Robust estimation in the large random matrix regime

Romain Couillet
Assistant Professor at SUPÉLEC
Joint work with J. W. Silverstein, F. Pascal, and A. Kammoun.

Random Matrices and their Applications

08-10-2012

Supélec

Outline

Outline

Parameter estimation and sample covariance matrix

→ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

[1] V. Girko, "Then years of general statistical analysis", Chapter 14, online.

[2] Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

[3] P.J. Huber, "Robust Statistics", 1981.

Parameter estimation and sample covariance matrix

→ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

- The main reasons are:
  - Assuming $E[x] = 0$, $E[xx^*] = C_N$, with $X = [x_1, \ldots, x_n]$, by the LLN

$$\hat{S}_N \triangleq \frac{1}{n} XX^* \xrightarrow{\text{a.s.}} C_N \text{ as } n \to \infty.$$

→ Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.

[1] V. Girko, "Then years of general statistical analysis", Chapter 14, online.

[2] Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

[3] P.J. Huber, "Robust Statistics", 1981.

Parameter estimation and sample covariance matrix

→ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

- The main reasons are:
  - Assuming $E[x] = 0$, $E[xx^*] = C_N$, with $X = [x_1, \ldots, x_n]$, by the LLN

$$\hat{S}_N \triangleq \frac{1}{n} XX^* \xrightarrow{\text{a.s.}} C_N \text{ as } n \to \infty.$$

    → Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.
  - The SCM $\hat{S}_N$ is the ML estimate of C_N for Gaussian x
    → One therefore expects $\hat{\theta}$ to closely approximate θ for all finite n.

[1] V. Girko, "Then years of general statistical analysis", Chapter 14, online.

[2] Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

[3] P.J. Huber, "Robust Statistics", 1981.

Parameter estimation and sample covariance matrix

→ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

- The main reasons are:
  - Assuming $E[x] = 0$, $E[xx^*] = C_N$, with $X = [x_1, \ldots, x_n]$, by the LLN

$$\hat{S}_N \triangleq \frac{1}{n} XX^* \xrightarrow{\text{a.s.}} C_N \text{ as } n \to \infty.$$

    → Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.
  - The SCM $\hat{S}_N$ is the ML estimate of C_N for Gaussian x
    → One therefore expects $\hat{\theta}$ to closely approximate θ for all finite n.
- This approach however has two limitations:
  - if N, n are of the same order of magnitude,

$$\|\hat{S}_N - C_N\| \not\to 0 \text{ as } N, n \to \infty, \; N/n \to c > 0, \text{ so that in general } |\hat{\theta} - \theta| \not\to 0$$

    → This motivated the introduction of G-estimators.[1][2]
  - if x is not Gaussian, but has heavier tails, $\hat{S}_N$ is a poor estimator for C_N.
    → This motivated the introduction of robust estimators.[3]

---

[1] V. Girko, "Then years of general statistical analysis", Chapter 14, online.
[2] Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.
[3] P.J. Huber, "Robust Statistics", 1981.

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by another estimate $\hat{C}_N$ of C_N which:
  - rejects (or downscales) observations deterministically
  - or rejects observations inconsistent with the full set of observations

→ **Example**: Huber estimator, $\hat{C}_N$ defined as solution of

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} \beta_i x_i x_i^* \text{ with } \beta_i = \alpha \min\left\{1, \frac{k^2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i}\right\} \text{ for some } \alpha > 1, k^2 \text{ function of } \hat{C}_N.$$

[4] J. T. Kent, D. E. Tyler, "Redescending M-estimates of multivariate location and scatter", 1991.

[5] Y. Chitour, F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis", 2008.

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by another estimate $\hat{C}_N$ of C_N which:
  - rejects (or downscales) observations deterministically
  - or rejects observations inconsistent with the full set of observations

→ **Example**: Huber estimator, $\hat{C}_N$ defined as solution of

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} \beta_i x_i x_i^* \text{ with } \beta_i = \alpha \min\left\{1, \frac{k^2}{\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i}\right\} \text{ for some } \alpha > 1, k^2 \text{ function of } \hat{C}_N.$$

- provide scale-free estimators of C_N:

→ **Example**: Tyler's estimator[4]: if one observes $x_i = \tau_i z_i$ for unknown scalars τ_i,

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

  - existence and uniqueness of $\hat{C}_N$ defined up to a constant.
  - few constraints on $x_1, \ldots, x_n$ ($N+1$ of them must be linearly independent)[5]

[4] J. T. Kent, D. E. Tyler, "Redescending M-estimates of multivariate location and scatter", 1991.

[5] Y. Chitour, F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis", 2008.

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.

  → **Example**: Maronna's estimator[6] for elliptical x

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

  with $u(s)$ such that

  (i) $u(s)$ is continuous and non-increasing on $[0, \infty)$

  (ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.

  (note that Huber's estimator is compliant with Maronna's estimators)

[6] R. A. Maronna, "Robust M-estimators of multivariate location and scatter", 1976.

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.
  → **Example**: Maronna's estimator[6] for elliptical x

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

with $u(s)$ such that

(i) $u(s)$ is continuous and non-increasing on $[0, \infty)$

(ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.

(note that Huber's estimator is compliant with Maronna's estimators)

  - existence is not too demanding
  - uniqueness imposes constraints on $N, n, u(s)$, e.g. $\phi_\infty > \frac{n}{n-N}$. Inconsistent with random matrix regime!
  - consistency result: $\hat{C}_N \to C_N$ **if** $u(s)$ meets the ML estimator for C_N.

[6] R. A. Maronna, "Robust M-estimators of multivariate location and scatter", 1976.

Results in RMT

$\rightarrow$ So far, RMT has mostly focused on the SCM $\hat{S}_N$.

