A Concentration of Measure and Random Matrix Approach to Large Dimensional Robust Statistics

author names withheld

Editor: Under Review for COLT 2020

Abstract

This article studies the *robust covariance matrix estimation* of a data collection $X = (x_1, \ldots, x_n)$ with $x_i = \sqrt{\tau}_i z_i + m$, where $z_i \in \mathbb{R}^p$ is a *concentrated vector* (e.g., an elliptical random vector), $m \in \mathbb{R}^p$ a deterministic signal and $\tau_i \in \mathbb{R}$ a scalar perturbation of possibly large amplitude, under the assumption where both n and p are large. This estimator is defined as the fixed point of a function which we show is contracting for a so-called *stable semi-metric*. We exploit this semi-metric along with concentration of measure arguments to prove the existence and uniqueness of the robust estimator as well as evaluate its limiting spectral distribution.

Keywords: Robust Estimation – Concentration of Measure – Random Matrix Theory.

1. Introduction

Robust estimators of covariance (or scatter) are necessary ersatz for he classical sample covariance when the dataset $X = (x_1, \ldots, x_n)$ present some diverging statistical properties, such as unbounded second moments of the x_i's. We study here the M-estimator of scatter $\hat{C}$ initially introduced in (Huber, 1964) defined as the solution (if it exists) to the following fixed point equation:

$$\hat{C} = \frac{1}{n} \sum_{i=1}^{n} u \left(\frac{1}{n} x_i^T (\hat{C} + \gamma I_p)^{-1} x_i \right) x_i x_i^T, \tag{1}$$

where $\gamma > 0$ is a regularization parameter and $u : \mathbb{R}^+ \to \mathbb{R}^+$ a mapping that tends to zero at $+\infty$, and whose object is to control outlying data. The literature in this domain has so far divided the study of $\hat{C}$ into (i) a first exploration of conditions for its existence and uniqueness as a *deterministic* solution to (1) (e.g., (Huber, 1964; Maronna, 1976; Tyler, 1987)) and (ii) an independent analysis of its statistical properties when seen as a random object (in the large n regime (Chitour and Pascal, 2008) or in the large n, p regime (Couillet and McKay, 2014; Zhang et al., 2014)).

In the present article, we study the large dimensional (n, p large) spectral properties of $\hat{C}$ in an original joint framework based on concentration of measure theory and on a new stable semi-metric argument. This joint framework has the multiple advantages of (i) relaxing the assumptions of independence in the entries of x_i made in (Couillet and McKay, 2014; Zhang et al., 2014), (ii) consistently articulating the "Lipschitz and stable semi-metric" properties of the model to propagate concentration. The major tool allowing this articulation is provided in the paper by Theorem 19. In passing, we further relax some of the assumptions made in the above articles, particularly on the constraints on the mapping u. Specifically, we require here that u be 1-Lipschitz for the stable semi-metric, to be introduced next. This semi-metric naturally arises when studying the so-called resolvent $(\hat{C} + \gamma I_p)^{-1}$ of $\hat{C}$, which is at the core of our large p, n random matrix analysis of $\hat{C}$.

© 2020 .

In detail, the data model under study decomposes as $x_i = \sqrt{\tau_i} z_i + m$ where the $z_1, \ldots, z_n$ are independent random vectors satisfying a concentration of measure hypothesis (in particular, the z_i's could arise from a very generic generative model, such as $z_i = h(\tilde{z}_i)$ for $\tilde{z}_i \sim \mathcal{N}(0, I_p)$ and $h : \mathbb{R}^p \to \mathbb{R}^p$ 1-Lipschitz), m is a deterministic vector (a signal) and τ_i are arbitrary (possibly large) deterministic values.[1] This setting naturally arises in many engineering applications, such as in antenna array processing where the τ_i's model noise impulsiveness and m is a sought-for signal (Ovarlez et al., 2011) or in statistical finance where x_i's model asset returns with high volatility and m is the market leading direction (Yang et al., 2014).

2. Preliminaries for the study of the resolvent

Let us note $\mathbb{R}^+ = \{x \in \mathbb{R}, x > 0\}$; $\mathcal{M}_{p,n}$, the set of real matrices of size $p \times n$, endowed with the spectral norm $\|M\| = \sup\{|Mu|, u \in \mathbb{R}^n, \|u\| \leq 1\}$, for $M \in \mathcal{M}_n$ and the Frobenius norm $\|M\|_F = \sqrt{\sum_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}} M_{i,j}}$. We further note $\mathcal{D}_n = \{\Delta \in \mathcal{M}_n \mid i \neq j \Leftrightarrow \Delta_{i,j} = 0\}$, the set of diagonal matrices endowed with the spectral norm of $\mathcal{M}_n$. Given $\Delta \in \mathcal{D}_n$, we let $\Delta_1, \ldots, \Delta_n \in \mathbb{R}$, be its diagonal elements, $\Delta = \mathrm{Diag}(\Delta_i)_{1 \leq i \leq n}$ so that $\|\Delta\| = \sup\{|\Delta_i|, i \in [n]\}$ (where $[n] = \{1, \ldots, n\}$). We let $\mathcal{S}_p$ be the set of symmetric matrices of size p and $\mathcal{S}_p^+$ the set of symmetric nonnegative matrices. Given $S, T \in \mathcal{S}_p$, we denote $S \leq T$ iif $T - S \in \mathcal{S}_p^+$. We will extensively work with the set $(\mathcal{S}_p)^n$ that we will note for simplicity $\mathcal{S}_p^n$. Given $S \in \mathcal{S}_p^n$, we finally denote $S_1, \ldots, S_n \in \mathcal{S}_p$ its components.

Given two sequences $(u_s)_{s \in \mathbb{N}}$ and $(v_s)_{s \in \mathbb{N}}$, we will write $u_s \leq O(v_s)$ to signify that there exists a constant $K > 0$ such that for all $s \in \mathbb{N}$, $u_s \leq K v_s$. We will also use the notations:

$$u_s \geq O(v_s) \Leftrightarrow \exists K > 0, \forall s \in \mathbb{N}, u_s \leq v_s; \qquad u_s \sim O(v_s) \Leftrightarrow O(v_s) \leq u_s \leq O(v_s).$$

We extend those characterizations to diagonal matrices: given $\Delta \in \mathcal{D}_n^+$, $\Delta \leq O(1)$ indicates that $\|\Delta\| \leq O(1)$ while $\Delta \geq O(1)$ means that $\|\frac{1}{\Delta}\| \leq O(1)$ and $\Delta \sim O(1)$ means that $O(1) \leq \Delta \leq O(1)$.

The different assumptions leading to our major results are presented progressively throughout the paper so that the reader easily understands easily their importance and direct implications. A full recollection of all these assumptions is given at the beginning of the appendix.

2.1. The resolvent behind robust statistics and its contracting properties

Given $\gamma > 0$ and $S \in \mathcal{S}_p^n$, we introduce the resolvent function at the core of our study :

$$\begin{array}{rccc} Q_\gamma : & \mathcal{S}_p^n \times \mathcal{D}_n^+ & \longrightarrow & \mathcal{M}_p \\ & (S, \Delta) & \longmapsto & \left(\frac{1}{n} \sum_{i=1}^n \Delta_i S_i + \gamma I_p \right)^{-1} . \end{array}$$

Given a dataset $X = (x_1, \ldots, x_n) \in \mathcal{M}_{p,n}$, if we note $X \cdot X^T = (x_i x_i^T)_{1 \leq i \leq n} \in \mathcal{S}_p^n$, the robust estimation of the scatter matrix then reads (if well defined):

$$\hat{C} = \gamma I_p + \frac{1}{n} X u(\hat{\Delta}) X^T \qquad \text{with} \qquad \hat{\Delta} = \mathrm{Diag}\left(\frac{1}{n} x_i Q_\gamma(X \cdot X^T, u(\hat{\Delta})) x_i) x_i^T \right)_{1 \leq i \leq n} . \tag{2}$$

1. We may alternatively assume the τ_i random independent of $Z = (z_1, \ldots, z_n)$

In the following, we will denote for simplicity $Q_\gamma^X \equiv Q_\gamma(X \cdot X^T, u(\hat{\Delta}))$. To understand the behavior (structural, spectral, statistical) of $\hat{C}$, one needs first to try and understand the resolvent $Q_\gamma(S, \Delta)$ for general $S \in \mathcal{S}_p^n$ and $\Delta \in \mathcal{D}_n^+$. Specifically, we list in this subsection its contracting properties.

We will sometimes allow ourselves to omit the index γ since this parameter will rarely change.

Lemma 1 *Given $\gamma > 0$, $S \in \mathcal{S}_p^n$, $M \in \mathcal{M}_{p,n}$ and $\Delta \in \mathcal{D}_n^+$:*

$$\|Q_\gamma(S,\Delta)\| \leq \frac{1}{\gamma}; \quad \left\| \frac{1}{\sqrt{n}} Q_\gamma(M \cdot M^T, \Delta) M \Delta^{\frac{1}{2}} \right\| \leq \frac{1}{\sqrt{\gamma}}; \quad \left\| \frac{1}{n} Q_\gamma(S,\Delta) \sum_{l=1}^k \Delta_l S_l \right\| \leq 1$$

Given $M \in \mathcal{M}_{p,n}$, and $S \in \mathcal{S}_p^n$, let us introduce the mapping $I_\gamma : \mathcal{S}_p^n \times \mathcal{D}_n^+ \to \mathcal{D}_n^+$ satisfying for all $S \in \mathcal{S}_p^n$ and $\Delta \in \mathcal{D}_n^+$:

$$I(S,\Delta) = \operatorname{Diag}\left(\frac{1}{n} \operatorname{Tr}\left(S_i Q_\gamma(S,\Delta) \right) \right)_{1 \leq i \leq n}$$

With the notation $I_\gamma^X(\Delta) \equiv I(X \cdot X^T, \Delta)$ the fixed point $\hat{\Delta}$ defined in (2) satisfies $\hat{\Delta} = I_\gamma^X(u(\hat{\Delta}))$.

Lemma 2 *Given $S \in \mathcal{S}_p^n$ and $\Delta, \Delta' \in \mathcal{D}_n^+$, we can bound (we omitted the index γ):*

$$\left\| \frac{I(S,\Delta) - I(S,\Delta')}{\sqrt{I(S,\Delta) I(S,\Delta')}} \right\| < \sup\left\{ \|1 - \gamma Q_\gamma(S,\Delta)\|, \Delta \in \mathcal{D}_n^+, \right\} \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta \Delta'}} \right\|.$$

The proof of this lemma is left to the appendix: it is a simple application of the Cauchy-Schwartz inequality. If one sees the term $\left\| \frac{\Delta - \Delta'}{\sqrt{\Delta\Delta'}} \right\|$ as a distance between Δ and Δ', then Lemma 2 sets the 1-Lipschitz character of $I(S, \cdot) : \Delta \mapsto I(S, \Delta)$, which is a fundamental property in what follows. We present in next subsection a precise description of such functions that will be called *stable mappings*.

