# OPTIMAL CLEANING FOR SINGULAR VALUES OF CROSS-COVARIANCE MATRICES

FLORENT BENAYCH-GEORGES, JEAN-PHILIPPE BOUCHAUD, AND MARC POTTERS



Abstract. We give a new algorithm for the estimation of the cross-covariance matrix $\mathbb{E}\,XY'$ of two large dimensional signals $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^p$ in the context where the number $T$ of observations of the pair $(X, Y)$ is itself large, but with $T \not\gg n, p$. This algorithm is optimal among *rotationally invariant estimators*, i.e. estimators derived from the *empirical estimator* by *cleaning* the singular values, while letting singular vectors unchanged. We give an interpretation of the singular value cleaning in terms of overfitting ratios.

## 1. Introduction

1.1. **Content of the article.** We give a new algorithm for the estimation of the cross-covariance matrix $\mathbb{E}\,XY'$ of two large dimensional signals $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^p$ in the context where the number $T$ of observations of the pair $(X, Y)$ is itself large, but with $T \not\gg n, p$. We prove that this algorithm, presented in Section 2.3, is optimal among *rotationally invariant estimators*, i.e. estimators derived from the *empirical estimator* given at (1) below by *cleaning* the singular values, while letting singular vectors unchanged. Algorithm efficiency is measured through simulations in Section 3. We also give an interpretation of the cleaning in terms of overfitting in Section 2.5.

1.2. **Context.** In high-dimensional statistics, it is well known that the standard *empirical estimator* (the one based over an average over the sample) has little efficiency when the sample size is not much larger than the dimension of the object we want to estimate. For example, the spectrum of the empirical covariance matrix of a sample of $T$ independent observations of a signal with covariance $I_n$ is not concentrated in the neighbourhood of 1 when $T \not\gg n$, but distributed according to a Marchenko-Pastur law. In the same way, for $(X(t), Y(t))_{t=1,\ldots,T}$ a sample of observations of a pair $(X, Y) \in \mathbb{R}^n \times \mathbb{R}^p$ of random vectors, the singular values of the empirical estimator

$$\mathbf{C}_{XY} := \frac{1}{T}\sum_{t=1}^{T} X(t)Y(t)' \qquad \text{with } Y(t)' := \text{transpose of the column } Y(t) \quad (1)$$

---

*Date*: January 18, 2019.

2010 *Mathematics Subject Classification.* 60B20;62G05;15B52.

*Key words and phrases.* Random matrices;Cross-covariance matrices;Rotationally Invariant Estimator.

of the true cross-covariance matrix are not distributed as the singular values of the true cross-covariance matrix when $T \not\gg n, p$ (see Figure 1, where we plot both the true singular values density and the histogram of the empirical singular values, which do not look alike at all).

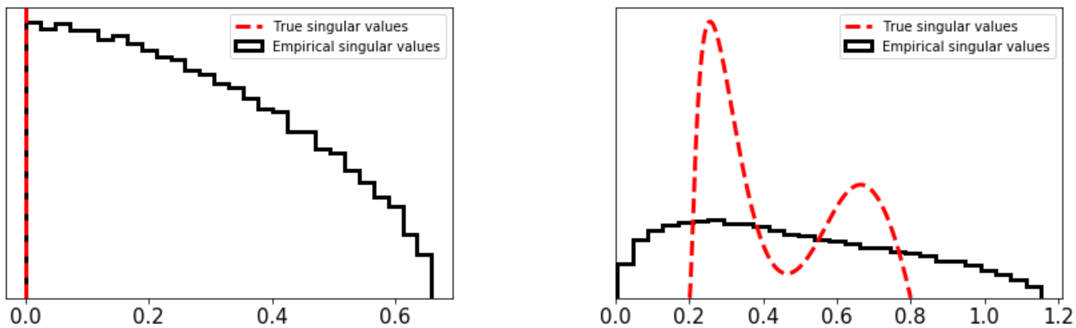


FIGURE 1. Singular values of $\frac{1}{T}\sum_t X(t)Y(t)'$ vs true singular values. **Left:** $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, I_{n+p})$. **Right:** $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$ for $\Sigma = \begin{pmatrix} I_n & \mathcal{C} \\ \mathcal{C}' & I_p \end{pmatrix}$ with $\mathcal{C}$ having singular values with density given by the red dashed curve. In both cases, $T/n = T/p = 10$ and $T = 25000$. The total lack of fit of the red curve by the histogram on the right and the spread between the true value 0 and most of the histogram on the left show that the empirical estimator works poorly (even though $T$ is 10 times higher than $n$ and $p$).

In the case of covariance estimation, several methods have been developed to circumvent these difficulties and improve the empirical estimator, based on regularization [12, 5, 14], shrinkage [21, 24, 23, 7, 8], specific sparsity or low-rank assumptions on the true covariance matrix [13, 17, 19, 18], robust statistics [10, 11] or fixed-point analysis [1]. Many applications exist, for example in finance [20, 22, 7, 8].

However, the problem of the estimation of cross-covariance matrices has, to our knowledge, not been addressed so far, despite its numerous applications in various fields (see e.g. [6], where the null model is studied).

Of course, cross-covariance estimation can formally be considered as a sub-problem of covariance estimation, as any pair of random vectors $(X, Y) \in \mathbb{R}^n \times \mathbb{R}^p$ can be concatenated in a vector $Z = \begin{pmatrix} X \\ Y \end{pmatrix} \in \mathbb{R}^{n+p}$ whose covariance matrix has upper-right corner the cross-covariance of $X$ and $Y$. The problem with this trick is that the above methods, when they are not specific to covariance matrices that are sparse or low-rank or essentially supported by a neighborhood of their diagonal (which makes them usually un-adapted to cross-covariance estimation), are *rotationally invariant estimators*, which means that they are justified in the Bayesian framework where the true covariance matrix of $Z$ has been chosen at random, with a prior distribution that is invariant under the action of the orthogonal group by conjugation.

Concretely, this means that the entries of $Z$ can naturally be blended in linear combinations. This clearly does not make sense when $X$ and $Y$ are of different nature, for example if $X$ contains stock returns and $Y$, say, weather data. However, an analogue notion exists for cross-covariance matrices, that we also call *rotationally invariant estimator*: it corresponds to estimators which clean the singular values of the empirical estimator but let its singular vectors unchanged, i.e. estimators relevant to the Bayesian framework where the true cross-covariance matrix has been chosen at random, with a prior distribution that is invariant under the actions of the orthogonal groups by multiplication on the left and on the right.

1.3. **Optimality.** The purpose of this text is precisely to compute the optimal rotationally invariant estimator for the true cross-covariance in the regime where we have at disposal a large number $T$ of observations of the pair $(X, Y)$, but where $T \not\gg n, p$. It is *optimal* in the sense that for Gaussian data, it is the solution of

$$\operatorname{argmin}_{\text{estimators}} \| \text{ Estimator} - \text{True cross-covariance}\|_{\mathrm{F}} \tag{2}$$

among the estimators whose singular vectors are those of the *empirical estimator* $\mathbf{C}_{XY}$ given at (1) above. Here, $\|\cdot\|_{\mathrm{F}}$ denotes the *Frobenius norm* , i.e. the standard Euclidean norm on matrices:

$$\|M\|_{\mathrm{F}} := \sqrt{\operatorname{Tr} MM'}. \tag{3}$$

Let us introduce the SVD of the empirical estimator $\mathbf{C}_{XY}$ from (1):

$$\mathbf{C}_{XY} = \sum_k s_k \mathbf{u}_k \mathbf{v}_k',$$

with $s_k$ the singular values and $\mathbf{u}_k$ (resp. $\mathbf{v}_k$) the left (resp. right) singular vectors. One easily gets (see (13) below) that optimality rewrites

$$\text{Estimator} = \sum_k s_k^{\text{cleaned}} \mathbf{u}_k \mathbf{v}_k' \quad \text{with} \quad s_k^{\text{cleaned}} = \mathbf{u}_k'(\text{True cross-covariance})\mathbf{v}_k. \tag{4}$$

The numbers

$$\mathbf{u}_k'(\text{True cross-covariance})\mathbf{v}_k,$$

called *oracle estimates*, are of course unknown, and the main problem is to estimate them. Here, we encode them via to an *oracle function* $L(z)$ (Proposition 2.1), which is then estimated in terms of observable variables only (Theorems 2.2 and 2.4).

1.4. **Organisation of the paper.** Model, main results and algorithms are presented in Section 2, where we also give an interpretation of the cleaning in terms of overfitting (Subsection 2.5). Then, algorithm accuracy is assessed and illustrated in Section 3, devoted to numerical simulations. Proofs are then given in Section 4.

1.5. **Notations.** Throughout this text, for $M$ a matrix, $M'$ denotes the transpose of $M$. For $Z$ a random variable, $\mathbb{E}\, Z$ denotes the expectation of $Z$.

Here, error terms in approximations depend on the parameters $n$, $p$, $T$ and $\Sigma$ of the problem, on the complex number $z$ and on the randomness. We will suppose that $\frac{n}{T}$, $\frac{p}{T}$, the operator norm of $\Sigma$ and $|z|$ are bounded by a constant $\mathfrak{M}$ and use the notation

$$O\left(\frac{1}{T\,|\mathfrak{Im} z|^k}\right)$$

to denote an error term which rewrites $\frac{1}{T\,|\mathfrak{Im} z|^k}$ times a bounded constant plus a centered Sub-Gaussian term with bounded Sub-Gaussian norm (the bound on the constant and on the Sub-Gaussian norm depending only on the constant $\mathfrak{M}$). Definition and basic properties of Sub-Gaussian variables can be found in [25, Sec. 2.5].

## 2. Main results and Algorithms

2.1. **Model.** Let $n \le p$ and let $(X, Y) \in \mathbb{R}^n \times \mathbb{R}^p$ be a pair of random vectors such that

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$$

for a given $\Sigma = \begin{pmatrix} \mathcal{A} & \mathcal{C} \\ \mathcal{C}' & \mathcal{B} \end{pmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}$ symmetric and non negative definite.

We are interested in the estimation of the true cross-covariance matrix

$$\mathcal{C} = \mathbb{E}\, XY'$$

out of its empirical version

$$\mathbf{C}_{XY} := \frac{1}{T}\mathbf{X}\mathbf{Y}' \in \mathbb{R}^{n\times p}, \tag{5}$$

where

$$\mathbf{X} := \begin{bmatrix} X(1) & \cdots & X(T) \end{bmatrix} \in \mathbb{R}^{n\times T} \qquad \text{and} \qquad \mathbf{Y} := \begin{bmatrix} Y(1) & \cdots & Y(T) \end{bmatrix} \in \mathbb{R}^{p\times T} \tag{6}$$

are defined thanks to a sequence

$$(X(1), Y(1)), \ldots, (X(T), Y(T)) \tag{7}$$

of independent copies of $(X, Y)$.

More precisely, we are looking for a *Rotationally Invariant Estimator* $\mathbf{C}_{XY,\mathrm{RIE}}$ of $\mathcal{C}$, i.e. an estimator constructed out of $\mathbf{X}$ and $\mathbf{Y}$ from (7) such that for any $V, W$ orthogonal matrices, if $\mathbf{X}$ and $\mathbf{Y}$ are respectively changed into $V\mathbf{X}$ and $W\mathbf{Y}$, then $\mathbf{C}_{XY,\mathrm{RIE}}$ is changed into $V\mathbf{C}_{XY,\mathrm{RIE}}W'$. If we also ask the estimator to be diagonal non negative definite when $\mathbf{C}_{XY}$ is so, then we need to define $\mathbf{C}_{XY,\mathrm{RIE}}$ as a matrix with the same singular vectors as $\mathbf{C}_{XY}$. Thus all we have to do is to *clean* the singular values of $\mathbf{C}_{XY}$.

Let us introduce the SVD of $\mathbf{C}_{XY}$. We set

$$\mathbf{C}_{XY} = \sum_{k=1}^{n} s_k \mathbf{u}_k \mathbf{v}_k' = \underbrace{\begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix}}_{:=\mathbf{U}} \operatorname{diag}(s_1, \ldots, s_n) \underbrace{\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}'}_{:=\mathbf{V}'} \tag{8}$$

for some $s_1, \ldots, s_n \geq 0$ and two orthonormal column vectors systems $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^n$, and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^p$.

