Eigen(singular)-vector distribution of anisotropic random matrices

Jun Yin

UW-Madison

HKU, Jan-2015

Joint work with A. Knowles

Anisotropic matrix:

case 1:

$$Y := TX$$

where T is deterministic matrix and X is i.i.d random matrix such that $\mathbb{E}X_{ij} = 0$.

case 2:

$$Y := X + A$$

where A is deterministic symmetric matrix and X is a Wigner random matrix.

Asymptotically close

We say that $\{\mathbf{u}_n\}_{n=1}^{\infty}$ is asymptotic close to $\{\mathbf{v}_n\}_{n=1}^{\infty}$, $n \to \infty$ if

1. $\mathbf{u}_n \in \mathbb{R}^{d_n}$, $\mathbf{v}_n \in \mathbb{R}^{d_n}$, $d_n \to \infty$

2. for any sequence of fixed dimensional subspace

$$\left\{ S_n \subset \mathbb{R}^{d_n} : \quad \mathrm{dim}(S_n) = k \right\}_{n=1}^{\infty}$$

we have

$$\lim_n f\left(\mathbf{u}_n\Big|_{S_n}\right) - f\left(\mathbf{v}_n\Big|_{S_n}\right) = 0$$

for any continuous bounded function $f : \mathbb{R}^k \to \mathbb{R}$.

Isotropic matrix:

Wigner matrix: Let W_n be a $n \times n$ Wigner matrix and $\mathbf{u}_{k,n}$ be the normalized k-th eigenvector of W_n. For sequence $k_n \in [[1, n]]$

$$\sqrt{n}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}(0, I_n)$$

Edge case: $k_n = o(n)$, [Knowles, Y, 2010] Green's function comparison method.

Bulk case: [Yau, Bourgade, 2014] Moment flow method.

Note: the "limit" is independent of the entry-distribution of W_n.

Isotropic "limit"

For Wigner matrix, any eigenvector is asymptotically close to

$$\mathbf{v}_n \sim \mathcal{N}(0, I_n)$$

which is orthogonal transform invariant, i.e,

$$\mathbf{v}_n \sim O\mathbf{v}_n$$

It is easy to understand this result in the Gaussian case, since

$$W^G \sim O^* W^G O$$

I.i.d matrix $X_{m\times n}$:

Recall: non-zero singular values of X are non-zero eigenvalues of XX^* and X^*X.

Non-zero-left-singular vectors of X are eigenvectors of XX^* whose eigenvalues are not zero.

Non-zero-right-singular vectors of X are eigenvectors of X^*X whose eigenvalues are not zero.

Let $\mathbf{u}_{k,n}$ be the k-th non-zero-left-singular vector of $X_{m\times n}$, $\log m \sim \log n$. For sequence k_n

$$\sqrt{m}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}(0, I_m)$$

Let $\mathbf{u}_{k,n}$ be the k-th non-zero-right-singular vector of $X_{m\times n}$, $\log m \sim \log n$.

$$\sqrt{n}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}(0, I_n)$$

Singular vector of anisotropic matrix:

Let $Y_{m\times n} = TX$, where T is $m \times \hat{m}$ deterministic matrix and X is i.i.d. $(\hat{m} \times n)$ random matrix such that $\mathbb{E}X_{ij} = 0$. We assume $\log m \sim \log n$.

Let $\mathbf{u}_{k,n}$ be the k-th right singular vector of $Y_{m\times n}$.

$$\sqrt{n}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}(0, I_n)$$

Let $\mathbf{u}_{k,n}$ be the k-th left singular vector of $Y_{m\times n}$.

$$\sqrt{m}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}\left(0, \frac{m/n}{\gamma_{k_n,n}} \left|\frac{T}{1 + m(\gamma_{k_n,n})|T|^2}\right|^2\right)$$

where $\lambda_{k_n,n}$ is the singular value w.r.t. $\mathbf{u}_{k_n,n}$ and $\gamma_{k_n,n}$ is its classical location. And

$$m(x) \in \mathbb{C} : m(x)^{-1} = -x + \frac{1}{n} \operatorname{Tr} \frac{|T|^2}{1 + m(x)|T|^2}, \quad \operatorname{Im} m(x) \geqslant 0$$