2.2. The stable semi-metric

The stable semi-metric which we define here is a convenient object that allows us to set Piquard-like fixed point theorems. It has a capital importance to set the existence and uniqueness of $\hat{C}$ but also to demonstrate rapidly some random matrix results, such as the estimation of the spectral distribution of sample covariance matrices with a variance profile (provided in the appendix).

Definition 3 *We call the* stable semi-metric *on $\mathcal{D}_n^+ = \{D \in \mathcal{D}_n, \forall i \in [n],\ D_i > 0\}$ the function:*

$$\forall \Delta, \Delta' \in \mathcal{D}_n^+ : \quad d_s(\Delta, \Delta') \equiv \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta \Delta'}} \right\|. \tag{3}$$

In particular, this semi-metric can be defined on $\mathbb{R}^+$, identifying $\mathbb{R}^+$ with $\mathcal{D}_1^+$. The function d_s is not a metric because it does not verify the triangular inequality: a counter-example is given in the appendix. The semi-metric d_s has some very interesting stability results.

Property 1 *Given* $\Delta, \Delta' \in \mathcal{D}_n^+$ *and* $\Lambda \in \mathcal{D}_n^+$*:*

$$d_s\left(\Lambda\Delta, \Lambda\Delta'\right) = d_s\left(\Delta, \Delta'\right) \qquad \textit{and} \qquad d_s\left(\Delta^{-1}, \Delta'^{-1}\right) = d_s\left(\Delta, \Delta'\right).$$

Definition 4 *The set of* 1*-Lipschitz functions for the stable semi-metric is called the stable class. We denote it:*

$$\mathcal{S}\left(\mathcal{D}_n^+\right) \equiv \left\{f : \mathcal{D}_n^+ \to \mathcal{D}_n^+ \mid \forall \Delta, \Delta' \in \mathcal{D}_n^+,\ \Delta \neq \Delta' :\ d_s(f(\Delta), f(\Delta')) \leq d_s(\Delta, \Delta')\right\}.$$

The elements of $\mathcal{S}\left(\mathcal{D}_n^+\right)$ *are called the stable mappings.*

This class has a very simple interpretation when $n = 1$. Given a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ we introduce two functions $f_/, f_\cdot : \mathbb{R}^+ \to \mathbb{R}^+$ that will help us to characterize the stable class:

$$f_/\ :\ x \mapsto \frac{f(x)}{x} \qquad \text{and} \qquad f_\cdot\ :\ x \mapsto xf(x).$$

Property 2 *A function* $f : \mathbb{R}^+ \to \mathbb{R}^+$ *is a stable mapping if and only if* $f_/$ *is nonincreasing and* $f_\cdot$ *is nondecreasing.*

We leave the proof to the appendix: it directly unfolds from the definition (3). Finally, we provide the properties which justify why we call $\mathcal{S}\left(\mathcal{D}_n^+\right)$ a *stable* class: this class indeed satisfies far more stability properties than the usual Lipschitz mappings (for a given norm).

Property 3 *Given* $\Lambda \in \mathcal{D}_n^+$ *and* $f, g \in \mathcal{S}(\mathcal{D}_n^+)$*:*

$$\Lambda f \in \mathcal{S}(\mathcal{D}_n^+) \qquad \frac{1}{f} \in \mathcal{S}(\mathcal{D}_n^+) \qquad f \circ g \in \mathcal{S}(\mathcal{D}_n^+) \qquad f + g \in \mathcal{S}(\mathcal{D}_n^+).$$

2.3. Fixed Point theorem for stable mappings

The Picard fixed point theorem states that a contracting function on a complete space admits a unique fixed point. The extension of this result to contracting mappings on $\mathcal{D}_n^+$, for the semi-metric d_s, is not obvious: first because d_s does not verify the triangular inequality and second because the completeness needs be proven. Most of the proofs here are left to the appendix.

Property 4 *The semi-metric space* $(\mathcal{D}_n^+, d_s)$ *is complete.*

Theorem 5 *Given a mapping* $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$*, contracting for the stable semi-metric* d_s *and bounded from below and above, there exists a unique fixed point* $\Delta^* \in \mathcal{D}_n^+$ *satisfying* $\Delta^* = f(\Delta^*)$*.*

Let us now state our fixed point results for $I_\gamma(S, \cdot) = \mathrm{Diag}(\mathrm{Tr}(S_i Q(S, \cdot)))_{1 \leq i \leq n}$ for $S \in \mathcal{M}_p^k$. We need for that a preliminary lemma.

Lemma 6 *Given* $S \in \mathcal{S}_p^n$ *and a function* $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ *bounded by* $f_0 \in \mathcal{D}_k^+$*,*

$$\forall \Delta \in \mathcal{D}_n^+ :\ \frac{1}{f_0 \left\|S\right\| + \gamma} \leq \left\|Q(S, f(\Delta))\right\| \leq \frac{1}{\gamma} \qquad \textit{where} \qquad \left\|S\right\| \equiv \frac{1}{n}\left\|\sum_{i=1}^n S_a\right\|.$$

Combined with Lemma 2, this result allows us to build a family of contracting stable mappings with the composition $\tilde{I}(S,\cdot) \circ f$ when $f \in \mathcal{S}(\mathcal{D}_n^+)$ is bounded from below. We thus obtain the following corollary to Theorem 5.

Corollary 7 *Given $f, g \in \mathcal{S}(\mathcal{D}_n^+)$ with f bounded, and a family of positive definite symmetric matrices $S = (S_1, \dots, S_k) \in \mathcal{S}_p^n$, the fixed point equation*

$$\Delta = g\left(I^S(f(\Delta))\right)$$

admits a unique solution in $\mathcal{D}_n^+$.

We will thus suppose from here on that u is a stable function to be able to use Corollary 7 and set the existence and uniqueness of $\hat{\Delta}$ and $\hat{C}$ as defined in (2).

Assumption 1 $u \in \mathcal{S}(\mathbb{R}^+)$*, and there exists* $u^\infty > 0$ *such that* $\forall t \in \mathbb{R}^+,\ u(t) \leq u^\infty$.

Proposition 8 *For $X \in \mathcal{M}_{p,n}$, there exists a unique diagonal matrix $\hat{\Delta} \in \mathcal{D}_n^+$ such that*

$$\hat{\Delta} = I^X\left(u(\hat{\Delta})\right).$$

2.4. The concentration of measure framework

Having proved the existence and uniqueness of $\hat{C}$, we now introduce statistical conditions on X to study $\hat{C}$ in the large dimensional $n, p \to \infty$ limit. We first define n p-dimensional random vectors $(z_1, \dots, z_n) \in \mathbb{R}^p$.

Assumption 2 *The random vectors $z_1, \dots, z_n$ are all independent.*

We denote their means $\mu_i \equiv \mathbb{E}[z_i] \in \mathbb{R}^p$ and their covariance matrices $\Sigma_i = \mathbb{E}[z_i z_i^T] - \mu_i \mu_i^T \in \mathcal{M}_p$. In the following, the number of data n and their size p must be thought of as large integers of the same order of magnitude. To place ourselves under this setting we suppose that all studied objects depend on an underlying asymptotic quantity s tending to ∞ and that $n = n(s)$ and $p = p(s)$ tend to ∞ as $s \to \infty$. Unless stated otherwise, we implicitly assume a large s limit; for instance, $C \leq O(s)$ means that $\forall s > 0:\ |C| = |C(s)| \leq Ks$ for a constant $K > 0$ (recall the notations at the beginning of Section 2). All quantities diverging to $+\infty$ or converging to 0 will be call "asymptotic quantities" and should be distinguished from the "constants" that stay unmodified when s increases.

Assumption 3 $p \sim O(n)$.

The matrix $Z = [z_1, \dots, z_n] \in \mathcal{M}_{p,n}$ depends on s under our formalism but we do not further specify this dependence to simplify the notations.

Let us now introduce the fundamental definition of a so-called *concentrated random vector* that will allow us to set our estimations and concentration rates. The global idea is that a concentrated vector $Z \in E$ is not "concentrated" around any point (visualize for instance Gaussian vectors which, while concentrated vectors, rather lie close to a sphere) but has concentrated "observations", that is random outputs $f(Z)$ for *any* 1-Lipschitz map $f : E \to \mathbb{R}$.

Definition 9 *Given a sequence of normed vector spaces $(E_s, \|\cdot\|_s)_{s\geq 0}$, a sequence of random vectors $(Z_s)_{s\geq 0} \in \prod_{s\geq 0} E_s$, a sequence of positive reals $(\sigma_s)_{s\geq 0} \in \mathbb{R}_+^{\mathbb{N}}$ and a parameter $q > 0$, we say that Z_s is q-*exponentially concentrated *with an* observable diameter *of order $O(\sigma_s)$ iff, for any sequence of 1-Lipschitz functions $f_s : E_s \to \mathbb{R}$, one of the following two equivalent assertions is verified (for the norms $\|\cdot\|_s$) :*

- *there exist $c_s \leq O(\sigma_s)$ and $C > 0$, such that, for all $s \in \mathbb{N}$, and for all $t > 0$,*
$$\mathbb{P}\left(\left|f_s(Z_s) - f_s(Z_s')\right| \geq t\right) \leq Ce^{(t/c_s)^q}$$
- *there exist $c_s \leq O(\sigma_s)$ and $C > 0$ such that, for all $s \in \mathbb{N}$ and for all $t > 0$,*
$$\mathbb{P}\left(|f_s(Z_s) - \mathbb{E}[f_s(Z_s)]| \geq t\right) \leq Ce^{(t/c_s)^q}$$

where Z_s' is an independent copy of Z_s. We denote in that case $Z_s \propto \mathcal{E}_q(\sigma_s)$ (or more simply $Z \propto \mathcal{E}_q(\sigma)$). If $\sigma \leq O(1)$, one can further write $Z_s \propto \mathcal{E}_q$.

The essential result that gives motivation to the definition is the concentration of Gaussian vectors.

Theorem 10 (Ledoux (2005)) *Given a deterministic vector $\mu \in \mathbb{R}^p$, if $Z \sim \mathcal{N}(\mu, I_p)$ then $Z \propto \mathcal{E}_2$*

Now that we have a concentrated vector, let us give four important properties to keep in mind when dealing with them. First, the class of random vectors is stable through Lipschitz maps:

Proposition 11 *Given two normed vector spaces $(E_1, \|\cdot\|_1)$ and $(E_2, \|\cdot\|_2)$, a random vector $Z \in E_1$, two sequences $\sigma, \lambda \in \mathbb{R}_+^{\mathbb{N}}$ and a $O(\lambda)$-Lipschitz function $\phi : E_1 \to E_2$:*[2]

$$Z \propto \mathcal{E}_q(\sigma) \qquad\Longrightarrow\qquad \Phi(Z) \propto \mathcal{E}_q(\lambda\sigma).$$

The concentration of a random vector can be alternatively understood through a controlled decreasing rate of the moments of its observations.

