Submitted to the Annals of Applied Probability

A CONCENTRATION OF MEASURE AND RANDOM MATRIX APPROACH TO LARGE DIMENSIONAL ROBUST STATISTICS

BY COSME LOUART[1,2] AND ROMAIN COUILLET[1,*]

[1]GIPSA-lab, Université Grenoble-Alpes, France, [*]romain.couillet@gipsa-lab.grenoble-inp.fr

[2]CEA, LIST, site Nano-innov, France cosme.louart@gipsa-lab.grenoble-inp.fr

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.

1. Introduction. Robust estimators of covariance (or scatter) are necessary ersatz for the classical sample covariance matrix 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 [4] 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., [4, 8, 12]) and (ii) an independent analysis of its statistical properties when seen as a random object (in the large n regime [1] or in the large n, p regime [2, 14]).

In the present article, we claim that the study of the conditions of existence (i) and statistical behavior (ii) of $\hat{C}$ can be conveniently carried out jointly. Specifically, by means of a flexible framework based on concentration of measure theory and on a new stable semi-metric argument, we simultaneously explore the existence and large dimensional (n, p large) spectral properties of $\hat{C}$. Our findings may be summarized as the following three main contributions to robust statistics and more generally to large dimensional statistics.

First, the proposed concentration of measure framework has the advantage of relaxing the assumptions of independence in the entries of x_i made in previous works [2, 14], thereby allowing for possibly complex and quite realistic data models. In detail, our data model decomposes x_i 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, .e.g, $z_i = h(\tilde{z}_i)$ for $\tilde{z}_i \sim \mathcal{N}(0, I_q)$ and $h : \mathbb{R}^q \to \mathbb{R}^p$ a 1-Lipschitz mapping), m is a deterministic vector (a signal or information common to all data) and τ_i

Keywords and phrases: Robust Estimation, Concentration of Measure, Random Matrix Theory
MSC 2010 subject classifications: Primary 60B20; secondary 62F35

are arbitrary (possibly large) deterministic values.[1] This setting naturally arises in many engineering applications, such as in antenna array processing (radar, brain signal processing, etc.) where the τ_i's model noise impulsiveness and m is an informative signal to be detected by the experimenter [11], or in statistical finance where the x_i's model asset returns with high volatility and m is the market leading direction [13]. Besides, the hypothesis made on z_i is adapted to the generative modelling of possibly extremely complex data: it in particular encompasses all data models produced by generative neural networks, such as the now popular GANs (generative adversarial neural networks [3]).

Second, as compared to previous works in the field [8, 5, 10, 9, 2], our frameworks allows for the relaxation of some of the classically posed constraints on the mapping u made. Specifically, u is here only required to be 1-Lipschitz with respect to the "stable semi-metric" (defined in the course of the article), which is equivalent to assuming that $t \mapsto tu(t)$ is non-decreasing and that $t \mapsto u(t)/t$ is non-increasing. The semi-metric naturally arises when studying the *resolvent* $(\hat{C} + \gamma I_p)^{-1}$ of $\hat{C}$, which is at the core of our large p, n analysis of $\hat{C}$, using modern tools from random matrix theory. To establish concentration properties in the large dimensional regime on $\hat{C}$, under our framework, the function u is nonetheless further requested to be such that $t \to tu(t)$ is strictly smaller than 1 ($\hat{C}$ is however still defined without this condition). Yet, and most importantly, u needs not be a non-increasing function, as demanded by most works in the field.

Third, the "Lipschitz and stable semi-metric" properties of the model are consistently articulated so as to propagate the concentration properties from Z to the robust scatter matrix $\hat{C}$. The core technical result allowing for this articulation is Theorem 4.1. This combined framework provides the rate of convergence of the Stieltjes transform of the spectral distribution of $\hat{C}$ to its large n, p limit along with conditions guaranteeing the possibility to recover the signal m from the asymptotic statistical properties of $\hat{C}$.

2. Main Result.

Let us note, for $k \in \mathbb{N}$, $[k] \equiv \{1, \ldots, k\}$; $\mathbb{R}^+ \equiv \{x \in \mathbb{R}, x > 0\}$; $\mathcal{M}_{p,n}$, 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 $M \in \mathcal{M}_{p,n}$ and the Frobenius norm $\|M\|_F = \sqrt{\sum_{\substack{1\leq i\leq p\\ 1\leq j\leq n}} M_{i,j}^2}$. We further note $\mathcal{D}_n \equiv \{\Delta \in \mathcal{M}_n \mid i \neq j \Leftrightarrow \Delta_{i,j} = 0\}$, the set of diagonal matrices endowed with the spectral norm of $\mathcal{M}_n \equiv \mathcal{M}_{n,n}$. 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 define then $\mathcal{D}_n^+ \equiv \{\Delta \in \mathcal{D}_n, \forall i \in [n], \Delta_i > 0\}$.

We place ourselves under the random matrix regime where p, the size of data $x_1, \ldots, x_n \in \mathbb{R}^p$ is of the same order as n, the number of data – for practical use, imagine that $10^{-2} \leq \frac{p}{n} \leq 10^2$. The convergence results will be expressed as functions of the quasi asymptotic quantities p and n that are thought of as tending to infinity (in practice our results are extremely accurate already for $p, n \geq 100$). We will then work with the notations $a_{n,p} \leq O(b_{n,p})$ or $a_{n,p} \geq O(b_{n,p})$ to signify that there exists a constant K independent of p and n such that $a_{n,p} \leq K b_{n,p}$ or $a_{n,p} \geq K b_{n,p}$, respectively, and to simplify the notation, most of the time, the indices n, p will be omitted. In particular we have $O(n) \leq p \leq O(n)$. Our hypotheses concern four central objects:

- $Z = (z_1, \ldots z_n) \in \mathcal{M}_{p,n}$ satisfies the concentration of measure phenomenon (to be presented later); all the random vectors $z_1, \ldots, z_n$ are independent and $\sup_{1\leq i\leq n} \|\mathbb{E}[z_i]\| \leq O(1)$;
- $\tau = \mathrm{Diag}(\tau_1, \ldots, \tau_n) \in \mathcal{D}_n^+$ satisfy $\forall i \in [n], \tau_i > 0$ and $\frac{1}{n}\sum_{i=1}^n \tau_i \leq O(1)$;
- $m \in \mathbb{R}^p$ and $\|m\| \leq O(1)$;

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

- $u : \mathbb{R}^+ \to \mathbb{R}^+$ is bounded, $t \mapsto tu(t)$ is non-decreasing, $t \mapsto \frac{u(t)}{t}$ is non-increasing and $\forall t > 0$: $tu(t) < 1$.

Those conditions are sufficient to retrieve part of the statistical properties of Z and of the signal m from the data matrix

$$X = Z\sqrt{\tau} + m\mathbb{1}^T$$

through the robust scatter matrix $\hat{C}$ defined in Equation (1). The standard sample covariance matrix $\frac{1}{n}XX^T$ instead inefficiently estimates some of these statistics due to the presence of possibly large (outlying) τ_i's (although $\frac{1}{n}\sum_{i=1}^n \tau_i \leq O(1)$, it is allowed for some τ_i's to be of order $\tau_i \geq O(n)$). The robust scatter matrix controls this outlying behavior by mitigating the impact of the high energy data x_i with the tapering action of the mapping u induced by the hypothesis $tu(t) < 1$ (see Figure 1).

Introducing the diagonal matrix $\hat{\Delta}$ solution to the fixed point equation:

$$\hat{\Delta} = \frac{1}{n} x_i^T \left(\frac{1}{n} X^T u(\hat{\Delta}) X + \gamma I_p \right)^{-1} x_i,$$

(with $u(\cdot)$ operating entry-wise on the diagonal elements of $\hat{\Delta}$) the robust scatter matrix is simply $\hat{C} = \frac{1}{n}Xu(\hat{\Delta})X^T$, and the tapering action is revealed by low values of $u(\tilde{\Delta})_{ii}$ when τ_i is large. As shown on the central display of Figure 1, compared to $\frac{1}{n}XX^T$, $\hat{C} = \frac{1}{n}Xu(\hat{\Delta})X^T$ has a cleaner spectral behavior which lets appear the signal induced by m as an isolated eigenvalue-eigenvector pair. This eigenvector can then be exploited to estimate m (this is a classical random matrix inference problem, which however is beyond the scope of the present article).