- $x = A_N y$, y having i.i.d. zero-mean unit variance entries,
    - Limiting spectrum of $\hat{S}_N$ known as $N, n \to \infty$, $N/n \to c$.[7][8]
    - No eigenvalues outside the support.[9]
    - Exact separation of eigenvalues.[10]
    - Statistical inference methods based on the SCM (improved subspace estimators).[11]

---

[7] V. A. Marcenko, L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices", 1967.

[8] J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices", 1995

[9] Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices", 1998.

[10] Z. D. Bai, J. W. Silverstein, "Exact separation of eigenvalues of large dimensional sample covariance matrices", 1999.

[11] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

[12] N. El Karoui, "Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond", 2009.

Results in RMT

→ So far, RMT has mostly focused on the SCM $\hat{S}_N$.

- $x = A_N y$, y having i.i.d. zero-mean unit variance entries,
  - Limiting spectrum of $\hat{S}_N$ known as $N, n \to \infty$, $N/n \to c$.[7][8]
  - No eigenvalues outside the support.[9]
  - Exact separation of eigenvalues.[10]
  - Statistical inference methods based on the SCM (improved subspace estimators).[11]
- x satisfies concentration inequalities,
  - Limiting spectrum of $\hat{S}_N$ for elliptically distributed x.[12]

[7] V. A. Marcenko, L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices", 1967.

[8] J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices", 1995

[9] Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices", 1998.

[10] Z. D. Bai, J. W. Silverstein, "Exact separation of eigenvalues of large dimensional sample covariance matrices", 1999.

[11] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

[12] N. El Karoui, "Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond", 2009.

Results in RMT

$\rightarrow$ So far, RMT has mostly focused on the SCM $\hat{S}_N$.

- $x = A_N y$, y having i.i.d. zero-mean unit variance entries,
  - Limiting spectrum of $\hat{S}_N$ known as $N, n \rightarrow \infty$, $N/n \rightarrow c$.[7][8]
  - No eigenvalues outside the support.[9]
  - Exact separation of eigenvalues.[10]
  - Statistical inference methods based on the SCM (improved subspace estimators).[11]
- x satisfies concentration inequalities,
  - Limiting spectrum of $\hat{S}_N$ for elliptically distributed x.[12]

Robust RMT estimation

Can we study the performance of estimators based on the $\hat{C}_N$?

- what are the spectral properties of $\hat{C}_N$?
- can we generate RMT-based estimators relying on $\hat{C}_N$?

---

[7] V. A. Marcenko, L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices", 1967.

[8] J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices", 1995

[9] Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices", 1998.

[10] Z. D. Bai, J. W. Silverstein, "Exact separation of eigenvalues of large dimensional sample covariance matrices", 1999.

[11] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

[12] N. El Karoui, "Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond", 2009.

Some first answers

→ We recall that

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

for some i.i.d. $x_1,\ldots,x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

$$F^{\alpha\hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N,n\to\infty,\ N/n\to c>0$$

so in particular, for $C_N = I_N$,

$$F^{\alpha\hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

Some first answers

→ We recall that

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

for some i.i.d. $x_1, \ldots, x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

$$F^{\alpha \hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N, n \to \infty,\ N/n \to c > 0$$

so in particular, for $C_N = I_N$,

$$F^{\alpha \hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

- Zooming in on the eigenvalues suggests also that

$$\lambda_i(\alpha \hat{C}_N) - \lambda_i(\hat{S}_N) \Rightarrow 0$$

Some first answers

→ We recall that

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

for some i.i.d. $x_1,\ldots,x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

$$F^{\alpha\hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N,n\to\infty,\ \ N/n\to c>0$$

so in particular, for $C_N = I_N$,

$$F^{\alpha\hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

- Zooming in on the eigenvalues suggests also that

$$\lambda_i(\alpha\hat{C}_N) - \lambda_i(\hat{S}_N) \Rightarrow 0$$

→ This behavior seems to be linked to a concentration result on

$$\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i,\ \ i=1,\ldots,n$$

⇒ This is what we are going to prove.

Some first answers

→ We recall that

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

for some i.i.d. $x_1, \ldots, x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

  $$F^{\alpha \hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N, n \to \infty,\ N/n \to c > 0$$

  so in particular, for $C_N = I_N$,

  $$F^{\alpha \hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

- Zooming in on the eigenvalues suggests also that

  $$\lambda_i(\alpha \hat{C}_N) - \lambda_i(\hat{S}_N) \Rightarrow 0$$

→ This behavior seems to be linked to a concentration result on

$$\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i,\ i = 1, \ldots, n$$

⇒ This is what we are going to prove.

Then, what happens to $\hat{C}_N$ when no concentration result occurs?

⇒ So far, we have no general answer to this question!

Some first answers (2)

→ Main difficulties for handling $\hat{C}_N$:

- $\hat{C}_N$ does not always exist/is not always unique.
  - sometimes, uniqueness results inconsistent with random matrix regime

Some first answers (2)

$\rightarrow$ Main difficulties for handling $\hat{C}_N$:

- $\hat{C}_N$ does not always exist/is not always unique.
  - sometimes, uniqueness results inconsistent with random matrix regime
- Contrary to classical RMT, the column vectors $\sqrt{u(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i)}x_i$ are not independent
  - difficult to find an angle to reuse previous results
- In general, it is already difficult to show that both $\|\hat{C}_N\|$ and $\|\hat{C}_N^{-1}\|$ remain bounded as $N, n \rightarrow \infty$, $N/n \rightarrow c > 0$.

