Fixed Rank Perturbations of Large Random Matrices: Methodology and Some Statistical Applications

Walid Hachem

CNRS; Telecom ParisTech

Joint work with
R. Couillet, Ph. Loubaton, X. Mestre and J. Najim

Signal model

$$\underset{\substack{N \times T \\ \text{Rcv signal}}}{Y} = \underset{\substack{N \times r \\ \text{Channel}}}{H} \; \underset{\substack{r \times T \\ \text{Src signal}}}{S^*} + \underset{\substack{N \times T \\ \text{Noise}}}{X}$$

- N-dimensional time series observed during a time window of length T, source signal with dimension r.
- Channel and source signals assumed deterministic.
- Noise matrix X has i.i.d. centered elements with variance σ^2/T.

Second order based methods

Second order methods used to detect the number of sources, to estimate the channel H (subspace methods), etc., rely on an estimate of

$$R = \mathbb{E} YY^* = HS^*SH^* + \sigma^2 I_N$$

Usually, this estimate is simply YY^*. When $T \to \infty$ (classical asymptotic regime), $\|YY^* - R\| \xrightarrow{\text{a.s.}} 0$ by the law of large numbers where $\| \cdot \|$ is the spectral norm.

As an example, assuming the problem is to know whether $r = 0$ or 1 (presence or absence of a source), a known test statistic is based on

$$\frac{\|YY^*\|}{N^{-1} \operatorname{tr}(YY^*)}$$

What asymptotic regime ?

- Classical asymptotic regime assumption is often questionable in practice. Window length T and observed signal dimension N are often **of the same order of magnitude**.

- We consider here the asymptotic regime where window length and observed signal dimension are both **large** and of the **same order**, while number of sources is **not large**.

- Formally,

$$N, T \to \infty, \;\; N/T \to c > 0, \;\; r \text{ is fixed}$$

In this case, $\|XX^* - \sigma^2 I_N\| \not\to 0$ and $\|YY^* - (HS^*SH^* + \sigma^2 I)\| \not\to 0$.

Fixed rank perturbations of large random matrices

Problem:

- Behavior of the extreme eigenvalues of large random matrices subjected to fixed rank $(=r)$ additive or multiplicative perturbations.
- Behavior of projections on their associated eigenspaces.

Some fields of application:

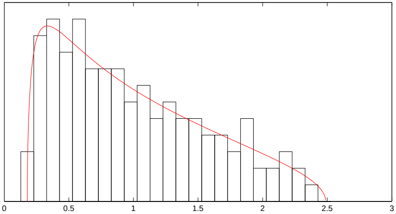
- Statistics (Principal Component Analysis),
- Wireless communications,
- Fault diagnosis,
- Finance (portfolio management),
- Chemometrics,
- ...

Limit spectral measure of XX^*

Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of XX^* with X as above, and let

$$L_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{\lambda_n}$$

be the random **spectral measure** of this matrix. It is well known that L_N converges to the **Marchenko-Pastur** (MP) probability distribution μ_c:



An eigenvalue histogram for $N = 128$, $T = 3N$
with the MP density for $c = 1/3$.

Limit spectral measure and extreme eigenvalue of XX^*

Put

$$\lambda_- = \sigma^2 \left(1 - \sqrt{c}\right)^2 \quad \lambda_+ = \sigma^2 \left(1 + \sqrt{c}\right)^2 \ .$$

Then the MP law has the expression

$$\mu_c(d\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi c \sigma^2 \lambda} \mathbb{1}_{[\lambda_-, \lambda_+]}(\lambda) d\lambda + \left(1 - \frac{1}{c}\right)_+ \delta_0(d\lambda).$$

Moreover, under some assumptions mainly on moments of elements of X,

$$\lambda_1 \xrightarrow[N\to\infty]{\text{a.s.}} \lambda_+,$$

$$T^{2/3} \frac{\lambda_1 - \lambda_+}{\sigma^2(1+\sqrt{c})(1+1/\sqrt{c})^{1/3}} \xrightarrow[N\to\infty]{\mathcal{L}} TW$$

where TW is the **Tracy-Widom** probability distribution.