Thus our estimator will have the form

$$\mathbf{C}_{XY,\mathrm{RIE}} = \mathbf{U} \operatorname{diag}(s_1^{\mathrm{cleaned}}, \ldots, s_n^{\mathrm{cleaned}}) \mathbf{V}'$$

and the *cleaned singular values*

$$s_1^{\mathrm{cleaned}}, \ldots, s_n^{\mathrm{cleaned}} \tag{9}$$

will be considered optimal when solving the optimization problem

$$\min_{s_1^{\mathrm{clean}}, \ldots, s_n^{\mathrm{clean}}} \|\mathbf{U} \operatorname{diag}(s_1^{\mathrm{clean}}, \ldots, s_n^{\mathrm{clean}}) \mathbf{V}' - \mathcal{C}\|_{\mathrm{F}}, \tag{10}$$

where $\|\cdot\|_{\mathrm{F}}$ has been defined at (3). Let us introduce the (implicitly depending on $z \in \mathbb{C}\backslash\mathbb{R}$) random variables

$$G := \frac{1}{T} \operatorname{Tr} \mathbf{G}, \qquad\qquad L := \frac{1}{T} \operatorname{Tr} \mathbf{G} \mathbf{C}_{XY} \mathcal{C}' \tag{11}$$

for $\mathbf{G}$ (resp. $\widetilde{\mathbf{G}}$, that we shall also use below) the resolvent of $\mathbf{C}_{XY} {\mathbf{C}_{XY}}'$ (resp. of ${\mathbf{C}_{XY}}' \mathbf{C}_{XY}$) defined through

$$\mathbf{G} := \left(z^2 - \mathbf{C}_{XY} {\mathbf{C}_{XY}}'\right)^{-1}, \qquad \widetilde{\mathbf{G}} := \left(z^2 - {\mathbf{C}_{XY}}' \mathbf{C}_{XY}\right)^{-1}. \tag{12}$$

The function $L(z)$ is defined through the quantity $\mathcal{C}$ we want to estimate and encodes the cleaning procedure by (14). For this reason, it is called the *oracle function.*

**Proposition 2.1.** *The solutions of* (10) *satisfy*

$$s_k^{\mathrm{cleaned}} = \mathbf{u}_k' \mathcal{C} \mathbf{v}_k \approx \frac{\mathfrak{Im} L(z)}{\mathfrak{Im}(z G(z))} \quad \text{for } z = s_k + \mathrm{i}\eta, \text{ with } \eta \ll 1, \tag{13}$$

*where the functions $L(z)$ and $G(z)$ are defined at* (11)*. More precisely, for any $\varepsilon > 0$ such that $[s_k - \varepsilon, s_k + \varepsilon] \cap \{s_1, \ldots, s_n\} = \{s_k\}$,*

$$s_k^{\mathrm{cleaned}} = \lim_{\eta \to 0} \frac{\int_{s_k - \varepsilon}^{s_k + \varepsilon} \mathfrak{Im} L(x + \mathrm{i}\eta) \mathrm{d}x}{\int_{s_k - \varepsilon}^{s_k + \varepsilon} \mathfrak{Im}((x + \mathrm{i}\eta) G(x + \mathrm{i}\eta)) \mathrm{d}x}. \tag{14}$$

2.2. **Estimations of the oracle function** $L(z)$**.** The problem with Formula (13) is that while the function $G(z)$ is explicit from the data $\mathbf{X}, \mathbf{Y}$, the definition of the function $L(z)$ involves the unknown true cross-covariance matrix $\mathcal{C}$. In Theorems 2.2 and 2.4, we give asymptotic approximations of $L(z)$ that can be estimated from the data alone, as is the case of the Ledoit-Péché estimator for covariance matrices [24, 23, 7, 8].

Let us introduce the random variables

$$H := \frac{1}{T}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}', \quad A := \frac{1}{T}\operatorname{Tr}\mathbf{G}\mathbf{C}_X, \quad B := \frac{1}{T}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y, \quad \Theta := z^2\frac{AB}{1+H} \tag{15}$$

for $\mathbf{G}$, $\widetilde{\mathbf{G}}$ as in (12) and $\mathbf{C}_X, \mathbf{C}_Y$ the empirical covariance matrices of $X$ and $Y$ defined by

$$\mathbf{C}_X := \frac{1}{T}\mathbf{X}\mathbf{X}', \qquad \mathbf{C}_Y := \frac{1}{T}\mathbf{Y}\mathbf{Y}'. \tag{16}$$

The following result makes the function $L$ of (11) explicit from the data alone, allowing a practical implementation of Formula (13) for the RIE.

**Theorem 2.2** (Oracle function estimation I)**.** *The function $L$ of* (11) *satisfies*

$$L = \frac{H-\Theta}{1+H-\Theta} + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right). \tag{17}$$

**Remark 2.3** (Case where $T \gg n, p$)**.** In the case where, as $T$ tends to infinity, $n$ and $p$ stay bounded, it can easily be seen that $L \approx H$, so that

$$\mathbf{C}_{XY,\mathrm{RIE}} \approx \mathbf{C}_{XY}.$$

Indeed, the estimate $L \approx H$ follows for example from the formulas (true for large $|z|$):

$$\begin{aligned} \frac{T}{n}L &= \sum_{k\geq 1}\frac{z^{-2k}}{n}\operatorname{Tr}\left((\mathbf{C}_{XY}\mathbf{C}_{XY}')^{k-1}\mathbf{C}_{XY}\mathcal{C}'\right) \\ \frac{T}{n}H(z) &= \sum_{k\geq 1}\frac{z^{-2k}}{n}\operatorname{Tr}\left((\mathbf{C}_{XY}\mathbf{C}_{XY}')^{k-1}\mathbf{C}_{XY}\mathbf{C}_{XY}'\right) \end{aligned}$$

and from standard complex analysis.

In the particular case where the covariance matrices of $X$ and $Y$ are both identity matrices, $\mathbf{C}_{XY,\mathrm{RIE}}$ is in fact an estimator of the *cross-correlation* matrix of $X$ and $Y$, and (13) can be simplified in a formula that uses less computation time when implemented. For $z \in \mathbb{C}\backslash\mathbb{R}$, let

$$K := \left(\frac{p-n}{T} + z^2 G\right)G(1+H)^2. \tag{18}$$

**Theorem 2.4** (Oracle function estimation II)**.** *Suppose that the true covariance matrices of $X$ and $Y$ are respectively $I_n$ and $I_p$. Then, the function $L$ of* (11) *satisfies*

$$L = \frac{1+2H-\sqrt{1+4K}}{2(1+H)} + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right) \tag{19}$$

*for $\sqrt{\cdot}$ the analytic version of the square root on $\mathbb{C}\backslash(-\infty, 0]$ with value* 1 *at* 1.

2.3. **Algorithmic consequences.** Formula (13) gives an expression for the cleaned singular values of the cross-covariance matrix, i.e. for the RIE of this matrix. The function $G(z)$ is explicit from the data $\mathbf{X}, \mathbf{Y}$, as well as the approximation of $L(z)$ given by formulas (17) and (19) above. Choosing $\eta = (npT)^{-1/6}$ for the "small $\eta$" (and using the formula $H = z^2 G - n/T$, for $H$ as in (15)), it leads to the explicit implementation formula

$$s_k^{\text{cleaned}} \;:=\; s_k \times \frac{\mathfrak{Im} L(z)}{\mathfrak{Im} H(z)} \qquad \text{for } z = s_k + \mathrm{i}(npT)^{-1/6}. \tag{20}$$

**Remark 2.5.** Following strictly (17) and (19), we should have slightly increased the imaginary part of $z$ in (20) and in the algorithms below, which, in practice, works as well. However, we believe that, following the method developed by Erdős, Yau and co-authors (see e.g. [15, 16, 4]), our local laws in theorems 2.2 and 2.4 can be improved roughly up to the scale $T^{-1}$, i.e. that the error terms, in (17) and (19), are in fact controlled essentially by $(T\mathfrak{Im} z)^{-1}$.

Using the approximation of $L(z)$ given by formulas (17), we get the first algorithm below, whose complexity is kept reasonable thanks to the following trick. With

$$\mathbf{C}_{XY} \;=\; \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix} \operatorname{diag}(s_1, \ldots, s_n) \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}' \tag{21}$$

the SVD of $\mathbf{C}_{XY}$, where the orthonormal system $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{R}^p$ is completed to an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_p$, we have

$$H(z) = \frac{1}{T} \sum_{\ell=1}^{n} \frac{s_\ell^2}{z^2 - s_\ell^2}, \tag{22}$$

$$A(z) = \frac{1}{T} \sum_{\ell=1}^{n} \frac{\mathrm{Coeff}_{A,\ell}}{z^2 - s_\ell^2}, \qquad B(z) = \frac{1}{T} \left( \sum_{\ell=1}^{n} \frac{\mathrm{Coeff}_{B,\ell}}{z^2 - s_\ell^2} + z^{-2}\, \mathrm{Coeff}_{B,[n+1:p]} \right) \tag{23}$$

for

$$\mathrm{Coeff}_{A,\ell} := \mathbf{u}_\ell' \mathbf{C}_X \mathbf{u}_\ell, \qquad \mathrm{Coeff}_{B,\ell} := \mathbf{v}_\ell' \mathbf{C}_Y \mathbf{v}_\ell, \qquad \mathrm{Coeff}_{B,[n+1:p]} := \sum_{\ell=n+1}^{p} \mathbf{v}_\ell' \mathbf{C}_Y \mathbf{v}_\ell \tag{24}$$

so that the functions $H$, $A$, $B$ and $\Theta$ from (15) can be computed without any matrix inversion (nor any matrix product) once the SVD of $\mathbf{C}_{XY}$ has been computed, which has only to be done once in the algorithm.

**Algorithm 1: Optimal cleaning for cross-covariance matrices :**

*Input:* $\mathbf{X} \in \mathbb{R}^{n\times T}$, $\mathbf{Y} \in \mathbb{R}^{p\times T}$ with $n \le p$.
*Output:* cleaned singular values $s_1^{\text{cleaned}}, \ldots, s_n^{\text{cleaned}}$.

(1) Compute $\mathbf{C}_{XY} = \frac{1}{T}\mathbf{X}\mathbf{Y}'$, $\mathbf{C}_X = \frac{1}{T}\mathbf{X}\mathbf{X}'$, $\mathbf{C}_Y = \frac{1}{T}\mathbf{Y}\mathbf{Y}'$
(2) Compute the SVD of $\mathbf{C}_{XY}$
(3) Compute the vectors $(\text{Coeff}_{A,\ell})_{\ell=1,\ldots,n}$ and $(\text{Coeff}_{B,\ell})_{\ell=1,\ldots,n}$ and the number $\text{Coeff}_{B,[n+1:p]}$ thanks to (24)
(4) For each $k \in \{1, \ldots, n\}$,
- set $z = s_k + \mathrm{i}(npT)^{-1/6}$ for $s_k$ the $k$-th singular value of $\mathbf{C}_{XY}$
- compute $H$, $A$, $B$ thanks to (22) and (23)
- compute
$$\Theta = z^2\frac{AB}{1+H} \qquad \text{and} \qquad L = 1 - \frac{1}{1+H-\Theta}$$
- compute
$$s_k^{\text{cleaned}} = s_k \times \frac{\mathfrak{Im} L}{\mathfrak{Im} H}$$

(5) possibly: apply the isotonic regression algorithm to the $s_k^{\text{cleaned}}$

One can also write an algorithm based on (19) instead of (17), which uses a bit less computation time (but only works when the true covariance matrices of $X$ and $Y$ are both identities):

**Algorithm 2: Optimal cleaning for cross-correlation matrices :**

*Input:* singular values $s_1, \ldots, s_n$ of $\mathbf{C}_{XY} = \frac{1}{T}\mathbf{X}\mathbf{Y}'$ for $\mathbf{X} \in \mathbb{R}^{n\times T}$, $\mathbf{Y} \in \mathbb{R}^{p\times T}$.
*Output:* cleaned singular values $s_1^{\text{cleaned}}, \ldots, s_n^{\text{cleaned}}$.

For each $k \in \{1, \ldots, n\}$,

(1) set $z = s_k + \mathrm{i}(npT)^{-1/6}$
(2) compute
$$G = \frac{1}{T}\sum_{\ell=1}^{n}\frac{1}{z^2 - s_\ell^2}, \qquad H = z^2 G - n/T, \qquad K = \left(\frac{p-n}{T} + z^2 G\right) G(1+H)^2$$
and
$$L = \frac{1 + 2H - \sqrt{1+4K}}{2(1+H)}$$
(3) compute
$$s_k^{\text{cleaned}} = s_k \times \frac{\mathfrak{Im} L}{\mathfrak{Im} H}$$
(4) possibly: apply the isotonic regression algorithm to the $s_k^{\text{cleaned}}$

**Remark 2.6.** It is a known fact that Ledoit-Péché's RIE is not working when $q :=$ dimension$/T$ is too close to 1. This problem is totally absent in our Algorithms 1 and 2.

2.4. **Cleaned vs empirical vs true singular values: some exact formulas.** Proposition 2.7 below is the precise statement behind the idea that, roughly,

$$\text{Cleaned singular values} \ < \ \text{True singular values} \ < \ \text{Empirical singular values.} \tag{25}$$

The first inequality, which, at first sight, could seem to contradict the optimality of our cleaning, is discussed and explained in Section 3.2 below.

Let us introduce/recall the following notation:

- $s_k$ : singular values of the empirical cross-covariance matrix $\mathbf{C}_{XY}$ from (5),
- $s_k^{\text{cleaned}}$ : cleaned singular values, solution of the optimization problem (10),
- $s_k^{\text{true}}$ : singular values of the true cross-covariance matrix $\mathcal{C}$,
- $\lambda_k^{\text{true},X}$ and $\lambda_l^{\text{true},Y}$ : eigenvalues of the true covariance matrices of $X$ and $Y$.

**Proposition 2.7.** *We have*

$$\mathbb{E}\sum_k s_k s_k^{\text{cleaned}} \ = \ \sum_k (s_k^{\text{true}})^2, \tag{26}$$

$$\mathbb{E}\sum_k s_k^2 \ = \ \left(1+\frac{1}{T}\right)\sum_k (s_k^{\text{true}})^2 + \frac{2}{T}\sum_k \lambda_k^{\text{true},X}\sum_l \lambda_l^{\text{true},Y} \tag{27}$$

*and*

$$\sum_k (s_k^{\text{true}})^2 \ = \ \frac{1}{1+T^{-1}-2T^{-2}}\,\mathbb{E}\left[\sum_k s_k^2 - \frac{1}{T}\operatorname{Tr}\mathbf{C}_X \operatorname{Tr}\mathbf{C}_Y\right] \tag{28}$$

*where $\mathbf{C}_X$, $\mathbf{C}_Y$ are the empirical covariance matrices of $X$ and $Y$ from (16)*

Note that the rough estimate of (25) follows from (26) and (27): it follows from (27) that on average,

$$\text{True singular values} \ < \ \text{Empirical singular values,}$$

whereas it follows from (26) that on average,

$$\text{Empirical singular values} \times \text{Cleaned singular values} \ \approx \ \text{True singular values.}$$

**Remark 2.8** (Unbiased estimator of $\sum_k (s_k^{\text{true}})^2$)**.** It follows from (28) that

$$\frac{1}{1+T^{-1}-2T^{-2}}\left(\sum_k s_k^2 - \frac{1}{T}\operatorname{Tr}\mathbf{C}_X \operatorname{Tr}\mathbf{C}_Y\right)$$

is an unbiased estimator of

$$\sum_k (s_k^{\text{true}})^2.$$

This estimator could be used to add a final step to our estimator, where we would rescale the $s_k^{\text{cleaned}}$ by a constant factor to achieve the equality

$$\sum_k (s_k^{\text{cleaned}})^2 = \frac{1}{1+T^{-1}-2T^{-2}} \left( \sum_k s_k^2 - \frac{1}{T} \operatorname{Tr} \mathbf{C}_X \operatorname{Tr} \mathbf{C}_Y \right). \tag{29}$$

One problem is then that the random variable on the RHT of (29) can take negative values, so that it has to be replaced by its positive part.