Outline

Outline

Reminder of Assumptions

Assumptions

- Assumptions on $u(s)$,
  (i) $u(s)$ is continuous and non-increasing on $[0, \infty)$
  (ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.
- Assumptions on $x_1, \ldots, x_n$,
  - $x_i = A_N y_i \in \mathbb{C}^N$, $y_i \in \mathbb{C}^M$ has independent entries with
    - $E[y_{i,j}] = 0$
    - $E[y_{i,j}^2] = 0$, $E[|y_{i,j}|^2] = 1$
    - $\sup_{i,j} E[|y_{i,j}|^{8+\eta}] < \infty$.
  - With $c_N = N/n$, $\bar{c}_N = M/N \geqslant 1$,

$$0 < \liminf_n c_N \leqslant \limsup_n c_N < 1, \quad \limsup_n \bar{c}_N < \infty$$

  - Denoting $C_N = A_N A_N^*$,

$$0 < \liminf_N \{\lambda_1(C_N)\} \leqslant \limsup_N \{\lambda_N(C_N)\} < \infty$$

Robust SCM estimator in the RMT regime[13]

Theorem

Assume the above and consider the fixed-point equation in $Z \in \mathbb{C}^{N\times N}$,

$$Z = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^* Z^{-1} x_i\right) x_i x_i^*. \qquad (1)$$

Then,

(I) *Equation* (1) *has a unique solution* $\hat{C}_N$ *for all large* N *a.s., defined as*

$$\hat{C}_N = \lim_{t\to\infty} Z^{(t)}$$

where $Z^{(0)} = I_N$ *and, for* $t \in \mathbb{N}$,

$$Z^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^* (Z^{(t)})^{-1} x_i\right) x_i x_i^*.$$

[13] R. Couillet, F. Pascal, J. W. Silverstein, "Robust M-Estimation for Array Processing: A Random Matrix Approach", 2012.

Robust SCM estimator in the RMT regime[13]

Theorem

Assume the above and consider the fixed-point equation in $Z \in \mathbb{C}^{N \times N}$,

$$Z = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} x_i^* Z^{-1} x_i\right) x_i x_i^*. \tag{1}$$

Then,

(I) *Equation* (1) *has a unique solution* $\hat{C}_N$ *for all large* N *a.s., defined as*

$$\hat{C}_N = \lim_{t \to \infty} Z^{(t)}$$

where $Z^{(0)} = I_N$ *and, for* $t \in \mathbb{N}$,

$$Z^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} x_i^* (Z^{(t)})^{-1} x_i\right) x_i x_i^*.$$

(II) *Defining* $\hat{C}_N = I_N$ *when* (1) *does not have a unique solution,*

$$\left\| \phi^{-1}(1) \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0.$$

[13] R. Couillet, F. Pascal, J. W. Silverstein, "Robust M-Estimation for Array Processing: A Random Matrix Approach", 2012.

Robust statistical inference in RMT regime

→ From Theorem 1,

- Weak convergence results on $\hat{S}_N$ propagate to $\hat{C}_N$;
- No eigenvalues and exact separation results propagate to $\hat{S}_N$;
- First order results on spiked models as well, etc.
- Irrelevant of underlying distribution of x, as opposed to the finite N regime

Robust statistical inference in RMT regime

→ From Theorem 1,

- Weak convergence results on $\hat{S}_N$ propagate to $\hat{C}_N$;
- No eigenvalues and exact separation results propagate to $\hat{S}_N$;
- First order results on spiked models as well, etc.
- Irrelevant of underlying distribution of x, as opposed to the finite N regime

→ (Almost) immediate consequence:

- RMT-based statistical estimators using $\hat{S}_N$ can be replaced by identical estimators using $\hat{C}_N$
- e.g. Mestre's DoA estimator

Robust statistical inference in RMT regime

→ From Theorem 1,

- Weak convergence results on $\hat{S}_N$ propagate to $\hat{C}_N$;
- No eigenvalues and exact separation results propagate to $\hat{S}_N$;
- First order results on spiked models as well, etc.
- Irrelevant of underlying distribution of x, as opposed to the finite N regime

→ (Almost) immediate consequence:

- RMT-based statistical estimators using $\hat{S}_N$ can be replaced by identical estimators using $\hat{C}_N$
- e.g. Mestre's DoA estimator

→ Theorem 1 however does not say anything about second order results.

- Current investigation: CLT on linear statistics for $\hat{C}_N$, for x with i.i.d. entries.
- This should provide the asymptotic performance comparison between robust-RMT estimators and traditional RMT estimators.
- So far, it seems that limiting variance depends mostly on C_N, c, $u'(\phi^{-1}(1))$, and the kurtosis of the entries of x.

Outline

Robust G-MUSIC estimator

→ Consider the model

$$x_t = \sum_{k=1}^{K} \sqrt{p_k} s(\theta_k) z_{k,t} + \sigma w_t = A_N y_t, \quad A_N \triangleq \left[S(\Theta) P^{\frac{1}{2}} \quad \sigma I_N \right], \text{ with}$$

- $S(\Theta) = [s(\theta_1), \ldots, s(\theta_K)]$ deterministic bounded norm steering vectors,
- $P = \text{diag}(p_1, \ldots, p_K)$ diagonal of powers,
- $y_t = (z_{1,t}, \ldots, z_{K,t}, w_t^\mathsf{T})^\mathsf{T} \in \mathbb{C}^{N+K}$, signals and noise vector.

[14] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

Robust G-MUSIC estimator

→ Consider the model

$$x_t = \sum_{k=1}^{K} \sqrt{p_k} s(\theta_k) z_{k,t} + \sigma w_t = A_N y_t, \quad A_N \triangleq \left[S(\Theta) P^{\frac{1}{2}} \quad \sigma I_N \right], \text{ with}$$

- $S(\Theta) = [s(\theta_1), \ldots, s(\theta_K)]$ deterministic bounded norm steering vectors,
- $P = \mathrm{diag}(p_1, \ldots, p_K)$ diagonal of powers,
- $y_t = (z_{1,t}, \ldots, z_{K,t}, w_t^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{C}^{N+K}$, signals and noise vector.

