Main Theorem 1: isolate an **error term,** which should be easier to control than $\mathrm{Tr}(\boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}\boldsymbol{A}^{(N)})^k$.

Main Theorem 2

For $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ as in Main Theorem 1, suppose additionally that one of α or β is $\frac{1}{2}$ and that for $k = k(N) \gg \log(N)$,

$$\max_{1 \le a \le k} \left| \mathbb{E} \operatorname{Tr} \left(\frac{\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N}{\sqrt{\alpha(1-\alpha)}} \frac{\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N}{\sqrt{\beta(1-\beta)}} \right)^a \right| \le \exp(o(k)).$$

Then,

$$\lambda_{\max}(\boldsymbol{A}^{(N)} \boldsymbol{B}^{(N)} \boldsymbol{A}^{(N)}) \xrightarrow{(p)} \operatorname{edge}(\alpha, \beta).$$

Similar idea to Main Theorem 1: isolate an **error term**, which should be easier to control than $\operatorname{Tr}(\boldsymbol{A}^{(N)} \boldsymbol{B}^{(N)} \boldsymbol{A}^{(N)})^k$.

Conjecture: Extra condition on α, β not necessary.

Seems just a technical challenge—stay tuned for details.

Main Theorem 2: Intuition

Why do these normalizations appear?

$$\widehat{\boldsymbol{A}} := (\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N)/\sqrt{\alpha(1-\alpha)}, \quad \widehat{\boldsymbol{B}} := (\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N)/\sqrt{\beta(1-\beta)}$$

Main Theorem 2: Intuition

Why do these normalizations appear?

$$\widehat{\boldsymbol{A}} := (\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N)/\sqrt{\alpha(1-\alpha)}, \quad \widehat{\boldsymbol{B}} := (\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N)/\sqrt{\beta(1-\beta)}$$

Normalization whitens the spectrum: if $\lambda_i(\boldsymbol{A}^{(N)}) \overset{\text{iid}}{\sim} \mathsf{Ber}(\alpha)$ and $\lambda_i(\boldsymbol{B}^{(N)}) \overset{\text{iid}}{\sim} \mathsf{Ber}(\beta)$, then

$$\mathbb{E}\lambda_i(\widehat{\boldsymbol{A}}) = \mathbb{E}\lambda_i(\widehat{\boldsymbol{B}}) = 0,$$
$$\mathbb{E}\lambda_i(\widehat{\boldsymbol{A}})^2 = \mathbb{E}\lambda_i(\widehat{\boldsymbol{B}})^2 = 1.$$

$\rightsquigarrow$ $\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}$ are **orthogonal in expectation**.

Main Theorem 2: Intuition

Why do these normalizations appear?

$$\widehat{\boldsymbol{A}} := (\boldsymbol{A}^{(N)} - \alpha \boldsymbol{I}_N) / \sqrt{\alpha(1-\alpha)}, \quad \widehat{\boldsymbol{B}} := (\boldsymbol{B}^{(N)} - \beta \boldsymbol{I}_N) / \sqrt{\beta(1-\beta)}$$

Normalization whitens the spectrum: if $\lambda_i(\boldsymbol{A}^{(N)}) \overset{\text{iid}}{\sim} \mathsf{Ber}(\alpha)$ and $\lambda_i(\boldsymbol{B}^{(N)}) \overset{\text{iid}}{\sim} \mathsf{Ber}(\beta)$, then

$$\mathbb{E}\lambda_i(\widehat{\boldsymbol{A}}) = \mathbb{E}\lambda_i(\widehat{\boldsymbol{B}}) = 0,$$
$$\mathbb{E}\lambda_i(\widehat{\boldsymbol{A}})^2 = \mathbb{E}\lambda_i(\widehat{\boldsymbol{B}})^2 = 1.$$

$\rightsquigarrow$ $\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}$ are **orthogonal in expectation**. Actually orthogonal matrices would indeed satisfy

$$|\operatorname{Tr}(\widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}})^k| \le N = \exp(o(k)).$$

Proof Ideas for Main Theorem 2

Exact relation between centered and uncentered moments by "trace rewriting" using idempotence:

$$\mathsf{Tr}\left((\boldsymbol{A} - \alpha \boldsymbol{I}_N)(\boldsymbol{B} - \beta \boldsymbol{I}_N)\right)^k$$

$$= w\,\mathsf{Tr}(\boldsymbol{I}_N) + x\,\mathsf{Tr}(\boldsymbol{A}) + y\,\mathsf{Tr}(\boldsymbol{B}) + \sum_{\ell=1}^{k} z_\ell\,\mathsf{Tr}(\boldsymbol{ABA})^\ell.$$

