Research Program

Romain COUILLET

CentraleSupélec
Université Paris-Sud 11

11 février 2016

CentraleSupélec

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
- Axis 1 : Robust Estimation in Large Dimensions
- Axis 2 : Classification in Large Dimensions
- Axis 3 : Random Matrices and Neural Networks
- Axis 4 : Graphs

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
- Axis 1 : Robust Estimation in Large Dimensions
- Axis 2 : Classification in Large Dimensions
- Axis 3 : Random Matrices and Neural Networks
- Axis 4 : Graphs

Education and Professional Experience

Professional Experience

Full Professor since Jan. 2011
CentraleSupélec, Gif sur Yvette, France.
Telecom Department, LANEAS group, Dvision Signals & Stats

Diplomas

Habilitation à Diriger des Recherches Feb. 2015

Place	University Paris–Saclay, France
Topic	Robust Estimation in the Large Random Matrix Regime

PhD in Physics Nov. 2010

Place	CentraleSupélec, Gif sur Yvette, France
Topic	Application of Random Matrix Theory to Future Wireless Flexible Networks
Advis.	Mérouane Debbah

Engineer and Master Diplomas Mar. 2008

Place	Telecom ParisTech, Paris, France
Grade	Very Good (Très Bien)
Topic	Communications, embedded systems, computer science.

Teaching Activities and Research Projects

Teaching Activities

ENS Cachan, Cachan, France since 2013

Courses	Master MVA, 18 hrs/year

CentraleSupélec, Gif sur Yvette, France since 2011

Courses	PhD level, 18 hrs/year
	Master SAR, research seminars, 24 hrs/year
	Undergraduate, lectures + practical courses, 70 hrs/year
Advising	Interns, undergraduate projects, ∼8/year

Research : Projects

HUAWEI RMTin5G	100% (PI)	2015-2016
ANR RMT4GRAPH	100% (PI)	2014-2017
ERC MORE	50%	2012-2017
ANR DIONISOS	25%	2012-2016

Research : Community Life

Special Session organizations	4
IEEE Senior Member	since 2015
IEEE SPTM technical committee member	since 2014
IEEE TSP Associate Editor	since 2015
Member of GRETSI	since 2011

PhD students

✔ **Axel MÜLLER (research engineer at HUAWEI Labs, Paris)** **2011–2014**

Subject	Random matrix models for multi-cellular wireless communications
Advising	50%, with M. Debbah (CentraleSupélec)
Publications	3 articles in IEEE-JSTSP, -TIT, -TSP, 5 IEEE conferences
Awards	1 best student paper award.

✔ **Julia VINOGRADOVA (postdoc at Linköping University, Sweden)** **2011–2014**

Subject	Random matrix theory applied to detection and estimation in antenna arrays
Advising	50%, with W. Hachem (Telecom ParisTech)
Publications	2 articles in IEEE-TSP, 2 IEEE conferences

✔ **Azary ABBOUD (postdoc at INRIA, France)** **2012–2015**

Subject	Distributed optimization in smart grids
Advising	33%, with M. Debbah and H. Siguerdidjane (CentraleSupélec)
Publications	1 article in IEEE-TSP, 1 IEEE conference

✎ **Gil KATZ** **2013–2016**

Subject	Interactive communications for distributed computation
Advising	33%, with M. Debbah and P. Piantanida (CentraleSupélec)
Publications	1 IEEE conference

✎ **Hafiz TIOMOKO ALI** **2015–2018**

Subject	Random matrices in machine learning
Advising	100%
Publications	2 IEEE conferences.

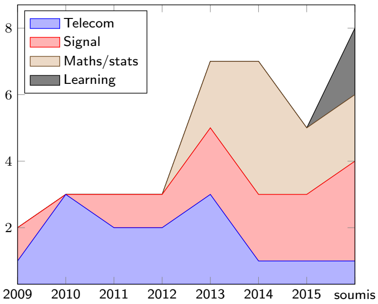
Publication Record (as of February 1st, 2016)

Publications	Books : 1, Chapters : 3, Journals : 36, Conferences : 53, Patents : 4.
Citations	1256 (five best : 282, 204, 84, 49, 33)
Indices	h-index : 17, i10-index : 25

Subjects

Mathematics	random matrix theory, statistics
Applications	machine learning, signal processing, communications

Journal Articles per Area



Prizes and Awards

IEEE Senior Member	2016
CNRS Bronze Medal (section INS2I)	2013
IEEE ComSoc Outstanding Young Researcher Award (EMEA region)	2013
EEA/GdR ISIS/GRETSI PhD thesis award	2011

Paper Awards

Second prize of the IEEE Australia Council Student Paper Contest	2013
Best Student Paper Award Final of the IEEE Asilomar Conference	2011
Best Student Paper Award of the ValueTools Conference	2008

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
- Axis 1 : Robust Estimation in Large Dimensions
- Axis 2 : Classification in Large Dimensions
- Axis 3 : Random Matrices and Neural Networks
- Axis 4 : Graphs

Context

The "BigData" Challenge

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)

Context

The “BigData” Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis
2. techniques relying on poorly robust empirical estimates or barely usable robust approaches

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis
2. techniques relying on poorly robust empirical estimates or barely usable robust approaches
3. bipolarity between :
   - powerful techniques based on models (signal processing approach)
   - ad-hoc techniques based on data (machine learning/stats approach)

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis
2. techniques relying on poorly robust empirical estimates or barely usable robust approaches
3. bipolarity between :
   - powerful techniques based on models (signal processing approach)
   - ad-hoc techniques based on data (machine learning/stats approach)

Our approach

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis
2. techniques relying on poorly robust empirical estimates or barely usable robust approaches
3. bipolarity between :
   - powerful techniques based on models (signal processing approach)
   - ad-hoc techniques based on data (machine learning/stats approach)

Our approach

1. development of methods and *mathematical* tools to handle large and numerous datasets

Context

The "BigData" Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis
2. techniques relying on poorly robust empirical estimates or barely usable robust approaches
3. bipolarity between :
   - powerful techniques based on models (signal processing approach)
   - ad-hoc techniques based on data (machine learning/stats approach)

Our approach

1. development of methods and *mathematical* tools to handle large and numerous datasets
2. revisit robust statistics in large dimensions

Context

The “BigData” Challenge

1. dramatic increase of data dimension (and number)
2. importance of outlying and missing data
3. data heterogeneity

Limitations of classical tools

1. classical statistics limited by small but numerous data hypothesis
2. techniques relying on poorly robust empirical estimates or barely usable robust approaches
3. bipolarity between :
   - powerful techniques based on models (signal processing approach)
   - ad-hoc techniques based on data (machine learning/stats approach)

Our approach

1. development of methods and *mathematical* tools to handle large and numerous datasets
2. revisit robust statistics in large dimensions
3. (for lack of better approach) better understand and improve ad-hoc techniques on simple but large dimensional models.