[Knowles, Y, 2014]

$$\sqrt{m}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}\left(0, \frac{m/n}{\gamma_{k_n,n}}\left|\frac{T}{1+m(\gamma_{k_n,n})|T|^2}\right|^2\right)$$

$$m(x) \in \mathbb{C} : m(x)^{-1} = -x + \frac{1}{n}\operatorname{Tr}\frac{|T|^2}{1+m(x)|T|^2}, \quad \operatorname{Im} m(x) \geqslant 0$$

Note: the "limit" only depends on TT^*.

Note: Since $T = UDV$, one only need to study the singular vector of $Y' = DVX$.

$$\sqrt{m}\mathbf{u}'_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}\left(0, \frac{m/n}{\gamma_{k_n,n}}\left|\frac{D}{1+m(\gamma_{k_n,n})|D|^2}\right|^2\right)$$

where the covariance matrix is diagonal, as people believed.

Let $Y_{n\times n} = X + A$, where A is $n \times n$ deterministic symmetric matrix and X is a Wigner matrix such that $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1/n$.

Let $\mathbf{u}_{k,n}$ be the k-th eigen-vector of $Y_{m\times n}$.

$$\sqrt{n}\mathbf{u}_{k_n,n} \to \mathbf{v}_n \sim \mathcal{N}\left(0, \quad \frac{m(\gamma_{k_n,n})^{-1}}{m(\gamma_{k_n,n}) + A - \gamma_{k_n,n}}\right)$$

where

$$m(x) \in \mathbb{C} : m(x) = \frac{1}{n}\,\mathrm{Tr}(-m(x) + A - x)^{-1}, \quad \mathrm{Im}\, m(x) \geqslant 0$$

[Knowles, Y, 2014]

Universality:

For $Y = TX$, the asymptotic behavior of singular-vectors are independent of

1. The distribution of entries of X,

2. The right singular vector of T, i.e., only depends on TT^*.

(Actually it depends on left singular vector in a very trivial way)

To understand this universality phenomenon, we developed a new comparison method: self-consistent comparison method.

Old: Lindeberg comparison method

For comparing two $m \times n$ random matrices X_{ij} and $\widetilde{X}_{ij}$, one constructs $mn+1$ new matrices $Y^{[k]}$. First, define an order for entries:

For example: We say (i, j) is the $\phi(i, j)$-th entry, with $\phi(i, j) = i \times n + j$.

Random matrices $Y^{[k]}$ are defined as: the "first" k entries of $Y^{[k]}$ have the same distribution as entries of X, the others have the same distribution as those of $\widetilde{X}$.

Advantage: The difference between $Y^{[k]}$ and $Y^{[k+1]}$ is just one entry.

$$F(X) - F(\widetilde{X}) = \sum_k \Big(F(Y^{[k]}) - F(Y^{[k+1]}) \Big)$$

Perturbation theory

Disadvantage:

1. The local law (or some prior bound) of $Y^{[k]}$ are needed for this method, i.e., one need to use other methods to derive the local law for $Y^{[k]}$ first.

2. The distance between $Y^{[k]}$ and $Y^{[k+1]}$ can not be arbitrary small.

We need a continuous self consistent version of comparison method.

