Principal components and linear mixed models

Zhou Fan

Yale University, Statistics and Data Science

(joint w/ Iain Johnstone, Yi Sun, Zhichao Wang)

Random Matrices and Related Topics
Korea Institute for Advanced Study, Seoul

May 6, 2019

Linear mixed models capture multiple “levels” of variation in data. They were introduced by R. A. Fisher in 1918 to study genetic and non-genetic components of variance in quantitative traits.

Linear mixed models capture multiple “levels” of variation in data. They were introduced by R. A. Fisher in 1918 to study genetic and non-genetic components of variance in quantitative traits.

This talk will describe some applications of random matrix theory to understand spectral behavior and principal components analysis for classical covariance estimates in these models.

Outline

Model and motivation

Example: A twin study

Measure p quantitative traits in $n/2$ pairs of twins. For $i = 1, \ldots, n/2$, model this with two "levels" of variation as

$$Y_{i,1} = \alpha_i + \varepsilon_{i,1} \in \mathbb{R}^p$$
$$Y_{i,2} = \alpha_i + \varepsilon_{i,2} \in \mathbb{R}^p$$

Example: A twin study

Measure p quantitative traits in $n/2$ pairs of twins. For $i = 1, \dots, n/2$, model this with two "levels" of variation as

$$Y_{i,1} = \alpha_i + \varepsilon_{i,1} \in \mathbb{R}^p$$
$$Y_{i,2} = \alpha_i + \varepsilon_{i,2} \in \mathbb{R}^p$$

Here, $\alpha_i \in \mathbb{R}^p$ is the shared genetic effect in the i^{th} twin pair, and $\varepsilon_{i,1}, \varepsilon_{i,2} \in \mathbb{R}^p$ are individual variations.

Example: A twin study

Measure p quantitative traits in $n/2$ pairs of twins. For $i = 1, \ldots, n/2$, model this with two "levels" of variation as

$$Y_{i,1} = \alpha_i + \varepsilon_{i,1} \in \mathbb{R}^p$$
$$Y_{i,2} = \alpha_i + \varepsilon_{i,2} \in \mathbb{R}^p$$

Here, $\alpha_i \in \mathbb{R}^p$ is the shared genetic effect in the i^{th} twin pair, and $\varepsilon_{i,1}, \varepsilon_{i,2} \in \mathbb{R}^p$ are individual variations.

Assume these are random and independent,

$$\alpha_i \overset{iid}{\sim} \mathcal{N}(0, \Sigma_A), \qquad \varepsilon_{i,j} \overset{iid}{\sim} \mathcal{N}(0, \Sigma_E)$$

Example: A twin study

Measure p quantitative traits in $n/2$ pairs of twins. For $i = 1, \ldots, n/2$, model this with two "levels" of variation as

$$Y_{i,1} = \alpha_i + \varepsilon_{i,1} \in \mathbb{R}^p$$
$$Y_{i,2} = \alpha_i + \varepsilon_{i,2} \in \mathbb{R}^p$$

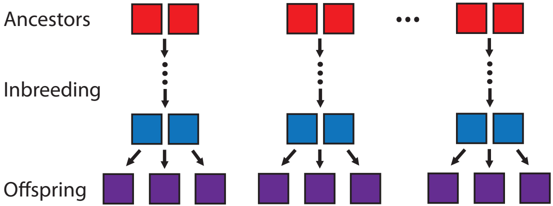
Here, $\alpha_i \in \mathbb{R}^p$ is the shared genetic effect in the i^{th} twin pair, and $\varepsilon_{i,1}, \varepsilon_{i,2} \in \mathbb{R}^p$ are individual variations.

Assume these are random and independent,

$$\alpha_i \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_A), \qquad \varepsilon_{i,j} \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_E)$$

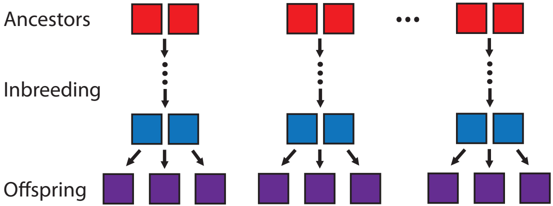
Only the $Y_{i,j}$'s (not the α_i's or $\varepsilon_{i,j}$'s) are observed. From this, we wish to separately understand Σ_A and Σ_E.

Example: Mutations in fruit flies [McGuigan et al '14]



In inbred lines of fruit lines, how much phenotypic variation arises due to genetic mutations across the generations?

Example: Mutations in fruit flies [McGuigan et al '14]



In inbred lines of fruit lines, how much phenotypic variation arises due to genetic mutations across the generations?

Model traits (gene expression measurements) in the j^{th} offspring of the i^{th} inbred line as

$$Y_{i,j} = \alpha_i + \varepsilon_{i,j}.$$

The covariance Σ_A of α_i's is the mutational variation of interest.

Example: Genome-wide association studies

In n individuals, measure:

- p quantitative traits, $Y \in \mathbb{R}^{n \times p}$
- genotypes $\{0, 1, 2\}$ at m SNPs, $X \in \mathbb{R}^{n \times m}$

Example: Genome-wide association studies

In n individuals, measure:

- p quantitative traits, $Y \in \mathbb{R}^{n \times p}$
- genotypes $\{0, 1, 2\}$ at m SNPs, $X \in \mathbb{R}^{n \times m}$

Fisher's infinitesimal model:

$$Y = XA + E$$

Example: Genome-wide association studies

In n individuals, measure:

- p quantitative traits, $Y \in \mathbb{R}^{n \times p}$
- genotypes $\{0, 1, 2\}$ at m SNPs, $X \in \mathbb{R}^{n \times m}$

Fisher's infinitesimal model:

$$Y = XA + E$$

- $A \in \mathbb{R}^{m \times p}$ has independent rows $\alpha_1, \ldots, \alpha_m$. Each $\alpha_i \in \mathbb{R}^p$ is the contribution of the i^{th} SNP to the observed traits.

Example: Genome-wide association studies

In n individuals, measure:

- p quantitative traits, $Y \in \mathbb{R}^{n \times p}$
- genotypes $\{0, 1, 2\}$ at m SNPs, $X \in \mathbb{R}^{n \times m}$

Fisher's infinitesimal model:

$$Y = XA + E$$

- $A \in \mathbb{R}^{m \times p}$ has independent rows $\alpha_1, \ldots, \alpha_m$. Each $\alpha_i \in \mathbb{R}^p$ is the contribution of the i^{th} SNP to the observed traits.
- $E \in \mathbb{R}^{n \times p}$ has independent rows $\varepsilon_1, \ldots, \varepsilon_n$. Each $\varepsilon_j \in \mathbb{R}^p$ is the residual trait variation in the j^{th} individual.

Example: Genome-wide association studies

In n individuals, measure:

- p quantitative traits, $Y \in \mathbb{R}^{n \times p}$
- genotypes $\{0, 1, 2\}$ at m SNPs, $X \in \mathbb{R}^{n \times m}$

Fisher's infinitesimal model:

$$Y = XA + E$$

- $A \in \mathbb{R}^{m \times p}$ has independent rows $\alpha_1, \ldots, \alpha_m$. Each $\alpha_i \in \mathbb{R}^p$ is the contribution of the i^{th} SNP to the observed traits.
- $E \in \mathbb{R}^{n \times p}$ has independent rows $\varepsilon_1, \ldots, \varepsilon_n$. Each $\varepsilon_j \in \mathbb{R}^p$ is the residual trait variation in the j^{th} individual.

The covariance Σ_A of α_i's is the (additive) genetic covariance.
The relative size of Σ_A to Σ_E provides a measure of heritability.