2.5. **Interpretation of the cleaning in terms of overfitting.** *Overfitting* is a very common issue in data analysis and machine learning. It refers to the problem that any model is fitted, *in sample*, on noisy data and apt to *learn noise*, which of course degrades its performance when assessed *out of sample*. In this section, we relate the cleaning procedure to the ratio

$$\text{Overfitting-factor} := \frac{\text{Out-of-sample-performance}}{\text{In-sample-performance}}$$

for a certain very elementary statistical learning algorithm (namely *Ridge regression with large* $\lambda$), proving (see (32)) that

$$\text{Overfitting-factor} \approx \frac{s_k^{\text{cleaned}}}{s_k}.$$

Suppose to be given a sample

$$(R(1), F(1)), \ldots, (R(T), F(T))$$

of observations of a pair $(R, F) \in \mathbb{R}^n \times \mathbb{R}^p$ of vectors, where $F$ is a collection of factors thanks to which we want to explain $R$.

Given this set of observations, if we observe an "out-of-sample" (oos) realization $F_{\text{oos}}$ of the factors and want to predict the corresponding $R_{\text{oos}}$, the Ridge predictor[1] with large $\lambda$ is given by

$$R^{\text{Ridge}} = \mathbf{C}_{RF} F_{\text{oos}} = \sum_k s_k (F'_{\text{oos}} \mathbf{v}_k) \mathbf{u}_k,$$

where $\sum_k s_k \mathbf{u}_k \mathbf{v}'_k$ is the SVD of the in-sample cross-covariance matrix $\mathbf{C}_{RF} := \frac{1}{T} \sum_t R(t) F(t)'$.

Each term of the previous sum defines a partial predictor

$$R^{\text{Ridge},k}(F_{\text{oos}}) := s_k (F'_{\text{oos}} \mathbf{v}_k) \mathbf{u}_k.$$

Let us now focus on the overlap of these predictors with the true values of $R$.

**Out of sample overlap:** it is given by

$$\begin{aligned} R_{\text{oos}} \cdot R^{\text{Ridge},k}(F_{\text{oos}}) &= s_k (F'_{\text{oos}} \mathbf{v}_k)(\mathbf{u}'_k R_{\text{oos}}) \\ &= s_k \mathbf{u}'_k R_{\text{oos}} F'_{\text{oos}} \mathbf{v}_k. \end{aligned}$$

[1] We could also consider the OLS predictor, but notations are lighter that way.

Over the out-of-sample time series

$$(R_{\text{oos}}(1), F_{\text{oos}}(1)), \dots, (R_{\text{oos}}(T_{\text{oos}}), F_{\text{oos}}(T_{\text{oos}})),$$

it averages out to

$$\begin{aligned}\text{Out-of-sample-overlap} &:= \frac{1}{T_{\text{oos}}} \sum_t s_k \mathbf{u}_k' R_{\text{oos}}(t) F_{\text{oos}}(t)' \mathbf{v}_k \\ &= s_k \mathbf{u}_k' \mathbf{C}_{RF}^{\text{out of sample}} \mathbf{v}_k\end{aligned}$$

for

$$\mathbf{C}_{RF}^{\text{out of sample}} := \frac{1}{T_{\text{oos}}} \sum_t R_{\text{oos}}(t) F_{\text{oos}}(t)'.$$

By (4) and the concentration of measure lemma 4.6, we get

$$\text{Out-of-sample-overlap} = s_k s_k^{\text{cleaned}} + O\left(\frac{1}{\sqrt{T_{\text{oos}}}}\right). \tag{30}$$

**In sample overlap:** it is given, at each date $t$ of the sample, by

$$R(t) \cdot R^{\text{Ridge},k}(F(t)) = s_k (F(t)' \mathbf{v}_k)(\mathbf{u}_k' R(t)) = s_k \mathbf{u}_k' R(t) F(t)' \mathbf{v}_k,$$

which, by (4), averages out, in sample, to

$$\text{In-sample-overlap} := \frac{1}{T} \sum_{t=1}^{T} s_k \mathbf{u}_k' R(t) F(t)' \mathbf{v}_k = s_k \mathbf{u}_k' \mathbf{C}_{RF} \mathbf{v}_k = s_k^2. \tag{31}$$

**Out of sample / in sample:** From (30) and (31), we deduce the following nice interpretation of the cleaning procedure, illustrated at Figure 2:

$$\text{Overfitting-factor} := \frac{\text{Out-of-sample-overlap}}{\text{In-sample-overlap}} \approx \frac{s_k^{\text{cleaned}}}{s_k}. \tag{32}$$

**Remark 2.9.** If, instead of considering the partial predictor $R^{\text{Ridge},k}(F_{\text{oos}})$, we consider sums, over $k$, of such predictors, then the previous ratio can still be expressed thanks to the numbers $s_k^{\text{cleaned}}$ and $s_k$.

## 3. Numerical simulations

3.1. **Oracle estimation.** The cornerstones of this work are (17) and (19) from Theorems 2.2 and 2.4: these formulas allow to approximate the (unknown) oracle function $L(z)$ by some functions that are explicit from the data. We made numerical simulations to verify these formulas for various models (i.e. various choices of $\Sigma$), all confirming their accuracy. In Figure 3, we present the *relative differences*

$$\frac{|L(z) - (\text{approximation of } L(z) \text{ given at } (17))|}{|L(z)|} \tag{33}$$

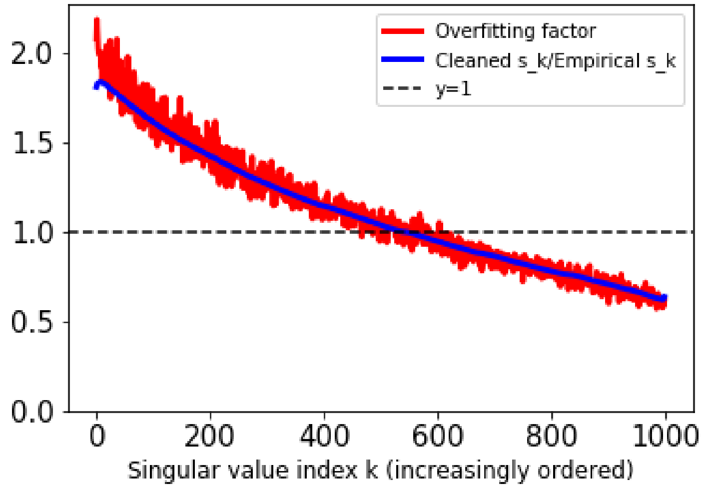


FIGURE 2. **Overfitting and cleaning**. Illustration of (32): LHT vs RHT. Here, $R = AF +$noise, with $F$ standard Gaussian vector, $A \in \mathbb{R}^{n\times p}$ fixed, $n = p = 1000$, $T = 10000$, $T_{\text{oos}} = 1000$.

and

$$\frac{|L(z) - (\text{approximation of } L(z) \text{ given at } (19))|}{|L(z)|} \tag{34}$$

for

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \Sigma) \quad \text{with} \quad \Sigma = \begin{pmatrix} I_n & \mathcal{C} \\ \mathcal{C}' & I_p \end{pmatrix} \tag{35}$$

with $\mathcal{C}$ a matrix with singular values distributed according to the bi-modal density from the right graph in Figure 1 and Haar-distributed singular vectors. We see that both approximations of $L(z)$ are very efficient and that the approximation of $L(z)$ given at (17) is a little bit better, which is confirmed by other simulations. An advantage of (19), however, is that it uses a little bit less computation time.

3.2. **Effect of the cleaning.** In Figure 4, we show the effect of the cleaning in the simulations from Figure 1.

Note that the fact that our estimator realizes the optimal of (2) does not imply that the cleaned singular values should be distributed as the true ones. Indeed, given the singular vectors of our estimator are not exactly those of the true cross-covariance matrix (but those of the empirical estimator from (1) above), the optimal of (2) has to be “cautious” in putting weight on these singular vectors (and thus to globally shrink the singular values). Precisely, we show in Section 2.4 that roughly,

$$\text{Cleaned singular values} \ < \ \text{True singular values} \ < \ \text{Empirical singular values}. \tag{36}$$

An analogous phenomenon for rotationally invariant estimators of covariance matrices is explained at Section 6.3 from [8].

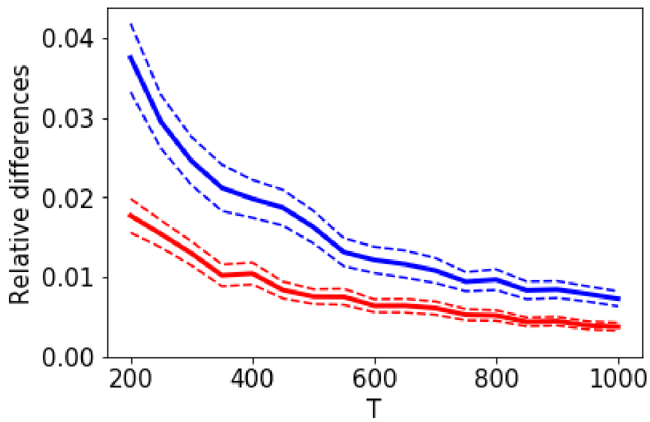


FIGURE 3. **Validity of our oracle estimations:** mean (out of 100 simulations, with 95% confidence interval given by dashed lines) relative differences from (33) (in red) and (34) (in blue) for the model of (35), for various values of $T$ (in abscissa) with fixed $n/T = 0.4$ and $p/T = 0.7$ and for $z = 0.5 + \mathrm{i}$.

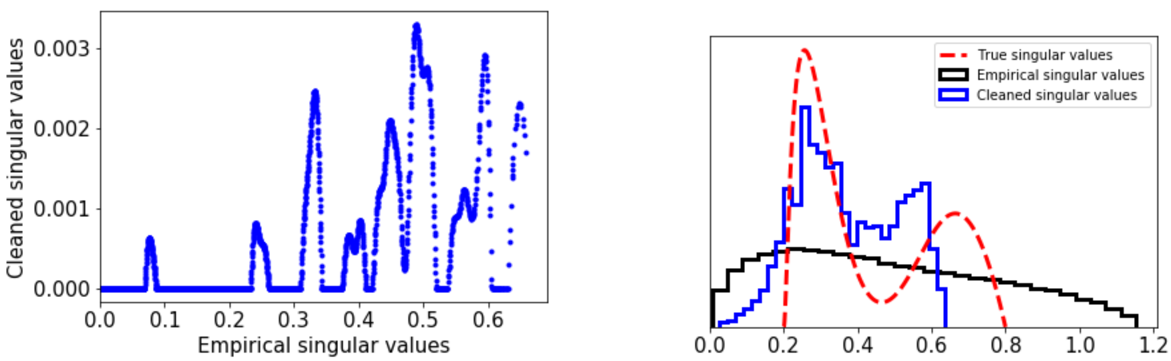


FIGURE 4. Cleaned vs empirical singular values for the simulation from Figure 1. **Left:** as one could expect from a good estimator, in the null model, most of the singular values are turned to approximately 0. **Right:** same as in Figure 1, with the cleaned singular values histogram added. The lack of monotonicity in the left graph is the reason why we added the isotonic regression as optional last step in our algorithm.

3.3. **Compared performance with empirical and Ledoit-Péché's estimators.** We have implemented Algorithms 1 and 2 from the present paper[2] for various models, i.e. various choices of the true total covariance matrix $\Sigma$ such that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \Sigma).$$

[2]Both give approximately the same result when $X$ and $Y$ have identity covariance matrices, so we shall focus on Algorithm 1 in this section.

We then compared their performance to that of the empirical estimator $\mathbf{C}_{XY}$, thanks to the distances percentages

$$100 \times \frac{\|(\text{Algorithm 1}) - (\text{True cross-covariance})\|_{\mathrm{F}}}{\|(\text{Empirical cross-covariance } \mathbf{C}_{XY}) - (\text{True cross-covariance})\|_{\mathrm{F}}}. \tag{37}$$

We also compared with the projection (on the $n \times p$ upper-right half corner) of Ledoit-Péché's RIE (the estimator from [24], which assumes $O(n+p)$-invariance), thanks to the distance percentages

$$100 \times \frac{\|(\text{Algorithm 1}) - (\text{True cross-covariance})\|_{\mathrm{F}}}{\|(\text{Projection of Ledoit-Péché's estimator}) - (\text{True cross-covariance})\|_{\mathrm{F}}} \tag{38}$$

The values of the quotients from (37) and (38) are reported in Table 1 (and discussed in Section 3.4 below) for the following models:

- Models (1) to (5):

$$X \sim \mathcal{N}(0, I_n) \ , \qquad Y = \mathcal{C}'X + \sigma \cdot (\text{Gaussian white noise in } \mathbb{R}^p \text{ independent of } X)$$

  where $\sigma^2 = 0.5$ and $\mathcal{C}$ has:
  - independent Haar-distributed left and right singular vectors,
  - 0%, 10%, 20%, 30% or 40% (for respectively Model (1),..., Model (5)) of non zero singular values, distributed uniformly in $[0.2, 0.5]$,

  so that $\Sigma = \begin{pmatrix} I_n & \mathcal{C} \\ \mathcal{C}' & \mathcal{C}'\mathcal{C} + \sigma^2 I_p \end{pmatrix}$,
- Model (6): $\Sigma = HH'/(2m)$ for $H$ an $m \times 2m$ matrix with i.i.d. standard Gaussian entries.

| Model | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Algo/Empirical | 1.1 ± 0.07 | 19 ± 0.01 | 26 ± 0.02 | 31 ± 0.03 | 35 ± 0.03 | 55.7 ± 0.1 |
| Algo/(Ledoit-P.) | 3.6 ± 0.2 | 54 ± 0.06 | 67.8 ± 0.07 | 75.6 ± 0.07 | 80.7 ± 0.05 | 96.3 ± 0.1 |

TABLE 1. Confidence intervals for (37) (first row) and (38) (second row) out of 100 simulations for $T = 500$, $n/T = 0.4$ and $p/T = 0.7$ (recall that (37) and (38) are percentages).

