An Optimal Rotational Invariant Estimator for General Covariance Matrices: the outliers

Preprint · February 2018

1 author:

Joel Bun
Capital Fund Management / Université Paris-Sud 11

An Optimal Rotational Invariant Estimator for General Covariance Matrices: the outliers

Joël Bun*

February 18, 2018

Abstract

The problem of estimating large covariance matrices is of crucial important in many very different fields. We investigate this problem for a general $p \times p$ population covariance matrix Σ whose spectrum contains a bounded number r of outliers. The observed data are represented through the $p \times p$ sample covariance matrix S coming from a large number n of independent realizations. We place ourselves in the high-dimensional regime where the number of variables p is comparable to n, and except for r, all quantities may depend on n in a arbitrary fashion. This problem has already been considered when Σ does not contain outliers so the focus is set on the outliers. We prove that with high probability, there exists a consistent, unified and optimal rotational invariant estimator of Σ that depends only on observable quantities. To that end, we depict the behaviour of the outliers and their associated eigenvectors for any anisotropic Σ. All the results hold at a local scale, thus providing a precise control of the estimation error. The proof mainly relies on the anisotropic local law for sample covariance matrices established in [23].

Introduction

1.1. Statistical problem. With the emergence of the "Big Data" trend, nowadays scientists try to exploit the large number of features they can collect leading to very high-dimensional problems. One consequence of this setting is that most results of classical multivariate statistics become untrustworthy. In that respect, the ability to handle large amount of data is one of the most challenging problem in a wide range of disciplines such as machine learning, genomics, wireless communications, climatology or economics.

Multivariate statistics focuses on the study of p variables which are thought to possess a certain degree of interdependences. The most basic phenomenon is that of covariances which measure the tendency of quantities to vary together. If we denote by $\mathbf{y} := (y_1, y_2, \ldots, y_p)^*$ the set of zero-mean variables, a typical question is to find the population covariance matrix

$$\Sigma \;:=\; \mathbb{E}\mathbf{y}\mathbf{y}^* = (\mathbb{E}\, y_i y_j)_{i,j=1}^{p} \,. \tag{1.1}$$

This problem is very crucial in multivariate statistics as Σ is a fundamental input in many applications such as regression/classification [11, 15, 36], wireless communications [8] or portfolio optimization [7, 28] to cite a few.

Obviously, the exact nature of the variables $\mathbf{y}$ is unknown so that Σ cannot be reached. A simple approach to bypass this problem is to collect n (independent) samples points for these p variables to produce a $p \times n$ matrix $Y := (Y_{i\mu}) \in \mathbb{R}^{p\times n}$ where $Y_{i\mu}$ denotes the μth sample of the variable y_i. We can then estimate the population covariance matrix (1.1) through its empirical mean, namely the sample covariance matrix

$$S \;:=\; \frac{1}{n} YY^* \;=\; \left(\frac{1}{n}\sum_{\mu=1}^{n} Y_{i\mu}Y_{j\mu}\right)_{i,j\in[\![p]\!]} . \tag{1.2}$$

*Independent researcher, joel.bun@gmail.com

If p is fixed and n tends to infinity, then the sample covariance matrix converges almost surely to the population covariance matrix Σ by the law of large numbers. However, we are rather interested here in the high-dimensional framework where the population size p is of the same order of magnitude than the number of samples n, i.e $\phi := p/n \in (0, \infty)$. In that case, we see that we are trying to estimate $O(p^2)$ quantities out of $O(np)$ observations, which is not sufficient. Hence, a large literature from different communities has flourished in order to find a more reliable estimator of Σ.

The inadequacy of the sample covariance matrix in the high-dimensional regime is now well-known since the so-called Stein's paradox [34] where the authors highlighted that (1.2) is not consistent as the dimension p increases. A nice manifestation of this "paradox" is to look at the spectrum of S which changes dramatically from the *true* one when p is comparable to n [30, 33]. Since then, the estimation of large population covariance matrices has been a classical topic in multivariate statistics for decades (see e.g. [9, 13, 18]) and that becomes even more crucial nowadays. To handle that problem, one typically tries to develop optimal strategies for estimation from a decision theoretic point of view. In that respect, perhaps one of the most influential contribution is the class of linear shrinkage estimators [18]. This estimator is fairly simple

$$\Xi^{lin} = \alpha S + (1-\alpha) I_p, \qquad \alpha \in [0,1], \tag{1.3}$$

and follows from an optimal Bayesian theory under a multivariate Gaussian likelihood function. It has encountered a lot of success in the aforementioned applications due to his robustness to the noise (see [21] for a recent review). However, the main question is how to estimate the parameter α. Different attempts are available in the literature and one famous solution has been given recently in [27]. Nonetheless, the procedure they proposed is sensitive to the presence of outliers, which often occurs in practice. A standard belief in Statistics is that outliers contain non-trivial informations about the exact nature of the underlying system and it is therefore essential to handle them correctly.

More recently, Random Matrix Theory (RMT) has revealed to be very useful for addressing many theoretical questions on high-dimensional statistics (see [31] for a review). As far as estimation of covariance matrices is concerned, RMT has provided many significant progress [4, 14, 25, 32]. These methods make use of the statistical features of the sample eigenvalues and are consistent under the high-dimensional framework. They also give a precise interpretation of the outliers which is clearly missing in the linear shrinkage. However, these procedures appear to be ad-hoc in the sense that they are not designed to be optimal under any decision theory with respect to Σ.

1.2. Rotational Invariant Estimators. In this paper, we attempt to construct an estimator that is optimal from a decision theoretic point of view and can deal with the presence of outliers. To that end, we restrain our study on the specific class of estimators that cannot be biased in any privileged direction(s). Differently said, we do not have any prior knowledge on the components of the population covariance matrix Σ. This can be seen as a pretty safe assumption in the general case.

In multivariate statistics, this class of estimators is known as *rotational invariant estimator (RIE)*, and has been subject to extensive studies [18, 19, 35]. More specifically, we shall only consider RIEs that depend on the data we have which is represented through the sample covariance matrix S. That being said, these estimators own the following properties: it shares the same eigenvectors than S and its eigenvalues are function of the sample ones. This allows to simplify a lot the problem as we now have order p quantities to estimate out of np observations. It is worth mentioning that all the estimators presented above exactly fall down into this class of estimators.

The problem of estimating Σ using RIEs reduces to find the optimal way to clean (or *shrink*) the sample eigenvalues $\lambda_1, \dots, \lambda_p$ in order to obtain a consistent estimator of Σ. In this paper, we use the term consistent to refer to an estimator that is optimal under the high dimensional framework. If we define by $\mathcal{M}(S)$ the set of matrices that share the sample eigenvectors, we define the optimal solution as the set of eigenvalues that minimizes the Hilbert-Schmidt distance with respect to Σ. What we get is an estimator $\tilde{\Xi} \in \mathcal{M}(S)$ whose eigenvalues $[\tilde{\xi}_i]_{i=1}^p$ require the knowledge of Σ itself

$$\tilde{\xi}_i = \langle \mathbf{u}_i , \Sigma \mathbf{u}_i \rangle, \qquad i \in [\![p]\!], \tag{1.4}$$

and thus, we will refer to this estimator as the *oracle* estimator. Clearly, such an estimator is pointless for

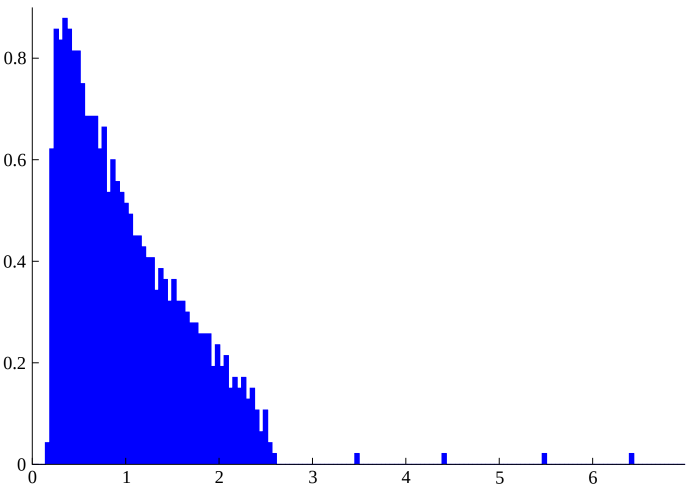


FIGURE 1.1. A histogram of the eigenvalue density of S with $p = 1000$ and $\phi = 0.3$. Here, we considered $\Sigma = I_p + W$ where W belongs to the Gaussian Orthogonal Ensemble of width $s = 0.25$ with 4 outliers located at $\{3, 4, 5, 6\}$.

practical applications, but it nonetheless tells us that any *optimal* estimator of Σ in $\mathcal{M}(S)$ (with respect to the Hilbert-Schmidt norm) has to perform as well as the oracle solution (1.4).

Recently, several works have investigated the asymptotic behaviour of the oracle eigenvalues in the limit of infinite dimension and this is where tools from RMT are useful [6,26]. More precisely, they showed using two different approaches that in the absence of outliers, the oracle estimator (1.4) converges to a limiting deterministic quantity that is independent from the population covariance matrix for any $i \in [\![p]\!]$. However the formalism used in those two papers suffer from two major weaknesses: (i) it only works at a global scale, meaning that it does not deal with outliers, neither give explicit error bounds; (ii) both use limiting statements, and for applications purposes, one is rather interested to the case where n is large but *bounded*.

1.3. Outline of the result. The aim of this paper is thus to provide a *consistent*, *unified* and *observable* optimal RIE for a *general* population covariance matrix Σ under the high-dimensional framework. We will prove that under mild assumptions, there exists an optimal RIE *with no hyperparameter* that is very close to the oracle estimator $\tilde{\Xi}$ with high probability.

Let us first fix some terminologies and state briefly the main assumptions in our model. Let S be defined as in (2.5). We assume that both p and n are comparable and that the entries of $Y = \sqrt{n}\Sigma^{1/2}X$ possess a sufficient number of bounded moments. Furthermore, we suppose that $\|\Sigma\|$ is bounded. Next, let us denote by $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$ the eigenvalues of S and we define by ϱ the asymptotic density of $X^*\Sigma X$ by ϱ which has the same nonzero eigenvalues than S. All our assumptions concerning the spectrum of Σ can be stated through the density ϱ. More specifically and motivated by statistical applications, we assume that ϱ consists of $r + 1$ connected components with r finite where:

(i) the top r components define the non-denegerated outliers which are well separated from the bulk and from each other;

(ii) the remaining component characterizes the *bulk* in which ϱ has a square root decay near its edges.

Our model encompasses the well-known *Spiked Covariance Matrix* model [20] and we give a simple example of it in Figure 1.1.

We are now ready to state the main result, which is a *local* and *fully observable* estimate of the oracle estimator $\tilde{\Xi}$. Let $\phi = p/n \in (0, \infty)$ and $\eta = n^{-1/2}$. We define $\hat{\Xi} \in \mathcal{M}(S)$ whose eigenvalues are defined by

$$\hat{\xi}_i := \frac{1}{\lambda_i |s(\lambda_i + \mathrm{i}\eta)|^2}, \qquad s(z) = \frac{1}{n}\sum_{i=1}^{p} \frac{1}{\lambda_i - z} - \frac{1-\phi}{z},$$

for any $i \in [\![p]\!]$ provided that $\lambda_i \geqslant \tau$. Note that $\hat{\Xi}$ is clearly a RIE while this is not the case for $\tilde{\Xi}$. Moreover, we see that the expression of $\hat{\Xi}$ is *universal* in the sense that we do not have to distinguish whether an eigenvalue is an outlier or not. In this paper, we establish the closeness between the outliers of $\hat{\Xi}$ and those of $\tilde{\Xi}$ as we shall prove that with high probability,

$$\hat{\xi}_i \approx \langle \mathbf{u}_i \,, \Sigma \mathbf{u}_i \rangle, \tag{1.5}$$

for any $i \in [\![r]\!]$ provided that $\lambda_i \geqslant \tau$ and n is large enough. The proof of (1.5) for the bulk eigenvalues will be given in a forthcoming paper as it involves different arguments.

The estimator (1.5) appears to be similar to the results of [26] or [6] but the differences are in fact substantials. First, (1.5) is a *local* estimate in the sense that the result holds for $\eta \gg n^{-1}$ while [6, 26] worked at a global scale ($\eta \sim O(1)$), hence ruling out outliers. Secondly, we emphasize that the result is purely *quantitative*, meaning that we do not need to consider the limit $n \to \infty$ in our definition of the high-dimensional limit compared to previous works [6, 26]. This allows us to provide a convergence rate of the result. All in all, we provide here a high probability optimal estimator for a general class of population covariance matrix Σ that relies only on the data we have. We emphasize that this result has already been mentioned in the PhD thesis of the author with a short sketch of the proof [5] and then in the review paper [7]. During the preparation of the current manuscript, an alternative proof of the same result has been proposed in [10].

The outline of the paper is as follows. In section 2, we give the precise definitions and assumptions of our model. We then present the class of rotational invariant estimators and discuss some of its properties. This is followed by the precise statement of the main theorem of this paper. We stress that the theorem we shall prove is performed under a Gaussian assumption on the entries of Y. However, this is made only in order to keep the presentation short and readable and we then discuss how to relax some assumptions (e.g. Gaussian entries or single bulk component for ϱ) without altering the result and the different techniques used in our proof. In section 3, we provide some theoretical and numerical applications of our result. In particular, the theoretical part focuses on the spectrum of $\hat{\Xi}$ and deals with some special cases where analytical expressions are possible. The rest of the paper will be dedicated to the proof of the theorem and we collect several proofs of subsidiary results in the appendices.

Conventions. The fundamental large parameter is n. Except for r, all quantities that are not explicitly constant may depend on n, and we usually omit this dependence from our notation. We use $C > 0$ to denote a generic large constant and $c > 0$ a generic small constant. We write $a \asymp b$ to mean $C^{-1}a \leqslant b \leqslant Ca$. We use capital letters for matrices and bold lowercase letters for vectors, which we regard as $p \times 1$ matrices. We denote by $\|\cdot\|$ and $\|\cdot\|_2$ the Euclidean operator norm and the Hilbert-Schmidt norm of a matrix, respectively. Finally, we use the abbreviations $[\![a, b]\!] := [a, b] \cap \mathbb{N}$ and $[\![1, a]\!] \equiv [\![a]\!]$ for $a, b \in \mathbb{N}$.

Acknowledgments. I wish to warmly thank Antti Knowles for his significant contribution in most results of this work and his advice. This work would not have been possible without his help. I am also grateful to Jean-Philippe Bouchaud who initiated this project and for his intuitions. Special thanks are addressed to Marc Potters and Jean-Yves Audibert for their careful reading of the manuscript and their comments. Finally, I thank the ETH Zürich for its hospitality.

Model and result

In this section we give the precise statements of our assumptions and our main result.

2.1. Model. Let $\mathbf{y} \in \mathbb{R}^p$ be a mean-zero random vector, and define the *population covariance matrix* of $\mathbf{y}$ through

$$\Sigma \;:=\; \mathbb{E}\mathbf{y}\mathbf{y}^* \,.$$

Hence, Σ is a real, symmetric, deterministic, nonnegative $p \times p$ matrix, and we use the notation

$$\Sigma \;=\; \sum_{i=1}^{p} \sigma_i \mathbf{v}_i \mathbf{v}_i^* \tag{2.1}$$

for the eigenvalues $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_p \geqslant 0$ and the associated normalized eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^p$ of Σ. We suppose that Σ is bounded,

$$\sigma_1 \;=\; \|\Sigma\| \;\leqslant\; C \,. \tag{2.2}$$

Let

$$\pi \;:=\; \frac{1}{p} \sum_{i=1}^{p} \delta_{\sigma_i} \tag{2.3}$$

denotes the empirical spectral density of Σ. We suppose that the spectrum of Σ is not concentrated at zero,

$$\pi([0,c]) \;\leqslant\; 1 - c. \tag{2.4}$$

Next, for $n \in \mathbb{N}$ let $Y = (Y_{i\mu}) \in \mathbb{R}^{p \times n}$ be a matrix whose columns $(Y_{i\mu})_{i \in [\![p]\!]}$, $\mu \in [\![n]\!]$, are i.i.d. copies of $\mathbf{y}$. We define the *sample covariance matrix*

$$S \;:=\; \frac{1}{n} Y Y^* \,. \tag{2.5}$$

Note that S is random and satisfies $\mathbb{E}S = \Sigma$. We use the notation

$$S \;=\; \sum_{i=1}^{p} \lambda_i \mathbf{u}_i \mathbf{u}_i^* \tag{2.6}$$

for the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$ and the associated normalized eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p \in \mathbb{R}^p$ of S.

We suppose that we are in the high-dimensional regime,

$$p \;\asymp\; n \,, \tag{2.7}$$

and define the dimensional ratio

$$\phi \;:=\; \frac{p}{n} \,. \tag{2.8}$$

We collect the above assumptions on our model for future reference.

Assumption 2.1. We suppose that (2.2), (2.4) and (2.7) hold.

Next, we define the Stieltjes transform of the empirical spectral measure of $n^{-1} Y^* Y$ through

$$s(z) \;:=\; \frac{1}{n} \operatorname{Tr} \left(n^{-1} Y^* Y - z\right)^{-1} \,. \tag{2.9}$$

The Stieltjes transform (2.9) contains all the information about the eigenvalues of the matrix $n^{-1} Y^* Y$, which has the same nonzero eigenvalues as S. Using that the nonzero eigenvalues of $Y^* Y$ and $Y Y^*$ coincide, we easily find

$$s(z) \;=\; \frac{1}{p} \sum_{i=1}^{p} \frac{\phi}{\lambda_i - z} - \frac{1 - \phi}{z}. \tag{2.10}$$

In order to state our precise assumptions on the spectrum of Σ, it is convenient to introduce the following deterministic function m that approximates s. Essentially, the function m is the Stieltjes transform of the asymptotic eigenvalue density of $n^{-1} Y^* Y$, which we denote by ϱ. Let $\mathbb{C}_+$ denote the complex upper-half plane.

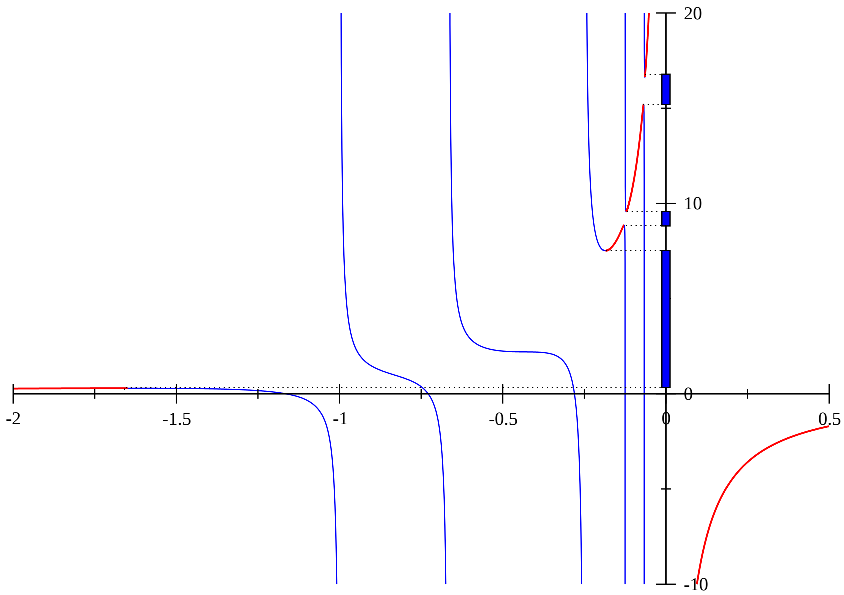


FIGURE 2.1. The function $f(x)$ associated with the population spectral measure $\frac{1}{p}\sum_{i=1}^{p}\delta_{\sigma_i} = 0.002\,\delta_{15}+0.002\,\delta_8+0.396\,\delta_3+0.3\,\delta_{1.5}+0.3\,\delta_1$. Here $p = 500$, $\phi = 0.3$ and we have three connected components. The vertical asymptotes are located at each $-\sigma^{-1}$ for $\sigma \in \{1, 1.5, 3, 8, 15\}$. The support of ϱ is indicated with thick blue lines on the vertical axis. The inverse of $m|_{\mathbb{R}\setminus\operatorname{supp}\varrho}$ is drawn in red.

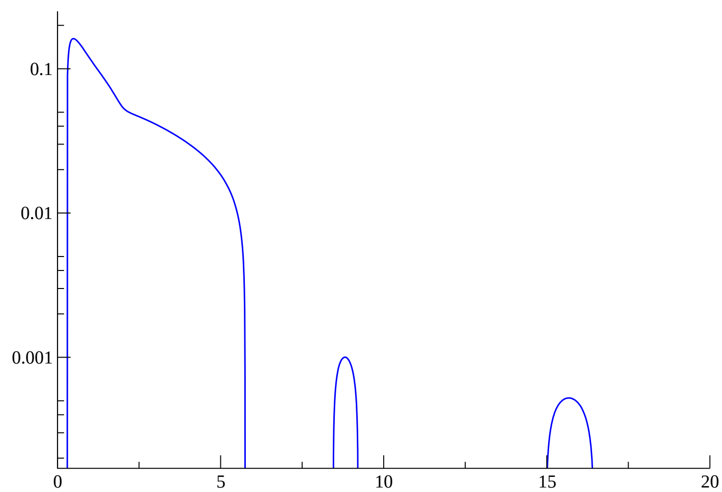


FIGURE 2.2. Asymptotic eigenvalue density ϱ of S for the example from Figure 2.1 using a logarithmic scale on the vertical axis. Both of the outliers have total mass 1, as required in Assumption 2.3 (i).

Lemma 2.2 (Asymptotic eigenvalue density)**.** *Define*

$$f(z) \;:=\; -\frac{1}{z}+\frac{1}{n}\sum_{i=1}^{p}\frac{1}{z+\sigma_i^{-1}}\,. \tag{2.11}$$

Then for each $z \in \mathbb{C}_+$, there is a unique solution $m \equiv m(z) \in \mathbb{C}_+$ of the equation

$$f(m) \;=\; z. \tag{2.12}$$

Moreover, m is the Stieltjes transform of a probability measure, denoted by ϱ:

$$m(z) \;=\; \int \frac{\varrho(\mathrm{d}x)}{x-z}\,. \tag{2.13}$$

Finally, ϱ has a continuous density supported on $[0,\infty)$.