The linear mixed model

A general model with k levels of variation is

$$Y = U_1 A_1 + \ldots + U_k A_k \in \mathbb{R}^{n \times p}$$

The linear mixed model

A general model with k levels of variation is

$$Y = U_1 A_1 + \ldots + U_k A_k \in \mathbb{R}^{n \times p}$$

- $A_1, \ldots, A_k$ are random and unobserved, with $n_1, \ldots, n_k$ independent rows distributed as $\mathcal{N}(0, \Sigma_1), \ldots, \mathcal{N}(0, \Sigma_k)$.

The linear mixed model

A general model with k levels of variation is

$$Y = U_1 A_1 + \ldots + U_k A_k \in \mathbb{R}^{n \times p}$$

- $A_1, \ldots, A_k$ are random and unobserved, with $n_1, \ldots, n_k$ independent rows distributed as $\mathcal{N}(0, \Sigma_1), \ldots, \mathcal{N}(0, \Sigma_k)$.
- $U_1, \ldots, U_k$ are known, deterministic, and specified by the experimental design. E.g. for the twin study, $k = 2$ and

$$U_1 = \begin{pmatrix} 1 & & \\ 1 & & \\ & \ddots & \\ & & 1 \\ & & 1 \end{pmatrix}, \qquad U_2 = \mathrm{Id}$$

The linear mixed model

A general model with k levels of variation is

$$Y = U_1 A_1 + \ldots + U_k A_k \in \mathbb{R}^{n \times p}$$

- $A_1, \ldots, A_k$ are random and unobserved, with $n_1, \ldots, n_k$ independent rows distributed as $\mathcal{N}(0, \Sigma_1), \ldots, \mathcal{N}(0, \Sigma_k)$.
- $U_1, \ldots, U_k$ are known, deterministic, and specified by the experimental design. E.g. for the twin study, $k = 2$ and

$$U_1 = \begin{pmatrix} 1 & & \\ 1 & & \\ & \ddots & \\ & & 1 \\ & & 1 \end{pmatrix}, \qquad U_2 = \mathrm{Id}$$

($k = 1$, $U_1 = \mathrm{Id}$ is the setting of n independent observations in $\mathbb{R}^p$)

The MANOVA covariance estimator

For $r \in \{1, \ldots, k\}$, a classical estimator for Σ_r is the MANOVA estimator. This is a matrix

$$\widehat{\Sigma} = Y^{\mathsf{T}} B Y.$$

Here, $B \in \mathbb{R}^{n \times n}$ is symmetric and chosen so that $\mathbb{E}[\widehat{\Sigma}] = \Sigma_r$.

The MANOVA covariance estimator

For $r \in \{1, \ldots, k\}$, a classical estimator for Σ_r is the MANOVA estimator. This is a matrix

$$\widehat{\Sigma} = Y^{\mathsf{T}} B Y.$$

Here, $B \in \mathbb{R}^{n \times n}$ is symmetric and chosen so that $\mathbb{E}[\widehat{\Sigma}] = \Sigma_r$.

Some examples:

- For $k = 1$ and independent observations, we take $B = \frac{1}{n} I$. This gives the usual sample covariance matrix $\widehat{\Sigma} = \frac{1}{n} Y^{\mathsf{T}} Y$.

The MANOVA covariance estimator

For $r \in \{1, \ldots, k\}$, a classical estimator for Σ_r is the MANOVA estimator. This is a matrix

$$\widehat{\Sigma} = Y^{\mathsf{T}} B Y.$$

Here, $B \in \mathbb{R}^{n \times n}$ is symmetric and chosen so that $\mathbb{E}[\widehat{\Sigma}] = \Sigma_r$.

Some examples:

- For $k = 1$ and independent observations, we take $B = \frac{1}{n} I$. This gives the usual sample covariance matrix $\widehat{\Sigma} = \frac{1}{n} Y^{\mathsf{T}} Y$.
- For $k = 2$ and the twin study, we take $B = \frac{1}{n}(\pi - \pi^{\perp})$ where $\pi, \pi^{\perp}$ are orthogonal projections onto the column span of U_1 and its complement.

The MANOVA covariance estimator

Substituting $Y = \sum_r U_r A_r$, we may express the estimator as

$$\widehat{\Sigma} = \sum_{r=1}^{k} \sum_{s=1}^{k} H_r^\mathsf{T} G_r^\mathsf{T} F_{rs} G_s H_s$$

- $H_r \equiv \Sigma_r^{1/2}$ and $F_{rs} \equiv U_r^\mathsf{T} B U_s$ are deterministic
- G_r are independent and random, with i.i.d. Gaussian entries

The MANOVA covariance estimator

Substituting $Y = \sum_r U_r A_r$, we may express the estimator as

$$\widehat{\Sigma} = \sum_{r=1}^{k} \sum_{s=1}^{k} H_r^\top G_r^\top F_{rs} G_s H_s$$

- $H_r \equiv \Sigma_r^{1/2}$ and $F_{rs} \equiv U_r^\top B U_s$ are deterministic
- G_r are independent and random, with i.i.d. Gaussian entries

This is the random matrix model that I'll discuss.

The MANOVA covariance estimator

Substituting $Y = \sum_r U_r A_r$, we may express the estimator as

$$\widehat{\Sigma} = \sum_{r=1}^{k} \sum_{s=1}^{k} H_r^\mathsf{T} G_r^\mathsf{T} F_{rs} G_s H_s$$

- $H_r \equiv \Sigma_r^{1/2}$ and $F_{rs} \equiv U_r^\mathsf{T} B U_s$ are deterministic
- G_r are independent and random, with i.i.d. Gaussian entries

This is the random matrix model that I'll discuss.

1. What is the bulk eigenvalue distribution for large $n, n_1, \ldots, n_k, p$?

The MANOVA covariance estimator

Substituting $Y = \sum_r U_r A_r$, we may express the estimator as

$$\widehat{\Sigma} = \sum_{r=1}^{k} \sum_{s=1}^{k} H_r^\top G_r^\top F_{rs} G_s H_s$$

- $H_r \equiv \Sigma_r^{1/2}$ and $F_{rs} \equiv U_r^\top B U_s$ are deterministic
- G_r are independent and random, with i.i.d. Gaussian entries

This is the random matrix model that I'll discuss.

1. What is the bulk eigenvalue distribution for large $n, n_1, \ldots, n_k, p$?
2. What is the behavior of principal components in spiked settings?

Aside: The case of isotropic noise

A simple statistical null hypothesis in this model is

$$\Sigma_r = \sigma_r^2 \, \mathrm{Id}$$

for every $r \in \{1, \ldots, k\}$, i.e. the distribution of every random effect is isotropic noise.

Aside: The case of isotropic noise

A simple statistical null hypothesis in this model is

$$\Sigma_r = \sigma_r^2 \,\mathrm{Id}$$

for every $r \in \{1, \ldots, k\}$, i.e. the distribution of every random effect is isotropic noise.

This setting is special: We may write $\widehat{\Sigma} = G^\top F G$ where

$$G = \begin{pmatrix} G_1 \\ \vdots \\ G_k \end{pmatrix}, \quad F = (\sigma_r F_{rs} \sigma_s)_{r,s=1}^{k}.$$

Aside: The case of isotropic noise

A simple statistical null hypothesis in this model is

$$\Sigma_r = \sigma_r^2 \, \mathrm{Id}$$

for every $r \in \{1, \ldots, k\}$, i.e. the distribution of every random effect is isotropic noise.

This setting is special: We may write $\widehat{\Sigma} = G^\top F G$ where

$$G = \begin{pmatrix} G_1 \\ \vdots \\ G_k \end{pmatrix}, \quad F = (\sigma_r F_{rs} \sigma_s)_{r,s=1}^k.$$

Here, $\widehat{\Sigma}$ is related to the sample covariance model. We've obtained more refined results in this setting than in the general case, due a known local law [Knowles, Yin '17].

