Asymptotic behavior of high-dimensional Gaussian correlated Wishart matrices

Ivan Nourdin
University of Luxembourg

XIV Brunel – Bielefeld Workshop on RMT
Dec. 14-15, 2018

This presentation is based on a joint work with my former PhD student **Guangqu Zheng** (University of Kansas).

1. Introduction to the problem

A REFERENCE

- Bubeck & Ganguly, 2015 : “Entropic CLT and phase transition in high-dimensional Wishart matrices”, to appear in *International Mathematics Research Notices*.
- In their section "open problems", they wrote “*A natural alternative route to prove Theorem 1 (or possibly a variant of it with a different metric) would be to use Stein’s method*”.

Preliminary setting

- Let

$$\mathcal{X}_{n,d} = (X_{ij})_{\substack{1\le i\le n \\ 1\le j\le d}} = \begin{pmatrix} X_{11} & \dots & X_{1d} \\ \vdots & & \vdots \\ X_{n1} & \dots & X_{nd} \end{pmatrix}$$

  be a **rectangular** random matrix $n \times d$ such that all the entries are independent $N(0,1)$ random variables.
- Consider the **Wishart matrix**

$$\boxed{\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T}.$$

- Caution : d = sample size, n = number of features.

ASYMPTOTIC BEHAVIOR WHEN n IS FIXED

- Suppose n is **fixed** (n = number of features)
- From the **law of large numbers** we deduce that, as $d \to \infty$,

$$\boxed{\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T \overset{\text{a.s.}}{\to} I_n}.$$

- From the multivariate **CLT** we deduce that, as $d \to \infty$,

$$\boxed{W_{n,d} := \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - I_n\right) \overset{\text{law}}{\to} M_n := \begin{pmatrix} G_{11} & \dots & G_{1n} \\ \vdots & & \vdots \\ G_{n1} & \dots & G_{nn} \end{pmatrix}},$$

where the G_{ij}, $j \geq i$ are independent, $G_{ii} \sim N(0,2)$ and $G_{ij} = G_{ji} \sim N(0,1), i \neq j$.

ASYMPTOTIC BEHAVIOR FOR LARGE n AND d

- In this talk, we are interested in the case when **BOTH** n **AND** d are large.
- What happens in such a situation ?

Marchenko-Pastur

- Let $\mathcal{X}_{n,d}$ be a $n \times d$ **matrix** whose entries are iid **random** variables that are centered with unit variance.
- Let $\lambda_1(n,d) \leq \ldots \leq \lambda_n(n,d)$ be the eigenvalues of $\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T$.
- The **spectral measure** of $\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T$ is $\boxed{\mu_{n,d} = \frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i(n,d)}}$.
- **Theorem** (Marchenko-Pastur, 1967) : If $n, d \to \infty$ are such that $n/d \to c \in]0, \infty[$, then $\mu_{n,d} \Rightarrow \mu_c$, where

$$\boxed{\mu_c := (1 - \frac{1}{c})_+\delta_0 + \frac{1}{c}\frac{1}{2\pi x}\sqrt{(a_+ - x)(x - a_-)}\mathbf{1}_{[a_-,a_+]}(x)dx},$$

  with $a_\pm = (1 \pm \sqrt{c})^2$.
- It is also possible to derive results at the **second order** (see, e.g., Chatterjee and references therein).
- What about $n/d \to 0$? $\to \infty$?

Another viewpoint

- Let $\mathcal{X}_{n,d}$ be a **random matrix** $n \times d$ whose entries are **independent** $N(0,1)$.
- Set $\boxed{W_{n,d} := \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - I_n\right)}$.
- Let M_n be of size $n \times n$ and distributed according to the **GOE distribution** (i.e. $M_n(i,j) = M_n(j,i)$, $M_n(i,i) \sim N(0,2)$, $M_n(i,j) \sim N(0,1)$, $i \neq j$, and independence)
- **Theorem** (Bubeck-Ding-Eldan-Rácz, Jiang-Li independently, 2015) :

$$\boxed{\text{If } \frac{n^3}{d} \to \left\{ \begin{array}{c} 0 \\ \infty \end{array} \right\} \text{ then } d_{TV}(W_{n,d}, M_n) \to \left\{ \begin{array}{c} 0 \\ 1 \end{array} \right\}.}$$

Two other references

- Bubeck & Ganguly (2015) : extension of the previous result to the log-concave case : $X_{ij} \sim e^{-\varphi}dx$ with φ convex. (Independence is crucial in their approach.)