Proof. This is a well-known result, which may be proved using a fixed point argument in $\mathbb{C}_+$. See e.g. [33, Section 5] and [1] for more details. $\square$

Since the work of Marčenko and Pastur [30], it is known that ϱ approximates the empirical spectral measure of $n^{-1}Y^*Y$ for large n. Using it, we may now state our precise assumptions on the spectrum of Σ.

Assumption 2.3. Suppose that there exists $r \in \mathbb{N}$ satisfying $r \leqslant C$ such that the following holds.

(i) The support of ϱ in $(0,\infty)$ consists of $r+1$ disjoint intervals, whose boundary points we label as $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_{2r+2} > 0$. We call the largest r intervals, $[a_{2k}, a_{2k-1}]$ for $k \in [\![r]\!]$, the *outliers* and the smallest interval, $[a_{2r+2}, a_{2r+1}]$, the *bulk*. It follows (see [23, Lemma A.1]) that $n\varrho([a_{2k}, a_{2k-1}]) \in \mathbb{N}$ for all $k \in [\![r+1]\!]$. We require that the outliers be simple, that is $n\varrho([a_{2k}, a_{2k-1}]) = 1$ for all $k \in [\![r]\!]$.

(ii) We suppose that the outliers are separated from each other and from the bulk, in the sense that for all $k \in [\![r]\!]$ we have

$$a_{2k} - a_{2k+1} \;\geqslant\; c\,. \tag{2.14}$$

(iii) We suppose that the bulk is *regular* in the sense that the density of ϱ has uniform square root decay on $[a_{2r+2}, a_{2r+1}] \cap [c,\infty)$: for all $x \in [a_{2r+2}, a_{2r+1}] \cap [c,\infty)$ we have

$$\frac{\mathrm{d}\varrho(x)}{\mathrm{d}x} \;\asymp\; \sqrt{(x-a_{2r+2}) \wedge (a_{2r+1}-x)}. \tag{2.15}$$

2.2. Rotational invariant estimators. In this subsection we introduce rotational invariant estimators and list their basic properties. We first give a precise definition of the class of RIEs.

Definition 2.4 (Rotational Invariance)**.** *(i) Denote by $\mathcal{M} \equiv \mathcal{M}_p$ the set of real symmetric nonnegative $p\times p$ matrices. For any $S \in \mathcal{M}$, define $\mathcal{M}(S)$ as the set of matrices in $\mathcal{M}$ with the same eigenvectors as S.*

(ii) A rotational invariant estimator (RIE) *is a function $\Xi \in C(\mathcal{M},\mathcal{M})$ such that for all $S \in \mathcal{M}$ and $O \in \mathrm{O}(p)$ we have*

$$O\Xi(S)O^* \;=\; \Xi(OSO^*)\,. \tag{2.16}$$

The following result is a simple and well-known (see e.g. [35] and references therein) characterization of rotational invariant estimators.

Lemma 2.5. *The function $\Xi \in C(\mathcal{M},\mathcal{M})$ is a RIE if and only if $\Xi(S) \in \mathcal{M}(S)$ for all $S \in \mathcal{M}$, and the eigenvalues of $\Xi(S)$ only depend on the eigenvalues of S.*

Our goal is to find a near-optimal RIE. In fact, we begin by determining the optimal (with respect to the Hilbert-Schmid norm) estimator Ξ subject to the condition that $\Xi(S) \in \mathcal{M}(S)$ for all $S \in \mathcal{M}$, which, by Lemma 2.5, is weaker than the condition that Ξ be a RIE.

Definition 2.6 (Oracle estimator)**.** *The* oracle estimator $\tilde{\Xi}(S)$ given the population covariance matrix Σ *is by definition the unique minimizer Ξ of the optimization problem*

$$\min\big\{\|\Xi - \Sigma\|_2 : \Xi \in \mathcal{M}(S)\big\}\,.$$

The following lemma gives the explicit form of the oracle estimator; we omit its trivial proof.

Lemma 2.7. *The oracle estimator $\tilde{\Xi}(S)$ given the population covariance matrix Σ is equal to*

$$\tilde{\Xi} \;=\; \sum_{i=1}^{p} \tilde{\xi}_i \, \mathbf{u}_i \mathbf{u}_i^* \,, \qquad \tilde{\xi}_i \;:=\; \langle \mathbf{u}_i \,, \Sigma \mathbf{u}_i \rangle \,, \tag{2.17}$$

where we used the notation (2.6)*.*

Note that the oracle estimator $\tilde{\Xi}$ satisfies $\tilde{\Xi} \in \mathcal{M}(S)$ by definition, but it is not an RIE because $\tilde{\xi}_i$ depends on the eigenvectors of S.

Remark 2.8. There exist other popular metric in statistics and each one can lead to another optimal decision theory, see e.g. [12]. We do not pursue such generalization here.

The eigenvalues $\tilde{\xi}_i$ of the oracle estimator (2.17) can be intepreted as the optimal shrinkage function that one should apply in order to clean the sample eigenvalues. However, this estimator is clearly not a RIE since each $\tilde{\xi}_i$ depends on the eigenvector $\mathbf{u}_i$. Moreover, and more importantly, this estimator is useless since it requires the a priori knowledge of the population covariance matrix Σ.

It has been shown in [26] that in the absence of outliers, $\tilde{\xi}_i$ converges to a limiting nonrandom function that does not involve the population matrix as $p \to \infty$ for any $i \in [\![p]\!]$. More precisely, let us define

$$\varphi(\lambda) \;:=\; \frac{1}{\lambda \left| \lim_{z \in \mathbb{C}^+ \to \lambda} m(z) \right|^2} \,,$$

then the Ledoit and Péché showed that ξ_i converges a.s. to $\varphi(\lambda)$ when $p \to \infty$ provided that λ corresponds to λ_i. Two remarks are in order. First, we see in this result that the spectral parameter z does not depend on p (neither n) meaning that the spectral resolution is of order 1 (*macroscopic* scale). Next, even though the result seems to depend only on the position of the sample eigenvalues, the Stieltjes transform $m(z)$ defined in (2.13) is not directly observable.

We are now ready to state the main result of the paper. Let us define the RIE

$$\hat{\Xi} \;:=\; \sum_{i=1}^{p} \hat{\xi}_i \, \mathbf{u}_i \mathbf{u}_i^* \,, \qquad \hat{\xi}_i \;:=\; \frac{1}{\lambda_i |s(\lambda_i + \mathrm{i}\eta)|^2} \tag{2.18}$$

where s is the *empirical* Stieltjes transform from (2.10). We note that the eigenvalues of (2.18) is very similar to the functional $\varphi(\lambda)$ except that the limiting Stieltjes transform is replaced by its empirical counterpart. Our result concerns the control of the convergence of the oracle eigenvalues at a *microscopic* scale that will enable us to conclude that the fully observable RIE (2.18) is close to (2.17) (in some sense to be made precise below). The following theorem establishes this result for the sample outliers, the case of the bulk eigenvalues will be tackled in a forthcoming paper as the proof involves different tools.

Theorem 2.9 (Oracle outlier eigenvalues local estimate)**.** *Suppose that Assumptions 2.1 and 2.3 hold. Suppose moreover that Y is Gaussian and let $\eta := n^{-1/2}$. Then, for all $i \in [\![r]\!]$ satisfying $\lambda_i \geqslant a_{2r+1} + c$, $\hat{\xi}_i$ is close to the oracle estimator $\tilde{\xi}_i$ in the following microscopic sense: for every fixed $\varepsilon, D > 0$ we have*

$$|\hat{\xi}_i - \tilde{\xi}_i| \;\leqslant\; n^{-1/2+\varepsilon} \tag{2.19}$$

with probability at least $1 - n^{-D}$ and large enough n (depending on ε and D).

A striking observation of (2.18) is that the formula is indifferent whether the sample eigenvalue is an outlier or not. This means that an observable and unified cleaning recipe for general covariance matrices is possible [7].

Remark 2.10 (Generalizations)**.** We stress that the Gaussian assumption on Y in Theorem 2.9 is made for convenience and to keep the presentation readable. In fact, we can extend our result to general entries that possess a sufficient number of bounded moments (see [23, Section 2.] for the precise condition). Moreover, we could also relax the condition (2.7) to $\log n \asymp \log p$ or consider several bulk and general outlier components like in [3]. We will not address these issues in this paper and refer the interested readers to [3, 23] for more details.

Applications

3.1. Position and monotonicity of the outliers of $\hat{\Xi}$. The understanding of the RIE (2.18) is very important in practice in order to understand the impact of the noise. In that regards, it has been noticed in [6] that in the absence of outliers, the limiting spectral density (LSD) of $\hat{\Xi}$ turns out to be narrower than the one of Σ which is itself narrower than the LSD of S. In other words, we should bring upward (downward) small (large) eigenvalues even more toward unity than their "true" locations. For outliers, a similar observation is obtained from Theorem 2.9. In addition, we can also show that the monotonicity of the outliers is preserved in the high-dimensional regime (2.7) which is an important feature whether we do not have prior views on the components of the covariance matrix.

Corollary 3.1. *Suppose that Assumptions 2.1 and 2.3 hold. Then, for all i satisfying $\lambda_i \geqslant a_{2r+1} + c$ and every fixed $\varepsilon, D > 0$, we have with probability at least $1 - n^{-D}$ and large enough n (depending on ε and D) the estimate*

$$\left| \tilde{\xi}_i - \frac{\sigma_i^2}{\lambda_i} \right| \leqslant n^{-1/2+\varepsilon} \tag{3.1}$$

that is monotone increasing as a function of λ_i.

We deduce from (3.1) that within Assumption 2.1, we have $c \leqslant \hat{\xi}_i \leqslant \sigma_i$ for any $i \in [\![r]\!]$ with high probability. This emphasizes that cleaning eigenvalues is *different* from moving the sample outliers to their true location. More generally, Corollary 3.1 can be understood as the fact that we shall also clean the outlier sample eigenvalues since we are stuck with a noisy version of the population eigenbasis.

3.2. Analytical examples and phase transition. Since the Marčenko-Pastur equation (2.12) cannot be solved explicitly for an arbitrary population covariance matrix Σ, analytical results are scarce even in the limit p and n both grow to infinity with the same rate. We provide in this section two insightful applications where the limiting value of (2.18) can be characterized exactly. These applications allow us to exhibit a phase transition phenomena for (2.18) between bulk and outliers eigenvalues even if the cleaning function stays identical.

3.2.1. Spiked covariance matrix. The most simple application is the case where Σ belongs to the spiked covariance matrix ensemble [20], i.e. its spectrum is made of an eigenvalue at unity with multiplicity $p-r$ and r distincts spikes. This model has a long history in high-dimensional statistics, especially to detect the number of relevant principal component in PCA [20, 24] or for covariance matrices estimation [12, 25, 32] (see Section 3.3 below).

Proposition 3.2 (Limiting optimal RIE for spiked covariance matrices)**.** *Let $\Sigma = I_p$ and suppose that Assumptions 2.1 and 2.3 hold. Let $n \to \infty$ with $\phi \in (0, \infty)$ and define the deterministic function*

$$\varphi_s(\lambda) \;:=\; \begin{cases} 1 & \text{if } \lambda \in [(1-\sqrt{\phi})^2, (1+\sqrt{\phi})^2] \\ \sigma_s^2(\lambda)/\lambda & \text{otherwise,} \end{cases} \tag{3.2}$$

with

$$\sigma_s(\lambda) \;:=\; \frac{1}{2}\left[(\lambda + 1 - \phi) + \sqrt{(\lambda + \phi - 1)^2 - 4\lambda\phi} \right]. \tag{3.3}$$

Then, the eigenvalues of (2.18) converges a.s. to $\varphi_s(\lambda)$ provided that λ corresponds to λ_i.

Proof. See Appendix B. □

All the eigenvalues that lie in the Marčenko-Pastur density are shrunk to unity but the main difference is that we also shrink the eigenvalues associated to the principal components to a value that is different from 1. This is because Theorem 2.9 takes into account that we use the sample eigenvectors.

3.2.2. Linear shrinkage. We consider another classical model in multivariate statistics, which will provide a genuine link between Theorem 2.9 and Bayesian statistics. Let us assume that the population covariance matrix Σ is an isotropic Inverse-Wishart matrix, or equivalently said, Σ^{-1} is an isotropic Wishart matrix. The Inverse Wishart distribution is well-known in Bayesian statistics for being the *conjugate prior* for the covariance matrix of a multivariate Gaussian distribution [18]. A remarkable result is that we can compute exactly the optimal Bayes estimator (with respect to the Hilbert-Schmidt norm) and this gives the so-called *linear shrinkage* estimator (1.3) of [18] presented in the introduction.

We may recover (1.3) using Theorem 2.9 and generalize it to the case where Σ contains a finite number of outliers. Indeed, following the parametrization of [6, Section IV.B] for the isotropic Inverse-Wishart matrix, one has

$$m(z) = -\frac{(1-\phi)}{z} + \phi\,\frac{\kappa(1-\phi) - z(\kappa+1) + \sqrt{(z-\gamma_+^{\mathrm{iw}})(z-\gamma_-^{\mathrm{iw}})}}{,} \qquad \kappa > 0, \tag{3.4}$$

where $\gamma_\pm^{\mathrm{iw}}$

$$\gamma_\pm^{\mathrm{iw}} = \frac{1}{\kappa}\Big[\kappa(1+\phi)+1 \pm \sqrt{(2\kappa+1)(2\phi\kappa+1)}\Big] \tag{3.5}$$

denotes the edges of the support of ϱ.

Proposition 3.3 (Limiting optimal RIE for Inverse Wishart matrices)**.** *Suppose that Assumptions 2.1 and 2.3 hold. Set Σ as an isotropic Inverse Wishart matrix parametrized as in (3.4) with $\sigma_p > 0$ and $\kappa > 0$. Let $n \to \infty$ with $\phi \in (0,\infty)$ and define the deterministic function*

$$\varphi_{iw}(\lambda) \;:=\; \begin{cases} 1+\alpha(\lambda-1) & \text{if } \lambda \in [\gamma_-^{iw}, \gamma_+^{iw}] \\ \sigma_{iw}^2(\lambda)/\lambda & \text{otherwise}, \end{cases} \tag{3.6}$$

where the shrinkage parameter α and the function $\sigma_{iw}(z)$ are given by

$$\begin{aligned} \alpha &:= \frac{1}{1+2\phi\kappa} \in [0,1], \\ \sigma_{iw}(z) &:= \frac{z(1-\phi\kappa) + \phi\kappa(1-\phi) + \phi\sqrt{z^2\kappa^2 - 2\kappa z(1+\kappa(1+\phi)) + \kappa^2(1-\phi)^2}}{1+2\phi\kappa}. \end{aligned}$$

Then, the eigenvalues of (2.18) converges a.s. to $\varphi_{iw}(\lambda)$ provided that λ corresponds to λ_i.

Proof. See Appendix B. ☐

Proposition 3.3 yields a simple way to exhibit that the linear shrinkage estimator (1.3) is not optimal in the presence of spikes. We give an illustration of this in Figure 3.1 with $p = 500$, $n = 1000$, $\kappa = 0.5$. The outliers are located at $\sigma \in \{8, 10, 12\}$. The result of Figure 3.1 comes from a single realization of S using a multivariate Gaussian distribution and the result is very satisfactory. We compare our estimate (2.18) with the linear shrinkage (1.3) and we see that the latter estimator fails for outliers.

Note that in practice, one can consistently estimate the shrinkage parameter α defined in (3.7) by using the method of [27].

3.3. Simulations. In this section, we shall consider other examples where the limiting value cannot be reached explicitly. This will allows us to show the robustness of Theorem 2.9. Note that we will also test the accuracy of the estimator (2.17) for the bulk eigenvalues with the same parameter $\eta = n^{-1/2}$ even if it is not proved in Theorem 2.9. Here are the four cases that we shall consider for the spectrum of Σ:

(i) Diagonal matrix composed of multiple sources (as in Figure 2.1) with spikes located at {8, 15};

(ii) Deformed GOE (as in Figure 1.1) with spikes located at {3, 3.5,4.5,6};

(iii) Toeplitz matrix with entries $\Sigma_{ij} = 0.6^{|i-j|}$ with spikes located at {7, 8, 10, 11};

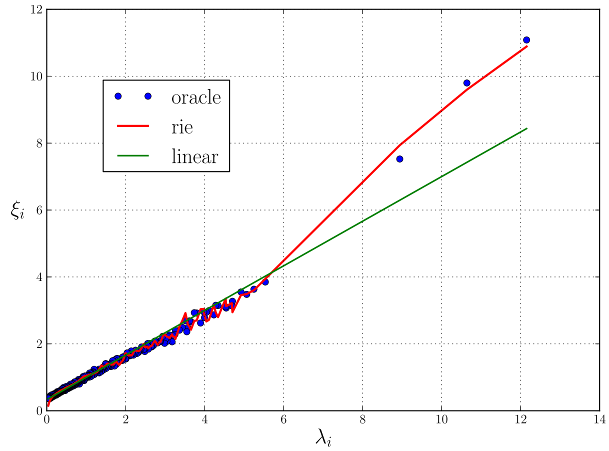


FIGURE 3.1. Comparison of our estimator (2.18) (red line) with the oracle estimator (2.17) (blue points) with Σ an isotropic Inverse Wishart with $p = 500$, $n = 1000$, $\kappa = 0.5$. The outliers are located at $\sigma \in \{8, 10, 12\}$. We also plot the linear shrinkage estimator (1.3) (green line).

(iv) Power-law distributed [4] with $\alpha = 0.35$ with

$$\sigma_i = 2\alpha - 1 + (1-\alpha)\sqrt{\frac{p}{i}} \qquad i \in [\![p]\!]. \tag{3.7}$$

Note that the power law distribution automatically generates a bounded number of outliers. Moreover, since we work with p and n bounded, we fortunately have $\|\Sigma\| \leqslant C$ for the case (iv) so that it is in adequation with Assumption (2.1). Note that the case (iv) and has found interesting applications in Finance [4]. For each case, the sample covariance matrix is generated using a multivariate Gaussian distribution. We plot the results in Figure 3.2 with $p = 500$ and $n = 1000$.

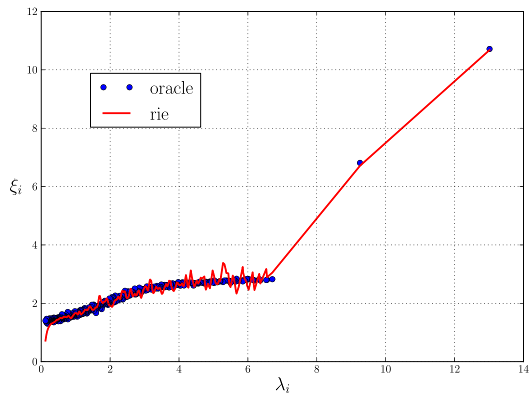
Overall, we see that the estimator (2.18) gives accurate predictions of either for the bulk eigenvalues or for outliers. We have considered several configurations of outliers. For the case (i), we see that the two isolated outliers are correctly estimated. For the deformed GOE or the Toeplitz case, the outliers are a little bit closer to each other and again, the predictions agrees with the oracle estimator. For the more complex case of a power law distributed spectrum, where there is no right edge, we see that (2.18) is again in balance with the oracle estimator. We nevertheless notice that the small eigenvalues (left edge) are underestimated compared to the oracle estimator.

We are now interested in the improvement that we get by using (2.18) a function of p with a fixed $\phi = 0.5$. To that end, we introduce the *relative percentage imporovement* (RPI) ratio,

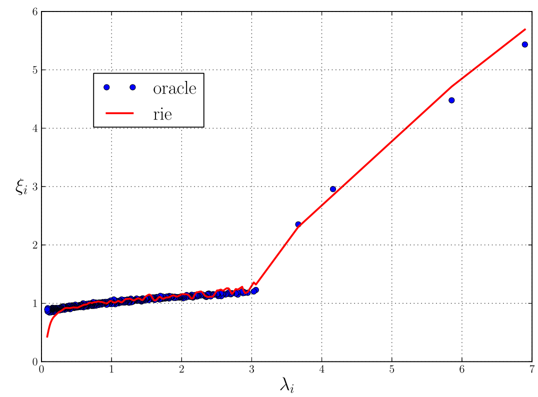
$$\mathrm{RPI}(\Xi) \;:=\; 100 \times \left(1 - \frac{\mathbb{E}\|\hat{\Xi} - \tilde{\Xi}\|_2^2}{\mathbb{E}\|\Xi - \tilde{\Xi}\|_2^2}\right), \tag{3.8}$$

for Ξ a given estimator of Σ. We shall consider the performance of $\hat{\Xi}$ with two different estimators Ξ: (i) sample covariance matrix S and (ii) eigenvalues substitution [14]. The first case will indicate what we gain by cleaning the eigenvalues. The second estimator is more subtle. Roughly, the substitution procedure replaces the sample eigenvalue by their true value, i.e. $\Xi^{\mathrm{sub}} := \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{u}_i^*$. This procedure is not trivial to implement in practice but here it is easy to test it since we know exactly Σ. Hence, this shall exhibit if we indeed improve our estimation by shrinking the eigenvalues even more toward unity than their true locations.

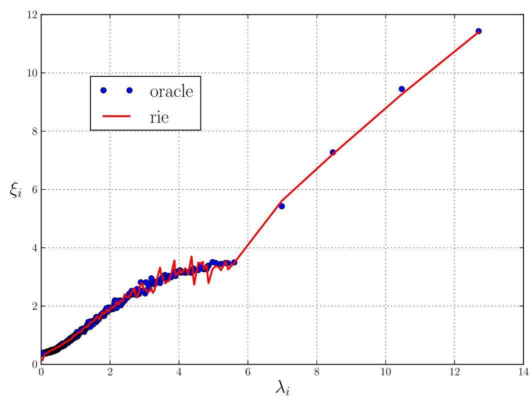
We reconsider the 4 examples above and we take average over 100 realizations of S for all $p \in \{50, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$. The result are given in Figure 3.3. For the sample covari-



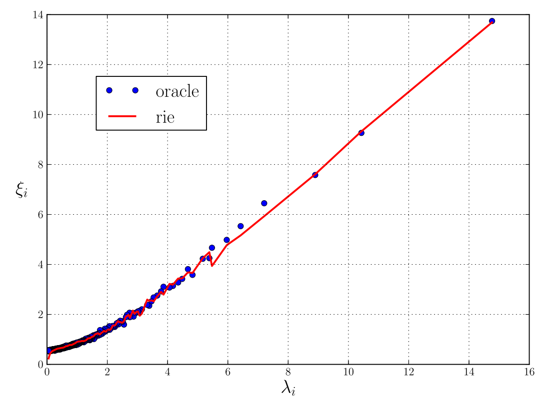
(A) Multiple sources (case (i)).