Results on spectral behavior

Empirical spectral distribution

Theorem (Fan, Johnstone '16)

As $n, n_1, \ldots, n_k, p \to \infty$ *proportionally, the e.s.d.* $\hat{\mu}$ *of* $\widehat{\Sigma}$ *is approximated (weakly a.s.) by a deterministic equivalent measure* μ_0*.*

Empirical spectral distribution

Theorem (Fan, Johnstone '16)

As $n, n_1, \ldots, n_k, p \to \infty$ proportionally, the e.s.d. $\hat{\mu}$ of $\widehat{\Sigma}$ is approximated (weakly a.s.) by a deterministic equivalent measure μ_0. This has Stieltjes transform $m_0(z)$ characterized by

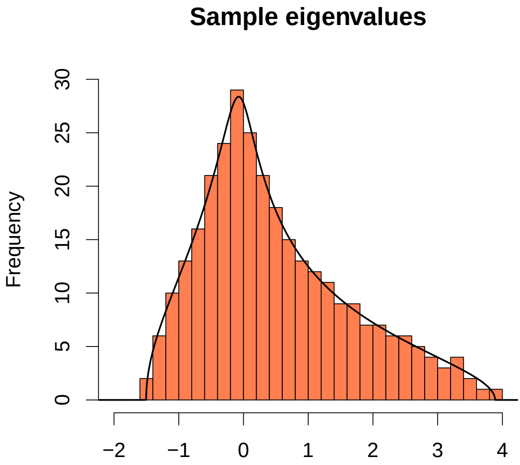
$$a_r(z) = -n_r^{-1} \operatorname{Tr} \left((z \operatorname{Id} + b \cdot \Sigma)^{-1} \Sigma_r \right)$$
$$b_r(z) = -n_r^{-1} \operatorname{Tr}_r \left((\operatorname{Id} + FD(a))^{-1} F \right)$$
$$m_0(z) = -p^{-1} \operatorname{Tr} \left((z \operatorname{Id} + b \cdot \Sigma)^{-1} \right)$$

for $r = 1, \ldots, k$. Here,

$$b \cdot \Sigma \equiv b_1(z)\Sigma_1 + \ldots + b_k(z)\Sigma_k, \qquad F \equiv (F_{rs})_{r,s=1}^k$$
$$D(a) \equiv \operatorname{diag}(a_1(z) \operatorname{Id}_{n_1}, \ldots, a_k(z) \operatorname{Id}_{n_k}),$$

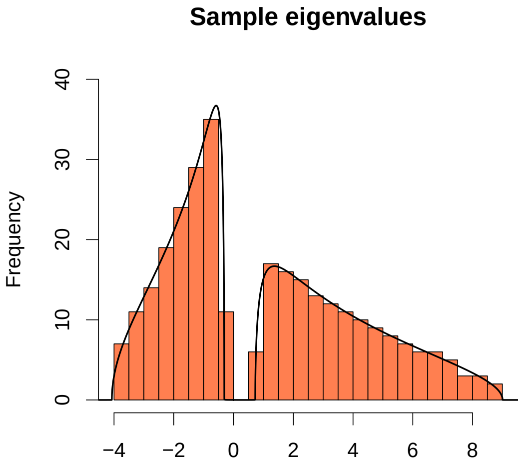
and Tr_r is the trace of the (r, r) block, of size n_r.

Empirical spectral distribution



Spectrum of $\widehat{\Sigma}_A$ in a twin study design, 300 twin pairs, 300 traits.

Empirical spectral distribution



Spectrum of $\widehat{\Sigma}_A$ in a twin study design, 150 twin pairs, 600 traits.

A free approximation

The measure μ_0 is the τ-law of an operator

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

in a non-commutative probability space $(\mathcal{A}, \tau)$:

A free approximation

The measure μ_0 is the τ-law of an operator

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

in a non-commutative probability space $(\mathcal{A}, \tau)$:

- $\{h_r\}_{r=1}^k$ and $\{f_{rs}\}_{r,s=1}^k$ have the same joint moments under τ as $\{H_r\}_{r=1}^k$ and $\{F_{rs}\}_{r,s=1}^k$ under the normalized matrix trace.

A free approximation

The measure μ_0 is the τ-law of an operator

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

in a non-commutative probability space $(\mathcal{A}, \tau)$:

- $\{h_r\}_{r=1}^k$ and $\{f_{rs}\}_{r,s=1}^k$ have the same joint moments under τ as $\{H_r\}_{r=1}^k$ and $\{F_{rs}\}_{r,s=1}^k$ under the normalized matrix trace.
- Each $g_r^* g_r$ has Marcenko-Pastur moments under τ.

A free approximation

The measure μ_0 is the τ-law of an operator

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

in a non-commutative probability space $(\mathcal{A}, \tau)$:

- $\{h_r\}_{r=1}^k$ and $\{f_{rs}\}_{r,s=1}^k$ have the same joint moments under τ as $\{H_r\}_{r=1}^k$ and $\{F_{rs}\}_{r,s=1}^k$ under the normalized matrix trace.
- Each $g_r^* g_r$ has Marcenko-Pastur moments under τ.
- The families $\{h_r : r = 1, \ldots, k\}$, $\{f_{rs} : r, s = 1, \ldots, k\}$, and elements $g_1, \ldots, g_k$ are free with amalgamation over a diagonal sub-algebra of projections. [Benaych-Georges '09]

A free approximation

The measure μ_0 is the τ-law of an operator

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

in a non-commutative probability space $(\mathcal{A}, \tau)$:

- $\{h_r\}_{r=1}^k$ and $\{f_{rs}\}_{r,s=1}^k$ have the same joint moments under τ as $\{H_r\}_{r=1}^k$ and $\{F_{rs}\}_{r,s=1}^k$ under the normalized matrix trace.
- Each $g_r^* g_r$ has Marcenko-Pastur moments under τ.
- The families $\{h_r : r = 1, \ldots, k\}$, $\{f_{rs} : r, s = 1, \ldots, k\}$, and elements $g_1, \ldots, g_k$ are free with amalgamation over a diagonal sub-algebra of projections. [Benaych-Georges '09]

We show that this approximation is asymptotically correct, due to the independence and rotational invariance of $G_1, \ldots, G_k$.

Computation of fixed-point equations

In $(\mathcal{A}, \tau)$, let $\tau^{\mathcal{H}} : \mathcal{A} \to \mathcal{H}$ be the conditional expectation onto the subalgebra $\mathcal{H} = \langle h_1, \ldots, h_k \rangle \subset \mathcal{A}$, and similarly for $\tau^{\mathcal{G}}, \tau^{\mathcal{F}}$.