3.4. **Conclusion and perspectives.** Our estimator is optimal, for the Euclidian distance, among *rotationally invariant estimators*, i.e. estimators derived from the empirical estimator by *cleaning* the singular values, while letting *singular vectors unchanged*. We see from Table 1 that:

- Our estimator performs way better than the empirical and Ledoit-Péché estimators for Models (1) to (5), which are all Bayesian models with prior distributions of the true total covariance matrix $\Sigma$ invariant under the action of $O(n) \times O(p)$ defined by

$$(U, V) \cdot \Sigma = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \Sigma \begin{pmatrix} U' & 0 \\ 0 & V' \end{pmatrix}. \tag{39}$$

- The advantage of our estimator over Ledoit-Péché's RIE is way higher when the prior on $\Sigma$ is $O(n) \times O(p)$-invariant rather than $O(n+p)$-invariant (case of Model (6)).

That being said, the $O(n) \times O(p)$-invariance from (39) implies that the singular vectors of $\mathcal{C}$ are Haar-distributed, but does not imply that the singular vectors of $\mathcal{C}$ are independent from the other observables (e.g. the eigenvectors of $\mathcal{A}$ and $\mathcal{B}$), hence does not define Bayesian models where the right way to estimate $\mathcal{C}$ is necessarily rotationally invariant[3]. This means that for Models (1) to (5), our estimator could be sub-optimal, and a cleaning of the singular vectors, based e.g. on the observation of the eigenvectors of $\mathcal{A}$ and $\mathcal{B}$, should possibly also be performed. Investigations about this singular vectors cleaning would be an interesting perspective, with certainly many applications.

## 4. Proofs

4.1. **Proof of Proposition 2.1.** Let $\tilde{\mathbf{V}}$ be a $p \times p$ orthogonal matrix with the same $n$ first columns as $\mathbf{V}$. Then

$$\mathbf{U} \operatorname{diag}(s_1^{\text{clean}}, \ldots, s_n^{\text{clean}}) \mathbf{V}' = \mathbf{U} \begin{bmatrix} \operatorname{diag}(s_1^{\text{clean}}, \ldots, s_n^{\text{clean}}) & 0_{n,p-n} \end{bmatrix} \tilde{\mathbf{V}}'$$

and, given the Frobenius norm is invariant by left and right multiplication by orthogonal matrices, the optimization problem (10) rewrites

$$\min_{s_1^{\text{clean}}, \ldots, s_n^{\text{clean}} \geq 0} \| \begin{bmatrix} \operatorname{diag}(s_1^{\text{clean}}, \ldots, s_n^{\text{clean}}) & 0_{n,p-n} \end{bmatrix} - \mathbf{U}' \mathcal{C} \tilde{\mathbf{V}} \|_{\mathrm{F}},$$

i.e.

$$\min_{s_1^{\text{clean}}, \ldots, s_n^{\text{clean}} \geq 0} \| \operatorname{diag}(s_1^{\text{clean}}, \ldots, s_n^{\text{clean}}) - \mathbf{U}' \mathcal{C} \mathbf{V} \|_{\mathrm{F}},$$

whose solution is given by

$$s_k^{\text{clean}} = (\mathbf{U}' \mathcal{C} \mathbf{V})_{kk} = \mathbf{u}_k' \mathcal{C} \mathbf{v}_k, \qquad k = 1, \ldots, n. \tag{40}$$

i.e. $s_k^{\text{clean}}$ can be expressed as the Radon-Nikodym derivative

$$s_k^{\text{clean}} = \frac{\mathrm{d}m_{\mathbf{C}_{XY}, \mathcal{C}}}{\mathrm{d}\nu_{\mathbf{C}_{XY}}}(s_k), \tag{41}$$

[3]Bayesian models where the right way to estimate $\mathcal{C}$ is necessarily rotationally invariant are those with prior distribution on $\Sigma$ invariant under the action of $O(n)^2 \times O(p)^2$ defined by $(U, W, V, K) \cdot \begin{pmatrix} \mathcal{A} & \mathcal{C} \\ \mathcal{C}' & \mathcal{B} \end{pmatrix} = \begin{pmatrix} U\mathcal{A}U' & W\mathcal{C}K' \\ K\mathcal{C}'W' & V\mathcal{B}V' \end{pmatrix}$.

for $m_{\mathbf{C}_{XY},\mathcal{C}}$ the null mass signed measure

$$m_{\mathbf{C}_{XY},\mathcal{C}} := \frac{1}{2n}\sum_{k=1}^{n} \mathbf{u}_k'\mathcal{C}\mathbf{v}_k \left(\delta_{s_k} - \delta_{-s_k}\right) \tag{42}$$

and $\nu_{\mathbf{C}_{XY}}$ the symmetrized empirical singular values distribution of $\mathbf{C}_{XY}$, defined by

$$\nu_{\mathbf{C}_{XY}} := \frac{1}{2n}\sum_{k=1}^{n} \left(\delta_{s_k} + \delta_{-s_k}\right). \tag{43}$$

Equation (41) allows, by (75), to express $s_k^{\text{clean}}$ thanks to the formula, true for any $\varepsilon > 0$ such that $[s_k-\varepsilon, s_k+\varepsilon]\cap\{s_1,\ldots,s_n\} = \{s_k\}$,

$$s_k^{\text{clean}} = \lim_{\eta\to 0} \frac{\int_{s_k-\varepsilon}^{s_k+\varepsilon} \mathfrak{Im}(\text{Stieltjes transform of } m_{\mathbf{C}_{XY},\mathcal{C}} \text{ at } x+\mathrm{i}\eta)\mathrm{d}x}{\int_{s_k-\varepsilon}^{s_k+\varepsilon} \mathfrak{Im}(\text{Stieltjes transform of } \nu_{\mathbf{C}_{XY}} \text{ at } x+\mathrm{i}\eta)\mathrm{d}x}. \tag{44}$$

Note that for any $z\in\mathbb{C}\backslash\mathbb{R}$ and any $s\geq 0$,

$$\frac{1}{2}\left(\frac{1}{z-s} - \frac{1}{z+s}\right) = \frac{s}{z^2-s^2},$$

so that the Stieltjes transform of $m_{\mathbf{C}_{XY},\mathcal{C}}$ is given by

$$\begin{aligned}
(\text{Stieltjes transform of } m_{\mathbf{C}_{XY},\mathcal{C}} \text{ at } z) &= \int \frac{\mathrm{d}m_{\mathbf{C}_{XY},\mathcal{C}}(s)}{z-s} \\
&= \frac{1}{2n}\sum_k \left(\frac{1}{z-s_k} - \frac{1}{z+s_k}\right)\mathbf{u}_k'\mathcal{C}\mathbf{v}_k \\
&= \frac{1}{n}\sum_k \frac{s_k}{z^2-s_k^2}\mathbf{u}_k'\mathcal{C}\mathbf{v}_k \\
&= \frac{1}{n}\sum_k \frac{s_k}{z^2-s_k^2}\operatorname{Tr}\mathbf{u}_k'\mathcal{C}\mathbf{v}_k \\
&= \frac{1}{n}\sum_k \frac{s_k}{z^2-s_k^2}\operatorname{Tr}\mathbf{v}_k\mathbf{u}_k'\mathcal{C} \\
&= \frac{1}{n}\sum_k \frac{s_k}{z^2-s_k^2}\operatorname{Tr}\mathcal{C}'\mathbf{u}_k\mathbf{v}_k' \\
&= \frac{1}{n}\operatorname{Tr}\left(\mathcal{C}'\sum_k \frac{s_k}{z^2-s_k^2}\mathbf{u}_k\mathbf{v}_k'\right) \\
&= \frac{1}{n}\operatorname{Tr}\left(\mathcal{C}'\left(z^2-\mathbf{C}_{XY}{\mathbf{C}_{XY}}'\right)^{-1}\mathbf{C}_{XY}\right) \\
&= \frac{1}{n}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathcal{C}'
\end{aligned} \tag{45}$$

Also, as for any $z \in \mathbb{C}\backslash\mathbb{R}$ and any $s \geq 0$,

$$\frac{1}{2}\left(\frac{1}{z-s}+\frac{1}{z+s}\right)=\frac{z}{z^2-s^2},$$

we have, for $G$ as in (11),

$$\text{(Stieltjes transform of } m_{\mathbf{C}_{XY},\mathcal{C}} \text{ at } z) = \frac{T}{n} zG \tag{46}$$

Then, (14) follows from (44), (45) and (46).

4.2. **Proof of Theorem 2.2.** Set

$$g := \mathbb{E}\, G,\ h := \mathbb{E}\, H,\ \ell := \mathbb{E}\, L,\ a := \mathbb{E}\, A,\ ,\widetilde{a} := \mathbb{E}\, \widetilde{A},\ b := \mathbb{E}\, B,\ \widetilde{b} := \mathbb{E}\, \widetilde{B}.$$

The following concentration of measure lemma can be proved using the Log-Sobolev inequality satisfied by the standard Gaussian law (a detailed proof is given in Section 4.6.1).

**Lemma 4.1.** *There is a constant $c > 0$, depending only on the bound $\mathfrak{M}$ of the hypothesis, such that for any $z \in \mathbb{C}\backslash\mathbb{R}$, we have, for any $t > 0$,*

$$\mathbb{P}\left(|G-g| \geq t\right) \quad \leq \quad 2\mathrm{e}^{-c(tT(\mathfrak{Im} z)^2)^2}.$$

*In other words, $G - g$ is a Sub-Gaussian random variable, with Sub-Gaussian norm*

$$O\left(\frac{1}{T(\mathfrak{Im} z)^2}\right).$$

*Besides, the same is true for any of the random variables $H - h$, $A - a$, $B - b$, $L - \ell$, $\Theta - \mathbb{E}\,\Theta$, $K - \mathbb{E}\, K$.*

By this lemma, using the decomposition

$$L = \ell + (L - \ell),$$

it suffices to prove that

$$\ell = \frac{h-\theta}{1+h-\theta} + O\left(\frac{1}{T\,(\mathfrak{Im} z)^2}\right). \tag{47}$$

Then, the key of the proof is the following proposition, whose proof, based on the multidimensional Stein formula for Gaussian vectors, is postponed to Section 4.4. Let us introduce the implicitly depending on $z \in \mathbb{C}\backslash\mathbb{R}$ random variables

$$\widetilde{A} := \frac{1}{T}\operatorname{Tr}\mathbf{G}\mathcal{A},\ \ \widetilde{a} := \mathbb{E}\,\widetilde{A}, \qquad \widetilde{B} := \frac{1}{T}\operatorname{Tr}\widetilde{\mathbf{G}}\mathcal{B},\ \ \widetilde{b} := \mathbb{E}\,\widetilde{B}. \tag{48}$$

**Proposition 4.2.** *We have*

$$h \;=\; \ell + \frac{z^2}{2}\left(a\widetilde{b} + b\widetilde{a}\right) + h\ell + O\left(\frac{1}{T\,(\mathfrak{Im} z)^2}\right) \tag{49}$$

$$a(1-\ell) \;=\; \widetilde{a}(1+h) + O\left(\frac{1}{T\,(\mathfrak{Im} z)^2}\right) \tag{50}$$

$$b(1-\ell) \;=\; \widetilde{b}(1+h) + O\left(\frac{1}{T\,(\mathfrak{Im} z)^2}\right) \tag{51}$$

Let us now conclude the proof of Theorem 2.2. Thus after multiplication of (49) by $1+h$, we have

$$h(1+h) \;=\; \ell(1+h) + z^2 ab(1-\ell) + h\ell(1+h) + O\left(\frac{1}{T\,(\mathfrak{Im} z)^2}\right)$$

Using the fact, following from lemma 4.1, Cauchy-Schwarz inequality and first part of Proposition 5.3, that

$$\mathbb{E}\,\Theta - \frac{z^2 ab}{1+h} = O\left(\frac{1}{T|\mathfrak{Im} z|^5}\right).$$

We get, for $\theta := \frac{z^2 ab}{1+h}$,

$$h \;=\; \ell + \theta(1-\ell) + h\ell + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right)$$

i.e.

$$\ell = \frac{h-\theta}{1+h-\theta} + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right). \tag{52}$$

Then, conclude that (47) is true using Lemma 4.1.

