Landscape of separable covariance matrices

Random matrices beyond sample covariance matrices

Xiucai Ding

Department of Mathematics
Duke University

Shanghai, 2019

Joint work with Fan Yang (Upenn)

Model from matrix normal distribution

- $Y = (y_{ij}), i = 1, 2, \cdots, p; j = 1, 2, \cdots n.$
- Matrix normal distribution (De Waal, 1985)

$$\mathbb{E}\,\mathrm{Vec}(Y) = \mathbf{0}_{pn},\ \mathrm{Cov}(\mathrm{Vec}(Y)) = A \otimes B.$$

- A : spatial; B : temporal.
- Spatial-temporal process: Climate, environmental sciences, medical sciences, brain-imaing.
- $Y = A^{1/2} X B^{1/2}, X \in \mathbb{R}^{p \times n}$ are of i.i.d. entries.
- $B = I \to$ sample covariance matrix .

Basic assumptions

- $X = (x_{ij}) \in \mathbb{R}^{p \times n}$ with i.i.d. entries and

$$\mathbb{E}x_{ij} = 0, \ \mathbb{E}x_{ij}^2 = n^{-1}.$$

- For some large integer $\varsigma > 0$ such that for $1 \leq k \leq \varsigma$

$$\mathbb{E}|x_{ij}|^k \leq C_k n^{-k/2}.$$

- For some constant $0 < \tau \leq 1$, we have

$$\tau \leq d := p/n \leq \tau^{-1}.$$

- $A \in \mathbb{R}^{p \times p}$ and $B \in \mathbb{R}^{n \times n}$ are some p.d.f. deterministic matrices satisfying regularity assumptions.

$$\operatorname{Spec}(A) = \{a_1, a_2, \cdots, a_p\}, \ \operatorname{Spec}(B) = \{b_1, b_2, \cdots, b_n\}.$$

Some notations

- $\mathcal{Q}_1 = A^{1/2} X B X^\top A^{1/2}, \mathcal{Q}_2 = B^{1/2} X^\top A X B^{1/2}$

$$\mathcal{G}_1(z) = (\mathcal{Q}_1 - z)^{-1},\ \mathcal{G}_2(z) = (\mathcal{Q}_2 - z)^{-1}.$$

- ρ : ESD of $\mathcal{Q}_1$. Stieltjes transform of ρ

$$m(z) = p^{-1} \operatorname{tr} \mathcal{G}_1(z).$$

$$m_1(z) = n^{-1} \sum_{i=1}^{p} a_i (\mathcal{G}_1(z))_{ii},\ m_2(z) = n^{-1} \sum_{i=1}^{n} b_i (\mathcal{G}_2(z))_{ii}(z).$$

- Self-consistent equations: $(m_{1c}, m_{2c}) \in \mathbb{C}_+^2$

$$m_{1c}(z) = d \int \frac{x}{-z(1 + x m_{2c}(z))} \pi_A(dx)$$

$$m_{2c}(z) = d \int \frac{x}{-z(1 + x m_{1c}(z))} \pi_B(dx)$$

Macroscopic picture

- Define $m_c(z)$

$$m_c(z) = d \int \frac{1}{-z(1 + x m_{2c}(z))} \pi_A(dx).$$

Theorem (Zhang, 2007)

For any $z \in \mathbb{C}_+$, there exists a unique solution $(m_{1c}, m_{2c}) \in \mathbb{C}_+^2$ to the systems of self-consistent equations. The function m_c is the Stieltjes transform of a probability measure μ_c supported on $\mathbb{R}_+$. Moreover, μ_c has a continuous derivative $\rho_c(x)$ on $(0, \infty)$.

- Edge behavior of ρ_c

$$f(z, m) := -m + \int \frac{x}{-z + xd \int \frac{t}{1+tm} \pi_A(dt)} \pi_B(dx).$$

Macroscopic picture: D. and Yang, 2019

- The densities ρ_c and $\rho_{1,2c}$ all have the same support on $(0, \infty)$, which is a union of intervals:

$$\operatorname{supp} \rho_c \cap (0, \infty) = \operatorname{supp} \rho_{1,2c} \cap (0, \infty) = \bigcup_{k=1}^{q} [\alpha_{2k}, \alpha_{2k-1}] \cap (0, \infty),$$

  where $q \in \mathbb{N}$ depends only on $\pi_{A,B}$.