Computation of fixed-point equations

In $(\mathcal{A}, \tau)$, let $\tau^{\mathcal{H}} : \mathcal{A} \to \mathcal{H}$ be the conditional expectation onto the subalgebra $\mathcal{H} = \langle h_1, \ldots, h_k \rangle \subset \mathcal{A}$, and similarly for $\tau^{\mathcal{G}}, \tau^{\mathcal{F}}$. Let

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

$$v = \sum_{r=1}^{k} \sum_{s=1}^{k} g_r^* f_{rs} g_s$$

$$u = \sum_{r=1}^{k} \sum_{s=1}^{k} f_{rs}$$

Computation of fixed-point equations

In $(\mathcal{A}, \tau)$, let $\tau^{\mathcal{H}} : \mathcal{A} \to \mathcal{H}$ be the conditional expectation onto the subalgebra $\mathcal{H} = \langle h_1, \ldots, h_k \rangle \subset \mathcal{A}$, and similarly for $\tau^{\mathcal{G}}, \tau^{\mathcal{F}}$. Let

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$$

$$v = \sum_{r=1}^{k} \sum_{s=1}^{k} g_r^* f_{rs} g_s$$

$$u = \sum_{r=1}^{k} \sum_{s=1}^{k} f_{rs}$$

We relate the $\tau^{\mathcal{H}}$ Stieltjes-transform of w, $\tau^{\mathcal{G}}$ Stietjes-transform of v, and $\tau^{\mathcal{F}}$ Stieltjes-transform of u using conditional cumulant relations, and compute $\tau \circ \tau^{\mathcal{H}}((z - w)^{-1})$. [Speicher, Vargas '12]

All eigenvalues stick to the support

Theorem (Fan, Sun, Wang '19)

For any fixed $\delta > 0$, *almost surely for all large* n

$$\mathrm{spec}(\widehat{\Sigma}) \subset \mathrm{supp}(\mu_0)_\delta \equiv \{x \in \mathbb{R} : \mathrm{dist}(x, \mathrm{supp}(\mu_0)) < \delta\}.$$

All eigenvalues stick to the support

Theorem (Fan, Sun, Wang '19)

For any fixed $\delta > 0$*, almost surely for all large* n

$$\mathrm{spec}(\widehat{\Sigma}) \subset \mathrm{supp}(\mu_0)_\delta \equiv \{x \in \mathbb{R} : \mathrm{dist}(x, \mathrm{supp}(\mu_0)) < \delta\}.$$

We get this from a strong asymptotic freeness result for GOE and deterministic matrices, by embedding $G_1, \ldots, G_k$ into GOE.

All eigenvalues stick to the support

Theorem (Fan, Sun, Wang '19)

For any fixed $\delta > 0$*, almost surely for all large* n

$$\mathrm{spec}(\widehat{\Sigma}) \subset \mathrm{supp}(\mu_0)_\delta \equiv \{x \in \mathbb{R} : \mathrm{dist}(x, \mathrm{supp}(\mu_0)) < \delta\}.$$

We get this from a strong asymptotic freeness result for GOE and deterministic matrices, by embedding $G_1, \ldots, G_k$ into GOE.

Theorem (Fan, Johnstone '17)

When $\Sigma_r = \sigma_r^2 \,\mathrm{Id}$ *for every* r*, the extremal eigenvalue at each "regular" edge of* $\mathrm{supp}(\mu_0)$ *has GOE Tracy-Widom fluctuations.*

All eigenvalues stick to the support

Theorem (Fan, Sun, Wang '19)

For any fixed $\delta > 0$, almost surely for all large n

$$\mathrm{spec}(\widehat{\Sigma}) \subset \mathrm{supp}(\mu_0)_\delta \equiv \{x \in \mathbb{R} : \mathrm{dist}(x, \mathrm{supp}(\mu_0)) < \delta\}.$$

We get this from a strong asymptotic freeness result for GOE and deterministic matrices, by embedding $G_1, \ldots, G_k$ into GOE.

Theorem (Fan, Johnstone '17)

When $\Sigma_r = \sigma_r^2 \,\mathrm{Id}$ for every r, the extremal eigenvalue at each "regular" edge of $\mathrm{supp}(\mu_0)$ has GOE Tracy-Widom fluctuations.

This is a distinct analytic argument specific to the isotropic setting. Establishing Tracy-Widom for general Σ_r is interesting and open.

Spiked model and outliers

For understanding principal components analysis in these models, we are interested in spiked settings

$$\Sigma_r = \mathring{\Sigma}_r + \Gamma_r^T \Gamma_r$$

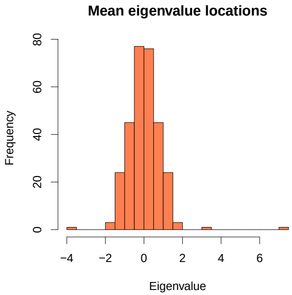
where $\Gamma_r \in \mathbb{R}^{p \times \ell_r}$, and $\Gamma_r^T \Gamma_r$ is a perturbation of fixed rank ℓ_r.

Spiked model and outliers

For understanding principal components analysis in these models, we are interested in spiked settings

$$\Sigma_r = \mathring{\Sigma}_r + \Gamma_r^T \Gamma_r$$

where $\Gamma_r \in \mathbb{R}^{p \times \ell_r}$, and $\Gamma_r^T \Gamma_r$ is a perturbation of fixed rank ℓ_r.



$\widehat{\Sigma}_A$ in a twin-study design, where each Σ_A and Σ_E is a rank-1 perturbation of Id. The spike-to-outlier mapping is not 1-to-1, and Σ_E produces outliers in $\widehat{\Sigma}_A$.

Spiked model and outliers

Define μ_0 by the bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$. Recall $b_1(z), \ldots, b_k(z)$ from the fixed-point equations for μ_0, and set

$$T(z) = \mathrm{Id} + \left(\Gamma_r^\mathsf{T} \left(z \, \mathrm{Id} + b \cdot \mathring{\Sigma} \right)^{-1} \Gamma_s b_s(z) \right)_{r,s=1}^{k} \in \mathbb{C}^{\ell_+ \times \ell_+}$$

where $\ell_+ = \ell_1 + \ldots + \ell_k$ and $b \cdot \mathring{\Sigma} = b_1 \mathring{\Sigma}_1 + \ldots + b_k \mathring{\Sigma}_k$.

Spiked model and outliers

Define μ_0 by the bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$. Recall $b_1(z), \ldots, b_k(z)$ from the fixed-point equations for μ_0, and set

$$T(z) = \mathrm{Id} + \left(\Gamma_r^{\mathsf{T}} \left(z\,\mathrm{Id} + b \cdot \mathring{\Sigma} \right)^{-1} \Gamma_s b_s(z) \right)_{r,s=1}^{k} \in \mathbb{C}^{\ell_+ \times \ell_+}$$

where $\ell_+ = \ell_1 + \ldots + \ell_k$ and $b \cdot \mathring{\Sigma} = b_1 \mathring{\Sigma}_1 + \ldots + b_k \mathring{\Sigma}_k$.

Theorem (Fan, Sun, Wang '19)

Eigenvalues $\widehat{\lambda}$ *of* $\widehat{\Sigma}$ *separated from* $\mathrm{supp}(\mu_0)$ *are in correspondence with roots of* $0 = \det T(\lambda)$*, such that* $\widehat{\lambda} - \lambda \to 0$.

Spiked model and outliers

Define μ_0 by the bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$. Recall $b_1(z), \ldots, b_k(z)$ from the fixed-point equations for μ_0, and set

$$T(z) = \mathrm{Id} + \left(\Gamma_r^\mathsf{T} \left(z\,\mathrm{Id} + b \cdot \mathring{\Sigma} \right)^{-1} \Gamma_s b_s(z) \right)_{r,s=1}^{k} \in \mathbb{C}^{\ell_+ \times \ell_+}$$

where $\ell_+ = \ell_1 + \ldots + \ell_k$ and $b \cdot \mathring{\Sigma} = b_1 \mathring{\Sigma}_1 + \ldots + b_k \mathring{\Sigma}_k$.

