Habilitation à diriger des Recherches

Robust Estimation Methods in the Large Random Matrix Regime

Romain COUILLET

CentraleSupélec
Université Paris-Sud 11

February 2, 2015

CentraleSupélec

Outline

Curriculum Vitae

Robust Estimation and Random Matrix Theory
- Robust estimates of scatter for elliptical and outlier data
- Robust shrinkage estimates of scatter
- Second-order statistics

Perspectives

Outline

Curriculum Vitae

Robust Estimation and Random Matrix Theory

- Robust estimates of scatter for elliptical and outlier data
- Robust shrinkage estimates of scatter
- Second-order statistics

Perspectives

Education and Professional Experience

Professional Experience

Assistant Professor Jan. 2011–Present
CentraleSupélec, Gif sur Yvette, France.
Telecom Department, Group LANEAS, Division Signals

Development Engineer and PhD student Sep. 2007–Dec. 2010
ST-Ericsson, Sophia Antipolis, France.

Education

Ph.D. in Physics (Telecommunications) Nov. 2010

Location	CentraleSupélec, Gif sur Yvette, France
Subject	Application of random matrix theory to future wireless flexible networks
Advisor	Mérouane Debbah

MSc. and Engineering Degree in Telecommunication Mar. 2008

Location	Telecom ParisTech, Paris, France
Grade	Very Good (Très Bien)
Topic	Mobile communications, embedded systems, computer science.

Teaching Activities and Research Projects

Teaching

ENS Cachan, Cachan, France. since 2013

Details Master 2, lectures, 18 hrs/year

CentraleSupélec, Gif sur Yvette, France. since 2011

Details PhD, lectures, 18 hrs/year
Master 2, seminar lectures, 24 hrs/year
Undergraduate, lectures + practical courses, 68 hrs/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

Tutorials in IEEE conferences	6
Workshops and special sessions	3
Editorship of journal special issues	1
Member of SPTM Technical Committee	2014
Associate Editor at IEEE TSP	2015

PhD Students

✔ **Axel MÜLLER (now Engineer at HUAWEI)** **2011–2014**

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

✔ **Julia VINOGRADOVA** **2011–2014**

Subject	Random matrices and applications to detection and estimation in array processing
Details	50%, with W. Hachem (Telecom ParisTech)
Publications	2 articles in IEEE-TSP, 2 IEEE conferences

✎ **Azary ABBOUD** **2012–2015**

Subject	Distributed optimization for smart grids
Details	33%, with M. Debbah and H. Siguerdidjane (CentraleSupélec)
Publications	1 article submitted, 1 IEEE conference

✎ **Gil KATZ** **2013–2016**

Subject	Interactive communication for distributed computing
Details	33%, with M. Debbah, P. Piantanida (CentraleSupélec)
Publications	1 IEEE conference

✎ **Evgeny KUSMENKO** **2015–2018**

Subject	Random matrix and machine learning
Details	80%, with M. Debbah (CentraleSupélec)

Research Activities

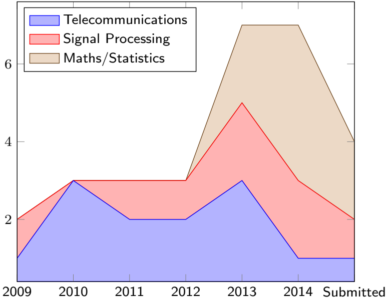
Publication Record (as of January 1st, 2015)

Publications	Book: 1, Book chapters: 3, Journals: 28, Conferences: 47, Patents: 4.
Citations	879 (five best: 187, 131, 72, 43, 24)
Indices	h-index: 15, i10-index: 23

Topics

Mathematics	random matrix theory (probability theory, complex analysis), statistics
Applications	signal processing (detection, estimation), wireless communications

Journal articles by field



Prizes and Awards

CNRS Bronze Medal (section INS2I)	2013
IEEE ComSoc Outstanding Young Researcher Award (EMEA Region)	2013
EEA/GdR ISIS/GRETSI 2011 Award of the Best 2010 Thesis	2011

Paper Awards

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

Outline

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

- **[Huber'67]** If $x_1 \sim (1-\varepsilon)\mathcal{N}(0, C_N) + \varepsilon G$, G unknown, robust estimator ($n > N$)

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} \max\left\{\ell_1, \frac{\ell_2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i}\right\} x_i x_i^* \text{ for some } \ell_1, \ell_2 > 0.$$

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

- **[Huber'67]** If $x_1 \sim (1-\varepsilon)\mathcal{N}(0, C_N) + \varepsilon G$, G unknown, robust estimator ($n > N$)

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} \max \left\{ \ell_1, \frac{\ell_2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i} \right\} x_i x_i^* \text{ for some } \ell_1, \ell_2 > 0.$$

- **[Maronna'76]** If x_1 elliptical (and $n > N$), ML estimator for C_N given by

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i \right) x_i x_i^* \text{ for some non-increasing } u.$$

Baseline scenario: $x_1,\ldots,x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1]=0$, $E[x_1x_1^*]=C_N$:

- If $x_1 \sim \mathcal{N}(0,C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

- **[Huber'67]** If $x_1 \sim (1-\varepsilon)\mathcal{N}(0,C_N)+\varepsilon G$, G unknown, robust estimator ($n>N$)

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^n \max\left\{\ell_1, \frac{\ell_2}{\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i}\right\} x_i x_i^* \text{ for some } \ell_1,\ell_2>0.$$

- **[Maronna'76]** If x_1 elliptical (and $n>N$), ML estimator for C_N given by

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^n u\left(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i\right) x_i x_i^* \text{ for some non-increasing } u.$$

- **[Pascal'13; Chen'11]** If $N>n$, x_1 elliptical or with outliers, shrinkage extensions

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

$$\check{C}_N(\rho) = \frac{\check{B}_N(\rho)}{\frac{1}{N}\operatorname{tr}\check{B}_N(\rho)},\quad \check{B}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N}x_i^*\check{C}_N^{-1}(\rho)x_i} + \rho I_N$$

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

We study such $\hat{C}_N$ in the regime

$$N, n \to \infty, \;\; N/n \to c \in (0, \infty).$$

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

We study such $\hat{C}_N$ in the regime

$$N, n \to \infty, \ N/n \to c \in (0, \infty).$$

- Math interest:
  - limiting eigenvalue distribution of $\hat{C}_N$
  - limiting values and fluctuations of functionals $f(\hat{C}_N)$

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

We study such $\hat{C}_N$ in the regime

$$N, n \to \infty, \ N/n \to c \in (0, \infty).$$

- Math interest:
  - limiting eigenvalue distribution of $\hat{C}_N$
  - limiting values and fluctuations of functionals $f(\hat{C}_N)$
- Application interest:
  - comparison between SCM and robust estimators
  - performance of robust/non-robust estimation methods
  - improvement thereof (by proper parametrization)

Outline of Theoretical Content

- First order convergence:

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

for some tractable random matrices $\hat{S}_N$.

- First order convergence:

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

for some tractable random matrices $\hat{S}_N$.

- Second order results:

$$N^{1-\varepsilon} \left(a^* \hat{C}_N^k b - a^* \hat{S}_N^k b \right) \xrightarrow{\text{a.s.}} 0$$

allowing transfer of CLT results.

Outline of Theoretical Content

- First order convergence:

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

for some tractable random matrices $\hat{S}_N$.

- Second order results:

$$N^{1-\varepsilon} \left(a^* \hat{C}_N^k b - a^* \hat{S}_N^k b \right) \xrightarrow{\text{a.s.}} 0$$

allowing transfer of CLT results.

- Applications:
  - improved robust covariance matrix estimation
  - improved robust tests / estimators
  - specific examples in statistics at large, array processing, statistical finance, etc.

Outline

Definition (Maronna's Estimator)

For $x_1, \ldots, x_n \in \mathbb{C}^N$ with $n > N$, $\hat{C}_N$ is the solution (upon existence and uniqueness) of

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

Definition (Maronna's Estimator)

For $x_1, \ldots, x_n \in \mathbb{C}^N$ with $n > N$, $\hat{C}_N$ is the solution (upon existence and uniqueness) of

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

where $u : [0, \infty) \to (0, \infty)$ is

- non-increasing
- such that $\phi(x) \triangleq x u(x)$ increasing of supremum ϕ_∞ with

$$1 < \phi_\infty < c^{-1}, \; c \in (0,1).$$

Definition (Maronna's Estimator)

For $x_1, \ldots, x_n \in \mathbb{C}^N$ with $n > N$, $\hat{C}_N$ is the solution (upon existence and uniqueness) of

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

where $u : [0, \infty) \to (0, \infty)$ is

- non-increasing
- such that $\phi(x) \triangleq x u(x)$ increasing of supremum ϕ_∞ with

$$1 < \phi_\infty < c^{-1}, \; c \in (0, 1).$$

Remark (Correlation Invariance)

For some $C_N \succ 0$, calling $\tilde{x}_i \triangleq C_N^{-\frac{1}{2}} x_i$, $\tilde{C}_N \triangleq C_N^{-\frac{1}{2}} \hat{C}_N C_N^{-\frac{1}{2}}$,

$$\tilde{C}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N} \tilde{x}_i^* \tilde{C}_N^{-1} \tilde{x}_i\right) \tilde{x}_i \tilde{x}_i^*$$

If $E[x_i x_i^*] = C_N$, sufficient to assume $E[x_i x_i^*] = I_N$.

Assumption ("Elliptical" Data)

$x_1, \ldots, x_n$ independent,

$$x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$$

- $w_i \in \mathbb{C}^N$ isotropic, $\|w_i\|^2 = N$
- $C_N \succ 0$, $\limsup_N \|C_N\| < \infty$
- $\tau_i > 0$ deterministic (or random independent of w_i)

Assumption ("Elliptical" Data)

$x_1, \ldots, x_n$ independent,

$$x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$$

- $w_i \in \mathbb{C}^N$ isotropic, $\|w_i\|^2 = N$
- $C_N \succ 0$, $\limsup_N \|C_N\| < \infty$
- $\tau_i > 0$ deterministic (or random independent of w_i)
- for $\tilde{\nu}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\tau_i}$ and some $m > 0$,

$$\tilde{\nu}_n([0, m)) < 1 - \phi_\infty^{-1} \text{ for all large } n \text{ (a.s.)}$$

- $\int \tau \tilde{\nu}_n(d\tau) \to 1$ (a.s.).

