

Grandes matrices aléatoires et traitement statistique du signal

Philippe Loubaton
Equipe Signal et Communications.

11-12 Décembre 2013

Contexte général.

Observation :

Série temporelle multivariable $(\mathbf{y}_n)_{n=1,\dots,N}$ avec $\dim(\mathbf{y}_n) = M$.

Faire de l'inférence statistique (détection, estimation,...) à partir de l'observation quand M et N sont grands et comparables.

Contexte souvent pertinent dans le cas d'observations de grandes dimensions (grands réseaux de capteurs, ,...)

Contexte modélisé par le régime asymptotique :

$M \rightarrow +\infty, N \rightarrow +\infty, \frac{M}{N} \rightarrow c$, avec $0 < c < +\infty$ qui fait apparaître naturellement la "grande" matrice aléatoire $M \times N$
 $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$.

Difficultés nouvelles et moins d'outils génériques que dans le cas où $N \rightarrow +\infty$ et M fixe.

Historique de la problématique.

Grandes matrices aléatoires et physique statistique :
depuis 1950

Grandes matrices aléatoires et probabilités : depuis 1990

Grandes matrices aléatoires et statistiques : depuis 2005

Grandes matrices aléatoires et traitement statistique du
signal : depuis 2008

Localisation de sources bande étroite par un grand réseau de capteurs.

$$\begin{array}{ccccccc} \text{Observation} & & \text{Channel} & \text{Source signals} & & & \text{Noise} \\ \left[\begin{array}{c} \mathbf{y}_1 \cdots \mathbf{y}_N \end{array} \right] & = & \left[\begin{array}{c} \mathbf{a}_1 \cdots \mathbf{a}_K \end{array} \right] & \left[\begin{array}{c} \mathbf{s}^1 \\ \cdots \\ \mathbf{s}^K \end{array} \right] & + & \left[\begin{array}{c} \mathbf{v}_1 \cdots \mathbf{v}_N \end{array} \right] \\ \mathbf{Y} & = & \mathbf{A} & \mathbf{S} & + & \mathbf{V} \\ M \times N & & M \times K & K \times N & & M \times N \end{array}$$

- ▶ $K < M < N$, \mathbf{A} et \mathbf{S} matrices déterministes non observables.
- ▶ \mathbf{V} matrice à éléments i.i.d. $\mathcal{N}_c(0, \sigma^2)$
- ▶ \mathbf{Y} modèle "Information plus Bruit"

La "méthode sous-espace".

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{V}$$

- ▶ $\mathbf{A} = (\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K))$
- ▶ La fonction $\theta \rightarrow \mathbf{a}(\theta)$ connue

Estimer les paramètres $\theta_1, \dots, \theta_K$

Π projecteur sur l'orthogonal de l'espace image de \mathbf{A}

Principe de base de l'approche sous-espace.

$$\mathbf{a}(\theta)^* \Pi \mathbf{a}(\theta) = 0 \text{ ssi } \theta \in \{\theta_1, \dots, \theta_K\}$$

Estimer Π par $\hat{\Pi}$ et minimiser $\theta \rightarrow |\mathbf{a}(\theta)^* \hat{\Pi} \mathbf{a}(\theta)|$.

Les problèmes en grande dimension.

Si $N \rightarrow +\infty$ et M reste fixe

$$\left\| \frac{\mathbf{Y}\mathbf{Y}^*}{N} - \mathbb{E} \left(\frac{\mathbf{Y}\mathbf{Y}^*}{N} \right) \right\| \rightarrow 0$$

Les problèmes en grande dimension.

Si $N \rightarrow +\infty$ et M reste fixe

$$\left\| \frac{\mathbf{Y}\mathbf{Y}^*}{N} - \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I} \right) \right\| \rightarrow 0$$

Les problèmes en grande dimension.

Si $N \rightarrow +\infty$ et M reste fixe

$$\left\| \frac{\mathbf{Y}\mathbf{Y}^*}{N} - \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I} \right) \right\| \rightarrow 0$$

$\hat{\Pi}_t =$ projecteur sur l'espace propre associé aux $M - K$ plus petites valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ fournit des estimateurs consistants $(\hat{\theta}_{1,t}, \dots, \hat{\theta}_{K,t})$.

Les problèmes en grande dimension.

Si $N \rightarrow +\infty$ et M reste fixe

$$\left\| \frac{\mathbf{Y}\mathbf{Y}^*}{N} - \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I} \right) \right\| \rightarrow 0$$

$\hat{\Pi}_t$ = projecteur sur l'espace propre associé aux $M - K$ plus petites valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ fournit des estimateurs consistants $(\hat{\theta}_{1,t}, \dots, \hat{\theta}_{K,t})$.

Si M et N comparables.

$\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ n'a pas le même comportement que $\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I}$

Les problèmes en grande dimension.