- Racz & Richey (2016) : the transition is smooth, namely :

$$\text{If } \frac{n^3}{d} \to c \text{ then } d_{TV}(W_{n,d}, M_n) \to \frac{2}{\sqrt{\pi}} \int_0^{4\sqrt{\frac{c}{3}}} e^{-t^2} dt.$$

2. My work with Guangqu

OUR STUDY

- We consider a matrix $\mathcal{X}_{n,d}$ of size $n \times d$ whose entries are **Gaussian** $N(0,1)$. We suppose that the raws are **independent**, but that the columns are **correlated**.
- More precisely : $\mathcal{X}_{n,d} = (X_{ij})_{\substack{1\le i\le n\\1\le j\le d}} = \begin{pmatrix} X_{11} & \dots & X_{1d} \\ \vdots & & \vdots \\ X_{n1} & \dots & X_{nd} \end{pmatrix}$,

  where $(X_{ij})_{i,j\geq 1}$ is a Gaussian family satisfying

$$\boxed{\mathbb{E}[X_{ij}X_{i'j'}] = \mathbf{1}_{\{i=i'\}} s(j-j')}, \quad \text{with } s(0)=1.$$

  (Note that *full independence* corresponds to $s(k) = \mathbf{1}_{\{k=0\}}$.)
- The corresponding (normalized) Wishart matrix is still

$$\boxed{W_{n,d} = \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - I_n\right)}.$$

n FIXED AND $d \to \infty$

Proposition : We fix n and we suppose $s \in \ell^2(\mathbb{Z})$. Then

$$\boxed{W_{n,d} \overset{\text{law}}{\to} G_n^{(s)} \quad \text{as } d \to \infty \text{ in } \mathcal{M}_n(\mathbb{R})},$$

where $G_n^{(s)} = (G_{ij}^{(s)})_{1 \le i,j \le n}$, with $G_{ij}^{(s)} = G_{ji}^{(s)}$, $G_{ii}^{(s)} \sim N(0, 2\|s\|^2_{\ell^2(\mathbb{Z})})$, $G_{ij}^{(s)} \sim N(0, \|s\|^2_{\ell^2(\mathbb{Z})})$, $i < j$, and independence.

Proof. We use Nualart-Peccati and Peccati-Tudor :

1. $W_{n,d}(i,j) = I_2(f_{ij}^{(d)})$ and $\mathbb{E}[W_{n,d}(i,j)] \to \|s\|^2_{\ell^2(\mathbb{Z})}(1 + \mathbf{1}_{\{i=j\}})$.
2. $\forall i, j : \quad \|f_{ij}^{(d)} \otimes_1 f_{ij}^{(d)}\| \to 0$ as $d \to \infty$.
3. $\mathbb{C}\mathrm{ov}(W_{n,d}) \to \mathbb{C}\mathrm{ov}(G_n^{(s)})$ by raw independence. $\square$

THEOREM 1

Theorem 1 (Nourdin, Zheng, 2018) : We suppose $s \in \ell^2(\mathbb{Z})$. Then there exists $c_s > 0$ (explicit and only depending of s) such that, for all $n, d \geq 1$:

$$\boxed{d_{\text{Wass}}\left(W_{n,d}, G_n^{(s)}\right) \leq c_s \sqrt{\frac{n^3}{d}\left(\sum_{k=1}^{d} |s(k)|^{\frac{4}{3}}\right)^3}}.$$

Remarks

- If $s \in \ell^{\frac{4}{3}}(\mathbb{Z})$ then one recover the quantity $\frac{n^3}{d}$ of Bubeck *et al* :

$$d_{\mathrm{Wass}}\left(W_{n,d}, G_n^{(s)}\right) = O\left(\sqrt{\frac{n^3}{d}}\right).$$

  In particular, it holds true for $s(k) = \mathbf{1}_{\{k=0\}}$, that is, in case of full independence.
- If

$$\boxed{s(k) = \frac{1}{2}\left(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}\right) \sim \mathrm{cst}|k|^{2H-2}}$$

  (**fractional Gaussian noise**), then

$$\frac{1}{d}\left(\sum_{k=1}^{d} |s(k)|^{\frac{4}{3}}\right)^3 \sim \mathrm{cst}\begin{cases} 1/d & \text{if } 0 < H < \frac{5}{8} \\ (\log d)^3/d & \text{if } H = \frac{5}{8} \\ d^{8H-6} & \text{if } \frac{5}{8} < H < 1 \end{cases}.$$