Self consistent comparison method

Suppose there are two collections $X^0 = (X_i^0)_{i\in\mathcal{I}}$ and $X^1 = (X_i^1)_{i\in\mathcal{I}}$ of random variables indexed by some finite set $\mathcal{I}$. We need to estimate

$$\mathbb{E}F(X^1) - \mathbb{E}F(X^0)$$

For example: we can derive large derivation bound by choosing

$$F(X) := (|\mathcal{I}|^{-1/2}\sum_{i\in\mathcal{I}} X_i)^{2p}$$

We construct a continuous family $(X^\theta)_{\theta\in[0,1]}$ such that

$$X_i^\theta \sim \chi_i^\theta X_i^1 + (1-\chi_i^\theta)X_i^0$$

where χ_i^θ are i.i.d r.v. such that $\mathbb{P}(\chi_i^\theta = 1) = \theta$ and $\mathbb{P}(\chi_i^\theta = 0) = 1-\theta$. X_i^θ has probability θ to have the same distribution as X_i^1, otherwise it has distribution of X_i^0.

$$\rho_{X_i^\theta} = \theta\rho_{X_i^1} + (1-\theta)\rho_{X_i^0}$$

With

$$X_i^\theta \sim \chi_i^\theta X_i^1 + (1 - \chi_i^\theta) X_i^0$$

we have

$$\mathbb{E}F(X^1) - \mathbb{E}F(X^0) = \int \frac{\partial}{\partial\theta} \mathbb{E}F(X^\theta) \mathrm{d}\theta$$

$$= \int \sum_{i \in \mathcal{I}} \left(\mathbb{E}F(X^{\theta,i,1}) - \mathbb{E}F(X^{\theta,i,0}) \right) \mathrm{d}\theta$$

where

$X^{\theta,i,1}$ is defined as X^θ except $X_i^{\theta,i,1} \sim X_i^1$

$X^{\theta,i,0}$ is defined as X^θ except $X_i^{\theta,i,0} \sim X_i^0$

With Taylor's expansion, we can replace $\mathbb{E}F(X^{\theta,i,1})$, $\mathbb{E}F(X^{\theta,i,0})$ back to $\mathbb{E}F(X^\theta)$.

We have an identity of the form

$$\frac{\partial}{\partial\theta}\mathbb{E}F(X^\theta) = \sum_{n\geqslant 1} K_n \sum_{i\in\mathcal{I}} \mathbb{E}\left(\frac{\partial}{\partial X_i^\theta}\right)^n F(X^\theta)\,, \tag{1}$$

where the constant K_n depends only on the first n moments of X^0 and X^1, $K_n = 0$ if the first n moments of X^1 and X^0 match.

Note: it is self-consistent in the sense that the right-hand side depends on the quantities to be estimated.

It is level-1: only X^θ appears on both sides.

Simple Example:

Let's prove CLT by choosing $F(X) := f\Big(|\mathcal{J}|^{-1/2}\sum_{i\in\mathcal{J}} X_i\Big)$.

Suppose $\mathbb{E}X_i^1 = \mathbb{E}X_i^0 = 0$ and $\mathbb{E}|X_i^1|^2 = \mathbb{E}|X_i^0|^2 = 1$.

$$\frac{\partial}{\partial\theta}\mathbb{E}F(X^\theta) \;=\; \sum_{1\leqslant n\leqslant C} K_n \sum_{i\in\mathcal{J}} \mathbb{E}\left(\frac{\partial}{\partial X_i^\theta}\right)^n F(X^\theta) + error\,, \tag{2}$$

$$= \sum_{3\leqslant n\leqslant C} K_n \sum_{i\in\mathcal{J}} \mathbb{E}\left(\frac{\partial}{\partial X_i^\theta}\right)^n F(X^\theta) + error$$

Since $\left(\frac{\partial}{\partial X_i^\theta}\right)^n F(X^\theta) = O(|\mathcal{J}|)^{-n/2}$, so we obtain that

$$\frac{\partial}{\partial\theta}\mathbb{E}F(X^\theta) = o(1)$$

which implies that

$$\mathbb{E}F(X^1) = \mathbb{E}F(X^0) + o(1)$$

Recall

$$\frac{\partial}{\partial\theta}\mathbb{E}F(X^\theta) \;=\; \sum_{n\geqslant 1} K_n \sum_{i\in\mathcal{I}} \mathbb{E}\left(\frac{\partial}{\partial X_i^\theta}\right)^n F(X^\theta)\,, \tag{3}$$