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
  Axis 1 : Robust Estimation in Large Dimensions
  Axis 2 : Classification in Large Dimensions
  Axis 3 : Random Matrices and Neural Networks
  Axis 4 : Graphs

Axis 1 : Robust Estimation in Large Dimensions

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_p$, but

- potentially heavy tailed
- existence of outliers

Axis 1 : Robust Estimation in Large Dimensions

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_p$, but

- potentially heavy tailed
- existence of outliers

Several Estimators for C_p

Axis 1 : Robust Estimation in Large Dimensions

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_p$, but

- potentially heavy tailed
- existence of outliers

Several Estimators for C_p

- ML estimator in Gaussian case : sample covariance matrix

$$\hat{C}_p = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^*.$$

  - very practical, used in many methods
  - but very sensitive to outliers

Axis 1 : Robust Estimation in Large Dimensions

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_p$, but

- potentially heavy tailed
- existence of outliers

Several Estimators for C_p

- ML estimator in Gaussian case : sample covariance matrix

$$\hat{C}_p = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^*.$$

  - very practical, used in many methods
  - but very sensitive to outliers

- [Huber'67 ; Maronna'76] Robust Estimators (outliers or heavy tails)

$$\hat{C}_p = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{p} x_i^* \hat{C}_p^{-1} x_i\right) x_i x_i^*.$$

Axis 1 : Robust Estimation in Large Dimensions

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_p$, but

- potentially heavy tailed
- existence of outliers

Several Estimators for C_p

- ML estimator in Gaussian case : sample covariance matrix

$$\hat{C}_p = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^*.$$

  - very practical, used in many methods
  - but very sensitive to outliers

- [Huber'67 ; Maronna'76] Robust Estimators (outliers or heavy tails)

$$\hat{C}_p = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{p} x_i^* \hat{C}_p^{-1} x_i \right) x_i x_i^*.$$

- [Pascal'13 ; Chen'11] Regularized Versions for Large Data (all n, p),

$$\hat{C}_p(\rho) = (1-\rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{p} x_i^* \hat{C}_p^{-1}(\rho) x_i} + \rho I_N.$$

Problem and Objectives

Problem and Objectives

- Ill-understood estimators, difficult to use (implicit definition of $\hat{C}_p$)

Problem and Objectives

- Ill-understood estimators, difficult to use (implicit definition of $\hat{C}_p$)
- Only known results for fixed p and $n \to \infty$: not appropriate in BigData.

Problem and Objectives

- Ill-understood estimators, difficult to use (implicit definition of $\hat{C}_p$)
- Only known results for fixed p and $n \to \infty$: not appropriate in BigData.
- We thus need :
  - study $\hat{C}_p$ as $n, p \to \infty$
  - exploit the double-concentration effect to better understand $\hat{C}_p$.

Axis 1 : Robust Estimation in Large Dimensions

Problem and Objectives

- Ill-understood estimators, difficult to use (implicit definition of $\hat{C}_p$)
- Only known results for fixed p and $n \to \infty$: not appropriate in BigData.
- We thus need :
  - study $\hat{C}_p$ as $n, p \to \infty$
  - exploit the double-concentration effect to better understand $\hat{C}_p$.

Results and Perspectives

- ✔ Asymptotic approximation of $\hat{C}_p$ by a tractable equivalent model
- ✔ Second order statistics (CLT type) for $\hat{C}_p$
- ✔ Study of elliptical cases, outliers, regularized or not.
- ✔ Applications :
  - radar array processing (impulsiveness due to clutter)
  - financial data processing

Axis 1 : Robust Estimation in Large Dimensions

Problem and Objectives

- Ill-understood estimators, difficult to use (implicit definition of $\hat{C}_p$)
- Only known results for fixed p and $n \to \infty$: not appropriate in BigData.
- We thus need :
  - study $\hat{C}_p$ as $n, p \to \infty$
  - exploit the double-concentration effect to better understand $\hat{C}_p$.

Results and Perspectives

✔ Asymptotic approximation of $\hat{C}_p$ by a tractable equivalent model

✔ Second order statistics (CLT type) for $\hat{C}_p$

✔ Study of elliptical cases, outliers, regularized or not.

✔ Applications :
- radar array processing (impulsiveness due to clutter)
- financial data processing

✎ Joint mean and covariance estimation

✎ Study of robust regression

✎ More generally, deeper study of iterative methods in large dimensions (such as AMP).

Theorem ([Couillet,Pascal,Silverstein'15] Maronna Estimator)

For $x_i = \sqrt{\tau_i} w_i$, with τ_i impulsive, w_i orthogonal and isotropic, $\|w_i\| = p$,

$$\left\| \hat{C}_p - \hat{S}_p \right\| \xrightarrow{\text{p.s.}} 0$$

in spectral norm, where

$$\hat{C}_p = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{p} x_i^* \hat{C}_p^{-1} x_i \right) x_i x_i^*$$

$$\hat{S}_p = \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_p) x_i x_i^*$$

with $v(t)$ similar to $u(t)$ and γ_p unique solution of

$$1 = \frac{1}{n} \sum_{j=1}^{n} \frac{\gamma_p v(\tau_i \gamma_p)}{1 + c\gamma_p v(\tau_i \gamma_p)}.$$

Consequences

Consequences

- Difficult object $\hat{C}_p$ made tractable thanks to $\hat{S}_p$

Consequences

- Difficult object $\hat{C}_p$ made tractable thanks to $\hat{S}_p$
- Analysis and optimization possible by replacing $\hat{C}_p$ by $\hat{S}_p$.

Consequences

- Difficult object $\hat{C}_p$ made tractable thanks to $\hat{S}_p$
- Analysis and optimization possible by replacing $\hat{C}_p$ by $\hat{S}_p$.

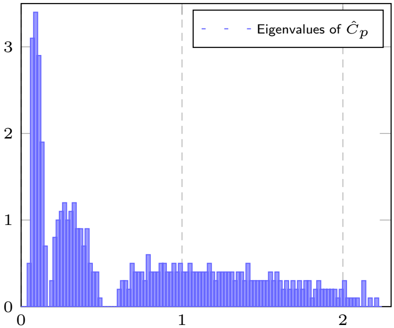


FIGURE – $n = 2500$, $p = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Consequences

- Difficult object $\hat{C}_p$ made tractable thanks to $\hat{S}_p$
- Analysis and optimization possible by replacing $\hat{C}_p$ by $\hat{S}_p$.

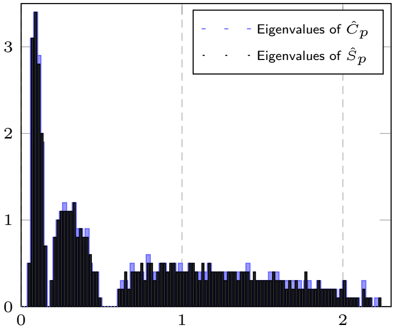


FIGURE – $n = 2500$, $p = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Consequences

- Difficult object $\hat{C}_p$ made tractable thanks to $\hat{S}_p$
- Analysis and optimization possible by replacing $\hat{C}_p$ by $\hat{S}_p$.

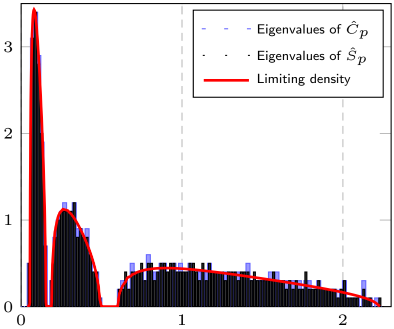


FIGURE – $n = 2500$, $p = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Application to Detection in Radars

Application to Detection in Radars

- Hypothesis testing under impulsive noise : purely noisy inputs $x_1, \ldots, x_n$, $x_i = \sqrt{\tau_i} w_i$, new datum

$$y = \begin{cases} \sqrt{\tau} w & , \mathcal{H}_0 \\ s + \sqrt{\tau} w & , \mathcal{H}_1 \end{cases}$$

Application to Detection in Radars

- Hypothesis testing under impulsive noise : purely noisy inputs $x_1, \ldots, x_n$, $x_i = \sqrt{\tau_i} w_i$, new datum

$$y = \begin{cases} \sqrt{\tau} w & , \mathcal{H}_0 \\ s + \sqrt{\tau} w & , \mathcal{H}_1 \end{cases}$$