4.3. **Proof of Theorem 2.4.** In the case where $\mathcal{A} = I_n$ and $\mathcal{B} = I_p$, the random variables $\widetilde{A}$ and $\widetilde{B}$ from (48) are respectively equal to $G$ and $(p-n)/(Tz^2) + G$, and rather than using (52) to estimate $\ell$, we shall solve (49) without using $a$ and $b$. Using (50) and (51), after multiplication by $1-\ell$, (49) rewrites

$$h(1-\ell) \;=\; \ell(1-\ell) + z^2 g((p-n)/(Tz^2) + g)(1+h) + h\ell(1-\ell) + O\left(\frac{1}{T\,(\mathfrak{Im} z)^2}\right)$$

for $g := \mathbb{E}\,G$. For $\kappa := -g((p-n)/T + z^2 g)(1+h)$, we get

$$(1+h)\,\ell^2 - (1+2h)\,\ell + h + \kappa + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right) = 0 \tag{53}$$

where we have used the fact, following from lemma 4.1, Cauchy-Schwarz inequality and first part of Proposition 5.3, that

$$\mathbb{E}\,K - \left(\frac{p-n}{T} + z^2 g\right) g(1+h)^2 = O\left(\frac{1}{T(\mathfrak{Im} z)^5}\right).$$

Second order polynomial equation (53) solves as

$$\ell \;=\; \frac{1+2h \pm \sqrt{1-4\kappa(1+h)}}{2(1+h)} + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right)$$

Considering the case where $\alpha$ and $\beta$ are small (where we should have $\ell \approx h$, as explained in Remark 2.3) and using analytic continuation, we have

$$\ell = \frac{1+2h-\sqrt{1-4\kappa(1+h)}}{2(1+h)} + O\left(\frac{1}{T\,|\mathfrak{Im} z|^5}\right) \tag{54}$$

for $\sqrt{\cdot}$ the analytic version of the square root on $\mathbb{C}\backslash(-\infty, 0]$ with value 1 at 1. Then, again, conclude using Lemma 4.1.

### 4.4. **Proof of Proposition 4.2.**

4.4.1. *Proof of* (49)*: expansion of* $\mathbb{E}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}'$. Using $\mathbf{C}_{XY} = \frac{1}{T}\sum_t X(t)Y(t)'$, we have

$$\begin{aligned}
\operatorname{Tr}\mathbf{C}_{XY}\mathbf{C}_{XY}'\mathbf{G}(z) &= \frac{1}{T}\sum_{t=1}^{T}\operatorname{Tr} X(t)Y(t)'\mathbf{C}_{XY}'\mathbf{G} \\
&= \frac{1}{T}\sum_{t=1}^{T} Y(t)'\mathbf{C}_{XY}'\mathbf{G}X(t) \\
&= \frac{1}{2T}\sum_{t=1}^{T} Y(t)'\mathbf{C}_{XY}'\mathbf{G}X(t) + X(t)'\mathbf{G}\mathbf{C}_{XY}Y(t) \\
&= \frac{1}{2T}\sum_{t=1}^{T} Z(t)'\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix} Z(t) \qquad (55)
\end{aligned}$$

for

$$Z(t) := \begin{pmatrix} X(t) \\ Y(t)\end{pmatrix}. \tag{56}$$

By (78) from Corollary 5.2, it follows that

$$\mathbb{E}\operatorname{Tr}\mathbf{C}_{XY}\mathbf{C}_{XY}'\mathbf{G}(z) = \tag{57}$$
$$\frac{1}{2}\mathbb{E}\operatorname{Tr}\Sigma\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix} + \frac{1}{2T}\mathbb{E}\sum_{t=1}^{T}\sum_{k=1}^{m}\left(\Sigma\left(\frac{\partial}{\partial Z(t)_k}\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}\right) Z(t)\right)_k$$

To distinguish between the $X$ components of $Z(t)$ (the $n$ first ones) and the $Y$ components (the $p$ last ones), we shall now rewrite the above sum as follows: for any $t$,

$$\begin{aligned}
&\sum_{k=1}^{m}\left(\Sigma\left(\frac{\partial}{\partial Z(t)_k}\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}\right) Z(t)\right)_k \qquad (58)\\
&= \sum_{k=1}^{n}\mathbf{e}_k'\Sigma\left(\frac{\partial}{\partial X(t)_k}\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}\right) Z(t) + \sum_{k=1}^{p}\mathbf{e}_{n+k}'\Sigma\left(\frac{\partial}{\partial Y(t)_k}\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}\right) Z(t),
\end{aligned}$$

where the $\mathbf{e}_i$'s denote the (column) vectors of the canonical basis of $\mathbb{R}^m$.

Let us now introduce the $m \times m$ matrix

$$\mathbf{F} := \begin{pmatrix} 0 & \mathbf{C}_{XY} \\ \mathbf{C}_{XY}' & 0 \end{pmatrix}. \tag{59}$$

Note that for any $k \geq 0$ integer, for we have

$$\mathbf{F}^{2k} = \begin{pmatrix} (\mathbf{C}_{XY}\mathbf{C}_{XY}')^k & 0 \\ 0 & (\mathbf{C}_{XY}'\mathbf{C}_{XY})^k \end{pmatrix}, \qquad \mathbf{F}^{2k+1} = \begin{pmatrix} 0 & (\mathbf{C}_{XY}\mathbf{C}_{XY}')^k\mathbf{C}_{XY} \\ (\mathbf{C}_{XY}'\mathbf{C}_{XY})^k\mathbf{C}_{XY}' & 0 \end{pmatrix}$$

so that for $|z|$ large enough,

$$\begin{aligned}
(z-\mathbf{F})^{-1} &= \sum_{k\geq 0} \frac{\mathbf{F}^k}{z^{k+1}} \\
&= \sum_{k\geq 0} z^{-(2k+1)} \begin{pmatrix} (\mathbf{C}_{XY}\mathbf{C}_{XY}')^k & 0 \\ 0 & (\mathbf{C}_{XY}'\mathbf{C}_{XY})^k \end{pmatrix} \\
&\quad + \sum_{k\geq 0} z^{-2(k+1)} \begin{pmatrix} 0 & (\mathbf{C}_{XY}\mathbf{C}_{XY}')^k\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'(\mathbf{C}_{XY}\mathbf{C}_{XY}')^k & 0 \end{pmatrix} \\
&= \begin{pmatrix} z(z^2-\mathbf{C}_{XY}\mathbf{C}_{XY}')^{-1} & (z^2-\mathbf{C}_{XY}\mathbf{C}_{XY}')^{-1}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'(z^2-\mathbf{C}_{XY}\mathbf{C}_{XY}')^{-1} & z(z^2-\mathbf{C}_{XY}'\mathbf{C}_{XY})^{-1} \end{pmatrix} \\
&= \begin{pmatrix} z\mathbf{G} & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & z\widetilde{\mathbf{G}} \end{pmatrix},
\end{aligned} \tag{60}$$

which is true for all $z \in \mathbb{C}\backslash\mathbb{R}$, by analytic continuation.

**Lemma 4.3.** *For any $t$, we have, for $k = 1, \ldots, n$,*

$$\begin{aligned}
\frac{\partial}{\partial X(t)_k} \begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0 \end{pmatrix} &= \frac{1}{2T}(z-\mathbf{F})^{-1} \begin{pmatrix} 0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0 \end{pmatrix} (z-\mathbf{F})^{-1} + \\
&\quad \frac{1}{2T}(z+\mathbf{F})^{-1} \begin{pmatrix} 0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0 \end{pmatrix} (z+\mathbf{F})^{-1}
\end{aligned}$$

*for $\mathbf{e}_k$ the $k$-th (column) vector of the canonical basis in $\mathbb{R}^n$ and we have, for $k = 1, \ldots, p$,*

$$\begin{aligned}
\frac{\partial}{\partial Y(t)_k} \begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0 \end{pmatrix} &= \frac{1}{2T}(z-\mathbf{F})^{-1} \begin{pmatrix} 0 & X(t)\mathbf{e}_k' \\ \mathbf{e}_k X(t)' & 0 \end{pmatrix} (z-\mathbf{F})^{-1} + \\
&\quad \frac{1}{2T}(z+\mathbf{F})^{-1} \begin{pmatrix} 0 & X(t)\mathbf{e}_k' \\ \mathbf{e}_k X(t)' & 0 \end{pmatrix} (z+\mathbf{F})^{-1}
\end{aligned}$$

*for $\mathbf{e}_k$ the $k$-th (column) vector of the canonical basis in $\mathbb{R}^p$*

*Proof.* We define the function

$$\varphi(s) := \frac{s}{z^2-s^2} = \frac{1}{2}\left(\frac{1}{z-s} - \frac{1}{z+s}\right) \qquad (s \in \mathbb{R}).$$

It is easy to see, by (60), that we have

$$\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0 \end{pmatrix} = \varphi(\mathbf{F})$$

We want to compute the derivatives, at $Z(1), \ldots, \widehat{Z(t)}, \ldots, Z(T)$ fixed, of the function

$$Z(t) \mapsto \varphi(\mathbf{F}).$$

The differential of the function $M \mapsto (z - M)^{-1}$ at the matrix $M$ is the operator

$$H \mapsto (z - M)^{-1} H (z - M)^{-1},$$

the differential of the function $M \mapsto (z + M)^{-1}$ at the matrix $M$ is the operator

$$H \mapsto -(z + M)^{-1} H (z + M)^{-1},$$

hence the differential of the function $M \mapsto \varphi(M)$ at the matrix $M$ is the operator

$$H \mapsto \frac{1}{2}\left((z - M)^{-1} H (z - M)^{-1} + (z + M)^{-1} H (z + M)^{-1}\right).$$

Besides, at $Z(1), \ldots, \widehat{Z(t)}, \ldots, Z(T)$ fixed, the differential of the function $Z(t) \mapsto \mathbf{C}_{XY}$ at $Z(t)$ is the operator

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{T}\left(X(t)y' + xY(t)'\right),$$

so that the differential of the function $Z(t) \mapsto \mathbf{F}$ at $Z(t)$ is the operator

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{T}\begin{pmatrix} 0 & X(t)y' + xY(t)' \\ yX(t)' + Y(t)x' & 0 \end{pmatrix}.$$

It follows that at $Z(1), \ldots, \widehat{Z(t)}, \ldots, Z(T)$ fixed, the differential of the function $Z(t) \mapsto \varphi(\mathbf{F})$ at $Z(t)$ is the operator

$$\begin{aligned}\begin{pmatrix} x \\ y \end{pmatrix} \mapsto & \frac{1}{2T}(z - \mathbf{F})^{-1}\begin{pmatrix} 0 & X(t)y' + xY(t)' \\ yX(t)' + Y(t)x' & 0 \end{pmatrix}(z - \mathbf{F})^{-1} + \\ & \frac{1}{2T}(z + \mathbf{F})^{-1}\begin{pmatrix} 0 & X(t)y' + xY(t)' \\ yX(t)' + Y(t)x' & 0 \end{pmatrix}(z + \mathbf{F})^{-1}\end{aligned}$$

The conclusion follows. □

We deduce that

$$\begin{aligned}\frac{\partial}{\partial X(t)_k}\begin{pmatrix} 0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0 \end{pmatrix} Z(t) \quad = \quad & \frac{1}{2T}(z - \mathbf{F})^{-1}\begin{pmatrix} 0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0 \end{pmatrix}(z - \mathbf{F})^{-1} Z(t) + \\ & \frac{1}{2T}(z + \mathbf{F})^{-1}\begin{pmatrix} 0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0 \end{pmatrix}(z + \mathbf{F})^{-1} Z(t)\end{aligned}$$

By (60),

$$(z-\mathbf{F})^{-1}\begin{pmatrix}0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0\end{pmatrix}(z-\mathbf{F})^{-1} =$$

$$\begin{pmatrix} z\mathbf{G} & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & z\widetilde{\mathbf{G}}\end{pmatrix}\begin{pmatrix}0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0\end{pmatrix}\begin{pmatrix} z\mathbf{G} & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & z\widetilde{\mathbf{G}}\end{pmatrix} =$$

$$\begin{pmatrix} z\mathbf{G}\mathbf{e}_kY(t)'\mathbf{C}_{XY}'\mathbf{G}+z\mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G} & z^2\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}}+\mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G}\mathbf{e}_kY(t)'\mathbf{C}_{XY}'\mathbf{G}+z^2\widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G} & z\mathbf{C}_{XY}'\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}}+z\widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}\end{pmatrix}$$

Thus

$$(z-\mathbf{F})^{-1}\begin{pmatrix}0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0\end{pmatrix}(z-\mathbf{F})^{-1}Z(t) =$$

$$\begin{pmatrix} z\mathbf{G}\mathbf{e}_kY(t)'\mathbf{C}_{XY}'\mathbf{G}X(t)+z\mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G}X(t)+z^2\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}}Y(t)+\mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}Y(t) \\ \mathbf{C}_{XY}'\mathbf{G}\mathbf{e}_kY(t)'\mathbf{C}_{XY}'\mathbf{G}X(t)+z^2\widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}X(t)+z\mathbf{C}_{XY}'\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}}Y(t)+z\widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}Y(t)\end{pmatrix}$$

Then, it is easy to see, by (60), that computing

$$(z+\mathbf{F})^{-1}\begin{pmatrix}0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0\end{pmatrix}(z+\mathbf{F})^{-1}Z(t)$$

amounts to take the same formula and change $\mathbf{C}_{XY}$ into $-\mathbf{C}_{XY}$. After, adding both and dividing by $2T$ amounts to keep only, in the previous formula, the terms which are even in $\mathbf{C}_{XY}$ (and divide by $T$). We get

$$\frac{\partial}{\partial X(t)_k}\begin{pmatrix}0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}Z(t) = \frac{1}{T}\begin{pmatrix} z^2\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}}Y(t)+\mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}Y(t) \\ z^2\widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}X(t)+\mathbf{C}_{XY}'\mathbf{G}\mathbf{e}_kY(t)'\mathbf{C}_{XY}'\mathbf{G}X(t)\end{pmatrix} \tag{61}$$