This paper precisely shows that the spectral distribution of $\hat{C}$ is asymptotical equivalent to the spectral distribution of $\frac{1}{n}Z^T UZ$ where U is a deterministic diagonal matrix satisfying $\|U\| \leq O(1)$. Interestingly, the definition of U merely depends on the second order moments of $z_1, \dots z_n$ which we denote, $\forall i \in [n]$, $C_i \equiv \mathbb{E}[z_i z_i^T]$, on the vector $\tau \in \mathbb{R}^n$ of the τ_i's, on the function u, but not on the signal m. The definition of U relies on the introduction of a function $\eta : \mathbb{R}^+ \to \mathbb{R}^+$ derived from u and defined as the solution to

$$\forall t \in \mathbb{R}^+ : \quad \eta(t) = \frac{t}{1 + tu(\eta(t))}$$

and on the diagonal matrix $\Lambda_z : \mathcal{D}_n^+ \to \mathcal{D}_n^+$. For any $z \in \mathbb{R}^+$ and $\Delta \in \mathcal{D}_n^+$, $\Lambda_z(\Delta)$ is defined as the unique solution to the n equations:

$$\forall i \in [n], \ \Lambda_z(\Delta)_i = \frac{1}{n} \operatorname{Tr} \left(C_i \left(\frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^n \frac{C_j \Delta_j}{1 + \Delta_j \Lambda_z(\Delta)_j} + zI_p \right)^{-1} \right).$$

Introducing the Stieltjes transform $m(z) = \frac{1}{p}\operatorname{Tr}((\hat{C} - zI_p)^{-1})$ of the spectral measure of $\hat{C}$, for $z < 0$, we have the concentration:

THEOREM 2.1. *For any $z \geq O(1)$, there exist two constants $C, c > 0$ ($C, c \sim O(1)$) such that, for any $\varepsilon > 0$, $\varepsilon \leq 1$,*

$$\mathbb{P}\left(\left| m(-z) - \frac{1}{p} \operatorname{Tr} \left(\frac{1}{n} \sum_{i=1}^n \frac{U_i C_i}{1 + \Lambda_z(U)_i U_i} + zI_p \right)^{-1} \right| \geq \varepsilon \right) \leq Ce^{-cn\varepsilon^2/\log(n)}$$

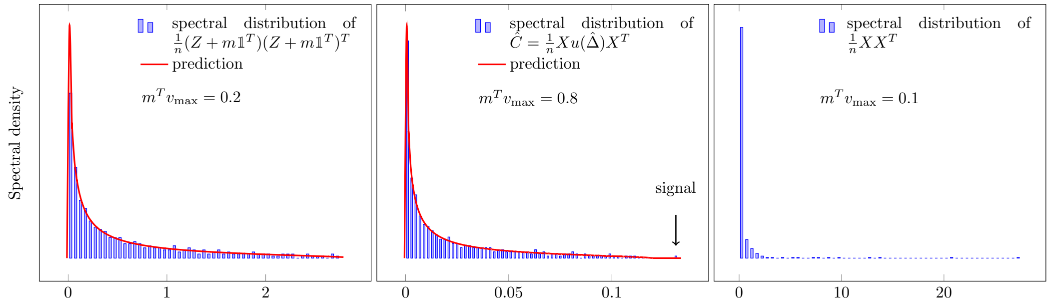

FIG 1. *Spectral distributions of the matrices $\frac{1}{n}(Z + m\mathbb{1}^T)(Z + m\mathbb{1}^T)^T$, $\hat{C}$ and $\frac{1}{n}XX^T$ against their large dimensional prediction; $p = 500$, $n = 400$ (null eigenvalues removed), $u : t \mapsto \min(t, \frac{1}{1+5t})$, the variables $\tau_1, \ldots, \tau_n$ are drawn independently from a Student distribution with 1 degree of freedom, $m = \mathbb{1} \in \mathbb{R}^p$; $Z = \sin(W)$ for $W \sim \mathcal{N}(0, AA^T)$ where $A \in \mathcal{M}_p$ is a fixed matrix whose entries are drawn from the Gaussian distribution with zero mean and unit variance ($Z \propto \mathcal{E}_2$ by construction). The population covariance and mean of Z are computed with a set of p^2 independent realizations of Z. The values of the projections of the signal m against the eigenvector v_{max} associated to the largest eigenvalue reveals that, with the robust scatter approach, the diverging action of τ in the model can be turned into an advantage to infer the signal m from the data. The choice of the mapping u is not optimized, our goal here is just to show that non monotonic functions are suited to robust statistics as long as they satisfy our assumptions.*

where $U = \mathrm{diag}(U_1, \ldots, U_n) \in \mathcal{D}_n^+$ satisfies $U \leq O(1)$ and is the unique solution to the equation:

$$U = \tau \cdot u \circ \eta\left(\tau \Lambda_\gamma(U)\right)$$

(the mappings u and η are applied entry-wise on the diagonal terms on $\mathcal{D}_n^+$).

The theorem states in particular that, under our present hypotheses, the robust scatter matrix is a concentrated object, the spectral distribution of which can be predicted. This is confirmed in Figure 1 which 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 scatter matrix (i.e., it is here the sample covariance matrix of the equivalent data $Xu(\hat{\Delta})^{1/2}$), and (iii) without any modification on X. For the two first spectral distributions, we displayed their estimation with the Stieltjes transform as per Theorem 2.1.

3. Preliminaries for the study of the resolvent. 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 $(\mathcal{S}_p)^n$ which will be denoted for simplicity $\mathcal{S}_p^n$. Given $S \in \mathcal{S}_p^n$, we finally let $S_1, \ldots, S_n \in \mathcal{S}_p$ be its n components.

Given two sequences of scalars $a_{n,p}, b_{n,p}$, the notation $a_{n,p} \sim O(b_{n,p})$ means that $a_{n,p} \leq O(b_{n,p})$ and $a_{n,p} \geq O(b_{n,p})$. 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 the main results are presented progressively throughout the article so the reader easily understands their importance and direct implications. A full recollection of all these assumptions is provided at the beginning of the appendix.

3.1. *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):

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

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 understand the behavior of the resolvent $Q_\gamma(S, \Delta)$ for general $S \in \mathcal{S}_p^n$ and $\Delta \in \mathcal{D}_n^+$. We document in this subsection its contracting properties.

As the scalar γ will rarely change in the remainder, it will be sometimes omitted for readability.

LEMMA 3.1. *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 $S \in \mathcal{S}_p^n$, further define the mapping $I_\gamma : \mathcal{S}_p^n \times \mathcal{D}_n^+ \to \mathcal{D}_n^+$,

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

With the notation $I_\gamma^X(\Delta) \equiv I(X \cdot X^T, \Delta)$, the fixed point $\hat{\Delta}$ defined in (2) is simply $\hat{\Delta} = I_\gamma^X(u(\hat{\Delta}))$. To prove the existence and uniqueness of $\hat{\Delta}$ we exploit the Banach fixed-point theorem to find contracting properties on the mapping $\Delta \mapsto I_\gamma^X(u(\Delta))$ for which $\hat{\Delta}$ is a fixed point. As we see in the following lemma, the contractive character does not appear relatively to the spectral norm on $\mathcal{D}_n^+$ but relatively to another metric which will be later referred to as the "stable semi-metric".

LEMMA 3.2. *Given $S \in \mathcal{S}_p^n$ and $\Delta, \Delta' \in \mathcal{D}_n^+$, we have (the index γ being omitted)*

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

PROOF. Given $a \in \{1, \ldots, k\}$, we can bound thanks to Cauchy Shwarz inequality:

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

$$\leq \frac{1}{n}\sqrt{\sum_{b=1}^{k}\mathrm{Tr}\left(S_a Q_\gamma(S,\Delta)\frac{S_b\left|\Delta'_b-\Delta_b\right|}{\sqrt{\Delta_b\Delta'_b}}\Delta_b Q_\gamma(S,\Delta)\right)}$$

$$\cdot\sqrt{\sum_{b=1}^{k}\mathrm{Tr}\left(S_a Q_\gamma(S,\Delta')\frac{S_b\left|\Delta'_b-\Delta_b\right|}{\sqrt{\Delta_b\Delta'_b}}\Delta'_b Q_\gamma(S,\Delta')\right)}$$

$$\leq \left\|\frac{\Delta'-\Delta}{\sqrt{\Delta\Delta'}}\right\|\sqrt{\frac{1}{n}\mathrm{Tr}\left(S_a Q_\gamma(S,\Delta)\left(1-\gamma Q_\gamma(S,\Delta)\right)\right)}$$

$$\cdot\sqrt{\frac{1}{n}\mathrm{Tr}\left(S_a Q_\gamma(S,\Delta')\left(1-\gamma Q_\gamma(S,\Delta')\right)\right)}$$

$$< \left\|\frac{\Delta'-\Delta}{\sqrt{\Delta\Delta'}}\right\|\sqrt{\tilde{I}(S,\Delta)_a\tilde{I}(S,\Delta')_a}$$

□

If one sees the term $\left\|\frac{\Delta-\Delta'}{\sqrt{\Delta\Delta'}}\right\|$ as a distance between Δ and Δ', then Lemma 3.2 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 the next subsection a precise description of such functions that will be called *stable mappings*.

3.2. *The stable semi-metric.* The stable semi-metric which we define here is a convenient object which allows us to set Banach-like fixed point theorems. It has a crucial importance to prove the existence and uniqueness of $\hat{C}$ but also to obtain some random matrix identities on $\hat{C}$, such as the estimation of its limiting spectral distribution.

DEFINITION 3.3. We call the *stable semi-metric* on $\mathcal{D}_n^+ = \{D\in\mathcal{D}_n, \forall i\in[n],\ D_i>0\}$ the function:

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

In particular, this semi-metric can be defined on $\mathbb{R}^+$, identifying $\mathbb{R}^+$ with $\mathcal{D}_1^+$.

REMARK 3.4. 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 $y_1,\ldots,y_{p-1}\in\mathbb{R}^+$, 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 $i\in\{1,\ldots,p-1\}$. 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).

The semi-metric d_s is called stable due to its many interesting stability properties.

PROPERTY 3.5. Given $\Delta, \Delta' \in \mathcal{D}_n^+$ and $\Lambda \in \mathcal{D}_n^+$:

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

DEFINITION 3.6. 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 $f_/, f_\cdot : \mathbb{R}^+ \to \mathbb{R}^+$ which will characterize the stable class:

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

PROPERTY 3.7. A function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a stable mapping if and only if $f_/$ is non-increasing and $f_\cdot$ is non-decreasing.