Proposition 12 *Let $Z \in E$. Then $Z \propto \mathcal{E}_q(\sigma)$ iff there exists a constant $C > 0$ such that, for any 1-Lipschitz mapping $f : E \to \mathbb{R}$,*

$$\forall r > q : \ \mathbb{E}\left[|f(Z) - \mathbb{E}[f(Z)]|^r\right] \leq C\left(\frac{r}{q}\right)^{\frac{r}{q}} \sigma^r.$$

Standard operations (addition, product) on concentrated random variables can be easily expressed through an intuitive "distributive rule" between concentration rates and expectations. We will mostly focus here on the case of scalar concentrated random vectors for which we introduce more telling notations: when $Z \in \mathbb{R}$ is a random *scalar* and satisfies $Z \propto \mathcal{E}_q(\sigma)$, we will use the notation $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$ if $|\tilde{Z} - \mathbb{E}Z| \leq O(\sigma)$ (of course, in particular $Z \in \mathbb{E}Z \pm \mathcal{E}_q(\sigma)$).

Proposition 13 *Let $Z_1, Z_2 \in \mathbb{R}$ be two random variables, $\sigma_1, \sigma_2 \in \mathbb{R}_+^{\mathbb{N}}$ two sequences of positive reals and $\tilde{Z}_1, \tilde{Z}_2 \in \mathbb{R}^{\mathbb{N}}$ two sequences of scalars. Then, if $Z_1 \in \tilde{Z}_1 \pm \mathcal{E}_q(\sigma_1)$ and $Z_2 \in \tilde{Z}_2 \pm \mathcal{E}_q(\sigma_2)$,*

$$Z_1 + Z_2 \in \tilde{Z}_1 + \tilde{Z}_2 \pm \mathcal{E}_q(\sigma_1 + \sigma_2); \qquad Z_1Z_2 \in \tilde{Z}_1\tilde{Z}_2 \pm \mathcal{E}_q(\sigma_1|\tilde{Z}_2| + \sigma_2|\tilde{Z}_1|) + \mathcal{E}_{q/2}(\sigma_1\sigma_2),$$

More over, $\forall f : \mathbb{R} \to \mathbb{R}$, 1-Lipschitz: $\quad f(Z_1) \propto f(\tilde{Z}_1) \pm \mathcal{E}_q(\sigma_1)$

2. The statement "ϕ is $O(\lambda)$-Lipschitz" means here that there exists $K \leq O(1)$ such that, for all $s \in \mathbb{N}$, ϕ_s is $(K\lambda_s)$-Lipschitz.

Concentration inequalities for operations on concentrated *vectors* express similarly but will not be needed in this work (more information is available in (Louart and Couillet, 2019)). We complete this short probabilistic introduction of concentration of measure theory with four results on the concentration of the norm; these results will be used continuously in the following to track the size of the various objects under study. We provide them here in the case $q = 2$, but similar inequalities exists in the general setting.

Lemma 14 *Let $Y \in \mathbb{R}^p$. Then, if $Y \propto \mathcal{E}_2$ in $(\mathbb{R}^p, \|\cdot\|)$,*

- $\|Y - \mathbb{E}Y\| \propto \mathcal{E}_2\left(p^{\frac{1}{2}}\right)$ *and* $\mathbb{E}\left[\|Y - \mathbb{E}Y\|\right] \leq O\left(\sqrt{p}\right)$
- $\|Y - \mathbb{E}Y\|_\infty \propto \mathcal{E}_2\left(\sqrt{\log p}\right)$ *and* $\mathbb{E}\left[\|Y - \mathbb{E}Y\|_\infty\right] \leq O\left(\sqrt{\log p}\right)$

and, conversely, if $\|Y - \mathbb{E}Y\| \propto \mathcal{E}_2$, then $Y \propto \mathcal{E}_2$. Let $Z \in \mathcal{M}_{p,n}$ be a random matrix. Then, if $Z \propto \mathcal{E}_2$ in $(\mathbb{R}^p, \|\cdot\|_F)$,

- $\|Y - \mathbb{E}Y\|_F \propto \mathcal{E}_2\left(\sqrt{pn}\right)$ *and* $\mathbb{E}\left[\|Y - \mathbb{E}Y\|_\infty\right] \leq O\left(\sqrt{pn}\right)$
- $\|Y - \mathbb{E}Y\| \propto \mathcal{E}_2\left(\sqrt{p+n}\right)$ *and* $\mathbb{E}\left[\|Y - \mathbb{E}Y\|\right] \leq O\left(\sqrt{p+n}\right)$.

In the following, we will thus assume that $Z = (z_1, \ldots, z_n)$ is concentrated.

Assumption 4 $Z \propto \mathcal{E}_2$.

We know from Assumption 4 that for all $i \in [n]$, $\|\Sigma_i\| \leq O(1)$ since $\forall u \in \mathbb{R}^p$ such that $\|u\| \leq 1$, $u^T \Sigma_i u = \mathbb{E}[u^T z_i z_i^T u - \mathbb{E}[u^T z_i]\mathbb{E}[z_i^T u]] \leq O(1)$ (from Propositions 12 and 13). But we also need to bound μ_i to avoid unbounded norms on $E[z_i z_i^T] = \Sigma_i + \mu_i \mu_i^T$.

Assumption 5 *For all $i \in \{1, \ldots, n\}$, $\|\mu_i\| \leq O(\sqrt{n})$.*

2.5. Deterministic equivalent of the resolvent

The resolvent $Q_\gamma^Z(I_n) = Q(Z \cdot Z^T, I_n)$ is a random matrix which exhibits interesting properties to understand the statistics of Z and more precisely its spectral behavior. In particular, the singular values of Z strongly relate to the well-known *Stieltjes transform* $m_Z(z) = \frac{1}{p}\operatorname{Tr}(Q_{-z}^Z(I_n)$. The function $m_Z(z)$ has been extensively studied in (Louart and Couillet, 2019) when Z is concentrated with identically distributed columns. As shown next, no major change occurs when the columns of Z have different distributions. We will study $Q_{-z}^Z(I_n)$ in the specific case where $z > 0$ and will denote $Q = Q_z^Z(I_n)$ for simplicity.

It can be shown that Q is a $\frac{2}{z^{3/2}\sqrt{n}}$-Lipschitz transformation of Z and, therefore, assuming that $\frac{1}{z} \leq O(1)$, we can deduce $Q \propto \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right)$. There exists an easily computed deterministic matrix $\tilde{Q}$, called the *deterministic equivalent* of Q such that $\|\mathbb{E}[Q] - \tilde{Q}\| \leq O(1/\sqrt{n})$. Matrix $\tilde{Q}$ thus verifies that, for any deterministic matrix $A \in \mathcal{M}_p$, such that $\|A\|_1 \equiv \operatorname{Tr}(\sqrt{AA^T}) \leq O(1)$ ($\|\cdot\|_1$ is the dual norm of $\|\cdot\|$ for the canonical scalar product on $\mathcal{M}_p$, $\langle\cdot,\cdot\rangle : A, B \mapsto \operatorname{Tr}(AB^T)$),

$$\operatorname{Tr}(AQ) \in \operatorname{Tr}(A\tilde{Q}) \pm \mathcal{E}_q\left(\frac{1}{\sqrt{n}}\right),$$

with the notation of concentrated random variables introduced before Proposition 13.

The deterministic equivalent of Q^Z is defined thanks to a diagonal matrix $\Lambda \in \mathcal{D}_n^+$.

Proposition 15 (Louart and Couillet (2019)) *For any $S \in \mathcal{S}_p^n$, the mapping $\Delta \mapsto \check{I}(S, \Delta)$ satisfying*

$$\check{I}(S,\Delta) = \frac{1}{n}\operatorname{Diag}(\operatorname{Tr}(S_i Q(S_{-i},\Delta))_{1\leq i\leq n}, \quad \text{for } S_{-i} = (S_1,\dots,S_{i-1},0,S_{i+1},\dots,S_n) \tag{4}$$

is stable and for any $\Delta \in \mathcal{D}_n^+$, the equation $\Lambda = \check{I}(C, \frac{I_n}{I_n+\Lambda})$ admits a unique solution $\Lambda^C \in \mathcal{D}_n^+$.

Proof The stability of $\check{I}(S,\cdot)$ is proven the same way as the stability of $I(S,\cdot)$ in the proof of Lemma 2. Then we apply an analog result to Corollary 7 (replacing I by $\check{I}$) with the mapping $f : \Lambda \mapsto \frac{I_n}{I_n+\Lambda}$ which is stable and bounded from above by $I_n \in \mathcal{D}_n^+$, and for $S_i = \frac{1}{n}C_i$ (for $i \in [n]$) to obtain the existence and uniqueness of $\Lambda \in \mathcal{D}_n^+$ satisfying $\Lambda = \check{I}_z(C, f(\Lambda))$. ■

The fixed point equation $\Lambda = \check{I}(C, \frac{I_n}{I_n+\Lambda})$ allows us to compute Λ iteratively. The deterministic equivalent $\tilde{Q}$ of Q^Z is then easily computed and is defined as follows.

Theorem 16 (Louart and Couillet (2019)) *Let $A \in \mathcal{M}_p$ be deterministic and such that $\|A\|_1 \leq O(1)$. Then,*

$$\operatorname{Tr}(AQ_z^Z) \in \operatorname{Tr}\left(AQ_z\left(C, \frac{I_n}{I_n+\Lambda^C}\right)\right) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right)$$

This theorem allows us to estimate the Stieltjes transform of the spectral distribution of $\hat{C}$ given at the end of the paper, but we need first the next corollary to predict the asymptotic behavior of $\hat{\Delta}$.

Corollary 17 *For all $\Delta \in \mathcal{D}_n^+$ with $\|\Delta\| \leq O(1)$, we have $\left\|\mathbb{E}[I^Z(\Delta)] - \frac{\Lambda^C(\Delta)}{I_n+\Delta\Lambda^C(\Delta)}\right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right)$, where:*

$$\Lambda^C(\Delta) = \check{I}\left(C, \frac{\Delta}{I_n + \Delta\Lambda^C(\Delta)}\right) \quad \textit{(see (4) for the definition of } \check{I}\textit{)} \tag{5}$$

($\Lambda^C(\Delta)/\Delta$ plays the same role as the diagonal matrix Λ^C defined in Proposition 15 to define the deterministic equivalent of Q^Z instead of $Q_z^{Z\Delta^{1/2}}$; we have indeed $I^Z(\Delta) = I^{Z\Delta^{1/2}}/\Delta$).

