RIE for cross-covariance matrices
(RIE = Rotationally Invariant Estimator)

Florent Benaych-Georges

Capital Fund Management

Joint work with Jean-Philippe Bouchaud (CFM) and Marc Potters (CFM)

April 30, 2019
Krakow

Data denoising and matrix estimation

- **Denoising** : estimate a (locally) fixed signal out of noisy observations

$$\mathrm{Observation}(t) = \mathrm{Signal} + \mathrm{Noise}(t) \qquad (t = 1, \dots, T)$$

Law of large numbers : in dimension 1, when $T \gg 1$,

$$\mathrm{Signal} \approx \frac{1}{T} \sum_t \mathrm{Observation}(t)$$

- **Matrix estimation** : case where the signal is **a matrix**

  - In small dimension, matrix estimation goes as classical estimation : error in

  $$\mathrm{Signal} \approx \frac{1}{T} \sum_t \mathrm{Observation}(t)$$

  is small for each entry and the number of entries is small

  - In large dimension, small errors on entries can add up to generate **global significative error**

Covariance matrices estimation

- $X(1)$, $X(2)$, $\ldots$, $X(T)$ independent observations of a (null mean) random vector $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$
- **Goal :** estimation of the covariance matrix $\mathrm{Cov}(X) \in \mathbb{R}^{n \times n}$
- *Law of large numbers :* **if** n **small** and T large, then

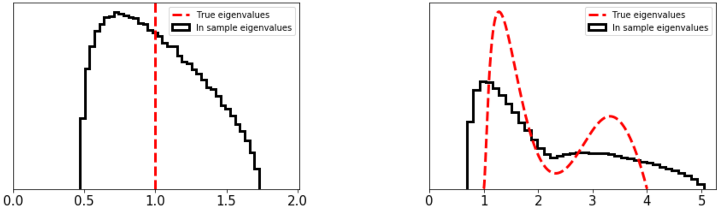
$$\mathrm{Cov}(X) \approx \frac{1}{T} \sum_t X(t) X(t)^*$$

- **What if** n **is not small ?**
  - Error term in each entry : $O(T^{-1/2})$

  $\hookrightarrow$ Global error term $n^2 T^{-1/2}$: small when $n^4 \ll T$

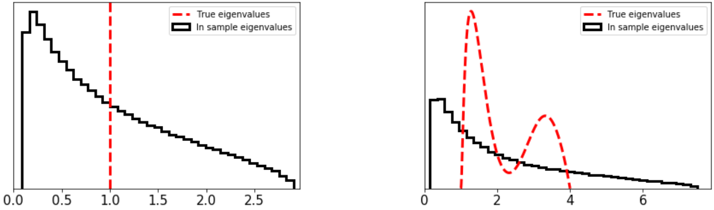
  $\hookrightarrow$ If $n = 100$ and t are (business) days : **400 000 years of (stationary) observations are needed !**
  - In fact, approximation above is right as soon as $T \gg n$ (Rudelson & Vershynin, 2000')
  - **... and what if** $T \not\gg n$ **?**

- **Covariance matrices** : $\text{Cov}(X) \not\approx \frac{1}{T} \sum_t X(t)X(t)^*$ when $T \not\gg n$



Spectrum of $\frac{1}{T} \sum_t X(t)X(t)^*$ histogram vs spectrum of $\text{Cov}(X)$ for $T/n = 10$
$\hookrightarrow$ ***in sample spectrum is widely dilated with respect to true one***

- When T/n decreases (below, $T/n = 2$), this dilation becomes still wider :



- **Consequences :** risk over/under estimation for linear combinations of the entries of X corresponding to largest/smallest eigenvectors

Cross-covariance matrices estimation

- For any $t = 1, \ldots, T$, $(X(t), Y(t))$ observation of a pair of (null mean) random vectors $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix}$
- **Goal :** estimation of the cross-covariance matrix

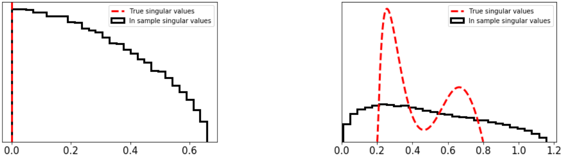
$$\mathrm{Cov}(X, Y) = \left[\mathrm{Cov}(X_i, Y_j)\right]_{i,j} \in \mathbb{R}^{n \times p}$$