PROOF. 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:

$$f(y) - f(x) \leq \frac{f(y)}{y}(y - x) \qquad \text{and} \qquad f(y) - f(x) \leq \frac{f(x)}{x}(y - x) \tag{4}$$

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

$$f(x) - f(y) \leq \frac{f(y)}{x}(y - x) \qquad \text{and} \qquad f(x) - f(y) \leq \frac{f(x)}{y}(y - x) \tag{5}$$

Now if $f(y) \geq f(x)$, we can take the root of the product of the two inequalities of (4) and if $f(y) \leq f(x)$, we take the root of the product of the two inequalities of (5), 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:

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

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 3.8. Given $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$, we can introduce the mappings $f_/, f_\cdot : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ defined with:

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

It is possible to inspire from Property 3.7 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 non-increasing and $f_\cdot$ is non-decreasing. 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 $I^M, \tilde{I}^S \in \mathcal{S}(\mathcal{D}_n^+)$.

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

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

Before proving Proposition 3.9, let us give two preliminary results.

LEMMA 3.10. *Given four positive numbers $a, b, c, d \in \mathbb{R}^+$:*

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

PROOF. For the first result, we deduce from the inequality $2ab\alpha\beta \le a\alpha + b\beta$:

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

For the second result, we simply bound:

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

$\square$

PROOF OF PROPERTY 3.9. 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 $f_\cdot$ given in Property 3.7 (for that reason, this property is easier to show on the set $S_w(\mathcal{D}_n^+)$ described in Remark 3.8). Nonetheless, given $f, g \in \mathcal{S}(\mathcal{D}_n^+)$ and $\Delta, \Delta' \in \mathbb{R}^+$ there exists $i_0 \in [n]$ such that:

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

thanks to Lemma 3.10 and the stable character of f and g. □

3.3. *Fixed Point theorem for stable mappings.* The Banach 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 since they rely on classical topological inferences.

PROPERTY 3.11. The semi-metric space $(\mathcal{D}_n^+, d_s)$ is complete.

We can now give the central result that will justify the definition of the robust scatter matrix C but also of the deterministic diagonal matrix U introduced in Section 2.

THEOREM 3.12. *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 (in $\mathcal{D}_n^+$), there exists a unique fixed point $\Delta^* \in \mathcal{D}_n^+$ satisfying $\Delta^* = f(\Delta^*)$.*

It is possible to relax a bit the contracting hypotheses on f if one supposes that f is monotonic. We express rigorously this result in next Theorem, but it will not be employed in our paper since we preferred to assume u bounded to obtain the contracting properties of the fixed point satisfied by $\hat{\Delta}$.

THEOREM 3.13. *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:*

$$\forall \Delta, \Delta' \in \mathcal{D}_n^+ : \ d_s(f(\Delta), f(\Delta')) < d_s(\Delta, \Delta') \tag{6}$$

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

PROOF. We first suppose that f is non-decreasing. 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 non-decreasing 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 (6).

Now if f is non-increasing 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 (6). □

To employ Theorem 3.12 to the fixed point equation satisfied by $\hat{\Delta}$, a first step is to look at $\Delta \mapsto I_\gamma(S, \Delta) = \mathrm{Diag}(\mathrm{Tr}(S_i Q(S, \Delta)))_{1\leq i\leq n}$ that we know to be stable from Lemma 3.2. The bounding assumption of Theorem 3.12 is issued from the following preliminary Lemma.

LEMMA 3.14. *Given $S \in \mathcal{S}_p^n$ and a function $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ bounded by $f_0 \in \mathcal{D}_k^+$,*

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

Combined with Lemma 3.2, 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 3.12.

COROLLARY 3.15. Given $f, g \in \mathcal{S}(\mathcal{D}_n^+)$ with f bounded, and a family of non-negative and non-zero 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 $\mathcal{D}_n^+$.

PROOF. We saw in Lemma 3.2 that:

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

Now, thanks to Lemma 3.14, we can bound:

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

and therefore $g \circ \tilde{Q}^S \circ f$ is contracting and bounded from below and above: we can employ Theorem 3.12 to set the existence and the uniqueness of a solution $\Delta \in \mathcal{D}_n^+$ to $\Delta = g \circ \tilde{Q}^S \circ f(\Delta)$. □

We will thus suppose from here on that u is a stable function to be able to use Corollary 3.15 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 3.16. For $X \in \mathcal{M}_{p,n}$, there exists a unique diagonal matrix $\hat{\Delta} \in \mathcal{D}_n^+$ such that

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

Now that $\hat{\Delta}$ is perfectly defined, let us introduce additional assumptions to be able to infer concentration properties on $\hat{\Delta}$.

3.4. *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 $\mu_i \equiv \mathbb{E}[z_i] \in \mathbb{R}^p$, their second order statistic matrix (or non-centered covariance matrix) $C_i \equiv \mathbb{E}[z_i z_i^T]$ and their covariance matrices $\Sigma_i = C_i - \mu_i \mu_i^T \in \mathcal{M}_p$. 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.

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

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

To measure the speed of concentration we generally express it with the dimension of the vector space E of W. That leads us to introducing an extra parameter s on which depends the random vector W (and the vector space E, most often $s = \dim(E)$ but not always). Precisely, we will not define the "concentration of a random vector" but the "concentration of a sequence of random vectors". The relevant parameter to express the concentration of the random matrix Z introduced in this section is $s = (n, p) \in \mathbb{N}^2$. We will also have to express the concentration of diagonal matrices for which the relevant parameter is the number of diagonal elements. For these reasons, we present a definition as general as possible with a set of indexes "S" to be specified depending on the applications.

DEFINITION 3.17. Given a set of indexes S, a sequence of normed vector spaces $(E_s, \|\cdot\|_s)_{s\in S}$, a sequence of random vectors $(Z_s)_{s\in S} \in \prod_{s\in S} E_s$, a sequence of positive reals $(\sigma_s)_{s\in S} \in \mathbb{R}_+^{\mathbb{N}}$ 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 two constants (i.e. independent of s) $C, c > 0$, such that, $\forall s \in S$, and $\forall t > 0$,
$$\mathbb{P}\left(\left|f_s(Z_s) - f_s(Z_s')\right| \geq t\right) \leq Ce^{(t/c\sigma_s)^q}$$
- there exist two constants $C, c > 0$ such that, for all $s \in S$ and for all $t > 0$,
$$\mathbb{P}\left(\left|f_s(Z_s) - \mathbb{E}[f_s(Z_s)]\right| \geq t\right) \leq Ce^{(t/c\sigma_s)^q}$$

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 $W \propto \mathcal{E}_q(\sigma)$). If $\sigma_s \leq O(1)$[2], one can further write $Z_s \propto \mathcal{E}_q$.

The essential result which motivates the definition is the concentration of Gaussian vectors.

THEOREM 3.18 ([6]). *Given a sequence of deterministic vectors $\mu_p \in \mathbb{R}^p$, if $W_p \sim \mathcal{N}(\mu_p, I_p)$ then $W_p \propto \mathcal{E}_2$.*

Having access to at least one class of concentrated vectors, let us state four important properties to keep in mind when dealing with these vectors. First, the class of random vectors is stable through Lipschitz maps:

PROPOSITION 3.19. Given two sequence of normed vector spaces $(E_1, \|\cdot\|_1)$ and $(E_2, \|\cdot\|_2)$, a sequence of random vectors $W \in E_1$, two sequences $\sigma, \lambda \in \mathbb{R}_+$ and a sequence of $O(\lambda)$-Lipschitz function $\phi : E_1 \to E_2$:[3]
$$W \propto \mathcal{E}_q(\sigma) \qquad \Longrightarrow \qquad \phi(W) \propto \mathcal{E}_q(\lambda\sigma).$$

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

PROPOSITION 3.20. Let $W \in E$. Then $W \propto \mathcal{E}_q(\sigma)$ iff there exists a constant $C > 0$ such that, for any 1-Lipschitz mapping $f : E \to \mathbb{R}$,
$$\forall r > q: \ \mathbb{E}\left[|f(W) - \mathbb{E}[f(W)]|^r\right] \leq C\left(\frac{r}{q}\right)^{\frac{r}{q}} \sigma^r.$$

---

[2]The notation $a_s = O(b_s)$ signifies that there exists a constant K (independent of s) such that $\forall s \in S$, $a_s \leq Kb_s$. The same way, $a_s \geq O(b_s)$ means that there exists a constant $\kappa > 0$ such that $a_s \geq \kappa b_s$ and the notation $a_s \sim O(b_s)$ is equivalent to $a_s \leq O(b_s)$ and $a_s \geq O(b_s)$

[3]The statement "ϕ is $O(\lambda)$-Lipschitz" means here that there exists $K \leq O(1)$ such that, for all $s \in S$, ϕ_s is $(K\lambda_s)$-Lipschitz.

Third, 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 $W \in \mathbb{R}$ is a random *scalar* and satisfies $W \propto \mathcal{E}_q(\sigma)$, we will use the notation $W \in \tilde{W} \pm \mathcal{E}_q(\sigma)$ if $|\tilde{W} - \mathbb{E}W| \leq O(\sigma)$ (of course, in particular $W \in \mathbb{E}W \pm \mathcal{E}_q(\sigma)$). Concentration inequalities for operations on concentrated *vectors* express similarly but will not be needed in this work (more information is available in [7]).

PROPOSITION 3.21. Let $W_1, W_2 \in \mathbb{R}$ be two sequence of random variables, $\sigma_1, \sigma_2 \in \mathbb{R}_+^S$ two sequences of positive reals and $\tilde{W}_1, \tilde{W}_2 \in \mathbb{R}^{\mathbb{N}}$ two sequences of scalars. Then, if $W_1 \in \tilde{W}_1 \pm \mathcal{E}_q(\sigma_1)$ and $W_2 \in \tilde{W}_2 \pm \mathcal{E}_q(\sigma_2)$,

$$W_1 + W_2 \in \tilde{W}_1 + \tilde{W}_2 \pm \mathcal{E}_q(\sigma_1 + \sigma_2)$$

$$W_1 W_2 \in \tilde{W}_1 \tilde{W}_2 \pm \mathcal{E}_q(\sigma_1 |\tilde{W}_2| + \sigma_2 |\tilde{W}_1|) + \mathcal{E}_{q/2}(\sigma_1 \sigma_2).$$

Forth, when dealing with a concentrated random variable W the deterministic scalar around which happens the concentration can be chosen indifferently in an interval of diameter equal to the observable diameter of W.