Proposition 18 *For any $S \in \mathcal{S}_p^n$, the mapping Λ^S defined in (5) is stable and satisfies the bounds*

$$\frac{\operatorname{Diag}(\frac{1}{n}\operatorname{Tr} S_i)_{1\leq i\leq n}}{z + \frac{1}{n}\|\sum S_i\Delta_i\|} \leq \Lambda^S(\Delta) \leq \frac{1}{z}\operatorname{Diag}\left(\frac{1}{n}\operatorname{Tr} S_i\right)_{1\leq i\leq n}.$$

3. Robust estimation of the scatter matrix

3.1. Setting and strategy of the proof

Having set up the necessary tools and preliminary results, we now concentrate on our target objective. Let $x_i = \sqrt{\tau_i}z_i + m$, $1 \leq i \leq n$, where τ_i is a deterministic positive variable, $m \in \mathbb{R}^p$ is a deterministic vector, and $z_1, \dots, z_n$ are the random vectors presented in the previous section. For $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$, we write $X = Z\tau^{\frac{1}{2}} + m\mathbb{1}^T$ where $\tau \equiv \operatorname{Diag}(\tau_i)_{1\leq i\leq n} \in \mathcal{D}_n^+$ and $\mathbb{1} \equiv (1, \dots, 1) \in \mathbb{R}^n$. The basic idea to estimate $\hat{\Delta}$, as a solution to the fixed point equation $\hat{\Delta} = I^X(u(\hat{\Delta}))$, consists in retrieving a deterministic equivalent also solution to a (now deterministic) fixed point equation. For this, we use the following central perturbation result.

Theorem 19 *Let f, f' be two stable functions of $\mathcal{D}_n^+$, each admitting a fixed point $\Delta, \Delta' \in \mathcal{D}_n^+$ as*

$$\Delta = f(\Delta) \qquad \text{and} \qquad \Delta' = f'(\Delta').$$

Further assume that $\Delta' \sim O(1)$, that f is contracting for the stable semi-metric around Δ' with a Lipschitz parameter $\lambda < 1$ satisfying $\frac{1}{1-\lambda} \leq O(1)$, and that $\|f(\Delta') - f'(\Delta')\| = o(1)$ ($\|f(\Delta') - f'(\Delta')\| \leq O(a_s)$ with $a_s \underset{s\to\infty}{\to} 0$). Then, there exists a constant $K \leq O(1)$ such that

$$\left\|\Delta - \Delta'\right\| \leq K \|f(\Delta') - f'(\Delta')\|.$$

Theorem 19 can be employed when Δ is random and Δ' is a deterministic equivalent (yet to be defined). If we let $f = I^X \circ u$ (and thus $\Delta = \hat{\Delta}$), it is not possible find a deterministic stable function f' close to f such that its fixed point Δ' would satisfy the core hypothesis $\Delta' \sim O(1)$. This is partly due to the fact that $I^X \circ u = \mathrm{Diag}(\frac{1}{n} x_i^T R^X \circ u x_i)_{1 \leq i \leq n}$ scales with τ which might be unbounded. For this reason, we will consider $D \equiv \frac{\hat{\Delta}}{\tau}$ rather than $\hat{\Delta}$ itself.

3.2. Definition of $\tilde{D}$, the deterministic equivalent of D

The matrix $D \equiv \frac{\hat{\Delta}}{\tau}$ satisfies the fixed point equation

$$D = I^{\mathring{Z}}(u^\tau(D)) \qquad \text{where} \qquad \mathring{z}_i \equiv z_i + \frac{m}{\sqrt{\tau_i}} \qquad \text{and} \qquad u^\tau : \Delta \mapsto \tau u(\tau \Delta).$$

Though, in order to apply Corollary 17 that relies on Assumption 5, we will need a bound on the energy of the signal and a "loose" control on the τ_i's.

Assumption 6 $\|m\| \leq O(1)$.

Assumption 7 $\|\tau\|_1, \|\frac{1}{\tau_i}\|_1 \leq O(n)$.

We have then indeed for $i \in [n]$ $\mathbb{E}[\mathring{z}_i] \leq \|\mu_i\| + \tau_i^{1/2}\|m\| \leq O(\sqrt{n})$ (later to bound D from below, we will need the bound on $\frac{1}{n}\|\tau\|_1$ and $\frac{1}{n}\|1/\tau\|_1$, that is why we adopted such a general assumption).

We can still not apply Corollary 17 since $\|u^\tau(D)\|$ is possibly unbounded. Still, let us assume for the moment that $\|u^\tau(D)\|$ is indeed bounded: then, following our strategy, we are led to introducing a deterministic diagonal matrix $\tilde{D}$ ideally approaching D and satisfying

$$\tilde{D} = \frac{\Lambda^{\mathring{C}}(u^\tau(\tilde{D}))}{I_n + u^\tau(\tilde{D})\Lambda^{\mathring{C}}(u^\tau(\tilde{D}))}, \quad \text{where we recall} \quad \Lambda^{\mathring{C}}(u^\tau(\tilde{D})) = \check{I}\left(\mathring{C}, \frac{u^\tau(\tilde{D})}{I_n + u^\tau(\tilde{D})\Lambda^{\mathring{C}}(u^\tau(\tilde{D}))}\right). \tag{6}$$

Before proving the validity of this estimate $\tilde{D}$ of D, let us justify the validity of its definition (i.e., the existence and uniqueness of $\tilde{D}$).

Proposition 20 *Let $x \in \mathbb{R}^+$. Then the equation*

$$\eta = \frac{1}{\frac{1}{x} + u(\eta)}, \eta \in \mathbb{R}^+.$$

admits a unique solution that we denote $\eta(x)$. The mapping $\eta : \mathbb{R} \to \mathbb{R}$ is stable.

Proof We already know from Proposition 18 that $D \mapsto \Lambda^{\mathring{C}}(u^\tau(D))$ is stable and bounded from above and below (since $u^\tau \leq \|\tau\| u^\infty$). The same is true for $\eta_\tau(\Lambda^{\mathring{C}}(u^\tau(\tilde{D})))$ since η is stable and, for all $x \in \mathbb{R}^+$, $\frac{1}{u^\infty + \frac{1}{x}} \leq \eta(x) \leq x$ (here x should be replaced by $\Lambda^{\mathring{C}}(u^\tau(xI_n))$ which is bounded from above and below). The existence and uniqueness of $\tilde{D}$ thus unfold from Theorem 5. ■

The first equation of (6) can be rewritten $\tilde{D} = \eta_\tau(\Lambda^{\mathring{C}}(u^\tau(\tilde{D})))$, with $\eta_\tau : x \mapsto \frac{\eta(\tau x)}{\tau}$, we are thus now allowed to define:

Proposition 21 *There exists a unique diagonal matrix $\tilde{D} \in \mathcal{D}_n^+$ satisfying (6).*

3.3. Concentration of D around $\tilde{D}$

Once $\tilde{D}$ defined, we can follow our strategy to bound $\|D - \tilde{D}\|$.The first step is to verify the core hypothesis of Theorem 19, namely $\tilde{D} \sim O(1)$. We will here need a supplementary assumption on η, which can be expressed through a condition on u, justified by the following lemma.

Lemma 22 *The mapping $\eta_/$ is bounded from below iff, for all $t \in \mathbb{R}^+$, $u_\cdot(t) = tu(t) < 1$.*

Assumption 8 *There exists $u_\cdot^\infty > 0$ such that $1 - u_\cdot^\infty \geq O(1)$ and, for all $t \in \mathbb{R}^+$, $u_\cdot(t) \leq u_\cdot^\infty$.*

Assumption 9 *For all $i \in [n]$, $\operatorname{Tr} C_i \geq O(n)$.*

With these assumptions, we then have:

Lemma 23 $\tilde{D} \sim O(1)$.

Proof We already know from our assumptions that $O(1) \leq \frac{1}{n}\operatorname{Tr}(C_i) + \frac{1}{n}\frac{m^T m}{\tau} = \frac{1}{n}\operatorname{Tr}\mathring{C}_i \leq O(1)$ and from Proposition 18 that, for any $\Delta \in \mathcal{D}_n^+$, $\frac{O(1)}{n(\gamma + \frac{1}{n}\|\sum C_i \tau_i u(\tau_i \Delta)\|)} \leq \Lambda^{\mathring{C}}(u^\tau(\Delta)) \leq O(1)$. Thus $\Lambda^{\mathring{C}}(u^\tau(\tilde{D})) \sim O(1)$, since $\|\sum C_i \tau_i u(\tau_i \tilde{D})\| \leq u^\infty \|C_i\| \|\tau\|_1 \leq O(n)$. As such, we can bound $\|\tilde{D}\| \leq \|\Lambda^{\mathring{C}}(u^\tau(\tilde{D}))\| \leq O(1)$ and $\tilde{D} = \eta_\tau(\Lambda^{\mathring{C}}(u^\tau(\tilde{D}))) \geq \eta_/^\infty \Lambda^{\mathring{C}}(u^\tau(\tilde{D})) \geq O(1)$. ■

Proposition 24 *There exist two constants $C, c > 0$ ($C, c \sim O(1)$) such that, for any $\varepsilon \in (0, 1]$,*

$$\mathbb{P}\left(\left\|D - \tilde{D}\right\| \geq \varepsilon\right) \leq Ce^{-cn\varepsilon^2/\log(n)}.$$

Proof Let us check the hypotheses of Theorem 19. We already know that $\tilde{D} \sim O(1)$ from Lemma 23. Let us now bound the Lipschitz parameter λ of $\check{I}^{\mathring{Z}} \circ u^\tau$ around $\tilde{D}$ defined as:

$$\forall \Delta \in \mathcal{D}_n^+ : \quad \left\| \frac{\check{I}^{\mathring{Z}}(u^\tau(\Delta)) - \check{I}^{\mathring{Z}}(u^\tau(\tilde{D}))}{\sqrt{\check{I}^{\mathring{Z}}(u^\tau(\Delta))\check{I}^{\mathring{Z}}(u^\tau(\tilde{D}))}} \right\| < \lambda \left\| \frac{\Delta - \tilde{D}}{\sqrt{\Delta D'}} \right\|.$$