1. $\mathbb{E}\big(\frac{\partial}{\partial X_i^\theta}\big)^n F(X^\theta)$ typically decays rapidly with increasing n – as is already apparent in the simple case $F(X) := \big(|\mathcal{I}|^{-1/2}\sum_{i\in\mathcal{I}} X_i\big)^p$ for $p \in 2\mathbb{N}$

2. We introduce a family of functions $(F_\alpha(X))_{\alpha\in\mathcal{A}}$ that

$$\frac{\partial}{\partial\theta}\mathbb{E}F_\alpha(X^\theta) \;\leqslant\; C \sup_{\beta\in\mathcal{A}} \mathbb{E}F_\beta(X^\theta) + (\text{small error terms})\,, \tag{4}$$

It is level-2: only $(F_\alpha(X))_{\alpha\in\mathcal{A}}$ appears on both sides.

All previous proofs with Lindeberg comparison method can be replaced with this new comparison method.

Application on random matrix: local anisotropic law

For anisotropic law: We choose

$$F_{\mathbf{v}}(X) := (\mathbf{v}, (G(X) - \Pi)\,\mathbf{v})^{2p}$$

and $\mathcal{A} = \{\mathbf{v}, \mathbf{e}_i, 1 \leqslant i \leqslant M\}$.

$$\frac{\partial}{\partial\theta}\mathbb{E}F_\alpha(X^\theta) \leqslant C \sup_{\beta\in\mathcal{A}} \mathbb{E}F_\beta(X^\theta) + (\text{small error terms})\,, \tag{5}$$

Since we can bound

$$\mathbb{E}F_{\mathbf{v}}(X^{Gaussian})$$

with the method mentioned in Knowles's talk, we can use self-consistent comparison method to bound $\mathbb{E}F_{\mathbf{v}}(X)$ for general X.

Comparison between $Y^0 = T^0 X$ and $Y^1 = T^1 X$.

Let T^0, T^1 be $m \times m'$ deterministic matrices and X be $m' \times n$ i.i.d. matrix.

$$Y^0 = \sum_{ij} (T^0 \cdot \mathbf{e}_{ij}) X_{ij}, \quad Y^1 = \sum_{ij} (T^1 \cdot \mathbf{e}_{ij}) X_{ij}$$

Here $\mathbf{e}_{ij}$ is $m' \times n$ matrix with all zero entries except that (i, j) entry equals to 1.

$$Y^\theta = \sum_{ij} \Big(\chi_{ij}^\theta (T^1 \cdot \mathbf{e}_{ij}) X_{ij} + (1 - \chi_{ij}^\theta)(T^0 \cdot \mathbf{e}_{ij}) X_{ij} \Big)$$

where χ_{ij}^θ are i.i.d r.v. such that $\mathbb{P}(\chi_{ij}^\theta = 1) = \theta$ and $\mathbb{P}(\chi_{ij}^\theta = 0) = 1 - \theta$.

Self consistent equation for $Y^0 = T^0 X$ and $Y^1 = T^1 X$.

$$\frac{\partial}{\partial\theta}\mathbb{E}F(Y^\theta) \tag{6}$$

$$= \sum_{n,m\geqslant 1} K_{n,m} \sum_{ij} \mathbb{E}\left(\left(P_{ij}^{T^0}\right)^n \left(P_{ij}^{T^1}\right)^m - \left(P_{ij}^{T^0}\right)^m \left(P_{ij}^{T^1}\right)^n\right) F(Y^\theta)\,, \tag{7}$$

where $K_{n,m}$ only depends on the first $n+m$ moment of X's entries.

$$P_{ij}^A = \left(A^T \cdot \nabla_Y\right)_{ij}$$

Application on random matrix: singular vector

Using moment flow, we can derive the distribution of singular vector of

$$TX + \sqrt{t}X^G = (T, \sqrt{t}I)(X, X^G)^T, \quad t \geqslant \min(m,n)^{-1+\varepsilon}$$

Suppose $T \geqslant 0$, with the new comparison method, we know the distribution of singular vector of

$$\widehat{T}X = (\widehat{T}, 0)(X, X^G)^T, \quad |\widehat{T}|^2 = |T|^2 + tI$$

Note: $\|(T, \sqrt{t}I) - (\widehat{T}, 0)\|$ is very small.