Theorem (Fan, Sun, Wang '19)

Eigenvalues $\widehat{\lambda}$ *of* $\widehat{\Sigma}$ *separated from* $\operatorname{supp}(\mu_0)$ *are in correspondence with roots of* $0 = \det T(\lambda)$, *such that* $\widehat{\lambda} - \lambda \to 0$. *If* λ *is an isolated root, then the eigenvector* $\widehat{v}$ *and unit vector* $u \in \ker T(\lambda)$ *satisfy*

$$(\Gamma_1^\mathsf{T} \widehat{v}, \ldots, \Gamma_k^\mathsf{T} \widehat{v}) - \alpha^{-1/2} u \to 0.$$

Spiked model and outliers

Define μ_0 by the bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$. Recall $b_1(z), \ldots, b_k(z)$ from the fixed-point equations for μ_0, and set

$$T(z) = \mathrm{Id} + \left(\Gamma_r^\top \left(z\,\mathrm{Id} + b \cdot \mathring{\Sigma} \right)^{-1} \Gamma_s b_s(z) \right)_{r,s=1}^{k} \in \mathbb{C}^{\ell_+ \times \ell_+}$$

where $\ell_+ = \ell_1 + \ldots + \ell_k$ and $b \cdot \mathring{\Sigma} = b_1 \mathring{\Sigma}_1 + \ldots + b_k \mathring{\Sigma}_k$.

Theorem (Fan, Sun, Wang '19)

Eigenvalues $\widehat{\lambda}$ *of* $\widehat{\Sigma}$ *separated from* $\mathrm{supp}(\mu_0)$ *are in correspondence with roots of* $0 = \det T(\lambda)$*, such that* $\widehat{\lambda} - \lambda \to 0$*. If* λ *is an isolated root, then the eigenvector* $\widehat{v}$ *and unit vector* $u \in \ker T(\lambda)$ *satisfy*

$$(\Gamma_1^\top \widehat{v}, \ldots, \Gamma_k^\top \widehat{v}) - \alpha^{-1/2} u \to 0.$$

For $\mathring{\Sigma}_r = \sigma_r^2\,\mathrm{Id}$, we established also in [Fan, Johnstone, Sun '18] a Gaussian CLT for $\widehat{\lambda} - \lambda$.

Bias and phase transition of eigenvalues

The number of outlier eigenvalues is predicted by

$$|\{\lambda \notin \mathrm{supp}(\mu_0) : 0 = \det T(\lambda)\}|.$$

This is 0 if all population spikes are small.

Bias and phase transition of eigenvalues

The number of outlier eigenvalues is predicted by

$$|\{\lambda \notin \mathrm{supp}(\mu_0) : 0 = \det T(\lambda)\}|.$$

This is 0 if all population spikes are small.

More generally, the number and locations of outliers depend on

- Alignments between bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$
- Their alignments with all spike directions $\Gamma_1, \ldots, \Gamma_k$

Bias and phase transition of eigenvalues

The number of outlier eigenvalues is predicted by

$$|\{\lambda \notin \mathrm{supp}(\mu_0) : 0 = \det T(\lambda)\}|.$$

This is 0 if all population spikes are small.

More generally, the number and locations of outliers depend on

- Alignments between bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$
- Their alignments with all spike directions $\Gamma_1, \ldots, \Gamma_k$

Qualitatively, for the twin-study design:

- Large spikes in Σ_E generate two outliers of opposite sign in $\widehat{\Sigma}_A$

Bias and phase transition of eigenvalues

The number of outlier eigenvalues is predicted by

$$|\{\lambda \notin \mathrm{supp}(\mu_0) : 0 = \det T(\lambda)\}|.$$

This is 0 if all population spikes are small.

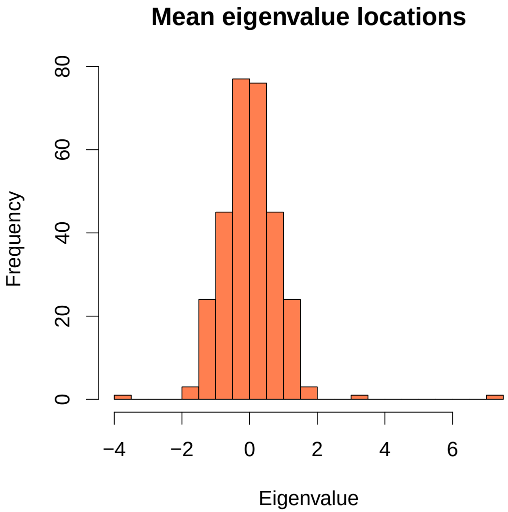
More generally, the number and locations of outliers depend on

- Alignments between bulk components $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$
- Their alignments with all spike directions $\Gamma_1, \ldots, \Gamma_k$

Qualitatively, for the twin-study design:

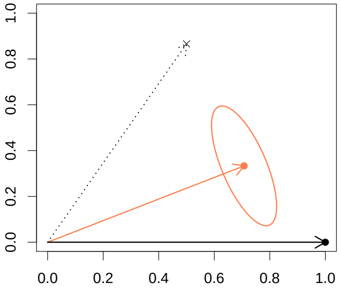
- Large spikes in Σ_E generate two outliers of opposite sign in $\widehat{\Sigma}_A$
- Large eigenvalues in Σ_A are observed with upward bias in $\widehat{\Sigma}_A$, where the bias is larger if this eigenvector of Σ_A is aligned with eigenvectors of Σ_E.

Bias and phase transition of eigenvalues



Bias of principal eigenvectors

In contrast to the sample covariance setting of $k = 1$, here there may also be bias in the outlier *eigenvectors*:



In high dimensions, principal component eigenvectors of $\widehat{\Sigma}_A$ may be biased towards eigenvectors of Σ_E.

Debiasing the principal components

For $\mathring{\Sigma}_r = \sigma_r^2 \,\mathrm{Id}$, we developed in [Fan, Johnstone, Sun '18] an algorithm to estimate the population eigenvalues, and also debias the estimated principal eigenvectors:

Debiasing the principal components

For $\mathring{\Sigma}_r = \sigma_r^2 \,\mathrm{Id}$, we developed in [Fan, Johnstone, Sun '18] an algorithm to estimate the population eigenvalues, and also debias the estimated principal eigenvectors:

1. Track the trajectories of outlier eigenvalues of $\widehat{\Sigma} = Y^\mathsf{T} BY$, as B varies within a $(k-1)$-dimensional family.

Debiasing the principal components

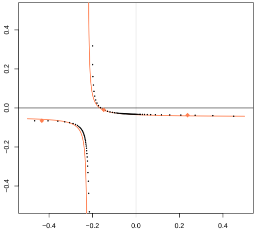
For $\mathring{\Sigma}_r = \sigma_r^2 \,\mathrm{Id}$, we developed in [Fan, Johnstone, Sun '18] an algorithm to estimate the population eigenvalues, and also debias the estimated principal eigenvectors:

1. Track the trajectories of outlier eigenvalues of $\widehat{\Sigma} = Y^\top BY$, as B varies within a $(k-1)$-dimensional family.
2. For B where an outlier $\widehat{\lambda}$ satisfies $b_2(\widehat{\lambda}) = \ldots = b_k(\widehat{\lambda}) = 0$, $\widehat{v}$ is unbiased for an eigenvector of Σ_1, and $\widehat{\lambda}$ is related to the eigenvalue of Σ_1.

Debiasing the principal components

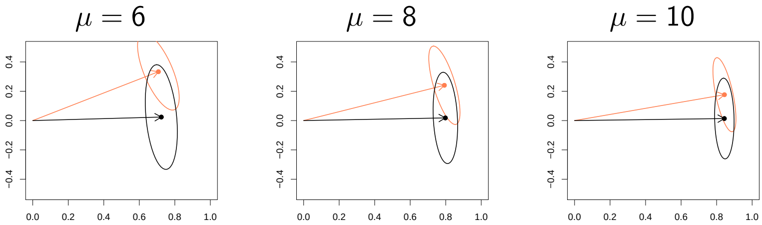
For $\mathring{\Sigma}_r = \sigma_r^2 \,\mathrm{Id}$, we developed in [Fan, Johnstone, Sun '18] an algorithm to estimate the population eigenvalues, and also debias the estimated principal eigenvectors:

1. Track the trajectories of outlier eigenvalues of $\widehat{\Sigma} = Y^\top BY$, as B varies within a $(k-1)$-dimensional family.
2. For B where an outlier $\widehat{\lambda}$ satisfies $b_2(\widehat{\lambda}) = \ldots = b_k(\widehat{\lambda}) = 0$, $\widehat{v}$ is unbiased for an eigenvector of Σ_1, and $\widehat{\lambda}$ is related to the eigenvalue of Σ_1. Use a grid search to find such B.