An inequality similar as in Lemma 2 gives us $\lambda \leq \sqrt{\|1 - \gamma Q^{\mathring{Z}}(\tilde{D})\|} \leq \frac{1}{1 + \frac{\gamma}{u^\infty \|\mathring{Z} \cdot \mathring{Z}^T\|}} < 1$ (thanks to Lemma 6). Now, since $\|\mathring{Z} \cdot \mathring{Z}^T\| = \frac{1}{n}\|\mathring{Z}\mathring{Z}^T\| \leq (\|\mathring{Z}\|/\sqrt{n})^2$, we know from Lemma 14 that, with probability bigger than $1 - Ce^{cn}$, (for some $C, c > 0$), $\|\mathring{Z}\| \leq K\sqrt{n}$. Thus under this highly probable event $\frac{1}{1-\lambda} \leq O(1)$. Eventually, we know from Assumption 4 that $\check{Z}u^\tau(\tilde{D}) \propto \mathcal{E}_2$ (since $u^\tau(\tilde{D}) \leq u^\infty \|\frac{1}{D}\| \leq O(1)$). We may thus employ Lemma **??** (in the appendix) to get $\check{I}^{\mathring{Z}}(u^\tau(\tilde{D}))_i = \frac{1}{n} z_i Q_{-i}^{\mathring{Z}}(u^\tau(\tilde{D}))z_i \propto \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_1(1/n)$ and Corollary 17 to state that $\|\mathbb{E}[\check{I}^{\mathring{Z}}(u^\tau(\tilde{D}))] - \tilde{D}\| \leq O(\sqrt{\log(n)/n})$. Thus there exist two constants $C, c > 0$ such that

$$\forall t > 0: \ \mathbb{P}\left(\left\|\check{I}^{\mathring{Z}}(u^\tau(\tilde{D})) - \tilde{D}\right\| \geq t\right) \leq C' e^{-c' nt^2/\log n}.$$

From the proof of Theorem 19, we see that it is sufficient to have with high probability $\check{I}^{\mathring{Z}}(u^\tau(\tilde{D})) - \tilde{D} \leq K$ for any $K \geq O(1)$ ($K \leq \frac{1}{\|\Delta\|}(\frac{1-\lambda}{2})^2 \geq O(1)$). This clearly holds with probability larger than $C' e^{-c' nK^2/\log n}$. Choosing C and c appropriately, we obtain the result of the proposition. ■

It is even possible to simplify the formulation of the deterministic equivalent $\tilde{D}$ under the supplementary assumption:

Assumption 10 *For all $i \in [n]$, $\tau_i \geq O(1/\sqrt{n})$.*

Proposition 25 *The fixed point equation $D = \eta_\tau \circ \Lambda^C \circ u^\tau(D)$ admits a unique solution, denoted $\tilde{D}_{-m} \in \mathcal{D}_n^+$, and which satisfies $\|\tilde{D} - \tilde{D}_{-m}\| \leq O\left(\frac{1}{\sqrt{n}}\right)$.*

Proof The existence and uniqueness of $\tilde{D}_{-m}$ are justified for the same reasons as for $\tilde{D}$ (just take $m = 0$). In order to use again Theorem 19, we know that $\tilde{D} \sim O(1)$ (as for $\tilde{D}_{-m}$) and we just need to bound $\left\|\eta_\tau \circ \Lambda^C \circ u^\tau(\tilde{D}) - \eta_\tau \circ \Lambda^{\mathring{C}} \circ u^\tau(\tilde{D})\right\|$. Note that η is 1-Lipschitz because, since it is stable, so that, for any $x, y \in \mathbb{R}^+$:

$$\frac{|\eta(x) - \eta(y)|}{|x - y|} \leq \sqrt{\frac{\eta(x)\eta(y)}{xy}} = \sqrt{\frac{1}{(1 + xu(\eta(x))\,(1 + yu(\eta(y)))}} \leq 1.$$

Thus η_τ is also 1-Lipschitz. We are then left to bounding the distance between $\Lambda^C \circ u^\tau(\tilde{D})$ and $\Lambda^{\mathring{C}} \circ u^\tau(\tilde{D})$, and we are naturally led to employing a second time Theorem 19 since those two values are both fixed points of stable mappings:

$$\Lambda^C(u^\tau(\tilde{D})) = \tilde{I}^C_{u^\tau(\tilde{D})}(\Lambda^C(u^\tau(\tilde{D}))) \qquad \text{and} \qquad \Lambda^{\mathring{C}}(u^\tau(\tilde{D})) = \tilde{I}^{\mathring{C}}_{u^\tau(\tilde{D})}(\Lambda^{\mathring{C}}(u^\tau(\tilde{D})))$$

where, for any $S \in \mathcal{S}_p^n$ and $\Delta \in \mathcal{D}_n^+$, $\tilde{I}^S_\Delta : \Lambda \mapsto \check{I}\left(S, \frac{\Delta}{I_n + \Delta\Lambda}\right)$. Once again, the first hypotheses are satisfied, $\Lambda^C(u^\tau(\tilde{D})) \sim O(1)$ and noting for simplicity $\Delta \equiv u^\tau(\tilde{D})$, $\Lambda \equiv \Lambda^C(\Delta)$ and $\tilde{Q}^S = \tilde{Q}^S(S, \frac{\Delta}{I_n + \Delta\Lambda})$ (for $S = \mathring{C}$ or $S = C$), and we are left to bounding, for any $i \in [n]$,

$$\begin{aligned}
\left|\tilde{I}^C_\Delta(\Lambda)_i - \tilde{I}^{\mathring{C}}_\Delta(\Lambda)_i\right| &\leq \frac{1}{n\tau_i} m^T \tilde{Q} m + \left|\frac{1}{n} \operatorname{Tr}\left(C_i \tilde{Q}^C \left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\tau_i} m m^T\right) \tilde{Q}\right)\right| \\
&\leq O\left(\frac{1}{\tau_i n}\right) + O\left(\frac{1}{n^2}\sum_{i=1}^n \frac{1}{\tau_i}\right) m^T \tilde{Q} C_i \tilde{Q}^C m \leq O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Applying twice Theorem 19, we retrieve the result of the proposition. ■

3.4. Spectral distribution of $\hat{C}$

As an immediate corollary of the previous results, a deterministic equivalent for the spectral distribution of $\hat{C} = \frac{1}{n}\mathring{Z}u^\tau(D)\mathring{Z}^T$ can be computed, through an estimation of the Stieltjes transform $m(z) = \frac{1}{p}\operatorname{Tr}((\hat{C} - zI_p)^{-1})$ for $z < 0$.

Theorem 26 *For any $z \geq O(1)$, there exist two constants $C, c > 0$ ($C, c \sim O(1)$) such that, for any $\varepsilon > 0$, $\varepsilon \leq 1$,*

$$\mathbb{P}\left(\left|m(-z) - \frac{1}{p}\operatorname{Tr}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{u^\tau(\tilde{D}_{-m})_i C_i}{1 + \Lambda^C(u^\tau(\tilde{D}_{-m}))_i u^\tau(\tilde{D}_{-m})_i} + zI_p\right)^{-1}\right| \geq \varepsilon\right) \leq Ce^{-cn\varepsilon^2/\log(n)}.$$

As a consequence, the *implicit* random matrix $\hat{C}$ actually has the same deterministic equivalent for its spectral distribution as the *explicit* random matrix $\frac{1}{n}Zu^\tau(\tilde{D}_{-m})Z^T$. Figure 1 depicts the eigenvalue distribution of the sample covariance of the data matrix X (i) deprived of the influence of τ (i.e. for $\tau = I_n$), (ii) corrected with the robust estimator of the scatter matrix (it is the sample covariance matrix of $Xu(\hat{\Delta})^{1/2}$) and (iii) without any modification on X. For the two first spectral distributions, we displayed their estimation with the Stieltjes transform as per Theorem 26.

Figure 1: Spectral distributions of the matrices $\frac{1}{n}(Z+m\mathbb{1}^T)(Z+m\mathbb{1}^T)^T$, $\hat{C}$ and $\frac{1}{n}XX^T$ with their prediction (when possible); $p = 400$, $n = 1200$, the variables $\tau_1, \ldots, \tau_n$ are taken from a Student distribution with 1 degree of freedom, $m \sim \mathcal{N}(0, I_p)$; $Z = \sin(AW)$ where $A \in \mathcal{M}_p$ is a fixed orthogonal matrix, $W \in \mathcal{M}_{p,n}$ is a matrix with zero-mean and unit variance Gaussian entries ($Z \propto \mathcal{E}_2$ by construction). The population covariance of Z is computed with a set of $100p$ drawings (the mean is 0 by construction).

4. Conclusion

In this article, we have developed an original framework to study the large dimensional behavior of the matrix solution to a fixed point equation, under a quite generic probabilistic data model (which notably does not enforce independence in the data entries). Recalling that most state-of-the-art statistical (machine) learning algorithms are optimization problems, having implicit solution, which are then applied to complex data models, our present work opens the path to a more systematic exploitation of concentration of measure theory for the large dimensional analysis of possibly complex machine learning algorithms and data models.

References

Y. Chitour and F. Pascal. Exact maximum likelihood estimates for sirv covariance matrix: Existence and algorithm analysis. *Signal Processing, IEEE Transactions on*, 56(10):4563–4573, 2008.

R. Couillet and M. McKay. Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, 131:99–120, 2014.

P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964.

Michel Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2005.

Cosme Louart and Romain Couillet. Concentration of measure and large random matrices with an application to sample covariance matrices. *submitted to Random Matrices: Theory and Applications*, 2019.

R. A. Maronna. Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, 4:51–67, 1976.

Jean Philippe Ovarlez, Sze Kim Pang, Frédéric Pascal, V Achard, and TK Ng. Robust detection using the sirv background modelling for hyperspectral imaging. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 4316–4319. IEEE, 2011.

D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15(1): 234–251, 1987.

L. Yang, R. Couillet, and M. McKay. Minimum variance portfolio optimization with robust shrinkage covariance estimation. In *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2014.

T. Zhang, X. Cheng, and A. Singer. Marchenko-Pastur Law for Tyler's and Maronna's M-estimators. *http://arxiv.org/abs/1401.3424*, 2014.

Appendix A. Assumptions

Assumption 1 $u \in \mathcal{S}(\mathbb{R}^+)$, $\exists u^\infty > 0$ *such that* $\forall t \in \mathbb{R}^+,\ u(t) \leq u^\infty$.

Assumption 2 *The random vectors* $z_1, \ldots, z_n$ *are all independents.*

Assumption 3 $p \sim O(n)$

Assumption 4 $Z \propto \mathcal{E}_2$

Assumption 5 $\forall i \in \{1, \ldots, n\} : \|\mu_i\| \leq O(1)$.

Assumption 6 $\|m\| \leq O(1)$

Assumption 7 $\|\tau\|_1, \|\frac{1}{\tau_i}\|_1 \leq O(n)$.

Assumption 8 $\exists u_\cdot^\infty > 0$ *such that* $1 - u_\cdot^\infty \geq O(1)$ *and* $\forall t \in \mathbb{R}^+, u_\cdot(t) = tu(t) \leq u_\cdot^\infty$.

Assumption 9 $\forall i \in [n] : \operatorname{Tr} C_i \geq O(n)$.

Assumption 10 $\forall i \in [n]$, $\tau_i \geq O(1/\sqrt{n})$.