(B) deformed GOE (case (ii)).



(C) Toeplitz (case (iii))



(D) Power law (case (iv))

FIGURE 3.2. Comparison of our estimator (2.18) (red line) with the oracle estimator (2.17) (blue points) for the four cases presented at the beginning of Section 3.3 with $p = 500$ and $n = 1000$. The results come from a single realization of S.

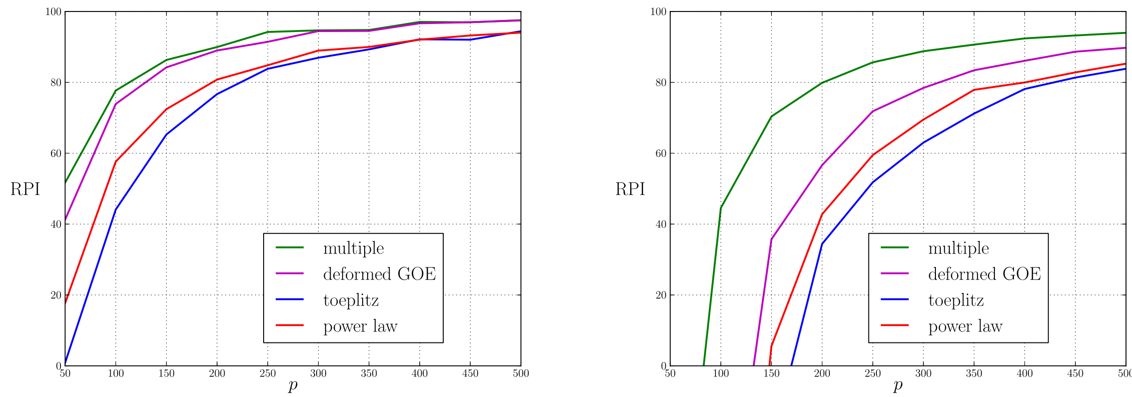
ance matrix, we see that we already get an improvement of 50% for all the cases by using Theorem 2.9 with $p = 100$ and $n = 200$. Hence, it shows that it is always better to apply (2.18) even for relatively small value of p and n. For $p = 500$, we reach a plateau of 90% of improvement for all the cases. Now, if we look at the RPI compared to the substitution estimator, using our RIE (2.18) becomes very relevant for any $p \geqslant 200$. Therefore, even if we know *exactly* the population eigenvalues $[\sigma_i]_{i\in[\![N]\!]}$, it is not optimal in the high-dimensional framework to use them to estimate Σ whether we use the sample eigenvectors. We again emphasize that our estimator depends only on the data we have with no hyperparameter to estimate.

Proof of Theorem 2.9

The rest of the paper is devoted to the proof of Theorem 2.9. To simplify the following arguments, we assume without loss of generality that

$$\sigma_p > 0. \tag{4.1}$$

Indeed, by a simple continuity argument for the limit $\sigma_p \to 0$, whose details we omit, our results remain true for $\sigma_p = 0$. The assumption (4.1) is purely qualitative, and needed for Σ^{-1} to make sense in the definition of G in (4.19) below.



(A) RPI compared to the sample estimate (no cleaning). (B) RPI compared to the substitution procedure.

FIGURE 3.3. Relative percentage improvement of our estimator (2.18) comapred to the sample covariance matrix (Figure 3.3a) and the substitution procedure (Figure 3.3b) as function of p with $\phi = 0.5$. The green line corresponds to the multiple sources (case (i)), the purple one to the deformed GOE (case (ii)), the blue one to Toeplitz (case (iii)) and the red one to the power law (case (iv)). The average has been taken over 100 realizations of S for each value of p.

This section is devoted to the proof of (2.19) for the outliers, i.e. for $i \in [\![r]\!]$. For the outliers, we do not need to assume Assumption 2.3 (iii). The following summarizes the assumptions of this section.

Assumption 4.1. We suppose that (4.1), Assumption 2.1, and Assumption 2.3 hold.

Throughout the proofs we frequently omit the spectral parameter z from our notation if this does not lead to confusion.

4.1. Tools and notations. In this section we introduce the basic tools and notations that we shall use in the proofs. Since we assume that Y is Gaussian, we may and shall always assume by invariance of the law of Y under the map $Y \mapsto OY$ for $O \in \mathrm{O}(p)$ that

$$\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_p), \tag{4.2}$$

so that $\mathbf{v}_i$ is the i-th standard basis vector of $\mathbb{R}^p$.

We rewrite $Y = n^{1/2}\Sigma^{1/2}X$, where $X = (X_{i\mu}) \in \mathbb{R}^{p\times n}$ has centered independent Gaussian entries satisfying

$$\mathbb{E}X_{i\mu}^2 = \frac{1}{n}. \tag{4.3}$$

We henceforth write

$$S \;=\; \Sigma^{1/2} X X^* \Sigma^{1/2}\,. \tag{4.4}$$

Throughout the following, the probability space is given by $\{X \in \mathbb{R}^{p\times n}\}$ with probability measure given by the law of X.

An important tool of our proof is the following *spikeless* population covariance matrix.

Definition 4.2 (Spikeless covariance matrix)**.** *We define the* spikeless population covariance matrix

$$\tilde{\Sigma} \;:=\; \sum_{i=1}^{p} \tilde{\sigma}_i \mathbf{v}_i \mathbf{v}_i^*\,, \qquad \tilde{\sigma}_i \;:=\; \begin{cases} \sigma_{r+1} & \text{if } i \leqslant r \\ \sigma_i & \text{if } i \geqslant r+1\,. \end{cases} \tag{4.5}$$

The choice of $\tilde{\sigma}_i = \sigma_{r+1}$ for any $i \in [\![r]\!]$ is made for definiteness but any other value σ_j with $j > r$ would do equally well (see Lemma C.2 for details). From (4.1), we have $\tilde{\Sigma} > 0$.

It is often convenient to parametrize the spikes σ_i using $\tilde{\sigma}_i$. To that end, for $i \in [\![r]\!]$ we introduce d_i satisfying

$$\sigma_i \;=\; \sigma_{r+1}(1+d_i)\,. \tag{4.6}$$

Since $\|\tilde{\Sigma}\| = \sigma_{r+1}$ by definition, we infer from Lemma C.1 and (2.2) that a spike is characterized by

$$d_i \asymp 1, \tag{4.7}$$

for any $i \in [\![r]\!]$ meaning that $D := \operatorname{diag}(d_1, \dots, d_r)$ is an invertible diagonal $r \times r$ matrix. Then, defining the $p \times r$ matrix $V := [\mathbf{v}_1, \dots, \mathbf{v}_r]$, we therefore have

$$\Sigma \;=\; \tilde{\Sigma}(I + VDV^*)\,. \tag{4.8}$$

Definition 4.3. *The sample covariance matrix associated with $\tilde{\Sigma}$ is*

$$\tilde{S} \;:=\; \tilde{\Sigma}^{1/2} X X^* \tilde{\Sigma}^{1/2}\,. \tag{4.9}$$

We denote the eigenvalues of $\tilde{S}$ by $\tilde{\lambda}_1 \geqslant \tilde{\lambda}_2 \geqslant \cdots \geqslant \tilde{\lambda}_p \geqslant 0$. Moreover, we denote by $\tilde{m}$, $\tilde{s}$, and $\tilde{\varrho}$ the quantities obtained from m, s, and ϱ by replacing Σ with $\tilde{\Sigma}$.

Throughout the following, we use the following notion of high-probability bound, which was introduced in [16].

Definition 4.4 (Stochastic domination)**.** *Let*

$$\xi \;=\; \left(\xi^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\right), \qquad \zeta \;=\; \left(\zeta^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\right)$$

be two families of nonnegative random variables, where $U^{(N)}$ is a possibly N-dependent parameter set.

(i) We say that ξ is stochastically dominated by ζ, uniformly in u, *if for all (small) $\varepsilon > 0$ and (large) $D > 0$ we have*

$$\sup_{u \in U^{(N)}} \mathbb{P}\Big[\xi^{(N)}(u) > n^{\varepsilon}\zeta^{(N)}(u)\Big] \;\leqslant\; n^{-D}$$

for large enough $n \geqslant n_0(\varepsilon, D)$. Throughout this paper the stochastic domination will always be uniform in all parameters (such as matrix indices, deterministic vectors, and spectral parameters z) that are not explicitly fixed. Note that $n_0(\varepsilon, D)$ may depend on quantities that are explicitly constant.

(ii) If ξ is stochastically dominated by ζ, uniformly in u, we use the notation $\xi \prec \zeta$. Moreover, if for some complex family ξ we have $|\xi| \prec \zeta$ we also write $\xi = O_\prec(\zeta)$.

(iii) We say that an event Ξ holds with high probability *if $\mathbf{1}(\Xi^c) \prec 0$.*

We now derive an upper bound on $\tilde{\lambda}_1$. To that end, we define $\tilde{\gamma}_+ := \sup(\operatorname{supp}(\tilde{\varrho}))$, the asymptotic right edge of the bulk spectrum.

Lemma 4.5. *Suppose that Assumption 4.1 holds. Then $\gamma_+ \asymp 1$.*

Proof. See Appendix C.2. □

Lemma 4.6. *Fix $\varepsilon > 0$ and suppose that Assumption 4.1 holds. Then $\tilde{\lambda}_1 \leqslant \tilde{\gamma}_+ + \varepsilon$ with high probability.*

Proof. The proof follows very closely that of [23, Lemma 10.1], with trivial modifications whose details we omit. □

We give some properties of $\tilde{m}(z)$ outside the spectrum of $\operatorname{supp} \tilde{\varrho}$.

Lemma 4.7. *Fix $\tau > 0$ and suppose that Assumption 4.1 holds. Then, we have for all $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$*

(i) $|\tilde{m}(z)| \asymp 1$ *and* $\tilde{m}'(z) \asymp 1$,

(ii) $\min_{k \in [\![p]\!]}[1 + \tilde{m}(z)\tilde{\sigma}_k] \asymp 1$.

Proof. See Appendix C.3. $\square$

Next, we define the function

$$w(z) \;:=\; 1 - \frac{1}{1 + \tilde{m}(z)\sigma_{r+1}}\,, \tag{4.10}$$

for which we have the following estimates.

Lemma 4.8. *Fix $\tau > 0$ and suppose that Assumption 4.1 holds. Then, for any $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$, we have*

$$|w(z)| \asymp 1, \qquad w'(z) \asymp 1. \tag{4.11}$$

Proof. See Appendix C.4. $\square$

In the following, it will be convenient to work with the functional inverse of $w(z)$ outside the spectrum of $\operatorname{supp} \tilde{\varrho}$. For this aim, we have the following result.

Lemma 4.9. *Suppose that Assumption 4.1 holds. Then, $w(z)$ is increasing, bounded and negative from $[\tilde{\gamma}_+, \infty)$ onto $[w(\tilde{\gamma}_+), 0)$.*

Proof. See Appendix C.5. $\square$

Next, thanks to Lemma 4.9, we may define for $\zeta \in [-1/w(\tilde{\gamma}_+), \infty)$ the function

$$\theta(\zeta) \;:=\; w^{-1}(-1/\zeta) \;=\; \tilde{f}\left(\frac{-1}{\sigma_{r+1}(1+\zeta)}\right), \tag{4.12}$$

with $-1/w(\tilde{\gamma}_+) > 0$ and $\tilde{f}$ defined as in (2.11) with Σ replaced by $\tilde{\Sigma}$. Furthermore, an application of the chain rule with $\zeta \geqslant -1/w(\tilde{\gamma}_+)$ and using Lemma 4.9 yields

$$\theta'(\zeta) = \frac{1}{w'(\theta(\zeta))\zeta^2} > 0\,, \tag{4.13}$$

so that θ is positive and increasing from $[-1/w(\tilde{\gamma}_+), \infty)$ onto $[\tilde{\gamma}_+, \infty)$.

Lemma 4.10. *Suppose that Assumption 4.1 holds. Then, for any fixed $\varepsilon > 0$ we have for all*

$$\zeta \in \left[-1/w(\tilde{\gamma}_+) + \varepsilon, \varepsilon^{-1}\right]$$

that

$$\theta(\zeta) \asymp 1, \quad \theta'(\zeta) \asymp 1. \tag{4.14}$$

Proof. See Appendix C.6. $\square$

We shall perform the bulk of the proof by working in the variable $\zeta = \theta^{-1}(z)$ instead of the variable $z = \theta(\zeta)$. The following result gives the basic separation bounds of the d_i's that lie in the ζ-plane. Its proof is given in Appendix C.8.

Lemma 4.11 (Outlier separation)**.** *Suppose that Assumption 4.1 holds. There exists a constant $c > 0$ such that*

(i) $d_i - (-1/w(\tilde{\gamma}_+)) \geqslant c$ *for all* $i \in [\![r]\!]$,

(ii) $|d_i - d_j| \geqslant c$ *for all distinct* $i, j \in [\![r]\!]$.

One notices from (4.12) and Lemma 4.11 that for any $i \in [\![r]\!]$, we have the identities

$$\theta(d_i) = \tilde{f}(-\sigma_i^{-1}), \qquad \theta'(d_i) = \frac{\sigma_{r+1}}{\sigma_i^2} \tilde{f}'(-\sigma_i^{-1}). \tag{4.15}$$

Moreover, using Lemma 4.11 and (C.33), there exists a constant $\tau > 0$ such that

$$\theta(d_i) \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]. \tag{4.16}$$

We conclude this subsection by introducing resolvents, which are the main tool in the proof of Proposition 4.23. Denote by

$$R(z) \;:=\; (S - z)^{-1}, \qquad \tilde{R}(z) \;:=\; (\tilde{S} - z)^{-1} \tag{4.17}$$

the resolvents of S and $\tilde{S}$. The following result, taken from [3, lemma 3.11], gives the relationship between $R(z)$ and $\tilde{R}(z)$.

Lemma 4.12. *We have*

$$V^* R(z) V \;=\; \frac{1}{z}\left[D^{-1} - \frac{\sqrt{I+D}}{D}\left(D^{-1} + I + zV^*\tilde{R}(z)V\right)^{-1}\frac{\sqrt{I+D}}{D}\right]. \tag{4.18}$$

Next, we introduce a linearizing block matrix from [23] that is a convenient tool in the analysis of $R(z)$ and $\tilde{R}(z)$. The following summarizes the relevant definitions from [23, Section 3.1 and 4].

Definition 4.13 (Matrix multiplication)**.** *We use matrices of the form $A = (A_{st} : s \in l(A), t \in r(A))$, whose entries are indexed by arbitrary finite subsets of $l(A), r(A) \subset \mathbb{N}$. Matrix multiplication AB is defined for $s \in l(A)$ and $t \in r(B)$ by*

$$(AB)_{st} \;:=\; \sum_{q \in r(A) \cap l(B)} A_{sq} B_{qt}\,.$$

We also need to define the sets of indices.

Definition 4.14 (Index sets)**.** *We introduce the index sets*

$$\mathcal{I}_p \;:=\; [\![1,p]\!]\,, \qquad \mathcal{I}_n \;:=\; [\![p+1,p+n]\!]\,, \qquad \mathcal{I} \;:=\; \mathcal{I}_p \cup \mathcal{I}_n \;=\; [\![1,p+n]\!]\,.$$

We consistently use the letters $i, j \in \mathcal{I}_p$, $\mu, \nu \in \mathcal{I}_n$, and $s, t \in \mathcal{I}$. We label the indices of the matrices according to

$$X \;=\; (X_{i\mu} : i \in \mathcal{I}_p, \mu \in \mathcal{I}_n)\,, \qquad \Sigma \;=\; (\Sigma_{ij} : i, j \in \mathcal{I}_p)\,.$$

Definition 4.15 (Linearizing block matrix)**.** *Under the condition (4.1), we define the $\mathcal{I} \times \mathcal{I}$ matrices*

$$H(z) \;:=\; \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -z \end{pmatrix}, \qquad G(z) \;:=\; H(z)^{-1}\,. \tag{4.19}$$

Definition 4.16 (Minors)**.** *For $T \subset \mathcal{I}$ we define the minor $H^{(T)} := (H_{st} : s, t \in \mathcal{I} \setminus T)$. We also write $G^{(T)} := (H^{(T)})^{-1}$. We abbreviate $(\{s\}) \equiv (s)$ and $(\{s,t\}) \equiv (st)$.*

Lemma 4.17 (Resolvent identities)**.** *(i) For $i, j \in \mathcal{I}_p$ and $\mu, \nu \in \mathcal{I}_n$ we have*

$$G_{ij} = (z\Sigma^{1/2} R\, \Sigma^{1/2})_{ij}\,, \qquad G_{\mu\nu} \;=\; \left((X^*\Sigma X - z)^{-1}\right)_{\mu\nu}. \tag{4.20}$$

(ii) For $u \in \mathcal{I}$ and $s, t \in \mathcal{I} \backslash \{u\}$ we have

$$G_{st}^{(u)} = G_{st} - \frac{G_{su}G_{ut}}{G_{uu}} \tag{4.21}$$

Proof. See [23, Lemma 4.4] and the references therein. ☐

The linearizing block matrix $\tilde{H}$, $\tilde{G}$ and its minors $\tilde{G}^{(T)}$ are defined in the same way as H, G and $G^{(T)}$, with Σ replaced by $\tilde{\Sigma}$.

Finally, we state the *anisotropic local law*, which compares $\tilde{G}$ to its *deterministic equivalent*.

Definition 4.18 (Deterministic equivalent of $\tilde{G}$)**.** *Define the $\mathcal{I} \times \mathcal{I}$ deterministic matrix*

$$\tilde{\Pi}(z) := \begin{pmatrix} -\tilde{\Sigma}(I + \tilde{m}(z)\tilde{\Sigma})^{-1} & 0 \\ 0 & \tilde{m}(z)I \end{pmatrix}. \tag{4.22}$$

Also, we extend $\tilde{\Sigma}$ to an $\mathcal{I} \times \mathcal{I}$ matrix through

$$\underline{\tilde{\Sigma}} := \begin{pmatrix} \tilde{\Sigma} & 0 \\ 0 & I \end{pmatrix}. \tag{4.23}$$

Throughout the following we use the spectral domain

$$\mathbf{D}_o := [\tilde{\gamma}_+ + \tau, \tau^{-1}] \times [-\tau^{-1}, \tau^{-1}]$$

for some fixed $\tau > 0$.

Lemma 4.19 (Anistropic local law outside of the spectrum of $\tilde{S}$)**.** *Fix $\varepsilon, \tau > 0$ and suppose that Assumption 4.1 holds. Then for all deterministic vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^p$, we have with high probability that*

$$\left\langle \mathbf{v}, \underline{\tilde{\Sigma}}^{-1}(\tilde{G}(z) - \tilde{\Pi}(z))\underline{\tilde{\Sigma}}^{-1}\mathbf{w} \right\rangle = O\big(|\mathbf{v}||\mathbf{w}|n^{-1/2+\varepsilon}\big) \tag{4.24}$$

for all $z \in \mathbf{D}_o$.

Proof. Theorem 4.19 follows from [23, Theorem 3.16 (i)] and Lemma 4.6. More precisely, the *simultaneous* bound for all $z \in \mathbf{D}_o$ follows from the pointwise bound in [23, Theorem 3.16 (i)] and the fact that, by Lemma 4.6, the left-hand side of (4.24) is with high probability C-Lipschitz for some constant C depending on τ. $\square$

The following result gives high-probability bounds on the locations of the outliers, and is proved in Appendix C.9.

Lemma 4.20 (Outlier locations)**.** *Suppose that Assumption 4.1 holds. Then for all $i \in [\![r]\!]$ we have*

$$|\lambda_i - \theta(d_i)| \prec n^{-1/2}. \tag{4.25}$$

We end this section with some estimates of $\tilde{m}(z)$ for $z \in \mathbf{D}_o$, whose easy proofs we omit.

Lemma 4.21. *Fix $\tau > 0$ and suppose that Assumption 4.1 holds. Then for any $z \in \mathbf{D}_o$, we have*

$$|\tilde{m}(z)| \asymp 1, \qquad |\tilde{m}'(z)| \asymp 1, \qquad |\tilde{m}''(z)| \leqslant C. \tag{4.26}$$

Using Lemma 4.21, we may extend θ away from the real line as a holomorphic function.

Lemma 4.22. *Suppose that Assumption 4.1 holds. Then, for any fixed $\varepsilon > 0$ the function θ extends to a holomorphic function in the domain*

$$\left[-1/w(\tilde{\gamma}_+) + \varepsilon, \varepsilon^{-1}\right] \times [-\varepsilon^{-1}, \varepsilon^{-1}],$$

where it satisfies $|\theta''(\zeta)| \leqslant C$.

Proof. See Appendix C.7. $\square$

4.2. The main estimate. In this subsection we state the key estimate behind the proof of (2.19) for $i \in [\![r]\!]$ – Proposition 4.23 below. For $i \in [\![r]\!]$, we split the oracle eigenvalue $\tilde{\xi}_i$ from (2.17) as

$$\tilde{\xi}_i \;=\; \alpha_i + \beta_i\,, \qquad \alpha_i \;:=\; \sum_{j=1}^{r} \sigma_j \langle \mathbf{u}_i\,, \mathbf{v}_j\rangle^2, \qquad \beta_i \;:=\; \sum_{j=r+1}^{p} \sigma_j \langle \mathbf{u}_i\,, \mathbf{v}_j\rangle^2\,. \tag{4.27}$$

The main work of this section is to prove Proposition 4.23 below, which characterizes the asymptotic behaviour of α_i and β_i. The remainder of this section is divided into three subsections. Sections 4.3 and 4.4 contain the proof of Proposition 4.23; they contain the estimates of α_i and β_i respectively, which are very different. Finally, in Section 4.8, using Proposition 4.23, we conclude the proof of (2.19) for $i \in [\![r]\!]$.