Model Description

Assumption ("Elliptical" Data)

$x_1, \ldots, x_n$ independent,

$$x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$$

- $w_i \in \mathbb{C}^N$ isotropic, $\|w_i\|^2 = N$
- $C_N \succ 0$, $\limsup_N \|C_N\| < \infty$
- $\tau_i > 0$ deterministic (or random independent of w_i)
- for $\tilde{\nu}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\tau_i}$ and some $m > 0$,

$$\tilde{\nu}_n([0, m)) < 1 - \phi_\infty^{-1} \text{ for all large } n \text{ (a.s.)}$$

- $\int \tau \tilde{\nu}_n(d\tau) \to 1$ (a.s.).

Fact (Existence and Uniqueness)

By **[Kent&Tyler'91]**, for each $n > N$, $\hat{C}_N$ is a.s. well-defined.

Assumption (Tail Control)

For each $a > b > 0$,

$$\frac{\limsup_n \tilde{\nu}_n([t, \infty))}{\phi(at) - \phi(bt)} \to 0$$

as $t \to \infty$.

Assumption (Tail Control)

For each $a > b > 0$,

$$\frac{\limsup_n \tilde{\nu}_n([t,\infty))}{\phi(at) - \phi(bt)} \to 0$$

as $t \to \infty$.

Example: If $u(x) = \frac{\alpha+1}{\alpha+x}$, τ_i i.i.d., sufficient to have $E[\tau_1^{1+\varepsilon}] < \infty$.

Assumption (Random Matrix Regime)

As $n \to \infty$,

$$c_N \triangleq \frac{N}{n} \to c \in (0,1).$$

Definition (v and ψ)

Letting $g(x) = x(1 - c\phi(x))^{-1}$ (on $\mathbb{R}_+$),

$$v(x) \triangleq (u \circ g^{-1})(x) \quad \text{non-increasing}$$
$$\psi(x) \triangleq xv(x) \quad \text{increasing and bounded by } \psi_\infty.$$

Definition (v and ψ)

Letting $g(x) = x(1 - c\phi(x))^{-1}$ (on $\mathbb{R}_+$),

$$\begin{aligned} v(x) &\triangleq (u \circ g^{-1})(x) && \text{non-increasing} \\ \psi(x) &\triangleq xv(x) && \text{increasing and bounded by } \psi_\infty. \end{aligned}$$

Lemma (Rewriting $\hat{C}_N$)

It holds (with $C_N = I_N$) that

$$\hat{C}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} \tau_i v\left(\tau_i d_i\right) w_i w_i^*$$

with $(d_1, \ldots, d_n) \in \mathbb{R}_+^n$ a.s. unique solution to

$$d_i = \frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i, \ i = 1, \ldots, n.$$

Remark (Quadratic Form close to Trace)

Random matrix insight: $(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j)w_j w_j^*)^{-1}$ “almost independent” of w_i, so

Remark (Quadratic Form close to Trace)

Random matrix insight: $(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*)^{-1}$ "almost independent" of w_i, so

$$d_i = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j\neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i \simeq \frac{1}{N} \text{tr} \left(\frac{1}{n} \sum_{j\neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} \simeq \gamma_N$$

for some deterministic sequence $(\gamma_N)_{N=1}^{\infty}$, irrespective of i.

Remark (Quadratic Form close to Trace)

Random matrix insight: $(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*)^{-1}$ "almost independent" of w_i, so

$$d_i = \frac{1}{N} w_i^* \left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*\right)^{-1} w_i \simeq \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*\right)^{-1} \simeq \gamma_N$$

for some deterministic sequence $(\gamma_N)_{N=1}^{\infty}$, irrespective of i.

Lemma (Key Lemma)

Letting $e_i \triangleq \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$ with γ_N unique solution to

$$1 = \int \frac{\psi(\tau\gamma_N)}{1 + c\psi(\tau\gamma_N)} \tilde{\nu}_n(d\tau),$$

we have

$$\max_{1\leq i\leq n} |e_i - 1| \xrightarrow{\text{a.s.}} 0.$$

(Proof in a few slides.)

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

$$\left[\textit{equivalently,} \quad \hat{S}_N = \gamma_N^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi(\tau_i \gamma_N) C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} \right]$$

Corollaries

- **Spectral measure**: $\mu_N^{\hat{C}_N} - \mu_N^{\hat{S}_N} \xrightarrow{\mathcal{L}} 0$ a.s. ($\mu_N^X \triangleq \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\delta}_{\lambda_i(X)}$)

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

$$\left[\textit{equivalently,} \quad \hat{S}_N = \gamma_N^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi(\tau_i \gamma_N) C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} \right]$$

Corollaries

- **Spectral measure**: $\mu_N^{\hat{C}_N} - \mu_N^{\hat{S}_N} \xrightarrow{\mathcal{L}} 0$ a.s. ($\mu_N^X \triangleq \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\lambda_i(X)}$)
- **Local convergence**: $\max_{1 \leq i \leq N} |\lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N)| \xrightarrow{\text{a.s.}} 0.$

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} v(\tau_i\gamma_N)x_i x_i^*.$$

$$\left[\textit{equivalently,}\quad \hat{S}_N = \gamma_N^{-1}\frac{1}{n}\sum_{i=1}^{n}\psi(\tau_i\gamma_N)C_N^{\frac{1}{2}}w_i w_i^* C_N^{\frac{1}{2}}\right]$$

Corollaries

- **Spectral measure**: $\mu_N^{\hat{C}_N} - \mu_N^{\hat{S}_N} \xrightarrow{\mathcal{L}} 0$ a.s. ($\mu_N^X \triangleq \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}_{\lambda_i(X)}$)
- **Local convergence**: $\max_{1\leq i\leq N}|\lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N)| \xrightarrow{\text{a.s.}} 0.$
- **Norm boundedness**: $\limsup_N \|\hat{C}_N\| < \infty$

$\longrightarrow$ Bounded spectrum (unlike SCM!)

Large dimensional behavior

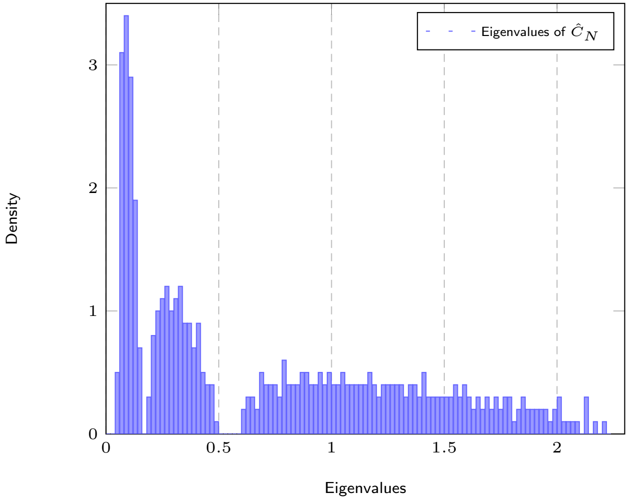


Figure: $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Large dimensional behavior

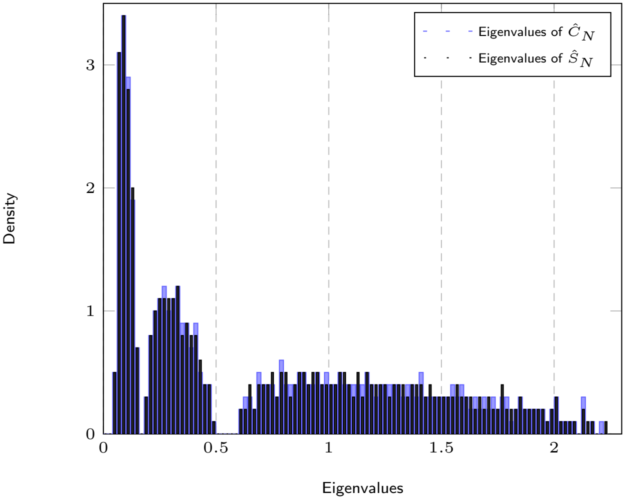


Figure: $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

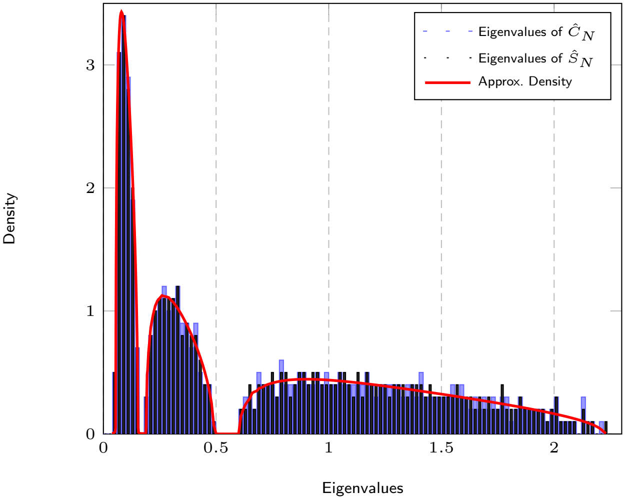


Figure: $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

Property (Quadratic form and γ_N)

$$\max_{1\leq i\leq n}\left|\frac{1}{N}w_i^*\left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j\gamma_N)w_j w_j^*\right)^{-1}w_i-\gamma_N\right|\xrightarrow{\text{a.s.}}0.$$

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

Property (Quadratic form and γ_N)

$$\max_{1\leq i\leq n} \left| \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j\neq i} \tau_j v(\tau_j \gamma_N) w_j w_j^* \right)^{-1} w_i - \gamma_N \right| \xrightarrow{\text{a.s.}} 0.$$

Proof of the Property

- Uniformity easy (moments of all orders for $[w_i]_j$).
- By a "quadratic form similar to trace" approach, we get

$$\max_{1\leq i\leq n} \left| \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j\neq i} \tau_j v(\tau_j \gamma_N) w_j w_j^* \right)^{-1} w_i - m(0) \right| \xrightarrow{\text{a.s.}} 0$$

  with $m(0)$ unique positive solution to **[MarPas'67; BaiSil'95]**

$$m(0) = \int \frac{\tau v(\tau \gamma_N)}{1 + c\tau v(\tau\gamma_N) m(0)} \tilde{\nu}_n(d\tau).$$

- γ_N precisely solves this equation, thus $m(0) = \gamma_N$.

