Non white sample covariance matrices.

S. Péché,
Université Grenoble 1,
joint work with O. Ledoit, Uni. Zurich

17-21/05/2010, Université Marne la Vallée
Workshop ”Probability and Geometry in High Dimensions”
Plan

• I. Eigenvectors of sample covariance matrices: problem and motivations.

• II. Review of known results (eigenvalues).

• III. Eigenvectors: The white case.

• IV. Eigenvectors: The non white case.

• V. Conclusion.
Model

We consider sample covariance matrices:

\[ M_N(\Sigma) = \frac{1}{p}YY^*, \quad \text{with} \quad Y = \Sigma^{1/2}X \]

where

- \( X \) is a \( N \times p \) random matrix s.t. the entries \( X_{ij} \) are i.i.d. complex (or real) random variables with distribution \( \mu \), \( \int xd\mu(x) = 0, \int |x|^2d\mu(x) = 1 \).

- \( p = p(N) \) with \( p/N \to \gamma \in (0, \infty) \) as \( N \to \infty \);

- \( \Sigma \) is a \( N \times N \) Hermitian deterministic (or random) matrix, \( \Sigma > 0 \) with bounded spectral radius. \( \Sigma \) is independent of \( X \).

What can be said about eigenvalues and eigenvectors as \( N \to \infty \)?
Motivations I.

Statistics Knowing $M_N(\Sigma)$ what can be said about $\Sigma$?
- if $N$ is fixed and $p \to \infty$ : $M_N(\Sigma)$ good estimator of $\Sigma$;
- in high dimensional setting (genetics, finance, ...)?
Understand e.g. the behavior of PCA in such a setting.

Density of the eigenvalues of $M_N(\Sigma)$ when $\Sigma = Id$.
Dispersion of the eigenvalues: $M_N(\Sigma)$ is NOT a good estimator of $\Sigma$ (smallest and largest eigenvalues e.g.)
Motivations II.

**Communication theory** “CDMA”: received signal \( r = \sum_{k=1}^{K} b_k s_k + w \), with \( K \) number of users, \( s_k \in \mathbb{C}^N \) the signature \( b_k \in \mathbb{C} \), \( \mathbb{E}|b_k|^2 = p_k \) transmitted signal, and \( w \in \mathbb{C}^N \) a Gaussian noise with i.i.d. \( \mathcal{N}(0, \sigma^2) \) components.

One has to decode/estimate the signal \( b_k \). A measure of the performance of the communication channel is the so-called “SIR” (Signal to Interference Ratio): linear receiver \( \hat{x}_1 = c_1^* r \)

\[
SIR = \frac{|C_1^* s_1|^2 p_1}{|c_1|^2 \sigma^2 + \sum_{i \geq 2} |c_i^* s_i|^2 p_i}.
\]

\( \Rightarrow \) as \( N, K \to \infty \), \( K/N \to \gamma \), the SIR depends on the eigenvalues AND the eigenvectors of \( S D S^* \) where \( S = [s_2, \ldots, s_K] \) is the signature matrix (random) and \( D = \text{diag}(p_2, \ldots, p_N) \).
**Eigenvalues I**

We denote by $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_N$ the eigenvalues of $\Sigma$ and suppose that

$$\rho_N(\Sigma) := \frac{1}{N} \sum_{i=1}^{N} \delta_{\pi_i} \xrightarrow{a.s.} H,$$

where $H$ is a probability measure.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of $M_N(\Sigma)$; $\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$.

**Theorem** Marchenko-Pastur (67)

A.s. $\lim_{N \to \infty} \mu_N = \rho_{MP}$, where the Stieltjes transform of $\rho_{MP}$ given by

$$\forall z \in \mathbb{C}, \Im(z) > 0, \quad m_{\rho}(z) := \int \frac{1}{\lambda - z} d\rho_{MP}(\lambda),$$

satisfies

$$m_{\rho}(z) = \int_{-\infty}^{+\infty} \left\{ \tau \left[ 1 - \gamma^{-1} - \gamma^{-1} z m_{\rho}(z) \right] - z \right\}^{-1} dH(\tau).$$
Eigenvalues II

If $\Sigma = Id$, one knows explicitly the density of the Marchenko-Pastur distribution

$$\gamma \geq 1, \quad \frac{d\rho_{MP}}{du} = \frac{\gamma}{2\pi u} \sqrt{(u_+ - u)(u - u_-)} 1_{[u_-,u_+]}(u),$$

with $u_\pm = (1 \pm \frac{1}{\sqrt{\gamma}})^2$.

Valid for both complex and real random matrices.

For general $H$, the relationship between $\rho_{MP}$ and $H$ is not “simple”, determining $H$ from $\rho_{MP}$ is not easy. El Karoui (2008) gives a consistent estimator (using convex approximation).