The main work in the proof of (2.19) for $i \in [\![r]\!]$ is the following result, from which (2.19) will follow easily (see Section 4.8 below).

Proposition 4.23. *Suppose that Assumption 4.1 holds. Suppose that Y is Gaussian. Then for all $i \in [\![r]\!]$ we have*

$$\left|\alpha_i - \frac{\tilde{f}'(-1/\sigma_i)}{\tilde{f}(-1/\sigma_i)}\right| \;\prec\; n^{-1/2} \tag{4.28}$$

and

$$\left|\beta_i - \frac{\sigma_i^2 - \tilde{f}'(-1/\sigma_i)}{\tilde{f}(-1/\sigma_i)}\right| \;\prec\; n^{-1/2}\,. \tag{4.29}$$

4.3. Proof of (4.28). The estimate (4.28) follows immediately from (2.2), (2.4), (4.15), Lemma 4.10 and the following lemma, which gives an estimate for the overlap between any outlier eigenvectors of S and Σ.

Lemma 4.24. *Under the assumptions of Proposition 4.23 we have for all $i, j \in [\![r]\!]$*

$$\left|\langle \mathbf{u}_i\,, \mathbf{v}_j\rangle^2 - \frac{\delta_{ij}}{\sigma_i}\frac{\tilde{f}'\,(-1/\sigma_i)}{\tilde{f}(-1/\sigma_i)}\right| \;\prec\; n^{-1/2}\,. \tag{4.30}$$

The rest of this subsection is devoted to the proof of Lemma 4.24, which is similar to that of [3, Theorem 2.3]. Fix $\varepsilon > 0$ and define

$$W(z) \;:=\; I + zV^*\tilde{R}(z)V\,, \tag{4.31}$$

and the event

$$\Omega \;\equiv\; \Omega_n(\varepsilon) \;:=\; \Big\{\|w(z)I - W(z)\| \leqslant n^{-1/2+\varepsilon} \text{ for all } z \in \mathbf{D}_o\Big\} \\ \cap \Big\{\tilde{\lambda}_1 \leqslant \tilde{\gamma}_+ + \varepsilon\Big\} \cap \Big\{|\lambda_i - \theta(d_i)| \leqslant n^{-1/2+\varepsilon} \text{ for all } i \in [\![r]\!]\Big\}\,. \tag{4.32}$$

We claim that Ω holds with high probability for any fixed $\varepsilon > 0$. Indeed, the first event in the definition of Ω has high probability by Theorem 4.19 and (4.20). The other two events in the definition of Ω have high probability by Lemmas 4.6 and 4.20 respectively. For the rest of this subsection, we fix a realization $X \in \Omega$. In particular, the following argument is entirely deterministic.

Our starting point is a contour representation of the overlap $\langle \mathbf{u}_i\,, \mathbf{v}_j\rangle^2$ for any $i, j \in [\![r]\!]$. To that end, we fix $\rho > 0$ and define for $i \in [\![r]\!]$

$$\Upsilon_i \;:=\; \overline{B_\rho(d_i)}\,, \qquad \Gamma_i \;:=\; \partial\Upsilon_i\,, \tag{4.33}$$

where $B_\rho(d_i)$ is the open disc of radius ρ around d_i, and regard Γ_i as a positively oriented contour. From Lemma 4.11 we easily deduce the following result.

Lemma 4.25. *Suppose that Assumption 4.1 holds. Then for small enough fixed $\rho > 0$ the set Υ_i contains d_j for $i, j \in [\![r]\!]$ if and only if $j = i$.*

For the following we abbreviate $\Upsilon := \bigcup_{i\in[\![r]\!]} \Upsilon_i$.

Lemma 4.26. *Suppose that Assumption 4.1 holds. Then for small enough fixed $\rho, \tau > 0$ the function θ is injective on Υ and $\theta(\Upsilon) \subset \mathbf{D}_o$.*

Proof. We know from Lemma 4.9 and (4.13) that θ is positive and increasing from $[-1/w(\tilde\gamma_+), \infty)$ onto $[\tilde\gamma_+, \infty)$. Moreover, for small enough fixed $\rho > 0$, we deduce from Lemma 4.11 that there exists a small constant $\omega \equiv \omega(\rho) > 0$ such that for any $\zeta \in \Upsilon$,

$$\zeta \in \big[-1/w(\tilde\gamma_+) + \omega, \omega^{-1}\big] \times \big[-\rho, \rho\big]. \tag{4.34}$$

Thus, we may invoke Lemma 4.10 and Lemma 4.22 to conclude that θ is injective on Υ for some small enough $\rho > 0$.

We shall now show that $\theta(\Upsilon) \subset \mathbf{D}_o$ for some small fixed $\tau > 0$. We begin with the upper bound and we get from (4.12) and that $\|\tilde\Sigma\| = \sigma_{r+1}$,

$$|\theta(\zeta)| \leqslant \big|\sigma_{r+1}(1+\zeta)\big| \left[1 + \frac{1}{n}\sum_{i=1}^{p} \frac{1}{\big|\tilde\sigma_i^{-1}\sigma_{r+1}(1+\zeta) - 1\big|}\right] \leqslant |\sigma_{r+1}(1+\zeta)| \left[1 + \frac{\phi}{|\zeta|}\right]. \tag{4.35}$$

The upper bound $|\theta(\zeta)| \leqslant C$ follows from (2.2) and (4.34) for any $\zeta \in \Upsilon$ and small enough ρ. It remains to prove the lower bound on $\operatorname{Re}\theta(\zeta)$. For small enough $\rho > 0$, we can conclude from (4.34), Lemma 4.10 and the mean value theorem that there exists a constant $c \equiv c(\rho, \omega)$ such that for any $\zeta \in \Upsilon$,

$$\operatorname{Re}\theta(\zeta) \geqslant \tilde\gamma_+ + c\big(\operatorname{Re}\zeta + \frac{1}{w(\tilde\gamma_+)}\big), \tag{4.36}$$

and the lower bound easily follows using (4.34). We therefore conclude that there exists a constant $\tau > 0$ such that $\theta(\Upsilon) \subset \mathbf{D}_o$ for some small enough $\rho > 0$. ☐

Throughout the following, we fix $\rho > 0$ to be a small enough positive constant. By analytic extension of (4.12), we have

$$w(\theta(\zeta)) = \zeta^{-1}, \tag{4.37}$$

for any $\zeta \in \Upsilon$.

By the residue theorem and Lemmas 4.25 and 4.26, for small enough $\rho > 0$ we have for any $i, j \in [\![r]\!]$

$$\langle \mathbf{u}_i \,, \mathbf{v}_j \rangle^2 \;=\; -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} \langle \mathbf{v}_j \,, R(z)\mathbf{v}_j \rangle \, \mathrm{d}z \,, \tag{4.38}$$

The integrand of (4.38) can be rewritten thanks to the identity (4.18) as

$$\langle \mathbf{u}_i \,, \mathbf{v}_j \rangle^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} \frac{1}{z} \left[d_j^{-1} - \frac{1+d_j}{d_j^2} (D^{-1} + W(z))^{-1}_{jj} \right] \mathrm{d}z. \tag{4.39}$$

Using Lemma 4.26 and the change of variables $z = \theta(\zeta)$ satisfying

$$w(\theta(\zeta)) \;=\; -\zeta^{-1} \,, \tag{4.40}$$

we obtain

$$\langle \mathbf{u}_i \,, \mathbf{v}_j \rangle^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\Gamma_i} \left[d_j^{-1} - \frac{1+d_j}{d_j^2} \big(D^{-1} + W(\theta(\zeta))\big)^{-1}_{jj} \right] \frac{\theta'(\zeta)}{\theta(\zeta)} \mathrm{d}\zeta. \tag{4.41}$$

Next, we define the $r \times r$ error matrix

$$\mathcal{E}(\theta(\zeta)) \;:=\; w(\theta(\zeta))I - W(\theta(\zeta)), \tag{4.42}$$

and we rewrite (4.41) as (omitting the identity I in our notation from now on)

$$\langle \mathbf{u}_i\,,\mathbf{v}_j\rangle^2 = \frac{1+d_j}{d_j^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}(D^{-1}+w(\theta(\zeta))-\mathcal{E}(\theta(\zeta)))^{-1}_{jj}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta\,,$$

where we used that $\theta'(\zeta)/\theta(\zeta)$ is holomorphic in Υ_i by Lemmas 4.22 and 4.10 provided ρ is chosen small enough. Using the resolvent identity

$$(D^{-1}+W)^{-1} = (D^{-1}+w)^{-1} + (D^{-1}+w)^{-1}\mathcal{E}(D^{-1}+W)^{-1}, \tag{4.43}$$

we obtain

$$\langle \mathbf{u}_i\,,\mathbf{v}_j\rangle^2 = \frac{1+d_j}{d_j^2}\left[\chi_j^{(0)}+\chi_j^{(1)}\right], \tag{4.44}$$

where

$$\begin{aligned}
\chi_j^{(0)} &:= \frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\frac{1}{d_j^{-1}+w(\theta(\zeta))}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta,\\
\chi_j^{(1)} &:= \frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\left[\frac{1}{D^{-1}+w(\theta(\zeta))}\mathcal{E}(\theta(\zeta))\frac{1}{D^{-1}+W(\theta(\zeta))}\right]_{jj}\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta.
\end{aligned} \tag{4.45}$$

Using (4.10) and Lemma 4.26, the integral $\chi_j^{(0)}$ for $i,j\in[\![r]\!]$ reads

$$\chi_j^{(0)} = \frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\frac{1}{d_j^{-1}-\zeta^{-1}}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta. \tag{4.46}$$

By Cauchy's theorem and Lemma 4.25, we get

$$\chi_j^{(0)} = \delta_{ij}\,d_i^2\frac{\theta'(d_i)}{\theta(d_i)}. \tag{4.47}$$

For the remaining error term $\chi_j^{(1)}$, residue calculations are unavailable. Recalling its definition from (4.45), we find for $\zeta\in\Gamma_i$ that

$$|\chi_j^{(1)}| \leqslant Cn^{-1/2+\varepsilon}\sup_{\zeta\in\Gamma_i}\frac{1}{|d_j^{-1}+w(\theta(\zeta))|}\left\|\frac{1}{D^{-1}+W(\theta(\zeta))}\right\|, \tag{4.48}$$

where we used (4.32) and Lemmas 4.26, 4.10, and 4.22. Then, using (4.40) we obtain

$$\frac{1}{|d_j^{-1}+w(\theta(\zeta))|}\left\|\frac{1}{D^{-1}+W(\theta(\zeta))}\right\| \leqslant \frac{1}{|d_j^{-1}-\zeta^{-1}|}\,\frac{1}{\min_{k\in[\![r]\!]}|d_k^{-1}-\zeta^{-1}|-\|W(\theta(\zeta))-w(\theta(\zeta))\|}. \tag{4.49}$$

Using (4.7), we find from Lemma 4.11 that

$$|d_k^{-1}-\zeta^{-1}| \;\geqslant\; c\rho \tag{4.50}$$

for all $k\in[\![r]\!]$ and $\zeta\in\Gamma_i$. Since ρ is a constant, we obtain from (4.50) and (4.32) that

$$|\chi_j^{(1)}| \leqslant Cn^{-1/2+\varepsilon}. \tag{4.51}$$

Therefore, by plugging (4.47) and (4.51) into (4.44), we obtain from (2.2) and (4.7) that

$$\left|\langle \mathbf{u}_i\,,\mathbf{v}_j\rangle^2-(1+d_i)\frac{\theta'(d_i)}{\theta(d_i)}\right| \leqslant Cn^{-1/2+\varepsilon}. \tag{4.52}$$

Since $i\in[\![r]\!]$, we get from Lemma 4.11, (4.15), and (4.6) that (4.52) implies

$$\left|\langle \mathbf{u}_i\,,\mathbf{v}_j\rangle^2-\frac{1}{\sigma_i}\,\frac{\tilde f'(-1/\sigma_i)}{\tilde f(-1/\sigma_i)}\right| \leqslant Cn^{-1/2+\varepsilon}. \tag{4.53}$$

Since $\varepsilon>0$ was arbitrary, we obtain (4.28).

4.4. Proof of (4.29)**.** The first step of the proof of the estimate (4.29) concerns the evaluation of the overlaps $\langle \mathbf{u}_i\,, \mathbf{v}_j\rangle^2$ for $i \in [\![r]\!]$ and $j \notin [\![r]\!]$. Since the definition of β_i consists of a sum of an order n such terms, these overlaps have to be computed very precisely, up to an error much smaller than n^{-1}. This computation is the content of the following result.

Lemma 4.27. *Under the assumptions of Proposition 4.23 we have for any $i \in [\![r]\!]$ and $j \in [\![r+1, p]\!]$*

$$\left|\langle \mathbf{u}_i\,, \mathbf{v}_j\rangle^2 - \frac{d_i^2\theta'(d_i)}{\sigma_{r+1}\tilde{\sigma}_j\theta(d_i)}\tilde{G}_{ij}(\theta(d_i))\tilde{G}_{ji}(\theta(d_i))\right| \prec n^{-3/2}. \tag{4.54}$$

Before proving this result, we explain how to prove (4.29) using Lemma 4.27. Indeed, we see if we take the sum over $j = r+1, \dots, p$ in Lemma 4.27, we get

$$\left|\beta_i - \frac{d_i^2\theta'(\sigma_i)}{\theta(d_i)\sigma_{r+1}}\sum_{j=r+1}^{p}\tilde{G}_{ij}(\theta(d_i))\tilde{G}_{ji}(\theta(d_i))\right| \prec n^{-1/2}\,, \tag{4.55}$$

using that r is bounded and (2.7). Next, let us fix $i \in [\![r]\!]$ and rename $\theta_i \equiv \theta(d_i)$ to unburden the notation. We then have the following results, whose proofs are postponed at the end of this section.

Lemma 4.28 (Concentration)**.** *Suppose that the assumptions of Proposition 4.23 hold. Then for every $i \in [\![r]\!]$ we have*

$$\left|\sum_{j=r+1}^{p}\tilde{G}_{ij}(\theta_i)\tilde{G}_{ji}(\theta_i) - \mathbb{E}\sum_{j=r+1}^{p}\tilde{G}_{ij}(\theta_i)\tilde{G}_{ji}(\theta_i)\right| \prec n^{-1/2}\,. \tag{4.56}$$

Lemma 4.29 (Expectation)**.** *Suppose that the assumptions of Proposition 4.23 hold. Then for every $i \in [\![r]\!]$ we have*

$$\left|\mathbb{E}\sum_{j=r+1}^{p}\tilde{G}_{ij}(\theta_i)\tilde{G}_{ji}(\theta_i) - \frac{1}{(\sigma_{r+1}^{-1}+\tilde{m}(\theta_i))^2}\left[\frac{\tilde{m}'(\theta_i)}{\tilde{m}^2(\theta_i)} - 1\right]\right| \prec n^{-1/2}\,. \tag{4.57}$$

With Lemmas 4.27–4.29, we may conclude the proof of (4.29). Indeed, by plugging the estimate of Lemmas 4.27–4.29 into (4.55) yields for every $i \in [\![r]\!]$,

$$\beta_i = \frac{d_i^2\theta_i'}{\theta_i\sigma_{r+1}(\sigma_{r+1}^{-1}+\tilde{m}(\theta_i))^2}\left[\frac{\tilde{m}'(\theta_i)}{\tilde{m}^2(\theta_i)} - 1\right] + O_\prec(n^{-1/2})\,. \tag{4.58}$$

Next, recall that $d_i = \sigma_i/\sigma_{r+1} - 1$ from (4.6) and also that $\tilde{m}(\theta_i) = -1/\sigma_i$ from (2.12) and (4.15) for $i \in [\![r]\!]$. This yields

$$\beta_i = \frac{\sigma_i^2\theta_i'}{\theta_i\sigma_{r+1}}\left[\sigma_i^2\tilde{m}'(\theta_i) - 1\right] + O_\prec(n^{-1/2})\,.$$

We also find from (2.12) that

$$\tilde{m}'(\theta_i) = \frac{1}{\tilde{f}'(\tilde{m}(\theta_i))} = \frac{1}{\tilde{f}'(-1/\sigma_i)}\,, \tag{4.59}$$

and by using the identities of (4.15), we therefore conclude that for any $i \in [\![r]\!]$

$$\left|\beta_i - \frac{\sigma_i^2 - \tilde{f}'(-1/\sigma_i)}{\tilde{f}(-1/\sigma_i)}\right| \prec n^{-1/2}\,,$$

which is (4.29).

The rest of this section is devoted to the proofs of Lemma 4.27, Lemma 4.28 and Lemma 4.29.

4.5. Proof of Lemma 4.27. Let us fix $i \in [\![r]\!]$. To simplify notation, we set without loss of generality $j \equiv r+1$. We extend the definition of the matrices V and D from (4.5) to

$$\widehat{V} \;:=\; [\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}], \qquad \widehat{D} \;\equiv\; \widehat{D}_\delta \;:=\; \operatorname{diag}(d_1, \dots, d_r, \delta), \tag{4.60}$$

with $\delta > 0$ is a free parameter that will eventually be sent to zero. Hence, $\widehat{D}$ is an invertible $(r+1)\times(r+1)$ diagonal matrix. We accordingly define the $(r+1)\times(r+1)$ matrix $\widehat{W}(z)$ by analogy to $W(z)$ as

$$\widehat{W}(z) \;:=\; I + z\widehat{V}^* \tilde{R}(z)\, \widehat{V}. \tag{4.61}$$

Fix $\varepsilon > 0$ and define the event

$$\widehat{\Omega} \;\equiv\; \widehat{\Omega}_n(\varepsilon) \;:=\; \Big\{ \|w(z)I - \widehat{W}(z)\| \leqslant n^{-1/2+\varepsilon} \text{ for all } z \in \mathbf{D}_o \Big\} \\ \cap \big\{ \tilde{\lambda}_1 \leqslant \tilde{\gamma}_+ + \varepsilon \big\} \cap \big\{ |\lambda_i - \theta(d_i)| \leqslant n^{-1/2+\varepsilon} \text{ for all } i \in [\![r]\!] \big\}. \tag{4.62}$$

We infer from Theorem 4.19, Lemma 4.6 and Lemma 4.20 respectively that $\widehat{\Omega}$ holds with high probability. For the remainder of the proof, we fix a realization $X \in \widehat{\Omega}$. Moreover, to unburden notation we use the abbreviation $\hat{r} := r+1$.

For any $i \in [\![r]\!]$, we use the contour Γ_i defined in (4.33), and Υ_i, the closed disc defined in (4.33). The contour integral representation of the overlap $\langle \mathbf{u}_i\,, \mathbf{v}_{\hat{r}} \rangle^2$ for any $i \in [\![r]\!]$ yields

$$\langle \mathbf{u}_i\,, \mathbf{v}_{\hat{r}} \rangle^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} [\widehat{V}^* R(z) \widehat{V}]_{\hat{r}\hat{r}}\, \mathrm{d}z, \tag{4.63}$$

and applying the resolvent identity (4.18) to $\widehat{D}$ instead of D yields

$$\langle \mathbf{u}_i\,, \mathbf{v}_{\hat{r}} \rangle^2 = \lim_{\delta \downarrow 0} -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} \frac{1}{z} \left[\widehat{D}^{-1} - \frac{\sqrt{I+\widehat{D}}}{\widehat{D}} \Big(\widehat{D}^{-1} + \widehat{W}(z) \Big)^{-1} \frac{\sqrt{I+\widehat{D}}}{\widehat{D}} \right]_{\hat{r}\hat{r}} \mathrm{d}z\,. \tag{4.64}$$

As in Section 4.3, we rename $z = \theta(\zeta)$ and henceforth work on the ζ-plane, which yields

$$\langle \mathbf{u}_i\,, \mathbf{v}_{\hat{r}} \rangle^2 = \lim_{\delta \downarrow 0} -\frac{1}{2\pi \mathrm{i}} \oint_{\Gamma_i} \left[\widehat{D}^{-1} - \frac{\sqrt{I+\widehat{D}}}{\widehat{D}} \Big(\widehat{D}^{-1} + \widehat{W}(\theta(\zeta)) \Big)^{-1} \frac{\sqrt{I+\widehat{D}}}{\widehat{D}} \right]_{\hat{r}\hat{r}} \frac{\theta'(\zeta)}{\theta(\zeta)}\, \mathrm{d}\zeta\,. \tag{4.65}$$

Next, we deduce from Lemma 4.22 that for any $\delta > 0$,

$$\frac{1}{2\pi \mathrm{i}} \oint_{\Gamma_i} \frac{\theta'(\zeta)}{\delta \theta(\zeta)} \mathrm{d}\zeta = 0, \tag{4.66}$$

so that the contribution of the term $\widehat{D}^{-1}$ in (4.65) vanishes. For the remaining term in (4.65), the main difference regarding the derivation performed in Section 4.3 is that the overlaps have to be computed exactly up to an error much smaller than n^{-1}. To this end, we do a resolvent expansion in the $\hat{r} \times \hat{r}$ matrix

$$\widehat{\mathcal{E}} \equiv \widehat{\mathcal{E}}(\zeta) := w(\theta(\zeta)) - \widehat{W}(\theta(\zeta)). \tag{4.67}$$

We also define the $\hat{r} \times \hat{r}$ diagonal matrix

$$\Phi \;\equiv\; \Phi(\zeta) \;:=\; (\widehat{D}^{-1} + w(\theta(\zeta)))^{-1} \;=\; (\widehat{D}^{-1} - \zeta^{-1})^{-1}\,, \tag{4.68}$$

where the last step follows from (4.12) and Lemma 4.22.