LEMMA 3.22. *Given a (sequence of) random variables W, three (sequences of) deterministic scalars $\tilde{W}_1, \tilde{W}_2 \in \mathbb{R}$, $\sigma > 0$, if $W \in \tilde{W}_1 \pm \mathcal{E}_q(\sigma)$ and $|\tilde{W}_1 - \tilde{W}_2| \leq O(\sigma)$ then $W \in \tilde{W}_2 \pm \mathcal{E}_q(\sigma)$.*

We finally complete this short probabilistic introduction of concentration of measure theory with key 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 3.23. *Let $W \in \mathbb{R}^p$. Then, if $W \propto \mathcal{E}_2$ in $(\mathbb{R}^p, \|\cdot\|)$,*

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

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

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

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

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

As a 1-Lipschitz projection of $Z \propto \mathcal{E}_2$, $z_i \propto \mathcal{E}_2$, and we can then conclude from Propositions 3.20 and 3.21 that $\sup_{1 \leq i \leq n} \|\Sigma_i\| \leq O(1)$ since $\forall u \in \mathbb{R}^p$ such that $\|u\| \leq 1$, $u^T \Sigma_i u = \mathbb{E}[u^T z_i z_i^T u - \mathbb{E}[u^T z_i]\mathbb{E}[z_i^T u]] \leq O(1)$. But we also need to bound μ_i to control $E[z_i z_i^T] = \Sigma_i + \mu_i \mu_i^T$.

ASSUMPTION 5. $\sup_{1 \leq i \leq n} \|\mu_i\| \leq O(1)$.

3.5. *Deterministic equivalent of the resolvent.* Given $\Delta \in \mathcal{D}_n^+$, the resolvent $Q_\gamma^Z(\Delta) = Q(Z \cdot Z^T, \Delta) = (\frac{1}{n} Z\Delta Z^T + \gamma I_p)^{-1}$ is a random matrix which exhibits useful properties to understand the statistics of Z and more importantly its spectral behavior. In particular, the distribution of the singular values of Z strongly relates to the well-known *Stieltjes transform* $m_Z(z) = \frac{1}{p}\operatorname{Tr}(Q_{-z}^Z(\Delta)$ where z in a complex value distinct from any of the singular values of Z. The function $m_Z(z)$ has been extensively studied in [7] when Δ is the identity matrix, but since set of assumptions allow the data $z_1, \dots, z_n$ to have different distributions, the results are easily adaptable, studying in the present case the data $\Delta_1 z_1, \dots, \Delta_n z_n$ in place of the data $z_1, \dots, z_n$. For the sake of completeness, the proof is reported in Appendix. We will study $Q_{-z}^Z(I_n)$ in the specific case where $z < 0$ and for that reason, we further look at $Q \equiv Q_z^Z(\Delta)$ for $z > 0$ and even $z \geq O(1)$ (for concentration issues).

It can be shown that Q is a $\frac{2\|\Delta\|^{1/2}}{z^{3/2}\sqrt{n}}$-Lipschitz transformation of Z and, therefore, assuming that $\frac{1}{z} \leq O(1)$, we can deduce that:

$$Q \propto \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \tag{7}$$

As shown subsequently, 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 $\|A\|_* \equiv \operatorname{Tr}(\sqrt{AA^T}) \leq O(1)$ ($\|\cdot\|_*$ 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 3.21.

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

PROPOSITION 3.24. For any $S \in \mathcal{S}_p^n$, the mapping $\Delta \mapsto \check{I}(S, \Delta)$ satisfying

$$\check{I}(S, \Delta) = \frac{1}{n}\operatorname{Diag}(\operatorname{Tr}(S_i Q(S_{-i}, \Delta))_{1 \leq i \leq n}, \quad \text{for } S_{-i} = (S_1, \dots, S_{i-1}, 0, S_{i+1}, \dots, S_n) \tag{8}$$

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

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 3.2. Then we apply an result analogous to Corollary 3.15 (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 with $S_i = \frac{1}{n} C_i$ (for $i \in [n]$) to obtain the existence and uniqueness of $\Lambda \in \mathcal{D}_n^+$ satisfying $\Lambda = \check{I}_z(C, f(\Lambda))$. $\square$

The fixed-point equation $\Lambda^C = \check{I}(C, \frac{\Delta}{\Delta + \Lambda^C})$, for $C = (C_1, \dots, C_n) \in \mathcal{S}_p^n$, allows us to compute $\Lambda^C(\Delta)$ iteratively via the standard fixed-point algorithm. The deterministic equivalent $\tilde{Q}_z^C(\Delta)$ of $Q_z^Z(\Delta)$ is then easily computed and is defined as follows:

$$\tilde{Q}_z^C(\Delta) \equiv Q_z\left(C, \frac{\Delta}{I_n + \Delta\Lambda^C}\right) = \left(\frac{1}{n}\sum_{i=1}^n \frac{\Delta_i C_i}{1 + \Delta_i \Lambda_i^C} + zI_p\right)^{-1}.$$

THEOREM 3.25. *Let $\Delta \in \mathcal{D}_n^+$ and $A \in \mathcal{M}_p$ be deterministic matrices such that $\|D\| \leq O(1)$ and $\|A\|_* \leq O(1)$, where $\|A\|_* = \mathrm{Tr}((AA^T)^{1/2})$. Then we have the concentration,*

$$\mathrm{Tr}(AQ_z^Z(\Delta)) \in \mathrm{Tr}\left(A\tilde{Q}_z^C(\Delta)\right) \pm \mathcal{E}_2\left(\sqrt{\frac{\log n}{n}}\right).$$

This theorem will later allow us to estimate the Stieltjes transform of the spectral distribution of $\hat{C}$ given at the beginning of the article. For this purpose, we need the next corollary to predict the asymptotic behavior of $\hat{\Delta}$ defined in (2). Recall that $I^X : \Delta \mapsto I(X \cdot X^T, \Delta)$. Then the following holds.

COROLLARY 3.26. For all $\Delta \in \mathcal{D}_n^+$ with $\|\Delta\| \leq O(1)$,

$$\left\|\mathbb{E}[I^Z(\Delta)] - \frac{\Lambda^C(\Delta)}{I_n + \Delta\Lambda^C(\Delta)}\right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right).$$

REMARK 3.27. It is possible to extend the results of Theorem 3.25 and Corollary 3.26 to the broader case (useful later) where each mean μ_i ($i \in [n]$) can be decomposed as the sum of a particular component $\mathring{\mu}_i$ of low energy (i.e. with a low norm) and a bigger component proportional to a general signal s of high energy as follows:

$$\mu_i = \mathring{\mu}_i + t_i s, \tag{9}$$

where $t_1, \ldots, t_n > 0$ are n scalars satisfying $\frac{1}{n}\sum_{i=1}^n t_i \geq 1$ and $\sup_{i \leq 1 \leq n} t_i \leq O(1)$. The concentration results are then exactly the same.

4. Estimation of the robust scatter matrix.

4.1. *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 $z_1, \ldots, z_n$ are the random vectors presented in the previous section. For $X = (x_1, \ldots, x_n) \in \mathcal{M}_{p,n}$, we write $X = Z\tau^{\frac{1}{2}} + m\mathbb{1}^T$ where $\tau \equiv \mathrm{Diag}(\tau_i)_{1 \leq i \leq n} \in \mathcal{D}_n^+$ and $\mathbb{1} \equiv (1, \ldots, 1) \in \mathbb{R}^n$. 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 4.1. *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 $\Delta' \sim O(1)$ (i.e. that $\Delta \geq O(1)$ and $\Delta \leq O(1)$), that f is contracting for the stable semi-metric around Δ' with a Lipschitz parameter $\lambda < 1$, and that[4]

$$1 - \lambda - \left\|\sqrt{\frac{f(\Delta') - f'(\Delta')}{\Delta'}}\right\| \geq O(1).$$

[4]In the application of the Theorem present in our paper, we are either in cases where $\|f(\Delta') - f'(\Delta')\| \underset{p,n\to\infty}{\to} 0$ for Δ and Δ' deterministic (Proposition 4.9) or in cases where Δ is random but for any $K \geq 0$, with very high probability, $\|f(\Delta') - f'(\Delta')\| \leq K$ (Proposition 4.8). In both cases, we are thus left to verifying that $1 - \lambda \geq O(1)$.