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

Substitution Trick (case $\tau_i \in [a,b] \subset (0,\infty)$)

Up to relabelling $e_1 \leq \ldots \leq e_n$, use

$$
\begin{aligned}
v(\tau_n \gamma_N) e_n = v(\tau_n d_n) &= v\left(\tau_n \frac{1}{N} w_n^* \left(\frac{1}{n} \sum_{i<n} \tau_i \underbrace{v(\tau_i d_i)}_{= v(\tau_i \gamma_N) e_i} w_i w_i^* \right)^{-1} w_n \right) \\
&\leq v\left(\tau_n e_n^{-1} \frac{1}{N} w_n^* \left(\frac{1}{n} \sum_{i<n} \tau_i v(\tau_i \gamma_N) w_i w_i^* \right)^{-1} w_n \right) \\
&\leq v\left(\tau_n e_n^{-1} (\gamma_N - \varepsilon_n) \right) \text{ a.s., } \varepsilon_n \to 0 \text{ (slow).}
\end{aligned}
$$

Substitution Trick (case $\tau_i \in [a,b] \subset (0,\infty)$)

Up to relabelling $e_1 \leq \ldots \leq e_n$, use

$$v(\tau_n\gamma_N)e_n = v(\tau_n d_n) = v\left(\tau_n \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i \underbrace{v(\tau_i d_i)}_{=v(\tau_i\gamma_N)e_i} w_i w_i^*\right)^{-1} w_n\right)$$

$$\leq v\left(\tau_n e_n^{-1} \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i v(\tau_i\gamma_N) w_i w_i^*\right)^{-1} w_n\right)$$

$$\leq v\left(\tau_n e_n^{-1}(\gamma_N - \varepsilon_n)\right) \text{ a.s., } \varepsilon_n \to 0 \text{ (slow).}$$

Use properties of ψ to get

$$\psi\left(\tau_n\gamma_N\right) \leq \psi\left(\tau_n e_n^{-1}\gamma_N\right)\left(1 - \varepsilon_n \gamma_N^{-1}\right)^{-1}$$

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

Substitution Trick (case $\tau_i \in [a,b] \subset (0,\infty)$)

Up to relabelling $e_1 \leq \ldots \leq e_n$, use

$$v(\tau_n \gamma_N) e_n = v(\tau_n d_n) = v\left(\tau_n \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i \underbrace{v(\tau_i d_i)}_{=v(\tau_i \gamma_N) e_i} w_i w_i^*\right)^{-1} w_n\right)$$

$$\leq v\left(\tau_n e_n^{-1} \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i v(\tau_i \gamma_N) w_i w_i^*\right)^{-1} w_n\right)$$

$$\leq v\left(\tau_n e_n^{-1} (\gamma_N - \varepsilon_n)\right) \text{ a.s., } \varepsilon_n \to 0 \text{ (slow).}$$

Use properties of ψ to get

$$\psi(\tau_n \gamma_N) \leq \psi\left(\tau_n e_n^{-1} \gamma_N\right)\left(1 - \varepsilon_n \gamma_N^{-1}\right)^{-1}$$

Conclusion: If $e_n > 1 + \ell$ i.o., as $\tau_n \in [a,b]$, on subsequence $\begin{cases} \tau_n \to \tau_0 > 0 \\ \gamma_N \to \gamma_0 > 0 \end{cases}$,

$$\psi(\tau_0 \gamma_0) \leq \psi\left(\frac{\tau_0 \gamma_0}{1+\ell}\right), \text{ a contradiction.}$$

General τ_i case

- Control of

$$\Delta_M = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i - \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{\substack{j \neq i \\ \tau_j \leq M}} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i.$$

- Rationale: Large M bring small Δ_M but (possibly) large τ_n
  $\longrightarrow$ Relative control between tail of $\tilde{\nu}_n$ and flattening of ψ.

General τ_i case

- Control of

$$\Delta_M = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i - \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{\substack{j \neq i \\ \tau_j \leq M}} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i.$$

- Rationale: Large M bring small Δ_M but (possibly) large τ_n
  $\longrightarrow$ Relative control between tail of $\tilde{\nu}_n$ and flattening of ψ.

This concludes the proof.

Assumption (Signal Model)

$x_1, \ldots, x_n$ independent,

$$x_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i$$

- $w_i \in \mathbb{C}^N$, τ_i as previously, (for simplicity) $\tilde{\nu}_n \to \tilde{\nu}$
- $s_{li} \in \mathbb{C}$ i.i.d., mean 0, variance 1
- $p_1 \geq \ldots \geq p_L \geq 0$
- $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic with $\sum_{l=1}^{L} p_l a_l a_l^* \to \mathrm{diag}(p_i)_{i=1}^{L}$.

Spiked Model Extension

Assumption (Signal Model)

$x_1, \ldots, x_n$ independent,

$$x_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i$$

- $w_i \in \mathbb{C}^N$, τ_i as previously, (for simplicity) $\tilde{\nu}_n \to \tilde{\nu}$
- $s_{li} \in \mathbb{C}$ i.i.d., mean 0, variance 1
- $p_1 \geq \ldots \geq p_L \geq 0$
- $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic with $\sum_{l=1}^{L} p_l a_l a_l^* \to \operatorname{diag}(p_i)_{i=1}^{L}$.

Theorem (Extension of pure-noise model [C'2014])

As $n \to \infty$, under previous assumptions,

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

where

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma) x_i x_i^*.$$

(same result but different model, $\gamma = \lim_N \gamma_N$)

Spiked Model Extension

$\longrightarrow \hat{S}_N$ follows a spiked random matrix model.

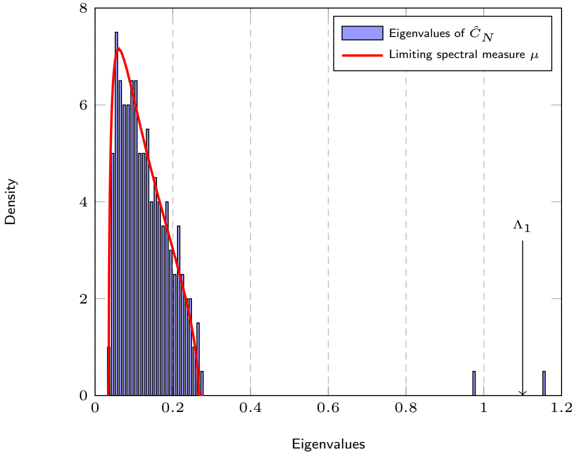


Figure: Eigenvalues of $\hat{C}_N$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

⟶ But eigenvalues allowed to wander away from limiting support.

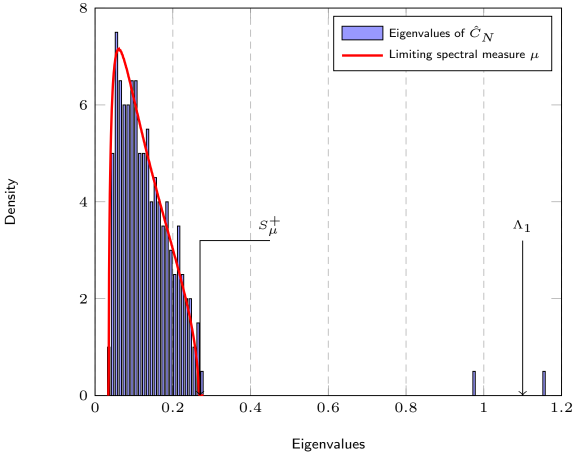


Figure: Eigenvalues of $\hat{C}_N$, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Spiked Model Extension

$\longrightarrow$ Noise eigenvalues are bounded by some S^+.

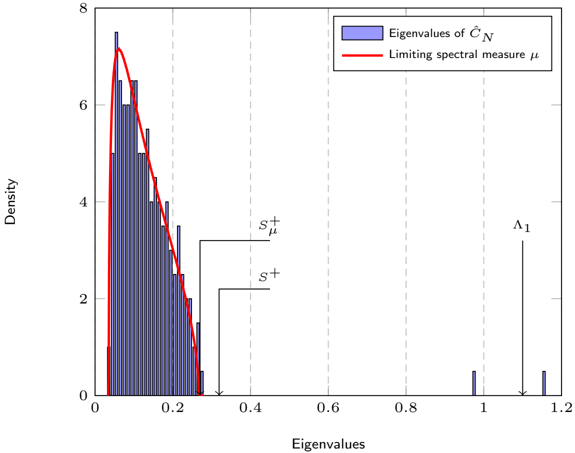


Figure: Eigenvalues of $\hat{C}_N$, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

$\longrightarrow$ To be compared versus SCM $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$

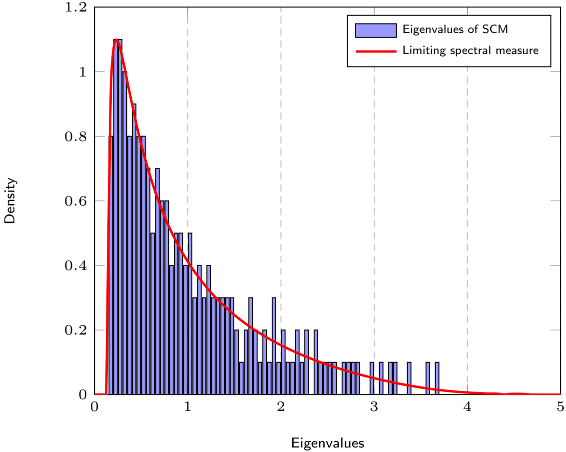


Figure: Eigenvalues of $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Sudent-t impulsions.