- $(x, m) = (\alpha_k, m_{2c}(\alpha_k))$ are the real solutions to the equations

$$f(x, m) = 0, \quad \text{and} \quad \frac{\partial f}{\partial m}(x, m) = 0.$$

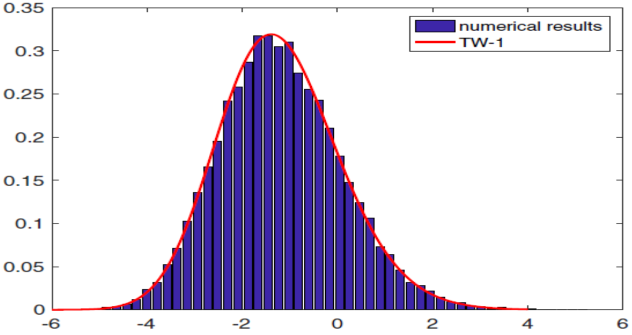
  Moreover, we have $m_{1c}(\alpha_1) \in (-b_1^{-1}, 0)$ and $m_{2c}(\alpha_1) \in (-a_1^{-1}, 0)$.

What do we expect?

- Setup: $d = 0.5$ and $n = 400$

$$\pi_A, \pi_B = 0.5\mathbf{1}_1 + 0.5\mathbf{1}_4$$

- Normalization: (El Karoui , AOP 2007)



Regularity assumption: D. and Yang, 2019

- $\lambda_r := \alpha_1$. To ensure TW, we expect: there exist constants $\beta_{1,2} > 0$ such that when $\operatorname{Im} z \geq 0$,

$$\rho_{1,2c}(\lambda_r - x) = \beta_{1,2} x^{1/2} + O(x), \quad x \downarrow 0,$$

$$m_{1,2c}(z) = m_{1,2c}(\lambda_r) + \pi a_{1,2}(z - \lambda_r)^{1/2} + O(|z - \lambda_r|), \; z \to \lambda_r.$$

- Taylor expansion: $\partial_z f(\lambda_r, m_r), \partial_\alpha^2 f(\lambda_r, m_r) = O(1)$.
- **Regularity assumption**

$$1 + m_{1c}(\lambda_r) b_1 \geq \tau, \; 1 + m_{2c}(\lambda_r) a_1 \geq \tau.$$

Separable sample covariance matrices without spikes

Theorem (D. and Yang, 2019+)

For the separable sample covariance matrices $\mathcal{Q}_1$, there exists some constant γ_r such that

$$n^{2/3}\gamma_r(\lambda_1(\mathcal{Q}_1)-\lambda_r) \Rightarrow \mathrm{TW}\,.$$

- Joint distributions: for some fixed constant $K>0$,

$$(n^{2/3}\beta_r(\lambda_1-\lambda_r),\cdots,n^{2/3}\beta_r(\lambda_K-\lambda_r)) \asymp (n^{2/3}(\lambda_1^{\mathrm{GOE}}-2),\cdots,n^{2/3}(\lambda_K^{\mathrm{GOE}}-2))$$

- The convergent limits and rates are first established in (Yang, EJP 2019). More generally, we have the rigidity results

$$|\lambda_i(\mathcal{Q}_1)-\gamma_i| \prec \min\{i, n\wedge p-i+1\}^{-1/3}n^{-2/3}.$$

To add spikes, we assume that there exist some fixed intergers $r, s \in \mathbb{N}$ and constants d_i^a, $1 \le i \le r$, and d_μ^b, $1 \le \mu \le s$, such that

$$\widetilde{A} = V^a \widetilde{\Sigma}^a (V^a)^\top, \ \widetilde{B} = V^b \widetilde{\Sigma}^b (V^b)^\top,$$

and

$$\widetilde{\Sigma}^a = \mathrm{diag}(\widetilde{a}_1, \ldots, \widetilde{a}_p), \ \widetilde{\Sigma}^b = \mathrm{diag}(\widetilde{b}_1, \ldots, \widetilde{b}_n).$$