→ From the above results and Mestre's DoA estimator[14],

Theorem (Robust G-MUSIC)

Denote $E_W \in \mathbb{C}^{N \times (N-K)}$ the "noise subspace" of C_N, $\hat{e}_k$ the eigenvector of $\hat{C}_N$ with eigenvalue $\hat{\lambda}_k \triangleq \lambda_k(\hat{C}_N)$. Then, as $N, n \to \infty$ and K fixed,

$$\gamma(\theta) - \hat{\gamma}(\theta) \xrightarrow{\text{a.s.}} 0, \quad \gamma(\theta) = s(\theta)^* E_W E_W^* s(\theta), \quad \hat{\gamma}(\theta) = \sum_{i=1}^{N} \beta_i s(\theta)^* \hat{e}_i \hat{e}_i^* s(\theta)$$

and

$$\beta_i = \begin{cases} 1 + \sum_{k=N-K+1}^{N} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_i - \hat{\mu}_k} \right) & , i \leqslant N-K \\ -\sum_{k=1}^{N-K} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_i - \hat{\mu}_k} \right) & , i > N-K \end{cases}$$

with $\hat{\mu}_1 \leqslant \ldots \leqslant \hat{\mu}_N$ the eigenvalues of $\mathrm{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{n}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^{\mathsf{T}}$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^{\mathsf{T}}$.

[14] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates", 2008.

Results

→ The interest of the above robust-DoA scheme is to:

- handle noise that is "only well-approximated by Gaussian"
- handle model based on bursts of errors on individual antennas
- handle noise distributions with heavier-than-Gaussian tails in radars with distributed antennas (e.g. MIMO radars)

Results

→ The interest of the above robust-DoA scheme is to:

- handle noise that is "only well-approximated by Gaussian"
- handle model based on bursts of errors on individual antennas
- handle noise distributions with heavier-than-Gaussian tails in radars with distributed antennas (e.g. MIMO radars)

→ Some strong limitations:

- cannot handle distributions with heavier-than-Gaussian tails in classical radars
  - this would impose to choose x e.g. elliptically distributed
  - our proof technique collapses here
- cannot handle scale-free detectors/estimators, with $u(s) = 1/s$

Simulation results: The Gaussian noise reference

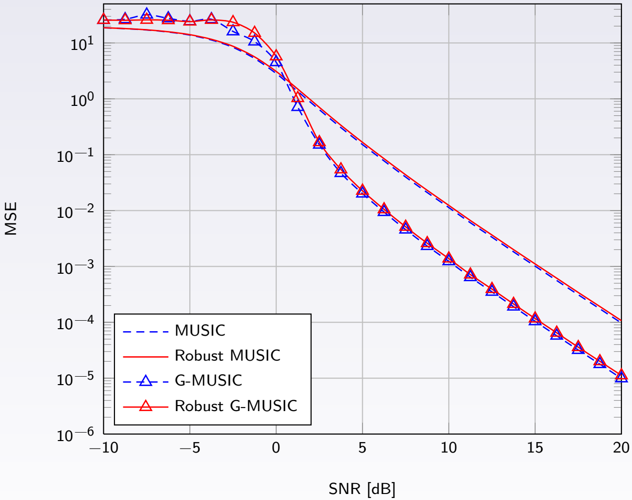

Figure : MSE performance of the various MUSIC estimators for $K = 1$, Gaussian noise, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Simulation results: Close-to-Gaussian noise with i.i.d. Student entries

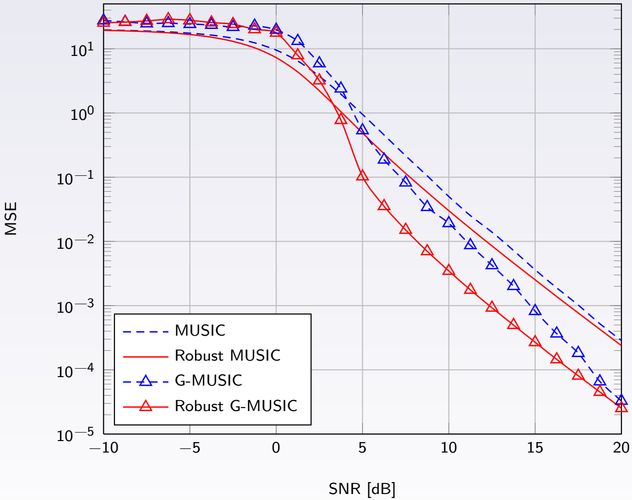

Figure : MSE performance of the various MUSIC estimators for $K = 1$, Student-t noise with $\nu = 5$, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Simulation results: Far-from-Gaussian noise with i.i.d. Student entries

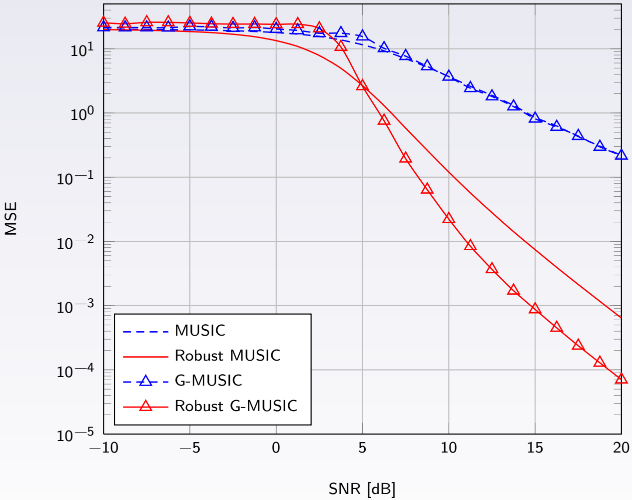

Figure : MSE performance of the various MUSIC estimators for $K = 1$, Student-t noise with $\nu = 2.5$, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Simulation results: Resolution power