Appendix B. The stable semi-metric

B.1. Stability of I

Proof [Proof of Lemma 2] Given $a \in \{1, \ldots, k\}$, we can bound thanks to Cauchy Shwarz inequality:

$$\begin{aligned}
\left|\tilde{I}(S,\Delta)_a - \tilde{I}(S,\Delta')_a\right| &= \left|\frac{1}{n}\operatorname{Tr}\left(S_a\left(Q_\gamma(S,\Delta') - Q_\gamma(S,\Delta)\right)\right)\right| \\
&= \left|\frac{1}{n}\sum_{b=1}^k \operatorname{Tr}\left(S_a Q_\gamma(S,\Delta') S_b \left(\Delta'_b - \Delta_b\right) Q_\gamma(S,\Delta)\right)\right| \\
&\leq \frac{1}{n}\sqrt{\sum_{b=1}^k \operatorname{Tr}\left(S_a Q_\gamma(S,\Delta)\frac{S_b\left|\Delta'_b - \Delta_b\right|}{\sqrt{\Delta_b\Delta'_b}}\Delta_b Q_\gamma(S,\Delta)\right)} \\
&\qquad \cdot \sqrt{\sum_{b=1}^k \operatorname{Tr}\left(S_a Q_\gamma(S,\Delta')\frac{S_b\left|\Delta'_b - \Delta_b\right|}{\sqrt{\Delta_b\Delta'_b}}\Delta'_b Q_\gamma(S,\Delta')\right)} \\
&\leq \left\|\frac{\Delta' - \Delta}{\sqrt{\Delta\Delta'}}\right\| \sqrt{\frac{1}{n}\operatorname{Tr}\left(S_a Q_\gamma(S,\Delta)\left(1 - \gamma Q_\gamma(S,\Delta)\right)\right)} \\
&\qquad \cdot \sqrt{\frac{1}{n}\operatorname{Tr}\left(S_a Q_\gamma(S,\Delta')\left(1 - \gamma Q_\gamma(S,\Delta')\right)\right)} \\
&< \left\|\frac{\Delta' - \Delta}{\sqrt{\Delta\Delta'}}\right\| \sqrt{\tilde{I}(S,\Delta)_a \tilde{I}(S,\Delta')_a}
\end{aligned}$$

■

B.2. General properties of the stable msemmimetric and of the stable class

Remark 27 *The function d_s is not a metric because it does not satisfy the triangular inequality, one can see for instance that:*

$$d_s(4,1) = \frac{3}{2} > \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = d_s(4,2) + d_s(2,1)$$

One can show that given $x, y \in \mathbb{R}^+$, for any $p \in \mathbb{N}^*$ and $y_1, \dots, y_{p-1} \in \mathbb{R}^+$, we have the inequality:

$$d_s(x, y_1) + \cdots + d_s(y_{p-1}, z) \geq d_s\left(x^{\frac{1}{p}}, z^{\frac{1}{p}}\right).$$

It is an equality in the case $y_i = x^{\frac{p-i}{p}} z^{\frac{i}{p}}$ for $i \in \{1, \dots, p-1\}$. We can not get interesting inferences to palliate the absence of a triangular inequality since the function $x \mapsto x^p$ is not Lipschitz for the semi-metric d_s (instead, one can show that $x \mapsto x^{\frac{1}{p}}$ is $\frac{1}{p}$-Lipschitz).

Proof [Proof of Proposition 2]

Let us consider $x, y \in \mathbb{R}^+$, such that, say, $x \leq y$. We suppose in a first time that $f_/$ is non-increasing and that $f_\cdot$ is non-decreasing. We know that $\frac{f(x)}{x} \geq \frac{f(y)}{y}$, and subsequently:

$$f(y) - f(x) \leq \frac{f(y)}{y}(y - x) \qquad \text{and} \qquad f(y) - f(x) \leq \frac{f(x)}{x}(y - x) \tag{7}$$

The same way, since $f(x)x \leq f(y)y$ we also have the inequalities:

$$f(x) - f(y) \leq \frac{f(y)}{x}(y - x) \qquad \text{and} \qquad f(x) - f(y) \leq \frac{f(x)}{y}(y - x) \tag{8}$$

Now if $f(y) \geq f(x)$, we can take the root of the product of the two inequalities of (7) and if $f(y) \leq f(x)$, we take the root of the product of the two inequalities of (8), to obtain, in both cases:

$$|f(x) - f(y)| \leq \sqrt{\frac{f(y)f(x)}{xy}}\,|x - y|$$

That means that $f \in \mathcal{S}(\mathbb{R}^+)$.

Reciprocally, we suppose that $f \in \mathcal{S}(\mathbb{R}^+)$, if $f(x) \leq f(y)$, then $f(x)x \leq f(y)y$ and:

$$\left(\frac{f(x)}{x} \leq \frac{f(y)}{y}\right) \quad \Rightarrow \quad \left(f(y) - f(x) \leq \frac{f(y)}{y}(y - x)\right) \quad \Rightarrow \quad \left(\frac{f(y)}{y} \leq \frac{f(x)}{x}\right),$$

and if $f(x) \leq f(y)$, then $\frac{f(x)}{x} \geq \frac{f(y)}{y}$ and:

$$(f(x)x \leq f(y)y) \quad \Rightarrow \quad \left(f(x) - f(y) \leq \frac{f(y)}{x}(y - x)\right) \quad \Rightarrow \quad (f(y)y \leq f(x)x).$$

In both cases ($f(x) \leq f(y)$ and $f(y) \leq f(x)$), we see that $f_/(x) \geq f_/(y)$ and $f_\cdot(x) \geq f_\cdot(y)$, we have thus proved our result.

■

Remark 28 *Given $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$, we can introduce as in subsection 2.2 $f_/, f. : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ defined with:*

$$f_/ \ : \ \Delta \mapsto \mathrm{Tr}\left(\frac{f(\Delta)}{\Delta}\right) \qquad \text{and} \qquad f. \ : \ \Delta \mapsto \mathrm{Tr}\left(\Delta f(\Delta)\right)$$

It is possible to inspire from Property 2 to define a similar class that can be called the weak stable class $\mathcal{S}_w(\mathcal{D}_n^+)$. A function $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ is in $\mathcal{S}_w(\mathcal{D}_n^+)$ if and only if $f_/$ is nonincreasing and $f.$ is nondecreasing. It can be showed that $I^M, \tilde{I}^S \in \mathcal{S}_w(\mathcal{D}_n^+)$ Although this definition does not rely on a metric (nor on a semi metric), it is quite convenient to show fixed point theorems, but we did not find any use in our paper since we already have $I^M, \tilde{I}^S \in \mathcal{S}(\mathcal{D}_n^+)$.

For the proof of Proposition 3, let us give two small results.

Lemma 29 *Given four positive numbers $a, b, c, d \in \mathbb{R}^+$:*

$$\sqrt{ab} + \sqrt{\alpha\beta} \leq \sqrt{(a+b)(\alpha+\beta)} \qquad \text{and} \qquad \frac{a+\alpha}{b+\beta} \leq \max\left(\frac{a}{b}, \frac{\alpha}{\beta}\right)$$

Proof For the first result, we deduce from the inequality $2ab\alpha\beta \leq a\alpha + b\beta$:

$$\left(\sqrt{ab} + \sqrt{\alpha\beta}\right)^2 = ab + \alpha\beta + 2\sqrt{ab\alpha\beta} \leq ab + \alpha\beta + a\alpha + b\beta = (a+b)(\alpha+\beta)$$

For the second result, we simply bound:

$$\frac{a+\alpha}{b+\beta} \leq \frac{a}{b}\frac{b}{b+\beta} + \frac{\alpha}{\beta}\frac{\beta}{b+\beta} \leq \max\left(\frac{a}{b}, \frac{\alpha}{\beta}\right)\left(\frac{b}{b+\beta} + \frac{\beta}{b+\beta}\right) = \max\left(\frac{a}{b}, \frac{\alpha}{\beta}\right)$$

■

Proof [Proof of Property 3] The three first properties are obvious, we are just left to show the stability through the sum. Note that this time, there is no characterization on $\mathcal{S}(\mathbb{R}^+)$ with the monotonicity of $f_/$ and $f.$ given in Property 2 (for that reason, this property is easier to show on the set $S_w(\mathcal{D}_n^+)$ described in Remark 28). Nonetheless, given $f, g \in \mathcal{S}(\mathcal{D}_n^+)$ and $\Delta, \Delta' \in \mathbb{R}^+$ there exists $i_0 \in [n]$ such that:

$$\begin{aligned} d_s(f(\Delta) + g(\Delta), f(\Delta') - g(\Delta')) &= \frac{\left|f(\Delta_{i_0}) - f(\Delta'_{i_0}) + g(\Delta_{i_0}) - g(\Delta'_{i_0})\right|}{\sqrt{(f(\Delta_{i_0}) + g(\Delta_{i_0}))(f(\Delta'_{i_0}) + g(\Delta'_{i_0}))}} \\ &\leq \frac{\left|f(\Delta_{i_0}) - f(\Delta'_{i_0})\right| + \left|g(\Delta_{i_0}) - g(\Delta'_{i_0})\right|}{\sqrt{f(\Delta_{i_0}) + f(\Delta_{i_0})} + \sqrt{g(\Delta'_{i_0}) + g(\Delta'_{i_0})}} \\ &\leq \max\left(\frac{\left|f(\Delta_{i_0}) - f(\Delta'_{i_0})\right|}{\sqrt{f(\Delta_{i_0}) + f(\Delta_{i_0})}}, \frac{\left|g(\Delta_{i_0}) - g(\Delta'_{i_0})\right|}{\sqrt{g(\Delta'_{i_0}) + g(\Delta'_{i_0})}}\right) \leq d_s(\Delta, \Delta') \end{aligned}$$

thanks to Lemma 29 and the stable character of f and g. ■

B.3. Topological properties of the stable semi-metric

Lemma 30 *Any Cauchy sequence of $(\mathcal{D}_n^+, d_s)$ is bounded from below and above*

Proof Considering a Cauchy sequence of diagonal matrices $\Delta^{(k)} \in \mathcal{D}_n^+$, we know that there exists $K \in \mathbb{N}$ such that:

$$\forall p, q \geq K,\ \forall i \in \{1, \dots, n\}: \ \ |\Delta_i^{(p)} - \Delta_i^{(q)})| \leq \sqrt{\Delta_i^{(p)} \Delta_i^{(q)}}$$

For $k \in \mathbb{N}$, let us introduce the indexes $i_M^k, i_m^k \in \mathbb{N}$, satisfying:

$$\Delta_{i_M^k}^{(k)} = \max\left(\Delta_i^{(k)}, 1 \leq i \leq n\right) \qquad \text{and} \qquad \Delta_{i_M^k}^{(k)} = \min\left(\Delta_i^{(k)}, 1 \leq i \leq n\right)$$