Here

$$\widetilde{a}_i = \begin{cases} a_i(1 + d_i^a), & 1 \le i \le r \\ a_i, & \text{otherwise} \end{cases}, \qquad \widetilde{b}_\mu = \begin{cases} b_\mu(1 + d_\mu^b), & 1 \le \mu \le s \\ b_\mu & \text{otherwise} \end{cases}.$$

Spike separable sample covariance matrices: outliers

- We assume the followings hold for all $1 \leq i \leq r$ and $1 \leq \mu \leq s$.

$$\widetilde{a}_i > -m_{2c}^{-1}(\lambda_r) \quad \text{or} \quad \widetilde{b}_i > -m_{1c}^{-1}(\lambda_r).$$

- We define the integers $0 \leq r^+ \leq r$ and $0 \leq s^+ \leq s$ such that

$$\widetilde{a}_i \geq -m_{2c}^{-1}(\lambda_r) + n^{-1/3} \quad \text{if and only if} \quad 1 \leq i \leq r^+,$$

and

$$\widetilde{b}_\mu \geq -m_{1c}^{-1}(\lambda_r) + n^{-1/3} \quad \text{if and only if} \quad 1 \leq \mu \leq s^+.$$

Spike separable sample covariance matrices: outliers

- Deterministic locations

$$\theta_1(\widetilde{a}_i) := g_{2c}\left(-a_i^{-1}\right) \quad \text{or} \quad \theta_2(\widetilde{b}_\mu) := g_{1c}\left(-\widetilde{b}_\mu^{-1}\right),$$

  where g_{1c}, g_{2c} are respectively the inverse functions of $m_{1c} : (\lambda_r, \infty) \to (m_{1c}(\lambda_r), 0), m_{2c} : (\lambda_r, \infty) \to (m_{2c}(\lambda_r), 0)$.
- We define the labelling functions $\alpha : \{1, \cdots, p\} \to \mathbb{N}$ and $\beta : \{1, \cdots, n\} \to \mathbb{N}$ as follows. For any $1 \leq i \leq r$, we assign to it a label $\alpha(i) \in \{1, \cdots, r+s\}$ if $\theta_1(\widetilde{a}_i)$ is the $\alpha(i)$-th largest element in $\{\theta_1(\widetilde{a}_i)\}_{i=1}^r \cup \{\theta_2(\widetilde{b}_\mu)\}_{\mu=1}^s$. We also assign to any $1 \leq \mu \leq s$ a label $\beta(\mu) \in \{1, \cdots, r+s\}$ in a similar way. Moreover, we define $\alpha(i) = i + s$ if $i > r$ and $\beta(\mu) = \mu + r$ if $\mu > s$.

Spike separable sample covariance matrices: outliers

- Index notations

$$\mathcal{O} := \{\alpha(i) : 1 \le i \le r\} \cup \{\beta(\mu) : 1 \le \mu \le s\},$$

$$\mathcal{O}^+ := \{\alpha(i) : 1 \le i \le r^+\} \cup \{\beta(\mu) : 1 \le \mu \le s^+\}.$$

- Fluctuation level

$$\Delta_1(\widetilde{a}_i) := \left(\widetilde{a}_i + m_{2c}^{-1}(\lambda_r)\right)^{1/2}, \ \Delta_2(b_\mu) := \left(\widetilde{b}_\mu + m_{1c}^{-1}(\lambda_r)\right)^{1/2}.$$

Outlier and extremal non-outlier eigenvalues

Theorem (D. and Yang, 2019)

$$\left|\widetilde{\lambda}_{\alpha(i)} - \theta_1(\widetilde{a}_i)\right| \prec n^{-1/2}\Delta_1(\widetilde{a}_i), \quad 1 \le i \le r^+,$$

$$\left|\widetilde{\lambda}_{\beta(\mu)} - \theta_2(\widetilde{b}_\mu)\right| \prec n^{-1/2}\Delta_2(\widetilde{b}_\mu), \quad 1 \le \mu \le s^+.$$

Furthermore, for any fixed $\varpi > r + s$, we have

$$|\widetilde{\lambda}_i - \lambda_r| \prec n^{-2/3}, \quad \text{for } \; i \notin \mathcal{O}^+ \text{ and } \; i \le \varpi.$$

- When Δ_1 changes from $n^{-1/3}$ to $O(1)$, we expect a transition for TW to Gaussian.
- (Bao-D.-Wang, 2018), (Bao-D.-Wang-Wang, 2019).
- In general, variance depends on 4th cumulants (both third and fourth moments).