Using the identity $(\widehat{D}^{-1} + \widehat{W})^{-1} = \Phi + (\widehat{D}^{-1} + \widehat{W})^{-1} \widehat{\mathcal{E}}\, \Phi$ we obtain the resolvent expansion

$$(\widehat{D}^{-1} + \widehat{W})^{-1} \quad = \quad \Phi + \Phi\widehat{\mathcal{E}}\Phi + \Phi\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}\Phi + \Phi\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}(\widehat{D}^{-1} + \widehat{W})^{-1}\widehat{\mathcal{E}}\Phi. \tag{4.69}$$

Going back to (4.65), we see that we have to compute

$$\langle \mathbf{u}_i\,, \mathbf{v}_{\hat{r}}\rangle^2 = \psi_{\hat{r}}^{(0)} + \psi_{\hat{r}}^{(1)} + \psi_{\hat{r}}^{(2)} + \psi_{\hat{r}}^{(3)}, \tag{4.70}$$

with

$$\begin{aligned}
\psi_{\hat{r}}^{(0)} &:= \lim_{\delta\downarrow 0} \frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i} [\Phi(\zeta)]_{\hat{r}\hat{r}}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta,\\
\psi_{\hat{r}}^{(1)} &:= \lim_{\delta\downarrow 0} \frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i} [\Phi(\zeta)\widehat{\mathcal{E}}(\zeta)\Phi(\zeta)]_{\hat{r}\hat{r}}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta,\\
\psi_{\hat{r}}^{(2)} &:= \lim_{\delta\downarrow 0} \frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i} [\Phi(\zeta)\widehat{\mathcal{E}}(\zeta)\Phi(\zeta)\widehat{\mathcal{E}}(\zeta)\Phi(\zeta)]_{\hat{r}\hat{r}}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta,\\
\psi_{\hat{r}}^{(3)} &:= \lim_{\delta\downarrow 0} \frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i} [\Phi(\zeta)\widehat{\mathcal{E}}(\zeta)\Phi(\zeta)\widehat{\mathcal{E}}(\zeta)(\widehat{D}^{-1}+\widehat{W}(\zeta))^{-1}\widehat{\mathcal{E}}(\zeta)\Phi(\zeta)]_{\hat{r}\hat{r}}\,\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta.
\end{aligned} \tag{4.71}$$

The rest of the proof is dedicated to the computation of (4.71). First, we need the following estimate on $\Phi_{\hat{r}\hat{r}}$ for any $\zeta\in\Upsilon$, which follows easily from Lemma 4.11.

Lemma 4.30. *Suppose that Assumption 4.1 holds. Then for a small enough constant $\rho>0$ and any small enough δ, we have for any $\zeta\in\Upsilon$ that*

$$|\Phi_{\hat{r}\hat{r}}(\zeta)| \leqslant C, \qquad |\Phi'_{\hat{r}\hat{r}}(z)| \leqslant C\,. \tag{4.72}$$

We may now evaluate $\psi^{(0)}$ and $\psi^{(1)}$.

Lemma 4.31. *Suppose that Assumption 4.1 holds and $X\in\widehat{\Omega}$. Then we have for any $z\in\Upsilon_i$ with $i\in[\![r]\!]$*

$$\psi_{\hat{r}}^{(0)} = \psi_{\hat{r}}^{(1)} = 0. \tag{4.73}$$

Proof. For small enough $\delta>0$, by Lemma 4.22 the integrand of $\psi_{\hat{r}}^{(0)}$ is holomorphic in Υ, so that by Cauchy's theorem we have $\psi_{\hat{r}}^{(0)}=0$.

For $\psi_{\hat{r}}^{(1)}$, we write using (4.37)

$$[\Phi(\zeta)\widehat{\mathcal{E}}(\zeta)\Phi(\zeta)]_{\hat{r}\hat{r}} \;=\; \Phi_{\hat{r}\hat{r}}(\zeta)^2\widehat{\mathcal{E}}_{\hat{r}\hat{r}}(\zeta) \;=\; -\Phi^2_{\hat{r}\hat{r}}(\zeta)\left[\theta(\zeta)\tilde{R}_{\hat{r}\hat{r}}(\theta(\zeta))+1+\zeta^{-1}\right].$$

From Lemmas 4.22 and 4.26, we conclude that the right-hand side is holomorphic for any $\zeta\in\Upsilon$. Hence, we get by dominated convergence,

$$\psi_{\hat{r}}^{(1)} = \frac{-1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\left[\theta(\zeta)\tilde{R}_{\hat{r}\hat{r}}(\theta(\zeta))+1+\zeta^{-1}\right]\frac{\theta'(\zeta)}{\theta(\zeta)}\,\mathrm{d}\zeta, \tag{4.74}$$

and since the term in the bracket is holomorphic in Υ, we deduce $\psi_{\hat{r}}^{(1)}=0$ from Lemma 4.22 and Cauchy's theorem. □

We now compute $\psi_{\hat{r}}^{(2)}$, which yields a nonzero contribution.

Lemma 4.32. *Suppose that Assumption 4.1 holds. Then we have for any $i\in[\![r]\!]$*

$$\psi_{\hat{r}}^{(2)} = \frac{d_i^2\theta'(d_i)}{\tilde{\sigma}_i\tilde{\sigma}_{\hat{r}}\theta(d_i)}\tilde{G}_{i\hat{r}}(\theta(d_i))\tilde{G}_{\hat{r}i}(\theta(d_i)). \tag{4.75}$$

Proof. Using the definition of the matrices Φ and $\widehat{\mathcal{E}}$ and decomposing the matrix multiplication into diagonal and offdiagonall terms, we get

$$[\Phi\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}\Phi]_{\hat{r}\hat{r}} \quad = \quad \Phi^2_{\hat{r}\hat{r}}\left[\Phi_{\hat{r}\hat{r}}\big(\theta(\zeta)\tilde{R}_{\hat{r}\hat{r}}(\theta(\zeta))+1+\zeta^{-1}\big)^2+\sum_{k=1}^{r}\frac{\theta(\zeta)^2\tilde{R}_{k\hat{r}}(\theta(\zeta))\tilde{R}_{\hat{r}k}(\theta(\zeta))}{d_k^{-1}-\zeta^{-1}}\right],$$

where we used (4.37) for $\zeta\in\Upsilon$. Since $\Phi_{\hat{r}\hat{r}}(\zeta)=O(\delta)$ for $\zeta\in\Gamma_i$, we obtain by dominated convergence that

$$\lim_{\delta\downarrow 0}\frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\Phi^3_{\hat{r}\hat{r}}(\zeta)\big(\theta(\zeta)\tilde{R}_{\hat{r}\hat{r}}(\theta(\zeta))+1+\zeta^{-1}\big)^2\frac{\theta'(\zeta)}{\theta(\zeta)}\mathrm{d}\zeta=0\,. \tag{4.76}$$

Thus,

$$\psi^{(2)}_{\hat{r}}=\lim_{\delta\downarrow 0}\frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\left[\sum_{k=1}^{r}\frac{\tilde{R}_{k\hat{r}}(\theta(\zeta))\tilde{R}_{\hat{r}k}(\theta(\zeta))}{d_k^{-1}-\zeta^{-1}}\right]\Phi^2_{\hat{r}\hat{r}}(\zeta)\theta(\zeta)\theta'(\zeta)\mathrm{d}\zeta. \tag{4.77}$$

From Lemma 4.26 we find that $\tilde{R}(\theta(\zeta))$ is holomorphic on Υ. Moreover, we infer from Lemma 4.30 and Lemma 4.22 that $\Phi^2_{\hat{r}\hat{r}}(\zeta)\theta(\zeta)\theta'(\zeta)$ is also holomorphic on Υ. Finally, we invoke Lemma 4.25 and Cauchy's theorem to deduce that the integral reduces to

$$\psi^{(2)}_{\hat{r}}=\lim_{\delta\downarrow 0}\frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\frac{\tilde{R}_{i\hat{r}}(\theta(\zeta))\tilde{R}_{\hat{r}i}(\theta(\zeta))}{d_i^{-1}-\zeta^{-1}}\Phi^2_{\hat{r}\hat{r}}(\zeta)\theta(\zeta)\theta'(\zeta)\mathrm{d}\zeta. \tag{4.78}$$

Using Lemma 4.25, this last integral can be computed using the residue theorem, and we find

$$\begin{aligned}\psi^{(2)}_{\hat{r}} &= \lim_{\delta\downarrow 0}\frac{1+\delta}{\delta^2}\Phi^2_{\hat{r}\hat{r}}(d_i)d_i^2\theta(d_i)\theta'(d_i)\tilde{R}_{i\hat{r}}(d_i)\tilde{R}_{\hat{r}i}(d_i),\\ &= d_i^2\theta(d_i)\theta'(d_i)\tilde{R}_{i\hat{r}}(d_i)\tilde{R}_{\hat{r}i}(d_i),\end{aligned} \tag{4.79}$$

where we used that $\Phi^2_{\hat{r}\hat{r}}(d_i)=(\delta^{-1}-d_i)^{-2}$ in the last step. The conclusion follows by noticing $\tilde{G}_{i\hat{r}}\tilde{G}_{\hat{r}i}=\theta^2(d_i)\tilde{\sigma}_i\tilde{\sigma}_{\hat{r}}\tilde{R}_{i\hat{r}}\tilde{R}_{\hat{r}i}$ from (4.20). □

It remains to estimate $\psi^{(3)}_{\hat{r}}$.

Lemma 4.33. *Suppose that Assumption 4.1 holds. Then*

$$|\psi^{(3)}_{\hat{r}}|\leqslant Cn^{-3/2+3\varepsilon}. \tag{4.80}$$

Proof. The proof is similar to the estimation (4.51). Residue calculations are unavailable and we need to estimate the integrand of $\psi^{(3)}_{\hat{r}}$ pointwise in $\zeta\in\Gamma_i$. To that end, we first rewrite

$$\psi^{(3)}_{\hat{r}}=\lim_{\delta\downarrow 0}\frac{1+\delta}{\delta^2}\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_i}\left[\Phi^2_{\hat{r}\hat{r}}\big(\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}(\widehat{D}^{-1}+\widehat{W})\mathcal{E}\big)_{\hat{r}\hat{r}}\right]\frac{\theta'}{\theta}\mathrm{d}\zeta. \tag{4.81}$$

We notice from Lemma 4.26, Lemma 4.30 and Lemma 4.22 that $\Phi^2_{\hat{r}\hat{r}}\theta'/\theta$ satisfies the estimate

$$\big|\Phi^2_{\hat{r}\hat{r}}\theta'(\zeta)/\theta(\zeta)\big|\leqslant C\delta^2 \tag{4.82}$$

for any $\zeta\in\Gamma_i$ and small enough δ. Next, we have from (4.62) that

$$\big|\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}(\widehat{D}^{-1}+\widehat{W})\widehat{\mathcal{E}}\big|_{\hat{r}\hat{r}}\leqslant n^{-3/2+3\varepsilon}\|\Phi\|\|(\widehat{D}^{-1}+\widehat{W})^{-1}\|. \tag{4.83}$$

Then, we infer from (4.7) and (4.50) that for a small enough constant $\rho>0$ and for small enough $\delta>0$, we have for all $\zeta\in\Gamma_i$ that $\|\Phi\|\leqslant C$. Similarly, we estimate

$$\|(\widehat{D}^{-1}+\widehat{W})^{-1}\| \;\leqslant\; \min_{k\in[\![\hat{r}]\!]}\frac{1}{|\widehat{D}_{kk}-\zeta^{-1}|-\|\widehat{\mathcal{W}}\|} \;\leqslant\; C$$

for $\zeta \in \Gamma_i$, for a small enough constant $\rho > 0$ and for small enough $\delta > 0$.

Summarizing, we have, for a small enough constant $\rho > 0$ and for small enough $\delta > 0$,

$$\big|\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}(\widehat{D}^{-1} + \widehat{W})\widehat{\mathcal{E}}\big|_{\hat{r}\hat{r}} \leqslant Cn^{-3/2+3\varepsilon} \tag{4.84}$$

for $\zeta \in \Gamma_i$. Therefore, we conclude from (4.82), (4.84) and by dominated convergence that

$$|\psi_{\hat{r}}^{(3)}| \leqslant C \sup_{z\in\Gamma_i}\big|\widehat{\mathcal{E}}\Phi\widehat{\mathcal{E}}(\widehat{D}^{-1} + \widehat{W})\widehat{\mathcal{E}}\big|_{\hat{r}\hat{r}} \leqslant Cn^{-3/2+3\varepsilon}, \tag{4.85}$$

where we used $\|\widehat{\mathcal{E}}\| \leqslant Cn^{-1/2+\varepsilon}$ by definition of $\widehat{\Omega}$. Now the claim follows. $\square$

Since $\varepsilon > 0$ was arbitrary in (4.62), the claim of Lemma 4.27 follows by applying Lemmas 4.31, 4.32, and 4.33 to (4.70). This concludes the proof of Lemma 4.27.

4.6. Proof of Lemma 4.28. Fix $i \in [\![r]\!]$ and abbreviate $z = \theta(d_i) = \theta_i$. The strategy of the proof is to show that with high probability the function

$$h(X) := \sum_{j=r+1}^{p} \tilde{G}_{ij}\tilde{G}_{ji} \equiv z^2 \sum_{j=r+1}^{p} \Big[\tilde{\Sigma}^{1/2}(\tilde{\Sigma}^{1/2}XX^*\tilde{\Sigma}^{1/2} - z)^{-1}\tilde{\Sigma}^{1/2}\Big]_{ij} \Big[\tilde{\Sigma}^{1/2}(\tilde{\Sigma}^{1/2}XX^*\tilde{\Sigma}^{1/2} - z)^{-1}\tilde{\Sigma}^{1/2}\Big]_{ji} \tag{4.86}$$

is a C-Lipschitz function, in order to use the Herbst inequality [29] to deduce concentration. To that end, we first introduce the orthogonal projection

$$P := \sum_{i=r+1}^{p} \mathbf{v}_i\mathbf{v}_i^*, \tag{4.87}$$

which is diagonal by Assumption 4.1 and commutes with $\tilde{\Sigma}$. Hence, using (4.5), (4.20) and (4.87), we rewrite

$$h(X) = z^2\sigma_{r+1}\Big[\tilde{R}(X)\,P\,\tilde{\Sigma}\,\tilde{R}(X))\Big]_{ii}, \tag{4.88}$$

where we renamed $\tilde{R}(X) \equiv \tilde{R}$, defined in (4.17), as a function of the matrix X while omitting the argument z.

Next, we define a truncation of X.

Lemma 4.34 (Truncation)**.** *Suppose that the assumptions of Proposition 4.23 hold. For $\chi > 0$ define the function*

$$T(X) \equiv T_\chi(X) := \frac{X}{\frac{\|\tilde{\Sigma}^{1/2}X\|}{\sqrt{\tilde{\gamma}_+ + \chi^{-1}}} \vee \frac{\|X\|}{\chi} \vee 1}. \tag{4.89}$$

Then for a large enough constant $\chi > 0$ we have $X = T(X)$ with high probability.

Proof. See Appendix C.10. $\square$

Note that, by construction, we have $\|T(X)\| \leqslant \chi$ and $\|\tilde{\Sigma}^{1/2}T(X)\|^2 \leqslant \tilde{\gamma}_+ + \chi^{-1}$.

Lemma 4.35. *Suppose that the assumptions of Proposition 4.23 hold. Then there is a constant C depending on χ such that*

$$\|T(X) - T(X')\| \leqslant C\|X - X'\| \tag{4.90}$$

for all $X, X' \in \mathbb{R}^{p\times n}$.

Proof. See Appendix C.10. $\square$

As a result, we may define the truncated version of $h(X)$ as

$$\hat{h}(X) \;:=\; z^2\sigma_{r+1}\left[\tilde{R}(T(X))\,P\,\tilde{\Sigma}\,\tilde{R}(T(X))\right]_{ii}\,, \tag{4.91}$$

and we see from Lemma 4.34 that $h(X) = \hat{h}(X)$ with high probability and we shall prove the Lipschitz continuity for $\hat{h}$. Hence, the following arguments are deterministic. First, we abbreviate $\tilde{R}_T \equiv \tilde{R}(T(X))$ and $\tilde{R}'_T \equiv \tilde{R}(T(X'))$. Then, recall that we work with $z = \theta(d_i)$ for $i \in [\![r]\!]$ and we thus infer from (2.2), (4.91) and Lemma 4.10 that

$$\left|\hat{h}(X) - \hat{h}(X')\right| \leqslant C\left|\left[\tilde{R}_T\,P\,\tilde{\Sigma}\,\tilde{R}_T\right]_{ii} - \left[\tilde{R}'_T\,P\,\tilde{\Sigma}\,\tilde{R}'_T\right]_{ii}\right|, \tag{4.92}$$

from which we deduce

$$\left|\hat{h}(X) - \hat{h}(X')\right| \leqslant C\left|\left\|\tilde{R}_T\,P\,\tilde{\Sigma}\,(\tilde{R}_T - \tilde{R}'_T)\right\| + \left\|(\tilde{R}_T - \tilde{R}'_T)\,P\,\tilde{\Sigma}\,\tilde{R}'_T\right\|\right| \leqslant C\|\tilde{R}_T - \tilde{R}'_T\|\big(\|\tilde{R}_T\| + \|\tilde{R}'_T\|\big), \tag{4.93}$$

where we used that $\|P\tilde{\Sigma}\| \leqslant C$ from (4.5) and (2.2). Next, using (4.16), there exists $\tau > 0$ such that $\theta(d_i) \geqslant \tilde{\gamma}_+ + \tau$ for $i \in [\![r]\!]$ and we deduce from Lemma 4.34 that $\|\tilde{R}_T\| \leqslant \varepsilon^{-1}$ for $\varepsilon > 0$ a constant that depends on χ and τ. Thus, this implies

$$\left|\hat{h}(X) - \hat{h}(X')\right| \leqslant C\|\tilde{R}_T - \tilde{R}'_T\|\,,$$

where C now depends on ε. Using the resolvent identity, we obtain

$$\tilde{R}_T - \tilde{R}'_T = \tilde{R}_T\tilde{\Sigma}^{1/2}\big[T(X)T(X)^* - T(X')T(X')^*\big]\tilde{\Sigma}^{1/2}\tilde{R}'_T, \tag{4.94}$$

so that we have

$$\begin{aligned}\left|\hat{h}(X) - \hat{h}(X')\right| \;&\leqslant\; C\left\|\tilde{R}_T\,\tilde{\Sigma}^{1/2}\right\|\left\|\tilde{R}'_T\,\tilde{\Sigma}^{1/2}\right\|\left\|T(X)T(X)^* - T(X')T(X')^*\right\|\\ &=\; C\left\|\tilde{R}_T\,\tilde{\Sigma}^{1/2}\right\|\left\|\tilde{R}'_T\,\tilde{\Sigma}^{1/2}\right\|\left\|\big[T(X) - T(X')\big]T(X)^* + T(X')\big[T(X)^* - T(X')^*\big]\right\|.\end{aligned}$$

Using the same arguments as above, we find

$$\left|\hat{h}(X) - \hat{h}(X')\right| \leqslant C\Big(\|T(X)\| + \|T(X')\|\Big)\|T(X) - T(X')\|\,. \tag{4.95}$$

Next, recall that $\|\cdot\|$ denotes the Euclidean operator norm so that one has from Lemma 4.34 that $\|T(X)\| \leqslant C$ for some constant $C > 0$. Hence, this leads to

$$\left|\hat{h}(X) - \hat{h}(X')\right| \leqslant C\left\|T(X) - T(X')\right\|\,, \tag{4.96}$$

and we now use Lemma 4.35 to conclude that $\left\|T(X) - T(X')\right\| \leqslant C\left\|X - X'\right\|_2$. All in all, we conclude that $\hat{h}$ is a Lipschitz continuous function

$$\left|\hat{h}(X) - \hat{h}(X')\right| \leqslant C\left\|X - X'\right\|_2 \tag{4.97}$$

and Herbst inequality with (4.3) tell us that

$$\mathbb{P}\Big(\big|\hat{h}(X) - \mathbb{E}\,\hat{h}(X')\big| > t\Big) \leqslant \exp\big(-C^{-1}t^2 n\big) \tag{4.98}$$

which is the claim.

4.7. Proof of Lemma 4.29. Fix $i \in [\![r]\!]$ and abbreviate $\theta_i \equiv \theta(d_i)$ throughout the following. We omit the argument θ_i in our notation as soon as there is no risk of confusion. The proof of Lemma 4.29 relies on a resolvent expansion of $\tilde{G}_{ij}$ for any $j \in [\![r+1, p]\!]$.

Using Definition 4.16, we introduce the conditional expectation

$$\mathbb{E}^{(ij)}[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\tilde{H}^{(ij)}]\,. \tag{4.99}$$

Without loss of generality, we may fix $j \in [\![r+1, p]\!]$ throughout the following. Next, using Schur's complement formula, we can write $\tilde{G}_{ij} = (B^{-1})_{ij}$ where $B = (B_{kl})_{k,l\in\{i,j\}}$ is a 2×2 matrix defined by

$$B_{kl} \;:=\; -\frac{\delta_{kl}}{\tilde{\sigma}_k} - (X\tilde{G}^{(kl)}X^*)_{kl}\,. \tag{4.100}$$

We rewrite this latter quantity as

$$B = A - (Z + \Delta), \tag{4.101}$$

where A, Z and Δ are 2×2 matrices defined for any $k, l \in \{i, j\}$ by

$$A_{kl}(\theta_i) \;\equiv\; A_{kl} \;:=\; -\delta_{kl}\big(\tilde{\sigma}_k^{-1} + \tilde{m}\big), \tag{4.102}$$

and

$$Z_{kl}(\theta_i) \equiv Z_{kl} \;:=\; \big(1 - \mathbb{E}^{(ij)}\big)\big(X\tilde{G}^{(kl)}X^*\big)_{kl}, \qquad \Delta_{kl}(\theta_i) \equiv \Delta_{kl} \;:=\; \mathbb{E}^{(ij)}\big(X\tilde{G}^{(kl)}X^*\big)_{kl} - \delta_{kl}\tilde{m}. \tag{4.103}$$

Recall from (4.15) that $\tilde{m}(\theta(d_i)) = -\sigma_i^{-1}$ for $i \in [\![r]\!]$. We therefore obtain from (2.2), (2.4), (4.5) and Lemma C.1 that

$$|A_{mm}| \asymp 1, \tag{4.104}$$

for any $m \in \{k, l\}$. We therefore deduce that A is a diagonal invertible 2×2 matrix. It remains to control the entries of Z and Δ. For this purpose, we introduce the following result.