If we suppose that there exists an extraction $(\Delta_{i_M^k}^{(\phi(k))})_{k\geq 0}$ such that $\Delta_{i_M^k}^{(\phi(k))} \underset{k\to\infty}{\longrightarrow} \infty$ then:

$$\sqrt{\Delta_{i_M^{\phi(k)}}^{(\phi(k))}} \leq \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)}} + \frac{\Delta_{i_M^{\phi(k)}}^{(N)}}{\sqrt{\Delta_{i_M^{\phi(k)}}^{(\phi(k))}}} \underset{k\to\infty}{\longrightarrow} \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)}} < \infty$$

which is absurd. Therefore $(\Delta_{i_M^k}^{(k)})_{k\geq 0}$ and thus also $(\Delta^{(k)})_{k\geq 0}$ are bounded superiorly. For the inferior bound, we consider the same way an extraction $(\Delta_{i_m^k}^{(\psi(k))})_{k\geq 0}$ such that $\Delta_{i_m^k}^{(\psi(k))} \underset{k\to\infty}{\longrightarrow} 0$. We have:

$$\Delta_{i_M^{\phi(k)}}^{(\phi(k))} \geq \Delta_{i_M^{\phi(k)}}^{(N)} - \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)} \Delta_{i_M^{\phi(k)}}^{(\phi(k))}} \underset{k\to\infty}{\longrightarrow} \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)}} > 0.$$

which is once again absurd. ■

Proof [Proof of Property 4] Given a Cauchy sequence of diagonal matrices $\Delta^{(k)} \in \mathcal{D}_n^+$, we know from preceding lemma that there exists $\delta_M, \delta_m \in \mathbb{R}^+$ such that $\forall k \geq 0 : \delta_m I_n \leq \Delta^{(k)} \leq \delta_M I_n$. Thanks to the Cauchy hypothesis:

$$\forall \varepsilon > 0, \exists K \geq 0 \mid \forall p, q \geq K : \ \forall i \in \{1, \dots, n\} : \left|\Delta_i^{(p)} - \Delta_i^{(p)}\right| \leq \varepsilon \delta_M$$

and as a consequence, $(\Delta^{(k)})_{k\geq 0}$ is a Cauchy sequence in the complete space $(\mathcal{D}_n^{0,+}, \|\cdot\|)$ it converges to a matrix $\Delta^{(\infty)} \in \mathcal{D}_n^{0,+}$. Moreover, $\Delta^{(\infty)} \geq \delta_k I_n$ as any $\Delta^{(k)}$, for all $k \in \mathbb{N}$, thus $\Delta^{(\infty)} \in \mathcal{D}_n^+$ and we are left to show that $\Delta^{(k)} \underset{k\to\infty}{\longrightarrow} \Delta^{(\infty)}$ for the semi-metric d_s. It suffices to write:

$$d_s(D^{(k)}, D^{(\infty)}) = \left\| \frac{D^{(k)} - D^{(\infty)}}{\sqrt{D^{(k)} D^{(\infty)}}} \right\| \leq \delta_m \left\| D^{(k)} - D^{(\infty)} \right\| \underset{k\to\infty}{\longrightarrow} 0.$$

■

Proof [Proof of Theorem 5] There exists $\lambda \in (0,1)$ and a constant $\delta_M, \delta_m \in \mathbb{R}^+$ such that $\forall \Delta, \Delta' \in \mathcal{D}_n^+$, $d_s(f(\Delta), f(\Delta')) \leq \lambda d_s(\Delta, \Delta')$ and $\delta_m I_n \leq f(\Delta) \leq \delta_M I_n$. The sequence $(\Delta^{(k)})_{k\geq 0}$ satisfying:

$$\Delta^{(0)} = I_n \qquad \text{and} \qquad \forall k \geq 1 : \ \Delta^{(k)} = f(\Delta^{(k-1)})$$

is a Cauchy sequence. Given $\epsilon > 0$, we have indeed for $K \geq \frac{\log(\epsilon\delta_m/2\delta_M)}{\log(\lambda)}$:

$$\forall p, q > K : d_s(\Delta^{(q)}, \Delta^{(p)}) \leq \lambda^K d_s\left(f^{q-K}(\Delta^{(q-K)}), f^{p-K}(\Delta^{(p)})\right) \leq \frac{2\delta_M \lambda^K}{\delta_m} \leq \varepsilon$$

We know thanks to Property 4 that there exists $\Delta^* \in \mathcal{D}_n^+$ such that $f(\Delta^*) = \Delta^*$ (since f is continuous) and the contracting character of f ensures that it is the unique fixed point. ■

Remark 31 *It is possible to relax a bit the contracting hypothese on f if one supposes that f is monotonic. Let us consider a weakly monotonic mapping $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ bounded from below and above. If we suppose that f is stable and verifies:*

$$\forall \Delta, \Delta' \in \mathcal{D}_n^+ : \ d_s(f(\Delta), f(\Delta')) < d_s(\Delta, \Delta') \tag{9}$$

then there exists a unique fixed point $D \in \mathcal{D}_n^+$ satisfying $\Delta^* = f(\Delta^*)$.

Proof We first suppose that f is nondecreasing. As before, let us consider $\delta_M, \delta_m \in \mathbb{R}^+$ such that $\forall \Delta \in \mathcal{D}_n^+ \ \delta_m I_n \leq f(\Delta) \leq \delta_M I_n$. The sequence $(\Delta^{(k)})_{k\geq 0}$ satisfying $\Delta^{(0)} = \Delta_m I_n$, and for all $k \geq 1$, $\Delta^{(k)} = f(\Delta^{(k-1)})$ is a nondecreasing sequence bounded superiorly with δ_M, thus it converges to $\Delta^* \in \mathcal{D}_n^+$ and $\Delta^* = f(\Delta^*)$. This fixed point is clearly unique thanks to (9).

Now if f is nonincreasing then $\Delta \mapsto f^2(\Delta)$ is non decreasing and bounded inferiorly and superiorly thus it admits a unique fixed point $\Delta^* \in \mathcal{D}_n^+$ satisfying $\Delta^* = f^2(\Delta^*)$. We can deduce that $f(\Delta^*) = f^2(f(\Delta^*))$ which implies by uniqueness of the fixed point that $f(\Delta^*) = \Delta^*$ and the uniqueness of such a Δ^* is again a consequence of (9). ■

Proof [Proof of Corollary 7] We saw in Lemma 2 that:

$$d_s\left(I(S,\Delta), I(S,\Delta')\right) < \lambda d_s\left(\Delta, \Delta'\right) \qquad \text{with} \qquad \lambda = \sup\left\{\left\|1 - \gamma\tilde{Q}^S(\Delta)\right\|, \Delta \in \mathcal{D}_n^+\right\}.$$

Now, thanks to Lemma 6, we can bound:

$$\lambda \leq \frac{1}{1 + \frac{\gamma}{f_0\|S\|}} < 1 \qquad \text{and} \qquad \frac{\inf_{1\leq a\leq k}(\operatorname{Tr} S_a) I_p}{f_0\|S\| + \gamma} \leq \|I(S, f(\Delta))\| \leq \frac{\sup_{1\leq a\leq k}(\operatorname{Tr} S_a) I_p}{\gamma}$$

and therefore $g \circ \tilde{Q}^S \circ f$ is contracting and bounded from below and above: we can employ Theorem 5 to set the existence and the uniqueness of a solution $\Delta \in \mathcal{D}_n^+$ to $\Delta = g \circ \tilde{Q}^S \circ f(\Delta)$. ■

Appendix C. Concentration and estimation of the resolvent

Given $S \in \mathcal{S}_p^n$, we note $S_{-i} \equiv (S_1, \dots, S_{i-1}, 0, S_i, \cdots S_n)$ then for $\Delta \in \mathcal{D}_n^+$ and $i \in [n]$, we note:

$$Q_{-i}(S, \Delta) \equiv Q_z(S_{-i}, \Delta) = \left(\frac{1}{n} \sum_{i=1}^n S_i \Delta + z I_n \right)^{-1}$$

We have the first simple identity:

$$Q(S, \Delta) - Q_{-i}(S, \Delta) = \frac{1}{n} Q(S, \Delta) S_i Q_{-i}(S, \Delta), \tag{10}$$

Now if we consider a matrix $M = (m_1, \dots, m_n) \in \mathcal{M}_{p,n}$, we note $M_{-i} \equiv (m_1, \cdots, m_{i-1}, 0, m_{i+1}, \dots, m_n)$ then $(M \cdot M^T)_{-i} = (M_{-i} \cdot M_{-i}^T)$ and noting for simplicity $Q^M \equiv Q_z(M, \cdot)$, we see that $\check{I}^M(\Delta) \equiv \check{I}_z(M \cdot M^T, \Delta) = \operatorname{Diag}(\frac{1}{n} m_i^T Q_{-i} m_i)_{1 \le i \le n}$ and we can deduce from (10) the so-called "Schur identity":

$$Q^M(\Delta) m_i = \frac{Q_{-i}^M(\Delta) m_i}{1 + \frac{\Delta_i}{n} m_i^T Q_{-i}^M(\Delta) m_i} \qquad \text{and} \qquad I^M(\Delta) = \frac{\check{I}^M(\Delta)}{I_n + \Delta \check{I}^M(\Delta)} \tag{11}$$

Proof [Proof of Corollary 17]

It is shown in Louart and Couillet (2019) that $\frac{1}{n} z_i Q_{-i}(\Delta) z_i \in \frac{1}{n} \operatorname{Tr}(C_i \mathbb{E}[Q_{-i}]) \pm \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_1(1/n)$ and since $\|\frac{1}{n} C_i\|_1 = \frac{1}{n} \operatorname{Tr}(C_i) = O(1)$, we know from Theorem 16 applied to the random matrix $Z\Delta^{1/2}$ that $\|\frac{1}{n} \operatorname{Tr}(C_i \mathbb{E}[Q_{-i}]) - \frac{1}{n} \operatorname{Tr}\left(C_i Q(C, \frac{\Delta}{1+\Delta\Lambda^C(\Delta)})] \right) \| \le O\left(\sqrt{\log n / n}\right)$. Thus with the identity $\Lambda^C(\Delta)_i = \frac{1}{n} \operatorname{Tr}\left(C_i Q(C, \frac{\Delta}{1+\Delta\Lambda^C(\Delta)})] \right)$, we deduce that:

$$\frac{1}{n} z_i Q_{-i}(\Delta) z_i \in \Lambda^C(\Delta)_i \pm \mathcal{E}_1\left(\sqrt{\frac{\log n}{n}}\right).$$

Now we can employ Proposition 13 to the $O(1)$-Lipschitz mapping $f : t \mapsto \frac{1}{1+\Delta_i t}$ to set that $w_i \equiv \frac{1}{1+\Delta \frac{1}{n} z_i Q_{-i}(\Delta) z_i} \in \frac{1}{1+\frac{1}{n} \operatorname{Tr}(C_i \mathbb{E}[Q_{-i}])} \pm \mathcal{E}_1(\sqrt{\log n / n})$ and even (since $|w_i| \le 1$ we have a stronger result):

$$\frac{\frac{1}{n} z_i Q_{-i}(\Delta) z_i}{1 + \Delta \frac{1}{n} z_i Q_{-i}(\Delta) z_i} \in \frac{\Lambda^C(\Delta)_i}{1 + \Delta \Lambda^C(\Delta)_i} \pm \mathcal{E}_1\left(\sqrt{\frac{\log n}{n}}\right).$$