Eigenvalue sticking

Theorem (D. and Yang, 2019)

$$\alpha_+ := \min\left\{ \min_i \left|\widetilde{a}_i + m_{2c}^{-1}(\lambda_r)\right|, \min_\mu \left|\widetilde{b}_\mu + m_{1c}^{-1}(\lambda_r)\right| \right\}.$$

Fix any sufficiently small constant $\tau > 0$. We have that for $1 \le i \le \tau n$,

$$\left|\widetilde{\lambda}_{i+r+s} - \lambda_i\right| \prec \frac{1}{n\alpha_+}, \quad 1 \le i \le \tau n.$$

- Tracy-Widom.

Outlier singular vectors

For $1 \leq i \leq r^+$, $1 \leq j \leq p$ and $1 \leq \nu \leq n$, we define

$$\delta^a_{\alpha(i),\alpha(j)} := |\widetilde{a}_j - \widetilde{a}_i|, \quad \delta^a_{\alpha(i),\beta(\nu)} := \left|\widetilde{b}_\nu + m_{1c}^{-1}(\theta_1(\widetilde{a}_i)\right|.$$

Similarly, for $1 \leq \mu \leq s^+$, $1 \leq j \leq p$ and $1 \leq \nu \leq n$, we define

$$\delta^b_{\beta(\mu),\alpha(j)} := |\widetilde{a}_j + m_{2c}^{-1}(\theta_2(\widetilde{b}_\mu))|, \quad \delta^b_{\beta(\mu),\beta(\nu)} := |\widetilde{b}_\nu - \widetilde{b}_\mu|.$$

Denote
$\delta_{\alpha(i)} = \left(\min_{k:\alpha(k)\neq\alpha(i)} \delta^a_{\alpha(i),\alpha(k)}\right) \wedge \left(\min_{\mu:\beta(\mu)\neq\alpha(i)} \delta^a_{\alpha(i),\beta(\mu)}\right)$

Outlier singular vectors

Theorem (D. and Yang, 2019)

$$|\langle \mathbf{v}_i^a, \widetilde{\boldsymbol{\xi}}_i\rangle|^2 =$$

$$\frac{1}{\widetilde{a}_i}\frac{g_{2c}'(-(\widetilde{a}_i)^{-1})}{g_{2c}(-(\widetilde{a}_i)^{-1})} + O_\prec\left(\frac{1}{n^{1/2}(\widetilde{a}_i + (m_{2c}^{-1}(\lambda_r))^{1/2}} + \frac{1}{n\delta_{\alpha(i)}^2}\right).$$

- Non-overlapping condition

$$\widetilde{a}_i + m_{2c}^{-1}(\lambda_r) \gg n^{-1/3}, \quad \delta_{\alpha(i)} \gg (\widetilde{b}_i + m_{2c}^{-1}(\lambda_r))^{-1/2} n^{-1/2}.$$

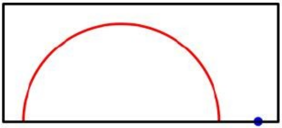
Theorem (D. and Yang, 2019)

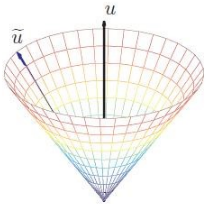
If $\alpha(i) \notin \mathcal{O}^+$, we have

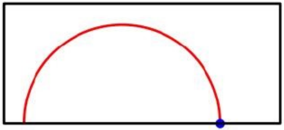
$$|\langle \mathbf{v}_j^a, \widetilde{\boldsymbol{\xi}}_{\alpha(i)}\rangle|^2 \prec \frac{1}{n(|\widetilde{a}_j + m_{2c}^{-1}(\lambda_r)|^2 + n^{-2/3})}.$$

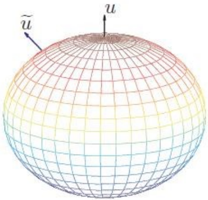
- Right singular vectors. General components.