Spiked Model Extension

Some important remarks:

- If $p_1 = \ldots = p_L = 0$, noise-only model and

$$\limsup_N \|\hat{C}_N\| = \limsup_N \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{\gamma} w_i w_i^* \right\| \leq S^+ \triangleq \frac{\phi_\infty (1+\sqrt{c})^2}{(1-c\phi_\infty)\gamma}.$$

Some important remarks:

- If $p_1 = \ldots = p_L = 0$, noise-only model and

$$\limsup_N \|\hat{C}_N\| = \limsup_N \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{\gamma} w_i w_i^* \right\| \leq S^+ \triangleq \frac{\phi_\infty (1+\sqrt{c})^2}{(1-c\phi_\infty)\gamma}.$$

- If $p_1 \geq \ldots \geq p_L > 0$, informative spikes if $\det(\hat{S}_N - xI_N)$ has solutions beyond S^+ (and not S_μ^+!), i.e., if

$$p_l > p_- \triangleq \lim_{x \downarrow S^+} -c \left(\int \frac{\delta(x) v(t\gamma)}{1+\delta(x) t v(t\gamma)} \tilde{\nu}(dt) \right)^{-1}$$

with $\delta(x)$, $x > S_\mu^+$, unique solution to

$$\delta(x) = c \left(-x + \int \frac{t v(t\gamma)}{1+\delta(x) t v(t\gamma)} \tilde{\nu}(dt) \right)^{-1}.$$

Theorem (Spiked estimation, known $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$ *(*$\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$*),*

Extreme eigenvalues. *For each* j *with* $p_j > p_-$*,*

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+ \text{ a.s., where } -c\left(\int \frac{\delta(\Lambda_j)v(\tau\gamma)}{1+\delta(\Lambda_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} = p_j.$$

Theorem (Spiked estimation, known $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$ *(*$\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$*),*

Extreme eigenvalues. *For each* j *with* $p_j > p_-$*,*

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+ \text{ a.s., where } -c\left(\int \frac{\delta(\Lambda_j)v(\tau\gamma)}{1+\delta(\Lambda_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} = p_j.$$

Power estimation. *For each* j *with* $p_j > p_-$*,*

$$-c\left(\int \frac{\delta(\hat{\lambda}_j)v(\tau\gamma)}{1+\delta(\hat{\lambda}_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

Theorem (Spiked estimation, known $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$ $(\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N)$*,*

Extreme eigenvalues. *For each* j *with* $p_j > p_-$*,*

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+ \text{ a.s., where } -c\left(\int \frac{\delta(\Lambda_j)v(\tau\gamma)}{1+\delta(\Lambda_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} = p_j.$$

Power estimation. *For each* j *with* $p_j > p_-$*,*

$$-c\left(\int \frac{\delta(\hat{\lambda}_j)v(\tau\gamma)}{1+\delta(\hat{\lambda}_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

Bilinear form estimation. *For* $a, b \in \mathbb{C}^N$, $\|a\| = \|b\| = 1$*, and* j *with* $p_j > p_-$*,*

$$\sum_{k, p_k = p_j} a^* u_k u_k^* b - \sum_{k, p_k = p_j} w_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{a.s.}} 0$$

where

$$w_k \triangleq \int \frac{v(t\gamma)\tilde{\nu}(dt)}{\left(1+\delta(\hat{\lambda}_k)tv(t\gamma)\right)^2}\left[\int \frac{v(t\gamma)\tilde{\nu}(dt)}{1+\delta(\hat{\lambda}_k)tv(t\gamma)}\left(1 - \frac{1}{c}\int \frac{\delta(\hat{\lambda}_k)^2 t^2 v(t\gamma)^2 \tilde{\nu}(dt)}{\left(1+\delta(\hat{\lambda}_k)tv(t\gamma)\right)^2}\right)\right]^{-1}$$

Theorem (Spiked estimation, unknown $\tilde{\nu}$ [C'2014])
With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$,

Empirical estimates.

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0, \;\; \hat{\gamma}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i$$

$$\max_{\tau_j < M} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0, \;\; \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i.$$

Theorem (Spiked estimation, unknown $\tilde{\nu}$ [C'2014])
With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$,

Empirical estimates.

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0,\ \hat{\gamma}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i$$

$$\max_{\tau_j < M} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0,\ \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i.$$

Power estimation. *For each* j *with* $p_j > p_-$,

$$-\left(\hat{\delta}(\hat{\lambda}_j) \frac{1}{N} \sum_{i=1}^{n} \frac{v(\hat{\tau}_i \hat{\gamma}_n)}{1 + \hat{\delta}(\hat{\lambda}_j)\hat{\tau}_i v(\hat{\tau}_i \hat{\gamma}_n)} \right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

Theorem (Spiked estimation, unknown $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$,

Empirical estimates.

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0, \;\; \hat{\gamma}_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i$$

$$\max_{\tau_j < M} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0, \;\; \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i.$$

Power estimation. *For each* j *with* $p_j > p_-$,

$$-\left(\hat{\delta}(\hat{\lambda}_j) \frac{1}{N} \sum_{i=1}^{n} \frac{v(\hat{\tau}_i \hat{\gamma}_n)}{1 + \hat{\delta}(\hat{\lambda}_j) \hat{\tau}_i v(\hat{\tau}_i \hat{\gamma}_n)} \right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

Bilinear form estimation. *For* $a, b \in \mathbb{C}^N$, $\|a\| = \|b\| = 1$, *and* j *with* $p_j > p_-$,

$$\sum_{k, p_k = p_j} a^* u_k u_k^* b - \sum_{k, p_k = p_j} \hat{w}_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{a.s.}} 0$$

for the corresponding $\hat{w}_k = f(\{\hat{\tau}_i\}, \hat{\delta}(\hat{\lambda}_k))$.

Spiked Model Extension

$\longrightarrow$ Application to angle estimation with

$$a_l = a(\theta_l),\ \theta_l \in [0, 2\pi)$$

Spiked Model Extension

$\longrightarrow$ Application to angle estimation with

$$a_l = a(\theta_l), \;\; \theta_l \in [0, 2\pi)$$

Corollary (Robust G-MUSIC)

Define $\hat{\eta}_{\rm RG}(\theta)$ *and* $\hat{\eta}_{\rm RG}^{\rm emp}(\theta)$ *as*

$$\hat{\eta}_{\rm RG}(\theta) = 1 - \sum_{k=1}^{|\{p_j > p_-\}|} w_k a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta)$$

$$\hat{\eta}_{\rm RG}^{\rm emp}(\theta) = 1 - \sum_{k=1}^{|\{p_j > p_-\}|} \hat{w}_k a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta)$$

Then, for each j *with* $p_j > p_-$*,*

$$\hat{\theta}_j \xrightarrow{\rm a.s.} \theta_j$$

$$\hat{\theta}_j^{\rm emp} \xrightarrow{\rm a.s.} \theta_j$$

where

$$\hat{\theta}_j \triangleq \operatorname{argmin}_{\theta \in V(\theta_j)} \left\{ \hat{\eta}_{\rm RG}(\theta) \right\}$$

$$\hat{\theta}_j^{\rm emp} \triangleq \operatorname{argmin}_{\theta \in V(\theta_j)} \left\{ \hat{\eta}_{\rm RG}^{\rm emp}(\theta) \right\}.$$

Spiked Model Extension

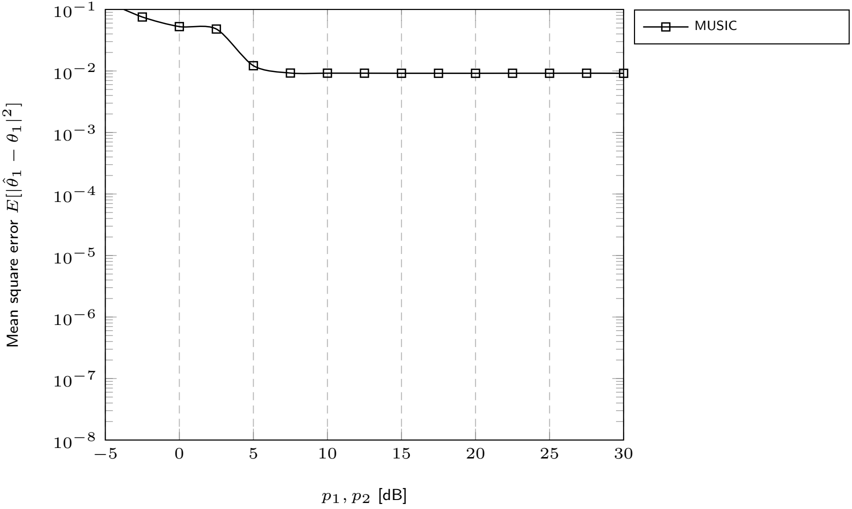


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

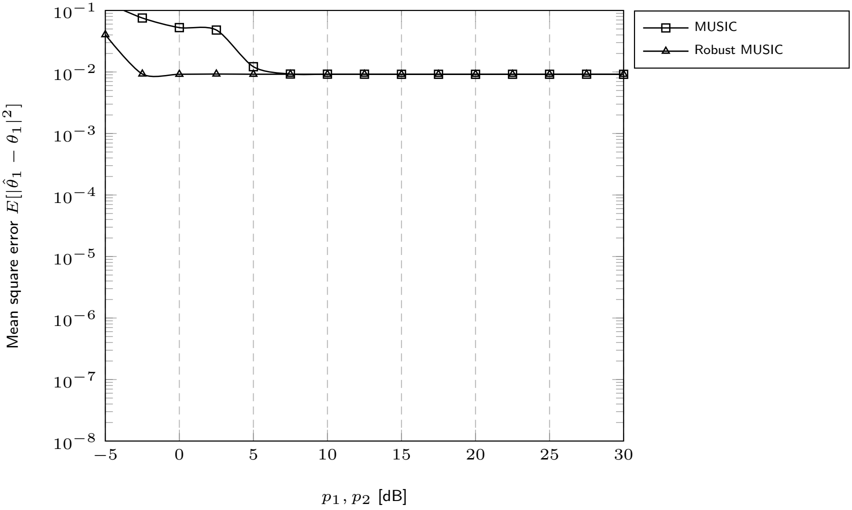


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $p_1 = p_2$.