- Robust detector $T_p(\rho)$ given by

$$T_p(\rho) \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \frac{\gamma}{\sqrt{p}}$$

where

$$T_p(\rho) = \frac{|y^* \hat{C}_p^{-1}(\rho) s|}{\sqrt{y^* \hat{C}_p^{-1}(\rho) y} \sqrt{p^* \hat{C}_p^{-1}(\rho) p}}$$

$$\hat{C}_p(\rho) = (1-\rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{p} x_i^* \hat{C}_p(\rho)^{-1} x_i} + \rho I_p.$$

Application to Detection in Radars

- Hypothesis testing under impulsive noise : purely noisy inputs $x_1, \ldots, x_n$, $x_i = \sqrt{\tau_i} w_i$, new datum

$$y = \begin{cases} \sqrt{\tau} w & , \mathcal{H}_0 \\ s + \sqrt{\tau} w & , \mathcal{H}_1 \end{cases}$$

- Robust detector $T_p(\rho)$ given by

$$T_p(\rho) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\gamma}{\sqrt{p}}$$

where

$$T_p(\rho) = \frac{|y^* \hat{C}_p^{-1}(\rho) s|}{\sqrt{y^* \hat{C}_p^{-1}(\rho) y} \sqrt{p^* \hat{C}_p^{-1}(\rho) p}}$$

$$\hat{C}_p(\rho) = (1-\rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{p} x_i^* \hat{C}_p(\rho)^{-1} x_i} + \rho I_p.$$

Objectives

- performance analysis
- find optimal regularization ρ parameter.

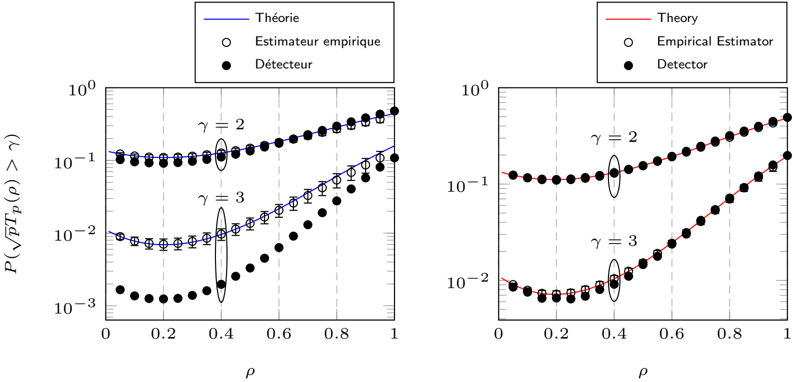


FIGURE – False alarm rate $P(\sqrt{p}T_p(\rho) > \gamma)$, for $p = 20$ (left), $p = 100$ (right), $s = p^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_p]_{ij} = 0.7^{|i-j|}$, $p/n = 1/2$.

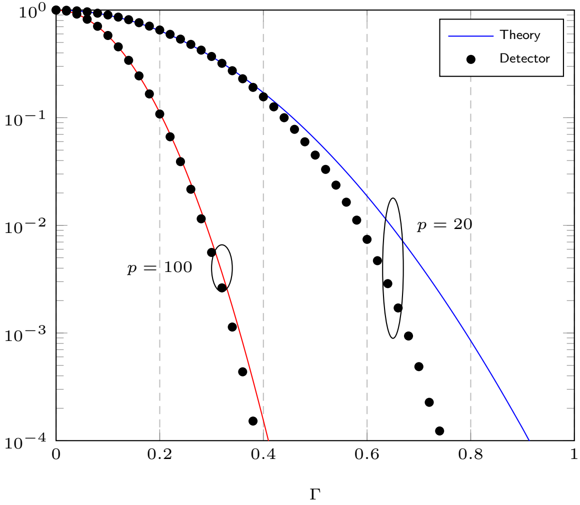


FIGURE – False alarm rate $P(T_p(\hat{\rho}_p^*) > \Gamma)$, $\hat{\rho}_p^*$ best estimated ρ, for $p = 20$ and $p = 100$, $s = p^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $p/n = 1/2$ and $[C_p]_{ij} = 0.7^{|i-j|}$.

Axis 1 : Robust Estimation in Large Dimensions

Theoretical Results (clickable title links)

- **R. Couillet**, M. McKay, "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators", Elsevier Journal of Multivariate Analysis, vol. 131, pp. 99-120, 2014.
- **R. Couillet**, F. Pascal, J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples", Elsevier Journal of Multivariate Analysis, vol. 139, pp. 56-78, 2015.
- D. Morales-Jimenez, **R. Couillet**, M. McKay, "Large Dimensional Analysis of Robust M-Estimators of Covariance with Outliers", IEEE Transactions on Signal Processing, vol. 63, no. 21, pp. 5784-5797, 2015.
- **R. Couillet**, A. Kammoun, F. Pascal, "Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals", Elsevier Journal of Multivariate Analysis, vol. 143, pp. 249-274, 2016.

Applications (clickable title links)

- **R. Couillet**, A. Kammoun, F. Pascal, "Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals", Elsevier Journal of Multivariate Analysis, vol. 143, pp. 249-274, 2016.
- **R. Couillet**, "Robust spiked random matrices and a robust G-MUSIC estimator", Elsevier Journal of Multivariate Analysis, vol. 140, pp. 139-161, 2015.

Axis 1 : Robust Estimation in Large Dimensions

L. Yang, **R. Couillet**, M. McKay, "A Robust Statistics Approach to Minimum Variance Portfolio Optimization" IEEE Transactions on Signal Processing, vol. 63, no. 24, pp. 6684–6697, 2015.

A. Kammoun, **R. Couillet**, F. Pascal, M.-S. Alouini, "Optimal Design of the Adaptive Normalized Matched Filter Detector" (submitted to) IEEE Transactions on Information Theory, 2015, arXiv Preprint 1504.01252.

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
- Axis 1 : Robust Estimation in Large Dimensions
- Axis 2 : Classification in Large Dimensions
- Axis 3 : Random Matrices and Neural Networks
- Axis 4 : Graphs

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ belonging to k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ to identify

- in supervised manner : numerous labelled data (e.g., support vector machine)
- in unsupervised manner : no labelled data (e.g., kernel spectral clustering)
- in semi-supervised manner : (few) labelled data (e.g., harmonic function method).

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ belonging to k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ to identify

- in supervised manner : numerous labelled data (e.g., support vector machine)
- in unsupervised manner : no labelled data (e.g., kernel spectral clustering)
- in semi-supervised manner : (few) labelled data (e.g., harmonic function method).

Spectral Algorithms

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ belonging to k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ to identify

- in supervised manner : numerous labelled data (e.g., support vector machine)
- in unsupervised manner : no labelled data (e.g., kernel spectral clustering)
- in semi-supervised manner : (few) labelled data (e.g., harmonic function method).

Spectral Algorithms

- Data often non linearly separable

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ belonging to k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ to identify

- in supervised manner : numerous labelled data (e.g., support vector machine)
- in unsupervised manner : no labelled data (e.g., kernel spectral clustering)
- in semi-supervised manner : (few) labelled data (e.g., harmonic function method).

Spectral Algorithms

- Data often non linearly separable
- Numerous methods based on kernel matrices $K \in \mathbb{R}^{n\times n}$, with (for instance)

$$K_{ij} = f\left(\|x_i - x_j\|^2\right)$$

and f some function (often decreasing).

Axis 2 : Classification in Large Dimensions

Baseline Scenario : $x_1, \ldots, x_n \in \mathbb{R}^p$ belonging to k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ to identify

- in supervised manner : numerous labelled data (e.g., support vector machine)
- in unsupervised manner : no labelled data (e.g., kernel spectral clustering)
- in semi-supervised manner : (few) labelled data (e.g., harmonic function method).