General picture



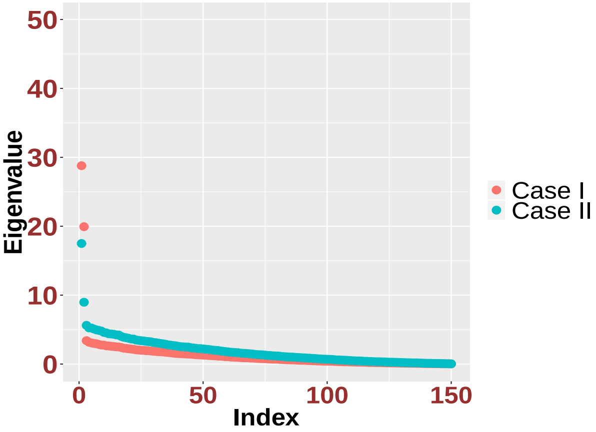




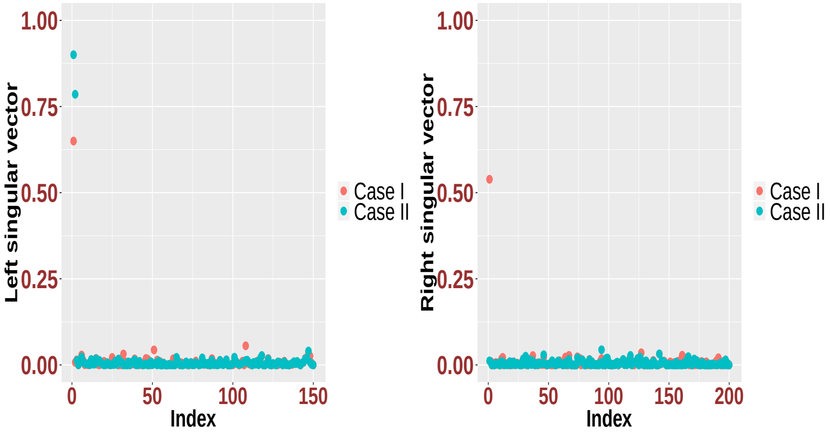


Some statistical remarks: which model?

$$\widetilde{\Sigma}^a = \text{diag}(5, 1, \cdots, 1), \quad \widetilde{\Sigma}^b = \text{diag}(5, 1, \cdots, 1), \qquad \text{(Case I)}$$
$$\widetilde{\Sigma}^a = \text{diag}(3, 2, 1, \cdots, 1), \quad \widetilde{\Sigma}^b = \text{diag}(1, 1, \cdots, 1). \qquad \text{(Case II)}$$



Some statistical remarks: which model?



Some statistical remarks: adaptive estimation

$$\widehat{a}_i := -\left(\frac{1}{n}\sum_{\nu=r+s+1}^{n}\frac{1}{\widetilde{\lambda}_\nu(\widetilde{\mathcal{Q}}_2)-\widetilde{\lambda}_{\alpha(i)}}\right)^{-1}, \quad 1 \leq i \leq r+s.$$

$$\widehat{b}_\mu := -\left(\frac{1}{n}\sum_{k=r+s+1}^{p}\frac{1}{\widetilde{\lambda}_k(\widetilde{\mathcal{Q}}_1)-\widetilde{\lambda}_{\beta(\mu)}}\right)^{-1}, \quad 1 \leq \mu \leq r+s.$$

- Suppose $\widetilde{B} = I_n + \mathcal{M}_n$, where $\mathcal{M}_n$ is a matrix of rank l_n. Then we have that for $1 \leq i \leq r$,

$$\widetilde{a}_i = \widehat{a}_i + O_\prec(l_n n^{-1/2}).$$

  Similarly, if $\widetilde{A}$ is an l_n-rank perturbation of the identity matrix, then for $1 \leq \mu \leq s$,

$$\widetilde{b}_\mu = \widehat{b}_\mu + O_\prec(l_n n^{-1/2}).$$

Anisotropic local laws: (Yang, 2019) and (D. and Yang, 2019)

- $(p+n)\times(p+n)$ self-adjoint block matrix (linear function of X):

$$H \equiv H(X,z) := z^{1/2}\begin{pmatrix} 0 & A^{1/2}XB^{1/2} \\ B^{1/2}X^*A^{1/2} & 0 \end{pmatrix}, \quad z \in \mathbb{C}_+.$$

$$G \equiv G(X,z) := \left(H(X,z) - z\right)^{-1}.$$

- Schur complement formula

$$G(z) = \begin{pmatrix} \mathcal{G}_1 & z^{-1/2}\mathcal{G}_1 Y \\ z^{-1/2}Y^*\mathcal{G}_1 & \mathcal{G}_2 \end{pmatrix} = \begin{pmatrix} \mathcal{G}_1 & z^{-1/2}Y\mathcal{G}_2 \\ z^{-1/2}\mathcal{G}_2 Y^* & \mathcal{G}_2 \end{pmatrix}$$

where $Y := A^{1/2}XB^{1/2}$.

