

Rotational invariant estimator for general noisy matrices

Joël Bun,^{1,2,3} Romain Allez,⁴ Jean-Philippe Bouchaud,¹ and Marc Potters¹

¹*Capital Fund Management, 23 rue de l'Université, 75 007 Paris, France*

²*CNRS, LPTMS, Batiment 100, Université d'Orsay, 91405 Orsay Cedex, France*

³*Leonard de Vinci Pôle Universitaire, Finance Lab, 92916 Paris La Défense, France*

⁴*Weierstrass Institute, Mohrenstr. 39, 10117 Berlin, Germany*

We investigate the problem of estimating a given real symmetric signal matrix \mathbf{C} from a noisy observation matrix \mathbf{M} in the limit of large dimension. We consider the case where the noisy measurement \mathbf{M} comes either from an arbitrary additive or multiplicative rotational invariant perturbation. We establish, using the Replica method, the asymptotic global law estimate for three general classes of noisy matrices, significantly extending previously obtained results. We give exact results concerning the asymptotic deviations (called *overlaps*) of the perturbed eigenvectors away from the true ones, and we explain how to use these overlaps to “clean” the noisy eigenvalues of \mathbf{M} . We provide some numerical checks for the different estimators proposed in this paper and we also make the connection with some well known results of Bayesian statistics.

I. INTRODUCTION

One of the most challenging problem in modern statistical analysis is to extract a true signal from noisy observations in data sets of very large dimensionality. Be it in physics, genomics, engineering or finance, scientists are confronted with datasets where the sample size T and the number of variables N are both very large, but with an observation ratio $q = N/T$ that is not small compared to unity. This setting is known in the literature as the high-dimensional limit and differs from the traditional large T , fixed N situation (i.e. $q \rightarrow 0$), meaning that the classical results of multivariate statistics do not necessarily apply.

However, when one deals with very large random matrices (such as covariance matrices), one expects the spectral measure of the matrix under scrutiny to exhibit some universal properties, which are independent of the specific realization of the matrix itself. This property is at the core of Random Matrix Theory (RMT), which provides a very precise description of the convergence of the spectral measure for a very large class of random matrices. Perhaps the two most influential results are Wigner’s semicircle law [1] and Marčenko and Pastur’s theorem [2]. As far as inference is concerned, the latter result is arguably the cornerstone result of RMT in the sense that it gives theoretical tools to understand why classical estimators are insufficient and is now at the heart of many applications in this field (for reviews, see e.g. [3], [4], [5] or [6] and references therein).

In this paper, we consider the statistical problem of a $N \times N$ matrix \mathbf{C} which stands for the unknown signal that one would like to estimate from the noisy measurement of a $N \times N$ matrix \mathbf{M} in the limit of large dimension $N \rightarrow \infty$. A natural question in statistics is to find an estimator $\widehat{\mathbf{C}}(\mathbf{M})$ of the true signal \mathbf{C} that depends on the dataset \mathbf{M} we have. The true matrix \mathbf{C} is unknown and we do not have any particular insights on its components (the eigenvectors). Therefore we would like our estimator $\widehat{\mathbf{C}}(\mathbf{M})$ to be constructed in a rotationally invariant way from the noisy observation \mathbf{M} that we have. In simple terms, this means that there is no privileged direction in the N -dimensional space that would allow one to bias the eigenvectors of the estimator $\widehat{\mathbf{C}}(\mathbf{M})$ in some special directions. More formally, the estimator construction must obey:

$$\Omega \widehat{\mathbf{C}}(\mathbf{M}) \Omega^\dagger = \widehat{\mathbf{C}}(\Omega \mathbf{M} \Omega^\dagger), \quad (\text{I.1})$$

for any rotation matrix Ω . Any estimator satisfying Eq. (I.1) will be referred to as a Rotational Invariant Estimator (RIE). It is not difficult to see that in this case the eigenvectors of the estimator $\widehat{\mathbf{C}}(\mathbf{M})$ have to be the same as those of the noisy matrix \mathbf{M} . As we will show in Section II, this implies that the best possible estimator $\widehat{\mathbf{C}}(\mathbf{M})$ depends on the *overlaps* (i.e. the squared scalar product) between the

eigenvectors of \mathbf{C} and those of \mathbf{M} . These overlaps turn out to be fully computable in the large N limit, using tools from Random Matrix Theory.

The study of the eigenvectors for statistical purposes is in fact quite a recent topic in random matrices. For sample covariance matrices, such considerations have been studied in [7] and [8]. In the latter paper, the notion of overlap and optimal (oracle) estimator are treated in great details. As far as we know, this is the only paper in the literature where the oracle estimator is related to random matrices. Besides the sample covariance matrix, the problem of the overlap for a Gaussian matrix with an external source (also named as deformed Wigner ensemble) has been treated first in [9] and then reconsidered in a more general setting in [10] using Dyson Brownian motions. However, no mention on how to clean the 'noisy' matrix has been given in [9, 10] and this is the gap we hope to fill for a broader class of random perturbations in the present paper.

The outline of this paper is organized as follows. We introduce in Section II A some notations and show that the optimal (oracle) RI estimator involves the overlaps between the eigenvectors of the signal matrix \mathbf{C} and its noisy estimate \mathbf{M} . In section II B, we observe that a convergence result on the resolvent of \mathbf{M} not only gives us all the information about the eigenvalues, but also the eigenvectors. After motivating the study of the resolvent of the measurement matrix \mathbf{M} , we provide in section III explicit expressions for three different perturbation processes. The first one is the case where we add a noisy matrix that is free with respect to the signal \mathbf{C} . The second model concerns multiplicative perturbations and includes the sample covariance matrix of (elliptically distributed) random variables. We also reconsider the case of the so-called 'Information-Plus-Noise' matrix that deals with sample covariance matrices constructed from rectangular Gaussian matrices with an external source. The evaluation of the resolvent for each model is based on the powerful but non-rigorous replica method (which has been extremely successful in various contexts, including RMT or disordered systems— see [11], or [12] for a more recent review). We will see that the derivation of our results using replicas can be done without too much effort and one can certainly imagine that our results can be proven rigorously, as was done in [13] for the resolvent or [8–10] for the overlaps of covariance matrices and Gaussian matrices with external sources. We relegate all these technicalities in various appendices and only give our final results and their numerical verifications in section III. Note in passing that we obtain using replicas the multiplication law of the S-transforms for product of free matrices (see Appendix B 3), a derivation that we have not seen in the literature before. In section IV, we come back to the problem of statistical inference and apply the results of section III to derive the optimal RIE for each considered model. In the multiplicative case, we recover and generalize the estimator recently derived by Ledoit and P ech e for covariance matrices [8]. Each estimator is illustrated by numerical simulations, and we also provide some analytical formulas that can be of particular interest for real life problems. We then conclude this work with some open problems and possible applications of our results.

II. ROTATIONALLY INVARIANT ESTIMATORS, EIGENVECTOR OVERLAPS AND THE RESOLVENT

A. The oracle estimator and the overlaps

Throughout this work, we will consider the signal matrix \mathbf{C} to be a symmetric matrix of dimension N with N that goes to infinity. We denote by $c_1 \geq c_2 \geq \dots \geq c_N$ its eigenvalues and by $|V_1\rangle, |V_2\rangle, \dots, |V_N\rangle$ their corresponding eigenvectors. The perturbed matrix \mathbf{M} will be assumed to be symmetric with eigenvalues denoted by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ associated to the eigenvectors $|U_1\rangle, |U_2\rangle, \dots, |U_N\rangle$. In the limit of large dimension, it is often more convenient to index the eigenvectors of both matrices by their corresponding eigenvalues, i.e. $|U_i\rangle \rightarrow |U_{\lambda_i}\rangle$ and $|V_i\rangle \rightarrow |V_{c_i}\rangle$ for any integer $1 \leq i \leq N$, and this is the convention that we adopt henceforth.

We now attempt to construct an estimator $\widehat{\mathbf{C}}(\mathbf{M})$ of the true signal \mathbf{C} that relies on the given dataset \mathbf{M} at our disposal. It is well known that an estimator is optimal with respect to a specific loss function

(e.g. the distance) and a standard metric is to consider the (squared) Euclidean (or Frobenius) norm

$$\text{Tr} \left[(\mathbf{C} - \widehat{\mathbf{C}}(\mathbf{M}))^2 \right].$$

The best estimator with respect to this loss function is the solution of the following minimization problem

$$\widehat{\mathbf{C}}(\mathbf{M}) = \underset{\text{RI } \mathbf{C}(\mathbf{M})}{\text{argmin}} \sum_{i,j} [\mathbf{C}_{i,j} - \mathbf{C}(\mathbf{M})_{i,j}]^2, \quad (\text{II.1})$$

considered over the set of all possible RI estimators $\mathbf{C}(\mathbf{M})$. We have seen that the RI estimators $\mathbf{C}(\mathbf{M})$ constructed from a given noisy observation matrix \mathbf{M} restrict to the symmetric matrices that have the same eigenvectors as \mathbf{M} . Therefore, the only free variables in the constrained optimization problem (II.1) are the eigenvalues of $\mathbf{C}(\mathbf{M})$ and we can rewrite

$$\widehat{\mathbf{C}}(\mathbf{M}) = U \widehat{\Lambda} U^\dagger, \quad (\text{II.2})$$

where the eigenvalues $\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_N$ are associated with the corresponding perturbed eigenvectors $|U_{\lambda_1}\rangle, |U_{\lambda_2}\rangle, \dots, |U_{\lambda_N}\rangle$. We seek for the $\widehat{\lambda}_i$ that solve the above optimization program:

$$\widehat{\Lambda} = \underset{\{\widehat{\lambda}_k\}_k \in \mathbb{R}^N}{\text{argmin}} \sum_{i,j=1}^N \left(\mathbf{C}_{i,j} - \sum_{k=1}^N U_{i,k} \widehat{\lambda}_k U_{j,k} \right)^2, \quad (\text{II.3})$$

where the $U_{i,k}$ denote the entries of the eigenvectors of \mathbf{M} . A simple computation leads to the following formulas for the $\widehat{\lambda}_i$ in terms of the overlaps between the perturbed $|U_i\rangle$, the non-perturbed eigenvectors $|V_j\rangle$ and the eigenvalues of the true matrix \mathbf{C} :

$$\widehat{\lambda}_i = \text{Tr} \left[|U_i\rangle \langle U_i | \mathbf{C} \right] \equiv \sum_{j=1}^N \langle U_i | V_j \rangle^2 c_j. \quad (\text{II.4})$$

A few comments on this estimator are in order. First, the estimator $\widehat{\lambda}_i$ is designed to construct the best RI estimator $\widehat{\mathbf{C}}(\mathbf{M})$ given in (II.2). The consequence is that if we restrict our estimator to have the eigenvectors of the noisy matrix \mathbf{M} , then the naive approach that consists in *substituting*¹ the eigenvalues $\{\widehat{\lambda}_i\}_{i=1}^N$ with the true ones $\{c_i\}_{i=1}^N$ yields to a spectrum that is too wide. Indeed, it is not hard to see from (II.4) that the top eigenvalues are shrunk downward while the bottom ones are shrunk upward. In other words, the empirical spectral density (ESD) of the $\widehat{\lambda}_i$ is narrower than the true one which shows that the RI estimator cannot be attained by the ‘‘eigenvalues substitution’’ procedure independently proposed in [5, 14], aside from the trivial case $\mathbf{C} = \mathbb{I}_N$.

We will also see that the estimator $\widehat{\lambda}_i$ is self-averaging in the large N -limit (in the sense that it converges almost surely, see section IIB) and can thus be approximated with its expected value

$$\widehat{\lambda}_i \approx \sum_{j=1}^N [\langle U_i | V_j \rangle^2] c_j.$$

where $[\cdot]$ denotes the expected value with respect to the random eigenvectors $(|U_i\rangle)_i$ of the matrix \mathbf{M} . We will sometimes use the following notation for the (rescaled) mean square overlaps

$$O(\lambda_i, c_j) := N [\langle U_i | V_j \rangle^2]. \quad (\text{II.5})$$

Eqs. (II.4) and (II.5) are the quantities of interest in this paper. In Statistics, the optimal RI estimator (II.4) is sometimes called the *oracle* estimator because it depends explicitly on the knowledge of the true signal \mathbf{C} . The ‘‘miracle’’ is that in the large N limit, and for a large class of problems, one can actually express the oracle estimator in terms of the (observable) limiting spectral density (LSD) of \mathbf{M} only.

¹ Remember that we have ranked the eigenvalues.

B. Relation between the resolvent and the overlaps

A convenient way to work out the overlap (II.5) is to study the resolvent of \mathbf{M} , defined as $\mathcal{G}_{\mathbf{M}}(z) := (z\mathbb{I}_N - \mathbf{M})^{-1}$. The claim is that for z not too close to the real axis, the matrix $\mathcal{G}_{\mathbf{M}}(z)$ is *self-averaging* in the large N limit so that its value is independent of the specific realization of \mathbf{M} . More precisely, this means that $\mathcal{G}_{\mathbf{M}}(z)$ converges to a deterministic matrix for any fixed value (i.e. independent of N) of $z \in \mathbb{C} \setminus \mathbb{R}$ when $N \rightarrow \infty$. We will refer to this deterministic limit as the global law of $\mathcal{G}_{\mathbf{M}}(z)$ in the following.