Then, there exists a constant $K \leq O(1)$ such that

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

PROOF. Let us first bound:

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

(one must be careful here that the stable semi-metric does not satisfy the triangular inequality). Besides:

$$\begin{aligned} \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta \Delta'}} \right\| &\leq \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta}} \left(\frac{\sqrt{f(\Delta')} - \sqrt{f'(\Delta')}}{\sqrt{\Delta'}\sqrt{f(\Delta')}} \right) \right\| + \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta f(\Delta')}} \right\| \\ &\leq \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta f(\Delta')}} \right\| \left(1 + \left\| \sqrt{\frac{f(\Delta') - f'(\Delta')}{\Delta'}} \right\| \right). \end{aligned}$$

Thus, by hypothesis, setting $K' = \frac{1}{1-\lambda-\varepsilon} \leq O(1)$, we have the inequality:

$$\left\| \frac{\Delta - \Delta'}{\sqrt{\Delta f(\Delta')}} \right\| \leq K' \left\| \frac{f(\Delta') - \Delta'}{\sqrt{\Delta f(\Delta')}} \right\|.$$

Thus, as $O(1) \leq \|\Delta'\| - O(a_s) \leq f(\Delta') \leq \|\Delta'\| + O(a_s) \leq O(1)$, we obtain the bound:

$$\left\| \frac{\Delta - \Delta'}{\sqrt{\Delta}} \right\| \leq K'' \left\| \frac{f(\Delta') - \Delta'}{\sqrt{\Delta}} \right\| \tag{10}$$

for some constant $K'' > 0$. We are left to bound from below and above $\|\Delta\|$ to recover the result of the theorem from (10). Considering the index i_0 such that $\Delta_{i_0} = \min(\Delta_i)_{1 \leq i \leq n}$, we have:

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

so that $\Delta_{i_0} \geq \Delta'_{i_0} - O(a_s) \geq O(1)$. On the other hand, one can bound again from (10):

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

As a consequence, $\Delta \sim O(1)$, and we can conclude from (10). □

Theorem 4.1 can be employed when Δ is random and Δ' is a deterministic equivalent (yet to be defined). If we let $f = I^X \circ u(\cdot)$ (and thus $\Delta = \hat{\Delta}$), it is not possible to state that $\Delta \sim O(1)$ since $I^X \circ u(\cdot) = \mathrm{Diag}(\frac{1}{n} x_i^T R^X \circ u(\cdot) x_i)_{1 \leq i \leq n}$ scales with τ which might be unbounded. For this reason, in place of $\hat{\Delta}$, we will consider $\hat{D} \equiv \frac{\hat{\Delta}}{\underline{\tau}}$ where:

$$\underline{\tau} \equiv \mathrm{Diag}(\max(\tau, 1)) \geq I_n.$$

We similarly denote $\bar{\tau} \equiv \mathrm{Diag}(\min(\tau, 1)) \leq I_n$.

4.2. *Definition of $\tilde{D}$, the deterministic equivalent of $\hat{D}$.* The matrix $\hat{D} \equiv \frac{\hat{\Delta}}{\underline{\tau}}$ satisfies the fixed point equation

$$\hat{D} = I^{\bar{Z}}(u^\tau(\hat{D})), \quad \text{where} \quad \bar{z}_i \equiv \frac{x_i}{\sqrt{\underline{\tau}_i}} = \sqrt{\overline{\tau}_i} z_i + \frac{m}{\sqrt{\underline{\tau}_i}} \quad \text{and} \quad u^\tau : \Delta \mapsto \underline{\tau} u(\underline{\tau}\Delta).$$

We will note from now on $\bar{m}_i = \mathbb{E}[\bar{z}_i]$ and $\bar{C}_i = \mathbb{E}[\bar{z}_i \bar{z}_i^T]$. In order to apply Corollary 3.26 with the hypothesis described in Remark 3.27, we will need a bound on the energy of the signal and on the τ_i's.

ASSUMPTION 6. $\|m\| = O(\sqrt{n})$.

ASSUMPTION 7. $\frac{1}{n}\sum_{i=1}^n \tau_i \leq O(1)$.

PROPOSITION 4.2. $\bar{Z} \propto \mathcal{E}_2$, $\sup_{1\leq i\leq n} \|\bar{C}_i\| \leq O(1)$ and $\bar{m}_i$ satisfies the hypotheses discribed in Remark 3.27.

PROOF. The concentration of $\bar{Z}$ is just a consequence of Assumption 4 and the 1-Lipschitz character of the mapping $M \mapsto M\sqrt{\overline{\tau}_i} + m\mathbb{1}^T \underline{\tau}_i^{-1/2}$. The bounds $\sup_{1\leq i\leq n} \|\bar{C}_i\| \leq O(1)$, $\|\mathring{\bar{m}}_i\| \leq O(1)$ and $\|m\| \leq O(\sqrt{n})$ are immediate from Assumption 5 and 6 and we have, by convexity of $t \mapsto \frac{1}{t}$, the bound $\frac{1}{n}\sum \frac{1}{\sqrt{\underline{\tau}_i}} \geq \frac{n}{\sum \underline{\tau}} \geq O(1)$ thanks to Assumption 7 ($\sum \underline{\tau} \leq \sum \tau + n \leq O(n)$). $\square$

We still cannot apply Corollary 3.26 since $\|u^\tau(\hat{D})\|$ is possibly unbounded. Still, let us assume for the moment that $\|u^\tau(\hat{D})\|$ is indeed bounded: then, following our strategy, we are led to introducing a deterministic diagonal matrix $\tilde{D}$ ideally approaching $\hat{D}$ and satisfying

$$\tilde{D} = \frac{\Lambda^{\bar{C}}(u^\tau(\tilde{D}))}{I_n + u^\tau(\tilde{D})\Lambda^{\bar{C}}(u^\tau(\tilde{D}))}, \tag{11}$$

(where we recall $\Lambda^{\bar{C}}(u^\tau(\tilde{D})) = \check{I}(\bar{C}, \frac{u^\tau(\tilde{D})}{I_n + u^\tau(\tilde{D})\Lambda^{\bar{C}}(u^\tau(\tilde{D}))}))$ Before proving the validity of the estimate $\tilde{D}$ of $\hat{D}$, let us justify the validity of its definition (i.e., the existence and uniqueness of $\tilde{D}$). To this end, we first introduce a stable auxiliary mapping $\eta : \mathbb{R}_+ \to \mathbb{R}_+$.

PROPOSITION 4.3. 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. It is a simple application of Theorem 3.12. 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 3.12 since f is contracting for the stable semi-metric:

$$d_s(f(\eta), f(\eta')) = d_s\left(\frac{1}{f(\eta)}, \frac{1}{f(\eta')}\right) = \frac{|u(\eta) - u(\eta')|}{\sqrt{\left(\frac{1}{x} + u(\eta)\right)\left(\frac{1}{x} + u(\eta')\right)}}$$

$$\leq \sqrt{\frac{u(\eta)u(\eta')}{\left(\frac{1}{x}+u(\eta)\right)\left(\frac{1}{x}+u(\eta')\right)}}d_s(u(\eta),u(\eta'))$$
$$\leq \sqrt{\frac{1}{\frac{1}{u(\eta)u(\eta')x^2}+\frac{1}{u(\eta')x}+\frac{1}{u(\eta)x}+1}}d_s(u(\eta),u(\eta'))$$
$$\leq \frac{1}{1+\frac{1}{u^\infty x}\left(1+\frac{1}{u^\infty x}\right)}d_s(\eta,\eta').$$

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 3.7. 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 non-decreasing,

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

Similarly, 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 3.7, $\eta \in \mathcal{S}(\mathbb{R}^+)$. □

The first equation of (11) can be rewritten $\tilde{D} = \eta_\tau(\Lambda^{\bar{C}}(u^\tau(\tilde{D})))$, with $\eta_\tau : x \mapsto \frac{\eta(\tau x)}{\tau}$. To define $\tilde{D}$ properly, we thus need to show that $\Lambda^{\bar{C}}$ is stable (with the aim of employing Theorem 3.12 again).

PROPOSITION 4.4. For any $S \in \mathcal{S}_p^n$, the mapping $\Lambda^S : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ is stable.

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

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

Thanks to Lemma 3.2, the stability rules given in Property 3.9, and the extra tools given by Lemma 3.10 (already used to prove Property 3.9). 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. □

We are thus now allowed to define $\tilde{D}$.

PROPOSITION 4.5. There exists a unique diagonal matrix $\tilde{D} \in \mathcal{D}_n^+$ satisfying (11).

PROOF. We already know from Proposition 4.4 that $D \mapsto \Lambda^{\bar{C}}(u^\tau(D))$ is stable and bounded from above and below (since $u^\tau \leq \|\underline{\tau}\| u^\infty$). The same is true for $\eta_\tau(\Lambda^{\bar{C}}(u^\tau(\tilde{D})))$ 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 $\Lambda^{\bar{C}}(u^\tau(xI_n))$ which is bounded from above and below). The existence and uniqueness of $\tilde{D}$ thus unfold from Theorem 3.12. □

4.3. *Concentration of $\hat{D}$ around $\tilde{D}$.* In order to establish the concentration of $\hat{D}$, we need an assumption on η to be able to bound $\tilde{D} = \eta_\tau(\Lambda^{\bar{C}}(u^\tau(\tilde{D})))$. This assumption is expressed through a condition on u, justified by the following lemma.

LEMMA 4.6. *The mapping $\eta_/$ is bounded from below iff, $\forall t \in \mathbb{R}^+$, $u_\cdot(t) = tu(t) < 1$.*

PROOF. 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 $\lim_{t\to\infty} \frac{\eta(t)}{t} = 0 < \alpha$), there exists a sequence $(x_n)_{n\geq 0} \in \mathbb{R}_+^S$ such that $\eta(x_n) \to \infty$. Thus ($u_\cdot$ being non-decreasing), $\forall t > 0, u_\cdot(t) \leq \lim_{n\to\infty} u(\eta(x_n))\eta(x_n) \leq 1-\alpha$. Conversely, if $\forall t > 0, u_\cdot(t) < 1$, $\forall x \in \mathbb{R}^+$:

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

□

ASSUMPTION 8. There exists $u_\cdot^\infty > 0$ such that for all $t \in \mathbb{R}^+$, $u_\cdot(t) \leq u_\cdot^\infty$.[5]

We complete this extra assumption with an additional "light" condition.

ASSUMPTION 9. $\inf_{1\leq i\leq n} \frac{1}{n} \operatorname{Tr} C_i \geq O(1)$.

These assumptions imply the following important control.