In the same way,

$$\frac{\partial}{\partial Y(t)_k}\begin{pmatrix}0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}Z(t) = \frac{1}{T}\begin{pmatrix} z^2\mathbf{G}X(t)\mathbf{e}_k'\widetilde{\mathbf{G}}Y(t)+\mathbf{G}\mathbf{C}_{XY}\mathbf{e}_kX(t)'\mathbf{G}\mathbf{C}_{XY}Y(t) \\ z^2\widetilde{\mathbf{G}}\mathbf{e}_kX(t)'\mathbf{G}X(t)+\mathbf{C}_{XY}'\mathbf{G}X(t)\mathbf{e}_k'\mathbf{C}_{XY}'\mathbf{G}X(t)\end{pmatrix} \tag{62}$$

Let us write

$$\Sigma = \begin{pmatrix}\mathcal{A} & \mathcal{C} \\ \mathcal{C}' & \mathcal{B}\end{pmatrix}, \quad \mathcal{A}=\operatorname{Cov}(X),\ \mathcal{C}=\operatorname{Cov}(X,Y),\ \mathcal{B}=\operatorname{Cov}(Y).$$

By (58) , (61) and (62) (and using (76), (77) and the facts that $\mathbf{G}'=\mathbf{G}$ and $\widetilde{\mathbf{G}}'=\widetilde{\mathbf{G}}$), we have

$$\begin{aligned}&\sum_{k=1}^m\left(\Sigma\left(\frac{\partial}{\partial Z(t)_k}\begin{pmatrix}0 & \mathbf{G}\mathbf{C}_{XY} \\ \mathbf{C}_{XY}'\mathbf{G} & 0\end{pmatrix}\right)Z(t)\right)_k \\ =\;& \frac{1}{T}\sum_{k=1}^n \mathbf{e}_k'\mathcal{A}\left(z^2\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}}Y(t)+\mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}Y(t)\right)+\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{T}\sum_{k=1}^{n}\mathbf{e}_k'\mathcal{C}\left(z^2\widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}X(t)+\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{e}_kY(t)'\mathbf{C}_{XY}{}'\mathbf{G}X(t)\right)+\\
&\frac{1}{T}\sum_{k=1}^{p}\mathbf{e}_k'\mathcal{C}'\left(z^2\mathbf{G}X(t)\mathbf{e}_k'\widetilde{\mathbf{G}}Y(t)+\mathbf{G}\mathbf{C}_{XY}\mathbf{e}_kX(t)'\mathbf{G}\mathbf{C}_{XY}Y(t)\right)+\\
&\frac{1}{T}\sum_{k=1}^{p}\mathbf{e}_k'\mathcal{B}\left(z^2\widetilde{\mathbf{G}}\mathbf{e}_kX(t)'\mathbf{G}X(t)+\mathbf{C}_{XY}{}'\mathbf{G}X(t)\mathbf{e}_k'\mathbf{C}_{XY}{}'\mathbf{G}X(t)\right)\\
=\;&\frac{1}{T}\left(z^2(\operatorname{Tr}\mathcal{A}\mathbf{G})Y(t)'\widetilde{\mathbf{G}}Y(t)+Y(t)'\mathbf{C}_{XY}{}'\mathbf{G}\mathcal{A}\mathbf{G}\mathbf{C}_{XY}Y(t)\right)+\\
&\frac{2}{T}\left(z^2X(t)'\mathbf{G}\mathcal{C}\widetilde{\mathbf{G}}Y(t)+(\operatorname{Tr}\mathcal{C}\mathbf{C}_{XY}{}'\mathbf{G})Y(t)'\mathbf{C}_{XY}{}'\mathbf{G}X(t)\right)+\\
&\frac{1}{T}\left(z^2(\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}})X(t)'\mathbf{G}X(t)+X(t)'\mathbf{G}\mathbf{C}_{XY}\mathcal{B}\mathbf{C}_{XY}{}'\mathbf{G}X(t)\right)
\end{aligned}
\tag{63}
$$

Let us now sum (63) over $t=1,\dots,T$. Having in mind that

$$
\frac{1}{T}\sum_t X(t)Y(t)'=\mathbf{C}_{XY},\qquad \mathbf{C}_X:=\frac{1}{T}\sum_t X(t)X(t)'\qquad\text{and}\qquad \mathbf{C}_Y:=\frac{1}{T}\sum_t Y(t)Y(t)',
$$

we get

$$
\begin{aligned}
&\sum_t\sum_{k=1}^{m}\left(\Sigma\left(\frac{\partial}{\partial Z(t)_k}\begin{pmatrix}0 & \mathbf{G}\mathbf{C}_{XY}\\ \mathbf{C}_{XY}{}'\mathbf{G} & 0\end{pmatrix}\right)Z(t)\right)_k=\\
&z^2\operatorname{Tr}\mathcal{A}\mathbf{G}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y+\operatorname{Tr}\mathbf{C}_{XY}{}'\mathbf{G}\mathcal{A}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_Y+2z^2\operatorname{Tr}\mathbf{G}\mathcal{C}\widetilde{\mathbf{G}}\mathbf{C}_{XY}{}'+\\
&2\operatorname{Tr}\mathcal{C}\mathbf{C}_{XY}{}'\mathbf{G}\operatorname{Tr}\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{C}_{XY}+z^2\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}}\operatorname{Tr}\mathbf{G}\mathbf{C}_X+\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathcal{B}\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{C}_X
\end{aligned}
\tag{64}
$$

Joining (57) and (64), we get

$$
\begin{aligned}
&\mathbb{E}\operatorname{Tr}\mathbf{C}_{XY}\mathbf{C}_{XY}{}'\mathbf{G}(z)=\\
&\mathbb{E}\operatorname{Tr}\mathcal{C}\mathbf{C}_{XY}{}'\mathbf{G}+\frac{1}{2T}\mathbb{E}\Big[z^2\operatorname{Tr}\mathcal{A}\mathbf{G}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y+\operatorname{Tr}\mathbf{C}_{XY}{}'\mathbf{G}\mathcal{A}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_Y+2z^2\operatorname{Tr}\mathbf{G}\mathcal{C}\widetilde{\mathbf{G}}\mathbf{C}_{XY}{}'+\\
&2\operatorname{Tr}\mathcal{C}\mathbf{C}_{XY}{}'\mathbf{G}\operatorname{Tr}\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{C}_{XY}+z^2\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}}\operatorname{Tr}\mathbf{G}\mathbf{C}_X+\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathcal{B}\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{C}_X\Big],
\end{aligned}
\tag{65}
$$

which allows to conclude.

4.4.2. *Proof of* (50)*: expansion of* $\mathbb{E}\operatorname{Tr}\mathbf{G}\mathbf{C}_X$. For $\mathbf{F}$ as in (59), by (60), we have

$$
\begin{pmatrix}\mathbf{G} & 0\\ 0 & \widetilde{\mathbf{G}}\end{pmatrix}\;=\;\psi(\mathbf{F})
\tag{66}
$$

for

$$
\psi(s)\;:=\;\frac{1}{z^2-s^2}\;=\;\frac{1}{2z}\left(\frac{1}{z-s}+\frac{1}{z+s}\right).
\tag{67}
$$

It follows that for $P := \begin{pmatrix} I_n & 0 \\ 0 & 0 \end{pmatrix}$,

$$\begin{pmatrix} \mathbf{G} & 0 \\ 0 & 0 \end{pmatrix} = P\psi(\mathbf{F}).$$

For $Z(t)$ as defined in (56), we have, by (78) of Corollary 5.2,

$$\begin{aligned}
\mathbb{E}\operatorname{Tr}\mathbf{G}\mathbf{C}_X &= \frac{1}{T}\sum_t \mathbb{E}\, X(t)'\mathbf{G}X(t) \\
&= \frac{1}{T}\sum_t \mathbb{E}\, Z(t)'\begin{pmatrix} \mathbf{G} & 0 \\ 0 & 0 \end{pmatrix} Z(t) \\
&= \mathbb{E}\operatorname{Tr}\mathcal{A}\mathbf{G} + \frac{1}{T}\mathbb{E}\sum_t \sum_{k=1}^m \mathbf{e}_k'\Sigma\frac{\partial}{\partial Z(t)_k}\left(P\psi(\mathbf{F})\right)Z(t) \\
&= \mathbb{E}\operatorname{Tr}\mathcal{A}\mathbf{G} + \frac{1}{T}\mathbb{E}\sum_t\left(\sum_{k=1}^n \mathbf{e}_k'\Sigma\frac{\partial}{\partial X(t)_k}\left(P\psi(\mathbf{F})\right)Z(t) + \sum_{l=1}^p \mathbf{e}_{n+l}'\Sigma\frac{\partial}{\partial Y(t)_l}\left(P\psi(\mathbf{F})\right)Z(t)\right)
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{\partial}{\partial X(t)_k}P\psi(\mathbf{F}) &= \frac{1}{2zT}P(z-\mathbf{F})^{-1}\begin{pmatrix} 0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0 \end{pmatrix}(z-\mathbf{F})^{-1} \\
&\quad -\frac{1}{2zT}P(z+\mathbf{F})^{-1}\begin{pmatrix} 0 & \mathbf{e}_k Y(t)' \\ Y(t)\mathbf{e}_k' & 0 \end{pmatrix}(z+\mathbf{F})^{-1} \\
&= \frac{1}{T}\begin{pmatrix} \mathbf{G}\mathbf{e}_k Y(t)'\mathbf{C}_{XY}{}'\mathbf{G} + \mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G} & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned}$$

and in the same way,

$$\frac{\partial}{\partial Y(t)_l}P\psi(\mathbf{F}) = \frac{1}{T}\begin{pmatrix} \mathbf{G}X(t)\mathbf{e}_l'\mathbf{C}_{XY}{}'\mathbf{G} + \mathbf{G}\mathbf{C}_{XY}\mathbf{e}_l X(t)'\mathbf{G} & 0 \\ 0 & 0 \end{pmatrix}$$

We deduce that

$$\begin{aligned}
\mathbb{E}\operatorname{Tr}\mathbf{G}\mathbf{C}_X &= \mathbb{E}\operatorname{Tr}\mathcal{A}\mathbf{G} + \frac{1}{T^2}\mathbb{E}\sum_t\sum_{k=1}^n \mathbf{e}_k'\mathcal{A}\left(\mathbf{G}\mathbf{e}_k Y(t)'\mathbf{C}_{XY}{}'\mathbf{G} + \mathbf{G}\mathbf{C}_{XY}Y(t)\mathbf{e}_k'\mathbf{G}\right)X(t) \\
&\quad + \frac{1}{T^2}\mathbb{E}\sum_t\sum_{l=1}^p \mathbf{e}_l'\mathcal{C}'\left(\mathbf{G}X(t)\mathbf{e}_l'\mathbf{C}_{XY}{}'\mathbf{G} + \mathbf{G}\mathbf{C}_{XY}\mathbf{e}_l X(t)'\mathbf{G}\right)X(t) \\
&= \mathbb{E}\Big[\operatorname{Tr}\mathcal{A}\mathbf{G} + \frac{1}{T}\operatorname{Tr}\mathcal{A}\mathbf{G}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}{}' + \frac{1}{T^2}\sum_t X(t)'\mathbf{G}\mathcal{A}\mathbf{G}\mathbf{C}_{XY}Y(t) \\
&\quad + \frac{1}{T^2}\sum_t X(t)'\mathbf{G}\mathbf{C}_{XY}\mathcal{C}'\mathbf{G}X(t) + \frac{1}{T}\operatorname{Tr}\mathcal{C}'\mathbf{G}\mathbf{C}_{XY}\operatorname{Tr}\mathbf{G}\mathbf{C}_X\Big]
\end{aligned}$$

$$= \mathbb{E}\Big[\operatorname{Tr}\mathcal{A}\mathbf{G} + \frac{1}{T}\operatorname{Tr}\mathcal{A}\mathbf{G}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}{}' + \frac{1}{T}\operatorname{Tr}\mathbf{G}\mathcal{A}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}{}' + \frac{1}{T}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathcal{C}'\mathbf{G}\mathbf{C}_X + \frac{1}{T}\operatorname{Tr}\mathcal{C}'\mathbf{G}\mathbf{C}_{XY}\operatorname{Tr}\mathbf{G}\mathbf{C}_X\Big],$$

which allows to conclude.