Assume that $H$ has been estimated, can we improve our knowledge of $\Sigma$? (even if $\Sigma = Id$, the sample covariance matrix is not a good estimator of $\Sigma$).
Eigenvectors: the white case.
**Gaussian sample**

Suppose that $\Sigma = Id$ and $X_{ij}$ i.i.d. $\mathcal{N}(0,1)$ complex or real.

$M_N = M_N(Id)$ is a so-called “white Wishart matrix”.

Let $(U, D)$ be a diagonalization of $M_N$: $M_N = UDU^*$ with $U \in \mathbb{U}(N)$ and $D$ a real diagonal matrix.

$U$ is Haar distributed.

Proof: Gram-Schmidt+ rotationnal invariance of the Gaussian distribution.

**Conjecture:** if $\Sigma = Id$ and if $X$ has non-Gaussian entries with $\mathbb{E}|X_{ij}|^4 < \infty$, the matrix of eigenvectors of $M_N$ shall “asymptotically be Haar distributed”.

Idea: neither direction is preferred.

**Question:** how to define “asymptotically Haar distributed”?
Non Gaussian matrices I.

Silverstein’s idea ('95): $U$ is asymptotically Haar distributed if, given an arbitrary vector $x \in S^{N-1} = \{x \in \mathbb{C}^N, |x| = 1\}$, $y = U^*x$ is asymptotically uniformly distributed on the unit sphere. Or setting

$$Y_N(t) := \sqrt{\frac{N}{2}} \sum_{i=1}^{[Nt]} (|y_i|^2 - 1/N),$$

$Y_N(t)$ shall converge in distribution to a Brownian bridge if $y$ is uniformly distributed ($y = Z/|Z|^2$ with $Z$ Gaussian).

Consider instead $X_N(t) = Y_N(F^N(t)) = \sqrt{\frac{N}{2}} (F_1^N(t) - F_N(t))$ with $F^N(t) = \frac{1}{N} \sum_{i=1}^{N} 1_{\lambda_i \leq t}$ cumulative distribution function (c.d.f.) of the spectral measure of $M_N(\Sigma)$ and

$$F_1^N(t) = \frac{1}{N} \sum_{i=1}^{N} |y_i|^2 1_{\lambda_i \leq t}, \text{ with } y = U^*x$$

also a c.d.f. (but combining the eigenvectors).
Non Gaussian matrices II.

Let

$$G_N(t) = \sqrt{N} \left( F_1^N(t) - F_*^N(t) \right)$$

where $F_*^N$ is the c.d.f. of $\rho_{MP}$ when $\gamma \to p/N$ and $H \to \rho_N(\Sigma)$ spectral measure of $\Sigma$.

Here $G_N \simeq X_N$ and should be close to $B(F(t))$ if $B$ is a Brownian bridge.

Let also $g$ be analytic on $[u_-, u_+]$.

**Theorem** Bai-Miao-Pan (2007)

Assume also that $\mathbb{E}|X_{ij}|^4 = 2$ and $x^*(\Sigma - zI)^{-1}x \to \int \frac{1}{\lambda - z} dH(\lambda)$. Then as $N \to \infty$,

$$\int g(x) dG_N(x) \to \text{a Gaussian random variable (centered and with known variance).}$$

Remark: extension to non-white matrices but with the additional assumption on $x^*(\Sigma - zI)^{-1}x$. 

Spikes in the covariance

Let $\Sigma = \text{diag}(\pi_1, \pi_2, \ldots, \pi_r, 1, \ldots, 1)$, $\pi_i \geq \pi_{i+1} \geq 1$, $i \leq r - 1$, $r$ independent of $N$. $\Sigma$ is a finite rank perturbation of the identity matrix: $H = \delta_1$.

$\mu$ is a centered distribution with variance 1 and finite fourth moment.

Let $\lambda_1$ be the largest eigenvalue of $M_N(\Sigma)$.


\[
\begin{align*}
\text{If } \pi_1 &< 1 + \frac{1}{\sqrt{\gamma}}, \quad \lambda_1 \rightarrow u_+ = (1 + \frac{1}{\sqrt{\gamma}})^2, \\
\text{If } \pi_1 &> 1 + \frac{1}{\sqrt{\gamma}}, \quad \lambda_1 \rightarrow \pi_1 \left(1 + \frac{\gamma^{-1}}{\pi_1 - 1}\right).
\end{align*}
\]

Remark: “Spikes” in the true covariance can be detected if they are large enough. Fluctuation theorems have been established: Bai-Yao (2008) and Féral-Péché (2008).
Eigenvectors for a spiked covariance


\[ \Sigma = \text{diag}(\pi_1, 1, \ldots, 1) \text{ with } \pi_1 > 1 + 1/\sqrt{\gamma}. \]