Si $N \rightarrow +\infty$ et M reste fixe

$$\left\| \frac{\mathbf{Y}\mathbf{Y}^*}{N} - \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I} \right) \right\| \rightarrow 0$$

$\hat{\Pi}_t$ = projecteur sur l'espace propre associé aux $M - K$ plus petites valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ fournit des estimateurs consistants $(\hat{\theta}_{1,t}, \dots, \hat{\theta}_{K,t})$.

Si M et N comparables.

$\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ n'a pas le même comportement que $\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I}$

- ▶ Les estimateurs $(\hat{\theta}_{1,t}, \dots, \hat{\theta}_{K,t})$ restent-ils consistants ?
- ▶ Peut-on les améliorer ?

Les problèmes en grande dimension.

Si $N \rightarrow +\infty$ et M reste fixe

$$\left\| \frac{\mathbf{Y}\mathbf{Y}^*}{N} - \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I} \right) \right\| \rightarrow 0$$

$\hat{\Pi}_t$ = projecteur sur l'espace propre associé aux $M - K$ plus petites valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ fournit des estimateurs consistants $(\hat{\theta}_{1,t}, \dots, \hat{\theta}_{K,t})$.

Si M et N comparables.

$\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ n'a pas le même comportement que $\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I}$

- ▶ Les estimateurs $(\hat{\theta}_{1,t}, \dots, \hat{\theta}_{K,t})$ restent-ils consistants ?
- ▶ Peut-on les améliorer ?

Travaux préalables de X. Mestre (2008) dans le cas de sources i.i.d. gaussiennes.

Les grandes étapes.

- ▶ Il est nécessaire que les K plus grandes valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ et les $M - K$ plus petites se trouvent dans des intervalles bien séparés : à quelles conditions ?
- ▶ Dans ce cas, combiner les espaces propres correspondants pour estimer de façon consistante $\theta \rightarrow \mathbf{a}(\theta)^* \mathbf{\Pi} \mathbf{a}(\theta)$

Nécessité d'améliorer des résultats existant concernant le modèle "Information plus Bruit", peu étudié auparavant.

Les cas $K \ll N$ et K et N du même ordre de grandeur ne présentent pas les mêmes difficultés.

Localisation des valeurs propres : illustration.

$$K = M/2, \sigma^2 = 2, \lambda_k \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* \right) = 5 \text{ pour } k = 1, \dots, M/2.$$

Si $\frac{M}{N} \simeq 0$:

$$\frac{\mathbf{Y}\mathbf{Y}^*}{N} \simeq \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* + \sigma^2 \mathbf{I} \right)$$

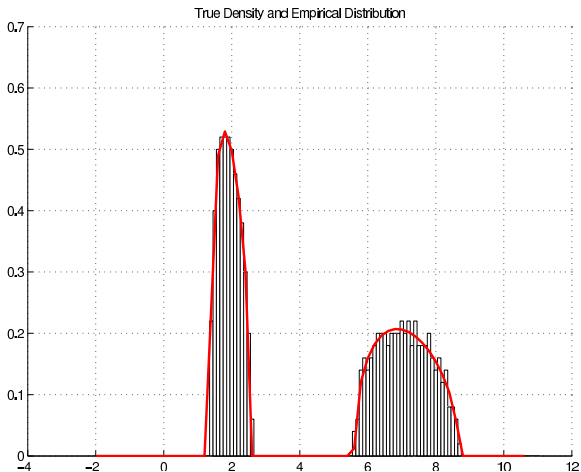
Valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$:

- ▶ $K = M/2$ valeurs propres très proches de 2
- ▶ $M - K = M/2$ valeurs propres très proches de $5+2=7$

Localisation des valeurs propres : illustration.

$$K = M/2, \sigma^2 = 2, \lambda_k \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* \right) = 5 \text{ pour } k = 1, \dots, M/2.$$

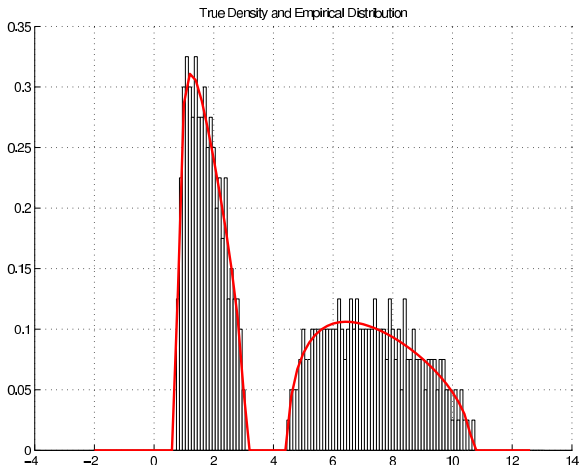
Si $\frac{M}{N} = 0.05$



Localisation des valeurs propres : illustration.