Anisotropic local laws: (Yang, 2019) and (D. and Yang, 2019)

- Spectral domains

$$S(\varsigma_1, \varsigma_2) := \{z = E + \mathrm{i}\eta : \lambda_r - \varsigma_1 \le E \le \varsigma_2 \lambda_r,\ 0 < \eta \le 1\},$$

$$S_0(\varsigma_1, \varsigma_2, \epsilon) := S(\varsigma_1, \varsigma_2) \cap \left\{z = E + \mathrm{i}\eta : \eta \ge n^{-1+\epsilon}\right\},$$

$$z \in S_{out}(\varsigma_2, \epsilon) := \left\{E + \mathrm{i}\eta : \lambda_r + n^{-2/3+\varepsilon} \le E \le \varsigma_2 \lambda_r, \eta \in [0, 1]\right\}.$$

- Convergent limits and control parameters

$$\Pi(z) := \begin{pmatrix} \Pi_1 & 0 \\ 0 & \Pi_2 \end{pmatrix}, \ \Psi(z) := \sqrt{\frac{\operatorname{Im} m_{2c}(z)}{n\eta}} + \frac{1}{n\eta}.$$

$$\Pi_1 := -z^{-1}\left(1 + m_{2c}(z) A\right)^{-1}, \quad \Pi_2 := -z^{-1}(1 + m_{1c}(z) B)^{-1}.$$

Anisotropic local laws: (Yang, 2019) and (D. and Yang, 2019)

- For $z \in S_0(\varsigma_1, \varsigma_2, \epsilon)$, we have

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec \Psi(z).$$

$$|m(z)-m_c(z)|+|m_1(z)-m_{1c}(z)|+|m_2(z)-m_{2c}(z)| \prec (n\eta)^{-1}.$$

- For $z \in S_{out}(\varsigma_2, \epsilon), \kappa = |\operatorname{Re} z - \lambda_r|$

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec \sqrt{\frac{\operatorname{Im} m_{2c}(z)}{n\eta}} \asymp n^{-1/2}(\kappa+\eta)^{-1/4},$$

$$\begin{aligned} &|m(z)-m_c(z)| + |m_1(z)-m_{1c}(z)| + |m_2(z)-m_{2c}(z)| \\ &\prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}. \end{aligned}$$

Technical proofs—Universality

- Three ways to to prove universality (Erdos-Yau)
  1. Comparison method (Erdos-Yau-Yin, Adv, 2010), (Bao-Pan-Zhou, AOS 2015; D.-Yang, 2018, AOAP)
  2. Continuous interpolation (Schelli-Lee, AOAP, 2016) (Johnstone-Zhou, 2018, "discrete version", swapping pair)
  3. Dyson Brownian motion (Erodos-Yau, JAMS;Landu-Yau, 2018)
- Three-step-strategy using DM
  1. Local laws for the random matrix ensemble H+rigidity of eigenvalues: **Proved**
  2. University of $H_t = H + \sqrt{t}G$, G is GOE, $t = o(1)$. t is the time such that **local** eigenvalue statistics reach equilibrium. Needs some basic discussion from free probability theory.
  3. A density argument comparing the eigenvalue statistics between Step 2 and the random matrix ensemble.