- *Law of large numbers :* **if** n, p **small** and T large, then

$$\mathrm{Cov}(X, Y) \approx \frac{1}{T} \sum_t X(t) Y(t)^*$$

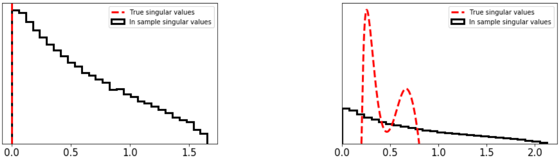
- But **if** n, p **not small**, same kind of problems as for covariance matrices

- **Cross-covariance matrices** : $\operatorname{Cov}(X,Y) \not\approx \frac{1}{T}\sum_t X(t)Y(t)^*$ when $T \not\gg n, p$



Singular values of $\frac{1}{T}\sum_t X(t)Y(t)^*$ vs singular values of $\operatorname{Cov}(X,Y)$ for $X, Y \sim \mathcal{N}(0, I_n)$ and $T/n = T/p = 10$: ***in sample singular values are widely dilated with respect to true ones***

- When $T/n, T/p$ decrease (below, $T/n = T/p = 2$), this dilation becomes still wider :



- **Consequences :** correlations over/under estimation between linear combinations of the entries of X and Y corresponding to largest/smallest singular vectors

Estimation technics

- **Regularization** : for structured covariance matrices (e.g. band matrix) only (cf Bose et al.)
- **Shrinkage** : average the empirical estimate $\frac{1}{T}\sum_t$ with a guess of the true value
- **Clipping** : turn all but a few eigenvalues (the largest ones) to a constant value chosen so that a certain quantity (like the trace) is preserved
- **Cleaning** : clean the eigenvalues (resp. singular values) to make the estimated matrix the closest as possible to the true one

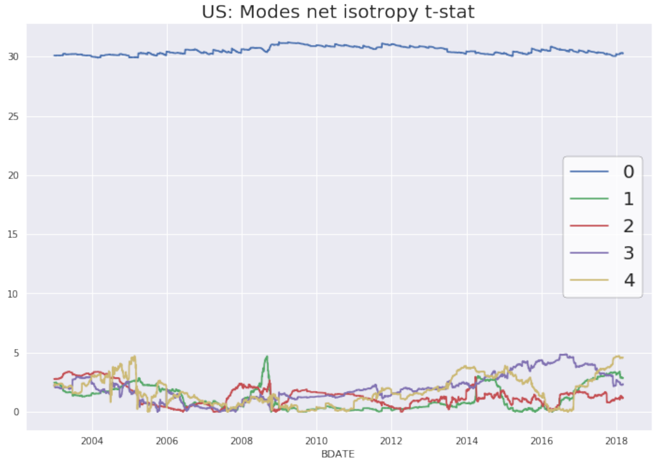
Isotropy

Shrinkage, clipping and cleaning : eigenvectors (resp. singular vectors) **unchanged !**

$\hookrightarrow$ **Agnostic** on the directions of largest/smallest risk (resp. of largest/smallest correlation)

$\hookrightarrow$ Implicitely suppose that we have a prior on the true covariance (resp. cross-covariance) that is *Rotationally Invariant*, *i.e.* invariant, in law, under conjugation by any orthogonal matrix (resp. by multiplication on the left and on the right by any orthogonal matrix) : **Rotationally Invariant Estimator** (**RIE**)

Market eigenvectors are not isotropic...



...so ideally, in financial risk management, we should also **clean eigenvectors**

Problem : we do not know how to do that...