The relation between the resolvent and the overlaps $O(\lambda_j, c_i)$ is relatively straightforward. For $z = \lambda - i\eta$ with $\lambda \in \mathbb{R}$ and $\eta \gg N^{-1}$, we have

$$\mathcal{G}_{\mathbf{M}}(\lambda - i\eta) = \sum_{k=1}^N \left[\frac{\lambda}{(\lambda - \lambda_k)^2 + \eta^2} + i \frac{\eta}{(\lambda - \lambda_k)^2 + \eta^2} \right] |U_k\rangle \langle U_k|.$$

If we take the trace of the above quantity, and take the limit $\eta \rightarrow 0$ (after $N \rightarrow \infty$), it is well known that one obtains the “density of states” (i.e. the LSD) $\rho_{\mathbf{M}}$:

$$\text{Im } G_{\mathbf{M}}(\lambda - i\eta) \equiv \text{Im } \frac{1}{N} \text{Tr} \mathcal{G}_{\mathbf{M}}(\lambda - i\eta) = \pi \rho_{\mathbf{M}}(\lambda), \quad (\text{II.6})$$

(see Appendix A). Similarly, the elements of $\text{Im } \mathcal{G}_{\mathbf{M}}(\lambda - i\eta)$ can be written for $\eta > 0$ as

$$\langle V_i | \text{Im } \mathcal{G}_{\mathbf{M}}(\lambda - i\eta) | V_i \rangle = \sum_{k=1}^N \frac{\eta}{(\lambda - \lambda_k)^2 + \eta^2} \langle V_i | U_k \rangle^2. \quad (\text{II.7})$$

This latter quantity is also self-averaging in the large N limit (the overlaps $\langle V_i | U_k \rangle^2$, $k = 1, \dots, N$ with i fixed display asymptotic independence when $N \rightarrow \infty$ so that the law of large number applies here) and we have

$$\langle V_i | \text{Im } \mathcal{G}_{\mathbf{M}}(\lambda - i\eta) | V_i \rangle \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}} \frac{\eta}{(\lambda - \mu)^2 + \eta^2} O(\mu, c_i) \rho_{\mathbf{M}}(\mu) d\mu.$$

where the overlap function $O(\mu, c_i)$ is extended (continuously) to arbitrary values of μ inside the support of $\rho_{\mathbf{M}}$ in the large N limit. Sending $\eta \rightarrow 0$ in this latter equation, we finally obtain the following formula valid in the large N limit

$$\langle V_i | \text{Im } \mathcal{G}_{\mathbf{M}}(\lambda - i\eta) | V_i \rangle \approx \pi \rho_{\mathbf{M}}(\lambda) O(\lambda, c_i). \quad (\text{II.8})$$

Eq. (II.8) will thus enable us to investigate the overlaps $O(\lambda, c_i)$ in great details through the calculation of the elements of the resolvent $\mathcal{G}_{\mathbf{M}}(z)$. This is what we aim for in the next section. We emphasize that the different equations of the mean square overlaps $O(\lambda, c_i)$ below will be expressed in the basis where \mathbf{C} is diagonal without loss of generality (see Appendix B for more details).

III. OVERLAPS: SOME EXACT RESULTS

A. Free additive noise

The first model of noisy measurement that we consider is the case where the true signal \mathbf{C} is corrupted by a free additive noise, that is to say

$$\mathbf{M} = \mathbf{C} + \mathbf{O} \mathbf{B} \mathbf{O}^\dagger, \quad (\text{III.1})$$

where \mathbf{B} is a fixed matrix with eigenvalues $b_1 > b_2 > \dots > b_N$ with limiting spectral density $\rho_{\mathbf{B}}$ and \mathbf{O} is a random matrix chosen uniformly in the Orthogonal group $O(N)$ (i.e. according to the Haar measure).

This family of models has found several applications in statistical physics of disordered systems subject to an external perturbation where the matrix \mathbf{M} is interpreted as the Hamiltonian of the system, given by the sum of a deterministic term and a random term [15]. A simple example is when the noisy matrix \mathbf{OBO}^\dagger is a symmetric Gaussian random matrix with independent and identically distributed (i.i.d.) entries, corresponding to the so-called GOE (Gaussian Orthogonal Ensemble). By construction, the eigenvectors of a GOE matrix are invariant under rotation.

It is now well known that the spectral density of \mathbf{M} can be obtained from that of \mathbf{C} and \mathbf{B} using free addition, see [16] and, in the language of statistical physics, [17]. The statistics of the eigenvalues of \mathbf{M} has therefore been investigated in great details, see [18] and [19] for instance. However, the question of the eigenvectors has been much less studied, except recently in [9, 10] in the special case where \mathbf{OBO}^\dagger belongs to the GOE (see below).

For a general free additive noise, we show in Appendix B-2 that the global law estimate for the resolvent reads in the large N limit:

$$\mathcal{G}_{\mathbf{M}}(z) = \mathcal{G}_{\mathbf{C}}(Z(z)) \quad (\text{III.2})$$

where the function $Z(z)$ is given by

$$Z(z) = z - R_{\mathbf{B}}(G_{\mathbf{M}}(z)), \quad (\text{III.3})$$

and $R_{\mathbf{B}}$ is the so-called R-transform of \mathbf{B} (see Appendix A for a reminder of the definition of the different useful spectral transforms).

Note that Eq. (III.2) is a matrix relation, that simplifies when written in the basis where \mathbf{C} is diagonal, since in this case $\mathcal{G}_{\mathbf{C}}(Z)$ is also diagonal. Therefore, the evaluation of the overlap $O(\lambda, c)$ is straightforward using Eq. (II.8). Let us define the Hilbert transform $H_{\mathbf{M}}(\lambda)$ which is simply the real part of the Stieltjes transform $G_{\mathbf{M}}(\lambda - i\eta)$ in the limit $\eta \rightarrow 0$. Then the overlap for the free additive noise is given by:

$$O(\lambda, c) = \frac{\beta_1(\lambda)}{(\lambda - c - \alpha_1(\lambda))^2 + \pi^2 \beta_1(\lambda)^2 \rho_{\mathbf{M}}(\lambda)^2}, \quad (\text{III.4})$$

where c is the corresponding eigenvalue of the unperturbed matrix \mathbf{C} , and where we have defined:

$$\begin{cases} \alpha_1(\lambda) := \text{Re} [R_{\mathbf{B}}(H_{\mathbf{M}}(\lambda) + i\pi\rho_{\mathbf{M}}(\lambda))], \\ \beta_1(\lambda) := \frac{\text{Im} [R_{\mathbf{B}}(H_{\mathbf{M}}(\lambda) + i\pi\rho_{\mathbf{M}}(\lambda))]}{\pi\rho_{\mathbf{M}}(\lambda)}. \end{cases} \quad (\text{III.5})$$

As a first check of these results, let us consider the normalized trace of Eq. (III.2) and then set $u = G_{\mathbf{M}}(z) = G_{\mathbf{C}}(Z(z))$. One can find by using the Blue transform that we indeed retrieve the free addition formula $R_{\mathbf{M}}(u) = R_{\mathbf{C}}(u) + R_{\mathbf{B}}(u)$ when $N \rightarrow \infty$, as it should be.

Deformed GOE

As a second verification, we specialize our result to the case where \mathbf{OBO}^\dagger is a GOE matrix such that the entries have a variance equal to σ^2/N . It is then well known that in this case $R_{\mathbf{B}}(z) = \sigma^2 z$, meaning that Eq. (III.3) simply becomes $Z(z) = z - \sigma^2 G_{\mathbf{M}}(z)$. This allows us to get a simpler expression for the overlap:

$$O(\lambda, c) = \frac{\sigma^2}{(c - \lambda + \sigma^2 H_{\mathbf{M}}(\lambda))^2 + \sigma^4 \pi^2 \rho_{\mathbf{M}}(\lambda)^2}, \quad (\text{III.6})$$

which is exactly the result derived in [9, 10] using other methods. In Fig (1), we illustrate this formula in the case where \mathbf{C} is an isotropic Wishart matrix of parameter q , by taking e.g. $\mathbf{C} = T^{-1} \mathbf{H} \mathbf{H}^\dagger$ where \mathbf{H} is a symmetric matrix of size $N \times T$ filled with i.i.d. standard Gaussian entries and $q = N/T$. We set $N = 500$, $T = 1000$, and take \mathbf{OBO}^\dagger as a GOE matrix with variance $1/N$. For a fixed \mathbf{C} , we generate 1000 samples of \mathbf{M} given by Eq. (III.1) for which we can measure numerically the overlap quantity. We see that the theoretical prediction (III.6) agrees remarkably with the numerical simulations.

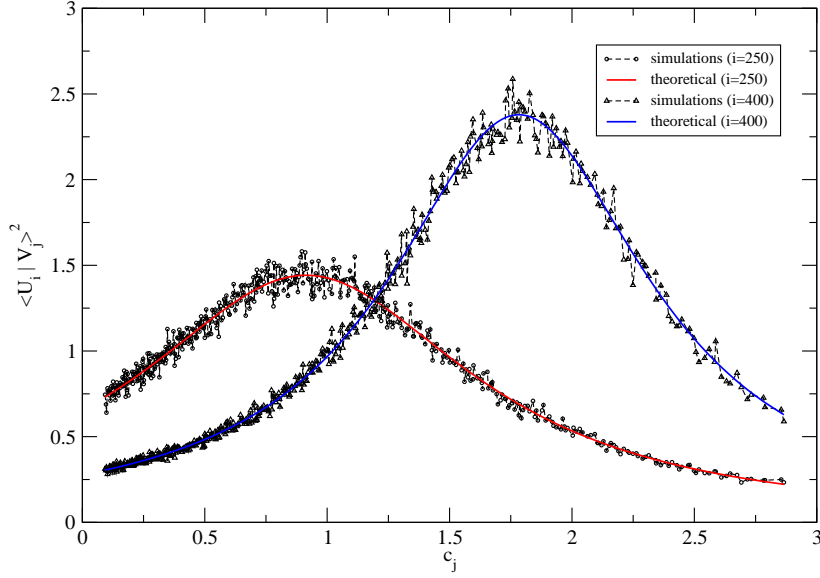


FIG. 1: Computations of the rescaled overlap $O(\lambda, c)$ as a function of c in the free addition perturbation. We chose $i = 250$, \mathbf{C} a Wishart matrix with parameter $q = 0.5$ and \mathbf{B} a Wigner matrix with $\sigma^2 = 1$. The black dotted points are computed using numerical simulations and the plain red curve is the theoretical predictions Eq. (III.4). The agreement is excellent. For $i = 250$, we have $c_i \approx 0.83$ and we see that the peak of the curve is in that region. The same observation holds for $i = 400$ where $c_i \approx 1.66$. The numerical curves display the empirical mean values of the overlaps over 1000 samples of \mathbf{M} given by Eq. (III.1) with \mathbf{C} fixed.

B. Free multiplicative noise and empirical covariance matrices

Our second model deals with *multiplicative* noise in the following sense: we consider that the noisy measurement matrix \mathbf{M} can be written as

$$\mathbf{M} = \sqrt{\mathbf{C}}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger\sqrt{\mathbf{C}}, \quad (\text{III.7})$$

where again \mathbf{C} is the signal, \mathbf{B} is a fixed matrix with eigenvalues $b_1 > b_2 > \dots > b_N$ with limiting density $\rho_{\mathbf{B}}$ and \mathbf{O} is a random matrix chosen in the Orthogonal group $O(N)$ according to the Haar measure. Note that we implicitly requires that \mathbf{C} is positive definite with Eq. (III.7), so that the square root of \mathbf{C} is well defined.

An explicit example of such a problem is provided by sample covariance matrices (namely the Wishart Ensemble [20]), which is of particular interest in multivariate statistical analysis. We shall come back later to this application. The Replica analysis leads to the following systems of equations (see Appendix B 3) for the general problem of a free multiplicative noise above, Eq. (III.7):

$$z\mathcal{G}_{\mathbf{M}}(z) = Z(z)\mathcal{G}_{\mathbf{C}}(Z(z)), \quad (\text{III.8})$$

with:

$$Z(z) = zS_{\mathbf{B}}(zG_{\mathbf{M}}(z) - 1), \quad (\text{III.9})$$

where $S_{\mathbf{B}}$ is the so-called S-transform of \mathbf{B} (see Appendix A) and $G_{\mathbf{M}}$ is the normalized trace of $\mathcal{G}_{\mathbf{M}}(z)$. The latter obeys, from Eq. (III.8), the self-consistent equation:

$$zG_{\mathbf{M}}(z) = Z(z)G_{\mathbf{C}}(Z(z)). \quad (\text{III.10})$$

Again, Eq. (III.8) is a matrix relation, that simplifies when written in the basis where \mathbf{C} is diagonal. Note that Eqs. (III.10) and (III.9) allow us to retrieve the usual free multiplicative convolution, that is to say:

$$S_{\mathbf{M}}(u) = S_{\mathbf{C}}(u)S_{\mathbf{B}}(u). \quad (\text{III.11})$$

This result is thus the analog of our result (III.2) in the multiplicative case. We refer the reader to the appendix B 3 for more details. Note that for technical reasons we restrict \mathbf{B} to have a normalized trace that differs from zero.