Remarques

As a result, since $s \in \ell^2(\mathbb{Z})$ iff $0 < H < \frac{3}{4}$:

- If $0 < H < \frac{5}{8}$: $W_{n,d} \approx G_n^{(s)}$ as soon as $\frac{n^3}{d} \to 0$;
- If $H = \frac{5}{8}$: $W_{n,d} \approx G_n^{(s)}$ as soon as $\frac{n^3 \log^3 d}{d} \to 0$;
- If $\frac{5}{8} < H < \frac{3}{4}$: $W_{n,d} \approx G_n^{(s)}$ as soon as $\frac{1}{d}\, n^{\frac{3}{6-8H}} \to 0$.

What about $H > \frac{3}{4}$?

Theorem 2

- Let $H \in (\frac{3}{4}, 1)$ and let $(X_{ij})_{i,j\geq 1}$ be a centered Gaussian family such that

$$\mathbb{E}[X_{ij}X_{i'j'}] = \mathbf{1}_{\{i=i'\}} \times \frac{1}{2}\left(|j-j'+1|^{2H} + |j-j'-1|^{2H} - 2|j-j'|^{2H}\right).$$

- We set

$$\mathcal{X}_{n,d} = (X_{ij})_{\substack{1\leq i\leq n \\ 1\leq j\leq d}} = \begin{pmatrix} X_{11} & \dots & X_{1d} \\ \vdots & & \vdots \\ X_{n1} & \dots & X_{nd} \end{pmatrix}.$$

- The corresponding normalized Wishart matrix becomes

$$\boxed{\widehat{W}_{n,d} = d^{2-2H}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^{T} - I_n\right)}.$$

THEOREM 2

- Let $R_n^{(H)}$ be the so-called **Wishart-Rosenblatt matrix** of size $n \times n$ and parameter H, defined as follows

$$R_n^{(H)} = L^2 - \lim_{d\to\infty} \begin{pmatrix} S_{11}(d) & \dots & S_{1n}(d) \\ \vdots & & \vdots \\ S_{n1}(d) & \dots & S_{nn}(d) \end{pmatrix},$$

where

$$S_{ij}(d) = \begin{cases} d\sum_{p=0}^{d-1} \left(B^i_{\frac{p+1}{d}} - B^i_{\frac{p}{d}}\right)\left(B^j_{\frac{p+1}{d}} - B^j_{\frac{p}{d}}\right) & \text{si } i \neq j \\ d\sum_{p=0}^{d-1} \left(\left(B^i_{\frac{p+1}{d}} - B^i_{\frac{p}{d}}\right)^2 - d^{-2H}\right) & \text{si } i = j \end{cases}$$

with $B^1, \dots, B^n$ independent fractional Brownian motions of Hurst index H.

THEOREM 2

Theorem 2 (Nourdin, Zheng, 2018). There exists $c_H > 0$ such that, for all $n, d \geq 1$:

$$\boxed{d_{\text{Wass}}\left(\widehat{W}_{n,d}, R_n^{(H)}\right) \leq c_H\, n\, d^{\frac{3-4H}{2}}}.$$

Sketch of the proof of Theorem 1, that gives the existence of $c_s > 0$ (only depending on s) such that, for all $n, d \geq 1$:

$$\boxed{d_{\text{Wass}}\left(W_{n,d}, G_n^{(s)}\right) \leq c_s \sqrt{\frac{n^3}{d}\left(\sum_{k=1}^{d} |s(k)|^{\frac{4}{3}}\right)^3}}, \quad \text{where}$$

$$W_{n,d} = \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - I_n\right) \quad \text{with} \quad \mathcal{X}_{n,d} = (X_{ij})_{\substack{1\leq i\leq n \\ 1\leq j\leq d}}$$

for $(X_{ij})_{i,j\geq 1}$ a centered Gaussian family of the form

$$\mathbb{E}[X_{ij}X_{i'j'}] = \mathbf{1}_{\{i=i'\}}s(j-j'), \quad \text{with } s(0) = 1,$$

and where $G_n^{(s)} = (G_{ij}^{(s)})_{1\leq i,j\leq n}$, with $G_{ij}^{(s)} = G_{ji}^{(s)}$, $G_{ii}^{(s)} \sim N(0, 2\|s\|^2_{\ell^2(\mathbb{Z})})$, $G_{ij}^{(s)} \sim N(0, \|s\|^2_{\ell^2(\mathbb{Z})})$, $i < j$, and independence.