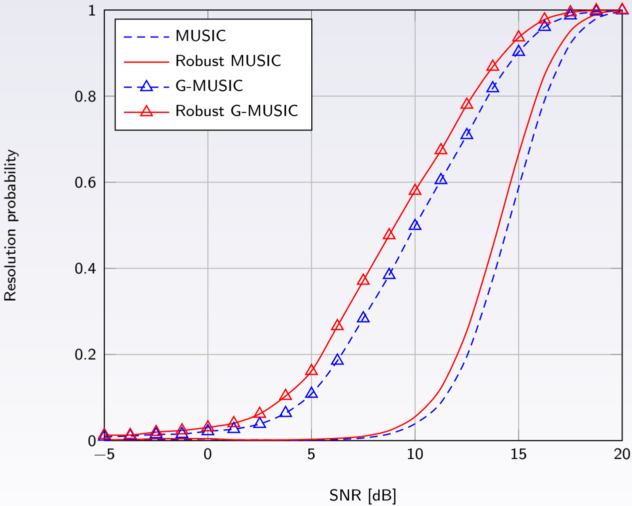

Figure : Resolution performance of the various MUSIC estimators, $\theta_1 = 10^\circ$, $\theta_2 = 15^\circ$, Student-t noise with $\nu = 5$, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Outline

Outline

Standard interference functions

Definition (Standard Interference Function)

Function $h = (h_1, \ldots, h_n) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is a standard interference function if for all j:

1. *Positivity:* for $q_1, \ldots, q_n \geqslant 0$, $h_j(q_1, \ldots, q_n) > 0$;
2. *Monotonicity:* $q_1 \geqslant q_1', \ldots, q_n \geqslant q_n'$ implies $h_j(q_1, \ldots, q_n) \geqslant h_j(q_1', \ldots, q_n')$;
3. *Scalability:* for all $\alpha > 1$, $\alpha h_j(q_1, \ldots, q_n) \geqslant h_j(\alpha q_1, \ldots, \alpha q_n)$.

[15] Yates, "A framework for uplink power control in cellular radio systems", 1995.

Standard interference functions

Definition (Standard Interference Function)

Function $h = (h_1, \ldots, h_n) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is a standard interference function if for all j:

1. *Positivity:* for $q_1, \ldots, q_n \geqslant 0$, $h_j(q_1, \ldots, q_n) > 0$;
2. *Monotonicity:* $q_1 \geqslant q_1', \ldots, q_n \geqslant q_n'$ implies $h_j(q_1, \ldots, q_n) \geqslant h_j(q_1', \ldots, q_n')$;
3. *Scalability:* for all $\alpha > 1$, $\alpha h_j(q_1, \ldots, q_n) \geqslant h_j(\alpha q_1, \ldots, \alpha q_n)$.

Theorem

If $h(q_1, \ldots, q_n)$ is a standard interference function and there exists $(q_1, \ldots, q_n)$ such that for all j, $q_j \geqslant h_j(q_1, \ldots, q_n)$, then

$$q_j = h_j(q_1, \ldots, q_n)$$

for $j = 1, \ldots, n$, has at least one solution, given by $\lim_{t\to\infty}(q_1^{(t)}, \ldots, q_n^{(t)})$, where

$$q_j^{(t+1)} = h_j(q_1^{(t)}, \ldots, q_n^{(t)})$$

for $t \geqslant 1$ and any initial values $q_1^{(0)}, \ldots, q_n^{(0)} \geqslant 0$. Moreover, if for all $\alpha > 1$, all $q_1, \ldots, q_n$, and all j, $\alpha h_j(q_1, \ldots, q_n) > h_j(\alpha q_1, \ldots, \alpha q_n)$, then this solution is unique.

[15] Yates, "A framework for uplink power control in cellular radio systems", 1995.

Standard interference functions

Definition (Standard Interference Function)

Function $h = (h_1, \ldots, h_n) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is a standard interference function if for all j:

1. *Positivity:* for $q_1, \ldots, q_n \geqslant 0$, $h_j(q_1, \ldots, q_n) > 0$;
2. *Monotonicity:* $q_1 \geqslant q_1', \ldots, q_n \geqslant q_n'$ implies $h_j(q_1, \ldots, q_n) \geqslant h_j(q_1', \ldots, q_n')$;
3. *Scalability:* for all $\alpha > 1$, $\alpha h_j(q_1, \ldots, q_n) \geqslant h_j(\alpha q_1, \ldots, \alpha q_n)$.

Theorem

If $h(q_1, \ldots, q_n)$ is a standard interference function and there exists $(q_1, \ldots, q_n)$ such that for all j, $q_j \geqslant h_j(q_1, \ldots, q_n)$, then

$$q_j = h_j(q_1, \ldots, q_n)$$

for $j = 1, \ldots, n$, has at least one solution, given by $\lim_{t \to \infty}(q_1^{(t)}, \ldots, q_n^{(t)})$, where

$$q_j^{(t+1)} = h_j(q_1^{(t)}, \ldots, q_n^{(t)})$$

for $t \geqslant 1$ and any initial values $q_1^{(0)}, \ldots, q_n^{(0)} \geqslant 0$. Moreover, if for all $\alpha > 1$, all $q_1, \ldots, q_n$, and all j, $\alpha h_j(q_1, \ldots, q_n) > h_j(\alpha q_1, \ldots, \alpha q_n)$, then this solution is unique.