With the global law estimate for the resolvent given by Eqs. (III.8) and (III.9) above, we can obtain a general overlap formula for the free multiplicative noise case. Let us define the following functions

$$\begin{cases} \alpha_2(\lambda) := \lim_{z \rightarrow \lambda - i0^+} \operatorname{Re} \left[\frac{1}{S_{\mathbf{B}}(zG_{\mathbf{M}}(z) - 1)} \right] \\ \beta_2(\lambda) := \lim_{z \rightarrow \lambda - i0^+} \operatorname{Im} \left[\frac{1}{S_{\mathbf{B}}(zG_{\mathbf{M}}(z) - 1)} \right] \frac{1}{\pi \rho_{\mathbf{M}}(\lambda)}, \end{cases} \quad (\text{III.12})$$

then the overlap $O(\lambda, c)$ between the eigenvectors of \mathbf{C} and \mathbf{M} are given by:

$$O(\lambda, c) = \frac{c\beta_2(\lambda)}{(\lambda - c\alpha_2(\lambda))^2 + \pi^2 c^2 \beta_2(\lambda)^2 \rho_{\mathbf{M}}(\lambda)^2}. \quad (\text{III.13})$$

In order to give more insights on our results, we will now specify these results to some well-known applications of multiplicative models in RMT.

Empirical covariance matrix

As mentioned previously, the most famous application of a model of the form (III.7) is given by the sample covariance estimator that we recall briefly. Let us define the $N \times T$ observation matrix \mathbf{R} that comes from T independent and identically distributed samples $\mathbf{R}^t \equiv (\mathbf{R}_1^t, \dots, \mathbf{R}_N^t)$ with $t \in [1, T]$ and we assume that each sample have zero mean. The N elements of \mathbf{R}^t generally display some degree of interdependence, that is often represented by the *true* (or also *population*) covariance matrix \mathbf{C} , defined as $\langle \mathbf{R}_i^t \mathbf{R}_j^{t'} \rangle = \mathbf{C}_{i,j} \delta_{t,t'}$, where $\delta_{t,t'}$ is the Kronecker symbol. As the signal \mathbf{C} is unknown, the classical estimator for the covariance matrix is to compute the empirical (or sample) covariance matrix thanks to the Pearson estimator

$$\mathbf{M} = \frac{1}{T} \mathbf{R} \mathbf{R}^\dagger = \sqrt{\mathbf{C}} \frac{1}{T} \mathbf{X} \mathbf{X}^\dagger \sqrt{\mathbf{C}},$$

where \mathbf{X} is a $N \times T$ matrix where all elements are i.i.d. random variables (i.e. their true covariance matrix is the identity matrix). So this model is a particular case of the model (III.7) with $\mathbf{B} := T^{-1} \mathbf{X} \mathbf{X}^\dagger$. The S transform of $\mathbf{B} = T^{-1} \mathbf{X} \mathbf{X}^\dagger$ has an explicit form

$$S_{\mathbf{B}}(x) = \frac{1}{1 + qx}, \quad q = \frac{N}{T}. \quad (\text{III.14})$$

Using our general results Eqs. (III.8) and (III.9), we obtain, in the basis where \mathbf{C} is diagonal:

$$z\mathcal{G}_{\mathbf{M}}(z) = Z(z)\mathcal{G}_{\mathbf{C}}(Z(z)), \quad \text{with} \quad Z(z) = \frac{z}{1 - q + qzG_{\mathbf{M}}(z)}. \quad (\text{III.15})$$

which is exactly the result found in [21] and also in [13] at leading order. We can therefore recover the well-known Marčenko-Pastur equation [2] which gives a fixed point equation satisfied by the resolvent of \mathbf{M} in term of the resolvent of the true matrix \mathbf{C}

$$zG_{\mathbf{M}}(z) = Z(z)G_{\mathbf{C}}(Z(z)), \quad \text{with} \quad Z(z) = \frac{z}{1 - q + qzG_{\mathbf{M}}(z)}. \quad (\text{III.16})$$

The expression of the limiting overlaps can be further simplified in this particular case

$$O(\lambda, c) = \frac{qc\lambda}{(c(1-q) - \lambda + qc\lambda H_{\mathbf{M}}(\lambda))^2 + q^2\lambda^2 c^2 \pi^2 \rho_{\mathbf{M}}(\lambda)^2}, \quad (\text{III.17})$$

and we recover, as expected, the Ledoit & P ech e result established in [8]. As a conclusion, our result generalizes the standard Mar cenko & Pastur formalism to an arbitrary multiplicative noise term $O\mathbf{B}O^\dagger$.

Elliptical ensemble

A slightly more general application of the model (III.7) is when we assume that the entries of \mathbf{R}_i^t can be written as the product of two independent sources $\mathbf{R}_i^t = \sigma_t \xi_{i,t}$. The $\{\xi_{i,t}\}$ are characterized by the true signal $\mathbf{C}_{i,j} = \langle \xi_{i,t} \xi_{j,t'} \rangle \delta_{t,t'}$ and are generated independently from the same distribution at time t that will be assumed to be Gaussian in our case. The $\{\sigma_t\}$ are such that $\langle \sigma^2 \rangle = 1$ and allows to add a time-dependent volatility with a factor σ_t that is common to all variables at time t . This defines the class of elliptical distributions and the most famous application is when the $\{\sigma_t\}$ are drawn from a inverse-gamma distribution which leads to the multivariate Student distribution [22] (see Sec. (IV B) below). The corresponding empirical correlation matrix can be written as

$$\mathbf{M} = \sqrt{\mathbf{C}} \frac{1}{T} \mathbf{X} \Sigma \mathbf{X}^\dagger \sqrt{\mathbf{C}}, \quad (\text{III.18})$$

where $\Sigma := \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_T^2)$. This model has been subject to several studies in RMT, see e.g. [23] [24], [25] or [26]. In all these works, the expression of the limiting Stieltjes transform of the spectral density is quite complex, except for the case where \mathbf{C} is the identity matrix. We find here that we can in fact obtain a self-consistent expression for the global law estimate of the corresponding resolvent by introducing the appropriate transforms. Our result not only generalizes the time-independent result of [21] or [13], but it also provides a tractable equation for the limiting eigenvalues density.

Before stating the result for the elliptical model (III.18), one has to be careful with the S-transform of \mathbf{B} . Indeed, it is in fact more convenient to work with the ‘‘dual’’ matrix $\mathbf{B}_* := T^{-1} \sqrt{\Sigma} \mathbf{X}^\dagger \mathbf{X} \sqrt{\Sigma}$ in order to use the free multiplication formula. We then obtain the S-transform of \mathbf{B}_* to finally express the Stieltjes transform of $G_{\mathbf{B}_*}$ as a function $G_{\mathbf{B}}$, simply by noticing that \mathbf{B}_* has the same eigenvalues as \mathbf{B} and the additional zero eigenvalue with multiplicity $T - N$. The final result reads, after elementary manipulations of the T-transform,

$$S_{\mathbf{B}_*}(x) = \frac{x+1}{x+q} S_{\mathbf{B}}\left(\frac{x}{q}\right). \quad (\text{III.19})$$

In a nutshell, applying the result Eq. (III.9) to the elliptical case leads to the result

$$Z(z) = \frac{z}{1-q+qzG_{\mathbf{M}}(z)} S_{\Sigma}(q(zG_{\mathbf{M}}(z)-1)) \quad (\text{III.20})$$

with Eq. (III.8) and (III.10) unchanged. With Eq. (III.20), our general result (III.10) extends the Mar cenko-Pastur to a time-dependent² framework. We also notice that we have a self-consistent equation in $G_{\mathbf{M}}(z)$ in Eq. (III.10) contrary to previous studies [24–26]. One can easily specialize the result of the overlaps $O(\lambda, c)$ to any λ and c as a function of the spectral measure of Σ . However, we do not find an expression as tractable as Eq. (III.17).

Finally, let us now show that Eq. (III.10) can be useful for practical purposes in order to construct non-trivial models. Suppose that \mathbf{C} is an inverse-Wishart matrix (see section IV.B. for the definition

² in the sense that the volatility depends on the observation time t .

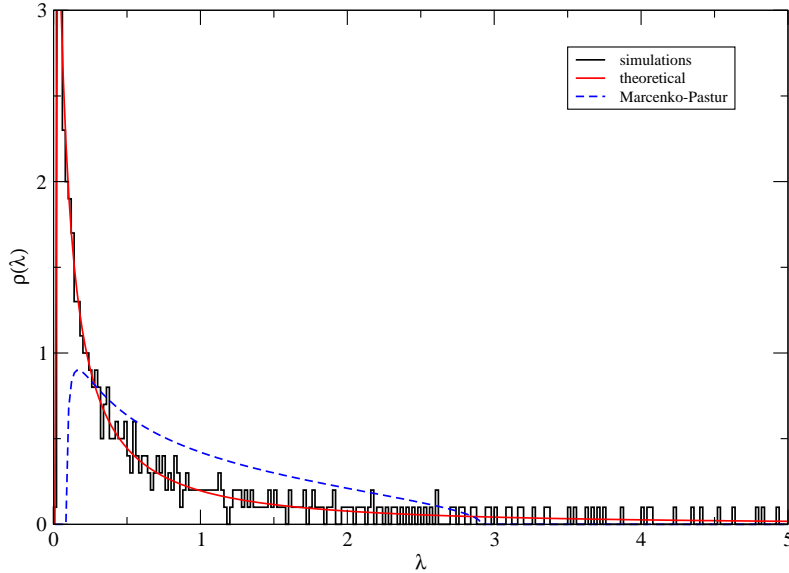


FIG. 2: Theoretical predictions of the density of states from Eq. (III.8) (red line) compared to simulated data when \mathbf{C} is a 500×500 inverse-Wishart matrix (parameter $\kappa = 0.2$) and Σ is a white Wishart with $q_0 = 0.6$. The agreement of our theoretical estimate is excellent and differs strongly from the classical Marčenko-Pastur density (blue dotted curve)

of this law) with parameter $\kappa = 0.2$ and define Σ to be a Wishart matrix of size $T \times T$ and parameter $q_0 = 0.6$. We follow the same numerical procedure as in the free additive noise case. We compare in Fig. 2 our theoretical result Eq. (III.10) with empirical simulations and the agreement is remarkable. The same conclusion holds for the overlap (see Fig. 3).

Note that the results obtained in this section could be extended to the case where the diagonal matrix Σ is not positive definite, for applications in regression analysis (see for example [27]).

C. Information-Plus-Noise matrix

The last model we will treat here is the so-called ‘Information-Plus-Noise’ family of matrix. In this model, we suppose that at each time t , we observe a N -dimensional vector $\mathbf{R}^t = A_t + \sigma \mathbf{X}_t$ where the signal is contained in the vector $A_t \in \mathbb{R}^N$ which is perturbed by an additive noise $\sigma \mathbf{X}_t \in \mathbb{R}^N$. We will assume that the entries of \mathbf{X}_t are i.i.d. Gaussian random variables with zero mean and unit variance. In the case where the number of samples $T \gg N$, the empirical covariance matrix defined as

$$\mathbf{M} = \frac{1}{T} \mathbf{R} \mathbf{R}^\dagger = \frac{1}{T} (\mathbf{A} + \sigma \mathbf{X})(\mathbf{A} + \sigma \mathbf{X})^\dagger \quad (\text{III.21})$$

is a good estimator of $\frac{1}{T} \mathbf{A} \mathbf{A}^\dagger + \sigma^2 \mathbb{I}_N$. This model is of particular interest in signal processing, in order to detect the number of sources and their direction of arrival [28]. Another example of application of this model comes from Finance where one may want to estimate the integrated covariance matrix from high-frequency noisy observation $\mathbf{R}^t = \mathbf{A}_t + \mathbf{X}_t$, where \mathbf{X}_t plays the role of the *microstructure* noise [29].

As usual, in the case where $T \sim \mathcal{O}(N)$, the empirical estimator cannot be fully trusted. The main assumption of the model is as usual the convergence of the empirical density of eigenvalues of $\mathbf{C} := T^{-1} \mathbf{A} \mathbf{A}^\dagger$ towards a limiting density $\rho_{\mathbf{C}}$. The anisotropic global law of the Information-Plus-Noise matrix reads, in a matrix sense:

$$\mathcal{G}_{\mathbf{M}}(z) = ((zZ(z) - \sigma^2(1 - q)) - Z(z)^{-1} \mathbf{C})^{-1}, \quad (\text{III.22})$$

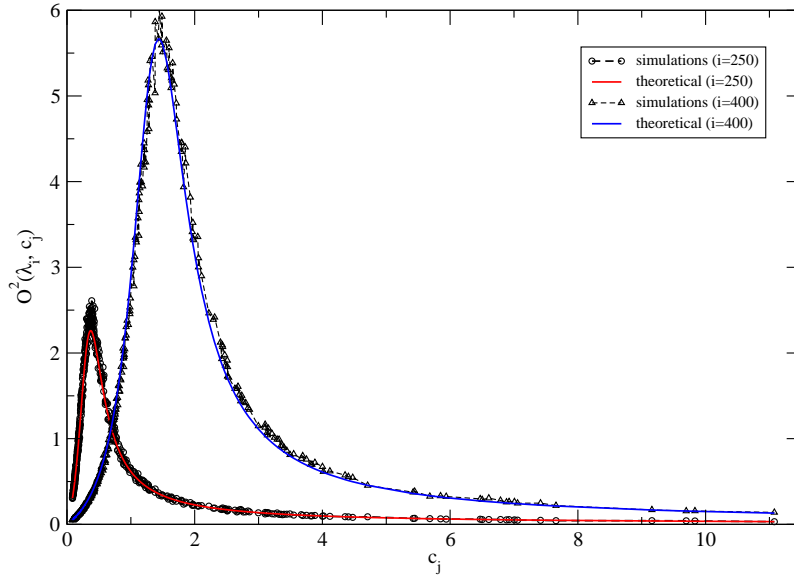


FIG. 3: Rescaled overlap $NO(\lambda, c)$ as a function of c_j in the free addition perturbation with $N = 500$. We chose \mathbf{C} as an inverse-Wishart matrix with parameter $\kappa = 0.2$ and Σ a Wishart matrix with $q_0 = 0.6$. The black dotted points are computed using numerical simulations and the plain curves are the theoretical predictions Eq. (III.13). For $i = 250$ (resp. $i = 400$), we have $c_i \approx 0.37$ (resp. $c_i \approx 1.48$) and we see that the peak of the curve is in that region for both value of i .

where we have defined

$$Z(z) = 1 - q\sigma^2 G_{\mathbf{M}}(z). \quad (\text{III.23})$$

This global law result has already been obtained in a mathematical context by [30] for applications in wireless communications and signal processing. However, it is satisfactory to see that the replica method is able to reproduce this result. If we take the normalized trace of the above equation, we find that the Stieltjes transform reads

$$G_{\mathbf{M}}(z) = \int \frac{dc \rho_{\mathbf{C}}(c)}{z(1 - q\sigma^2 G_{\mathbf{M}}(z)) - \sigma^2(1 - q) - \frac{c}{1 - q\sigma^2 G_{\mathbf{M}}(z)}}, \quad (\text{III.24})$$

which is the result obtained in [31]. As far as we understand, the authors of [30] did not discuss the overlaps in the present context. The final expression for $O(\lambda, c)$ is quite cumbersome, but again completely explicit, and reads:

$$O(\lambda, c) = \frac{q\sigma^2 \alpha_3(\lambda)(\lambda \alpha_3(\lambda) + c)}{[(1 - q\sigma^2 H_{\mathbf{M}}(\lambda))(\lambda \alpha_3(\lambda) - c) - \alpha_3(\lambda)\sigma^2(1 - q)]^2 + (\lambda \alpha_3(\lambda) + c)^2 q^2 \sigma^4 \pi^2 \rho_{\mathbf{M}}^2(\lambda)}, \quad (\text{III.25})$$

where now $\alpha_3(\lambda) := (1 - q\sigma^2 H_{\mathbf{M}}(\lambda))^2 + q^2 \sigma^4 \pi^2 \rho_{\mathbf{M}}^2(\lambda)$.