The Stieltjes Transform

The **Stieltjes Transform** (ST) is one of the many transforms associated to a measure. It is particularly well-suited to study large random matrices. The ST of a probability measure ν is the complex function

$$m_\nu(z) = \int \frac{1}{\lambda - z} \nu(d\lambda)$$

analytical on $\mathbb{C} - \text{support}(\nu)$.
Important example: let

$$M = U \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} U^*,$$

be a $N \times N$ Hermitian matrix with spectral measure

$$L_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{\lambda_n}$$

Stieltjes Transform and resolvent

Let

$$Q(z) = (M - zI_N)^{-1}$$

is the **resolvent** of M.
Then

$$m_{L_N}(z) = \int \frac{1}{\lambda - z} L_N(d\lambda) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\lambda_n - z} = \frac{1}{N} \operatorname{tr} Q(z)$$

Existence and characterization of the limit spectral measure of a random matrix can be established thanks to the asymptotic study of $N^{-1} \operatorname{tr} Q(z)$.

Stieltjes Transform and resolvent

When studying the asymptotic behavior of the spectral measure of a Gram matrix XX^* where $X \in \mathbb{C}^{N \times T}$, a common technique consists in considering the resolvents

$$Q(z) = (XX^* - zI_N)^{-1} \quad \text{and} \quad \widetilde{Q}(z) = (X^*X - zI_T)^{-1}$$

and by showing that

$$\frac{1}{N} \operatorname{tr} Q(z) \xrightarrow[n\to\infty]{\text{a.s.}} m(z) \quad \text{and} \quad \frac{1}{T} \operatorname{tr} \widetilde{Q}(z) \xrightarrow[n\to\infty]{\text{a.s.}} \tilde{m}(z)$$

where $m(z)$ (resp. $\tilde{m}(z) = cm(z) - (1-c)/z$) shows to be the Stieltjes Transform of the limit spectral measure of XX^* (resp. of X^*X).
Often, $m(z)$ is defined as the solution of an implicit equation. Can be solved only in a few particular cases. The MP case is one of these.

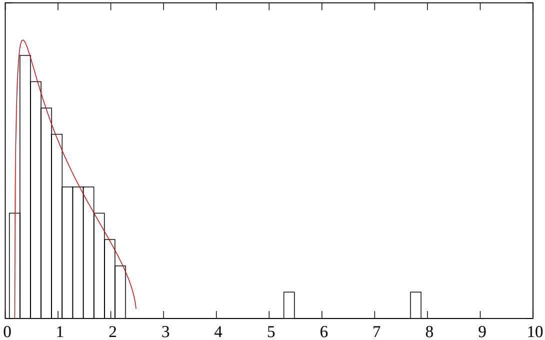
1. The context
   - An application example
   - The unperturbed case
   - Basic tools: Stieltjes transforms and resolvents
   - Fixed rank perturbations

Global *vs* local behavior

- Consider the $N \times T$ matrix $Y = X + P$ where X is random, XX^* has a limit spectral measure ν, and P has a fixed rank r.
- By the interlacing inequality, we can show that P does not impact the global spectral behavior of YY^*: spectral measure of YY^* still converges to ν.
- However, YY^* might have **isolated eigenvalues** which stay out of the support of ν.

A spectrum example for YY^*

- X has iid centered elements with variance $1/T$,
- P is deterministic with rank 2 and singular values 2 and 2.5.



An eigenvalue histogram for $(X + P)(X + P)^*$
with $N = 64$ and $T = 3N$

Overview of perturbed large random matrix models

Purpose: study isolated eigenvalues and possibly their eigenspaces when a large random matrix X is perturbed with the fixed rank matrix P.

- $(I+P)^{1/2}XX^*(I+P)^{1/2}$ where P is Hermitian and X has centered iid elements ("population covariance matrix" is $I+P$): Johnstone'01, Baik *et.al.*'05, Baik Silverstein'06, ...
- $X+P$ where X and P are hermitian and X is a Wigner matrix: Capitaine *et.al.*'09.
- $(X+P)(X+P)^*$ where X is rectangular: Benaych-Georges Nadakuditi'11, HLMNV'11, CCHM'12.