Lemma 4.36. *Suppose that the assumptions of Proposition 4.23 hold. Then, we have for any $i \in [\![r]\!]$ and $t \in \mathcal{I}$,*

$$|\tilde{G}_{tt}(\theta_i)| \asymp 1, \qquad \max_{s\neq t\in\mathcal{I}}|\tilde{G}_{st}(\theta_i)| \prec n^{-1/2}. \tag{4.105}$$

Moreover, let $u, v \in \mathcal{I}$ with $u \neq v$. We have for any $i \in [\![r]\!]$ and any $s, t \in \mathcal{I}\backslash\{u, v\}$ that

$$\big|\tilde{G}_{st}^{(u)}(\theta_i) - \tilde{G}_{st}(\theta_i)\big| \prec n^{-1}, \qquad \big|\tilde{G}_{st}^{(uv)}(\theta_i) - \tilde{G}_{st}(\theta_i)\big| \prec n^{-1}. \tag{4.106}$$

Proof. Let $i \in [\![r]\!]$. From (4.16), we use Lemma 4.19 to find $|\tilde{G}_{st} - \tilde{\Pi}_{st}| \prec n^{-1/2}$ for any $s, t \in \mathcal{I}$ where $\tilde{\Pi}$ is defined in (4.22). We then obtain both estimates of (4.105) by using (4.22), (4.16) and Lemma 4.7. Next, we prove the first estimate of (4.106). Let us fix $u \in \mathcal{I}$ and from (4.21), we have for any $s, t \in \mathcal{I}\backslash\{u\}$

$$\tilde{G}_{st}^{(u)} = \tilde{G}_{st} - \frac{\tilde{G}_{su}\tilde{G}_{ut}}{G_{uu}} = \tilde{G}_{st} + O_{\prec}(n^{-1}), \tag{4.107}$$

where we used (4.105) in the last step. The second estimate of (4.106) follows by applying (4.21), (4.105) and (4.107). Indeed, let $v \neq u \in \mathcal{I}$, one has from (4.21) that for any $s, t \in \mathcal{I}\backslash\{u, v\}$,

$$\tilde{G}_{st}^{(uv)} = \tilde{G}_{st}^{(u)} - \frac{\tilde{G}_{sv}^{(u)}\tilde{G}_{vt}^{(u)}}{\tilde{G}_{vv}^{(u)}}. \tag{4.108}$$

We deduce from (4.107) and (4.105) that $|\tilde{G}_{vv}^{(u)}| \geqslant c$ and therefore, the conclusion follows from (4.107). $\square$

Let us now record some immediate consequences of Lemma 4.36. First, we deduce from (2.2), (2.4), Lemma 4.19 and Lemma 4.36 that for any $u \in \mathcal{I}$ and $s, t \in \mathcal{I}\backslash\{u\}$ with $s \neq t$,

$$\big|\tilde{G}_{tt}^{(u)}\big| \asymp 1\,, \qquad \big|\tilde{G}_{st}^{(u)}\big| \prec n^{-1/2}\,. \tag{4.109}$$

Using the same arguments, we obtain for any $u, v \in \mathcal{I}$ with $u \neq v$ and any $s, t \in \mathcal{I}\backslash\{u, v\}$ with $s \neq t$ that

$$\big|\tilde{G}_{tt}^{(uv)}\big| \asymp 1\,, \qquad \big|\tilde{G}_{st}^{(uv)}\big| \prec n^{-1/2}\,. \tag{4.110}$$

The control of the entries of Δ and Z follows immediately from the following lemmas.

Lemma 4.37. *Let $i \in [\![r]\!]$ and $j \in [\![p]\!]$ and suppose that the assumptions of Proposition 4.23 hold. Then, we have for any $k, l \in \{i, j\}$ that*

$$|\Delta_{kl}(\theta_i)| \prec n^{-1/2}. \tag{4.111}$$

Proof. Let $i \in [\![r]\!]$, $j \in [\![p]\!]$ and we fix $k, l \in \{i, j\}$ throughout the following. From (4.99), we find

$$\mathbb{E}^{(ij)}(X\tilde{G}^{(kl)}X^*)_{kl} = \frac{\delta_{kl}}{n} \sum_{\mu \in \mathcal{I}_n} \tilde{G}_{\mu\mu}^{(kl)}\,. \tag{4.112}$$

Next, we invoke Lemma 4.36 and then Lemma 4.19 to obtain

$$\mathbb{E}^{(ij)}(X\tilde{G}^{(kl)}X^*)_{kl} = \frac{\delta_{kl}}{n} \sum_{\mu \in \mathcal{I}_n} G_{\mu\mu} + O_\prec(n^{-1}) = \delta_{kl}\tilde{m} + O_\prec(n^{-1/2}). \tag{4.113}$$

□

Lemma 4.38. *Let $i \in [\![r]\!]$ and $j \in [\![p]\!]$ and suppose that the assumptions of Proposition 4.23 hold. Then, we have for any $k, l \in \{i, j\}$ that*

$$|Z_{kl}(\theta_i)| \prec n^{-1/2}. \tag{4.114}$$

Proof. Let $i \in [\![r]\!]$, $j \in [\![p]\!]$ and we fix $k, l \in \{i, j\}$ throughout the following. We begin with the case $k \neq l$. In that case, we infer from (4.103) and (4.112) that

$$Z_{kl} = \big(X\tilde{G}^{(kl)}X^*\big)_{kl} = \sum_{\mu,\nu \in \mathcal{I}_n} X_{k\mu}\tilde{G}_{\mu\nu}^{(kl)}X_{l\nu}. \tag{4.115}$$

Then, we can use a large deviation estimate like that of [2, Lemma 3.1] to get

$$Z_{kl} \prec \left(\frac{1}{n^2}\sum_{\mu,\nu \in \mathcal{I}_n} \big|\tilde{G}_{\mu\nu}^{(kl)}\big|^2\right)^{1/2}. \tag{4.116}$$

It is then easy to obtain the desired bound from (4.110).

It now remains to tackle the case $k = l$. We have from (4.103),

$$\begin{aligned} Z_{kk} &= \sum_{\mu,\nu \in \mathcal{I}_n} \big(1 - \mathbb{E}^{(ij)}\big)X_{k\mu}\tilde{G}_{\mu\nu}^{(k)}X_{k\nu} \\ &= \sum_{\mu \in \mathcal{I}_n} \big(1 - \mathbb{E}^{(ij)}\big)X_{k\mu}^2\tilde{G}_{\mu\mu}^{(k)} + \sum_{\mu \neq \nu \in \mathcal{I}_n} X_{k\mu}\tilde{G}_{\mu\nu}^{(k)}X_{k\nu}, \end{aligned} \tag{4.117}$$

where we used (4.3) in the last line. The two terms on the right-hand side can be estimated using large deviation estimates (see [2, Lemma 3.1]). For the first term of (4.117) we get

$$\left|\sum_{\mu \in \mathcal{I}_n} \big(1 - \mathbb{E}^{(ij)}\big)X_{k\mu}^2\tilde{G}_{\mu\mu}^{(k)}\right| \prec \left(\frac{1}{n^2}\sum_{\mu \in \mathcal{I}_n} \big|\tilde{G}_{\mu\mu}^{(k)}\big|^2\right)^{1/2} \prec n^{-1/2} \tag{4.118}$$

where we used (4.109) in the last step. For the second term in the right-hand side of (4.117), the large estimate of [2, Lemma 3.1] and (4.109) yields

$$\sum_{\mu\neq\nu\in\mathcal{I}_n} X_{k\mu}\tilde{G}^{(k)}_{\mu\nu}X_{k\nu} \prec \left(\frac{1}{n^2}\sum_{\mu\neq\nu\in\mathcal{I}_n}\big|\tilde{G}^{(k)}_{\mu\nu}\big|^2\right)^{1/2} \prec n^{-1/2} \tag{4.119}$$

and this ends the proof.

□

We now turn on the resolvent expansion. Since A is invertible, one obtains from (4.101)

$$\tilde{G}_{ij} = (B^{-1})_{ij} = \Big[A^{-1} + A^{-1}(Z+\Delta)A^{-1} + Q\Big]_{ij}, \tag{4.120}$$

where

$$Q(\theta_i) \;\equiv\; Q \;:=\; A^{-1}(Z+\Delta)(A-(Z+\Delta))^{-1}(Z+\Delta)A^{-1}. \tag{4.121}$$

Hence, using that A is a diagonal invertible matrix from (4.102), then the fact that $i \neq j$ since $i \in [\![r]\!]$ and $j \in [\![r+1,p]\!]$, we deduce from (4.120) that

$$\begin{aligned}\tilde{G}_{ij}\tilde{G}_{ji} &= \Big[(A^{-1}(Z+\Delta)A^{-1})_{ij} + Q_{ij}\Big]\Big[(A^{-1}(Z+\Delta)A^{-1})_{ji} + Q_{ji}\Big]\,,\\ &= (A^{-1}_{ii})^2(A^{-1}_{jj})^2(Z+\Delta)_{ij}(Z+\Delta)_{ji} + \Theta_{ij}\,,\end{aligned} \tag{4.122}$$

where we defined the 2×2 matrix

$$\Theta_{kl} \;:=\; (A^{-1}(Z+\Delta)A^{-1})_{kl}Q_{lk} + Q_{kl}(A^{-1}(Z+\Delta)A^{-1})_{lk} + Q_{kl}Q_{lk}\,, \tag{4.123}$$

for any $k,l \in \{i,j\}$. Let us focus on the remainder term Θ_{ij} for which we have the following estimate.

Lemma 4.39. *Suppose that the assumptions of Lemma 4.29 hold. Then we have for any $i \in [\![r]\!]$,*

$$\|Q(\theta_i)\| \prec n^{-1}\,. \tag{4.124}$$

Proof. For any $i \in [\![r]\!]$, we infer from (4.121) and Lemmas 4.37-4.38 that

$$\|Q\| \prec n^{-1}\big\|A^{-1}\big\|^2\big\|(A-(Z+\Delta))^{-1}\big\|\,. \tag{4.125}$$

Next, we invoke (4.104) to find for any θ_i with $i \in [\![r]\!]$,

$$\big\|A^{-1}\big\| \leqslant C\,. \tag{4.126}$$

Finally, we get from (4.104) and Lemmas 4.37–4.38 that for any θ_i with $i \in [\![r]\!]$,

$$\big\|(A-(Z+\Delta))^{-1}\big\| \leqslant \frac{1}{\min_{k\in[\![r]\!]}|\tilde{\sigma}_k^{-1}+\tilde{m}| - \|Z+\Delta\|} \leqslant C, \tag{4.127}$$

and this proves our claim. □

Using that A is diagonal, (4.104), Lemmas 4.37, 4.38 and 4.39, we conclude that

$$\big\|\Theta\big\| \leqslant 2\big\|A^{-2}\big\|\big\|Z+\Delta\big\|\big\|Q\big\| + \big\|Q\big\|^2 \prec n^{-3/2}\,, \tag{4.128}$$

so that we obtain from (4.122),

$$\mathbb{E}\tilde{G}_{ij}\tilde{G}_{ji} = (A^{-1}_{ii})^2(A^{-1}_{jj})^2\mathbb{E}\Big[(Z+\Delta)_{ij}(Z+\Delta)_{ji}\Big] + O_\prec(n^{-3/2})\,, \tag{4.129}$$

where we used that A is a deterministic matrix. Since Δ is a diagonal matrix from Lemma 4.37 and using that $i \neq j$ for any $i \in [\![r]\!]$ and $j \in [\![r+1,p]\!]$, we have

$$(Z+\Delta)_{ij}(Z+\Delta)_{ji} = Z_{ij}Z_{ji} + \Delta_{ij}\Delta_{ji} + Z_{ij}\Delta_{ji} + \Delta_{ji}Z_{ij} = Z_{ij}Z_{ji}\,. \tag{4.130}$$

As a result, we conclude from (4.129) that for any $i \in [\![r]\!]$ and $j \in [\![r+1,p]\!]$,

$$\mathbb{E}\tilde{G}_{ij}\tilde{G}_{ji} = (A_{ii}^{-1})^2(A_{jj}^{-1})^2\mathbb{E}\Big[Z_{ij}Z_{ji}\Big] + O_\prec(n^{-3/2})\,. \tag{4.131}$$

We then need the following estimate.

Lemma 4.40. *Suppose that the assumptions of Proposition 4.23 hold. Then for any θ_i with $i \in [\![r]\!]$ and any $j \in [\![r+1,p]\!]$, we have*

$$\left|\mathbb{E}Z_{ij}Z_{ji} - \frac{1}{n^2}\sum_{\mu,\nu\in\mathcal{I}_n}\tilde{G}_{\mu\nu}(\theta(d_i))\tilde{G}_{\nu\mu}(\theta(d_i))\right| \prec n^{-3/2}. \tag{4.132}$$

Proof. We fix throughout the following $i \in [\![r]\!]$ and $j \in [\![r+1,p]\!]$. Using (4.103) and then (4.112), we obtain

$$\begin{aligned}
\mathbb{E}Z_{ij}Z_{ji} &= \mathbb{E}\left[\Big((1-\mathbb{E}^{(ij)})(X\tilde{G}^{(ij)}X^*)_{ij}(1-\mathbb{E}^{(ji)})(X\tilde{G}^{(ji)}X^*)_{ji}\Big)\right] \\
&= \mathbb{E}\left[(X\tilde{G}^{(ij)}X^*)_{ij}(X\tilde{G}^{(ji)}X^*)_{ji} - \delta_{ij}\left(\frac{1}{n}\sum_{\mu\in\mathcal{I}_n}\tilde{G}^{(ij)}_{\mu\mu}\right)^2\right] \\
&= \mathbb{E}\Big[(X\tilde{G}^{(ij)}X^*)_{ij}(X\tilde{G}^{(ji)}X^*)_{ji}\Big]\,,
\end{aligned} \tag{4.133}$$

where we used that $i \neq j$ in the last step. By the law of iterated expectation and Wick's theorem, one finds

$$\begin{aligned}
\mathbb{E}\Big[(X\tilde{G}^{(ij)}X^*)_{ij}(X\tilde{G}^{(ji)}X^*)_{ji}\Big] &= \mathbb{E}\left[\delta_{ij}\left(\frac{1}{n}\sum_{\mu\in\mathcal{I}_n}\tilde{G}^{(ij)}_{\mu\mu}\right)^2 + \frac{\delta_{ij}}{n^2}\sum_{\mu,\nu\in\mathcal{I}_n}|\tilde{G}^{(ij)}_{\mu\nu}|^2 + \frac{1}{n^2}\sum_{\mu,\nu\in\mathcal{I}_n}\tilde{G}^{(ij)}_{\mu\nu}\tilde{G}^{(ji)}_{\nu\mu}\right] \\
&= \mathbb{E}\left[\frac{1}{n^2}\sum_{\mu,\nu\in\mathcal{I}_n}\tilde{G}^{(ij)}_{\mu\nu}\tilde{G}^{(ji)}_{\nu\mu}\right],
\end{aligned} \tag{4.134}$$

using once again that $i \neq j$ in the last step. Hence, by plugging (4.134) into (4.133), we have

$$\mathbb{E}Z_{ij}Z_{ji} = \frac{1}{n^2}\sum_{\mu,\nu\in\mathcal{I}_n}\mathbb{E}\tilde{G}^{(ij)}_{\mu\nu}\tilde{G}^{(ji)}_{\nu\mu}\,, \tag{4.135}$$

and we conclude from Lemma 4.36 that

$$\mathbb{E}Z_{ij}Z_{ji} = \frac{1}{n^2}\sum_{\mu,\nu\in\mathcal{I}_n}\mathbb{E}\tilde{G}_{\mu\nu}\tilde{G}_{\nu\mu} + O_\prec(n^{-3/2})\,, \tag{4.136}$$

which is the claim.

□

By plugging the estimate of Lemma 4.40 into (4.131), then recalling that $\tilde{\sigma}_i = \sigma_{r+1}$ for any $i \in [\![r]\!]$ from (4.5), we infer from (4.102), (4.104) and Assumption 2.3 that

$$\sum_{j=r+1}^{p}\mathbb{E}\tilde{G}_{ij}\tilde{G}_{ji} = \frac{1}{(\sigma_{r+1}^{-1}+\tilde{m})^2}\left[\frac{1}{n^2}\sum_{j=r+1}^{p}\frac{\mathrm{Tr}(\tilde{G}_n^2)}{(\tilde{\sigma}_j^{-1}+\tilde{m})^2}\right] + O_\prec(n^{-1/2})\,, \tag{4.137}$$

where we defined $\tilde{G}_n \equiv \tilde{G}_n(z) := (X^*\tilde{\Sigma}X - z)^{-1}$. Next, using Lemma 4.19 and [23, Theorem 3.12], we get

$$\left|\frac{1}{n}\operatorname{Tr}(\tilde{G}_n^2) - m'(z)\right| \prec n^{-1/2}\,, \tag{4.138}$$

so that we obtain using (4.102) and (4.104) that

$$\sum_{j=r+1}^{p} \mathbb{E}\tilde{G}_{ij}\tilde{G}_{ji} = \frac{1}{(\sigma_{r+1}^{-1}+\tilde{m})^2}\left[\frac{1}{n}\sum_{j=r+1}^{p}\frac{\tilde{m}'(\theta_i)}{(\tilde{\sigma}_j^{-1}+\tilde{m}(\theta_i))^2}\right] + O_\prec(n^{-1/2})\,. \tag{4.139}$$

It remains to estimate the term in the bracket in this latter equation. To that end, we invoke Assumption 2.3, (4.16) and Lemma 4.7 to deduce that for any $i \in [\![r]\!]$

$$\frac{1}{n}\sum_{j=r+1}^{p}\frac{\tilde{m}'(\theta_i)}{(\tilde{\sigma}_j^{-1}+\tilde{m}(\theta_i))^2} = \frac{1}{n}\sum_{j=1}^{p}\frac{\tilde{m}'(\theta_i)}{(\tilde{\sigma}_j^{-1}+\tilde{m}(\theta_i))^2} + O_\prec(n^{-1})\,. \tag{4.140}$$

Then, one finds after taking the derivative with respect to z in (2.12) that

$$\frac{1}{n}\sum_{j=1}^{p}\frac{\tilde{m}'(z)}{(\tilde{\sigma}_j^{-1}+\tilde{m}(z))^2} = \frac{\tilde{m}'(z)}{\tilde{m}^2(z)} - 1\,, \tag{4.141}$$

so we conclude by plugging this last two estimates into (4.139) that for $z = \theta_i$ with $i \in [\![r]\!]$,

$$\sum_{j=r+1}^{p} \mathbb{E}\tilde{G}_{ij}(\theta_i)\tilde{G}_{ji}(\theta_i) = \frac{1}{(\sigma_{r+1}^{-1}+\tilde{m}(\theta_i))^2}\left[\frac{\tilde{m}'(\theta_i)}{\tilde{m}^2(\theta_i)} - 1\right] + O_\prec(n^{-1/2})\,, \tag{4.142}$$

which is exactly (4.57).

4.8. Conclusion of the proof of (2.19) for outliers. Using the decomposition (4.27), (4.15) and the results of Proposition 4.23, we see that the oracle estimator (2.17) may be approximated with high probability by a deterministic quantity for any $i \in [\![r]\!]$,

$$\left|\tilde{\xi}_i - \frac{\sigma_i^2}{\theta(d_i)}\right| \prec n^{-1/2},$$

where we recall that $\theta(d_i) \asymp 1$ from (4.16) and Lemma 4.5. The remainder of this section is dedicated to the rewriting of this last estimate as a function of observable and random quantities. We first need the following estimate.

Lemma 4.41. *Suppose that the assumptions of Proposition 4.23 hold. Then for every $i \in [\![r]\!]$ we have*

$$\left|\tilde{\xi}_i - \frac{1}{\lambda_i\tilde{m}^2(\lambda_i)}\right| \prec n^{-1/2}\,. \tag{4.143}$$

Proof. Without loss of generality, we may fix $i \in [\![r]\!]$ throughout the following. Let $\varepsilon > 0$ and define the event

$$\bar{\Omega} \;\equiv\; \bar{\Omega}(\varepsilon) \;:=\; \left\{|\lambda_i - \theta(d_i)| \leqslant n^{-1/2+\varepsilon} \text{ for all } i \in [\![r]\!]\right\} \cap \left\{\lambda_i \geqslant \tilde{\gamma}_+ + \varepsilon \text{ for all } i \in [\![r]\!]\right\}. \tag{4.144}$$

We claim that the event $\bar{\Omega}$ holds with high probability for any fixed $\varepsilon > 0$. Indeed, we see that the first event of (4.144) follows from Lemma 4.20 while the second one is obtained from (4.16) and Lemma 4.20. For the rest of this proof, we fix a realization of $X \in \bar{\Omega}$ so that the following arguments are entirely deterministic.

Using (4.144), we infer from (4.16) and Lemma 4.5 that

$$\left|\tilde{\xi}_i - \frac{\sigma_i^2}{\lambda_i}\right| \leqslant n^{-1/2+\varepsilon}\,. \tag{4.145}$$

Moreover, using (4.6), (4.10) and Lemma 4.11, we have $\sigma_i = -1/\tilde{m}(\theta(d_i))$. Then, we deduce from (4.144), (4.16) and Lemma 4.5 that

$$\big|\tilde{m}(\theta(d_i)) - \tilde{m}(\lambda_i)\big| = \left|\int \frac{(\lambda_i - \theta(d_i))\tilde{\varrho}(\mathrm{d}x)}{(\theta(d_i)-x)(\lambda_i - x)}\right| \leqslant n^{-1/2+\varepsilon}\,,$$

so that we conclude with (4.144) and Lemma 4.7 that

$$\left|\sigma_i - \frac{-1}{\tilde{m}(\lambda_i)}\right| \leqslant n^{-1/2+\varepsilon}\,. \tag{4.146}$$

It then suffices to plug (4.146) into (4.145) to get the desired result.