Esquisse de preuve du théorème 1

- Let $\mathfrak{H}$ be a real separable Hilbert space (scalar product : $\langle \cdot, \cdot \rangle$, norm : $\| \cdot \|$)
- Let $X = \{X(h) : h \in \mathfrak{H}\}$ be an isonormal Gaussian process (i.e. $\mathbb{E}[X(g)X(h)] = \langle g, h \rangle_{\mathfrak{H}} \; \forall g, h \in \mathfrak{H}$)
- Let $\{e_{ij}\}_{i,j \geq 1} \subset \mathfrak{H}$ be such that $\langle e_{ij}, e_{i'j'} \rangle = \mathbf{1}_{\{i=i'\}} s(j-j')$.
- We set $X_{ij} = X(e_{ij})$.
- We then have

$$W_{n,d}(i,j) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{d}} \sum_{k=1}^{d} \left(X_{ik}^2 - 1 \right) & \text{if } i = j \\ \frac{1}{\sqrt{d}} \sum_{k=1}^{d} X_{ik} X_{jk} & \text{if } i \neq j \end{array} \right\} = I_2(f_{ij}^{(d)})$$

where

$$f_{ij}^{(d)} = \frac{1}{2\sqrt{d}} \sum_{k=1}^{d} \left(e_{ik} \otimes e_{jk} + e_{jk} \otimes e_{ik} \right).$$

Sketch of the proof of Theorem 1

1st step :

- If W is a *symmetric* matrix of size n, we set

$$W^{\text{half}} = (W_{11}, \ldots, W_{1n}, W_{22}, \ldots, W_{2n}, \ldots, W_{nn})^T \in \mathbb{R}^{\frac{n(n+1)}{2}}.$$

We then have, for two symetric random matrices of size n :

$$d_{\text{Wass}}(X, Y) \leq \sqrt{2}\, d_{\text{Wass}}(X^{\text{half}}, Y^{\text{half}}).$$

- Multivariate Malliavin-Stein (Nourdin-Peccati-Réveillac) :
if $H = (H_1, \ldots, H_m) = (I_2(h_1), \ldots, I_2(h_m))$ and if $Z \sim N_m(0, \Sigma)$ where $\Sigma = \mathbb{C}\text{ov}(H)$, then

$$\boxed{d_{\text{Wass}}\,(H, Z) \leq \text{cst}\, \|\Sigma^{-1/2}\|_{\text{op}} \sqrt{\sum_{i,j=1}^{m} \|h_i \otimes_1 h_j\|^2}}.$$

Sketch of the proof of Theorem 1

2nd step :

- $\{i,j\} \cap \{k,l\} = \emptyset \Longrightarrow f_{ij}^{(d)} \otimes_1 f_{kl}^{(d)} = 0.$
- $\#\big\{(i,j,k,l) \in \{1,\ldots,n\} : \{i,j\} \cap \{k,l\} \neq \emptyset\big\} = O(n^3).$
- one has

$$\begin{aligned}
\|f_{ij}^{(d)} \otimes_1 f_{kl}^{(d)}\|^2 &= \langle f_{ij}^{(d)} \otimes_1 f_{kl}^{(d)}, f_{ij}^{(d)} \otimes_1 f_{kl}^{(d)} \rangle \\
&= \langle f_{ij}^{(d)} \otimes_1 f_{ij}^{(d)}, f_{kl}^{(d)} \otimes_1 f_{kl}^{(d)} \rangle \\
&\leq \|f_{ij}^{(d)} \otimes_1 f_{ij}^{(d)}\| \, \|f_{kl}^{(d)} \otimes_1 f_{kl}^{(d)}\|.
\end{aligned}$$

Sketch of the proof of Theorem 1

3rd step :

- One has, after setting $s_d(k) = |s(k)|\mathbf{1}_{\{|k|<d\}}$,

$$\begin{aligned}
\|f_{ij}^{(d)} \otimes_1 f_{ij}^{(d)}\|^2 &= \frac{8}{d^2}\sum_{i,j,k,l=1}^{d} s(i-j)s(j-k)s(k-l)s(l-i) \\
&\leq \frac{8}{d^2}\sum_{i,l=0}^{d-1}\sum_{j,k\in\mathbb{Z}} s_d(i-j)s_d(j-k)s_d(k-l)s_d(l-i) \\
&= \frac{8}{d^2}\sum_{i,l=0}^{d-1}(s_d * s_d)(l-i)^2 \leq \frac{8}{d}\sum_{k\in\mathbb{Z}}(s_d * s_d)(k)^2 \\
&= \frac{8}{d}\|s_d * s_d\|_{\ell^2(\mathbb{Z})}^2.
\end{aligned}$$