Benaych-Georges and Nadakuditi devised a generic and powerful method for studying the three models.

Model and notations

We get back to the rectangular model $Y = P + X \in \mathbb{C}^{N\times T}$ where X has centered iid elements with variance σ^2/T and where P is deterministic with fixed rank r.

Singular value decompositions: $P = U\sqrt{\Omega}\widetilde{U}^*$ and $Y = W\sqrt{\hat{\Lambda}}\widetilde{W}^*$,

$$U = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \in \mathbb{C}^{N\times r}, \quad \Omega = \begin{bmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_r \end{bmatrix},$$

$$W = \begin{bmatrix} w_1 & \cdots & w_N \end{bmatrix} \in \mathbb{C}^{N\times N}, \quad \hat{\Lambda} = \begin{bmatrix} \hat{\lambda}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_N \end{bmatrix}$$

where $\omega_1 \geq \cdots \geq \omega_r$ are assumed not to depend on N, and where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_N$.

Main result on the eigenvalues

Theorem 1: Consider the previous model. Assume $N, T \to \infty$ with $N/T \to c > 0$. Let $i \leq r$ be the maximum index for which $\omega_i > \sigma^2\sqrt{c}$. Then for $k = 1, \ldots, i$,

$$\hat{\lambda}_k \xrightarrow[N\to\infty]{\text{a.s.}} \rho_k = \frac{\left(\sigma^2 c + \omega_k\right)\left(\omega_k + \sigma^2\right)}{\omega_k} > \lambda_+ = \sigma^2(1+\sqrt{c})^2$$

while

$$\hat{\lambda}_{i+1} \xrightarrow[N\to\infty]{\text{a.s.}} \lambda_+ .$$

Main result on the eigenvectors

Theorem 2: Assume the setting of Theorem 1. Assume in addition that $\omega_1 > \omega_2 > \cdots > \omega_i \; (> \sigma^2\sqrt{c})$. For $k = 1, \ldots, i$, let

$$\Pi_k = u_k u_k^* \quad \text{and} \quad \widehat{\Pi}_k = w_k w_k^*.$$

Then for any sequence a_N of deterministic $N \times 1$ vectors with bounded Euclidean norms,

$$a^* \widehat{\Pi}_k a - h(\rho_k) a^* \Pi_k a \xrightarrow[N\to\infty]{\text{a.s.}} 0, \quad h(x) = \frac{x m(x)^2 \tilde{m}(x)}{(x m(x) \tilde{m}(x))'}$$

where $m(z)$ is the ST of the MP law μ_c and where $\tilde{m}(z) = cm(z) - (1-c)/z$.

Generalization to the case where P has eigenspaces with dimensions > 1 is possible.

2 A case study with some applications

- Results
- Proof technique
- Some applications
- A word about fluctuations

Eigenvalues: principle of the proof of Theorem 1

We follow the approach of Benaych-Georges and Nadakuditi'2011.
We study the isolated eigenvalues of YY^*, or equivalently, the isolated singular values of Y.

A matrix algebraic lemma: Let A be a $N \times T$ matrix. Then $\sigma_1, \ldots, \sigma_{N\wedge T}$ are the singular values of A if and only if

$$\sigma_1, \ldots, \sigma_{n\wedge N}, -\sigma_1, \ldots, -\sigma_{n\wedge N}, \underbrace{0, \ldots, 0}_{|N-T|}$$

are the eigenvalues of

$$\mathbf{A} = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$$

Eigenvalues: principle of the proof of Theorem 1

Recall the SVD $P = U\sqrt{\Omega}\widetilde{U}^*$. Write

$$\mathbf{Y} = \begin{bmatrix} 0 & Y \\ Y^* & 0 \end{bmatrix} = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix} + \begin{bmatrix} U & 0 \\ 0 & \widetilde{U}\sqrt{\Omega} \end{bmatrix} \begin{bmatrix} 0 & I_r \\ I_r & 0 \end{bmatrix} \begin{bmatrix} U^* & 0 \\ 0 & \sqrt{\Omega}\widetilde{U}^* \end{bmatrix} = \mathbf{X} + CJC^*$$