For the sake of completeness, we provide a numerical example for the overlap (III.25) where \mathbf{A} is a Gaussian matrix of size $N \times T$ with $q = 0.5$ and $N = 500$ with variance 1. The perturbation is a Gaussian noise of same size with $\sigma = 1$. The procedure is the same than in the previous section and we give a numerical example in Fig. 4.

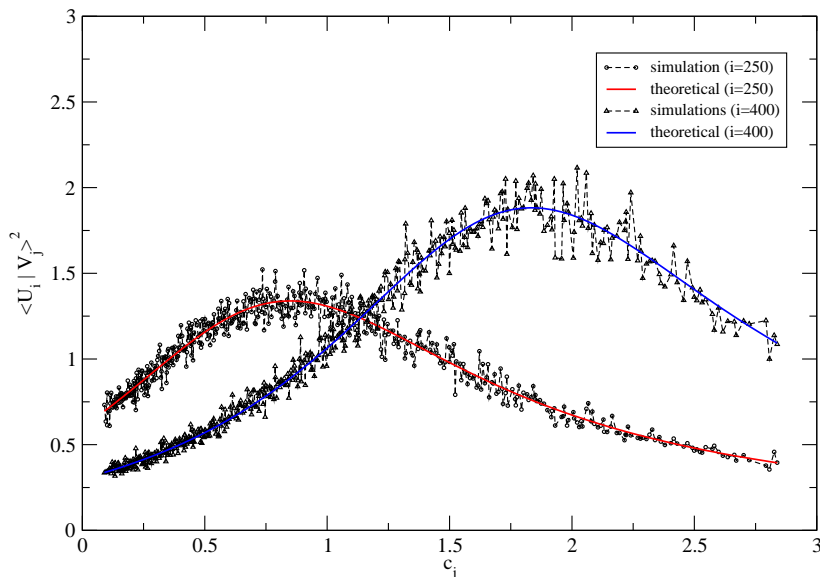


FIG. 4: Rescaled overlap $O(\lambda, c)$ as a function of c_j in the information-plus-noise model with $N = 500$. We chose \mathbf{A} and \mathbf{X} to be a $N \times T$ Gaussian matrix and $T = 2N$. The black dotted points are computed using numerical simulations and the plain curves are the theoretical predictions Eq. (III.25). For $i = 250$ (resp. $i = 400$), we have $c_i \approx 0.83$ (resp. $c_i \approx 1.66$) and we see that the peak of the curve is in that region for both value of i .

IV. OPTIMAL ROTATIONAL INVARIANT ESTIMATOR

The above resolvent and overlap formulas for various models of random matrices are the central results of this study. Equipped with these results, we can now tackle the problem of the optimal RIE of the signal \mathbf{C} . Indeed, the high-dimensional limit $N \rightarrow \infty$, that allows one to reach some degree of universality. First, we rewrite the RIE ((II.4)) as:

$$\widehat{\lambda}_i \stackrel{N \rightarrow \infty}{=} \frac{1}{N} \sum_{j=1}^N O(\lambda_i, c_j) c_j \approx \int dc \rho_{\mathbf{C}}(c) O(\lambda_i, c) c.$$

Quite remarkably, as we show below, the optimal RIE can be expressed, in the three above cases, as a function of the spectral measure of the observable (noisy) \mathbf{M} only.

Let us however stress that the nonlinear “shrinkage” estimators $\widehat{\lambda}_i$ we obtain below are *a priori* valid in the support of \mathbf{M} only. An interesting problem for future research would be to extend the results obtained here for the bulk eigenvalues to the spiked eigenvalues, also called *outliers*. Here, we assume that there are no spikes and we perform the optimal RIE for each models.

A. Free additive noise

We now specialize the RIE and we begin with the free additive noise case for which the noisy measurement is given by

$$\mathbf{M} = \mathbf{C} + \mathbf{O} \mathbf{B} \mathbf{O}^\dagger.$$

It is easy to see from Eqs. (II.8) and (III.2) that:

$$\widehat{\lambda}_i = \frac{1}{\pi \rho_{\mathbf{M}}(\lambda_i)} \lim_{z \rightarrow \lambda_i - i0^+} \text{Im} \int dc \frac{\rho_{\mathbf{C}}(c) c}{Z(z) - c} = \frac{1}{N \pi \rho_{\mathbf{M}}(\lambda_i)} \lim_{z \rightarrow \lambda_i - i0^+} \text{Im} \text{Tr} [\mathcal{G}_{\mathbf{M}}(z) \mathbf{C}], \quad (\text{IV.1})$$

where $Z(z)$ is given by Eq. (III.3). From Eq. (III.2) one also has $\text{Tr}(\mathcal{G}_{\mathbf{M}}(z)\mathbf{C}) = N(Z(z)G_{\mathbf{M}}(z) - 1)$, and using Eqs. (III.3) and (III.5), we end up with:

$$\lim_{z \rightarrow \lambda - i0^+} \text{Im Tr} [\mathcal{G}_{\mathbf{M}}(z)\mathbf{C}] = N\pi\rho_{\mathbf{M}}(\lambda) [\lambda - \alpha(\lambda) - \beta(\lambda)H_{\mathbf{M}}(\lambda)].$$

We therefore find the following optimal RIE nonlinear ‘‘shrinkage’’ function F_1 :

$$\hat{\lambda}_i = F_1(\lambda_i); \quad F_1(\lambda) = \lambda - \alpha_1(\lambda) - \beta_1(\lambda)H_{\mathbf{M}}(\lambda), \quad (\text{IV.2})$$

where α_1, β_1 are defined in Sect. III.A, Eq. (III.5). This result states that if we consider a model where the signal \mathbf{C} is perturbed with an additive noise (that is free with respect to \mathbf{C}), the optimal way to ‘clean’ the eigenvalues of \mathbf{M} in order to get $\hat{\mathbf{C}}(\mathbf{M})$ is to keep the eigenvectors of \mathbf{M} and apply the nonlinear shrinkage formula (IV.2).

Deformed GOE

Let us consider the case where \mathbf{OBO}^\dagger is a GOE matrix. Using the definition of α_1 and β_1 given in Eq. (III.5), the nonlinear shrinkage function is given by

$$F_1(\lambda) = \lambda - 2\sigma^2 H_{\mathbf{M}}(\lambda). \quad (\text{IV.3})$$

Moreover, suppose that \mathbf{C} is also a GOE matrix so that \mathbf{M} is also a GOE matrix with variance $\sigma_{\mathbf{M}}^2 = \sigma_{\mathbf{C}}^2 + \sigma^2$. As a consequence, the Hilbert transform of \mathbf{M} can be computed straightforwardly from the Wigner semicircle law and we find

$$H_{\mathbf{M}}(\lambda) = \frac{\lambda}{2\sigma_{\mathbf{M}}^2}.$$

The optimal cleaning scheme to apply in this case is then given by:

$$F_1(\lambda) = \lambda \left(\frac{\sigma_{\mathbf{C}}^2}{\sigma_{\mathbf{C}}^2 + \sigma^2} \right), \quad (\text{IV.4})$$

where one can see that the optimal cleaning is given by rescaling the empirical eigenvalues by the signal-to-noise ratio. This result is expected in the sense that we perturb a Gaussian signal by adding a Gaussian noise. We know in this case that the optimal estimator of the signal is given, element by element, by the Wiener filter [32], and this is exactly the result that we have obtained with (IV.4). We can also notice that the ESD of the cleaned matrix is narrower than the true one. Indeed, let us define the signal-to-noise ratio $\text{SNR} = \sigma_{\mathbf{C}}^2/\sigma_{\mathbf{M}}^2 \in [0, 1]$, and it is obvious from (IV.4) that $\hat{\mathbf{C}}(\mathbf{M})$ is a Wigner matrix with variance $\sigma_{\mathbf{C}}^2 \times \text{SNR}$ which leads to

$$\sigma_{\mathbf{M}}^2 \geq \sigma_{\mathbf{C}}^2 \geq \sigma_{\mathbf{C}}^2 \times \text{SNR}, \quad (\text{IV.5})$$

as it should be.

As a second example, we now consider a less trivial case and suppose that \mathbf{C} is a white Wishart matrix with parameter q_0 . For any q_0 , it is well known that the Wishart matrix has nonnegative eigenvalues. However, we expect that the noisy effect coming from the GOE matrix pushes some true eigenvalues towards the negative side of the real axis. In Fig 5, we clearly observe this effect and a good cleaning scheme should bring these negative eigenvalues back to positive values. In order to use Eq. (IV.3), we use once again the free addition formula to find the following equation for the Stieltjes transform of \mathbf{M} :

$$-q_0\sigma^2 G_{\mathbf{M}}(z)^3 + (\sigma^2 + q_0 z)G_{\mathbf{M}}(z)^2 + (1 - q - z)G_{\mathbf{M}}(z) + 1 = 0,$$

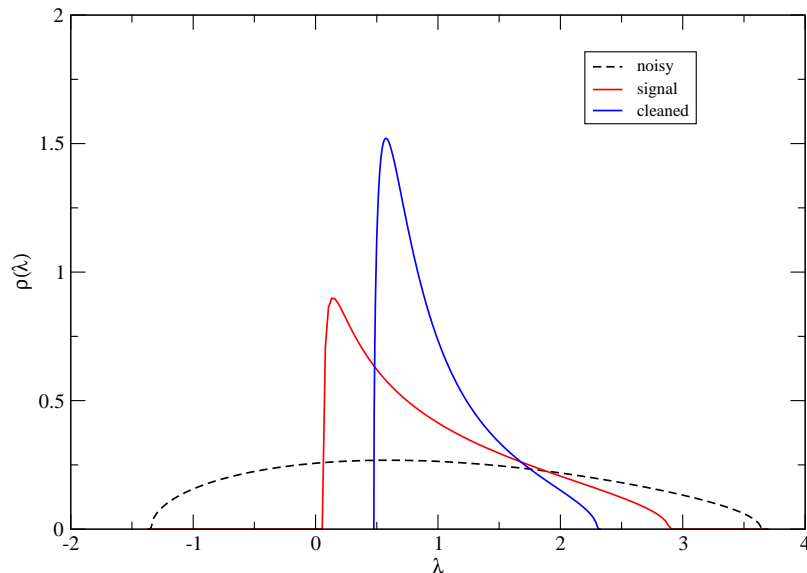


FIG. 5: Eigenvalues of the noisy measurement \mathbf{M} (black dotted line) compared to the true signal \mathbf{C} drawn from a 500×500 Wishart matrix of parameter $q_0 = 0.5$ (red line). We have corrupted the signal by adding a GOE matrix with radius 1. The eigenvalues density of \mathbf{M} allows negative values while the true one has only positive values. The blue line is the LSD of the optimally cleaned matrix. We clearly notice that the cleaned eigenvalues are all positive and its spectrum is narrower than the true one, while preserving the trace.

for any $z = \lambda - i\eta$ with $\eta \rightarrow 0$. It then suffices to take the real part of the Stieltjes transform $G_{\mathbf{M}}(z)$ that solves this equation³ to get the Hilbert transform. In order to check formula Eq. (IV.2) using numerical simulations, we have generated a matrix of \mathbf{M} given by Eq. (III.1) with \mathbf{C} a fixed white Wishart matrix with parameter q_0 and OBO^\dagger a GOE matrix with radius 1. As we know exactly \mathbf{C} , we can compute numerically the oracle estimator as given in (II.4) for each sample. In Fig. (6), we see that our theoretical prediction in the large N limit compares very nicely with the mean values of the empirical oracle estimator computed from the sample. We can also notice in Fig. 5 that the spectrum of the cleaned matrix (represented by the ESD in green) is narrower than the standard Marčenko-Pastur density. This confirms the observation made in Sec. II A.