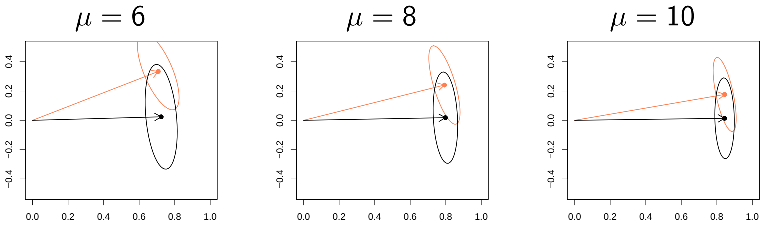
Debiasing the principal components

Mean and 90%-ellipsoid of MANOVA and debiased principal eigenvector estimates, for true eigenvector $(1, 0)$ and eigenvalue μ:



Debiasing the principal components

Mean and 90%-ellipsoid of MANOVA and debiased principal eigenvector estimates, for true eigenvector $(1, 0)$ and eigenvalue μ:

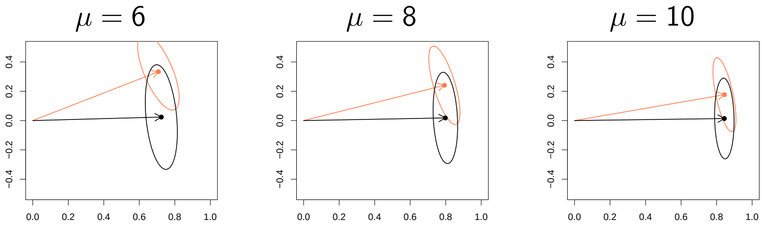


Mean and st. dev. of MANOVA and debiased eigenvalue estimates:

	$\mu = 6$	$\mu = 8$	$\mu = 10$
MANOVA	10.57 (1.74)	11.98 (1.85)	13.59 (1.99)
Estimated	6.28 (1.56)	8.21 (1.72)	10.15 (1.91)

Debiasing the principal components

Mean and 90%-ellipsoid of MANOVA and debiased principal eigenvector estimates, for true eigenvector $(1, 0)$ and eigenvalue μ:



Mean and st. dev. of MANOVA and debiased eigenvalue estimates:

	$\mu = 6$	$\mu = 8$	$\mu = 10$
MANOVA	10.57 (1.74)	11.98 (1.85)	13.59 (1.99)
Estimated	6.28 (1.56)	8.21 (1.72)	10.15 (1.91)

Developing estimation procedures for general $\mathring{\Sigma}_r \neq \sigma_r^2 \,\mathrm{Id}$ is open.

A few general tools

1. ℓ_2 fluctuation averaging

In the setting $\Sigma_r = \sigma_r^2 \,\mathrm{Id}$, recall that $\widehat{\Sigma} = G^\mathsf{T} F G$.

1. ℓ_2 fluctuation averaging

In the setting $\Sigma_r = \sigma_r^2 \,\mathrm{Id}$, recall that $\widehat{\Sigma} = G^\top F G$. The resolvent

$$R(z) = (\widehat{\Sigma} - z\,\mathrm{Id})^{-1}$$

has a deterministic approximation $R_0(z)$ in the form of an anisotropic local law [Knowles, Yin '17].

For z separated from the spectral support, $\mathrm{dist}(z, \mathrm{supp}(\mu_0)) > \delta$, and any deterministic unit vectors u and v, this says

$$u^*(R(z) - R_0(z))v \prec 1/\sqrt{n}.$$

1. ℓ_2 fluctuation averaging

In the setting $\Sigma_r = \sigma_r^2 \operatorname{Id}$, recall that $\widehat{\Sigma} = G^\top F G$. The resolvent

$$R(z) = (\widehat{\Sigma} - z \operatorname{Id})^{-1}$$

has a deterministic approximation $R_0(z)$ in the form of an anisotropic local law [Knowles, Yin '17].

For z separated from the spectral support, $\operatorname{dist}(z, \operatorname{supp}(\mu_0)) > \delta$, and any deterministic unit vectors u and v, this says

$$u^*(R(z) - R_0(z))v \prec 1/\sqrt{n}.$$

To approximate other combinations of resolvent entries, e.g. $\operatorname{Tr} R(z)$, one may apply weak dependence of these entries and "fluctuation averaging" techniques [Erdös, Yau, Yin '11].

1. ℓ_2 fluctuation averaging

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random quantities. For $\mathcal{Y}$ a scalar function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denote

$$\mathcal{P}_i[\mathcal{Y}] = \mathbb{E}_{\mathbf{x}_i}[\mathcal{Y}], \quad \mathcal{Q}_i[\mathcal{Y}] = \mathcal{Y} - \mathcal{P}_i[\mathcal{Y}], \quad \mathcal{Q}_S[\mathcal{Y}] = \left(\prod_{i \in S} \mathcal{Q}_i \right) [\mathcal{Y}]$$

1. ℓ_2 fluctuation averaging

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random quantities. For $\mathcal{Y}$ a scalar function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denote

$$\mathcal{P}_i[\mathcal{Y}] = \mathbb{E}_{\mathbf{x}_i}[\mathcal{Y}], \quad \mathcal{Q}_i[\mathcal{Y}] = \mathcal{Y} - \mathcal{P}_i[\mathcal{Y}], \quad \mathcal{Q}_S[\mathcal{Y}] = \left(\prod_{i \in S} \mathcal{Q}_i\right)[\mathcal{Y}]$$

Lemma (Fan, Johnstone, Sun '18)

Suppose $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ satisfy $\mathcal{P}_i[\mathcal{Y}_i] = 0$, $\mathcal{Y}_i \prec 1$, and the weak dependence $\mathcal{Q}_S[\mathcal{Y}_i] \prec n^{-|S|/2}$ for $i \notin S$.

1. ℓ_2 fluctuation averaging

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random quantities. For $\mathcal{Y}$ a scalar function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denote

$$\mathcal{P}_i[\mathcal{Y}] = \mathbb{E}_{\mathbf{x}_i}[\mathcal{Y}], \quad \mathcal{Q}_i[\mathcal{Y}] = \mathcal{Y} - \mathcal{P}_i[\mathcal{Y}], \quad \mathcal{Q}_S[\mathcal{Y}] = \left(\prod_{i \in S} \mathcal{Q}_i\right)[\mathcal{Y}]$$

Lemma (Fan, Johnstone, Sun '18)

Suppose $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ satisfy $\mathcal{P}_i[\mathcal{Y}_i] = 0$, $\mathcal{Y}_i \prec 1$, and the weak dependence $\mathcal{Q}_S[\mathcal{Y}_i] \prec n^{-|S|/2}$ for $i \notin S$. Then

$$\sum_{i=1}^n u_i \mathcal{Y}_i \prec \left(\sum_{i=1}^n |u_i|^2\right)^{1/2}.$$

1. ℓ_2 fluctuation averaging

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random quantities. For $\mathcal{Y}$ a scalar function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denote

$$\mathcal{P}_i[\mathcal{Y}] = \mathbb{E}_{\mathbf{x}_i}[\mathcal{Y}], \quad \mathcal{Q}_i[\mathcal{Y}] = \mathcal{Y} - \mathcal{P}_i[\mathcal{Y}], \quad \mathcal{Q}_S[\mathcal{Y}] = \left(\prod_{i \in S} \mathcal{Q}_i\right)[\mathcal{Y}]$$

Lemma (Fan, Johnstone, Sun '18)

Suppose $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ satisfy $\mathcal{P}_i[\mathcal{Y}_i] = 0$, $\mathcal{Y}_i \prec 1$, and the weak dependence $\mathcal{Q}_S[\mathcal{Y}_i] \prec n^{-|S|/2}$ for $i \notin S$. Then

$$\sum_{i=1}^n u_i \mathcal{Y}_i \prec \left(\sum_{i=1}^n |u_i|^2\right)^{1/2}.$$

Compared with existing results that I'm aware of, this does not require $u_i \equiv 1/n$ or use an upper bound on $\|u\|_\infty$.