Let \( u_1 \) (resp. \( e_1 \)) be the normalized eigenvector of \( M_N(\Sigma) \) (resp. of \( \Sigma \)) associated to \( \lambda_1 \) (resp. \( \pi_1 \)):

\[
\lim_{N \to \infty} | < u_1, e_1 > | = \sqrt{\frac{1 - \gamma/(\pi_1 - 1)^2}{1 + \gamma/(\pi_1 - 1)}} \text{ a.s. .}
\]

Idea: perturbation of the eigenvector associated to \( \pi_1 \) (the largest eigenvalue of \( \Sigma \)) by a random matrix.
Eigenvectors: the non-white case.
Another approach (Ledoit-Péché (2009))

Even for a Gaussian sample, the distribution of the eigenvectors is unknown if $\Sigma \neq Id$. It is NOT expected that the matrix of eigenvectors is Haar distributed.

The idea is to study functionals:

$$\theta_N(g) := \frac{1}{N} \text{Tr} \left( g(\Sigma)(M_N(\Sigma) - zI)^{-1} \right),$$

with $z \in \mathbb{C}^+ = \{ z \in \mathbb{C}, \Im z > 0 \}$,

$g$ is a regular function (bounded with a finite number of discontinuities or analytic),

$g(\Sigma) = V \text{diag}(g(\pi_1), \ldots, g(\pi_N))V^*$ if $V$ is the matrix of eigenvectors of $\Sigma$.

**Aim**: understand how the eigenvectors of $M_N(\Sigma)$ project onto those of $\Sigma$.

**Remark**: If $\Sigma \propto Id$ useless. We thus concentrate on the case where $H \neq \delta_1$. 

:: The non white case
A theoretical result

Assume that the support of $H$ is included in $[a_1, a_2]$ with $a_1 > 0$ and

$$\mathbb{E}|X_{ij}|^{12} < \infty$$

independent of $N$ and $p$.

**Theorem:** Ledoit-Péché (2009)

Let $g$ be a bounded function with a finite number of discontinuities on $[a_1, a_2]$. Then $\theta_N(g) \to \theta(g)$ a.s. as $N \to \infty$ where

$$\forall z \in \mathbb{C}^+, \quad \Theta^g(z) = \int_{-\infty}^{+\infty} \left\{ \tau \left[ 1 - \gamma^{-1} - \gamma^{-1}zm_\rho(z) \right] - z \right\}^{-1} g(\tau)dH(\tau).$$

Remark: the same kernel

$$\left\{ \tau \left[ 1 - \gamma^{-1} - \gamma^{-1}zm_\rho(z) \right] - z \right\}^{-1}$$

arises as in the Marchenko-Pastur theorem.
Corrolary 1.

Question: How much do the eigenvectors of $M_N(\Sigma)$ deviate from those of $\Sigma$?

We set $g = 1_{(-\infty, \tau)}$ and $\Phi_N(\lambda, \tau) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |u_i^* v_j|^2 1_{[\lambda_i, +\infty)}(\lambda) \times 1_{[\tau_j, +\infty)}(\tau)$.

Let $v_j$ be the normalized eigenvector of $\Sigma$ associated to $\pi_j$. The average of $N|u_i^* v_j|^2$ bearing on the eigenvectors associated to sample eigenvalues (resp. eigenvalues of the true covariance) in the interval $[\lambda, \bar{\lambda}]$ (resp. $[\tau, \bar{\tau}]$) is:

$$\frac{\Phi_N(\bar{\lambda}, \tau) - \Phi_N(\bar{\lambda}, \bar{\tau}) - \Phi_N(\lambda, \bar{\tau}) + \Phi_N(\lambda, \tau)}{[F_N(\lambda) - F_N(\bar{\lambda})] \times [H_N(\tau) - H_N(\bar{\tau})]}$$

if $F_N$ (resp. $H_N$) is the c.d.f. of $M_N(\Sigma)$ (resp. $\Sigma$).

If one can choose $\lambda, \bar{\lambda}$ and $\tau, \bar{\tau}$ arbitrarily close, then one gets precise information!
Corollary 1.

**Theorem:** $\Phi_N(\lambda, \tau) \xrightarrow{a.s.} \Phi(\lambda, \tau)$ at any point of continuity of $\Phi$. And $\forall (\lambda, \tau) \in \mathbb{R}^2$, $\Phi(\lambda, \tau) = \int_{-\infty}^{\lambda} \int_{-\infty}^{\tau} \varphi(l, t) \, dH(t) \, d\rho_{MP}(l)$, where

$$\varphi(l, t) = \begin{cases} 
\frac{\gamma^{-1}lt}{(at - l)^2 + b^2t^2}, & \text{if } l > 0 \\
1 - \frac{1}{\gamma} - \frac{l\tilde{m}_\rho(l)}{\gamma} =: a + ib, & \text{if } l = 0 \text{ and } \gamma < 1 \\
\frac{1}{(1 - \gamma)[1 + \tilde{m}_\rho(0)t]}, & \text{if } l = 0 \text{ and } \gamma > 1 \\
0, & \text{otherwise}
\end{cases}$$

Here $\tilde{m}_\rho(0) = \lim_{z \to 0} m_\rho(z)$ and $m_\rho$ is the limiting Stieltjes transform of $X^*\Sigma X/N$.