Spectral Algorithms

- Data often non linearly separable
- Numerous methods based on kernel matrices $K \in \mathbb{R}^{n \times n}$, with (for instance)

$$K_{ij} = f\left(\|x_i - x_j\|^2\right)$$

  and f some function (often decreasing).
- Spectral methods consist in :
  - extracting dominating eigenvectors of K (spectral clustering)
  - solve optimization problem based on K (support vector machine)
  - linear functional of K (semi-supervised methods)

Problems and Objectives

Problems and Objectives

- Matrix K very difficult to analyze for small dimensional x_i (even Gaussian)
- Qualitative understanding of the tools, difficult to optimize.

Problems and Objectives

- Matrix K very difficult to analyze for small dimensional x_i (even Gaussian)
- Qualitative understanding of the tools, difficult to optimize.
- We need here :
  - study K as $n, p \to \infty$
  - exploit the double-concentration to better understand K
  - deduce quantitative performance of learning methods
  - improve performances as well as methods.

Problems and Objectives

- Matrix K very difficult to analyze for small dimensional x_i (even Gaussian)
- Qualitative understanding of the tools, difficult to optimize.
- We need here :
  - study K as $n, p \to \infty$
  - exploit the double-concentration to better understand K
  - deduce quantitative performance of learning methods
  - improve performances as well as methods.

Results and Perspectives

- ✔ Development of tools for large dimensional analysis of kernel matrices.
- ✔ Thorough analysis of spectral clustering performance in (large) Gaussian mixtures.
- ✔ Optimization for subspace clustering (new approach, undergoing patent by HUAWEI).

Axis 2 : Classification in Large Dimensions

Problems and Objectives

- Matrix K very difficult to analyze for small dimensional x_i (even Gaussian)
- Qualitative understanding of the tools, difficult to optimize.
- We need here :
  - study K as $n, p \to \infty$
  - exploit the double-concentration to better understand K
  - deduce quantitative performance of learning methods
  - improve performances as well as methods.

Results and Perspectives

- ✔ Development of tools for large dimensional analysis of kernel matrices.
- ✔ Thorough analysis of spectral clustering performance in (large) Gaussian mixtures.
- ✔ Optimization for subspace clustering (new approach, undergoing patent by HUAWEI).
- ✎ Generalization to semi-supervised case.
- ✎ Study of support vector machines in this context.
- ✎ Generalization to more realistic models, deeper comparison to real datasets.

Model : Consider Laplacian matrix (core of Ng–Weiss–Jordan algorithm)

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}} - n\frac{D^{\frac{1}{2}}1_n1_n^\mathsf{T}D^{\frac{1}{2}}}{1_n^\mathsf{T}D1_n}$$

where $D = \operatorname{diag}(K1_n)$ and $x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, \frac{1}{p}C_a)$, $|\mathcal{C}_a| = n_a$.

Model : Consider Laplacian matrix (core of Ng–Weiss–Jordan algorithm)

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}} - n\frac{D^{\frac{1}{2}}1_n1_n^\mathsf{T}D^{\frac{1}{2}}}{1_n^\mathsf{T}D1_n}$$

where $D = \mathrm{diag}(K1_n)$ and $x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, \frac{1}{p}C_a)$, $|\mathcal{C}_a| = n_a$.

Theorem ([Couillet,Benaych'16] Equivalent to Laplacian Matrix)

As $n, p \to \infty$, under appropriate hypotheses, $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ with

$$\hat{L} = -2\frac{f'(\tau)}{f(\tau)}\left[\frac{1}{p}PW^\mathsf{T}WP + UBU^\mathsf{T}\right] + \alpha(\tau)I_n$$

where $\tau = 2\sum_a \frac{n_a}{np}\mathrm{tr}\,C_a$, $W = [w_1, \ldots, w_n]$ ($x_i = \mu_a + w_i$), $P = I_n - \frac{1}{n}1_n1_n^\mathsf{T}$,

$$U = \left[\frac{1}{\sqrt{p}}J, \Phi, \psi\right], \quad B = \begin{bmatrix} B_{11} & * \\ * & * \end{bmatrix}$$

$$B_{11} = M^\mathsf{T}M + \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)}\right)tt^\mathsf{T} - \frac{f''(\tau)}{f'(\tau)}T + \frac{p}{n}\frac{f(\tau)\alpha(\tau)}{2f'(\tau)}1_k1_k^\mathsf{T}.$$

Model : Consider Laplacian matrix (core of Ng–Weiss–Jordan algorithm)

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}} - n\frac{D^{\frac{1}{2}}1_n1_n^\mathsf{T}D^{\frac{1}{2}}}{1_n^\mathsf{T}D1_n}$$

where $D = \operatorname{diag}(K1_n)$ and $x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, \frac{1}{p}C_a)$, $|\mathcal{C}_a| = n_a$.

Theorem ([Couillet,Benaych'16] Equivalent to Laplacian Matrix)

As $n, p \to \infty$, under appropriate hypotheses, $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ with

$$\hat{L} = -2\frac{f'(\tau)}{f(\tau)}\left[\frac{1}{p}PW^\mathsf{T}WP + UBU^\mathsf{T}\right] + \alpha(\tau)I_n$$

where $\tau = 2\sum_a \frac{n_a}{np}\operatorname{tr} C_a$, $W = [w_1, \ldots, w_n]$ ($x_i = \mu_a + w_i$), $P = I_n - \frac{1}{n}1_n1_n^\mathsf{T}$,

$$U = \left[\frac{1}{\sqrt{p}}J, \Phi, \psi\right], \quad B = \begin{bmatrix} B_{11} & * \\ * & * \end{bmatrix}$$

$$B_{11} = M^\mathsf{T}M + \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)}\right)tt^\mathsf{T} - \frac{f''(\tau)}{f'(\tau)}T + \frac{p}{n}\frac{f(\tau)\alpha(\tau)}{2f'(\tau)}1_k1_k^\mathsf{T}.$$

Important Notations :

$\frac{1}{\sqrt{p}}J = [j_1, \ldots, j_k]$, j_a canonical vector of class $\mathcal{C}_a$.

Model : Consider Laplacian matrix (core of Ng–Weiss–Jordan algorithm)

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}} - n\frac{D^{\frac{1}{2}}1_n 1_n^{\mathsf{T}} D^{\frac{1}{2}}}{1_n^{\mathsf{T}} D 1_n}$$

where $D = \mathrm{diag}(K1_n)$ and $x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, \frac{1}{p}C_a)$, $|\mathcal{C}_a| = n_a$.

Theorem ([Couillet,Benaych'16] Equivalent to Laplacian Matrix)

As $n, p \to \infty$, under appropriate hypotheses, $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ with

$$\hat{L} = -2\frac{f'(\tau)}{f(\tau)}\left[\frac{1}{p}PW^{\mathsf{T}}WP + UBU^{\mathsf{T}}\right] + \alpha(\tau)I_n$$

where $\tau = 2\sum_a \frac{n_a}{np} \mathrm{tr}\, C_a$, $W = [w_1, \ldots, w_n]$ ($x_i = \mu_a + w_i$), $P = I_n - \frac{1}{n}1_n 1_n^{\mathsf{T}}$,

$$U = \left[\frac{1}{\sqrt{p}}J, \Phi, \psi\right], \quad B = \begin{bmatrix} B_{11} & * \\ * & * \end{bmatrix}$$

$$B_{11} = M^{\mathsf{T}}M + \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)}\right)tt^{\mathsf{T}} - \frac{f''(\tau)}{f'(\tau)}T + \frac{p}{n}\frac{f(\tau)\alpha(\tau)}{2f'(\tau)}1_k 1_k^{\mathsf{T}}.$$

Important Notations :
$M = [\mu_1^\circ, \ldots, \mu_k^\circ]$, $\mu_a^\circ = \mu_a - \sum_{b=1}^k \frac{n_b}{n}\mu_b$.