Assume

$$\hat{\lambda} \notin \text{spectrum}(XX^*), \quad \hat{\lambda} \in \text{spectrum}(YY^*)$$

or equivalently

$$\det\left(\mathbf{X} - \sqrt{\hat{\lambda}} I_{N+T}\right) \neq 0, \quad \det\left(\mathbf{Y} - \sqrt{\hat{\lambda}} I_{N+T}\right) = 0.$$

We have

$$\det(\mathbf{Y} - xI) = \det(\mathbf{X} - xI + CJC^*)$$
$$= \det(\mathbf{X} - xI)\det\left(I_{2r} + JC^*(\mathbf{X} - xI)^{-1} C\right)$$

Eigenvalues: principle of the proof of Theorem 1

Using inversion formula for partitioned matrices,

$$(\mathbf{X} - xI)^{-1} = \begin{bmatrix} -xI & X \\ X^* & -xI \end{bmatrix}^{-1} = \begin{bmatrix} xQ(x^2) & X\widetilde{Q}(x^2) \\ \widetilde{Q}(x^2)X^* & x\widetilde{Q}(x^2) \end{bmatrix}$$

where $Q(x) = (XX^* - xI)^{-1}$ and $\widetilde{Q}(x) = (X^*X - xI)^{-1}$ are the usual resolvents.

Hence $\sqrt{\hat{\lambda}}$ is a zero of

$$\det\left(I_{2r} + JC^*\left(\mathbf{X} - xI\right)^{-1} C\right)$$
$$= (-1)^r \det \underbrace{\begin{bmatrix} xU^*Q(x^2)U & I_r + U^*X\widetilde{Q}(x^2)\widetilde{U}\sqrt{\Omega} \\ I_r + \sqrt{\Omega}\widetilde{U}^*\widetilde{Q}(x^2)X^*U & x\sqrt{\Omega}\widetilde{U}^*\widetilde{Q}(x^2)\widetilde{U}\sqrt{\Omega} \end{bmatrix}}_{\widehat{H}(x)}$$

Eigenvalues: principle of the proof of Theorem 1

When $x > \sqrt{\lambda_+}$, $Q(x^2)$ and $\widetilde{Q}(x^2)$ are well defined for large N, because $\|XX^*\| \xrightarrow{\text{a.s.}} \lambda_+$.

An essential part consists in proving that for $x > \sqrt{\lambda_+}$,

$$U^* Q(x^2) U \xrightarrow[N\to\infty]{\text{a.s.}} m(x^2)\mathbf{I}_r, \quad \widetilde{U}^* \widetilde{Q}(x^2) \widetilde{U} \xrightarrow[N\to\infty]{\text{a.s.}} \tilde{m}(x^2)\mathbf{I}_r, \text{ and}$$
$$\widetilde{U}^* \widetilde{Q}(x^2) X^* U \xrightarrow[N\to\infty]{\text{a.s.}} \mathbf{0},$$

Traditionally, random matrix techniques deal with the **normalized traces** of the resolvents. Here we are interested in **bilinear forms** involving these resolvents. In the MP case, this can be done easily.

Eigenvalues: principle of the proof of Theorem 1

Thanks to these results,

$$\widehat{H}(x) \xrightarrow[n\to\infty]{\text{a.s.}} H(x) = \begin{bmatrix} xm(x^2)I_r & I_r \\ I_r & x\tilde{m}(x^2)\Omega \end{bmatrix}$$

outside the support of μ_c, *i.e.*, the eigenvalue bulk.

So YY^* should have isolated eigenvalues near the zeros of equation $\det H(\sqrt{x})$ which **lie outside the support of** μ_c.