→ This is an adaptation of a result from Yates[15]
→ Since $su(s)$ may have a flat region, scalability may not be ensured!

[15] Yates, "A framework for uplink power control in cellular radio systems", 1995.

Proving existence (1)

→ Consider the vector-function $h = (h_1, \ldots, h_n)$ with

$$h_j(q_1, \ldots, q_n) \triangleq \frac{1}{N} x_j^* \left(\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \right)^{-1} x_j$$

Proving existence (1)

→ Consider the vector-function $h = (h_1, \ldots, h_n)$ with

$$h_j(q_1, \ldots, q_n) \triangleq \frac{1}{N} x_j^* \left(\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \right)^{-1} x_j$$

→ We have:

- **Positivity**: $h_j > 0$ for all large n a.s.
- **Monotonicity**: Let $(q_1', \ldots, q_n') \geqslant (q_1, \ldots, q_n)$, then
  - From u non-increasing,

$$\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \succeq \frac{1}{n} \sum_{i=1}^{n} u(q_i') x_i x_i^*$$

  - From Corollary 7.7.4 in Horn and Johnson, "Matrix Analysis"

$$\left(\frac{1}{n} \sum_{i=1}^{n} u(q_i') x_i x_i^* \right)^{-1} \succeq \left(\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \right)^{-1}$$

  - Hence $h_j(q_1', \ldots, q_n') \geqslant h_j(q_1, \ldots, q_n)$
- **Scalability**: Let $\alpha > 1$,
  - From ϕ non-decreasing,

$$\phi(\alpha q_i) \geqslant \phi(q_i) \text{ so that } u(\alpha q_i) \geqslant \frac{1}{\alpha} u(q_i)$$

  - With same arguments as above, $\alpha h_j(q_1, \ldots, q_n) \geqslant h_j(\alpha q_1, \ldots, \alpha q_n)$

Proving existence (1)

→ Consider the vector-function $h = (h_1, \ldots, h_n)$ with

$$h_j(q_1, \ldots, q_n) \triangleq \frac{1}{N} x_j^* \left(\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \right)^{-1} x_j$$

→ We have:

- **Positivity**: $h_j > 0$ for all large n a.s.
- **Monotonicity**: Let $(q_1', \ldots, q_n') \geqslant (q_1, \ldots, q_n)$, then
  - From u non-increasing,

$$\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \succeq \frac{1}{n} \sum_{i=1}^{n} u(q_i') x_i x_i^*$$

  - From Corollary 7.7.4 in Horn and Johnson, "Matrix Analysis"

$$\left(\frac{1}{n} \sum_{i=1}^{n} u(q_i') x_i x_i^* \right)^{-1} \succeq \left(\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \right)^{-1}$$

  - Hence $h_j(q_1', \ldots, q_n') \geqslant h_j(q_1, \ldots, q_n)$
- **Scalability**: Let $\alpha > 1$,
  - From ϕ non-decreasing,

$$\phi(\alpha q_i) \geqslant \phi(q_i) \text{ so that } u(\alpha q_i) \geqslant \frac{1}{\alpha} u(q_i)$$

  - With same arguments as above, $\alpha h_j(q_1, \ldots, q_n) \geqslant h_j(\alpha q_1, \ldots, \alpha q_n)$

→ h is therefore a standard interference function.

Proving existence (2)

$\rightarrow$ Now, does there exist $(q_1, \ldots, q_n)$ such that

$$q_j \geqslant h_j(q_1, \ldots, q_n) \text{ ?}$$

Proving existence (2)

→ Now, does there exist $(q_1, \ldots, q_n)$ such that

$$q_j \geqslant h_j(q_1, \ldots, q_n) \ ?$$

⇒ In general, NO!

Proving existence (2)

→ Now, does there exist $(q_1, \ldots, q_n)$ such that

$$q_j \geqslant h_j(q_1, \ldots, q_n) \ ?$$

⇒ In general, NO!

→ To pursue, we place ourselves in the random matrix regime and prove the parallel lemma:

Lemma (Concentration of quadratic forms)

$$\max_{i \leqslant n} \left\{ \left| \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i - 1 \right| \right\} \xrightarrow{\text{a.s.}} 0.$$

Proving existence (2)

$\rightarrow$ Now, does there exist $(q_1, \dots, q_n)$ such that

$$q_j \geqslant h_j(q_1, \dots, q_n) \; ?$$

$\Rightarrow$ In general, NO!

$\rightarrow$ To pursue, we place ourselves in the random matrix regime and prove the parallel lemma:

Lemma (Concentration of quadratic forms)

$$\max_{i \leqslant n} \left\{ \left| \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i - 1 \right| \right\} \xrightarrow{\text{a.s.}} 0.$$

Proof.

- Standard results on convergence of $\frac{1}{N} x_i^* \left(X_{(i)} X_{(i)}^* \right)^{-1} x_i$, $X_{(i)} = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$.
- Main difficulty lies in the control of the norm of $X_{(i)} X_{(i)}^*$. For this, we show:

Lemma (No eigenvalue close to zero)

Denote $X_{(i)} = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$. *Then, there exists* $\varepsilon > 0$ *such that*

$$\min_{i \leqslant n} \left\{ \lambda_1 \left(\frac{1}{n} X_{(i)} X_{(i)}^* \right) \right\} > \varepsilon$$

for all large n a.s. (convention: $\lambda_1 \leqslant \dots \leqslant \lambda_N$*)*

Proof of Lemma (No eigenvalue close to zero)

→ Classical result but not known in general case with y having independent entries.