LEMMA 4.7. $\tilde{D} \sim O(1)$.

[5] We implicitly assume here that $1 - u_\cdot^\infty \geq O(1)$, since we introduced our model in such a way that u does not scale with n or p, i.e., no assumption links u with either n or p. For the same reasons, we implicitly assume that $\gamma \sim O(1)$.

PROOF. We already know from our assumptions that $O(1) \leq \frac{1}{n}\operatorname{Tr}(C_i) + \frac{1}{n}\frac{m^T m}{\tau} = \frac{1}{n}\operatorname{Tr}\bar{C}_i \leq O(1)$ and from Proposition 4.4 that, for any $\Delta \in \mathcal{D}_n^+$:

$$\frac{O(1)}{n(\gamma + \frac{1}{n}\|\sum C_i \tau_i u(\tau_i \Delta)\|)} \leq \Lambda^{\bar{C}}(u^\tau(\Delta)) \leq O(1).$$

Thus $\Lambda^{\bar{C}}(u^\tau(\tilde{D})) \sim O(1)$, since $\|\frac{1}{n}\sum C_i \tau_i u(\tau_i \tilde{D})\| \leq u^\infty \|C_i\| \frac{1}{n}\sum_{i=1}^n \tau_i \leq O(1)$. As such, we can bound $\|\tilde{D}\| \leq \|\Lambda^{\bar{C}}(u^\tau(\tilde{D}))\| \leq O(1)$ and:

$$\tilde{D} = \eta_\tau(\Lambda^{\bar{C}}(u^\tau(\tilde{D}))) \geq \eta_/^\infty \Lambda^{\bar{C}}(u^\tau(\tilde{D})) \geq O(1).$$

□

This control allows us to establish the concentration of $\hat{D}$:

PROPOSITION 4.8. There exist two constants $C, c > 0$ ($C, c \sim O(1)$) such that, for any $\varepsilon \in (0,1]$,

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

PROOF. Let us check the hypotheses of Theorem 4.1. Let us first bound the Lipschitz parameter (for the stable semi-metric) λ of $\check{I}^{\bar{Z}} \circ u^\tau$ around $\tilde{D}$ defined as:

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

An inequality similar as in Lemma 3.2 gives us:

$$\lambda \leq \sqrt{\|1 - \gamma Q^{\bar{Z}}(u^\tau(\tilde{D})\|} \leq 1 - \frac{\gamma}{\gamma + \frac{1}{n}\|u^\tau(\tilde{D})\|\|\bar{Z}\bar{Z}^T\|}$$

(thanks to Lemma 3.14). Now, from Proposition 4.7, $u^\tau(\tilde{D}) \leq \frac{u^\infty}{\tilde{D}} \leq O(1)$ and since $\frac{1}{n}\|\bar{Z}\bar{Z}^T\| \leq (\|\bar{Z}\|/\sqrt{n})^2$, we know from Lemma 3.23 that, with probability larger than $1 - Ce^{-cn}$ (for some constants $C, c > 0$), $\|\bar{Z}\| \leq K\sqrt{n}$. Thus there exists a constant $K' > 0$, such that under this highly probable event $1 - \lambda \geq K'$.

We know from Proposition 4.2 that $\bar{Z}u^\tau(\tilde{D}) \propto \mathcal{E}_2$. We may thus employ Lemma C.1 (in the appendix) to get $\check{I}^{\bar{Z}}(u^\tau(\tilde{D}))_i = \frac{1}{n}z_i Q_{-i}^{\bar{Z}}(u^\tau(\tilde{D}))z_i \propto \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_1(1/n)$ and Corollary 3.26 to state that $\|\mathbb{E}[\check{I}^{\bar{Z}}(u^\tau(\tilde{D}))] - \tilde{D}\| \leq O(\sqrt{\log(n)/n})$. Thus there exist two constants $C', c' > 0$ such that

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

We can then choose t small enough ($t = \frac{K'^2}{4\|1/\tilde{D}\|}$) such that on an event of probability larger than $1 - C'e^{-c'' n/\log n}$ ($c'' > 0$), we have:

$$1 - \lambda - \sqrt{\left\|\frac{\check{I}^{\bar{Z}}(u^\tau(\tilde{D})) - \tilde{D}}{\tilde{D}}\right\|} \geq \frac{K'}{2}.$$

We can then apply Theorem 4.1 and choose C and c appropriately to obtain the result of the proposition. □

It is even possible to provide a deterministic equivalent for $\hat{D}$ independent of the signal m if its norm is small enough:

ASSUMPTION 6 bis. $\|m\| \leq O(1)$.

PROPOSITION 4.9. The fixed-point equation $D = \eta_\tau \circ \Lambda^{\bar{\tau}C} \circ u^\tau(D)$ 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 $m = 0$). We want to employ again Theorem 4.1, with the deterministic mappings:

$$f = \eta_\tau \circ \Lambda^{\bar{\tau}C} \circ u^\tau \qquad \text{and} \qquad f' = \eta_\tau \circ \Lambda^{\bar{C}} \circ u^\tau,$$

and with $\Delta = \tilde{D}_{-m}$ and $\Delta' = \tilde{D}$. We note that $\tilde{D} \sim O(1)$ and the Lipschitz parameter λ of f for the semi-metric satisfies a similar inequality as in the proof of Proposition 4.8:

$$1 - \lambda \geq \frac{\gamma}{\gamma + u_\cdot^\infty \|\frac{1}{\tilde{D}}\| \sup \|C_i\|} \geq O(1).$$

We then need to bound the spectral norm $\|\eta_\tau \circ \Lambda^{\bar{\tau}C} \circ u^\tau(\tilde{D}) - \eta_\tau \circ \Lambda^{\bar{C}} \circ u^\tau(\tilde{D})\|$. Note that η is 1-Lipschitz for the absolute value because, for any $x, y \in \mathbb{R}^+$, the stability of η implies:

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

Thus η_τ is also 1-Lipschitz. We are then left to bounding the distance (in spectral norm) between $\Lambda^{\bar{\tau}C} \circ u^\tau(\tilde{D})$ and $\Lambda^{\bar{C}} \circ u^\tau(\tilde{D})$, and we are naturally led to employing a second time Theorem 4.1 since those two values are both fixed points of stable mappings:

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

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 hypothesis is satisfied, $\Lambda^C(u^\tau(\tilde{D})) \sim O(1)$ and λ', the Lipschitz parameter of $\tilde{I}^{\bar{C}}_\Delta$ satisfies $1 - \lambda' \geq O(1)$. Noting for simplicity $\Delta \equiv u^\tau(\tilde{D})$, $\Lambda \equiv \Lambda^{\bar{\tau}C}(\Delta)$ and $\tilde{Q}^S = \tilde{Q}^S(S, \frac{\Delta}{I_n + \Delta\Lambda})$ (for $S = \bar{C}$ or $S = C$), we are left to bounding, for any $i \in [n]$,

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

since $\sum_{i=1}^n \frac{1}{\underline{\tau}_i} \leq n$ ($\forall i \in [n]$, $1/\underline{\tau}_i \leq 1$). Applying twice Theorem 4.1, we retrieve the result of the proposition. □

We can then deduce from this corollary and Theorem 3.25 the estimation of the Stieltjes transform of the spectral distribution of $\hat{C}$ given in Theorem 2.1 setting $U = \bar{\tau} u^\tau(\tilde{D})$ ($\|U\| \leq \frac{\|\bar{\tau}\| u_\cdot^\infty}{\tilde{D}} \leq O(1)$).

5. Conclusion. In this article, we have developed an original framework to study the large dimensional behavior of a family of matrices 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 solutions, which are then applied to complex data models, this 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

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

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

[3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, A.C.Y., 2014.
Generative adversarial nets.
NIPS .

[4] Huber, P.J., 1964.
Robust estimation of a location parameter.
The Annals of Mathematical Statistics 35, 73–101.

[5] Kent, J.T., Tyler, D.E., 1991.
Redescending M-estimates of multivariate location and scatter.
The Annals of Statistics , 2102–2119.

[6] Ledoux, M., 2005.
The concentration of measure phenomenon.
89, American Mathematical Soc.

[7] Louart, C., Couillet, R., 2019.
Concentration of measure and large random matrices with an application to sample covariance matrices.
submitted to Random Matrices: Theory and Applications .

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

[9] Ollila, E., Palomar, D.P., Pascal, F., 2020.
M-estimators of scatter with eigenvalue shrinkage.
arXiv preprint:2002.04996v1 .

[10] Ollila, E., Tyler, D.E., 2014.
Regularized m-estimators of scatter matrix.
IEEE Transactions on Signal Processing 62, 6059–6070.

[11] Ovarlez, J.P., Pang, S.K., Pascal, F., Achard, V., Ng, T., 2011.
Robust detection using the sirv background modelling for hyperspectral imaging, in: Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, IEEE. pp. 4316–4319.

[12] Tyler, D.E., 1987.
A distribution-free M-estimator of multivariate scatter.
The Annals of Statistics 15, 234–251.

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

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

APPENDIX A: ASSUMPTIONS

We recollect here all the assumptions introduced in the core of the article.

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. $\|m\| \leq O(n)$

ASSUMPTION 6 bis. $\|m\| \leq O(1)$.

ASSUMPTION 7. $\frac{1}{n}\sum_{i=1}^n \tau_i \leq O(1)$.

ASSUMPTION 8. There exists $u_\cdot^\infty > 0$ such that for all $t \in \mathbb{R}^+$, $u_\cdot(t) \leq u_\cdot^\infty$.

ASSUMPTION 9. $\inf_{1\leq i\leq n} \frac{1}{n}\operatorname{Tr} C_i \geq O(1)$.