Model : Consider Laplacian matrix (core of Ng–Weiss–Jordan algorithm)

$$L = nD^{-\frac{1}{2}} K D^{-\frac{1}{2}} - n \frac{D^{\frac{1}{2}} 1_n 1_n^\mathsf{T} D^{\frac{1}{2}}}{1_n^\mathsf{T} D 1_n}$$

where $D = \mathrm{diag}(K 1_n)$ and $x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, \frac{1}{p} C_a)$, $|\mathcal{C}_a| = n_a$.

Theorem ([Couillet,Benaych'16] Equivalent to Laplacian Matrix)

As $n, p \to \infty$, under appropriate hypotheses, $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ with

$$\hat{L} = -2 \frac{f'(\tau)}{f(\tau)} \left[\frac{1}{p} P W^\mathsf{T} W P + U B U^\mathsf{T} \right] + \alpha(\tau) I_n$$

where $\tau = 2 \sum_a \frac{n_a}{np} \mathrm{tr}\, C_a$, $W = [w_1, \ldots, w_n]$ ($x_i = \mu_a + w_i$), $P = I_n - \frac{1}{n} 1_n 1_n^\mathsf{T}$,

$$U = \left[\frac{1}{\sqrt{p}} J, \Phi, \psi \right], \quad B = \begin{bmatrix} B_{11} & * \\ * & * \end{bmatrix}$$

$$B_{11} = M^\mathsf{T} M + \left(\frac{5 f'(\tau)}{8 f(\tau)} - \frac{f''(\tau)}{2 f'(\tau)} \right) t t^\mathsf{T} - \frac{f''(\tau)}{f'(\tau)} T + \frac{p}{n} \frac{f(\tau) \alpha(\tau)}{2 f'(\tau)} 1_k 1_k^\mathsf{T}.$$

Important Notations :

$t = \left[\frac{1}{\sqrt{p}} \mathrm{tr}\, C_1^\circ, \ldots, \frac{1}{\sqrt{p}} \mathrm{tr}\, C_k^\circ \right]$, $C_a^\circ = C_a - \sum_{b=1}^k \frac{n_b}{n} C_b$.

Model : Consider Laplacian matrix (core of Ng–Weiss–Jordan algorithm)

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}} - n\frac{D^{\frac{1}{2}}1_n1_n^\mathsf{T}D^{\frac{1}{2}}}{1_n^\mathsf{T}D1_n}$$

where $D = \operatorname{diag}(K1_n)$ and $x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, \frac{1}{p}C_a)$, $|\mathcal{C}_a| = n_a$.

Theorem ([Couillet,Benaych'16] Equivalent to Laplacian Matrix)

As $n, p \to \infty$, under appropriate hypotheses, $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ with

$$\hat{L} = -2\frac{f'(\tau)}{f(\tau)}\left[\frac{1}{p}PW^\mathsf{T}WP + UBU^\mathsf{T}\right] + \alpha(\tau)I_n$$

where $\tau = 2\sum_a \frac{n_a}{np}\operatorname{tr} C_a$, $W = [w_1, \ldots, w_n]$ ($x_i = \mu_a + w_i$), $P = I_n - \frac{1}{n}1_n1_n^\mathsf{T}$,

$$U = \left[\frac{1}{\sqrt{p}}J, \Phi, \psi\right], \quad B = \begin{bmatrix} B_{11} & * \\ * & * \end{bmatrix}$$

$$B_{11} = M^\mathsf{T}M + \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)}\right)tt^\mathsf{T} - \frac{f''(\tau)}{f'(\tau)}T + \frac{p}{n}\frac{f(\tau)\alpha(\tau)}{2f'(\tau)}1_k1_k^\mathsf{T}.$$

Important Notations :

$T = \left\{\frac{1}{p}\operatorname{tr} C_a^\circ C_b^\circ\right\}_{a,b=1}^k$, $C_a^\circ = C_a - \sum_{b=1}^k \frac{n_b}{n}C_b$.

Consequences :

Consequences :

- thorough understanding of eigenvector structure

Consequences :

- thorough understanding of eigenvector structure
- important consequences to kernel choice (depends on derivatives $f^{(\ell)}(\tau)$)

Axis 2 : Classification in Large Dimensions

Consequences :

- thorough understanding of eigenvector structure
- important consequences to kernel choice (depends on derivatives $f^{(\ell)}(\tau)$)

Application to real data : MNIST database (μ_a, C_a evaluated from full database)

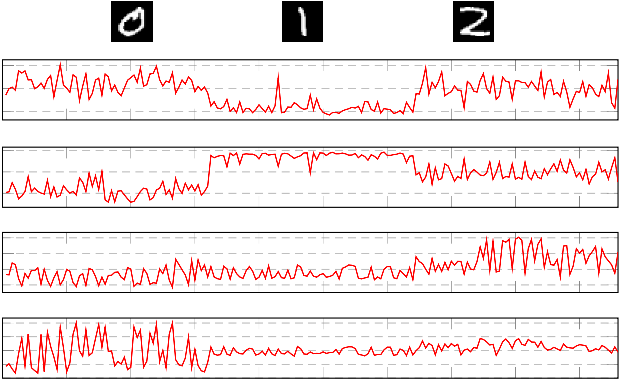


FIGURE – Four leading eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST dataset (**red**), equivalent Gaussian model (**black**), and asymptotic results (**blue**).

Consequences :

- thorough understanding of eigenvector structure
- important consequences to kernel choice (depends on derivatives $f^{(\ell)}(\tau)$)

Application to real data : MNIST database (μ_a, C_a evaluated from full database)

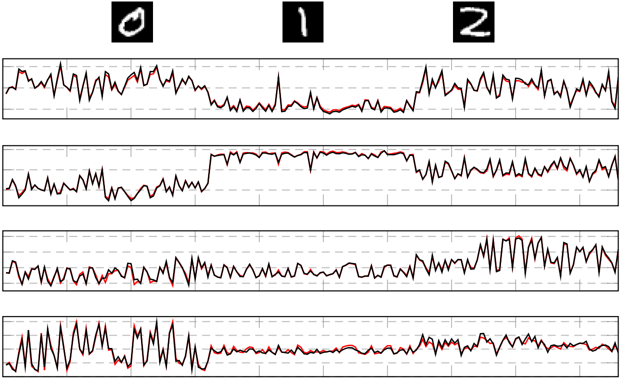


FIGURE – Four leading eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST dataset (**red**), equivalent Gaussian model (**black**), and asymptotic results (**blue**).

Consequences :

- thorough understanding of eigenvector structure
- important consequences to kernel choice (depends on derivatives $f^{(\ell)}(\tau)$)

Application to real data : MNIST database (μ_a, C_a evaluated from full database)

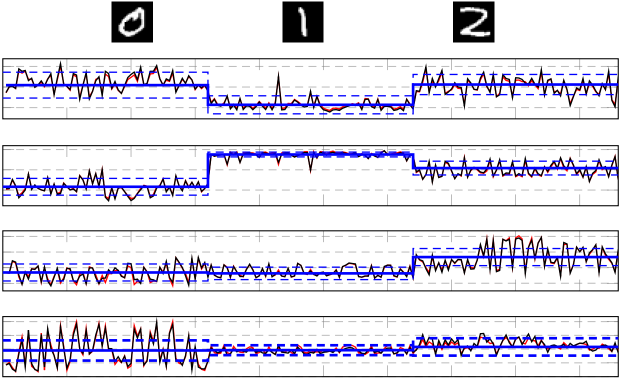


FIGURE – Four leading eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST dataset (**red**), equivalent Gaussian model (**black**), and asymptotic results (**blue**).