Technical proofs–Universality

$$H_t := \begin{pmatrix} 0 & W+\sqrt{t}X \\ (W+\sqrt{t}X)^\top & 0 \end{pmatrix},$$

$$W = \begin{pmatrix} D & 0 \end{pmatrix}, D^2 = \mathrm{diag}(d_1, \cdots, d_p).$$

- Denote μ_i as the unique strong solution to the SDE

$$d\mu_i = 2\sqrt{\mu_i}\frac{dB_i}{\sqrt{n}} + \left(\frac{1}{N} \sum_{l \neq k} \frac{\lambda_k + \lambda_l}{\lambda_k - \lambda_l} + \frac{p}{n} \right) dt,$$

  $\mu_i(0)$: Wishart matrices.
- For some i_0, we denote λ_i as

$$d\mu_i = 2\sqrt{\mu_i}\frac{dB_{i-i_0+1}}{\sqrt{n}} + \left(\frac{1}{N} \sum_{l \neq k} \frac{\lambda_k + \lambda_l}{\lambda_k - \lambda_l} + \frac{p}{n} \right) dt,$$

  with $\lambda_i(0) = \lambda_i(\gamma_0 H_{t_0})$

Techincal proofs-Universality

Theorem (D. and Yang, 2019+)

For $t_0 = n^{-1/3+\epsilon_0}$ *and* $t_1 = n^{-1/3+\epsilon_1}, 0 < \epsilon_1 < \epsilon_0/200,$

$$|(\lambda_{i_0+i-1}(t_1) - E_\lambda(t_1)) - (\mu_i - E_\mu(t_1))| \prec n^{-2/3-\delta},$$

for any bounded i.

- $E_\lambda(\cdot)$ is the edge. Clearly, $E_\mu(t) = (1+\sqrt{d})^2\sqrt{1+t}$.
- To find γ and $E_\mu()$, we need to study the macroscopic structure of signal-plus-noise matrix.
- Free probability: rectangular free convolution and subordiation function $\Rightarrow$ square root behavior.

Technical proofs–Outliers

$$\widetilde{H}(X,z) = z^{1/2}\begin{pmatrix} 0 & \widetilde{A}^{1/2}X\widetilde{B}^{1/2} \\ \widetilde{B}^{1/2}X^*\widetilde{A}^{1/2} & 0 \end{pmatrix}, \quad z \in \mathbb{C}_+ \cup \mathbb{R}.$$

$$\mathbf{U} = \begin{pmatrix} V_o^a & 0 \\ 0 & V_o^b \end{pmatrix}, \quad \mathcal{D} = \begin{pmatrix} D^a(D^a+1)^{-1} & 0 \\ 0 & D^b(D^b+1)^{-1} \end{pmatrix}.$$

- Master equation

$$\det\left(\mathcal{D}^{-1} + x\mathbf{U}^* G(x)\mathbf{U}\right) = 0.$$

- Anisotropic local laws $\Rightarrow$

$$\det(\mathcal{D}^{-1} + x\mathbf{U}^*\Pi(x)\mathbf{U}) = 0 \quad \Rightarrow$$

$$\prod_{i=1}^{r}\left(\frac{d_i^a+1}{d_i^a} - \frac{1}{1+m_{2c}(x)\sigma_i^a}\right)\prod_{\mu=1}^{s}\left(\frac{d_\mu^b+1}{d_\mu^b} - \frac{1}{1+m_{1c}(x)\sigma_\mu^b}\right) = 0.$$

Outlier eigenvectors

$$\langle \mathbf{v}_i^a, \widetilde{\boldsymbol{\xi}}_i \mathbf{v}_j^a \rangle = -\frac{1}{2\pi \mathrm{i}} \oint_{g_{2c}(\Gamma)} \langle \mathbf{v}_i^a, \widetilde{G}(z) \mathbf{v}_j^a \rangle dz,$$

$$\mathbf{U}^* \widetilde{G}(z) \mathbf{U} = \widetilde{\mathcal{D}}^{1/2} \left[\mathbf{U}^* G(z) \mathbf{U} - z \mathbf{U}^* G(z) \mathbf{U} \frac{1}{\mathcal{D}^{-1} + z \mathbf{U}^* G(z) \mathbf{U}} \mathbf{U}^* G(z) \mathbf{U} \right] \widetilde{\mathcal{D}}^{1/2},$$

where

$$\widetilde{\mathcal{D}} := \begin{pmatrix} (1 + D^a)^{-1} & 0 \\ 0 & (1 + D^b)^{-1} \end{pmatrix}.$$

References

Xiucai Ding and Fan Yang (2019)
Spiked seperable sample covariance matrices and principal components

Xiucai Ding and Fan Yang (2019+)
Tracy-Widom distribution of seperable sample covariance matrices