$$K = M/2, \sigma^2 = 2, \lambda_k \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* \right) = 5 \text{ pour } k = 1, \dots, M/2.$$

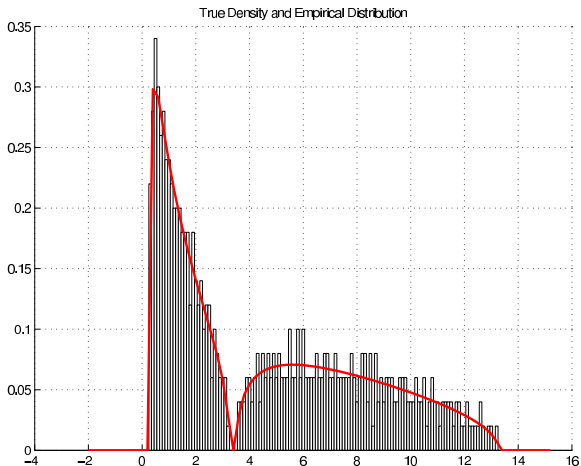
Si $\frac{M}{N} = 0.2$



Localisation des valeurs propres : illustration.

$$K = M/2, \sigma^2 = 2, \lambda_k \left(\mathbf{A} \frac{\mathbf{S}\mathbf{S}^*}{N} \mathbf{A}^* \right) = 5 \text{ pour } k = 1, \dots, M/2.$$

Si $\frac{M}{N} = 0.5$



Exemple de résultats.

$K = 2, \frac{SS^*}{N} \rightarrow \mathbf{I}_2$, antenne linéaire à capteurs équidistants.

Difficulté dépend de $|\theta_2 - \theta_1|$

$$|\theta_2 - \theta_1| \gg \frac{1}{M}$$

Si $\text{SNR} > \sqrt{\frac{M}{N}}$, l'approche traditionnelle fonctionne, et a les mêmes performances que les techniques adaptées aux grandes dimensions.

$$(\theta_2 - \theta_1) \simeq \frac{\alpha}{M}$$

Si $\text{SNR}(1 - |\frac{\sin \alpha}{\alpha}|) > \sqrt{\frac{M}{N}}$:

- ▶ $\hat{\theta}_{i,t} - \theta_i = \mathcal{O}(\frac{1}{N})$
- ▶ $\hat{\theta}_i - \theta_i = \mathcal{O}(\frac{1}{N^{3/2}})$

Travaux futurs (ANR DIONISOS).

Problèmes mettant en jeu des signaux large bande.

Observations générées par $\mathbf{y}_n = \sum_{p=0}^{P-1} \mathbf{a}_p s_{n-p} + \mathbf{v}_n$

Usuel d'introduire, pour L bien choisi, la matrice $ML \times N$ $\mathbf{Y}_N^{(L)}$ définie

$$\mathbf{Y}_N^L = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_N \\ \mathbf{y}_2 & \mathbf{y}_3 & \dots & \mathbf{y}_{N+1} \\ \mathbf{y}_3 & \dots & \dots & \mathbf{y}_{N+2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \dots & \mathbf{y}_{N+L-1} \end{pmatrix}$$

Propriétés asymptotiques des éléments propres de $\frac{\mathbf{Y}_N^L \mathbf{Y}_N^{L*}}{N}$

Travaux futurs (ANR DIONISOS).

Matrices de covariances empiriques associées à des moments d'ordre 4 (avec L. Pastur)

$$\frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n \otimes \mathbf{y}_n) (\mathbf{y}_n \otimes \mathbf{y}_n)^*$$

- ▶ M et N comparables : El-Karoui (2008)
- ▶ M^2 et N comparables : plus intéressant.

Participation au développement d'outils génériques utiles.

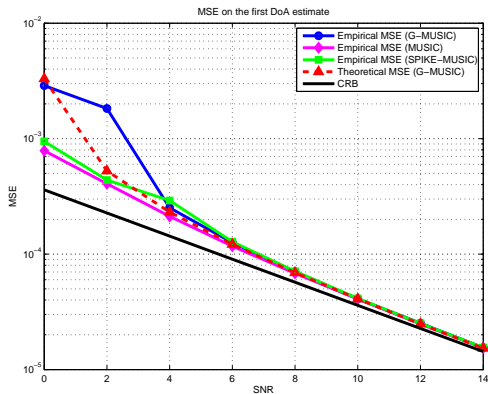
- ▶ Etude des fluctuations de fonctionnelles des valeurs propres / vecteurs propres.
- ▶ Grandes déviations pour les grandes matrices aléatoires.

Collaborateurs.