- Known in Gaussian case using Nash–Poincaré inequality[16] / Haagerup approach[17]
- Extensible to distributions satisfying Poincaré–Nash inequality
- Different tools must be used in generic case.

[16] L. A. Pastur, "A simple approach to the global regime of Gaussian ensembles of random matrices", 2005.

[17] U. Haagerup, S. Thorbjornsen, "A New Application of Random Matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ Is Not a Group", 2005.

Proof of Lemma (No eigenvalue close to zero) (2)

→ We start from: λ is an eigenvalue of $\frac{1}{n}X_{(i)}X_{(i)}^*$ iif

$$\det\left(\frac{1}{n}X_{(i)}X_{(i)}^* - \lambda I_N\right) = 0$$

- Adding/Subtracting $\frac{1}{n}x_i x_i^*$ and after some manipulations, this is

$$\det Q(\lambda)\left(1 - \frac{1}{n}x_i^* Q(\lambda)^{-1}x_i\right) = 0 \text{ with } Q(x) = (\frac{1}{n}XX^* - xI_N)^{-1}.$$

- For $\lambda > \varepsilon$, $\det Q(\lambda) > 0$ for all large n a.s.

Proof of Lemma (No eigenvalue close to zero) (2)

→ We start from: λ is an eigenvalue of $\frac{1}{n}X_{(i)}X_{(i)}^*$ iif

$$\det\left(\frac{1}{n}X_{(i)}X_{(i)}^* - \lambda I_N\right) = 0$$

- Adding/Subtracting $\frac{1}{n}x_i x_i^*$ and after some manipulations, this is

$$\det Q(\lambda)\left(1 - \frac{1}{n}x_i^* Q(\lambda)^{-1}x_i\right) = 0 \text{ with } Q(x) = (\frac{1}{n}XX^* - xI_N)^{-1}.$$

- For $\lambda > \varepsilon$, $\det Q(\lambda) > 0$ for all large n a.s.
- We study $f_{n,i}(x) = 1 - \frac{1}{n}x_i^* Q(x)^{-1}x_i$.
  - For $x < 0$, from standard trace lemma, using $\sup_{i,j} E[|y_{ij}|^{8+\eta}] < \infty$

$$\max_{i\leqslant n}\left|f_{n,i}(x) - \bar{f}_n(x)\right| \xrightarrow{\text{a.s.}} 0 \text{ for } \bar{f}_n(x) \text{ the det. eq. for } f_{n,i}(x).$$

  - With $\bar{f}_n(0) = c_N < 1$, $\bar{f}_n$ and $f_{n,i}$ non-decreasing on $\mathbb{R}^-$, for any $\zeta > 0$, for all large n a.s.

$$\sup_{i\leqslant n} f_{n,i}(-\zeta) < c_N$$

Proof of Lemma (No eigenvalue close to zero) (2)

$\rightarrow$ We start from: λ is an eigenvalue of $\frac{1}{n}X_{(i)}X_{(i)}^*$ iif

$$\det\left(\frac{1}{n}X_{(i)}X_{(i)}^* - \lambda I_N\right) = 0$$

- Adding/Subtracting $\frac{1}{n}x_i x_i^*$ and after some manipulations, this is

$$\det Q(\lambda)\left(1 - \frac{1}{n}x_i^* Q(\lambda)^{-1}x_i\right) = 0 \text{ with } Q(x) = (\frac{1}{n}XX^* - xI_N)^{-1}.$$

- For $\lambda > \varepsilon$, $\det Q(\lambda) > 0$ for all large n a.s.
- We study $f_{n,i}(x) = 1 - \frac{1}{n}x_i^* Q(x)^{-1}x_i$.
  - For $x < 0$, from standard trace lemma, using $\sup_{i,j} E[|y_{ij}|^{8+\eta}] < \infty$

$$\max_{i \leqslant n}\left|f_{n,i}(x) - \bar{f}_n(x)\right| \xrightarrow{\text{a.s.}} 0 \text{ for } \bar{f}_n(x) \text{ the det. eq. for } f_{n,i}(x).$$

  - With $\bar{f}_n(0) = c_N < 1$, $\bar{f}_n$ and $f_{n,i}$ non-decreasing on $\mathbb{R}^-$, for any $\zeta > 0$, for all large n a.s.

$$\sup_{i \leqslant n} f_{n,i}(-\zeta) < c_N$$

  - For $x \geqslant 0$, previous reasoning breaks, but $\sup_{x \in [-\zeta, \varepsilon]} \|Q(x)^{-2}\| \leqslant M$ for all large n a.s. $\Rightarrow$ Boundedness of $\sup_{i \leqslant n} \sup_{x \in [-\zeta, \varepsilon]} |f_{i,n}'(x)|$.
  - By continuity of $f_{n,i}$, this implies $f_{n,i}(\varepsilon) < M(\varepsilon + \zeta) + c_N$
  - Since ζ, ε are arbitrary, independent of i,

$$\sup_{i \leqslant n} f_{n,i}(\varepsilon) < 1 \text{ for some } \varepsilon > 0.$$

  - Since $f_{n,i}$ increases, no eigenvalue is found in $[0, \varepsilon]$ for all large n a.s.

Proving existence (3)

→ Let $q_1 = \ldots = q_n = q$. Then

$$h_i(q, \ldots, q) = \frac{1}{u(q)} \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i = \frac{q}{\phi(q)} \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i$$

Proving existence (3)

→ Let $q_1 = \ldots = q_n = q$. Then

$$h_i(q, \ldots, q) = \frac{1}{u(q)} \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i = \frac{q}{\phi(q)} \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i$$

→ From previous lemma, and $\phi_\infty > 1$, we can take q large enough so that

$$\frac{1}{q} h_i(q, \ldots, q) < 1 \text{ for all } i$$

for all large n a.s.