Proof Ideas for Main Theorem 2

Exact relation between centered and uncentered moments by "trace rewriting" using idempotence:

$$\operatorname{Tr}\left((\boldsymbol{A} - \alpha \boldsymbol{I}_N)(\boldsymbol{B} - \beta \boldsymbol{I}_N)\right)^k$$

$$= w \operatorname{Tr}(\boldsymbol{I}_N) + x \operatorname{Tr}(\boldsymbol{A}) + y \operatorname{Tr}(\boldsymbol{B}) + \sum_{\ell=1}^{k} z_\ell \operatorname{Tr}(\boldsymbol{ABA})^\ell.$$

Inverting this, the "main term" is the limiting MANOVA moment:

$$\operatorname{Tr}(\boldsymbol{ABA})^k = N \mathop{\mathbb{E}}_{\lambda \sim \mu_{\alpha,\beta}} \lambda^k + x'(\operatorname{Tr}(\boldsymbol{A}) - \alpha N) + y'(\operatorname{Tr}(\boldsymbol{B}) - \beta N)$$

$$+ \sum_{\ell=1}^{k} z'_\ell \operatorname{Tr}\left((\boldsymbol{A} - \alpha \boldsymbol{I}_N)(\boldsymbol{B} - \beta \boldsymbol{I}_N)\right)^k.$$

Proof Ideas for Main Theorem 2

First result: an explicit recursion for the moment errors

$$\Delta_k = \mathrm{Tr}(\boldsymbol{ABA})^k - N \mathop{\mathbb{E}}_{\lambda \sim \mu_{\alpha,\beta}} \lambda^k$$
$$\approx \sum_{\ell=1}^{k-1} c_{k,\ell} \Delta_\ell + c'_k \, \mathrm{Tr}\left((\boldsymbol{A} - \alpha \boldsymbol{I}_N)(\boldsymbol{B} - \beta \boldsymbol{I}_N)\right)^k.$$

Still not easy to analyze for large k...

Proof Ideas for Main Theorem 2

First result: an explicit recursion for the moment errors

$$\begin{aligned}\Delta_k &= \mathsf{Tr}(\boldsymbol{ABA})^k - N \mathop{\mathbb{E}}_{\lambda\sim\mu_{\alpha,\beta}} \lambda^k \\ &\approx \sum_{\ell=1}^{k-1} c_{k,\ell}\Delta_\ell + c_k' \,\mathsf{Tr}\left((\boldsymbol{A}-\alpha\boldsymbol{I}_N)(\boldsymbol{B}-\beta\boldsymbol{I}_N)\right)^k.\end{aligned}$$

Still not easy to analyze for large k...

Saving grace: if α or β is $\frac{1}{2}$, can solve this recursion in closed form!

Identify **Riordan arrays** in recursion: triangular matrices with special generating functions allowing formal inversion.

“Application:” New Proof for Invariant Model

New, arguably **more robust**, proof of edge limit theorem for $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ unitarily invariant projections.

"Application:" New Proof for Invariant Model

New, arguably **more robust**, proof of edge limit theorem for $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ unitarily invariant projections.

Model equivalent to:

$$\begin{aligned} \boldsymbol{A} \text{ diagonal with } A_{ii} &\overset{\text{iid}}{\sim} \mathsf{Ber}(\alpha), \\ \boldsymbol{D} \text{ diagonal with } D_{ii} &\overset{\text{iid}}{\sim} \mathsf{Ber}(\beta), \\ \boldsymbol{U} &\sim \mathsf{Haar}(\mathcal{U}(N)), \\ \boldsymbol{B} &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*. \end{aligned}$$

"Application:" New Proof for Invariant Model

New, arguably **more robust**, proof of edge limit theorem for $\boldsymbol{A}^{(N)}, \boldsymbol{B}^{(N)}$ unitarily invariant projections.