1. ℓ_2 fluctuation averaging

Lemma (Fan, Johnstone, Sun '18)

Suppose $(\mathcal{Y}_{ij})_{i\neq j}$ *satisfy* $\mathcal{P}_i[\mathcal{Y}_{ij}] = \mathcal{P}_j[\mathcal{Y}_{ij}] = 0$, $\mathcal{Y}_{ij} \prec 1$, *and* $\mathcal{Q}_S[\mathcal{Y}_{ij}] \prec n^{-|S|/2}$ *for* $i, j \notin S$. *Then*

$$\sum_{i\neq j} u_{ij}\mathcal{Y}_{ij} \prec \left(\sum_{i\neq j} |u_{ij}|^2\right)^{1/2}.$$

1. ℓ_2 fluctuation averaging

Lemma (Fan, Johnstone, Sun '18)

Suppose $(\mathcal{Y}_{ij})_{i \neq j}$ *satisfy* $\mathcal{P}_i[\mathcal{Y}_{ij}] = \mathcal{P}_j[\mathcal{Y}_{ij}] = 0$, $\mathcal{Y}_{ij} \prec 1$, *and* $\mathcal{Q}_S[\mathcal{Y}_{ij}] \prec n^{-|S|/2}$ *for* $i, j \notin S$. *Then*

$$\sum_{i \neq j} u_{ij} \mathcal{Y}_{ij} \prec \left(\sum_{i \neq j} |u_{ij}|^2 \right)^{1/2} .$$

For z with $\operatorname{dist}(z, \operatorname{supp}(\mu_0)) > \delta$, we obtain from these and the entrywise local law, for any deterministic matrix M, that

$$\operatorname{Tr}(R(z) - R_0(z))M \prec \|M\|_{\mathrm{HS}}/\sqrt{n}.$$

1. ℓ_2 fluctuation averaging

Lemma (Fan, Johnstone, Sun '18)

Suppose $(\mathcal{Y}_{ij})_{i \neq j}$ *satisfy* $\mathcal{P}_i[\mathcal{Y}_{ij}] = \mathcal{P}_j[\mathcal{Y}_{ij}] = 0$, $\mathcal{Y}_{ij} \prec 1$, *and* $\mathcal{Q}_S[\mathcal{Y}_{ij}] \prec n^{-|S|/2}$ *for* $i, j \notin S$. *Then*

$$\sum_{i \neq j} u_{ij} \mathcal{Y}_{ij} \prec \left(\sum_{i \neq j} |u_{ij}|^2 \right)^{1/2}.$$

For z with $\operatorname{dist}(z, \operatorname{supp}(\mu_0)) > \delta$, we obtain from these and the entrywise local law, for any deterministic matrix M, that

$$\operatorname{Tr}(R(z) - R_0(z))M \prec \|M\|_{\mathrm{HS}}/\sqrt{n}.$$

This recovers the anisotropic local law for rank-one $M = vu^*$, but provides a strengthened guarantee when $\operatorname{rank}(M) \gg 1$. We use this to establish the CLT for fluctuations of outlier eigenvalues.

2. Anisotropic resolvent approximation

For $\Sigma_r \neq \sigma_r^2 \operatorname{Id}$, we still wish to approximate $u^* R(z) v$ to analyze outliers by a "master equation" approach [Benaych-Georges, Nadakuditi '12], but we currently don't have a local law.

2. Anisotropic resolvent approximation

For $\Sigma_r \neq \sigma_r^2 \, \mathrm{Id}$, we still wish to approximate $u^* R(z) v$ to analyze outliers by a "master equation" approach [Benaych-Georges, Nadakuditi '12], but we currently don't have a local law.

We show using free probability techniques, for general self-adjoint polynomial matrix models, that for $\mathrm{dist}(z, \mathrm{supp}(\mu_0)) > \delta$ we have

$$u^*(R(z) - R_0(z))v \to 0$$

almost surely, for a certain deterministic approximation R_0.

2. Anisotropic resolvent approximation

For $\Sigma_r \neq \sigma_r^2 \,\mathrm{Id}$, we still wish to approximate $u^* R(z) v$ to analyze outliers by a "master equation" approach [Benaych-Georges, Nadakuditi '12], but we currently don't have a local law.

We show using free probability techniques, for general self-adjoint polynomial matrix models, that for $\mathrm{dist}(z, \mathrm{supp}(\mu_0)) > \delta$ we have

$$u^*(R(z) - R_0(z))v \to 0$$

almost surely, for a certain deterministic approximation R_0.

Note: R_0 is not necessarily isotropic. The approximation is not $u^* R(z) v \approx u^* v \cdot m_0(z)$ if u, v are aligned with $R_0(z)$.

2. Anisotropic resolvent approximation

Let

- $H_1, \ldots, H_q \in \mathbb{C}^{n\times n}$ be deterministic matrices,
- $G_1, \ldots, G_p \in \mathbb{C}^{n\times n}$ be random and jointly orthogonally invariant in law.

So $\{H_1, \ldots, H_q\}$ is asymptotically free of $\{G_1, \ldots, G_p\}$.

2. Anisotropic resolvent approximation

Let

- $H_1, \ldots, H_q \in \mathbb{C}^{n \times n}$ be deterministic matrices,
- $G_1, \ldots, G_p \in \mathbb{C}^{n \times n}$ be random and jointly orthogonally invariant in law.

So $\{H_1, \ldots, H_q\}$ is asymptotically free of $\{G_1, \ldots, G_p\}$.

Define the von Neumann algebras

- $(\mathcal{A}_1, \tau_1) \equiv (\mathbb{C}^{n \times n}, n^{-1} \operatorname{Tr})$, containing $H_1, \ldots, H_q$,
- $(\mathcal{A}_2, \tau_2)$ containing $\{g_1, \ldots, g_p\}$ which approximate $\{G_1, \ldots, G_p\}$ in joint law.

Let $(\mathcal{A}, \tau)$ be the free deterministic equivalent model defined by their von Neumann free product.

2. Anisotropic resolvent approximation

Let

- $H_1, \ldots, H_q \in \mathbb{C}^{n \times n}$ be deterministic matrices,
- $G_1, \ldots, G_p \in \mathbb{C}^{n \times n}$ be random and jointly orthogonally invariant in law.

So $\{H_1, \ldots, H_q\}$ is asymptotically free of $\{G_1, \ldots, G_p\}$.

Define the von Neumann algebras

- $(\mathcal{A}_1, \tau_1) \equiv (\mathbb{C}^{n \times n}, n^{-1}\,\mathrm{Tr})$, containing $H_1, \ldots, H_q$,
- $(\mathcal{A}_2, \tau_2)$ containing $\{g_1, \ldots, g_p\}$ which approximate $\{G_1, \ldots, G_p\}$ in joint law.