□

We see from Lemma 4.41 that the eigenvalues of the oracle estimator (2.17) can be estimated with high probability by a function that depends only of the sample eigenvalues. However, the asymptotic Stieltjes $\tilde{m}$ is not measurable empirically which prevent this formula to be useful in practice. In order to obtain a fully observable estimator, we define for $i \in [\![r]\!]$, $z_i := \lambda_i + \mathrm{i}\eta$ with $\eta > 0$. We first notice from Lemma 4.20 and (4.16) that $z_i \in \mathbf{D}_o$ with high probability for any $i \in [\![r]\!]$ and small enough $\eta > 0$. Consequently, we invoke Lemma 4.21 to conclude for small enough η that

$$|\tilde{m}(\lambda_i) - \tilde{m}(z_i)| \leqslant |\mathrm{i}\,\eta\,\tilde{m}'(\lambda_i) + O(\eta^2)| \prec \eta. \tag{4.147}$$

We deduce from this last equation and Lemma 4.19 that

$$|\tilde{m}(\lambda_i) - \tilde{s}(z_i)| \prec n^{-1/2} + \eta, \tag{4.148}$$

from which, we also infer

$$|\tilde{s}(z_i)| \asymp 1 \tag{4.149}$$

using Lemma 4.21 for small enough η and $i \in [\![r]\!]$. As a result, we obtain from (4.148), (4.149) and Lemma 4.41 that for small enough $\eta > 0$, the oracle estimator (2.17) satisfies with high probability,

$$\left|\tilde{\xi}_i - \frac{1}{\lambda_i\,|\tilde{s}(z_i)|^2}\right| \prec n^{-1/2} + \eta. \tag{4.150}$$

Finally, the following result allows to replace the fictitious stieltjes transform $\tilde{s}$ by the *observable* one.

Lemma 4.42. *Suppose that the assumptions of Proposition 4.23 hold. Then, for any $\eta > 0$ and $i \in [\![r]\!]$ such that $z_i = \lambda_i + \mathrm{i}\eta \in \mathbf{D}_o$, we have*

$$|s(z_i) - \tilde{s}(z_i)| \prec \frac{1}{n\eta}. \tag{4.151}$$

Proof. This result is obtained from the deterministic eigenvalue interlacing property (see [3, Lemma 4.1]) and then, by following closely the arguments of [17, Lemma 7.1] whose details we omit.

□

In a nutshell, plugging the result of Lemma 4.42 and (4.149) into (4.150) allows us to find,

$$\left|\tilde{\xi}_i - \frac{1}{\lambda_i\,|s(\lambda_i + \mathrm{i}\eta)|^2}\right| \prec n^{-1/2} + \eta + \frac{1}{n\eta}.$$

The only non observable parameter here is η that we choose to minimize the error bound. This yields $\eta = n^{-1/2}$ and we therefore conclude for any $i \in [\![r]\!]$ that

$$\left|\tilde{\xi}_i - \frac{1}{\lambda_i\,|s(\lambda_i + \mathrm{i}/\sqrt{n})|^2}\right| \prec n^{-1/2},$$

as announced in Theorem 2.9.

Appendices

Proof of Section 3.1

Proof of Corrolary 3.1 . The result (3.1) follows from the estimate (4.145). To prove the monotonicity with respect to the sample outliers λ_i, let us define

$$\vartheta(\lambda) \;:=\; \frac{1}{\lambda \tilde{m}^2(\lambda)} \tag{A.1}$$

that is well defined and positive for any $\lambda \geqslant \tilde{\gamma}_+ + c$. By taking the derivative, we obtain

$$\vartheta'(\lambda) \;=\; -\tilde{m}(\lambda)\vartheta^2(\lambda)[\tilde{m}(\lambda) + 2\lambda\tilde{m}'(\lambda)] \tag{A.2}$$

Using the identity

$$\lambda\tilde{m}'(\lambda) = \int \frac{u\tilde{\varrho}(u)\mathrm{d}u}{(u-\lambda)^2} - \tilde{m}(\lambda)\,, \tag{A.3}$$

we obtain

$$\vartheta'(\lambda) \;=\; -\tilde{m}(\lambda)\vartheta^2(\lambda)\Big[2\int \frac{u\tilde{\varrho}(u)\mathrm{d}u}{(u-\lambda)^2} - \tilde{m}(\lambda)\Big]. \tag{A.4}$$

Using that $\operatorname{supp}[\tilde{\varrho}] \subset \mathbb{R}^+$, $\tilde{m}(\lambda)$ is negative and Lemma 4.7, we deduce that $\vartheta'(\lambda) > 0$. The conclusion then follows from the estimate of Lemma 4.41.

□

Proofs of Section 3.2

Proof of Proposition 3.2 . Let us consider the case $\lambda \in [(1-\sqrt{\phi})^2, (1+\sqrt{\phi})^2]$ as the other one is trivial. Since we consider $n \to \infty$, we can replace the random quantity $s(z)$ by its limiting value $m(z)$ using e.g. [23, Theorem 3.6]. Next, we have for $n \to \infty$:

$$\lim_{\eta\downarrow 0} m(\lambda + \mathrm{i}\eta) = \frac{\phi - 1 - \lambda + i\sqrt{4\lambda\phi - (\lambda+\phi-1)^2}}{2\lambda} \tag{B.1}$$

that comes from Marčenko-Pastur law. Hence, this leads to

$$\lim_{\eta\downarrow 0} |m(\lambda + \mathrm{i}\eta)|^2 = \frac{1}{4\lambda^2}\Big[(\phi-1-\lambda)^2 + \Big(4\lambda\phi - (\lambda+\phi-1)^2\Big)\Big], \tag{B.2}$$

and one can easily find that $\lim_{\eta\downarrow 0}|m(\lambda+\mathrm{i}\eta)|^2 = \lambda^{-1}$ which yields the desired result. □

Proof of Proposition 3.3. The proof relies on a elementary analysis of (3.4) and is very similar to the proof of Proposition 3.2 whose details we omit. □

Additional proofs of Section 4.

C.1. Support of $\tilde{\varrho}$ and proof of (4.7). The spectum of $\tilde{S}$ can be analyzed using (2.11) and (2.12). We define the distinct eigenvalues of Σ by $[s_i]_{i\leqslant q}$ ranked in decreasing order with $r+1 \leqslant q \leqslant p$. By Assumption 2.3(i), we know that the limiting eigenvalues density ϱ has $(r+1)$ components that are represented through the critical values $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_{2(r+1)}$. We define by $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_{2(r+1)}$ critical points associated to the critical values through $a_k = f(x_k)$ for $k = 1, \ldots, 2(r+1)$.

Much of the analysis of the structure of ϱ can be made using the function f, defined in (2.11), that we rewrite as

$$z \;=\; f(m), \qquad \operatorname{Im} m > 0, \tag{C.1}$$

with

$$f(x) \;:=\; -\frac{1}{x} + \phi \sum_{i=1}^{q} \frac{\pi_p(\{s_i\})}{x + s_i^{-1}}, \qquad \pi_p \;:=\; \frac{1}{p}\sum_{i=1}^{p} \delta_{\sigma_i}. \tag{C.2}$$

Indeed, we see that the function f is smooth on $(q+1)$ open intervals

$$I_1 \;:=\; (-s_1^{-1}, 0),\; I_j \;:=\; (-s_j^{-1}; -s_{j-1}^{-1}) \quad (j \in [\![2,q]\!]),\; I_0 \;:=\; \breve{\mathbb{R}} \backslash \bigcup_{j=1}^{q} I_j, \tag{C.3}$$

where $\breve{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$. If we define by $\mathcal{C}$ the multiset of critical points of f, we get from Assumption 2.3(ii) and [23, Appendix A] for $r \geqslant 1$

$$|\mathcal{C} \cap I_0| = |\mathcal{C} \cap I_1| = 1, \qquad |\mathcal{C} \cap I_k| \;=\; \begin{cases} 2 & \text{if } k \in [\![2, r+1]\!], \\ 0 & \text{if } k \in [\![r+2, q]\!]. \end{cases} \tag{C.4}$$

This can be well observed in Figure 2.1. Next, one deduces from (2.11) that

$$f'(x) = \frac{1}{x^2}\left[1 - \frac{1}{n}\sum_{i=1}^{p} \frac{x^2}{(x+\sigma_i^{-1})^2}\right], \qquad f''(x) = \frac{2}{x^3}\left[-1 + \frac{1}{n}\sum_{i=1}^{p} \frac{x^3}{(x+\sigma_i^{-1})^3}\right]. \tag{C.5}$$

The estimate (4.7) follows from (4.6) and the following lemma.

Lemma C.1. *Suppose that Assumption 4.1 holds. Then we have for any $k \in [\![r]\!]$ that $s_k = \sigma_k$ with $\sigma_k \geqslant \sigma_{k+1} + c$ for $c > 0$.*

Proof. Using Assumption 2.3(i) and [23, Lemma A.1], we infer that the top $r+1$ eigenvalues of Σ are distinct so that we have $s_k = \sigma_k$ for any $k \in [\![1, r+1]\!]$. Next, invoking Assumption 2.3-(i), (C.3), (C.4) and [23, Appendix A.1], we deduce that $\{a_{2(k-1)+1}, a_{2(k-1)}\} = f(\partial \mathcal{H}_k)$ with $\mathcal{H}_k := \{x \in I_k : f'(x) > 0\} \neq \emptyset$ for any $k \in [\![2, r+1]\!]$. The different sets $\{\mathcal{H}_k\}_k$ are depicted in red in Figure 2.1. Moreover, we get from (C.5) that

$$0 < f'(x) \leqslant \frac{1}{x^2} \leqslant C \tag{C.6}$$

for any $x \in \cup_{k=2}^{r+1} \mathcal{H}_k$ since $\sigma_{r+1} \geqslant c$ from (2.4). Consequently, since $a_{2k} - a_{2k+1} \geqslant c$ from Assumption 2.3-(ii), we deduce from the definition (C.3) that $\sigma_{k+1}^{-1} - \sigma_k^{-1} \geqslant c'$ for any $k \in [\![r+1]\!]$. We then conclude using (2.4). □

We now study the support of $\tilde{\varrho}$. First, let us define by $[\tilde{s}_i]_{i \in [\![\tilde{q}]\!]}$ the set of distinct eigenvalues of $\tilde{\Sigma}$ with $\tilde{q} = q - r$ since the set $[\tilde{s}_i]_{i \in [\![\tilde{q}]\!]}$ is by definition characterized by the $q - r$ distinct bulk eigenvalues of Σ while replacing all the spikes[1] $[\sigma_i]_{i \in [\![r]\!]}$ by $\sigma_{r+1} = s_{r+1}$ with multiplicity r.

Next, we define $\tilde{f}$ as (C.2) while replacing $[s_k]_{k \in [\![q]\!]}$ by $[\tilde{s}_k]_{k \in [\![\tilde{q}]\!]}$ and π_p by $\tilde{\pi}_p = p^{-1}\sum_{i=1}^{p} \delta_{\tilde{\sigma}_i}$. The function $\tilde{f}$ is smooth on the set $\bigcup_{i=1}^{\tilde{q}} \tilde{I}_i$ defined by

$$\tilde{I}_1 \;:=\; (-\tilde{s}_1^{-1}, 0), \qquad \tilde{I}_i \;:=\; (-\tilde{s}_i^{-1}; -\tilde{s}_{i-1}^{-1}) \quad (i \in [\![2, \tilde{q}]\!]), \qquad \tilde{I}_0 \;:=\; \breve{\mathbb{R}} \backslash \bigcup_{i=1}^{\tilde{q}} \tilde{I}_i. \tag{C.7}$$

Note that the derivatives $\tilde{f}'$ and $\tilde{f}''$ are obtained from (C.5) by replacing $[\sigma_i]_{i \in [\![p]\!]}$ with $[\tilde{\sigma}_i]_{i \in [\![p]\!]}$.

Lemma C.2. *Suppose that Assumption 4.1. Then, the asymptotic spectral density $\tilde{\varrho}$ of $\tilde{S}$ consists of a single bulk component.*

[1] Note that $\alpha = \sigma_k$ for any $k > r+1$ would do equally well.

Proof. Let $\tilde{\mathcal{C}}$ denotes the set of critical points of $\tilde{f}$. We shall show that

$$|\widetilde{\mathcal{C}} \cap \tilde{I}_0| = |\widetilde{\mathcal{C}} \cap \tilde{I}_1| = 1, \qquad |\widetilde{\mathcal{C}} \cap \tilde{I}_k| = 0 \quad k \in [\![2, \tilde{q}\,]\!], \tag{C.8}$$

which would allow us to prove the claim. The first equality of (C.8) can be inferred from [23, Appendix A.1]. The second one is obtained by contraposition. Firstly, we notice that both f and $\tilde{f}$ are smooth on $\bigcup_{k=2}^{\tilde{q}} \tilde{I}_k$. If there was a given $l \in [\![2, r+1]\!]$ such that $|\tilde{\mathcal{C}} \cap \tilde{I}_l| = 2$, then there exist a small $\tau' > 0$ and $u \in \tilde{I}_l$ satisfying $\operatorname{dist}(u, \partial \tilde{I}_l) \geqslant \tau'$ and $|f(u) - \tilde{f}(u)| \geqslant c$ from (C.4). Using (2.11) and (4.5), we have for $u \in \tilde{I}_l$,

$$|f(u) - \tilde{f}(u)| = \left| \frac{1}{n} \sum_{i=1}^{r} \left[\frac{1}{u + \sigma_i^{-1}} - \frac{1}{u + \sigma_{r+1}^{-1}} \right] \right| \tag{C.9}$$

Then, using Lemma C.1, we see for any $u \in \tilde{I}_l$ satisfying $\operatorname{dist}(u, \partial \tilde{I}_l) \geqslant \tau'$ that

$$|u + \sigma_i^{-1}| \geqslant c', \qquad i \leqslant r+1. \tag{C.10}$$

Therefore, we conclude for any $u \in \tilde{I}_l$ satisfying $\operatorname{dist}(u, \partial \tilde{I}_l) \geqslant \tau'$,

$$|f(u) - \tilde{f}(u)| \leqslant C n^{-1}, \tag{C.11}$$

and this yields the contradiction. □

C.2. Proof of Lemma 4.5. The upper bound $\tilde{\gamma}_+ \leqslant C$ follows from [23, Lemma 2.5]. For the lower bound, since $\tilde{m}(\tilde{\gamma}_+) \in \tilde{I}_1$ and $\tilde{f}$ is defined as in (C.2) while replacing $[s_k]_{k \in [\![q]\!]}$ by $[\tilde{s}_k]_{k \in [\![\tilde{q}]\!]}$ and π_p by $\tilde{\pi}_p = p^{-1} \sum_{i=1}^{p} \delta_{\tilde{\sigma}_i}$, one has

$$\tilde{\gamma}_+ = \tilde{f}(\tilde{m}(\tilde{\gamma}_+)) \geqslant -\frac{1}{\tilde{m}(\gamma_+)} \geqslant c \tag{C.12}$$

where we used that $\tilde{m}(\tilde{\gamma}_+) > -\sigma_{r+1}^{-1}$ in the second step and then (2.4) in the last one.

C.3. Proof of Lemma 4.7. Let $\tau > 0$. Since $|x - z| \asymp 1$ for any $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$ and any $x \in \operatorname{supp} \tilde{\varrho}$, we easily get $|\tilde{m}(z)| \asymp 1$. Then, using that

$$\tilde{m}'(z) = \int \frac{\tilde{\varrho}(\mathrm{d}x)}{(x - z)^2}, \tag{C.13}$$

it is not hard to see that $\tilde{m}'(z) \asymp 1$ for any $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$.

For the second estimate, note from Appendix C.1 and (C.7) that for $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$,

$$\min_{k \in [\![p]\!]} [1 + \tilde{m}(z) \tilde{\sigma}_k] = 1 + \tilde{m}(z) \sigma_{r+1}, \tag{C.14}$$

since $\tilde{m}(z)$ is negative and satisfies $\tilde{m}(\tilde{\gamma}_+) > -\sigma_{r+1}^{-1}$. Using the triangle inequality, (2.2) and Lemma 4.7(i), we get the uniform upper bound. The lower bound follows closely the arguments of [23, Appendix A.4] with slight modifications whose details we omit.

C.4. Proof of Lemma 4.8. We rewrite (4.10) as

$$w(z) = \frac{\tilde{m}(z) \sigma_{r+1}}{1 + \tilde{m}(z) \sigma_{r+1}}, \tag{C.15}$$

and it is not hard to see from (2.2), (2.4) and Lemma 4.7 that $|w(z)| \asymp 1$ for any $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$. Next, we have

$$w'(z) = \frac{\tilde{m}'(z) \sigma_{r+1}}{(1 + \tilde{m}(z) \sigma_{r+1})^2}, \tag{C.16}$$

which is positive for any $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$ using Lemma 4.7. The conclusion thus follows from (2.2), (2.4) and Lemma 4.7 for any $z \in [\tilde{\gamma}_+ + \tau, \tau^{-1}]$.

C.5. Proof of Lemma 4.9. Let $J := [\tilde\gamma_+, \infty)$ and we first prove that $\tilde m(z)$ is increasing, bounded and negative on J. We know from [23, Appendix A.1] that if $z \in J$, then $z \notin \operatorname{supp}\tilde\varrho$. Consequently, we have that $\tilde m(z)$ is bounded and negative on J. Moreover, we deduce from (C.13) that $m'(z) > 0$ for any $z \in J$ and this yields the desired result.

Next, we claim that $1 + \tilde m(z)\sigma_{r+1} \in (0,1)$ for any $z \in J$. Since this function is obviously increasing in J from the previous paragraph, let us begin with $z \to \infty$. First, notice that $\sigma_{r+1} > 0$ from (2.4) and $\tilde m(z) < 0$ on J. Then, we have by dominated convergence,

$$\tilde m(z) = \frac{-1 + o(1)}{z}, \tag{C.17}$$

where $o(1)$ is an expression that, for any fixed $\tilde\varrho$, goes to zero as z goes to infinity. This gives the upper bound. For the lower bound, we deduce from the first paragraph and from (C.17) that $\tilde m$ is injective and increasing from J onto $[\tilde m(\tilde\gamma_+), 0)$ with inverse function $\tilde f$ as observed in [23, Appendix A.1]. Thus, $\tilde f$ is increasing on $[\tilde m(\tilde\gamma_+), 0)$ and we get from (C.7) and (C.8) that $\operatorname{dist}(\tilde m(z), \partial \tilde I_1) > 0$ with $\tilde I_1 := (-\sigma_{r+1}^{-1}, 0)$ for any $z \in J$. This implies for any $z \in J$ that

$$\sigma_{r+1}^{-1} + \tilde m(z) > 0 \tag{C.18}$$

and we obtain the lower bound by invoking (2.2).

All in all, we conclude from the second paragraph that $w(z)$ is negative for any $z \in J$ with lower bound given by $w(\tilde\gamma+)$. Moreover, using (2.4), (2.2) and that $\tilde m'(z) > 0$ on J, we infer from (C.16) that $w(z)$ is increasing. Next, we get from (C.17) that $w(z) \to 0$ as $z \to \infty$. We therefore conclude that $w(z)$ is increasing, bounded and negative from $[\tilde\gamma_+, \infty)$ onto $[w(\tilde\gamma_+), 0)$ and this ends the proof.

C.6. Proof of Lemma 4.10. Let $\varepsilon > 0$ and define $\mathcal{J} := \big[-1/w(\tilde\gamma_+) + \varepsilon, \varepsilon^{-1}\big]$. Let us first show the upper bound on $\theta(\zeta)$ for any $\zeta \in \mathcal{J}$. From the definition (4.12), we have

$$\theta(\zeta) \;=\; \sigma_{r+1}(1+\zeta)\left[1 + \frac{1}{n}\sum_{i=1}^{p} \frac{1}{\tilde\sigma_i^{-1}\sigma_{r+1}(1+\zeta) - 1}\right] \tag{C.19}$$

Since $\zeta \geqslant \varepsilon$ for any $\zeta \in \mathcal{J}$ from Lemmas 4.8–4.9 and $\|\tilde\Sigma\| = \sigma_{r+1}$ from (4.5), we deduce that

$$\min_i \big[\tilde\sigma_i^{-1}\sigma_{r+1}(1+\zeta) - 1\big] = \zeta \geqslant \varepsilon\,. \tag{C.20}$$

It is now easy to deduce the upper bound from (C.19) by using (C.20), (2.2) and that $\zeta \leqslant \varepsilon^{-1}$ for any $\zeta \in \mathcal{J}$. The lower bound is also immediate from (C.19) and (C.20) since we have

$$\theta(\zeta) \geqslant \sigma_{r+1}(1+\zeta), \tag{C.21}$$

and the conclusion follows from (2.4) and the fact that $\zeta \geqslant \varepsilon$ for any $\zeta \in \mathcal{J}$.

Next, we compute the derivative from (4.12) to obtain

$$\theta'(\zeta) = \sigma_{r+1}\left(\frac{1}{\sigma_{r+1}(1+\zeta)}\right)^2 \tilde f'\left(\frac{-1}{\sigma_{r+1}(1+\zeta)}\right), \tag{C.22}$$

that we rewrite as

$$\theta'(\zeta) = \sigma_{r+1}\big[1 + \kappa(\zeta)\big], \qquad \kappa(\zeta) \;:=\; -\frac{1}{n}\sum_{i=1}^{p} \frac{1}{\big(\tilde\sigma_i^{-1}\sigma_{r+1}(1+\zeta) - 1\big)^2}. \tag{C.23}$$

We notice that $\kappa(\zeta) \leqslant 0$ for any $\zeta \in \mathcal{J}$ so that we conclude

$$\theta'(\zeta) \leqslant \sigma_{r+1} \leqslant C, \tag{C.24}$$

where we used (2.2) in the last step. It remains to prove the lower bound in order to conclude the proof and to that end, we claim that there exists a constant $c>0$ that depends on ε such that

$$\kappa(\zeta) \geqslant -1+c. \tag{C.25}$$

Before proving (C.25), let us first conclude on the lower bound of $\theta'(\zeta)$. We see from (C.23) that by applying (C.25) and (2.4) yields for any $\zeta \in \mathcal{J}$,

$$\theta'(\zeta) \geqslant \sigma_{r+1} c \geqslant c', \tag{C.26}$$

for some small constant $c'>0$.