Rotationally Invariant Estimators

- 2 estimation problems :
  - Estimate the **true covariance** $\mathrm{Cov}(X)$ given the **empirical covariance** $\frac{1}{T}\sum_t X(t)X(t)^*$
  - Estimate the **true cross-covariance** $\mathrm{Cov}(X,Y)$ given the **empirical covariance** $\frac{1}{T}\sum_t X(t)Y(t)^*$
- Constraints : let eigenvectors/singular vectors **unchanged**
- **Optimality :** given this constraint, realize **minimum distance** to target
- **Surprise :** both problems have a **computable solution** !
  - **Covariance :** Ledoit-Péché 2011 (+improvements by Bun-Bouchaud-Potters-Knowles)
  - **Cross-covariance :** B-Bouchaud-Potters 2018

- **Starting point :**
  - **Covariance estimation** : write decompositions

$$\mathrm{Cov}(X)^{\mathrm{emp}} = \frac{1}{T}\sum_t X(t)X(t)^* = \sum_k \lambda_k \mathbf{u}_k \mathbf{u}_k^*$$

$$\mathrm{Cov}(X)^{\mathrm{emp,cleaned}} = \sum_k \lambda_k^{\mathrm{cleaned}} \mathbf{u}_k \mathbf{u}_k^*$$

  Optimality : $\| \mathrm{Cov}(X)^{\mathrm{emp,cleaned}} - \mathrm{Cov}(X)^{\mathrm{true}} \|_{\text{Frobenius}}$ minimal, which rewrites :

$$\boxed{\lambda_k^{\mathrm{cleaned}} = \mathbf{u}_k^* \,\mathrm{Cov}(X)^{\mathrm{true}} \mathbf{u}_k}$$

  - **Cross-covariance estimation** : in the same way, **optimality** rewrites as follows : for

$$\mathrm{Cov}(X,Y)^{\mathrm{emp}} = \frac{1}{T}\sum_t X(t)Y(t)^* = \sum_k s_k \mathbf{u}_k \mathbf{v}_k^*,$$

  Optimality : $\| \mathrm{Cov}(X,Y)^{\mathrm{emp,cleaned}} - \mathrm{Cov}(X,Y)^{\mathrm{true}} \|_{\text{Frobenius}}$ minimal, which rewrites :

$$\boxed{s_k^{\mathrm{cleaned}} = \mathbf{u}_k^* \,\mathrm{Cov}(X,Y)^{\mathrm{true}} \mathbf{v}_k}$$

- **Problem :** optimal solutions $\lambda_k^{\mathrm{cleaned}}$, s_k^{cleaned} expressed in terms of the unknown $\mathrm{Cov}(X)^{\mathrm{true}}$, $\mathrm{Cov}(X,Y)^{\mathrm{true}}$

$\hookrightarrow$ **RMT**

RIE formulas

- **Covariance matrices :**
  - Define the function

$$m(z) := \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\lambda_k - z} \qquad z\in\mathbb{C}$$

    (empirical covariance matrix Stieltjes transform)
  - Then

$$\lambda_k^{\text{cleaned}} = \frac{\lambda_k}{|1-\frac{n}{T}-\frac{n}{T}\lambda m(z)|^2} \qquad \text{for} \qquad z := \lambda_k + \frac{\mathrm{i}}{(nT)^{1/4}}$$

- **Cross-covariance matrices :**

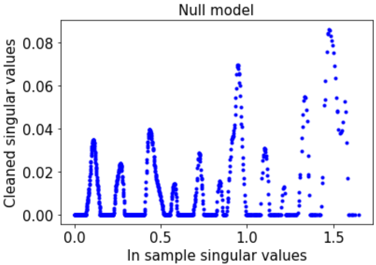
$$s_k^{\text{cleaned}} := \frac{\Im(L(z))}{\Im(H(z))}\, s_k \quad \text{for } z = s_k + \frac{\mathrm{i}}{(npT)^{1/6}}$$

  with $L(z) = \cdots$, $H(z) = \cdots$ (complicated formulas)

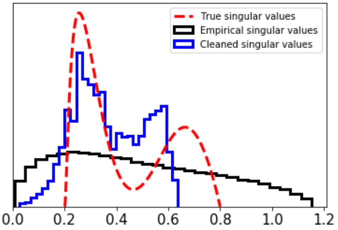
- **Performance accuracy :** each of both RIEs is way better performing than *naive estimation* on models with the good isotropy property

Cross-covariance RIE at work

- **Null model** : $X, Y \sim \mathcal{N}(0, I_n)$, $\mathrm{Cov}(X,Y)^{\mathrm{true}} = 0$:



- **Non null model** : In sample (= Empirical) vs cleaned vs true singular values



Cross-covariance RIE : non converging estimator !