Sketch of the proof of Theorem 1

3rd step, continued :

- We recall Young's inequality :

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{s} \Longrightarrow \|u * v\|_{\ell^s(\mathbb{Z})} \leq \|u\|_{\ell^p(\mathbb{Z})} \|v\|_{\ell^q(\mathbb{Z})}$$

- We deduce

$$\begin{aligned} \|f_{ij}^{(d)} \otimes_1 f_{ij}^{(d)}\|^2 &\leq \frac{8}{d} \|s_d * s_d\|^2_{\ell^2(\mathbb{Z})} \leq \frac{8}{d} \|s_d\|^4_{\ell^{4/3}(\mathbb{Z})} \\ &= \frac{8}{d} \left(\sum_{k=1}^{d} |s(k)|^{\frac{4}{3}} \right)^3 . \quad \square \end{aligned}$$

Sketch of the proof of Theorem 2

Sketch of the proof of Theorem 2, according to which there exists $c_H > 0$ (only depending on $H > \frac{3}{4}$) such that, for all $n, d \geq 1$:

$$\boxed{d_{\text{Wass}}\left(\widehat{W}_{n,d}, R_n^{(H)}\right) \leq c_H \frac{n}{\sqrt{d^{4H-3}}}}, \quad \text{where}$$

$\widehat{W}_{n,d} = d^{2-2H}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - I_n\right)$ with $\mathcal{X}_{n,d} = (X_{ij})_{\substack{1\leq i\leq n\\ 1\leq j\leq d}}$ for $(X_{ij})_{i,j\geq 1}$ a Gaussian centered family satisfying

$$\mathbb{E}[X_{ij}X_{i'j'}] = \mathbf{1}_{\{i=i'\}} \times \frac{1}{2}\left(|j-j'+1|^{2H} + |j-j'-1|^{2H} - 2|j-j'|^{2H}\right),$$

and where $R_n^{(H)}$ denotes the Rosenblatt-Wishart matrix of size n.

Esquisse de preuve du théorème 2

1st step :

- By selfsimilarity,

$$\widehat{W}_{n,d} \overset{\text{law}}{=} S_{n,d},$$

where $S_{n,d} = (S_{ij})_{1 \leq i,j \leq n}$, for

$$S_{ij} = \begin{cases} d \sum_{p=0}^{d-1} \left(B^i_{\frac{p+1}{d}} - B^i_{\frac{p}{d}} \right) \left(B^j_{\frac{p+1}{d}} - B^j_{\frac{p}{d}} \right) & \text{if } i \neq j \\ d \sum_{p=0}^{d-1} \left(\left(B^i_{\frac{p+1}{d}} - B^i_{\frac{p}{d}} \right)^2 - d^{-2H} \right) & \text{if } i = j \end{cases}$$

with $B^1, \dots, B^n$ independent fractional Brownian motions of Hurst index H.

Sketch of the proof of Theorem 2

- **2nd step** : By definition,

$$
\begin{aligned}
d_{\text{Wass}}\left(\widehat{W}_{n,d}, R_n^{(H)}\right) &\leq \sqrt{2}\, d_{\text{Wass}}\left(S_{n,d}, R_n^{(H)}\right) \\
&\leq \sqrt{2 \sum_{1\leq i\leq j\leq n} \mathbb{E}[(S_{ij} - R_{ij}^{(H)})^2]}
\end{aligned}
$$

- **3rd step** : in one of my papers published in ECP with J.C. Breton, for completely different reasons I have shown that

$$
\mathbb{E}[(S_{ij} - R_{ij}^{(H)})^2] = O(d^{3-4H}). \quad \square
$$

ADVERTISEMENT ;-)

Should you want to learn more about the **Malliavin-Stein approach**, I recommend you the following two references :

- I. Nourdin (2012) : Lectures on Gaussian approximations with Malliavin calculus. *Sém. Probab. XLV*, pp. 3-89.
- I. Nourdin and G. Peccati (2012) : *Normal Approximations with Malliavin Calculus : from Stein's Method to Universality*. Cambridge Tracts in Mathematics. Cambridge University Press.