We can then conclude with identity 11 the result of the Corollary. ■

Proof [Proof of Proposition 18] Given $S \in \mathcal{S}_p^n$ and $\Delta, \Delta' \in \mathcal{D}_n^+$, there exists $i_0 \in [n]$ such that:

$$\begin{aligned}
d_s(\Lambda^S(\Delta), \Lambda^S(\Delta')) &= d_s\left(\check{I}\left(S, \frac{\Delta}{I_n + \Delta\Lambda^S(\Delta)}\right), \check{I}\left(S, \frac{\Delta'}{I_n + \Delta'\Lambda^S(\Delta')}\right)\right) \\
&< d_s\left(\frac{\Delta}{I_n + \Delta\Lambda^S(\Delta)}, \frac{\Delta'}{I_n + \Delta'\Lambda^S(\Delta')}\right) \\
&= d_s\left(\frac{I_n}{\Delta} + \Lambda^S(\Delta), \frac{I_n}{\Delta'} + \Lambda^S(\Delta')\right) \\
&= \frac{\left|\frac{1}{\Delta_{i_0}} + \Lambda^S(\Delta)_{i_0} - \frac{1}{\Delta'_{i_0}} + \Lambda^S(\Delta')_{i_0}\right|}{\sqrt{\left(\frac{1}{\Delta_{i_0}} + \Lambda^S(\Delta)_{i_0}\right)\left(\frac{1}{\Delta'_{i_0}} + \Lambda^S(\Delta')_{i_0}\right)}} \\
&\leq \max\left(\frac{\left|\frac{1}{\Delta_{i_0}} - \frac{1}{\Delta'_{i_0}}\right|}{\sqrt{\frac{1}{\Delta_{i_0}}\frac{1}{\Delta'_{i_0}}}}, \frac{\left|\Lambda^S(\Delta)_{i_0} - \Lambda^S(\Delta')_{i_0}\right|}{\sqrt{\Lambda^S(\Delta)_{i_0}\Lambda^S(\Delta')_{i_0}}}\right) \\
&\leq \max\left(d_s(\Delta, \Delta'), d_s(\Lambda^S(\Delta), \Lambda^S(\Delta'))\right)
\end{aligned}$$

Thanks to Lemma 2, the stability rules given in Property 3, and the little tools given by Lemma 29 already used to prove Property 3. As a conclusion:

$$d_s(\Lambda^S(\Delta), \Lambda^S(\Delta')) < \max\left(d_s(\Delta, \Delta'), d_s(\Lambda^C(\Delta), \Lambda^C(\Delta'))\right),$$

which directly implies that $d_s(\Lambda^S(\Delta), \Lambda^S(\Delta')) < d_s(\Delta, \Delta')$, in other words, Λ^S is stable.

The upper and lower bounds on $\Lambda^S(\Delta)$ for $S \in \mathcal{S}_p^n$ and $\Delta \in \mathcal{D}_n^+$ are direct consequences of the bounds on I (the same are true for $\check{I}$) given in Lemma 6 ■

Appendix D. Estimation of the robust scatter

D.1. Strategical Theorem

Proof [Proof of Theorem 19] Let us first bound (be careful that the stable semi-metric does not satisfy the triangular inequality):

$$\left\|\frac{\Delta - \Delta'}{\sqrt{\Delta f(\Delta')}}\right\| \leq d_s(f(\Delta), f(\Delta')) + \left\|\frac{f(\Delta') - \Delta'}{\sqrt{\Delta f(\Delta')}}\right\| \leq \lambda\left\|\frac{\Delta - \Delta'}{\sqrt{\Delta\Delta'}}\right\| + \left\|\frac{f(\Delta') - \Delta'}{\sqrt{\Delta f(\Delta')}}\right\|.$$

Then, since $\left\|\frac{\Delta-\Delta'}{\sqrt{\Delta\Delta'}}\right\| \leq \left\|\frac{\Delta-\Delta'}{\sqrt{\Delta f(\Delta')}}\right\|\left(1 + \left\|\frac{\sqrt{f(\Delta')}-\sqrt{f'(\Delta')}}{\sqrt{\Delta'}}\right\|\right)$ and $\varepsilon \equiv \left\|\frac{\sqrt{|f(\Delta')-f'(\Delta')|}}{\sqrt{\Delta'}}\right\| \leq O(\sqrt{a_s})$ by hypothesis, setting $K' = \frac{1}{1-\lambda-\varepsilon} \leq O(1)$, we have the inequality:

$$\left\|\frac{\Delta - \Delta'}{\sqrt{\Delta f(\Delta')}}\right\| \leq K'\left\|\frac{f(\Delta') - \Delta'}{\sqrt{\Delta f(\Delta')}}\right\| \quad \text{which implies} \quad \left\|\frac{\Delta - \Delta'}{\sqrt{\Delta}}\right\| \leq K''\left\|\frac{f(\Delta') - \Delta'}{\sqrt{\Delta}}\right\|, \quad (12)$$

for some constant $K'' > 0$, thanks to the inequality $\|f(\Delta') - \Delta'\| \leq O(a_s)$ leading to:

$$O(1) \leq \|\Delta'\| - O(a_s) \leq f(\Delta') \leq \|\Delta'\| + O(a_s) \leq O(1).$$

We are left to bound from below and above $\|\Delta\|$ to recover the result of the Theorem from (12). Considering the index i_0 such that $\Delta_{i_0} = \min(\Delta_i)_{1\leq i\leq n}$, we have:

$$\left|\Delta_{i_0} - \Delta'_{i_0}\right| \leq K''\sqrt{\Delta_{i_0}}\left\|\frac{f(\Delta') - \Delta'}{\sqrt{\Delta_{i_0}}}\right\| \leq O(a_s),$$

therefore, $\Delta_{i_0} \geq \Delta'_{i_0} - O(a_s) \geq O(1)$. On the other hand, one can bound again from (12):

$$\|\sqrt{\Delta}\| \leq \left\|\frac{\Delta'}{\sqrt{\phi}}\right\| + K''\left\|\frac{f(\Delta') - \Delta'}{\sqrt{\Delta}}\right\| \leq O(1)$$

As a consequence $\Delta \sim O(1)$, and we can conclude from (12). ■

D.2. Stability properties of η

Proof [Proof of Proposition 20] It is a simple application of Theorem 5. If we note $f : \eta \mapsto \frac{1}{\frac{1}{x} + u(\eta)}$, we know that f is bounded from below and above, for all $\eta \in \mathbb{R}^+$:

$$\frac{1}{\frac{1}{x} + u^\infty} \leq f(\eta) \leq x.$$

We can then employ Theorem 5 since f is contracting for the stable semi-metric:

$$\begin{aligned}
d_s(f(\eta), f(\eta')) = d_s\left(\frac{1}{f(\eta)}, \frac{1}{f(\eta')}\right) &= \frac{|u(\eta) - u(\eta')|}{\sqrt{\left(\frac{1}{x} + u(\eta)\right)\left(\frac{1}{x} + u(\eta')\right)}} \\
&\leq \sqrt{\frac{u(\eta)u(\eta')}{\left(\frac{1}{x} + u(\eta)\right)\left(\frac{1}{x} + u(\eta')\right)}}\, d_s(u(\eta), u(\eta')) \\
&\leq \sqrt{\frac{1}{\frac{1}{u(\eta)u(\eta')x^2} + \frac{1}{u(\eta')x} + \frac{1}{u(\eta)x} + 1}}\, d_s(u(\eta), u(\eta')) \\
&\leq \frac{1}{1 + \frac{1}{u^\infty x}\left(1 + \frac{1}{u^\infty x}\right)} d_s(\eta, \eta')
\end{aligned}$$

To prove the stability of η, we are going to use the characterization with the monotonicity of the functions $\eta_/ : x \mapsto \frac{\eta(x)}{x}$ and $\eta_\cdot \mapsto x\eta(x)$ presented in Property 2. Let us consider $x, y \in \mathbb{R}^+$ such that $x \leq y$, if $\eta(x) \leq \eta(y)$, then $\eta_\cdot(x) \leq \eta_\cdot(y)$, besides since, in addition, $u_/$ is nondecreasing:

$$\eta_/(x) = \frac{1}{1 + xu(\eta(x))} \geq \frac{1}{1 + \frac{y\eta(y)u(\eta(x))}{\eta(x)}} \geq \frac{1}{1 + yu(\eta(y))} = \eta_/(y).$$

The same way, if $\eta(x) \geq \eta(y)$, then $\eta_{/}(x) \geq \eta_{/}(y)$ and:

$$\eta_{\cdot}(x) = \frac{1}{\frac{1}{x^2} + \frac{u(\eta(x))}{x}} \leq \frac{1}{\frac{1}{y^2} + \frac{\eta(x)}{x}\frac{u(\eta(y))}{\eta(y)}} \leq \frac{1}{\frac{1}{y^2} + \frac{u(\eta(y))}{y}} = \eta_{\cdot}(y)$$

We see that in both cases, $\eta_{/}(x) \geq \eta_{/}(y)$ and $\eta_{\cdot}(x) \leq \eta_{\cdot}(y)$, therefore, thanks to Property 2, $\eta \in \mathcal{S}(\mathbb{R}^+)$. ■

Proof [Proof of Lemma 22] If there exists $\alpha > 0$ (and $\alpha < 1$) such that $\forall x \in \mathbb{R}^+$, $\frac{\eta(x)}{x} \geq \alpha$, then

$$\frac{\eta(x)}{x} + (1-\alpha) \geq 1 \qquad \text{and therefore:} \qquad \frac{1}{\frac{1}{x} + u(\eta(x))} = \eta(x) \geq \frac{1}{\frac{1}{x} + \frac{1-\alpha}{\eta(x)}},$$

which implies that $u(\eta(x))\eta(x) \leq 1-\alpha$. But since η is not bounded (otherwise $\lim_{t\to\infty} \frac{\eta(t)}{t} = 0 < \alpha$), η takes all the values around ∞ and in particular ($u_{\cdot}$ being nondecreasing), $u_{\cdot} \leq 1-\alpha$. Conversely, if $u_{\cdot}^{\infty} < 1$, $\forall x \in \mathbb{R}^+$:

$$\frac{\eta(x)}{x} \geq \frac{1}{1 + u_{\cdot}^{\infty}\frac{x}{\eta(x)}} \qquad \text{thus} \qquad \frac{\eta(x)}{x} \geq 1 - u_{\cdot}^{\infty} > 0$$

■