The general $\widehat{T}$ case can be derived similarly with some linear algebra argument.

Moment flow method

First used by Bourgade and Yau for Wigner matrix.

The basic idea is with extended Dyson Brownian motion, one can derive a dynamic equation about

$$f_{k_1k_2\cdots k_m}^{a_1a_2\cdots a_m} := C_{a_1a_2\cdots a_m}\mathbb{E}\left\langle \mathbf{u}_{k_1}(t), \mathbf{v}\right\rangle^{2a_1}\left\langle \mathbf{u}_{k_2}(t), \mathbf{v}\right\rangle^{2a_2}\cdots\left\langle \mathbf{u}_{k_m}(t), \mathbf{v}\right\rangle^{2a_m}$$

where $\mathbf{u}_k(t)$ is the k-the eigenvector of $W + \sqrt{t}W^G$.

With maximum principle:

$$\mathcal{F}_a(t) := \max_{\sum_i a_i = a} \max_{k_1k_2\cdots k_m} f_{k_1k_2\cdots k_m}^{a_1a_2\cdots a_m}, \quad a \in \mathbb{N}$$

$$\partial_t \mathcal{F}_a(t) \leqslant -N^{1/3}\left(\mathcal{F}_a(t) - 1\right)$$

Then $\mathcal{F}_a(t) \leqslant 1 + o(1)$ for $t \gg N^{-1/3}$. Similarly with minimum principle, one obtains that

$$f_{k_1k_2\cdots k_m}^{a_1a_2\cdots a_m} = 1 + o(1), \quad t \gg N^{-1/3}$$

For anisotropic matrices,

$$f^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m} \not\to 1, \quad f^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m} \to g^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m} \quad (deterministic)$$

$$\mathcal{F}_a(t) := \sum_{\sum_i a_i = a} \sum_{k_1 k_2 \cdots k_m} \left(f^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m} - g^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m} \right)^{2p}, \quad p \in \mathbb{N}$$

Goal:

$$\partial_t \mathcal{F}_a(t) \leqslant -N^{1/3} \mathcal{F}_a(t)$$

Unfortunately $g^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m}$ in $\mathcal{F}_t$ can not be deterministic, since there is a singular term in $\partial_t \mathcal{F}_a(t)$,

$$\frac{g^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k_m} - g^{a_1 a_2 \cdots a_m}_{k_1 k_2 \cdots k'_m}}{\lambda_{k_m} - \lambda_{k'_m}}$$

So we define

$$g_{k_1 k_2 \cdots k_m}^{a_1 a_2 \cdots a_m} := g_{k_1 k_2 \cdots k_m}^{a_1 a_2 \cdots a_m}\left(\lambda_{k_1}, \lambda_{k_2}, \cdots, \lambda_{k_m}\right)$$

which solve the issue of the following singular term

$$\frac{g_{k_1 k_2 \cdots k_m}^{a_1 a_2 \cdots a_m} - g_{k_1 k_2 \cdots k_m'}^{a_1 a_2 \cdots a_m}}{\lambda_{k_m} - \lambda_{k_m'}}$$

Then differential equation becomes stochastic differential equation.

$$\partial_t \mathcal{F}_a(t) \leqslant -N^{1/3} \mathcal{F}_a(t) + \text{error term} + O(1)\mathrm{d}B$$

Advantage: The limit of $f_{k_1 k_2 \cdots k_m}^{a_1 a_2 \cdots a_m}(t)$ is allowed to depend on $k_1 k_2 \cdots k_m$ and $a_1 a_2 \cdots a_m$.

Further applications will appear in future work.

Thank you