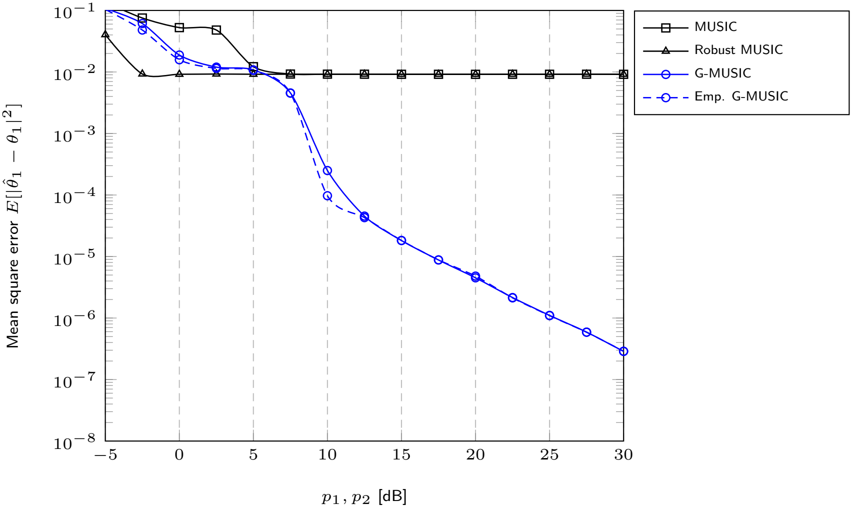


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

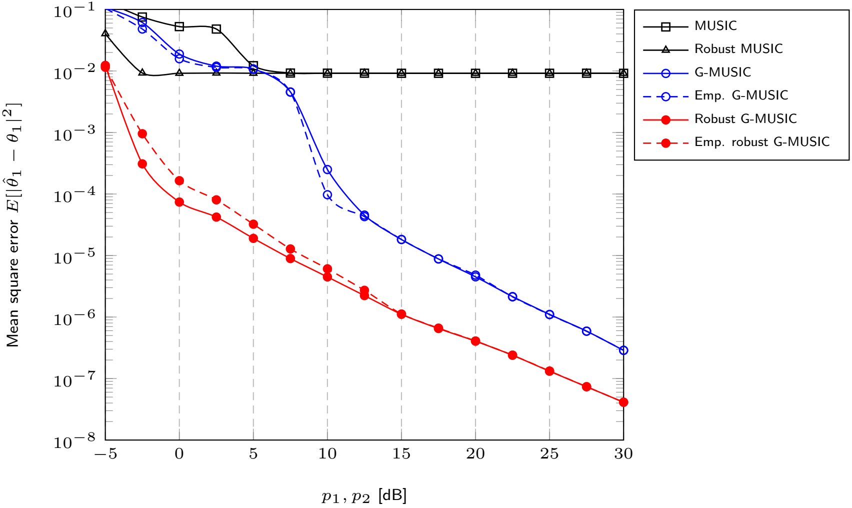


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Spiked Model Extension

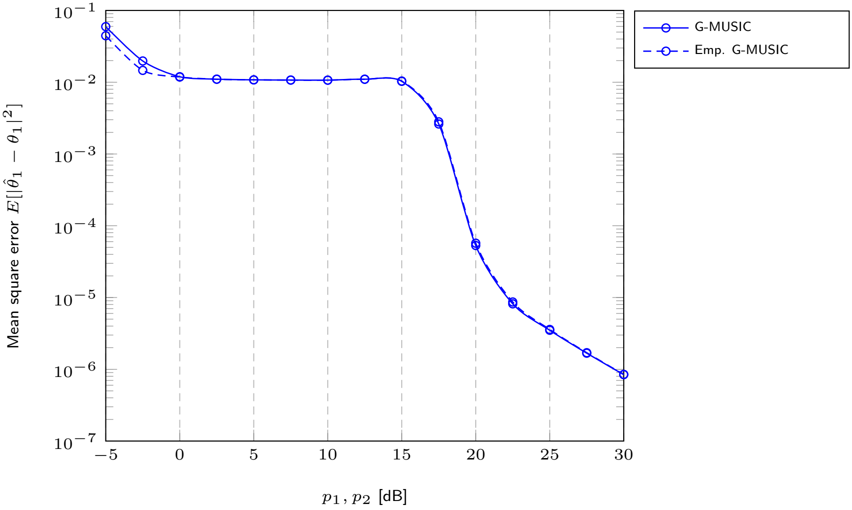


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

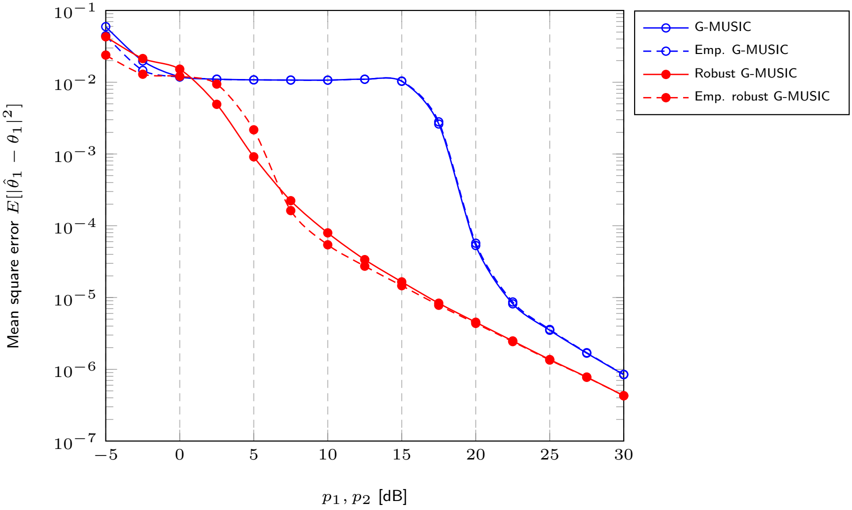


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Original setting of Huber

Assumption (Outlying Data)

Observation set

$$X = \left[x_1, \ldots, x_{(1-\varepsilon_n)n}, a_1, \ldots, a_{\varepsilon_n n}\right]$$

where $x_i \sim \mathcal{CN}(0, C_N)$ and $a_1, \ldots, a_{\varepsilon_n n} \in \mathbb{C}^N$ deterministic with

$$\max_i \limsup_n \frac{\|a_i\|}{\sqrt{N}} < \infty$$

(or only a.s. if a_i random).

Theorem (Outlier Rejection [Morales-Jimenez,C,McKay'14])

As $n \to \infty$,

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

where

$$\hat{S}_N \triangleq v(\gamma_N) \frac{1}{n} \sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n}) a_i a_i^*$$

with γ_N and $\alpha_{1,n}, \ldots, \alpha_{\varepsilon_n n,n}$ unique positive solutions to

$$\gamma_N = \frac{1}{N} \mathit{tr} C_N \left(\frac{(1-\varepsilon)v(\gamma_N)}{1+cv(\gamma_N)\gamma_N} C_N + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n}) a_i a_i^* \right)^{-1}$$

$$\alpha_{i,n} = \frac{1}{N} a_i^* \left(\frac{(1-\varepsilon)v(\gamma_N)}{1+cv(\gamma_N)\gamma_N} C_N + \frac{1}{n} \sum_{j \neq i}^{\varepsilon_n n} v(\alpha_{j,n}) a_j a_j^* \right)^{-1} a_i, \ \ i = 1, \ldots, \varepsilon_n n.$$

Outlier Data

- For $\varepsilon_n n = 1$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^* + \left(v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}a_1^* C_N^{-1} a_1\right) + o(1)\right)a_1 a_1^*$$

Outlier rejection relies on $\frac{1}{N}a_1^* C_N^{-1} a_1 \lessgtr 1$.

Outlier Data

- For $\varepsilon_n n = 1$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^* + \left(v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}a_1^* C_N^{-1} a_1\right) + o(1)\right) a_1 a_1^*$$

Outlier rejection relies on $\frac{1}{N}a_1^* C_N^{-1} a_1 \lessgtr 1$.

- For $a_i \sim \mathcal{CN}(0, D_N)$,

$$\hat{S}_N = v(\gamma_n)\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + v(\alpha_n)\frac{1}{n}\sum_{i=1}^{\varepsilon_n n} a_i a_i^*$$

$$\gamma_n = \frac{1}{N}\operatorname{tr} C_N\left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n}D_N\right)^{-1}$$

$$\alpha_n = \frac{1}{N}\operatorname{tr} D_N\left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n}D_N\right)^{-1}.$$

Outlier Data

- For $\varepsilon_n n = 1$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^* + \left(v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}a_1^* C_N^{-1} a_1\right) + o(1)\right) a_1 a_1^*$$

Outlier rejection relies on $\frac{1}{N} a_1^* C_N^{-1} a_1 \lessgtr 1$.

- For $a_i \sim \mathcal{CN}(0, D_N)$,

$$\hat{S}_N = v(\gamma_n)\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + v(\alpha_n)\frac{1}{n}\sum_{i=1}^{\varepsilon_n n} a_i a_i^*$$

$$\gamma_n = \frac{1}{N}\operatorname{tr} C_N\left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n}D_N\right)^{-1}$$

$$\alpha_n = \frac{1}{N}\operatorname{tr} D_N\left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n}D_N\right)^{-1}.$$

For $\varepsilon_n \to 0$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}\operatorname{tr} D_N C_N^{-1}\right) a_i a_i^*$$

Outlier rejection relies on $\frac{1}{N}\operatorname{tr} D_N C_N^{-1} \lessgtr 1$.