B. Free multiplicative noise

By proceeding in the same way as in the additive case, we can derive formally a nonlinear shrinkage estimator that depends on the observed eigenvalues λ of \mathbf{M} defined by

$$\mathbf{M} = \sqrt{\mathbf{C}}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger\sqrt{\mathbf{C}}.$$

Following the computations done above, we can find after some manipulations of the global law estimate (III.8):

$$\text{Tr}(\mathcal{G}_{\mathbf{M}}(z)\mathbf{C}) = N(zG_{\mathbf{M}}(z) - 1)S_{\mathbf{B}}(zG_{\mathbf{M}}(z) - 1). \quad (\text{IV.6})$$

³ We take the solution which has a strictly nonnegative imaginary part

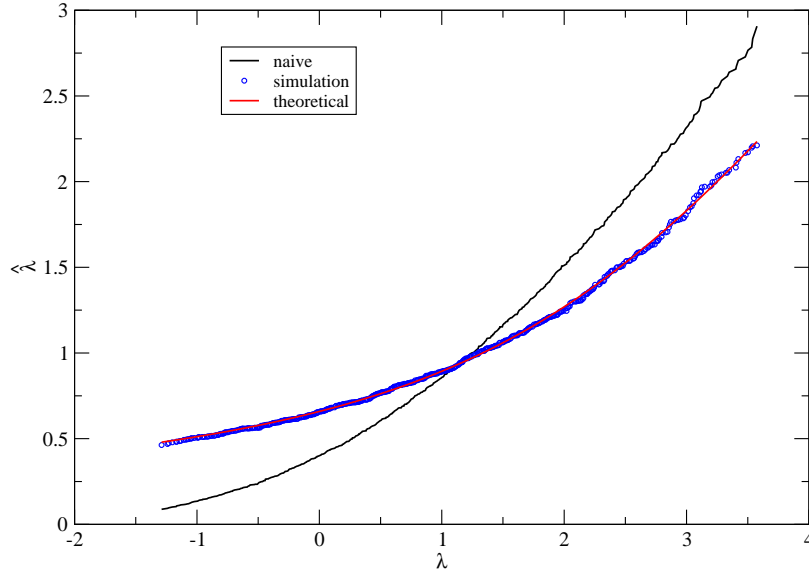


FIG. 6: Eigenvalues according to the optimal cleaning formula (IV.4) (red line) as a function of the observed noisy eigenvalues λ . The parameter are the same as in Fig. 5. We also provide a comparison against the naive eigenvalues substitution method (black line) and we see that the optimal cleaning scheme indeed narrows the spacing between eigenvalues.

Using the analyticity of the S-transform, we define the function $\gamma_{\mathbf{B}}$ and $\omega_{\mathbf{B}}$ such that:

$$\lim_{z \rightarrow \lambda - i0^+} S_{\mathbf{B}}(zG_{\mathbf{M}}(z) - 1) := \gamma_{\mathbf{B}}(\lambda) + i\pi\rho_{\mathbf{M}}(\lambda)\omega_{\mathbf{B}}(\lambda), \quad (\text{IV.7})$$

As a consequence, the optimal RIE (or nonlinear shrinkage formula) for the free multiplicative noise model (III.7) reads:

$$\hat{\lambda}_i = F_2(\lambda_i); \quad F_2(\lambda) = \lambda\gamma_{\mathbf{B}}(\lambda) + (\lambda H_{\mathbf{M}}(\lambda) - 1)\omega_{\mathbf{B}}(\lambda), \quad (\text{IV.8})$$

and this is the analog of the estimator (IV.2) in the multiplicative case.

Empirical covariance matrix

As a first application of the general result Eq. (IV.8), we reconsider the homogeneous Marčenko-Pastur setting where $\mathbf{B} = \frac{1}{T}\mathbf{X}\mathbf{X}^\dagger$. We trivially find from the definition of the S-transform that (IV.7) yields in this case:

$$\gamma_{\mathbf{B}}(\lambda) = \frac{1 - q + q\lambda H_{\mathbf{M}}(\lambda)}{|1 - q + q\lambda \lim_{z \rightarrow \lambda - i0^+} G_{\mathbf{M}}(z)|^2} \quad \text{and} \quad \omega_{\mathbf{B}}(\lambda) = -\frac{q\lambda}{|1 - q + q\lambda \lim_{z \rightarrow \lambda - i0^+} G_{\mathbf{M}}(z)|^2}. \quad (\text{IV.9})$$

The nonlinear shrinkage function F_2 thus becomes:

$$F_2(\lambda) = \frac{\lambda}{(1 - q + q\lambda H_{\mathbf{M}}(\lambda))^2 + q^2\lambda^2\pi^2\rho_{\mathbf{M}}^2(\lambda)}, \quad (\text{IV.10})$$

which is precisely the Ledoit-Péché estimator derived in [8]. Let us insist once again on the fact that this is the oracle estimator, but it can be computed without the knowledge of \mathbf{C} itself, but only with its noisy

version \mathbf{M} . This “miracle” is of course only possible thanks to the $N \rightarrow \infty$ limit that allows the spectral properties of \mathbf{M} and \mathbf{C} to become deterministically related one to the other.

Like in the additive case, we can give a pretty insightful application of the formula (IV.8) based on Bayesian statistics once again. Let us suppose that \mathbf{C} is a white inverse-Wishart matrix (i.e. \mathbf{C}^{-1} is a white Wishart matrix of parameter q). The eigenvalue distribution of \mathbf{C} can then be computed exactly by performing the following change of variable⁴ $c = ((1 - q)c)^{-1}$ in the Marčenko-Pastur density function to get

$$\rho_{\mathbf{C}}(c) = \frac{\kappa}{\pi c^2} \sqrt{(c_+ - c)(c - c_-)}, \quad \text{with } c_{\pm} = \frac{1}{\kappa} [\kappa + 1 \pm \sqrt{2\kappa + 1}] \quad (\text{IV.11})$$

with $q = (2\kappa + 1)^{-1}$ and κ the hyper-parameter which is positive. From there, one can compute the corresponding Stieltjes transform of \mathbf{C}

$$G_{\mathbf{C}}(z) = \frac{(1 + \kappa)z - \kappa \pm \kappa \sqrt{(z - c_+)(z - c_-)}}{z^2}, \quad (\text{IV.12})$$

and we can also compute the Stieltjes transform of the perturbed matrix \mathbf{M} thanks to the Marčenko-Pastur equation:

$$G_{\mathbf{M}}(z) = \frac{z(1 + \kappa) - \kappa(1 - q) \pm \sqrt{(\kappa(1 - q) - z(1 + \kappa))^2 - z(z + 2q\kappa)(2\kappa + 1)}}{z(z + 2q\kappa)}. \quad (\text{IV.13})$$

The reason why we insist on this matrix ensemble is that it plays a special role in multivariate statistics, especially for estimating the covariance matrix because the famous linear shrinkage estimator [33] turns out to be exact in this case, in the sense that it corresponds to the RIE as defined in the introduction. We can recover this result within the present formalism. Indeed, the use of Eq. (IV.13) in the estimator (IV.10) leads, after some computations, to:

$$F_2(\lambda) = \alpha\lambda + (1 - \alpha), \quad \text{with } \alpha = \frac{1}{1 + 2q\kappa}. \quad (\text{IV.14})$$

This is the linear shrinkage estimator that tells us to replace the noisy eigenvalues by a linear combination of the noisy eigenvalues and unity. Equivalently said, the estimator can be formally written in matrix form as $\hat{\Lambda} = \alpha\Lambda + (1 - \alpha)\mathbb{I}_N$.

Elliptical Ensemble

In this subsection, we now consider the elliptical model i.e.

$$\mathbf{B} = \frac{1}{T} \mathbf{X} \Sigma \mathbf{X}^\dagger,$$

for an arbitrary diagonal $T \times T$ matrix Σ . One immediately sees that the optimal shrinkage formula (IV.8) will now depends on $q = N/T$ and on the spectral measure of Σ , which prevents us to get a tractable form as in the homogeneous Marčenko-Pastur case (IV.10). However, we we can definitely expect to find a nonlinear shrinkage formula even when the signal is given by an Inverse-Wishart matrix. The optimal cleaning scheme is therefore given by Eq. (IV.8) where we can explicitly compute the S-transform of \mathbf{B} using the free multiplication and Eq. (III.19) for any Σ . We illustrate this in Fig. (7) where the eigenvalues of Σ are generated following Marčenko-Pastur density and we see that the

⁴ The factor $1 - q$ is such that $\text{Tr} \mathbf{C} = N$, which follows from the Marčenko-Pastur equation.

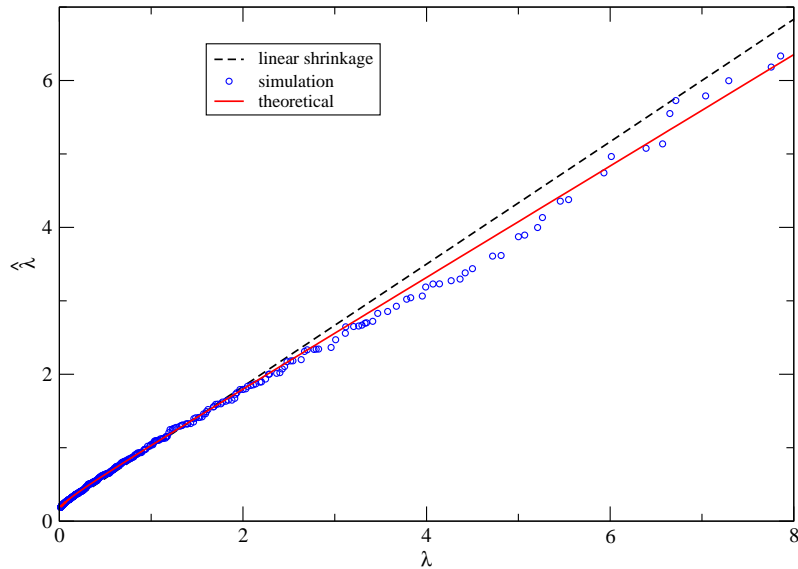


FIG. 7: Eigenvalues according to the optimal cleaning formula (IV.8) (red line) as a function of the observed noisy eigenvalues λ when \mathbf{C} is an inverse-Wishart matrix (with parameter $\kappa = 0.2$ and the $\Sigma = \text{diag}\{\{\sigma_t^2\}_t\}$ is distributed according the Marčenko-Pastur density (with parameter $q_0 = 0.5$). We compare the result against numerical simulations (blue points) and the agreement is excellent. We furthermore compute the optimal cleaning scheme when $\Sigma = \mathbb{I}_T$ (black dotted line) and we see that Σ allows one to go from linear to nonlinear shrinkage.

estimator (IV.8) clearly deviates from the linear shrinkage (IV.14).

As a second example, we consider the *Student* ensemble of correlation matrices [23] which has encountered some success in quantitative finance because it allows one to construct non-Gaussian correlated data with a clear interpretation of the matrix Σ . We impose the $\{\sigma_t^2\}_{t=1}^T$ to be distributed according an inverse-gamma distribution

$$\rho_{\Sigma}(\sigma^2) = \frac{1}{\Gamma\left(\frac{\mu}{2}\right)} \exp\left[-\frac{\sigma_0^2}{\sigma^2}\right] \frac{\sigma_0^{\mu}}{\sigma^{1+\mu}} \quad (\text{IV.15})$$

where we set $\sigma_0^2 := (\mu - 2)/2$ and $\mu > 2$ in order to have $\langle \sigma^2 \rangle = 1$. Within such prescription, the sample data $\mathbf{R}_i^t := \sigma_t \xi_{i,t}$, with $\langle \xi_{i,t} \xi_{j,t'} \rangle \delta_{t,t'} = \mathbf{C}_{i,j}$, is characterized by the multivariate Student distribution of parameters μ and N [22]. From a financial perspective, this parametrization can be useful as a model where all individual stock returns are impacted by the same, time dependent scale factor σ_t that represents the “market volatility” (see [34] for a discussion of this assumption). From empirical studies, one possible choice that matches quite well the data is to choose Eq. (IV.15) with $\mu \approx 3 - 5$. The results above allow us to compute numerically either the LSD or the RIE for an arbitrary “true” signal \mathbf{C} , thus generalizing the work done in [23].

We plot in Fig. 8 the RIE (IV.8) when the eigenvalues of Σ are generated following the inverse-gamma distribution with $\mu = 6$ and \mathbf{C} is still an inverse-Wishart matrix of parameter $\kappa = 0.2$. The numerical procedure is the same as for the previous example. The results we obtain are quite convincing, especially in the bulk. The noisy fluctuations for the largest eigenvalues in Fig. 8 can be explained by the difficulty to solve Eq. (III.10) and (III.20) outside of the bulk, most notably due to the inversion of the T-transform of Σ . However, we see that these large eigenvalues still have the right behaviour in the sense that they are shrunk downward compared to the “naive” substitution procedure.