APPENDIX B: TOPOLOGICAL PROPERTIES OF THE STABLE SEMI-METRIC

LEMMA B.1. *Any Cauchy sequence of $(\mathcal{D}_n^+, d_s)$ is bounded from below and above (in $\mathcal{D}_n^+$).*

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:

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

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 a subsequence $(\Delta_{i_M^k}^{(\phi(k))})_{k\geq 0}$ 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 from above. For the lower bound, we consider the same way a subsequence $(\Delta_{i_m^k}^{(\psi(k))})_{k\geq 0}$ 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 OF PROPERTY 3.11. Given a Cauchy sequence of diagonal matrices $\Delta^{(k)} \in \mathcal{D}_n^+$, we know from the 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:

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

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 $k \in \mathbb{N}$, so that $\Delta^{(\infty)} \in \mathcal{D}_n^+$ and we are left to showing 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 OF THEOREM 3.12. There exist $\lambda \in (0,1)$ and two constants $\delta_M, \delta_m \in \mathbb{R}^+$ 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 3.11 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. □

APPENDIX C: CONCENTRATION AND ESTIMATION OF THE RESOLVENT

C.1. Some results on the resolvent. Given $S \in \mathcal{S}_p^n$, we introduce the notation $S_{-i} \equiv (S_1, \ldots, S_{i-1}, 0, S_i, \cdots S_n)$ and for $\Delta \in \mathcal{D}_n^+$ and $i \in [n]$:

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

We have the first simple identity:

$$Q(S, \Delta) - Q_{-i}(S, \Delta) = \frac{1}{n} Q(S, \Delta) S_i Q_{-i}(S, \Delta). \tag{12}$$

Now, for a matrix $M = (m_1, \ldots, m_n) \in \mathcal{M}_{p,n}$, if we note M_{-i} the matrix M with a zero in the i^{th} column, 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 $\check{I}^M(\Delta) \equiv \check{I}_z(M \cdot M^T, \Delta) = \text{Diag}(\frac{1}{n} m_i^T Q_{-i} m_i)_{1 \leq i \leq n}$ and we can deduce from (12) the so-called "Schur identity":

$$Q^M(\Delta) m_i = \frac{Q_{-i}^M(\Delta) m_i}{1 + \frac{\Delta_i}{n} m_i^T Q_{-i}^M(\Delta) m_i} \qquad \text{and} \qquad I^M(\Delta) = \frac{\check{I}^M(\Delta)}{I_n + \Delta \check{I}^M(\Delta)}. \tag{13}$$

That is reminiscent of Corollary 3.26. The two next subsections establish this link.

C.2. A first deterministic equivalent. We work in the larger setting presented in Remark 3.27 where, for any $i \in [n]$, the mean $\mu_i = \mathbb{E}[z_i]$ decomposes as $\mu_i = \mathring{\mu}_i + t_i s$ with $\|\mathring{\mu}_i\| \leq O(1)$, $\|s\| \leq O(\sqrt{n})$, $\sum_{i=1}^n t_i \geq O(1)$ and $\sup_{1\leq i\leq n} t_i \leq O(1)$. The next results, shown in [7], will be crucial for our estimation of $\mathbb{E}[Q(\Delta)]$ in Proposition C.4.

LEMMA C.1. *Given $\Delta \in \mathcal{D}_n^+$ such that $\|\Delta\| \leq O(1)$ and $u \in \mathbb{R}^p$ such that $\|u\| = 1$:*

$$\check{I}^Z(\Delta) = \operatorname{Diag}\left(\frac{1}{n} z_i^T Q_{-i}(\Delta) z_i\right)_{1\leq i\leq n} \propto \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) + \mathcal{E}_1\left(\frac{1}{n}\right) \qquad \textit{in } (\mathcal{D}_n^+, \|\cdot\|).$$

With this concentration in mind, we obtain a first deterministic equivalent for $Q(\Delta)$ depending on a deterministic diagonal matrix $\Theta \in \mathcal{D}_n^+$ having as diagonal elements:

$$\Theta(\Delta) \equiv \mathbb{E}\left[\check{I}^Z(\Delta)\right] = \operatorname{Diag}\left(\mathbb{E}\left[\frac{1}{n}\operatorname{Tr}\left(C_i Q_{-i}(\Delta)\right)\right]\right)_{1\leq i\leq n}. \tag{14}$$

In the next lemma and in the remainder of the subsection, we will often write $\Theta = \Theta(\Delta)$ for simplicity.

LEMMA C.2. *If $\|\Delta\| \leq O(1)$, then $\Theta \sim O(1)$.*

PROOF. Let us introduce $\kappa > 0$ such that $D \geq \kappa I_n$. On the one hand, $\sup_{1\leq i\leq n} \Theta_i \leq \sup_{1\leq i\leq n} \frac{\operatorname{Tr}(C_i)}{n\gamma} \leq O(1)$. On the other hand, since $Z\Delta Z^T \leq \kappa\|Z\|^2 I_p$, we know that $Q_{-i} \geq \frac{I_p}{\gamma + \frac{\kappa}{n}\|Z\|^2}$ and, therefore, the concentration $\frac{1}{n}\|Z\|^2 \propto \mathcal{E}_2(\frac{1}{\sqrt{n}}) + \mathcal{E}_1(\frac{1}{n})$ combined with Lemma 3.23 implies:

$$\inf_{1\leq i\leq n} \Theta_i \geq \mathbb{E}\left[\frac{\inf_{1\leq i\leq n} \frac{1}{n}\operatorname{Tr}(C_i)}{\gamma + \frac{\kappa}{n}\|Z\|^2}\right] \geq O(1).$$

$\square$

We are then going to show that

$$\tilde{Q}_1 \equiv Q_z\left(C, \frac{\Delta}{1 + \Delta\Theta}\right) = \left(\frac{1}{n}\sum_{i=1}^n \frac{\Delta_i C_i}{1 + \Delta_i \Theta_i} + \gamma I_p\right)^{-1}$$

is a deterministic equivalent of $Q(\Delta)$.

LEMMA C.3. $\sup_{1\leq i\leq n} \|C_i \tilde{Q}_1\| \leq O(1)$.

PROOF. Let us note $\mathring{C}_i = C_i - (\mu_i + t_i s)(\mu_i + t_i s)^T \geq 0$ (as a symmetric matrix), $\mu = \frac{1}{n}\sum_{i=1}^n \frac{\mu_i}{\sqrt{1/\Delta_i + \Theta_i}}$ and $\nu = \frac{1}{n}\sum_{i=1}^n \frac{t_i}{\sqrt{1/\Delta_i + \Theta_i}}$. We can decompose:

$$\frac{1}{n}\sum_{i=1}^n \frac{C_i}{1/\Delta_i + \Theta_i} = \frac{1}{n}\sum_{i=1}^n \frac{\mathring{C}_i}{1/\Delta_i + \Theta_i} + \frac{1}{n}\sum_{i=1}^n \frac{\mu_i \mu_i^T}{1/\Delta_i + \Theta_i} - \mu\mu^T + (\mu + \nu s)(\mu + \nu s)^T,$$

where we note that $\frac{1}{n}\sum_{i=1}^n \frac{\mu_i\mu_i^T}{1/\Delta_i + \Theta_i} - \mu\mu^T \geq 0$ (as a covariance matrix). Therefore

$$Q_{-s} \equiv \left(\frac{1}{n}\sum_{i=1}^n \frac{\mathring{C}_i}{1/\Delta_i + \Theta_i} + \frac{1}{n}\sum_{i=1}^n \frac{\mu_i\mu_i^T}{1/\Delta_i + \Theta_i} - \mu\mu^T + zI_p\right)^{-1} \geq 0$$

and we can employ the Schur formula to have

$$(\mu+\nu s)^T\tilde{Q}_1(\mu+\nu s) = \frac{(\mu+\nu s)^T\tilde{Q}_{-s}(\mu+\nu s)}{1+(\mu+\nu s)^T\tilde{Q}_{-s}(\mu+\nu s)} < 1.$$

Thus since we know from Lemma C.2 that $\Theta \sim 1$, $\frac{1}{\nu} \geq O(\frac{n}{\sum \tau} \geq O(1)$, $\|\mathring{\mu}\| = O(1)$ and $\|\tilde{Q}_1\| \leq O(1)$ (see Lemma 3.1):

$$t_i s^T\tilde{Q}_1 s \leq \frac{2t_i}{\nu^2}(\mu+\nu s)^T\tilde{Q}_1(\mu+\nu s) + \frac{2t_i}{\nu^2}\mu^T\tilde{Q}_1\mu \leq O(1),$$

and similarly $\mu_i^T\tilde{Q}_1 s, \mu_i^T\tilde{Q}_1\mu_i \leq O(1)$, from which we conclude since $\|\mathring{C}_i\| \leq O(1)$. □

PROPOSITION C.4. Given $\Delta \in \mathcal{D}_n^+$ such that $\|\Delta\| \leq O(1)$:

$$\left\|\mathbb{E}\left[Q(\Delta)\right] - \tilde{Q}_1\right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right).$$

PROOF. It suffices to bound for any $u, v \in \mathbb{R}^p$ such that $\|u\|, \|v\| \leq 1$:

$$\begin{aligned}
&\frac{1}{n}\left|\mathbb{E}\left[u^T\left(Q(\Delta) - \tilde{Q}_1\right)v\right]\right| \\
&\quad \leq \frac{1}{n}\left|\mathbb{E}\left[u^TQ(\Delta)\left(\sum_{i=1}\frac{\Delta_i C_i}{1+\Delta_i\Theta_i} - \Delta_i z_i z_i^T\right)\tilde{Q}_1 v\right]\right| \\
&\quad \leq \frac{1}{n}\left|\mathbb{E}\left[\sum_{i=1}^n \frac{\Delta_i u^TQ(\Delta)C_i\tilde{Q}_1 v}{1+\Delta_i\Theta_i} - \frac{\Delta_i u^TQ_{-i}(\Delta)z_i z_i^T\tilde{Q}_1 v}{1+\Delta_i\frac{1}{n}z_i^TQ_{-i}(\Delta)z_i}\right]\right| \\
&\quad \leq \varepsilon_1 + \varepsilon_2, \quad \text{with:}
\end{aligned}$$