Thus in principle one can obtain precise information on the eigenvectors (but this assumes that one knows the c.d.f. of $H_N$).
Corollary 2.

Question: how does $M_N(\Sigma)$ differ from $\Sigma$ and how can we improve the initial estimator of $\Sigma$ given by $M_N(\Sigma)$?

We get a better estimator by choosing $g(x) = x$.

One seeks an estimator of $\Sigma$ of the kind $UD_N U^*$, $D_N$ diagonal i.e. an estimator which has the same eigenvectors as $M_N(\Sigma)$.

The best estimator (Frobenius norm) is

$$\tilde{D}_N = \text{diag}(\tilde{d}_1, \ldots, \tilde{d}_N) \quad \text{where} \quad \forall i = 1, \ldots, N \quad \tilde{d}_i = u_i^* \Sigma_N u_i.$$

Can we say a few things on the $\tilde{d}_i$'s:

yes asymptotically by choosing $g(x) = x$.  

- 18 -
Corollary 2.

We set

\[ \forall x \in \mathbb{R}, \quad \Delta_N(x) = \frac{1}{N} \sum_{i=1}^{N} \tilde{d}_i \cdot 1_{[\lambda_i, +\infty)}(x) = \frac{1}{N} \sum_{i=1}^{N} u_i^* \sum_N u_i \times 1_{[\lambda_i, +\infty)}(x). \]

Then one has

\[ \forall i = 1, \ldots, N \quad \tilde{d}_i = \lim_{\varepsilon \to 0^+} \frac{\Delta_N(\lambda_i + \varepsilon) - \Delta_N(\lambda_i - \varepsilon)}{F_N(\lambda_i + \varepsilon) - F_N(\lambda_i - \varepsilon)}. \]

**Theorem:** For all \( x \neq 0 \), \( \Delta_N(x) \to \Delta(x) \). Moreover

\[ \Delta(x) = \int_{-\infty}^{x} \delta(\lambda) \, dF(\lambda), \]

with

\[ \delta(\lambda) = \begin{cases} \frac{\lambda}{|1 - \gamma^{-1} - \gamma^{-1} \lambda \, \check{m}_\rho(\lambda)|^2} & \text{if } \lambda > 0 \\ \frac{\gamma}{(1 - \gamma) \check{m}_\rho(0)} & \text{if } \lambda = 0 \text{ and } \gamma < 1 \\ 0 & \text{otherwise.} \end{cases} \]
An improved estimator

We consider the “improved” estimator $\tilde{S}_N := UD'U^*$, where

$$D'_i = \frac{\lambda_i}{|1 - \gamma^{-1} - \gamma^{-1}\lambda_i \tilde{m}_\rho(\lambda_i)|^2}.$$  

We ran 10,000 simulations with $\rho_N(\Sigma) = 0.2\delta_1 + 0.4\delta_3 + 0.4\delta_{10} , \quad \gamma = 2$ and increasing the number of variables $p$ from 5 to 100. For each simulation, we calculate the “Percentage Relative Improvement in Average Loss” (PRIAL):

if $M$ is an estimator of $\Sigma_N$ and if $|A|^2_F = \text{Tr}AA^*$ (Frobenius norm),

$$PRIAL(M) = 100 \times \left[ 1 - \frac{\mathbb{E}\left\| M - U_N\tilde{D}_NU_N^* \right\|^2_F}{\mathbb{E}\left\| M_N(\Sigma) - U_N\tilde{D}_NU_N^* \right\|^2_F} \right].$$
Simulations

Even for small sizes, $p = 40$, the PRIAL is $95\%$. 

![Graph showing relative improvement in average loss for different sample sizes. The y-axis represents the relative improvement in average loss ranging from 86% to 100%, and the x-axis represents sample size ranging from 10 to 200. The line indicates an increasing trend as sample size increases.]

The non white case
Concluding remarks

- $\theta_N(g)$ is a new tool that allows to study the average behavior of the eigenvectors: for instance we cannot recover D. Paul’s result for the eigenvector associated to the largest eigenvalue separating from the bulk.

- in general we cannot say anything on the eigenvectors associated to extreme eigenvalues: average behavior of the eigenvectors.

- for the moment theoretical results only: one has to define first appropriate estimators for $\tilde{m}_N, H_N \ldots$