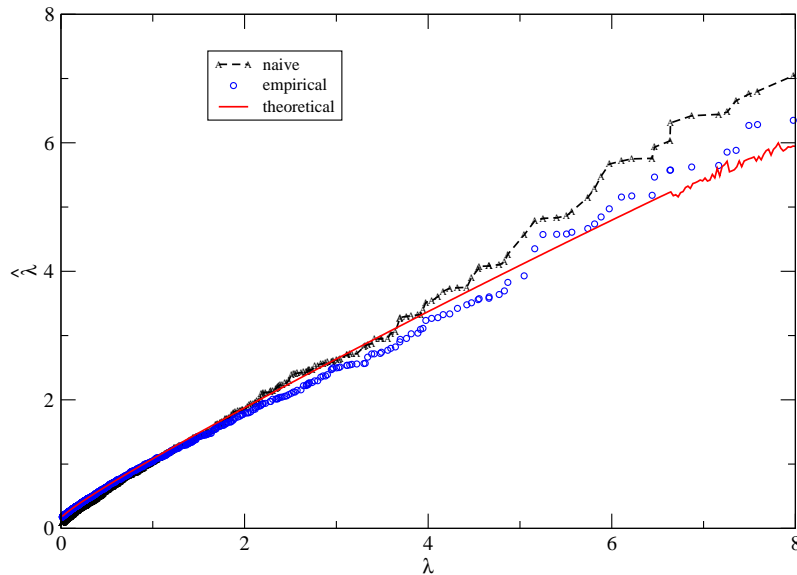


FIG. 8: Eigenvalues according to the optimal cleaning formula (IV.8) as a function of the noisy observed eigenvalues λ when \mathbf{C} is an inverse-Wishart matrix (with parameter $\kappa = 0.2$) and the $\Sigma = \text{diag}(\{\sigma_t^2\}_t)$ is generated according an inverse-gamma distribution (IV.15) (with parameter $\mu = 6$). We compare the RIE (IV.8) (red line) against numerical simulations (blue points) and the agreement is quite convincing, especially in the bulk. We compare it with the substitution procedure (black dotted line) which leads to a wider spectrum.

C. Information-Plus-Noise matrix

The derivation of the asymptotic RI estimator for the Information-Plus-Noise model is a bit more tedious compared to the previous cases but one can follow the same route to find the desired result. We leave the complete derivation for the reader; the final formula for the corresponding shrinkage function F_3 reads:

$$F_3(\lambda) = (1 - q\sigma^2 H_{\mathbf{M}}(\lambda))(\lambda - \sigma^2(1 - q) - 2q\sigma^2 \lambda_i H_{\mathbf{M}}(\lambda)) + q\sigma^2(1 - \gamma_{\mathbf{M}}(\lambda)), \quad (\text{IV.16})$$

where $H_{\mathbf{M}}(\lambda)$ is as before the Hilbert transform of the probability density $\rho_{\mathbf{M}}$, and the function $\gamma_{\mathbf{M}}$ is defined by

$$\gamma_{\mathbf{M}}(\lambda) = H_{\mathbf{M}}(\lambda)(\lambda - \sigma^2(1 - q)) + q\sigma^2 \lambda (\pi^2 \rho_{\mathbf{M}}^2(\lambda) - H_{\mathbf{M}}^2(\lambda)).$$

If we consider the trivial case of zero noise (i.e. $\sigma = 0$), we have by definition that $\mathbf{M} = \mathbf{C}$ and we indeed see this in Eq. (IV.16) where the optimal shrinkage formula becomes $\hat{\lambda}_i = \lambda_i$. The other limit that can be studied without much effort is when the sample size becomes much larger than the number of variable (i.e. $q = 0$). In this case, we know that $\mathbf{M} = \mathbf{C} + \sigma^2 \mathbb{I}_N$ by the law of large number. The optimal shrinkage (IV.16) gives in that case $\hat{\lambda}_i = \lambda_i - \sigma^2$ which was expected because the observation matrix \mathbf{M} is simply a shift of the signal by a factor σ^2 . Let us now reconsider the same numerical example of Sec. (III C) and we apply the same procedure to test the RIE Eq. (IV.16) than the last two sections. We clearly see in Fig. 9 that the agreement is remarkable.

V. CONCLUSION AND OPEN PROBLEMS

As we recalled in the introduction, RMT is already at the heart of many significant contributions when it comes to reconstructing a true signal matrix \mathbf{C} of large dimension from a noisy measurement.

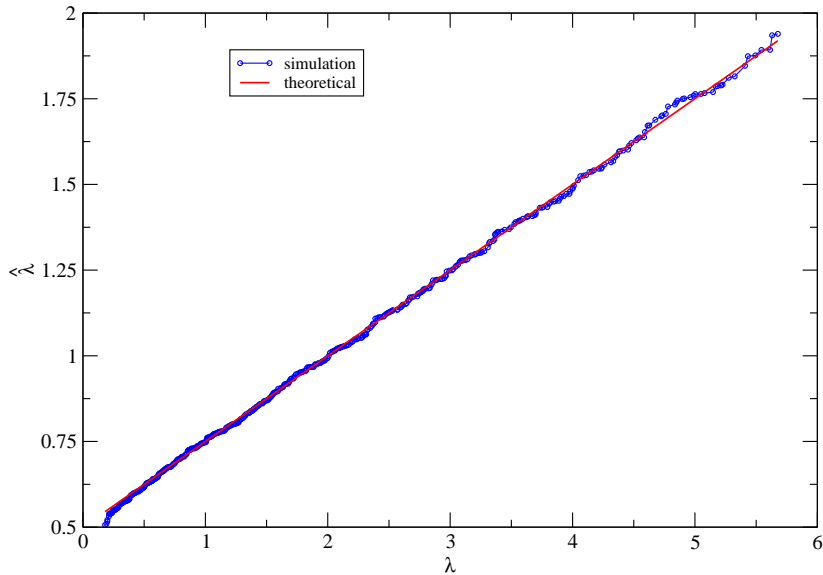


FIG. 9: Eigenvalues according to the optimal cleaning formula (IV.16) as a function of the noisy observed eigenvalues λ with the same setting than in Sec. (III C). We compare the estimator (IV.16) (red line) against numerical simulations coming from a *single* sample with $N = 500$ (blue points), and the agreement is excellent.

In this paper, we have revisited this statistical problem and considered the so-called oracle estimators which is optimal with respect to the Euclidean norm. In particular, we have established the global resolvent law for three distinct ensembles of random matrices that embrace well-known models like the deformed Wigner or the sample covariance matrix. These results on the asymptotic convergence have two important applications: (i) they allow us to find exact results on the overlap between the eigenvectors of the signal matrix with the corrupted ones; (ii) most importantly, they lead to the ‘miracle’ which allows the oracle estimator to be expressed *without any knowledge* of the signal \mathbf{C} in the large N limit. This last observation, that generalizes the work of Ledoit and Péché [8], should be of particular interest in practical cases.

Although our computations are based on the *non-rigorous* Replica method, the comparison between our theoretical formulas and empirical simulations demonstrates the robustness of each proposed estimator. Hence, one can certainly think of possible extensions of this work based on the same method. For instance, a natural extension for our free additive perturbation model would be given by

$$\mathbf{M} = \mathbf{C} + O_q \mathbf{B} O_q^\dagger \quad (\text{V.1})$$

where the law of the matrix $O_q \in O(N)$ interpolates between the Haar measure on the Orthogonal group $O(N)$ when $q = 0$ and a given measure on the permutation group when $q = +\infty$. Differently said, \mathbf{M} interpolates between the free and the classical addition. A natural prescription would be to imagine that O_q is the result of Biane’s Brownian motion [35] on $O(N)$. Another natural possibility is to assume that O_q is distributed according to the probability measure with the Harish-Chandra-Itzykson-Zuber (HCIZ) weight [36], [37]:

$$\mathcal{P}_q(dO) \propto \exp [qN \text{Tr} \mathbf{C} O \mathbf{B} O^\dagger] dO \quad (\text{V.2})$$

which has the right limits when $q \rightarrow 0$ (Haar measure on the orthogonal group) and $q \rightarrow \infty$ (deterministic measure rearranging the spectrum of B in non-increasing order). Hence, considering a replica method for this specific case might give us access to the global law of the resolvent of \mathbf{M} which should enable

us to express the correlation function of angular integrals, leading to an alternative expression of the Morosov-Shatashvili formula ([38], [39]) expressed in terms of the free energy of HCIZ integrals.

We emphasize that the proposed estimators are *optimal* (in the \mathbb{L}^2 norm sense) when the dimension of the problem becomes very large and under no particular prior beliefs on the eigenvector structure of the true matrix \mathbf{C} . However, it happens in practice that one could have a prior structure on the eigenvectors of \mathbf{C} (factor models), and it would be interesting to see how can we rewrite our problem in a non-RI framework. This natural extension is left for future work.

As a challenging open problem, we think it would be very interesting to extend the results obtained here for the bulk eigenvalues to the spiked eigenvalues. To solve this problem, one would need to compute the asymptotic overlaps between the perturbed outlier eigenvectors with the non-perturbed eigenvectors (the corresponding outlier one and the bulk eigenvectors). This question appears to be rather difficult and very few results are available at the moment of writing this paper (to the best of our knowledge). The case of the Gaussian matrix with an external source has been investigated in [10] where one can find the overlaps between the non-perturbed eigenvector associated to the spike and the perturbed eigenvectors (isolated or in the bulk). This solves only partially the problem here as one should compute the “dual” overlaps, i.e. projecting the perturbed spike state onto the non-perturbed states (and not the other way around as in [10]). The main component between the perturbed or non-perturbed spikes is known from [10]. Let us mention that the method used in [10] permits one to handle the information plus noise matrix model as well, using the Bru process introduced in [40]. The case of isotropic covariance matrices has also been considered in [41] where the authors obtained precise delocalization bounds on the overlaps between the non-perturbed spiked eigenvectors and the perturbed bulk eigenvectors in a universal framework but do not compute the limiting explicit value of those mean square overlaps. To the best of our knowledge, the asymptotic of the dual overlaps between the outlier perturbed eigenvectors and the non-perturbed bulk eigenvectors have not yet been obtained in this case.

Finally, we established a connexion between our work with some famous result of Bayesian statistics. For instance, we found out that our results generalize the Wiener filter [32] (additive case) but also the linear shrinkage [33, 42] (multiplicative case), and both have encountered many successes in practical cases. Moreover, the Bayesian theory has found several applications in modern statistical analysis, especially because the large amount of data may allow one to identify a pattern in the data which could be used as a prior. We therefore believe that this work could be the starting point of a *Bayesian random matrix theory* by introducing a notion of prior distribution on the signal \mathbf{C} which is consistently estimated from the data. Indeed, even with this slight change of point of view, the mechanism presented here provides the *optimal* way to clean a large class of noisy measurement such as very large covariance matrices [43].

Acknowledgments

We are indebted to M. Abeille and S. Péché for fruitful discussions. RA received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement nr. 258237 and thanks the Statslab in DPMMS, Cambridge for its hospitality.

-
- [1] E. P. Wigner, in *Mathematical Proceedings of the Cambridge Philosophical Society* (Cambridge Univ Press, 1951), vol. 47, pp. 790–798.
 - [2] V. A. Marchenko and L. A. Pastur, *Matematicheskii Sbornik* **114**, 507 (1967).
 - [3] A. M. Tulino and S. Verdú, *Communications and Information theory* **1**, 1 (2004).
 - [4] I. M. Johnstone and D. M. Titterton, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **367**, 4237 (2009).

- [5] J.-P. Bouchaud and M. Potters, in *The Oxford handbook of Random Matrix Theory* (Oxford University Press, 2011).
- [6] D. Paul and A. Aue, *Journal of Statistical Planning and Inference* **150**, 1 (2014).
- [7] Z. Bai, B. Miao, and G. Pan, *The Annals of Probability* **35**, 1532 (2007).
- [8] O. Ledoit and S. Péché, *Probability Theory and Related Fields* **151**, 233 (2011).
- [9] R. Allez and J.-P. Bouchaud, *Random Matrices: Theory and Applications* **03**, 1450010 (2014).
- [10] R. Allez, J. Bun, and J.-P. Bouchaud, arXiv preprint arXiv:1412.7108 (2014).
- [11] M. Mézard, M. A. Virasoro, and G. Parisi, *Spin glass theory and beyond* (World scientific, 1987).
- [12] F. Morone, F. Caltagirone, E. Harrison, and G. Parisi, arXiv preprint arXiv:1409.2722 (2014).
- [13] A. Knowles and J. Yin, arXiv preprint arXiv:1410.3516 (2014).
- [14] N. El Karoui, *The Annals of Statistics* **36**, 2757 (2008).
- [15] E. Brézin and A. Zee, *Physical Review E* **49**, 2588 (1994).
- [16] D. Voiculescu, K. Dykema, and A. Nica, *Free random variables*, 1 (American Mathematical Soc., 1992).
- [17] A. Zee, *Nuclear Physics B* **474**, 726 (1996).
- [18] E. Brézin and A. Zee, *Nuclear Physics B* **453**, 531 (1995).
- [19] E. Brézin, S. Hikami, and A. Zee, *Physical Review E* **51**, 5442 (1995).
- [20] J. Wishart, *Biometrika* **20A**, pp. 32 (1928).
- [21] Z. Burda, A. Görlich, A. Jarosz, and J. Jurkiewicz, *Physica A: Statistical Mechanics and its Applications* **343**, 295 (2004).
- [22] J.-P. Bouchaud and M. Potters, *Theory of financial risk and derivative pricing: from statistical physics to risk management* (Cambridge university press, 2003).
- [23] G. Biroli, J.-P. Bouchaud, and M. Potters, *Acta Physica Polonica B* **38**, 4009 (2007).
- [24] Z. Burda, J. Jurkiewicz, and B. Waclaw, *Physical Review E* **71**, 026111 (2005).
- [25] L. Zhang, National University of Singapore PHD Thesis (2006).
- [26] N. El Karoui, *The Annals of Applied Probability* **19**, 2362 (2009).
- [27] P.-A. Reigerson, R. Allez, and J.-P. Bouchaud, *Physica A: Statistical Mechanics and its Applications* **390**, 3026 (2011).
- [28] R. O. Schmidt, *Antennas and Propagation*, *IEEE Transactions on* **34**, 276 (1986).
- [29] N. Xia and X. Zheng, arXiv preprint arXiv:1409.2121 (2014).
- [30] W. Hachem, P. Loubaton, J. Najim, and P. Vallet, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **49**, 36 (2013).
- [31] R. B. Dozier and J. W. Silverstein, *Journal of Multivariate Analysis* **98**, 678 (2007).
- [32] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series*, vol. 2 (MIT press Cambridge, MA, 1949).
- [33] L. Haff, *The Annals of Statistics* **8**, 586 (1980).
- [34] R. Chicheportiche and J.-P. Bouchaud, *International Journal of Theoretical and Applied Finance* **15**, 1 (2012).
- [35] P. Biane, in *Free probability theory (Waterloo, ON, 1995)* (Amer. Math. Soc., 1997), vol. 12, pp. 1–19.
- [36] Harish-Chandra, *American Journal of Mathematics* **79**, 87 (1957).
- [37] C. Itzykson and J.-B. Zuber, *Journal of Mathematical Physics* **21**, 411 (1980).
- [38] A. Morozov, *Modern Physics Letters A* **7**, 3503 (1992).
- [39] S. L. Shatashvili, *Communications in mathematical physics* **154**, 421 (1993).
- [40] M.-F. Bru, *Journal of Theoretical Probability* **4**, 725 (1991).
- [41] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin, arXiv preprint arXiv:1404.0788 (2014).
- [42] O. Ledoit and M. Wolf, *Journal of multivariate analysis* **88**, 365 (2004).
- [43] J. Bun, J. P. Bouchaud, and M. Potters, in preparation (2015).
- [44] R. Speicher, in *The Oxford handbook of Random Matrix Theory* (Oxford University Press, 2011).
- [45] Z. Burda, *Journal of Physics: Conference Series* **473**, 012002 (2013).
- [46] G. Parisi, *Journal of Physics A: Mathematical and General* **13**, L115 (1980).
- [47] E. Marinari, G. Parisi, and F. Ritort, *Journal of Physics A: Mathematical and General* **27**, 7647 (1994).
- [48] T. Tanaka, *Journal of Physics: Conference Series* **95**, 012002 (2008).
- [49] A. Guionnet and M. Maïda, *Journal of Functional Analysis* **222**, 435 (2005), ISSN 0022-1236.