Outlier Data

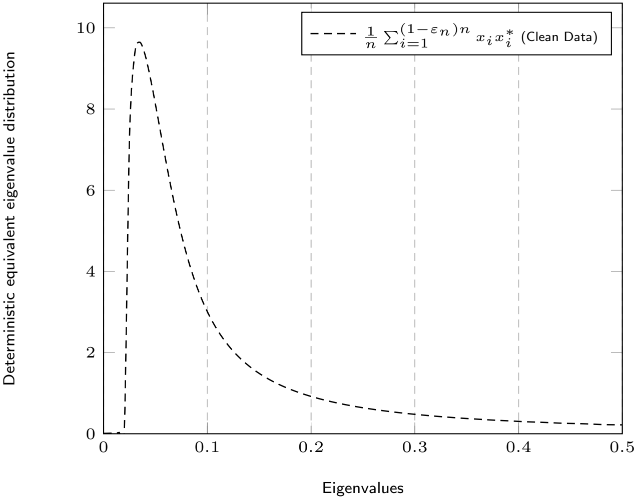


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

Outlier Data

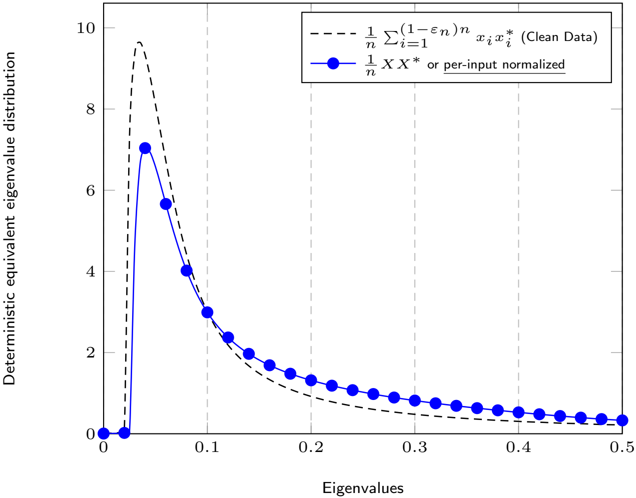


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

Outlier Data

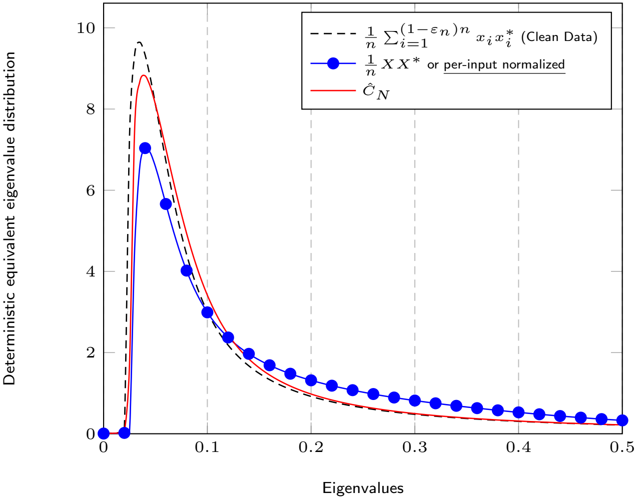


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

Outline

Context

- Generalize robust estimators to $N > n$

Shrinkage Estimators

Context

- Generalize robust estimators to $N > n$
- Exploit Ledoit–Wolf type shrinkage ($+\rho I_N$ regularization)

Context

- Generalize robust estimators to $N > n$
- Exploit Ledoit–Wolf type shrinkage ($+\rho I_N$ regularization)
- Optimize degree of freedom ρ.

Context

- Generalize robust estimators to $N > n$
- Exploit Ledoit–Wolf type shrinkage ($+\rho I_N$ regularization)
- Optimize degree of freedom ρ.

Assumption (Pure-noise model)

Independent $x_1, \ldots, x_n \in \mathbb{C}^N$,

$$x_i = \sqrt{\tau}_i z_i$$

with

- $\tau_i > 0$ arbitrary
- $z_i \sim \mathcal{CN}(0, C_N)$, $\limsup_N \|C_N\| < \infty$
- $\nu_n \triangleq \frac{1}{N} \sum_{i=1}^N \boldsymbol{\delta}_{\lambda_i(C_N)} \to \nu$.

Two estimators in the literature

Definition (Abramovich–Pascal estimate)

For $\rho \in (\max\{0, 1 - n/N\}, 1]$, unique solution $\hat{C}_N(\rho)$ to

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

Property: $\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho) = 1$.

Shrinkage Estimators

Two estimators in the literature

Definition (Abramovich–Pascal estimate)

For $\rho \in (\max\{0, 1 - n/N\}, 1]$, unique solution $\hat{C}_N(\rho)$ to

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

Property: $\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho) = 1$.

Definition (Chen estimate)

For $\rho \in (0, 1]$, unique solution $\check{C}_N(\rho)$ to

$$\check{C}_N(\rho) = \frac{\check{B}_N(\rho)}{\frac{1}{N}\operatorname{tr}\check{B}_N(\rho)}$$

$$\check{B}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\frac{x_i x_i^*}{\frac{1}{N}x_i^*\check{C}_N^{-1}(\rho)x_i} + \rho I_N.$$

Property: $\frac{1}{N}\operatorname{tr}\check{C}_N(\rho) = 1$.

Theorem (Abramovich–Pascal estimator [C,McKay'14])

For $\hat{\mathcal{R}}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$ *as* $n \to \infty$ *with* $N/n \to c \in (0, \infty)$,

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \xrightarrow{\text{a.s.}} 0$$

with

$$\hat{S}_N(\rho) = \frac{1}{\hat{\gamma}(\rho)} \frac{1-\rho}{1-(1-\rho)c} \frac{1}{n} \sum_{i=1}^{n} z_i z_i^* + \rho I_N$$

and $\hat{\gamma}(\rho)$ *unique positive solution to*

$$1 = \int \frac{t}{\rho \hat{\gamma}(\rho) + (1-\rho)t} \nu(dt).$$

Theorem (Chen estimator [C,McKay'14])

Letting $\check{\mathcal{R}}_\varepsilon = [\varepsilon, 1]$, *as* $n \to \infty$ *with* $N/n \to c \in (0, \infty)$,

$$\sup_{\rho \in \check{\mathcal{R}}_\varepsilon} \left\| \check{C}_N(\rho) - \check{S}_N(\rho) \right\| \xrightarrow{\text{a.s.}} 0$$

where

$$\check{S}_N(\rho) = \frac{1-\rho}{1-\rho+T_\rho} \frac{1}{n} \sum_{i=1}^{n} z_i z_i^* + \frac{T_\rho}{1-\rho+T_\rho} I_N$$

in which $T_\rho = \rho \check{\gamma}(\rho) F(\check{\gamma}(\rho); \rho)$ *with, for all* $x > 0$,

$$F(x;\rho) = \frac{1}{2}\left(\rho - c(1-\rho)\right) + \sqrt{\frac{1}{4}\left(\rho - c(1-\rho)\right)^2 + (1-\rho)\frac{1}{x}}$$

and $\check{\gamma}(\rho)$ *unique positive solution to*

$$1 = \int \frac{t}{\rho \check{\gamma}(\rho) + \frac{1-\rho}{(1-\rho)c + F(\check{\gamma}(\rho);\rho)} t} \nu(dt).$$

Corollary (Model Equivalence)

For $\rho \in (0,1]$, there exists a unique $(\hat{\rho}, \check{\rho})$ such that

$$\frac{\hat{S}_N(\hat{\rho})}{\lim_N \frac{1}{N} \operatorname{tr} \hat{S}_N(\hat{\rho})} = \check{S}_N(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} z_i z_i^* + \rho I_N.$$

Besides, $\rho \mapsto \hat{\rho}$ and $\rho \mapsto \check{\rho}$ are continuously increasing and onto.

Consequence: both estimators equivalent in limit to Ledoit–Wolf on z_i (not x_i).

Optimal asymptotic shrinkage

Uniform convergence allows for optimization over ρ.

Proposition (Optimal Frobenius-norm Shrinkage)

For each ρ, define

$$\hat{D}_N(\rho) = \frac{1}{N}\operatorname{tr}\left(\frac{\hat{C}_N(\rho)}{\frac{1}{N}\operatorname{tr}\hat{C}_N(\rho)} - C_N\right)^2$$

$$\check{D}_N(\rho) = \frac{1}{N}\operatorname{tr}\left(\check{C}_N(\rho) - C_N\right)^2$$

$$D^\star = c\frac{M_{\nu,2}-1}{c+M_{\nu,2}-1} \quad (M_{\nu,2},\ \text{order-2 moment})$$

$$\rho^\star = \frac{c}{c+M_{\nu,2}-1}$$

and $\hat{\rho}^\star, \check{\rho}^\star$ unique solutions to

$$\frac{\hat{\rho}^\star}{\frac{1}{\hat{\gamma}(\hat{\rho}^\star)}\frac{1-\hat{\rho}^\star}{1-(1-\hat{\rho}^\star)c} + \hat{\rho}^\star} = \frac{T_{\check{\rho}^\star}}{1-\check{\rho}^\star + T_{\check{\rho}^\star}} = \rho^\star.$$

Then,

$$\inf_{\rho\in\hat{\mathcal{R}}_\varepsilon} \hat{D}_N(\rho) \xrightarrow{\text{a.s.}} D^\star, \quad \inf_{\rho\in\check{\mathcal{R}}_\varepsilon} \check{D}_N(\rho) \xrightarrow{\text{a.s.}} D^\star$$

$$\hat{D}_N(\hat{\rho}^\star) \xrightarrow{\text{a.s.}} D^\star, \quad \check{D}_N(\check{\rho}^\star) \xrightarrow{\text{a.s.}} D^\star.$$

Proposition (Optimal Frobenius-norm shrinkage estimate)

Let $\hat{\rho}_N, \check{\rho}_N$ be solutions to

$$\frac{\hat{\rho}_N}{\frac{1}{N}\operatorname{tr}\hat{C}_N(\hat{\rho}_N)} = \frac{c_N}{\frac{1}{N}\operatorname{tr}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N}\|x_i\|^2}\right)^2\right]-1}$$

$$\frac{\check{\rho}_N \frac{1}{n}\sum_{i=1}^n \frac{x_i^* \check{C}_N(\check{\rho}_N)^{-1} x_i}{\|x_i\|^2}}{1-\check{\rho}_N+\check{\rho}_N \frac{1}{n}\sum_{i=1}^n \frac{x_i^* \check{C}_N(\check{\rho}_N)^{-1} x_i}{\|x_i\|^2}} = \frac{c_N}{\frac{1}{N}\operatorname{tr}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N}\|x_i\|^2}\right)^2\right]-1}.$$

Then

$$\hat{D}_N(\hat{\rho}_N) \xrightarrow{\text{a.s.}} D^\star$$

$$\check{D}_N(\check{\rho}_N) \xrightarrow{\text{a.s.}} D^\star.$$

Optimal asymptotic shrinkage

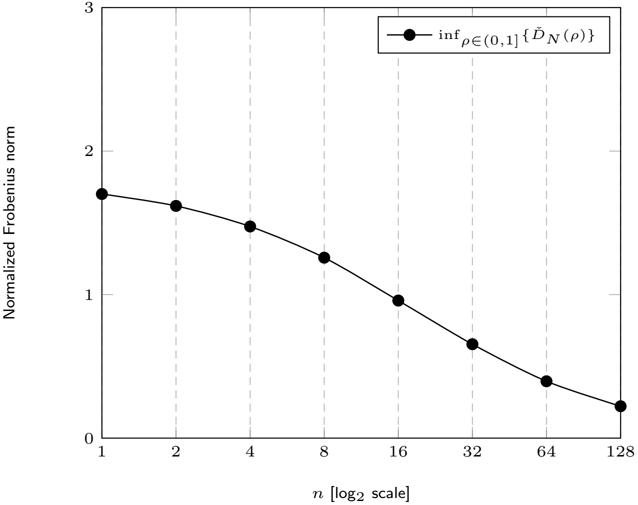


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N{}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