Let $(\mathcal{A}, \tau)$ be the free deterministic equivalent model defined by their von Neumann free product.

Let $\tau^{\mathcal{H}} : \mathcal{A} \to \mathcal{H} \equiv \langle H_1, \ldots, H_q \rangle$ be the conditional expectation.
Note: For any $a \in \mathcal{A}$, $\tau^{\mathcal{H}}(a) \in \mathcal{H} \subset \mathcal{A}_1$ is an $n \times n$ matrix!

2. Anisotropic resolvent approximation

Theorem (Fan, Sun, Wang '19)

Fix a self-adjoint $*$-polynomial Q and $\delta > 0$. Let

$$W = Q(G_1, \ldots, G_p, H_1, \ldots, H_q) \in \mathbb{C}^{n \times n},$$
$$w = Q(g_1, \ldots, g_p, H_1, \ldots, H_q) \in \mathcal{A}.$$

Theorem (Fan, Sun, Wang '19)

Fix a self-adjoint $*$-polynomial Q and $\delta > 0$. Let

$$W = Q(G_1, \ldots, G_p, H_1, \ldots, H_q) \in \mathbb{C}^{n \times n},$$
$$w = Q(g_1, \ldots, g_p, H_1, \ldots, H_q) \in \mathcal{A}.$$

Let $R(z) = (W - z\,\mathrm{Id})^{-1}$, and define a deterministic approximation

$$R_0(z) = \tau^{\mathcal{H}}((w - z)^{-1}) \in \mathbb{C}^{n \times n}.$$

Then for any deterministic unit vectors $u, v \in \mathbb{C}^n$, as $n \to \infty$,

$$u^*(R(z) - R_0(z))v \to 0$$

uniformly over $\{z \in \mathbb{C} : \mathrm{dist}(z, \mathrm{spec}(W) \cup \mathrm{spec}(w)) > \delta\}$.

3. Augmented Cauchy and $\mathcal{R}$-transforms

Our computations in the approximating free model use relations between (conditional) Cauchy and $\mathcal{R}$-transforms:

3. Augmented Cauchy and $\mathcal{R}$-transforms

Our computations in the approximating free model use relations between (conditional) Cauchy and $\mathcal{R}$-transforms:

Let $\kappa_\ell(a_1, \ldots, a_\ell)$ be the ℓ^{th} order non-crossing cumulant, and

$$G_a(z) = \tau((z-a)^{-1}) = \sum_{\ell=0}^{\infty} z^{-(\ell+1)}\tau(a^\ell),$$

$$\mathcal{R}_a(z) = \sum_{\ell=1}^{\infty} z^{\ell-1}\kappa_\ell(a, \ldots, a).$$

The moment-cumulant relations for non-crossing partitions give

$$G_a(z) = (z - \mathcal{R}_a(G_a(z)))^{-1}.$$

3. Augmented Cauchy and $\mathcal{R}$-transforms

In addition to approximating $n^{-1} \operatorname{Tr} R(z)$, we also need

$$n^{-1} \operatorname{Tr} AR(z)$$

for certain (random) matrices A.

3. Augmented Cauchy and $\mathcal{R}$-transforms

In addition to approximating $n^{-1}\,\mathrm{Tr}\,R(z)$, we also need

$$n^{-1}\,\mathrm{Tr}\,AR(z)$$

for certain (random) matrices A. We introduce augmented transforms and some associated moment-cumulant relations:

$$G_{a,w}(z) = \tau(a(z-w)^{-1}) = \sum_{\ell=0}^{\infty} z^{-(\ell+1)}\tau(aw^{\ell}),$$

$$\mathcal{R}_{a,w}(z) = \sum_{\ell=1}^{\infty} z^{\ell-1}\kappa_{\ell}(a,w,\ldots,w).$$

Lemma (Fan, Sun, Wang '19)

$$G_{a,w}(z) = \mathcal{R}_{a,w}(G_w(z))G_w(z)$$

4. Strong asymptotic freeness for GOE

In the setting $\Sigma_r \neq \sigma_r^2 \,\mathrm{Id}$, we use a strong asymptotic freeness result to show that eigenvalues stick to $\mathrm{supp}(\mu_0)$:

4. Strong asymptotic freeness for GOE

In the setting $\Sigma_r \neq \sigma_r^2\,\mathrm{Id}$, we use a strong asymptotic freeness result to show that eigenvalues stick to $\mathrm{supp}(\mu_0)$:

Theorem (Fan, Sun, Wang '19)

Let $W_1, \ldots, W_p$ be independent GOE, $H_1, \ldots, H_q$ deterministic with bounded norm, and $w_1, \ldots, w_p$ free semicircular elements. For any fixed self-adjoint $*$-polynomial Q and $\delta > 0$, a.s. for large n,

$$\mathrm{spec}(Q(\{W_i\}_{i=1}^p, \{H_j\}_{j=1}^q)) \subset \mathrm{spec}(Q(\{w_i\}_{i=1}^p, \{H_j\}_{j=1}^q))_\delta.$$

4. Strong asymptotic freeness for GOE

In the setting $\Sigma_r \neq \sigma_r^2 \,\mathrm{Id}$, we use a strong asymptotic freeness result to show that eigenvalues stick to $\mathrm{supp}(\mu_0)$:

Theorem (Fan, Sun, Wang '19)

Let $W_1, \dots, W_p$ be independent GOE, $H_1, \dots, H_q$ deterministic with bounded norm, and $w_1, \dots, w_p$ free semicircular elements. For any fixed self-adjoint $*$-polynomial Q and $\delta > 0$, a.s. for large n,

$$\mathrm{spec}(Q(\{W_i\}_{i=1}^p, \{H_j\}_{j=1}^q)) \subset \mathrm{spec}(Q(\{w_i\}_{i=1}^p, \{H_j\}_{j=1}^q))_\delta.$$

The GUE analogue was proven in [Male '12], and extended to complex Wigner matrices in [Belinschi, Capitaine '17].

4. Strong asymptotic freeness for GOE

In the setting $\Sigma_r \neq \sigma_r^2 \,\mathrm{Id}$, we use a strong asymptotic freeness result to show that eigenvalues stick to $\mathrm{supp}(\mu_0)$:

Theorem (Fan, Sun, Wang '19)

Let $W_1, \ldots, W_p$ be independent GOE, $H_1, \ldots, H_q$ deterministic with bounded norm, and $w_1, \ldots, w_p$ free semicircular elements. For any fixed self-adjoint $*$-polynomial Q and $\delta > 0$, a.s. for large n,

$$\mathrm{spec}(Q(\{W_i\}_{i=1}^p, \{H_j\}_{j=1}^q)) \subset \mathrm{spec}(Q(\{w_i\}_{i=1}^p, \{H_j\}_{j=1}^q))_\delta.$$

The GUE analogue was proven in [Male '12], and extended to complex Wigner matrices in [Belinschi, Capitaine '17].

We follow proof ideas of these works, and of [Schultz '05] in the pure GOE setting.

References

Zhou Fan, Iain M. Johnstone, "Eigenvalue distributions of variance components estimators in high-dimensional random effects models," *The Annals of Statistics*, 2019+.

Zhou Fan, Yi Sun, Zhichao Wang, "Principal components in linear mixed models with general bulk," arXiv:1903.09592, 2019.

Zhou Fan, Iain M. Johnstone, "Tracy-Widom at each edge of real covariance and MANOVA estimators," arXiv:1707.02352, 2018.

Zhou Fan, Iain M. Johnstone, Yi Sun, "Spiked covariances and principal components analysis in high-dimensional random effects models," arXiv:arXiv:1806.09529, 2018.

Thank you!