Appendix A: Reminder on transforms in RMT

We give in this first appendix a short reminder on the different transforms that are useful in the study of the statistics of eigenvalues in RMT due to their link with free probability theory (see e.g. [44] or [45] for a review). We recall that the resolvent of \mathbf{M} is defined by:

$$\mathcal{G}_{\mathbf{M}}(z) := (z\mathbb{I}_N - \mathbf{M})^{-1}, \quad (\text{A.1})$$

and the *Stieltjes* (or sometimes *Cauchy*) transform is the normalized trace of the Resolvent:

$$\begin{aligned} G_{\mathbf{M}}(z) := \frac{1}{N} \text{Tr} \mathcal{G}_{\mathbf{M}}(z) &= \frac{1}{N} \sum_{k=1}^N \frac{1}{z - \lambda_k}, \\ &\underset{N \rightarrow \infty}{\sim} \int \frac{d\lambda \rho_{\mathbf{M}}(\lambda)}{z - \lambda}. \end{aligned} \quad (\text{A.2})$$

The Stieltjes transform can be interpreted as the average law and is very convenient in order to describe the convergence of the eigenvalues density $\rho_{\mathbf{M}}$. If we set $z = \lambda - i\eta$ and take the limit $\eta \rightarrow 0$, we have in the large N limit

$$G_{\mathbf{M}}(\lambda_i - i\eta) = \text{P.V.} \int \frac{d\lambda' \rho_{\mathbf{M}}(\lambda')}{\lambda - \lambda'} + i\pi \rho_{\mathbf{M}}(\lambda)$$

where the real part is often called the *Hilbert* transform $H_{\mathbf{M}}(\lambda)$ and the imaginary part leads to the eigenvalues density.

When we consider the case of adding two random matrices that are (asymptotically) free with each other, it is suitable to introduce the functional inverse of the Stieltjes transform known as the *Blue* transform

$$B_{\mathbf{M}}(G_{\mathbf{M}}(z)) = z. \quad (\text{A.3})$$

This allows us to define the so-called R-transform

$$R_{\mathbf{M}}(z) := B_{\mathbf{M}}(z) - \frac{1}{z}, \quad (\text{A.4})$$

which can be seen as the analogue in RMT of the logarithm of the Fourier transform for free additive convolution. More precisely, if \mathbf{A} and \mathbf{B} are two $N \times N$ independent invariant symmetric random matrices, then in the large N limit, the spectral measure of $\mathbf{M} = \mathbf{A} + \mathbf{B}$ is given by

$$R_{\mathbf{M}}(z) = R_{\mathbf{A}}(z) + R_{\mathbf{B}}(z), \quad (\text{A.5})$$

known as the free addition formula [16]. In this case, we note by $\rho_{\mathbf{A} \boxplus \mathbf{B}}$ the eigenvalues density of \mathbf{M} .

We can do the same for the free multiplicative convolution. In this case, we rather have to define the so-called *T* (or sometimes η [3]) transform given by

$$T_{\mathbf{M}}(z) = \int \frac{d\lambda \rho_{\mathbf{M}}(\lambda) \lambda}{z - \lambda} \equiv z G_{\mathbf{M}}(z) - 1, \quad (\text{A.6})$$

which can be seen as the moment generating function of \mathbf{M} . The S-transform of \mathbf{M} is then defined as

$$S_{\mathbf{M}}(z) := \frac{z + 1}{z T_{\mathbf{M}}^{-1}(z)} \quad (\text{A.7})$$

where $T_{\mathbf{M}}^{-1}(z)$ is the functional inverse of the *T*-transform. Before showing why the S-transform is important in RMT, one has to be careful about the notion of product of free matrices. Indeed, if we reconsider the two $N \times N$ independent symmetric random matrices \mathbf{A} and \mathbf{B} , the product $\mathbf{A}\mathbf{B}$ is in general

not self-adjoint even if \mathbf{A} and \mathbf{B} are self-adjoint. However, if \mathbf{A} is positive definite, then the product $\sqrt{\mathbf{A}\mathbf{B}}\sqrt{\mathbf{A}}$ makes sense and share the same moments than the product $\mathbf{A}\mathbf{B}$. We can thus study the spectral measure of $\mathbf{M} = \sqrt{\mathbf{A}\mathbf{B}}\sqrt{\mathbf{A}}$ in order to get the distribution of the free multiplicative convolution $\rho_{\mathbf{A}\boxtimes\mathbf{B}}$. The result, first obtained in [16], reads:

$$S_{\mathbf{A}\boxtimes\mathbf{B}}(z) := S_{\mathbf{M}}(z) = S_{\mathbf{A}}(z)S_{\mathbf{B}}(z). \quad (\text{A.8})$$

The S-transform is therefore the analogue of the Fourier transform for free multiplicative convolution.

Appendix B: Derivation of the global law estimate

1. The replica method

The starting point of our approach is to rewrite the entries of the resolvent $\mathcal{G}_{\mathbf{M}}(z)$ by the Gaussian integral representation of an inverse matrix

$$\mathcal{G}_{\mathbf{M}}(z)_{i,j} = \frac{\int \left(\prod_{k=1}^N d\eta_k \right) \eta_i \eta_j \exp \left\{ -\frac{1}{2} \sum_{k,l=1}^N \eta_k (z\delta_{k,l} - \mathbf{M}_{k,l}) \eta_l \right\}}{\int \left(\prod_{k=1}^N d\eta_k \right) \exp \left\{ -\frac{1}{2} \sum_{k,l=1}^N \eta_k (z\delta_{k,l} - \mathbf{M}_{k,l}) \eta_l \right\}}. \quad (\text{B.1})$$

We recall that the claim is that for a complex z not too close to the real axis, we expect the resolvent to be self-averaging in the large N limit, that is to say independent of the specific realization of the matrix itself. Therefore we can study the resolvent $\mathcal{G}_{\mathbf{M}}(z)$ through its ensemble average (denoted by $\langle \cdot \rangle$ in the following) given by:

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \left\langle \frac{1}{\mathcal{Z}} \int \left(\prod_{k=1}^N d\eta_k \right) \eta_i \eta_j \exp \left\{ -\frac{1}{2} \sum_{k,l=1}^N \eta_k (z\delta_{k,l} - \mathbf{M}_{k,l}) \eta_l \right\} \right\rangle, \quad (\text{B.2})$$

where \mathcal{Z} is the partition function, i.e. the denominator in Eq. (B.1). The computation of the average value is highly non trivial in the general case. The replica method tells us that the expectation value can be handled thanks to the following identity

$$\begin{aligned} \langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle &= \lim_{n \rightarrow 0} \left\langle \mathcal{Z}^{n-1} \int \left(\prod_{k=1}^N d\eta_k \right) \eta_i \eta_j \exp \left\{ -\frac{1}{2} \sum_{k,l=1}^N \eta_k (z\delta_{k,l} - \mathbf{M}_{k,l}) \eta_l \right\} \right\rangle \\ &= \lim_{n \rightarrow 0} \int \left(\prod_{k=1}^N \prod_{\alpha=1}^n d\eta_k^\alpha \right) \eta_i \eta_j \left\langle \exp \left\{ -\frac{1}{2} \sum_{\alpha=1}^n \sum_{k,l=1}^N \eta_k^\alpha (z\delta_{k,l} - \mathbf{M}_{k,l}) \eta_l^\alpha \right\} \right\rangle. \end{aligned} \quad (\text{B.3})$$

We have thus transformed our problem to the computation of n replicas of the initial system (B.1). So when we have computed the average value in (B.2), it suffices to perform an analytical continuation of the result to real values of n and finally takes the limit $n \rightarrow 0$. The main concern of this non-rigorous approach is that we assume that the analytical continuation can be done with only n different set of points which could lead to uncontrolled approximation in some cases [46].

2. Free additive noise

We consider a model of the form

$$\mathbf{M} = \mathbf{C} + \mathbf{O}\mathbf{B}\mathbf{O}^\dagger$$

where \mathbf{B} is a fixed matrix with eigenvalues $b_1 > b_2 > \dots > b_N$ with spectral $\rho_{\mathbf{B}}$ and O is a random matrix chosen in the Orthogonal group $O(N)$ according to the Haar measure. Clearly, the noise term is invariant under rotation so that we expect the resolvent of \mathbf{M} to be in the same basis than \mathbf{C} . We therefore set without loss of generality that \mathbf{C} is diagonal. In order to derive the global law estimate for the resolvent of the matrix \mathbf{M} , we have to consider the ensemble average value of the resolvent over the Haar measure for the $O(N)$ group, which can be written as follow

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \int \left(\prod_{\alpha=1}^n \prod_{k=1}^N d\eta_k^\alpha \right) \eta_i \eta_j \prod_{\alpha=1}^n e^{-\frac{1}{2} \sum_{k=1}^N (\eta_k^\alpha)^2 (z - c_k)} \left\langle e^{-\frac{1}{2} \sum_{k,l=1}^N \eta_k^\alpha (O\mathbf{B}O^\dagger)_{k,l} \eta_l^\alpha} \right\rangle_O. \quad (\text{B.4})$$

The evaluation of the later equation can be done straightforwardly if we set the measure dO to be a flat measure constrained to the fact that $OO^\dagger = \mathbb{I}_N$, or equivalently said:

$$\mathcal{D}O \propto \prod_{i,j=1}^N dO_{i,j} \prod_{i,j=1}^N \delta \left(\sum_k O_{i,k} O_{j,k} - \delta_{i,j} \right)$$

where $\delta(\cdot)$ is the Dirac delta function and $\delta_{i,j}$ is Kronecker delta. In the case where n is finite (and independent of N), one can notice that Eq. (B.4) is the Orthogonal low-rank version of the Harish-Chandra-Itzykson-Zuber integrals ([36], [37]). The result is known for all symmetry groups ([47, 48] or [49] for a more rigorous derivation), and this reads for the rank- n case

$$\int \mathcal{D}O \exp \left[\text{Tr} \left(\frac{1}{2} \sum_{\alpha=1}^n \eta^\alpha (\eta^\alpha)^\dagger O\mathbf{B}O^\dagger \right) \right] = \exp \left[\frac{N}{2} \sum_{\alpha=1}^n W_{\mathbf{B}}((\eta^\alpha)^\dagger \eta^\alpha) \right], \quad (\text{B.5})$$

with W_B the primitive of the R-transform of B . The computation of the resolvent (B.4) becomes:

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \int \left(\prod_{k=1}^N d\eta_k \right) \eta_i \eta_j \exp \left\{ \frac{N}{2} \sum_{\alpha=1}^n \left[W_{\mathbf{B}}((\eta^\alpha)^\dagger \eta^\alpha) - \frac{1}{2} \sum_{k=1}^N (\eta_k^\alpha)^2 (z - c_k) \right] \right\},$$

where we have introduced a Lagrange multiplier $p^\alpha = \frac{1}{N} (\eta^\alpha)^\dagger \eta^\alpha$ which gives using Fourier transform (renaming $\zeta^\alpha = 2i\zeta^\alpha/N$)

$$\begin{aligned} \langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle &\propto \int \int \left(\prod_{\alpha=1}^n dp^\alpha d\zeta^\alpha \right) \exp \left\{ \frac{N}{2} \sum_{\alpha=1}^n [W_{\mathbf{B}}(p^\alpha) + p^\alpha \zeta^\alpha] \right\} \\ &\times \int \left(\prod_{\alpha=1}^n \prod_{k=1}^N d\eta_k^\alpha \right) \eta_i \eta_j \exp \left\{ -\frac{1}{2} \sum_{\alpha=1}^n \sum_{k=1}^N (\eta_k^\alpha)^2 (z + \zeta - c_k) \right\}. \end{aligned}$$