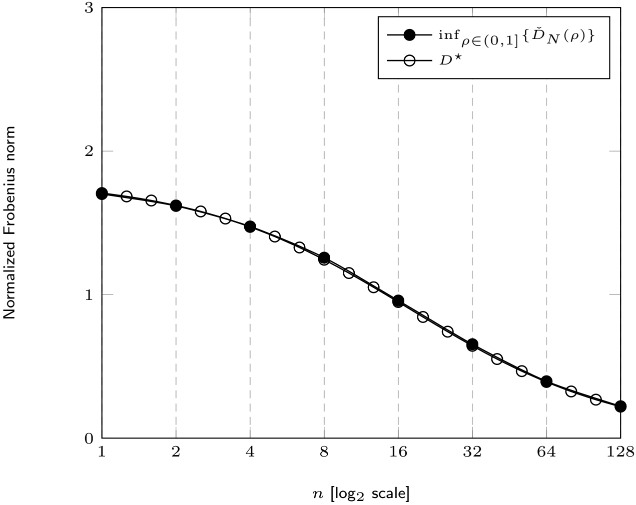


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N{}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

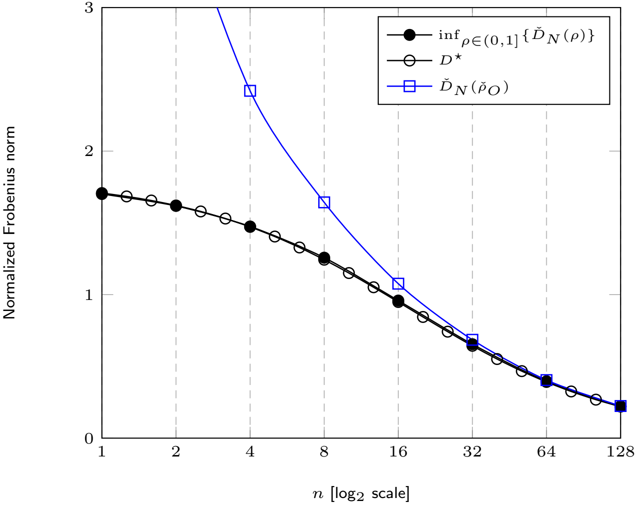


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

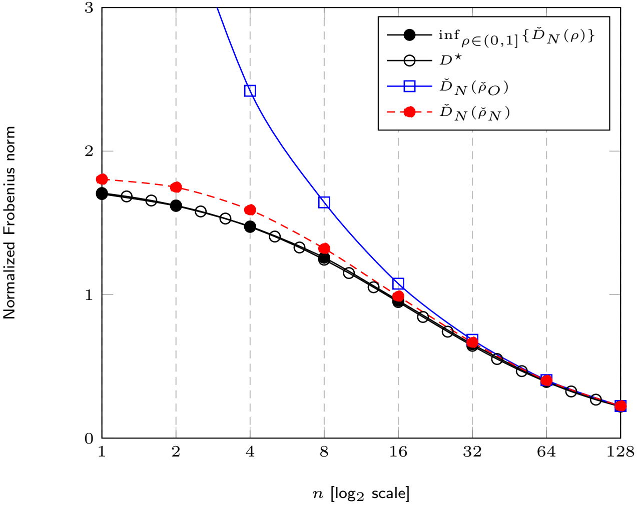


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N{}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

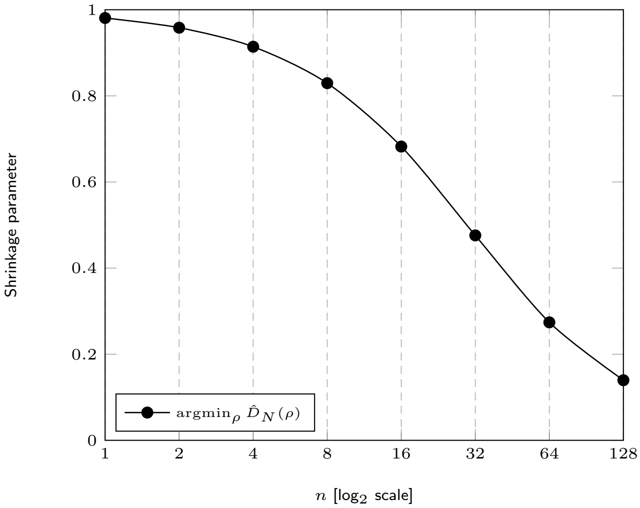


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

Optimal asymptotic shrinkage

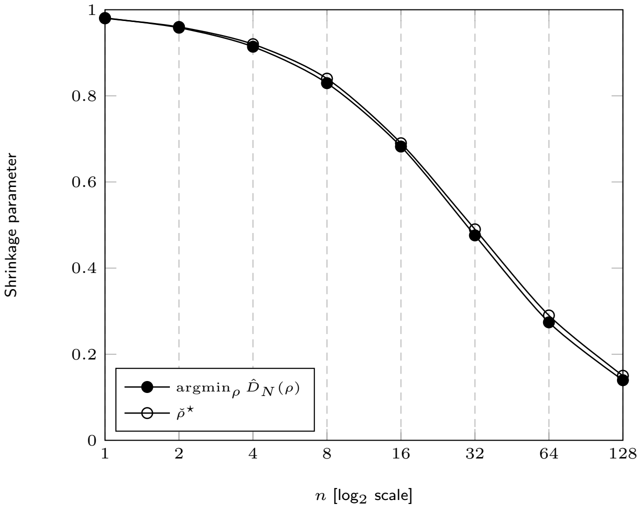


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

Optimal asymptotic shrinkage

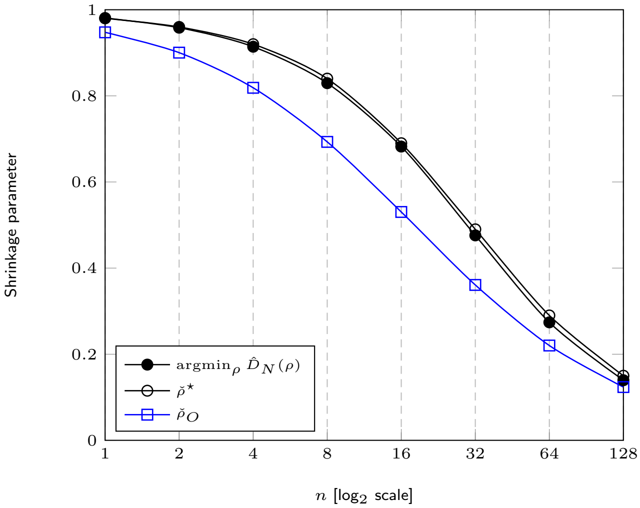


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

Optimal asymptotic shrinkage

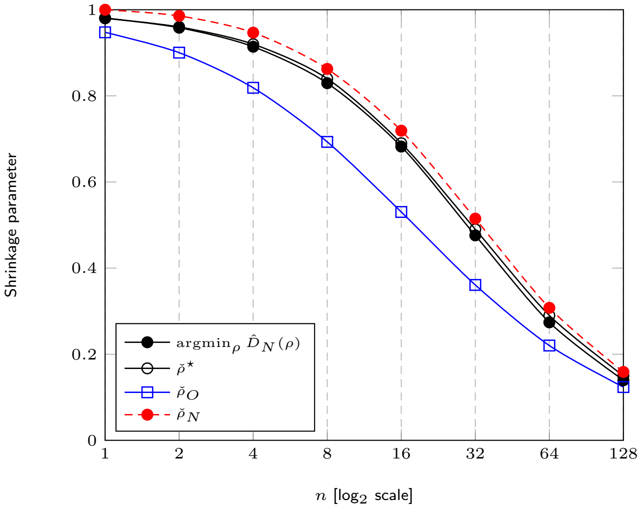


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

Outline

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0, \sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \ \ ?$$

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0, \sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \ \ ?$$

Conjectures

- From simulations, it seems that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| = O(N^{-\frac{1}{2}})$. Weak result.

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0, \sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \ \ ?$$

Conjectures

- From simulations, it seems that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| = O(N^{-\frac{1}{2}})$. Weak result.
- Because of self-averaging, we hope: $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b = o(N^{-\frac{1}{2}})$

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0, \sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \ \ ?$$

Conjectures

- From simulations, it seems that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| = O(N^{-\frac{1}{2}})$. Weak result.
- Because of self-averaging, we hope: $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b = o(N^{-\frac{1}{2}})$
- Since $\sqrt{N}a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, this would imply

$$\sqrt{N}a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \to \mathcal{N}(0, \sigma^2).$$

Theorem (Fluctuation of bilinear forms [C,Kammoun,Pascal'14])

Let $a, b \in \mathbb{C}^N$ *with* $\|a\| = \|b\| = 1$. *Then, as* $n \to \infty$, $N/n \to c \in (0, \infty)$, *for all* $\varepsilon > 0$, $k \in \mathbb{Z}$,

$$\sup_{\rho \in \mathcal{R}_\kappa} N^{1-\varepsilon} \left| a^* \hat{C}_N^k(\rho) b - a^* \hat{S}_N^k(\rho) b \right| \xrightarrow{\text{a.s.}} 0.$$

Theorem (Fluctuation of bilinear forms [C,Kammoun,Pascal'14])

Let $a, b \in \mathbb{C}^N$ *with* $\|a\| = \|b\| = 1$. *Then, as* $n \to \infty$, $N/n \to c \in (0, \infty)$, *for all* $\varepsilon > 0$, $k \in \mathbb{Z}$,

$$\sup_{\rho \in \mathcal{R}_\kappa} N^{1-\varepsilon} \left| a^* \hat{C}_N^k(\rho) b - a^* \hat{S}_N^k(\rho) b \right| \xrightarrow{\text{a.s.}} 0.$$

(with $\varepsilon < \frac{1}{2}$, desired result)

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

- We prove easily (classical proof but with speed)

$$\max_{1 \leq i \leq n} N^{\frac{1}{2}-\varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$$

Not good enough.