The remaining of this section is dedicated to the proof of the estimate (C.25). First, we use that $\tilde{f}'(\tilde{m}(\tilde{\gamma}_+))=0$ which allows to find from (C.22) and (C.23) that

$$\kappa(-1/w(\tilde{\gamma}_+)) = -1. \tag{C.27}$$

Next, we write

$$\kappa(\zeta) - \kappa(-1/w(\tilde{\gamma}_+)) = \int_{-1/w(\tilde{\gamma}_+)}^{\zeta} \kappa'(x)\mathrm{d}x \tag{C.28}$$

where

$$\kappa'(x) = \frac{2}{n}\sum_{i=1}^{p} \frac{\sigma_{r+1}}{\sigma_i\big(\tilde{\sigma}_i^{-1}\tilde{\sigma}_{r+1}(1+x)-1\big)^3}. \tag{C.29}$$

Since $\|\tilde{\Sigma}\| = \sigma_{r+1}$, we deduce that

$$\min_i \big[(\sigma_i^{-1}\sigma_{r+1}(1+x)-1)^3\big] \leqslant C \tag{C.30}$$

for any $x \in [-1/w(\tilde{\gamma}_+), \zeta]$ and $\zeta \in \mathcal{J}$. Therefore by using (C.30) and (2.4), we find for any $x \in [-1/w(\tilde{\gamma}_+), \zeta]$ and $\zeta \in \mathcal{J}$ that

$$\kappa'(x) \geqslant \tau, \tag{C.31}$$

with $\tau > 0$ a small constant that depends on ε. As a result, we obtain from this last equation and (C.28) that there exists a small constant $c>0$ that depends on ε and τ such that for any $\zeta \in \mathcal{J}$,

$$\kappa(\zeta) - \kappa(-1/w(\tilde{\gamma}_+)) \geqslant \omega\big[\zeta + 1/w(\tilde{\gamma}_+)\big] \geqslant c. \tag{C.32}$$

By invoking (C.27), we retrieve (C.25) and this ends the proof.

C.7. Proof of Lemma 4.22. Let $\varepsilon > 0$ and define $\mathcal{S} := \big[-1/w(\tilde{\gamma}_+)+\varepsilon, \varepsilon^{-1}\big] \times \big[-\varepsilon^{-1}, \varepsilon^{-1}\big]$. It is easy to see from Lemma 4.10 that $|\theta(\zeta)| \leqslant C$ for any $\zeta \in \mathcal{S}$. From Lemma 4.10, we may use the mean value theorem which implies that there exist some constants $c, c' > 0$ that depends on ε such that for any $\zeta \in \mathcal{S}$,

$$\operatorname{Re}\theta(\zeta) \geqslant \tilde{\gamma}_+ + c\left[\operatorname{Re}\zeta - \frac{-1}{w(\tilde{\gamma}_+)}\right] \geqslant \tilde{\gamma}_+ + c', \tag{C.33}$$

so that $\theta(\zeta) \in \mathbf{D}_o$ for any $\zeta \in \mathcal{S}$.

As a consequence, we can conclude from Lemma 4.8 and (C.33) that for any $\zeta \in \mathcal{S}$,

$$\big|\theta'(\zeta)\big| = \frac{1}{\big|w'(\theta(\zeta))\big|} \asymp 1. \tag{C.34}$$

Therefore, the function θ extends to a holomorphic function in $\mathcal{S}$. It remains to show the estimate for the second derivative, and one has

$$|\theta''(\zeta)| = |\theta'(\zeta)|^2 \big|\tilde{\sigma}_{r+1}(1+\zeta)\big(\tilde{m}''(\theta(\zeta)\theta'(\zeta)(1+\zeta) + 2\tilde{m}'(z)\big)\big|. \tag{C.35}$$

The conclusion follows by using (C.33), (2.2), Lemma 4.21, (C.34) and $|\zeta| \leqslant \varepsilon^{-1}$ for any $\zeta \in \mathcal{S}$.

C.8. Proof of Lemma 4.11. The second estimate is a trivial consequence of Lemma C.1 and (2.4) as we have

$$|d_i - d_j| = \frac{|\sigma_i - \sigma_j|}{\sigma_{r+1}} \geqslant c, \tag{C.36}$$

for any $i, j \in [\![r]\!]$ with $i \neq j$. For the first estimate, we claim that

$$\sigma_i + \frac{1}{\tilde{m}(\tilde{\gamma}_+)} \geqslant c \tag{C.37}$$

for any $i \in [\![r]\!]$. Before proving (C.37), let us show how it implies the first estimate of Lemma 4.11. From (4.6) and (4.10), we have

$$d_i + \frac{1}{w(\tilde{\gamma}_+)} = \frac{1}{\sigma_{r+1}}\left(\sigma_i + \frac{1}{\tilde{m}(\tilde{\gamma}_+)}\right) \geqslant c, \tag{C.38}$$

where we used (2.2) and (C.37) in the last step.

The end of this section is dedicated to the proof of (C.37). Using Assumption 2.3-(iii), we get from [23, Section 2.2] and references therein that

$$\min_{i \in [\![p]\!]} |\sigma_i^{-1} + x_{2r+1}| \geqslant c. \tag{C.39}$$

Next, we claim that

$$|\tilde{m}(\tilde{\gamma}_+) - x_{2r+1}| \leqslant C n^{-1}, \tag{C.40}$$

and it is easy to find (C.37) from (C.39) and (C.40). It remains to show (C.40), and to that end we first use that $x_{2r+1} \in \tilde{I}_1$, (C.9) and (C.39) to deduce,

$$\left|\tilde{f}'(x_{2r+1})\right| = \left|\tilde{f}'(x_{2r+1}) - f'(x_{2r+1})\right| \leqslant C n^{-1}, \tag{C.41}$$

for C a constant that depends on r and where we used that $f'(x_{2r+1}) = 0$ by definition in the first step. Next we write

$$\tilde{f}'(x_{2r+1}) = \int_{\tilde{m}(\tilde{\gamma}_+)}^{x_{2r+1}} y^{-3}\big(y^3 \tilde{f}''(y)\big)\mathrm{d}y. \tag{C.42}$$

In the following, we assume that $x_{2r+1} \leqslant \tilde{m}(\tilde{\gamma}_+)$ as the other case may be treated similarly by adding a minus in the right-hand side of the latter equation. Using (C.5), we obtain $\big(y^3 \tilde{f}''(y)\big)' > 0$ for any $y \in \tilde{I}_1$ so that we have

$$\tilde{f}'(x_{2r+1}) \geqslant x_{2r+1}^3 \tilde{f}''(x_{2r+1}) \int_{\tilde{m}(\tilde{\gamma}_+)}^{x_{2r+1}} \frac{\mathrm{d}y}{y^3} = \frac{x_{2r+1}\tilde{f}''(x_{2r+1})}{2\tilde{m}^2(\tilde{\gamma}_+)}\big(x_{2r+1} - \tilde{m}(\tilde{\gamma}_+)\big)\big(x_{2r+1} + \tilde{m}(\tilde{\gamma}_+)\big). \tag{C.43}$$

Using Lemma 4.5, we may invoke [23, Lemma 4.10] that yields $|\tilde{m}(\tilde{\gamma}_+)| \asymp 1$. Furthermore, using Assumption 2.3-(iii) and then [23, Lemma A.3], we have $|x_{2r+1}| \asymp 1$ and $|\tilde{f}''(x_{2r+1})| \asymp 1$. Therefore, since x_{2r+1} and $\tilde{m}(\tilde{\gamma}_+)$ have the same sign, we conclude that

$$\left|\tilde{f}'(x_{2r+1})\right| \geqslant c|x_{2r+1} - \tilde{m}(\tilde{\gamma}_+)|, \tag{C.44}$$

and we retrieve (C.40) by plugging (C.41) into this last equation.

C.9. Proof of Lemma 4.20. Let $\tau > 0$ and $\varepsilon > 0$. We will show that the outliers $[\lambda_i]_{i \in [\![r]\!]}$ of S stick with high probability to $[\theta(d_i)]_{i \in [\![r]\!]}$. The first step is to prove that with high probability there are no outliers outside a neighbourhood of $\theta(d_i)$. More presicely, let us define for each $i \in [\![r]\!]$ the interval

$$\mathcal{L}_i(D) \;:=\; \big[\theta(d_i) - n^{-1/2+\varepsilon}, \theta(d_i) + n^{-1/2+\varepsilon}\big], \tag{C.45}$$

and $\mathcal{L}_0 := [0, \theta(-1/w(\tilde{\gamma}_+) + \tau/2)]$. Moreover, we define $\mathcal{L}(D) := \mathcal{L}_0 \cup \bigcup_{i=1}^r \mathcal{L}_i(D)$ and $\mathcal{L}^c(D)$ its complement set. We claim that the set $\mathcal{L}^c(D)$ contains no eigenvalue of S with high probability. To prove our claim, we first use the decomposition (4.8) of Σ and the definition of the sample matrix $\tilde{S}$ to obtain,

$$\det(S - z) = \det(\tilde{S}(I + VDV^*) - z), \tag{C.46}$$

which follows from the identity $\det(I_p + AB) = \det(I_n + BA)$ for any matrices $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{n\times p}$. Using the fact that $D > 0$ that follows from (4.7), we rewrite this last equation as

$$\det(S - z) = \det(\tilde{S} - z)\det(D^{-1} + I + zV^*\tilde{R}(z)V)\det(D). \tag{C.47}$$

Hence, z is an eigenvalue of S but not of $\tilde{S}$ if $D^{-1} + I + zV^*\tilde{R}(z)V$ becomes singular. We define

$$W(z) \;:=\; I + zV^*\tilde{R}(z)V, \tag{C.48}$$

and the event

$$\breve{\Omega} \;\equiv\; \breve{\Omega}_n(\varepsilon) \;:=\; \Big\{\|w(z)I - W(z)\| \leqslant n^{-1/2+\varepsilon} \text{ for all } z \notin \mathcal{L}_0\Big\} \cap \big\{\tilde{\lambda}_1 \leqslant \tilde{\gamma}_+ + \varepsilon\big\}, \tag{C.49}$$

where w is defined in (4.10). We deduce from Theorem 4.19 and Lemma 4.6 that the event $\breve{\Omega}$ holds with high probability, and we henceforth fix a realization $X \in \breve{\Omega}$ for the following. We can conclude from (C.47) and (C.49) that $z \notin \mathcal{L}_0$ is an eigenvalue of S if and only if

$$D^{-1} + W(z) = D^{-1} + w(z)I + O(n^{-1/2+\varepsilon}) \tag{C.50}$$

becomes singular. Using (4.12), it suffices to show that for any $z \in \mathcal{L}^c(D)$,

$$\min_{i\in[\![r]\!]}\big|w(z) - d_i\big| \gg n^{-1/2+\varepsilon}. \tag{C.51}$$

This is obtained using that w is monotone increasing in $\mathcal{L}^c(D)$ from Lemma 4.8 as we have

$$w'(z) \geqslant c, \tag{C.52}$$

for any $z \in \mathcal{L}^c(D)$. We omit further details, which may be found in [22, Section 6] for instance and this yields the estimate (C.51) for $z \in \mathcal{L}^c(D)$. The second and last step of the proof consists in showing that the allowed neighborhoods $\mathcal{L}_i(D)$ for each $i \in [\![r]\!]$ contains exactly one outlier. The counting argument follows closely that of [22, Proposition 6.6] whose details we omit.

C.10. Proof of Lemmas 4.34 and 4.35.

Proof of Lemma 4.34. Fix $\varepsilon > 0$. The proof consists in showing that there exists a large constant $\chi > 0$ that depends on ε such that

$$\frac{\|\tilde{\Sigma}^{1/2}X\|}{\sqrt{\tilde{\gamma}_+ + \chi^{-1}}} \leqslant 1, \qquad \frac{\|X\|}{\chi} \leqslant 1, \tag{C.53}$$

hold with high probability. To prove this, we deduce from the very definition of the Euclidean operator norm and Lemma 4.6 that

$$\|\tilde{\Sigma}^{1/2}XX^*\tilde{\Sigma}^{1/2}\| \leqslant \tilde{\gamma}_+ + \varepsilon \tag{C.54}$$

with high probability. Furthermore, since we have

$$\|A^*A\| = \sup_{\|x\|=1}\langle x\,, A^*Ax\rangle = \sup_{\|x\|=1}\|A\|^2,$$

it allows us to write with high probability

$$\|\tilde{\Sigma}^{1/2}X\| \leqslant \sqrt{\tilde{\gamma}_+ + \varepsilon}. \tag{C.55}$$

Next, we see by taking $\tilde{\Sigma} = I_p$ into (C.55) that we obtain with high probability

$$\|X\| \leqslant C_0, \tag{C.56}$$

where $C_0 \equiv C_0(\varepsilon) > 0$ is a constant independent of X and n. We conclude from Lemma 4.5, (C.55) and (C.56) that there exists a large constant $\chi \equiv \chi(\varepsilon) > 0$ such that (C.53) holds. □

Proof of Lemma 4.35. In order to lighten the notations, let us define

$$\varsigma(X) := \frac{\|\tilde{\Sigma}^{1/2}X\|}{\sqrt{\tilde{\gamma}_+ + \chi^{-1}}}, \qquad \vartheta(X) := \frac{\|X\|}{\chi}, \tag{C.57}$$

with $\chi > 0$ fixed. Then, we have from Lemma 4.34 that

$$\|T(X) - T(X')\| = \left\| \frac{X}{\varsigma(X) \vee \vartheta(X) \vee 1} - \frac{X'}{\varsigma(X') \vee \vartheta(X') \vee 1} \right\|.$$

The trick is to rewrite this last equation as

$$\begin{aligned}\|T(X) - T(X')\| \leqslant\ & \left\| \frac{X}{\varsigma(X) \vee \vartheta(X) \vee 1} - \frac{X'}{\varsigma(X) \vee \vartheta(X) \vee 1} \right\| \\ &+ \left\| \frac{X'}{\varsigma(X) \vee \vartheta(X) \vee 1} - \frac{X'}{\varsigma(X) \vee \vartheta(X') \vee 1} \right\| \\ &+ \left\| \frac{X'}{\varsigma(X) \vee \vartheta(X') \vee 1} - \frac{X'}{\varsigma(X') \vee \vartheta(X') \vee 1} \right\|.\end{aligned} \tag{C.58}$$

By choosing correctly the denominator in each terms of the right-hand side, we obtain with the inverse triangle inequality:

$$\left\| \frac{X}{\varsigma(X) \vee \vartheta(X) \vee 1} - \frac{X'}{\varsigma(X) \vee \vartheta(X) \vee 1} \right\| \leqslant \|X - X'\|,$$

$$\begin{aligned}\left\| \frac{X'}{\varsigma(X) \vee \vartheta(X) \vee 1} - \frac{X'}{\varsigma(X) \vee \vartheta(X') \vee 1} \right\| &\leqslant \left\| \frac{X'}{\vartheta(X')} \left[\vartheta(X') - \vartheta(X)\right] \right\| \\ &= \big| \|X\| - \|X'\| \big| \\ &\leqslant \|X - X'\|,\end{aligned} \tag{C.59}$$

and there exists $C \equiv C(\chi) > 0$,

$$\begin{aligned}\left\| \frac{X'}{\varsigma(X) \vee \vartheta(X') \vee 1} - \frac{X'}{\varsigma(X') \vee \vartheta(X') \vee 1} \right\| &\leqslant \left| \frac{X'\left[\varsigma(X') - \varsigma(X)\right]}{\vartheta(X')} \right| \\ &= \left\| \frac{\chi}{\sqrt{\tilde{\gamma}_+ + \chi^{-1}}} \frac{X' \left[\|\tilde{\Sigma}^{1/2}X'\| - \|\tilde{\Sigma}^{1/2}X\| \right]}{\|X'\|} \right\| \\ &\leqslant \left| \frac{\chi}{\sqrt{\tilde{\gamma}_+ + \chi^{-1}}} \right| \|\tilde{\Sigma}^{1/2}X' - \tilde{\Sigma}^{1/2}X\| \\ &\leqslant C\|X - X'\|,\end{aligned} \tag{C.60}$$

where we used (2.2) and Lemma 4.5 in the last step. Summing up all these estimates, and using the fact that $\|X\| \leqslant \|X\|_2$, we conclude that there exists a constant $C > 0$ independent of n such that

$$\|T(X) - T(X')\| \leqslant C\|X - X\|_2. \tag{C.61}$$

□

References

[1] Zhidong Bai and Jack W Silverstein, *Spectral analysis of large dimensional random matrices*, vol. 20, Springer, 2010.

[2] Alex Bloemendal, László Erdos, Antti Knowles, Horng-Tzer Yau, and Jun Yin, *Isotropic local laws for sample covariance and generalized wigner matrices*, Electron. J. Probab **19** (2014), no. 33, 1–53.

[3] Alex Bloemendal, Antti Knowles, Horng-Tzer Yau, and Jun Yin, *On the principal components of sample covariance matrices*, Probability theory and related fields **164** (2016), no. 1-2, 459–552.

[4] Jean-Philippe Bouchaud and Marc Potters, *Financial applications of random matrix theory: a short review*, arXiv preprint arXiv:0910.1205 (2009).

[5] Joël Bun, *Application of Random Matrix Theory to High Dimensional Statistics*, Theses, Université Paris-Saclay, September 2016.

[6] Joël Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters, *Rotational invariant estimator for general noisy matrices*, IEEE Transactions on Information Theory **62** (2016), no. 12, 7475–7490.

[7] Joël Bun, Jean-Philippe Bouchaud, and Marc Potters, *Cleaning large correlation matrices: tools from random matrix theory*, Physics Reports **666** (2017), 1–109.

[8] Romain Couillet and Merouane Debbah, *Random matrix methods for wireless communications*, Cambridge University Press, 2011.

[9] Dipak K Dey and C Srinivasan, *Estimation of a covariance matrix under stein's loss*, The Annals of Statistics (1985), 1581–1591.

[10] Xiucai Ding, *Asymptotics of empirical eigen-structure for high dimensional sample covariance matrices of general form*, arXiv preprint arXiv:1708.06296 (2017).

[11] Edgar Dobriban and Stefan Wager, *High-dimensional asymptotics of prediction: Ridge regression and classification*, arXiv preprint arXiv:1507.03003 (2015).

[12] David L Donoho, Matan Gavish, and Iain M Johnstone, *Optimal shrinkage of eigenvalues in the spiked covariance model*, arXiv preprint arXiv:1311.0851 (2013).

[13] Bradley Efron and Carl Morris, *Multivariate empirical bayes and estimation of covariance matrices*, The Annals of Statistics (1976), 22–32.

[14] Noureddine El Karoui, *Spectrum estimation for large dimensional covariance matrices using random matrix theory*, The Annals of Statistics (2008), 2757–2790.

[15] Khalil Elkhalil, Abla Kammoun, Romain Couillet, Tareq Y Al-Naffouri, and Mohamed-Slim Alouini, *A large dimensional analysis of regularized discriminant analysis classifiers*, arXiv preprint arXiv:1711.00382 (2017).

[16] László Erdős, Antti Knowles, and Horng-Tzer Yau, *Averaging fluctuations in resolvents of random band matrices*, Annales Henri Poincaré **14** (2013), no. 8, 1837–1926.

[17] László Erdős, Antti Knowles, Horng-Tzer Yau, Jun Yin, et al., *Spectral statistics of erdős–rényi graphs i: local semicircle law*, The Annals of Probability **41** (2013), no. 3B, 2279–2375.

[18] LR Haff, *Empirical bayes estimation of the multivariate normal covariance matrix*, The Annals of Statistics (1980), 586–597.

[19] William James and Charles Stein, *Estimation with quadratic loss*, Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, vol. 1, 1961, pp. 361–379.

[20] Iain M Johnstone, *On the distribution of the largest eigenvalue in principal components analysis*, Annals of statistics (2001), 295–327.

[21] Noureddine El Karoui and Holger Kösters, *Geometric sensitivity of random matrix results: consequences for shrinkage estimators of covariance and related statistical methods*, arXiv preprint arXiv:1105.1404 (2011).

[22] Antti Knowles and Jun Yin, *The isotropic semicircle law and deformation of wigner matrices*, Communications on Pure and Applied Mathematics **66** (2013), no. 11, 1663–1749.

[23] ———, *Anisotropic local laws for random matrices*, Probability Theory and Related Fields **169** (2017), no. 1-2, 257–352.

[24] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters, *Noise dressing of financial correlation matrices*, Physical review letters **83** (1999), no. 7, 1467.

[25] Laurent Laloux, Pierre Cizeau, Marc Potters, and Jean-Philippe Bouchaud, *Random matrix theory and financial correlations*, International Journal of Theoretical and Applied Finance **3** (2000), no. 03, 391–397.

[26] Olivier Ledoit and Sandrine Péché, *Eigenvectors of some large sample covariance matrix ensembles*, Probability Theory and Related Fields **151** (2011), no. 1-2, 233–264.

[27] Olivier Ledoit and Michael Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of multivariate analysis **88** (2004), no. 2, 365–411.

[28] ———, *Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks*, The Review of Financial Studies **30** (2017), no. 12, 4349–4388.

[29] Michel Ledoux, *Concentration of measure and logarithmic sobolev inequalities*, Seminaire de probabilites XXXIII, Springer, 1999, pp. 120–216.

[30] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur, *Distribution of eigenvalues for some sets of random matrices*, Matematicheskii Sbornik **114** (1967), no. 4, 507–536.

[31] Debashis Paul and Alexander Aue, *Random matrix theory in statistics: A review*, Journal of Statistical Planning and Inference **150** (2014), 1–29.

[32] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luis A Nunes Amaral, Thomas Guhr, and H Eugene Stanley, *Random matrix approach to cross correlations in financial data*, Physical Review E **65** (2002), no. 6, 066126.

[33] Jack W Silverstein and ZD Bai, *On the empirical distribution of eigenvalues of a class of large dimensional random matrices*, Journal of Multivariate analysis **54** (1995), no. 2, 175–192.

[34] Charles Stein et al., *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proceedings of the Third Berkeley symposium on mathematical statistics and probability, vol. 1, 1956, pp. 197–206.

[35] Akimichi Takemura, *An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population*, Tech. report, DTIC Document, 1983.

[36] Jianfeng Yao, Shurong Zheng, and ZD Bai, *Sample covariance matrices and high-dimensional data analysis*, Cambridge University Press, 2015.