4.4.3. *Proof of* (51)*: expansion of* $\mathbb{E}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y$*.* For $\mathbf{F}$ as in (59), by (60), we have

$$\begin{pmatrix}\mathbf{G} & 0\\ 0 & \widetilde{\mathbf{G}}\end{pmatrix} = \psi(\mathbf{F})$$

for

$$\psi(s) := \frac{1}{z^2 - s^2} = \frac{1}{2z}\left(\frac{1}{z-s} + \frac{1}{z+s}\right).$$

It follows that for $Q := \begin{pmatrix}0 & 0\\ 0 & I_p\end{pmatrix}$,

$$\begin{pmatrix}0 & 0\\ 0 & \widetilde{\mathbf{G}}\end{pmatrix} = P\psi(\mathbf{F}).$$

For $Z(t)$ as defined in (56), we have, by (78) of Corollary 5.2,

$$\begin{aligned}
\mathbb{E}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y &= \frac{1}{T}\sum_t \mathbb{E}\,Y(t)'\widetilde{\mathbf{G}}Y(t)\\
&= \frac{1}{T}\sum_t \mathbb{E}\,Z(t)'\begin{pmatrix}0 & 0\\ 0 & \widetilde{\mathbf{G}}\end{pmatrix}Z(t)\\
&= \mathbb{E}\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}} + \frac{1}{T}\mathbb{E}\sum_t\sum_{k=1}^m \mathbf{e}_k'\Sigma\frac{\partial}{\partial Z(t)_k}\left(P\psi(\mathbf{F})\right)Z(t)\\
&= \mathbb{E}\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}} + \frac{1}{T}\mathbb{E}\sum_t\left(\sum_{k=1}^n \mathbf{e}_k'\Sigma\frac{\partial}{\partial X(t)_k}\left(P\psi(\mathbf{F})\right)Z(t) + \right.\\
&\qquad\left.\sum_{l=1}^p \mathbf{e}_{n+l}'\Sigma\frac{\partial}{\partial Y(t)_l}\left(P\psi(\mathbf{F})\right)Z(t)\right)
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{\partial}{\partial X(t)_k}P\psi(\mathbf{F}) &= \frac{1}{2zT}P(z-\mathbf{F})^{-1}\begin{pmatrix}0 & \mathbf{e}_kY(t)'\\ Y(t)\mathbf{e}_k' & 0\end{pmatrix}(z-\mathbf{F})^{-1}\\
&\quad -\frac{1}{2zT}P(z+\mathbf{F})^{-1}\begin{pmatrix}0 & \mathbf{e}_kY(t)'\\ Y(t)\mathbf{e}_k' & 0\end{pmatrix}(z+\mathbf{F})^{-1}\\
&= \frac{1}{T}\begin{pmatrix}0 & 0\\ 0 & \mathbf{C}_{XY}{}'\mathbf{G}\mathbf{e}_kY(t)'\widetilde{\mathbf{G}} + \widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}\end{pmatrix}
\end{aligned}$$

and in the same way,

$$\frac{\partial}{\partial Y(t)_l} P\psi(\mathbf{F}) \quad = \quad \frac{1}{T}\begin{pmatrix} 0 & 0 \\ 0 & \mathbf{C}_{XY}{}'\mathbf{G}X(t)\mathbf{e}_l'\widetilde{\mathbf{G}} + \widetilde{\mathbf{G}}\mathbf{e}_l X(t)'\mathbf{G}\mathbf{C}_{XY} \end{pmatrix}$$

We deduce that

$$\begin{aligned}
\mathbb{E}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y \quad &= \quad \mathbb{E}\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}} + \frac{1}{T^2}\mathbb{E}\sum_t\sum_{k=1}^n \mathbf{e}_k'\mathcal{C}\left(\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{e}_k Y(t)'\widetilde{\mathbf{G}} + \widetilde{\mathbf{G}}Y(t)\mathbf{e}_k'\mathbf{G}\mathbf{C}_{XY}\right)Y(t) \\
&\quad + \frac{1}{T^2}\mathbb{E}\sum_t\sum_{l=1}^p \mathbf{e}_l'\mathcal{B}\left(\mathbf{C}_{XY}{}'\mathbf{G}X(t)\mathbf{e}_l'\widetilde{\mathbf{G}} + \widetilde{\mathbf{G}}\mathbf{e}_l X(t)'\mathbf{G}\mathbf{C}_{XY}\right)Y(t) \\
&= \quad \mathbb{E}\Big[\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}} + \frac{1}{T}\operatorname{Tr}\mathcal{C}\mathbf{C}_{XY}{}'\mathbf{G}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y + \frac{1}{T^2}\sum_t Y(t)'\mathbf{C}_{XY}{}'\mathbf{G}\mathcal{C}\widetilde{\mathbf{G}}Y(t) \\
&\quad + \frac{1}{T^2}\sum_t Y(t)'\widetilde{\mathbf{G}}\mathcal{B}\mathbf{C}_{XY}{}'\mathbf{G}X(t) + \frac{1}{T}\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}{}'\Big] \\
&= \quad \mathbb{E}\Big[\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}} + \frac{1}{T}\operatorname{Tr}\mathcal{C}\mathbf{C}_{XY}{}'\mathbf{G}\operatorname{Tr}\widetilde{\mathbf{G}}\mathbf{C}_Y + \frac{1}{T}\operatorname{Tr}\mathbf{C}_{XY}{}'\mathbf{G}\mathcal{C}\widetilde{\mathbf{G}}\mathbf{C}_Y \\
&\quad + \frac{1}{T}\operatorname{Tr}\widetilde{\mathbf{G}}\mathcal{B}\mathbf{C}_{XY}{}'\mathbf{G}\mathbf{C}_{XY} + \frac{1}{T}\operatorname{Tr}\mathcal{B}\widetilde{\mathbf{G}}\operatorname{Tr}\mathbf{G}\mathbf{C}_{XY}\mathbf{C}_{XY}{}'\Big],
\end{aligned}$$

which allows to conclude.

## 4.5. **Proof of Proposition 2.7.**

4.5.1. *Proof of* (26). A simple application of equality $s_k^{\text{cleaned}} = \mathbf{u}_k'\mathcal{C}\mathbf{v}_k$ from (13) gives:

$$\mathbb{E}\sum_k s_k s_k^{\text{cleaned}} = \mathbb{E}\sum_k s_k\mathbf{u}_k'\mathcal{C}\mathbf{v}_k = \mathbb{E}\operatorname{Tr}\mathbf{C}_{XY}{}'\mathcal{C} = \operatorname{Tr}(\mathbb{E}\,\mathbf{C}_{XY})'\mathcal{C} = \operatorname{Tr}\mathcal{C}'\mathcal{C} = \sum_k (s_k^{\text{true}})^2.$$

4.5.2. *Proof of* (27) *and* (28). We start with the following Gaussian integrals:

**Lemma 4.4.** *Let $m, T \geq 1$, $P, Q \in \mathbb{R}^{m\times m}$ and let $Z \in \mathbb{R}^{m\times T}$ be a matrix whose entries are independent standard Gaussian variables. Then we have*

$$\mathbb{E}\operatorname{Tr} ZZ'QZZ'P = T^2\operatorname{Tr} PQ + T\operatorname{Tr} P'Q + T\operatorname{Tr} P\operatorname{Tr} Q \tag{68}$$

*and*

$$\mathbb{E}\operatorname{Tr} ZZ'Q\operatorname{Tr} ZZ'P = T^2\operatorname{Tr} P\operatorname{Tr} Q + T\operatorname{Tr} PQ + T\operatorname{Tr} P'Q. \tag{69}$$

*Proof.* We have

$$\mathbb{E}\operatorname{Tr} ZZ'QZZ'P \quad = \quad \mathbb{E}\sum_{i,j,k,l,r,s}\mathbb{E}\, Z_{ij}Z_{kj}Q_{kl}Z_{lr}Z_{sr}P_{si}$$

Using then the fact that the entries of $Z$ are even and independent, we get

$$\mathbb{E}\operatorname{Tr} ZZ'QZZ'P \quad = \quad \sum_{i,j,l,r}\mathbb{E}\, Z_{ij}Z_{ij}Q_{il}Z_{lr}Z_{lr}P_{li} + \sum_{i,j,k}\mathbb{E}\, Z_{ij}Z_{kj}Q_{ki}Z_{ij}Z_{kj}P_{ki} +$$

$$
\begin{aligned}
&\sum_{i,j,k} \mathbb{E}\, Z_{ij} Z_{kj} Q_{kk} Z_{kj} Z_{ij} P_{ii} - 2 \sum_{i,j} \mathbb{E}\, Z_{ij} Z_{ij} Q_{ii} Z_{ij} Z_{ij} P_{ii} \\
= &\sum_{(i,j)\neq(l,r)} Q_{il} P_{li} + 3 \sum_{i,j} Q_{ii} P_{ii} + \sum_{i\neq k,j} Q_{ki} P_{ki} + 3 \sum_{i,j} Q_{ii} P_{ii} + \\
&\sum_{i\neq k,j} Q_{kk} P_{ii} + 3 \sum_{i,j} Q_{ii} P_{ii} - 6 \sum_{i,j} Q_{ii} P_{ii} \\
= &\sum_{i,j,l,r} Q_{il} P_{li} - \sum_{i,j} Q_{ii} P_{ii} + 3 \sum_{i,j} Q_{ii} P_{ii} + \\
&\sum_{i,k,j} Q_{ki} P_{ki} - \sum_{i,j} Q_{ii} P_{ii} + 3 \sum_{i,j} Q_{ii} P_{ii} + \\
&\sum_{i,k,j} Q_{kk} P_{ii} - \sum_{i,j} Q_{ii} P_{ii} + 3 \sum_{i,j} Q_{ii} P_{ii} - 6 \sum_{i,j} Q_{ii} P_{ii} \\
= &\, T^2 \operatorname{Tr} PQ + T \operatorname{Tr} P'Q + T \operatorname{Tr} P \operatorname{Tr} Q
\end{aligned}
$$

In the same way,

$$
\mathbb{E} \operatorname{Tr} ZZ'Q \operatorname{Tr} ZZ'P \;=\; \mathbb{E} \sum_{i,j,k,l,r,s} \mathbb{E}\, Z_{ij} Z_{kj} Q_{ki} Z_{lr} Z_{sr} P_{sl}
$$

and using again the fact that the entries of $Z$ are even and independent, we get

$$
\begin{aligned}
\mathbb{E} \operatorname{Tr} ZZ'Q \operatorname{Tr} ZZ'P \;=\; &\sum_{i,j,l,r} \mathbb{E}\, Z_{ij} Z_{ij} Q_{ii} Z_{lr} Z_{lr} P_{ll} + \sum_{i,j,k} \mathbb{E}\, Z_{ij} Z_{kj} Q_{ki} Z_{ij} Z_{kj} P_{ki} \\
&+ \sum_{i,j,k} \mathbb{E}\, Z_{ij} Z_{kj} Q_{ki} Z_{kj} Z_{ij} P_{ik} - 2 \sum_{i,j} \mathbb{E}\, Z_{ij} Z_{ij} Q_{ii} Z_{ij} Z_{ij} P_{ii} \\
= &\sum_{(i,j)\neq(l,r)} Q_{ii} P_{ll} + 3 \sum_{i,j} Q_{ii} P_{ii} + \sum_{i\neq k,j} Q_{ki} P_{ki} + 3 \sum_{i,j} Q_{ii} P_{ii} \\
&+ \sum_{i\neq k,j} Q_{ki} P_{ik} + 3 \sum_{i,j} Q_{ii} P_{ii} - 6 \sum_{i,j} Q_{ii} P_{ii} \\
= &\sum_{i,j,l,r} Q_{ii} P_{ll} - \sum_{i,j} Q_{ii} P_{ii} + 3 \sum_{i,j} Q_{ii} P_{ii} + \sum_{i,k,j} Q_{ki} P_{ki} - \sum_{i,j} Q_{ii} P_{ii} \\
&+ \sum_{i,k,j} Q_{ki} P_{ik} - \sum_{i,j} Q_{ii} P_{ii} \\
= &\sum_{i,j,l,r} Q_{ii} P_{ll} + \sum_{i,k,j} Q_{ki} P_{ki} + \sum_{i,k,j} Q_{ki} P_{ik} \\
= &\, T^2 \operatorname{Tr} P \operatorname{Tr} Q + T \operatorname{Tr} PQ + T \operatorname{Tr} P'Q
\end{aligned}
$$

□

Let us now prove (27) and (28). Recall that

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$$

for $\Sigma = \begin{pmatrix} \mathcal{A} & \mathcal{C} \\ \mathcal{C}' & \mathcal{B} \end{pmatrix} \in \mathbb{R}^{m \times m}$ for $m = n + p$.

Equations (27) and (28) follow directly from the following lemma.

**Lemma 4.5.** *We have*

$$\mathbb{E} \operatorname{Tr} \mathbf{C}_{XY} \mathbf{C}_{XY}' = \frac{T+1}{T} \operatorname{Tr} \mathcal{C}\mathcal{C}' + T^{-1} \operatorname{Tr} \mathcal{A} \operatorname{Tr} \mathcal{B}$$

*and*

$$\mathbb{E} \operatorname{Tr} \mathbf{C}_X \operatorname{Tr} \mathbf{C}_Y = \operatorname{Tr} \mathcal{A} \operatorname{Tr} \mathcal{B} + 2T^{-1} \operatorname{Tr} \mathcal{C}\mathcal{C}'$$

*Proof.* Let $Z \in \mathbb{R}^{m \times T}$ be a matrix whose entries are independent standard Gaussian variables and

$$R = \begin{bmatrix} \mathcal{A} & \mathcal{C} \end{bmatrix} \in \mathbb{R}^{n \times m} \qquad \text{and} \qquad S = \begin{bmatrix} \mathcal{C}' & \mathcal{B} \end{bmatrix} \in \mathbb{R}^{p \times m}, \tag{70}$$

so that $\mathbf{X}$, $\mathbf{Y}$ can be realized by

$$\mathbf{X} = RZ \qquad \text{and} \qquad \mathbf{Y} = SZ. \tag{71}$$

Recall the notation from (5) and (16):

$$\mathbf{C}_X = \frac{1}{T}\mathbf{X}\mathbf{X}', \quad \mathbf{C}_Y = \frac{1}{T}\mathbf{Y}\mathbf{Y}', \quad \mathbf{C}_{XY} = \frac{1}{T}\mathbf{X}\mathbf{Y}'. \tag{72}$$

By Lemma 4.4, we have, for $P := R'R$ and $Q := S'S$,

$$\begin{aligned} T^2 \,\mathbb{E} \operatorname{Tr} \mathbf{C}_{XY} \mathbf{C}_{XY}' &= \mathbb{E} \operatorname{Tr} RZZ'S'SZZ'R' \\ &= \mathbb{E} \operatorname{Tr} ZZ'QZZ'P \end{aligned}$$

and

$$\begin{aligned} T^2 \,\mathbb{E} \operatorname{Tr} \mathbf{C}_X \operatorname{Tr} \mathbf{C}_Y &= \mathbb{E} \operatorname{Tr} RZZ'R' \operatorname{Tr} SZZ'S' \\ &= \mathbb{E} \operatorname{Tr} ZZ'Q \operatorname{Tr} ZZ'P \end{aligned}$$

Then, we conclude noting that $\mathcal{C} = RS'$, $\mathcal{A} = RR'$ and $\mathcal{B} = SS'$. □

4.6. **Proof of concentration results.**

4.6.1. *Proof of Lemma 4.1.* Let $Z \in \mathbb{R}^{m\times T}$ be a matrix whose entries are independent standard Gaussian variables, so that $\mathbf{X}$, $\mathbf{Y}$ can be realized as in (71): $\mathbf{X} = RZ$, $\mathbf{Y} = SZ$, for $R, S$ as in (70). This leads, by (72), to

$$\mathbf{C}_{XY} = \frac{1}{T}RZZ'S', \quad \mathbf{C}_X = \frac{1}{T}RZZ'R \qquad \text{and} \qquad \mathbf{C}_Y = \frac{1}{T}SZZ'S.$$

By the second part of Proposition 5.3, what we have to prove is that the maps associating the variables $G, H, A, B, L, \Theta, K \in \mathbb{C}$ to $Z \in \mathbb{R}^{m\times T}$ are all Lipschitz for the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ from (3) on $\mathbb{R}^{m\times T}$, with Lipschitz constant

$$O\left(\frac{1}{T(\mathfrak{Im} z)^2}\right).$$

As this argument is quite standard (close to e.g. [2, Sec. 2.3.1] or [9, Lem. 7.1] with [3, Lem. B.2] instead of [9, Lem. A.2]), we only give the main lines. Consider a variation $\delta_Z$ of $Z$, and then:

(1) Use the *resolvant formula*: for all square matrices $M, \delta_M$,

$$(z-(M+\delta_M))^{-1} - (z-M)^{-1} \;=\; (z-(M+\delta_M))^{-1}\delta_M(z-M)^{-1}$$

to expand the variations of the matrices $\mathbf{G}$ and $\widetilde{\mathbf{G}}$ at first order in $\delta_Z$.