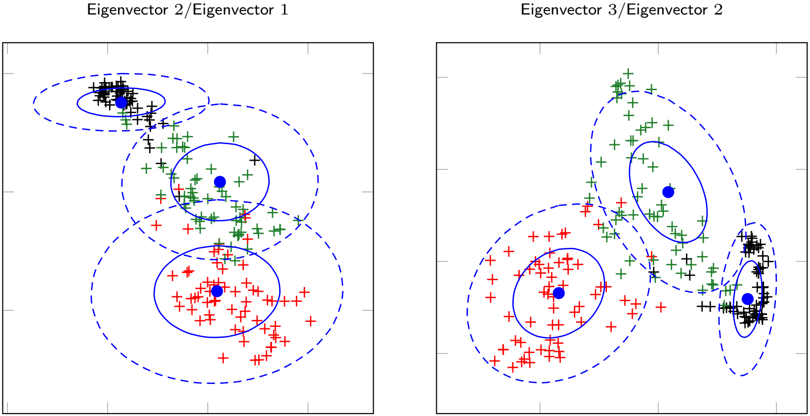


FIGURE – 2D plot of eigenvectors of L, MNIST database. Theoretical 1-σ and 2-σ standard deviations in **blue**. Classes 0, 1, 2 in colors.

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
- Axis 1 : Robust Estimation in Large Dimensions
- Axis 2 : Classification in Large Dimensions
- Axis 3 : Random Matrices and Neural Networks
- Axis 4 : Graphs

Baseline Scenario : Study of large recurrent neural nets (RNN and ESN)

Axis 3 : Random Matrices and Neural Networks

Baseline Scenario : Study of large recurrent neural nets (RNN and ESN)

- dynamical model

$$x_t = S\left(Wx_{t-1} + mu_t + \eta\varepsilon_t\right)$$

  with
  - n-node network with connectivity $W \in \mathbb{R}^{n\times n}$
  - activation function S
  - internal noise ε_t (biological model essentially)

Baseline Scenario : Study of large recurrent neural nets (RNN and ESN)

- dynamical model

$$x_t = S\left(Wx_{t-1} + mu_t + \eta\varepsilon_t\right)$$

  with
  - n-node network with connectivity $W \in \mathbb{R}^{n\times n}$
  - activation function S
  - internal noise ε_t (biological model essentially)
- readout $\omega \in \mathbb{R}^n$ training only (depth-1 NN) by LS regression

$$\omega = (XX^\mathsf{T})^{-1}Xr$$

  for T-long training $u \leftrightarrow r \in \mathbb{R}^T$ and $X = [x_1, \ldots, x_T]$.

Axis 3 : Random Matrices and Neural Networks

Baseline Scenario : Study of large recurrent neural nets (RNN and ESN)

- dynamical model

$$x_t = S\left(W x_{t-1} + m u_t + \eta \varepsilon_t\right)$$

  with
  - n-node network with connectivity $W \in \mathbb{R}^{n \times n}$
  - activation function S
  - internal noise ε_t (biological model essentially)
- readout $\omega \in \mathbb{R}^n$ training only (depth-1 NN) by LS regression

$$\omega = (XX^{\mathsf{T}})^{-1} X r$$

  for T-long training $u \leftrightarrow r \in \mathbb{R}^T$ and $X = [x_1, \ldots, x_T]$.

Performance Measures : quadratic errors in training and testing

- memory, training

$$\mathrm{MSE} = \left\| r - X^{\mathsf{T}} \omega \right\|$$

  for training couples $u \leftrightarrow r \in \mathbb{R}^T$.

Baseline Scenario : Study of large recurrent neural nets (RNN and ESN)

- dynamical model

$$x_t = S\left(Wx_{t-1} + mu_t + \eta\varepsilon_t\right)$$

  with
  - n-node network with connectivity $W \in \mathbb{R}^{n \times n}$
  - activation function S
  - internal noise ε_t (biological model essentially)
- readout $\omega \in \mathbb{R}^n$ training only (depth-1 NN) by LS regression

$$\omega = (XX^{\mathsf{T}})^{-1}Xr$$

  for T-long training $u \leftrightarrow r \in \mathbb{R}^T$ and $X = [x_1, \ldots, x_T]$.

Performance Measures : quadratic errors in training and testing

- memory, training

$$\mathrm{MSE} = \left\| r - X^{\mathsf{T}}\omega \right\|$$

  for training couples $u \leftrightarrow r \in \mathbb{R}^T$.
- generalization, test

$$\mathrm{MSE} = \left\| \hat{r} - \hat{X}^{\mathsf{T}}\omega \right\|$$

  for test couples $\hat{u} \leftrightarrow \hat{r} \in \mathbb{R}^{\hat{T}}$, $\omega = \omega(u, r)$.

Problems and Objectives

Problems and Objectives

- Performance evaluation essentially qualitative

Problems and Objectives

- Performance evaluation essentially qualitative
- Difficulty linked to randomness in W and ε_t

Problems and Objectives

- Performance evaluation essentially qualitative
- Difficulty linked to randomness in W and ε_t
- We need here :
  - Study asymptotic performances as $n, T, \hat{T} \to \infty$
  - Understand effects of W-defining hyper-parameters
  - Generalize study to more advanced models.

Problems and Objectives

- Performance evaluation essentially qualitative
- Difficulty linked to randomness in W and ε_t
- We need here :
  - Study asymptotic performances as $n, T, \hat{T} \to \infty$
  - Understand effects of W-defining hyper-parameters
  - Generalize study to more advanced models.

Results and Perspectives

Axis 3 : Random Matrices and Neural Networks

Problems and Objectives

- Performance evaluation essentially qualitative
- Difficulty linked to randomness in W and ε_t
- We need here :
  - Study asymptotic performances as $n, T, \hat{T} \to \infty$
  - Understand effects of W-defining hyper-parameters
  - Generalize study to more advanced models.

Results and Perspectives

- ✔ Asymptotic deterministic equivalents for training and testing MSE
- ✔ Multiples new consequences and intuitions
- ✔ Proposition of improved structures for W

Axis 3 : Random Matrices and Neural Networks

Problems and Objectives

- Performance evaluation essentially qualitative
- Difficulty linked to randomness in W and ε_t
- We need here :
  - Study asymptotic performances as $n, T, \hat{T} \to \infty$
  - Understand effects of W-defining hyper-parameters
  - Generalize study to more advanced models.

Results and Perspectives

- ✔ Asymptotic deterministic equivalents for training and testing MSE
- ✔ Multiples new consequences and intuitions
- ✔ Proposition of improved structures for W
- ✎ Generalization to non-linear setting
- ✎ Introduction of external memory, back-propagation
- ✎ Analogous study of deep networks, extreme ML, auto-encoders, etc.

Theorem ([Couillet,Wainrib'16] Training MSE for fixed W)

As $n, T \to \infty$, with $n/T \to c < 1$,

$$\mathrm{MSE} = \frac{1}{T} r^{\mathsf{T}} \left(I_T + \mathcal{R} + \frac{1}{\eta^2} U^{\mathsf{T}} \left\{ m^{\mathsf{T}} (W^i)^{\mathsf{T}} \tilde{\mathcal{R}}^{-1} W^j m \right\}_{i,j=0}^{T-1} U \right)^{-1} r + o(1)$$

where $U_{ij} = u_{i-j}$ and $\mathcal{R}$, $\tilde{\mathcal{R}}$ are solutions to

$$\mathcal{R} = c \left\{ \frac{1}{n} \mathrm{tr} \left(S_{i-j} \tilde{\mathcal{R}}^{-1} \right) \right\}_{i,j=1}^{T}, \quad \tilde{\mathcal{R}} = \sum_{q=-\infty}^{\infty} \frac{1}{T} \mathrm{tr} \left(J^q (I_T + \mathcal{R})^{-1} \right) S_q$$

with $[J^q]_{ij} \equiv \boldsymbol{\delta}_{i+q,j}$ and $S_q \equiv \sum_{k \geq 0} W^{k+(-q)^+} (W^{k+q^+})^{\mathsf{T}}$.