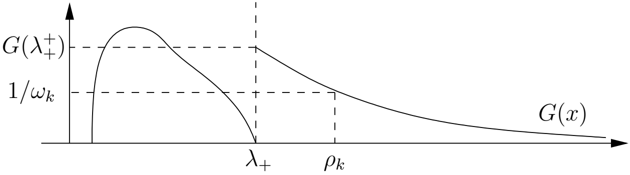
Eigenvalues: principle of the proof of Theorem 1

Consider the equation

$$\det H(\sqrt{x}) = \prod_{k=1}^{r} (x m(x) \tilde{m}(x) \omega_k - 1) = 0 \; . \tag{1}$$

- Recall $\omega_1 \geq \cdots \geq \omega_r$. Arrange the zeros of (1) in decreasing order, similarly to the eigenvalues $\hat{\lambda}_k$ of YY^*.
- From the general properties of the Stieltjes Transforms, function $G(x) = x m(x) \tilde{m}(x)$ decreases from $G(\lambda_+^+)$ to zero for $x \in (\lambda_+, \infty)$.
- Assume $\omega_\ell > 1/G(\lambda_+^+)$. Then the ℓ^{th} zero ρ_ℓ of (1) (which satisfies $G(\rho_\ell) = 1/\omega_\ell$) will satisfy $\rho_\ell > \lambda_+$.
- In that situation, due to $\det \widehat{H} \xrightarrow{\text{a.s.}} \det H$ outside the eigenvalue bulk, we infer that $\hat{\lambda}_\ell \xrightarrow{\text{a.s.}} \rho_\ell$. Otherwise, $\hat{\lambda}_\ell \xrightarrow{\text{a.s.}} \lambda_+$.

Illustration



Exploiting the expressions of $m(z)$ and $\tilde{m}(z)$ (Stieltjes Transforms of MP distributions), condition $\omega_k > 1/G(\lambda_+^+)$ can be rewritten $\omega_k > \sigma^2\sqrt{c}$.
In this case, solving $G(\rho_k) = 1/\omega_k$ gives $\rho_k = \left(\sigma^2 c + \omega_k\right)\left(\omega_k + \sigma^2\right)/\omega_k$.
Hence Theorem 1.

Theorem 2 is proven with similar arguments.

2 A case study with some applications

- Results
- Proof technique
- **Some applications**
- A word about fluctuations

Passive Signal Detection

- $Y = P + X$, non observable signal + AWGN. Noise variance unknown.
- P is a rank one matrix ($r = 1$ source) such that $\|P\|^2 \xrightarrow[N\to\infty]{} \omega > 0$.

Hypothesis test: $\begin{cases} \textbf{H0} & : \quad Y = X & \text{(Noise)} \\ \textbf{H1} & : \quad Y = P + X & \text{(Info+Noise)} \end{cases}$

Generalized Likelihood Ratio Test (GLRT):

$$\xi = \frac{\hat{\lambda}_1}{N^{-1}\,\mathrm{tr}\,(YY^*)}$$

Asymptotic behavior of this statistic ?

Passive signal detection and perturbed model

- Under either **H0** or **H1**, $N^{-1}\operatorname{tr}(YY^*) \xrightarrow[N\to\infty]{\text{a.s.}} \sigma^2$.
- Under **H1** (consequence of main result on eigenvalues):
  - If $\omega > \sigma^2\sqrt{c}$, then

$$\hat{\lambda}_1 \xrightarrow[N\to\infty]{\text{a.s.}} \rho = \frac{(\sigma^2 c + \omega)(\omega + \sigma^2)}{\omega} > \sigma^2(1+\sqrt{c})^2,$$
$$\hat{\lambda}_2 \xrightarrow[N\to\infty]{\text{a.s.}} \sigma^2(1+\sqrt{c})^2.$$

  - If $\omega \le \sigma^2\sqrt{c}$, then

$$\hat{\lambda}_1 \xrightarrow[N\to\infty]{\text{a.s.}} \sigma^2(1+\sqrt{c})^2.$$

Passive Signal Detection and perturbed model

We therefore have

- Under **H0**,

$$\xi_N \xrightarrow[N\to\infty]{\text{a.s.}} (1+\sqrt{c})^2.$$

- Under **H1**,
  - If $\omega > \sigma^2\sqrt{c}$, then

$$\xi_N \xrightarrow[N\to\infty]{\text{a.s.}} \frac{\left(\sigma^2 c+\omega\right)\left(\omega+\sigma^2\right)}{\sigma^2\omega} > (1+\sqrt{c})^2$$

  - If $\omega \le \sigma^2\sqrt{c}$, then

$$\xi_N \xrightarrow[N\to\infty]{\text{a.s.}} (1+\sqrt{c})^2.$$

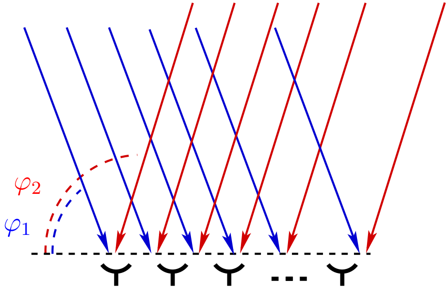
$\omega > \sigma^2\sqrt{c}$ provides the **limit of detectability** by the GLRT.