Proving existence (3)

→ Let $q_1 = \ldots = q_n = q$. Then

$$h_i(q, \ldots, q) = \frac{1}{u(q)} \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i = \frac{q}{\phi(q)} \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i$$

→ From previous lemma, and $\phi_\infty > 1$, we can take q large enough so that

$$\frac{1}{q} h_i(q, \ldots, q) < 1 \text{ for all } i$$

for all large n a.s.

→ By the theorem on standard interference function, this proves existence of a solution to

$$q_i = h_i(q_1, \ldots, q_n)$$

for all i.

→ Since the q_i define uniquely Z, this defines a solution to

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

Proving uniqueness (1)

→ We now take $(d_1, \ldots, d_n)$, $d_i = \frac{1}{N} x_i^* \hat{C}_N^{-1} x_i$ with $\hat{C}_N$ one of the solutions

- We assume $d_1 \leqslant \ldots \leqslant d_n$;
- We also define $D = \mathrm{diag}(u(d_1), \ldots, u(d_n))$.

→ $u(s)$ is non-increasing, so

$$XDX^* \succeq u(d_n) XX^*$$

so that

$$\frac{1}{u(d_n)} \hat{S}_N^{-1} \succeq \hat{C}_N^{-1}$$

and then

$$\frac{1}{u(d_n)} \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n \geqslant d_n$$

from which

$$\phi(d_n) \leqslant \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n$$

→ Proceeding similarly for d_1, and using ϕ non-decreasing, we conclude, for all i

$$\frac{1}{N} x_1^* \hat{S}_N^{-1} x_1 \leqslant \phi(d_1) \leqslant \phi(d_i) \leqslant \phi(d_n) \leqslant \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n$$

Proving uniqueness (2)

→ From our main Lemma, for all large n a.s., for all $\varepsilon > 0$,

$$\max_{i \leqslant n} |\phi(d_i) - 1| \leqslant \varepsilon$$

→ But $\phi_\infty > 1$ so that ϕ invertible in a neighborhood of 1, and

$$\max_{i \leqslant n} |d_i - \phi^{-1}(1)| \leqslant \varepsilon'$$

Proving uniqueness (2)

→ From our main Lemma, for all large n a.s., for all $\varepsilon > 0$,

$$\max_{i \leqslant n} |\phi(d_i) - 1| \leqslant \varepsilon$$

→ But $\phi_\infty > 1$ so that ϕ invertible in a neighborhood of 1, and

$$\max_{i \leqslant n} |d_i - \phi^{-1}(1)| \leqslant \varepsilon'$$

→ Take $(d_i^{(1)})_{i=1}^n$ and $(d_i^{(2)})_{i=1}^n$ two solutions. Then we can find $\alpha > 1$ such that

$$\alpha d_k^{(1)} = d_k^{(2)} \text{ and } \alpha d_i^{(1)} \geqslant d_i^{(2)}, \ i \neq k$$

- From the above,
$$\alpha = d_k^{(1)} / d_k^{(2)} \leqslant 1 + \varepsilon''$$
- Hence, setting ε properly
$$\alpha d_i^{(1)} < \phi^{-1}(\phi_\infty -)$$
so that, from ϕ increasing on $[0, \phi^{-1}(\phi_\infty -))$,
$$\phi(\alpha d_i^{(1)}) > \phi(d_i^{(1)}), \text{ i.e. } \alpha u(\alpha d_i^{(1)}) > u(d_i^{(1)}) \text{ for all } i$$
- Hence $\alpha h_i(d_1^{(1)}, \ldots, d_n^{(1)}) > h_i(\alpha d_1^{(1)}, \ldots, \alpha d_n^{(1)})$.
- With monotonicity of h,
$$d_k^{(2)} = h_k(d_1^{(2)}, \ldots, d_n^{(2)}) \leqslant h_k(\alpha d_1^{(1)}, \ldots, \alpha d_n^{(1)}) < \alpha h_i(d_1^{(1)}, \ldots, d_n^{(1)}) = \alpha d_k^{(1)}$$
a contradiction.

Outline

Finishing the proof

→ Recall that

$$\max_{i \leqslant n} \left| d_i - \phi^{-1}(1) \right| \xrightarrow{\text{a.s.}} 0$$

so that

$$\max_{i \leqslant n} \left| u(d_i) - \frac{1}{\phi^{-1}(1)} \right| \xrightarrow{\text{a.s.}} 0$$

(note that $\phi^{-1}(1)u(\phi^{-1}(1)) = 1$)

Finishing the proof

→ Recall that

$$\max_{i\leqslant n}\left|d_i-\phi^{-1}(1)\right|\xrightarrow{\text{a.s.}}0$$

so that

$$\max_{i\leqslant n}\left|u(d_i)-\frac{1}{\phi^{-1}(1)}\right|\xrightarrow{\text{a.s.}}0$$

(note that $\phi^{-1}(1)u(\phi^{-1}(1))=1$)

→ We then conclude with

$$\min_{i\leqslant n}\left\{u(d_i)-\frac{1}{\phi^{-1}(1)}\right\}\frac{1}{n}XX^*\preceq\frac{1}{n}\sum_{i=1}^{n}\left(u(d_i)-\frac{1}{\phi^{-1}(1)}\right)x_ix_i^*\preceq\max_{i\leqslant n}\left\{u(d_i)-\frac{1}{\phi^{-1}(1)}\right\}\frac{1}{n}XX^*$$

which entails, along with the a.s. boundedness of $\|\frac{1}{n}XX^*\|$,

$$\left\|\hat{C}_N-\frac{1}{\phi^{-1}(1)}\hat{S}_N\right\|\xrightarrow{\text{a.s.}}0$$