- $$\begin{aligned}
\varepsilon_1 &= \frac{1}{n}\left|\sum_{i=1}^n \mathbb{E}\left[\frac{\Delta_i u^T(Q(\Delta) - Q_{-i}(\Delta))C_i\tilde{Q}_1 v}{1+\Delta_i\Theta_i}\right]\right| \\
&= \frac{1}{n^2}\left|\sum_{i=1}^n \mathbb{E}\left[\frac{\Delta_i u^TQ(\Delta)z_i z_i^TQ_{-i}(\Delta)C_i\tilde{Q}_1 v}{1+\Delta_i\Theta_i}\right]\right| \\
&\leq \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\Delta_i\check{I}^Z(\Delta)_i u^TQz_i z_i^TQu\right]\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[v^T\tilde{Q}_1C_iQ_{-i}C_i\tilde{Q}_1 v\right]} \\
&\leq O\left(\frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}\mathbb{E}\left[u^TQZ\Delta\check{I}^Z(\Delta)Z^TQu\right]}\right) \;\leq\; O\left(\frac{\sqrt{\mathbb{E}[\|\check{I}^Z(\Delta)\|]}}{\sqrt{n}}\right),
\end{aligned}$$

  thanks to Lemmas C.3 and 3.1 (and the bound on Δ). We can then conclude from Lemmas C.1 and 3.23 that $\mathbb{E}[\|\check{I}^Z(\Delta)\|] \leq \|\Theta\| + O(\sqrt{\log(n)/n})$ which entices $\varepsilon_1 \leq O\left(1/\sqrt{n}\right)$.

- $$\begin{aligned}\varepsilon_2 &= \frac{1}{n}\left|\sum_{i=1}^n \mathbb{E}\left[\frac{\Delta_i^2 u^T Q_{-i}(\Delta) z_i z_i^T \tilde Q_1 v\left(\frac{1}{n} z_i^T Q_{-i}(\Delta) z_i - \Theta_i\right)}{\left(1+\Delta_i \frac{1}{n} z_i^T Q_{-i}(\Delta) z_i\right)(1+\Delta_i\Theta_i)}\right]\right| \\ &= \frac{1}{n}\sqrt{\mathbb{E}\left[u^T Q(\Delta) Z \Delta^{1/2}(\check I^Z(\Delta)-\Theta)^2\Delta^{1/2} Z^T Q(\Delta) u\right]\sum_{i=1}^n \mathbb{E}\left[\frac{\Delta_i v \tilde Q_1 z_i z_i^T \tilde Q_1 v}{1+\Delta_i\Theta_i}\right]} \\ &\leq O\left(\sqrt{\mathbb{E}\left[\left\|\check I^Z(\Delta)-\Theta\right\|^2\right]}\right),\end{aligned}$$

  with the same justifications as previously. Again, we conclude with Lemmas C.1 and 3.23 that $\varepsilon_2 \leq O(\sqrt{\log n/n})$.

$\square$

C.3. A Second Deterministic equivalent for the resolvent. We now suppose that $\Delta \in \mathcal{D}_n^+$ is fixed and note for simplicity $\Lambda \equiv \Lambda(\Delta)$ defined in Proposition 3.24. A first result completely similar to Lemma C.2 allows us to bound Λ from below and above.

LEMMA C.5. *If $\|\Delta\| \leq O(1)$, then $\Lambda \sim O(1)$.*

Theorem 3.25 is just a consequence of the concentration of the resolvent and the following proposition.

PROPOSITION C.6. If $\|\Delta\| \leq O(1)$:

$$\|\Lambda - \Theta\| \leq O\left(\sqrt{\frac{\log n}{n}}\right) \qquad \text{and} \qquad \left\|\mathbb{E}\left[Q(\Delta)\right] - \tilde Q_2\right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right)$$

with the notation $\tilde Q_2 \equiv Q_z(C, \frac{\Delta}{1+\Delta\Lambda})$.

PROOF. With the notation $\tilde I_\Theta = \check I\left(C, \frac{\Delta}{1+\Delta\Theta}\right)$, and thanks to the identity $\Lambda = \check I\left(C, \frac{\Delta}{1+\Delta\Lambda}\right)$, we can employ Lemma 3.2 and the stability properties given in Property 3.9 to bound:

$$\begin{aligned}\left\|\frac{\Lambda-\Theta}{\sqrt{\Lambda\tilde I_\Theta}}\right\| &\leq \lambda\left\|\frac{\Lambda-\Theta}{\sqrt{\Lambda\Theta}}\right\| + \left\|\frac{\tilde I_\Theta-\Theta}{\sqrt{\Lambda\tilde I_\Theta}}\right\| \\ &\leq \lambda\left\|\frac{\Lambda-\Theta}{\sqrt{\Lambda\tilde I_\Theta}}\right\| + \left\|\frac{\Lambda-\Theta}{\sqrt{\Lambda\tilde I_\Theta}}\right\|\left\|\frac{\sqrt{\tilde I_\Theta}-\sqrt{\Theta}}{\sqrt{\Theta}}\right\| + \left\|\frac{\tilde I_\Theta-\Theta}{\sqrt{\Lambda\tilde I_\Theta}}\right\| \qquad (15)\end{aligned}$$

for $\lambda < 1$ satisfying $\frac{1}{1-\lambda} = O(1)$ (see Lemma 3.2). Now, we can deduce from Proposition C.4 that:

$$\left\|\tilde I_\Theta - \Theta\right\| = \sup_{1\leq i\leq n}\left\|\frac{1}{n}\operatorname{Tr}\left(C_i\left(\tilde Q_1 - \mathbb{E}[Q]\right)\right)\right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right),$$

thanks to Assumptions 4 and 5 ($\|C_i\| = \|\Sigma_i + \mu_i\mu_i^T\| \leq O(n)$). Then we deduce from Lemma C.2:

$$\left\|\frac{\sqrt{\tilde I_\Theta}-\sqrt{\Theta}}{\sqrt{\Theta}}\right\| \leq \sqrt{\left\|\tilde I_\Theta - \Theta\right\|}(c_\Theta+\gamma) \leq O\left(\left(\frac{\log n}{n}\right)^{1/4}\right).$$

Therefore, $1/(1-\lambda-\|\frac{\sqrt{\tilde I_\Theta}-\sqrt{\Theta}}{\sqrt{\Theta}}\|) \leq O(1)$ and we deduce from (15):

$$\left\|\frac{\Lambda-\Theta}{\sqrt{\Lambda \tilde I_\Theta}}\right\| \leq \frac{\left\|\frac{\tilde I_\Theta-\Theta}{\sqrt{\Lambda \tilde I_\Theta}}\right\|}{1-\lambda-\left\|\frac{\sqrt{\tilde I_\Theta}-\sqrt{\Theta}}{\sqrt{\Theta}}\right\|} = O\left(\sqrt{\frac{\log n}{n}}\right).$$

The upper bounds given by Lemmas C.2 and C.5 allow us to conclude that $\|\Lambda-\Theta\| \leq O(\sqrt{\log n/n})$.

We can already bound $\left\|\mathbb{E}\left[Q(\Delta)\right]-\tilde Q_1\right\|$ thanks to Proposition C.4, and we are left to bound:

$$\begin{aligned}
\left\|\tilde Q_1-\tilde Q_2\right\| &= \frac{1}{n}\left\|\tilde Q_1 \sum_{i=1}^n \frac{\Delta_i C_i(\Theta_i-\Lambda_i)}{(1+\Delta_i\Lambda_i)(1+\Delta_i\Theta_i)}\tilde Q_2\right\| \\
&= \left\|\frac{\Theta-\Lambda}{\sqrt{(1+\Delta_i\Lambda_i)(1+\Delta_i\Theta_i)}}\right\| \sqrt{\left\|\tilde Q_1 \frac{1}{n}\sum_{i=1}^n \frac{\Delta_i C_i}{1+\Delta_i\Theta_i}\tilde Q_1\right\| \left\|\tilde Q_2 \frac{1}{n}\sum_{i=1}^n \frac{\Delta_i C_i}{1+\Delta_i\Lambda_i}\tilde Q_2\right\|} \\
&\leq O\left(\sqrt{\frac{\log n}{n}}\right).
\end{aligned}$$

□

Proposition 3.5 combined with Proposition 4.9 imply that

$$\left\|\mathbb{E}\left[Q(\Delta)\right]-\tilde Q_2\right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right). \tag{16}$$

Theorem 3.25 is then deduced from the concentration of Q_z^Z given in (7) and Lemma 3.22 that sets that under (7) and (16), for any deterministic matrix $A \in \mathcal{M}_p$ such that $\|A\|_* \leq O(1)$, a concentration of $\mathrm{Tr}(AQ(\Delta))$ around $\mathrm{Tr}(A\mathbb{E}\left[Q(\Delta)\right])$ is equivalent to a concentration around $\mathrm{Tr}(A\tilde Q_2)$. Corollary 3.26 is a consequence of the concentration of $\frac{1}{n}z_iQ_{-i}(D)z_i$ given by Lemma C.1, relation (13) and from Theorem 3.25 and Lemma 3.23.