(2) By non-commutative Hölder inequalities (see e.g. [2, Appendix A.3]), for any product $M_1 \cdots M_k$ of matrices with any size and any $i = 1, \dots, k$,

$$\|M_1\cdots M_k\|_{\mathrm{F}} \le \|M_1\|_{\mathrm{op}} \cdots \widehat{\|M_i\|_{\mathrm{op}}} \cdots \|M_k\|_{\mathrm{op}}\|M_i\|_{\mathrm{F}}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm. This has to be used with the fact that $\mathbf{G}$ and $\widetilde{\mathbf{G}}$ have operator norms $\le \operatorname{dist}(z^2, [0, +\infty))^{-1}$.

(3) On any square matrices space endowed with the Frobenius norm, the trace is the scalar product with the identity matrix, hence is Lipschitz with Lipschitz constant the Frobenius norm of the identity matrix (which depends on the dimension).

4.6.2. *Concentration lemma for Section 2.5.*

**Lemma 4.6.** *With the notation from Section 2.1, for any deterministic vectors* $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^p$, *the random variable* $\mathbf{u}'\mathbf{C}_{XY}\mathbf{v} - \mathbf{u}'\mathcal{C}\mathbf{v}$ *is centered with* $L^2$*-norm* $\le \sqrt{2}\|\Sigma\|_{op}/\sqrt{T}$.

**Remark 4.7.** Using Hanson-Wright inequality [25], one could improve the variance bound into an exponential control on the tail.

*Proof.* Let us use the same notation as in the previous proof (Section 4.6.1). First,

$$\mathbb{E}\,\mathbf{u}'\mathbf{C}_{XY}\mathbf{v} = \mathbf{u}'RS'\mathbf{v} = \mathbf{u}'\mathcal{C}\mathbf{v}. \tag{73}$$

Secondly, we have

$$\mathbf{u}'\mathbf{C}_{XY}\mathbf{v} = \frac{1}{T}\mathbf{u}'RZZ'S'\mathbf{v} = \frac{1}{T}\operatorname{Tr} ZZ'S'\mathbf{v}\mathbf{u}'R$$

so that, by (69),

$$\begin{aligned}\mathbb{E}(\mathbf{u}'\mathbf{C}_{XY}\mathbf{v})^2 &= (\operatorname{Tr} S'\mathbf{v}\mathbf{u}'R)^2 + \frac{1}{T}\operatorname{Tr} S'\mathbf{v}\mathbf{u}'RS'\mathbf{v}\mathbf{u}'R + \frac{1}{T}\operatorname{Tr} R'\mathbf{u}\mathbf{v}'SS'\mathbf{v}\mathbf{u}'R \\ &= (\mathbf{u}'RS'\mathbf{v})^2 + \frac{1}{T}(\mathbf{u}'RS'\mathbf{v})^2 + \frac{1}{T}(\mathbf{u}'RR'\mathbf{u})(\mathbf{v}'SS'\mathbf{v}),\end{aligned}$$

which, by (73), allows to conclude. □

## 5. Appendix

5.1. **Stieltjes transform inversion.** Any signed measure $\mu$ on $\mathbb{R}$ can be recovered out of its Stieltjes transform

$$g_\mu(z) := \int \frac{\mathrm{d}\mu(t)}{z-t}, \quad z \in \mathbb{C}\backslash\mathbb{R} \tag{74}$$

by the formula

$$\mu = -\frac{1}{\pi}\lim_{\eta\to 0^+}(\mathfrak{Im} g_\mu(x+\mathrm{i}\eta)\mathrm{d}x), \tag{75}$$

where the limit holds in the weak topology (see e.g. [2, Th. 2.4.3] and use the decomposition of any signed measure as a difference of finite positive measures).

5.2. **Linear algebra.** We notify some formulas frequently used (and referred to) in the paper: for $(\mathbf{e}_k)$ an orthonormal basis, for any matrices $M, N$,

$$\mathbf{e}_k'M\mathbf{e}_l = \mathbf{e}_l'M'\mathbf{e}_k, \quad \sum_k \mathbf{e}_k'M\mathbf{e}_k = \operatorname{Tr} M, \quad \sum_{k,l}\mathbf{e}_k'M\mathbf{e}_l\mathbf{e}_k'N\mathbf{e}_l = \operatorname{Tr} MN' \tag{76}$$

and for any column vectors $u, v$,

$$\sum_k \mathbf{e}_k'u\mathbf{e}_k'v = \sum_k u'\mathbf{e}_k v'\mathbf{e}_k = v'u. \tag{77}$$

5.3. **Stein formula for Gaussian random vectors.**

**Proposition 5.1.** *Let $X = (X_1, \ldots, X_d)$ be a centered Gaussian vector with covariance $\Sigma$ and $f : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{C}^1$ function with derivatives having at most polynomial growth. Then for all $i_0 = 1, \ldots, d$,*

$$\mathbb{E}\, X_{i_0} f(X_1, \ldots, X_d) = \sum_{k=1}^d \Sigma_{i_0 k}\, \mathbb{E}(\partial_k f)(X_1, \ldots, X_d).$$

(see e.g. [3, Lem. A.1])

**Corollary 5.2.** *With the same notation, considering $X$ as a column vector, for $F : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ a matrix-valued function, we have*

$$\mathbb{E}\, X'F(X)X = \operatorname{Tr}\Sigma\,\mathbb{E}\, F(X) + \sum_{k=1}^d (\mathbb{E}\,\Sigma(\partial_k F)(X)X)_k\,. \tag{78}$$

*Proof.* We have, by Proposition 5.1,

$$\begin{aligned}
\mathbb{E}\, X'F(X)X &= \sum_{ij} \mathbb{E}\, X_i X_j F(X)_{ij} \\
&= \sum_{ijk} \mathbb{E}\, \Sigma_{ik} \frac{\partial}{\partial X_k} X_j F(X)_{ij} \\
&= \sum_{ijk} \mathbb{E}\, \Sigma_{ik} \left(\delta_{j=k} F(X)_{ij} + X_j (\partial_k F)(X)_{ij}\right) \\
&= \operatorname{Tr}\Sigma\, \mathbb{E}\, F(X) + \sum_{ijk} \mathbb{E}\, \Sigma_{ik} X_j (\partial_k F)(X)_{ij} \\
&= \operatorname{Tr}\Sigma\, \mathbb{E}\, F(X) + \sum_{k} \left(\mathbb{E}\, \Sigma (\partial_k F)(X) X\right)_k
\end{aligned}$$

□

5.4. **Concentration of measure for Gaussian vectors.** The following proposition can be found e.g. in [2, Sec. 4.4.1] or [25, Th. 5.2.2].

**Proposition 5.3.** *Let $X = (X_1, \ldots, X_d)$ be a standard real Gaussian vector and $f : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{C}^1$ function with gradient $\nabla f$. Then we have*

$$\operatorname{Var}(f(X)) \;\leq\; \mathbb{E}\, \|\nabla f(X)\|^2, \tag{79}$$

*where $\|\cdot\|$ denotes the standard Euclidian norm.*

*Besides, if $f$ is $k$-Lipschitz, then for any $t > 0$, we have*

$$\mathbb{P}(|f(X) - \mathbb{E}\, f(X)| \geq t) \;\leq\; 2\mathrm{e}^{-\frac{t^2}{2k^2}}, \tag{80}$$

*i.e. $f(X) - \mathbb{E}\, f(X)$ is Sub-Gaussian with Sub-Gaussian norm $\leq k$, up to a universal constant factor.*

## References


[1] Amsalu, S., Duan, J., Matzinger, H., Popescu, I. *Recovery of spectrum from estimated covariance matrices and statistical kernels for machine learning and big data*, arXiv.

[2] Anderson, G., Guionnet, A., Zeitouni, O. *An Introduction to Random Matrices.* Cambridge studies in advanced mathematics, 118 (2009).

[3] Benaych-Georges, F., Couillet, R. *Spectral analysis of the Gram matrix of mixture models*, ESAIM Probab. Statist., Vol. 20 (2016), 217–237.

[4] Benaych-Georges, F., Knowles, A. *Local semicircle law for Wigner matrices. Advanced topics in random matrices*, 1–90, Panor. Synthèses, 53, Soc. Math. France, Paris, 2017.

[5] Bose, A., Bhattacharjee, M. *Large Covariance and Autocovariance Matrices*, 2018, Chapman and Hall/CRC.

[6] Bouchaud, J.-P., Laloux, L., Miceli, M.A., Potters M. *Large dimension forecasting models and random singular value spectra*, Eur. Phys. J. B (2007) 55: 201.

[7] Bun, J., Bouchaud, J.-P., Potters, M. *Cleaning correlation matrices*, Risk magazine, 2016.

[8] Bun, J., Bouchaud, J.-P., Potters, M. *Cleaning large correlation matrices: Tools from Random Matrix Theory*, Physics Reports Volume 666, Review article, 1–109, 2017.
[9] Capitaine, M. *Additive/multiplicative free subordination property and limiting eigenvectors of spiked additive deformations of Wigner matrices and spiked sample covariance matrices*. J. Theoret. Probab. 26 (2013), no. 3, 595–648.
[10] Couillet, R., Pascal, F., Silverstein, J., *Robust estimates of covariance matrices in the large dimensional regime*. IEEE Trans. Inform. Theory 60 (2014), no. 11, 7269–7278.
[11] Couillet, R., Pascal, F., Silverstein, J., *The random matrix regime of Maronna's M-estimator with elliptically distributed samples*. J. Multivariate Anal. 139 (2015), 56–78.
[12] El Karoui, N. *Spectrum estimation for large dimensional covariance matrices using random matrix theory*. Annals of Statistics, 2008, 36(6):2757–2790.
[13] El Karoui, N. *Operator norm consistent estimation of large dimensional sparse covariance matrices*, Annals of Statistics, 2008, 36(6): 2717–2756
[14] El Karoui, N. *Random matrices and high-dimensional statistics: beyond covariance matrices.* Proceedings of the International Congress of Mathematicians, 2018.
[15] Erdős, L., Schlein, B., Yau, H.-T. *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*. Ann. Probab. 37 (2009), no. 3, 815–852.
[16] Erdős, L., Yau, H.-T. *A dynamical approach to random matrix theory*. Courant Lecture Notes in Mathematics, 28, New York; American Mathematical Society, Providence, RI, 2017.
[17] Klopp, O., Lounici, K., Tsybakov, A. B. *Robust matrix completion*. Probab. Theory Related Fields 169 (2017), no. 1–2, 523–564.
[18] Klopp, O., Y. Lu, A. Tsybakov, A. B., Zhou, H. *Structured Matrix Estimation and Completion*, preprint.
[19] Koltchinskii, V., Lounici, K., Tsybakov, A. B. *Estimation of low-rank covariance function*, Stochastic Process. Appl. 126 (2016), no. 12, 3952–3967.
[20] Laloux, L., Cizeau, P., Bouchaud, J.-P., Potters, M. *Noise Dressing of Financial Correlation Matrices*, Phys. Rev. Lett. 83, 1467, 1999.
[21] Ledoit, O., Wolf, M. *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis 88, 2004, 365–411.
[22] Ledoit, O., Wolf, M. *Honey, I shrunk the sample covariance matrix*. Journal of Portfolio Management, 30, Volume 4, 2004, 110–119.
[23] Ledoit, O., Wolf, M. *Nonlinear shrinkage estimation of large-dimensional covariance matrices*. Annals of Statistics 40, 2012, 1024–1060.
[24] Ledoit, O., Péché, S. *Eigenvectors of some large sample covariance matrix ensembles* Probability Theory and Related Fields, 2011, 151.1, 233–264
[25] Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, 2018.

(FBG, JPB, MP) CFM, 23 rue de l'Université, 75007 Paris, France

*E-mail address*: florent.benaych-georges@cfm.fr, jean-philippe.bouchaud@cfm.fr, marc.potters@cfm.fr