Model equivalent to:

$$\begin{aligned} \boldsymbol{A} \text{ diagonal with } A_{ii} &\overset{\text{iid}}{\sim} \mathsf{Ber}(\alpha), \\ \boldsymbol{D} \text{ diagonal with } D_{ii} &\overset{\text{iid}}{\sim} \mathsf{Ber}(\beta), \\ \boldsymbol{U} &\sim \mathsf{Haar}(\mathcal{U}(N)), \\ \boldsymbol{B} &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*. \end{aligned}$$

Proof: expand; use non-asymptotic bounds for Weingarten function [Collins, Matsumoto '17] to control moments

$$\mathbb{E}[U_{i_1 j_1} \cdots U_{i_k j_k} \overline{U_{i'_1 j'_1}} \cdots \overline{U_{i'_{k'} j'_{k'}}}].$$

Obstacles to General Models

Consider same $\boldsymbol{A}$ but $\boldsymbol{B}$ deterministic (like in frame application) and real symmetric.

Obstacles to General Models

Consider same $\boldsymbol{A}$ but $\boldsymbol{B}$ deterministic (like in frame application) and real symmetric.

No expectation over $\boldsymbol{B}$; expanding traces leads to trying to analyze **"graphical moments"** of $\boldsymbol{B}$:

From a graph $G = ([k], E)$ and $\boldsymbol{B} \in \mathbb{R}^{N \times N}_{\mathsf{sym}}$, compute

$$\sum_{i_1, \dots, i_k \in [N] \text{ distinct}} \prod_{\{a,b\} \in E} B_{i_a i_b}.$$

Scalar version of **graph matrices** appearing in literature on **sum-of-squares optimization**.

Techniques to analyze for general G and deterministic $\boldsymbol{B}$?

Paley Graphs Revisited

Numerical experiments $\rightsquigarrow$ Edge Conjecture should hold for Paley graph construction.

Paley Graphs Revisited

Numerical experiments $\rightsquigarrow$ Edge Conjecture should hold for Paley graph construction.

Proof would probably require intricate analysis of character sums ($\approx$ graphical moments in ± 1 adjacency matrix).

Paley Graphs Revisited

Numerical experiments $\rightsquigarrow$ Edge Conjecture should hold for Paley graph construction.

Proof would probably require intricate analysis of character sums ($\approx$ graphical moments in ± 1 adjacency matrix).

But, by enhanced spectral bound on clique number, would yield progress on long-standing open problem:

Conjecture: Largest clique in G_p is $O(\mathrm{polylog}(p))$.

Best known upper bound: Largest clique in G_p is $\leq \sqrt{p/2}$.

Paley Graphs Revisited

Numerical experiments $\rightsquigarrow$ Edge Conjecture should hold for Paley graph construction.

Proof would probably require intricate analysis of character sums ($\approx$ graphical moments in ± 1 adjacency matrix).

But, by enhanced spectral bound on clique number, would yield progress on long-standing open problem:

Conjecture: Largest clique in G_p is $O(\text{polylog}(p))$.

Best known upper bound: Largest clique in G_p is $\leq \sqrt{p/2}$.

Theorem: [K '23] If Edge Conjecture holds for Paley graph construction, then largest clique in G_p is $\leq o(\sqrt{p})$.

Open Questions

Open Questions

- Prove $\lambda_{\max}$ limit theorem for **any** non-invariant model.
    - Invariant model with discrete Fourier matrix instead of $\boldsymbol{U}$: unsolved since [Farrell '11].
    - Deterministic subsets of Paley frames $\Rightarrow$ improved bounds for clique number of Paley graph.

Open Questions

- Prove λ_{max} limit theorem for **any** non-invariant model.
  - Invariant model with discrete Fourier matrix instead of $\boldsymbol{U}$: unsolved since [Farrell '11].
  - Deterministic subsets of Paley frames $\Rightarrow$ improved bounds for clique number of Paley graph.
- Other universal features:
  - Local laws and spacing? [Farrell, Nadakuditi '15]
  - Rate of convergence of e.s.d.?

Open Questions

- Prove $\lambda_{\max}$ limit theorem for **any** non-invariant model.
  - Invariant model with discrete Fourier matrix instead of $\boldsymbol{U}$: unsolved since [Farrell '11].
  - Deterministic subsets of Paley frames $\Rightarrow$ improved bounds for clique number of Paley graph.
- Other universal features:
  - Local laws and spacing? [Farrell, Nadakuditi '15]
  - Rate of convergence of e.s.d.?
- Adapt to analyze RIP (**all** small submatrices)?

Thank you!