This additional constraint allows one to retrieve a Gaussian integral over the $\{\eta_j\}$ which can be computed exactly. Ignoring normalization terms, we obtain

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle \propto \int \int \left(\prod_{\alpha=1}^n dp^\alpha d\zeta^\alpha \right) \frac{\delta_{i,j}}{z + \sigma \zeta^\alpha - c_i} \exp \left\{ -\frac{Nn}{2} F_0(p^\alpha, \zeta^\alpha) \right\}$$

where the ‘free energy’ F_0 is given by

$$F_0(p, \zeta) = \frac{1}{Nn} \sum_{\alpha=1}^n \left[\sum_{k=1}^N \log(z + \zeta^\alpha - c_k) - W_{\mathbf{B}}(p^\alpha) - p^\alpha \zeta^\alpha \right]. \quad (\text{B.6})$$

In the large N limit, the integral can be evaluated by considering the saddle-point of the free energy F_0 as the other term is obviously sub-leading. We now use the *replica symmetric* ansatz that tells us if the

free energy is invariant under the action of the symmetry group $O(N)$, then we expect a saddle-point which is also invariant. This implies that we have at the saddle-point

$$p^\alpha = p \quad \text{and} \quad \zeta^\alpha = \zeta, \quad \forall \alpha \in \{1, \dots, n\}, \quad (\text{B.7})$$

and hence, we have to solve the following set of equations:

$$\begin{cases} \zeta^* = -R_{\mathbf{B}}(p) \\ p^* = G_{\mathbf{C}}(z + \zeta). \end{cases}$$

The trick is to see that we can get rid off one variable by taking the normalized trace of the (average) resolvent which gives the following relation for the Stieltjes transform: $G_{\mathbf{M}}(z) = G_{\mathbf{C}}(z - R_{\mathbf{B}}(p^*)) = p^*$, that is to say

$$G_{\mathbf{M}}(z) = G_{\mathbf{C}}(z - R_{\mathbf{B}}(G_{\mathbf{M}}(z))). \quad (\text{B.8})$$

In conclusion, by taking the limit $n \rightarrow 0$, we have obtained the following global law estimate

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = (Z(z)\mathbb{I}_N - \mathbf{C})_{i,i}^{-1} \delta_{i,j} \quad (\text{B.9})$$

with

$$Z(z) = z - R_{\mathbf{B}}(G_{\mathbf{M}}(z)), \quad (\text{B.10})$$

which are exactly the result stated in Eq. (III.2) and (III.3).

3. Free multiplicative noise

Let us set the measurement matrix \mathbf{M} as:

$$\mathbf{M} = \sqrt{\mathbf{C}}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger\sqrt{\mathbf{C}} \quad (\text{B.11})$$

where \mathbf{O} is still a rotation matrix over the Orthogonal group, \mathbf{C} is a positive definite matrix and \mathbf{B} is such that $\text{Tr}\mathbf{B} \neq 0$. Note that we can assume without loss of generality that \mathbf{C} is diagonal because the argument of the previous subsection still applies. The replica method allows us to write the entries of the resolvent of \mathbf{M} as follow

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \int \left(\prod_{\alpha=1}^n \prod_{k=1}^N d\eta_k^\alpha \right) \eta_i \eta_j e^{-\frac{z}{2} \sum_{\alpha=1}^n \sum_{k=1}^N (\eta_k^\alpha)^2} \left\langle e^{\frac{1}{2} \sum_{k,l=1}^N \sum_{\alpha=1}^n \eta_k^\alpha (\sqrt{\mathbf{C}}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger\sqrt{\mathbf{C}})_{k,l} \eta_l^\alpha} \right\rangle_{\mathbf{O}}. \quad (\text{B.12})$$

We can notice that the matrix $\sum_{\alpha=1}^n (\sqrt{\mathbf{C}}\eta^\alpha) (\sqrt{\mathbf{C}}\eta^\alpha)^\dagger$ is a symmetric rank- n matrix, with n finite and independent of N . Therefore, the expectation over the Haar measure still leads to a rank- n Orthogonal version of HCIZ integral, and the result reads

$$\left\langle \exp \left\{ \frac{1}{2} \sum_{k,l=1}^N \sum_{\alpha=1}^n \eta_k^\alpha (\sqrt{\mathbf{C}}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger\sqrt{\mathbf{C}})_{k,l} \eta_l^\alpha \right\} \right\rangle_{\mathbf{O}} = \exp \left[\frac{N}{2} \sum_{\alpha=1}^n W_{\mathbf{B}} \left(\frac{1}{N} \sum_{i=1}^N (\eta_i^\alpha)^2 c_i \right) \right]. \quad (\text{B.13})$$

As in the free addition case, let us defined the auxiliary variable $p^\alpha = \frac{1}{N} \sum_{i=1}^N (\eta_i^\alpha)^2 c_i$ that we enforce by a Dirac delta function. This allows us to get a Gaussian integral over the $\{\eta_k^\alpha\}$ that yields

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle \propto \int \int \left(\prod_{\alpha=1}^n dp^\alpha d\zeta^\alpha \right) \frac{\delta_{i,j}}{z - \zeta c_i} \exp \left\{ \frac{Nn}{2} F_0(p^\alpha, \zeta^\alpha) \right\} \quad (\text{B.14})$$

where the free energy is given by

$$F_0(p^\alpha, \zeta^\alpha) = \frac{1}{n} \sum_{\alpha=1}^n \left[\frac{1}{N} \sum_{k=1}^N \log(z - \zeta^\alpha c_k) + \zeta^\alpha p^\alpha - W_{\mathbf{B}}(p^\alpha) \right]. \quad (\text{B.15})$$

We now assume that the saddle-point solution can be computed using the replica symmetry ansatz presented in the previous section, so that the free energy becomes

$$F_0(p^\alpha, \zeta^\alpha) \equiv F_0(p, \zeta) = \frac{1}{N} \sum_{k=1}^N \log(z - \zeta c_k) + \zeta p - W_{\mathbf{B}}(p). \quad (\text{B.16})$$

We first consider the derivative w.r.t. p which leads to

$$\zeta^* = R_{\mathbf{B}}(p). \quad (\text{B.17})$$

The other derivative gives

$$p^* = \frac{T_{\mathbf{C}}\left(\frac{z}{R_{\mathbf{B}}(p^*)}\right)}{R_{\mathbf{B}}(p^*)}. \quad (\text{B.18})$$

Hence, we see that the resolvent is given in the large N limit and the limit $n \rightarrow 0$ by

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \frac{\delta_{i,j}}{z - R_{\mathbf{B}}(p^*)c_i}. \quad (\text{B.19})$$

We can find a genuine simplification of the last expression using the connexion with the free multiplication convolution. By taking the normalized trace of $\mathcal{G}_{\mathbf{M}}(z)$, we see that we have

$$zG_{\mathbf{M}}(z) = ZG_{\mathbf{C}}(Z), \quad \text{with} \quad Z = \frac{z}{R_{\mathbf{B}}(p^*)}, \quad (\text{B.20})$$

which can be easily rewrite as

$$T_{\mathbf{M}}(z) = T_{\mathbf{C}}(Z).$$

Let us define $x = T_{\mathbf{M}}(z) = T_{\mathbf{C}}(Z)$ which implies that $p^* = x/R_{\mathbf{B}}(p^*)$. We recall that we aim to prove that we indeed get the free multiplicative convolution in the large N limit. Hence, we rewrite the last equation as

$$zT_{\mathbf{M}}(z) = ZT_{\mathbf{C}}(Z)R_{\mathbf{B}}(p^*).$$

It is trivial to see that putting x in this latter expression yields $xT_{\mathbf{M}}^{-1}(x) = xT_{\mathbf{C}}^{-1}(x)R_{\mathbf{B}}(p^*)$, and by definition of the S-transform, we have

$$S_{\mathbf{M}}(x) = S_{\mathbf{C}}(x) \frac{1}{R_{\mathbf{B}}(p^*)}. \quad (\text{B.21})$$

In order to retrieve the desire result, we use the following relation

$$\frac{1}{R_{\mathbf{B}}(p^*)} = S_{\mathbf{B}}(p^* R_{\mathbf{B}}(p^*)), \quad (\text{B.22})$$

which comes from the very definition of the S-transform of \mathbf{B} . But recalling that $p^* = x/R_{\mathbf{B}}(p^*)$, we conclude that the spectral density of \mathbf{M} is given by the free multiplication

$$S_{\mathbf{M}}(x) = S_{\mathbf{C}}(x)S_{\mathbf{B}}(x), \quad (\text{B.23})$$

as expected and it proves that the replica symmetry ansatz holds in this model. So, going back to the resolvent, we can characterize its large N behaviour by a deterministic quantity which reads:

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \frac{\delta_{i,j}}{z - \frac{c_i}{S_{\mathbf{B}}(T_{\mathbf{M}}(z))}}. \quad (\text{B.24})$$

where we have used the definition of x and the relation (B.22). All in all, the global law of the resolvent of \mathbf{M} in the case of the product of free matrices \mathbf{C} and \mathbf{B} is given by

$$z \langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \delta_{i,j} Z(z) (Z(z) - c_i)^{-1}, \quad \text{with} \quad Z = z S_{\mathbf{B}}(z G_{\mathbf{M}}(z) - 1) \quad (\text{B.25})$$

which is exactly the stated in Eq. (III.8).

4. Information-Plus-Noise matrix

The computation of the global law estimate for this model is pretty similar to the sample covariance [22]. The noisy measurement matrix is given by

$$\mathbf{M} = \frac{1}{T} (\mathbf{A} + \sigma \mathbf{X}) (\mathbf{A} + \sigma \mathbf{X})^\dagger$$

with \mathbf{A} a fixed $N \times T$ matrix such that $T^{-1} \mathbf{A} \mathbf{A}^\dagger = \mathbf{C}$. As we posit that \mathbf{X} is a Gaussian matrix, we can work in the basis where \mathbf{C} is once again diagonal. Moreover, we can in fact show that interchanging the integral and the average leads to the same result. We hence consider directly the annealed average in order to lighten the notations. Let us compute

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \int \left(\prod_{k=1}^N d\eta_k \right) \eta_i \eta_j e^{-\frac{z}{2} \sum_{k=1}^N \eta_k^2} \left\langle e^{\frac{1}{2T} \sum_{k,l=1}^N \sum_{t=1}^T \sigma_t^2 \eta_k c_k^{1/2} Y_{k,t} Y_{l,t} c_l^{1/2} \eta_l} \right\rangle,$$

where we have defined $Y_t := \mathbf{A}_t + \sigma \mathbf{X}_t$ which is still a Gaussian vector. One can readily compute the average value over the measure of Y to find

$$\left\langle \exp \left\{ \frac{1}{2T} \sum_{k,l=1}^N \sum_{t=1}^T \sigma_t^2 \eta_k c_k^{1/2} Y_{k,t} Y_{l,t} c_l^{1/2} \eta_l \right\} \right\rangle \propto \left(1 - \frac{\sigma^2}{T} \eta^\dagger \eta \right)^{-\frac{T}{2}} \exp \left\{ \frac{1}{2} \left(1 - \frac{\sigma^2}{T} \eta^\dagger \eta \right)^{-1} \sum_{k=1}^N c_k \eta_k^2 \right\},$$

where we have omitted all constants terms and used Sherman-Morrison formula in the exponential term. Rewriting $p = \sigma^2 T^{-1} \eta^\dagger \eta$ that we enforced by a Dirac delta function, we can therefore compute the integral over $\{\eta_k\}$ to find

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \int dp \int d\zeta \frac{\delta_{i,j}}{z - q\zeta\sigma^2 - \frac{c_i}{1-p^*}} \exp \left\{ -\frac{N}{2} F_0(p, \zeta) \right\},$$

and we can once again compute the integral in the large N limit by performing the saddle-point of the following free energy

$$F_0(p, \xi) = \frac{1-q}{q} \log(1-p) + p\xi + \frac{1}{N} \sum_{k=1}^N \log [(z - q\sigma^2\xi)(1-p) - c_k]. \quad (\text{B.26})$$

The derivation over ξ gives the following equation $p^* = q\sigma^2(1-p^*)G_{\mathbf{C}}((z - q\zeta\sigma^2)(1-p^*))$, and by taking the normalized trace of the resolvent, we see that we have $G_{\mathbf{M}}(z) = \frac{p^*}{q\sigma^2}$. The other derivative leads to $\xi^* = \frac{1-q}{q} + zG_{\mathbf{M}}(z)$. Therefore, the global law estimate reads:

$$\langle \mathcal{G}_{\mathbf{M}}(z)_{i,j} \rangle = \delta_{i,j} \left((zZ(z) - \sigma^2(1-q)) - Z(z)^{-1} \mathbf{C} \right)_{i,i}^{-1}, \quad (\text{B.27})$$

with

$$Z(z) = 1 - q\sigma^2 G_{\mathbf{M}}(z). \quad (\text{B.28})$$

This is exactly the result announced in Eqs. (III.22, III.23) and in [30], showing that considering directly the annealed average is correct.