Theorem ([Couillet,Wainrib'16] Training MSE for fixed W)

As $n, T \to \infty$, with $n/T \to c < 1$,

$$\mathrm{MSE} = \frac{1}{T} r^{\mathsf{T}} \left(I_T + \mathcal{R} + \frac{1}{\eta^2} U^{\mathsf{T}} \left\{ m^{\mathsf{T}} (W^i)^{\mathsf{T}} \tilde{\mathcal{R}}^{-1} W^j m \right\}_{i,j=0}^{T-1} U \right)^{-1} r + o(1)$$

where $U_{ij} = u_{i-j}$ and $\mathcal{R}$, $\tilde{\mathcal{R}}$ are solutions to

$$\mathcal{R} = c \left\{ \frac{1}{n} \mathrm{tr} \left(S_{i-j} \tilde{\mathcal{R}}^{-1} \right) \right\}_{i,j=1}^{T}, \quad \tilde{\mathcal{R}} = \sum_{q=-\infty}^{\infty} \frac{1}{T} \mathrm{tr} \left(J^q (I_T + \mathcal{R})^{-1} \right) S_q$$

with $[J^q]_{ij} \equiv \boldsymbol{\delta}_{i+q,j}$ and $S_q \equiv \sum_{k \geq 0} W^{k+(-q)^+} (W^{k+q^+})^{\mathsf{T}}$.

Corollaries :

- for $c = 0$ ($S_0 = \sum_{k \geq 0} W^k (W^k)^{\mathsf{T}}$),

$$\mathrm{MSE} = \frac{1}{T} r^{\mathsf{T}} \left(I_T + \frac{1}{\eta^2} U^{\mathsf{T}} \left\{ m^{\mathsf{T}} (W^i)^{\mathsf{T}} S_0^{-1} W^j m \right\}_{i,j=0}^{T-1} U \right)^{-1} r + o(1)$$

Theorem ([Couillet,Wainrib'16] Training MSE for fixed W)

As $n, T \to \infty$, with $n/T \to c < 1$,

$$\mathrm{MSE} = \frac{1}{T} r^\mathsf{T} \left(I_T + \mathcal{R} + \frac{1}{\eta^2} U^\mathsf{T} \left\{ m^\mathsf{T} (W^i)^\mathsf{T} \tilde{\mathcal{R}}^{-1} W^j m \right\}_{i,j=0}^{T-1} U \right)^{-1} r + o(1)$$

where $U_{ij} = u_{i-j}$ and $\mathcal{R}$, $\tilde{\mathcal{R}}$ are solutions to

$$\mathcal{R} = c \left\{ \frac{1}{n} \mathrm{tr} \left(S_{i-j} \tilde{\mathcal{R}}^{-1} \right) \right\}_{i,j=1}^{T}, \quad \tilde{\mathcal{R}} = \sum_{q=-\infty}^{\infty} \frac{1}{T} \mathrm{tr} \left(J^q (I_T + \mathcal{R})^{-1} \right) S_q$$

with $[J^q]_{ij} \equiv \boldsymbol{\delta}_{i+q,j}$ and $S_q \equiv \sum_{k \geq 0} W^{k+(-q)^+} (W^{k+q^+})^\mathsf{T}$.

Corollaries :

- for $c = 0$ ($S_0 = \sum_{k \geq 0} W^k (W^k)^\mathsf{T}$),

$$\mathrm{MSE} = \frac{1}{T} r^\mathsf{T} \left(I_T + \frac{1}{\eta^2} U^\mathsf{T} \left\{ m^\mathsf{T} (W^i)^\mathsf{T} S_0^{-1} W^j m \right\}_{i,j=0}^{T-1} U \right)^{-1} r + o(1)$$

- for $W = \sigma Z$ with Z Haar, $\|m\| = 1$ independent of W,

$$\mathrm{MSE} = (1-c) \frac{1}{T} r^\mathsf{T} \left(I_T + \frac{1}{\eta^2} U^\mathsf{T} \mathrm{diag} \left\{ (1-\sigma^2) \sigma^{2(i-1)} \right\}_{i=1}^{T} U \right)^{-1} r + o(1).$$

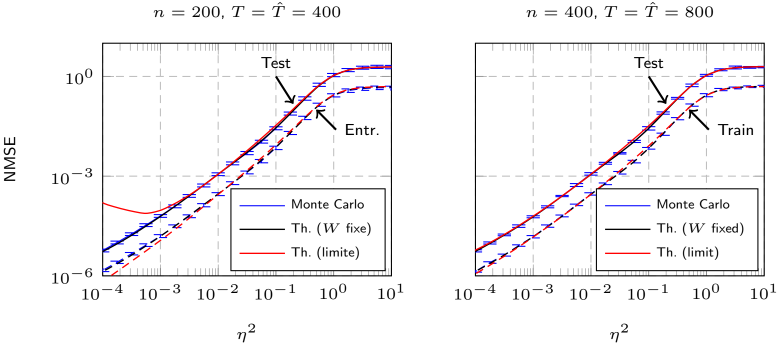


FIGURE – Prediction for the Mackey Glass model, $W = \sigma Z$, $\sigma = .9$, Z Haar.

Consequences : Analysis suggests choice $W = \mathrm{diag}(W_1, \ldots, W_k)$, $W_j = \sigma_j Z_j$, $Z_j \in \mathbb{R}^{n_j \times n_j}$ Haar, leading to change

$$(1-\sigma^2)\sigma^{2\tau} \leftrightarrow \mathrm{MC}(\tau) \equiv \frac{\sum_{j=1}^k c_j \sigma_j^{2\tau}}{\sum_{j=1}^k c_j (1-\sigma_j^2)^{-1}}.$$

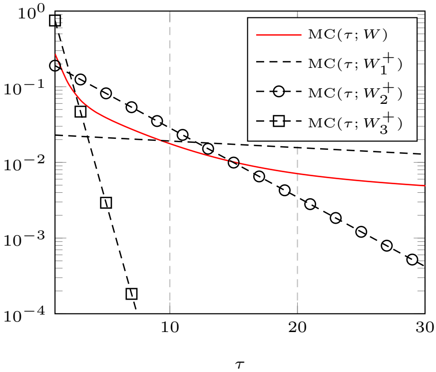


FIGURE – Memory curve (MC) for $W = \mathrm{diag}(W_1, W_2, W_3)$, $W_j = \sigma_j Z_j$, $Z_j \in \mathbb{R}^{n_j \times n_j}$ Haar, $\sigma_1 = .99$, $n_1/n = .01$, $\sigma_2 = .9$, $n_2/n = .1$, and $\sigma_3 = .5$, $n_3/n = .89$. Matrices W_i^+ defined by $W_i^+ = \sigma_i Z_i^+$, with $Z_i^+ \in \mathbb{R}^{n \times n}$ Haar.