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

- We prove easily (classical proof but with speed)

$$\max_{1 \leq i \leq n} N^{\frac{1}{2}-\varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$$

Not good enough.

- **IDEA 1**: Exploit self-averaging

$$a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

$$= \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right]$$

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

- We prove easily (classical proof but with speed)

$$\max_{1 \leq i \leq n} N^{\frac{1}{2}-\varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$$

Not good enough.

- **IDEA 1**: Exploit self-averaging

$$a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

$$= \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right]$$

- But too hard. Since d_i implicit.

Proof idea

- **IDEA 2**: Introduce intermediate quantity

$$\tilde{d}_i(\rho) = \frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i = \frac{1}{N} z_i^* \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{j \neq i}^{n} \frac{z_j z_j^*}{\gamma_N} + \rho I_N \right)^{-1} z_i$$

and write

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = \frac{1-\rho}{1-(1-\rho)c_N} \underbrace{\frac{1}{n} \sum_{i=1}^{n} a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right]}_{\text{Term (A)}}$$

$$+ \frac{1-\rho}{1-(1-\rho)c_N} a^* \hat{C}_N^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \underbrace{\left[\frac{1}{\tilde{d}_i} - \frac{1}{d_i} \right]}_{\text{Term (B)}} z_i z_i^* \right) \hat{S}_N^{-1} b.$$

Proof idea

- **IDEA 2**: Introduce intermediate quantity

$$\tilde{d}_i(\rho) = \frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i = \frac{1}{N} z_i^* \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{j \neq i}^{n} \frac{z_j z_j^*}{\gamma_N} + \rho I_N \right)^{-1} z_i$$

and write

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = \underbrace{\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^{n} a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right]}_{\text{Term (A)}}$$

$$+ \frac{1-\rho}{1-(1-\rho)c_N} a^* \hat{C}_N^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \underbrace{\left[\frac{1}{\tilde{d}_i} - \frac{1}{d_i} \right]}_{\text{Term (B)}} z_i z_i^* \right) \hat{S}_N^{-1} b.$$

- Key lemma for both Terms (A)-(B):

Lemma (Key Lemma, Self-averaging)

$$E\left[\left| \frac{1}{n} \sum_{i=1}^{n} a^* \hat{S}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left(\frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i - \gamma_N \right) \right|^{2p} \right] = O\left(N^{-2p}\right)$$

Application to GLRT detection

Context (Hypothesis Test)

We observe $x_1, \ldots, x_n$, $x_i = \sqrt{\tau}_i w_i$, $\|w_i\|^2 = N$ isotropic, and receive

$$y = \begin{cases} x & , \mathcal{H}_0 \\ \alpha p + x & , \mathcal{H}_1 \end{cases}$$

with $\alpha > 0$ unknown, $p \in \mathbb{C}^N$ known.

Context (Hypothesis Test)

We observe $x_1, \ldots, x_n$, $x_i = \sqrt{\tau}_i w_i$, $\|w_i\|^2 = N$ isotropic, and receive

$$y = \begin{cases} x & , \mathcal{H}_0 \\ \alpha p + x & , \mathcal{H}_1 \end{cases}$$

with $\alpha > 0$ unknown, $p \in \mathbb{C}^N$ known.

Definition (GLRT Detector)

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

with

$$T_N(\rho) \triangleq \frac{|y^* \hat{C}_N^{-1}(\rho) p|}{\sqrt{y^* \hat{C}_N^{-1}(\rho) y} \sqrt{p^* \hat{C}_N^{-1}(\rho) p}}.$$

Theorem (Asymptotic detector performance [C,Kammoun,Pascal'14])

Under $\mathcal{H}_0$, as $n \to \infty$ with $N/n \to c \in (0,\infty)$,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}} \right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\rho)} \right) \right| \to 0$$

where

$$\sigma_N^2(\rho) \triangleq \frac{1}{2} \frac{p^* C_N Q_N^2(\underline{\rho}) p}{p^* Q_N(\underline{\rho}) p \cdot \frac{1}{N} \mathit{tr} C_N Q_N(\underline{\rho}) \cdot \left(1 - c(1-\underline{\rho})^2 m(-\underline{\rho})^2 \frac{1}{N} \mathit{tr} C_N^2 Q_N^2(\underline{\rho}) \right)}$$

with $Q_N(\underline{\rho}) \triangleq (I_N + (1-\underline{\rho}) m(-\underline{\rho}) C_N)^{-1}$ *and* $\underline{\rho} = \rho(\rho + \frac{1}{\gamma_N(\rho)} \frac{1-\rho}{1-(1-\rho)c})^{-1}$.

Proposition (Empirical performance optimum)

Let

$$\hat{\sigma}_N^2(\rho) \triangleq \frac{1}{2} \frac{1 - \rho \frac{p^* \hat{C}_N^{-2}(\rho) p}{p^* \hat{C}_N^{-1}(\rho) p}}{\left(1 - c_N + c_N \rho\right)(1 - \rho)}.$$

Then,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| \sigma_N^2(\rho) - \hat{\sigma}_N^2(\rho) \right| \xrightarrow{\text{a.s.}} 0.$$

Proposition (Empirical performance optimum)

Let

$$\hat{\sigma}_N^2(\rho) \triangleq \frac{1}{2} \frac{1 - \rho \frac{p^* \hat{C}_N^{-2}(\rho) p}{p^* \hat{C}_N^{-1}(\rho) p}}{(1 - c_N + c_N \rho)(1 - \rho)}.$$

Then,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| \sigma_N^2(\rho) - \hat{\sigma}_N^2(\rho) \right| \xrightarrow{\text{a.s.}} 0.$$

Besides, let

$$\hat{\rho}_N^* \in \operatorname{argmin}_{\rho \in \mathcal{R}_\kappa} \left\{ \hat{\sigma}_N^2(\rho) \right\}.$$

Then, for every $\gamma > 0$,

$$P\left(\sqrt{N} T_N(\hat{\rho}_N^*) > \gamma \right) - \inf_{\rho \in \mathcal{R}_\kappa} \left\{ P\left(\sqrt{N} T_N(\rho) > \gamma \right) \right\} \to 0.$$

Application to GLRT detection

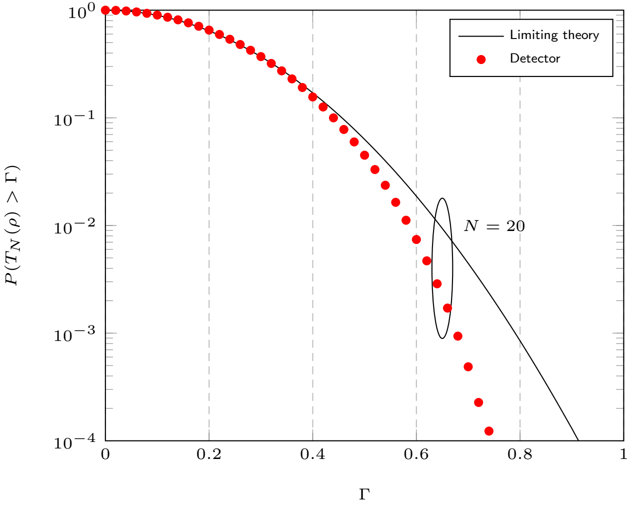


Figure: False alarm rate $P(T_N(\hat{\rho}_N^*) > \Gamma)$, $N = 20$ or $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_N]_{ij} = .7^{|i-j|}$, $N/n = 1/2$.

Application to GLRT detection

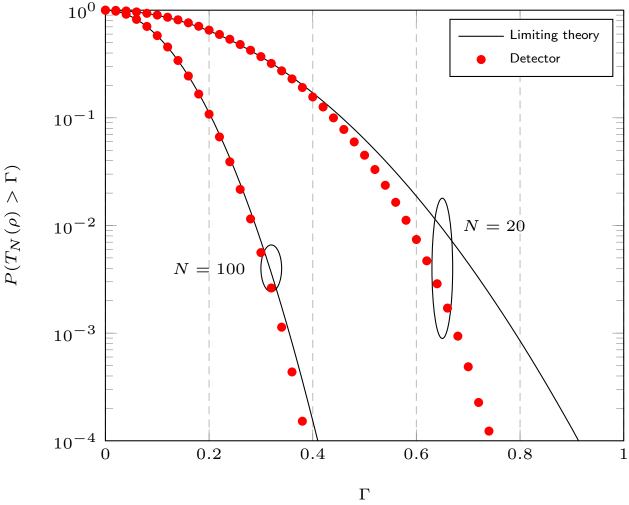


Figure: False alarm rate $P(T_N(\hat{\rho}_N^*) > \Gamma)$, $N = 20$ or $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_N]_{ij} = .7^{|i-j|}$, $N/n = 1/2$.

Outline

Takeaway messages

- Robust estimators of scatter: $\left\{ \begin{array}{l} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{array} \right.$

Takeaway messages

- Robust estimators of scatter: $\left\{ \begin{array}{l} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{array} \right.$
- Approximations useful up to second-order fluctuations.

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \mathrm{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \; y_i \in \mathbb{R}, \; x_i \in \mathbb{R}^N.$$

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \; y_i \in \mathbb{R}, \; x_i \in \mathbb{R}^N.$$

More general framework: BigData RMT

- Kernel random matrices (application to spectral clustering, kernel PCA)

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \; y_i \in \mathbb{R}, \; x_i \in \mathbb{R}^N.$$

More general framework: BigData RMT

- Kernel random matrices (application to spectral clustering, kernel PCA)
- (A)symmetric graphical models (adjacency, Laplacian matrices), neural nets

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \mathrm{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \ y_i \in \mathbb{R}, \ x_i \in \mathbb{R}^N.$$

More general framework: BigData RMT

- Kernel random matrices (application to spectral clustering, kernel PCA)
- (A)symmetric graphical models (adjacency, Laplacian matrices), neural nets
- Sparsity considerations (sparse PCA).

Thank you.