Question : do we have :

$$\text{Optimality} \quad \Longrightarrow \quad s_k^{\text{cleaned}} \approx s_k \text{ at the limit ?}$$

Reply : NO. On average,

Cleaned singular values $<$ True singular values $<$ Empirical singular values

Precisely :

$$\mathbb{E} \sum_k s_k^2 \;=\; \left(1 + \frac{1}{T}\right) \sum_k (s_k^{\text{true}})^2 + \frac{2}{T} \operatorname{Tr} \operatorname{Cov}(X)^{\text{emp}} \operatorname{Tr} \operatorname{Cov}(Y)^{\text{emp}}$$

and

$$\mathbb{E} \sum_k s_k s_k^{\text{cleaned}} \;=\; \sum_k (s_k^{\text{true}})^2$$

How come ? Optimal *matrix estimator* keeping *empirical singular vectors unchanged* takes into account the fact that these vectors are noisy versions of the true ones, hence reduces their weights by *shrinking the singular values*.

RIE formula : optimality condition

Optimality : for

$$\operatorname{Cov}(X,Y)^{\text{emp}} = \frac{1}{T}\sum_t X(t)Y(t)^* = \sum_k s_k \mathbf{u}_k \mathbf{v}_k^*,$$

s_k^{cleaned} are such that

$$\left\| \sum_k s_k^{\text{cleaned}} \mathbf{u}_k \mathbf{v}_k^* - \operatorname{Cov}(X,Y)^{\text{true}} \right\|_{\text{Frobenius}} \qquad \text{is minimal},$$

i.e.

$$\left\| [\mathbf{u}_1 \cdots] \operatorname{diag}(s_1^{\text{cleaned}}, \ldots)[\mathbf{v}_1 \cdots]^* - \operatorname{Cov}(X,Y)^{\text{true}} \right\|_{\text{Frobenius}} \quad \text{minimal},$$

i.e., using the left and right invariance of the Frobenius norm,

$$\left\| \operatorname{diag}(s_1^{\text{cleaned}}, \ldots) - [\mathbf{u}_1 \cdots]^* \operatorname{Cov}(X,Y)^{\text{true}}[\mathbf{v}_1 \cdots] \right\|_{\text{Frobenius}} \quad \text{minimal},$$

i.e.

$$s_k^{\text{cleaned}} = \left([\mathbf{u}_1 \cdots]^* \operatorname{Cov}(X,Y)^{\text{true}}[\mathbf{v}_1 \cdots]\right)_{kk} = \underbrace{\mathbf{u}_k^* \operatorname{Cov}(X,Y)^{\text{true}} \mathbf{v}_k}_{\text{To estimate!!}}$$

RIE formula : estimation of $\mathbf{u}_k^* \,\mathrm{Cov}(X,Y)^{\mathrm{true}} \mathbf{v}_k$

Let

$$\mu = \sum_k (\delta_{s_k} + \delta_{-s_k}) \quad \text{(singular values distribution of } \mathrm{Cov}(X,Y)^{\mathrm{emp}}\text{)}$$

and

$$\nu = \sum_k \mathbf{u}_k^* \,\mathrm{Cov}(X,Y)^{\mathrm{true}} \mathbf{v}_k \, (\delta_{s_k} - \delta_{-s_k}) \quad \text{(the same with some weights)}$$

Then

$$\mathbf{u}_k^* \,\mathrm{Cov}(X,Y)^{\mathrm{true}} \mathbf{v}_k = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(s_k) = \frac{\Im(\text{Stieltjes transf. of } \nu \text{ at } s_k + \mathrm{i}0^+)}{\Im(\text{Stieltjes transf. of } \mu \text{ at } s_k + \mathrm{i}0^+)}$$

Problem : estimate, out of **observable variables**, the Stieltjes transf. of ν, which happens to rewrite

$$\mathrm{Tr}(\underbrace{\mathrm{Cov}(X,Y)^{\mathrm{true}}}_{\text{unknown !!}})^* \left(z^2 - \mathrm{Cov}(X,Y)^{\mathrm{emp}} (\mathrm{Cov}(X,Y)^{\mathrm{emp}})^*\right)^{-1} \mathrm{Cov}(X,Y)^{\mathrm{emp}}$$

$\hookrightarrow$ **Stein formula**, **concentration of measure**