Outline

Curriculum Vitae

Research Project : Learning in Large Dimensions
Axis 1 : Robust Estimation in Large Dimensions
Axis 2 : Classification in Large Dimensions
Axis 3 : Random Matrices and Neural Networks
Axis 4 : Graphs

Baseline Scenario : Analysis of inference methods for large graphs

Axis 4 : Graphs

Baseline Scenario : Analysis of inference methods for large graphs

- community detection on realistic graphs (spectral methods)

Axis 4 : Graphs

Baseline Scenario : Analysis of inference methods for large graphs

- community detection on realistic graphs (spectral methods)
- analysis of signal processing on graphs methods

Axis 4 : Graphs

Baseline Scenario : Analysis of inference methods for large graphs

- community detection on realistic graphs (spectral methods)
- analysis of signal processing on graphs methods

Tools :

Axis 4 : Graphs

Baseline Scenario : Analysis of inference methods for large graphs

- community detection on realistic graphs (spectral methods)
- analysis of signal processing on graphs methods

Tools :

- methods based on adjacency A, Laplacian L, or modularity M

$$L = D - A$$
$$M = A - E[A].$$

e.g., if $A_{ij} \sim \mathrm{Bern}(q_i q_j)$, $M = A - qq^{\mathsf{T}}$.

Baseline Scenario : Analysis of inference methods for large graphs

- community detection on realistic graphs (spectral methods)
- analysis of signal processing on graphs methods

Tools :

- methods based on adjacency A, Laplacian L, or modularity M

$$L = D - A$$
$$M = A - E[A].$$

e.g., if $A_{ij} \sim \mathrm{Bern}(q_i q_j)$, $M = A - qq^{\mathsf{T}}$.

- spectrum and eigenvectors of A, L fundamental to inference methods

Axis 4 : Graphs

Baseline Scenario : Analysis of inference methods for large graphs

- community detection on realistic graphs (spectral methods)
- analysis of signal processing on graphs methods

Tools :

- methods based on adjacency A, Laplacian L, or modularity M

$$L = D - A$$
$$M = A - E[A].$$

e.g., if $A_{ij} \sim \mathrm{Bern}(q_i q_j)$, $M = A - qq^{\mathsf{T}}$.

- spectrum and eigenvectors of A, L fundamental to inference methods
- optimization, regression, PCA, etc., based on spectral properties.

Problems and Objectives

Problems and Objectives

- Community detection based on homogeneous graph methods

Axis 4 : Graphs

Problems and Objectives

- Community detection based on homogeneous graph methods
- Signal processing oh graphs purely deterministically studied

Axis 4 : Graphs

Problems and Objectives

- Community detection based on homogeneous graph methods
- Signal processing oh graphs purely deterministically studied
- We need here :
  - develop and analyze community detection algorithms for realistic graphs
  - analyze performances of signal processing on graphs methods

Axis 4 : Graphs

Problems and Objectives

- Community detection based on homogeneous graph methods
- Signal processing oh graphs purely deterministically studied
- We need here :
  - develop and analyze community detection algorithms for realistic graphs
  - analyze performances of signal processing on graphs methods

Results and Perspectives

Axis 4 : Graphs

Problems and Objectives

- Community detection based on homogeneous graph methods
- Signal processing oh graphs purely deterministically studied
- We need here :
  - develop and analyze community detection algorithms for realistic graphs
  - analyze performances of signal processing on graphs methods

Results and Perspectives

✔ New algorithms (and their analysis) for community detection with heterogeneous nodes

Axis 4 : Graphs

Problems and Objectives

- Community detection based on homogeneous graph methods
- Signal processing oh graphs purely deterministically studied
- We need here :
  - develop and analyze community detection algorithms for realistic graphs
  - analyze performances of signal processing on graphs methods

Results and Perspectives

✔ New algorithms (and their analysis) for community detection with heterogeneous nodes

✎ Exportation to signal processing on graphs problems.

Axis 4 : Graphs

Model : graph G with n nodes and k classes, with for $i \in \mathcal{C}_a$, $j \in \mathcal{C}_b$,

$$A_{ij} \sim \mathrm{Bern}\left(q_i q_j C_{ab}\right)$$

where $C_{ab} = 1 + n^{-\frac{1}{2}} \Gamma_{ab}$, $\Gamma_{ab} = O(1)$.

Model : graph G with n nodes and k classes, with for $i \in \mathcal{C}_a$, $j \in \mathcal{C}_b$,

$$A_{ij} \sim \mathrm{Bern}\left(q_i q_j C_{ab}\right)$$

where $C_{ab} = 1 + n^{-\frac{1}{2}} \Gamma_{ab}$, $\Gamma_{ab} = O(1)$.

Limitations of classical approaches : Normalized modularity ($\hat{q} = \frac{1}{n} A 1_n$)

$$L = \frac{1}{\sqrt{n}} \mathrm{diag}(\hat{q})^{-1} \left[A - \frac{\hat{q}\hat{q}^\mathsf{T}}{\frac{1}{n} 1_n^\mathsf{T} \hat{q}} \right] \mathrm{diag}(\hat{q})^{-1}.$$

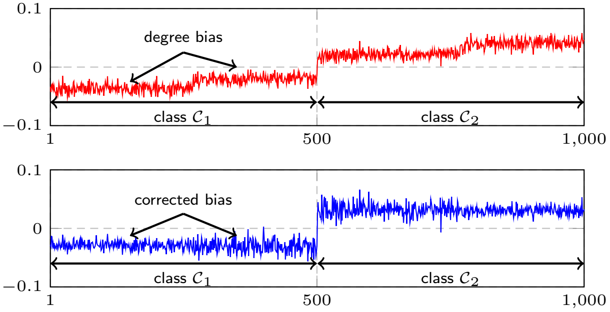


FIGURE – 2nd eigenvector of A (top) and 1st eigenvector of L (bottom) with bimodal q_i, 2 classes, $n = 1000$.

Theorem ([Tiomoko Ali,Couillet'16] Limiting Deterministic Equivalent)

As $n \to \infty$, $\|L - \tilde{L}\| \xrightarrow{\text{p.s.}} 0$ *with*

$$\tilde{L} = \frac{1}{m_q^2}\left[\frac{1}{\sqrt{n}} D^{-1} X D^{-1} + U \Lambda U^\mathsf{T}\right]$$

where $D = \operatorname{diag}(\{q_i\})$, $m_q = \lim_n \frac{1}{n}\sum_i q_i$ *and*

$$U = \begin{bmatrix} \frac{J}{\sqrt{n}} & \frac{1}{n m_q} D^{-1} X 1_n \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} (I_k - 1_k c^\mathsf{T})\Gamma(I_k - c 1_k^\mathsf{T}) & -1_k \\ 1_k & 0 \end{bmatrix}$$

$J = [j_1, \ldots, j_k]$, $j_a = [0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^n$ *canonical vector of class* $\mathcal{C}_a$*.*

Theorem ([Tiomoko Ali,Couillet'16] Limiting Deterministic Equivalent)
As $n \to \infty$, $\|L - \tilde{L}\| \xrightarrow{\text{p.s.}} 0$ *with*

$$\tilde{L} = \frac{1}{m_q^2}\left[\frac{1}{\sqrt{n}}D^{-1}XD^{-1} + U\Lambda U^{\mathsf{T}}\right]$$

where $D = \operatorname{diag}(\{q_i\})$, $m_q = \lim_n \frac{1}{n}\sum_i q_i$ *and*

$$U = \begin{bmatrix} \frac{J}{\sqrt{n}} & \frac{1}{nm_q}D^{-1}X1_n \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} (I_k - 1_k c^{\mathsf{T}})\Gamma(I_k - c1_k^{\mathsf{T}}) & -1_k \\ 1_k & 0 \end{bmatrix}$$

$J = [j_1, \ldots, j_k]$, $j_a = [0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0]^{\mathsf{T}} \in \mathbb{R}^n$ *canonical vector of class* $\mathcal{C}_a$.

Consequences :

- detection based on eigenvalues of Γ
- alignment of eigenvectors to j_a

Thank you.