- ▶ W. Hachem (LTCI)
- ▶ R. Couillet, M. Debbah (Supelec)
- ▶ P. Vallet (IMS)
- ▶ X. Mestre (CTTC, Espagne)
- ▶ L. Pastur (Institut de Physique des Basses Températures, Ukraine)
- ▶ J.W. Silverstein (Univ North Carolina, Raleigh).

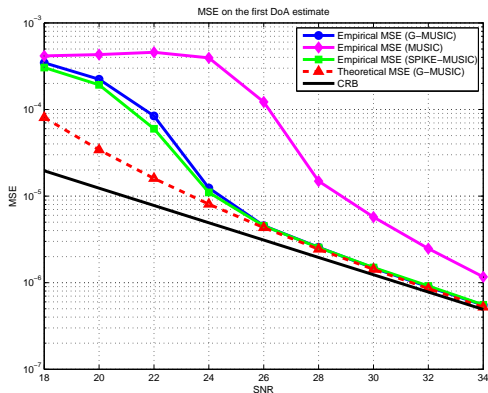
Illustrations.

$$K = 2, M = 20, N = 40, \theta_2 - \theta_1 = \frac{\pi}{4}.$$



Erreur quadratique moyenne de l'estimateur de θ_1 en fonction du rapport signal sur bruit.

$$K = 2, M = 40, N = 80, \theta_2 - \theta_1 = \frac{\pi}{2M}.$$



Erreur quadratique moyenne de l'estimateur de θ_1 en fonction du rapport signal sur bruit.

Publications associées.

- ▶ P. Loubaton, P. Vallet, "Almost sure localization of the eigenvalues in a gaussian information plus noise model. Applications to the spiked models", *Electronic J. on Probability*, October 2011, pp. 1934-1959.
- ▶ P. Vallet, P. Loubaton, X. Mestre, "Improved Subspace Estimation for Multivariate Observations of High Dimension : The Deterministic Signal Case", *IEEE Trans. IT*, vol. 58, no. 2, pp. 1043-1068, 2012.
- ▶ W. Hachem, P. Loubaton, J. Najim, X. Mestre, P. Vallet "Large information plus noise random matrix models and consistent subspace estimation in large sensor networks ", *Random Matrices, Theory and Applications (RMTA)*, vol. 1, no. 2 (2012).
- ▶ W. Hachem, P. Loubaton, J. Najim, X. Mestre, P. Vallet, "A subspace estimator for fixed rank perturbations of large random matrices", *JMVA*, 114 (2013), pp. 427-447.

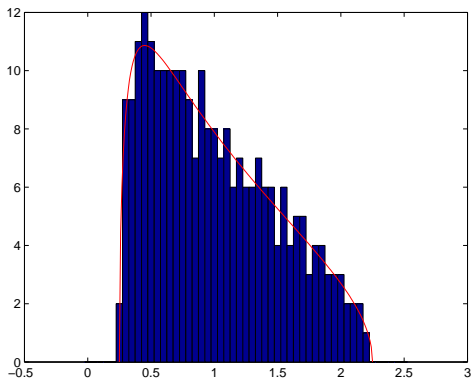
Exemple classique de difficulté : estimation de la matrice de covariance de l'observation.

$(\mathbf{y}_n)_{n=1,\dots,N}$ i.i.d, estimer $\mathbf{R} = \mathbb{E}(\mathbf{y}_n \mathbf{y}_n^*)$.

Matrice de covariance empirique $\hat{\mathbf{R}}_N = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^* = \frac{\mathbf{Y} \mathbf{Y}^*}{N}$

- ▶ Si $N \rightarrow +\infty$ et M fixe, $\|\mathbf{R} - \hat{\mathbf{R}}_N\| \rightarrow 0$
- ▶ Si $M \rightarrow +\infty, N \rightarrow +\infty, \frac{M}{N} \rightarrow c$, propriété inexacte, difficile d'estimer \mathbf{R} sans information a priori (parcimonie, matrice structurée,...)
- ▶ $M \rightarrow +\infty, N \rightarrow +\infty, \frac{M}{N} \rightarrow c$ et $\mathbf{R} = \mathbf{I}$, les valeurs propres de $\hat{\mathbf{R}}_N$ ne se concentrent pas autour de 1.

$\mathbf{R} = \mathbf{I}$, histogramme des valeurs propres de $\hat{\mathbf{R}}_N$ pour $M = 256, \frac{M}{N} = \frac{1}{4}$.



- La distribution empirique des valeurs propres de $\hat{\mathbf{R}}_N$ converge vers la distribution de Marcenko-Pastur (en rouge) portée par l'intervalle $[(1 - \sqrt{\frac{M}{N}})^2, (1 + \sqrt{\frac{M}{N}})^2]$.
- Les valeurs propres de $\hat{\mathbf{R}}_N$ restent au voisinage de $[(1 - \sqrt{\frac{M}{N}})^2, (1 + \sqrt{\frac{M}{N}})^2]$