- False Alarm Probability can be approximated with the help of the Tracy-Widom law.

Source localization

Problem: r radio sources send their signals to a uniform array of N antennas during T signal snapshots.

Estimate arrival angles $\varphi_1, \ldots, \varphi_r$



Example with two sources

Source localization with a subspace method (MUSIC)

Model: $Y = \underbrace{T^{-1/2}AS^*}_{P} + X$ with

- $A = \begin{bmatrix} a(\varphi_1) & \cdots & a(\varphi_r) \end{bmatrix} \in \mathbb{C}^{N\times r}$ with $a(\varphi) = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ e^{\imath\pi \sin\varphi} \\ \vdots \\ e^{\imath(N-1)\pi \sin\varphi} \end{bmatrix}$
- S is deterministic, $\operatorname{rank}(S) = r$.

Let Π be the orthogonal projection matrix on the span of A, or equivalently, on the eigenspace of $\mathbb{E}YY^* = PP^* + \sigma^2 I$ associated with the eigenvalues $> \sigma^2$ ("signal subspace"). Notice that $\Pi = UU^*$.

MUSIC algorithm principle:

$$a(\varphi)^*(I - \Pi)a(\varphi) = 0 \quad \Leftrightarrow \quad \varphi \in \{\varphi_1, \ldots, \varphi_K\}.$$

MUSIC algorithm

Traditional MUSIC: angles are estimated as local minima of

$$a(\varphi)^*(I - \widehat{\Pi})a(\varphi)$$

where $\widehat{\Pi}$ is the orthogonal projection matrix on the eigenspace associated with the r largest eigenvalues of YY^*. Equivalently, local maxima of $a(\varphi)^*\widehat{\Pi}a(\varphi)$.

Notice that $\widehat{\Pi} = \left[w_1 \cdots w_r\right]\left[w_1 \cdots w_r\right]^*$.

- Behavior of $a(\varphi)^*\widehat{\Pi}a(\varphi)$ in our asymptotic regime ?
- Is it possible to **improve** the traditional estimator and to adapt it to our asymptotic regime ?

Modification of the traditional MUSIC algorithm

Modified MUSIC estimator: Application of Theorem 2

Assume that $\liminf_N \omega_r > \sigma^2\sqrt{c}$. Then

$$a(\varphi)^*\Pi a(\varphi) - \sum_{k=1}^{r} \frac{|a(\varphi)^* w_k|^2}{h(\hat{\lambda}_k)} \xrightarrow[N\to\infty]{\text{a.s.}} 0$$

uniformly on $\varphi \in [0,\pi]$.

$\Rightarrow$ find local maxima of $\displaystyle\sum_{k=1}^{r} \frac{|a(\varphi)^* w_k|^2}{h(\hat{\lambda}_k)}$.

2 A case study with some applications

- Results
- Proof technique
- Some applications
- A word about fluctuations

Isolated eigenvalues fluctuations

Fluctuations of the isolated eigenvalues and the projections on associated eigenspaces have been studied for some instances of the three structures $(I+P)^{1/2}XX^*(I+P)^{1/2}$, $X+P$ and $(X+P)(X+P)^*$ introduced above. (Bai-Yao'08, Capitaine *et.al.*'09, Benaych *et.al.*'11, HLMNV'11, CH'11, CCHM'12, ...)

In general,

$$\sqrt{N}\left(\hat{\lambda}_i - \rho_i\right) = \mathcal{O}_P(1)$$

However, the Gaussian limit is not universal

Large deviations of the isolated eigenvalues have been studied in some simple cases (Bianchi *et.al.*'11).